

University of Southampton Research Repository ePrints Soton

Copyright © and Moral Rights for this thesis are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given e.g.

AUTHOR (year of submission) "Full thesis title", University of Southampton, name of the University School or Department, PhD Thesis, pagination

UNIVERSITY OF SOUTHAMPTON

Pose Invariant Gait Analysis And Reconstruction

by

Nicholas M. Spencer

A thesis submitted in partial fulfillment for the
degree of Doctor of Philosophy

in the
Faculty of Engineering, Science and Mathematics
School of Electronics and Computer Science

December 2005

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

Faculty of Engineering, Science and Mathematics
School of Electronics and Computer Science

Doctor of Philosophy

by Nicholas M. Spencer

One of the unique advantages of human gait is that it can be perceived from a distance. A varied range of research has been undertaken within the field of gait recognition. However, in almost all circumstances subjects have been constrained to walk fronto-parallel to the camera with a single walking speed. In this thesis we show that gait has sufficient properties that allows us to exploit the structure of articulated leg motion within single view sequences, in order to remove the unknown subject pose and reconstruct the underlying gait signature, with no prior knowledge of the camera calibration.

Articulated leg motion is approximately planar, since almost all of the perceived motion is contained within a single limb swing plane. The variation of motion out of this plane is subtle and negligible in comparison to this major plane of motion. Subsequently, we can model human motion by employing a cardboard person assumption. A subject's body and leg segments may be represented by repeating spatio-temporal motion patterns within a set of bilaterally symmetric limb planes.

The static features of gait are defined as quantities that remain invariant over the full range of walking motions. In total, we have identified nine static features of articulated leg motion, corresponding to the fronto-parallel view of gait, that remain invariant to the differences in the mode of subject motion. These features are hypothetically unique to each individual, thus can be used as suitable parameters for biometric identification.

We develop a stratified approach to linear trajectory gait reconstruction that uses the rigid bone lengths of planar articulated leg motion in order to reconstruct the fronto-parallel view of gait. Furthermore, subject motion commonly occurs within a fixed ground plane and is imaged by a static camera. In general, people tend to walk in straight lines with constant velocity. Imaged gait can then be split piecewise into natural segments of linear motion. If two or more sufficiently different imaged trajectories are available then the calibration of the camera can be determined. Subsequently, the total pattern of gait motion can be globally parameterised for all subjects within an image sequence. We present the details of a sparse method that computes the maximum likelihood estimate of this set of parameters, then conclude with a reconstruction error analysis corresponding to an example image sequence of subject motion.

Contents

Acknowledgements	viii
1 Introduction	1
1.1 Biometrics	2
1.2 Human movement	3
1.3 Prior work	4
1.4 Research outline	8
1.5 Thesis structure	10
2 Stratified Recovery of Planar Gait Motion	14
2.1 Introduction	14
2.2 Representation of 2D primitives	16
2.2.1 Points	16
2.2.2 Lines	17
2.2.3 Homogeneous scale of points and lines	20
2.2.4 Ideal points and the line at infinity	20
2.2.5 Conics	21
2.2.6 Polarity and conjugacy of conics	23
2.3 Camera model	24
2.3.1 Linear projection	24
2.3.2 Radial distortion	25
2.3.3 Camera extrinsic parameters	27
2.4 Planar transformations	27
2.4.1 Planar homography	28
2.4.2 Image mosaicing	29
2.4.3 Transformation of points, lines and conics	29
2.4.4 Vanishing points and lines	30
2.4.5 Repeating planar patterns	32
2.4.6 Direct linear transformation	34
2.5 Planar geometry	35
2.5.1 Absolute conic	35
2.5.2 Image of the absolute conic	36
2.5.3 Camera calibration	37
2.5.3.1 Constraints on the intrinsic parameters	37
2.5.3.2 Recovering the extrinsic parameters	39
2.5.3.3 Maximum likelihood estimation	40
2.5.3.4 Stereo calibration	40

2.5.4	Stratified rectification of images of planes	42
2.5.4.1	Recovery of affine properties	43
2.5.4.2	Recovery of metric properties	43
2.5.4.3	Known ratios of lengths	45
2.6	Epipolar geometry	46
2.6.1	Point correspondences between views.	46
2.6.2	Pure translation	48
2.6.3	Stereopsis transformation	49
2.7	Multiple view geometry	50
2.8	Geometry of gait	52
2.8.1	Identification of the vanishing line	52
2.8.2	Recovering metric structure	58
2.9	Conclusions	61
3	Static Features of Human Gait	64
3.1	Introduction	64
3.2	Gait Cycle	67
3.2.1	Stance and Swing	67
3.2.2	Phases of Gait	68
3.2.2.1	Weight acceptance	68
3.2.2.2	Single limb support	69
3.2.2.3	Limb advancement	70
3.2.3	Energy conservation	70
3.3	Motion marker systems	72
3.4	Experimental set-up	73
3.4.1	Markers	74
3.4.2	Cameras	75
3.5	Planar limb swing assumption	78
3.5.1	Overground walking	80
3.6	Treadmill experiments	83
3.6.1	Limb angle function	85
3.6.2	Maximum likelihood estimation	88
3.6.3	Reconstruction error analysis	88
3.6.4	Analysis of the limb function	92
3.6.5	Motion model discussion	101
3.6.5.1	Gait reconstruction function	103
3.6.5.2	Stride and cadence	105
3.6.5.3	Hip joint motion	106
3.6.5.4	Marker configuration	108
3.6.6	Biometric identification	109
3.6.7	Intra and inter-class variation	111
3.6.7.1	Discrimination between individual parameter features	112
	Normalized limb segment lengths.	112
	Normalized amplitude components.	113
	Normalized phase components.	115
3.6.7.2	Uniqueness of the proposed biometric	117
3.7	Experimental procedure	120

3.8	Conclusions	122
4	Pose Invariant Gait Reconstruction	126
4.1	Introduction	126
4.2	A stratified approach to linear trajectory gait reconstruction	128
4.2.1	Periodicity and the imaged direction of motion	130
4.2.1.1	Self-similarity error metric	132
4.2.1.2	Recovering periodicity from self-similarity	137
4.2.2	Recovering affine structure	139
4.2.3	Recovering metric structure	144
4.2.4	Recovering gait dynamics	147
4.2.5	Stratified reconstruction analysis	151
4.3	Maximum likelihood estimation	155
4.4	Imposing a rigid motion transform model	159
4.5	Reconstruction analysis	162
4.5.1	Intra and inter-class variation	166
4.5.1.1	Variation of reconstructed limb motion from different camera views	170
4.5.1.2	Variation of parameters between treadmill and overground walking	171
4.6	Reconstruction discussion	176
4.6.1	Lateral displacement	176
4.6.2	Fourier coefficients	179
4.6.3	Periodic gait series function	180
4.6.4	Propagation of covariance in initial stratified reconstruction	180
4.7	Conclusions	181
5	Total Parameterisation of Generalized Gait Motion	186
5.1	Introduction	186
5.2	Imaged ground plane motion	188
5.3	Experimental set-up	192
5.3.1	Markers	193
5.3.2	Motion segmentation	194
5.4	System parameterisation	196
5.5	Biometric parameterisation	197
5.6	Fusion of subject biometric parameters	200
5.6.1	Linear computation of biometric parameters	200
5.6.2	Maximum likelihood estimation	201
5.7	Fusion of system parameters	202
5.7.1	Recovering the camera intrinsic parameters	203
5.7.2	Known principal point	205
5.7.3	Known aspect ratio	206
5.7.4	Pose decomposition	207
5.7.5	Maximum likelihood estimation	209
5.8	Fusion discussion	215
5.9	Global optimization of parameters	217
5.9.1	General sparse LM minimization	218

5.9.2	Covariance matrix	220
5.9.3	Block sparse LM method	222
5.10	Reconstruction analysis	228
5.10.1	System parameters	229
5.10.2	Subject and pose parameters	236
5.11	Conclusions	240
5.12	Tables of parameter estimates	243
6	Conclusions	247
6.1	Restatement of hypotheses	247
6.2	Contributions	248
6.3	Further work	253
6.4	Publications associated with this thesis	255
A	Projective Geometry	256
A.1	Classes of planar transformations	256
A.1.1	Euclidean Transformation	256
A.1.2	Similarity Transformation	257
A.1.3	Affine Transformation	259
A.1.4	Perspective Transformation	262
A.1.5	Cross Ratio	263
B	Matrices and Linear Systems	265
B.1	Orthogonal regression	265
B.1.1	Constrained orthogonal regression	267
B.1.2	Orthogonal projection of points onto a line	268
B.2	Singular value decomposition	269
B.2.1	Pseudo-inverse	269
B.2.2	Least squares solution of linear equations	270
B.2.3	Least squares solution of homogeneous equations	271
B.2.4	Approximating a 3×3 matrix by a rotation matrix	271
B.3	Parameterising a rotation matrix	272
B.3.1	Rodrigues Formula	273
C	Non-Linear Minimization	274
C.1	Newtonian iteration	274
C.2	Levenberg-Marquardt minimization	275
C.3	Confidence limits	280
C.4	Sparse Levenberg-Marquardt minimization	283
C.4.1	General sparse LM method	283
C.4.2	Covariance matrix	285
C.4.3	Block sparse LM method	285
D	Implementation and Software	289
D.1	Introduction	289
D.2	Testing	289
D.3	Geometry Tools	290
D.3.1	Primitive Visualisation	290

D.3.2 Planar geometry	291
D.3.3 Multiple view geometry	293
D.4 Gait system	296
Bibliography	297

For my parents, Mike and Sally Spencer.

Acknowledgements

Thanks to the usual suspects - Pete Myerscough, Lee Middleton and Rob Boston for their infectious enthusiasm, time and help within the laboratory. I am indebted to my supervisors, Dr. John Carter and Prof. Mark Nixon for their invaluable guidance. I would also like to thank my colleagues and friends whom have made my time at Southampton an enjoyable and rewarding experience.

I am grateful for the tenure of a studentship from the Engineering and Physical Sciences Research Council (EPSRC). I must also acknowledge partial support from the European research office of the US Army under contract No. N68171-01-C9002 during the 'Human ID at a Distance' programme.

Chapter 1

Introduction

This thesis answers many of the important questions with regard to the reconstruction and identification of human motion. The bulk of the work concentrates on answering three simple questions:

- What is gait and how accurately can we represent it?
- What are the unique features of motion and how do they vary over a range of walking speeds?
- How can we reconstruct the fronto-parallel view of gait from an oblique trajectory of imaged subject motion?

We use a passive marker system to identify interest feature locations on the human body, thus removing the necessity for a complex vision system to track and automatically label the anatomical landmark features. Correspondingly, all interest features within an image sequence are manually labelled. The emphasis throughout this thesis is on the geometric properties of human gait and the projective geometry techniques used to reconstruct a consistent view of the limb motion.

This chapter introduces the subject of biometrics, identifies the important properties required for a successful biometric, then gives a brief discussion on the current and future role for biometrics within the UK. We then introduce a number of possible application domains for the analysis of human movement and gait recognition.

Each chapter is self contained and addresses a clearly defined problem. Subsequently, the critical research material and prior work is reviewed within each of the relevant chapters. We give here a brief literature review of prior work that is similar in application or has influenced the development of this thesis in some way. We then state the hypotheses that were subsequently investigated. Finally, we provide a breakdown of the thesis structure and give a brief overview of the content within each of the chapters.

1.1 Biometrics

With the growing importance of applications requiring human identification, the demand for adequate security measures has increased dramatically. In response to this demand, new technologies are being introduced aimed to ensure that the requisite level of security can be achieved. One of these technologies is often referred to as biometrics. Biometrics concern any human physiological or behavioural characteristics [51] which are: (i) *Universal*, every person should have that characteristic; (ii) *Unique*, no two people should be the same in terms of that characteristic; (iii) *Permanent*, invariant with time; (iv) *Collectable*, can be measured quantitatively; (v) *Reliable*, must be safe and operate at a satisfactory performance level; (vi) *Acceptable*, non-invasive and socially tolerable; and (vii) *Non-circumventable*, how easily the system is fooled into granting access to impostors.

An important advantage of biometrics lies in the fact that physical or behavioural traits cannot be transferred to other individuals. Examples of biological characteristics that have been explored for their potential as biometrics so far are face, fingerprints, DNA, hand geometry, vein geometry, iris and retinal patterns, signature, voice, gait and ear. Unfortunately, most biometrics are not perfect. Fingerprints and hand geometry are reliable but require physical contact. Although signatures based on face and iris are non-intrusive in nature, the applicability of these methods is restricted to very controlled environments.

With the current global drive towards biometric enabled identity and travel documents, many experts feel that extensive research and development is needed to ensure the successful use of biometrics for large-scale identification projects. The scale of the ID card project in Britain is certainly daunting in terms of the history of past Whitehall computer projects. A report released by the European Commission on 30 March 2005 warned that, on the technological side, there is currently a lack of independent empirical data.

The UK government has announced plans for a compulsory national identity card scheme. The scheme is to be phased in over a number of years and will include basic personal information and biometric identifiers. These unique biometric identifiers (such as iris pattern or fingerprints) will help prevent people's identities being stolen and will also securely confirm a person's identity when a card is checked. The introduction of the first identity cards will, on current plans, start from 2008.

The Home Secretary also announced on 11 November 2003 that the UK Passport Service (UKPS) would run a biometric pilot. The aim of the pilot was to evaluate issues surrounding biometric recording using facial recognition, iris pattern and fingerprints. The pilot ran between April and December 2004.

While 10,000 volunteers participated in the trial, the detailed research released by the

UKPS [81] focused on 2,000 quota people, picked to match the general population, and 750 disabled people. The UKPS stressed that the trial's aim was to measure people's reaction to having biometric data collected, not the technology's effectiveness at gathering facial, iris and fingerprint information. Nevertheless, the results showed that there were still problems with individual types of biometrics.

The test covered both the initial enrolment and verification for each biometric. While there were high rates of success, 90% or more for enrolment on all three types of biometric, the results for the verification varied. For the facial biometric, which measures the distance between various points on the face, the success rate was 69% for the quota group and 48% for the disabled people who took part. It should however be noted that disability was not a failure factor, as the majority of disabled participant verifications took place in a mobile enrolment centre where lighting conditions adversely affected all facial verifications. Iris verification was better, with a 96% quota group success rate and 91% amongst the disabled volunteers. Fingerprint verification was successful in 81% of the quota group and 80% of the disabled group, with problems occurring with the machine not recording enough detail from the fingers.

Another potential headache for the UKPS is the amount of time it takes to physically gather the biometric information from individuals. According to the report, it took an average of 7 minutes, 56 seconds for people in the quota group to be enrolled, and an average of 8 minutes and 15 seconds for those in the disabled group.

According to the UKPS, the trial results have highlighted several issues that require further investigation. Among other things, further trials are needed, specifically targeted towards those disabled groups that have experienced enrolment difficulties due to environment design, biometric device design, or to specific group problems.

1.2 Human movement

Traditionally there has been keen interest in human movement from a wide variety of disciplines. There is a rich literature, including medical and psychological studies, indicating the potential of gait in personal identification [78]. In psychology, there have been the classic studies on human perception by Johansson [52]. Experiments with moving light displays attached to body parts showed that human observers can almost instantly recognize biological motion patterns, even when presented with only a few of these moving dots. Further studies confirmed that we can correctly discriminate between the gender of a walker [57]. Studies of human locomotion found that male walkers tend to swing their shoulders more, while female walkers their hips [69]. In kinesiology the goal has been to develop models of the human body that explain how it functions mechanically and how one might increase its movement efficiency. A typical procedure

involves obtaining 3D joint data, performing kinematic analysis, and computing the corresponding forces and torques for a movement of interest. 3D data is typically obtained in an intrusive manner, by placing markers on the human body. Typical clinical analysis of gait data is performed on a case by case basis, with concerted efforts towards understanding the bio-mechanical significance of particular deviations from normal gait patterns. The aim of medical research has been to classify the components of gait for the treatment of pathologically abnormal patients. Early medical studies suggest that if all gait movements are considered then gait is unique. Murray's work [76, 75] indicates that gait contains 24 different components giving it the richness necessary for a successful biometric.

An important application domain of human motion recognition is in smart surveillance. Here the system does more than motion detection, a straightforward task that is prone to false alarms (the system is easily fooled by the presence of moving vehicles, animals and natural effects such as wind producing motion in trees). The first capability is to sense that humans are indeed present. This is followed by person tracking and biometric identification for the purpose of access control. A typical application may be to identify suspicious behaviour of individuals walking around in a car park, repeatedly looking through car windows. The benefits of such applications need in some cases to be balanced with possible drawbacks e.g. regarding privacy.

Biometrics are not perfect and many suffer from social and practical problems. One may need to make physical contact with systems, for example fingerprinting, or suffer social embarrassment when interrogating a voice recognition system. Biometrics that need no physical contact, such as face recognition, are more acceptable to users, but can be limited by practical issues such as face visibility. Gait is one of the newest biometrics and has the potential of overcoming many of these problems. One of the unique advantages of using gait as a biometric is that it can be perceived from a distance, making acquisition non-invasive and convenient. Biometrics such as iris and retinal patterns, ear and face recognition require high resolution images. A typical surveillance camera often captures poor quality, low resolution images. Gait suffers less from this shortcoming because the body has a proportionally larger area compared with the eyes or face. Furthermore, it cannot be easily disguised without impeding one's natural gait.

1.3 Prior work

An informative survey of the current analysis techniques to date, regarding human movement has been outlined by Gavrilu [36]. He surveys the work on visual analysis of gestures and whole body movement, and identifies the various areas of research current approaches have taken. The various application domains can be categorized into three main sections.

- 2D approaches without explicit shape models.
- 2D approaches with explicit shape models.
- 3D approaches.

Here we describe some of the techniques adopted by researchers in the field of human gait analysis and identification.

Statistical based approaches derive a unique signature by computing a spatio-temporal motion pattern based on a sequence of segmented images of a moving person. Typically, the shape of the body is reduced to a binary silhouette and some statistical measures are taken from this sequence of silhouettes. Techniques such as Principal Components Analysis and Linear Discriminant Analysis have been used to provide a statistical description of the sequence [49, 48]. These techniques are fairly successful on small subject populations, achieving high recognition rates. However, it is not clear exactly which features of gait contribute to the recognition process.

Little and Boyd [64, 65] develop a model-free description of instantaneous motion, the shape of motion, that varies with the type of moving figure and the type of motion. Subjects are discriminated by periodic variations in the shape of their motion. The image flow of a moving figure varies both spatially and temporally. For each image in a sequence, dense optical flow is derived. Scale independent features based on moments of the moving point characterize the spatial distribution of the flow. The periodic structure of these sequences of scalars is analysed, which have the same fundamental period but differ in phase. Some phase features are consistent for one person and show significant statistical variation among people, enabling individuals to be identified by the shape of their motion.

Polana and Nelson [85] develop a shape free technique that uses a periodicity measure based on grey level signals extracted along spatio-temporal reference curves. They define the periodicity measure of a signal as the normalized difference between the sum of the spectral energy at the highest amplitude frequency and its multiples, and the sum of the energy at the frequencies halfway between. Similar work by Picard [66, 67] develops this idea further by consideration of harmonic energy ratios and applying better spectral thresholding. Both algorithms consist of two main parts: i) Object tracking by frame alignment, which transforms data into a form in which periodicity can be easily detected and measured; ii) Fourier spectral harmonic peak detection and energy computation, to identify regions of periodicity and measure its strength. Their methods allow simultaneous detection, segmentation and characterization of spatio-temporal periodicity, and is computationally efficient.

Cutler and Davis [21, 22] detect periodic motion by first segmenting the motion and tracking objects in the foreground. Objects are aligned along the temporal axis and

the self-similarity of the object is computed as it evolves in time. Periodic motions result in the self-similarity metric also being periodic, thus time-frequency analysis is applied to robustly detect and characterize this inherent periodicity within the 2D lattice structures.

Hayfron-Acquah and Nixon [43, 44] describe a method for automatic gait recognition based on analysing the symmetry of human motion by using the generalized symmetry operator. This operator, rather than relying on the edges of a shape or on general appearance, locates features by their symmetrical properties. Analysis also suggests that symmetry offers practical advantages, such as relative immunity to noise, missing frames and the capability to handle occlusion. Promising recognition rates of over 95% have been achieved, with the ability to discriminate between human and animal motions.

Foster and Nixon [34] present a statistical area based metric, called gait masks. Gait masks aim to combine holistic and model based approaches by using statistical data from human silhouettes that is intimately related to the nature of gait. Each gait mask aims to isolate a portion of the image and measure the change within that area. The masks are intuitively chosen to capture meaningful information about the shape of motion of a subject's gait. Early results show promising results with a recognition rate of 90% on a small database of human subjects. Furthermore, gait masks can also be used to provide information about the periodicity of gait.

Another method is to base recognition on a physical model of human motion. Perhaps the earliest approach to gait recognition was to derive a gait signature from a spatio-temporal motion pattern [79, 80]. The initial model for the walker is simple. A walker is a translating blob which has braided spatio-temporal patterns in the lower half of the subject's body. By recognizing these spatio-temporal signatures, a model can be imposed for subsequent analysis. A fixed length feature vector is composed by interpolating the set of angles formed by fitting a stick figure model to the motion. This vector is then used for classification by the simple K-nearest neighbour of Euclidean distance.

Other researchers have chosen to model hip rotation by a simple pendulum [19], whose motion is approximately described as simple harmonic, thus can be expressed by a Fourier series. This has also been extended to include the lower leg by using a bilaterally symmetric and coupled oscillator model [117]. The gait signature is created from the phase weighted magnitude of the lower order Fourier components of both the thigh and lower leg rotations. These signatures can be used for recognition either by running or walking [118, 119], though large differences between both running and walking modes of motion mean that there is no unified biometric signature [116].

Tanawongsuwan and Bobick [103, 102] explore the spatio-temporal gait parameters of stride length and cadence over a number of controlled walking speeds. They give a detailed study of motion obtained from treadmill walking and show that there is a linear relationship between stride length, cadence and measured gait speed which can

be used to normalize subject motions to a common framework for further comparison.

Constrained motion in laboratories is often not the same as seen in the real world. In particular, subject pose and camera positions are all well controlled in laboratory experiments, usually by forcing the subject to walk fronto-parallel to the camera. Before these techniques can be of practical use, the many different effects that may perturb and influence recognition rates must be quantified. Clearly some form of invariance or correction is required to normalize the signatures of walking subjects to be independent of the imaged effects of pose.

BenAbdelkader and Cutler [5, 6] describe two different gait recognition methods: a non-parametric method that uses the self-similarity plot of a walking sequence as the input feature for classification; and a parametric method that estimates the spatio-temporal parameters of gait (the cadence and stride length) and exploits their linear relationship as a cue for identification. Normalizing for the variation in camera viewpoint is not possible, except for very small changes. This is because a different (planar) projection of gait dynamics is captured in the image from any one camera viewpoint. Hence, it is necessary to index the gait recognition method by different ranges of camera viewpoint.

Carter and Nixon [11, 12] use a model based perspective technique to correct measured gait angles from geometrically marked limb positions, imaged under oblique trajectory angles. They demonstrate that normalized gait signatures based on phase and high order amplitude measurements provide some invariance to changes in pose.

Kale and Chowdhury [55] propose a method for synthesizing arbitrary views of planar objects. They apply this technique to recover the fronto-parallel motion of subjects walking at arbitrary angles to the camera. Their method uses a perspective projection model and an optical flow based structure from motion strategy for estimating the azimuth angle of the original view from monocular image sequences. However, they require some information about the camera calibration in order to compute these synthesized views.

Taylor [104] investigates the problem of recovering information about the configuration of an articulated object, such as a human figure, from point correspondences within a single image. The reconstruction method does not assume that the camera is calibrated and uses a weak perspective projection model. Correspondingly, there are a family of solutions to this reconstruction problem, parameterised by a single scalar variable. A simple algorithm is developed for recovering the entire set of solutions by considering the foreshortening of limb segments within the image.

Liebowitz and Carlsson [61] describe an algorithm for 3D reconstruction of dynamic articulated structures, such as humans, from uncalibrated, multiple views of subject motion. The reconstruction exploits constraints associated with a dynamic articulated structure, specifically the conservation over time of lengths between rotational joints.

These constraints allow metric reconstruction from at least two different images, in each of two uncalibrated orthographic projection cameras. The method also extends to pairs of cameras that are zooming, where calibration of the cameras allows compensation for the changing scale factor in a scaled orthographic camera.

1.4 Research outline

A varied and large range of research has been undertaken within the field of automatic gait recognition. However, in almost all circumstances the situations have been constrained for particular motions. Most research has been restricted to subjects walking fronto-parallel to the camera with a single natural gait speed. Recognition rates obtained by these techniques, when processing sequences of motion acquired with different geometric pose and speed configurations, are typically quite low. Currently proposed gait features have poor intra-person reliability due to their dynamic nature, being dependent on various physiological, psychological and external factors such as footwear, clothing, load carrying, surface of walking, mood, illness and fatigue. Realistically, people will always walk with varying speeds along different motion trajectories to the camera. It is this particular problem that requires further attention, in order to better understand and recover suitable biometric measurements for the purpose of identification at a distance.

Any successful biometric has static features that are invariant to the day to day circumstantial variations in subject motion. Researchers [103, 102] have already made some headway into removing variations between intra-person gait motion, by choosing to map subject locomotion to a common walking speed in order to facilitate further comparison. Gait is though not a one to one function. A person can achieve a desired gait speed in a number of different ways. There exists a mapping between cadence and stride length that allows us to alter the mode of walking whilst maintaining the required velocity. Normalization to a common gait speed then provides some invariance over a range of statistically average gait motions, though does not completely remove the problem.

What we should really be looking for are geometric or algebraic properties of gait that are perforce unchanged by the circumstances of gait motion. Limb segment lengths correspond to rigid bone sections, thus remain fixed over the entire image sequence. Since we are unable to gauge depth from monocular views and are unable to determine the true length measures of any of the subject's limbs, the fronto-parallel reconstruction corresponding to an imaged sequence of subject motion can only be determined up to scale. Length ratios are invariant to changes in scale and can easily be extracted from the reconstructed view. Provided we can recover the motion dynamics corresponding to the fronto-parallel view, then the ratios between limb segment lengths make ideal biometric parameters.

We must be mindful of the usefulness of any technique that we wish to develop. It may be

possible to provide good classification results from a system involving dozens of cameras. In practice this may not be feasible due to either or both, the cost involved in camera resources and the circumstances in which motion sequences are acquired. A typical surveillance application makes maximum use of the hardware, i.e. placement of cameras in a convenience store is usually designed to maximize the coverage of the entire shop. In these circumstances redundant coverage of the same areas by cameras is wasteful and is typically avoided, thus multi-view methods are often impractical. This is not to say that we should discount these methods altogether, but that identification from monocular sequences is more commonly apparent. We take this as a basis for our investigation into viewpoint invariance, and to this end devote the remaining chapters of this thesis to the theory and development of invariance from monocular motion sequences.

We hypothesize that gait has the following features and properties.

- **Human locomotion can be modelled as a collection of dynamically moving, articulated limb segments.** Each limb is connected to the trunk and is composed of a number of inter-connected bone and joint structures. Each bone segment is rigid and of fixed length. These bone segments are allowed to freely pivot about the corresponding joint positions, although only within a constrained arc of motion.
- **Articulated leg motion is approximately planar.** While in reality the displacement of leg motion is within all three Euclidean directions, almost all the of the perceived motion is contained within a single plane. The variation of motion out of this plane is subtle and negligible in comparison to this major motion plane.
- **Normal gait is bilaterally symmetric with a half phase shift.** Walking uses a repetitious sequence of leg motion to move the body forward. This series of events is repeated by each leg with reciprocal timing. The stance period of one leg equals the swing of the other, thus motion on one leg swing plane is related to the motion of the other by a period of half the gait cycle.
- **Natural gait motion is piecewise linear.** In general, people tend to walk in straight lines with constant velocity. Deviation from this assumption infers inconsistent, non-repetitious limb motion and consequently suggests unnatural gait. Imaged gait can then be split piecewise into natural segments of gait motion.
- **Each individual has a set of possibly unique static features.** The static geometric features of gait that remain invariant over time are based on the fixed length measurements of limb segments. Similarly, there are static motion features that are derived from the representation of the dynamic leg motion function. These features are invariant to the circumstantial changes in subject motion such as stride length, cadence and consequently gait speed.

In this thesis we show that gait has sufficient properties that allows us to reconstruct the fronto-parallel view of gait motion. We identify the static, geometric and motion features of gait and show that these features remain invariant over a range of circumstantial gait motions for a small trial group of four people. Furthermore, if two or more sufficiently different imaged trajectories of gait motion are available, though not necessarily of the same person, then the calibration of the camera can be determined. Subsequently, the total pattern of gait motion can be globally parameterised for all subjects within an image sequence. We present the details of a sparse method that computes the maximum likelihood estimate of this set of parameters, then conclude with a reconstruction error analysis corresponding to an example image sequence of generalized subject motion.

1.5 Thesis structure

The remainder of this thesis is arranged as follows.

- **Chapter 2: Stratified Recovery of Planar Gait Motion.**

This chapter introduces the main geometric ideas and notation that are required to understand the remaining material covered in this thesis. In particular, the chapter covers camera models, image projection and planar geometry. Geometric properties of the plane can be classified into three main groups of transformation: perspective, affine and similarity transformations. Identification of specific entities within the image allows us to employ a stratified technique to map them back to their canonical positions. Metric structure of the scene plane is typically recovered in a two step process: i) Identification of the imaged vanishing line of the scene plane allows us to compute the perspective transformation that recovers the affine properties of the plane. ii) Identification of the imaged circular points then allows us to compute the affine transformation that recovers the metric properties of the plane.

We show that gait has sufficient properties that allows us to exploit the structure of planar leg motion in order to remove the unknown subject pose and reconstruct the canonical motion pattern, with no prior knowledge of the camera calibration. As an example, this stratified reconstruction technique is applied to a synthesized image of an obliquely viewed human motion figure pattern, in order to recover the canonical fronto-parallel view.

- **Chapter 3: Static Features of Human Gait.**

This chapter is concerned with identifying the features of gait which remain invariant to the circumstantial changes in gait motion. We first give a brief overview of the terminology and biomechanics of subject motion from the medical literature, in order to better understand the nature of gait. We then provide a quantitative

verification and analysis of the planar leg motion assumption. We first compute the true 3D positions of marked leg joints over an image sequence, via triangulation from multiple calibrated camera views. A scene plane is fitted to these reconstructed points, and a statistical analysis of the orthogonal deviation between sample points and the motion plane is then given.

We proceed by defining a suitable limb angle function, based on a modified Fourier series, that is able to represent the dynamics of subject leg motion. Finally, we analyse the behaviour of these motion parameters over a range of controlled walking speeds, for a number of subjects who walk on a treadmill. We emphasize parameter properties that remain invariant over these speeds and outline a biometric feature vector suitable for recognition purposes.

Throughout the chapter, we consider more closely four out of the seven properties of biometrics: *Universal*, *Unique*, *Collectable* and *Reliable*. Where possible, we identify the major sources of error between the motion model and imaged data. We quantify the level of reconstruction error over a range of walking speeds, with a small trial set of four subjects. An analysis of the intra and inter class variance is given for each of the proposed biometric features of gait motion that allows us to comment on their capability to discriminate between subjects under different motion conditions.

- **Chapter 4: Pose Invariant Gait Reconstruction.**

The work in this chapter develops the geometric properties and biometric features identified within chapters 2 and 3, though with respect to real human motion sequences. The human body is inherently a dynamically moving, articulated 3D motion structure. The articulated leg motion of both left and right sides of the body can be approximated by motion within two separate planes. We can apply further constraints on the form of the articulated leg motion by simultaneous consideration of the bilateral symmetry between left and right leg. This allows us to combine and maximize the utilization of the imaged sample data in order to compute a more robust estimation of gait motion.

We develop a novel method that uses the geometric properties of articulated leg motion to compute a stratified reconstruction of the fronto-parallel dynamics of gait motion. After recovering the fronto-parallel structure of subject gait, we can then determine the representation of the bilateral leg angle function by fitting a modified Fourier series to the data.

Parameterisation of subject motion is split into two phases: i) *Limb stance*, limb pose positions on the metric plane are determined by evaluating the biometric limb angle function. ii) *Pose projection*, the orientation and displacement of a subject's worldspace limb swing plane is determined by the homography transformation that projects metric plane structure into the image plane.

A method is presented to compute the maximum likelihood estimate of these recon-

struction parameters. An analysis of the resultant reconstructions over a number of different subject trajectories is given for a small trial set of four people. We outline the intra and inter class variance of the proposed biometric feature vector for the set of reconstructed subject motion trajectories. We also cross compare the intra and inter class variation of these reconstructed overground walking motions with the set of treadmill walking motions from chapter 3.

The chapter concludes with a brief discussion on a number of possible improvements and considerations that can be made to the stratified reconstruction technique.

- **Chapter 5: Total Parameterisation of Generalized Gait Motion.**

Since people tend to walk from point to point in straight lines, any generalized gait motion can be approximated by a set of straight line motion segments. Each piecewise linear segment of reconstructed gait has a set of biometric motion parameters and a corresponding set of pose projection parameters. Common to all subject pose projections are the intrinsic coefficients of the camera. Similarly, each subject has a common set of underlying biometric gait parameters within each of the reconstructed linear motion segments. The work presented in this chapter is concerned with recovering the intrinsic parameters of the camera and the underlying limb function of subject motion. Subsequently, we can determine the set of worldspace subject poses from the parameterised set of limb swing plane mappings. Subjects typically walk within a flat ground plane, hence the configuration and parameterisation of subject motion is specialized further. Details of a sparse minimization technique are then given that computes the maximum likelihood estimate of the set of partitioned model parameters corresponding to subject motion over the entire image sequence. The chapter concludes with an analysis of the reconstruction results obtained from a real image sequence of subject motion around multiple trajectory segments of a test track. A discussion on the measured uncertainties within the parameterisation is then given and a number of explanations for the major sources of error considered further.

- **Chapter 6: Conclusions.**

We summarize and put into context the ideas and results of all previous chapters with respect to real articulated human motions. We highlight the successes of the many practical methods developed throughout the project and discuss areas that warrant further improvement. We finish by outlining the novel contributions made within this work to the field of gait analysis and recognition.

- **Appendices.**

Further background material and mathematical proofs that are not intimately related to the main body of work are presented within the four appendices at the end of the thesis. Appendix A outlines the transformation properties of the set of perspective, affine and similarity classes of planar geometry. Appendix B is

concerned with linear systems of equations and describes a number of numerical algorithms that are frequently used throughout this work. Appendix C reviews the mathematics of non-linear optimization methods that are used in order to compute the maximum likelihood estimates of a set of model parameters. In particular, we discuss the form of the sparse minimization techniques that are frequently encountered throughout this thesis. Finally, appendix D outlines the details of the software developed during this period of study. We give details of a number of software tools that were implemented by the author to enable manual marking of point features within an image sequence, camera calibration, triangulation of worldspace structure and image rectification. This chapter demonstrates the significant practical contribution made by the author in conjunction with this work.

Chapter 2

Stratified Recovery of Planar Gait Motion

2.1 Introduction

This chapter introduces the main geometric ideas and notation that are required to understand the remaining material covered within this thesis. In particular, the chapter covers the geometry of projective transformations of the plane. These transformations model the geometric distortion which arises when a plane is imaged by a perspective camera. Under perspective imaging certain geometric properties are preserved, such as collinearity, while others are not. Parallel worldspace lines are not in general imaged as parallel lines. Projective geometry models this imaging and also provides a mathematical basis that is appropriate for representing the transformation of worldspace structure.

A detailed modern description of the mathematics of geometry is given in [9], which describes a number of different geometries including affine, projective, inversive, non-Euclidean and spherical geometries. For a good overview and background into perspective imaging see the appendix of [74], or for a complete study to date of projective geometry within computer vision we refer you to the texts of Hartley and Zisserman [40], Cipolla [13], and Faugeras and Luong [27].

We review here the basic properties, nomenclature, and essential details of projective geometry. We begin by describing the representation of 2D points, lines and conics in homogeneous notation, and how these entities map under transformation of the image plane. This is followed by a discussion on central projection and the development of the camera model.

The camera is a remarkable measuring device. It not only captures a realistic picture of a scene, but also provides information from which geometric properties of imaged structure can be measured. Reconstructing scene geometry from images is one of the most active

areas in computer vision. Reconstruction techniques all require some form of calibration; through direct methods prior to scene capture or auto-calibration methods that use scene and motion constraints between views, such as parallelism and orthogonality of scene lines.

We first describe the simple linear pinhole projection model and relate the intrinsic camera coefficients with the physical properties of the camera projection. We then turn our attention to more complex camera models and emphasize how to represent non-linear distortion effects such as radial distortion. We finalize our discussion on cameras with the details of various camera calibration methods. We describe the mathematics and procedures required to reliably compute accurate worldspace measurements through the process of back projection from multiple camera views.

In a world crafted with flat surfaces, worldspace structure and detail commonly occurs within single planes, e.g. printed text on a book page. Projection of planar worldspace structure into the image then has a much simpler form. In general, transformation of structure from one plane to another is achieved by a 3×3 matrix mapping known as a homography. We review the mathematics and essential properties of planar geometry then give examples of some commonly occurring specialized planar transformations, such as image mosaicing [99, 100] and repetition of planar patterns [89, 90].

The homography that maps structure from one plane to another may be directly computed by using linear methods with only the knowledge of known point correspondences between both planes. On the other hand, we may not have physical point correspondences between the canonical reference plane and the image plane. We may though have some additional knowledge about the structure on the reference plane (orthogonality of vectors, known ratios of lengths, etc.) that allows us to reconstruct the canonical plane.

There has recently been much interest in Euclidean reconstruction from uncalibrated views [41, 16]. Liebowitz and Zisserman [62, 63] describe a stratified approach to scene plane reconstruction that allows us to combine a number of different constraints within a unified framework, in order to compute the set of transformations that recover the metric properties of the scene plane. Geometric properties of the plane can be classified into three main groups of transformation: perspective, affine and similarity transformations. Each class of transformation distorts image structure differently, thus there are geometric entities that are only affected by a specific class of transformation. Identification of these entities within the image then allows us to compute the transformation that rectifies them back to their canonical positions.

We conclude the chapter by putting this planar geometry theory into practice, in order to validate our assumptions about the nature of gait motion. Most current research assumes that imaged gait motion is fronto-parallel. This may be viable in certain constrained situations where we are able to force subjects to walk in such a manner. In general, subject motion is never truly fronto-parallel, thus a correction transformation is required

to reconstruct the canonical dynamics of gait locomotion.

We hypothesize that human limb motion during natural gait is approximately planar and that people tend to walk, over a number of gait cycles, along linear trajectories with constant velocity. The periodicity of planar limb motion is then analogous to the specialized geometry of repeating planar patterns, and allows us to recover the imaged direction of subject motion and the vanishing line of the limb swing plane. Within the worldspace, articulated limb segments remain fixed in length, thus form exactly the type of planar constraints required to compute a stratified reconstruction of the limb swing plane.

We set up an experiment in which several different gait poses within a gait cycle are represented by planar skeletal figures printed onto a sheet of paper. An oblique motion trajectory is synthesized by placing the pattern on a flat surface and taking a picture with a camera orientated at an angle to this scene plane. A stratified reconstruction technique is then applied to the marked limb segment endpoints in order to recover the fronto-parallel view of the figure motion pattern. We demonstrate that gait has sufficient properties that allows us to exploit the structure of articulated limb motion within single view sequences, in order to remove the unknown subject pose and reconstruct the underlying gait signature, with no prior knowledge of the camera calibration.

2.2 Representation of 2D primitives

Geometric primitives that lie on the image plane, such as points, lines and conics, can be represented in a number of different ways. They often arise as the projections of three dimensional primitives within the worldspace, i.e. points, planes and quadrics. We give here a mathematical description of the homogeneous representation of these entities.

The image plane coordinate system is customarily aligned with the X axis to the right and the Y axis downward. Following this convention, the worldspace coordinate system is also aligned with the X axis to the right and the Y axis downward. Subsequently, the Z axis is aligned into or toward the viewing plane, as shown in figure 2.1. To disambiguate between lines in the image plane and projective construction lines that pass through the image, we refer to these projection lines as rays.

2.2.1 Points

A point on a plane may be represented by the pair of coordinates (x, y) in \mathbb{R}^2 . We can also represent the same point homogeneously as (x_1, x_2, x_3) , a ray through the origin in \mathbb{R}^3 , i.e. the line that passes through the Euclidean coordinates $(0, 0, 0)$ and (x_1, x_2, x_3) .

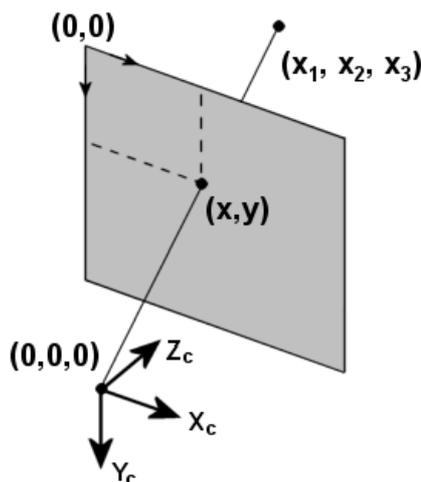


FIGURE 2.1: The homogeneous representation of a point (x, y) in \mathbb{R}^2 is a point (x_1, x_2, x_3) in \mathbb{R}^3 . The ray that passes through the origin and (x_1, x_2, x_3) intersects the $z = 1$ plane in the inhomogeneous coordinate (x, y) .

This homogeneous ray vector is related to the coordinate pair (x, y) in the image plane of \mathbb{R}^2 by the mapping.

$$(x, y) \mapsto \left(\frac{x_1}{x_3}, \frac{x_2}{x_3}, 1 \right) \quad (2.1)$$

In general, if the point in \mathbb{R}^3 representing the inhomogeneous point (x, y) is written as (x_1, x_2, x_3) then $(\lambda x_1, \lambda x_2, \lambda x_3)$ also represents the same (x, y) coordinate, with $\lambda \neq 0$. This implies that a homogeneous point represents an \mathbb{R}^2 coordinate up to scale.

The homogeneous representation of a point in \mathbb{R}^n can be expressed as a vector \mathbb{R}^{n+1} . We therefore represent the inhomogeneous Euclidean 3D point (x, y, z) homogeneously by a vector (x_1, x_2, x_3, x_4) in \mathbb{R}^4 .

2.2.2 Lines

A line on the image plane in \mathbb{R}^2 is represented homogeneously by a plane in \mathbb{R}^3 passing through both the image plane line L and the origin point O .

Since a line is represented by a plane in \mathbb{R}^3 , it must consist of the set of Euclidean points (x, y, z) that satisfy the equation $ax + by + cz = 0$, where (a, b, c) are real and not all zero. It has the homogeneous scaling property $\lambda ax + \lambda by + \lambda cz = 0$, with $\lambda \neq 0$, and represents the line L on the image plane in \mathbb{R}^2 simply by the coefficients of the plane normal vector $N = (a, b, c)$.

Any two distinct points in the image plane \mathbb{R}^2 lie on a unique line. The condition for a point to lie on the line can then be written homogeneously as the inner product between

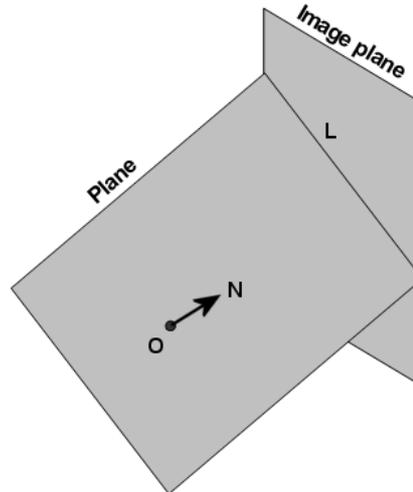


FIGURE 2.2: The homogeneous representation of a line L in \mathbb{R}^2 is a plane N in \mathbb{R}^3 which passes through both the origin O and the line L .

the point X and the line L .

$$X \cdot L = 0 \quad (2.2)$$

$$(\lambda_1 x, \lambda_1 y, \lambda_1) \cdot (\lambda_2 a, \lambda_2 b, \lambda_2 c) = 0$$

$$\lambda_1 \lambda_2 \cdot (ax + by + c) = 0$$

We introduce here the vector notation for representing homogeneous entities. In the usual manner, the product of a matrix and a vector is another vector, which brings up the distinction between column and row vectors. Since a matrix may pre-multiply a column vector to its right, it is conventional to represent geometric entities as column vectors. A bold face symbol \mathbf{x} will always represent a column vector, and its transpose \mathbf{x}^\top a row vector. Using this convention, we can represent column vectors as $(x_1, x_2, x_3)^\top$ within the body of a text paragraph. The condition for a point \mathbf{x} to lie on a line \mathbf{l} in this notation can be represented by the inner product.

$$\mathbf{x}^\top \mathbf{l} = \mathbf{l}^\top \mathbf{x} = 0 \quad (2.3)$$

If we use the fact that two distinct points define a line and our homogeneous representation of a line is a plane in \mathbb{R}^3 , then any further points on this plane in \mathbb{R}^3 must also lie on the line in \mathbb{R}^2 .

Consider finding the line between the two homogeneous points $\mathbf{x}_1 = (1, -1, 1)^\top$ and $\mathbf{x}_2 = (2, -1, 4)^\top$. Any third point $\mathbf{x}_3 = (x, y, z)^\top$ on the line also lies on the plane in \mathbb{R}^3

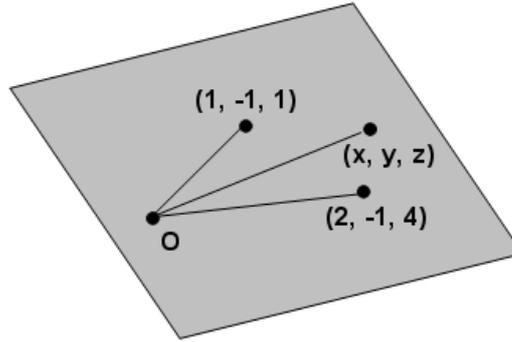


FIGURE 2.3: Any point on the plane in \mathbb{R}^3 also lies on the line in \mathbb{R}^2 . All other points on the plane can be expressed as linear combinations of the two homogeneous position vectors.

and must be represented as a linear combination of the two homogeneous position vectors \mathbf{x}_1 and \mathbf{x}_2 . Writing out the matrix determinant:

$$\begin{vmatrix} x & y & z \\ 2 & -1 & 4 \\ 1 & -1 & 1 \end{vmatrix} = 0 \quad (2.4)$$

and expanding for the first row:

$$x \begin{vmatrix} -1 & 4 \\ -1 & 1 \end{vmatrix} - y \begin{vmatrix} 2 & 4 \\ 1 & 1 \end{vmatrix} + z \begin{vmatrix} 2 & -1 \\ 1 & -1 \end{vmatrix} = 0 \quad (2.5)$$

$$3x + 2y - z = 0$$

is the equation of the plane in \mathbb{R}^3 or by writing out only the coefficients as a vector $(3, 2, -1)^\top$, is our representation of the homogeneous line. Checking this inhomogeneously, we see that the line in \mathbb{R}^2 is $3x + 2y - 1 = 0$ and the inhomogeneous points are $(\frac{2}{4}, -\frac{1}{4})$ and $(1, -1)$, then by substitution into the equation of the line.

$$3 \left(\frac{2}{4} \right) + 2 \left(-\frac{1}{4} \right) - 1 = 0$$

$$3(1) + 2(-1) - 1 = 0$$

We note from the determinant property that homogeneously, the line between two homogeneous points $\mathbf{x}_1, \mathbf{x}_2$ is determined from the cross product $\mathbf{l} = \mathbf{x}_1 \times \mathbf{x}_2$. The dual to this is that two lines meet at a point, which is defined by interchanging the role of the points and lines, i.e. $\mathbf{x} = \mathbf{l}_1 \times \mathbf{l}_2$.

2.2.3 Homogeneous scale of points and lines

Since points and lines are represented up to scale, we must remove the homogeneous scale factor in order to determine if two representations are similar. We can remove the homogeneous scale ambiguity by taking the cross product. Two parallel vectors \mathbf{u}, \mathbf{v} have zero cross product $\mathbf{u} \times \mathbf{v} = \mathbf{0}$, regardless of any differences in homogeneous scale. This provides an essential means for comparing any two similar representations. The cross product may also be written in a convenient matrix form.

$$\mathbf{u} \times \mathbf{v} = [\mathbf{u}]_{\times} \mathbf{v} \quad (2.6)$$

where the expansion $[\mathbf{u}]_{\times}$ corresponds to the 3×3 skew-symmetric matrix.

$$[\mathbf{u}]_{\times} = \begin{pmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{pmatrix} \quad (2.7)$$

This property will be exploited later in order to provide linear constraints between worldspace and imaged entities that allows us to compute the corresponding transformation mapping.

2.2.4 Ideal points and the line at infinity

An important property of representing 2D primitives homogeneously is that we can easily describe the set of points infinitely far away. This is not true in the inhomogeneous geometry where the set of infinite points form a special case. This set of points are known as the ideal points and can be written $\mathbf{x} = (x_1, x_2, 0)^{\top}$, i.e. a set of vectors on the $z = 0$ plane. Since the $z = 0$ plane and the image plane are parallel then any such vector \mathbf{x} will only cut the image plane at infinity. The set of ideal points can be encoded by a single angle θ , in a one parameter family of vectors, as illustrated in figure 2.4(a).

$$\mathbf{x} = \lambda(\cos \theta, \sin \theta, 0)^{\top} \quad (2.8)$$

where λ is an arbitrary homogeneous scaling factor.

An arbitrary plane passing through the origin cuts the image plane in a line \mathbf{l} . The set of points on this line can subsequently be encoded by a one parameter family of vectors on the cutting plane. This construction is shown in figure 2.4(b) and illustrates that any set of co-planar homogeneous vectors represent points on a single image line. The

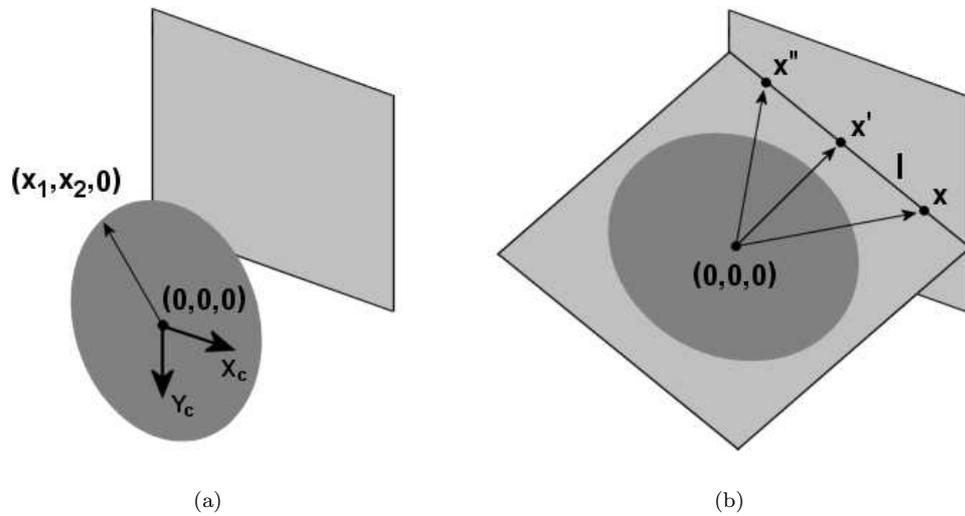


FIGURE 2.4: (a) The set of ideal points lie on the $z = 0$ plane and are encoded by an angle θ , in a one parameter family of vectors $\mathbf{x} = \lambda(\cos \theta, \sin \theta, 0)^\top$. (b) Points on an image line \mathbf{l} can be encoded by a one parameter family of vectors on the plane cutting this line.

set of ideal points must therefore lie on a single line that is located at infinity. This line is fixed up to scale and is known as the line at infinity \mathbf{l}_∞ .

$$\mathbf{l}_\infty = \lambda(0, 0, 1)^\top \quad (2.9)$$

where λ is an arbitrary homogeneous scaling factor.

The ideal points and line at infinity form an important set of constraints on a workspace scene plane. Identification of the images of these quantities allows us to recover metric structure from the imaged plane.

2.2.5 Conics

A conic is a curve described by a second degree equation in the plane. In Euclidean geometry, conics are of three main types: hyperbola, ellipse, and parabola (apart from the degenerate conics). These three types arise as conic sections, generated by planes of differing orientation, as illustrated in figure 2.5. The degenerate conics (single or double lines) arise from planes which contain the cone vertex.

The equation of a conic in inhomogeneous coordinates can be written:

$$ax^2 + bxy + cy^2 + dx + ey + f = 0 \quad (2.10)$$

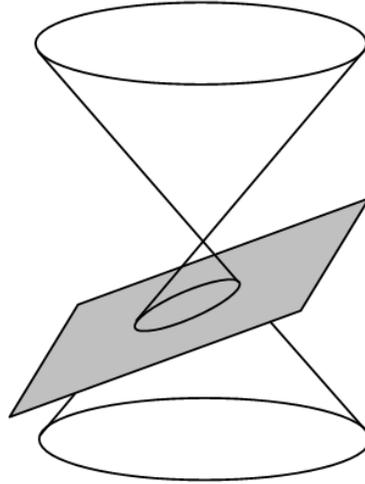


FIGURE 2.5: Elliptical conic section formed from the intersection of a plane through the cone.

This can be homogenized by replacing the coefficients $x \mapsto x_1/x_3$ and $y \mapsto x_2/x_3$.

$$\frac{ax_1^2}{x_3^2} + \frac{bx_1x_2}{x_3^2} + \frac{cx_2^2}{x_3^2} + \frac{dx_1}{x_3} + \frac{ex_2}{x_3} + f = 0 \quad (2.11)$$

$$ax_1^2 + bx_1x_2 + cx_2^2 + dx_1x_3 + ex_2x_3 + fx_3^2 = 0 \quad (2.12)$$

or in matrix form:

$$\mathbf{x}^\top \mathbf{C} \mathbf{x} = 0 \quad (2.13)$$

where the conic coefficient matrix \mathbf{C} is given by:

$$\mathbf{C} = \begin{pmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & f \end{pmatrix} \quad (2.14)$$

The coefficient matrix is symmetric and like the point and line homogeneous representations is defined up to scale, giving it five degrees of freedom. A minimum of five points are then required to compute the coefficients \mathbf{C} of a conic. From the inhomogeneous representation of the conic equation.

$$ax_i^2 + bx_iy_i + cy_i^2 + dx_i + ey_i + f = 0 \quad (2.15)$$

Each point places one constraint on the coefficients $\mathbf{c} = (a, b, c, d, e, f)^\top$, hence the constraints can be stacked to form the design matrix.

$$\begin{bmatrix} x_1^2 & x_1 y_1 & y_1^2 & x_1 & y_1 & 1 \\ & & \vdots & & & \\ x_5^2 & x_5 y_5 & y_5^2 & x_5 & y_5 & 1 \end{bmatrix} \mathbf{c} = \mathbf{0} \quad (2.16)$$

The set of coefficients \mathbf{c} are defined up to scale. The solution to this homogeneous system of equations of the form $\mathbf{Ax} = \mathbf{0}$ is then solved by singular value decomposition with the constraint that $\|\mathbf{c}\| = 1$, see appendix B.2.3 for further details.

2.2.6 Polarity and conjugacy of conics

The polar of a point \mathbf{x} with respect to a conic \mathbf{C} is the line $\mathbf{l}_x = \mathbf{Cx}$. Geometrically, the polar is the line through the points of tangency to \mathbf{C} , corresponding to the pair of lines through \mathbf{x} . Dually, $\mathbf{x} = \mathbf{C}^{-1}\mathbf{l}_x$ is the pole of \mathbf{l}_x with respect to \mathbf{C} .

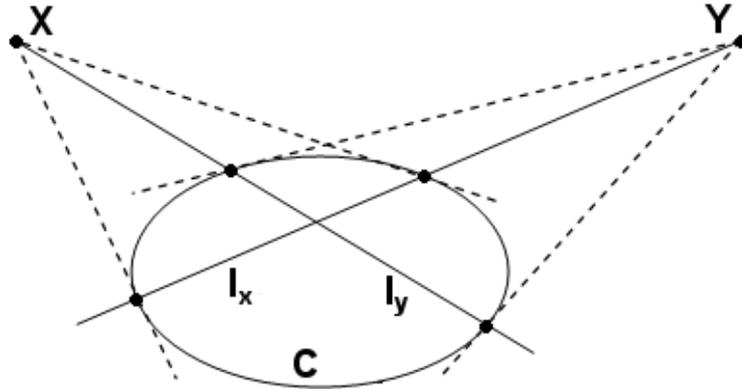


FIGURE 2.6: The polar of \mathbf{x} is the line \mathbf{l}_x . Points \mathbf{x} and \mathbf{y} are conjugate if \mathbf{x} lies on the polar \mathbf{l}_y of \mathbf{y} , and \mathbf{y} lies on the polar \mathbf{l}_x of \mathbf{x} .

A pair of points \mathbf{x} and \mathbf{y} are conjugate with respect to a conic \mathbf{C} if \mathbf{x} lies on the polar \mathbf{l}_y of \mathbf{y} and \mathbf{y} lies on the polar \mathbf{l}_x of \mathbf{x} . The conjugate points and lines satisfy the condition.

$$\mathbf{x}^\top \mathbf{C} \mathbf{y} = 0 \quad (2.17)$$

$$\mathbf{l}_x^\top \mathbf{C}^{-1} \mathbf{l}_y = 0 \quad (2.18)$$

2.3 Camera model

A camera consists of an image plane and a lens which provides a transformation between the world and image spaces. This transformation can not be perfectly described by perspective projection because of distortions which occur through the lens. These distortions can be modelled. However, the model may only be an approximation to the real relationship. How closely the model conforms to reality will depend on the functional representation and how well the corresponding parameters can be estimated.

The basic pinhole camera model projects worldspace points onto the image plane. The centre of projection \mathbf{C} lies at the origin of a right handed Euclidean coordinate system. Following the convention corresponding to most image formats, the X pixel axis is aligned to the right and the Y pixel axis downward. It is also customary to place the image plane in front of the camera centre. Subsequently, the camera Z axis (principal axis) is then aligned into or toward the image plane, as shown in figure 2.7.

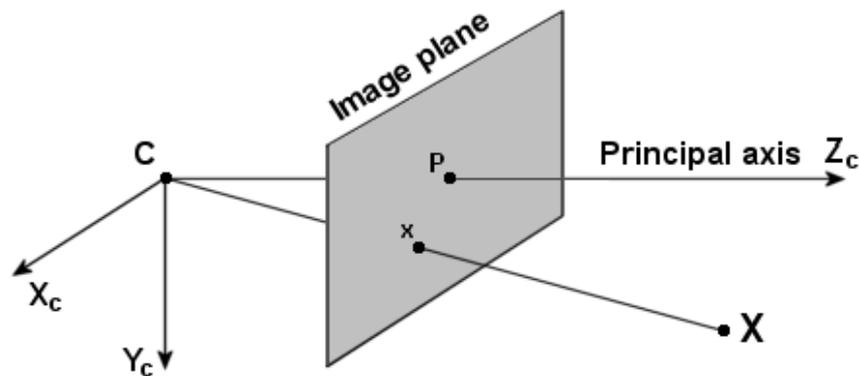


FIGURE 2.7: Pinhole camera geometry. \mathbf{C} is the camera centre of projection and \mathbf{P} the principal point (the point where the principal axis meets the image plane).

Under central projection, an inhomogeneous worldspace point $\mathbf{X} = (x, y, z)^\top$ is mapped to the corresponding point on the image plane by a ray that passes through both \mathbf{X} and the centre of projection \mathbf{C} . By using the arrangement of similar triangles, the worldspace point $(x, y, z)^\top$ is mapped to the image plane point as $(fx/z, fy/z, f)^\top$, where f is the camera focal length. Ignoring the final projected z coordinate of the focal plane, the first two coefficients then correspond to the required $(u, v)^\top$ image coordinates.

2.3.1 Linear projection

If the world and image points are represented by homogeneous vectors, then central projection is simply expressed as a linear mapping between their homogeneous coordinates.

$$\begin{pmatrix} fx \\ fy \\ z \end{pmatrix} = \begin{bmatrix} f & & 0 \\ & f & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (2.19)$$

This matrix expression can be written $\mathbf{x} = \mathbf{P}\mathbf{X}$, where \mathbf{P} is the camera projection matrix. The camera projection matrix can then be decomposed.

$$\begin{aligned} \mathbf{P} &= \begin{pmatrix} f & & \\ & f & \\ & & 1 \end{pmatrix} \begin{bmatrix} 1 & & 0 \\ & 1 & 0 \\ & & 1 & 0 \end{bmatrix} \\ \mathbf{P} &= \mathbf{K}_f [\mathcal{I} \mid \mathbf{0}] \end{aligned} \quad (2.20)$$

where $[\mathcal{I} \mid \mathbf{0}]$ represents the 3×3 identity matrix augmented with a column 3-vector of zeros. Up to this point, the image plane coordinate system is coincident with the principal point. A further set of scaling and translation transformations are then applied in order to model the relative XY size differences within the camera CCD elements and the coordinate shift between the optical and image coordinate systems. The projection of homogeneous worldspace points into the image is then given by the mapping.

$$\mathbf{P} = \mathbf{K} [\mathcal{I} \mid \mathbf{0}] \quad (2.21)$$

where the camera calibration matrix \mathbf{K} has the form.

$$\mathbf{K} = \begin{pmatrix} m_x & \alpha & u_0 \\ & m_y & v_0 \\ & & 1 \end{pmatrix} \quad (2.22)$$

The parameter α is referred to as the *skew* parameter. The skew parameter will be zero for most normal cameras, however in certain unusual instances it can take non zero values.

2.3.2 Radial distortion

In reality, light passing through the lens is distorted differently over the surface of the image. The main lens distortion effect is seen radially from the camera principal point. Tangential distortion effects may also be modelled, but in general are negligible in comparison to the radial components. We use here a symmetric distortion model

$\mathbf{w}' = f(r) \cdot \mathbf{w}$ with two distortion coefficients k_1, k_2 to facilitate the warping of light through the camera lens.

$$f(r) = 1 + k_1 \cdot r^2 + k_2 \cdot r^4 \quad (2.23)$$

where the radial component r is the distance of the lens ray point $\mathbf{w} = (w_x, w_y)^\top$ from the camera principal point, $r^2 = (w_x)^2 + (w_y)^2$.

The projection of worldspace structure onto the camera CCD sensor plane is achieved through a three step process, illustrated in figure 2.8.

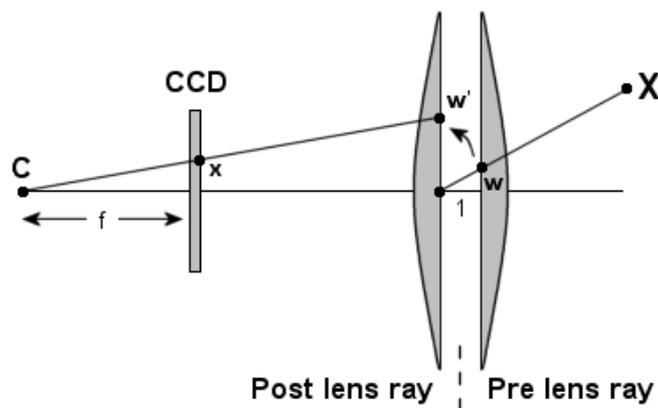


FIGURE 2.8: Three stage process of the non-linear projection of worldspace structure into image by modelling lens distortion effects.

1. The first step uses the canonical camera to determine the projection of the light ray entering the lens $\mathbf{w} = [\mathcal{Z} \mid \mathbf{0}] \mathbf{X}$.
2. The symmetric, non-linear lens distortion function $\mathbf{w}' = f(r) \cdot \mathbf{w}$ is then applied to the inhomogeneous lens plane point.
3. The final focal projection step of the ray leaving the lens onto the camera CCD elements is then modelled by the pin hole projection $\mathbf{x} = \mathbf{K}\mathbf{u}$, where $\mathbf{u} = (w'_x, w'_y, 1)^\top$ is the distorted post lens ray point.

The final step includes the mapping from the optical to image coordinate systems. Other distortion functions can be used [68, 45, 114]. The distortion function is essentially a Taylor's series corresponding to the radial distance of the lens ray from the principal point. More complex distortion functions choose to model the centre of distortion and the principal point separately, and may also include components of tangential lens distortion.

2.3.3 Camera extrinsic parameters

We have assumed thus far that the camera is located at the origin of the Euclidean coordinate system. In general, points in space will be expressed in terms of a different coordinate frame, known as the world coordinate frame. Camera and world coordinate frames are related via a rotation \mathbf{R} and a translation \mathbf{t} . Figure 2.9 illustrates this coordinate frame mapping from the world space system to the local camera system.

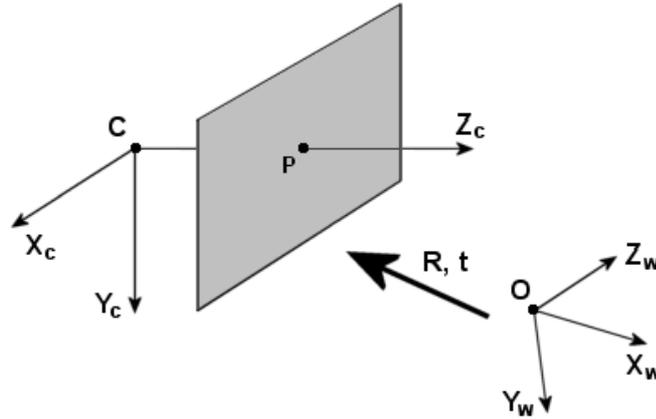


FIGURE 2.9: Euclidean transformation between the world and camera coordinate frames.

If $\widetilde{\mathbf{W}}$ is an inhomogeneous 3-vector representing the coordinates of a point in the world coordinate frame, and $\widetilde{\mathbf{X}}$ represents the same point in the camera coordinate frame then the transformation from world to local camera coordinates can be expressed as $\widetilde{\mathbf{X}} = \mathbf{R}\widetilde{\mathbf{W}} + \mathbf{t}$. The camera projection matrix \mathbf{P} that maps homogeneous points \mathbf{X} from the worldspace coordinate frame into the image can then be expressed by the homogeneous projection transformation $\mathbf{x} = \mathbf{P}\mathbf{X}$. The projection matrix \mathbf{P} is a 3×4 matrix with a total of 11 degrees of freedom.

$$\mathbf{P} = \mathbf{K}[\mathbf{R} \mid \mathbf{t}] \quad (2.24)$$

2.4 Planar transformations

Structure and detail is often confined to lie on a scene plane within the worldspace, e.g. an image on a billboard or markings on a road surface. In such cases, projection of points from the scene plane to the image plane has a simplified form, and is characterized by a 3×3 matrix mapping known as a homography.

2.4.1 Planar homography

Points on the workspace scene plane can be identified by a set of Euclidean $(u, v)^\top$ coordinates. We can choose to align the world coordinate frame with the scene plane such that any 3D point \mathbf{X} on this plane has zero z component and can be represented by the homogeneous vector $\mathbf{X} = (u, v, 0, w)^\top$.

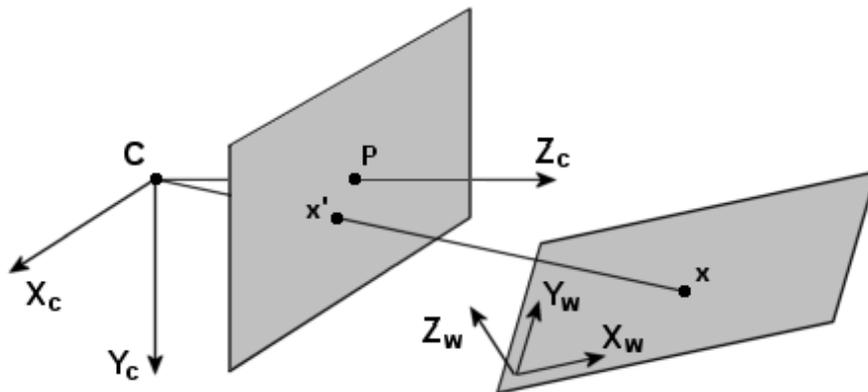


FIGURE 2.10: Homography mapping between the workspace scene plane and the image plane.

The corresponding projection $\mathbf{x}' = \mathbf{P}\mathbf{X}$ of points into the image then has a reduced form. Writing the columns of the world to camera coordinate frame rotation matrix \mathbf{R} as \mathbf{r}_i then the projection of planar structure into the image is given by

$$\mathbf{x}' = \mathbf{K} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \mathbf{t} \end{bmatrix} \begin{pmatrix} u \\ v \\ 0 \\ w \end{pmatrix} \quad (2.25)$$

$$\mathbf{x}' = \mathbf{K} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix} \begin{pmatrix} u \\ v \\ w \end{pmatrix} \quad (2.26)$$

$$\mathbf{x}' = \mathbf{H}\mathbf{x} \quad (2.27)$$

The form of the homography mapping \mathbf{H} is determined by the orientation of the workspace plane and the intrinsic parameters of the calibration matrix. There are a number of specific classes of planar transformation that arise as a result of the various orientations and possible camera parameters within the projection. The main class types can be categorized as perspective, affine and similarity transformations. Each class of planar transformation plays an important part within the context of stratified planar reconstruction. Further important properties and geometric invariants of these transformation classes

are summarized within appendix A.1.

2.4.2 Image mosaicing

Another plane to plane transformation that occurs in practice is when a camera with fixed intrinsic parameters rotates freely about its origin. This situation is common within many CCTV surveillance systems, where a degree of rotational mobility is available to a remote operator within a control room.

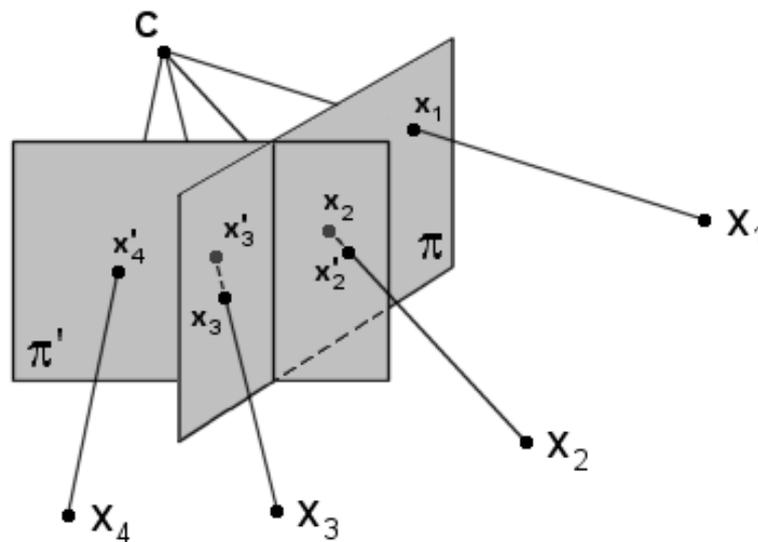


FIGURE 2.11: Planar correspondence that occurs as a result of a freely rotating camera with fixed intrinsic parameters about its origin point.

Fixed worldspace points are collinear with the camera centre of projection \mathbf{C} and the corresponding projected image plane points within different views, as illustrated in figure 2.11. The image correspondences enable us to compute the planar homographies between rotationally different camera planes and build a panorama of images within a single common reference plane.

2.4.3 Transformation of points, lines and conics

If a point \mathbf{x} lies on the line \mathbf{l} then $\mathbf{l}^\top \mathbf{x} = 0$. We can then apply a homography transformation $\mathbf{x}' = \mathbf{H}\mathbf{x}$ that maps points to a second plane. From this we can deduce the corresponding planar transformation rule that similarly transfers lines onto the second plane.

$$\begin{aligned} \mathbf{l}^\top \mathbf{x} &= 0 \\ (\mathbf{l}^\top \mathbf{H}^{-1}) (\mathbf{H}\mathbf{x}) &= 0 \\ \mathbf{l}'^\top \mathbf{x}' &= 0 \end{aligned}$$

Hence, points and lines are transformed:

$$\mathbf{x}' = \mathbf{H}\mathbf{x} \quad (2.28)$$

$$\mathbf{l}' = \mathbf{H}^{-\top} \mathbf{l} \quad (2.29)$$

Similarly, by using the point transformation rule, point conics are transformed:

$$\begin{aligned} \mathbf{x}^\top \mathbf{C}\mathbf{x} &= 0 \\ (\mathbf{x}^\top \mathbf{H}^\top) (\mathbf{H}^{-\top} \mathbf{C}\mathbf{H}^{-1}) (\mathbf{H}\mathbf{x}) &= 0 \\ \mathbf{x}'^\top \mathbf{C}' \mathbf{x}' &= 0 \end{aligned}$$

Hence, the point conic \mathbf{C} is transformed as:

$$\mathbf{C}' = \mathbf{H}^{-\top} \mathbf{C}\mathbf{H}^{-1} \quad (2.30)$$

2.4.4 Vanishing points and lines

Ideal points and the line at infinity are important entities of a worldspace scene plane and can be used as a source of constraints for image rectification. In general, the ideal points and line at infinity are mapped to finite entities after image projection. Writing \mathbf{h}_i as the column vectors of the homography transformation \mathbf{H} then the set of ideal points $(u, v, 0)^\top$ are mapped to the corresponding set of imaged points.

$$\begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3 \end{bmatrix} \begin{pmatrix} u \\ v \\ 0 \end{pmatrix} = u \cdot \mathbf{h}_1 + v \cdot \mathbf{h}_2 \quad (2.31)$$

Providing that the first two coefficients of the third row of the homography matrix \mathbf{H} are non zero then the ideal points are mapped to finite points in the image. Figure 2.12 shows geometrically the distortion of parallel scene lines and the corresponding imaged vanishing point within the projected plane.

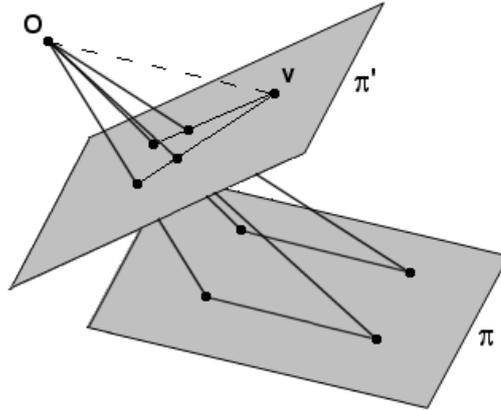


FIGURE 2.12: Parallelism of lines is not preserved by projective transformation. The ideal point defined by the intersection of the parallel lines on the scene plane π is imaged as a finite point \mathbf{v} in the image plane π' .

The corresponding transformation rule for lines $\mathbf{l}' = \mathbf{H}^{-\top} \mathbf{l}$ maps scene plane lines into the image. Writing \mathbf{m}_i as the column vectors of the matrix $\mathbf{H}^{-\top}$ then the line at infinity $\mathbf{l}_\infty = (0, 0, 1)^\top$ is mapped into the image as:

$$\begin{bmatrix} \mathbf{m}_1 & \mathbf{m}_2 & \mathbf{m}_3 \end{bmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \mathbf{m}_3 \quad (2.32)$$

Providing that the first two coefficients of the third row of the homography matrix \mathbf{H} are non zero then the line at infinity \mathbf{l}_∞ is mapped to a finite line \mathbf{l}'_∞ within the image.

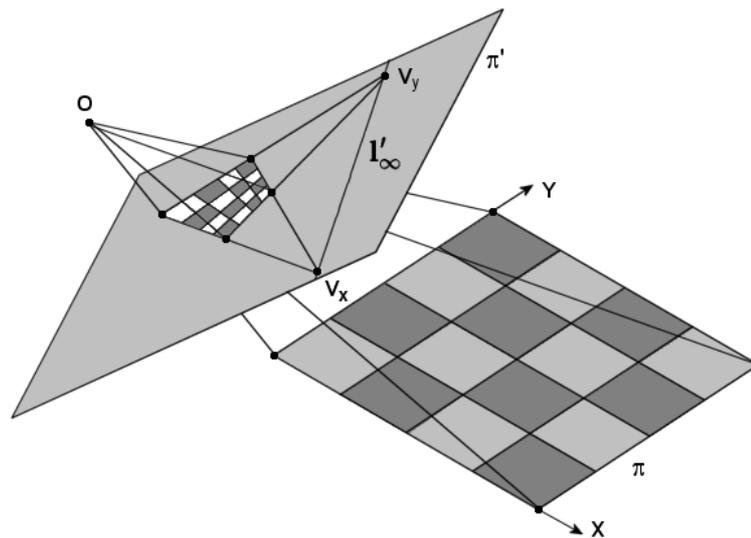


FIGURE 2.13: The two ideal points defined by intersection of parallel lines on the scene plane π are imaged as the finite points \mathbf{v}_x and \mathbf{v}_y in the image plane π' . These points lie on the image of the vanishing line of the scene plane \mathbf{l}'_∞ .

Figure 2.13 shows the geometric distortion of the ideal line of the scene plane after projection into the image. Identification of the corresponding set of imaged vanishing points and vanishing line of the scene plane enables us to compute the homography transformation that, in part or fully restores the metric properties of the scene plane structure.

2.4.5 Repeating planar patterns

Repetition of planar structure is common-place within the real world, e.g. bricks in a house wall or lamp-posts down the side of a street. In such cases, the form of the planar homography that maps imaged points across the set of repeated correspondences is specialized [89, 90]. Consider the planar mapping $\mathbf{x}'_i = \mathbf{H}\mathbf{x}_i$ that transforms points from one plane π_1 to a second plane π_2 . A Euclidean translation of a point on the first plane is mapped to a corresponding conjugate translation of the point on the second plane.

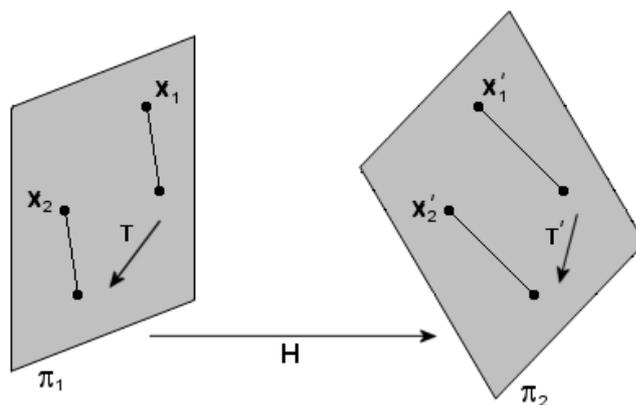


FIGURE 2.14: Repeated planar structure corresponding to a conjugate Translation of points.

Figure 2.14 shows the mapping of points from one plane π_1 to a second plane π_2 , and similarly the respective translation and conjugate translation between the repeated point correspondences on these planes. The Euclidean translation that maps the first point \mathbf{x}_1 to the second \mathbf{x}_2 on the scene plane π_1 can be represented by the homography \mathbf{T} . There is then a corresponding conjugate translation \mathbf{T}' on plane π_2 between the imaged points \mathbf{x}'_1 and \mathbf{x}'_2 .

$$\begin{aligned}
\mathbf{x}'_1 &= \mathbf{H}\mathbf{x}_1 \\
\mathbf{x}_2 &= \mathbf{T}\mathbf{x}_1 \\
\mathbf{x}'_2 = \mathbf{H}\mathbf{x}_2 &= \mathbf{HT}(\mathbf{H}^{-1}\mathbf{x}'_1) \\
\mathbf{T}' &= \mathbf{HTH}^{-1}
\end{aligned} \tag{2.33}$$

The point transfer \mathbf{T} on the first plane can be thought of as a translation in the X-axis direction with distance λ and can therefore be written.

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & \lambda \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \mathcal{I} + \lambda \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \tag{2.34}$$

It follows from the result $\mathbf{T}' = \mathbf{HTH}^{-1}$ that the corresponding conjugate translation can be written:

$$\mathbf{T}' = \mathbf{H} \left[\mathcal{I} + \lambda \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \right] \mathbf{H}^{-1} \tag{2.35}$$

$$\mathbf{T}' = \mathbf{H}\mathcal{I}\mathbf{H}^{-1} + \lambda\mathbf{H} \left[\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \right] \mathbf{H}^{-1} \tag{2.36}$$

$$\mathbf{T}' = \mathcal{I} + \lambda\mathbf{v}\mathbf{l}'_{\infty}{}^{\top} \tag{2.37}$$

Where \mathbf{v} is the imaged X-axis direction of motion on π_2 , \mathbf{l}'_{∞} the imaged vanishing line of the first plane in the second π_2 and λ a scalar constant. Further conjugate repeated points on the second plane are determined at integer multiples of λ within the transformation \mathbf{T}' , i.e. $(\lambda, 2\lambda, \dots, N\lambda)$.

Here, we have described the form of the conjugate translation with respect to point transformation between planes. Without too much effort the conjugate translations can be derived for both the repeated line and conic planar transformations. We will though not discuss them further here.

2.4.6 Direct linear transformation

If we know the correspondence points between two planes $\mathbf{x}_i \mapsto \mathbf{x}'_i$ then we can directly compute the homography transformation \mathbf{H} that maps one set to the other. We can remove the homogeneous scaling ambiguity by taking the cross product between the transferred point $\mathbf{H}\mathbf{x}_i$ and the corresponding point \mathbf{x}'_i on the second plane.

$$\mathbf{x}'_i \times \mathbf{H}\mathbf{x}_i = \mathbf{0} \quad (2.38)$$

We can then rearrange the set of equations into the form $\mathbf{A}\mathbf{h} = \mathbf{0}$, where elements of the column 9-vector \mathbf{h} are made up from the rows of the homography mapping matrix \mathbf{H} . Writing \mathbf{h}^j as the column 3-vector corresponding to the j^{th} row of \mathbf{H} , then the transferred points from the first plane $\mathbf{H}\mathbf{x}_i$ can be written.

$$\mathbf{H}\mathbf{x}_i = \begin{pmatrix} \mathbf{h}^{1\top} \mathbf{x}_i \\ \mathbf{h}^{2\top} \mathbf{x}_i \\ \mathbf{h}^{3\top} \mathbf{x}_i \end{pmatrix} \quad (2.39)$$

Writing the measured points on the second plane as $\mathbf{x}'_i = (u'_i, v'_i, w'_i)^\top$, then the equations formed from the cross product have the form.

$$[\mathbf{x}'_i]_{\times} \mathbf{H}\mathbf{x}_i = \begin{pmatrix} v'_i \mathbf{h}^{3\top} \mathbf{x}_i - w'_i \mathbf{h}^{2\top} \mathbf{x}_i \\ w'_i \mathbf{h}^{1\top} \mathbf{x}_i - u'_i \mathbf{h}^{3\top} \mathbf{x}_i \\ u'_i \mathbf{h}^{2\top} \mathbf{x}_i - v'_i \mathbf{h}^{1\top} \mathbf{x}_i \end{pmatrix} \quad (2.40)$$

The inner product can be commuted such that $\mathbf{h}^{j\top} \mathbf{x}_i = \mathbf{x}_i^\top \mathbf{h}^j$, then by substitution the system of equations of the form $\mathbf{A}\mathbf{h} = \mathbf{0}$ can be computed.

$$\begin{bmatrix} \mathbf{0}^\top & -w'_i \mathbf{x}_i^\top & v'_i \mathbf{x}_i^\top \\ w'_i \mathbf{x}_i^\top & \mathbf{0}^\top & -u'_i \mathbf{x}_i^\top \\ -v'_i \mathbf{x}_i^\top & u'_i \mathbf{x}_i^\top & \mathbf{0}^\top \end{bmatrix} \begin{pmatrix} \mathbf{h}^1 \\ \mathbf{h}^2 \\ \mathbf{h}^3 \end{pmatrix} = \mathbf{0} \quad (2.41)$$

Only two of the three equations are linearly independent. It is customary to only include the first two sets of equations from each point correspondence [98], within the design matrix \mathbf{A} . Each point correspondence then forms two individual constraints on the elements of \mathbf{h} . The solution vector is defined up to scale (8 degrees of freedom) thus a total of four correspondences are required to compute a minimal solution. The solution of \mathbf{h} is given by computing the singular value decomposition of the homogeneous system of equations of the form $\mathbf{A}\mathbf{h} = \mathbf{0}$, see appendix B.2.3 for further details.

The result of the Direct Linear Transformation (DLT) algorithm for computing planar homographies depends on the coordinate system within which points are expressed. Coordinate systems may vary due to the nature of the problem, with some being better than others with respect to the computed result. Improved accuracy of the results can be achieved by performing an initial data normalization step to transform the set of points into a common coordinate system.

Similarity transforms (isotropic scale and translation) \mathbf{T} and \mathbf{T}' are first applied to each of the respective point sets to normalize them to lie within the unit square. The DLT algorithm is then performed to compute $\tilde{\mathbf{H}}$, the planar mapping between normalized point sets $\tilde{\mathbf{x}}_i$ and $\tilde{\mathbf{x}}'_i$. The homography mapping between the original point sets is then given by undoing the effects of the data normalization transformations.

$$\tilde{\mathbf{x}}'_i = \tilde{\mathbf{H}}\tilde{\mathbf{x}}_i \quad (2.42)$$

$$\mathbf{T}'^{-1}\tilde{\mathbf{x}}'_i = \mathbf{T}'^{-1}\tilde{\mathbf{H}}(\mathbf{T}\mathbf{x}_i) \quad (2.43)$$

$$\mathbf{x}'_i = (\mathbf{T}'^{-1}\tilde{\mathbf{H}}\mathbf{T})\mathbf{x}_i \quad (2.44)$$

$$\mathbf{H} = \mathbf{T}'^{-1}\tilde{\mathbf{H}}\mathbf{T} \quad (2.45)$$

2.5 Planar geometry

A 3D point \mathbf{X} is represented by a homogeneous four vector $\mathbf{X} = (X_1, X_2, X_3, X_4)^\top$ and a plane π described by $\pi = (\pi_1, \pi_2, \pi_3, \pi_4)^\top$. Points that lie on the plane satisfy the condition $\pi^\top \mathbf{X} = 0$. Correspondingly, three points define a unique plane and three planes define a point.

Analogously to the ideal points and the line at infinity of any worldspace scene plane, the ideal entities of the 3D space correspond to the set of ideal points $\mathbf{X}_\infty = (x, y, z, 0)^\top$ and the plane at infinity $\pi_\infty = (0, 0, 0, 1)^\top$. Identification of the images of these entities enables us to recover, in part or fully, the metric worldspace structure. Since π_∞ is indeed a plane, structure on it is mapped into the image via a 3×3 planar homography transformation.

2.5.1 Absolute conic

The absolute conic was introduced within the vision literature by Faugeras [30] and plays an important role within camera calibration and scene reconstruction. The absolute conic $\mathbf{\Omega}_\infty$ lies on the plane at infinity π_∞ and is often expressed by the equation.

$$X_1^2 + X_2^2 + X_3^2 + X_4^2 = 0, \quad X_4 = 0 \quad (2.46)$$

This may be re-written to resemble the more familiar conic form $\mathbf{x}^\top \mathbf{C} \mathbf{x} = 0$.

$$(X_1 \ X_2 \ X_3) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = 0 \quad (2.47)$$

$$(X_1 \ X_2 \ X_3) \mathbf{\Omega}_\infty (X_1 \ X_2 \ X_3)^\top = 0 \quad (2.48)$$

Correspondingly, $\mathbf{\Omega}_\infty$ is a circular point conic with complex radius i [28] of purely imaginary points on the plane at infinity π_∞ . Every scene plane in space intersects π_∞ in a line, and each of these lines intersects $\mathbf{\Omega}_\infty$ in the circular points of the ideal plane $\mathbf{I} = (1, i, 0)^\top$ and $\mathbf{J} = (1, -i, 0)^\top$. The absolute conic and circular points of the plane at infinity π_∞ remain fixed under any similarity transformation, see appendix A.1.2.

2.5.2 Image of the absolute conic

Any ideal 3D point can be written $\mathbf{X} = (\mathbf{W}^\top, 0)^\top$ and is projected into the image via the perspective transformation $\mathbf{x} = \mathbf{P}\mathbf{X}$, where the projection matrix \mathbf{P} has the form $\mathbf{P} = \mathbf{K}[\mathbf{R} \mid \mathbf{t}]$.

$$\begin{aligned} \mathbf{x} &= \mathbf{K}[\mathbf{R} \mid \mathbf{t}]\mathbf{X} \\ \mathbf{x} &= \mathbf{K}\mathbf{R}\mathbf{W} \end{aligned} \quad (2.49)$$

A point lies on the absolute conic $\mathbf{\Omega}_\infty = \mathcal{I}$ if $\mathbf{W}^\top \mathbf{\Omega}_\infty \mathbf{W} = 0$. Using equation 2.49 we can make the substitution $\mathbf{W} = \mathbf{R}^\top \mathbf{K}^{-1} \mathbf{x}$ within the conic constraint condition.

$$(\mathbf{R}^\top \mathbf{K}^{-1} \mathbf{x})^\top \mathbf{\Omega}_\infty \mathbf{R}^\top \mathbf{K}^{-1} \mathbf{x} = 0 \quad (2.50)$$

$$\mathbf{x}^\top \omega \mathbf{x} = 0 \quad (2.51)$$

where the image of the absolute conic (IAC) is given by

$$\omega = \mathbf{K}^{-\top} \mathbf{K}^{-1} \quad (2.52)$$

The IAC is only dependent on the coefficients of the camera intrinsic parameters \mathbf{K} .

2.5.3 Camera calibration

The history and development of camera calibration methods has been well documented [14]. Experiences during the first world war had demonstrated the benefits of aerial surveying and, linked with developments in early stereoscopic plotting instruments, it soon became obvious that to achieve higher accuracies in stereo photogrammetric measurements, some knowledge or calibration of the lens system was necessary. The first aerial camera to be calibrated was in Canada in 1920, and the important constants determined were the focal length and location of the principal point. Today, cheap computing and the widespread use of self-calibration using bundle adjustment methods has meant that a high level of performance has become common-place.

Conventional calibration methods [109, 94, 97] determine the intrinsic parameters of the camera from images of a known calibration object such as a Tsai grid, or from properties of the scene such as vanishing points of orthogonal directions.

Auto-calibration methods differ in the sense that a camera can be calibrated directly from an image sequence, despite unknown structure and motion [107, 72, 42]. Such methods use the fact that a camera moves rigidly, so the absolute conic is fixed under the motion.

2.5.3.1 Constraints on the intrinsic parameters

Every worldspace scene plane intersects the plane at infinity π_∞ in a line, and each of these lines intersects the absolute conic Ω_∞ in the circular points $\mathbf{I} = (1, i, 0)^\top$ and $\mathbf{J} = (1, -i, 0)^\top$. Every scene plane thus contains these fixed circular points. Since \mathbf{I} and \mathbf{J} lie on the absolute conic Ω_∞ , the projections of these points \mathbf{I}' and \mathbf{J}' into the image must also lie on the image of the absolute conic ω . Each pose projection of the circular points, corresponding to a different orientation of the scene plane, ensures separate and distinct values for their imaged positions \mathbf{I}'_i and \mathbf{J}'_i . Similarly to \mathbf{I} and \mathbf{J} , the set of imaged circular points are complex, and thus are not physically realizable. Identification of a number of these imaged circular point positions enables us to compute the IAC and consequently the calibration of the camera.

If we can compute the homography transformation \mathbf{H} between the metric reference plane and the corresponding image pose then the imaged circular points are given by $\mathbf{I}' = \mathbf{H}\mathbf{I}$ and $\mathbf{J}' = \mathbf{H}\mathbf{J}$. Constraints on the IAC can then be formed from these transformed points.

$$\mathbf{I}'^\top \omega \mathbf{I}' = 0 \quad (2.53)$$

$$\begin{pmatrix} 1 & i & 0 \end{pmatrix} \mathbf{H}^\top \omega \mathbf{H} \begin{pmatrix} 1 \\ i \\ 0 \end{pmatrix} = 0 \quad (2.54)$$

$$(\mathbf{h}_1^\top + i \mathbf{h}_2^\top) \omega (\mathbf{h}_1 + i \mathbf{h}_2) = 0 \quad (2.55)$$

$$\mathbf{h}_1^\top \omega \mathbf{h}_1 - \mathbf{h}_2^\top \omega \mathbf{h}_2 + 2i \mathbf{h}_1^\top \omega \mathbf{h}_2 = 0 \quad (2.56)$$

where the column vectors of the homography matrix \mathbf{H} are written as \mathbf{h}_j . Equating real and imaginary parts to zero yields the result.

$$\mathbf{h}_1^\top \omega \mathbf{h}_1 - \mathbf{h}_2^\top \omega \mathbf{h}_2 = 0 \quad (2.57)$$

$$\mathbf{h}_1^\top \omega \mathbf{h}_2 = 0 \quad (2.58)$$

Note that exactly the same constraints are formed from the solution of \mathbf{J} , thus each independent circular point fully encodes the two Euclidean axis directions within a single complex entity.

$$\mathbf{I} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + i \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad (2.59)$$

A planar calibration target is used to compute the homography that transforms structure from the metric plane into the image. Non-planar calibration targets can also be used [109, 10, 112, 113], though are often difficult to build with any level of precision. A single planar calibration target is practical, easy to store and cheap to make with great accuracy. Any pattern can be placed onto the target, though a checker board pattern is readily identifiable within an image and provides an easy way to automatically extract the required corner correspondences. Figure 2.15 shows the calibration target used within our laboratory.

Planar based approaches to calibration have been developed both by Sturm and Maybank [97], and Zhang [120] that are based on orthogonality constraints formed from the columns of the world to image plane homography. Liebowitz [60] generalized further by developing the set of linear constraints on the IAC with respect to the circular points of the plane, the details of which are summarized within this section.



FIGURE 2.15: Tsai grid calibration pattern.

On computing the set of homographies, a linear system of the form $\mathbf{Ax} = \mathbf{0}$ can be generated from the constraint equations 2.57 and 2.58, where the elements of \mathbf{x} correspond to the six coefficients of the symmetric matrix ω . This set of homogeneous equations are solved by singular value decomposition, see appendix B.2.3. The camera calibration matrix \mathbf{K} can then be extracted from the computed IAC. Since $\omega = \mathbf{K}^{-\top} \mathbf{K}^{-1}$, we can make the substitution $\mathbf{L} = \mathbf{K}^{-\top}$ such that $\omega = \mathbf{LL}^{\top}$. Subsequently, the matrix ω can be factorized by Cholesky decomposition, and the calibration matrix computed as $\mathbf{K} = \mathbf{L}^{-\top}$.

2.5.3.2 Recovering the extrinsic parameters

The extrinsic pose parameters \mathbf{R}_i and \mathbf{t}_i can then be recovered from the set of computed homography matrices \mathbf{H}_i by applying the inverse of the calibration matrix \mathbf{K} .

$$\mathbf{H} = \mathbf{K} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix} \quad (2.60)$$

$$\tilde{\mathbf{r}}_1 = \lambda \mathbf{K}^{-1} \mathbf{h}_1 \quad (2.61)$$

$$\tilde{\mathbf{r}}_2 = \lambda \mathbf{K}^{-1} \mathbf{h}_2 \quad (2.62)$$

$$\tilde{\mathbf{r}}_3 = \tilde{\mathbf{r}}_1 \times \tilde{\mathbf{r}}_2 \quad (2.63)$$

$$\mathbf{t} = \lambda \mathbf{K}^{-1} \mathbf{h}_3 \quad (2.64)$$

with $\lambda = 1 / \|\mathbf{K}^{-1} \mathbf{h}_1\| = 1 / \|\mathbf{K}^{-1} \mathbf{h}_2\|$. Of course, because of measurement noise the

computed matrix $\tilde{\mathbf{R}} = [\tilde{\mathbf{r}}_1, \tilde{\mathbf{r}}_2, \tilde{\mathbf{r}}_3]$ does not in general satisfy the properties of a rotation matrix. The best rotation matrix \mathbf{R} corresponding to the 3×3 matrix $\tilde{\mathbf{R}}$ is achieved by setting $\mathbf{R} = \mathbf{U}\mathbf{V}^\top$, where $\tilde{\mathbf{R}} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$ is the singular value decomposition of matrix $\tilde{\mathbf{R}}$, see appendix B.2.4 for further details.

2.5.3.3 Maximum likelihood estimation

As a final step, a maximum likelihood estimation is made that minimizes the residual image reprojection error. Given n images of the planar calibration target which contains m points then the maximum likelihood estimate can be obtained by minimizing the functional:

$$\sum_{i=1}^n \sum_{j=1}^m \| \mathbf{m}_{ij} - \hat{\mathbf{m}}(\mathbf{K}_c, \mathbf{K}_r, \mathbf{R}_i, \mathbf{t}_i, \mathbf{M}_j) \|^2 \quad (2.65)$$

where $\hat{\mathbf{m}}(\mathbf{K}_c, \mathbf{K}_r, \mathbf{R}_i, \mathbf{t}_i, \mathbf{M}_j)$ is the projection of the point \mathbf{M}_j in image i . The parameters \mathbf{K}_c correspond to the intrinsic coefficients of the camera and \mathbf{K}_r the coefficients of radial distortion. Each of the rotation matrices \mathbf{R}_i is encoded by a Rodrigues 3-vector \mathbf{v}_i , see the discussion in appendix B.3.1. Correspondingly, the vector \mathbf{P} that parameterises the entire system can be partitioned.

$$\mathbf{P} = (\mathbf{k}_c^\top, \mathbf{k}_r^\top \mid \mathbf{v}_1^\top, \mathbf{t}_1^\top, \dots, \mathbf{v}_n^\top, \mathbf{t}_n^\top)^\top \quad (2.66)$$

The minimization is sparse, since each imaged pose is independent, and may be computed by using a partitioned Levenberg-Marquardt algorithm as described in appendix C.4.

2.5.3.4 Stereo calibration

Two synchronized cameras can be calibrated in much the same way. The only additional parameters that need to be determined are the worldspace pose rotation \mathbf{R}_c and translation \mathbf{t}_c between both cameras. The projection matrices corresponding to both cameras are of the form.

$$\mathbf{P} = \mathbf{K}[\mathcal{I} \mid \mathbf{0}] \quad (2.67)$$

$$\mathbf{P}' = \mathbf{K}'[\mathbf{R}_c \mid \mathbf{t}_c] \quad (2.68)$$

where \mathbf{K} and \mathbf{K}' correspond to the respective camera calibration matrices. In reality these transformation steps include the non-linear projection into the image planes caused

by radial distortion, and are parameterised by both sets of calibration coefficients $\mathbf{K}_c, \mathbf{K}_r$ and $\mathbf{K}'_c, \mathbf{K}'_r$.

The planar calibration target has a two-fold rotational ambiguity corresponding to the position of the origin for unequal numbers of tiles, and a four-fold rotational ambiguity for equal numbers of tiles. Image point correspondences must be labelled consistently within both views in order to determine the pose rotation \mathbf{R}_c and translation \mathbf{t}_c . In order to aid automated extraction, the calibration target pattern needs to be orientable with no rotational symmetry. Correspondingly, the origin can be identified by placing a circular marker, of different colour within one of the corner squares of the calibration grid.

Initially both cameras can be calibrated independently, such that $\mathbf{R}_i, \mathbf{t}_i$ and $\mathbf{R}'_i, \mathbf{t}'_i$ are the set of extrinsic pose parameters of the calibration grid with respect to each camera. The worldspace origin can then be constrained to lie at the first camera centre. Since we know the extrinsic parameters corresponding to the calibration target poses for both cameras, the worldspace camera pose mapping that transfers the first camera to the second can be found by first mapping back the first camera coordinate frame to the model coordinate frame, then from the model space to the second camera coordinate frame.

$$\mathbf{M}_i = \begin{pmatrix} \mathbf{R}'_i & \mathbf{t}'_i \\ \mathbf{0}^\top & 1 \end{pmatrix} \begin{pmatrix} \mathbf{R}_i & \mathbf{t}_i \\ \mathbf{0}^\top & 1 \end{pmatrix}^{-1} \quad (2.69)$$

$$\tilde{\mathbf{R}}_i = \mathbf{R}'_i \mathbf{R}_i^\top \quad (2.70)$$

$$\tilde{\mathbf{t}}_i = -\mathbf{R}'_i \mathbf{R}_i^\top \mathbf{t}_i + \mathbf{t}'_i \quad (2.71)$$

The true extrinsic pose parameters \mathbf{R}_c and \mathbf{t}_c between the first and second cameras are then computed from the set of reconstructed estimates $\tilde{\mathbf{R}}_i$ and $\tilde{\mathbf{t}}_i$.

$$\tilde{\mathbf{R}}_c = \frac{1}{n} \sum_{i=1}^n \tilde{\mathbf{R}}_i \quad (2.72)$$

$$\mathbf{t}_c = \frac{1}{n} \sum_{i=1}^n \tilde{\mathbf{t}}_i \quad (2.73)$$

Due to measurement noise, the computed matrix $\tilde{\mathbf{R}}_c$ does not in general satisfy the properties of a rotation matrix. The best rotation matrix \mathbf{R}_c corresponding to the 3×3 matrix $\tilde{\mathbf{R}}_c$ is achieved by setting $\mathbf{R}_c = \mathbf{U}\mathbf{V}^\top$, where $\tilde{\mathbf{R}}_c = \mathbf{U}\mathbf{D}\mathbf{V}^\top$ is the singular value

decomposition of matrix $\tilde{\mathbf{R}}_c$, as detailed in appendix B.2.4.

The worldspace pose rotation \mathbf{R}_c between cameras is then encoded by a Rodrigues 3-vector \mathbf{v}_c , see appendix B.3.1 for details. Correspondingly, the vector \mathbf{P} that parameterises the entire stereo system can be partitioned.

$$\mathbf{P} = (\mathbf{k}_c^\top, \mathbf{k}_r^\top, \mathbf{k}'_c^\top, \mathbf{k}'_r^\top, \mathbf{v}_c^\top, \mathbf{t}_c^\top \mid \mathbf{v}_1^\top, \mathbf{t}_1^\top, \dots, \mathbf{v}_n^\top, \mathbf{t}_n^\top)^\top \quad (2.74)$$

The minimization is sparse and may be computed by using a partitioned Levenberg-Marquardt algorithm as described in appendix C.4.

2.5.4 Stratified rectification of images of planes

Metric properties of the scene plane are recovered by determining the homography transformation, formed from the known scene plane structure constraints, that maps the imaged vanishing line and circular points back to their canonical values. The process of computing the rectification transformation can be split into three stages. Each stage of the rectification is designed to remove a number of degrees of freedom from the 8 required to compute the planar homography. Constraints are formed in each stage by use of properties and invariants associated with the particular class of projective transformation.

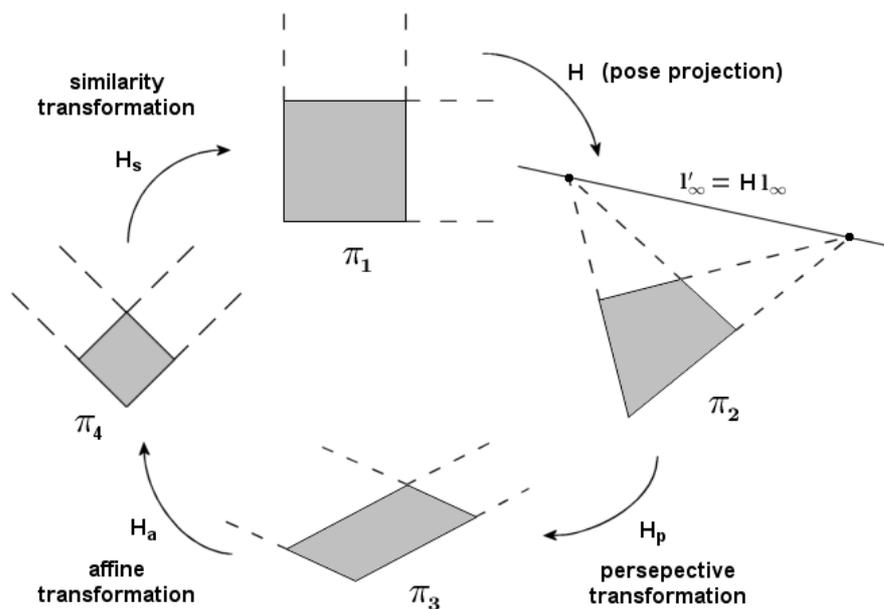


FIGURE 2.16: Stratification process of scene plane rectification.

Figure 2.16 shows the series of transformations, \mathbf{H}_p , \mathbf{H}_a and \mathbf{H}_s , that are applied to the image plane π_2 in order to reconstruct metric structure on the scene plane. Additional

properties and invariants associated with each of these classes of projective transformation are described within appendix A.1.

2.5.4.1 Recovery of affine properties

The vanishing line of the imaged scene plane encodes the perspective component of the projection transformation. The vanishing line is frequently computed from a number of corresponding vanishing points. These in turn are determined from the intersection of imaged parallel worldspace lines. The imaged scene plane π_2 within figure 2.16 shows such a construction of the vanishing line from the sides of an imaged rectangular figure.

Once the imaged line at infinity has been identified, it is then possible to recover the affine properties of the plane. It follows that the vanishing line $\mathbf{l}'_\infty = (l_1, l_2, l_3)^\top$ can be transformed back to its canonical position $\mathbf{l}_\infty = (0, 0, 1)^\top$ by a perspective transformation \mathbf{H}_p of the image plane. A suitable point transformation that restores the affine properties of the scene plane is given by:

$$\mathbf{H}_p = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ l_1 & l_2 & l_3 \end{pmatrix} \quad (2.75)$$

It can be easily verified, by the transformation rule for lines $\mathbf{l}' = \mathbf{H}^{-\top} \mathbf{l}$, that the imaged vanishing line \mathbf{l}'_∞ is mapped back to its canonical position. The transformed plane and subsequently any coplanar worldspace planes are affinely rectified, illustrated by the mapping between planes $\pi_2 \mapsto \pi_3$ within figure 2.16. After rectification parallelism of lines and ratios of lengths along straight line segments are restored, however angles and length ratios along non-collinear lines are not restored.

2.5.4.2 Recovery of metric properties

Having recovered the plane geometry up to an affine transformation by applying the matrix \mathbf{H}_p , which positions the imaged vanishing line back to its canonical position $\mathbf{l}_\infty = (0, 0, 1)^\top$, the final stage is the recovery of metric geometry. This requires an affine transformation of the plane \mathbf{H}_a , that will restore angles and length ratios for non-parallel lines. The set of ideal points on the line at infinity are not fixed pointwise after the application of \mathbf{H}_p . Subsequently, the imaged circular points \mathbf{I}_a and \mathbf{J}_a on the affine plane must then be transformed back to their canonical positions.

$$\mathbf{I}_a = (1, \mu + i\lambda, 0)^\top \quad (2.76)$$

$$\mathbf{J}_a = (1, \mu - i\lambda, 0)^\top \quad (2.77)$$

We must determine μ and λ , the real and imaginary parts of \mathbf{I}_a and \mathbf{J}_a which lie on the ideal line. This fully defines the affine transformation \mathbf{H}_a that restores the metric properties of the rectified plane.

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ & & 1 \end{pmatrix} \begin{pmatrix} 1 \\ \mu + i\lambda \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ i \\ 0 \end{pmatrix} \quad (2.78)$$

$$a_{11} + \mu a_{12} + i\lambda a_{12} = 1 \quad (2.79)$$

$$a_{21} + \mu a_{22} + i\lambda a_{22} = i \quad (2.80)$$

Equating real and imaginary parts, \mathbf{H}_a then has the form.

$$\mathbf{H}_a = \begin{pmatrix} 1 & 0 & 0 \\ -\mu/\lambda & 1/\lambda & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (2.81)$$

Note that there remains some ambiguity on the metric image plane since the circular points cannot be distinguished. Swapping \mathbf{I}_a and \mathbf{J}_a results in a reflection of the metric plane. In fact, there is a possible four-fold reflection ambiguity (flip and mirror) of the reconstructed metric plane.

Work by Liebowitz and Zisserman [62] describes how three types of constraints can be used to find the circular points of the affine plane. The constraints are quadratic in μ and λ and may be represented as circles within the (μ, λ) parameter space. The required solution values that represent the rectification coefficients are then computed by simple circle-circle intersection. These constraints can be combined and may be found from a number of different geometric properties.

- Known angles between lines
- Equality of unknown but equal angles
- Known ratios of lengths

The circular points are invariant to similarity transformations, see appendix A.1.2. Further constraints must be applied by the experimenter to recover structure within a common coordinate system. We may choose to place a known point at the origin, scale a common feature to unit length or align a known vector with one of the coordinate system axes.

2.5.4.3 Known ratios of lengths

One interesting case arises from the known ratios of lengths of line segments on a scene plane [62]. Solution of the constraints formed from these known length ratios between different line segment endpoints can be linearized, in order to compute the affine transformation \mathbf{H}_a that recovers the metric structure of the fronto-parallel view.

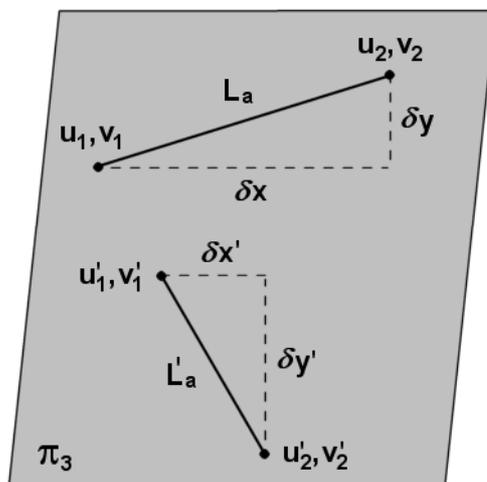


FIGURE 2.17: Length ratios between two non parallel line segments \mathbf{L}_a and \mathbf{L}'_a on the affinely corrected scene plane π_3 .

The squared distance between any two line segment endpoints, shown in figure 2.17, can be written as the inner product $d^2 = \Delta \mathbf{x}^\top \Delta \mathbf{x}$, where $\Delta \mathbf{x} = (u_2 - u_1, v_2 - v_1)^\top$ is the endpoint difference vector. If we know the length ratio s between two non parallel line segments \mathbf{L}_m and \mathbf{L}'_m on the metric scene plane, i.e. $\|\mathbf{L}'_m\| = s\|\mathbf{L}_m\|$, then an affine transformation can be computed that restores the metric properties of the fronto-parallel plane view. Since lengths between line segment end points are invariant to translations, we need only consider the upper-left 2×2 sub-matrix \mathbf{H} of the affine transformation matrix \mathbf{H}_a . Equating the squared distances of both rectified line segments on the metric plane.

$$\Delta \mathbf{x}'^\top \mathbf{H}^\top \mathbf{H} \Delta \mathbf{x}' = s^2 \Delta \mathbf{x}^\top \mathbf{H}^\top \mathbf{H} \Delta \mathbf{x} \quad (2.82)$$

If we write the endpoint difference vectors as $\Delta \mathbf{x} = (\delta x, \delta y)^\top$, $\Delta \mathbf{x}' = (\delta x', \delta y')^\top$ and the elements of the symmetric 2×2 matrix $\mathbf{M} = \mathbf{H}^\top \mathbf{H}$ as $\mathbf{m} = (\mathbf{M}_{11}, \mathbf{M}_{12}, \mathbf{M}_{22})^\top$ then the

set of linear constraints on \mathbf{m} can be written.

$$\begin{bmatrix} (s^2\delta x^2 - \delta x'^2) & 2(s^2\delta x^2\delta y^2 - \delta x'^2\delta y'^2) & (s^2\delta y^2 - \delta y'^2) \end{bmatrix} \mathbf{m} = \mathbf{0} \quad (2.83)$$

Since \mathbf{m} is defined up to scale (2 d.o.f.) then a minimum of two such corresponding pose constraints are required to fully determine \mathbf{m} . We stack all constraints formed from all known length ratio correspondences and solve the system of homogeneous equations of the form $\mathbf{Ax} = \mathbf{0}$ through singular value decomposition, see appendix B.2.3. The rectification matrix \mathbf{H}_a is then formed from the extracted parameters of $\mathbf{H}^\top \mathbf{H}$.

$$\mathbf{H}^\top \mathbf{H} = \xi \begin{pmatrix} 1 + (\mu/\lambda)^2 & -\mu/\lambda^2 \\ -\mu/\lambda^2 & 1/\lambda^2 \end{pmatrix} = \begin{pmatrix} m_1 & m_2 \\ m_2 & m_3 \end{pmatrix} \quad (2.84)$$

$$\mu = -\frac{m_2}{m_3} \quad (2.85)$$

$$\lambda = \sqrt{\frac{m_1}{m_3} - \mu^2} \quad (2.86)$$

After rectification metric properties such as angles and length ratios are restored. However structure is ambiguous up to a similarity transformation of the plane, i.e. rotation, translation and scale.

2.6 Epipolar geometry

Epipolar geometry describes the relationship between two cameras and their images. It is independent of scene structure, and only depends on the internal parameters and the relative pose between both cameras. The fundamental matrix \mathbf{F} is a 3×3 matrix of rank 2 that encapsulates the stereo correspondence geometry between views. Here, we define the epipolar geometry between two views and describe some of the geometric properties of the fundamental matrix.

2.6.1 Point correspondences between views.

Consider the camera correspondence shown in figure 2.18. An image point \mathbf{x} in the first view back-projects to a ray in the worldspace. This ray passes through the first camera centre \mathbf{C} , the image point \mathbf{x} and the corresponding worldspace point \mathbf{X} . The ray is imaged as a line \mathbf{l}' in the second view, hence the mapping $\mathbf{x} \mapsto \mathbf{l}'$ is encoded by the transform $\mathbf{l}' = \mathbf{F}\mathbf{x}$, where \mathbf{F} is a 3×3 matrix known as the fundamental matrix. The

set of lines that are mapped from points in one view to lines in the other are known as epipolar lines.

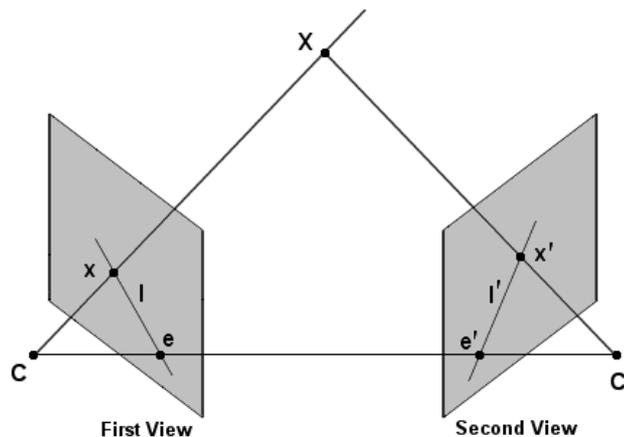


FIGURE 2.18: Epipolar geometry.

The worldspace point \mathbf{X} lies on the back projected ray, hence the image of \mathbf{X} , i.e. \mathbf{x}' , in the second view must also lie on the epipolar line l' . Consequently, the condition $l'^T \mathbf{x}' = 0$ between the images of the back projected ray and the worldspace point in the second image indicates the correspondence of points $\mathbf{x} \leftrightarrow \mathbf{x}'$ between views.

$$\begin{aligned} \mathbf{x}'^T l' &= 0 \\ \mathbf{x}'^T \mathbf{F} \mathbf{x} &= 0 \end{aligned} \quad (2.87)$$

The epipolar correspondence follows from the construction of the plane containing both camera centres \mathbf{C}, \mathbf{C}' and the worldspace point \mathbf{X} . The baseline between the cameras is fixed, as are the images of the cameras (epipoles \mathbf{e} and \mathbf{e}') within each of the other views. The set of correspondence planes that contain both camera centres and the worldspace points form a pencil of planes about the baseline. Consequently, the pencil of epipolar lines formed from the intersection of the correspondence and image planes are all coincident with the respective epipoles. All worldspace points that lie on the baseline identically satisfy the epipolar correspondence, thus providing the null space of the mapping. The fundamental matrix can be factored into the form.

$$\mathbf{F} = [\mathbf{e}']_{\times} \mathbf{K}' \mathbf{R} \mathbf{K}^{-1} \quad (2.88)$$

where \mathbf{K} and \mathbf{K}' are the camera calibration matrices of the first and second cameras respectively and \mathbf{R} is the extrinsic pose rotation between both cameras.

There are a number of important properties of the fundamental matrix which are sum-

marized below.

- If \mathbf{F} is the fundamental matrix of the pair of cameras $(\mathbf{P}, \mathbf{P}')$ then \mathbf{F}^\top is the fundamental matrix of the pair in the opposite order $(\mathbf{P}', \mathbf{P})$.
- For any point \mathbf{x} in the first image, the corresponding epipolar line in the second is $\mathbf{l}' = \mathbf{F}\mathbf{x}$. Similarly, $\mathbf{l} = \mathbf{F}^\top \mathbf{x}'$ represents the epipolar line in the first image corresponding to the point \mathbf{x}' from the second.
- For any point \mathbf{x} other than the epipole, the epipolar line $\mathbf{l}' = \mathbf{F}\mathbf{x}$ contains the epipole \mathbf{e}' , thus \mathbf{e}' satisfies $\mathbf{e}'^\top (\mathbf{F}\mathbf{x}) = (\mathbf{e}'^\top \mathbf{F})\mathbf{x} = 0$ for all \mathbf{x} . It follows that $\mathbf{e}'^\top \mathbf{F} = \mathbf{0}$ is the left null space of \mathbf{F} and similarly $\mathbf{F}\mathbf{e} = \mathbf{0}$ is the right null space.
- \mathbf{F} has seven degrees of freedom: there are nine coefficients but the homogeneous scaling is unimportant, which leaves eight degrees of freedom. \mathbf{F} also satisfies the constraint $\det(\mathbf{F}) = 0$ which removes one more degree of freedom, leaving a total of seven.

2.6.2 Pure translation

Consider the camera correspondence shown in figure 2.19, where both cameras have the same orientation $\mathbf{R} = \mathcal{I}$ and intrinsic parameters $\mathbf{K}' = \mathbf{K}$. The only difference is the translation \mathbf{t} between both cameras. In effect, this is the same situation that occurs when we have a single translating camera. Providing the scene remains static during the time the camera takes to translate to its new position, the two geometries are identical.

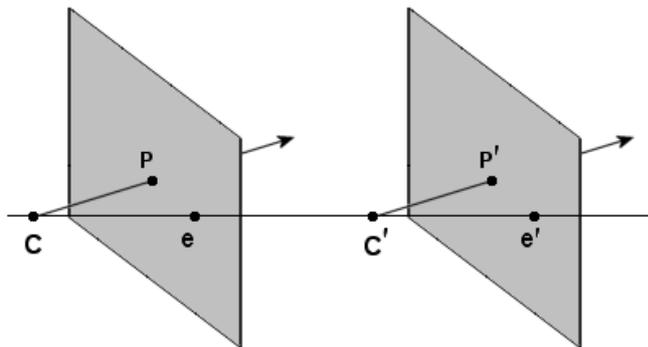


FIGURE 2.19: Auto-epipolar geometry.

Given that the internal camera parameters remain unchanged, with only a pure translation, then the fundamental matrix of equation 2.88 becomes.

$$\begin{aligned}
\mathbf{F} &= [\mathbf{e}']_{\times} \mathbf{K}' \mathbf{R} \mathbf{K}^{-1} \\
&= [\mathbf{e}']_{\times} \mathbf{K} \mathcal{I} \mathbf{K}^{-1} \\
&= [\mathbf{e}']_{\times}
\end{aligned} \tag{2.89}$$

The principal planes of both cameras are coplanar and the positions of the epipoles are coincident such that $\mathbf{e} = \mathbf{e}'$. This geometric construction is known as an auto-epipolar geometry as all epipolar lines form a pencil of lines about both coincident epipolar points.

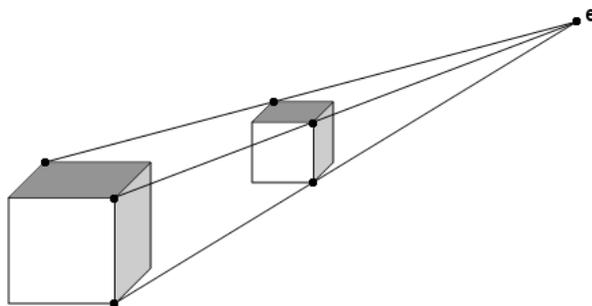


FIGURE 2.20: Imaged pure translational, auto-epipolar motion.

The fundamental matrix $\mathbf{F} = [\mathbf{e}]_{\times}$ corresponding to pure translation is skew-symmetric with 2 degrees of freedom. It follows that \mathbf{e} is the imaged direction of motion, as illustrated in figure 2.20. A dual situation to the translating camera geometry is if we have a stationary camera with scene structure that purely translates. Both geometries are identical except that the motion directions of the object and camera are opposite with time.

2.6.3 Stereopsis transformation

Within this section we discuss a suitable projective transformation $\mathbf{H}_{\mathbf{e}}$ that maps the epipole $\mathbf{e} = (e_1, e_2, e_3)^{\top}$ to the ideal point $(1, 0, 0)^{\top}$. This particular transformation then ensures that all corresponding epipolar lines are aligned parallel with the X axis.

We may choose a mapping that approximates a rigid motion transformation within the neighbourhood of any selected point $\mathbf{x}_0 = (u_0, v_0)^{\top}$. We proceed by first mapping this specified point \mathbf{x}_0 to the origin with the translation \mathbf{T} . Consequently, the epipole is also transformed as $\mathbf{e}' = \mathbf{T} \mathbf{e}$.

$$\begin{pmatrix} 1 & -u_0 \\ & 1 & -v_0 \\ & & 1 \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix} = \begin{pmatrix} e_1 - u_0 e_3 \\ e_2 - v_0 e_3 \\ e_3 \end{pmatrix} \quad (2.90)$$

Next we perform a rotation \mathbf{R} about the Z axis to align the transformed epipole \mathbf{e}' with the X axis.

$$\begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \\ & & 1 \end{pmatrix} \begin{pmatrix} e'_1 \\ e'_2 \\ e'_3 \end{pmatrix} = \begin{pmatrix} a \\ 0 \\ b \end{pmatrix} \quad (2.91)$$

We then need to apply a perspective transformation \mathbf{G} that maps the transformed epipole to the ideal point $(1, 0, 0)^\top$.

$$\begin{pmatrix} 1 & & \\ & 1 & \\ d & & 1 \end{pmatrix} \begin{pmatrix} a \\ 0 \\ b \end{pmatrix} = \begin{pmatrix} a \\ 0 \\ da + b \end{pmatrix} \quad (2.92)$$

where $d = -b/a$. Since the set of ideal points are invariant to translations then the origin point may be mapped back to the specified point \mathbf{x}_0 by the inverse transformation \mathbf{T}^{-1} . The full homography transformation \mathbf{H}_e that places the epipole back to its canonical position $(1, 0, 0)^\top$ is then given by

$$\mathbf{H}_e = \mathbf{T}^{-1} \mathbf{G} \mathbf{T} \quad (2.93)$$

The stereopsis transformation is useful, in that point correspondences can be made to lie on the same scan lines within the transformed images. In general, two such transformations may be chosen in such a way that the two corrected images match up as closely as possible, i.e. disparities between the images are in the X direction only.

2.7 Multiple view geometry

Within this section we describe a simple linear triangulation method. The triangulation method is a direct analogue of the planar DLT. The back projected rays formed through each set of corresponding image points should all meet in a single worldspace point \mathbf{X} . This worldspace point is then projected into each of the camera images as $\mathbf{x} = \mathbf{P}\mathbf{X}$, $\mathbf{x}' = \mathbf{P}'\mathbf{X}$, etc. Figure 2.21 shows the geometry of the linear triangulation process within multiple camera views.

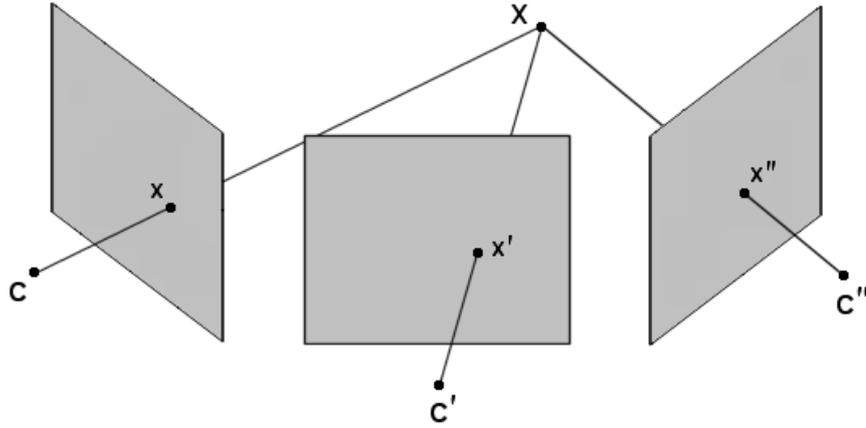


FIGURE 2.21: Triangulation of workspace structure in multiple views by back projection of imaged point correspondences.

Each set of projection equations can be combined into a linear system of the form $\mathbf{A}\mathbf{X} = \mathbf{0}$. We remove the homogenous scale factor by forming the cross product between the measured image and projected workspace points $\mathbf{x} \times \mathbf{P}\mathbf{X} = \mathbf{0}$.

Writing \mathbf{p}^j as the column 4-vector corresponding to the j^{th} row of \mathbf{P} then a projected workspace point can be written.

$$\mathbf{P}\mathbf{X} = \begin{pmatrix} \mathbf{p}^{1\top} \mathbf{X} \\ \mathbf{p}^{2\top} \mathbf{X} \\ \mathbf{p}^{3\top} \mathbf{X} \end{pmatrix} \quad (2.94)$$

If we then write the corresponding measured point in one of the images as $\mathbf{x} = (u, v, w)^\top$ then the equations formed from the cross product can be written.

$$[\mathbf{x}]_{\times} \mathbf{P}\mathbf{X} = \begin{pmatrix} v\mathbf{p}^{3\top} \mathbf{X} - w\mathbf{p}^{2\top} \mathbf{X} \\ w\mathbf{p}^{1\top} \mathbf{X} - u\mathbf{p}^{3\top} \mathbf{X} \\ u\mathbf{p}^{2\top} \mathbf{X} - v\mathbf{p}^{1\top} \mathbf{X} \end{pmatrix} \quad (2.95)$$

Only two of the three equations are linearly independent, thus it is normal to only include the equations from the first two rows within the design matrix \mathbf{A} . We then stack the set of projected point constraints from all corresponding camera views.

$$\begin{bmatrix} v\mathbf{p}^{3\top} - w\mathbf{p}^{2\top} \\ w\mathbf{p}^{1\top} - u\mathbf{p}^{3\top} \\ v'\mathbf{p}'^{3\top} - w'\mathbf{p}'^{2\top} \\ w'\mathbf{p}'^{1\top} - u'\mathbf{p}'^{3\top} \\ \vdots \end{bmatrix} \mathbf{X} = \mathbf{0} \quad (2.96)$$

In general, the lines formed from back projecting two imaged points do not precisely intersect in a single worldspace point. In order for the back projections to meet in a single worldspace point we need to compute the set of optimal point estimates that lie on the epipolar lines between views. There is an analytical solution to this problem within two views that requires solving a polynomial of degree 6, which is described within [40]. However, for more than two views no analytical solution has yet been determined to compute these optimal point correspondences.

2.8 Geometry of gait

We can make the gross assumption that the dynamics of gait are planar, thus a stratified approach can be employed to recover the metric structure of the limb swing plane. This planar motion assumption then allows us to use the properties of planar geometry, as described in the former part of this chapter, in order to identify the set of constraints required to compute the rectification transformations. Within this section, we show that gait has sufficient geometric properties that allows us to recover the true metric structure of the limb swing plane, with no prior knowledge of the subject's motion trajectory or the camera calibration.

2.8.1 Identification of the vanishing line

We must first identify the imaged vanishing line of the limb swing plane in order to compute the perspective transformation \mathbf{H}_p that recovers the affine properties of the motion plane. We further assume that people move along linear trajectories with constant velocity, and that the captured video stream contains at least two periods of subject motion.

Multiple periods of gait motion is then analogous to a single period of motion viewed from multiple cameras that are related by linear translation. Consequently, the geometry of point correspondences at similar positions of gait phase, over the range of motion periods is auto-epipolar. Figure 2.22 illustrates the duality between multiple view geometry and periods of gait motion.

The imaged direction of motion \mathbf{e} corresponds to the coincident set of camera epipoles, through which all epipolar lines must pass. All point correspondences of a single landmark point at similar positions of gait phase must lie on a single epipolar line over the set of gait periods. Figure 2.23 shows a planar test pattern consisting of seven instances of a repeating dot pattern. The pattern repetition is in one motion direction, and represents a single gait pose of repeated subject motion. The distortion effect of central projection ensures that the image of each dot pattern, over the repeated periods (0-6), is

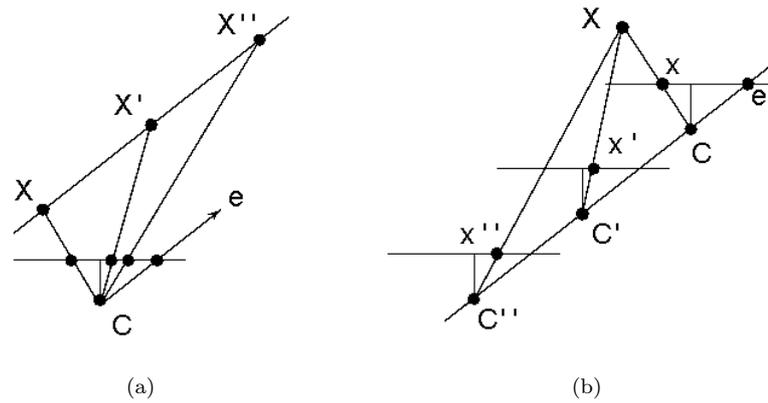


FIGURE 2.22: **Duality of multiple view geometry and periods of gait motion.** (a) Single camera view of multiple points that represent similar pose positions within a number of gait cycles. (b) Multiple cameras related by linear translation that image a pose point within a single gait cycle.

noticeably different over the camera field of view. Similar perspective distortion effects are evident within the repeated spatio-temporal motion structure of human gait.

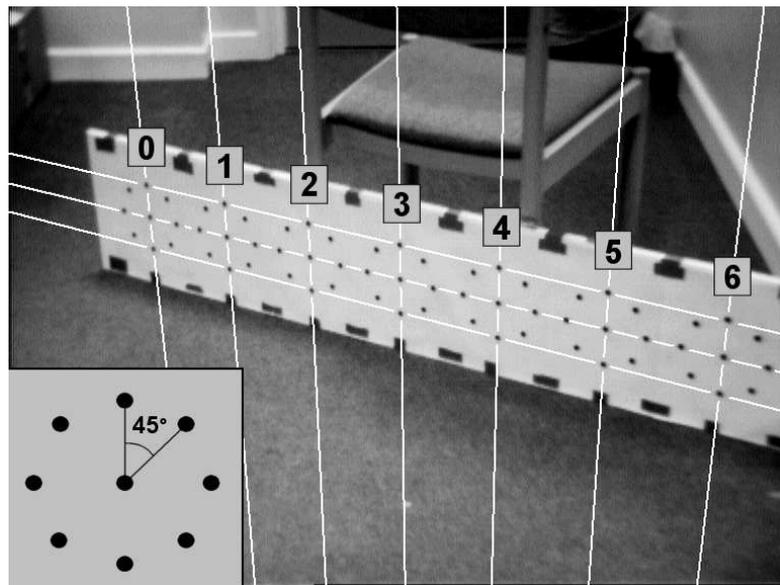


FIGURE 2.23: Repeated planar dot pattern corresponding to a single gait pose of simulated motion over a number of gait cycles.

In order to recover the canonical motion plane, the epipole e must be mapped back to the ideal point $(1, 0, 0)^T$ by applying the stereopsis transformation described within section 2.6.3. After application of this transformation, the set of epipolar lines are all aligned parallel to the X axis. Similarly, the transformed vanishing line of the gait plane is also parallel to the X axis and has the form $l_\infty = (0, l_2, l_3)^T$. Figure 2.24 corresponds to the image of the repeating planar pattern after applying the stereopsis transformation. The set of imaged vertical lines all meet in a single point on the vanishing line. Although

partially rectified, the set of repeated dot patterns are still dissimilar over the camera field of view.

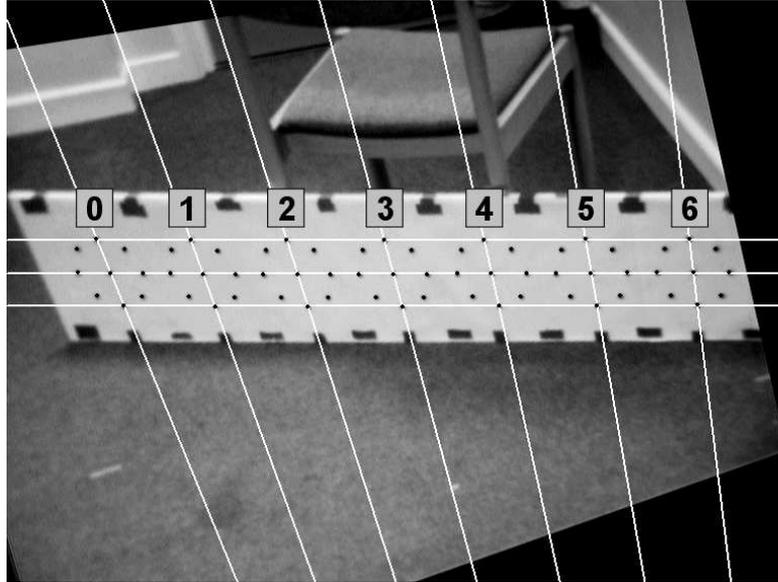


FIGURE 2.24: Rectified image after applying the stereopsis transformation. All epipolar lines corresponding to the direction of motion are aligned parallel with the X axis.

Having replaced the epipole back to its canonical position, the conjugate translation \mathbf{M} , that maps repeated planar patterns within the image, now has a much simpler form. Following the result derived in equation 2.37, the conjugate translation is determined by the imaged direction of motion \mathbf{v} , the vanishing line of the motion plane l'_∞ and a scalar constant λ that describes that apparent change in displacement.

$$\mathbf{M} = \mathcal{I} + \lambda \mathbf{v} l'_\infty{}^\top \quad (2.97)$$

$$\mathbf{M} = \mathcal{I} + \lambda \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & l_2 & l_3 \end{pmatrix} \quad (2.98)$$

$$\mathbf{M} = \begin{pmatrix} a & b & c \\ & a & \\ & & a \end{pmatrix} \quad (2.99)$$

Repeated application of the conjugate translation \mathbf{M} then generates the imaged point correspondences over the set of periods, i.e. $\mathbf{M}^k = \mathbf{M} \cdot \mathbf{M} \cdot (\dots) \cdot \mathbf{M}$.

$$\mathbf{M}^k = \begin{pmatrix} a & kb & kc \\ & a & \\ & & a \end{pmatrix} \quad (2.100)$$

The correspondence equations between the points $\mathbf{x}_k = (u_k, v, 1)^\top$, repeating on the image plane, form constraints on the elements of the transformation matrix \mathbf{M} . The coefficients of the conjugate translation can then be solved by solution of the Direct Linear Transformation.

$$[\mathbf{x}_k]_\times \mathbf{M}^k \mathbf{x}_0 = \mathbf{0} \quad (2.101)$$

Since the conjugate translation has a reduced form, the DLT similarly has a simpler representation.

$$\begin{pmatrix} 0 & -1 & v \\ 1 & 0 & -u_k \\ -v & u_k & 0 \end{pmatrix} \begin{pmatrix} a & kb & kc \\ & a & \\ & & a \end{pmatrix} \begin{pmatrix} u_0 \\ v \\ 1 \end{pmatrix} = \mathbf{0} \quad (2.102)$$

$$\begin{bmatrix} 0 \\ au_0 + kbv + kc - u_k a \\ -v(au_0 + kbv + kc) + vu_k a \end{bmatrix} = \mathbf{0} \quad (2.103)$$

Since we have placed the epipole back to its canonical position, all similar point correspondences have the same y coordinate, thus the first row of the constraint equations shown in 2.103 is identically zero. We also notice that the third row is a scalar multiple of the second, hence each point correspondence only provides a single constraint on the transformation parameters.

$$\begin{bmatrix} (u_0 - u_k) & kv & k \end{bmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \mathbf{0} \quad (2.104)$$

Length ratios on each epipolar line of a corresponding set of repeated points are preserved, thus distances between consecutive points $\Delta u = (u_{k-1} - u_k)$ are equal.

$$u_0 - u_k = k\Delta u \quad (2.105)$$

Equation 2.105 clearly shows that the constraints formed from all combinations of similar

point correspondences with common y coordinates are equal since k is a common scaling factor.

$$k \begin{bmatrix} \Delta u & v & 1 \end{bmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \mathbf{0} \quad (2.106)$$

A minimal solution to the system of homogeneous equations can be derived from two point correspondences with different y coordinates. There is a clear geometric interpretation for the construction of the vanishing line of the plane, illustrated in figure 2.25.

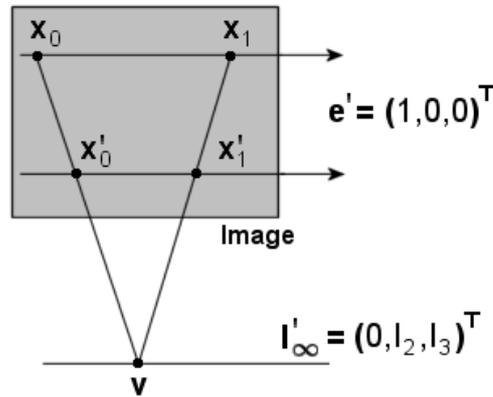


FIGURE 2.25: Construction of the vanishing line from two point correspondences with different y coordinates. The foreshortening of similar worldspace lengths, due to perspective distortion, along epipolar lines parallel with the X axis generates the corresponding vanishing point \mathbf{v} .

Given any two pairs of repeating points $(\mathbf{x}_0, \mathbf{x}_1)$ and $(\mathbf{x}'_0, \mathbf{x}'_1)$ that have a common translation within the worldspace motion plane, the intersection of the lines formed through the equivalent points $\mathbf{l}_0 = \mathbf{x}_0 \times \mathbf{x}'_0$ and $\mathbf{l}_1 = \mathbf{x}_1 \times \mathbf{x}'_1$ meet in a vanishing point \mathbf{v} of the imaged motion plane. This can be explained since lines \mathbf{l}_0 and \mathbf{l}_1 are parallel in the worldspace, thus the intersection generates an ideal point on the plane. Intersection of the images of these lines then identifies the corresponding imaged vanishing point \mathbf{v} . The vanishing line of the gait motion plane must be parallel to the X axis, thus the coefficients of this line are fully determined from the imaged vanishing point \mathbf{v} .

Constraints formed from all combinations of similar point correspondences with common y coordinates are identical. Where we have more than two correspondence periods, a least squares estimate of Δu may be computed. An inhomogeneous solution can be formed to solve the linear displacement $k \cdot \Delta u + u_0 = u_k$ of the set of repeating points on the line. The system of equations of the form $\mathbf{Ax} = \mathbf{b}$ are constructed by stacking the constraints.

$$\begin{bmatrix} k & 1 \end{bmatrix} \begin{pmatrix} \Delta u \\ u_0 \end{pmatrix} = u_k \quad (2.107)$$

One constraint equation of the form 2.108 can then be computed for each set of repeated point correspondences with different y coordinates. A minimum of two such constraint equations are required to solve this set of homogeneous equations.

$$\begin{bmatrix} \Delta u & v & 1 \end{bmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \mathbf{0} \quad (2.108)$$

The homography that maps the vanishing line back to its canonical position $(0, 0, 1)^\top$ is finally given by the perspective transformation \mathbf{H}_p .

$$\mathbf{H}_p = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & b & c \end{pmatrix} \quad (2.109)$$

Consequently, the affine properties of the plane are restored by this transformation. Figure 2.26 shows the image of the reconstructed scene plane after application of the perspective transformation \mathbf{H}_p .

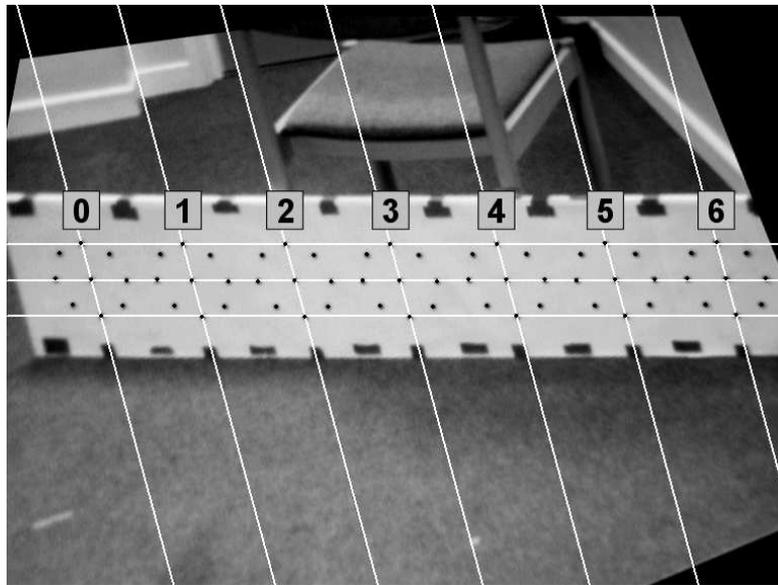


FIGURE 2.26: Affine rectification of the test pattern after mapping the vanishing line of the plane back to its canonical position.

Although metric properties such as angles are not recovered, structure is now similar across all repeated dot patterns. The self-similarity of the affinely reconstructed image structure is then a good cue for periodicity detection.

Other researchers have detected periodic activities using self-similarity measures. Cutler and Davis [20, 21, 22] have put forward an algorithm that consists of two stages. The first facilitates segmentation and object tracking in the foreground. The second then aligns each object along the temporal axis by using the object's centroid. A suitable scaling is also applied that ensures that all resampled objects are similarly sized. The self-similarity of these normalized image regions is periodic, thus time-frequency analysis is employed to detect and characterize this periodicity. The method put forward by Cutler and Davis does not make any assumptions about the nature of motion, only that these resampled image regions are periodically similar in appearance. However, image regions are matched without any attempt to remove the effects of perspective distortion. They report favourable detection of subject periodicities using this method.

With known landmark points, we have shown that perspective distortion can be removed through identification of the imaged vanishing line of the limb swing plane. The real problem is then how best to determine these landmark points, or indeed if they are really necessary in order to determine the perspective transformation that recovers the affine properties of the plane. By assuming that gait motion is piecewise linear with constant velocity, the resulting auto-epipolar motion geometry provides a number of strong constraints on the required scaling factors and repeated image structure that are not exploited within the work of Cutler and Davis. Any departure from this motion geometry signifies a change in the mode of gait, and potentially provides a suitable mechanism for segmenting the imaged motion into linear segments of natural locomotion.

Realistically, auto-correlation of self-similar subject poses provides a large source of pixel correspondences that can be used to simultaneously determine both the periodicity of gait, and the required transformation that restores the affine properties of the motion plane. The required landmark features could then be recovered by spatio-temporal template matching. We understand that this task of segmenting a moving subject from the background, tracking it through the image sequence, then determining the periodicity and corresponding landmark features is a hard problem and a research topic in its own right. Automated segmentation and extraction of these landmark features will not be discussed further within this thesis. Since the main area of interest within this project is concerned with the theoretical and geometric properties of gait, we continue with the understanding that this problem can be solved and proceed by manually marking landmark points within the captured image sequences.

2.8.2 Recovering metric structure

The motion of articulated human limbs is dynamic and dependent on the underlying skeleto-musculature interaction. Human bones are rigid with fixed lengths and provide one source of geometric constraints that remain invariant to the mode of gait motion. Identification of important joint landmark features between rigid bone segments enables

us to compute the affine transformation \mathbf{H}_a that recovers the metric properties of the limb swing plane. Since the size of each individual bone segment remains the same over the image sequence then the known length ratio between poses remains fixed (unity). We can therefore use Liebowitz's known ratio of length constraints described within section 2.5.4.3 to recover the metric properties of the limb swing plane.

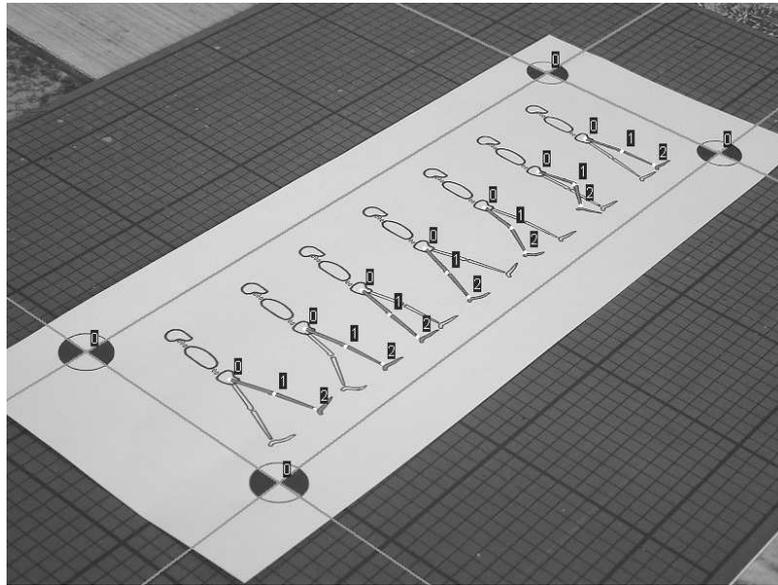


FIGURE 2.27: Synthesized image of a single gait cycle which has been sampled into seven distinct phase poses. The image obtained from the camera represents the motion dynamics of a subject walking at an oblique trajectory angle.

Figure 2.27 represents seven different poses of human gait within a single period of gait motion. The landmark points (hip, knee and ankle) between rigid length bone segments are used to determine the required transformation that recovers the metric motion structure. In the previous section we showed that it is possible to recover the affine properties of the plane by matching corresponding features within similar gait poses over a number of gait periods. A similar arrangement can be made by placing two or more of these printed patterns side by side, to represent multiple periods of gait motion. The detail of all motion periods is hard to see within a single image so we consider only a single planar pattern. We proceed to recover the imaged vanishing line and consequently the homography transformation \mathbf{H}_p that recovers the affine properties of the gait plane from additional markings placed on the test pattern.

Figure 2.28 shows the corresponding affinely rectified image. Length ratios on parallel lines are recovered, though the non isometric scaling and skew between the XY axes is most apparent within the measurement square markings behind the test pattern. The set of length ratio constraints are formed from combinations of each of the matching bone segments within all poses of the gait cycle.

Figure 2.29 shows the corresponding reconstructed structure of the metric plane. Properties of the metric plane are recovered up to a similarity transformation. The circular

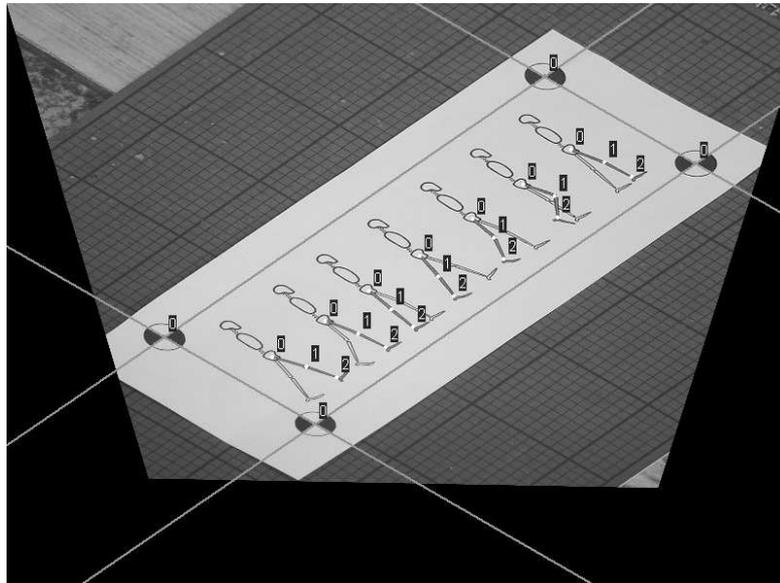


FIGURE 2.28: Affine rectification of the planar test pattern, obtained by mapping the identified vanishing line of the motion plane back to its canonical position.

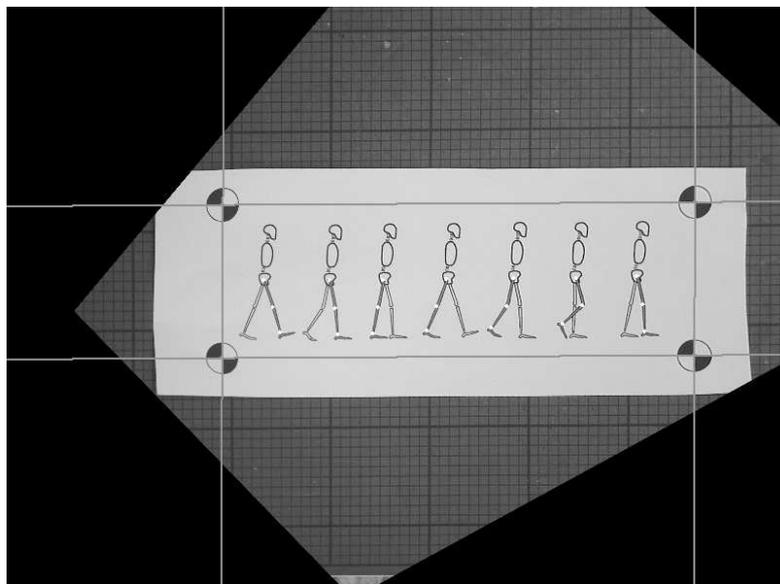


FIGURE 2.29: Metric reconstruction of the test pattern, obtained by applying the affine rectifying transformation that is determined from the consistent rigid bone length constraints within each sample pose of the image sequence.

points \mathbf{I}, \mathbf{J} are invariant to any further similarity transformations thus it is up to the experimenter to choose the alignment and scaling of the required coordinate system. Here, we have chosen the similarity transformation that aligns the motion direction with the X axis, with isotropic scaling that best fits the recovered structure into the image. We can clearly see that the recovered measurement square markings behind the test pattern are indeed square again after the rectification transformation.

2.9 Conclusions

Within this chapter we reviewed the background literature and geometric properties of projective geometry that are essential for understanding the remaining material covered in this thesis. We showed how planar and 3D primitives such as points, lines, planes and conics are represented and how they are transformed under projective imaging.

We gave an overview of the projection process of worldspace 3D structure into the image, and described both the simple linear camera model and more complex cameras with radial lens distortion. We outlined the epipolar geometry and correspondence conditions between points in two views and also considered the specialized auto-epipolar case which arises from a linearly translating camera with fixed internal parameters. We then reviewed the literature on camera calibration and subsequently the back projection and triangulation of worldspace structure.

Worldspace structure commonly occurs within single planes. Subsequently, projection of planar worldspace structure into the image has a simpler form. The transformation of structure from one plane to another is achieved by a 3×3 matrix mapping known as a homography. We reviewed the mathematics and essential properties of planar geometry then gave details of some commonly occurring specialized planar transformations, such as image mosaicing and repetition of planar patterns.

The homography transformation mapping one point set to another can be determined from the known correspondences of points between the two planes by using a direct linear transformation technique. On the other hand, we may not have physical point correspondences between the canonical reference plane and the image plane. We may though have some additional knowledge about the structure on the reference plane that allows us to employ a stratified approach to recover the metric structure.

Geometric properties of the plane can be classified into three main groups of transformation. Further details and properties of these classes of transformation can be found within the discussion in appendix A.1.

- **Perspective transformations.** Parallelism and orthogonality of lines are not preserved. Length ratios on lines are not preserved. The ideal points and the line

at infinity become finite after transformation.

- **Affine transformations.** Parallelism of lines is preserved, although due to skew, orthogonality is not. Ratios of lengths on parallel lines are preserved. The set of ideal points remain ideal, though are not fixed pointwise by the transformation. The line at infinity remains fixed.
- **Similarity transformations.** The circular points \mathbf{I}, \mathbf{J} , angles between lines and ratios of lengths all remain invariant under transformation.

Metric properties of the scene plane are recovered by determining the homography mapping, formed from the known scene plane structure constraints, that maps the imaged vanishing line and circular points back to their canonical positions. Each stage within the stratified rectification process is designed to remove a number of degrees of freedom from the eight required to compute the planar homography.

The set of ideal points all lie on the ideal line, thus identification of the imaged vanishing line of a plane allows us to compute the perspective transformation \mathbf{H}_p that recovers the affine properties of the plane.

The circular points \mathbf{I}, \mathbf{J} of the ideal plane encode the Euclidean coordinate axes within a single complex conjugate entity. Constraints can be formed on the circular points from prior knowledge of the scene structure, and enables us to compute the affine transformation \mathbf{H}_a that recovers the metric properties of the plane.

Since the circular points \mathbf{I}, \mathbf{J} remain invariant to similarity transforms, further constraints that recover structure within a common coordinate system must be explicitly chosen by the experimenter. A similarity transformation \mathbf{H}_s may be chosen to place a known point at the origin, scale a common feature to unit length or align a known vector with one of the coordinate system axes. The combined set of stratified transformations $\mathbf{H} = \mathbf{H}_s \mathbf{H}_a \mathbf{H}_p$ then recovers the required properties of the metric plane.

We hypothesize that articulated limb motion within human gait is approximately planar. Almost all of the perceived limb motion is contained within a single plane. Consequently, gait has sufficient properties that allows us to exploit the structure of planar articulated limb motion in order to recover the fronto-parallel motion dynamics, with no prior knowledge of the camera calibration. As an example, the stratified reconstruction technique was applied to a synthesized image of an obliquely viewed motion figure pattern in order to demonstrate that the canonical fronto-parallel view could be successfully recovered.

Our review of the camera calibration literature showed that a planar calibration target could be used to determine the intrinsic parameters of the camera. Two or more different orientations of the calibration target are sufficient to compute the camera parameters. A similar argument can be made for the motion of gait. Since we can recover the homography that transforms the fronto-parallel reference plane of subject motion to its image

then a minimum of two different imaged motion trajectories are sufficient to determine the intrinsic parameters of the camera. The work presented in this chapter provides the groundwork for what is essentially an “Auto-calibration from gait” algorithm (a close second candidate choice for the title of this thesis).

The later chapters of this thesis are then concerned with determining a suitable model representation of articulated limb motion, identifying the static features of gait that may serve as a useful biometric, and a practical validation of the theoretical auto-calibration method identified within this chapter.

Chapter 3

Static Features of Human Gait

3.1 Introduction

Within the context of human identification, subject gaits can be observed within various situations and from different viewpoints. Viewpoint and environment can be carefully controlled within a laboratory setting, while little or no control is possible for outdoor scenes [12, 54]. Many features are proposed in the literature for gait recognition including optical flow, joint angles, silhouette, etc. We can categorize them as static and dynamic features that evolve in time. Static features reflect instantaneous, geometry-based measurements such as stride length / cadence, limb lengths and height [54, 6, 7]. Dynamic measurements are sensitive to the temporal motion structure of subject activity, such as joint angles, optical flow, symmetry and self similarity [79, 20, 101, 23, 43]. Normal walking conditions (constant and natural walking speed, carrying no objects, level ground plane, etc.) are some of the fundamental assumptions made in most current techniques. Many proposed features and algorithms do not work well if these conditions are violated. Even though gait patterns are repeatable most of the time, changes in walking conditions affect these motion patterns. There are many factors, both physical and psychological, within our daily lives that can influence the variations between our motion patterns such as walking speed, cadence, ground surface, load carrying and state of mind. Understanding the characteristics of gait motion patterns under various walking conditions will help improve the techniques used in further gait research. Here, we are interested in human gaits across different speeds. In particular, we want to understand the patterns of gait motion parameters such as stride length and cadence, which are potentially measurable by computer vision techniques.

Bobick and Johnson [8, 54] develop a gait-recognition method that recovers static body and stride parameters of subjects as they walk. Their technique does not directly analyse the dynamic gait patterns, but uses the action of walking to extract relative body parameters. The set of static body parameters measured are four lengths: the vertical

distance between the head and foot, the distance between the head and pelvis, the distance between the foot and pelvis, and the distance between left and right feet. These distances are measured only at the maximal separation points of the feet during double support phases of the gait cycle. However, they consider step length by itself to be a static gait parameter, while in fact it varies considerably for any one individual over the range of walking speeds. The typical range of variation in step length for adults is about 30 cm [50], which is far from negligible. Their method for estimating the separation distance between feet does not exploit the periodicity of walking, and hence is not robust to tracking and calibration errors.

Davis and Taylor [24] develop an approach for recognizing human walking movements using low level motion regularities and constraints (gait period, stance/swing ratio and double support time). Biomechanical features for classification are automatically extracted from video sequences of walkers. A multiplicative classification rule using statistical distances is then used to determine whether an unknown motion is consistent with normal walking patterns.

BenAbdelkader and Cutler [5, 6] develop a correspondence free method to automatically estimate the spatio-temporal parameters of gait (stride length and cadence) of a walking person from imaged motion. Cadence is estimated using the periodicity of a walking person. Using a calibrated camera system and a known ground plane, the stride length is estimated by first tracking the person and estimating their distance travelled over a period of time. By counting the number of steps and assuming constant velocity during walking they are able to estimate the stride to within 1 cm for a typical outdoor surveillance configuration. They show that stride length and cadence are linearly related over a range of gait speeds. Their approach works with low-resolution images of people, is view-invariant, and robust to changes in lighting, clothing, and tracking errors. It achieves its accuracy by exploiting the nature of human walking, and computing parameters of stride and cadence over many steps.

Further work by Tanawongsuwan and Bobick [103, 102] explores the spatio-temporal gait parameters (stride length and cadence) across a number of controlled walking speeds. They give an in depth study of 15 people, with repeated measurements on different days, for motion obtained from treadmill walking. Their results agree closely with the findings of BenAbdelkader and Cutler.

We understand that the prior work by both BenAbdelkader and Tanawongsuwan has answered many of the relevant questions with regard to the parameters of stride length and cadence. Both their techniques require a calibrated camera and that the ground plane is known, in order to extract the required features for recognition. Our research differs from theirs, since we are interested in observing the dynamic behaviour of the articulated limb angle motion over a range of controlled walking speeds. We propose a suitable motion model that enables us to extract the features of gait without the need

to know the camera calibration, ground plane or subject trajectory. In this chapter, we outline a suitable representation to model the articulated limb motion and give a brief analysis for a small set of trial subjects. In essence, we provide enough information to justify our choice of suitable gait features and biometric representation of motion, that facilitates further development of the view invariant reconstruction methods described within later chapters. Although our set of gait features and motion model function differ from the work of BenAbdelkader and Tanawongsuwan, we show that the results are proportionally similar and correspond well with their findings.

We first give an overview of the terminology and biomechanics of subject motion from the medical literature [50, 84, 15, 77], in order to better understand the nature of gait. We describe the sequence of events within a gait cycle that allows a person to progress forward. Eight individual phases of gait have been identified, each with a different functional objective, that form six distinct motion patterns known as the determinants of gait.

Literature within the context of planar geometry [62] indicates that constraints can be formed from known ratios of lengths within an imaged scene plane in order to determine the transformation that recovers the true metric structure (angles and length ratios). We then hypothesize that articulated limb motion is approximately planar and proceed to verify this assumption by measuring the deviation of 3D joint positions from planarity. We give a brief outline of various motion marker systems that other researchers use in order to capture the dynamics of gait motion, then describe the marker system and experimental set-up used within our laboratory for the purpose of these experiments. The worldspace joint positions are computed by triangulation of imaged point correspondences over a number of camera views. We then give an analysis of the level of pixel reprojection error between the model and corresponding image measurements, caused by the limb swing plane assumption, to determine the validity of the proposed model.

The human skeletal structure is articulated but with fixed length limb segments. These limb segments provide a set of static parameters of gait which remain constant over the entire image sequence. We model the articulated limb motion with a suitable periodic function, hence motion parameterisation is determined from all available data within the image sequence. Techniques that look for specific key frames [8, 54], i.e. positions of maximal foot separation, may be susceptible to lost or occluded frames. The proposed limb motion model is robust to both noise and missing data.

Finally, we analyse the behaviour of the motion parameters over a range of controlled gait speeds for a small trial set of subjects. We emphasize the parameter properties that remain invariant over these speeds and outline a biometric feature vector suitable for recognition purposes. We show similar results to the works of BenAbdelkader and Tanawongsuwan. Finally, a brief discussion on the major sources of error and further

areas of development are addressed.

3.2 Gait Cycle

Walking uses a repetitious sequence of limb motion to move the body forward while simultaneously maintaining stability. As the body moves forward, one limb serves as a source of support while the other progresses to a new support site. The limbs then reverse their roles. For the transfer of body weight from one limb to the other, both feet are in contact with the ground. This series of events is repeated by each limb with reciprocal timing until the intended destination is achieved. A single sequence of events by one limb is called a *gait cycle* [76]. Any event within the sequence can be selected to represent the onset of the gait cycle. The initial moment of contact with the ground is the most readily defined event. People with a normal gait initiate floor contact with their heel (heel strike), though not everyone has this capability, hence the generic term initial contact is used to define the start of the gait cycle [1].

3.2.1 Stance and Swing

Each gait cycle is divided into two periods, stance and swing. *Stance* is used to describe the entire period during which the foot is on the ground and swing the period for which the foot is in the air. Stance is subdivided into three intervals according to the sequence of floor contact for both limbs. Both the start and end of stance involve a period of bilateral foot contact with the ground (*double stance*), while the middle portion has a period of single limb contact (*single limb stance*). The duration of single limb support for one limb equals the swing of the other.

The relative distribution of the periods of gait within a gait cycle is 60% for stance and 40% for swing [76]. The subdivision and distribution of the periods of gait motion for both limbs is more easily seen within figure 3.1. The precise duration of these intervals within the gait cycle varies with a person's walking velocity [2, 82].

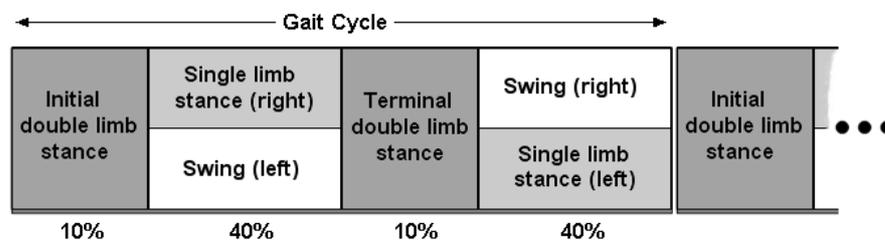


FIGURE 3.1: The subdivisions of stance and their relationship to the bilateral floor contact pattern.

3.2.2 Phases of Gait

In order to provide the basic functions required for walking, each stride involves a series of complex motion patterns performed by the hip, knee and ankle. The term stride is synonymous with a gait cycle. It is based on the actions of one limb, and is defined as the interval between two sequential initial contact positions by the same limb. The stance and swing periods of gait can be further divided into eight functional patterns known as the *phases* of gait. Analysis of a person's walking pattern by phases more directly identifies the functional significance of the different motions occurring at the individual joints. The phases of gait also provide a means for correlating the simultaneous actions of individual joints into patterns of total limb function.

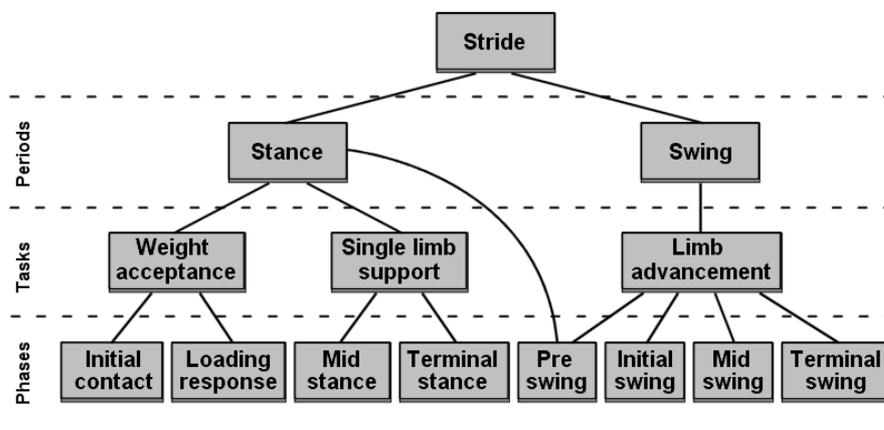


FIGURE 3.2: Divisions of a gait cycle.

Each of the eight phases has a functional objective and enables the limb to accomplish three basic tasks: weight acceptance, single limb support and limb advancement. Figure 3.2 shows the relationship between the periods, functional tasks and phases of gait.

3.2.2.1 Weight acceptance

This is the most demanding task within the gait cycle, as the person abruptly transfers body weight from the stance limb to the other limb that has just finished swinging forward. Two gait phases are involved, initial contact and loading response.

Initial contact. This phase includes the moment when the foot just touches the floor. The joint postures at this time determine the limb's loading response pattern. This phase accounts for the interval between 0 – 2% of the gait cycle.

Loading response. This is the initial double stance period, beginning with initial contact and continuing until the other foot is lifted for swing. The functional objectives

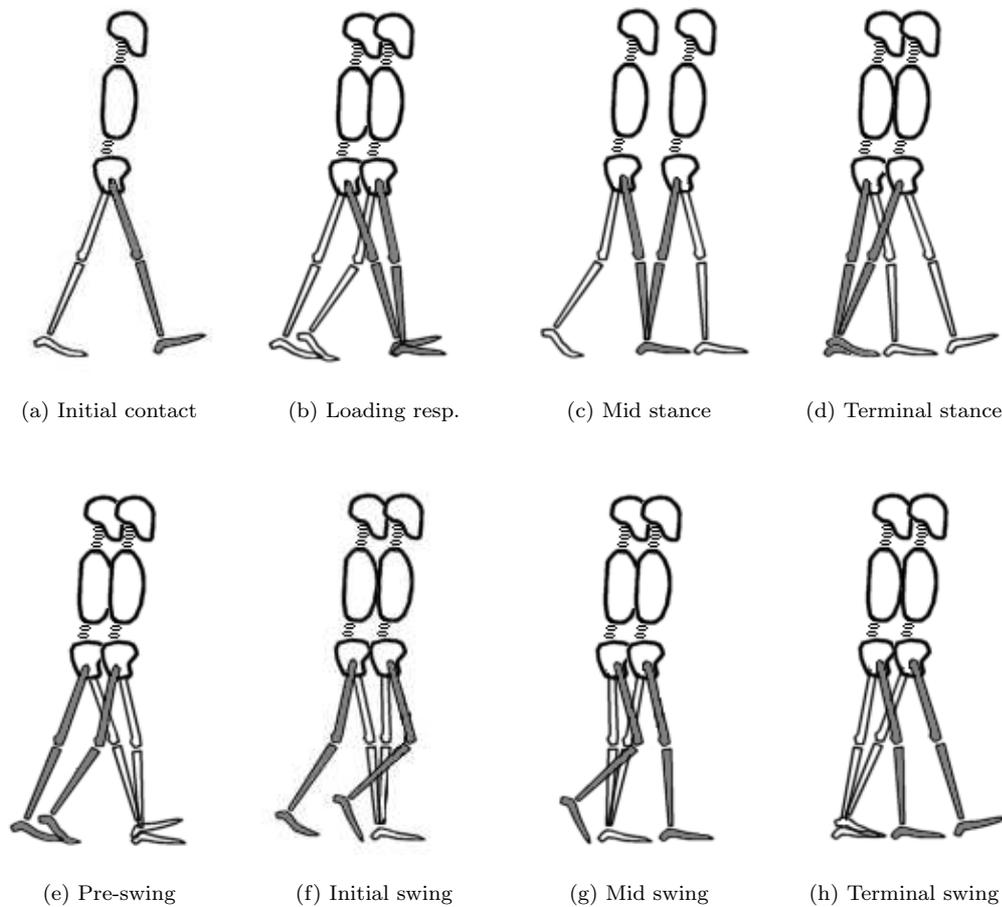


FIGURE 3.3: Limb posture during the eight phases of gait.

of loading response are to provide shock absorption, posture stability and preservation of progression. This phase accounts for the interval between 0 – 10% of the gait cycle.

3.2.2.2 Single limb support

Lifting the other foot for swing begins the single limb support interval of stance. This continues until the opposite foot again contacts the floor. Two gait phases are involved in single limb support, mid stance and terminal stance, and differ primarily by their mechanisms of progression.

Mid stance. This is the first half of the single limb support interval, beginning as the other foot is lifted and continues until the body weight is aligned over the forefoot. The functional objectives of mid stance are to provide progression over the stationary foot and stability of both the limb and trunk. This phase accounts for the interval between 10 – 30% of the gait cycle.

Terminal stance. This is the final phase of single limb support, beginning with heel rise and continuing until the other foot strikes the ground. The functional objective of

terminal stance is to ensure progression of the body beyond the supporting foot. This phase accounts for the interval between 30 – 50% of the gait cycle.

3.2.2.3 Limb advancement

To meet the demands of limb advancement, preparatory posturing begins at the end of stance. The limb then swings through three postures as it lifts, advances and prepares for the next stance interval. Four gait phases are involved in limb advancement: pre-swing (end of stance), initial swing, mid swing and terminal swing.

Pre-swing. This is the final phase of stance and represents the terminal double stance interval within the gait cycle. It begins with the contact of the opposite limb and ends with toe-off of the reference limb. The abrupt transfer of body weight promptly unloads the limb, though this action makes no active contribution to the event. Instead, the unloaded limb uses its freedom to prepare for the actions of swing. This phase accounts for the interval between 50 – 60% of the gait cycle.

Initial swing. This first swing phase accounts for about a third of the total swing period. It begins with lift of the foot from the floor and ends when the swinging foot is opposite the stance foot. The functional objective of the initial swing phase is to advance the trailing limb from its previous stance position. This phase accounts for the interval between 60 – 73% of the gait cycle.

Mid swing. This second phase of swing begins with the swinging limb opposite the other stance limb and ends when the swinging limb is ahead of the truck with the tibia aligned vertically. The functional objective of mid swing is to ensure limb advancement and provide adequate clearance for the foot from the floor. This phase accounts for the interval between 73 – 87% of the gait cycle.

Terminal swing. This final phase of swing begins with a vertical tibia and ends when the foot strikes the floor. The functional objective of terminal swing is to ready the limb for contact with the floor. This phase accounts for the interval between 87 – 100% of the gait cycle.

3.2.3 Energy conservation

Minimizing the amount that the body's centre of gravity is displaced from the line of progression is the main mechanism for reducing the muscular effort of walking. The dependence on reciprocal bipedal locomotion presents two costly situations during each stride. As the left and right limbs alternate their support roles, the body must shift laterally from one side to the other. The limbs also change their vertical alignment between double and single support, thus causing the pelvis and ultimately the body

mass to be elevated up and down. The body is at its lowest during periods of double stance, and at its highest while in mid stance when the supporting limb is vertical. Though a mixture of six motion patterns called the *determinants of gait*, the magnitude of these costly horizontal and vertical displacements is reduced to approximately 2.3 cm in each direction [88]. The resulting motion also avoids abrupt changes in direction and consequently the amount of energy expended on locomotion [50]. Figure 3.4 shows the arc of motion traced by the centre of gravity of the human body over a complete gait cycle.

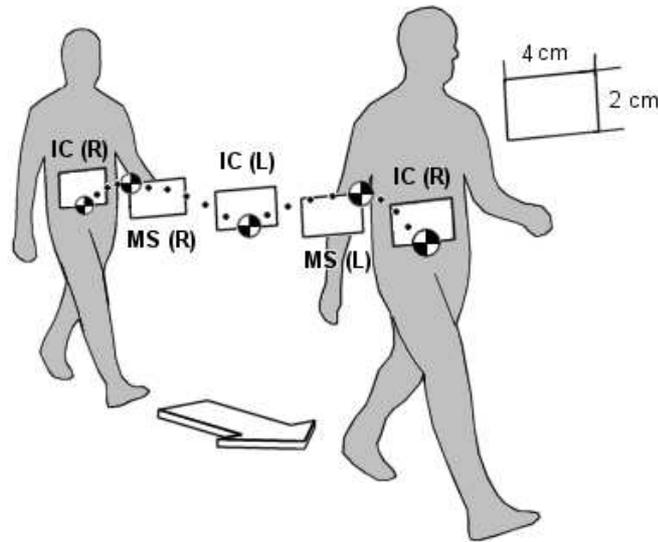


FIGURE 3.4: Motion arc traced by the centre of gravity of the human body over a complete gait cycle.

Limb motions make an active contribution to the smoothing of the vertical displacement path of the body. Vertical lift is lessened by the lateral tilt of the pelvis combined with ankle and knee flexion of the stance limb. Vertical drop is reduced through terminal stance heel rise, initial heel contact combined with knee extension and horizontal rotation of the pelvis. As a result, the body's centre of gravity follows a smooth three dimensional sinusoidal path.

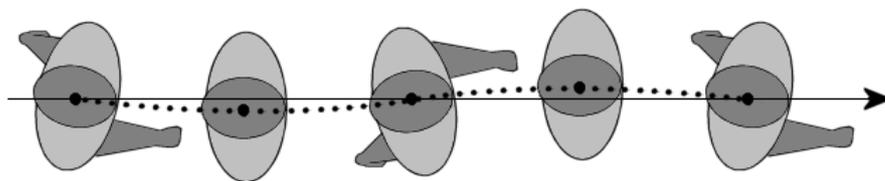


FIGURE 3.5: Lateral displacement.

Figure 3.5 shows the apparent lateral displacement and transverse rotation of the head and body during a single gait cycle. The head and body deviate laterally from the progression midline, defined though the set of head poses in each of the double stance intervals. All connected axial segments of the head, trunk and pelvis deviate by the same

amount. This deviation represents a single sinusoidal motion and averages about 4.6 cm for the total displacement arc between maximum left and right positions [111]. The transverse rotation of the pelvis and trunk averages about 10 degrees in each direction for normal gait motion [76, 88]. Changes in gait velocity alter the segment displacement pattern. Slow walking causes a 30% greater deviation, while the difference is 20% less for fast gait motions [105].

3.3 Motion marker systems

Cameras offer a remote, non-contact means of recording the motion of the entire body. Over the years, computers have become more powerful and analysis software more commercially available. There are now a number of automated motion analysis tools available. Two basic systems are used: video with enhanced passive markers (ViconTM, Peak PerformanceTM, United TechnologiesTM, Motion AnalysisTM) and optoelectrical active marker systems with light emitting diodes (SelspotTM, WhatsmartTM, OptitrackTM).

All motion systems depend on computing the arcs of motion and positions of individual limb joints numerically. The basic technique consists of placing markers on the skin surface in locations that accurately represent the actions of the underlying joints. Joint positions within the worldspace are determined through a process of triangulation of the imaged set of markers from each of the camera views. A minimum of two cameras are required to compute worldspace structure, though for a practical system, three to five cameras are required to reliably capture all the markers.

The basic approach has been to place three markers on each body segment to model the joint locations. There is no guideline for the exact placing of these markers, hence marker placement is dependent on the anatomical interpretation of the individual investigators.

To permit simultaneous measurement of sagittal, coronal and transverse motion of the hip, knee and ankle, multiple surface markers are used. Arcs of transverse rotation are often too small to capture visually with skin markers alone. To overcome this limitation a mid segment stick marker is used that visually amplifies the rotational arc of motion. Figure 3.6(a) shows a typical arrangement of markers required for medical gait analysis. To circumvent the difficulties of inconsistent movement and location of the skin markers, the use of a mid segment cluster of markers has also been introduced. These three dimensionally orientated markers are fixed to a common base plate that is strapped to the centre of the limb segment, illustrated in figure 3.6(b). It is assumed that the influence of skin motion between markers is then reduced.

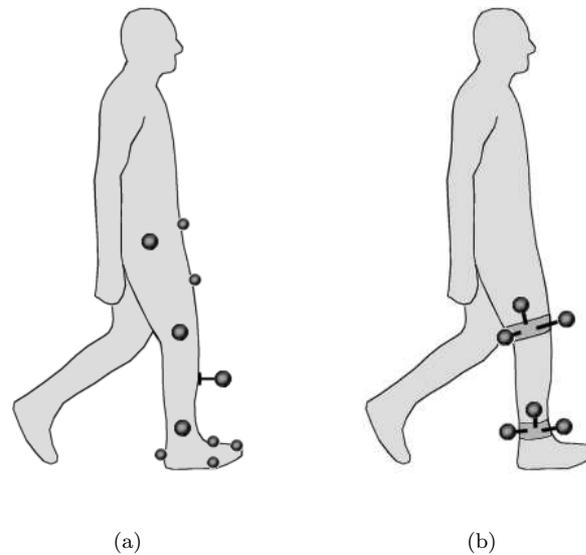


FIGURE 3.6: **Anatomical marker placement.** (a) Typical arrangement of markers required for medical gait analysis. (b) Mid-segment cluster of markers. Three markers on a common base define a unique motion plane.

3.4 Experimental set-up

Since the process of multiple view calibration takes some time, we need to design a practical experimental procedure that minimizes the total number of cameras and the number of required changes between subject capture sessions. A single capture session requires that we obtain motion data corresponding to treadmill and overground walking, with a consistent set of landmark features between both. Since we must acknowledge that people can not be detained indefinitely and the fact that we only have a total of 16 passive markers, all experiments for a single test subject must be run consecutively.

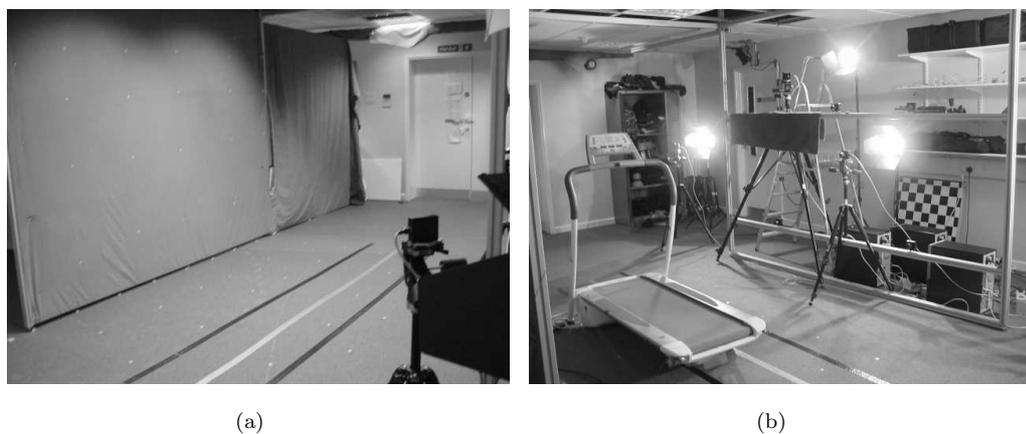


FIGURE 3.7: **Laboratory test track.** (a) View of the test track from behind the left camera. (b) Test track with the treadmill placed fronto-parallel to the central camera.

These experiments must be repeatable, thus a documented experimental procedure must be followed in order to achieve consistent and reliable results. We describe the aims, mathematical details and results of these experiments within the next few sections. A detailed discussion of the instructions required to set up and handle a motion capture session is then given, recipe style within section 3.7.

A common camera set-up is used to capture both treadmill and overground walking motion data for a number of volunteer subjects. Figure 3.7 shows the test track and placement of the treadmill within our laboratory for the purposes of the required experiments. The set of cameras are placed along a line, down one side of the test track. A series of lines have been marked onto the floor in order to provide a visual cue for subjects to follow. Subjects are then instructed to walk along a path 3.5 metres from the line of cameras.

3.4.1 Markers

We are only interested in the gross spatio-temporal motion structure of the main limb joints, in particular the leg motion. We hypothesize that limb motion is planar, thus it is necessary to place the set of markers as close as possible to the required anatomical landmark joint positions. The arrangement of motion markers used within our laboratory is shown in figure 3.8. We employ the use of thirteen similarly sized passive markers that are attached to the head, shoulders, elbows, wrists, hips, knees and ankles of a subject. Each landmark is labelled with a unique number (0-12) that can be used to index the tables of feature tracking positions.



FIGURE 3.8: Marker placement and labelling schema used to identify each individual landmark.

Flat surface markers that are stuck directly onto clothing do not have similar appearances within each camera view. These markers, attached to the surface of a limb, are easily

self occluded by the same limb within oblique views. In contrast, spherical markers are always imaged as circles within each camera view, and are far less easily occluded. The markers we use in our laboratory are 40 mm diameter balls, covered with retro-reflective tape. Light sources (TD beam 800W) placed behind each of the cameras are reflected from the markers and provide bright circular patches of high contrast within each of the images. Automation of marker extraction is very much a secondary concern within this research project, hence the corresponding landmark locations are manually segmented from the digitised video streams.

3.4.2 Cameras

There are many types of suitable video camera that are commercially available. In the past researchers have used analogue camera recording systems to capture image sequences onto tape. These tapes are subsequently digitised into a more useable format required for image processing. More recently, the availability and declining cost of digital camera systems and computer processing power has meant that large arrays of synchronized cameras have become a practical and viable alternative.

The camera system that we currently have within our laboratory consists of 8 Point Grey - Dragonfly digital cameras. These cameras can be synchronized and allow us to stream image data directly to hard disk by using a IEEE-1394 Firewire connection. Due to bandwidth issues required to stream the raw image data to disk (progressive scan CCD, 640×480 resolution : 8-bit colour, Bayer tiled at 30 fps), a maximum of two cameras is connected to any one storage computer. The four required storage clients are then connected to a central server via a Local Area Network (Gigabit network). Synchronization of the cameras between clients is handled by three Point Grey - Sync Unit devices which are chained across the set of IEEE-1394 Firewire busses.

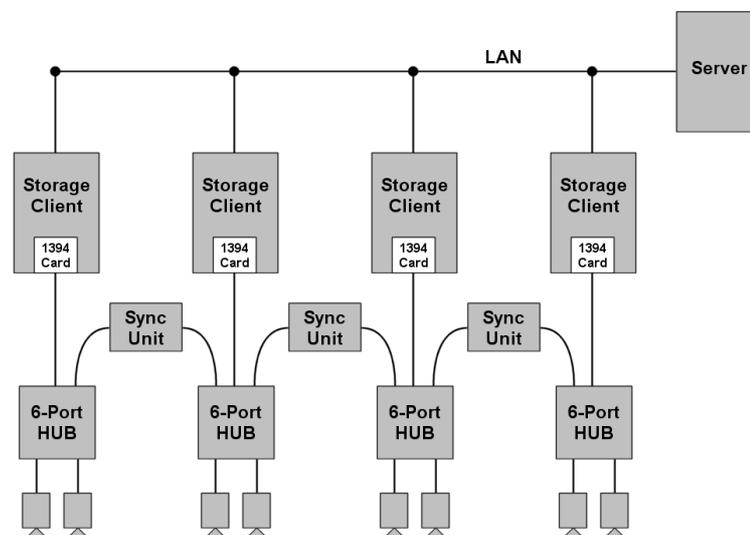


FIGURE 3.9: Camera synchronization connection diagram.

The server handles the image retrieval and requests to start and stop capturing image data from the cameras by using an in-house software system [73]. This software enables us to easily set the essential configuration details, such as shutter speed and colour balance, for each of the cameras in the system. A standard template of configuration parameters can be stored in a single file and may be broadcast to all cameras within the array.

The set-up we proposed for these experiments requires the use of four synchronized digital cameras. We then attached a number of different lenses (Cosmicar/Pentax - 2.8 mm, 4.0 mm and 6.0 mm) to each of the cameras. Figure 3.10 shows the corresponding experimental camera set-up used within our laboratory. All cameras are then accurately calibrated, enabling us to recover the true baseline measurements of worldspace joint positions by triangulation of the imaged landmark features in different views. Since the positions of these cameras are fixed throughout the entire set of capture sessions, we only need to perform the calibration process once.



(a) Camera array

(b) Dragonfly digital camera

FIGURE 3.10: **Laboratory camera set-up used for motion capture.** (a) Placement of cameras within the array. (b) Close up of a digital camera and rear light source.

Three of the cameras are fixed at the far left, middle and far right of the test track and are separated by a distance of at least three metres. Accuracy of the reconstructed worldspace structure is determined largely by the precision of the individual camera calibrations. Calibration of the cameras is typically achieved within a two stage process. First, each camera is individually calibrated using Zhang's method [120], over a large number of image correspondences. The set of imaged calibration target point features can be orientated and distributed differently over the entire field of view for each camera, enabling us to determine an accurate representation for the camera intrinsic parameters (linear projection and radial distortion coefficients). The set of intrinsic parameters are then fixed within further calculations to compute the extrinsic pose parameters between cameras. Since the baselines are quite wide between cameras, correspondences of similar structure appear very different and often within localized regions of the individual

images. Finding correspondences over the entire fields of view, within subsets or all cameras, is often impossible. Fixing the camera intrinsic parameters obviates some of the errors associated with matching the correspondences within these local image regions. The extrinsic pose parameters between cameras are then solved pairwise between cameras in the array. The corresponding back projection of worldspace structure is valid to within an uncertainty of ± 2 mm.

A fourth ‘Sky’ camera is placed on a fully extended mounting tripod and allowed to rotate on its spindle. The motion of the changing camera orientation is controlled by a human operator standing nearby on a step ladder. The camera operator is instructed to keep the walking subject as close as possible within the centre of the view by means of best guess line of sight. Although the locus of rotation is arbitrary, all of the camera motion is approximately about a single point.

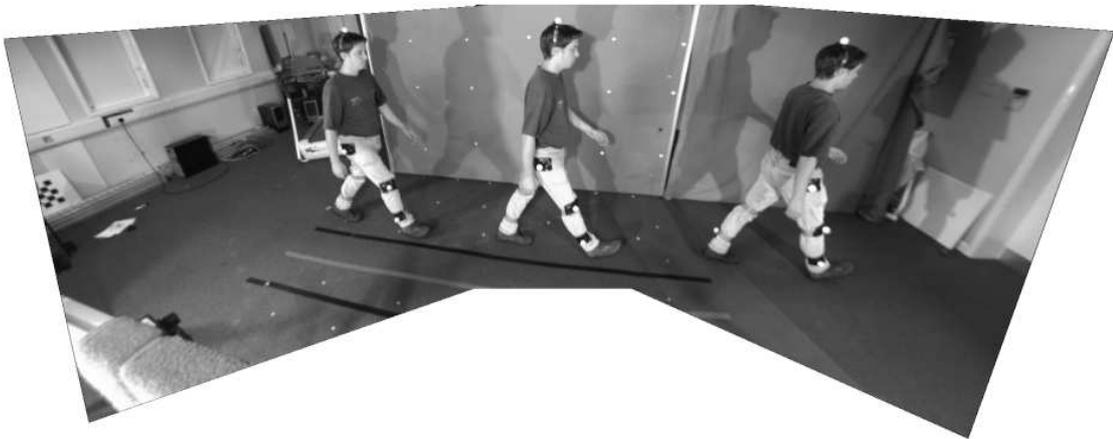


FIGURE 3.11: Panorama composed of three images obtained from a camera that is allowed to freely rotate about its origin.

The geometry is specialized for a camera that has constant intrinsic parameters but is allowed to rotate freely about its origin point. Static worldspace structure is related between images by a homography mapping, see section 2.4.2 for further details. We can recover the composition of all images within a single chosen reference frame by determining the planar mapping between these point correspondences. Figure 3.11 shows the panorama obtained through reconstruction of a set of three images within the motion sequence, with respect to the second of these images.

For a practical system, we need to determine these static image correspondences within the initial step of segmenting background and foreground objects. Here, the correspondences have been manually marked from small retro-reflective patches located on both the wall and the floor. We can clearly see that such a situation is similar to the single view case, only that the panorama generates a single but much larger image from the composition of the set of smaller views.

<i>Camera Position</i>	<i>Camera Model</i>	<i>Resolution (pixels)</i>	<i>Frame rate (fps)</i>	<i>Shutter speed (secs)</i>	<i>Lens (mm)</i>
Left	Dragonfly	640 × 480	30	1/250	2.8
Middle*	Dragonfly	640 × 480	30	1/250	2.8
Right	Dragonfly	640 × 480	30	1/250	4.0
Sky	Dragonfly	640 × 480	30	1/250	6.0

TABLE 3.1: Camera information within the experimental set-up. All cameras are static and fixed with the exception of the ‘Sky’ camera which is allowed to rotate on its tripod.

Table 3.1 shows the camera and lens information corresponding to the experimental set-up used to capture the image sequences of subject gait motion. The middle camera, denoted with the asterisk in the table, is used to record the treadmill motion data. All four cameras are used to capture the imaged motion sequences of subject over-ground walking.

3.5 Planar limb swing assumption

One of fundamental assumptions that the majority of researchers use is that human subjects can be modelled as cardboard figures and that all limb motion lies within a single vertical plane [79, 20, 101, 23, 43]. In reality, the limbs move within small arcs of motion that is economical with the amount of energy required to preserve the progression of motion. It is common, within the context of single view gait recognition, to ignore the six major motion patterns (determinants of gait) and proceed with the cardboard person assumption. We can better model articulated limb motion by assuming that each limb swings within a separate plane. Corresponding left and right limbs are bilaterally symmetric about the mid plane. This partly resolves the spatial configuration of body parts, though separate limbs are still modelled with planar motion. We describe here the effects of applying this planar motion assumption and give a quantitative estimate of the size of error induced through image reprojection.

We first assume that a person is walking fronto-parallel to the camera, i.e. the z coordinate is modelled as being constant throughout the trajectory path. In reality, the true depth of the limb point deviates from this constant value depending on the phase of the gait cycle. Over a suitably large human population, there will be some average error deviation Δz from the planar limb motion assumption. We can then investigate the image reprojection error caused by this average deviation from the planar motion assumption. Figure 3.12 shows the perspective projection of a single assumed planar motion point $(x, z)^\top$ and the corresponding putative error deviate point $(x, z + \Delta z)^\top$ imaged with a camera of focal length f .

These worldspace points are projected to image points u, u' by the pin hole camera model. The absolute image error $\epsilon = |u - u'|$ between the projected points gives us a

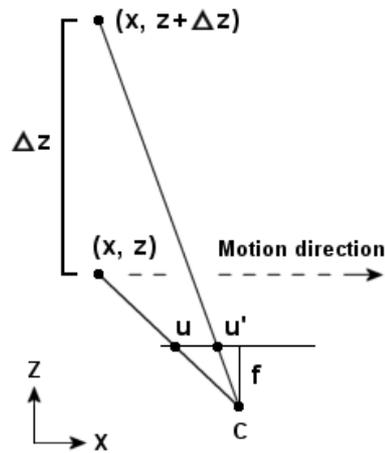


FIGURE 3.12: Image reprojection error of planar point u and error deviate point u' .

quantitative estimate of the contribution of error, caused by the deviation from the planar motion assumption. We can then give some indication of the goodness of fitting associated with the planar motion model.

$$\frac{u}{f} = \frac{x}{z} \quad \text{and} \quad \frac{u'}{f} = \frac{x}{z + \Delta z} \quad (3.1)$$

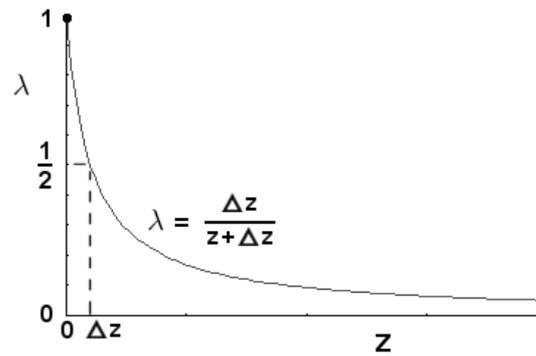
$$\epsilon = |u - u'| \quad (3.2)$$

$$\epsilon = \left| \frac{\Delta z \cdot fx}{z(z + \Delta z)} \right| = \left| \frac{\Delta z}{(z + \Delta z)} \cdot u \right| \quad (3.3)$$

$$\epsilon = |\lambda(z) \cdot u| \quad \text{where} \quad \lambda(z) = \frac{\Delta z}{(z + \Delta z)} \quad (3.4)$$

Equation 3.4 shows that the absolute projection error ϵ is dependent on the imaged position of the point u from the optical centre and the projection function $\lambda(z)$. This function $\lambda(z)$ is dependent on two quantities: the distance z of the subject from the camera origin and the observable deviation factor Δz of a subject's limb swing from planarity. The distance of the subject from the camera is variable but the deviation factor Δz will have some measurable statistical value throughout the population. This measurable deviation may though be influenced by sex, age, weight and pathology.

Assuming that Δz is constant then the projection error function $\lambda(z)$ is proportional to a $1/z$ curve plot. Figure 3.13 shows this relationship with varying distance from the camera. We can clearly see that there is a sharp initial drop off of error value that then tails slowly away. The question we must then ask is: What value of distance z from the camera gives an acceptably small value of absolute reprojection error ϵ , such that it is comparable to the experimental noise tolerance? With so many circumstantial factors that may influence the calculation, i.e. camera calibration and subject position, there

FIGURE 3.13: Projection error function $\lambda(z)$.

may not be a single optimal estimate. We may only quantify the level of error pertaining to individual experiments, thus can only give an empirical estimate of the contribution of error caused by the planar motion assumption.

3.5.1 Overground walking

We need to compute the apparent deviation Δz of the set of articulated leg joint positions from the putative limb swing plane \mathbf{V} . First, the experimental set-up is put in place with the synchronized four camera system described in section 3.3. The cameras are calibrated so that we know the intrinsic parameters of all cameras and the extrinsic parameters describing their positions in space, with respect to the first camera. This then allows us to back project any corresponding set of imaged points in order to compute the best worldspace point \mathbf{X} that minimizes the reprojection error within all camera views.

A small group of four volunteer subjects, shown in figure 3.14, who are willing to endure the experience of having reflective markers taped to them, were told to walk along the path of the test track through the field of view of all cameras.



FIGURE 3.14: The usual subjects: Image of the experimental volunteer group.

For each subject, the set of imaged joint markers in all views corresponding to the closest leg were back projected to form the workspace point tracks \mathbf{X}_i . We first fit the limb swing plane \mathbf{V} to the set of point tracks \mathbf{X}_i , by solution of the set of homogeneous linear equations $\mathbf{X}_i^T \mathbf{V} = \mathbf{0}$. Figure 3.15(a) shows a typical frame pose of back projected workspace points from the set of cameras. The workspace viewer allows us to use a virtual camera in order to construct a novel view that shows the positions of both cameras and subject landmark points. Figure 3.15(b) shows a virtual view of the merged set of workspace point tracks \mathbf{X}_i , corresponding to the selected leg joint positions over the complete image sequence. The view is oblique, and is designed to show the planarity of the set of point tracks. The fitted limb plane \mathbf{V} is shown by the projection of the shaded rectangular region.

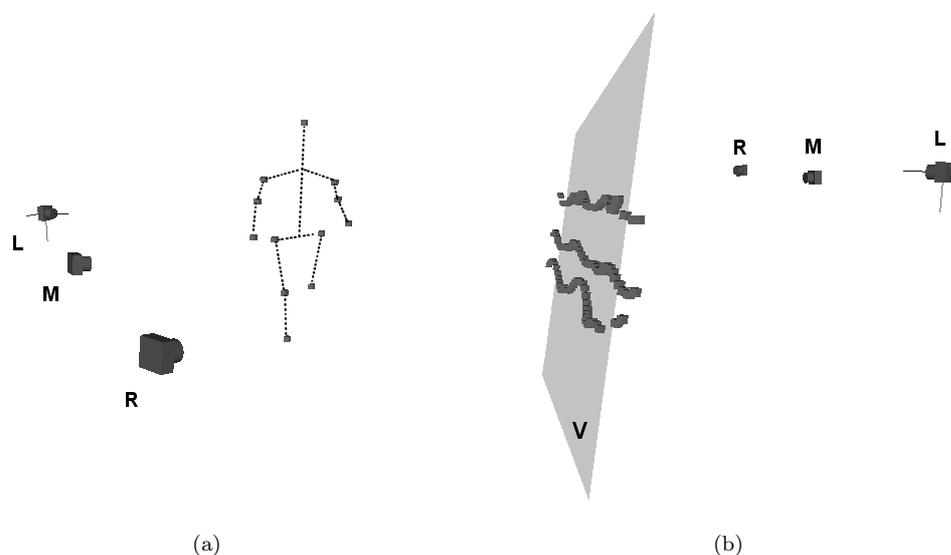


FIGURE 3.15: **Novel virtual camera view of the reconstructed workspace motion structure.** (a) A typical frame pose of a subject's joint positions. (b) A view of the merged joint tracks \mathbf{X}_i and the corresponding fitted limb swing plane \mathbf{V} .

The corresponding set of reprojected point tracks within the image views are shown in figure 3.16. We can clearly see the cycloidal nature of human gait within the fronto-parallel view.

After computing the limb swing plane $\mathbf{V} = (v_1, v_2, v_3, v_4)^T$, we can determine the Euclidean distance d between inhomogeneous representations of the plane and point tracks $\mathbf{X} = (x, y, z, 1)^T$.

$$d = \frac{v_1 x + v_2 y + v_3 z + v_4}{\sqrt{v_1^2 + v_2^2 + v_3^2}} \quad (3.5)$$

We can compute the mean and standard deviation of the set of absolute distance measures $|d_i|$, over all point tracks \mathbf{X}_i from the fitted limb plane. The 3σ standard deviation

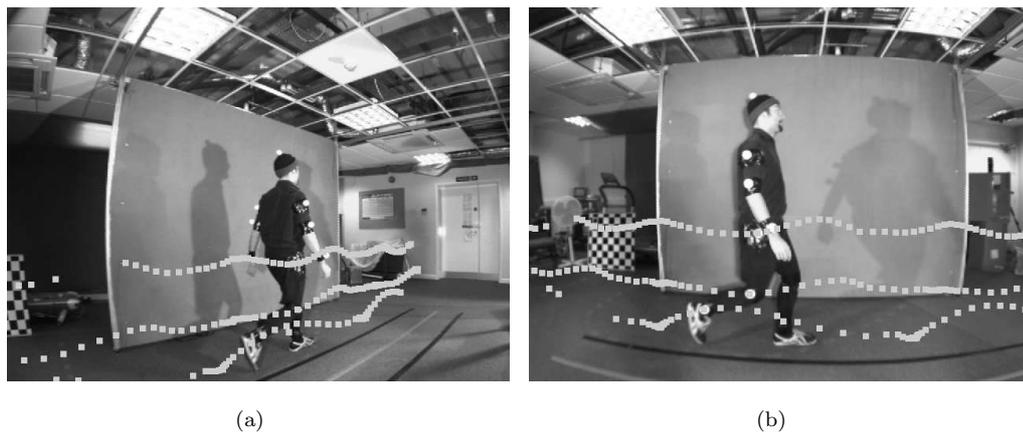


FIGURE 3.16: Set of reprojected joint motion tracks.

then gives a good indication for the size of the error deviation Δz from the planar limb motion assumption. Table 3.2 shows the fitting information for all four test subjects, along with the average deviation estimates over all subjects.

<i>Subject</i>	<i>min</i>	<i>max</i>	<i>mean</i>	3σ
00	0.0078	54.09	12.17	31.01
01	0.035	68.65	15.26	38.25
02	0.024	88.44	16.21	43.08
03	0.059	40.85	10.04	26.72
<i>All</i>	0.0078	88.44	13.42	34.77

TABLE 3.2: Absolute error deviation, in millimetres, of the point tracks \mathbf{X}_i from the fitted limb swing plane \mathbf{V} .

The estimated error Δz corresponds to a 34.77 mm deviation of the limb joints from the planar motion assumption. The mean distance of subject motion from the fronto-parallel camera view is calculated as 3212.3 mm, hence the putative reprojection error factor $\lambda(z)$ from equation 3.4 is estimated as 0.010708. Assuming an image size of 640 pixels then the worst pixel reprojection error, seen at the extents of the image, corresponds to a projected deviation Δu of approximately 3.4 pixels. The average pixel reprojection error will be less, though still provides a sizeable contribution of systematic error, unaccounted for by the model, to the reconstruction process.

We can make a better estimate of the probable projected deviation error by assuming that most imaged motion occurs within the 68.3 percentile (1σ) of the total image error. The reprojection error function is linear, $\epsilon = |\lambda(z) \cdot u|$ with fixed $\lambda(z)$ coefficient over the visible range $[-w : w]$, as shown in figure 3.17.

The area under the graph represents the total error over the field of view. The 68.3% area is shown by both shaded regions in the graph. Note that the field of view and consequently the error function are symmetric about the optical centre. We seek to find

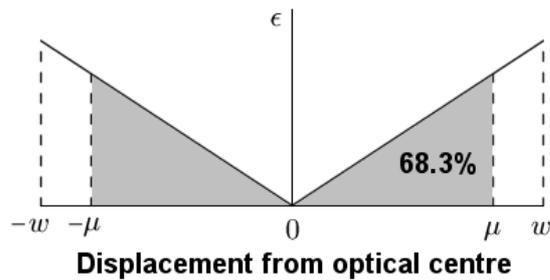


FIGURE 3.17: Standard pixel reprojection error.

the value $u = \mu$ that corresponds to the $\sigma = 68.3$ percentile boundary as a suitable measure of the standard pixel reprojection error. Equating the required areas under the graph.

$$\sigma \cdot \lambda(z) \cdot w^2 = \lambda(z) \cdot \mu^2 \quad (3.6)$$

$$\mu = \sqrt{\sigma} \cdot w \quad (3.7)$$

The standard pixel reprojection error corresponding to the imaged subject motion at a distance of 3212.3 mm from the camera is then computed at 2.8 pixels.

Most CCTV camera systems are positioned to capture human gait motion at far greater distances, of at least 10 metres. A similar argument for this distance generates a maximum pixel reprojection error of the order 1.1 pixels at the extents of the image. The corresponding standard pixel error is then approximately 0.9 pixels, and consequently is within the experimental landmark measurement error ± 1 pixel. The further the person is from the camera the smaller they appear in the image. To ensure that people at a considerable distance appear suitably sized within the image, the focal length of the camera must be increased. As a result the field of view is narrowed, such that a smaller percentage of the total spatial gait motion is visible. One would then assume that there is some form of optimal configuration that satisfies both the spatial and planarity requirements.

3.6 Treadmill experiments

The action of natural gait serves to progress the person forward with constant velocity. A simple and effective way to study the dynamics of gait is to remove the progressional component of locomotion by placing the person on a treadmill. This gives us the ability to accurately control the speed of gait and capture the subject motion with a fixed camera. Many medical studies in biomechanics and human movement use a treadmill to observe the characteristics of human locomotion [25, 70, 59, 47, 83]. In [59] a treadmill

is used to control both walking and running motions, for inter-subject comparison at similar speed and stride frequencies. The symmetry of gait is studied using a treadmill based system in [25]. Researchers have identified subtle differences between treadmill motion and overground walking on a level surface. [115] reports small differences between treadmill and overground walking but observes no effect on the symmetry. [95] finds significant differences in sagittal knee joint motion between treadmill and overground walking when subjects are given between 1-2 minutes to familiarise themselves on the treadmill. [70] also reports that subjects should be given at least 6 minutes to familiarise themselves with treadmill walking in order to obtain limb joint kinematics and spatio-temporal gait measurements similar to overground walking.

In a study of adults between 20 to 60 years of age, who were unaware that they were being observed, the mean walking speed for males averaged 4.92 km/h while the average for females was 4.44 km/h [32]. The measured slow and fast walking speeds in adults between 20 to 59 years of age averaged 2.22 km/h and 5.94 km/h respectively [110]. The functional range of customary walking speeds in adults ranges from approximately 2.2 km/h up to 6 km/h. At speeds above 6 km/h there is then a choice between walking or running. A study of adult males found that the transition speed between walking and running averaged 6.8 km/h [106]. Research into the expenditure of energy from locomotion has shown that running becomes more efficient than walking at speeds above 7.98 km/h [26]. Several works have shown that increases in velocity are normally achieved by increasing both cadence and stride length [58, 56].

We describe here a suitable periodic function that models the planar limb swing motion of gait. For each test subject, we analyse the behaviour of the corresponding motion parameters over a number of controlled walking speeds within the customary range 3 - 6 km/h. Any periodic function can be used to encode the limb angle function $\theta(t)$, though the obvious choice is to use a Fourier series representation.

$$\theta(t) = a_0 + \sum_{k=1}^n a_k \cos(2\pi k f_0 t + \phi_k) \quad (3.8)$$

The question we must first ask is: How many Fourier harmonics are required to sufficiently model the dynamics of gait? Researchers have previously suggested [3] that the maximum frequency content of human walking is ~ 5 Hz and that the fundamental frequency of normal gait is ~ 1 Hz. This suggests that $n = 5$ Fourier harmonics is sufficient to model the limb angle function. We validate this assumption by giving a quantitative analysis of the image reprojection error from the gait motion function over a range of Fourier harmonic coefficients. We then determine a suitable number of harmonics n that enables good reconstruction.

3.6.1 Limb angle function

We have previously shown that the motion of a single articulated limb is approximately planar. A good initial estimate for this limb swing plane can be computed by aligning a calibration grid pattern with the subject's leg on the treadmill, while in the quiet standing posture. Figure 3.18 shows the configuration of a calibration grid aligned with a subject's leg plane.



FIGURE 3.18: Vertical reference plane of the subject's leg.

The projection of metric points \mathbf{u} from the canonical reference plane to the imaged leg swing plane can be computed by the planar homography mapping as $\hat{\mathbf{x}} = \mathbf{H}\mathbf{u}$. The correspondence between known grid points \mathbf{u} on the metric plane and the set of imaged points \mathbf{x} allows us to compute the elements of the homography matrix \mathbf{H} by solution of the Direct Linear Transformation $[\mathbf{x}]_{\times}\mathbf{H}\mathbf{u} = \mathbf{0}$. These equations can be rearranged into the form $\mathbf{A} \cdot \mathbf{h} = \mathbf{0}$, where \mathbf{h} is the flattened set of homography matrix coefficients. Further details of the Direct Linear Transformation can be found in section 2.4.6.

The putative set of limb joint positions \mathbf{w} on the metric plane are first recovered by applying the inverse mapping $\mathbf{w} = \mathbf{H}^{-1}\mathbf{x}$ to the set of imaged marker points \mathbf{x} . The set of articulated leg segments are rigid and have fixed lengths over all frames in the sequence. We can accumulate the vector \mathbf{d} , of mean upper and lower leg segment lengths over the image sequence, by computing the distances between recovered joint marker endpoints. Since we are only interested in limb length ratios, we compute the normalization transformation \mathbf{K}_n , that maps the centroid of all joint marker positions \mathbf{w} over the sequence to the origin, with isotropic scaling such that the first limb segment has unit length. The set of metric plane points are then transformed as $\mathbf{w}' = \mathbf{K}_n\mathbf{w}$. Consequently, the updated homography that maps these normalized points into the image is given by $\mathbf{H}' = \mathbf{H} \cdot \mathbf{K}_n^{-1}$. Angles between joint marker endpoints are invariant to the similarity transformation \mathbf{K}_n , hence the transformation serves as a convenient way to normalize the data into a consistent format over the set of test subjects.

We first need to find an initial estimate of the fundamental frequency of gait f_0 . A cost function, that is dependent on the self-similarity of limb direction vectors between putative values of gait period, is then a good indicator of the subject's periodicity. The dot product \mathcal{C} between limb angle unit vectors is maximal (unity) when the two vectors are similar. The cost $1 - \mathcal{C}$ is then minimal between similar positions of pose. A vector of root mean squared self-similarity costs is then accumulated over a suitable range of putative periodicities. We know the camera frame rate accurately and can use the 1 Hz estimate of natural gait to determine this sensible range of putative periodicities. To eliminate any false local minima caused by measurement noise within the cost vector, we first apply a (1, 4, 6, 4, 1) smoothing filter to the vector of self-similarity cost errors. We then assume that there is a single minima within the chosen range of putative periodicities that represents the true period of subject gait. The true (discrete) periodicity estimate is found by performing gradient descent on the vector of self-similarity cost errors from a 1 Hz initial periodicity estimate. A sub-temporal estimate of the period T is then found by fitting a quadratic curve to the data about this computed discrete periodicity estimate. A more detailed discussion on periodicity detection is given later within chapter 4, as part of the gait reconstruction algorithm.

The set of valid limb angles and their corresponding time sample vectors are computed for each of the normalized leg segments. The Fourier series representation of each limb angle function is then determined, with fixed fundamental frequency $f_0 = 1/T$. The set of minimized coefficients for each limb segment are then stored in a biometric reconstruction vector \mathbf{V}_i , where the coefficients of \mathbf{V} have the form:

$$\mathbf{V} = (a_0, a_1, \phi_1, \dots, a_n, \phi_n)^\top \quad (3.9)$$

With the knowledge of the normalized leg lengths \mathbf{D} , we can find by back substitution the best set of hip points \mathbf{X}_0 consistent with the limb angle functions.

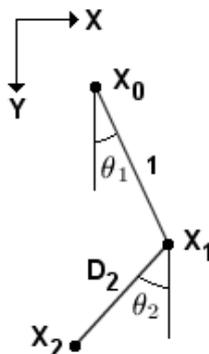


FIGURE 3.19: Articulated limb segment model. The hip point \mathbf{X}_0 is defined by a set of Cartesian $(x, y)^\top$ coordinates. The remaining articulated limb endpoints are defined by a connected set of polar coordinates $(d, \theta)^\top$. The first limb segment length is canonically normalized to unit length.

Figure 3.19 shows the model of articulated limb connections. The hip point \mathbf{X}_0 is given by the Cartesian coordinates $(x, y)^\top$ while the remaining limb endpoints are defined by the connected set of polar coordinates $(d, \theta)^\top$, where the first limb segment has been normalized to unit length. The Cartesian coordinates $(x_i, y_i)^\top$ of any limb point with index i in the model is then given by the equation:

$$(x_i, y_i)^\top = \begin{cases} (u, v)^\top & i = 0 \\ (u, v)^\top + \sum_{j=1}^i \mathbf{D}_j \cdot (\sin \theta_j, \cos \theta_j)^\top & i \geq 1 \end{cases} \quad (3.10)$$

where the pose angles θ_j are given by evaluating the Fourier series function $\theta(t)$ at the current pose frame with the biometric coefficients \mathbf{V}_j and fundamental frequency $f_0 = 1/T$.

Given any endpoint in the articulated limb set, we can compute the putative position of the hip point $(u, v)^\top$ by back substitution. Since a limb segment endpoint is computed relative to its predecessor, measurement fitting errors will be compounded within the back substitution process. A weighted putative hip point $(u'_i, v'_i, w'_i)^\top$, where w'_i is the associated weighting factor, is computed from each of the valid metric plane data points $(\tilde{x}_i, \tilde{y}_i)^\top$ of the articulated leg pose. The putative hip point, from any indexed metric plane data point $i = 0 \cdots m$ within the current leg pose, is given by the set of equations.

$$(u'_i, v'_i, w'_i)^\top = \begin{cases} (m+1) \cdot (\tilde{x}_i, \tilde{y}_i, 1)^\top & i = 0 \\ (m+1-i) \cdot \left[(\tilde{x}_i, \tilde{y}_i, 1)^\top - \sum_{j=1}^i \mathbf{D}_j \cdot (\sin \theta_j, \cos \theta_j, 0)^\top \right] & i \geq 1 \end{cases} \quad (3.11)$$

Where m is the total number of segments within the articulated leg model, i.e. two for a model of upper and lower legs. The fitted hip point $(u, v)^\top$ is then given by the summation of all valid weighted points.

$$(u, v)^\top = \left(\frac{\sum_{i=0}^m u'_i}{\sum_{i=0}^m w'_i}, \frac{\sum_{i=0}^m v'_i}{\sum_{i=0}^m w'_i} \right)^\top \quad (3.12)$$

We only require a minimum of one metric plane data point within a leg pose to compute the associated hip point. This resolves the problem that arises when the swinging arm occludes the hip point marker. The set of computed hip points \mathbf{X}_i on the metric plane encode the instantaneous positions of hip motion.

3.6.2 Maximum likelihood estimation

We have computed an initial estimate of the set of parameters that model the articulated leg motion over the image sequence. As a final step, we optimize the motion parameters \mathbf{P} in order that we minimize image reprojection error. The parameter vector \mathbf{P} can be partitioned into two sections. The first contains the set of coefficients common over the entire sequence $(\mathbf{h}, \mathbf{D}, f_0, \mathbf{V})$. The second contains the subsidiary set of instantaneous hip positions $\mathbf{X}_i = (u_i, v_i)^\top$ corresponding to each of the individual frames.

$$\mathbf{P} = \left(\mathbf{h}^\top, \mathbf{D}^\top, f_0, \mathbf{V}^\top \mid \mathbf{X}_1^\top, \dots, \mathbf{X}_N^\top \right)^\top \quad (3.13)$$

Where \mathbf{h} is the vector of homography coefficients that map points on the metric plane into the image, \mathbf{D} is the vector of normalized limb lengths, f_0 is the fundamental frequency of gait and \mathbf{V} contains the sets of articulated limb segment Fourier coefficients.

The set of parameters \mathbf{P} are optimized by performing the Levenberg-Marquardt minimization method. The form of the Jacobian is sparse and consequently the minimization procedure is similar to that described within appendix C.4.

3.6.3 Reconstruction error analysis

Each of the four test subjects are recorded walking at different speeds on the treadmill (Tunturi - J6). The set of captured sequences are then manually marked and the reconstruction process performed for values of Fourier harmonics $n = 1 \dots 10$. The root mean squared reprojection error computed for each reconstruction gives us an indication of the goodness of fitting between the motion model and the imaged subject motion.

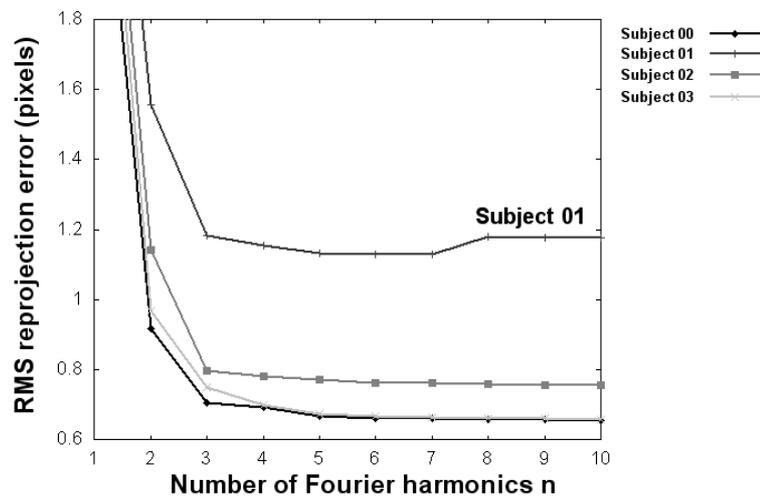


FIGURE 3.20: Root mean squared reprojection error generated from the gait reconstruction function with a varying number of Fourier harmonics n .

Figure 3.20 shows the means of these reprojection errors for the four test subjects, over the entire range of walking speed motion sequences. Three of the reprojection error curves are quite similar. The reprojection error starts quite high $\simeq 2.8$ pixels, by modelling the motion as simple harmonic, i.e. $n = 1$. This quickly falls below the ± 1 pixel measurement error deviation, then slowly trails off to a level of $\simeq 0.7$ pixels. The reprojection errors corresponding to subject 01 level out at a value above the ± 1 pixel measurement error, suggesting that here the model inadequately represents the dynamics of gait motion. There are a number of reasons why this level of error should differ from the others.

Care has been taken to ensure that all manually marked joint positions lie as close to the true positions as possible, to within a ± 1 pixel tolerance. An analysis of the rigid articulated leg segment lengths over the range of walking speeds shows that the variation in limb lengths for subject 01 is much higher than in the others, illustrated in figure 3.21. This suggests that the markers may have been improperly placed for subject 01.

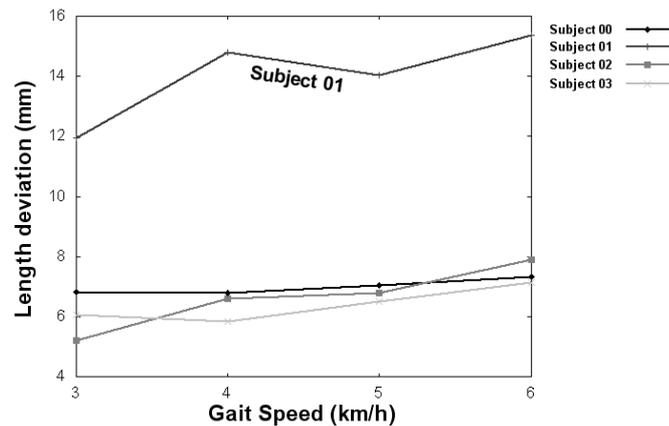


FIGURE 3.21: Analysis of the rigid articulated leg segment lengths over the range of controlled walking speeds.

Figure 3.22 shows the set of subjects in similar double stance postures. In this pose, the joint marker positions on the reference limb should be almost aligned in a single straight line. Natural gait alignment appears slightly flexed in this pose. Figures 3.22(a), 3.22(c) and 3.22(d) correspond well with this alignment while figure 3.22(b), corresponding to the pose of subject 01, shows a high degree of misalignment.

Accurate placement of markers is often quite difficult, especially over clothing. Markers are usually attached to a subject while they adopt a quiet standing posture. Since subjects remain still during the attachment procedure, markers taped over clothing retain their position. Clothing can tend to slip around during periods of locomotion in order to better fit with the body's shape and motion, thus changing the positions of markers that were originally well placed. Experimenters often have difficulty attaching markers to certain positions of the body. Large markers within the hip area may be knocked by the swinging action of the arms. Baggy clothing can also allow the markers

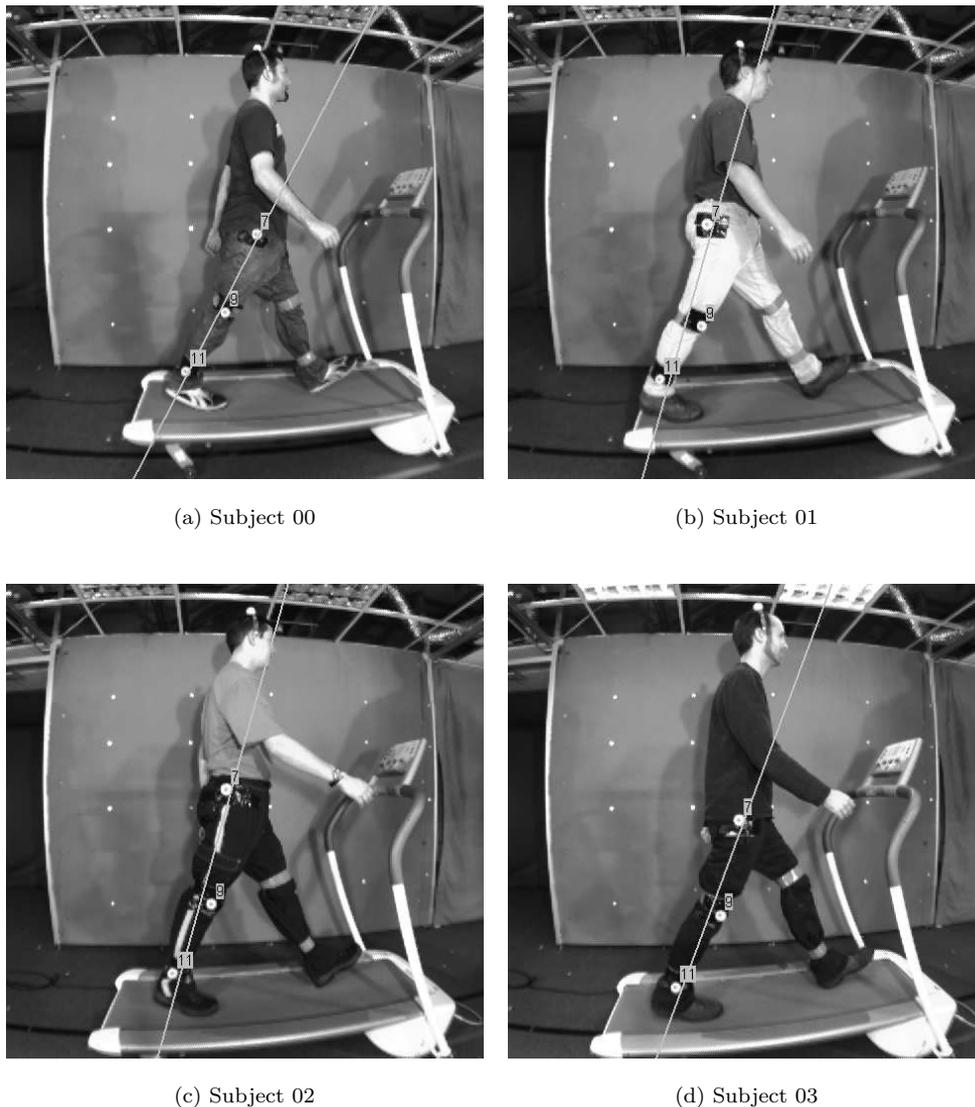


FIGURE 3.22: Alignment of markers in similar double stance postures. Natural gait joint alignment should appear slightly flexed from the line of fit. This fitting line is computed via orthogonal regression through the set of joint markers. The alignment of subject 01 is significantly different from the norm, indicating that marker placement is poor.

to drift during motion. Experimenters must ensure that markers are taped firmly to each of the limb segments. On the other hand, skin and internal muscle structures need to be able to move in order to facilitate locomotion. Taping markers too firmly to joints causes stiffness within the limbs, and consequently subjects complain of unnatural walking motions.

Improper placement of markers also affects reconstruction. Placing markers over actual joint regions of the body is difficult since these areas undergo the largest changes in deformation. Figure 3.23 shows the geometric effect of placing a marker, in error, a small distance from the true joint position.

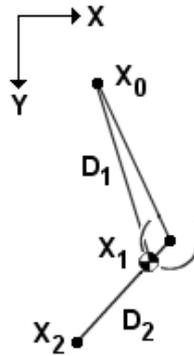


FIGURE 3.23: Effect of placing a marker a small distance away from the true joint position on the locus of motion.

The locus of motion of the marked knee point \mathbf{X}_1 contains components of motion from both the upper and lower leg segments. Consequently, the lower limb length \mathbf{D}_2 remains fixed while the upper limb length \mathbf{D}_1 varies with the phase of gait.

Table 3.3 shows the resulting root mean squared reprojection errors corresponding to the number of Fourier harmonics n used to model the limb motion. The total summation cost $\sum \mathbf{r}$, of the r.m.s. reprojection errors over all subjects and walking speeds, gives an indication of the ability to represent the dynamics of motion by using the required number of Fourier harmonics. The corresponding mean experimental r.m.s. pixel reprojection error $\epsilon = \sum \mathbf{r}/N$, where N is the total number of experiments, then gives an estimate of the level of pixel fitting error within any experiment.

n	$\sum \mathbf{r}$	ϵ	$\Delta \mathbf{r}$	$\Delta \epsilon$	%error
1	33.4084	2.784	-	-	-
2	12.1011	1.0084	21.3073	1.778	63.7783
3	8.9985	0.7499	3.1026	0.2586	9.2868
4	8.6889	0.7241	0.3096	0.0258	0.9267
5	8.4443	0.7037	0.2446	0.020383	0.7321
6	8.362	0.6968	0.0823	6.8583×10^{-3}	0.2465
7	8.334	0.6945	0.028	2.3333×10^{-3}	0.0838
8	8.2992	0.6916	0.0348	2.9×10^{-3}	0.1042
9	8.2861	0.6905	0.0131	1.0917×10^{-3}	0.0391
10	8.2732	0.6894	0.0129	1.075×10^{-3}	0.0386

TABLE 3.3: Resulting root mean squared reprojection errors corresponding to the number of Fourier harmonics n used to model the limb motion. Reprojection errors are computed over all walking speeds and from all valid subjects in the experiments (Subjects 00, 02 and 03).

The reduction in r.m.s. fitting errors $\Delta \mathbf{r} = \sum \mathbf{r}(n) - \sum \mathbf{r}(n-1)$, caused by increasing the number of harmonics used to represent the dynamics of gait, is also shown within table 3.3. The mean experimental error reduction $\Delta \epsilon = \Delta \mathbf{r}/N$ then gives an indication of the reduction in pixel error within an experiment, caused by increasing the number

of Fourier harmonics. The percentage error reduction estimates $100 \times \Delta \mathbf{r}(n) / \sum \mathbf{r}(1)$ then correspond to the level of error reduction as a percentage of the simple harmonic experimental reconstruction error $\sum \mathbf{r}(1)$.

We can clearly see that the quoted values of $n = 5$ represent a suitable choice for the number of Fourier harmonics required to model the limb motion. Here, the percentage error reduction estimate is less than 1% and any further increase in the number of harmonics $n \geq 6$ used to represent the dynamics of motion have root mean squared pixel error reduction levels $\Delta \epsilon$ of the order 1×10^{-3} .

3.6.4 Analysis of the limb function

It is easy enough to represent the dynamics of gait over a small time period by fitting a Fourier series to the limb angles. We assume an arbitrary, constant subject velocity over each experimental sample period. There are then a number of important questions that arise, which may preclude the use of gait as a suitable biometric for identification at a distance.

- How can we compare two reconstructed limb angle functions from a single subject, taken at two different times?
- Though we may be able to compare two similar gaits, is it possible to match gait motion across a range of different speeds?
- Is there an underlying biometric motion function that is unique to the individual, which we can subtly alter through a number of parameter modifiers in order to generate the range of possible gait motions?
- Clearly gait is not a simple one to one function, since the range of gait speeds can be achieved by varying both cadence and stride length. How many factors, both bio-mechanical and psychological, influence the total pattern of gait motion?

Let the Fourier series representation of the limb angle function $\theta(t)$ be defined by the equation:

$$\theta(t) = a_0 + \sum_{k=1}^n a_k \cos(2\pi k f_0 t + \phi_k) \quad (3.14)$$

A time shifted signal $\theta(t - t_s)$ only updates the coefficients of phase within the Fourier series representation.

$$\theta(t - t_s) = a_0 + \sum_{k=1}^n a_k \cos(2\pi k f_0 \cdot (t - t_s) + \phi_k) \quad (3.15)$$

$$\theta(t - t_s) = a_0 + \sum_{k=1}^n a_k \cos(2\pi k f_0 t + \psi_k) \quad (3.16)$$

$$\psi_k = \phi_k - 2\pi k f_0 t_s \quad (3.17)$$

It is unclear which features of gait represent the origin pose position within a gait cycle. It is then natural to align the Fourier signals by computing the time shift that zeros the first coefficient of phase $\psi_1 = 0$, i.e. $t_s = \phi_1/2\pi f_0$. The normalized set of phase coefficients ψ_k can then be written:

$$\psi_k = \phi_k - k\phi_1 \quad (3.18)$$

The normalized set of coefficients ψ_k are then restricted to lie within the range $(-\pi : \pi)$ by finding the suitable corresponding set of angles. Transformation of the captured motion dynamics into a consistent coordinate framework then allows us to compare two representations of limb motion taken at different times. Other researchers use prominent features of gait itself, such as the heel strike with the floor, to determine the onset of the gait cycle. This entails looking for key frames within the gait sequence to determine the required position. The disadvantage of these methods is that they are often prone to the effects of noise and occlusion. The zero phase alignment method uses a simple property of the Fourier series to determine the start of the gait cycle.

A change in walking speed affects the set of computed Fourier coefficients, corresponding to the limb angle function $\theta(t)$ of a test subject. Most notably, the cadence of gait is proportionally similar to the fundamental frequency term f_0 of the Fourier function. As subjects increase their speed, the rate of reciprocal foot contact with the floor also increases. Subjects also lengthen their stride to increase their speed. An increase in stride length denotes an increase in limb swing amplitude, which is proportional to the first harmonic Fourier amplitude term a_1 . These two factors, rate of foot contact with the floor ($\sim f_0$) and subject stride length ($\sim a_1$), can be simultaneously altered to achieve the desired progressional velocity.

We examine the relationship between controlled walking speed v_x , the fundamental frequency f_0 and amplitude coefficients a_1 for walking motions that the test subjects report as natural. Figure 3.24 shows the result of subject motion at different walking speeds on the fundamental frequency term f_0 of the Fourier reconstruction function.

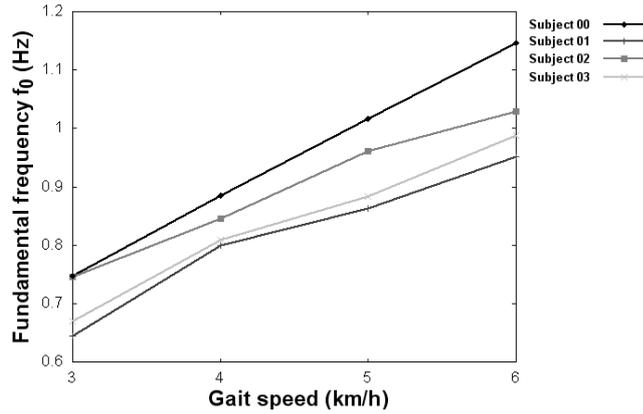


FIGURE 3.24: Change in fundamental frequency f_0 for each of the four test subjects over a range of walking speeds.

We can see that the relationship between fundamental frequency f_0 and walking speed is approximately linear. The gradients and offsets corresponding to each of the linear plots differ between subjects, and similarly correspond to the results of cadence / speed within [103, 102, 5, 6]. The mapping between walking speed, cadence and stride length is influenced by the size of a subject's legs. Subjects with smaller legs need higher cadences to achieve the required walking speeds, thus accounting for the different line offsets between subject plots.

Figures 3.25 and 3.26 show the resulting behaviour of subject upper and lower leg motion at different walking speeds on the a_1 terms of the Fourier reconstruction functions $\theta(t)$. The results for all four subjects are plotted, and show a linear trend with increasing walking speed. The results for subject 01 are significantly different to the others. As previously indicated, poor reconstruction for the trend curve corresponding to subject 01 can be accounted for by the misalignment of marker positions. However, it is interesting to see how much of a difference marker position can make to the reconstruction trend. It may then be worth investigating the potential sensitivity that marker placement has on the accuracy of reconstruction.

The amplitude plots for the lower leg segments shown in figure 3.26 are similar, most notably in initial offset. This may indicate that the lower leg arc of motion is more constrained and similar between people. On the other hand, the amplitude plots for the upper leg segments shown in figure 3.25 are significantly different in initial offset. We can attribute these differences in magnitude offset between subjects to the variation in the size of their limbs.

The apparent natural coupling between walking speed and cadence/stride is evident in the differing line gradients within both the fundamental frequency f_0 and amplitude a_1 plots. The trends for both cadence and stride length are both linear, thus we can make the first order approximation that all limb angle reconstructions $\theta(t)$ are similar, though have different temporal and angular scalings that are dependent on the speed

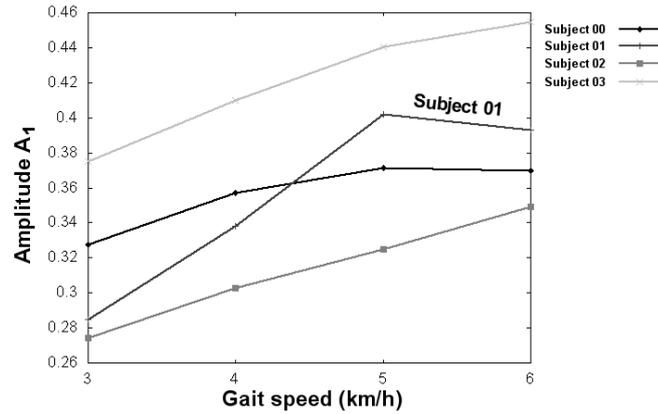


FIGURE 3.25: Change in upper leg fundamental amplitude a_1 for each of the four test subjects over a range of walking speeds.

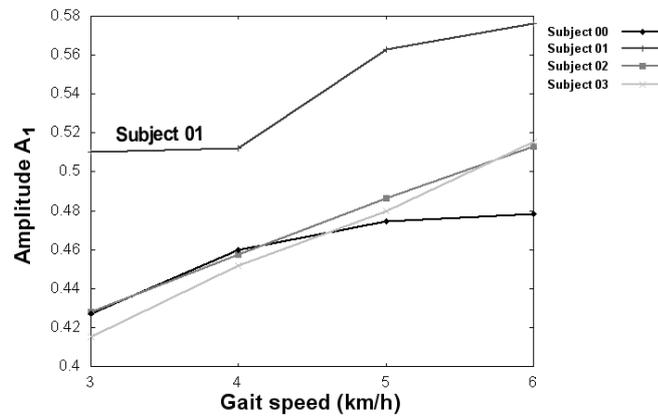


FIGURE 3.26: Change in lower leg fundamental amplitude a_1 for each of the four test subjects over a range of walking speeds.

and mode of walking motion. This allows us to make the reconstructed limb angle functions invariant to walking speed by applying scalings that map the fundamental frequency f_0 and amplitude coefficients a_1 , in both articulated leg segments, to unity. The corresponding set of normalized amplitudes b_k are given by $b_k = a_k/a_1$. The modified Fourier series representation of the original limb angle function $\theta(t)$ can then be written.

$$\theta(t) = a_0 + a_1 \cos(2\pi f_0 \cdot (t + t_s)) + a_1 \cdot \sum_{k=2}^n b_k \cos(2\pi k f_0 \cdot (t + t_s) + \psi_k) \quad (3.19)$$

The set of normalized coefficients $\tilde{\mathbf{v}} = (b_2, \psi_2, \dots, b_n, \psi_n)^\top$ then form the basis for a biometric parameter vector. The remaining parameters $\tilde{\mathbf{w}} = (f_0, a_0, a_1, t_s)^\top$ of the modified Fourier series function form the set of circumstantial parameters of gait motion.

The normalized limb angle function $\tilde{\theta}(t)$ formed from the set of biometric parameters $\tilde{\mathbf{v}}$ alone describes the unique underlying limb dynamics of gait motion, and is invariant to initial subject position, stride length and cadence.

$$\tilde{\theta}(t) = \cos(2\pi t) + \sum_{k=2}^n b_k \cos(2\pi kt + \psi_k) \quad (3.20)$$

Figures 3.27 and 3.28 show the reconstructed normalized leg angle functions $\tilde{\theta}(t)$ corresponding to subject 00 over a range of walking speeds. The set of reconstructed plots are almost identical, even though the captured image sequences correspond to subject motion at different speeds, with different initial poses.

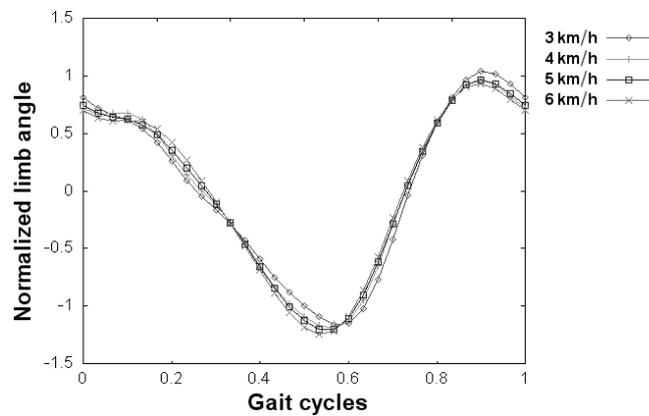


FIGURE 3.27: Reconstructed normalized upper leg angle function $\tilde{\theta}(t)$ corresponding to subject 00.

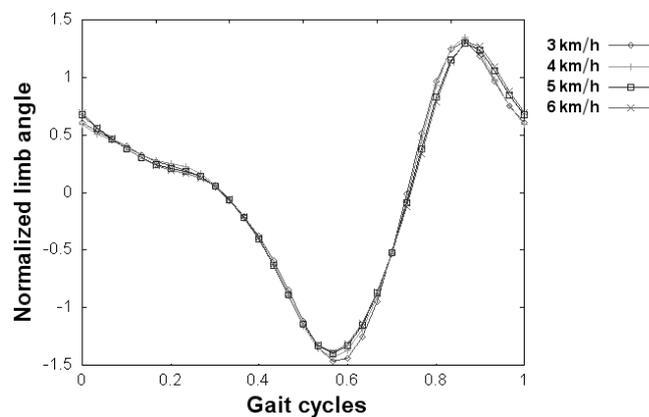
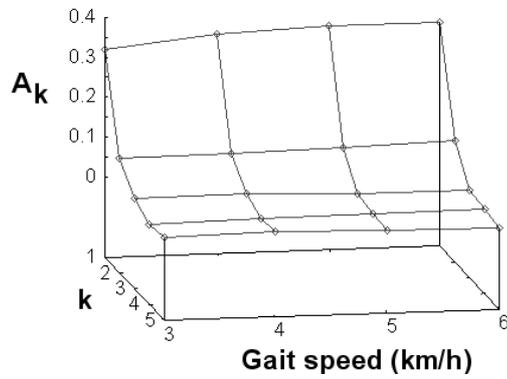


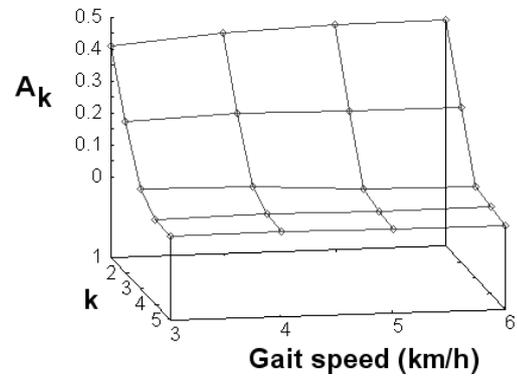
FIGURE 3.28: Reconstructed normalized lower leg angle function $\tilde{\theta}(t)$ corresponding to subject 00.

Figures 3.29 to 3.32, corresponding to all four test subjects, show the original amplitudes, normalized amplitudes and aligned phases for the captured gait dynamics of upper and lower legs over a range of walking speeds.

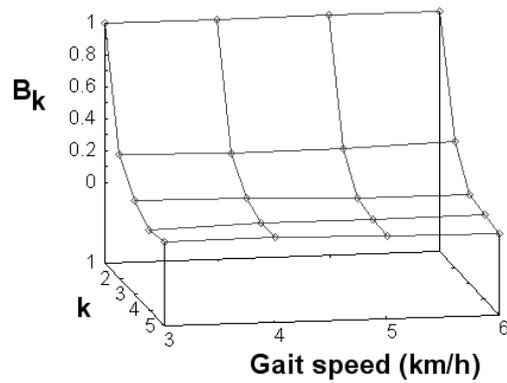
We can clearly see the linear trend corresponding to amplitude / speed changes within the original amplitude plots. The set of amplitude harmonics have an exponential trend,



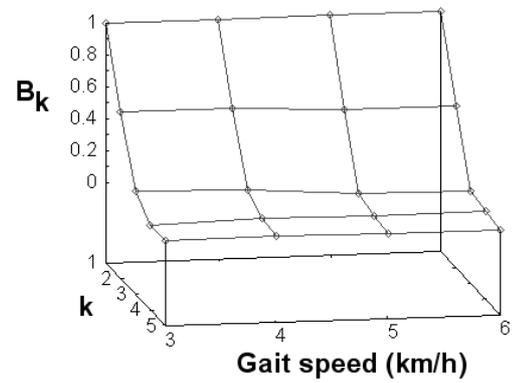
(a) Upper leg amplitudes : a_k



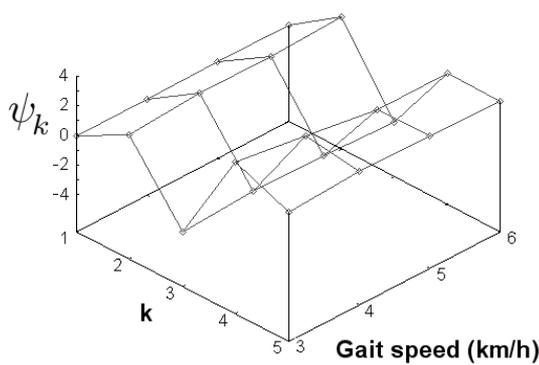
(b) Lower leg amplitudes : a_k



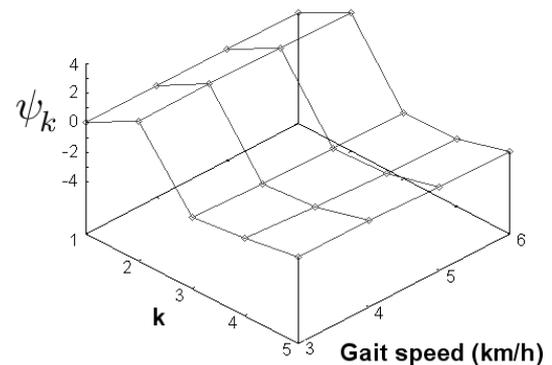
(c) Upper leg normalized amplitudes : b_k



(d) Lower leg normalized amplitudes : b_k

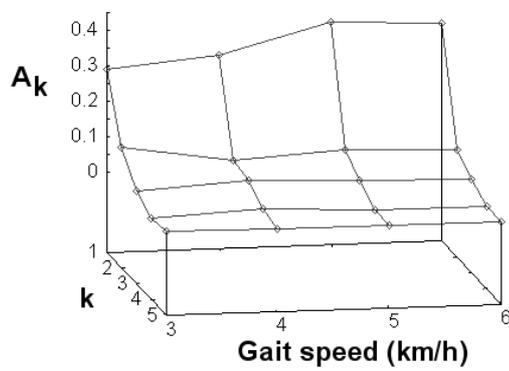


(e) Upper leg aligned phases : ψ_k

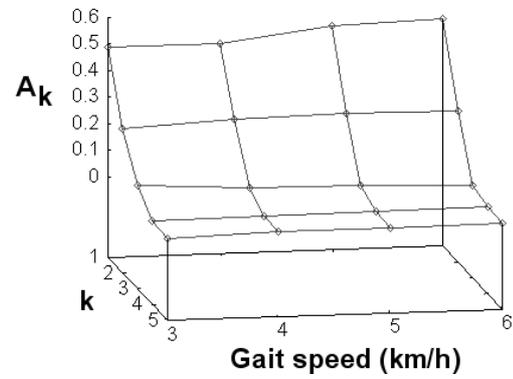


(f) Lower leg aligned phases : ψ_k

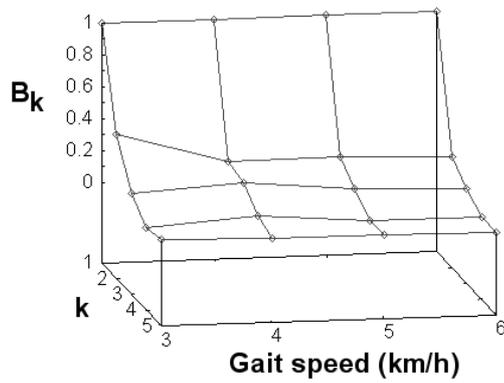
FIGURE 3.29: Subject 00: Reconstructed leg angle functions for amplitude, normalized amplitude and aligned phase plots of both upper (a,c,e) and lower legs (b,d,f) at a number of different walking speeds.



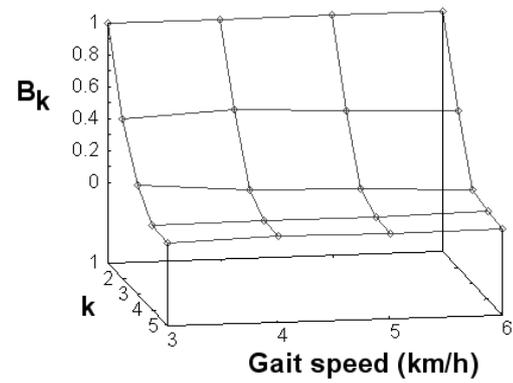
(a) Upper leg amplitudes : a_k



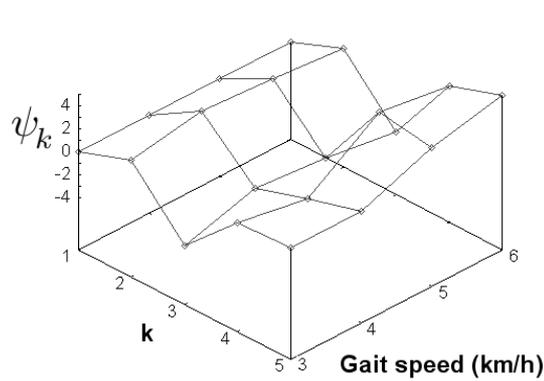
(b) Lower leg amplitudes : a_k



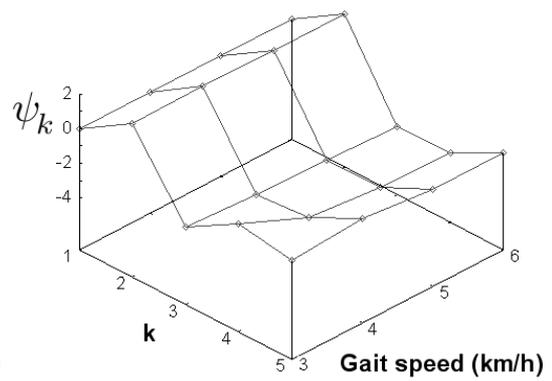
(c) Upper leg normalized amplitudes : b_k



(d) Lower leg normalized amplitudes : b_k



(e) Upper leg aligned phases : ψ_k



(f) Lower leg aligned phases : ψ_k

FIGURE 3.30: Subject 01: Reconstructed leg angle functions for amplitude, normalized amplitude and aligned phase plots of both upper (a,c,e) and lower legs (b,d,f) at a number of different walking speeds.

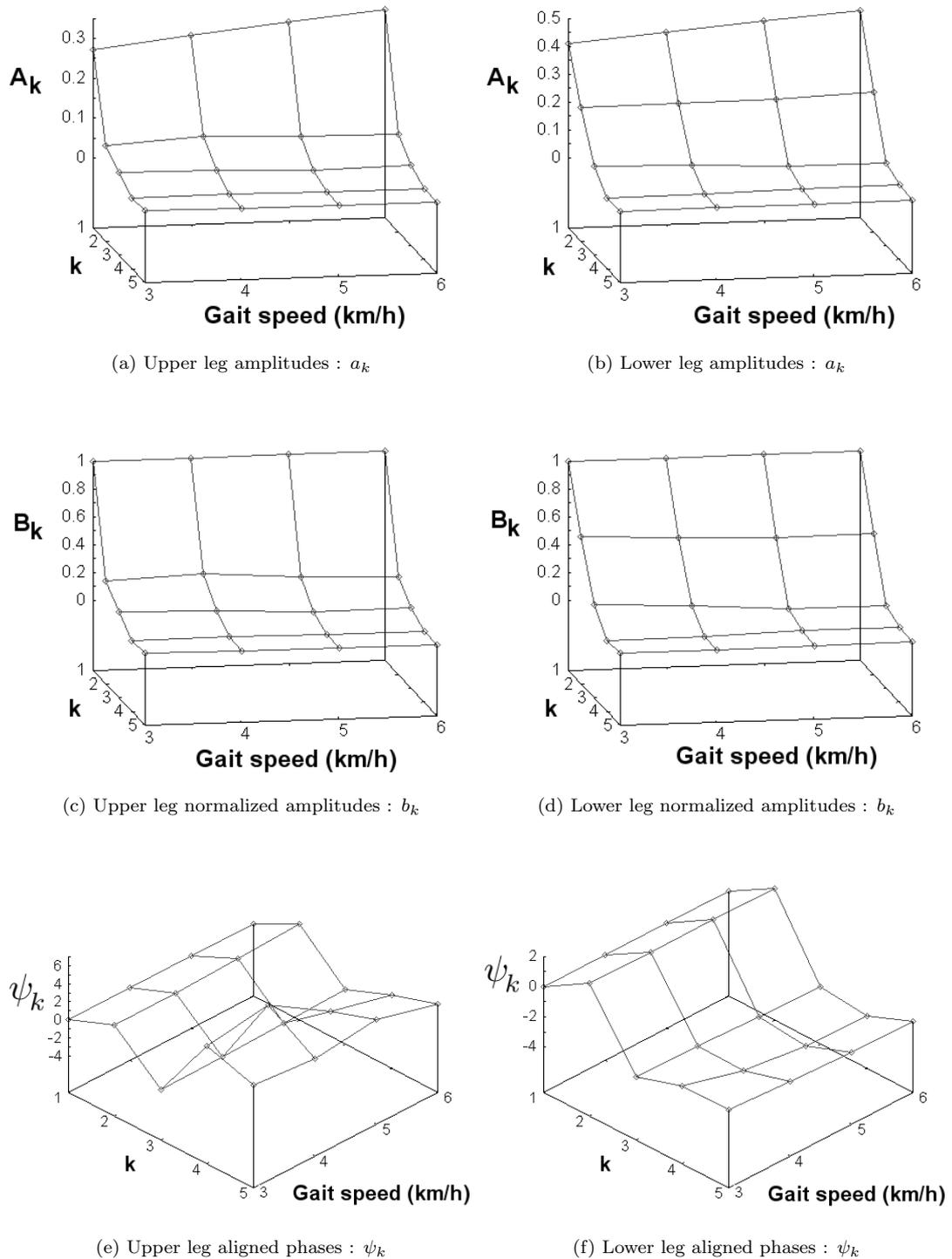


FIGURE 3.31: Subject 02: Reconstructed leg angle functions for amplitude, normalized amplitude and aligned phase plots of both upper (a,c,e) and lower legs (b,d,f) at a number of different walking speeds.

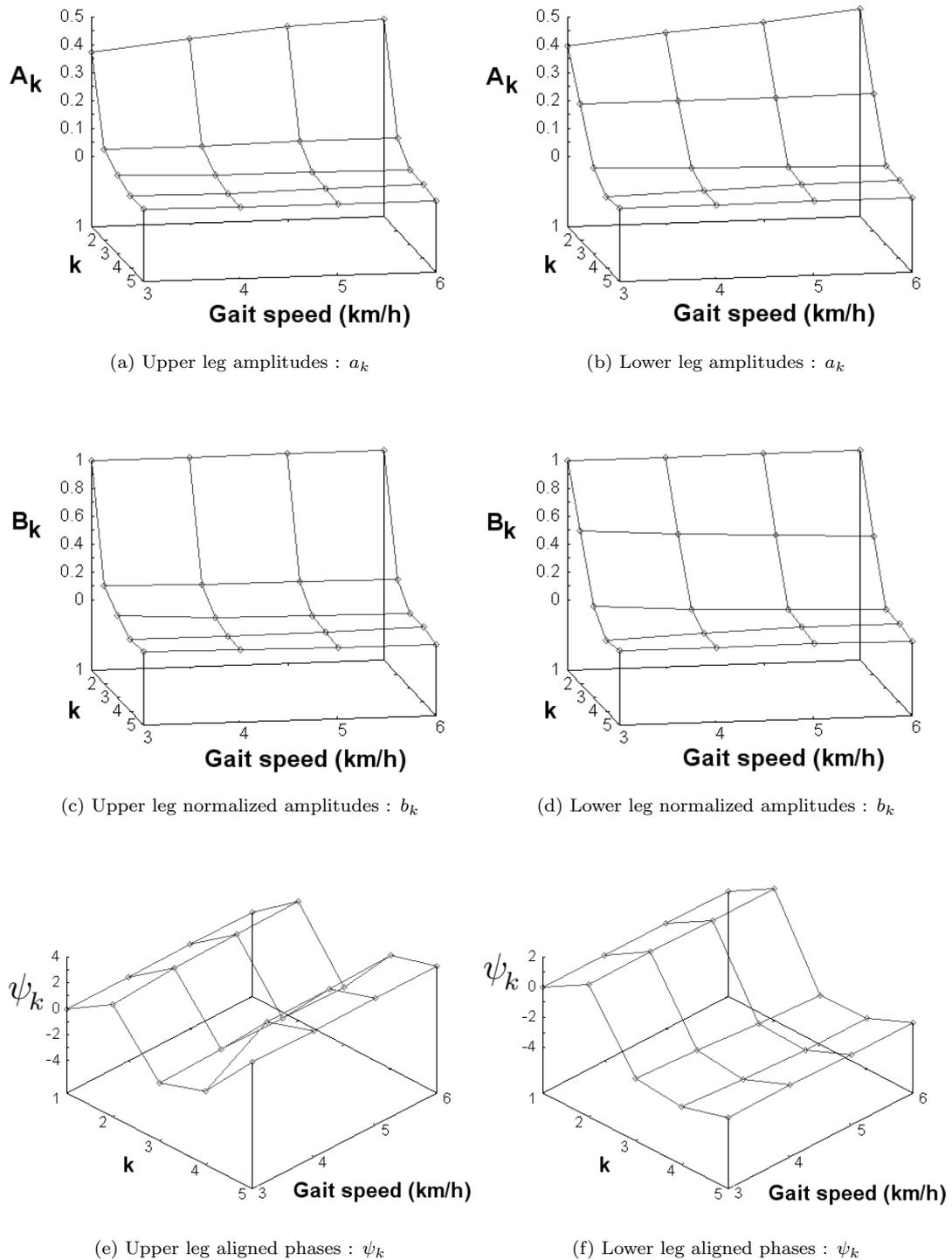


FIGURE 3.32: Subject 03: Reconstructed leg angle functions for amplitude, normalized amplitude and aligned phase plots of both upper (a,c,e) and lower legs (b,d,f) at a number of different walking speeds.

such that the higher order coefficient magnitudes are comparatively smaller than the fundamental. The detail within the higher order harmonics is not visible within the linear scale shown. Figure 3.33 shows the set of normalized Fourier components plotted for each of the test subjects, with a logarithmic scale. We can clearly see the poor reconstruction for subject 01, since the second normalized harmonic terms b_2 for the upper leg are substantially different across all walking speeds. Reconstruction of the lower leg angle function is though reasonable, leading us to believe that the knee joint marker may be placed similarly to that shown in figure 3.23.

The plots for the other subjects approximate a log-linear relationship, illustrated by the straight line trends within the logarithmic plots in figure 3.33. The uncertainty within the higher order harmonics becomes more apparent as k increases. This first order approximation for the dynamics of gait motion over the range of walking speeds gives us a fairly accurate feature vector for the second harmonic normalized amplitude and phase terms $(b_2, \psi_2)^\top$. The remaining coefficients are less accurate, leading us to believe that the relationship between leg functions over different walking speeds is more complicated than just a simple scaling of a baseline waveform.

3.6.5 Motion model discussion

In summary, the modified form of the Fourier series function $\tilde{\theta}(t)$ offers us a way to represent the underlying biometric limb function through a set of normalized harmonic coefficients $(b_2, \psi_2, \dots, b_5, \psi_5)^\top$. The circumstantial parameters of gait motion $(f_0, a_0, a_1, t_s)^\top$ allow us to distort this underlying biometric limb function by applying a series of linear deformations:

- f_0 - scale waveform within the temporal axis.
- a_1 - scale waveform within the θ axis.
- t_s - offset waveform along the temporal axis.
- a_0 - offset waveform along the θ axis.

The modified Fourier series function encodes the perceived angular motion path of the dynamics of gait. However, there is no clear relationship between the underlying muscle motion and the resulting representation of the limb angle function. Most of the muscle energy is expended within the loading response (10% GC) and pre-swing phases (10% GC) of the gait cycle. These phases account for only 20% of the total gait cycle and are located in regions of high curvature within the limb angle function. Figures 3.34 and 3.35 show the reconstructed normalized leg angle functions $\tilde{\theta}(t)$ with corresponding marked positions of gait phase, for subject 00 over the range of walking speeds.

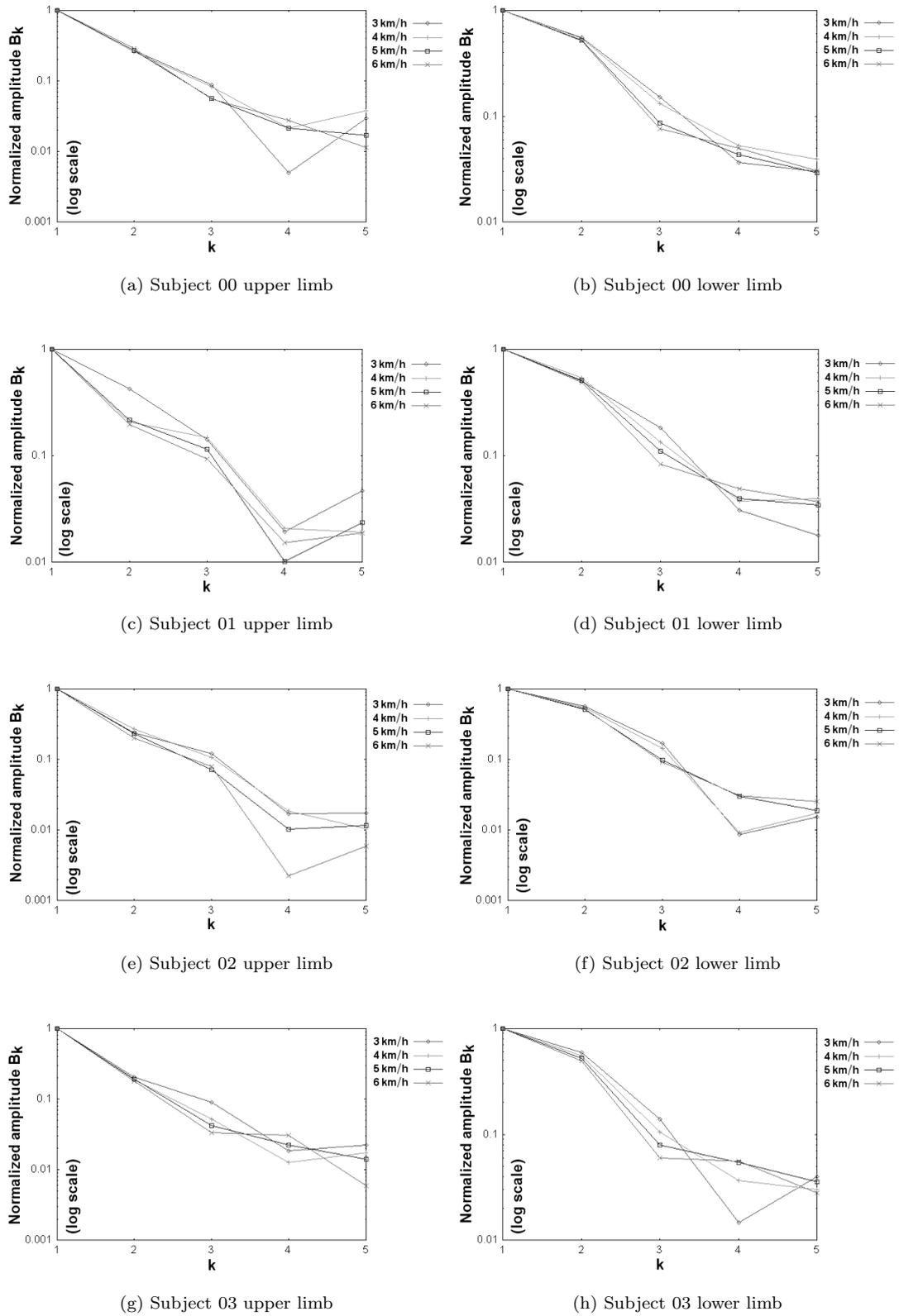


FIGURE 3.33: Reconstructed normalized amplitude coefficients for upper and lower legs corresponding to each of the four test subjects over a range of different walking speeds. The plots are shown with a logarithmic scale.

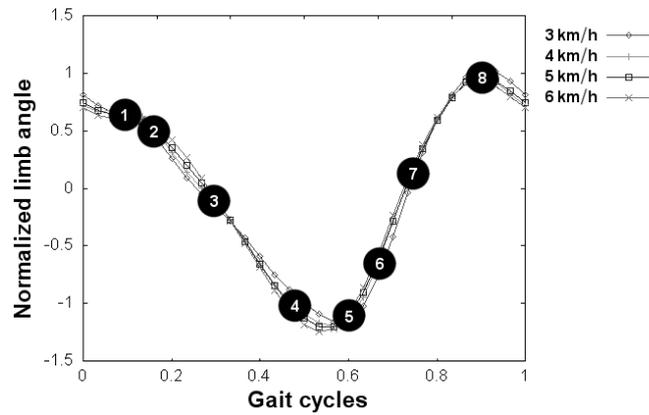


FIGURE 3.34: Reconstructed normalized upper leg angle function $\tilde{\theta}(t)$ corresponding to subject 00 with marked phases of gait.

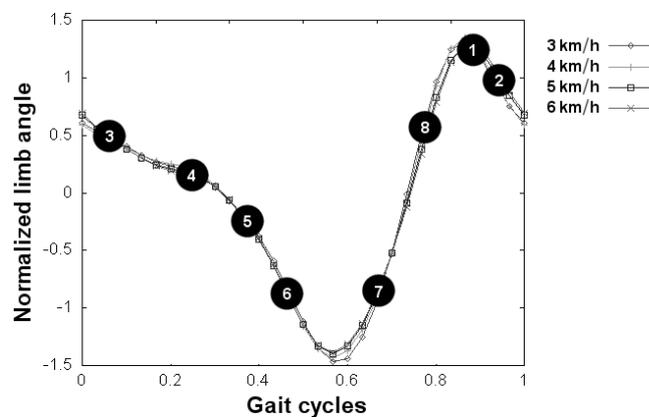


FIGURE 3.35: Reconstructed normalized lower leg angle function $\tilde{\theta}(t)$ corresponding to subject 00 with marked phases of gait.

We have used a linear similarity transformation to deform the baseline limb angle waveform to approximate the appropriate shape. The deformation characteristics for second order and above approximations are unknown and impossible to deduce from the small available data set.

3.6.5.1 Gait reconstruction function

The Fourier representation does a respectable job of representing the underlying motion structure of the limbs, whilst maintaining its invariance to walking speed. It does though tend to capture the dynamics of swing phases more since stance/swing periods account for 40% of the complete gait cycle each, with high rates of motion predominantly apparent within limb swing. It may then be advantageous to treat each phase of gait independently by modelling each section of the reconstructed limb angle function with a different basis function. The duration and position of each of these sections is putatively fixed within the gait cycle, see the review of the medical literature in section 3.2.2. Of the eight phases of gait described in the literature, that represent the complex pattern

of muscle motions, only seven actively contribute to the visible motion. The initial contact phase accounts for only 2% of the gait cycle and is more representable as an instantaneous state rather than a contributory functional phase of gait. Each of these seven phases of gait has a functional objective that requires a series of suitable muscle motions in order to facilitate locomotion. By separating the complex motion pattern into the seven constituent parts, each phase of gait can be accurately modelled by a simple function. We can also apply deformation transformations \mathbf{P}_k independently to each of the phases, in order to better model gait motion across different speeds. A subject's periodicity and initial pose can be adjusted by modifying the temporal parameters \mathbf{T} of motion (period and time shift). The summation of all independent gait phase functions generates the required limb angle function. Figure 3.36 shows a block diagram of the functional model that generates the complex limb angle motion from the seven active phases of gait.

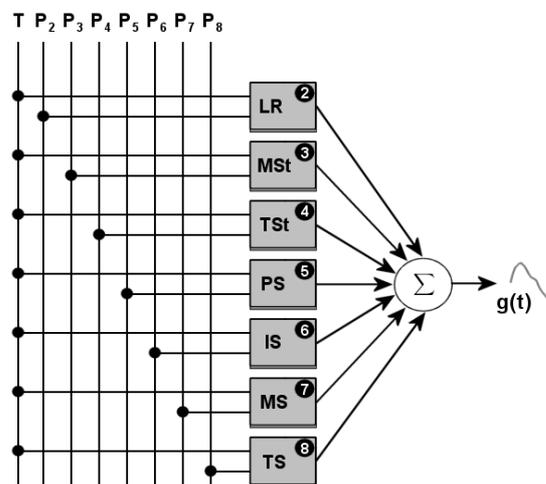


FIGURE 3.36: Block diagram of the functional model that generates the leg angle function from the seven active phases of gait. The temporal \mathbf{T} parameters adjust the periodicity and initial pose of a subject, while the set of individual deformation transformations \mathbf{P}_k adjust each of the functional contributions of gait phase.

The basis functions can be different for each of the active phase blocks. A suitable function that can be used to represent the dynamics of motion, e.g. an n^{th} degree polynomial, is a topic for further investigation. There are many more questions that need to be addressed, though we have no time to look into them further.

- What degree of polynomial is sufficient to represent the motion function?
- We are building up a representation of the limb angle function within each of the functional phases of gait. If we assume no interaction between neighbouring gait phases then in general, the reconstructed waveform is discontinuous. Do we really need to enforce the continuity constraints across the phases of gait?
- Could we model the set of component phase functions by using Splines?

- How do we extract an invariant measure from the resulting polynomial function?

3.6.5.2 Stride and cadence

A person can achieve a desired walking speed in a number of different ways. There exists a mapping between cadence and stride length that allows us to alter the mode of walking whilst maintaining a required velocity. A similar walking speed can be achieved by taking either small strides with a high stepping rate, or large strides with a low stepping rate. The ranges of cadence and stride length are physically limited by the size of a subject's limbs and the effort required to maintain locomotion. These natural ranges of cadence and stride length, coupled with experimental error, generate elliptical clusters of measured coefficients for each controlled walking speed. Figure 3.37(a) shows the theoretical clustering of gait motion parameters f_0 and a_1 over a number of walking speeds.

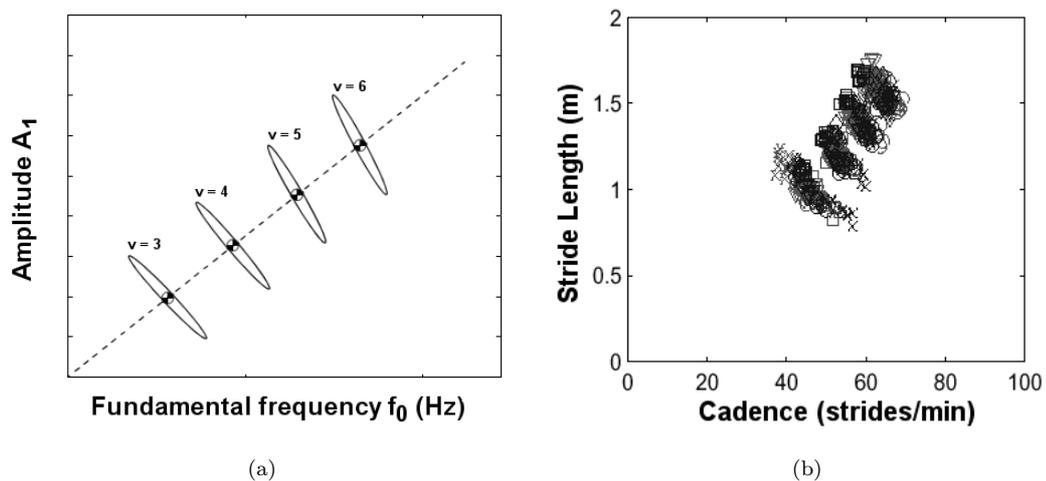


FIGURE 3.37: **Clustering of gait motion coefficients over a range of speeds.** (a) Theoretical clustering of the dynamic motion coefficients (f_0, a_1). (b) A similar plot of stride length and cadence taken from the results of Tanawongsuwan and Bobick [102].

The major axis of each elliptical parameter cluster identifies the linearity between stride length and cadence (s, c), required to achieve the target walking speed. The choice of coefficients (s, c) along the axis line corresponds to the mode of walking. Selected values of cadence and stride length, that are required to maintain a constant velocity, appear normally distributed about the subject's natural choice of motion coefficients (s_n, c_n). The ratio s_n/c_n remains approximately constant over the range of customary walking speeds.

Changes in the motion trends over the range of walking speeds is indicated by the different gradients between cluster axes and the sizes of the corresponding bounding ellipses. The gradients between cluster axes differ as the higher order Fourier harmonics

become more significant, due to the increased energy transfer requirements at greater walking speeds. The sizes of the elliptical clusters decrease with increasing walking speed, since the nature of motion becomes more constrained.

Figure 3.37(b) shows the results of a similar study by Tanawongsuwan and Bobick [102] for the relationship between cadence and stride length parameters over a range of walking speeds. They give an in depth study of 15 people, with repeated measurements on different days, for motions obtained from treadmill walking. We have already shown the linear relationship between cadence and fundamental frequency f_0 , stride length and fundamental amplitude a_1 . Consequently, we expect the same clustering patterns for motion parameters f_0 and a_1 . We cite their results as evidence that validates the theoretical analysis of our representation of the dynamic parameters of gait motion. Due to time constraints and the author's reticence in providing further proofs requiring vast amounts of manual marking, we omit further experimental verification of the dynamic parameters of gait motion over different walking speeds.

An interesting question then arises as to where the stride length / cadence relationship breaks down. There is a fundamental limit on the maximum stride length of a subject during walking, that is dependent on their limb sizes. Cadence can be increased ad-infinitum until the limit of energy transfer in the muscles is reached. The natural way to increase stride length is to break into a run, so that there are periods where both limbs are not in contact with the ground. There is then a choice between walking or running at speeds above 6 km/h. A study of adult males found that the transition speed between walking and running averaged 6.8 km/h [106]. Researchers have also shown that running becomes more efficient than walking at speeds above 7.98 km/h [26].

3.6.5.3 Hip joint motion

Assuming that subject motion is natural on the treadmill then the path of the hip point is repetitious, closed and possibly unique to the individual. Figure 3.38(a) shows the path traced by the reconstructed hip joint marker over the sequence during natural walking for subject 00. Figure 3.38(b) shows the marked positions of gait phase within the corresponding gait cycle.

Subjects attain natural gait motion on a treadmill after a few minutes. It can be quite difficult to maintain a precise speed on a treadmill without visual cues. The disparity between controlled and measured walking speeds can cause a degree of progressional displacement of the subject on the treadmill over time. A subject's gait is often natural, though the velocity is either lower or higher than the controlled speed. Figure 3.39(a) shows the effects of progressional displacement on the reconstructed motion of subject 00, over a period of three gait cycles at 6 km/h.

Figure 3.39(b) shows similar subject double stance poses at the beginning and after

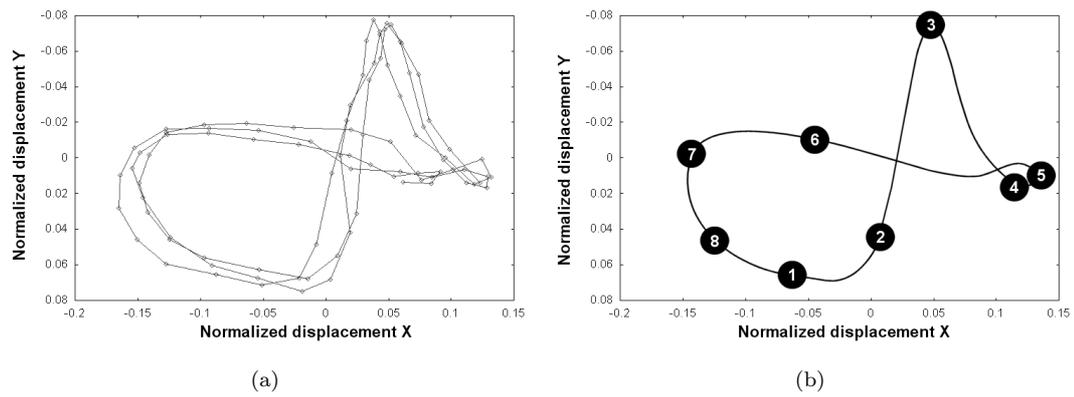


FIGURE 3.38: **Normalized origin limb point displacement of subject 00 at a walking speed of 5 km/h.** (a) Reconstructed motion path of the hip point over three gait cycles. (b) Trend of hip joint motion shown with marked positions of gait phase.

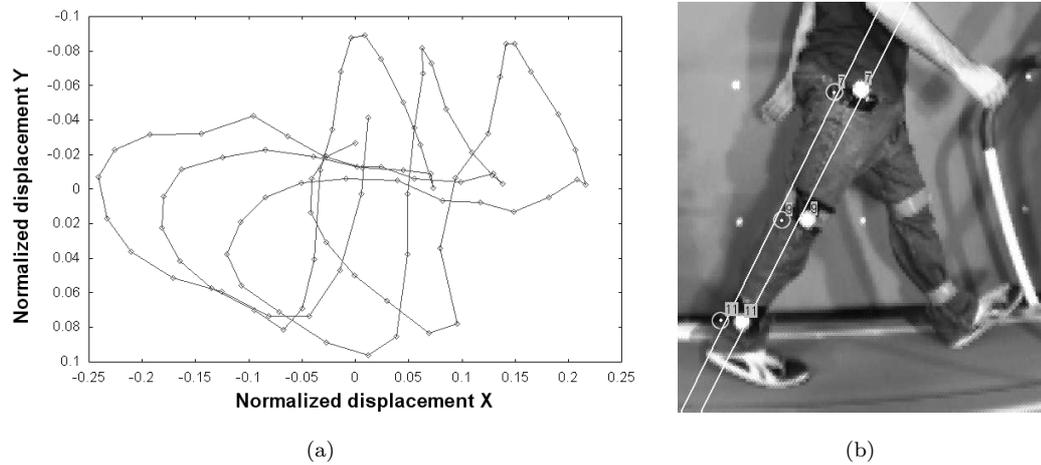


FIGURE 3.39: **Progressional displacement of treadmill gait motion.** (a) Reconstructed hip joint motion path. (b) Similar poses during double stance periods at the beginning (visible frame) and after three gait cycles of motion (augmented line with circular markers).

three gait cycles of motion. We can clearly see that the subject's speed is less than the controlled treadmill speed, resulting in the backward progressional motion drift. The resulting reconstructed hip joint motion shown in figure 3.39(a) is similar over gait cycles, though has a linear velocity component corresponding to the residual between measured and controlled speeds. This displacement effect may explain some of the visible curvature corresponding to subject 00 within the fundamental amplitude plots shown in figures 3.25 and 3.26.

The path of the hip joint is also periodic and is instrumental in the resulting pattern of limb motion. We can model the hip joint displacement motion with modified Fourier series functions in the X and Y axis directions. The two displacement functions differ only in the additional velocity term within the progressional X motion direction.

$$\begin{aligned}
x(t) &= v_x \cdot t + a_0 + a_1 \cos(2\pi f_0 \cdot (t + t_s)) + \\
&\quad a_1 \cdot \sum_{k=2}^n b_k \cos(2\pi k f_0 \cdot (t + t_s) + \psi_k)
\end{aligned} \tag{3.21}$$

$$\begin{aligned}
y(t) &= a'_0 + a'_1 \cos(2\pi f_0 \cdot (t + t'_s)) + \\
&\quad a'_1 \cdot \sum_{k=2}^n b'_k \cos(2\pi k f_0 \cdot (t + t'_s) + \psi'_k)
\end{aligned} \tag{3.22}$$

The underlying dynamics of hip displacement motion, that remain invariant to changes in walking speed, are represented by the normalized coefficients $\tilde{\mathbf{x}} = (b_2, \psi_2, \dots, b_n, \psi_n)^\top$ and $\tilde{\mathbf{y}} = (b'_2, \psi'_2, \dots, b'_n, \psi'_n)^\top$. These coefficients can be used in conjunction with the parameters of the normalized leg angle functions $\tilde{\theta}(t)$ to form a suitable biometric feature vector. The remaining coefficients within the displacement functions form the set of circumstantial parameters of hip motion $\tilde{\mathbf{w}} = (f_0, v_x, a_0, a_1, t_s, a'_0, a'_1, t'_s)^\top$, and encode the subject cadence, gait speed, cyclic displacements and initial pose positions.

3.6.5.4 Marker configuration

Reliance on accurate placement of markers is often critical for motion reconstruction. Placing markers over actual joint regions of the body is difficult, since these areas undergo the largest changes in deformation. Measurement errors of shared joint markers affect the results of both connecting limb segments. If we are only interested in limb angles, then increasing the number of markers used to model the mid-line of each limb segment reduces the sensitivity to sources of experimental noise.

We can model each limb segment better by using three markers. Each limb segment mid-line is computed as the line through all three joint markers, by the process of orthogonal regression. The knee joint location is then determined as the intersection of both upper and lower mid-line approximations. The added attraction of using three markers is that there is a distinctive cross ratio configuration between the three imaged limb markers and the vanishing point on the limb swing plane. The corresponding set of imaged vanishing points over all frames then defines the vanishing line of the limb swing plane. We can use this information along with the static length constraints [62], between the three markers to compute a stratified reconstruction of the limb angle motion. This removes the need to use a planar calibration target placed against the leg during quiet standing, in order to find a similar planar reconstruction homography.

3.6.6 Biometric identification

Researchers have previously used magnitude weighted phase as a suitable feature vector [18, 17, 19, 117, 119] for biometric identification. Their results suggest that magnitude weighted phase achieves better discrimination between subjects, when compared to components of magnitude or phase alone. Lower order harmonics seem to be more significant, encode the gross features of cyclic motion and are less susceptible to experimental noise than those of higher orders. Random measurement noise and natural fluctuations in day to day subject motion patterns account for the majority of these differences within the higher order harmonics.

Values of phase are normalized modulo 2π to the range $(-\pi : \pi)$, hence phases distributed around both extremes of the range must be treated with caution. Naively matching coefficients of magnitude weighted phase in order to facilitate recognition is asking for trouble. Each phase direction vector corresponding to each of the harmonic contributions is of unit length, hence all vectors carry an equal weighting. Two putatively similar phase vectors \mathbf{p} and \mathbf{q} can be matched by computing a cost error \mathcal{C} based on the dot product between both direction vectors: $\mathcal{C} = \frac{1}{2}(1 - \mathbf{p}^\top \mathbf{q})$. This cost error lies within the range $(0 : 1)$. Since all phase vectors have equal weighting, then the sum of all residual costs, between the set of phases in two gait feature vectors, is susceptible to a level of noise contamination from the higher order harmonics. We can assign a probabilistic weighting factor of significance ω_k to each of the harmonics. The obvious choice is to assign the significance weights from the tail of some suitable Gaussian (an exponential function). However, should we use the same Gaussian for each person? What happens if the amplitude of a phasor happens to be zero?

The first thing we notice about the distribution of the set of normalized amplitude coefficients b_k is that they are log-linear, illustrated by the straight line plots in figure 3.33, hence are of the required exponential form. Since we have also normalized the set of coefficients so that the first component b_1 is unity, then the normalized amplitude coefficients b_k can be used directly as the required set of significance weights ω_k . These weighting factors are perforce invariant to changes in gait speed, stride length and cadence, thus are ideal for comparative purposes over the range of required walking conditions.

We have identified a suitable number of invariant features that can be used for subject identification. The static parameters of articulated leg motion include the normalized limb segment lengths (d_2, \dots) , and corresponding normalized amplitude and phase coefficients of the modified Fourier series leg functions (b_2, ψ_2, \dots) . We can then compute a Euclidean distance metric $\rho = \|\hat{\mathbf{P}} - \mathbf{P}\|$, between the measured biometric feature vector \mathbf{P} of subject motion and a feature vector $\hat{\mathbf{P}}$ stored in the database.

$$\rho^2 = \lambda_1 \sum_{j=2}^m (\hat{d}_j - d_j)^2 + \lambda_2 \sum_{k=2}^n (\hat{b}_k - b_k)^2 + \lambda_3 \sum_{k=2}^n \left(\frac{b_k}{2} (1 - \hat{\mathbf{v}}_k^\top \mathbf{v}_k) \right)^2 + \dots \quad (3.23)$$

where the phase direction vectors $\hat{\mathbf{v}}$ and \mathbf{v} are commuted from polar to Euclidean form, m is the number of segments in the articulated limb model, and n is the number of Fourier harmonics used to represent the limb motion. We can also pick the respective weighting factors λ_i to bias the fitting error in favour of any particular set of feature coefficients.

$$\hat{\mathbf{v}}_k = (\cos \hat{\psi}_k, \sin \hat{\psi}_k)^\top \quad (3.24)$$

$$\mathbf{v}_k = (\cos \psi_k, \sin \psi_k)^\top \quad (3.25)$$

$$\sum_{i=1}^N \lambda_i = 1 \quad (3.26)$$

Note that the order of magnitude of each set of coefficients is similar, since all biometric parameters have been normalized so that the set of first coefficients are unity: $d_1 = 1$, $b_1 = 1$ and $\hat{\mathbf{v}}_1 \cdot \mathbf{v}_1 = 1$ (first phase angles are zero).

Our results are similar to the works of BenAbdelkader and Tanawongsuwan in the fact that the set of fundamental amplitude terms a_1 is proportionally similar to stride length, and fundamental frequency f_0 is proportional to cadence. We choose to unit normalize these parameters in order to retain a consistent scaling over different walking speeds. The first order approximation for the dynamics of gait motion over the range of walking speeds gives us a fairly accurate set of invariant features, corresponding to the second harmonic normalized amplitude and phase terms (b_2, ψ_2) within figure 3.33. The similarity between the remaining harmonic coefficients is less accurate, since they are more dynamically related to the mode of motion and range of customary walking speeds. For this reason, it is better to use only the second harmonic terms within the feature vector of static gait parameters.

If we model the normalized limb angle functions $\tilde{\theta}(t)$ of both upper and lower leg segments, then we have one parameter for the limb length ratio d_2 and two parameters (b_2, ψ_2) for each modified Fourier series limb function. Overall, we have five distinct static parameters of gait motion that we can use for biometric identification. In addition, if we model the hip joint displacement motion $x(t)$ and $y(t)$ with modified Fourier functions then the number of static parameters of gait motion can be increased to nine measurements. Although theoretically sound, further research needs to be done in order to validate the usefulness of these features to discriminate between subjects over the

customary range of different gait motions.

3.6.7 Intra and inter-class variation

We have discussed in previous sections the variation of parameters within the articulated leg motion function, for individual subjects over a range of walking speeds. We have outlined a suitable Euclidean distance metric $\rho = \|\hat{\mathbf{P}} - \mathbf{P}\|$ between a known $\hat{\mathbf{P}}$ and unknown \mathbf{P} gait feature vector. Here, we wish to study the variation of ρ , in order to provide a suitable measure of whether $\hat{\mathbf{P}}$ and \mathbf{P} belong to the same subject. Specifically, we wish to examine the variation in two ways: The variation that arises from differences in measurements from the same subject (intra-class variation), and the variation resulting from the differences between measurements of different subjects (inter-class variation).

To describe this variation we take the set of measured parameter vectors \mathbf{P}_i , for the three valid test subjects walking over a range of speeds. These vectors form the sample parameter set that we wish to study. We then compute the matrix \mathcal{D} of difference measurements, by evaluation of the biometric distances $\mathcal{D}_{i,j} = \|\mathbf{P}_i - \mathbf{P}_j\|$ between parameter vectors, as described in equation 3.23. The difference matrix is symmetric and has the block form:

$$\mathcal{D} = \begin{pmatrix} \mathcal{D}_{aa} & \mathcal{D}_{ab} & \mathcal{D}_{ac} \\ \mathcal{D}_{ab}^\top & \mathcal{D}_{bb} & \mathcal{D}_{bc} \\ \mathcal{D}_{ac}^\top & \mathcal{D}_{bc}^\top & \mathcal{D}_{cc} \end{pmatrix} \quad (3.27)$$

where each sub-block is a 4×4 matrix, corresponding to the set of biometric difference measures for the four walking speeds 3 – 6 km/h. The diagonal blocks represent the set of parameter differences for individual subjects over the four walking speeds, while the off-diagonal blocks represent the set of parameter differences between different subjects.

The intra and inter-class difference measurement sets are then extracted from the elements of \mathcal{D} . Since \mathcal{D} is symmetric, we form the class data sets from values within the upper triangular portion of \mathcal{D} . The intra-class data set \mathcal{M}_v corresponds to values from the upper triangular portions of blocks \mathcal{D}_{aa} , \mathcal{D}_{bb} and \mathcal{D}_{cc} (excluding values on each of the major diagonals, since they are zero). The inter-class data set \mathcal{M}_c corresponds to values within the blocks \mathcal{D}_{ab} , \mathcal{D}_{ac} and \mathcal{D}_{bc} .

Having created both class data sets, we can find the mean μ and variance σ^2 of each set, i.e. the intra and inter-class means and variances μ_v, σ_v^2 and μ_c, σ_c^2 .

$$\mu = \frac{1}{N} \sum_{i=1}^N \rho_i \quad \forall \rho \in \mathcal{M} \quad (3.28)$$

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (\rho_i - \mu)^2 \quad \forall \rho \in \mathcal{M} \quad (3.29)$$

We then compare and contrast both the intra and inter-class variances for our valid set of test subjects. We give a quantitative assessment of the discriminatory capability of each of the biometric features within the proposed parameter vector \mathbf{P} , and demonstrate that gait is sufficiently rich to be useful as a potential biometric.

3.6.7.1 Discrimination between individual parameter features

We first investigate the variation within individual parameters in the proposed biometric feature vector \mathbf{P} . We therefore assume that the feature vector only contains a single parameter, thus the distance metric $\rho = \|\widehat{\mathbf{P}} - \mathbf{P}\|$ describes the intra and inter-class variances, σ_v^2 and σ_c^2 , for that parameter. Furthermore, we can quantitatively assess the ability of this parameter to discriminate between people, by evaluating two properties that are dependent on these class variances.

The class distinction quantity γ describes the percentage of the intra-class variance compared with that of the inter-class variance, i.e. $\gamma = 100 \times \sigma_v^2 / \sigma_c^2$. Low percentages indicate good discrimination between different subjects, while high percentages highlight the inability of the biometric parameter to distinguish between people at all.

The class distinction says nothing about the magnitude of the parameter variation, only the ratio of variation between classes. The parameter distinction quantity β describes the intra-class parameter deviation as a percentage of the mean subject parameter estimate μ_p , i.e. $\beta = 100 \times \sigma_v / \mu_p$. This percentage gives us an indication of the relative magnitude of the intra-class parameter deviation.

Normalized limb segment lengths. The variance of the normalized limb segment lengths d_j is examined over the set of biometric parameter vectors \mathbf{P}_i . The gait distance metric $\rho = \|\widehat{\mathbf{P}} - \mathbf{P}\|$, corresponding to the normalized limb segment lengths d_j , can be written.

$$\rho^2 = \sum_{j=2}^m (\widehat{d}_j - d_j)^2 \quad (3.30)$$

Figure 3.40 shows the difference matrix \mathcal{D} corresponding to the normalized limb segment length parameter d_2 . The normalized limb lengths can be described as the static geometric parameters of gait motion, which remain invariant to any changes in walking motion. Consequently, we expect the variance within these parameters to be quite low.

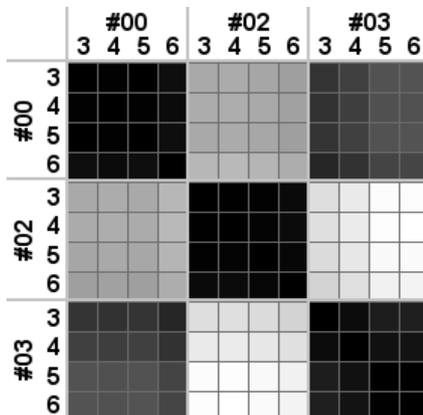


FIGURE 3.40: Difference matrix \mathcal{D} corresponding to the normalized limb segment length parameter d_2 . Intra and inter-class variances: max = 0.224398, intra = 6.3693e-05, inter = 0.003951, $\gamma = 1.61\%$, $\beta = 0.91\%$.

The figure confirms this prediction, and shows good distinction between different subjects. The intra-class deviation of parameter d_2 is estimated at 0.91% of the mean subject limb segment length, and the magnitude of the intra-class variance at 1.61% of the inter-class variance level. This demonstrates that d_2 is a well defined biometric parameter that has good discrimination between subjects.

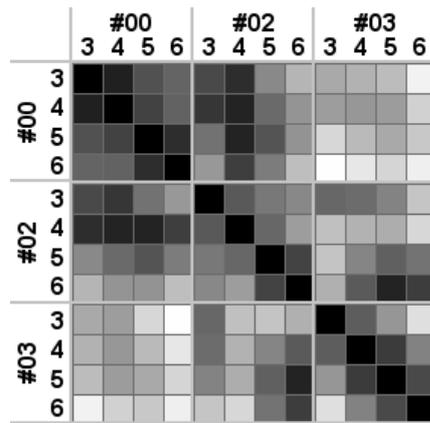
Normalized amplitude components. The variance within each of the normalized amplitude components b_k is examined over the set of biometric parameter vectors \mathbf{P}_i . The corresponding gait distance metric $\rho = \|\hat{\mathbf{P}} - \mathbf{P}\|$ has the form:

$$\rho^2 = \sum_{k=2}^n (\hat{b}_k - b_k)^2 \quad (3.31)$$

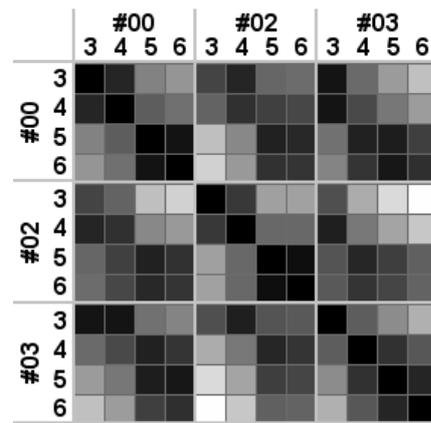
Figure 3.41 shows the difference matrices \mathcal{D} corresponding to the individual normalized amplitude components b_k of the leg angle function (components of both upper and lower limb segments). We have previously shown within section 3.6.4, that the intra-class variance of the second order amplitudes b_2 is reasonably static.

The intra-class deviation of parameter b_2 is estimated at 3.78% of the mean normalized amplitude, and demonstrates that the coefficient b_2 remains relatively constant over the range of customary walking speeds.

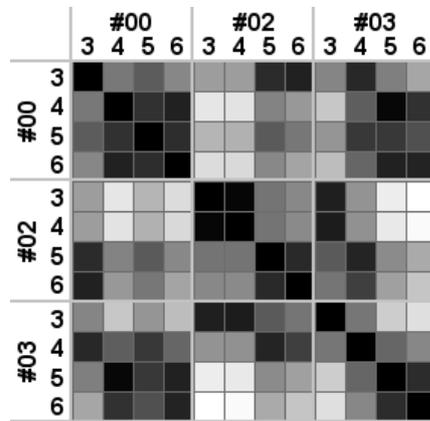
The magnitude of the intra-class variance is estimated at 57.64% of the inter-class variance level. The intra-class variance then has almost twice the level of discrimination



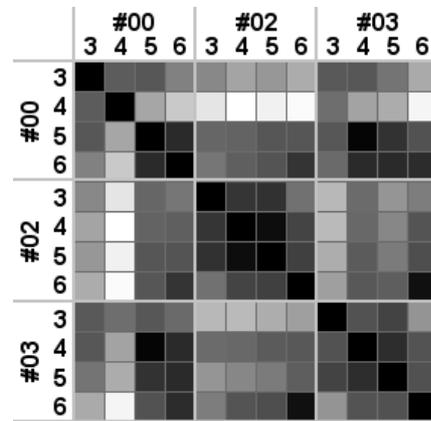
(a) b_2 : max = 0.082147, intra = 0.000211, inter = 0.000365, $\gamma = 57.64\%$, $\beta = 3.78\%$



(b) b_3 : max = 0.099194, intra = 0.000373, inter = 0.000534, $\gamma = 69.7\%$, $\beta = 20.96\%$



(c) b_4 : max = 0.034871, intra = 6.03734e-05, inter = 8.96055e-05, $\gamma = 67.38\%$, $\beta = 29.55\%$



(d) b_5 : max = 0.024746, intra = 2.01419e-05, inter = 3.36887e-05, $\gamma = 59.79\%$, $\beta = 19.95\%$

FIGURE 3.41: Difference matrices \mathcal{D} corresponding to the individual amplitude components b_k of the normalized leg angle function.

over the inter-class matches. As a biometric feature, the normalized amplitude component b_2 appears relatively weak in comparison to the geometric static parameter of gait d_2 , which has over sixty times the level of discrimination.

The higher order amplitudes are less reliable and show significant levels of parameter deviation $\beta > 20\%$ from their mean estimates. Discrimination between the intra and inter-class variances is also poor, with $\gamma > 60\%$. Table 3.4 shows the result of combining ranges of normalized amplitudes b_k within the distance metric ρ . The distinction between class variances appears to worsen as more components are combined. The table suggests that we are not able to distinguish between people at all if we include these higher order components.

\sum_k	<i>Intra-class variance</i>	<i>Inter-class variance</i>	γ
$k = 2$	0.000210425	0.000365082	57.64%
$k = 2 \cdots 3$	0.000500881	0.000503108	99.56%
$k = 2 \cdots 4$	0.00052302	0.000467122	111.97%
$k = 2 \cdots 5$	0.000518958	0.000444299	116.8%

TABLE 3.4: Discrimination over all normalized amplitude components b_k . Combining the higher order amplitude components results in equal intra and inter-class variances.

Normalized phase components. The variance within each of the aligned phase components ψ_k is examined over the set of biometric parameter vectors \mathbf{P}_i . We first study the variance of the dot product between unit vector representations of these phase angles. The gait distance metric $\rho = \|\widehat{\mathbf{P}} - \mathbf{P}\|$ can then be written in terms of the corresponding phase angle unit vectors $\widehat{\mathbf{v}}_k$ and \mathbf{v}_k .

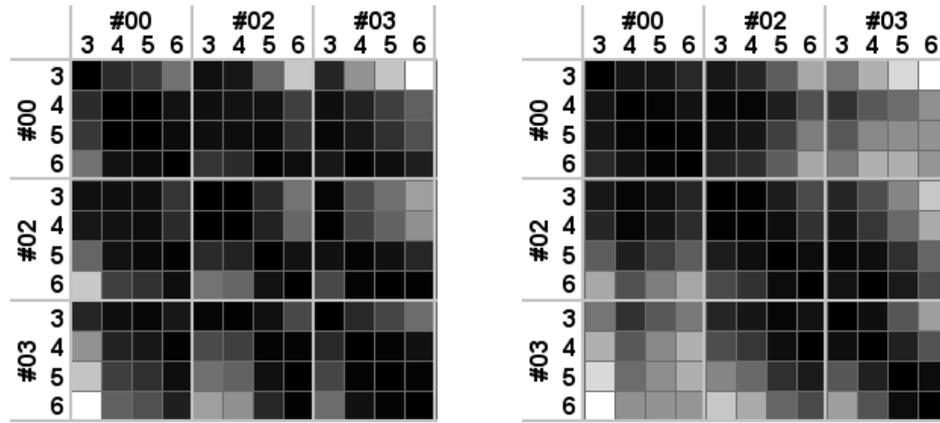
$$\rho^2 = \sum_{k=2}^n \left(\frac{1}{2} (1 - \widehat{\mathbf{v}}_k^\top \mathbf{v}_k) \right)^2 \quad (3.32)$$

Figure 3.42 shows the difference matrices \mathcal{D} corresponding to the individual phase components ψ_k of the normalized leg angle function (components of both upper and lower limb segments). Figures 3.29 to 3.32 on page 97 showed that the lower order intra-subject phase angles remained fairly consistent over the range of walking speeds. This is also reflected by the quantitative assessment of the intra-class variance shown within figure 3.42. The dot product measure between phase vectors lies within the range (0 : 1), therefore we express the parameter distinction quantity β as the percentage of intra-class deviation over this unit range, i.e. $\beta = 100 \times \sigma_v$. This percentage deviation remains quite low $\simeq 2\%$ for the first two phase components then quickly becomes unstable. This is also reflected by the corresponding poor class distinction percentages γ between the intra and inter-class variances. The higher order components have equal intra and inter-class variances, thus are unsuitable as potential biometric features.

As biometric parameters, the phase components are similar to the normalized amplitudes, in that they appear relatively weak in comparison to the geometric static parameter of gait d_2 . This is not unexpected, since leg motion is highly dynamic and we have only approximated a set of consistent gait features over the range of walking speeds.

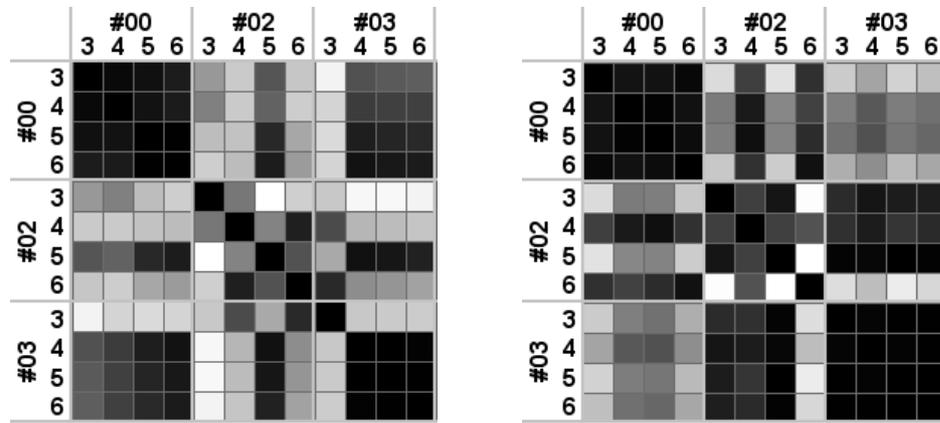
The level of variance within each phase component increases with higher order. We can then choose to weight the contribution of each phase component with the corresponding normalized amplitude, in order to increase the significance of the lower order phases. Subsequently, the magnitude weighted phase version of the gait distance metric $\rho = \|\widehat{\mathbf{P}} - \mathbf{P}\|$ can be defined as:

$$\rho^2 = \sum_{k=2}^n \left(\frac{b_k}{2} (1 - \widehat{\mathbf{v}}_k^\top \mathbf{v}_k) \right)^2 \quad (3.33)$$



(a) ψ_2 : max = 0.130181, intra = 0.000412, inter = 0.000901, $\gamma = 45.69\%$, $\beta = 2.03\%$

(b) ψ_3 : max = 0.156812, intra = 0.000572, inter = 0.001564, $\gamma = 36.58\%$, $\beta = 2.39\%$



(c) ψ_4 : max = 0.883111, intra = 0.093758, inter = 0.069765, $\gamma = 134.39\%$, $\beta = 30.62\%$

(d) ψ_5 : max = 0.32412, intra = 0.009619, inter = 0.008877, $\gamma = 108.37\%$, $\beta = 9.81\%$

FIGURE 3.42: Difference matrices \mathcal{D} corresponding to the individual phase components ψ_k of the normalized leg angle function. No magnitude weighting is performed within the distance metric ρ .

Table 3.5 compares the class distinction percentages γ between magnitude weighted phase and normal phase variances. The class distinction between higher order phase components is improved by the magnitude weighting, while the lower order components remain similar.

k	Intra-class variance	Inter-class variance	γ (MWP)	β	γ
2	2.50272e-05	6.23271e-05	40.15%	2.03%	45.69%
3	1.09812e-05	2.48628e-05	44.17%	2.39%	36.58%
4	2.44339e-05	4.06368e-05	60.13%	30.62%	134.39%
5	2.02181e-06	6.00553e-06	33.67%	9.81%	108.37%

TABLE 3.5: The magnitude weighted phase (MWP) components ψ_k of the normalized limb angle function. The class distinction between higher order components is improved by the magnitude weighting.

Table 3.6 shows the result of combining ranges of magnitude weighted phase components within the distance metric ρ . The first two phase components ψ_2 and ψ_3 individually remain relatively constant over the range of walking speeds. The overall distinction between intra and inter-class variance is then improved by combining both of these phase components, for both non-weighted and magnitude weighted distance measures.

\sum_k	<i>Intra-class variance</i>	<i>Inter-class variance</i>	γ (MWP)	γ
$k = 2$	2.50272e-05	6.23271e-05	40.15%	45.69%
$k = 2 \cdots 3$	3.17207e-05	8.04965e-05	39.41%	39.82%
$k = 2 \cdots 4$	4.19072e-05	8.8079e-05	47.58%	139.57%
$k = 2 \cdots 5$	4.18024e-05	8.7462e-05	47.79%	172.11%

TABLE 3.6: Discrimination over all magnitude weighted phase (MWP) components.

Individually, the higher order phase components ψ_4 and ψ_5 vary significantly within their unit range, and have similar intra and inter-class variances. Consequently, the level of class distinction is worsened by combining these higher order phase components within the distance metric. Table 3.6 shows that the use of magnitude weighted phase clearly alleviates the impact of including these noisy higher order phase terms.

3.6.7.2 Uniqueness of the proposed biometric

We first examine the intra and inter-class variances for the combined components of amplitude and phase, corresponding to the upper and lower leg angle functions. Subsequently, the gait distance metric $\rho = \|\widehat{\mathbf{P}} - \mathbf{P}\|$ has the form:

$$\rho^2 = \sum_{k=2}^n (\widehat{b}_k - b_k)^2 + \sum_{k=2}^n \left(\frac{b_k}{2} (1 - \widehat{\mathbf{v}}_k^\top \mathbf{v}_k) \right)^2 \quad (3.34)$$

Figure 3.43 shows the difference matrices \mathcal{D} corresponding to increasing numbers of normalized amplitude and phase components b_k and ψ_k within the biometric feature vector. The magnitude of the intra-class variance for the second order parameters b_2 and ψ_2 is estimated at 56.23% of the inter-class variance level. However, the intra and inter-class variances are similar when further normalized amplitude and phase components are combined within the distance metric. This leads us to believe that only the second order components b_2 and ψ_2 are sufficiently similar across walking speeds to be useful as potential biometric features.

The static geometric parameters of gait are invariant to the dynamic changes between walking motions, thus significantly contribute to the ability of the biometric to discriminate between people. The final proposed biometric feature vector contains five components $\mathbf{P} = (d_2, b_2, \psi_2, b'_2, \psi'_2)^\top$, where d_2 is the normalized lower leg segment length, (b_2, ψ_2) are the normalized amplitude and phase terms of the upper leg angle function,

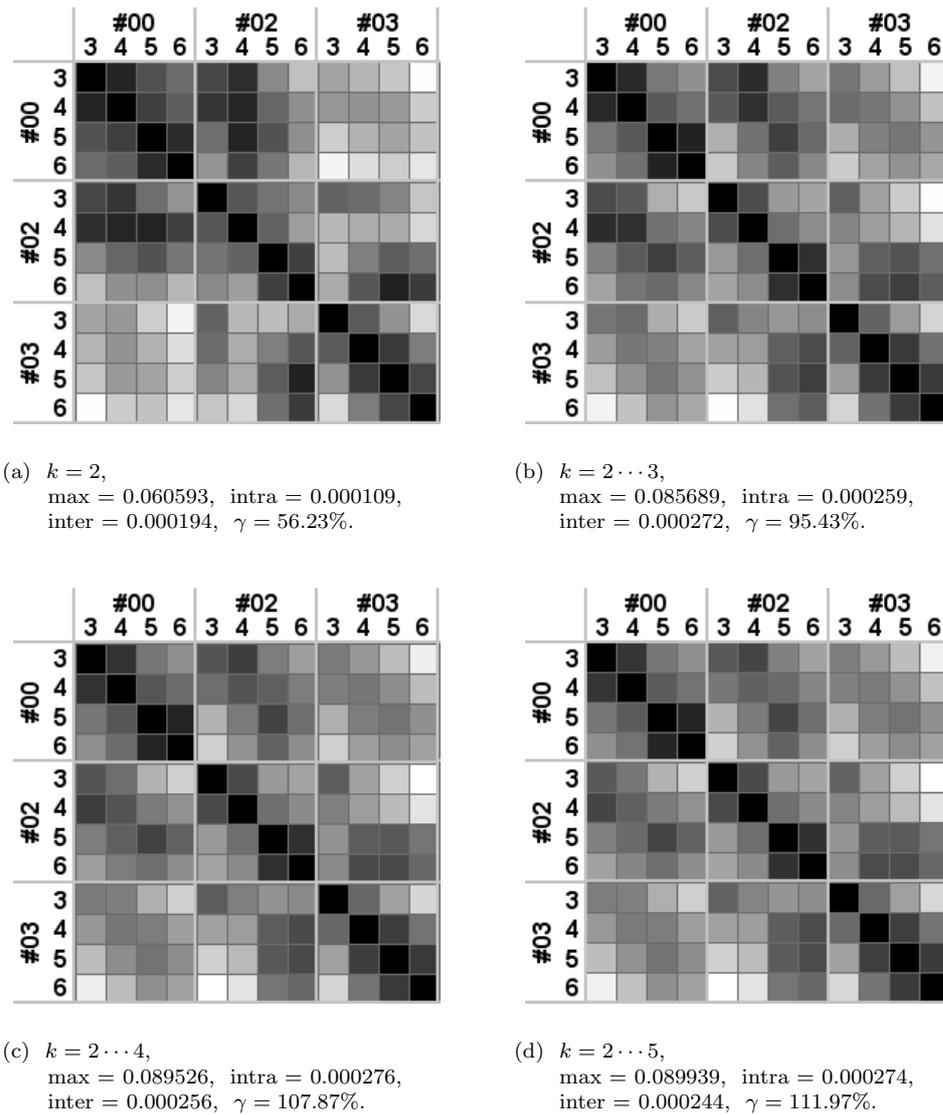


FIGURE 3.43: Difference matrices \mathcal{D} corresponding to the combined amplitude and phase components (b_k, ψ_k) of the normalized limb angle function. Magnitude weighted phase terms are used within the distance metric ρ .

and (b'_2, ψ'_2) the corresponding coefficients of the lower leg angle function. The proposed gait distance metric $\rho = \|\widehat{\mathbf{P}} - \mathbf{P}\|$, suitable for subject identification, can then be written:

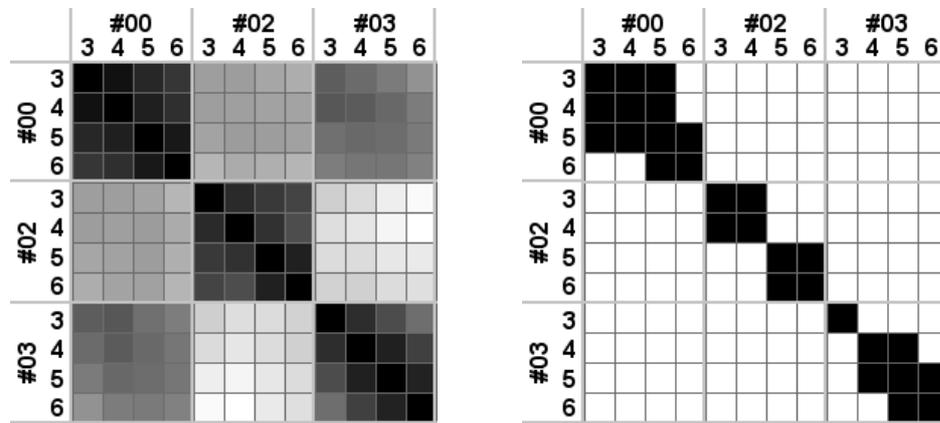
$$\begin{aligned} \rho^2 = & \left(\widehat{d}_2 - d_2 \right)^2 + \left(\widehat{b}_2 - b_2 \right)^2 + \left(\frac{b_2}{2} (1 - \widehat{\mathbf{v}}_2^\top \mathbf{v}_2) \right)^2 \\ & + \\ & \left(\widehat{b}'_2 - b'_2 \right)^2 + \left(\frac{b'_2}{2} (1 - \widehat{\mathbf{v}}_2'^\top \mathbf{v}'_2) \right)^2 \end{aligned} \quad (3.35)$$

where the phase direction vectors $\widehat{\mathbf{v}}$ and \mathbf{v} are commuted from polar to Euclidean form.

$$\hat{\mathbf{v}}_2 = (\cos \hat{\psi}_2, \sin \hat{\psi}_2)^\top \quad \hat{\mathbf{v}}'_2 = (\cos \hat{\psi}'_2, \sin \hat{\psi}'_2)^\top \quad (3.36)$$

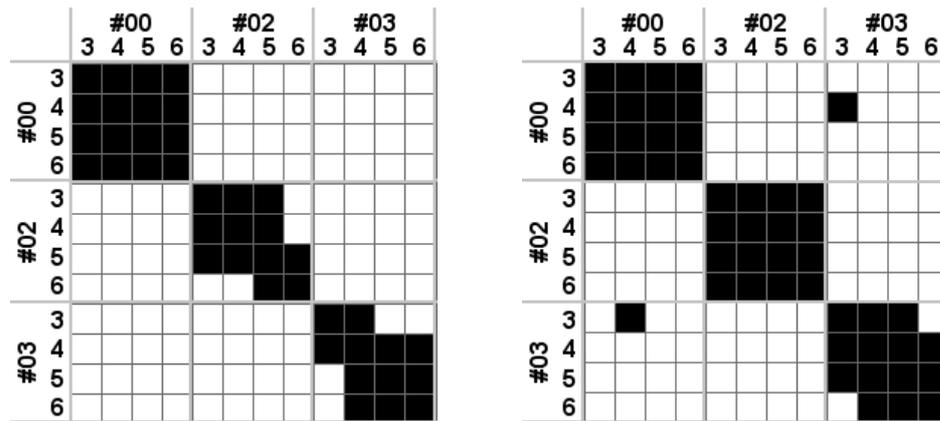
$$\mathbf{v}_2 = (\cos \psi_2, \sin \psi_2)^\top \quad \mathbf{v}'_2 = (\cos \psi'_2, \sin \psi'_2)^\top \quad (3.37)$$

Figure 3.44(a) shows the difference matrix \mathcal{D} corresponding to the proposed biometric feature vector. The magnitude of the intra-class variance is estimated at 21.03% of the inter-class variance level, i.e. the variation of the intra-subject biometric feature vectors measured over the range of walking speeds is almost five times smaller than the variation between different people.



(a) max = 0.11027, intra = 9.13742e-05,
inter = 0.000435, $\gamma = 21.03\%$.

(b) thresh = $2\sigma_v$, false accept = 0%,
false reject = 41.67%



(c) thresh = $3\sigma_v$, false accept = 0%,
false reject = 16.67%

(d) thresh = $4\sigma_v$, false accept = 2.08%,
false reject = 4.17%

FIGURE 3.44: Difference matrices \mathcal{D} corresponding to the proposed biometric feature vector $\mathbf{P} = (d_2, b_2, \psi_2, b'_2, \psi'_2)^\top$. (a) The computed difference matrix \mathcal{D} . (b-d) The thresholded difference matrices at a number of integer levels of the intra-class deviation σ_v .

We then classify each subject by thresholding the difference matrix \mathcal{D} at a suitable level. We choose this level depending on the intra-class deviation σ_v . Figures 3.44(b) to 3.44(d) show the difference matrix \mathcal{D} thresholded at a number of different levels. We quantify the uniqueness of the biometric parameterisation by determining the false acceptance and false rejection rates of subject classification. Specifically, we are interested in the equal error rate of subject classification as a suitable uniqueness measure. Figure 3.44(d) shows an equal error rate of $\simeq 3\%$ at a threshold of $4\sigma_v$. This level of classification error demonstrates the uniqueness of gait and its potential to be a reliable biometric.

We have shown that the most discriminating biometric feature is the normalized limb segment length, with a distinction level of $\gamma = 1.61\%$. Magnitude weighted phase has the best discrimination of the limb angle function features, with $\gamma = 40.15\%$. Normalized amplitude components of the limb angle function are then the least discriminating, with $\gamma = 57.64\%$. There is an order of magnitude difference in the discrimination between the static geometric and limb angle motion features, that can be attributed to the dynamic and multi-modal nature of articulated limb motion.

There are other similar static features that can be used within gait motion. We have considered here only the leg motion, but we could just as easily use the arm motion and the fixed length between the head and pelvis as sources of additional biometric features. We have also made a linear approximation to the biometric mapping, over the range of customary walking speeds. The intra-class variance can be reduced if we model the articulated limb motion more precisely and determine a better biometric mapping over the range of walking motions. This may subsequently improve the discrimination between subjects, and could also allow us to include more of the limb angle function coefficients within the biometric feature vector. Further work needs to be done to establish whether this is possible.

3.7 Experimental procedure

A pilot capture session run as a precursor to the experimental phase of this project highlighted some important practical concerns. We must stress the importance of following proper scientific procedure, especially when the data capture and subsequent manual marking processes can take days or even weeks to perform. We describe here the planned tasks that need to be followed in order to achieve a successful set of experiments.

In the previous sections we described a number of experiments that required us to capture subject motion data corresponding to both treadmill and overground walking. The treadmill data was used to analyse the leg motion function over a range of controlled walking speeds, and validate our assumptions about the static features of gait. The captured motion data corresponding to overground walking from each camera view was

used to triangulate the positions of a subject's joint features, thus enabling us to validate the planarity assumption of limb motion. The captured sequences from each camera view can also be used independently as different imaged trajectories of similar gait motion. We discuss the reconstruction and comparison of these different imaged motion sequences within chapter 4, though give here the details of how we conduct each set of required subject treadmill and overground walking experiments.

While the cameras in the array only need to be calibrated once, each subject capture session needs to be consistent and coordinated in a controlled manner. The events of a subject capture session can be summarised as follows:

- **Marker placement.** The set of feature markers are attached to the principal joint features of the skeletal system with adhesive tape. Subjects are then instructed to pace up and down the test track. This firstly gives them a chance to familiarise themselves with the test track and the starting cues by making a few dummy runs. Secondly, the walking motion ensures that the markers are aligned naturally with the clothing and joint positions. Any deviation, caused by the effects of clothing, on the original placement of markers can be corrected here before any data is collected. Similarly, any feedback from the subject regarding any restriction in movement, due to markers being attached too tightly, can also be addressed.
- **Overground walking.** Subjects are instructed to make passes along the test track from both directions and at three different walking speeds (slow, natural and fast). These speeds are entirely dependent on the interpretation of the individual. The actual walking speed of a subject is recovered by triangulation of the head feature marker, and the velocity computed based on the distance travelled along the line of progression within the captured sequence of frames.

The operator must allow the subject to make a couple of passes, to ensure that they are walking naturally, before starting to capture the image sequence. The operator should never inform the subject of when they intend to start the data capture, since it may unduely affect the subject's concentration and natural rhythm.

- **Treadmill walking.** The treadmill is then placed in the centre of the test track so that it is fronto-parallel to the camera array. The positions of the treadmill feet are previously marked onto the floor so that we can repeat its exact placement to within a millimetre. The subject is then told to stand on the treadmill in the *quiet standing* posture, while the calibration grid is held fronto-parallel against their closest leg. A single frame is then captured and stored for later analysis. This enables us to determine the articulated limb swing plane of the required reference leg. Subjects are then given six minutes [70] to familiarise themselves with treadmill walking while the operators set-up for the data capture.

The subject is then recorded walking over a range of gait speeds (3.0 km/h,

4.0 km/h, 5.0 km/h and 6.0 km/h) on the treadmill. Each captured image sequence is four seconds in length, thus contains approximately four periods of motion.

Communication with the subject while walking on the treadmill often breaks their rhythm and concentration. Talking should be kept to a minimum, and silence strictly observed during the data acquisition, both for subjects and operators. The operator first asks the subject if they are ready, i.e. if they feel that their walking is natural. The operator must then leave a time gap of at least 15 seconds before starting the data capture in order that the subject regains their natural rhythm. The operator must in no way inform the subject when they intend to start capturing the data.

The operator must increase the belt speed themselves after each captured image sequence to ensure that the correct settings are selected. The subject is then left to familiarise themselves with the new gait speed for a period of not less than two minutes, while the set of image sequences are downloaded from the capture computers to the main storage database. This procedure is repeated until all of the required walking speeds have been completed.

3.8 Conclusions

We have presented the material in this chapter not as a definitive piece of work, but as an interesting conceit into the suitability of subject motion as a source of features for biometric identification. This study goes some way towards answering many of the fundamental questions that are necessary for a practical system: What is gait? Which features of motion, suitable for biometric identification, remain invariant to changes in walking speed.

In order to better understand the nature of gait we first gave an overview of gait motion from the medical literature. Researchers have identified eight distinct phases of motion within the gait cycle. Each phase has a functional objective that requires the action of specific muscle motions in order to align and progress the limbs forward. There are a total of six major motion patterns, known as the determinants of gait, that result from the underlying limb and muscle actions.

The human body is made up of a series of articulated limb segments. We hypothesized that each articulated limb can be modelled by motion within a single swing plane. We then proceeded to validate this assumption by reconstructing the worldspace motion structure of specific anatomical landmark points during periods of overground walking. A synchronized, calibrated three camera system was used to capture the motion of a set of four test subjects. Retro-reflective markers were attached to the principal joint features of the skeletal system of each subject. The corresponding landmark features

were manually marked within each camera view and back projected to form the set of worldspace points.

For each subject, the entire set of reconstructed worldspace limb markers corresponding to their nearest leg were then fitted to a single motion plane. The residual variance between worldspace markers and limb swing plane was observed as a measure of the departure from motion planarity. We determined that the 3σ confidence interval accounted for a gross motion deviation of approximately 3.5 cm over the set of test subjects. A brief analysis of the level of image reprojection error, caused by assuming this empirical planarity deviation, was given by simulating fronto-parallel gait motions at a number of different distances from the camera. Image reprojection error is dependent on both object depth and displacement within the view. In general, subject motion is never localized to any one single region of the image alone. For this reason we proposed a measure, the standard pixel reprojection error, that gives a practical estimate of the probable mean error over a sequence of imaged motion.

The standard pixel reprojection error for fronto-parallel subject motion at a distance of 3.2 metres from the camera is approximately 2.8 pixels. This level of error is reasonably high and provides a sizeable contribution of systematic error, unaccounted for by the motion model, to the reconstruction process. However, the corresponding standard pixel reprojection error computed for a practical set-up similar to most CCTV systems with subject motion at a distance of 10 metres is approximately 0.9 pixels. This level of error is well within the landmark measurement deviation tolerance and consequently the planar limb swing motion assumption is valid.

Most of the deviation from planarity can be explained by the lateral sinusoidal deviation of subject motion over the gait cycle. Our results agree with the findings of medical studies that suggest that all parts of the body are similarly displaced from the mid-line of progression by approximately 4 cm.

Static features of gait are defined as quantities that remain constant over the full spectrum of subject walking motions. The bone segments within an articulated limb are rigid and of fixed length throughout the entire image sequence. Though we may not know the actual physical sizes of any of the segment lengths, i.e. we are unable to determine the scaling transformation that maps the canonical imaged limb lengths (pixels) to worldspace measurements (mm), the length ratio between upper and lower leg lengths is geometrically invariant to changes in scale. We then normalize the set of leg segments such that the first segment has unit length. The resulting normalized lengths represent the required set of limb length ratio invariants, and form the static parameters of gait.

Each articulated limb pose is represented by a Euclidean hip displacement $(x, y)^\top$ position followed by a series of connected rigid length bone segments, defined by polar coordinates $(d, \theta)^\top$. The angular motion of each limb segment $\theta(t)$ can be represented by a Fourier series function. Each test subject was told to walk on a treadmill at a

number of controlled walking speeds. The set of leg angle functions were then reconstructed with increasing numbers of Fourier harmonics, in order to determine a suitable value that enables us to accurately model the limb motion. The results suggest that increasing the number of harmonics beyond five has little significant benefit on the level of pixel reprojection error. Our results agree with the findings of previous studies by Angeloni [3] who also suggests that five harmonics is sufficient to model subject limb motion.

The set of reconstructed limb angle Fourier coefficients vary significantly over the range of walking speeds. Within each reconstruction, the amplitude harmonic coefficients are exponentially related such that the first amplitude is the most significant, and encodes the gross angular variation. The higher order amplitude harmonics are the least significant and are more responsive to measurement noise and changes in walking speed. The trends between both first harmonic amplitude and fundamental frequency, versus walking speed are approximately linear. Similar findings have been shown by BenAbdelkader [5, 6] and Tanawongsuwan [103, 102] who both show linear relationships between both stride length and cadence, versus gait speed. Similarities can be inferred between our works in the fact that the first harmonic amplitude coefficients encode the gross periodic angular swing and are thus proportional to subject stride length. Subject cadence (rate of stepping) is similarly proportional to the fundamental frequency of the Fourier series.

If we normalize the amplitude coefficients such that the first coefficient is unity, then the remaining parameters represent the ratio between selected and first harmonic amplitudes. Analysis of the reconstructed leg angle functions for each subject shows that the normalized second order amplitude harmonics remain constant over the customary range of walking speeds. The remaining coefficients are less consistent over different walking speeds, which leads us to believe that the characteristics of articulated leg motion are changed more subtly than by a simple first order linear scaling. These normalized second order amplitude harmonics then form a further set of static parameters of gait. We can also remove the differences in initial pose by computing the time shifts that zero the first coefficients of phase corresponding to the upper and lower leg angle functions. Analysis similarly shows that the normalized second order harmonic coefficients of phase remain constant over the range of gait speeds, thus provide the final set of static parameters of gait motion. A modified form of the Fourier series function offers us a way to represent the underlying biometric leg function $\tilde{\theta}(t)$, through the set of normalized harmonic coefficients $(b_2, \psi_2, \dots, b_5, \psi_5)^\top$. The circumstantial parameters of gait motion $(f_0, a_0, a_1, t_s)^\top$ then allow us to distort this underlying leg function to approximate an arbitrary gait motion, by applying a series of linear deformations (scale and offset) in both the temporal and angular coordinate axes.

The hip displacement motion can also be modelled by using similar modified Fourier series functions. In addition, a velocity term is included with the X displacement function in order to model the linear progression of subject motion. The full articulated motion

model is parameterised by using five Fourier harmonics. Consequently, a total of 45 coefficients are required to represent the complete dynamics of articulated leg motion. Since normal gait is bilaterally symmetric with a half phase shift, these 45 parameters simultaneously encode the motion of both left and right legs.

Other static parameters of gait have been suggested. Tanawongsuwan and Bobick [103, 102] exploit the linear relationship between stride length and cadence over different velocities in order to map subject motion to a common walking speed. The two parameters of normalized stride length and cadence are then used as the static motion parameters of gait, allowing us to compare similar subject motions. However, where the constant of proportionality is not known between any individual subject's range of walking speeds (the norm), a generic global estimate is used to map the parameters to the reference speed (one size fits all).

Throughout the chapter we have identified nine static features of articulated leg motion that remain invariant to differences in the mode of subject motion. These features are hypothetically unique to each individual, thus can be used as suitable parameters for biometric identification. Where other techniques look for specific key frames, our method maximizes the utilization of the measurement data to make a robust estimate of the gait motion parameters, even in the presence of occlusion and image noise.

Many of the questions put forward have been addressed either theoretically or still remain unanswered. This study is a positive step towards understanding the dynamic nature of gait in order to extract features that are invariant over the range of customary walking motions.

Chapter 4

Pose Invariant Gait Reconstruction

4.1 Introduction

Since the gait of a person is readily identified when extracted from a canonical side view, most gait recognition algorithms work with the premise that the motion is fronto-parallel in nature [117, 43], or at least require some knowledge of the camera calibration in order to reconstruct the motion [93, 55]. Others have used projective factorization techniques [39], in the case of unknown and varying camera focal lengths, to reconstruct the static scene for rigidly structured objects that move with linear velocity. Realistically, people will always walk along different trajectories to the camera. We hypothesize that articulated limb motion within human gait is approximately planar, since almost all of the perceived limb motion is contained within a single plane. The nature of plane perspective distortion is well understood. The mapping between a worldspace scene plane and the image plane is a 2D homography [33, 92, 98].

Geometric properties of the plane can be classified into three main groups of transformation: perspective, affine and similarity transformations. Identification of specific entities within the image allows us to employ a stratified technique [41, 16] to map them back to their canonical positions. Metric structure of the scene plane is typically recovered in a two step process: i) Identification of the imaged vanishing line of the scene plane allows us to compute the perspective transformation that recovers the affine properties of the plane. ii) Identification of the imaged circular points \mathbf{I}, \mathbf{J} then allows us to compute the affine transformation that recovers the metric properties of the plane.

Liebowitz and Zisserman [62] have shown that metric structure on a plane can be recovered via a stratified technique that uses the constraints formed from known angles, equal but unknown angles and known ratios of lengths. These constraints are quadratic

and may be easily combined to recover the appropriate rectification transformation that restores the metric properties of the plane. Liebowitz and Carlsson [61] have also shown that affine reconstruction of 3D human motion is possible from multiple synchronized views. They use the fact that articulated limb segments are rigid and of fixed length in all views.

The work presented in this chapter details a stratified approach to pose rectification in single view sequences that makes use of the rigid limb segments of articulated human motion. We have already laid out much of the groundwork and mathematical theory within section 2.8. We have demonstrated that the metric properties corresponding to a planar test pattern of synthesized human motion can be recovered by using a stratified reconstruction technique. This chapter develops the geometric properties and biometric features identified within chapters 2 and 3, though with respect to real human motion sequences. We develop a novel method that exploits the geometric properties of articulated leg motion in order to compute a stratified reconstruction of the fronto-parallel dynamics of gait motion. We assume no prior knowledge of the camera calibration, only that people walk in straight lines with constant velocity and legs that swing in planes. We assume that the computer vision task of finding limb landmark points and tracking them over all frames in the sequence is solved and that the camera sampling rate is high enough to capture the dynamics of gait.

The articulated leg motion of both left and right sides of the body can be approximated by motion within two separate planes. We can apply further constraints on the form of the articulated leg motion by simultaneous consideration of the bilateral symmetry between left and right limbs. Consequently, human motion can be modelled by using a cardboard person assumption. A subject's body and leg parts are then modelled as a set of repeating spatio-temporal motion patterns within separate planes. Repetition of structure on a plane is defined by a specialized homography transformation known as a conjugate translation [89, 90]. Identification of subject periodicity and point correspondences over different gait cycles is solved simultaneously by computing the self-similarity of structure [21, 22] within the image sequence. Determination of this conjugate translation provides enough constraints that enables us to recover the vanishing line of the limb swing plane. We describe a method to robustly recover the subject periodicity from the self-similarity of landmark feature points, then further develop the stratification process outlined within chapter 2 that allows us to reconstruct and parameterise the fronto-parallel dynamics of articulated leg motion.

Projection of the planarized human motion representation into the image is achieved by a parameterised set of planar homography transformations. We detail two different reconstruction parameterisations that model human gait. Consequently, two methods are presented to compute the maximum likelihood estimates of the set of reconstruction parameters. Both of these reconstruction parameterisations are then compared to see which performs best. An analysis of the resultant reconstructions over a number of

different subject trajectories is given for a small trial set of four people. Finally, we conclude the chapter with a brief discussion on a number of possible improvements and considerations that the author did not have the time to look into further.

4.2 A stratified approach to linear trajectory gait reconstruction

The effect of varying a subject's trajectory is quite pronounced on the set of measured limb angles. Figure 4.1 shows a single position of gait pose that is imaged from three different camera viewpoints. The cameras are synchronized so that the underlying limb motion is sampled consistently within each view.

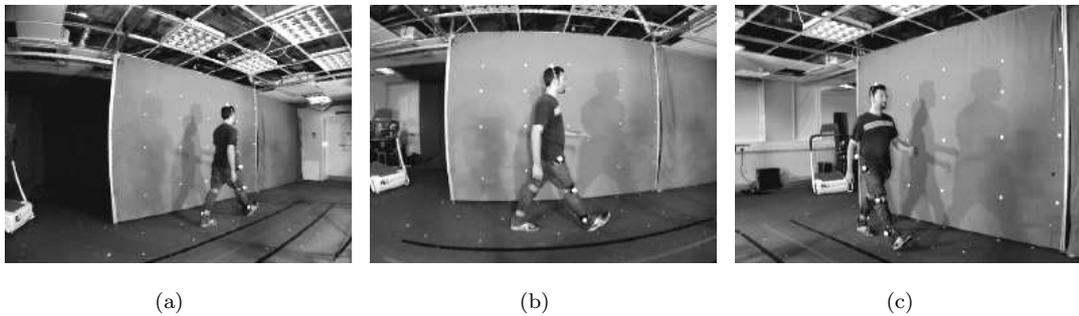


FIGURE 4.1: Three synchronized views of a subject walking along a linear trajectory with constant velocity.

Figures 4.2 and 4.3 show the corresponding set of extracted leg angles after removing the camera calibration. Since the camera distortion is removed and the gait dynamics are consistent in all three image sequences, the only set of parameters that differ between views are the camera extrinsic parameters, or analogously the subject motion trajectory. The extracted set of leg angles appear vary different. Clearly some form of correction is required to remove the differences between subject poses.

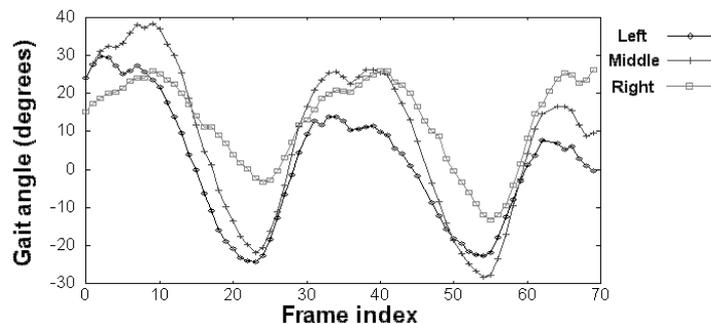


FIGURE 4.2: **Extracted upper leg angles.** Comparison of the extracted limb angle plots from all three views after removing the camera calibration.

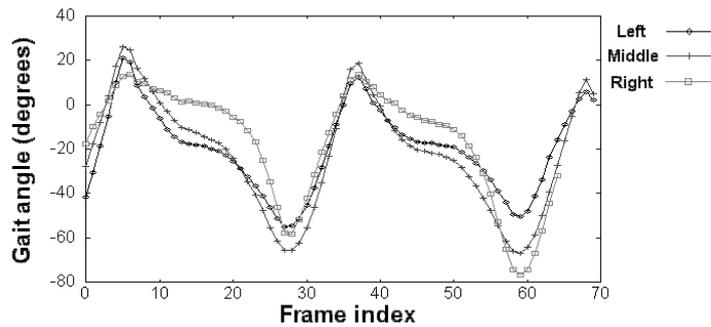


FIGURE 4.3: **Extracted lower leg angles.** Comparison of the extracted limb angle plots from all three views after removing the camera calibration.

Gait can be modelled by using a cardboard person assumption, with articulated limb motion contained within three similarly spaced parallel planes. The central mid-plane contains the head and torso sections, while the bilateral planes contain the left and right legs. The arms may also be modelled by adding a further two planes. However, the independence and large freedom of motion available to the arms precludes their use as any meaningful repeating motion pattern. The shape of the mid-plane body parts remain relatively static throughout an image sequence and can be modelled by translational motion, with displacement components containing linear velocity in X and sinusoidal oscillation in Y. Both bilateral left and right leg swing planes contain dynamic articulated limb motion. The set of rigid leg segment poses are related by both translation and rotation transformations throughout the image sequence. Figure 4.4 shows this arrangement of cardboard motion planes used to model subject gait.

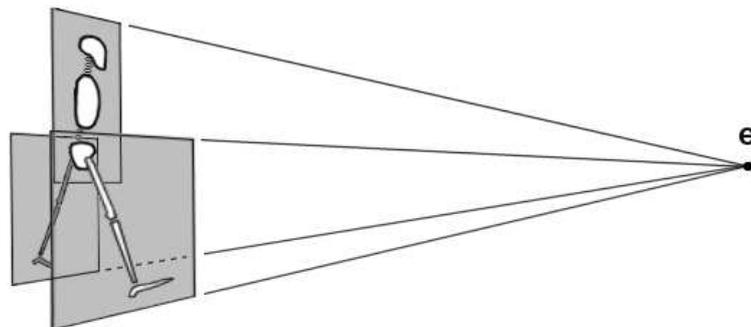


FIGURE 4.4: Cardboard human motion model containing three similarly spaced, parallel limb planes.

Since all body part planes are parallel within the worldspace, they share a common vanishing line l'_∞ in the image. Similarly, the progressional component of gait motion is common to all body parts over the image sequence. The imaged direction of motion e is then common to all limb motion planes within the image, thus also lies on the vanishing line l'_∞ .

In order to reconstruct the canonical dynamics of gait, we need to identify the 2D ho-

mography mappings that transform structure from the imaged limb swing planes back to the fronto-parallel motion plane. The stratification process of computing this rectification transformation can be split into three stages: \mathbf{H}_p , \mathbf{H}_a and \mathbf{H}_s . Each stage of the rectification removes a number of degrees of freedom (d.o.f.) from a total of eight that completely parameterises a planar homography. Identification of properties and invariants within each particular class of projective transformation enables us to remove the associated distortion effects. A more detailed discussion on the properties and transformation characteristics of these transformation classes can be found in appendix A.1.

The proposed algorithm uses planar projective geometry on the imaged entities of gait motion alone, in order to compute a stratified reconstruction that recovers the canonical view of the underlying limb motion. Subsequently, we do not need to know any prior information about the camera calibration or the pose of a subject. We use only our knowledge, that human limb segments have an unknown but fixed length throughout the captured motion sequence, as a source of constraints for the reconstruction process.

We first compute the projective transformation \mathbf{H}_p that maps the imaged vanishing line \mathbf{l}'_∞ of the limb swing plane back to its canonical position $(0, 0, 1)^\top$. Structure on the rectified plane is then defined up to an affine ambiguity (6 d.o.f.). Computation of the affine transformation \mathbf{H}_a is formed from Liebowitz's known ratio of length constraints [61], on the set of rigid articulated limb body segments. Application of \mathbf{H}_a restores angles and length ratios of the metric plane. The combined set of transformations $\mathbf{H}' = \mathbf{H}_a \mathbf{H}_p$ effectively map the imaged circular points \mathbf{I}', \mathbf{J}' back to their canonical positions $(1, \pm i, 0)^\top$, thus structure on the metric plane is defined up to a similarity transformation (4 d.o.f.). Since the circular points are invariant to similarity transformations, we are free to choose any particular scale and orientation of the subject on the metric plane. We choose to rectify the subject motion to a consistent coordinate system via the similarity transformation \mathbf{H}_s , such that subject gait is left to right, trajectory aligned with the X axis, sky upward and the initial subject pose positioned at the origin. We remove the scale ambiguity of the subject by constraining the upper leg limb segment to have unit length. The set of stratified transformations then define the complete 2D homography mapping $\mathbf{H} = \mathbf{H}_s \mathbf{H}_a \mathbf{H}_p$ that restores imaged structure back to the canonical coordinate frame.

4.2.1 Periodicity and the imaged direction of motion

Knowing that an object's motion is periodic is a strong cue for action recognition [52, 53, 37]. Furthermore, the periodic motion of people can be used to recognize individuals [65]. Methods for detecting periodicity can be categorized into those requiring point correspondences [108, 91], those analysing periodicities of pixels [66, 67, 86], those analysing features of periodic motion [79, 80, 35, 46] and those analysing the periodicities of object similarities [20, 21, 22, 91]. Here, we assume that limb landmark points

have already been found, thus the task of finding the subject's periodicity is somewhat simpler. Realistically, periodicity is a good cue for both subject classification and salient feature detection. Computation of landmark points and subject periodicity may be best solved simultaneously. We acknowledge that the segmentation, tracking and classification aspect of human motion is a difficult problem and a further research topic in its own right. A full treatment of periodicity and human motion segmentation is beyond the scope of this project. We describe here a minimal solution required to solve the periodicity and correspondence problems.

One point that lies on the vanishing line l'_∞ of the limb swing plane is the imaged direction of motion \mathbf{e} . Multiple periods of linear gait motion is analogous to a single period viewed from many cameras that are related by linear translation. The pure translational nature of the camera geometry leads to an auto-epipolar configuration, as shown in figure 4.5.

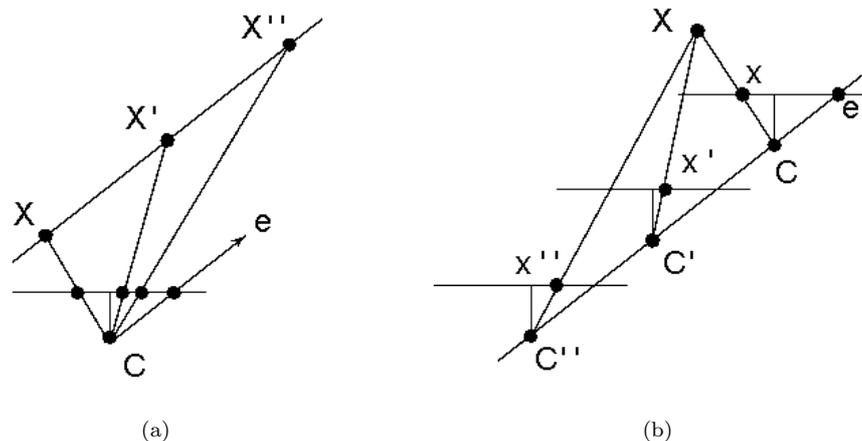


FIGURE 4.5: **Duality of multiple view geometry and periods of gait motion.** (a) Single camera view of multiple points that represent similar pose positions within a number of gait cycles. (b) Multiple cameras related by linear translation that image a single pose point within an assumed single cycle of gait.

Matching landmark correspondence points, at integer multiples of the gait period T , lie in an auto-epipolar configuration with the imaged motion direction \mathbf{e} . Figure 4.6 shows four poses within an image sequence, with each pose taken at half period intervals. We can clearly see the correspondence matches between every other pose and the half phase bilateral symmetry of adjacent poses between each side of the body.

In order to compute the landmark correspondence matches we need to know the gait cycle period T . Since this period is unknown, we must simultaneously determine both the landmark correspondences and subject periodicity by accumulation of the self-similarity error metric over a range of suitable putative periodicities.

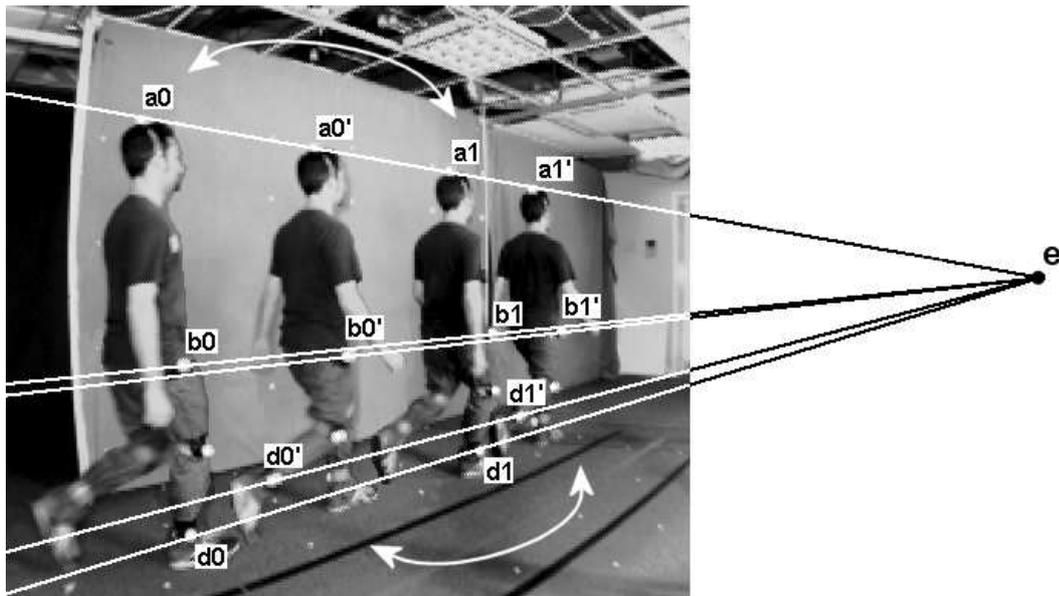


FIGURE 4.6: Similar phase poses of imaged subject motion. Landmark points $\mathbf{a} - \mathbf{d}$ shown with matching pose positions. Each landmark point has T phase positions $\mathbf{a}, \mathbf{a}', \dots, \mathbf{a}^*$, with each set of matching phase clusters $(\mathbf{a}_0, \dots, \mathbf{a}_n), (\mathbf{a}'_0, \dots, \mathbf{a}'_n), \dots$ in auto-epipolar configuration with the motion epipole \mathbf{e} .

4.2.1.1 Self-similarity error metric

The work presented in section 2.8.1 showed that the periodicity of planar limb motion is analogous to the specialized geometry of repeating planar patterns. Subsequently, the knowledge of landmark point correspondences across similar positions of gait phase allows us to recover the imaged direction of motion and the vanishing line of the limb swing plane. This then enables us to compute the transformation \mathbf{H}_p that replaces the imaged vanishing line back to its canonical position, thus recovering the affine properties of the plane. Since these affine properties are invariant to changes in displacement, the apparent translation between matching limb poses within different gait cycles has no effect on the self-similarity between corresponding limb angles. The squared residual fitting error, between all recovered matching limb angle poses over the image sequence, is then a good cost functional that describes the similarity between subject poses.

Since the nature of the motion is auto-epipolar $\mathbf{F} = [\mathbf{e}]_x$, each set of similar pose landmark points formed from matches at integer multiples of the periodicity defines a *cluster* of point correspondences. For example, the points $(\mathbf{a}_0, \dots, \mathbf{a}_n)$ within figure 4.6 define a single cluster of correspondence matches. A cluster of point correspondences must all lie on a common epipolar line, thus a robust estimate is computed by fitting all points within this cluster to a single line, via a process of orthogonal regression. Further details of line fitting and orthogonal regression can be found in appendix B.1.

Each landmark point then has T (frames) different phase positions $(\mathbf{a}, \mathbf{a}', \dots, \mathbf{a}^*)$ of articulated limb pose within a gait cycle. Consequently, each phase position has a corre-

sponding cluster of point matches over all the imaged motion periods of gait, with each cluster collinear with the motion epipole \mathbf{e} . This argument applies to all the landmark points $\mathbf{a} - \mathbf{d}$ within the image sequence of subject motion, and provides an initial means for identifying the unknown motion epipole \mathbf{e} . Figure 4.7 shows the relationship between subject pose positions and matching landmark clusters within the image sequence.

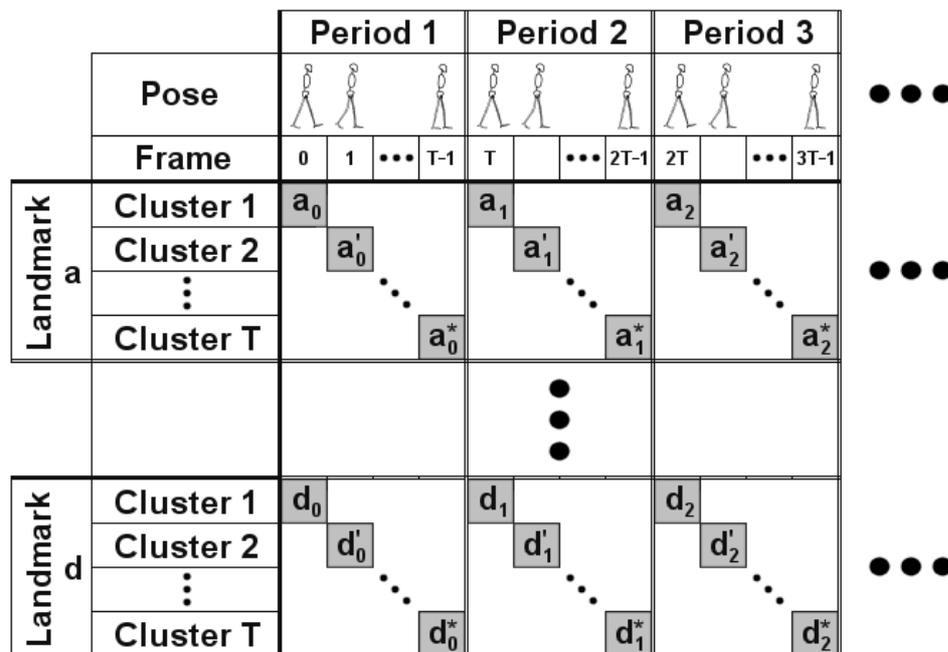


FIGURE 4.7: Relationship of cluster point correspondences within the image sequence.

For each putative value of periodicity T , we compute the set of epipolar lines \mathbf{l}_i through the supposed landmark point clusters. The combined set of fitted epipolar lines must pass through the imaged direction of motion \mathbf{e} , hence an estimate for the epipole \mathbf{e} is obtained by solving the homogeneous set of linear equations $\mathbf{l}_i^\top \mathbf{e} = 0$.

We then apply the stereopsis transformation \mathbf{H}_e that maps the imaged epipole \mathbf{e} to the ideal point $(1, 0, 0)^\top$. An inappropriate choice of homography mapping may cause severe projective distortion of the image. We can insist that the transformation should act as far as possible like a rigid motion transformation (rotation and translation only), in the neighbourhood of any selected point $\mathbf{x}_c = (u, v)^\top$ within the image. The chosen point \mathbf{x}_c is selected as the centroid of all landmark points over the image sequence. Further details of the stereopsis transformation were given in section 2.6.3. Having replaced the epipole back to its canonical position, the conjugate translation \mathbf{M} that maps repeated planar patterns within the image now has the simplified form shown in equation 4.2.

$$\mathbf{M} = \mathcal{I} + \lambda \mathbf{v} \mathbf{l}'_{\infty}{}^{\top} \quad (4.1)$$

$$\mathbf{M} = \begin{pmatrix} a & b & c \\ & a & \\ & & a \end{pmatrix} \quad (4.2)$$

After applying the stereopsis transformation, all epipolar lines are aligned parallel with the X axis. The transformed vanishing line of the limb planes is also aligned parallel to the X axis and has the form $\mathbf{l}'_{\infty} = (0, b, c)^{\top}$. All similar landmark point correspondences within a cluster then have the same y coordinate $\mathbf{x}_k = (u_k, v)^{\top}$, which is determined by computing the mean \bar{y} of the cluster of points. Length ratios on each epipolar line corresponding to a set of repeated points are preserved, thus distances between consecutive points $\Delta u = (u_{k-1} - u_k)$ are equal. Constraints formed from all combinations of similar point correspondences with common y coordinates have been shown within section 2.8.1 to be the same. Given more than two correspondence periods, a least squares estimate of Δu may be computed. An inhomogeneous solution can be formed to solve the linear displacement function $k \cdot \Delta u + u_0 = u_k$ of the set of repeating points on the epipolar line. A system of equations of the form $\mathbf{A}\mathbf{x} = \mathbf{b}$ is then generated by stacking the constraints.

$$\begin{bmatrix} k & 1 \end{bmatrix} \begin{pmatrix} \Delta u \\ u_0 \end{pmatrix} = u_k \quad (4.3)$$

One constraint equation on the coefficients of the conjugate translation \mathbf{M} can then be computed for each cluster of repeated point correspondences.

$$\begin{bmatrix} \Delta u & v & 1 \end{bmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \mathbf{0} \quad (4.4)$$

The 3D motion structure of a subject's legs is approximated by repeated planar motion patterns within two separate planes. The coefficients of both leg plane conjugate translations \mathbf{M} and \mathbf{M}' must then be solved by computing both direct linear transformations.

$$[\mathbf{x}_j]_{\times} \mathbf{M}^j \mathbf{x}_0 = \mathbf{0} \quad (4.5)$$

$$[\mathbf{x}'_k]_{\times} \mathbf{M}'^k \mathbf{x}'_0 = \mathbf{0} \quad (4.6)$$

However, the repeating limb plane transformations share a common vanishing line $\mathbf{l}'_{\infty} = (0, b, c)^{\top}$. The coefficients of both independent conjugate translations \mathbf{M} and \mathbf{M}' can be combined into a single set of simultaneous linear equations of the form shown in equation 4.7.

$$\begin{bmatrix} \Delta u & 0 & v & 1 \\ 0 & \Delta u' & v' & 1 \end{bmatrix} \begin{pmatrix} a \\ a' \\ b \\ c \end{pmatrix} = \mathbf{0} \quad (4.7)$$

A minimal solution to this homogeneous set of equations can be obtained from three repeated point constraints, each with different y coordinates and at least one constraint taken from each limb plane. Having computed the elements of the vanishing line $\mathbf{l}'_{\infty} = (0, b, c)^{\top}$, a perspective transformation \mathbf{H}_1 can be computed that maps this line back to its canonical position $\mathbf{l}_{\infty} = (0, 0, 1)^{\top}$.

$$\mathbf{H}_1 = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & b & c \end{pmatrix} \quad (4.8)$$

The combined perspective homography $\mathbf{H}_p = \mathbf{H}_1 \mathbf{H}_e$ then recovers the affine properties of all parallel motion planes. Figure 4.8 shows the recovered affine motion structure corresponding to the image sequence shown in figure 4.6.

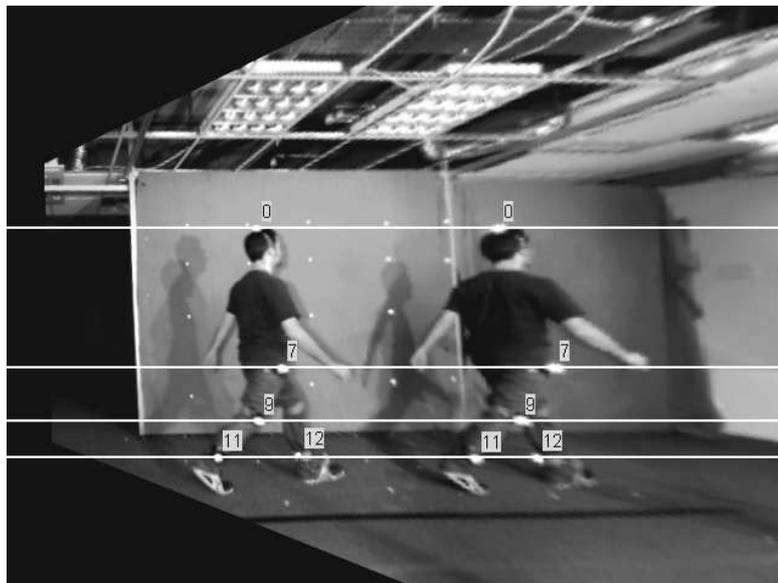


FIGURE 4.8: Rectified image of similar poses by applying the perspective transformation $\mathbf{H}_p = \mathbf{H}_1 \mathbf{H}_e$ that recovers the affine properties of all parallel motion planes within the cardboard person model. After transformation, all epipolar lines are aligned parallel to the X axis.

Although metric properties such as angles are not recovered by the transformation \mathbf{H}_p , structure is now similar across all repeated gait poses. The self-similarity of the affinely reconstructed limb pose vectors \mathbf{p}_i between landmark points is then a good cue for periodicity detection.

For each position of phase within the gait cycle, we compute the corresponding set of leg segment direction vectors \mathbf{p}_i over all periods of affinely recovered motion, i.e. \mathbf{p}_i are the set of direction vectors formed from the recovered limb angles from the i^{th} period. Each pose vector is normalized to unit length $\|\mathbf{p}_i\| = 1$ so that all poses are weighted equally. We can then determine an orthogonal vector \mathbf{v} to this set of pose unit vectors, that minimizes the residual distances $d_{\perp} = \mathbf{p}_i^{\top} \mathbf{v}$. The least squares residual cost error \mathcal{C} , corresponding to the sum of squared inner products between the set of similar limb poses \mathbf{p}_i and the orthogonal vector \mathbf{v} , then has the form.

$$\mathcal{C} = \sum_{i=1}^n (\mathbf{p}_i^{\top} \mathbf{v})^2 \quad (4.9)$$

$$\mathcal{C} = \mathbf{v}^{\top} \left(\sum_{i=1}^n \mathbf{p}_i \mathbf{p}_i^{\top} \right) \mathbf{v} \quad (4.10)$$

$$\mathcal{C} = \mathbf{v}^{\top} \mathbf{M} \mathbf{v} \quad (4.11)$$

We can then make a symmetric Eigen-decomposition of the moment matrix \mathbf{M} via the substitution $\mathbf{M} \mathbf{v} = \lambda \mathbf{v}$. The moment matrix \mathbf{M} is symmetric, positive-definite hence all Eigenvalues are real and non-negative. The two Eigenvectors of the decomposition correspond to the columns of a rotation matrix, thus are orthogonal and of unit norm $\|\mathbf{v}\| = 1$. Correspondingly, we see that the residual cost error is given by $\mathcal{C} = \mathbf{v}^{\top} \lambda \mathbf{v} = \lambda$. The minimum of the least squares cost function occurs with least Eigenvalue λ , thus the Eigenvector corresponding to the smallest Eigenvalue is the solution we require for the best orthogonal vector \mathbf{v} .

The self-similarity error ϵ , corresponding to the chosen putative value of periodicity T , is then given by computing the root mean square of all T gait phase positions of residual fitting errors λ_k . We accumulate these self-similarity errors ϵ within a vector, corresponding to increasing values of putative periodicity.

$$\epsilon = \sqrt{\frac{1}{N} \sum_{k=1}^T (\lambda_k + \lambda'_k)} \quad (4.12)$$

Although only one measurement pose vector \mathbf{p}_1 is required to compute an orthogonal vector \mathbf{v} , we are not able to infer any meaningful information about the periodicity.

Subsequently, we classify any correspondence sets with only a single pose vector as invalid and do not include them in the self-similarity cost function. The correspondence errors are determined from the matching poses of both the upper and lower legs λ_k, λ'_k . There are then a total of N valid pose vectors contributing to the cost summation of the self-similarity error ϵ . Due to occlusion and missing data, the number of matches is not fixed over the range of putative periods. The root mean squared residual error ϵ , shown in equation 4.12, lies within the range $(0 : 1)$ and gives a good measure of how much the poses deviate from similarity (0 - similar, 1 - dissimilar).

4.2.1.2 Recovering periodicity from self-similarity

The vector of valid self-similarity cost errors is first smoothed by a $(1, 4, 6, 4, 1)$ filter. The errors within the smoothed vector are then normalized to lie in the range $(-1 : 1)$, by finding the means of the two largest and two smallest errors and applying the required translation and scaling transformations. The resulting normalized cost fitting curve is cyclic in nature with the same period as the subject's gait. We must also note, that for long vectors there will be many minima located at integer values of the fundamental period, of which the global minimum may not be the fundamental.

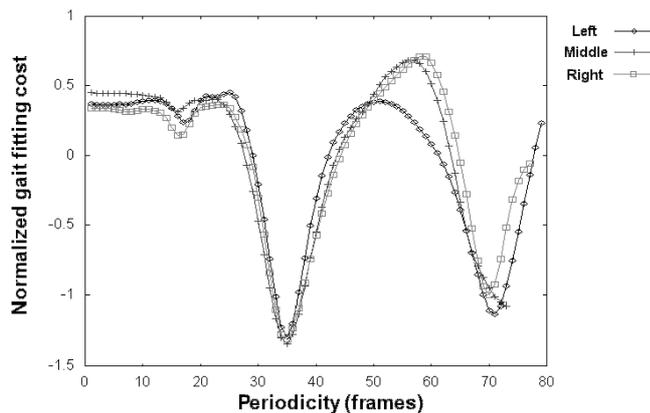


FIGURE 4.9: Normalized gait periodicity fitting cost vector. The true periodicity of gait is approximately 35 frames.

The vector of normalized fitting costs shown in figure 4.9 clearly shows two periods of gait motion. Even after smoothing there may still be many local minima within the head section of the cost vector, which occur as a result of poor fitting. Most poorly fitted values occur in both the head and tail segments of the cost vector, so the periodicity function is best computed by first multiplying the data vector by a Gaussian envelope and fitting the resulting data to a first order harmonic series with Gaussian envelope.

$$x(t) = A_1 \cdot e^{-\frac{(t-t_0)^2}{2\sigma^2}} \cdot \cos(2\pi f_0 t + \phi) \quad (4.13)$$

The envelope is centred about the middle of the cost vector $t_0 = N/2$. The width

constant σ is computed such that the envelope has a small empirical cut-off tolerance of $\lambda = 0.0001$ at the extents of the vector.

$$y(t) = e^{-\frac{(t-t_0)^2}{2\sigma^2}} \quad (4.14)$$

$$\lambda = e^{-\frac{(t_0)^2}{2\sigma^2}} \quad (4.15)$$

$$\ln(\lambda) = -\frac{(t_0)^2}{2\sigma^2} \quad (4.16)$$

$$\sigma = \frac{t_0}{\sqrt{-2\ln(\lambda)}} \quad (4.17)$$

Figure 4.10 shows the corresponding normalized gait periodicity fitting cost vector multiplied with a Gaussian envelope. The significance of the noisy cost errors within the head and tail of the vector is dramatically reduced.

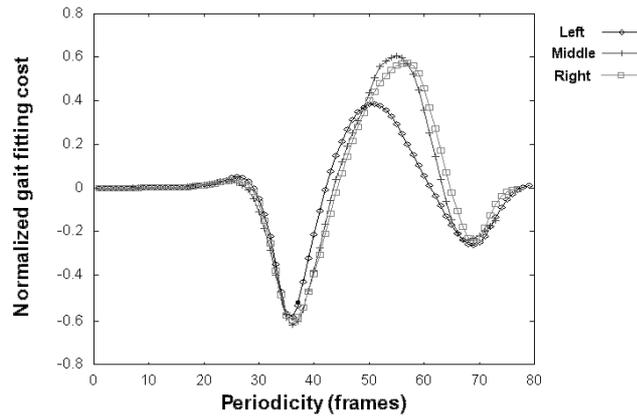


FIGURE 4.10: Normalized gait periodicity fitting cost vector multiplied with a Gaussian envelope.

Since the subject's period is the same as the normalized cost errors, determining the fundamental frequency f_0 of a first order harmonic series that fits the data then gives a good indication of the required periodicity. There are potentially many local minima within the parameter space of unknowns $\mathbf{P} = (f_0, A_1, \phi)^\top$, not all of which are located at integer multiples of the fundamental frequency. The best strategy is to accumulate the χ^2 residual fitting errors of equation 4.13 over a range of putative periodicities T (frames), while minimizing the set of subsidiary parameters $\mathbf{P} = (A_1, \phi)^\top$.

For each putative value of periodicity T , we first initialise the fundamental frequency $f_0 = 1/T$ and the set of subsidiary parameters \mathbf{P} . Since we have normalized the data to the range $(-1 : 1)$, we can set the A_1 parameter to unity. We then compute the residual χ^2 errors at three sample phase positions $(-\frac{2\pi}{3}, 0, \frac{2\pi}{3})$ and set the initial ϕ parameter to the phase associated with the smallest residual error. Minimization is then

performed on the subsidiary parameters $\mathbf{P} = (A_1, \phi)^\top$ using the Levenberg-Marquardt method with a maximum of 10 iterations. The computed χ^2 residual error and the corresponding subsidiary parameters \mathbf{P} are then added to the periodicity accumulation vector. Once all residual fitting errors have been computed within the vector of putative periods, the initial frequency estimate f_0 is determined by the periodicity value corresponding to the smallest residual error.

An optimal solution may then be given by minimization until convergence of all three parameter estimates $\mathbf{P} = (f_0, A_1, \phi)^\top$, by using the Levenberg-Marquardt method. This boot strapping technique robustly removes the potential of encountering most, if not all false minima within the parameter space by determining the set of initial parameter estimates that lie sufficiently close to the true values.

The last optimization step may be considered optional, since the fitting process of a simple harmonic series to the data is designed only to determine which of the local minima corresponds to the fundamental periodicity of gait. The accuracy of determining f_0 is far in excess of identifying the required local minimum within the normalized cost error vector.

The normalized periodicity cost errors shown in figure 4.9 are based on the geometric properties of subject self-similarity. Fitting a simple harmonic series to these cost errors reliably determines which of the local minima corresponds to the fundamental periodicity of gait. Finding this minimum within the cost vector then gives an accurate estimate of the true periodicity of gait. As such, we can use the calculated periodicity value $\tilde{T} = 1/f_0$ from the minimization of equation 4.13 as an initial starting point to find the corresponding local 1D minimum within the normalized cost vector. The true gait periodicity T is found to sub-temporal accuracy by performing gradient descent then fitting a quadratic curve to the data at the local minimum.

4.2.2 Recovering affine structure

Since we have found the periodicity of gait to sub-time sample accuracy, the tracked landmark points are in general not aligned with frame boundaries, at integer multiples of the period T . Correspondingly, we must interpolate the set of landmark tracks by assuming linear velocity between consecutive frames. From these interpolated correspondence points of similar gait pose, we recompute the motion epipole \mathbf{e} consistent with the epipolar lines \mathbf{l}_i formed through each of the landmark point clusters.

We first apply the normalization transformation \mathbf{K}_n that best maps the set of landmark point tracks to the unit square with isotropic scaling, $\mathbf{x}' = \mathbf{K}_n \mathbf{x}$. We then transform the motion epipole $\mathbf{e}' = \mathbf{K}_n \mathbf{e}$ and re-normalize to unit norm $\|\mathbf{e}'\| = 1$.

$$\mathbf{K}_n = \begin{pmatrix} s & t_x \\ & s & t_y \\ & & 1 \end{pmatrix} \quad (4.18)$$

For each correspondence set of landmark point clusters, we then proceed to compute the optimal point estimates $\hat{\mathbf{x}}_i$ that lie on the epipolar line satisfying the condition $\hat{\mathbf{x}}_i^\top [\mathbf{e}']_\times \hat{\mathbf{x}}_j = 0$. We first fit the epipolar line corresponding to a landmark point cluster by using orthogonal regression, with the constraint that this line passes through the motion epipole \mathbf{e}' . Details of line fitting and constrained orthogonal regression are described within appendix B.1.1. Having computed the epipolar line consistent with the auto-epipolar geometry, we can then find the optimal set of points $\hat{\mathbf{x}}_i$ that lie on this line, closest to each of the corresponding imaged landmark points \mathbf{x}'_i . Further details of orthogonal projection of points onto a line can be found in appendix B.1.2.

From the principle of duality between multiple views and periods of imaged gait motion, see figure 4.5, the back projected rays formed from a set of optimal point estimates $\hat{\mathbf{x}}_i$ intersect in a single worldspace point \mathbf{X} . These optimal point estimates lie on the plane π that passes through the set of camera centres and the imaged epipolar line consistent within all views, as illustrated in figure 4.11.

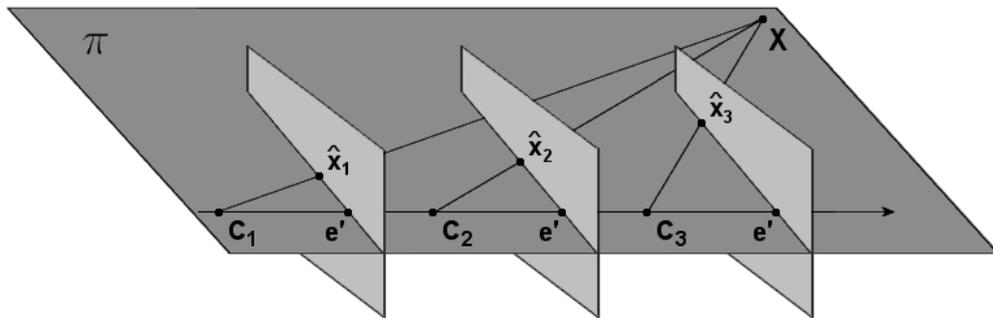


FIGURE 4.11: Back projection of the set of optimal point estimates, that lie on the plane which passes through all camera centres and the corresponding imaged epipolar line, meets in a single worldspace point \mathbf{X} .

Back projection of all corresponding landmark point clusters then generates the set of 3D point tracks for an assumed single period of reconstructed gait motion. We use the Direct Linear Transform (DLT) to triangulate each of the worldspace points \mathbf{X} .

$$([\hat{\mathbf{x}}_k]_\times \mathbf{P}_k) \cdot \mathbf{X} = 0 \quad (4.19)$$

with the set of camera projection matrices:

$$\mathbf{P}_k = [\mathbf{R}_e^\top \mid -k \cdot \mathbf{e}'] \quad (4.20)$$

where $\hat{\mathbf{x}}_k$ is the image of the worldspace point \mathbf{X} in the k^{th} period image, \mathbf{R}_e is the 3×3 rotation matrix that aligns the epipolar vector \mathbf{e}' with the X axis, and k is an integer describing the camera periodicity translation.

We can now apply the assumption that articulated leg motion is approximately planar, and proceed to fit the set of recovered 3D limb points to a set of planes. Since we have aligned the epipolar vector \mathbf{e}' with the X axis, one such point that must lie on each of the worldspace planes is the ideal point $(1, 0, 0, 0)^\top$. The pencil of planes that intersect this ideal point have the form $\pi = (0, v_2, v_3, v_4)^\top$, hence the problem reduces to that of finding two lines within the YZ plane cross section data.

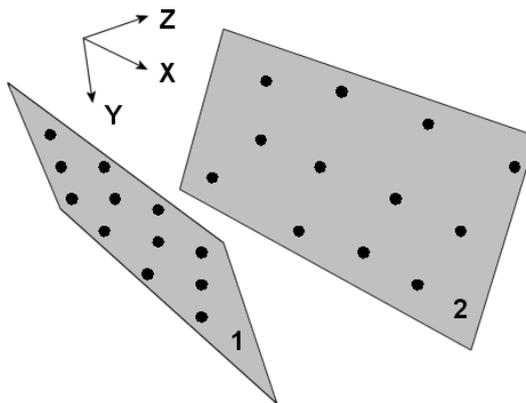


FIGURE 4.12: Worldspace limb swing planes. The YZ cross section plane contains two lines $\mathbf{l}_1, \mathbf{l}_2$ of data points.

We evaluate the mean $(\bar{y}, \bar{z})^\top$ of the cross section point distribution and apply a translation \mathbf{H}_t that maps this point to the origin. The two cross section plane lines $\mathbf{l}_1, \mathbf{l}_2$ are then each computed by orthogonal regression, see appendix B.1 for details. The intersection point \mathbf{u} of the two lines, given by the cross product $\mathbf{u} = \mathbf{l}_1 \times \mathbf{l}_2$, is then aligned with the positive Y axis by applying a rotation \mathbf{H}_r . Consequently, the pair of transformed lines are mapped to $\mathbf{l}'_i = \mathbf{H}_r^{-\top} \mathbf{l}_i = \mathbf{H}_r \mathbf{l}_i$. The transformed point \mathbf{u}' is then mapped to the ideal point $(1, 0, 0)^\top$ by application of the perspective transformation \mathbf{H}_α .

$$\mathbf{H}_\alpha = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \alpha & 0 & 1 \end{pmatrix} \quad (4.21)$$

Since \mathbf{u}' lies on the Y axis and has the form $(y, 0, w)^\top$ then the transformation $\mathbf{H}_\alpha \mathbf{u}'$ gives us $\alpha = -w/y$, and the corresponding line mapping $\mathbf{H}_\alpha^{-\top} \mathbf{l}'_i$ effectively zeros the first component of both line normals. The lines are now parallel and can be re-normalized such that $\mathbf{l}''_1 = (0, 1, -c_1)^\top$ and $\mathbf{l}''_2 = (0, 1, -c_2)^\top$ in order that we can find the points at which they cut the Z axis (c_1, c_2) . We then apply a further similarity transformation \mathbf{H}_s that translates the mid-point $(c_1 + c_2)/2$ to the origin and scales in the Z direction

to rectify the lines to the form $\mathbf{l} = (0, 1, \pm 1)^\top$. Application of a plane selection transformation \mathbf{H}_β then translates by ± 1 , mapping the selected set of points onto the $z = 0$ plane. The combined set of transformations then form the limb plane transformation $\mathbf{H}_\mathbf{v} = \mathbf{H}_\beta \mathbf{H}_\mathbf{s} \mathbf{H}_\alpha \mathbf{H}_\mathbf{r} \mathbf{H}_\mathbf{t}$. A similar set of transformations can be constructed that allows us to change the matrix order.

$$\begin{aligned}\mathbf{H}_\mathbf{v} &= \mathbf{H}_\beta \mathbf{H}_\alpha (\mathbf{H}_\alpha^{-1} \mathbf{H}_\mathbf{s} \mathbf{H}_\alpha) \mathbf{H}_\mathbf{r} \mathbf{H}_\mathbf{t} \\ \mathbf{H}_\mathbf{v} &= \mathbf{H}_\beta \mathbf{H}_\alpha \mathbf{H}_\mathbf{s}' \mathbf{H}_\mathbf{r} \mathbf{H}_\mathbf{t}\end{aligned}\quad (4.22)$$

The projection transformation mapping the back projected workspace points \mathbf{X} into the image can then be decomposed into block form:

$$\hat{\mathbf{x}}(k) = [\mathbf{R}_\mathbf{e}^\top | -k \cdot \mathbf{e}'] \begin{pmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{H}_\mathbf{v}^{-1} \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{H}_\mathbf{v} \end{pmatrix} \mathbf{X} \quad (4.23)$$

$$\hat{\mathbf{x}}(k) = [\mathbf{R}_\mathbf{e}^\top | -k \cdot \mathbf{e}'] \tilde{\mathbf{H}}_\mathbf{v}^{-1} \mathbf{W} \quad (4.24)$$

Where \mathbf{W} is the transformed workspace point on the $z = 0$ plane and the augmented 4×4 matrix $\tilde{\mathbf{H}}_\mathbf{v}^{-1}$ has the form:

$$\tilde{\mathbf{H}}_\mathbf{v}^{-1} = \left[\begin{array}{c|ccc} 1 & \mathbf{m}_2 & \mathbf{m}_3 & \mathbf{m}_4 \\ \hline \mathbf{0} & 0 & 0 & 1 \end{array} \right] \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -\beta \\ 0 & -\alpha & 0 & 1 \end{pmatrix} \quad (4.25)$$

The corresponding projection transformation of workspace points $\mathbf{W} = (u, v, 0, w)^\top$ into the image is then given by the 3×3 homography mapping $\hat{\mathbf{x}} = \mathbf{H}_\mathbf{p} \cdot (u, v, w)^\top$.

$$\mathbf{H}_\mathbf{p} = \left[\begin{array}{ccc} \mathbf{e}' & \mathbf{m}'_2 - \alpha \cdot (\mathbf{m}'_4 - k \cdot \mathbf{e}') & (\mathbf{m}'_4 - k \cdot \mathbf{e}') - \beta \cdot \mathbf{m}'_3 \end{array} \right] \quad (4.26)$$

where $\mathbf{m}'_i = \mathbf{R}_\mathbf{e}^\top \mathbf{m}_i$ and $\mathbf{e}' = \mathbf{R}_\mathbf{e}^\top \cdot (1, 0, 0)^\top$

We finally find both sets of optimal $z = 0$ plane points by solution of the planar Direct Linear Transform for each point $\hat{\mathbf{u}}$, in order to minimize image reprojection error.

$$([\hat{\mathbf{x}}_{k,\beta}]_\times \mathbf{H}_\mathbf{p}(k, \beta)) \cdot \hat{\mathbf{u}} = \mathbf{0} \quad (4.27)$$

The individual homography matrices that map structure from the $z = 0$ plane to their corresponding imaged leg swing planes are given by setting $k = 0$ and $\beta = \mp 1$ within equation 4.26.

$$\mathbf{H}_{\mathbf{p}1} = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 \end{bmatrix} \quad (4.28)$$

$$\mathbf{H}_{\mathbf{p}2} = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}'_3 \end{bmatrix} \quad (4.29)$$

In practice, fitting the YZ cross section data to two individual lines can be unreliable due to the effect of resampling and measurement noise within the back projection process. Since both worldspace leg swing planes are approximately parallel, both sets of cross section data points are best fitted to two lines with a common normal $\mathbf{N} = (a, b)^\top$. The sum of squares fitting cost function \mathcal{C} of orthogonal regression, corresponding to both parallel lines then has the form.

$$\mathcal{C} = \mathbf{N}^\top \mathbf{M}_1 \mathbf{N} + \mathbf{N}^\top \mathbf{M}_2 \mathbf{N} \quad (4.30)$$

$$\mathcal{C} = \mathbf{N}^\top (\mathbf{M}_1 + \mathbf{M}_2) \mathbf{N} \quad (4.31)$$

$$\mathcal{C} = \mathbf{N}^\top \mathbf{M}_{12} \mathbf{N} \quad (4.32)$$

The 2×2 moment matrix \mathbf{M} is generated from the set of component differences $\Delta \mathbf{x}_i$ between the data points \mathbf{x}_i and the corresponding centroid $\bar{\mathbf{x}}$ of the distribution. Further details are given in the discussion on orthogonal regression within appendix B.1.

$$\mathbf{M} = \sum_{i=1}^n \Delta \mathbf{x}_i \cdot \Delta \mathbf{x}_i^\top \quad \text{where} \quad \Delta \mathbf{x}_i = \mathbf{x}_i - \bar{\mathbf{x}} \quad (4.33)$$

We can then make a symmetric Eigen-decomposition of the combined moment matrix \mathbf{M}_{12} via the substitution $\mathbf{M}_{12} \cdot \mathbf{N} = \lambda \mathbf{N}$, and note that $\mathcal{C} = \mathbf{N}^\top \lambda \mathbf{N} = \lambda$ gives the cost error corresponding to both point distributions. The minimum of this least squares cost function occurs with least Eigenvalue λ , thus the Eigenvector corresponding to the smallest Eigenvalue is the solution we require for the normal vector \mathbf{N} of both fitted lines.

Subsequently, the perspective homographies $\mathbf{H}_{\mathbf{p}}$ that transform points on the $z = 0$ plane to the images of the limb swing planes then have a much simpler form, where $k = 0$ and $\alpha = 0$ in equation 4.26.

$$\mathbf{H}_p = \begin{bmatrix} \mathbf{e}' & \mathbf{m}'_2 & (\mathbf{m}'_4 - \beta \cdot \mathbf{m}'_3) \end{bmatrix} \quad (4.34)$$

4.2.3 Recovering metric structure

Structure on the $z = 0$ plane has been recovered up to an affine ambiguity. We need to find the affine transformation \mathbf{H}_a that maps the imaged circular points $(1, \mu \pm i \cdot \lambda, 0)^\top$ back to their canonical positions $(1, \pm i, 0)^\top$.

$$\mathbf{H}_\mu = \begin{pmatrix} 1 & 0 & 0 \\ -\mu/\lambda & 1/\lambda & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (4.35)$$

We can recover metric structure on the plane by using the known ratios of lengths [62] between articulated limb landmark points over all reconstructed frame poses. The skeletal structure is rigid, hence the length ratio of a limb segment should remain fixed (unity) over all frames. The squared distance between any two limb segment endpoints $\mathbf{x}_1 = (u_1, v_1)^\top$ and $\mathbf{x}_2 = (u_2, v_2)^\top$ can be written as the inner product $d^2 = \Delta \mathbf{x}^\top \Delta \mathbf{x}$, where $\Delta \mathbf{x} = (u_2 - u_1, v_2 - v_1)^\top$ is the endpoint difference vector. If $\Delta \mathbf{x}$ and $\Delta \mathbf{x}'$ are the endpoint difference vectors for a corresponding limb segment within two different frames, then an affine transformation \mathbf{H}_μ can be computed that restores the metric properties of the plane. Since lengths between limb segment end points are invariant to translations, we need only consider the upper-left 2×2 sub-matrix \mathbf{H} of the affine transformation matrix \mathbf{H}_μ .

$$\Delta \mathbf{x}'^\top \mathbf{H}^\top \mathbf{H} \Delta \mathbf{x}' = \Delta \mathbf{x}^\top \mathbf{H}^\top \mathbf{H} \Delta \mathbf{x} \quad (4.36)$$

If we write the endpoint difference vectors as $\Delta \mathbf{x} = (\delta x, \delta y)^\top$, $\Delta \mathbf{x}' = (\delta x', \delta y')^\top$ and the elements of the symmetric 2×2 matrix $\mathbf{M} = \mathbf{H}^\top \mathbf{H}$ as $\mathbf{m} = (\mathbf{M}_{11}, \mathbf{M}_{12}, \mathbf{M}_{22})^\top$ then the set of linear constraints on \mathbf{m} can be written.

$$\begin{bmatrix} (\delta x^2 - \delta x'^2) & 2(\delta x^2 \delta y^2 - \delta x'^2 \delta y'^2) & (\delta y^2 - \delta y'^2) \end{bmatrix} \mathbf{m} = \mathbf{0} \quad (4.37)$$

Since \mathbf{m} is defined up to scale (2 d.o.f.) then a minimum of two such corresponding pose constraints are required to fully determine \mathbf{m} . We stack all constraints formed from all known length ratio correspondences and solve the system of homogeneous equations of the form $\mathbf{A} \mathbf{x} = \mathbf{0}$ through singular value decomposition, see appendix B.2.3. The rectification matrix \mathbf{H}_μ is then formed from the extracted parameters of $\mathbf{H}^\top \mathbf{H}$.

$$\mathbf{H}^\top \mathbf{H} = \xi \begin{pmatrix} 1 + (\mu/\lambda)^2 & -\mu/\lambda^2 \\ -\mu/\lambda^2 & 1/\lambda^2 \end{pmatrix} = \begin{pmatrix} m_1 & m_2 \\ m_2 & m_3 \end{pmatrix} \quad (4.38)$$

$$\mu = -\frac{m_2}{m_3} \quad (4.39)$$

$$\lambda = \sqrt{\frac{m_1}{m_3} - \mu^2} \quad (4.40)$$

The ideal epipole $(1, 0, 0)^\top$ is mapped by \mathbf{H}_μ to $(1, -\mu/\lambda, 0)^\top$ so we must also apply a rotation \mathbf{H}_r to align the epipole back along the X axis, such that $\mathbf{H}_a = \mathbf{H}_r \mathbf{H}_\mu$ is the required affine transformation that recovers metric angles and length ratios on both limb planes. Since we have ensured that the ideal epipole remains fixed, the transformation \mathbf{H}_a is upper triangular and correspondingly so is its inverse \mathbf{H}_a^{-1} . Points on the metric plane $\hat{\mathbf{w}}$ are then mapped into the image as:

$$\begin{aligned} \hat{\mathbf{x}} &= \mathbf{H}_p \mathbf{H}_a^{-1} (\mathbf{H}_a \hat{\mathbf{u}}) \\ \hat{\mathbf{x}} &= \mathbf{H}_p \mathbf{H}_a^{-1} \hat{\mathbf{w}} \end{aligned} \quad (4.41)$$

Writing \mathbf{p}_i as the columns of \mathbf{H}_p and the coefficients (a, b, c) of the affine matrix \mathbf{H}_a^{-1} , then the set of homography mapping matrices \mathbf{H}_q , that project points on the metric plane to the images of the limb swing planes, can be written:

$$\mathbf{H}_{q1} = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 \end{bmatrix} \begin{pmatrix} a & b \\ c & 1 \end{pmatrix} = \begin{bmatrix} a \cdot \mathbf{p}_1 & (b \cdot \mathbf{p}_1 + c \cdot \mathbf{p}_2) & \mathbf{p}_3 \end{bmatrix} \quad (4.42)$$

$$\mathbf{H}_{q2} = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}'_3 \end{bmatrix} \begin{pmatrix} a & b \\ c & 1 \end{pmatrix} = \begin{bmatrix} a \cdot \mathbf{p}_1 & (b \cdot \mathbf{p}_1 + c \cdot \mathbf{p}_2) & \mathbf{p}'_3 \end{bmatrix} \quad (4.43)$$

We are only interested in limb length ratios, so scalings are applied to both planes in order to transform each upper leg segment to unit length. In practice, many data points may be missing due to occlusion. Even in the ideal case where motion is fronto-parallel, the hip point on the occluded side of the body may never be imaged. To robustly compute the scaling transforms, we first compute \mathbf{H}_τ the scaling between both leg swing planes. We evaluate the mean set of limb lengths \mathbf{d} and \mathbf{d}' for both planes, along with the corresponding binary vectors \mathbf{v} and \mathbf{v}' (0 or 1 values) of validly computed

length flags. The set of mean leg segment lengths are then related by the inter-plane scaling $\mathbf{d}_i = \tau \cdot \mathbf{d}'_i$. A minimal solution to this trivial set of linear equations requires at least one valid mean length correspondence within the set of limb segments, i.e. the hip point on the occluded side of the body may never be seen, but the corresponding lower leg segments are visible for most of the sequence, thus enabling us to determine the inter-plane scaling factor τ . The solution to this set of inhomogeneous equations of the form $\mathbf{A}\mathbf{x} = \mathbf{b}$ is computed by forming the set of normal equations $\mathbf{x} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b}$.

$$(\mathbf{v} \cdot \mathbf{v}' \cdot \mathbf{d}') \tau = (\mathbf{v} \cdot \mathbf{v}' \cdot \mathbf{d}) \quad (4.44)$$

$$(\mathbf{v} \cdot \mathbf{v}' \cdot \mathbf{d}')^\top (\mathbf{v} \cdot \mathbf{v}' \cdot \mathbf{d}') \tau = (\mathbf{v} \cdot \mathbf{v}' \cdot \mathbf{d}')^\top (\mathbf{v} \cdot \mathbf{v}' \cdot \mathbf{d}) \quad (4.45)$$

$$\tau = \frac{\sum_{i=1}^m (\mathbf{v}_i \mathbf{v}'_i) \cdot \mathbf{d}'_i \mathbf{d}_i}{\sum_{i=1}^m (\mathbf{v}_i \mathbf{v}'_i) \cdot (\mathbf{d}'_i)^2} \quad (4.46)$$

Where m is the number of limb segments within the articulated limb model (two for upper and lower leg). With the scaling transformation \mathbf{H}_τ between both leg planes now known, we can determine the optimal upper leg segment length $\hat{\mathbf{d}}_1$ on the first leg swing plane.

$$\hat{\mathbf{d}}_1 = \frac{\mathbf{v}_1 \mathbf{d}_1 + \tau \mathbf{v}'_1 \mathbf{d}'_1}{\mathbf{v}_1 + \mathbf{v}'_1} \quad (4.47)$$

We then compute the isotropic scaling transformation \mathbf{H}_s that maps $\hat{\mathbf{d}}_1$ to unit length and update both sets of points and projection homographies.

$$\mathbf{H}_{m1} = \mathbf{H}_{\mathbf{q}1} \mathbf{H}_s^{-1} = \begin{bmatrix} \mathbf{q}_1/s & \mathbf{q}_2/s & \mathbf{q}_3 \end{bmatrix} \quad (4.48)$$

$$\mathbf{H}_{m2} = \mathbf{H}_{\mathbf{q}2} \mathbf{H}_\tau^{-1} \mathbf{H}_s^{-1} = \begin{bmatrix} \mathbf{q}_1/s.\tau & \mathbf{q}_2/s.\tau & \mathbf{q}'_3 \end{bmatrix} \quad (4.49)$$

$$\mathbf{H}_{m2} = \begin{bmatrix} \mathbf{q}_1/s & \mathbf{q}_2/s & \tau.\mathbf{q}'_3 \end{bmatrix} \quad (4.50)$$

where $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$ and \mathbf{q}'_3 are the column vectors of the homography mapping matrices $\mathbf{H}_{\mathbf{q}}$ in equations 4.42 and 4.43, s is the isotropic scaling factor that maps $\hat{\mathbf{d}}_1$ to unit length and τ is the inter-plane scaling coefficient between both leg swing planes.

4.2.4 Recovering gait dynamics

Having identified both homography transformations which map points on the metric plane to the set of imaged leg swing planes, structure on the metric plane is ambiguous only by a isometric Euclidean transformation (four-fold reflection about X,Y axes and translation). Since the circular points \mathbf{I}, \mathbf{J} remain fixed under any similarity transformation, there are no further constraints that can be obtained solely from point correspondences alone. Any remaining ambiguity must be resolved as a function of the dynamics of gait.

Since we have identified both imaged leg swing planes, albeit through interpolation of spatio-temporal motion structure, we can now recompute both sets of leg plane points \mathbf{w}_i and \mathbf{w}'_i on the metric plane that are sampled at frame boundaries. The points on the metric plane are computed by applying the inverse mappings $\mathbf{w}_i = (\mathbf{H}_{\mathbf{m}1})^{-1} \mathbf{x}'_i$ and $\mathbf{w}'_i = (\mathbf{H}_{\mathbf{m}2})^{-1} \mathbf{x}_i$ to the set of normalized image points \mathbf{x}'_i .

The four-fold X,Y reflection ambiguity of the metric plane is resolved by consideration of the gross spatio-temporal motion structure. Two smoothed data vectors $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{u}}'$, generated from the mean X coordinate positions of articulated limb points by using a centred three frame filter, are computed and fitted to a linear velocity model with a pair of simultaneous equations.

$$\tilde{u}_i = v_x \cdot i + u_0 \quad (4.51)$$

$$\tilde{u}'_i = v_x \cdot i + u'_0 \quad (4.52)$$

We choose to normalize gait sequences to emulate a left to right walk, so we ensure that v_x is positive by applying a reflection about the Y axis. We then update both sets of points $\mathbf{w}_i, \mathbf{w}'_i$ and homography mappings accordingly. The reflection about the X axis, to ensure that the sky is upward, is achieved by determining the Y coordinate ordering (*hip* \rightarrow *knee* \rightarrow *ankle*) from the means of each tracked limb point over all frames. The only remaining ambiguity is then the Euclidean translation between both sets of metric plane points.

Normal gait is bilaterally symmetric with a half phase shift. Since angles on the metric plane are invariant to changes in translation, then the set of leg angles can be used directly to compute the angular reconstruction function $\theta(t)$. Any periodic function can be used to encode the leg angle function $\theta(t)$, though the obvious choice is to use a Fourier series representation.

$$\theta(t) = a_0 + \sum_{k=1}^n a_k \cos(2\pi k f_0 t + \phi_k) \quad (4.53)$$

For each articulated leg segment, we compute both sets of valid leg swing plane angles \mathbf{a}, \mathbf{a}' and their corresponding time sample vectors \mathbf{t}, \mathbf{t}' . We concatenate both angle vectors $\mathbf{A} = (\mathbf{a}^\top, \mathbf{a}'^\top)^\top$ and time sample vectors $\mathbf{S} = (\mathbf{t}^\top, \mathbf{t}'^\top + \frac{1}{2}\mathbf{T}^\top)^\top$, where \mathbf{T} is the N-vector of gait period values T that facilitate the half phase bilateral shift. We then determine, with fixed fundamental frequency $f_0 = 1/T$, the Fourier series representation of the limb angle function $\theta(t)$. The set of minimized Fourier coefficients for each leg segment are stored in a biometric reconstruction vector \mathbf{V}_i , where the coefficients of \mathbf{V} have the form:

$$\mathbf{V} = (a_0, a_1, \phi_1, \dots, a_n, \phi_n)^\top \quad (4.54)$$

With the knowledge of the normalized leg lengths \mathbf{D} we can find by back substitution the set of hip points \mathbf{X}_0 and \mathbf{X}'_0 on the metric leg swing planes.

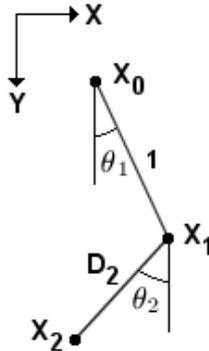


FIGURE 4.13: Articulated limb segment model. The hip point \mathbf{X}_0 is defined by a set of Cartesian $(x, y)^\top$ coordinates. The remaining articulated limb endpoints are defined by a connected set of polar coordinates $(d, \theta)^\top$. The first limb segment length is canonically normalized to unit length.

Figure 4.13 shows the model of articulated limb connections. The hip point \mathbf{X}_0 is given by the Cartesian coordinates $(x, y)^\top$, while the remaining limb endpoints are defined by a connected set of polar coordinates $(d, \theta)^\top$, where the first limb segment has been normalized to unit length. The Cartesian coordinates $(x_i, y_i)^\top$ of any limb point with index i in the model is then given by the equation:

$$(x_i, y_i)^\top = \begin{cases} (u, v)^\top & i = 0 \\ (u, v)^\top + \sum_{j=1}^i \mathbf{D}_j \cdot (\sin \theta_j, \cos \theta_j)^\top & i \geq 1 \end{cases} \quad (4.55)$$

where the pose angles θ_j are given by evaluating the Fourier series functions $\theta(t)$ and $\theta(t + T/2)$ at the current pose frame with the biometric coefficients \mathbf{V}_j and fundamental frequency $f_0 = 1/T$.

Given any endpoint in the articulated limb set, we can compute the putative position

of the hip point $(u, v)^\top$ by back substitution. Since a limb segment endpoint is computed relative to its predecessor, measurement fitting errors will be compounded within the back substitution process. A weighted putative hip point $(u'_i, v'_i, w'_i)^\top$, where w'_i is the associated weighting factor, is computed from each of the valid metric plane data points $(\tilde{x}_i, \tilde{y}_i)^\top$ of the articulated leg pose. The putative hip point, from any indexed metric plane data point $i = 0 \cdots m$ within the current leg pose, is given by the set of equations.

$$(u'_i, v'_i, w'_i)^\top = \begin{cases} (m+1) \cdot (\tilde{x}_i, \tilde{y}_i, 1)^\top & i = 0 \\ (m+1-i) \cdot \left[(\tilde{x}_i, \tilde{y}_i, 1)^\top - \sum_{j=1}^i \mathbf{D}_j \cdot (\sin \theta_j, \cos \theta_j, 0)^\top \right] & i \geq 1 \end{cases} \quad (4.56)$$

Where m is the total number of segments within the articulated leg model, i.e. two for a model of upper and lower legs. The fitted hip point $(u, v)^\top$ is then given by the summation of all valid weighted points.

$$(u, v)^\top = \left(\frac{\sum_{i=0}^m u'_i}{\sum_{i=0}^m w'_i}, \frac{\sum_{i=0}^m v'_i}{\sum_{i=0}^m w'_i} \right)^\top \quad (4.57)$$

We only require a minimum of one metric plane landmark point within a leg pose to compute the hip point. Even if the hip point itself is never imaged, the back substitution process will generate the required Euclidean position. This is a common case within fronto-parallel motion, where the hip point on the occluded side of the body is never visible.

We recover both swing plane vectors \mathbf{o} and \mathbf{o}' of hip point displacement positions, through the back substitution of all leg poses within the motion sequence. The hip point displacement function is separable in both X, Y Euclidean directions and can be parameterised by fitting a modified Fourier series with an additional velocity component v_x to the recovered hip displacement vectors.

$$u(t) = v_x \cdot t + u_0 + \sum_{k=1}^n a_k \cos(2\pi k f_0 t + \phi_k) \quad (4.58)$$

$$u'(t) = v_x \cdot t + u'_0 + \sum_{k=1}^n a_k \cos(2\pi k f_0 (t + \frac{1}{2}T) + \phi_k) \quad (4.59)$$

We first compute two smoothed data vectors $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{u}}'$, generated from the mean X coordinate positions of hip points by using a centred three frame filter, then fit the

linear velocity model to the pair of simultaneous equations in t . This gives a reasonable estimate of the linear velocity component v_x and initial X pose displacements (u_0, u'_0) of gait on the metric plane. Having recovered the linear motion parameters we can remove them from the hip displacement vectors \mathbf{o}, \mathbf{o}' and generate both vectors \mathbf{z}, \mathbf{z}' of purely oscillatory motion.

We fit a partitioned bilateral Fourier series representation of the hip displacement function to the sample data \mathbf{o}, \mathbf{o}' . A first order simple harmonic approximation is first fitted by partitioning the parameter vector as:

$$\mathbf{P}_1 = (v_x, a_1, \phi_1 \mid u_0, u'_0)^\top \quad (4.60)$$

Initial estimates of (a_1, ϕ_1) are computed by finding a_1 based on the means of the two largest and two smallest amplitudes within both oscillatory motion vectors \mathbf{z}, \mathbf{z}' . The initial estimate of phase ϕ_1 is chosen by evaluating a first order simple harmonic series at three different phase positions $\phi_1 = (-\frac{2\pi}{3}, 0, \frac{2\pi}{3})$ and choosing the value with smallest residual fitting error.

Minimization with fixed $f_0 = 1/T$ of the partitioned first order simple harmonic series \mathbf{P}_1 gives a good initial estimate of the gross motion structure of the origin limb points on both swing planes. The computed estimates of \mathbf{P}_1 are then used to bootstrap the full partitioned parameterisation.

$$\mathbf{P} = (v_x, a_1, \phi_1, \dots, a_n, \phi_n \mid u_0, u'_0)^\top \quad (4.61)$$

The parameterisation is similar for the Y component hip displacement function, except that v_y is held fixed (zero). Both are computed using a partitioned Levenberg-Marquardt algorithm with fixed fundamental frequency f_0 . See appendix C.4 for further details on Levenberg-Marquardt minimization and parameter partitioning.

Having found the initial set of pose displacements $(u_0, v_0)^\top$ and $(u'_0, v'_0)^\top$ of the hip motion functions on both swing planes, we can apply the translations $\mathbf{H}_\mathbf{o}$ and $\mathbf{H}'_\mathbf{o}$ that map them back to the coordinate system origin. We then update the homography mapping matrices accordingly.

$$\mathbf{H}_1 = \begin{bmatrix} \mathbf{m}_1 & \mathbf{m}_2 & \mathbf{m}_3 \end{bmatrix} \begin{pmatrix} 1 & -u_0 \\ & 1 & -v_0 \\ & & 1 \end{pmatrix} \quad (4.62)$$

$$\mathbf{H}_2 = \begin{bmatrix} \mathbf{m}_1 & \mathbf{m}_2 & \mathbf{m}'_3 \end{bmatrix} \begin{pmatrix} 1 & -u'_0 \\ & 1 & -v'_0 \\ & & 1 \end{pmatrix} \quad (4.63)$$

where \mathbf{m}_i are the column vectors of the reconstruction homography matrices after having been updated by the suitable set of reflections in the X, Y axes that restores the canonical gait coordinate system. We finally apply the inverse normalization transformation $\mathbf{K}_n^{-1}\mathbf{H}_i$ in order that we map metric plane points to real image points. The homography mapping functions, that map articulated leg motion from the metric plane to the image plane, then have the form:

$$\mathbf{x}_1(t) = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3 \end{bmatrix} g(t : f_0, \mathbf{D}, \mathbf{X}, \mathbf{Y}, \mathbf{V}) \quad (4.64)$$

$$\mathbf{x}_2(t) = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}'_3 \end{bmatrix} g(t + T/2 : f_0, \mathbf{D}, \mathbf{X}, \mathbf{Y}, \mathbf{V}) \quad (4.65)$$

where $g(t)$ is the bilateral Fourier series function of articulated leg motion, \mathbf{X} and \mathbf{Y} are the velocity and Fourier coefficients of the metric plane hip displacement functions, \mathbf{V} the Fourier coefficients of the upper and lower leg angle functions, and \mathbf{D} is the vector of normalized leg lengths.

4.2.5 Stratified reconstruction analysis

A synchronized, three camera system is set up as part of the laboratory experiments described within section 3.4. The cameras are positioned at least three metres apart, down one side of an indoor test track. Each camera is equipped with a different type of lens and is orientated to fit the entire test track within its field of view. Four subjects, who have retro-reflective marker balls attached to the principal joint features of the skeletal system, are told to walk along a linear trajectory through the field of view of all cameras. The captured set of image sequences are manually marked and the reconstruction process performed. Since image acquisition is synchronous and gait dynamics consistent in all three views, the only parameters that differ between image sequences are the camera intrinsic and extrinsic parameters.

Figures 4.14, 4.15 and 4.16 show three different camera views of reconstructed gait

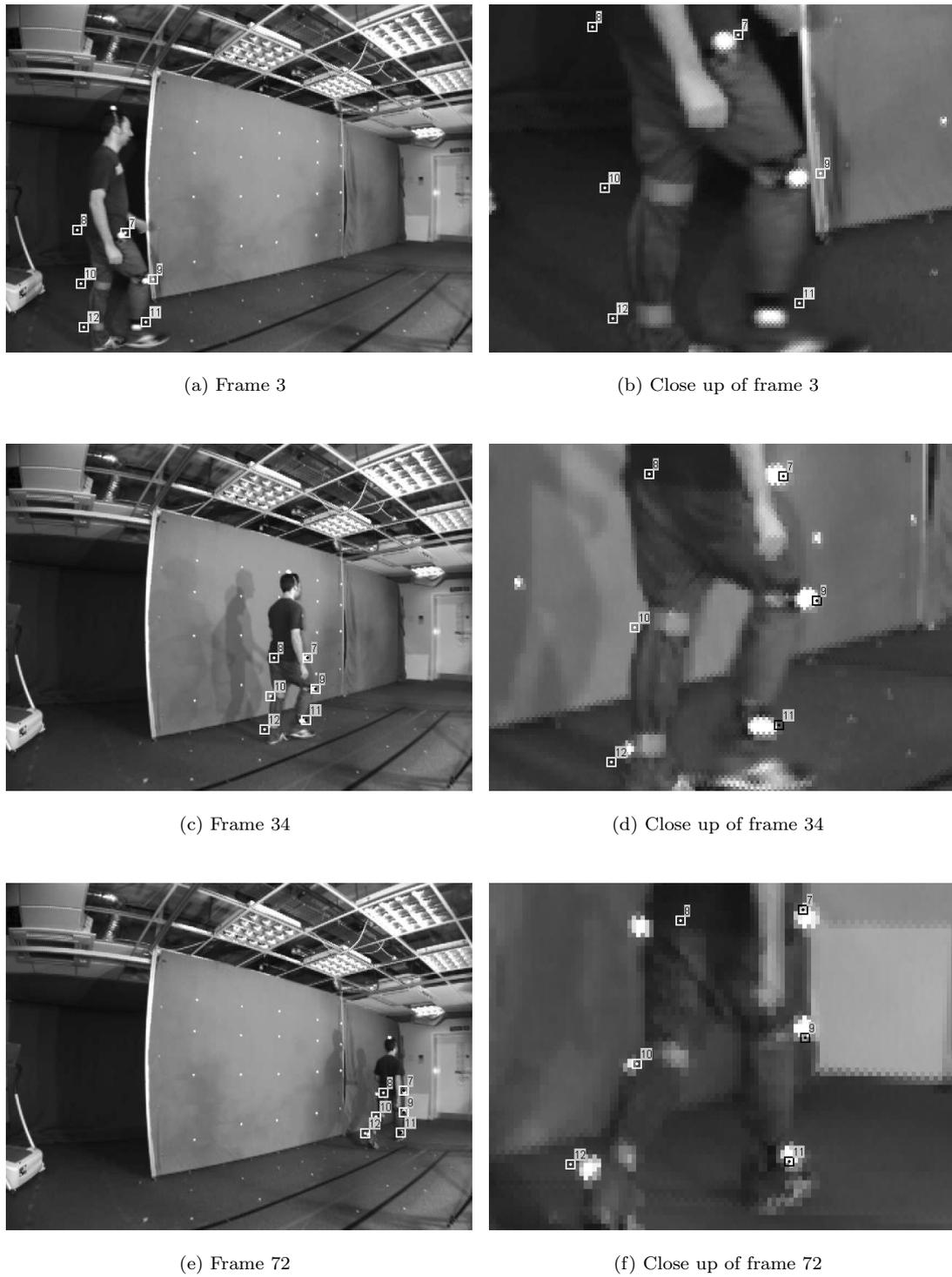


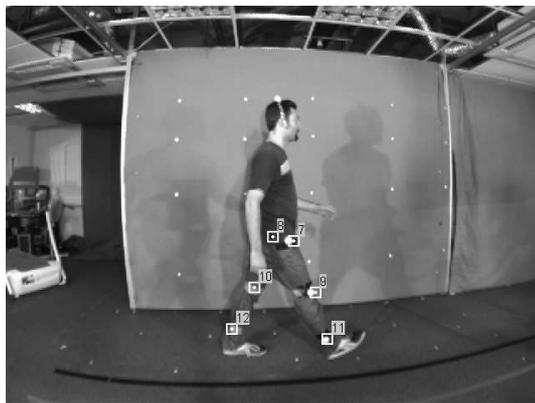
FIGURE 4.14: Left camera stratified reconstruction sequence: Three poses of a subject at the beginning, middle and end of the sequence with reprojected landmark points.



(a) Frame 11



(b) Close up of frame 11



(c) Frame 37



(d) Close up of frame 37



(e) Frame 67



(f) Close up of frame 67

FIGURE 4.15: Middle camera stratified reconstruction sequence: Three poses of a subject at the beginning, middle and end of the sequence with reprojected landmark points.

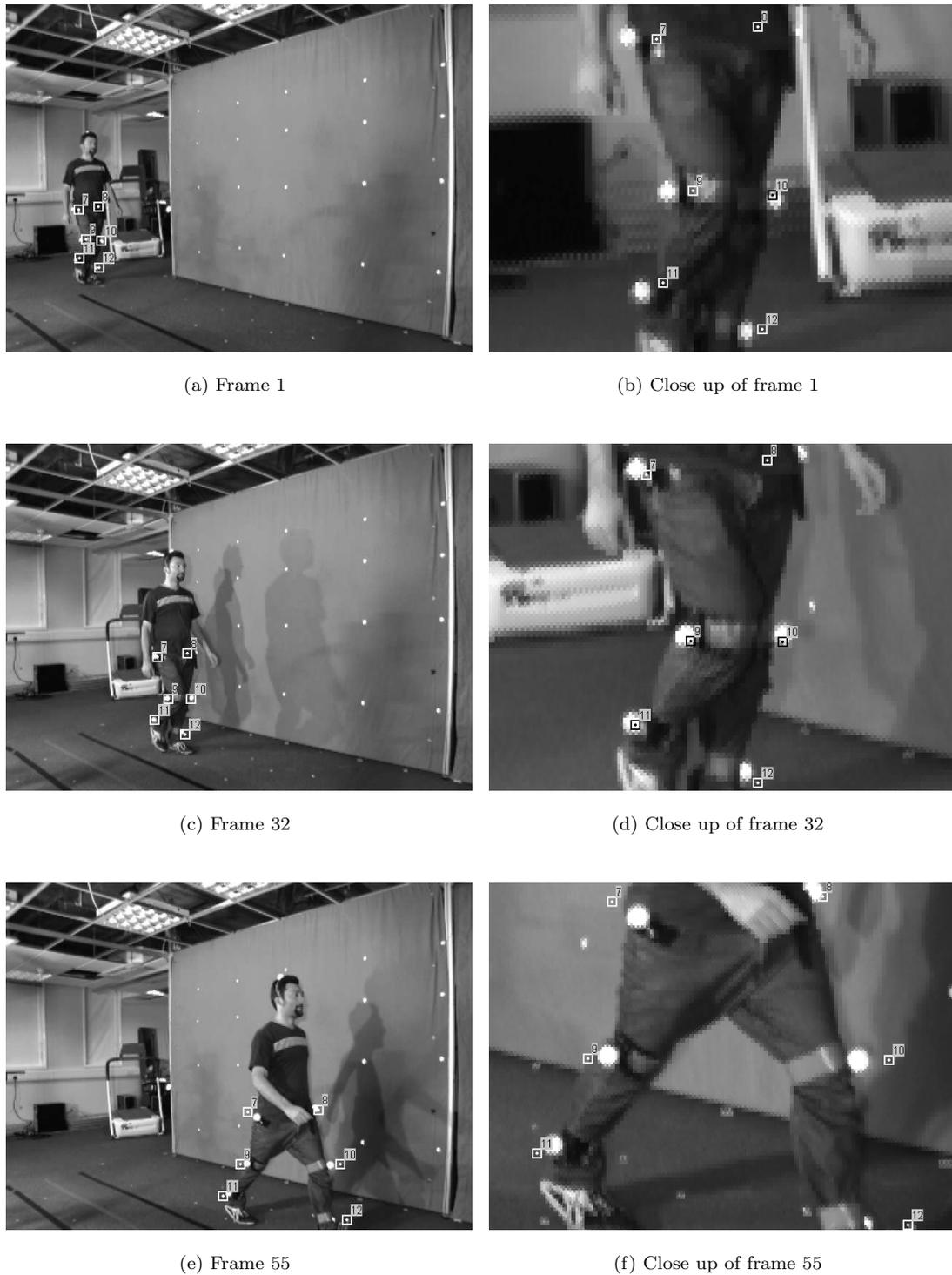


FIGURE 4.16: Right camera stratified reconstruction sequence: Three poses of a subject at the beginning, middle and end of the sequence with reprojected landmark points.

motion, with each figure depicting three pose positions taken at the start, middle and end of the image sequence. Reconstruction for fronto-parallel gait motion is reasonably accurate while the fitting error due to projective distortion is most apparent in both oblique views. Since most gait motion is observed closer to the focus of expansion (epipole) both spatially and temporally, the stratified technique of warping back to the canonical frame favours fitting to these data points. The fitting process on the metric plane does not minimize reprojection error within the image, hence reconstruction error is most apparent in points furthest from the focus of expansion, as shown in figures 4.14(a) and 4.16(e).

A further optimization step, to improve the fitting of the origin limb displacement function parameters \mathbf{X} and \mathbf{Y} with all other parameters fixed, was evaluated that minimizes the image reprojection error of equation 4.64. Since metric plane points are mapped into the image via the homography transformations \mathbf{H}_1 and \mathbf{H}_2 , minimization of the parameters \mathbf{X} , \mathbf{Y} is no longer separable. The resulting improvement in fitting is marginal, suggesting that most of the error is due to the coefficients of the homography mapping matrices. This is hardly surprising, given that the stratified reconstruction technique is derived from interpolated spatio-temporal motion structure and projective transformations of the plane. The majority of the reconstruction parameters are defined within the elements of the bilateral Fourier motion functions \mathbf{X} , \mathbf{Y} , \mathbf{V} , thus further optimization to minimize reprojection error on these parameters alone requires a large computational overhead for very little improvement. We conclude that any further minimization, other than a full bundle adjustment, at this stage is unnecessary.

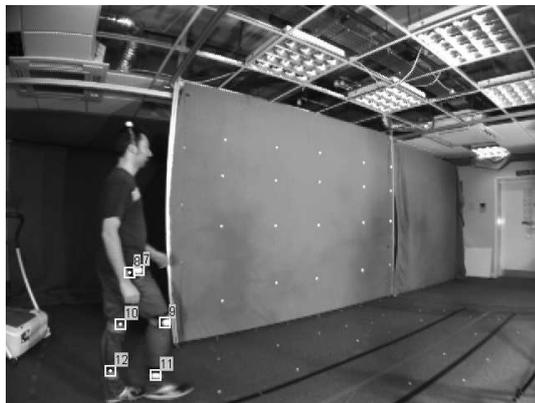
4.3 Maximum likelihood estimation

As a final optimization step we perform a bundle adjustment procedure that minimizes image reprojection error with respect to all parameters of the gait projection function. The set of parameters can be partitioned as:

$$\mathbf{P} = (f_0, \mathbf{D}^\top, \mathbf{X}^\top, \mathbf{Y}^\top, \mathbf{V}^\top, \mathbf{h}_1^\top, \mathbf{h}_2^\top \mid \mathbf{h}_3^\top, \mathbf{h}'_3^\top)^\top \quad (4.66)$$

The set of parameters are optimized by performing a Levenberg-Marquardt minimization on the partitioned vector \mathbf{P} . For a discussion on non-linear minimization and the Levenberg-Marquardt method see appendix C.2. The parameter partitioning is apparent since both planes are independent. Refer to the section in appendix C.4 for details on parameter partitioning and sparse methods.

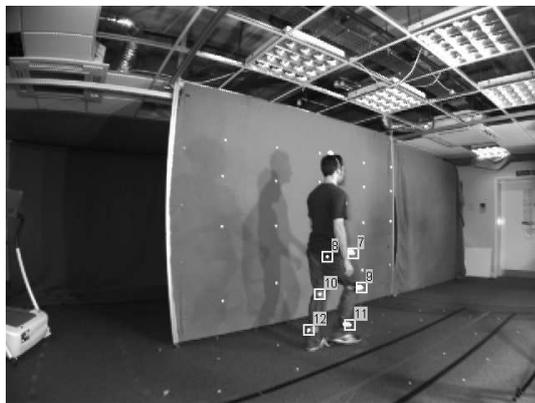
Figures 4.17, 4.18 and 4.19 show three different camera views of reconstructed gait motion with each figure depicting three pose positions taken at the start, middle and



(a) Frame 3



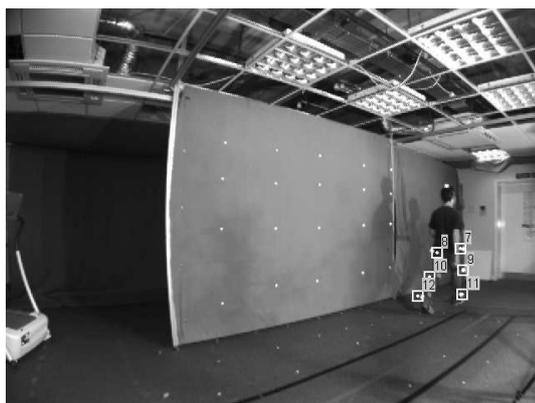
(b) Close up of frame 3



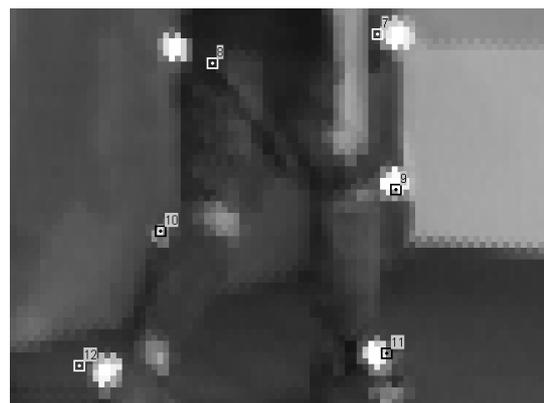
(c) Frame 34



(d) Close up of frame 34



(e) Frame 72



(f) Close up of frame 72

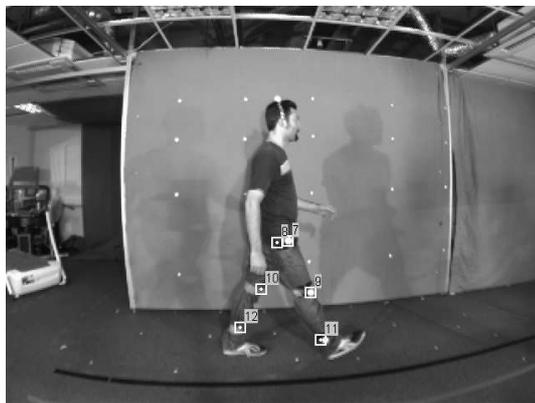
FIGURE 4.17: Left camera reconstruction sequences using the pose parameterisation $\mathbf{H} = (\mathbf{h}_1^\top, \mathbf{h}_2^\top, \mathbf{h}_3^\top, \mathbf{h}'_3^\top)^\top$: Three poses of a subject at the beginning, middle and end of the sequence with reprojected landmark points.



(a) Frame 11



(b) Close up of frame 11



(c) Frame 37



(d) Close up of frame 37



(e) Frame 67



(f) Close up of frame 67

FIGURE 4.18: Middle camera reconstruction sequences using the pose parameterisation $\mathbf{H} = (\mathbf{h}_1^\top, \mathbf{h}_2^\top, \mathbf{h}_3^\top, \mathbf{h}'_3^\top)^\top$: Three poses of a subject at the beginning, middle and end of the sequence with reprojected landmark points.



(a) Frame 1



(b) Close up of frame 1



(c) Frame 32



(d) Close up of frame 32



(e) Frame 55



(f) Close up of frame 55

FIGURE 4.19: Right camera reconstruction sequences using the pose parameterisation $\mathbf{H} = (\mathbf{h}_1^\top, \mathbf{h}_2^\top, \mathbf{h}_3^\top, \mathbf{h}'_3^\top)^\top$: Three poses of a subject at the beginning, middle and end of the sequence with reprojected landmark points.

end of the image sequence. We can clearly see a marked improvement in the fitting over the initial stratified reconstruction.

Since all image sequences are acquired from a set of synchronized cameras, the dynamics of subject gait are consistent in all views. The reconstructed set of gait functions $g(t : f_0, \mathbf{D}, \mathbf{X}, \mathbf{Y}, \mathbf{V})$ should match for each of the camera sequences. Figures 4.20 and 4.21 show all three camera views of the reconstructed leg angle motion. The leg angle motion for both upper and lower limbs is reasonably well aligned.

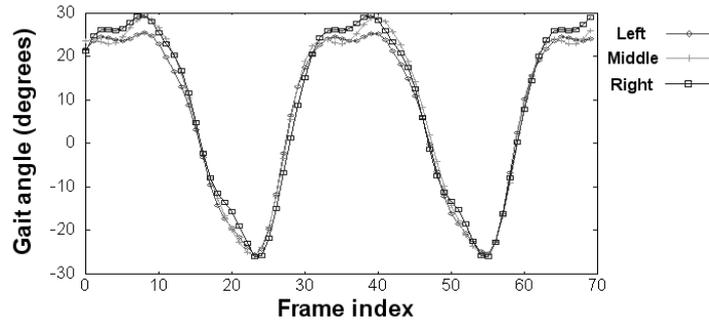


FIGURE 4.20: **Metric reconstruction of the upper leg limb angles.** The corresponding unrectified set of angles are shown in figure 4.2.

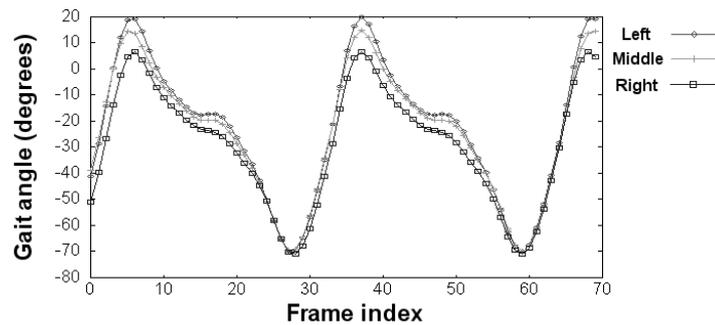


FIGURE 4.21: **Metric reconstruction of the lower leg limb angles.** The corresponding unrectified set of angles are shown in figure 4.3.

4.4 Imposing a rigid motion transform model

The parameterisation of the leg swing plane, specifically the limb inclination angle to the vertical, has up until now been defined by a projectivity that maps the ideal line at infinity to the line formed from the join of both worldspace leg swing planes. Another way to parameterise the leg pose is to constrain the mapping to consist of only rigid motion transformations (scaling, rotation and translation). We can compute a subject limb plane pose by first applying a rotation \mathbf{H}_α about the X axis to facilitate the leg swing plane inclination to the vertical, then apply the leg plane selection translation \mathbf{H}_β to map the required hip point to ∓ 1 . This is followed by a scaling \mathbf{H}_τ in the Z direction that

generates the correct distance between both hip points. The subject pose projection \mathbf{P} then maps the set of workspace points \mathbf{X} into the image as $\mathbf{x}' = \mathbf{P}\mathbf{X}$.

$$\mathbf{P} = \mathbf{K} [\mathbf{R} \mid \mathbf{t}] \mathbf{H}_\tau \mathbf{H}_\beta \mathbf{H}_\alpha \quad (4.67)$$

The leg plane inclination angles α must have opposite signs because of the bilateral symmetry of the human posture. From the basic properties of trigonometric functions, $\cos(-\alpha) = \cos(\alpha)$ and $\sin(-\alpha) = -\sin(\alpha)$, the set of rotation transformation coefficients in conjunction with the required limb plane selection $\beta = \mp 1$ are defined by $(\cos \alpha, \beta \sin \alpha)$. The coefficients of the subject pose projection \mathbf{P} can then be written.

$$\mathbf{P} = \mathbf{K} [\mathbf{R} \mid \mathbf{t}] \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \tau & \\ & & & 1 \end{pmatrix} \begin{pmatrix} 1 & & & \\ & \cos \alpha & -\beta \sin \alpha & \\ & \beta \sin \alpha & \cos \alpha & \beta \\ & & & 1 \end{pmatrix} \quad (4.68)$$

Subsequently, the set of $z = 0$ plane workspace points $\mathbf{X} = (u, v, 0, w)^\top$ are projected into the image as $\mathbf{x}' = \mathbf{P}\mathbf{X}$.

$$\mathbf{x}' = \mathbf{K} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \mathbf{t} \end{bmatrix} \begin{pmatrix} 1 & & & \\ & \cos \alpha & -\beta \sin \alpha & \\ & \tau \beta \sin \alpha & \tau \cos \alpha & \tau \beta \\ & & & 1 \end{pmatrix} \begin{pmatrix} u \\ v \\ 0 \\ w \end{pmatrix} \quad (4.69)$$

Projection of planar workspace structure into the image is then achieved by the corresponding leg plane homography transformations.

$$\mathbf{H} = \begin{bmatrix} \mathbf{K} \cdot \mathbf{r}_1 & (\cos \alpha \cdot \mathbf{K} \cdot \mathbf{r}_2 + \tau \beta \sin \alpha \cdot \mathbf{K} \cdot \mathbf{r}_3) & (\tau \beta \cdot \mathbf{K} \cdot \mathbf{r}_3 + \mathbf{K} \cdot \mathbf{t}) \end{bmatrix} \quad (4.70)$$

The individual left and right leg plane homography transformations \mathbf{H}_1 and \mathbf{H}_2 then have the form:

$$\mathbf{H}_1 = \begin{bmatrix} \mathbf{m}_1 & (\cos \alpha \cdot \mathbf{m}_2 - \sin \alpha \cdot \mathbf{m}_3) & (\mathbf{m}_4 - \mathbf{m}_3) \end{bmatrix} \quad (4.71)$$

$$\mathbf{H}_2 = \begin{bmatrix} \mathbf{m}_1 & (\cos \alpha \cdot \mathbf{m}_2 + \sin \alpha \cdot \mathbf{m}_3) & (\mathbf{m}_4 + \mathbf{m}_3) \end{bmatrix} \quad (4.72)$$

where $\mathbf{m}_1 = \mathbf{K} \cdot \mathbf{r}_1$, $\mathbf{m}_2 = \mathbf{K} \cdot \mathbf{r}_2$, $\mathbf{m}_3 = \tau \cdot \mathbf{K} \cdot \mathbf{r}_3$ and $\mathbf{m}_4 = \mathbf{K} \cdot \mathbf{t}$

The reconstruction algorithm is performed in a similar fashion to the previous method. We first compute the initial stratified reconstruction of the two leg swing planes, with the assumption that both are parallel $\alpha = 0$.

$$\mathbf{H}_1 = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3 \end{bmatrix} \quad (4.73)$$

$$\mathbf{H}_2 = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}'_3 \end{bmatrix} \quad (4.74)$$

The corresponding set of column vectors \mathbf{m} are then related to those of \mathbf{h} by:

$$\mathbf{m}_1 = \mathbf{h}_1 \quad (4.75)$$

$$\mathbf{m}_2 = \mathbf{h}_2 \quad (4.76)$$

$$\mathbf{m}_3 = \frac{1}{2}(\mathbf{h}'_3 - \mathbf{h}_3) \quad (4.77)$$

$$\mathbf{m}_4 = \frac{1}{2}(\mathbf{h}'_3 + \mathbf{h}_3) \quad (4.78)$$

The elements of the pose transformation mapping \mathbf{M} can be parameterised by a single 13-vector.

$$\mathbf{M} = (\mathbf{m}_1^\top, \mathbf{m}_2^\top, \mathbf{m}_3^\top, \mathbf{m}_4^\top, \alpha)^\top \quad (4.79)$$

The bundle adjustment procedure that minimizes reprojection error within the image sequence is then parameterised by the vector:

$$\mathbf{P} = (f_0, \mathbf{D}^\top, \mathbf{X}^\top, \mathbf{Y}^\top, \mathbf{V}^\top, \mathbf{M}^\top)^\top \quad (4.80)$$

The set of parameters are optimized by performing a Levenberg-Marquardt minimization on the vector \mathbf{P} . See appendix C.2 for details on Levenberg-Marquardt minimization. The elements of the pose projection vector \mathbf{M} over parameterise the planar transformation that maps structure on the metric plane to the imaged limb swing planes, i.e. there are 13 parameters that encode the set of transformations that have a total of 12 degrees of freedom. The Jacobian matrix \mathbf{J} is therefore rank deficient, as is the Hessian matrix $\mathbf{J}^\top \Sigma_{\mathbf{x}}^{-1} \mathbf{J}$. The corresponding covariance matrix $\Sigma_{\mathbf{p}}$ of the set of estimated parameters \mathbf{P} is then determined by computing the pseudo-inverse, see appendix B.2.1.

$$\Sigma_{\mathbf{p}} = (\mathbf{J}^{\top} \Sigma_{\mathbf{x}}^{-1} \mathbf{J})^+ \quad (4.81)$$

4.5 Reconstruction analysis

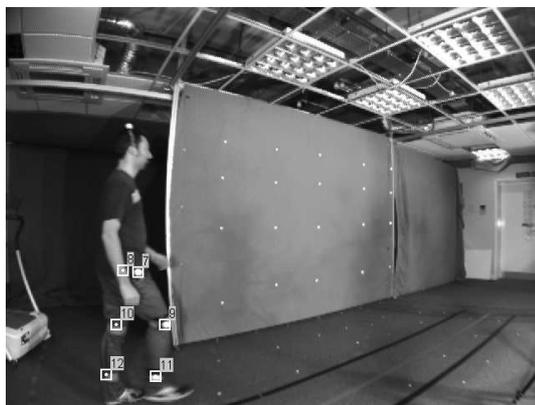
Figures 4.22, 4.23 and 4.24 show three different camera views of reconstructed gait motion using the rigid motion transform parameterisation $\mathbf{M} = (\mathbf{m}_1^{\top}, \mathbf{m}_2^{\top}, \mathbf{m}_3^{\top}, \mathbf{m}_4^{\top}, \alpha)^{\top}$. Each series of three images within the figure depicts pose positions taken at the start, middle and end of the image sequence. The set of reprojected landmark points corresponding to the optimized parameters of gait motion have also been marked within each of the image frames.

The improvement in fitting over the previous method is most apparent in landmark correspondences across the bilateral symmetry plane. Minimizing reprojection error using the first method favours optimizing residual error with respect to the greatest density of point distribution. The majority of the occluded points occur on one side of the subject within the set of image sequences. Minimization then favours fitting the inclined swing planes with a bias towards the set of points on the right side of the body. Figure 4.17(e) clearly shows a fitting disparity between points on both sides of the body. Since the left side hip point (landmark no. 08) is imaged in very few frames then it is unsurprising that the residual error of this point is large when compared to the others. In fact, because of the nature of the projective transformation model, the root mean square residual fitting errors between imaged and reprojected points on both swing planes are individually quite different. This is clearly seen from the comparison of the individual root mean square image reprojection errors, shown within table 4.1, corresponding to each separate leg swing plane.

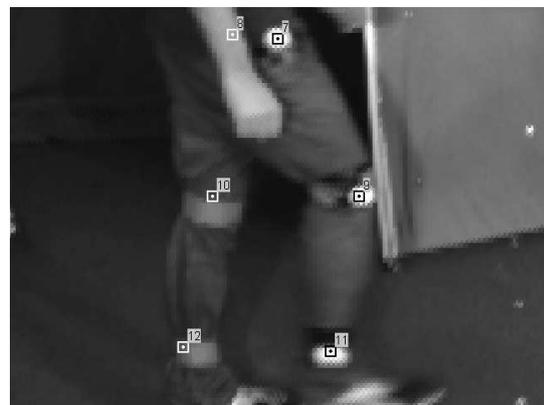
Minimizing reprojection error using the rigid motion transform model \mathbf{M} produces a much closer fitting match than that of the first method. This is due to the more realistic parameterisation of the leg pose. Subsequently, the residual error is distributed more evenly over both leg planes, as imaged in figure 4.22(e). The difference between parameterisations is shown by the root mean square pixel reprojection errors of both independent leg swing planes, shown in table 4.1.

<i>Method</i>	<i>left</i> π_1	<i>left</i> π_2	<i>middle</i> π_1	<i>middle</i> π_2	<i>right</i> π_1	<i>right</i> π_2
H	1.735	3.48	2.524	4.11	2.894	4.996
M	1.438	1.984	2.345	2.501	1.985	2.746

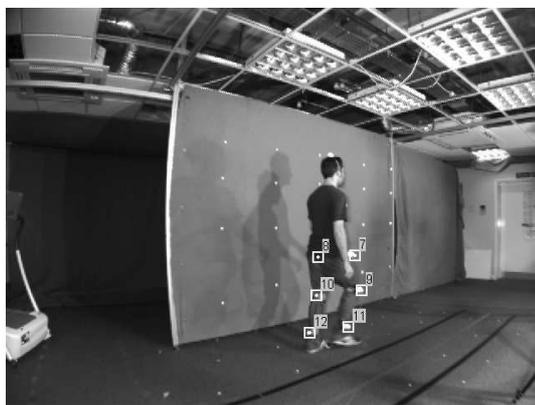
TABLE 4.1: Comparison of root mean square pixel reprojection errors between both independent leg swing planes (π_1, π_2) of subject 00 for both methods of reconstruction. The bundle adjustment methods parameterise the leg swing planes by using a projectivity $\mathbf{H} = (\mathbf{h}_1^{\top}, \mathbf{h}_2^{\top}, \mathbf{h}_3^{\top}, \mathbf{h}'_3)^{\top}$, and a set of rigid motion transforms $\mathbf{M} = (\mathbf{m}_1^{\top}, \mathbf{m}_2^{\top}, \mathbf{m}_3^{\top}, \mathbf{m}_4^{\top}, \alpha)^{\top}$.



(a) Frame 3



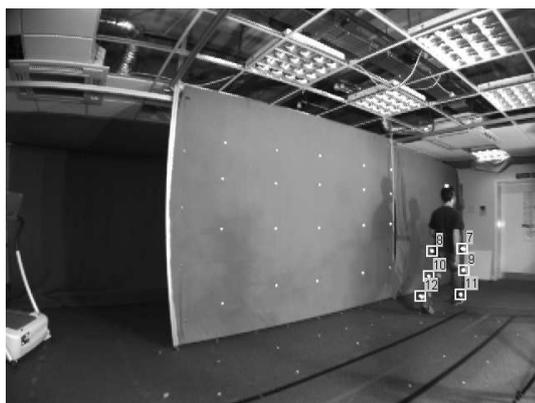
(b) Close up of frame 3



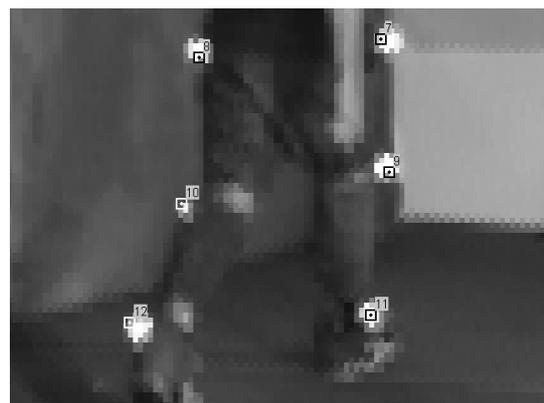
(c) Frame 34



(d) Close up of frame 34



(e) Frame 72



(f) Close up of frame 72

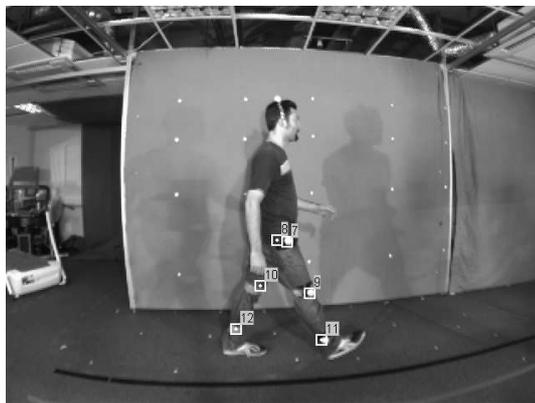
FIGURE 4.22: Left camera reconstruction sequences using the pose parameterisation $\mathbf{M} = (\mathbf{m}_1^\top, \mathbf{m}_2^\top, \mathbf{m}_3^\top, \mathbf{m}_4^\top, \alpha)^\top$: Three poses of a subject at the beginning, middle and end of the sequence with reprojected landmark points.



(a) Frame 11



(b) Close up of frame 11



(c) Frame 37



(d) Close up of frame 37



(e) Frame 67



(f) Close up of frame 67

FIGURE 4.23: Middle camera reconstruction sequences using the pose parameterisation $\mathbf{M} = (\mathbf{m}_1^\top, \mathbf{m}_2^\top, \mathbf{m}_3^\top, \mathbf{m}_4^\top, \alpha)^\top$: Three poses of a subject at the beginning, middle and end of the sequence with reprojected landmark points.



(a) Frame 1



(b) Close up of frame 1



(c) Frame 32



(d) Close up of frame 32



(e) Frame 55



(f) Close up of frame 55

FIGURE 4.24: Right camera reconstruction sequences using the pose parameterisation $\mathbf{M} = (\mathbf{m}_1^\top, \mathbf{m}_2^\top, \mathbf{m}_3^\top, \mathbf{m}_4^\top, \alpha)^\top$: Three poses of a subject at the beginning, middle and end of the sequence with reprojected landmark points.

A comparison, for each of the four test subjects, of the root mean square residual fitting errors between the three reconstruction methods is summarised in table 4.2. Naturally, we expect the initial stratified reconstruction to have the largest fitting error. The improvement in reprojection error is quite pronounced between the two MLE methods, and is reflected in the comparison between reconstructed leg angle functions on the metric plane. The set of reconstructed leg angles shown in figures 4.25 and 4.26, match considerably more closely by using the rigid motion transform parameterisation \mathbf{M} , over those of the projectivity parameterisation \mathbf{H} shown in figures 4.20 and 4.21.

<i>Method</i>	<i>left</i>	<i>middle</i>	<i>right</i>
Stratified	7.489	6.171	9.033
\mathbf{H}	2.349	2.802	3.719
\mathbf{M}	1.609	2.479	2.317

(a) Subject 00

<i>Method</i>	<i>left</i>	<i>middle</i>	<i>right</i>
Stratified	8.545	5.485	8.373
\mathbf{H}	2.042	3.276	4.667
\mathbf{M}	1.496	2.811	2.377

(b) Subject 01

<i>Method</i>	<i>left</i>	<i>middle</i>	<i>right</i>
Stratified	8.794	9.893	9.895
\mathbf{H}	2.775	2.821	5.411
\mathbf{M}	2.063	2.246	2.481

(c) Subject 02

<i>Method</i>	<i>left</i>	<i>middle</i>	<i>right</i>
Stratified	3.503	3.709	6.732
\mathbf{H}	2.285	2.874	3.424
\mathbf{M}	1.399	1.895	2.868

(d) Subject 03

TABLE 4.2: Comparison of root mean square pixel reprojection errors between the three different reconstruction methods. Errors are compared for subjects in each camera view. The bundle adjustment methods parameterise the leg swing planes by using a projectivity $\mathbf{H} = (\mathbf{h}_1^\top, \mathbf{h}_2^\top, \mathbf{h}_3^\top, \mathbf{h}'_3^\top)^\top$, and a set of rigid motion transforms $\mathbf{M} = (\mathbf{m}_1^\top, \mathbf{m}_2^\top, \mathbf{m}_3^\top, \mathbf{m}_4^\top, \alpha)^\top$.

The reconstructed set of gait functions $g(t : f_0, \mathbf{D}, \mathbf{X}, \mathbf{Y}, \mathbf{V})$ for each of the four test subjects are shown in figures 4.25 to 4.32. Each reconstructed set of leg angle plots, corresponding to the three different views, match reasonably well. This demonstrates that the dynamics of gait can be recovered, irrespective of the camera parameters and the subject's motion trajectory.

4.5.1 Intra and inter-class variation

We have discussed in the previous chapter, the intra and inter-class variation of the proposed set of biometric parameters \mathbf{P} , for subjects walking over a range of speeds on a treadmill. We have outlined a suitable Euclidean distance metric $\rho = \|\hat{\mathbf{P}} - \mathbf{P}\|$ between a known $\hat{\mathbf{P}}$ and unknown \mathbf{P} gait feature vector. The parameters within the feature vector \mathbf{P} correspond to the coefficients b_k and ψ_k of the modified Fourier series that represents the articulated limb angle motion.

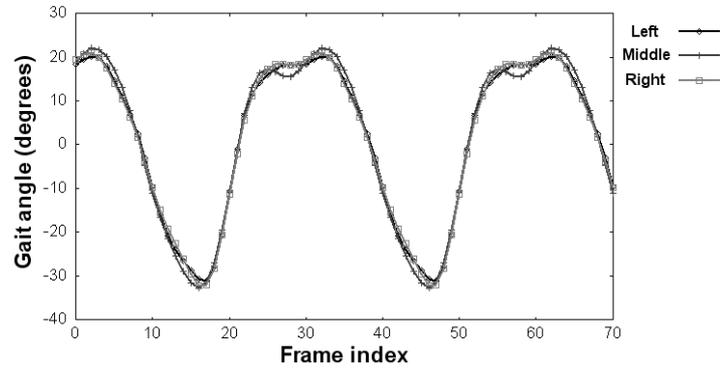


FIGURE 4.25: Reconstruction of upper leg motion for subject 00 using the rigid motion transform pose parameterisation $\mathbf{M} = (\mathbf{m}_1^\top, \mathbf{m}_2^\top, \mathbf{m}_3^\top, \mathbf{m}_4^\top, \alpha)^\top$.

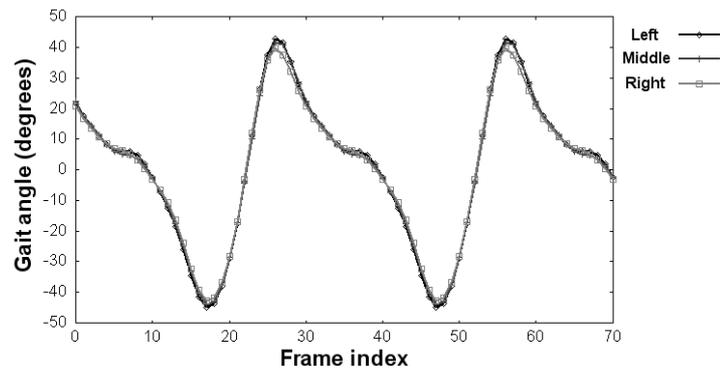


FIGURE 4.26: Reconstruction of lower leg motion for subject 00 using the rigid motion transform pose parameterisation $\mathbf{M} = (\mathbf{m}_1^\top, \mathbf{m}_2^\top, \mathbf{m}_3^\top, \mathbf{m}_4^\top, \alpha)^\top$.

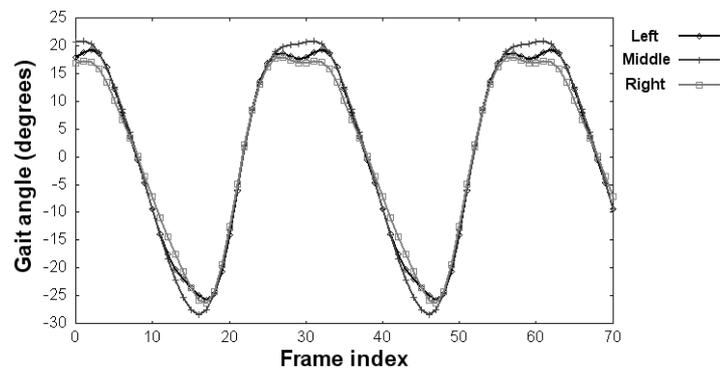


FIGURE 4.27: Reconstruction of upper leg motion for subject 01 using the rigid motion transform pose parameterisation $\mathbf{M} = (\mathbf{m}_1^\top, \mathbf{m}_2^\top, \mathbf{m}_3^\top, \mathbf{m}_4^\top, \alpha)^\top$.

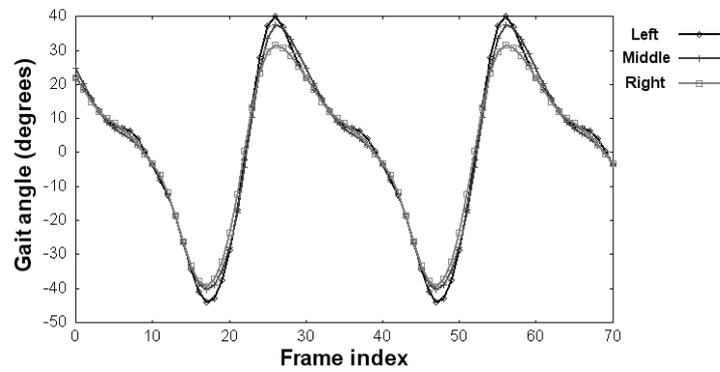


FIGURE 4.28: Reconstruction of lower leg motion for subject 01 using the rigid motion transform pose parameterisation $\mathbf{M} = (\mathbf{m}_1^\top, \mathbf{m}_2^\top, \mathbf{m}_3^\top, \mathbf{m}_4^\top, \alpha)^\top$.

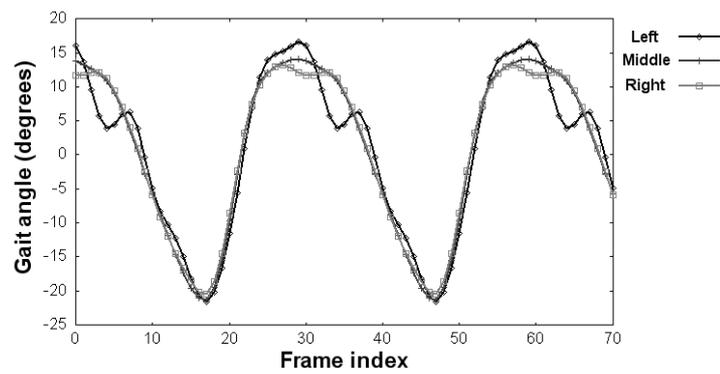


FIGURE 4.29: Reconstruction of upper leg motion for subject 02 using the rigid motion transform pose parameterisation $\mathbf{M} = (\mathbf{m}_1^\top, \mathbf{m}_2^\top, \mathbf{m}_3^\top, \mathbf{m}_4^\top, \alpha)^\top$.

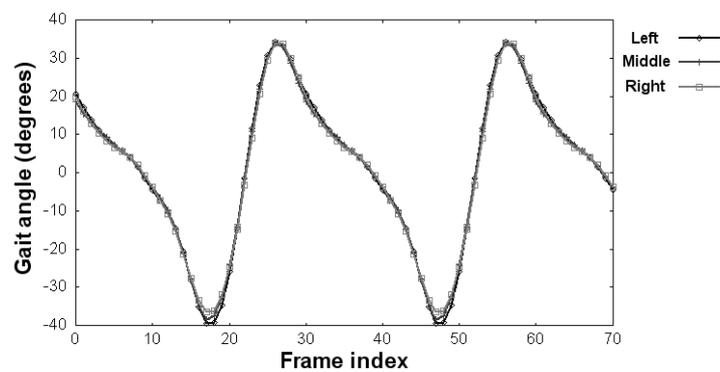


FIGURE 4.30: Reconstruction of lower leg motion for subject 02 using the rigid motion transform pose parameterisation $\mathbf{M} = (\mathbf{m}_1^\top, \mathbf{m}_2^\top, \mathbf{m}_3^\top, \mathbf{m}_4^\top, \alpha)^\top$.

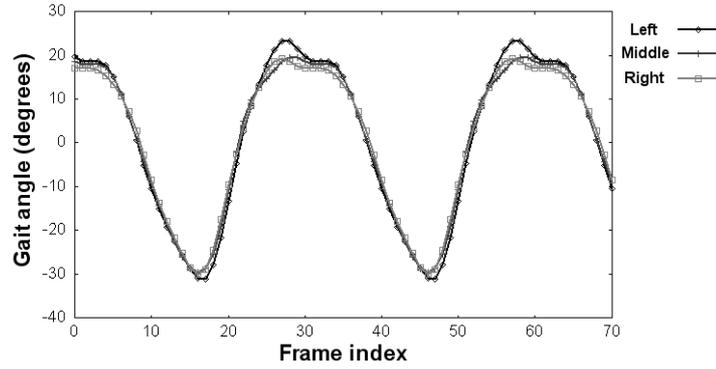


FIGURE 4.31: Reconstruction of upper leg motion for subject 03 using the rigid motion transform pose parameterisation $\mathbf{M} = (\mathbf{m}_1^\top, \mathbf{m}_2^\top, \mathbf{m}_3^\top, \mathbf{m}_4^\top, \alpha)^\top$.

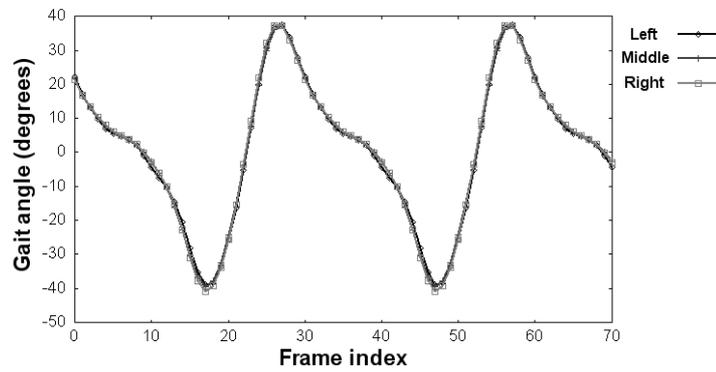


FIGURE 4.32: Reconstruction of lower leg motion for subject 03 using the rigid motion transform pose parameterisation $\mathbf{M} = (\mathbf{m}_1^\top, \mathbf{m}_2^\top, \mathbf{m}_3^\top, \mathbf{m}_4^\top, \alpha)^\top$.

$$\theta(t) = a_0 + a_1 \cos(2\pi f_0 \cdot (t + t_s)) + a_1 \cdot \sum_{k=2}^n b_k \cos(2\pi k f_0 \cdot (t + t_s) + \psi_k) \quad (4.82)$$

where $b_k = a_k/a_1$ are the normalized amplitude coefficients, $\psi_k = \phi_k - k\phi_1$ are the aligned phase coefficients of the articulated leg angle function and $t_s = \phi_1/2\pi f_0$ the required coordinate system time shift. See the discussion of the limb angle function, in section 3.6.4 on page 92, for further details.

The proposed biometric feature vector contains five components $\mathbf{P} = (d_2, b_2, \psi_2, b'_2, \psi'_2)^\top$, where d_2 is the normalized lower leg segment length, (b_2, ψ_2) are the normalized amplitude and phase terms of the upper leg angle function, and (b'_2, ψ'_2) the corresponding coefficients of the lower leg angle function. The proposed gait distance metric $\rho = \|\hat{\mathbf{P}} - \mathbf{P}\|$, that is suitable for subject identification, can then be written:

$$\begin{aligned} \rho^2 = & \left(\widehat{d}_2 - d_2\right)^2 + \left(\widehat{b}_2 - b_2\right)^2 + \left(\frac{b_2}{2}(1 - \widehat{\mathbf{v}}_2^\top \mathbf{v}_2)\right)^2 \\ & + \\ & \left(\widehat{b}'_2 - b'_2\right)^2 + \left(\frac{b'_2}{2}(1 - \widehat{\mathbf{v}}_2'^\top \mathbf{v}'_2)\right)^2 \end{aligned} \quad (4.83)$$

where the phase direction vectors $\widehat{\mathbf{v}}$ and \mathbf{v} are commuted from polar to Euclidean form.

$$\widehat{\mathbf{v}}_2 = (\cos \widehat{\psi}_2, \sin \widehat{\psi}_2)^\top \quad \widehat{\mathbf{v}}_2' = (\cos \widehat{\psi}'_2, \sin \widehat{\psi}'_2)^\top \quad (4.84)$$

$$\mathbf{v}_2 = (\cos \psi_2, \sin \psi_2)^\top \quad \mathbf{v}'_2 = (\cos \psi'_2, \sin \psi'_2)^\top \quad (4.85)$$

We describe the variation between the measurements ρ in two ways. The variation of differences in measurements from the same subject (intra-class variation), and the variation of differences between measurements of different subjects (inter-class variation). We first compute the matrix \mathcal{D} of difference measurements ρ , by determining the set of biometric distances $\mathcal{D}_{i,j} = \|\mathbf{P}_i - \mathbf{P}_j\|$ between the reconstructed gait feature vectors from different camera views. The difference matrix is symmetric and is formed from the same valid set of test subjects that correspond to the treadmill experiments described in section 3.6. The difference matrix then has the block form:

$$\mathcal{D} = \begin{pmatrix} \mathcal{D}_{aa} & \mathcal{D}_{ab} & \mathcal{D}_{ac} \\ \mathcal{D}_{ab}^\top & \mathcal{D}_{bb} & \mathcal{D}_{bc} \\ \mathcal{D}_{ac}^\top & \mathcal{D}_{bc}^\top & \mathcal{D}_{cc} \end{pmatrix} \quad (4.86)$$

where each sub-block is a 3×3 matrix, corresponding to the set of biometric difference measures for the motion reconstructions within the left, middle and right camera views. The diagonal blocks \mathcal{D}_{aa} , \mathcal{D}_{bb} and \mathcal{D}_{cc} represent the set of parameter differences for individual subjects over the three camera views, while the off-diagonal blocks \mathcal{D}_{ab} , \mathcal{D}_{ac} and \mathcal{D}_{bc} represent the set of parameter differences between different subjects.

The intra and inter-class difference measurement sets are then extracted from the elements of \mathcal{D} . Having created both class data sets, we can then find the intra-class mean and variance μ_v, σ_v^2 and the inter-class mean and variance μ_c, σ_c^2 .

4.5.1.1 Variation of reconstructed limb motion from different camera views

We first investigate the variation within individual parameters of the biometric feature vector $\mathbf{P} = (d_2, b_2, \psi_2, b'_2, \psi'_2)^\top$, by evaluation of two properties that are dependent on the class variances σ_v^2 and σ_c^2 .

The class distinction quantity γ describes the percentage of the intra-class variance compared with that of the inter-class variance, i.e. $\gamma = 100 \times \sigma_v^2 / \sigma_c^2$. Low percentages indicate good discrimination between different subjects, while high percentages highlight the inability of the biometric parameter to distinguish between people at all.

The parameter distinction quantity β describes the intra-class parameter deviation as a percentage of the mean subject parameter estimate μ_p , i.e. $\beta = 100 \times \sigma_v / \mu_p$. This percentage gives us an indication of the relative magnitude of the intra-class parameter deviation.

Table 4.3 shows the intra and inter-class variance along with the corresponding levels of class γ and parameter β distinction for the set of reconstructed overground motion parameters. The order of discrimination between parameters is comparably similar to the results of treadmill walking. The static geometric parameter d_2 , corresponding to the normalized lower leg length, is the most discriminating. The phase components have the best discrimination between the normalized limb angle function coefficients.

<i>Parameter</i>	<i>Intra-class</i>	<i>Inter-class</i>	γ	β	γ (TM)	β (TM)
d_2	1.77829e-05	0.007025	0.25%	0.45%	1.61%	0.91%
b_2, b'_2	1.83886e-05	0.000193	38.86%	1.12%	57.64%	3.78%
ψ_2, ψ'_2	2.16145e-06	1.08406e-05	19.94%	0.64%	40.15%	2.03%

TABLE 4.3: Individual discrimination of parameters within the biometric feature vector \mathbf{P} . The class and parameter distinction estimates γ (TM) and β (TM) corresponding to the treadmill motion experiments are also shown for comparison.

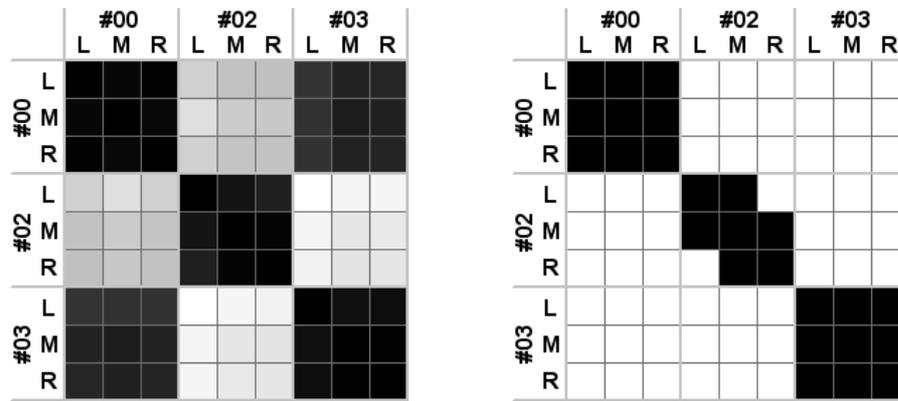
Since image acquisition is synchronous between views and the underlying gait motion is the same, the parameter discrimination quantity β is much better for overground walking than for treadmill motion. The variation between reconstructions in the left, middle and right views can then be attributed to the geometric error associated with the planar limb swing model of leg motion.

Figure 4.33(a) shows the difference matrix \mathcal{D} corresponding to the combined parameters of the proposed biometric feature vector $\mathbf{P} = (d_2, b_2, \psi_2, b'_2, \psi'_2)^\top$. The magnitude of the intra-class variance is estimated at 1.09% of the inter-class variance level. This indicates that the set of reconstructed subject biometric feature vectors are well separated.

We then classify each subject by thresholding the difference matrix \mathcal{D} at the $3\sigma_v$ deviation level. The thresholded difference matrix shown in figure 4.33(b) indicates that the pose reconstruction process is good, with an equal error rate in the region of $\simeq 3.7\%$.

4.5.1.2 Variation of parameters between treadmill and overground walking

We have discussed the intra and inter-class variation of biometric parameters from different camera views of reconstructed motion. We have also discussed within section 3.6.7.2

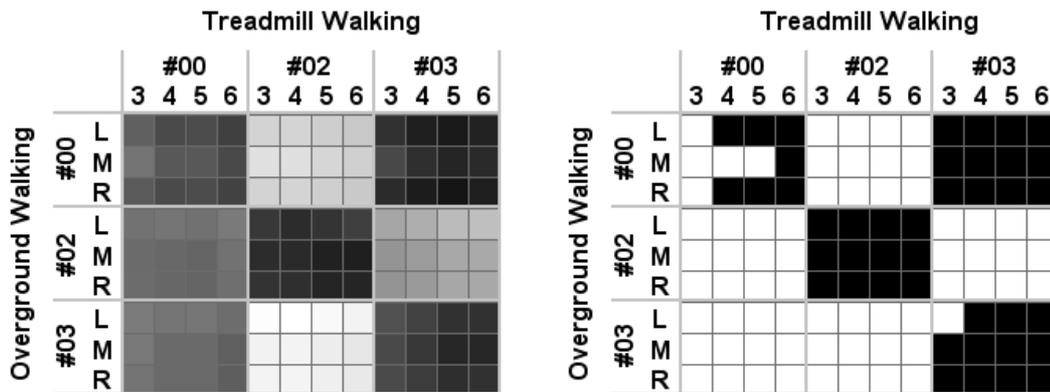


(a) max = 0.142023, intra = 2.56206e-05,
inter = 0.00235811, $\gamma = 1.09\%$.

(b) thresh = $3\sigma_v$, false accept = 0%,
false reject = 7.41%

FIGURE 4.33: Difference matrices \mathcal{D} corresponding to the proposed biometric feature vector $\mathbf{P} = (d_2, b_2, \psi_2, b'_2, \psi'_2)^\top$. (a) The computed difference matrix \mathcal{D} . (b) The thresholded difference matrix at the $3\sigma_v$ deviation level.

the uniqueness of the proposed biometric feature vector, for subject treadmill motion over a number of customary walking speeds. Here, we cross compare the recovered biometric feature vectors for treadmill and overground walking. Since the same set of limb markers were used within both experiments, and these experiments run consecutively as described within the motion capture procedure outlined in section 3.7, we expect the underlying limb motion to be similar for both experiments.



(a) max = 0.151327, intra = 0.000230645,
inter = 0.00152755, $\gamma = 15.1\%$.

(b) thresh = $3\sigma_v$, false accept = 16.67%,
false reject = 16.67%

FIGURE 4.34: Difference matrices \mathcal{D} corresponding to the proposed biometric feature vector $\mathbf{P} = (d_2, b_2, \psi_2, b'_2, \psi'_2)^\top$. (a) The computed difference matrix \mathcal{D} between treadmill and overground walking motions. (b) The thresholded difference matrix at the $3\sigma_v$ deviation level.

Figure 4.34(a) shows the difference matrix \mathcal{D} corresponding to the proposed biometric feature vector $\mathbf{P} = (d_2, b_2, \psi_2, b'_2, \psi'_2)^\top$ between treadmill and overground walking mo-

tions. The magnitude of the intra-class variance is estimated at 15.1% of the inter-class variance level. This level of discrimination is better than the treadmill experiments alone, where the class distinction had a level of $\gamma = 21.03\%$, however the corresponding classification error is higher. This may indicate that the separation between the most dissimilar subject is greater, though the separation between the closest different subject is small enough to cause a significant increase in classification error.

The thresholded difference matrix shown in figure 4.34(b) shows a relatively high degree of classification error, with an equal error rate of $\simeq 16.7\%$. The corresponding equal errors rates for the treadmill and overground walking motions are individually much lower, with errors of $\simeq 3.1\%$ and $\simeq 3.7\%$ respectively. This indicates that the reconstructed biometric features are individually consistent but are different between the two experiments.

There are a number of possible reasons why the intra-class gait features may differ between experiments: i) Gait motions may display differences for treadmill walking because of the unnatural walking surface. ii) The placement of markers changed between experiments. iii) The bilateral limb swing motion model caused a consistent but significant difference in the reconstructed limb lengths between experiments.

The experimental procedure described within section 3.7 was designed to minimize fluctuations in the naturalness of gait motion, by ensuring that subjects have enough time to familiarise themselves with treadmill walking. However, we have little further control over the circumstances that enable us to ensure that motion is natural.

Since the overground walking experiments are run consecutively after those of the treadmill, changes in limb marker positions are unlikely and would only present as a change in the normalized limb length d_2 of the subject in question. Table 4.4 shows the set of normalized limb length parameters d_2 for the subjects during the series of treadmill experiments. The variance σ_d^2 of the normalized limb lengths d_2 remains low for each subject. The individual parameter distinctions $\beta = 100 \times \sigma_d / \mu_d$, that describe the magnitude of the parameter deviation as a percentage of the mean μ_d , indicate that the normalized limb lengths deviate within an order of only 1% of their values.

<i>Motion</i>	<i>Subject 00</i>	<i>Subject 02</i>	<i>Subject 03</i>
3 km/h	0.900003	0.750785	0.946243
4 km/h	0.902166	0.749397	0.957006
5 km/h	0.900241	0.753096	0.972601
6 km/h	0.912581	0.759761	0.973795
mean μ_d	0.903748	0.75326	0.962411
variance σ_d^2	2.67145e-05	1.58323e-05	0.000131007
distinction β	0.57%	0.53%	1.19%

TABLE 4.4: Normalized leg length parameters d_2 for treadmill walking.

Table 4.5 shows the corresponding set of normalized limb lengths d_2 for the overground walking experiments. The variance σ_d^2 of the normalized limb lengths d_2 remains low for each subject. However, the mean parameter estimates μ_d are significantly different between treadmill and overground walking experiments, suggesting that this is either a marker or modelling problem.

<i>Motion</i>	<i>Subject 00</i>	<i>Subject 02</i>	<i>Subject 03</i>
left	0.977539	0.78118	1.026218
middle	0.984784	0.789767	1.012254
right	0.979328	0.792153	1.014623
mean μ_d	0.980551	0.7877	1.0177
variance σ_d^2	9.49426e-06	2.22037e-05	3.72289e-05
distinction β	0.31%	0.6%	0.6%

TABLE 4.5: Normalized leg length parameters d_2 for overground walking.

The variance σ_d^2 of the normalized limb length parameters d_2 is similar for each subject in both experiments. We can then compute the parameter difference Δd_2 between experiments, and express this as a percentage of the mean subject limb length deviation $\sigma_s = 5.74805\text{e-}03$ (all subjects from both experiments), i.e. $\tau = 100 \times \Delta d_2 / \sigma_s$. Table 4.6 shows the mean parameter differences Δd_2 between the treadmill and overground walking experiments. The parameter differences Δd_2 are relatively large and are within the order of $600\% < \tau < 1300\%$, for all subjects between experiments. This indicates that the difference between the parameters of treadmill and overground walking is between six and thirteen times higher than the intra-experimental deviation alone.

	<i>Subject 00</i>	<i>Subject 02</i>	<i>Subject 03</i>
Δd_2	0.076803	0.03444	0.055289
τ	1336.15%	599.16%	961.87%

TABLE 4.6: Parameter differences Δd_2 between treadmill and overground walking.

One difference between the experiments lies in the fact that the treadmill experiments use a single set of limb markers, from the swing plane closest to the camera. Both bilateral limb swing planes are used within the reconstruction of the overground walking motions. It is often difficult to mark the joint features consistently on both sides of the body, therefore the difference between marked limb segment lengths on both planes could account for the dissimilar normalized limb length parameters d_2 in tables 4.4 and 4.5.

The articulated limb motion within the treadmill experiments outlined in section 3.6.1 can be modelled by the partitioned parameter vector:

$$\mathbf{P} = \left(\mathbf{h}^\top, d_2, f_0, \mathbf{V}^\top \mid \mathbf{X}_1^\top, \dots, \mathbf{X}_N^\top \right)^\top \quad (4.87)$$

where \mathbf{h} is the vector of homography coefficients, corresponding to the limb swing plane,

that maps metric plane points into the image. Parameter d_2 is the normalized lower leg segment length, f_0 the fundamental frequency of gait and \mathbf{V} contains the Fourier series coefficients of the upper and lower limb angle functions. The set of hip positions within the sequence are encoded by the subsidiary parameters $\mathbf{X}_i = (u_i, v_i)^\top$.

The algorithm described within section 3.6.1, to compute the representation of leg motion during treadmill walking, requires that we supply an initial estimate of the leg swing plane homography matrix. This was provided by placing a calibration grid against the subject's leg plane while in the quiet standing posture. We can use the computed leg swing plane \mathbf{H}_1 , determined from the coefficients of \mathbf{M} in equation 4.71, as a similar initial estimate in the case of overground walking. We then perform without modification, exactly the same algorithm on the marked joint features of the closest leg, for the overground walking sequences. Table 4.7 shows the corresponding set of normalized limb lengths d_2 for the overground walking experiments, with only a single reconstructed leg swing plane. The results are similar to the bilateral parameterisation of the overground walking shown in table 4.5. This suggests that the landmark features have been marked appropriately on both sides of the body and are not responsible for the parameter differences between treadmill and overground walking experiments.

<i>Motion</i>	<i>Subject 00</i>	<i>Subject 02</i>	<i>Subject 03</i>
left	0.985986	0.793071	1.016425
middle	0.98625	0.789874	1.00853
right	1.000314	0.801851	1.009163
mean μ_d	0.990851	0.794933	1.01137
variance σ_d^2	4.47978e-05	2.56405e-05	1.28316e-05

TABLE 4.7: Normalized leg length parameters d_2 for treadmill walking. A single limb plane is used to reconstruct the motion.

One of the advantages of parameterising the set of frame hip positions $\mathbf{X}_i = (u_i, v_i)^\top$ independently, is that we can disassociate the error caused by modelling the hip displacement motion from the reconstruction process. Table 4.8 shows the root mean squared re-projection error corresponding to the bilateral $\mathbf{P}_m = (\mathbf{M}^\top, d_2, f_0, \mathbf{X}^\top, \mathbf{Y}^\top, \mathbf{V}^\top)^\top$ and the unilateral $\mathbf{P}_h = (\mathbf{h}^\top, d_2, f_0, \mathbf{V}^\top | \mathbf{X}_1^\top, \dots, \mathbf{X}_N^\top)^\top$ parameterisations of articulated limb motion.

We can clearly see a marked reduction in re-projection error for the unilateral parameterisation of motion. This suggests that a significant proportion of the error is due to the mis-modelling of the hip motion. Specifically, the measured difference between treadmill and overground experimental parameters is an artifact of the lateral depth deviation and departure from planarity, of the fixed leg swing plane assumption.

The location of joint markers, the camera set-up and the style of walking were all kept the same for both treadmill and overground walking experiments. In order to make

<i>Method</i>	<i>left</i>	<i>middle</i>	<i>right</i>
\mathbf{P}_m	1.438	2.501	1.985
\mathbf{P}_h	0.804	1.413	0.834

(a) Subject 00

<i>Method</i>	<i>left</i>	<i>middle</i>	<i>right</i>
\mathbf{P}_m	1.405	3.23	2.235
\mathbf{P}_h	0.996	2.383	1.074

(b) Subject 01

<i>Method</i>	<i>left</i>	<i>middle</i>	<i>right</i>
\mathbf{P}_m	1.948	2.041	2.409
\mathbf{P}_h	0.787	1.247	1.114

(c) Subject 02

<i>Method</i>	<i>left</i>	<i>middle</i>	<i>right</i>
\mathbf{P}_m	1.268	1.919	1.743
\mathbf{P}_h	0.849	1.092	1.021

(d) Subject 03

TABLE 4.8: Comparison of root mean square pixel reprojection errors between the bilateral $\mathbf{P}_m = (\mathbf{M}^\top, d_2, f_0, \mathbf{X}^\top, \mathbf{Y}^\top, \mathbf{V}^\top)^\top$ (reprojection error of right leg joint features only) and unilateral $\mathbf{P}_h = (\mathbf{h}^\top, d_2, f_0, \mathbf{V}^\top \mid \mathbf{X}_1^\top, \dots, \mathbf{X}_N^\top)^\top$ reconstruction methods. Errors are compared for subjects in each camera view.

full use of the available resolution, the treadmill was placed half a metre closer to the camera than the path for overground walking. This allows us to fill the camera view with each of the subject’s limb motions, thus allowing us to make accurate observations. The greater depth of the overground walking allows us to image the subject’s motion in all camera views. The disparity between depths, though only relatively small, is the likely cause for the difference in parameters between treadmill and overground walking experiments.

4.6 Reconstruction discussion

Most of the differences between reconstructed gait motions occur in regions of high curvature. This raises several interesting questions about the sources of error in the fitting model, the minimization strategy and indeed the model itself. The discussion given here is intended to show that the author has given some thought to possible improvements that can be made to the reconstruction process, though due to time restrictions was not able to explore any of these avenues further.

4.6.1 Lateral displacement

The abrupt changes in the vertical displacement of the body, caused by reciprocal bipedal locomotion, are minimized by a series of limb motions. These motions reduce the overall muscular effort required to progress the person forward. As a result, the body’s centre of gravity follows a smooth three dimensional sinusoidal path. Figure 4.35 illustrates the resulting lateral deviation of the head and body segments from the progression mid-line of motion.

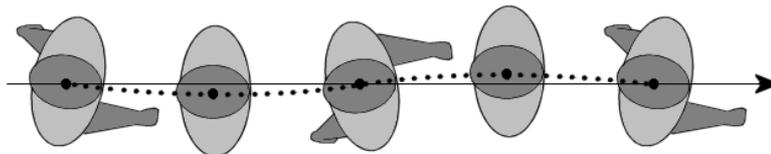


FIGURE 4.35: Lateral displacement.

The standard pixel projection error was discussed within section 3.5.1 to quantify the level of error over the field of view. The standard pixel error of imaged motion at a distance of 3212.3 mm from the camera was computed at 2.8 pixels. This level of error is similar to the root mean square pixel projection errors, shown in table 4.2, after optimization of the gait parameters.

We may model this sinusoidal lateral displacement of subject motion within the parameterisation of gait. We can still assume that the motion dynamics of both limbs is planar and inclined at an angle to the vertical. Each planar subject pose within the sequence is then laterally displaced in the Z direction by the sinusoidal displacement function. Figure 4.36 shows the apparent limb sweep generated by this sinusoidal lateral motion model.

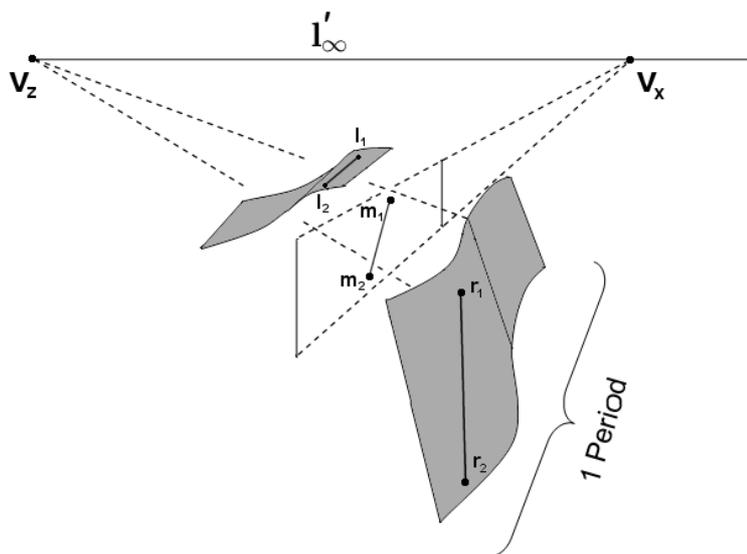


FIGURE 4.36: Sinusoidal lateral displacement of planar limb motion.

The period of the sinusoidal lateral motion is the same as the gait cycle T , thus it can be modelled by a simple harmonic function $z(t) = \gamma \sin(2\pi f_0 t + \psi)$. The amplitude coefficient γ represents the extent of the lateral displacement, while the phase term ψ directly encodes the phase of gait from the neutral double stance position. Double stance poses then occur at zero positions of the simple harmonic function $z(t) = 0$.

$$t = \frac{kT}{2} - \frac{\psi}{2\pi f_0} \quad k = 0, 1, 2, \dots, \infty \quad (4.88)$$

We can compute a subject leg plane pose by first applying a rotation \mathbf{H}_α about the X axis to facilitate the limb swing plane inclination to the vertical, then apply the limb plane selection translation \mathbf{H}_β to map the required hip point to ∓ 1 . We then apply the sinusoidal lateral displacement translation \mathbf{H}_γ in the Z direction. This is followed by a scaling \mathbf{H}_τ in the Z direction that generates the correct distance between both hip points for a subject. The subject pose projection \mathbf{P} then maps the set of worldspace points \mathbf{X} into the image as $\mathbf{x}' = \mathbf{P}\mathbf{X}$.

$$\mathbf{P} = \mathbf{K} [\mathbf{R} \mid \mathbf{t}] \mathbf{H}_\tau \mathbf{H}_\gamma \mathbf{H}_\beta \mathbf{H}_\alpha \quad (4.89)$$

The transformation matrices \mathbf{H}_τ , \mathbf{H}_β and \mathbf{H}_α are identical to those described within the previous sections, while the lateral displacement translation \mathbf{H}_γ has the form:

$$\mathbf{H}_\gamma = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \gamma \sin(2\pi f_0 t + \psi) \\ & & & 1 \end{pmatrix} \quad (4.90)$$

Subsequently, the set of $z = 0$ plane worldspace points $\mathbf{X} = (u, v, 0, w)^\top$ are projected into the image as $\mathbf{x}' = \mathbf{P}\mathbf{X}$.

$$\mathbf{x}' = \mathbf{K} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \mathbf{t} \end{bmatrix} \begin{pmatrix} 1 & & & \\ & \cos \alpha & -\beta \sin \alpha & \\ & \tau \beta \sin \alpha & \tau \cos \alpha & \tau(\beta + \gamma \sin(2\pi f_0 t + \psi)) \\ & & & 1 \end{pmatrix} \begin{pmatrix} u \\ v \\ 0 \\ w \end{pmatrix} \quad (4.91)$$

Projection of planar worldspace structure into the image is then achieved by the corresponding leg plane homography transformations.

$$\mathbf{H} = \mathbf{K} \begin{bmatrix} \mathbf{r}_1 & (\cos \alpha \cdot \mathbf{r}_2 + \tau \beta \sin \alpha \cdot \mathbf{r}_3) & \tau(\beta + \gamma \sin(2\pi f_0 t + \psi)) \cdot \mathbf{r}_3 + \mathbf{t} \end{bmatrix} \quad (4.92)$$

The individual left and right leg plane homography transformations \mathbf{H}_1 and \mathbf{H}_2 then have the form:

$$\mathbf{H}_1 = \begin{bmatrix} \mathbf{m}_1 & (\cos \alpha \cdot \mathbf{m}_2 - \sin \alpha \cdot \mathbf{m}_3) & (\mathbf{m}_4 + \gamma \sin(2\pi f_0 t + \psi) \cdot \mathbf{m}_3 - \mathbf{m}_3) \end{bmatrix} \quad (4.93)$$

$$\mathbf{H}_2 = \begin{bmatrix} \mathbf{m}_1 & (\cos \alpha \cdot \mathbf{m}_2 + \sin \alpha \cdot \mathbf{m}_3) & (\mathbf{m}_4 + \gamma \sin(2\pi f_0 t + \psi) \cdot \mathbf{m}_3 + \mathbf{m}_3) \end{bmatrix} \quad (4.94)$$

where $\mathbf{m}_1 = \mathbf{K} \cdot \mathbf{r}_1$, $\mathbf{m}_2 = \mathbf{K} \cdot \mathbf{r}_2$, $\mathbf{m}_3 = \tau \cdot \mathbf{K} \cdot \mathbf{r}_3$ and $\mathbf{m}_4 = \mathbf{K} \cdot \mathbf{t}$

The elements of the pose transformation mapping \mathbf{M} can be parameterised by a single 15-vector.

$$\mathbf{M} = (\mathbf{m}_1^\top, \mathbf{m}_2^\top, \mathbf{m}_3^\top, \mathbf{m}_4^\top, \alpha, \gamma, \psi)^\top \quad (4.95)$$

4.6.2 Fourier coefficients

Following medical studies [3], which suggest that the maximum frequency content of human walking is ~ 5 Hz and that the fundamental frequency of normal gait is ~ 1 Hz, then a choice of five Fourier harmonics has been employed to model the limb motion dynamics of gait. We have demonstrated that good reconstruction results can be achieved with this number of harmonics. The number of parameters required to fully model the dynamics of the upper and lower leg motion $\mathbf{P}_g = (f_0, \mathbf{D}^\top, \mathbf{X}^\top, \mathbf{Y}^\top, \mathbf{V}_1^\top, \mathbf{V}_2^\top)^\top$ on the metric plane is shown in table 4.9. Together with the 13 parameters of \mathbf{M} required to encode the subject pose, a total of 58 parameters are needed to model the dynamics of leg motion.

<i>Parameters</i>	<i>coefficients</i>	<i>total</i>
f_0	f_0	1
\mathbf{D}	d_2	1
\mathbf{X}	$(v_x, a_1, \phi_1, \dots, a_n, \phi_n)^\top$	11
\mathbf{Y}	$(a_1, \phi_1, \dots, a_n, \phi_n)^\top$	10
\mathbf{V}_1	$(a_0, a_1, \phi_1, \dots, a_n, \phi_n)^\top$	11
\mathbf{V}_2	$(a_0, a_1, \phi_1, \dots, a_n, \phi_n)^\top$	11
		45

TABLE 4.9: Breakdown of parameters required to model the upper and lower legs of gait motion with $n = 5$ Fourier harmonics.

Reconstruction of gait could benefit from a greater number of Fourier harmonics $n \geq 6$ in order to obviate the effect of poor fitting in regions of high curvature. It may even be beneficial to parameterise the hip displacement coefficients \mathbf{X} , \mathbf{Y} with a different number of Fourier harmonics than the leg angle function coefficients \mathbf{V} .

4.6.3 Periodic gait series function

The dynamics of gait have been captured by using a modified Fourier series function $g(t)$ to approximate the articulated leg motion. The Fourier function is formed from an evenly distributed mixture of harmonic contributions. The power spectra of gait harmonics show that most of the energy is associated with the lower order harmonics. The unique features of gait are mostly apparent in the higher order components, where fast changes in musculature structure causes rapid changes in the limb function. This is true, specifically during the pre-swing phase of gait where the body weight is rapidly transferred to the other support limb. The Fourier representation does a respectable job of representing the general motion structure of the limbs. However, it does tend to capture the dynamics of the swing phase, since stance/swing phases (stance phase of one limb is the swing phase of the other) account for 80% of the complete gait cycle. This is illustrated in the limb angle plots within figures 4.25 to 4.32 which show good fitting during periods of rapidly changing angle. The gait function during the loading response and pre-swing phases of gait account for only 20% of the gait cycle but contain a substantial percentage of the expended energy. The apparent smoothing and subsequent misrepresentation of the gait function during such an important phase of gait may possibly be avoided by changing the gait series function.

4.6.4 Propagation of covariance in initial stratified reconstruction

After computing the set of homography mappings \mathbf{H}_{m1} and \mathbf{H}_{m2} , that project structure from the metric plane to the imaged leg swing planes, we may recompute the set of points \mathbf{w}_i and \mathbf{w}'_i on the metric plane that are sampled at frame boundaries, by applying the inverse mappings $\mathbf{w}_i = (\mathbf{H}_{m1})^{-1} \mathbf{x}_i$ and $\mathbf{w}'_i = (\mathbf{H}_{m2})^{-1} \mathbf{x}_i$. We have shown that minimization of the gait function $g(t : f_0, \mathbf{D}, \mathbf{X}, \mathbf{Y}, \mathbf{V})$ on the metric plane does not optimize image reprojection error, hence fitting is biased towards the greatest density of imaged points closest to the epipole. The minimization described in section 4.2.4 to recover the parameters of gait dynamics assumes that all recovered metric plane points have equal measurement deviation errors. Subsequently, the normal equations required to solve the Newton iteration step can be written $\mathbf{J}^\top \mathbf{J} \cdot \delta \mathbf{a} = \mathbf{J}^\top \mathbf{r}$.

If we know the measurement error deviations $\Sigma_{\mathbf{x}}$ of the set of imaged points, then assuming that there is no uncertainty in any of the homography transformations $\mathbf{H}_{\mathbf{m}}$, we can propagate the set of image measurement errors to the metric plane by computing the covariance matrix corresponding to each of the metric plane points \mathbf{w}_i .

$$\Sigma_{\mathbf{w}} = \mathbf{J}_{\mathbf{x}} \Sigma_{\mathbf{x}} \mathbf{J}_{\mathbf{x}}^\top \quad (4.96)$$

where the Jacobian matrix is $\mathbf{J}_{\mathbf{x}} = \partial \mathbf{w} / \partial \mathbf{x}$, computed with the inverse homography

mapping matrices $\mathbf{H}'_{\mathbf{m}} = \mathbf{H}_{\mathbf{m}}^{-1}$, that map the imaged points \mathbf{x} onto the metric plane as $\mathbf{w} = \mathbf{H}'_{\mathbf{m}} \mathbf{x}$. The set of normal equations required to solve the Newton iteration step then have the form.

$$\mathbf{J}^{\top} \Sigma_{\mathbf{w}}^{-1} \mathbf{J} \cdot \delta \mathbf{a} = \mathbf{J}^{\top} \Sigma_{\mathbf{w}}^{-1} \mathbf{r} \quad (4.97)$$

This should obviate some of the bias within the fitting procedure, in order that we obtain a better initial estimate of the gait parameters. However, we have shown that most of the source of error corresponds to the set of computed homography matrices, due in part to the interpolation of spatio-temporal motion structure and the process of back projection.

4.7 Conclusions

Parameterisation of subject motion is split into two phases: i) *Limb stance*, non-linear modelling of the articulated leg motion within the canonical motion plane; ii) *Pose projection*, linear projection of the worldspace subject motion structure into the image.

- **Limb stance.** Human motion is modelled by using a cardboard person assumption. A subject's body and limb parts are represented as a set of repeating spatio-temporal motion patterns within separate planes. The canonical representation of leg motion is approximated by an articulated limb function $g(t)$ on two bilateral swing planes, which are inclined at an angle with the vertical. The dynamics of the gait function on both leg planes are related by a half phase shift, such that one leg undergoes exactly the same motion as the other, only half a gait cycle later. Corresponding left and right leg poses on the metric plane are then determined by evaluation of the biometric limb functions at $g(t)$ and $g(t + T/2)$ respectively, where T is the period of gait.

The canonical representation of gait assumes that a person walks from left to right with constant velocity. Since we are unable to gauge depth from monocular motion sequences, the scale ambiguity of a subject's height is resolved by normalizing the upper leg segment to unit length. We represent the non-linear articulated leg function by the modified Fourier series $g(t : f_0, \mathbf{D}, \mathbf{X}, \mathbf{Y}, \mathbf{V})$, where \mathbf{D} is the vector of normalized leg lengths, \mathbf{X} and \mathbf{Y} are the velocity and Fourier coefficients of the metric plane hip displacement functions, and \mathbf{V} the Fourier coefficients of the upper and lower leg angle functions.

- **Pose projection.** Projection of this planarized human motion model into the image is achieved by a parameterised set of homography transformations \mathbf{M} that encode both the individual leg plane homography mappings \mathbf{H}_1 and \mathbf{H}_2 . Each

planar homography consists of a set of rigid motion transformations (scaling, rotation and translation). A subject's leg plane pose is computed by applying a rotation \mathbf{H}_α about the X axis to facilitate the limb swing plane inclination to the vertical, then applying the leg plane selection translation \mathbf{H}_β to map the required hip point to ∓ 1 . This is followed by a scaling \mathbf{H}_τ in the Z direction that generates the correct distance between both hip points for a subject. The worldspace orientation, subject displacement and subsequent projection into the image is then achieved via the linear pin-hole projection transformation $\mathbf{K}[\mathbf{R} \mid \mathbf{t}]$, where \mathbf{K} is the camera calibration matrix and \mathbf{R}, \mathbf{t} are the camera extrinsic pose matrices. Since the canonical spatio-temporal motion structure of gait is modelled on the metric $z = 0$ plane, projection of articulated leg points into the image is achieved by the homography matrices formed from the first, second and fourth columns of the pose projection transformation.

These two phases of subject motion projection are independent. The *pose projection* step is based on the linear projection of planar geometry into the image, thus enables us to determine an inverse transformation that recovers the canonical motion structure of subject gait from the corresponding imaged features.

We have demonstrated that gait has sufficient properties that allows us to exploit the structure of articulated limb motion, in order to remove the unknown camera and pose ambiguities and reconstruct the underlying gait signature. We have assumed only that a consistent set of limb landmark points can be tracked within an image sequence and that people walk in straight lines, over at least two gait cycles.

We developed a stratified approach to linear trajectory gait reconstruction that uses the constraints of articulated leg motion in order to recover the fronto-parallel view of gait dynamics. The stratification process of computing this rectification transformation is split into three stages: \mathbf{H}_p , \mathbf{H}_a and \mathbf{H}_s . Each stage of the rectification is designed to remove a number of degrees of freedom from the 8 required to fully determine a planar homography. Constraints are formed in each stage by use of properties and invariants associated with the particular class of projective transformation.

- **Perspective transformation.** We first compute the perspective transformation \mathbf{H}_p that recovers the affine properties of the leg swing planes. We initially assume that all of a subject's cardboard limb planes are parallel, thus they all share a common vanishing line within the image. Identification of landmark correspondences and subject periodicity is solved simultaneously by computing the self similarity of structure over the image sequence. The imaged positions of repeated gait poses are related by a conjugate translation of the leg swing plane. We combine constraints from left and right leg planes in order to determine the coefficients of both transformations. We then extract the shared vanishing line

of the limb swing planes from these coefficients and subsequently compute the required perspective transformation \mathbf{H}_p that restores the affine properties of the limb planes.

- **Affine transformation.** Metric properties of the plane are then recovered by identifying the images of the circular points \mathbf{I}' and \mathbf{J}' from the fixed lengths of the tracked leg segments throughout the image sequence. Linear constraints on the elements of the affine transformation \mathbf{H}_a , that maps the imaged circular points back to their canonical values, are computed from pairs of corresponding leg segments taken at different frame positions.
- **Similarity transformation.** The scale ambiguity on both leg planes is removed by constraining the upper leg segment to be of unit length. The circular points \mathbf{I}, \mathbf{J} remain fixed under any similarity transformation \mathbf{H}_s , thus the remaining translational ambiguity is resolved by enforcing the bilateral symmetry constraint between both planes of recovered gait motion. We then compute a robust estimation of the gait motion function by fitting the articulated leg motion on both limb planes to a single modified Fourier series function.

Details of two different reconstruction methods were given that compute the maximum likelihood estimates of the corresponding set of parameters. The methods differ only in the way that the metric plane points are projected to the images of the leg swing planes.

The first uses a projectivity to facilitate the leg swing plane inclination angle. Experimental analysis shows that the fitting procedure is sensitive to the distribution of points between both planes. The unbalanced numbers of points, due to missing or occluded points on one of the planes, biases the reconstruction in favour of the data on the other plane.

The second uses only rigid motion transformations (rotation, translation and scale) to parameterise the plane poses. This offers a more realistic parameterisation of the mapping between metric and image planes. Correspondingly, this method demonstrates better reconstruction results over those obtained from the first method. An analysis of the reconstructed gait functions, corresponding to a trial set of four people, shows that each subject's signature matches quite well over a number of different camera viewpoints.

We have identified a number of issues that need to be considered further. A discussion was given within the previous section that addressed many of these concerns, though only from a theoretical perspective. More work needs to be done in order to validate the usefulness of this research and demonstrate its practical significance within the field of biometric identification.

- **Periodicity from self-similarity of pixel correspondences.** The task of determining periodicity and point correspondences over the image sequence has been

made much simpler by the fact that we have manually marked the landmark interest features. There is a wide range of literature on periodicity detection and motion classification. The self-similarity based periodicity detection method outlined within this chapter is most closely related to the work of Cutler and Davis [21, 22]. Their work compares re-scaled image regions corresponding to a tracked subject, in order to determine the periodicity of self-similar pixel structures. Our method extends this work further by enforcing the geometric constraints of repeating planar motions, through identification of the imaged conjugate translations corresponding to subject motion. Our self-similarity method could easily be developed to enable periodicity detection and correspondence matching from pixel regions alone. Identification of the conjugate translation enables us to recover the affine properties of the subject motion. To be of any practical interest, further investigation is required to enable segmentation of the required landmark features from the affinely recovered image regions of subject motion.

- **Lateral displacement.** A series of limb motions are employed to smooth abrupt changes in the vertical displacement of the body. As a result, the head and body deviate laterally from the progression mid-line. A more realistic motion model can be developed to account for this type of displacement. We still assume that the motion dynamics of both legs is planar and inclined at an angle to the vertical. Each planar subject pose within the sequence is then laterally displaced in the Z direction by the simple harmonic lateral displacement function $z(t) = \gamma \sin(2\pi f_0 t + \psi)$. This parameterisation is non-linear and dependent on the position of the subject within the image sequence, thus can only be modelled within the maximum likelihood estimation procedure. Initial estimates for the motion model parameters are first computed via the stratified reconstruction method.
- **Periodic gait series function.** The dynamics of gait have been captured by using a modified Fourier series function $g(t)$ to approximate the articulated leg motion. The unique features of gait are mostly apparent in the higher order Fourier components, where abrupt changes in the musculature structures causes rapid changes in the limb function. The Fourier motion representation tends to capture the dynamics of the swing phases better, since they account for 80% of the complete gait cycle. The apparent smoothing and subsequent misrepresentation of the limb function during the loading response and pre-swing phases of gait may possibly be avoided by changing the periodic gait reconstruction basis function.
- **Reconstruction error analysis.** While the work presented in this chapter demonstrates that subject motion can be recovered from many viewpoints, little has been done to analyse the major sources of reconstruction error. Further work needs to be performed to test the robustness of the method to the presence of noise and other sources of imaging error such as camera radial lens distortion. In order for the method to have any practical application, we must be able to

quantify the level of uncertainty within each of the reconstruction parameters. We must then determine what level of landmark measurement error is acceptable in order for us to compute a reliable set of biometric motion features.

Further work needs to be done in order to validate the invariance of the reconstructed subject motion to changes in walking speed. The work presented within section 3.6 demonstrated that suitable biometric features could be extracted from the reconstructed motion of subjects walking on a treadmill, over a number of controlled walking speeds. We need to validate these same assumptions in light of the reconstructed subject motion of overground walking. Subsequently, we should also compare the reconstructions corresponding to both treadmill and overground walking, over a number of walking speeds.

Chapter 5

Total Parameterisation of Generalized Gait Motion

5.1 Introduction

We have shown in the previous chapter that the canonical motion parameters of gait can be reconstructed along linear trajectories, within single view image sequences. Since people tend to walk from point to point in straight lines then any generalized gait motion can be approximated by a set of straight line motion segments. Each piecewise linear segment of reconstructed gait motion has a canonical view leg motion function and the corresponding pose projection parameters. Common to all sets of subject pose parameters are the intrinsic parameters of the camera. Common to each subject, within all of their reconstructed linear motion segments, are the underlying biometric parameters of limb motion.

The work presented in this chapter is concerned with recovering the camera intrinsic parameters and subsequently the set of worldspace subject poses from the parameterised limb swing plane mappings. We first describe the specialized geometry of piecewise linear gait motion within a fixed ground plane, and develop a strategy to segment this generalized motion into linear segment blocks. Parameterisation of consistent worldspace motion is a three step process: i) *Reconstruction*, perform the algorithm described within chapter 4 to determine the motion parameterisation of each individual trajectory segment; ii) *Fusion*, combine the set of independent trajectory motions for a subject into a single representation of the underlying motion parameters, and decompose the set of pose projection homographies into a form consistent with the constrained geometry of ground plane motion; iii) *Optimization*, compute the maximum likelihood estimation corresponding to the parameterisation of worldspace motion and analyse the uncertainty within each of the recovered parameters.

Each of these steps forms an essential part of the complete gait reconstruction algorithm. We analyse the performance of each of the stages within this algorithm, by comparing the root mean squared image reprojection errors for a number of proposed methods. The reconstruction algorithm centres about a single example image sequence with multiple trajectories of subject motion. This image sequence is recorded from subject motion around a figure of eight test track in an outdoor environment. Consequently, we first describe the geometry of generalized ground plane motion and outline the details of the experimental set-up. Further details and subsequent analysis of the motion reconstruction then correspond to our overall attempt to recover the best parameterisation of worldspace gait consistent with this experiment.

There is an abundance of literature on camera calibration. Researchers have previously shown that the intrinsic parameters of the camera can be recovered by identifying the image of the absolute conic [29]. The IAC may be computed from a set of homographies that map structure from the metric reference plane to the corresponding projected image planes. Details of the IAC and camera calibration were previously described within section 2.5.3. In general, readily identifiable planar patterns such as chess board grids are used to compute these planar homographies. Calibration techniques that are based on such planar mappings have been described in the literature by Sturm and Maybank [97], Liebowitz and Zisserman [63], Trigs [107] and Zhang [120]. The reconstructed gait pose projection parameters \mathbf{M} , described in the previous chapter, encode the set of limb swing plane homographies. Subsequently, the familiar arguments and principles of the calibration literature can be followed in order to recover the required intrinsic camera parameters. Factorization of the set of subject pose projection mappings into camera intrinsic and extrinsic worldspace pose parameters is then possible for the constrained type of ground plane gait motion.

The form of the underlying biometric function that represents articulated leg motion remains somewhat of a research topic. The biometric should ideally be invariant to changes in stride, cadence and walking speed. Researchers have previously shown that a modified Fourier representation of the leg angle function offers some invariance to these changes within a range of walking speeds [19, 12, 117]. This mapping does not hold across all modes of gait, though there may be unique mappings for walking and running independently [119]. Reconstruction of each of the individual linear trajectory segments generates a corresponding set of similar underlying subject biometric parameters. We describe a suitable biometric motion function to model the dynamics of gait and develop a method to fuse the set of independent representations into a single parameterisation. We compare two methods that fuse the set of biometric parameters; one linear and another that computes the maximum likelihood estimation by minimizing image reprojection error.

Subjects typically walk on a flat ground plane, hence the configuration and parameterisation of subject motion is specialized further. The projective nature of planar motion

is well understood [4, 31, 96]. Parameterisation of the complete reconstruction model is then partitioned into three sections. The ground plane and camera intrinsic parameters are common to all frames within the image sequence and form the set of *system* parameters. Each subject has a unique underlying gait motion function that is encoded by the biometric limb coefficients, thus forming the set of *subject* parameters. Each subject then has an arbitrary number of *pose* parameter segments, where each pose segment in the piecewise linear motion sequence encodes: the angle, velocity and initial position of motion on the ground plane, and a number of other circumstantial parameters that describe the cadence and initial phase offset within the canonical gait cycle. Details of a sparse minimization technique is then given that computes the maximum likelihood estimate of the corresponding set of partitioned parameters over the entire motion sequence.

The chapter concludes with an analysis of the reconstruction results, corresponding to the example image sequence of subject gait containing multiple trajectories of motion. Subsequently, a discussion on the measured uncertainties within the parameterisation is given and a number of explanations for the major sources of error considered further. Finally, details for further development and possible improvements are outlined.

5.2 Imaged ground plane motion

A further specialization of the epipolar geometry occurs when the cameras are related by motion within a plane. This is the dual situation to a person walking with unconstrained motion in the ground plane. In this case, the rotation axis between views is orthogonal to the set of camera translation directions, as illustrated in figure 5.1. Orthogonality then imposes one constraint on the motion [71].

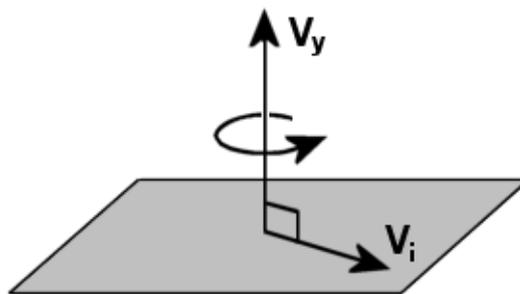


FIGURE 5.1: Unconstrained planar subject motion.

Consider the set of worldspace points \mathbf{X}_i that are projected to the same point $\mathbf{x}_i = \mathbf{x}'_i$ in two image views, i.e. $\mathbf{x}_i = \mathbf{P}\mathbf{X}_i = \mathbf{P}'\mathbf{X}_i$. These imaged points \mathbf{x}_i are then fixed under the camera motion $\mathbf{P} \mapsto \mathbf{P}'$. The corresponding fundamental matrix \mathbf{F} can be split into symmetric \mathbf{F}_s and anti-symmetric \mathbf{F}_a parts.

$$\mathbf{F} = \mathbf{F}_s + \mathbf{F}_a \quad (5.1)$$

$$\mathbf{F}_s = \frac{\mathbf{F} + \mathbf{F}^\top}{2} \quad (5.2)$$

$$\mathbf{F}_a = \frac{\mathbf{F} - \mathbf{F}^\top}{2} \quad (5.3)$$

The anti-symmetric part \mathbf{F}_a may be written $\mathbf{F}_a = [\mathbf{v}]_\times$ and is of rank 2, with 2 degrees of freedom identified by the homogeneous point \mathbf{v} . The mapping $\mathbf{x}^\top \mathbf{F}_a \mathbf{x}$ is identically zero.

$$\begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \begin{bmatrix} 0 & -a & b \\ a & 0 & -c \\ -b & c & 0 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad (5.4)$$

$$-ax_1x_2 + bx_1x_3 + ax_1x_2 - cx_2x_3 - bx_1x_3 + cx_2x_3 = 0 \quad (5.5)$$

Consequently, only the symmetric part of \mathbf{F} contributes to the correspondence condition $\mathbf{x}^\top \mathbf{F} \mathbf{x} = 0$. Geometrically, \mathbf{F}_s can be thought of as a conic in the image plane. The locus of all worldspace points \mathbf{X}_i for which $\mathbf{x}_i = \mathbf{x}'_i$ is known as the *horopter* curve, which is generally a twisted cubic in 3-space passing through the two camera centres.

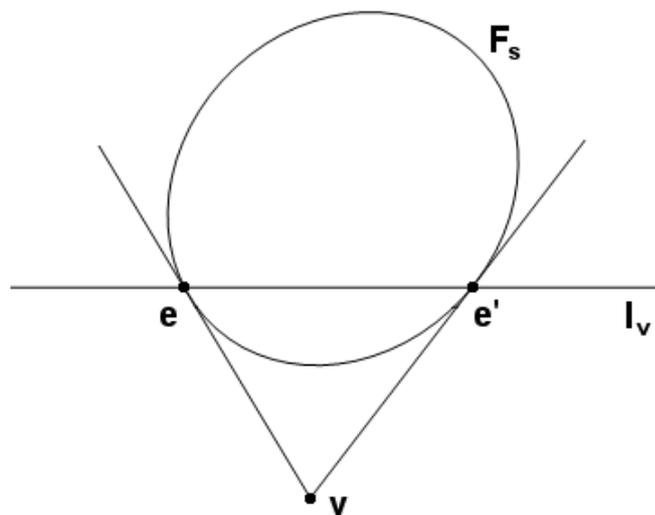


FIGURE 5.2: **Geometric representation of \mathbf{F} .** The conic \mathbf{F}_s represents the symmetric part of \mathbf{F} , and the apex point \mathbf{v} the skew-symmetric part. The conic \mathbf{F}_s is the locus of intersection of corresponding epipolar lines and represents the image of the *horopter* curve.

The image of the *horopter* is the conic defined by \mathbf{F}_s , called the *Steiner conic* [92], which

plays an important part in image auto-calibration. \mathbf{F}_s is symmetric and in general of rank 3, with a total of 5 degrees of freedom. Both epipoles \mathbf{e} and \mathbf{e}' lie on the conic \mathbf{F}_s and on the polar line \mathbf{l}_v of the apex point \mathbf{v} .

In planar camera motion the rotation axis is orthogonal to the translation direction, thus orthogonality imposes one constraint on the motion. It was shown by Maybank [71], that if the rotation axis direction is orthogonal or parallel to the translation direction then the symmetric part of \mathbf{F} is of rank 2, thus the *Steiner* conic is degenerate and is equivalent to two non-coincident lines \mathbf{l}_s and \mathbf{l}_v .

$$\mathbf{F}_s = \mathbf{l}_v \mathbf{l}_s^\top + \mathbf{l}_s \mathbf{l}_v^\top \quad (5.6)$$

Figure 5.4(a) shows the geometric construction of the degenerate *Steiner* conic corresponding to planar camera motion. An arbitrary camera rotation and translation within the worldspace motion plane can be decomposed into a single rotation about a single screw axis. Figure 5.3 illustrates the similarity between an arbitrary 2D rigid motion transformation of an object and its screw decomposition.

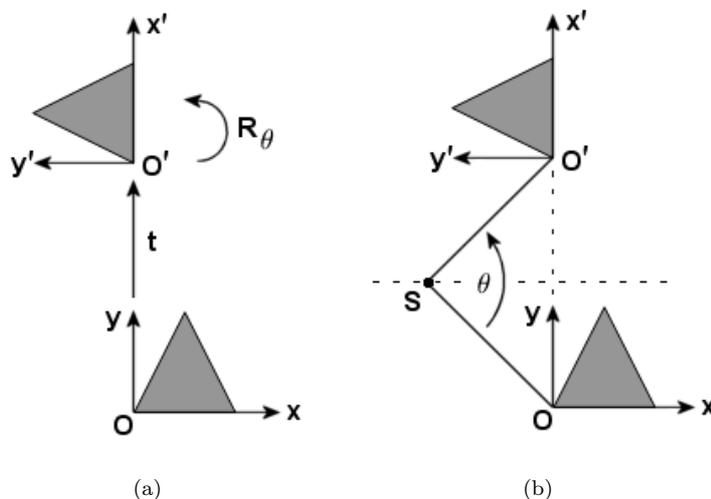


FIGURE 5.3: **A general 2D Euclidean motion and its screw decomposition.** (a) The object undergoes a translation \mathbf{t} and a rotation by an angle θ . The motion is in the plane orthogonal to the rotation axis. (b) This motion is equivalent to a single rotation about the screw axis \mathbf{S} .

The screw axis \mathbf{S} constitutes a line of fixed points within the worldspace, and consequently is imaged as the fixed line \mathbf{l}_s within both views. The screw axis \mathbf{S} is parallel to the normal of the motion ground plane, thus two such identified screw axes are also parallel and meet at the ideal worldspace point in the direction of the plane normal. The image of this ideal point is the apex point \mathbf{v} , which corresponds to the anti-symmetric part of \mathbf{F} , as shown in figure 5.4(b). The vanishing ground plane horizon line is also imaged as the fixed line \mathbf{l}_v between the two images, though is not a line of fixed points.

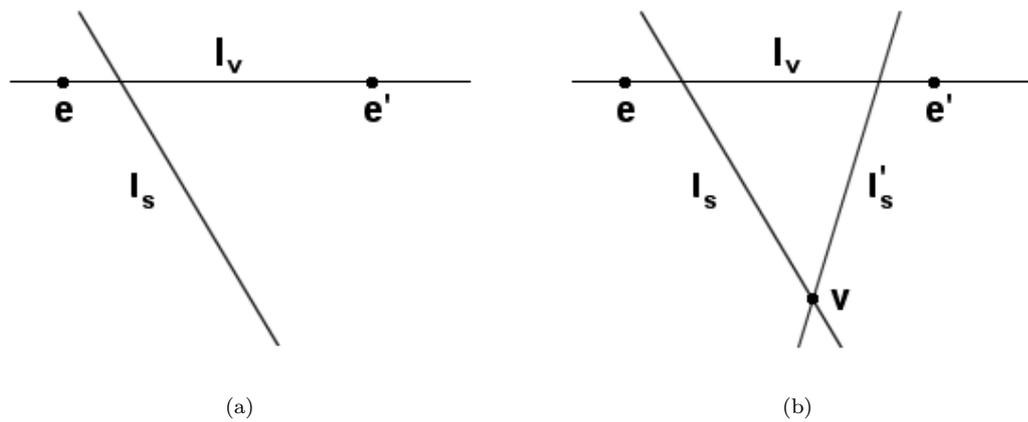


FIGURE 5.4: **Geometric representation of F for planar motion.** (a) Degenerate Steiner conic corresponding to planar motion consists of two lines l_v the image of the vanishing ground plane horizon line, and l_s the image of the screw axis of rotation between the two cameras. (b) Intersection of two different screw axes meets at the apex point v , which represents the anti-symmetric part of F .

The same geometric construction occurs for subject motion within the ground plane. The image of the vanishing line of the ground plane l_v and the imaged vertical axis direction v remain fixed throughout the image sequence.

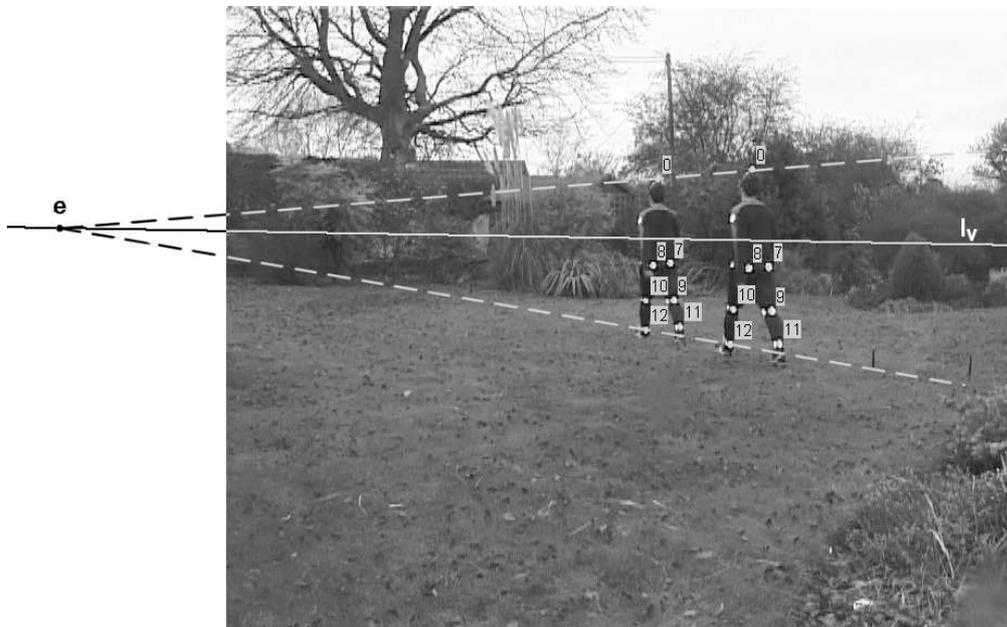


FIGURE 5.5: **Imaged ground plane motion.** Each imaged direction of subject motion e_i lies on the image of the ground plane l_v . This imaged ground plane line l_v and the imaged vertical axis direction v (not shown) are fixed throughout the image sequence.

Subject motion is assumed to be piecewise linear, thus the set of imaged subject motion trajectory epipoles e_i must all lie on the fixed ground plane vanishing line l_v . Consequently, the vanishing line l_v can be recovered from a minimum of two different linear

motion trajectories. Figure 5.5 shows a composite image of two similar subject poses, within a single linear motion trajectory segment. We can clearly see that the imaged subject motion is consistent with the geometric construction of planar motion.

5.3 Experimental set-up

A figure of eight test track, illustrated in figure 5.6, is marked out and a subject is recorded walking around it. The motion sequence is captured from a single static camera **C**. There are four marked way-points (0-3) in the track that define six possible straight line trajectory segments. Five of these linear trajectory segments have useful image data. Figure 5.7 shows a composite image of the corresponding figure of eight test track with a number of marked trajectory segments.

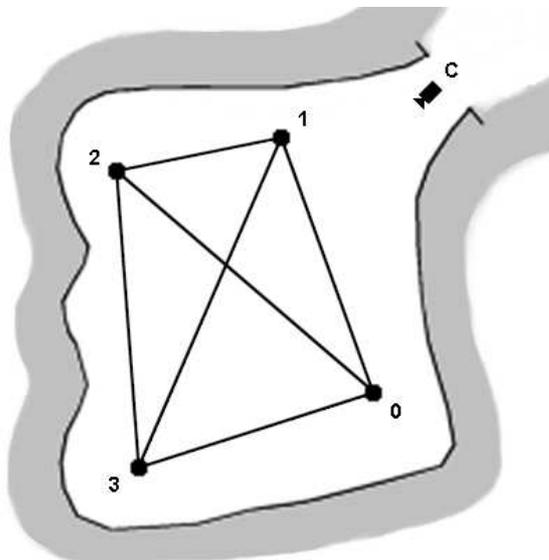


FIGURE 5.6: Plan view of the figure of eight test track with four way-point markers (0 - 3) and a single static camera **C** used within the experimental set-up.

The camera is placed a distance of at least ten metres from the closest point of approach within the linear trajectory segments. Two light sources (TD beam 800W) are placed behind the camera and orientated to fully illuminate the test track. The light reflected from retro-reflective marker patches, that are attached to the subject's limbs, provides suitable high contrast pixel regions that enables us to manually mark the required joint features within the set of images.

The imaged gait motion is captured by using a Sony DCR TRV 900e digital camcorder with a frame rate of 25 fps and shutter speed of 1/250 seconds. The frames are interlaced with a resolution of 720×576 pixels. Subsequently, we can de-interlace the set of captured images to produce a similar image sequence with an effective frame rate of 50 fps. Splitting an image into even and odd scan lines halves the vertical height, hence each image is resized and the missing scan line pixels interpolated from its neighbours.

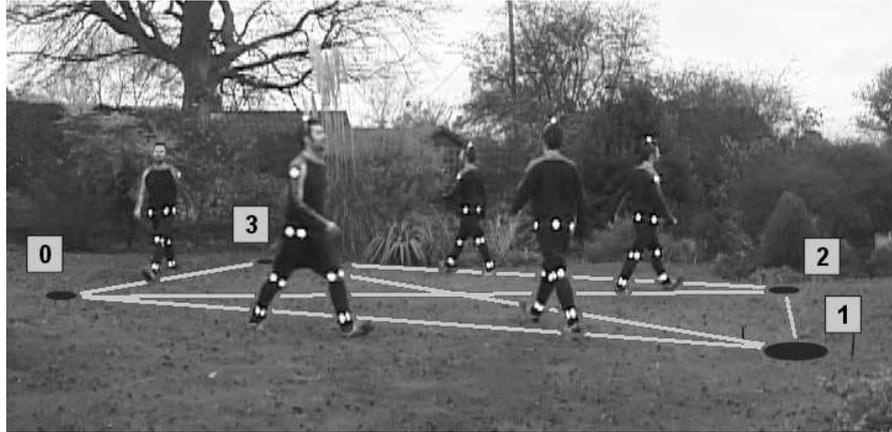


FIGURE 5.7: Image of the figure of eight test track with four way-point markers (0 - 3).

We take care to preserve the positions of the valid pixels, i.e. this is not simply: split into two half images then rescale them vertically by a factor of two. The interpolation strategy is dependent on the parity of the scan line fields.

We then calibrate the camera in order to accurately determine the set of true baseline calibration coefficients. We use Zhang's calibration algorithm [120], with at least thirty different imaged poses of the calibration target, in order to determine an accurate representation of the camera model.

5.3.1 Markers

A subject who walks around the test track is imaged with a full 360 degrees of pose orientation. Consequently, occlusion is a major problem when manually marking the set of joint features within the captured image sequence. Attaching a single set of retro-reflective markers to the principal joint sites on the outside of a subject's limbs is insufficient, since we are unable to see these markers at more oblique trajectories (self-occlusion from the same limb). In order for us to determine the set of landmark features over the full range of poses, we need to place markers on all four sides of each limb segment. Figure 5.8 shows this arrangement of markers viewed from a number of different subject poses.

Accuracy of the motion reconstruction is dependant on the placement of the set of joint markers. Since four markers are used to define a single joint location, then any misalignment due to poor placement will also be reflected in the computed reconstruction. Since the set of markers on each side of the limb have a distinct planar configuration, we can fairly easily ensure that the length ratios between joint markers is similar. However, while the set of markers on a limb share the same underlying gait motion, we must be aware that different marker sets on each side of the limb may be inclined at slightly different angles to the vertical, and have different separation distances between

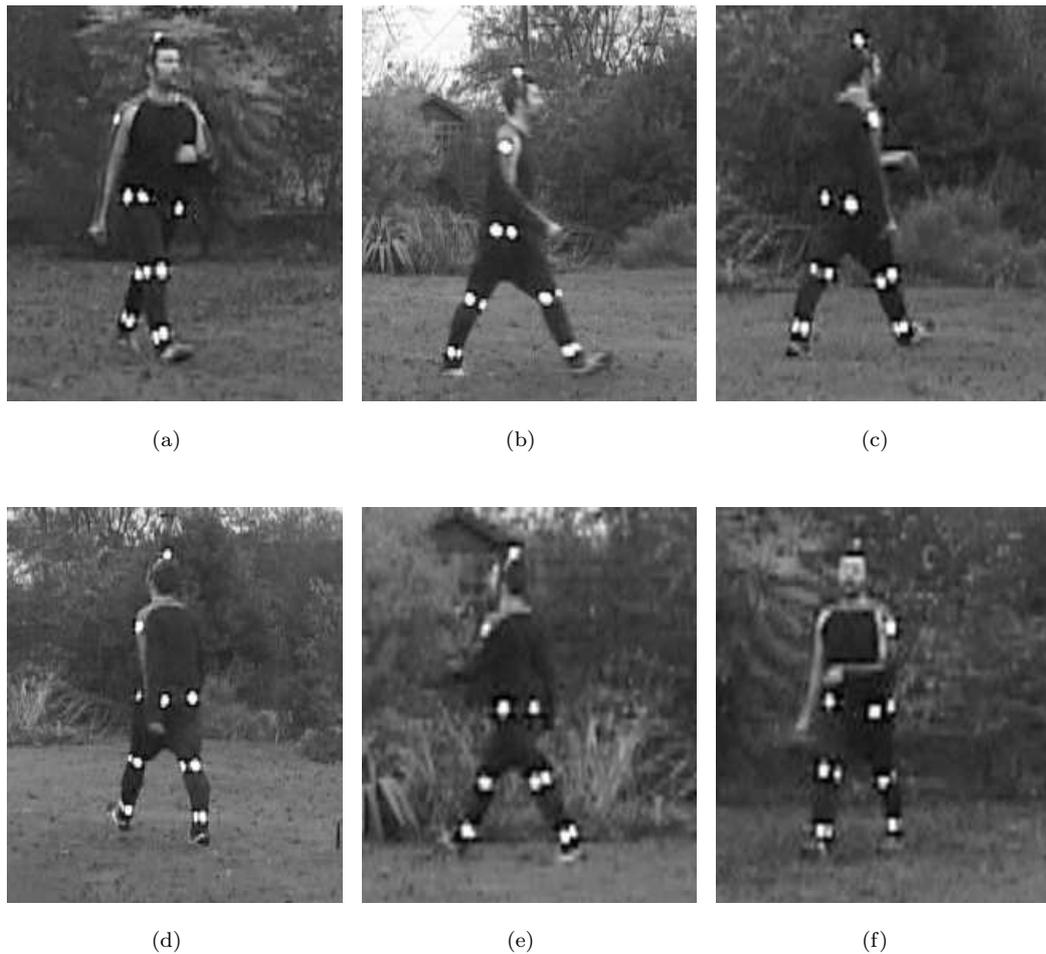


FIGURE 5.8: **Configuration of landmark points attached to the body.** To ensure that a limb joint is visible within the image sequence four retro-reflective patches are stuck to the opposing faces of each limb segment.

hip markers. We must then take care to pick joint features from the same side of each limb segment, within each piecewise linear trajectory of the image sequence, to ensure that we compute a reconstruction consistent with the planar limb motion assumption.

5.3.2 Motion segmentation

Each assumed linear trajectory segment requires a minimum of two complete gait cycles in order to compute a valid reconstruction. We choose to divide each linear motion segment into image sub-sequence blocks of between two to three seconds, in order that we preserve the linear trajectory assumption. An experimental analysis of normal gait motion patterns gives an empirical frequency value for the gait cycle close to 1 Hz, thus the size of each sequence data block is similarly proportional to the number of imaged gait cycles.

The reconstruction algorithm described in chapter 4 gives a constraint on the minimum

number of frames required to make a valid rectification. On the other hand, there is no such constraint on the sample separation of the set of sequence data blocks, i.e. each block of imaged data must be N frames long, though the number and sample spacing between each of these data segments within the M frames of the complete image sequence may be chosen by the experimenter.

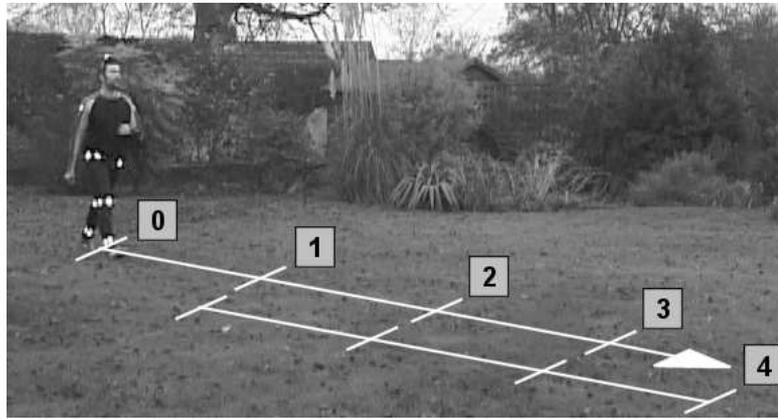


FIGURE 5.9: Marked spatial positions of gait at integer second intervals. Two trajectory segments are shown each with three second sub-sequence data blocks and a sample separation of one second. We can clearly see the overlapping region of data between the two segments.

The maximum number of sample data blocks is achieved with a sample separation of a single frame, such that $N - 1$ frames overlap with the preceding data blocks. The minimum number of sample data blocks is achieved with a spacing of N frames, with no corresponding shared frames. A choice of sample separation somewhere in between gives a good trade-off between reliable estimation of gait reconstruction within curved trajectories and the computational overhead required to process the data.

Figure 5.9 shows the segmentation strategy employed here. The sample separation is equivalent to one second of data, approximately every gait cycle. If we choose the sample data size N to be equivalent to three seconds of data then two gait cycles are shared between three consecutive data segment blocks.

The data segment blocks can then be analysed so that we can compute the true set of periodicity and imaged motion epipoles. Section 4.2.1 outlines the basic principle for determining the subject periodicity and motion direction. In order to find within the image sequence the set of linear motion segments, we need to break the image sequence at points of high pose curvature. The periodicity fitting function and the change of imaged motion direction can be used to determine these high curvature positions. For now we assume that this problem has been solved and proceed to pick our linear motion segments manually.

Following identification of the putative linear trajectory segments, we then re-section the set of sub-sequence data blocks. Tables 5.1 and 5.2 show the division of sub-sequence

data blocks corresponding to the five straight line trajectory segments between way-points in the test track.

<i>Name</i>	<i>Start Pt</i>	<i>End Pt</i>	<i>Num Frames (N)</i>	<i>Cycles (N/fps)</i>	<i>rem</i>
a	0	1	160	3	10
b	2	3	192	3	42
c	3	0	127	2	27
d	0	2	189	3	39
e	1	3	220	4	20

TABLE 5.1: Computation of the putative number of gait cycles within each of the five straight line trajectory segments. We assume that the periodicity of normal gait is approximately 1 second then compute the number of cycles and remaining frames available for division.

<i>Name</i>	<i>Num Frames</i>	<i>Block size</i>	<i>Cycles</i>	<i>trunc.</i>
a	160	53	3	1
b	192	64	3	0
c	127	63	2	1
d	189	63	3	0
e	220	55	4	0

TABLE 5.2: Each gait cycle frame block size is nominally fps frames in size. We then evenly distribute the remaining frames over the available gait cycles so that the block size $fps + (rem / Cycles)$ evenly divides the number of sequence frames with minimal truncation.

We then break the set of trajectories into piecewise linear motion segments (a, b, c, d, e1, e2) such that each motion segment has three or less cycle sample blocks. Table 5.3 shows the root mean square reprojection errors corresponding to the individual sub-sequence data blocks after performing the reconstruction algorithm outlined within chapter 4.

<i>Segment Name</i>	<i>Stratified</i>	<i>MLE</i>
a	4.675	3.102
b	2.853	1.048
c	2.648	1.321
d	1.526	1.398
e1	2.865	1.619
e2	2.068	1.189

TABLE 5.3: Piecewise linear segment reprojection error. Root mean squared pixel errors for both the initial stratified reconstruction and after the bundle adjustment procedure.

5.4 System parameterisation

The camera projection model can be parameterised in many ways. In many cases a linear pin hole model is good enough. One major source of uncertainty within reconstruction

is the component of radial distortion caused by the camera lens. We need to use a model that incorporates these distortion effects, in order to compute an accurate reconstruction. The camera model employed here has three distinct steps within the projection process.

Extrinsic worldspace pose. The camera and the worldspace coordinate system are related by a worldspace rotation \mathbf{R} and translation \mathbf{t} . The ideal camera projects a worldspace point \mathbf{X} as $\mathbf{w} = [\mathbf{R} \mid \mathbf{t}] \mathbf{X}$. Correspondingly, the ray entering the lens is represented by the inhomogeneous point $\tilde{\mathbf{w}} = (w_1/w_3, w_2/w_3)^\top$.

Lens distortion. The main distortion effect is seen radially from the camera principal point. Tangential distortion effects may also be modelled, but in general are negligible in comparison to these radial components. We use here a symmetric distortion model $\mathbf{u} = f(r) \cdot \tilde{\mathbf{w}}$ with two radial components $\mathbf{K}_r = (k_1, k_2)^\top$ to facilitate the warping of light through the camera lens.

$$f(r) = 1 + k_1 \cdot r^2 + k_2 \cdot r^4 \quad (5.7)$$

where the radial distance component r is a function of the ray point $\tilde{\mathbf{w}} = (\tilde{w}_x, \tilde{w}_y)^\top$ from the camera principal point, i.e. $r^2 = (\tilde{w}_x)^2 + (\tilde{w}_y)^2$.

Linear projection. The focal projection of the ray leaving the lens onto the camera CCD elements is modelled by the pin hole projection $\mathbf{x}' = \mathbf{K}\mathbf{x}$, where $\mathbf{x} = (u_x, u_y, 1)^\top$ is the post lens distortion point. We use a camera calibration matrix \mathbf{K} that has zero skew and can be parameterised by the 4-vector of focal and principal point coefficients $\mathbf{K}_c = (m_x, m_y, u_0, v_0)^\top$.

$$\mathbf{K} = \begin{pmatrix} m_x & & u_0 \\ & m_y & v_0 \\ & & 1 \end{pmatrix} \quad (5.8)$$

We assume that the camera intrinsic parameters remain fixed throughout the duration of the image sequence. We can also fix the worldspace pose so that camera and worldspace coordinate systems coincide, thus the extrinsic pose projection step within the camera model has the simpler form $\mathbf{w} = [\mathcal{I} \mid \mathbf{0}] \mathbf{X}$. The only further assumptions we employ are that the ground plane is flat and subject motion is piecewise linear within this plane.

5.5 Biometric parameterisation

A brief analysis of the fundamental properties of gait was given within section 3.6 for subjects walking on a treadmill. We demonstrated that the dynamics of gait can be

represented by a simple modified Fourier function that is pseudo-invariant to changes in stride length and cadence within a range of normal walking speeds.

Let the Fourier series representation of an arbitrary signal be defined by the equation.

$$x(t) = a_0 + \sum_{k=1}^n a_k \cos(2\pi k f_0 t + \phi_k) \quad (5.9)$$

A time shifted signal $x(t - t_s)$ only updates the coefficients of phase within the Fourier series representation.

$$x(t - t_s) = a_0 + \sum_{k=1}^n a_k \cos(2\pi k f_0 \cdot (t - t_s) + \phi_k) \quad (5.10)$$

$$x(t - t_s) = a_0 + \sum_{k=1}^n a_k \cos(2\pi k f_0 t + \psi_k) \quad (5.11)$$

$$\psi_k = \phi_k - 2\pi k f_0 t_s \quad (5.12)$$

It is unclear which features of gait represent the origin pose position within the gait cycle. It is then natural to align the Fourier signals by computing the time shift that zeros the first coefficient of phase $\psi_1 = 0$, i.e. $t_s = \phi_1 / 2\pi f_0$. Correspondingly, $\psi_k = \phi_k - k\phi_1$ are the aligned phase coefficients of the modified series. The signal can then be made invariant to scale by normalizing the first amplitude coefficient a_1 to unity, such that the set of normalized amplitudes b_k are given by $b_k = a_k / a_1$. The modified Fourier series representation of the original signal $x(t)$ with added linear velocity term v_x then has the form.

$$x(t) = v_x t + a_0 + a_1 \cos(2\pi f_0 \cdot (t + t_s)) + a_1 \cdot \sum_{k=2}^n b_k \cos(2\pi k f_0 \cdot (t + t_s) + \psi_k) \quad (5.13)$$

The set of modified coefficients $\tilde{\mathbf{v}} = (b_2, \psi_2, \dots, b_n, \psi_n)^\top$ form the basis for a biometric parameter vector. The remaining parameters $\tilde{\mathbf{w}} = (v_x, f_0, a_0, a_1, t_s)^\top$ of the modified Fourier series function then form the set of *circumstantial* parameters of motion.

The normalized gait function $\tilde{x}(t)$ formed from the set of biometric parameters $\tilde{\mathbf{v}}$ alone describes the unique underlying limb dynamics of gait motion, and is invariant to initial subject position, stride length and cadence.

$$\tilde{x}(t) = \cos(2\pi t) + \sum_{k=2}^n b_k \cos(2\pi kt + \psi_k) \quad (5.14)$$

Figures 5.10 and 5.11 show the set of normalized leg angle functions $\tilde{\theta}(t)$ generated from each of the individual piecewise linear trajectory segments. We can clearly see, that to a first order approximation, the signals match reasonably well. The uncertainty within the fitting of each independent trajectory segment, that is caused by the experimental marking error, the camera radial lens distortion and the approximation of the planar limb motion model, is evident from the detail within the peaks of the reconstructed signals.

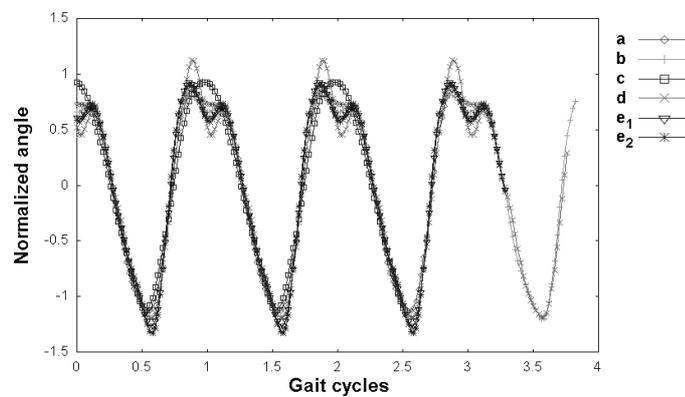


FIGURE 5.10: Upper normalized limb angle functions $\tilde{\theta}(t)$ shown for each of the individual reconstructed linear trajectory segments.

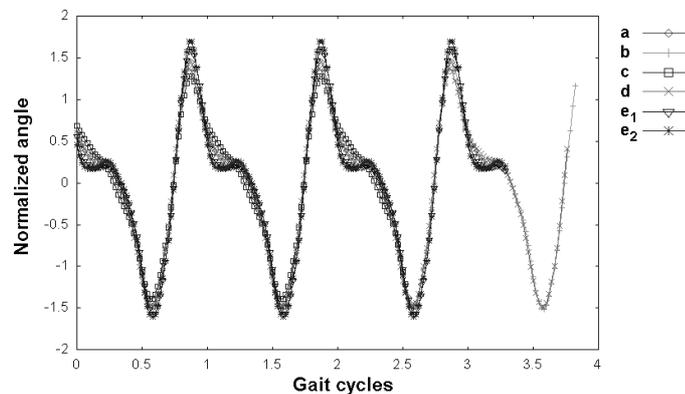


FIGURE 5.11: Lower normalized limb angle functions $\tilde{\theta}(t)$ shown for each of the individual reconstructed linear trajectory segments.

The set of reconstructed gait parameters $\mathbf{G} = (f_0, \mathbf{D}^\top, \mathbf{X}^\top, \mathbf{Y}^\top, \mathbf{V}^\top)^\top$ from the articulated limb function $g(t)$, that describes the components of hip displacement and angular leg motion, can then be decomposed into the underlying biometric \mathbf{B} and circumstantial \mathbf{W} parameter vectors of motion.

$$\mathbf{B} = (\mathbf{D}^\top, \tilde{\mathbf{X}}^\top, \tilde{\mathbf{Y}}^\top, \tilde{\mathbf{V}}^\top)^\top \quad (5.15)$$

$$\mathbf{W} = (v_x, f_0, \mathbf{p}^\top, \mathbf{q}^\top, \mathbf{r}^\top)^\top \quad (5.16)$$

where vector \mathbf{p} contains the a_0 offset coefficients of \mathbf{V} , vector \mathbf{q} contains the a_1 amplitude coefficients of $\mathbf{X}, \mathbf{Y}, \mathbf{V}$ and vector \mathbf{r} contains the t_s time shift coefficients of $\mathbf{X}, \mathbf{Y}, \mathbf{V}$.

5.6 Fusion of subject biometric parameters

Each subject within the image sequence has a single set of underlying biometric parameters \mathbf{B} and a number of circumstantial pose parameter vectors \mathbf{W}_i , corresponding to each motion trajectory segment. For each of the piecewise linear segments of reconstructed gait motion, we can compute the biometric and circumstantial parameters $\tilde{\mathbf{G}}_i = (\tilde{\mathbf{B}}_i^\top, \tilde{\mathbf{W}}_i^\top)^\top$. The problem then is to fuse together the set of biometric parameters $\tilde{\mathbf{B}}_i$ in such a way, that we can find a maximum likelihood estimate of \mathbf{B} that minimizes the image reprojection error corresponding to the measurement data.

5.6.1 Linear computation of biometric parameters

We first compute a putative set of biometric coefficients \mathbf{B} by using a linear algorithm. Given a set of N reconstructed biometric parameter $\tilde{\mathbf{B}}_i$ segments, we compute a weighted average of all the amplitude \mathbf{b} and limb length \mathbf{d} coefficient vectors.

$$\mathbf{d}' = \sum_{i=1}^N \omega_i \cdot \mathbf{d}_i \quad (5.17)$$

$$\mathbf{b}' = \sum_{i=1}^N \omega_i \cdot \mathbf{b}_i \quad (5.18)$$

The weights need not all be fixed as the value $1/N$ but can be derived from the diagonal elements of the covariance fitting matrices $\Sigma_{\mathbf{g}}$ of the piecewise linear reconstructions.

Values of phase are normalized modulo 2π to the range $(-\pi : \pi)$, hence phases distributed around both extremes of the range must be treated with caution. For each phase angle ψ'_k of \mathbf{B} , we recover the set of unit vectors \mathbf{v}_i that correspond to the required phase angles ψ_k within the set of \mathbf{B}_i . If we form the design matrix \mathbf{A} by stacking all the weighted vectors $\omega_i \cdot \mathbf{v}_i^\top$, then the Eigenvector \mathbf{x} with largest Eigenvalue of the 2×2 symmetric Eigen-system $\mathbf{A}^\top \mathbf{A} \cdot \mathbf{x} = \lambda \mathbf{x}$ is parallel to the required phase vector. We resolve the sign

ambiguity of the vector direction by computing the sum of all dot products between \mathbf{x} and \mathbf{v}_i .

$$\mathbf{y} = \text{sign}\left(\sum_{i=1}^N \mathbf{v}_i^\top \mathbf{x}\right) \cdot \mathbf{x} \quad (5.19)$$

Vector \mathbf{y} represents the best phase consistent with all corresponding biometric phase coefficients. The vector \mathbf{y} is then converted back into a phase angle within the required range $(-\pi : \pi)$, such that $\psi'_k = \angle \mathbf{y}$.

5.6.2 Maximum likelihood estimation

Having computed an initial estimate of the biometric parameters \mathbf{B} , we can then partition the parameters of the reconstruction into two segments.

$$\mathbf{P} = (\mathbf{B}^\top \mid \mathbf{W}_1^\top, \mathbf{W}_2^\top, \dots, \mathbf{W}_N^\top)^\top \quad (5.20)$$

The first partition contains the set of biometric parameters common to all piecewise linear subject motion segments. The second contains the set of subsidiary circumstantial parameters \mathbf{W}_i that describe the apparent changes in gait dynamics throughout the image sequence. In addition, we define the auxiliary vector \mathbf{Q} of parameterised pose projection mappings \mathbf{M}_i that remain fixed during the minimization process.

$$\mathbf{Q} = (\mathbf{M}_1^\top, \mathbf{M}_2^\top, \dots, \mathbf{M}_N^\top)^\top \quad (5.21)$$

In essence, each piecewise linear reconstruction computes the best set of limb swing planes parameterised by \mathbf{M}_i , along with initial estimates of the gait dynamics $\tilde{\mathbf{B}}_i, \tilde{\mathbf{W}}_i$. Identifying each of the limb swing planes allows us to compute the MLE of the partitioned gait dynamics within these fixed planes. Each set of \mathbf{W}_i parameters is independent, leading to a sparsely structured Jacobian. Minimization of the partitioned parameter vector \mathbf{P} is then computed via a sparse Levenberg-Marquardt algorithm, as described in appendix C.4.

The biometric fusion process is repeated for each subject within the image sequence. Table 5.4 shows the results of the biometric fusion process on the residual image reprojection errors of the reconstructed motion.

The set of pose projections \mathbf{M}_i are fixed for each motion trajectory segment, thus the optimal set of χ^2 residual errors are obtained by reprojection with the individual $\tilde{\mathbf{B}}_i$

<i>Segment Name</i>	$\tilde{\mathbf{B}}_i$	<i>Linear</i>	<i>MLE</i>
a	3.102	3.775 (21.7%)	3.315 (6.87%)
b	1.048	1.482 (41.41%)	1.286 (22.71%)
c	1.321	3.053 (131.11%)	1.878 (42.17%)
d	1.398	2.206 (57.8%)	1.605 (14.81%)
e1	1.619	2.974 (83.69%)	1.793 (10.75%)
e2	1.189	2.566 (115.81%)	1.348 (13.37%)

TABLE 5.4: Piecewise linear trajectory segment root mean square reprojection errors. $\tilde{\mathbf{B}}_i$ parameterises the gait dynamics independently for each trajectory segment while the reprojection errors attributed to both linear and MLE fusion methods reconstruct the gait motion with a common set of biometric parameters \mathbf{B} . Values in brackets are percentage errors $\Delta\epsilon$ between computed and $\tilde{\mathbf{B}}_i$ residual r.m.s. errors.

gait parameterisations. Perturbing the biometric parameters $\tilde{\mathbf{B}}_i$ away from these optimal estimates increases the χ^2 fitting errors. Consolidation of the set of biometric parameters \mathbf{B} through the fusion process can then only increase the χ^2 fitting errors. We quantify the goodness of the computed fusion parameterisation by measuring the difference in root mean square residual errors $\Delta\epsilon$, between the fused \mathbf{B} and individual $\tilde{\mathbf{B}}_i$ biometric reconstruction errors, and express this error difference as a percentage of the optimal $\tilde{\mathbf{B}}_i$ reprojection error.

We can clearly see that performing the maximum likelihood estimation of the subject biometric parameters reduces the percentage r.m.s. reconstruction errors $\Delta\epsilon$ by at least half, when compared to the linear fusion method alone. The mean percentage error increase $\Delta\epsilon$ after performing the MLE is of the order $\simeq 18\%$, indicating a good biometric parameterisation. Comparatively, the mean error increase $\Delta\epsilon$ caused by linear fusion alone is of the order $\simeq 75\%$, with a doubling of the reprojection error in places. We then recommend that the MLE step be performed to compute a good set of initial parameter estimates, before initiating a global optimization of all available parameters.

5.7 Fusion of system parameters

The set of leg swing plane pose homography mappings \mathbf{H}_1 and \mathbf{H}_2 are parameterised by the camera intrinsic parameters \mathbf{K} , the column vectors $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ of the worldspace pose rotation \mathbf{R} and a translation vector \mathbf{t} , inter-plane scaling factor τ and leg plane selection coefficient $\beta = \mp 1$.

$$\mathbf{H} = \mathbf{K} \begin{bmatrix} \mathbf{r}_1 & (\cos \alpha \cdot \mathbf{r}_2 + \tau\beta \sin \alpha \cdot \mathbf{r}_3) & (\tau\beta \cdot \mathbf{r}_3 + \mathbf{t}) \end{bmatrix} \quad (5.22)$$

Since all motion is within a common ground plane, each subject pose trajectory can be parameterised by the common pose rotation \mathbf{R}_N and an angular motion direction θ

within the canonical ground plane. We assume that the camera intrinsic parameters \mathbf{K} and ground plane rotation \mathbf{R}_N remain fixed throughout the image sequence.

$$\begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 \end{bmatrix} = \mathbf{R}_N \mathbf{R}_\theta \quad (5.23)$$

Each subject pose mapping can be parameterised by a 6-vector \mathbf{Z} with coefficients:

$$\mathbf{Z} = (\theta, \alpha, \tau, \mathbf{t}^\top)^\top \quad (5.24)$$

In order to recover these parameters of pose, we first need to compute the fixed system parameters \mathbf{K} and \mathbf{R}_N . Once found, the inverses may be applied to decompose the set of pose transformations \mathbf{M} into the required form.

5.7.1 Recovering the camera intrinsic parameters

Any ideal 3D point can be written $\mathbf{X} = (\mathbf{W}^\top, 0)^\top$ and is projected into the image via the perspective transformation $\mathbf{x} = \mathbf{P}\mathbf{X}$, where the projection matrix \mathbf{P} has the form $\mathbf{P} = \mathbf{K}[\mathbf{R} \mid \mathbf{t}]$.

$$\begin{aligned} \mathbf{x} &= \mathbf{K}[\mathbf{R} \mid \mathbf{t}]\mathbf{X} \\ \mathbf{x} &= \mathbf{K}\mathbf{R}\mathbf{W} \end{aligned} \quad (5.25)$$

A point lies on the absolute conic $\Omega = \mathcal{I}$ if $\mathbf{W}^\top \Omega \mathbf{W} = 0$. If we make the substitution $\mathbf{W} = \mathbf{R}^\top \mathbf{K}^{-1} \mathbf{x}$ then the conic constraint condition can be written as:

$$(\mathbf{R}^\top \mathbf{K}^{-1} \mathbf{x})^\top \Omega \mathbf{R}^\top \mathbf{K}^{-1} \mathbf{x} = 0 \quad (5.26)$$

$$\mathbf{x}^\top \omega \mathbf{x} = 0 \quad (5.27)$$

where the image of the absolute conic (IAC) is given by

$$\omega = \mathbf{K}^{-\top} \mathbf{K}^{-1} \quad (5.28)$$

The IAC is only dependent on the coefficients of the camera intrinsic parameters \mathbf{K} . Two special points known as the circular points $\mathbf{I} = (1, i, 0)^\top$ and $\mathbf{J} = (1, -i, 0)^\top$ lie on

the absolute conic Ω . The imaged circular points \mathbf{I}' and \mathbf{J}' of a planar transformation $\mathbf{H} \cdot (1, \pm i, 0)^\top$ must also lie on the IAC.

$$\begin{pmatrix} 1 & i & 0 \end{pmatrix} \mathbf{H}^\top \omega \mathbf{H} \begin{pmatrix} 1 \\ i \\ 0 \end{pmatrix} = 0$$

$$(\mathbf{h}_1^\top + i \mathbf{h}_2^\top) \omega (\mathbf{h}_1 + i \mathbf{h}_2) = 0 \quad (5.29)$$

$$\mathbf{h}_1^\top \omega \mathbf{h}_1 - \mathbf{h}_2^\top \omega \mathbf{h}_2 + 2i \mathbf{h}_1^\top \omega \mathbf{h}_2 = 0 \quad (5.30)$$

where the column vectors of the homography matrix \mathbf{H} are written as \mathbf{h}_j . Equating real and imaginary parts to zero yields the result.

$$\mathbf{h}_1^\top \omega \mathbf{h}_1 - \mathbf{h}_2^\top \omega \mathbf{h}_2 = 0 \quad (5.31)$$

$$\mathbf{h}_1^\top \omega \mathbf{h}_2 = 0 \quad (5.32)$$

It must be noted that the constrained nature of motion means that there is a one parameter family of solutions for the calibration because all the camera rotations are about the same axis. It is then necessary to make assumptions about the camera intrinsic parameters in order to resolve this ambiguity. In certain cases however, the zero skew assumption does not resolve the ambiguity. Such a situation arises when the rotation is about the camera X, Y or Z axes. These exceptions are described in more detail within [121]. If we restrict the form of the camera intrinsic parameters to have zero skew then the elements of the IAC are given by:

$$\mathbf{U} = \mathbf{K}^{-1} = \begin{pmatrix} a & b \\ & c & d \\ & & e \end{pmatrix} \quad (5.33)$$

$$\omega = \mathbf{U}^\top \mathbf{U} = \begin{pmatrix} a^2 & & ab \\ & c^2 & cd \\ ab & cd & b^2 + d^2 + e^2 \end{pmatrix} = \begin{pmatrix} a' & & b' \\ & c' & d' \\ b' & d' & e' \end{pmatrix} \quad (5.34)$$

The elements of the IAC can be written as a vector $\mathbf{w} = (a', b', c', d', e')^\top$. Subsequently, the real and imaginary parts of the constraint equation 5.30 can then be written in the form $\mathbf{A}\mathbf{w} = \mathbf{0}$, in terms of the entries of the homography matrix h_{ij} . The design

matrix \mathbf{A} is formed by stacking all such constraints from each of the pose homography transformations.

$$\begin{bmatrix} (h_{11}^2 - h_{12}^2) & 2(h_{31}h_{11} - h_{32}h_{12}) & (h_{21}^2 - h_{22}^2) & 2(h_{31}h_{21} - h_{32}h_{22}) & (h_{31}^2 - h_{32}^2) \\ (h_{11}h_{12}) & (h_{32}h_{11} + h_{12}h_{31}) & (h_{22}h_{21}) & (h_{32}h_{21} + h_{22}h_{31}) & (h_{32}h_{31}) \end{bmatrix} \mathbf{w} = \mathbf{0} \quad (5.35)$$

Since \mathbf{w} is defined up to scale (4 degrees of freedom) then a minimal solution to this set of homogeneous equations can be found from just two non parallel homography matrix constraints. The solution vector \mathbf{w} is found by computing the SVD of the homogeneous set of equations, see appendix B.2.3 for details. The elements of matrix \mathbf{U} are then extracted from the solution vector \mathbf{w} as:

$$\begin{aligned} a &= \sqrt{a'}, & c &= \sqrt{c'} \\ b &= b'/a, & d &= d'/c \\ e &= \sqrt{e' - (b^2 + d^2)} \end{aligned}$$

The calibration matrix \mathbf{K} is finally computed from the inverse transformation $\mathbf{K} = \mathbf{U}^{-1}$. Within the presence of noise, the solution to \mathbf{w} may have invalid elements of \mathbf{U} . Since a' must be positive, we can remove the homogeneous sign ambiguity by applying the scaling $\mathbf{w}' = \text{sign}(a') \cdot \mathbf{w}$. Consequently, if $c' \leq 0$ or $e' \leq b^2 + d^2$ then matrix \mathbf{U} is invalid.

The camera principal point is notoriously difficult to compute accurately. We can enforce an empirical tolerance on its position by assuming that it lies within 50 pixels from the centre of the image. If we fail to compute a valid calibration matrix then we must constrain the form of \mathbf{K} further.

5.7.2 Known principal point

The camera intrinsic parameter matrix \mathbf{K} can be decomposed into two parts \mathbf{K}_f and \mathbf{K}_t .

$$\mathbf{K} = \mathbf{K}_t \mathbf{K}_f = \lambda \begin{pmatrix} 1 & u_0 \\ & 1 & v_0 \\ & & 1 \end{pmatrix} \begin{pmatrix} m_x & & \\ & m_y & \\ & & 1 \end{pmatrix} \quad (5.36)$$

Subsequently, the planar homography transformation \mathbf{H} can be decomposed as:

$$\mathbf{H} = \mathbf{K}_t \mathbf{K}_f \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix} \quad (5.37)$$

If we know the camera principal point then we can remove \mathbf{K}_t from the set of homographies, by applying the transformation $\mathbf{H}' = \mathbf{K}_t^{-1}\mathbf{H}$. The transformed IAC $\omega' = \mathbf{K}_f^{-\top}\mathbf{K}_f^{-1}$ then has a much simpler form since \mathbf{K}_f is a diagonal matrix.

$$\mathbf{U} = \mathbf{K}_f^{-1} = \begin{pmatrix} a & & \\ & c & \\ & & e \end{pmatrix} \quad (5.38)$$

$$\omega' = \mathbf{U}^\top \mathbf{U} = \begin{pmatrix} a^2 & & \\ & c^2 & \\ & & e^2 \end{pmatrix} = \begin{pmatrix} a' & & \\ & c' & \\ & & e' \end{pmatrix} \quad (5.39)$$

The elements of the transformed IAC can be written as a vector $\mathbf{w} = (a', c', e')^\top$. Subsequently, the real and imaginary parts of the constraint equation 5.30 can be written in the form $\mathbf{A}\mathbf{w} = \mathbf{0}$, in terms of the entries of the transformed homography matrix h_{ij} . The design matrix \mathbf{A} is formed by stacking all such constraints from each of the transformed pose homography transforms.

$$\begin{bmatrix} (h_{11}^2 - h_{12}^2) & (h_{21}^2 - h_{22}^2) & (h_{31}^2 - h_{32}^2) \\ (h_{11}h_{12}) & (h_{21}h_{22}) & (h_{31}h_{32}) \end{bmatrix} \mathbf{w} = \mathbf{0} \quad (5.40)$$

The elements of matrix \mathbf{U} are then extracted from the solution vector \mathbf{w} as:

$$a = \sqrt{a'} \quad c = \sqrt{c'} \quad e = \sqrt{e'}$$

The calibration matrix \mathbf{K}_f is finally computed from the inverse transformation $\mathbf{K}_f = \mathbf{U}^{-1}$ and subsequently $\mathbf{K} = \mathbf{K}_t\mathbf{K}_f$. Again we can remove the homogeneous sign ambiguity and ensure that a' is positive. If any of the updated coefficients of \mathbf{w} are negative then the calibration matrix can not be computed. In such a circumstance we must question the validity of the set of homography matrices corresponding to each of the subject poses. If all pose mappings are close to parallel within measurement noise, then this leads to a degenerate configuration and subsequently there is a 1-parameter family of possible calibrations. We require at least two trajectories separated by an oblique angle to robustly compute the camera intrinsic parameters.

5.7.3 Known aspect ratio

If we know the camera aspect ratio in addition to the principal point then the only remaining free parameter describing the camera intrinsic parameters is the effective

focal length f . The camera calibration matrix can then be decomposed as:

$$\mathbf{K} = \mathbf{K}_t \mathbf{K}_a \mathbf{K}_f = \lambda \begin{pmatrix} 1 & u_0 \\ & 1 & v_0 \\ & & 1 \end{pmatrix} \begin{pmatrix} \gamma & & \\ & 1 & \\ & & 1 \end{pmatrix} \begin{pmatrix} f & & \\ & f & \\ & & 1 \end{pmatrix} \quad (5.41)$$

We then compute the set of transformed homography matrices $\mathbf{H}' = \mathbf{K}_a^{-1} \mathbf{K}_t^{-1} \mathbf{H}$ and solve the set of constraint equations of the form $\mathbf{A} \mathbf{w} = \mathbf{0}$.

$$\begin{bmatrix} (h_{11}^2 - h_{12}^2 + h_{21}^2 - h_{22}^2) & (h_{31}^2 - h_{32}^2) \\ (h_{11}h_{12} + h_{22}h_{21}) & (h_{32}h_{31}) \end{bmatrix} \mathbf{w} = \mathbf{0} \quad (5.42)$$

5.7.4 Pose decomposition

The limb swing plane pose vector \mathbf{M} computed from the linear trajectory reconstruction has the form:

$$\mathbf{M} = (\mathbf{m}_1^\top, \mathbf{m}_2^\top, \mathbf{m}_3^\top, \mathbf{m}_4^\top, \alpha)^\top \quad (5.43)$$

where $\mathbf{m}_1 = \mathbf{K} \cdot \mathbf{r}_1$, $\mathbf{m}_2 = \mathbf{K} \cdot \mathbf{r}_2$, $\mathbf{m}_3 = \tau \cdot \mathbf{K} \cdot \mathbf{r}_3$ and $\mathbf{m}_4 = \mathbf{K} \cdot \mathbf{t}$

We can then construct the set of normalized projection matrices of the form $\hat{\mathbf{P}} = [\mathbf{Q} \mid \mathbf{t}]$, by applying the transformation \mathbf{K}^{-1} to remove the camera intrinsic parameters. We remove the resulting homogeneous scaling ambiguity by ensuring that the first column vector has unit norm $\|\tilde{\mathbf{r}}_1\| = 1$. The sign of the ambiguity is removed by ensuring that the subject is in front of the camera $\mathbf{t}_z > 0$.

$$\hat{\mathbf{P}} = \begin{bmatrix} \tilde{\mathbf{r}}_1 & \tilde{\mathbf{r}}_2 & \tau \cdot \tilde{\mathbf{r}}_3 & \mathbf{t} \end{bmatrix} \quad (5.44)$$

We assume that all subjects walk on a level ground plane. Although the set of subject ground planes must be parallel, i.e. share a common normal \mathbf{N} , there is no restriction on the translation between them. This allows us to model subjects walking at street level, on raised platforms, different floors within a building, etc. The translation vector \mathbf{t} fully encodes the subject's worldspace displacement, while the normal \mathbf{N} encodes the orientation of the subject's plane of motion. The problem is then to decompose the normalized projection matrices into the form:

$$\begin{aligned} \hat{\mathbf{P}} &= [\mathbf{Q} \mid \mathbf{t}] \\ \hat{\mathbf{P}} &= [\mathbf{R}_N \mathbf{R}_\theta \mathbf{H}_\tau \mid \mathbf{t}] \end{aligned} \quad (5.45)$$

where the inter-plane scaling transformation \mathbf{H}_τ is given by:

$$\mathbf{H}_\tau = \begin{pmatrix} 1 & & \\ & 1 & \\ & & \tau \end{pmatrix} \quad (5.46)$$

The set of $\tilde{\mathbf{r}}_1$ motion direction vectors must all lie within a single plane, and the set of vertical directions $\tilde{\mathbf{r}}_2$ should all be parallel. Each set of constraints can be stacked to form a system of equations of the form $\mathbf{A}\mathbf{N} = \mathbf{0}$.

$$\begin{bmatrix} \tilde{\mathbf{r}}_1^\top \\ [\tilde{\mathbf{r}}_2]_\times \end{bmatrix} \mathbf{N} = \mathbf{0} \quad (5.47)$$

Note that only two of the three equations of $[\tilde{\mathbf{r}}_2]_\times \mathbf{N} = \mathbf{0}$ are linearly independent, hence it is customary to only include the first two within the design matrix \mathbf{A} . The solution vector \mathbf{N} is found by computing the SVD of the homogeneous set of equations, and is constrained to have unit norm $\|\mathbf{N}\| = 1$, see appendix B.2.3 for details. Having computed the ground plane normal vector \mathbf{N} , we can compute the rotation $\mathbf{H}_\mathbf{r}$ that maps this vector back onto the canonical Y axis $(0, 1, 0)^\top$. Consequently the inverse rotation $\mathbf{R} = \mathbf{H}_\mathbf{r}^\top$ transforms the canonical ground plane to the required pose.

The ground plane normal vector is constrained to have unit norm but has a sign ambiguity. There are then two possible vectors \mathbf{N} , $-\mathbf{N}$ with unit norm parallel to the ground plane normal. The updated matrix $\mathbf{Q}' = \mathbf{H}_\mathbf{r}\mathbf{Q}$ then has the form:

$$\mathbf{Q}' = \begin{pmatrix} a & & c \\ & \lambda_2 & \\ b & & d \end{pmatrix} \quad (5.48)$$

We ensure that λ_2 is positive, and thus resolve the sign ambiguity, by applying a further rotation $\mathbf{H}_\lambda = \text{diag}(-1, -1, 1)$ where necessary. The ground plane mapping \mathbf{R} is also updated $\mathbf{R}' = \mathbf{R}\mathbf{H}_\lambda$ as a consequence of applying this transformation.

We next compute the rotation \mathbf{H}_θ about the Y axis that aligns the vector $(a, 0, b)^\top$ with the positive X axis $(1, 0, 0)^\top$. The inverse rotation $\mathbf{R}_\theta = \mathbf{H}_\theta^\top$ then transforms the canonical X axis within the ground plane to the required position. The updated matrix $\mathbf{Q}'' = \mathbf{H}_\theta\mathbf{Q}'$ then has the form:

$$\mathbf{Q}'' = \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{pmatrix} \quad (5.49)$$

In general $\lambda_1 \neq \lambda_2$ due to noise, hence we compute the best homogeneous scaling factor $\lambda = (\lambda_1 + \lambda_2)/2$ consistent with the rotation matrix. The inter-plane scaling transformation \mathbf{H}_τ is recovered by removing this homogeneous scaling factor. The best absolute value of τ consistent with a right hand coordinate system is then given by:

$$\tau = \frac{|\lambda_3|}{\lambda} \quad (5.50)$$

and the updated pose translation vector by:

$$\mathbf{t}' = \frac{1}{\lambda} \cdot \mathbf{t} \quad (5.51)$$

We choose to align the canonical coordinate system with the first subject pose within the image sequence. The ground plane rotation matrix is then given by $\mathbf{R}_N = \mathbf{R}\mathbf{R}_\theta$ and the subsequent angle θ within the first pose vector \mathbf{Z}_1 is set to zero.

All further subject pose decompositions of the form $\hat{\mathbf{P}} = [\mathbf{Q} \mid \mathbf{t}]$ then start by removing the ground plane rotation \mathbf{R}_N , such that $\mathbf{Q}' = (\mathbf{R}_N)^\top \mathbf{Q}$. Consequently, the matrix \mathbf{Q}' has the same form as in equation 5.48. Following the same procedure; the motion direction rotation \mathbf{R}_θ , the inter plane scaling factor τ and the pose translation vector \mathbf{t} can all be computed.

$$\mathbf{R}_\theta = \begin{pmatrix} c & -s \\ & 1 \\ s & c \end{pmatrix} \quad (5.52)$$

The parameterised motion direction angle θ within the range $(-\pi : \pi)$ is then extracted from the rotation matrix \mathbf{R}_θ , shown in equation 5.52, and the elements of the subject pose vector \mathbf{Z}_i set.

$$\theta = \text{sign}(s) \cdot \arccos(c) \quad (5.53)$$

The set of the pose vectors \mathbf{Z}_i then parameterise the mappings \mathbf{M}_i consistent with the ground plane motion, relative to the first subject pose.

5.7.5 Maximum likelihood estimation

We have decomposed the set of pose mappings \mathbf{M}_i into a product of system parameters \mathbf{K}, \mathbf{R}_N and the set of subject poses \mathbf{Z}_i . We perform an additional step here to optimize the set of system parameters before computing the global MLE on all available parameters.

We assume that the computed piecewise linear gait function segments $g(t : f_0, \mathbf{D}, \mathbf{X}, \mathbf{Y}, \mathbf{V})$ accurately reflect the dynamics of gait. The reconstructed metric plane points \mathbf{X}_i for the entire set of subjects are then fixed, thus we proceed to optimize both the system and pose parameters. The parameter vector \mathbf{P} can be partitioned into two segments:

$$\mathbf{P} = \left(\mathbf{K}_c^\top, \mathbf{K}_r^\top, \mathbf{N}^\top \mid \mathbf{Z}_1^\top, \dots, \mathbf{Z}_n^\top \right)^\top \quad (5.54)$$

The first partition contains the system parameters. The camera model is encoded by the intrinsic parameters $\mathbf{K}_c = (m_x, m_y, u_0, v_0)^\top$ and the components of radial lens distortion $\mathbf{K}_r = (k_1, k_2)^\top$. The ground plane pose rotation \mathbf{R}_N is represented by the Rodrigues vector $\mathbf{N} = (n_1, n_2, n_3)^\top$, see appendix B.3.1 for details. The values of the camera radial distortion \mathbf{K}_r are initialised to zero. The second partition contains the set of independent piecewise linear subject poses $\mathbf{Z}_i = (\theta_i, \alpha_i, \tau_i, \mathbf{t}_i^\top)^\top$. In addition, the auxiliary vector \mathbf{Q} of fixed metric plane landmark points \mathbf{X}_i is defined as:

$$\mathbf{Q} = (\mathbf{X}_1^\top, \dots, \mathbf{X}_n^\top)^\top \quad (5.55)$$

Minimization of the partitioned parameter vector is then computed via a sparse Levenberg-Marquardt algorithm, see appendix C.4 for details. Table 5.5 shows the calibration coefficients of \mathbf{K}_c and \mathbf{K}_r , computed linearly from the image of the absolute conic and subsequently the set of minimized parameters from the sparse MLE method.

<i>Method</i>	m_x	m_y	u_0	v_0	k_1	k_2
IAC*	811.621	930.129	360.0	288.0	0	0
MLE (\mathbf{K}_c)	813.033	913.49	369.14	288.851	0	0
MLE ($\mathbf{K}_c, \mathbf{K}_r$)	705.328	878.789	306.946	289.492	-0.211599	0.0904326
Zhang	865.281	947.937	355.645	274.194	-0.224416	0.305238

TABLE 5.5: Comparison of the camera calibration coefficients \mathbf{K}_c and \mathbf{K}_r computed from the image of the absolute conic (known principal point) and the maximum likelihood estimate of parameters \mathbf{P} . Accurate baseline values for each of the coefficients are obtained by performing Zhang’s calibration algorithm.

The full IAC method fails to compute the principal point within a 50 pixel tolerance of the centre of the image. We then assume that the camera principal point lies at the image centre and compute the remaining coefficients using the reduced form of the IAC. The camera principal point is often the hardest quantity to estimate using linear methods. It is no surprise then that the distribution of subject trajectories weighs quite heavily on the computed calibration coefficients. We need a large number of disparate trajectories in order to compute an accurate estimate of the camera intrinsic parameters. In the presence of noise, it may then be beneficial to compute both sets of calibration coefficients using the known and unknown principal points. Then as an initial step

before computing the MLE, we choose the set of coefficients that generates the smallest residual image reprojection error.

Accurate camera coefficients are obtained by performing Zhang’s calibration algorithm [120]. As an initial estimate of the camera intrinsic and extrinsic parameters both IAC and MLE methods compare reasonably well with the baseline parameters.

<i>Segment</i>	$ \theta_c $	\mathbf{M}	<i>Linear</i>	<i>MLE</i> (\mathbf{K}_c)	<i>MLE</i> ($\mathbf{K}_c, \mathbf{K}_r$)
a	18.3	3.102	4.708 (51.77%)	3.629 (16.99%)	3.768 (21.47%)
b	29.0	1.048	3.84 (266.41%)	2.316 (120.99%)	1.837 (75.29%)
c	80.8	1.321	35.087 (2556.09%)	2.408 (82.29%)	2.348 (77.74%)
d	27.3	1.398	9.649 (590.2%)	2.304 (64.81%)	1.897 (35.69%)
e1	63.6	1.619	3.715 (129.46%)	2.753 (70.04%)	2.617 (61.64%)
e2	64.6	1.189	3.511 (195.29%)	2.451 (106.14%)	2.365 (98.91%)

TABLE 5.6: Piecewise linear trajectory segment root mean square reprojection errors. \mathbf{M} parameterises each trajectory segment independently, while the reprojection errors attributed to both linear decomposition and MLE reconstructions constrain the gait motion segments to lie within planes parallel to the worldspace ground plane. Values in brackets are percentage errors $\Delta\epsilon$ between computed and \mathbf{M} residual r.m.s. errors. $|\theta_c|$ is the true absolute pose trajectory angle (degrees) w.r.t. the camera X axis.

Table 5.6 shows the results of the pose decomposition process on the residual reprojection errors. The set of metric plane limb points \mathbf{X}_i are fixed, thus the optimal set of χ^2 residual errors are obtained by reprojection of the individual \mathbf{M} pose parameterisations. Perturbing the pose parameters \mathbf{M} away from these optimal estimates increases the χ^2 fitting error. The re-parameterisation of the individual \mathbf{M} vectors through the decomposition process can then only increase the χ^2 fitting errors. We quantify the goodness of the decomposition parameterisation by measuring the difference in root mean square residual errors $\Delta\epsilon$, between the decomposed and individual \mathbf{M} errors and express this error difference as a percentage of the optimal \mathbf{M} reprojection error.

Reprojection error is worst for oblique trajectories, with respect to the camera coordinate system, using linear decomposition alone. Removing the camera calibration matrix from the pose parameters \mathbf{M} should generate the matrix \mathbf{Q} with orthogonal columns. Given a poor estimate of the intrinsic parameters, the departure from orthogonality causes large errors in the computation of the ground plane normal and consequently the pose trajectory angles.

Table 5.7 shows a comparison of the percentage error increases $\Delta\epsilon$ of both the biometric and system parameter maximum likelihood estimation steps. The table clearly shows that the decomposition of the system and pose projection parameters accounts for most of the reconstruction uncertainty.

<i>Segment Name</i>	M	<i>Biometric</i>	<i>System</i>
a	3.102	6.87%	21.47%
b	1.048	22.71%	75.29%
c	1.321	42.17%	77.74%
d	1.398	14.81%	35.69%
e1	1.619	10.75%	61.64%
e2	1.189	13.37%	98.91%

TABLE 5.7: Piecewise linear trajectory segment root mean square reprojection percentage errors $\Delta\epsilon$ after performing maximum likelihood estimation of both biometric and system parameters.

<i>Segment Name</i>	θ_c	<i>Linear</i>	<i>MLE (\mathbf{K}_c)</i>	<i>MLE ($\mathbf{K}_c, \mathbf{K}_r$)</i>
a	-18.3	-33.3	-33.7	-27.2
b	151.0	153.9	149.8	148.4
c	-99.2	-137.8	-95.8	-101.6
d	27.3	26.0	25.8	21.4
e1	116.4	112.8	112.2	110.1
e2	115.4	112.8	112.1	110.1

TABLE 5.8: Computed trajectory angles for both linear and MLE pose decomposition methods. θ_c is the true pose trajectory angle (degrees) w.r.t. the camera X axis.

Table 5.8 shows the set of computed pose trajectory angles θ , for both linear and MLE methods. We can see quite clearly the disparity between true and decomposed trajectory angles. Figures 5.12 and 5.13 show the corresponding set of reprojected limb points for trajectory segments ‘a’ and ‘c’ respectively. Three different frames from the beginning, middle and end of these sequences are shown. The poorly estimated trajectory angle for segment ‘c’, using linear decomposition alone, has a noticeably bad set of reprojected limb points.

Fusion of the subject biometric parameters requires us to minimize $33 + 12n$ parameters per subject, where n is the number of linear trajectory segments for that subject. The Jacobian matrices also require extensive computation of trigonometric functions. Maximum likelihood estimation of the system parameters requires us to minimize a total of $9 + 6m$ parameters, where m is the complete number (all subjects) of linear trajectory segments in the system. Both minimizations are sparse and of the order $O(n)$, though evaluation of the gait functions $g(t : f_0, \mathbf{D}, \mathbf{X}, \mathbf{Y}, \mathbf{V})$ to compute the set of metric plane points \mathbf{X}_i need only be performed once for the system parameters. If we initialise a global parameter optimization with poor estimates of the system parameters, we risk undoing all the hard work required to compute the fused biometric parameters within the first few iterations of the global optimization step. For this reason, we recommend that the MLE of the system parameters is an essential and compulsory step required before computing the global parameter optimization.

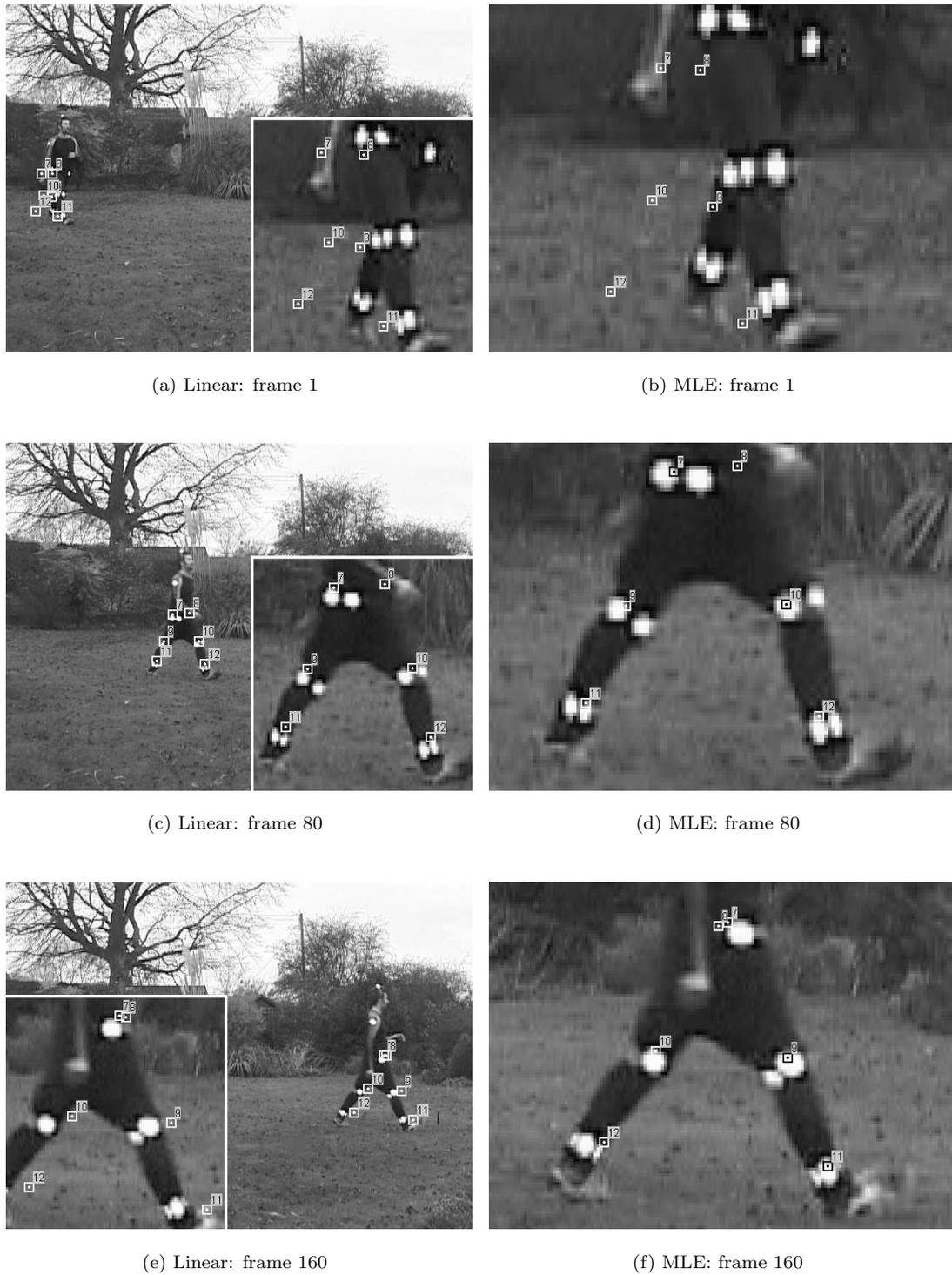
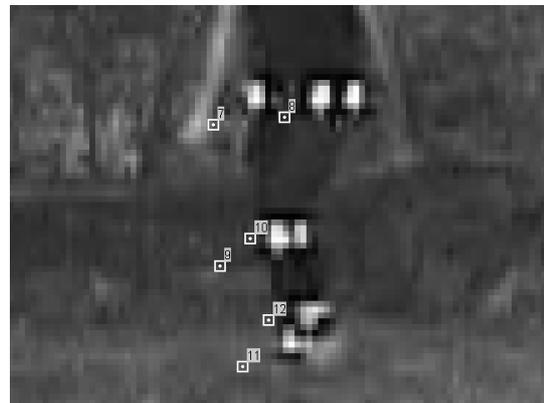


FIGURE 5.12: Reconstruction of trajectory segment 'a' after linear pose decomposition (a,c,e) and after maximum likelihood estimation of the pose parameters (b,d,f).



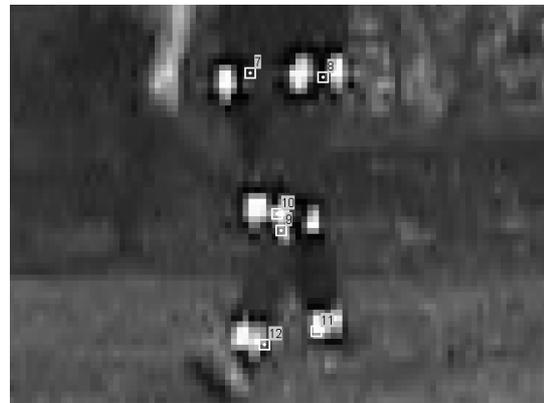
(a) Linear: frame 1



(b) MLE: frame 1



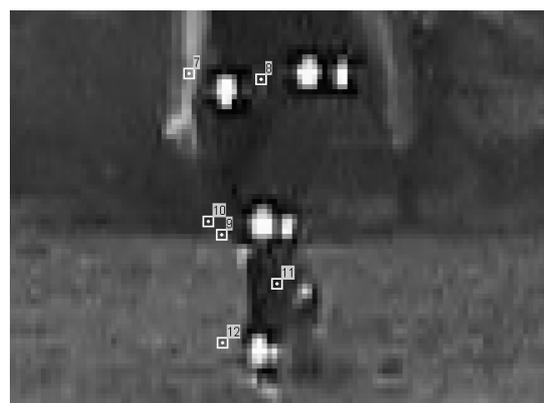
(c) Linear: frame 64



(d) MLE: frame 64



(e) Linear: frame 127



(f) MLE: frame 127

FIGURE 5.13: Reconstruction of trajectory segment ‘c’ after linear pose decomposition (a,c,e) and after maximum likelihood estimation of the pose parameters (b,d,f).

5.8 Fusion discussion

Biometric and system parameter fusion is a two step process. Each step is independent, can be performed simultaneously and requires only that we have access to the subject pose tree of reconstructed gait functions $g(t : f_0, \mathbf{D}, \mathbf{X}, \mathbf{Y}, \mathbf{V})$ and pose projection mappings \mathbf{M} .

Each subject pose is modelled by a linear motion trajectory with limbs that swing within bilateral planes. The reconstruction phase described in chapter 4 essentially identifies these bilateral planes and the representation of limb motion within them. The fusion step allows us to use the elements of these reconstructions independently. The set of consistent biometric parameters \mathbf{B} and \mathbf{W}_i are computed by fixing the bilateral limb planes and optimizing the biometric gait functions within them. The consistent set of system parameters $\mathbf{K}_c, \mathbf{K}_r, \mathbf{N}$ and the parameterised subject poses \mathbf{Z}_i are computed by fixing the reconstructed limb points within the metric plane and optimizing the set of bilateral limb planes.

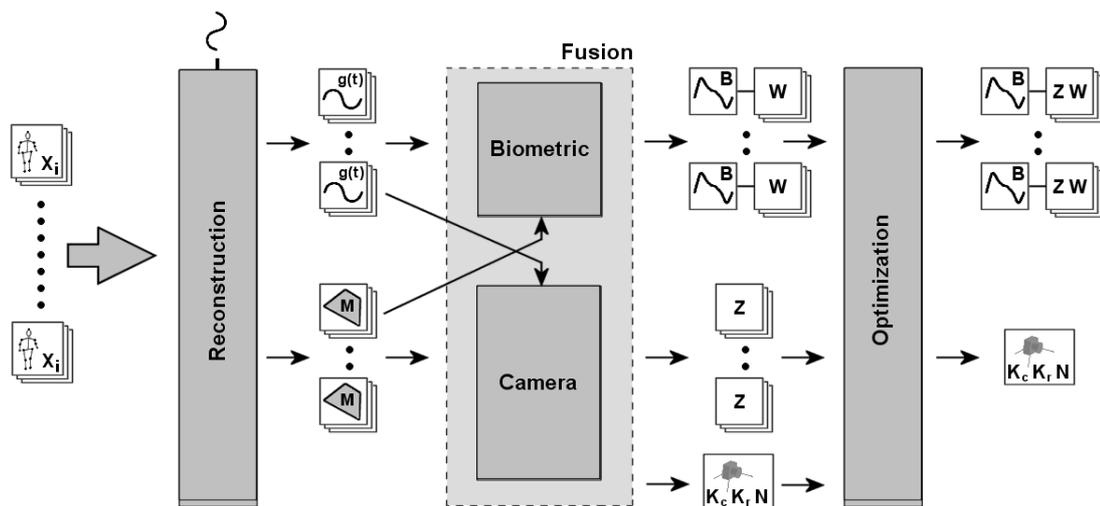


FIGURE 5.14: Data flow diagram of the gait reconstruction algorithm that shows the three main steps (reconstruction, fusion and optimization) within the processing chain.

The basis function used to represent the dynamics of limb motion is arbitrary, since the initial reconstruction phase is concerned only with identifying the set of bilateral limb swing planes and the intermediate representation of gait. A periodic function is chosen that accurately represents the dynamics of gait within a small number of parameters. The obvious choice is to use a Fourier series representation, and good reconstruction results have been obtained with five Fourier harmonics. Complicated basis functions risk introducing many local minima within the parameter space, with potentially disastrous consequences for minima far from the global optimum. The topology of the parameter space is often proportionally as complex as the basis function, and as a result convergence may be slow.

Fusion of the biometric parameters recovers an initial estimate of the unique properties of a subject's motion dynamics, that are consistent over a range of gait modes. The form of the biometric function and indeed the question of what are the unique features of a subject's gait remain largely unanswered, see the discussion in section 3.6.5. The function should ideally be invariant to changes in gait mode, i.e. speed, stride and cadence. We have outlined a modified Fourier function, that to first order approximates an invariant representation of gait.

If we intend to perform maximum likelihood estimation within the fusion stage, then we recommend that minimization should be performed on both biometric and system parameters. Performing the MLE within the fusion stage generates a good initial estimate of the global parameters and potentially widens the region of convergence. The fusion step then removes some of the risk of encountering outlying local minima and potential slow convergence far from the true minimum, within the global parameter space.

<i>Segment Name</i>	M	<i>Linear</i>	<i>MLE</i>
a	3.102	5.215 (68.12%)	3.925 (26.53%)
b	1.048	4.065 (287.88%)	1.935 (84.64%)
c	1.321	35.057 (2553.82%)	2.682 (103.03%)
d	1.398	9.832 (603.29%)	2.021 (44.56%)
e1	1.619	4.423 (173.19%)	2.717 (67.82%)
e2	1.189	4.19 (252.4%)	2.448 (105.89%)

TABLE 5.9: Piecewise linear trajectory segment root mean square reprojection errors of the decomposed parameters. Values in brackets are percentage errors $\Delta\epsilon$ between computed and **M** residual r.m.s. errors.

Table 5.9 shows the root mean square residual errors after performing the two methods of parameter fusion on both biometric and system parameters. The fusion process combines all piecewise linear motion segment information and forms it into a realistic worldspace motion model. The percentage errors $\Delta\epsilon$ generated from the reprojection of the fused parameters gives an indication of the quality of the worldspace model decomposition. The errors determined from computation of the parameters, using linear methods alone, clearly demonstrates a poor conversion between piecewise linear segments and the worldspace model. Where maximum likelihood estimation has been employed to refine the parameter estimates in the fusion process, the percentage errors are significantly reduced and distributed more evenly.

The accuracy of the computed initial worldspace parameters is then dependent on the uncertainty within each of the fitted linear motion trajectory segments. These segments are computed by using a simple pin hole projection model. Consequently, radial lens distortion is one of the dominant sources of error within the initial fitting process. We should take care to reduce as much as possible the distortion effects by modelling data across small patches of the image surface. This is another good reason for breaking linear motion sequences into three second blocks.

5.9 Global optimization of parameters

We have now computed some sensible initial estimates of both the subject biometric parameters \mathbf{B} , with corresponding circumstantial pose parameters $\mathbf{Z}_i, \mathbf{W}_i$, and the system parameters $\mathbf{K}_c, \mathbf{K}_r, \mathbf{N}$.

$$\mathbf{K}_c = (m_x, m_y, u_0, v_0)^\top \quad (5.56)$$

$$\mathbf{K}_r = (k_1, k_2)^\top \quad (5.57)$$

$$\mathbf{N} = (n_1, n_2, n_3)^\top \quad (5.58)$$

$$\mathbf{B} = (\mathbf{D}^\top, \tilde{\mathbf{X}}^\top, \tilde{\mathbf{Y}}^\top, \tilde{\mathbf{V}}^\top)^\top \quad (5.59)$$

$$\mathbf{Z} = (\theta, \alpha, \tau, \mathbf{t}^\top)^\top \quad (5.60)$$

$$\mathbf{W} = (v_x, f_0, \mathbf{p}^\top, \mathbf{q}^\top, \mathbf{r}^\top)^\top \quad (5.61)$$

These parameters are partitioned into three distinct categories, namely the *system*, *subject* and *pose* blocks. The *system* parameters encode the geometric properties of the camera and worldspace motion plane, and are fixed throughout the image sequence. The *subject* parameters encode the unique properties related to the dynamics of a subject's gait. These parameters are independent and fixed for each of the N subjects viewed throughout the image sequence. The *pose* parameters encode the circumstantial properties of each piecewise linear trajectory segment of a subject's motion. Each set of pose parameters is independent and describes the spatial position, velocity, cadence, initial phase of gait and other properties of the motion segment. We define these three entities by combining the corresponding parameter vectors.

$$\mathbf{a} = (\mathbf{K}_c^\top, \mathbf{K}_r^\top, \mathbf{N}^\top)^\top \quad (5.62)$$

$$\mathbf{b} = (\mathbf{D}^\top, \tilde{\mathbf{X}}^\top, \tilde{\mathbf{Y}}^\top, \tilde{\mathbf{V}}^\top)^\top \quad (5.63)$$

$$\mathbf{c} = (\mathbf{Z}^\top, \mathbf{W}^\top)^\top \quad (5.64)$$

There is then a hierarchical relationship between the system, subject and each of the subject's poses that is best described by a tree structure. This tree of parameters can be flattened into a single parameter vector \mathbf{P} with three partitions, and completely captures the properties of the imaged worldspace motion.

$$\mathbf{P} = (\mathbf{a}^\top \mid \mathbf{b}_1^\top, \dots, \mathbf{b}_n^\top \mid \mathbf{c}_1^\top, \dots, \mathbf{c}_m^\top)^\top \quad (5.65)$$

5.9.1 General sparse LM minimization

Many minimization problems within projective geometry can be partitioned into two sections. A detailed discussion on non-linear estimation of parameters that are partitioned in such a way is addressed in appendix C.4. Minimization of a set of parameters that are partitioned into three distinct sections can be treated in much the same way. The hierarchical structure of the relationship between parameter blocks will be reflected in the formation of the normal equations. The form of the Jacobian matrix is sparse, whose shape can be exploited in order to make great time savings when solving for a large number of subsidiary parameters. We describe here the general solution of a set of parameters partitioned into three sections using the Levenberg-Marquardt method. We first naively assume no independence between parameter blocks within each of the partitions.

In general, the form of the normal equations required to compute the parameter update vector $\delta\mathbf{p}$ can be written as shown in equation 5.66; where \mathbf{J} is the Jacobian matrix, $\Sigma_{\mathbf{x}}$ the diagonal covariance matrix of measurement error deviations and \mathbf{r} the residual fitting error vector between measured and reprojected model points.

$$\mathbf{J}^T \Sigma_{\mathbf{x}}^{-1} \mathbf{J} \cdot \delta\mathbf{p} = \mathbf{J}^T \Sigma_{\mathbf{x}}^{-1} \mathbf{r} \quad (5.66)$$

We augment the diagonal elements of the Hessian matrix $\mathbf{D} = \mathbf{J}^T \Sigma_{\mathbf{x}}^{-1} \mathbf{J}$ with the Levenberg-Marquardt scaling factor $(1 + \lambda)$, denoted by the updated blocks \mathbf{U}^* , \mathbf{Y}^* and \mathbf{V}^* . The normal equations of 5.66 can then be re-written in block form:

$$\begin{bmatrix} \mathbf{U}^* & \mathbf{X} & \mathbf{W} \\ \mathbf{X}^T & \mathbf{Y}^* & \mathbf{Z} \\ \mathbf{W}^T & \mathbf{Z}^T & \mathbf{V}^* \end{bmatrix} \begin{pmatrix} \delta\mathbf{a} \\ \delta\mathbf{b} \\ \delta\mathbf{c} \end{pmatrix} = \begin{pmatrix} \mathbf{e}_a \\ \mathbf{e}_b \\ \mathbf{e}_c \end{pmatrix} \quad (5.67)$$

We then proceed to apply a set of transformations to both sides of the equations that eliminate the upper triangle of blocks above the diagonal. We first apply the set of transformations to the left hand side of the equations.

$$\begin{bmatrix} \mathcal{I} & & \\ & \mathcal{I} & \\ & & \mathcal{I} \end{bmatrix} \begin{bmatrix} -\mathbf{W}\mathbf{V}^{*-1} \\ -\mathbf{Z}\mathbf{V}^{*-1} \\ \mathcal{I} \end{bmatrix} \begin{bmatrix} \mathbf{U}^* & \mathbf{X} & \mathbf{W} \\ \mathbf{X}^T & \mathbf{Y}^* & \mathbf{Z} \\ \mathbf{W}^T & \mathbf{Z}^T & \mathbf{V}^* \end{bmatrix} = \begin{bmatrix} \mathbf{U}' & \mathbf{X}' & \mathbf{0} \\ \mathbf{X}'^T & \mathbf{Y}' & \mathbf{0} \\ \mathbf{W}^T & \mathbf{Z}^T & \mathbf{V}^* \end{bmatrix} \quad (5.68)$$

$$\begin{bmatrix} \mathcal{I} & & \\ & \mathcal{I} & \\ & & \mathcal{I} \end{bmatrix} \begin{bmatrix} -\mathbf{X}'\mathbf{Y}'^{-1} \\ \mathcal{I} \\ \mathcal{I} \end{bmatrix} \begin{bmatrix} \mathbf{U}' & \mathbf{X}' & \mathbf{0} \\ \mathbf{X}'^T & \mathbf{Y}' & \mathbf{0} \\ \mathbf{W}^T & \mathbf{Z}^T & \mathbf{V}^* \end{bmatrix} = \begin{bmatrix} \mathbf{U}'' & \mathbf{0} & \mathbf{0} \\ \mathbf{X}''^T & \mathbf{Y}' & \mathbf{0} \\ \mathbf{W}^T & \mathbf{Z}^T & \mathbf{V}^* \end{bmatrix} \quad (5.69)$$

$$\mathbf{U}' = \mathbf{U}^* - \mathbf{W}\mathbf{V}^{*-1}\mathbf{W}^\top \quad (5.70)$$

$$\mathbf{X}' = \mathbf{X} - \mathbf{W}\mathbf{V}^{*-1}\mathbf{Z}^\top \quad (5.71)$$

$$\mathbf{Y}' = \mathbf{Y}^* - \mathbf{Z}\mathbf{V}^{*-1}\mathbf{Z}^\top \quad (5.72)$$

$$\mathbf{U}'' = \mathbf{U}' - \mathbf{X}'\mathbf{Y}'^{-1}\mathbf{X}'^\top \quad (5.73)$$

Then correspondingly apply the same transformations to the right hand side of the equations. The transformed right hand vector is updated as:

$$\begin{bmatrix} \mathcal{I} & -\mathbf{W}\mathbf{V}^{*-1} \\ & \mathcal{I} & -\mathbf{Z}\mathbf{V}^{*-1} \\ & & & \mathcal{I} \end{bmatrix} \begin{pmatrix} \mathbf{e}_a \\ \mathbf{e}_b \\ \mathbf{e}_c \end{pmatrix} = \begin{pmatrix} \mathbf{e}_a' \\ \mathbf{e}_b' \\ \mathbf{e}_c \end{pmatrix} \quad (5.74)$$

$$\begin{bmatrix} \mathcal{I} & -\mathbf{X}'\mathbf{Y}'^{-1} \\ & \mathcal{I} \\ & & \mathcal{I} \end{bmatrix} \begin{pmatrix} \mathbf{e}_a' \\ \mathbf{e}_b' \\ \mathbf{e}_c \end{pmatrix} = \begin{pmatrix} \mathbf{e}_a'' \\ \mathbf{e}_b' \\ \mathbf{e}_c \end{pmatrix} \quad (5.75)$$

$$\mathbf{e}_a' = \mathbf{e}_a - \mathbf{W}\mathbf{V}^{*-1}\mathbf{e}_c \quad (5.76)$$

$$\mathbf{e}_b' = \mathbf{e}_b - \mathbf{Z}\mathbf{V}^{*-1}\mathbf{e}_c \quad (5.77)$$

$$\mathbf{e}_a'' = \mathbf{e}_a' - \mathbf{X}'\mathbf{Y}'^{-1}\mathbf{e}_b' \quad (5.78)$$

The complete transformation matrix \mathbf{T} can be written as the product of the two individual matrices.

$$\begin{aligned} \mathbf{T} &= \begin{bmatrix} \mathcal{I} & -\mathbf{X}'\mathbf{Y}'^{-1} \\ & \mathcal{I} \\ & & \mathcal{I} \end{bmatrix} \begin{bmatrix} \mathcal{I} & -\mathbf{W}\mathbf{V}^{*-1} \\ & \mathcal{I} & -\mathbf{Z}\mathbf{V}^{*-1} \\ & & & \mathcal{I} \end{bmatrix} \\ \mathbf{T} &= \begin{bmatrix} \mathcal{I} & -\mathbf{X}'\mathbf{Y}'^{-1} & (\mathbf{X}'\mathbf{Y}'^{-1}\mathbf{Z}\mathbf{V}^{*-1} - \mathbf{W}\mathbf{V}^{*-1}) \\ & \mathcal{I} & -\mathbf{Z}\mathbf{V}^{*-1} \\ & & & \mathcal{I} \end{bmatrix} \quad (5.79) \end{aligned}$$

We then proceed to solve the set of linear equations.

$$\begin{bmatrix} \mathbf{U}'' & & \\ \mathbf{X}'^\top & \mathbf{Y}' & \\ \mathbf{W}^\top & \mathbf{Z}^\top & \mathbf{V}^* \end{bmatrix} \begin{pmatrix} \delta \mathbf{a} \\ \delta \mathbf{b} \\ \delta \mathbf{c} \end{pmatrix} = \begin{pmatrix} \mathbf{e}_a'' \\ \mathbf{e}'_b \\ \mathbf{e}_c \end{pmatrix} \quad (5.80)$$

$$\begin{aligned} \mathbf{U}'' \cdot \delta \mathbf{a} &= \mathbf{e}_a'' \\ \delta \mathbf{a} &= \mathbf{U}''^{-1} \mathbf{e}_a'' \end{aligned} \quad (5.81)$$

We find the parameter update vector $\delta \mathbf{b}$ by back substitution of the computed update vector $\delta \mathbf{a}$. Having already computed \mathbf{Y}'^{-1} previously, this step consists of only a few matrix multiplications.

$$\begin{aligned} \mathbf{X}'^\top \cdot \delta \mathbf{a} + \mathbf{Y}' \cdot \delta \mathbf{b} &= \mathbf{e}'_b \\ \delta \mathbf{b} &= \mathbf{Y}'^{-1} \cdot (\mathbf{e}'_b - \mathbf{X}'^\top \cdot \delta \mathbf{a}) \end{aligned} \quad (5.82)$$

We can then perform a final back substitution step with both $\delta \mathbf{a}$ and $\delta \mathbf{b}$ in order to compute the update vector $\delta \mathbf{c}$. Again \mathbf{V}^{*-1} has already been computed before so this step consists of only a few matrix multiplications.

$$\begin{aligned} \mathbf{W}^\top \cdot \delta \mathbf{a} + \mathbf{Z}^\top \cdot \delta \mathbf{b} + \mathbf{V}^* \cdot \delta \mathbf{c} &= \mathbf{e}_c \\ \delta \mathbf{c} &= \mathbf{V}^{*-1} (\mathbf{e}_c - \mathbf{W}^\top \cdot \delta \mathbf{a} - \mathbf{Z}^\top \cdot \delta \mathbf{b}) \end{aligned} \quad (5.83)$$

We then compute the new putative set of parameters $\mathbf{P}' = (\mathbf{a}^\top + \delta \mathbf{a}^\top, \mathbf{b}^\top + \delta \mathbf{b}^\top, \mathbf{c}^\top + \delta \mathbf{c}^\top)^\top$ and test whether \mathbf{P}' decreases the χ^2 fitting function. We update the Levenberg-Marquardt factor λ and parameter vector \mathbf{P} accordingly. The iteration process continues until convergence of the parameters, or the maximum number of iterations is exceeded. Besides changing the solution step of the normal equations, the parameter update and termination conditions remain unchanged from the basic implementation given in appendix C.2.

5.9.2 Covariance matrix

The covariance matrix $\Sigma_{\mathbf{p}}$ is computed by inverting the Hessian matrix $\Sigma_{\mathbf{p}} = (\mathbf{J}^\top \Sigma_{\mathbf{x}}^{-1} \mathbf{J})^{-1}$. Note that the unmodified diagonal blocks of the Hessian matrix $\mathbf{U}, \mathbf{Y}, \mathbf{V}$ are used here

instead of those formed from augmentation by the Levenberg-Marquardt scaling factor $(1 + \lambda)$. We proceed to compute the covariance matrix by Gaussian elimination.

$$\begin{bmatrix} \mathbf{U} & \mathbf{X} & \mathbf{W} \\ \mathbf{X}^\top & \mathbf{Y} & \mathbf{Z} \\ \mathbf{W}^\top & \mathbf{Z}^\top & \mathbf{V} \end{bmatrix} \begin{bmatrix} \Sigma_{\mathbf{U}} & \Sigma_{\mathbf{X}} & \Sigma_{\mathbf{W}} \\ \Sigma_{\mathbf{X}}^\top & \Sigma_{\mathbf{Y}} & \Sigma_{\mathbf{Z}} \\ \Sigma_{\mathbf{W}}^\top & \Sigma_{\mathbf{Z}}^\top & \Sigma_{\mathbf{V}} \end{bmatrix} = \begin{bmatrix} \mathcal{I} & & \\ & \mathcal{I} & \\ & & \mathcal{I} \end{bmatrix} \quad (5.84)$$

$$\begin{bmatrix} \mathbf{U}'' & & \\ \mathbf{X}'^\top & \mathbf{Y}' & \\ \mathbf{W}^\top & \mathbf{Z}^\top & \mathbf{V} \end{bmatrix} \begin{bmatrix} \Sigma_{\mathbf{U}} & \Sigma_{\mathbf{X}} & \Sigma_{\mathbf{W}} \\ \Sigma_{\mathbf{X}}^\top & \Sigma_{\mathbf{Y}} & \Sigma_{\mathbf{Z}} \\ \Sigma_{\mathbf{W}}^\top & \Sigma_{\mathbf{Z}}^\top & \Sigma_{\mathbf{V}} \end{bmatrix} = \mathbf{T} \quad (5.85)$$

where

$$\mathbf{T} = \begin{bmatrix} \mathcal{I} & -\mathbf{X}'\mathbf{Y}'^{-1} & (\mathbf{X}'\mathbf{Y}'^{-1}\mathbf{Z}\mathbf{V}^{-1} - \mathbf{W}\mathbf{V}^{-1}) \\ & \mathcal{I} & -\mathbf{Z}\mathbf{V}^{-1} \\ & & \mathcal{I} \end{bmatrix} \quad (5.86)$$

$$\begin{aligned} \mathbf{U}'' \cdot \Sigma_{\mathbf{U}} &= \mathcal{I} \\ \Sigma_{\mathbf{U}} &= \mathbf{U}''^{-1} \end{aligned} \quad (5.87)$$

$$\begin{aligned} \mathbf{U}'' \cdot \Sigma_{\mathbf{X}} &= -\mathbf{X}'\mathbf{Y}'^{-1} \\ \Sigma_{\mathbf{X}} &= -\mathbf{U}''^{-1}\mathbf{X}'\mathbf{Y}'^{-1} \\ \Sigma_{\mathbf{X}} &= -\Sigma_{\mathbf{U}}\mathbf{X}'\mathbf{Y}'^{-1} \end{aligned} \quad (5.88)$$

$$\begin{aligned} \mathbf{U}'' \cdot \Sigma_{\mathbf{W}} &= \mathbf{X}'\mathbf{Y}'^{-1}\mathbf{Z}\mathbf{V}^{-1} - \mathbf{W}\mathbf{V}^{-1} \\ \Sigma_{\mathbf{W}} &= \mathbf{U}''^{-1}\mathbf{X}'\mathbf{Y}'^{-1}\mathbf{Z}\mathbf{V}^{-1} - \mathbf{U}''^{-1}\mathbf{W}\mathbf{V}^{-1} \\ \Sigma_{\mathbf{W}} &= -\Sigma_{\mathbf{X}}\mathbf{Z}\mathbf{V}^{-1} - \Sigma_{\mathbf{U}}\mathbf{W}\mathbf{V}^{-1} \end{aligned} \quad (5.89)$$

$$\begin{aligned} \mathbf{X}'^\top \cdot \Sigma_{\mathbf{X}} + \mathbf{Y}' \cdot \Sigma_{\mathbf{Y}} &= \mathcal{I} \\ \Sigma_{\mathbf{Y}} &= \mathbf{Y}'^{-1} - \mathbf{Y}'^{-1}\mathbf{X}'^\top \Sigma_{\mathbf{X}} \\ \Sigma_{\mathbf{Y}} &= \mathbf{Y}'^{-1} - (\mathbf{X}'\mathbf{Y}'^{-1})^\top \Sigma_{\mathbf{X}} \end{aligned} \quad (5.90)$$

$$\begin{aligned}
\mathbf{X}'^\top \cdot \boldsymbol{\Sigma}_\mathbf{W} + \mathbf{Y}' \cdot \boldsymbol{\Sigma}_\mathbf{Z} &= -\mathbf{ZV}^{-1} \\
\boldsymbol{\Sigma}_\mathbf{Z} &= -\mathbf{Y}'^{-1}\mathbf{ZV}^{-1} - \mathbf{Y}'^{-1}\mathbf{X}'^\top \boldsymbol{\Sigma}_\mathbf{W} \\
\boldsymbol{\Sigma}_\mathbf{Z} &= -\mathbf{Y}'^{-1}\mathbf{ZV}^{-1} - \left(\mathbf{X}'\mathbf{Y}'^{-1}\right)^\top \boldsymbol{\Sigma}_\mathbf{W} \quad (5.91)
\end{aligned}$$

$$\begin{aligned}
\mathbf{W}^\top \cdot \boldsymbol{\Sigma}_\mathbf{W} + \mathbf{Z}^\top \cdot \boldsymbol{\Sigma}_\mathbf{Z} + \mathbf{V} \cdot \boldsymbol{\Sigma}_\mathbf{V} &= \mathcal{I} \\
\boldsymbol{\Sigma}_\mathbf{V} &= -\mathbf{V}^{-1}\mathbf{W}^\top \boldsymbol{\Sigma}_\mathbf{W} - \mathbf{V}^{-1}\mathbf{Z}^\top \boldsymbol{\Sigma}_\mathbf{Z} \\
\boldsymbol{\Sigma}_\mathbf{V} &= -(\mathbf{WV}^{-1})^\top \boldsymbol{\Sigma}_\mathbf{W} - (\mathbf{ZV}^{-1})^\top \boldsymbol{\Sigma}_\mathbf{Z} \quad (5.92)
\end{aligned}$$

5.9.3 Block sparse LM method

We have thus far ignored the independence of the set of parameters within each of the partitions. The computational overhead of naively inverting the \mathbf{V}^* block is the most dominant factor, while solving the normal equations. The independence of each set of *pose* parameters allows us to solve the normal equations while only inverting each of the independent \mathbf{V}_i^* blocks, a computation of the order $O(n)$ rather than the $O(n^3)$ required for the naive approach. It is difficult to express the tree relationship mathematically using familiar matrix notation, without causing unnecessary confusion. The strategy outline is best visualised, more by way of example rather than rigorous mathematical notation. Let a parameter vector \mathbf{P} , partitioned into three segments, represent the worldspace model.

$$\mathbf{P} = (\mathbf{a}^\top \mid \mathbf{b}^\top, \mathbf{b}'^\top, \dots \mid \mathbf{c}_1^\top, \mathbf{c}_2^\top, \mathbf{c}_3^\top, \mathbf{c}'_1^\top, \mathbf{c}'_2^\top, \dots)^\top \quad (5.93)$$

Where \mathbf{b} represents the *subject* parameters for the first person, \mathbf{b}' the parameters for the second person, \mathbf{b}'' the third person, and so forth. The corresponding set of circumstantial subject *pose* parameters are then given by \mathbf{c}_i for the first person, \mathbf{c}'_j for the second person, and so forth. Note that each subject can have an arbitrary number of *pose* parameter blocks, hence the use of different indices (i, j, \dots) and summations without qualification of limits.

The Jacobian matrix \mathbf{J} is sparsely structured and has the form shown in equation 5.94.

$$\mathbf{J} = \left[\begin{array}{c|c|c} \mathbf{A}_1 & \mathbf{B}_1 & \mathbf{C}_1 \\ \mathbf{A}_2 & \mathbf{B}_2 & \mathbf{C}_2 \\ \mathbf{A}_3 & \mathbf{B}_3 & \mathbf{C}_3 \\ \hline \mathbf{A}'_1 & \mathbf{B}'_1 & \mathbf{C}'_1 \\ \mathbf{A}'_2 & \mathbf{B}'_2 & \mathbf{C}'_2 \\ \hline \vdots & \vdots & \ddots \end{array} \right] \quad (5.94)$$

$$\begin{aligned} \mathbf{A}_i &= \frac{\partial \mathbf{x}_i}{\partial \mathbf{a}} & \mathbf{B}_i &= \frac{\partial \mathbf{x}_i}{\partial \mathbf{b}} & \mathbf{C}_i &= \frac{\partial \mathbf{x}_i}{\partial \mathbf{c}_i} \\ \mathbf{A}'_j &= \frac{\partial \mathbf{x}'_j}{\partial \mathbf{a}} & \mathbf{B}'_j &= \frac{\partial \mathbf{x}'_j}{\partial \mathbf{b}'} & \mathbf{C}'_j &= \frac{\partial \mathbf{x}'_j}{\partial \mathbf{c}'_j} \\ & \vdots & & \vdots & & \vdots \end{aligned} \quad (5.95)$$

The block form of the normal equations, corresponding to the general case shown in equation 5.67, then has a more sparsely structured shape. Note that we have chosen to align the ground plane coordinate system with the first subject pose within the system. The θ parameter within the first *pose* block is then fixed (zero), thus the Jacobian \mathbf{C}_1 has one less column than the remaining blocks. Correspondingly, the Hessian blocks \mathbf{W}_1 and \mathbf{Z}_1 also have one fewer columns, \mathbf{V}_1 both one fewer rows and columns, and $\delta \mathbf{c}_1, \mathbf{e}_{c1}$ one less row each.

$$\left[\begin{array}{c|c|c} \mathbf{U} & \mathbf{X} & \mathbf{X}' & \cdots & \mathbf{W}_1 & \mathbf{W}_2 & \mathbf{W}_3 & \mathbf{W}'_1 & \mathbf{W}'_2 & \cdots \\ \mathbf{X}^\top & \mathbf{Y} & & & \mathbf{Z}_1 & \mathbf{Z}_2 & \mathbf{Z}_3 & & & \\ \mathbf{X}'^\top & & \mathbf{Y}' & & & & & \mathbf{Z}'_1 & \mathbf{Z}'_2 & \\ \vdots & & & \ddots & & & & & & \cdots \\ \hline \mathbf{W}_1^\top & \mathbf{Z}_1^\top & & & \mathbf{V}_1 & & & & & \\ \mathbf{W}_2^\top & \mathbf{Z}_2^\top & & & & \mathbf{V}_2 & & & & \\ \mathbf{W}_3^\top & \mathbf{Z}_3^\top & & & & & \mathbf{V}_3 & & & \\ \mathbf{W}'_1^\top & & \mathbf{Z}'_1{}^\top & & & & & \mathbf{V}'_1 & & \\ \mathbf{W}'_2^\top & & \mathbf{Z}'_2{}^\top & & & & & & \mathbf{V}'_2 & \\ \vdots & & & \vdots & & & & & & \ddots \end{array} \right] \begin{pmatrix} \delta \mathbf{a} \\ \delta \mathbf{b} \\ \delta \mathbf{b}' \\ \vdots \\ \delta \mathbf{c}_1 \\ \delta \mathbf{c}_2 \\ \delta \mathbf{c}_3 \\ \delta \mathbf{c}'_1 \\ \delta \mathbf{c}'_2 \\ \vdots \end{pmatrix} = \begin{pmatrix} \mathbf{e}_a \\ \mathbf{e}_b \\ \mathbf{e}_{b'} \\ \vdots \\ \mathbf{e}_{c1} \\ \mathbf{e}_{c2} \\ \mathbf{e}_{c3} \\ \mathbf{e}_{c'1} \\ \mathbf{e}_{c'2} \\ \vdots \end{pmatrix} \quad (5.96)$$

$$\begin{aligned}
\mathbf{U} &= \sum_i \mathbf{A}_i^\top \Sigma_{\mathbf{x}i}^{-1} \mathbf{A}_i + \sum_j \mathbf{A}'_j{}^\top \Sigma_{\mathbf{x}'j}^{-1} \mathbf{A}'_j + \dots \\
\mathbf{X} &= \sum_i \mathbf{A}_i^\top \Sigma_{\mathbf{x}i}^{-1} \mathbf{B}_i & \mathbf{X}' &= \sum_j \mathbf{A}'_j{}^\top \Sigma_{\mathbf{x}'j}^{-1} \mathbf{B}'_j & \dots \\
\mathbf{W}_i &= \mathbf{A}_i^\top \Sigma_{\mathbf{x}i}^{-1} \mathbf{C}_i & \mathbf{W}'_j &= \mathbf{A}'_j{}^\top \Sigma_{\mathbf{x}'j}^{-1} \mathbf{C}'_j & \dots \\
\mathbf{Y} &= \sum_i \mathbf{B}_i^\top \Sigma_{\mathbf{x}i}^{-1} \mathbf{B}_i & \mathbf{Y}' &= \sum_j \mathbf{B}'_j{}^\top \Sigma_{\mathbf{x}'j}^{-1} \mathbf{B}'_j & \dots \\
\mathbf{Z}_i &= \mathbf{B}_i^\top \Sigma_{\mathbf{x}i}^{-1} \mathbf{C}_i & \mathbf{Z}'_j &= \mathbf{B}'_j{}^\top \Sigma_{\mathbf{x}'j}^{-1} \mathbf{C}'_j & \dots \\
\mathbf{V}_i &= \mathbf{C}_i^\top \Sigma_{\mathbf{x}i}^{-1} \mathbf{C}_i & \mathbf{V}'_j &= \mathbf{C}'_j{}^\top \Sigma_{\mathbf{x}'j}^{-1} \mathbf{C}'_j & \dots
\end{aligned} \tag{5.97}$$

$$\begin{aligned}
\mathbf{e}_a &= \sum_i \mathbf{A}_i^\top \Sigma_{\mathbf{x}i}^{-1} \mathbf{r}_i + \sum_j \mathbf{A}'_j{}^\top \Sigma_{\mathbf{x}'j}^{-1} \mathbf{r}'_j + \dots \\
\mathbf{e}_b &= \sum_i \mathbf{B}_i^\top \Sigma_{\mathbf{x}i}^{-1} \mathbf{r}_i & \mathbf{e}_{b'} &= \sum_j \mathbf{B}'_j{}^\top \Sigma_{\mathbf{x}'j}^{-1} \mathbf{r}'_j & \dots \\
\mathbf{e}_{ci} &= \mathbf{C}_i^\top \Sigma_{\mathbf{x}i}^{-1} \mathbf{r}_i & \mathbf{e}_{c'j} &= \mathbf{C}'_j{}^\top \Sigma_{\mathbf{x}'j}^{-1} \mathbf{r}'_j & \dots
\end{aligned} \tag{5.98}$$

We augment the diagonal elements of the Hessian matrix with the Levenberg-Marquardt scaling factor $(1+\lambda)$, then apply the transformation matrix to both sides of the equations to eliminate the \mathbf{W} and \mathbf{Z} blocks.

$$\left[\begin{array}{c|ccc|cccc}
\mathcal{I} & & & & -\mathbf{W}_1 \mathbf{V}_1^{*-1} & -\mathbf{W}_2 \mathbf{V}_2^{*-1} & -\mathbf{W}_3 \mathbf{V}_3^{*-1} & -\mathbf{W}'_1 \mathbf{V}'_1{}^{*-1} & -\mathbf{W}'_2 \mathbf{V}'_2{}^{*-1} & \dots \\
\mathcal{I} & & & & -\mathbf{Z}_1 \mathbf{V}_1^{*-1} & -\mathbf{Z}_2 \mathbf{V}_2^{*-1} & -\mathbf{Z}_3 \mathbf{V}_3^{*-1} & & & \\
\mathcal{I} & & & & & & & -\mathbf{Z}'_1 \mathbf{V}'_1{}^{*-1} & -\mathbf{Z}'_2 \mathbf{V}'_2{}^{*-1} & \\
\vdots & & & & & & & & & \dots \\
\hline
\mathcal{I} & & & & & & & & & \\
& \mathcal{I} & & & & & & & & \\
& & \mathcal{I} & & & & & & & \\
& & & \mathcal{I} & & & & & & \\
& & & & \mathcal{I} & & & & & \\
& & & & & \mathcal{I} & & & & \\
& & & & & & \mathcal{I} & & & \\
& & & & & & & \mathcal{I} & & \\
& & & & & & & & \mathcal{I} & \\
& & & & & & & & & \ddots \\
& & & & & & & & & \dots
\end{array} \right] \tag{5.99}$$

The set of transformed normal equations are then of the form.

We can then perform a final back substitution step in order to compute the update vectors $(\delta \mathbf{c}_i, \delta \mathbf{c}'_j, \dots)$. Again the individual blocks $(\mathbf{V}_i^{*-1}, \mathbf{V}'_j{}^{*-1}, \dots)$ have already been computed before, so this step consists of only a few matrix multiplications.

$$\left. \begin{aligned} \mathbf{W}_i^\top \cdot \delta \mathbf{a} + \mathbf{Z}_i^\top \cdot \delta \mathbf{b} + \mathbf{V}_i^* \cdot \delta \mathbf{c}_i &= \mathbf{e}_{c_i} \\ \mathbf{W}'_j{}^\top \cdot \delta \mathbf{a} + \mathbf{Z}'_j{}^\top \cdot \delta \mathbf{b}' + \mathbf{V}'_j{}^* \cdot \delta \mathbf{c}'_j &= \mathbf{e}_{c'_j} \end{aligned} \right\} \dots \quad (5.111)$$

$$\left. \begin{aligned} \delta \mathbf{c}_i &= \mathbf{V}_i^{*-1} (\mathbf{e}_{c_i} - \mathbf{W}_i^\top \cdot \delta \mathbf{a} - \mathbf{Z}_i^\top \cdot \delta \mathbf{b}) \\ \delta \mathbf{c}'_j &= \mathbf{V}'_j{}^{*-1} (\mathbf{e}_{c'_j} - \mathbf{W}'_j{}^\top \cdot \delta \mathbf{a} - \mathbf{Z}'_j{}^\top \cdot \delta \mathbf{b}') \end{aligned} \right\} \dots \quad (5.112)$$

The covariance matrix is computed in much the same way. Note that the unmodified diagonal blocks of the Hessian matrix $\mathbf{U}, \mathbf{Y}, \mathbf{V}$ are used here instead of those formed from augmentation by the Levenberg-Marquardt scaling factor $(1 + \lambda)$.

Since the pose projection \mathbf{M} discussed within chapter 4 over parameterises the mapping from metric to image planes, computation of the covariance matrix for each set of individual linear motion segments requires inverting the Hessian by use of the pseudo inverse. Here, the parameterisation of the worldspace model \mathbf{P} is of full rank, thus inversion of all the required matrices can be performed by using any of the standard methods.

Only the diagonal elements of the covariance matrix are required for the computation of the confidence limits of each parameter independently. With a little extra effort the set of cross covariance blocks can also be computed, though this will not be discussed further. Without further working we give the equations required to compute the diagonal blocks of the covariance matrix.

$$\Sigma_{\mathbf{U}} = \ddot{\mathbf{U}}^{-1} \quad (5.113)$$

$$\left. \begin{aligned} \Sigma_{\mathbf{X}} &= -\Sigma_{\mathbf{U}} \tilde{\mathbf{X}} \tilde{\mathbf{Y}}^{-1} \\ \Sigma_{\mathbf{X}'} &= -\Sigma_{\mathbf{U}} \tilde{\mathbf{X}}' \tilde{\mathbf{Y}}'^{-1} \end{aligned} \right\} \dots$$

$$\left. \begin{aligned} \Sigma_{\mathbf{W}_i} &= -\Sigma_{\mathbf{X}} \mathbf{Z}_i \mathbf{V}_i^{-1} - \Sigma_{\mathbf{U}} \mathbf{W}_i \mathbf{V}_i^{-1} \\ \Sigma_{\mathbf{W}'_j} &= -\Sigma_{\mathbf{X}'} \mathbf{Z}'_j \mathbf{V}'_j{}^{-1} - \Sigma_{\mathbf{U}} \mathbf{W}'_j \mathbf{V}'_j{}^{-1} \end{aligned} \right\} \dots$$

$$\left. \begin{aligned} \Sigma_{\mathbf{Y}} &= \tilde{\mathbf{Y}} - (\tilde{\mathbf{X}} \tilde{\mathbf{Y}}^{-1})^\top \Sigma_{\mathbf{X}} \\ \Sigma_{\mathbf{Y}'} &= \tilde{\mathbf{Y}}' - (\tilde{\mathbf{X}}' \tilde{\mathbf{Y}}'^{-1})^\top \Sigma_{\mathbf{X}'} \end{aligned} \right\} \dots \quad (5.114)$$

$$\left. \begin{aligned} \Sigma_{\mathbf{Z}_i} &= - \left(\tilde{\mathbf{Y}}^{-1} \mathbf{Z}_i \mathbf{V}_i^{-1} \right) - \left(\tilde{\mathbf{X}} \tilde{\mathbf{Y}}^{-1} \right)^\top \Sigma_{\mathbf{W}_i} \\ \Sigma_{\mathbf{Z}'_j} &= - \left(\tilde{\mathbf{Y}}'^{-1} \mathbf{Z}'_j \mathbf{V}'_j{}^{-1} \right) - \left(\tilde{\mathbf{X}}' \tilde{\mathbf{Y}}'^{-1} \right)^\top \Sigma_{\mathbf{W}'_j} \end{aligned} \right\} \dots \\
\left. \begin{aligned} \Sigma_{\mathbf{V}_i} &= - \left(\mathbf{W}_i \mathbf{V}_i^{-1} \right)^\top \Sigma_{\mathbf{W}_i} - \left(\mathbf{Z}_i \mathbf{V}_i^{-1} \right)^\top \Sigma_{\mathbf{Z}_i} \\ \Sigma_{\mathbf{V}'_j} &= - \left(\mathbf{W}'_j \mathbf{V}'_j{}^{-1} \right)^\top \Sigma_{\mathbf{W}'_j} - \left(\mathbf{Z}'_j \mathbf{V}'_j{}^{-1} \right)^\top \Sigma_{\mathbf{Z}'_j} \end{aligned} \right\} \dots \quad (5.115)$$

5.10 Reconstruction analysis

Table 5.10 shows a comparison between the root mean squared residual errors for each individual linear trajectory segment \mathbf{M} , the fused parameter estimates and after global maximum likelihood estimation. Individual reprojection by \mathbf{M} does not take the components of radial lens distortion into account. Unsurprisingly, the reprojection errors after initial parameter fusion are greater than the individual \mathbf{M} errors. We can clearly see that the global MLE improves the residual fitting errors, when compared to the initial reconstruction using the fused parameters. The difference in residual errors with and without components of radial distortion is quite noticeable. The mean percentage error increase over each independent \mathbf{M} error, without radial components is $\simeq 24\%$, while inclusion of \mathbf{K}_r often makes significant reductions in the percentage fitting errors. The mean percentage error increase by including the coefficients \mathbf{K}_r is $\simeq 1\%$, an order of magnitude better than modelling the motion without radial lens distortion.

<i>Seg</i>	\mathbf{M}	<i>Fused</i> (\mathbf{K}_c)	<i>Fused</i> ($\mathbf{K}_c, \mathbf{K}_r$)	<i>MLE</i> (\mathbf{K}_c)	<i>MLE</i> ($\mathbf{K}_c, \mathbf{K}_r$)
a	3.102	3.802 (22.56%)	3.925 (26.53%)	3.213 (3.57%)	1.423 (-54.12%)
b	1.048	2.381 (127.19%)	1.935 (84.63%)	1.439 (37.30%)	1.016 (-3.05%)
c	1.321	2.733 (106.88%)	2.682 (103.02%)	2.292 (73.50%)	1.971 (49.20%)
d	1.398	2.425 (73.46%)	2.021 (44.56%)	1.626 (16.30%)	1.188 (-15.02%)
e1	1.619	2.846 (75.78%)	2.717 (67.82%)	1.729 (6.79%)	1.893 (16.92%)
e2	1.189	2.531 (112.86%)	2.448 (105.88%)	1.292 (8.66%)	1.342 (12.86%)

TABLE 5.10: Piecewise linear root mean square reprojection errors shown for each individual trajectory segment \mathbf{M} , after parameter fusion and maximum likelihood estimation with and without radial distortion coefficients \mathbf{K}_r . Values in brackets are percentage errors $\Delta\epsilon$ between computed and \mathbf{M} residual r.m.s. errors.

Figures 5.15 and 5.16 show the comparison of the reconstructed leg angle functions $\tilde{\theta}(t)$ between minimized subject biometric coefficients \mathbf{B} after global MLE, and each of the individually computed sets of coefficients $\tilde{\mathbf{B}}_i$.

We can clearly see the improvement in reprojection error by including the full camera distortion model $\mathbf{K}_c, \mathbf{K}_r$ within the optimization procedure. The trajectory segment of subject motion, shown within figure 5.17, covers a large proportion of the camera field

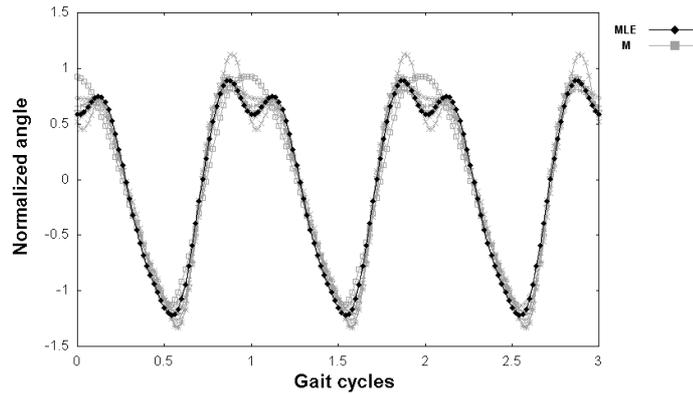


FIGURE 5.15: Comparison of upper normalized limb angle functions $\tilde{\theta}(t)$ shown for each of the individual reconstructed linear trajectory segments \mathbf{M} and after maximum likelihood estimation.

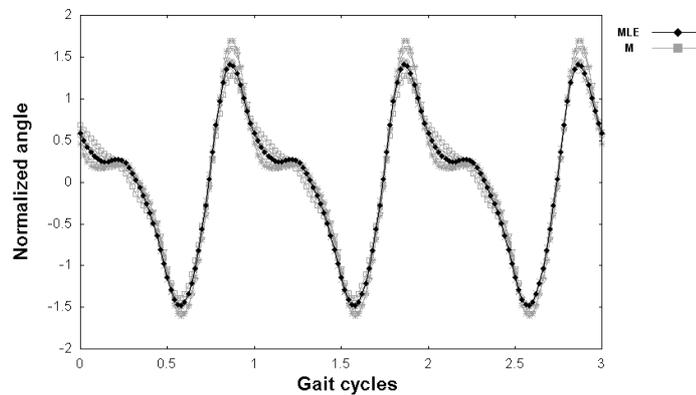


FIGURE 5.16: Comparison of lower normalized limb angle functions $\tilde{\theta}(t)$ shown for each of the individual reconstructed linear trajectory segments \mathbf{M} and after maximum likelihood estimation.

of view. The middle frames corresponding to the pose reprojection \mathbf{M} match quite well, while image frames at the beginning and end of the sequence show significant differences. Due to the effects of radial lens distortion, the residual image error is distributed unevenly, with an increasing level of error seen further from the camera principal point. The representation of imaged subject motion using the camera parameters $\mathbf{K}_c, \mathbf{K}_r$ is sufficiently well modelled. The corresponding residual errors are distributed evenly over the set of image sequence frames.

5.10.1 System parameters

We compare here the set of reconstructed camera parameters, found by global parameter optimization, to those obtained through an accurate calibration algorithm. The baseline camera coefficients are computed by performing Zhang's calibration algorithm [120], with a sufficiently large number of image frames. Table 5.11 shows the camera intrinsic parameters \mathbf{K}_c computed for the global optimization, with and without components

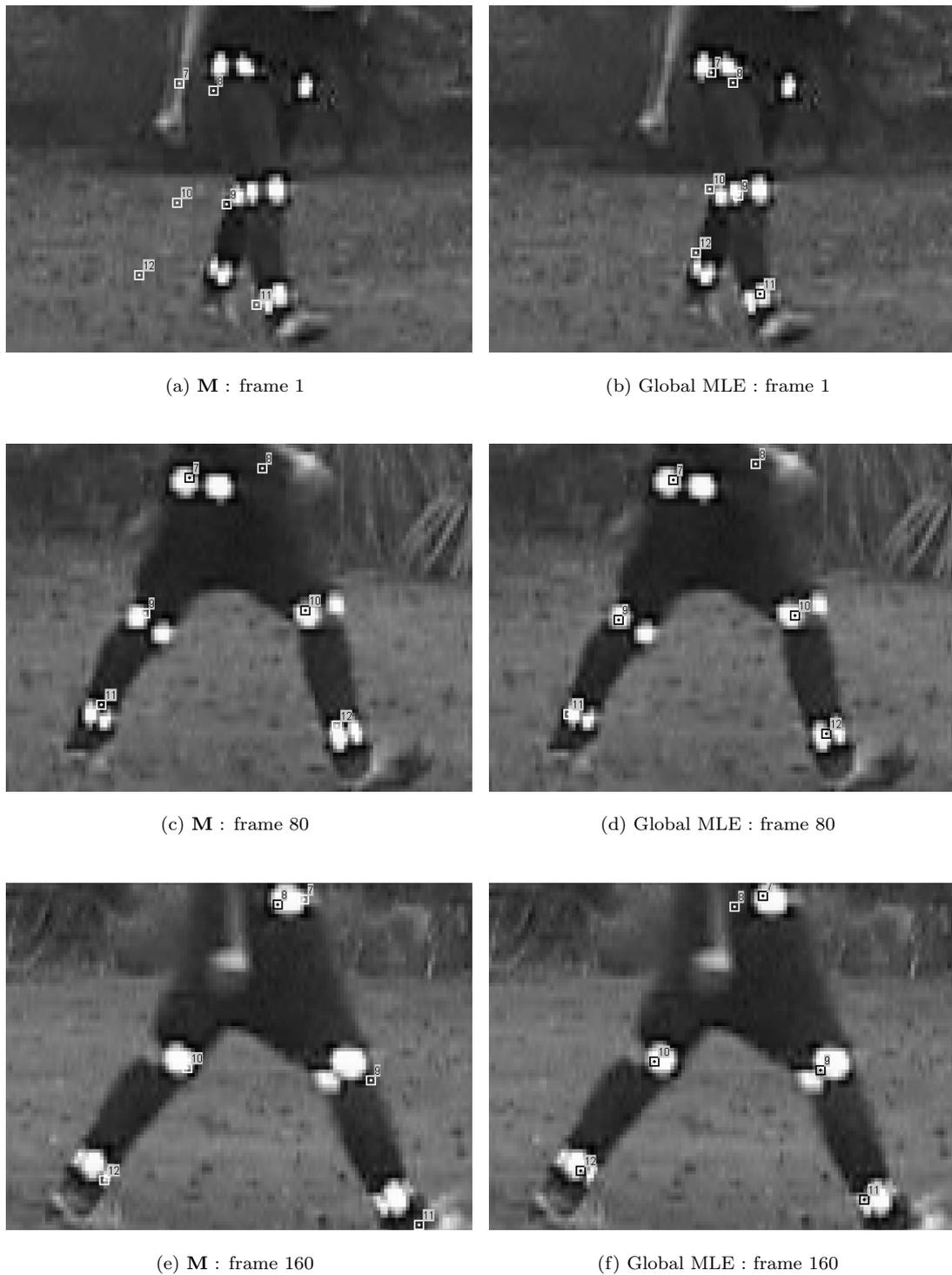


FIGURE 5.17: Comparison of the reprojected limb points for individual pose parameterisations M (a,c,e) and after global MLE with camera coefficients K_c, K_r (b,d,f). The image frames are taken from the beginning, middle and end of the motion trajectory segment 'a'.

of radial distortion \mathbf{K}_r . The levels of uncertainty within each parameterisation are indicative of what we might expect.

<i>Method</i>	m_x	m_y	u_0	v_0
MLE (\mathbf{K}_c)	830.60 ± 39.62	923.64 ± 35.71	230.48 ± 20.97	531.86 ± 35.61
MLE ($\mathbf{K}_c, \mathbf{K}_r$)	707.91 ± 25.59	815.05 ± 26.01	360.49 ± 4.03	221.98 ± 3.76
Zhang	865.28 ± 1.24	947.93 ± 1.37	355.64 ± 2.27	274.19 ± 2.44

TABLE 5.11: Camera intrinsic coefficients with corresponding 3σ confidence limits computed from the worldspace gait model. Accurate calibration coefficients are given by performing Zhang’s calibration algorithm.

The parameters found by modelling the worldspace motion without \mathbf{K}_r have much higher uncertainties than those with distortion coefficients. The computed principal point is also significantly different to the baseline values. The distribution of worldspace data is mostly within the X and Z directions of the ground plane coordinate system. The imaged data within these directions provides most of the fitting constraints, while the principal point will tend to drift within the imaged Y direction, in order to compensate for the apparent effect of radial distortion through projective warping. The worldspace subject t_y positions are then free to take whatever values that are necessary to achieve the required reprojection, and unsurprisingly are also significantly different from the baseline values.

The parameters found by modelling the worldspace motion with distortion coefficients \mathbf{K}_r have similar values to the baseline parameters. The difference in focal parameters can be explained by the close coupling with the lens distortion coefficients. Lens distortion scales the image radially from the principal point, thus the focal parameters will also be adjusted to compensate for the apparent difference between baseline and computed radial parameters. Table 5.12 shows the computed and baseline radial distortion coefficients \mathbf{K}_r and their corresponding uncertainties.

<i>Method</i>	k_1	k_2
MLE ($\mathbf{K}_c, \mathbf{K}_r$)	-0.135964 ± 0.049123	-1.856917 ± 0.352171
Zhang	-0.224416 ± 0.007648	0.305238 ± 0.045303

TABLE 5.12: Camera radial distortion coefficients with corresponding 3σ confidence limits computed from the worldspace gait model. Accurate radial coefficients are given by performing Zhang’s calibration algorithm.

Comparison of the radial coefficients are best visualised by plotting the distortion profile for both sets of parameters. Figure 5.18 shows a plot of the radial distortion function $f(r) = 1 + k_1 \cdot r^2 + k_2 \cdot r^4$ within the visible range of rays entering the lens. The corresponding plot shown in figure 5.19 shows the complete distortion mapping of rays entering the lens with projection onto the focal plane $x(r) = m_x \cdot r \cdot f(r)$, for the cross section of the lens X axis.

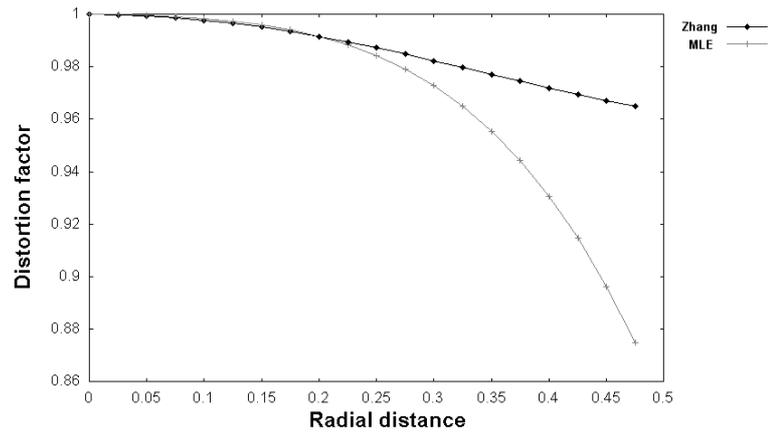


FIGURE 5.18: Comparison of lens distortions for MLE and baseline coefficients. Lens distortion function $f(r)$.

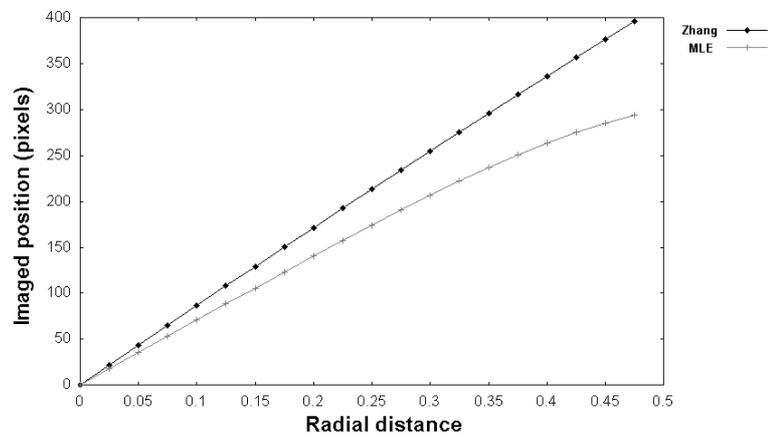


FIGURE 5.19: Comparison of lens distortions for MLE and baseline coefficients. Lens projection onto focal plane $x(r) = m_x \cdot r \cdot f(r)$.

The disparity between lens distortion mappings is evident in figure 5.19. The close coupling between parameters of the projection process $\mathbf{K}_c, \mathbf{K}_r$ and workspace points \mathbf{X} means that if we perturb one set of parameters away from the true values then the other two parameters must also be adjusted to compensate. Subsequently if the optimal χ^2 error is achieved with a set of radial distortion coefficients \mathbf{K}_r slightly different from the true baseline values then both \mathbf{K}_c and workspace points \mathbf{X} will also differ in value. Figure 5.20 shows the comparison of reconstructed workspace points between the computed optimal and baseline values.

As the experimenter, we are able to control many of the aspects of workspace motion that may influence the uncertainty within the modelling process. We can ensure that the ground plane is as flat as possible by choosing a suitable location, and that motion is piecewise linear by manually picking the trajectory segments within the image sequence. If we fix the camera intrinsic parameters during optimization with the true baseline values of \mathbf{K}_c and \mathbf{K}_r , then a substantial proportion of the residual errors can be attributed to the uncertainty within the underlying biometric and circumstantial parameters of

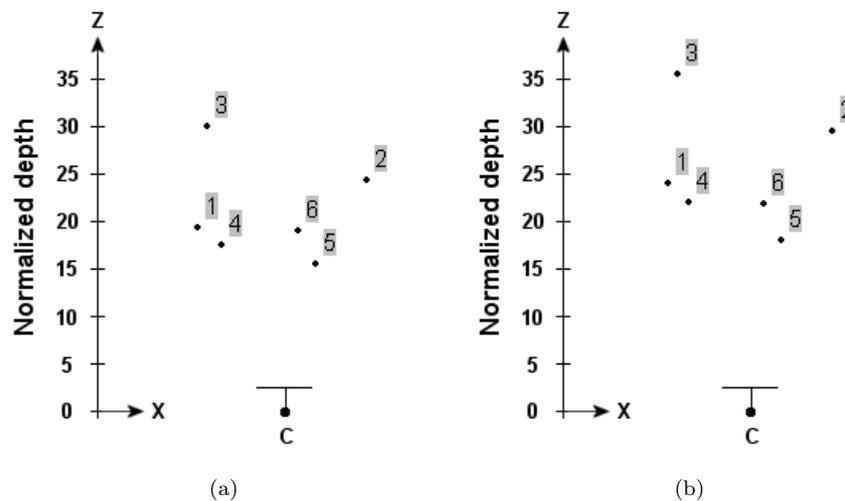


FIGURE 5.20: **Comparison of reconstructed workspace points.** The plotted workspace points represent the initial positions of the subject at the start of the trajectory segment within the XZ ground plane. (a) Positions computed from the optimized $\mathbf{K}_c, \mathbf{K}_r$ camera parameters. (b) Positions computed with the fixed baseline camera intrinsic parameters.

subject motion.

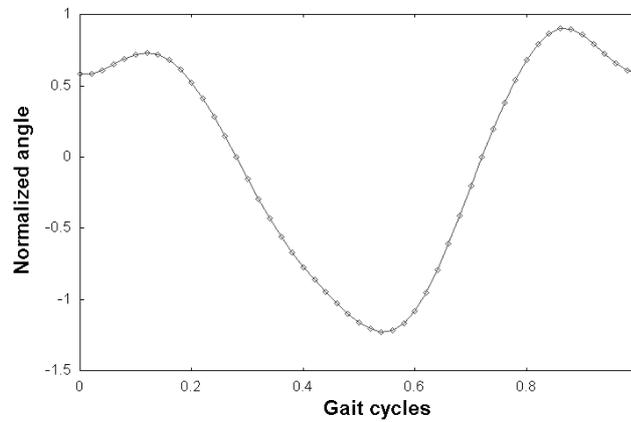
Table 5.13 shows that the greatest percentage error $\Delta\epsilon$ improvement in fitting is achieved for imaged subject motion covering a significant proportion of the field of view. The worst percentage error increase corresponds to the most oblique trajectory, with imaged motion within a localized region of the image space.

<i>Seg</i>	\mathbf{M}	<i>MLE</i>	<i>MLE</i> ($\mathbf{K}_c, \mathbf{K}_r$)
a	3.102	3.015 (-2.805%)	1.423 (-54.126%)
b	1.048	1.111 (6.011%)	1.016 (-3.053%)
c	1.321	2.392 (81.075%)	1.971 (49.205%)
d	1.398	1.543 (10.372%)	1.188 (-15.021%)
e1	1.619	1.909 (17.912%)	1.893 (16.924%)
e2	1.189	1.358 (14.214%)	1.342 (12.868%)

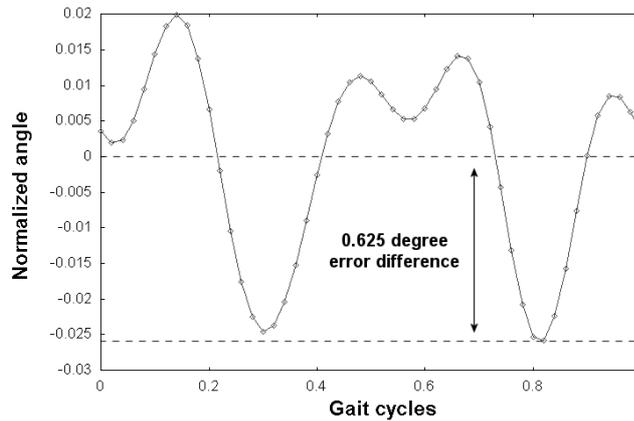
TABLE 5.13: Piecewise linear root mean square reprojection errors shown for each individual trajectory segment \mathbf{M} and maximum likelihood estimation with fixed baseline camera parameters (MLE) and computed intrinsic coefficients $\mathbf{K}_c, \mathbf{K}_r$. Values in brackets are percentage errors $\Delta\epsilon$ between computed and \mathbf{M} residual r.m.s. errors.

The underlying subject biometric parameters are similar for both sets of estimated parameters. Figures 5.21 and 5.22 show a comparison of both reconstructed upper and lower leg normalized angle functions $\tilde{\theta}(t)$ over a single period of gait. The maximum error difference between the optimized and baseline reconstructions of the upper normalized leg angle functions is 0.025, corresponding to a 0.625 degree limb angle difference. The maximum error difference between the optimized and baseline reconstructions of the

lower normalized leg angle functions is 0.02, which corresponds to a 0.8 degree limb angle difference.



(a)



(b)

FIGURE 5.21: **Comparison of the reconstructed upper normalized limb angle functions $\tilde{\theta}(t)$.** (a) Optimized parameters computed with the fixed baseline camera parameters (MLE). (b) Difference between reconstructed limb functions computed with the minimized camera intrinsic parameters $\mathbf{K}_c, \mathbf{K}_r$ and with the fixed baseline camera parameters.

Since the reconstruction algorithm is based on metric rectification of landmark points, the position and uncertainty within which these markers are measured is the major contributing factor within the accuracy of the reconstruction. To ensure that an imaged limb joint can be marked for any arbitrary pose within the entire image sequence, four retro-reflective marker patches were attached to the opposing faces of each joint site on the body. The choice of the four imaged landmarks for each limb joint was picked consistently over a linear trajectory segment. However, the use of all available trajectory segments in the global optimization circumvents a consistent choice of markers over the entire image sequence. We have chosen to robustly parameterise this difference in marker placement within the set of poses, by ensuring independent inter-plane scaling factors τ

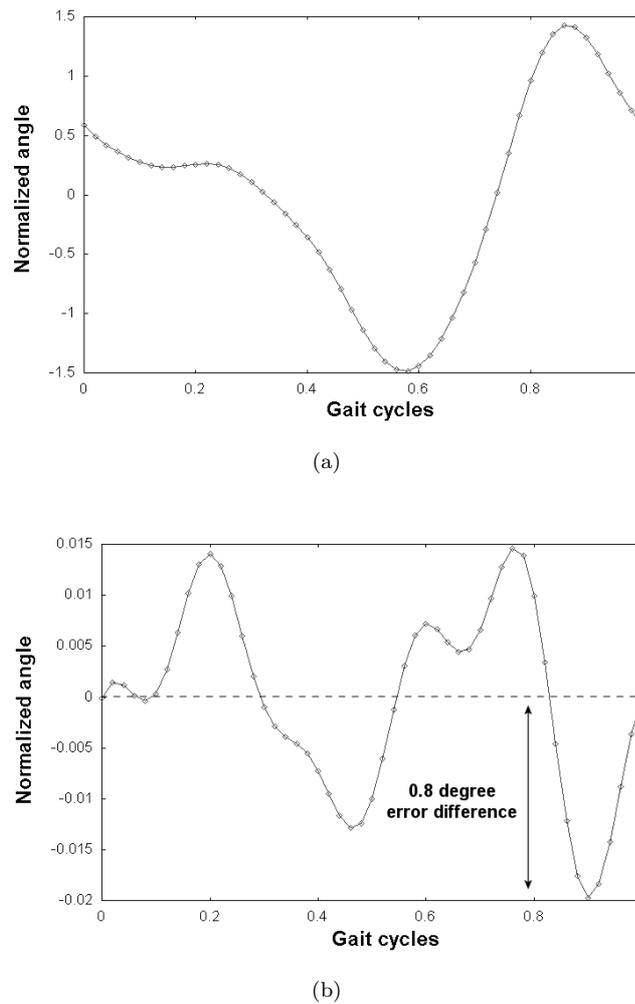


FIGURE 5.22: **Comparison of the reconstructed lower normalized limb angle functions $\tilde{\theta}(t)$.** (a) Optimized parameters computed with the fixed baseline camera parameters (MLE). (b) Difference between reconstructed limb functions computed with the minimized camera intrinsic parameters $\mathbf{K}_c, \mathbf{K}_r$ and with the fixed baseline camera parameters.

and inclination angles α within the circumstantial pose parameters \mathbf{Z} of subject motion. Symmetric markers on opposite sides of the body are not visible within the images. Joint markers on the outside of one leg coupled with markers on the inside of the other leg are used in all the fronto-parallel views, hence the mid symmetry plane between the marker set may not be entirely parallel with the ground plane normal, thus accounting for a significant proportion of the source of deviation between worldspace motion model and what we actually measure.

The assumption of constant velocity over a linear segment between way-points may not be entirely true. Subjects tend to slow down while approaching turning points and, in order to make a smooth transition between trajectories, the motion path becomes slightly elliptical. Restricting the camera intrinsic parameters by fixing \mathbf{K}_c and \mathbf{K}_r

with the baseline coefficients in effect reduces the degrees of freedom in which we can distort the motion to our advantage. For example, speed up/down effects of motion across the field of view can be removed by applying a suitable adjustment to the radial coefficients to approximate rectification back to linear motion. All but one of the linear trajectories account for imaged motion across the field of view. The set of turning points are all imaged within a small band at the edges of the view. It then seems reasonable to conclude that the apparent linearization of gait motion, by adjustment of the radial lens coefficients, accounts for some of the differences between baseline and computed camera parameters.

<i>Method</i>	v_1	v_2	v_3
MLE	0.007103 ± 0.004421	0.319089 ± 0.015115	0.016467 ± 0.001186

TABLE 5.14: Computed Rodrigues vector \mathbf{N} corresponding to the pose rotation of the worldspace ground plane. The corresponding uncertainties are computed at the 3σ confidence limits.

Table 5.14 shows the computed worldspace ground plane pose rotation, encoded as a Rodrigues 3-vector, and the corresponding uncertainties. We must note that the reconstructed worldspace is not the same as the true worldspace, since we have chosen to normalize the set of upper leg segments for each subject to unit length. Consequently, the trajectories are similar to those of the true metric worldspace, but distances from the camera are scaled. The properties of parallelism and distance ratio relationships hold, although intersection constraints do not. Without prior knowledge of each subject's height, that enables us to reconstruct the true metric worldspace, we are unable to find the intersection point where two people may meet in space. In essence, the reconstructed space may be termed a normalized space.

5.10.2 Subject and pose parameters

Tables 5.15 to 5.24 at the end of the chapter, show the computed subject biometric parameter estimates of \mathbf{B} and the corresponding confidence limits, after performing global optimization on all available parameters. Each table also contains the same parameter estimates \mathbf{B} and corresponding confidence limits, computed from each of the independent pose \mathbf{M} maximum likelihood estimations. We can calculate the mean uncertainties from this set of individual trajectory parameter estimates. The mean parameter uncertainties can then be compared to those obtained through global optimization. Figures 5.23 and 5.24 show the comparison between these estimated uncertainties. There is a clear reduction in the computed parameter uncertainties, achieved by performing the global optimization which can be attributed to the more realistic worldspace motion model and the fact that there is a larger supporting measurement data set. We can then quantify this *reduction* in uncertainty between both parameterisations by quoting the uncertainty differences $\Delta\sigma$, as a percentage of the computed individual mean uncer-

ainties. The mean percentage difference uncertainty reduction is of the order $\simeq 72\%$ between the finally estimated and mean of the linear trajectory uncertainty estimates.

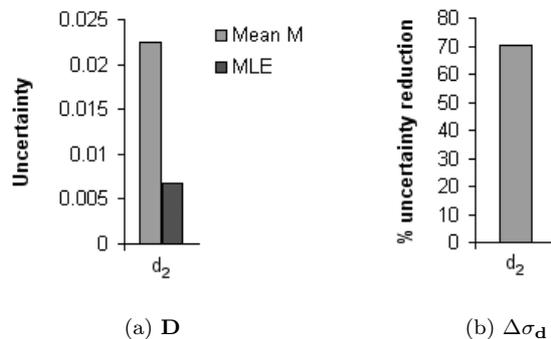


FIGURE 5.23: Estimated subject biometric parameter uncertainties computed at the 3σ confidence limits. Shown here are the mean uncertainties attributed to the independent set of pose projections \mathbf{M} and the uncertainties computed from the global optimization. The reduction in uncertainties between parameterisations is quantified by the uncertainty difference as a percentage of the mean individual uncertainties $\Delta\sigma$.

We have already briefly discussed the effects of non constant velocity on the intrinsic parameters of the camera. In order to correct for apparent changes in velocity, the radial lens distortion parameters are adjusted from their true values to effectively linearize the motion. We may then wish to consider replacing the simple linear displacement function $x(t) = at + b$ within the biometric motion equation $g(t)$ with a suitable quadratic function $x(t) = at^2 + bt + c$. The linear velocity assumption precludes that the dynamics of gait remain constant. If we add an acceleration term then we must also couple this with a change in cadence f_0 . We can then model this change in fundamental frequency by the linear function $f(t) = \omega t + f_0$. The modified Fourier series biometric motion function can then be modelled by the equation.

$$\begin{aligned}
 x(t) = & (at^2 + bt + c) + a_1 \cos(2\pi(\omega t + f_0) \cdot (t + t_s)) + \\
 & a_1 \cdot \sum_{k=2}^n b_k \cos(2\pi k \cdot (\omega t + f_0) \cdot (t + t_s) + \psi_k)
 \end{aligned} \tag{5.116}$$

Each linear trajectory reconstruction \mathbf{M} then computes the initial linear estimates of velocity and cadence. The higher order terms a, ω may then be initialised to zero before optimization, and form part of the subject circumstantial parameters \mathbf{W} .

$$\mathbf{W} = (a, b, \omega, f_0, \mathbf{p}^\top, \mathbf{q}^\top, \mathbf{r}^\top)^\top \tag{5.117}$$

where vector \mathbf{p} contains the c offset coefficients of \mathbf{V} , vector \mathbf{q} contains the a_1 amplitude coefficients of $\mathbf{X}, \mathbf{Y}, \mathbf{V}$ and vector \mathbf{r} contains the t_s time shift coefficients of $\mathbf{X}, \mathbf{Y}, \mathbf{V}$.

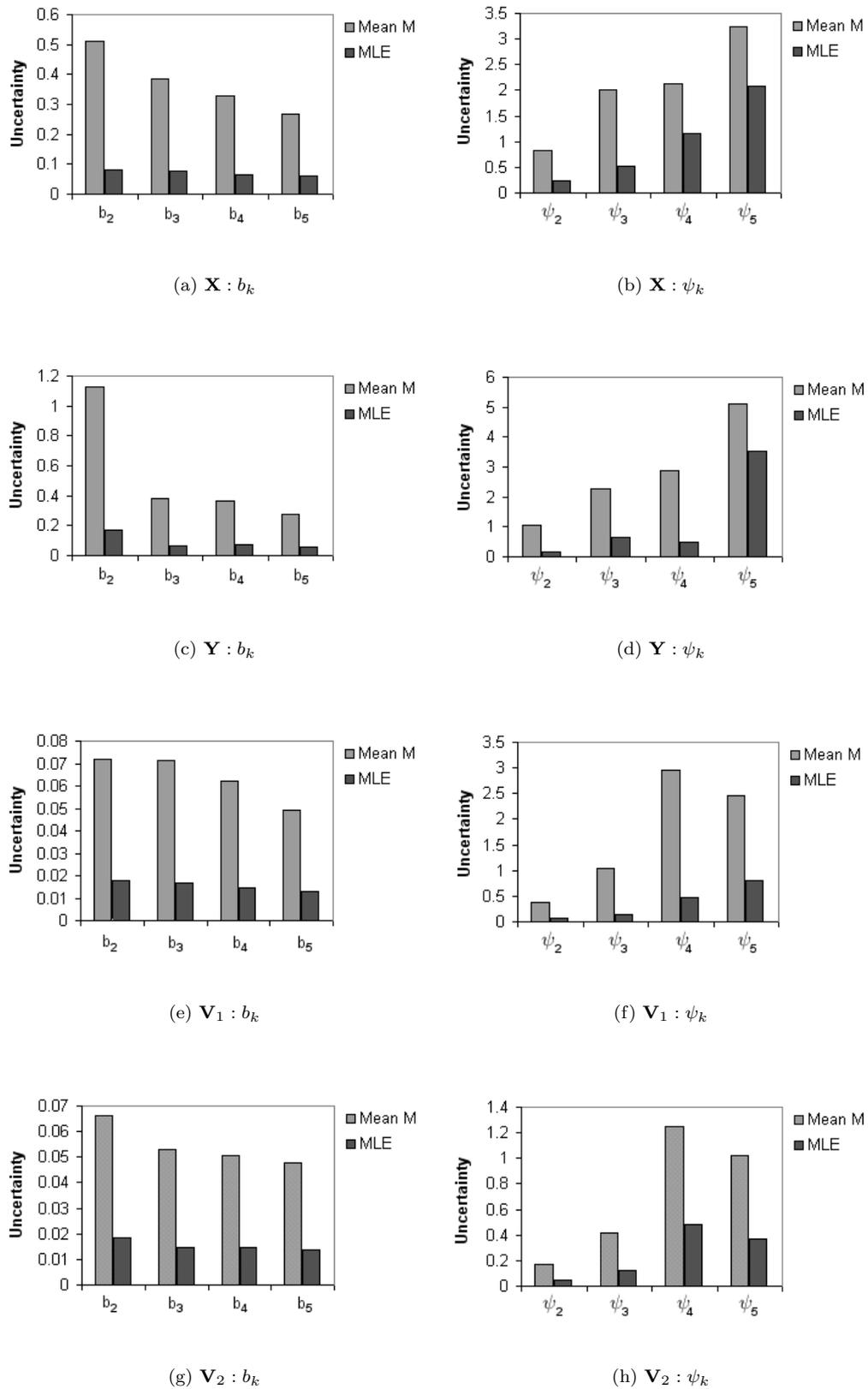


FIGURE 5.24: Estimated subject biometric parameter uncertainties computed at the 3σ confidence limits. Shown here are the mean uncertainties attributed to the independent set of pose projections \mathbf{M} and the uncertainties computed from the global optimization.

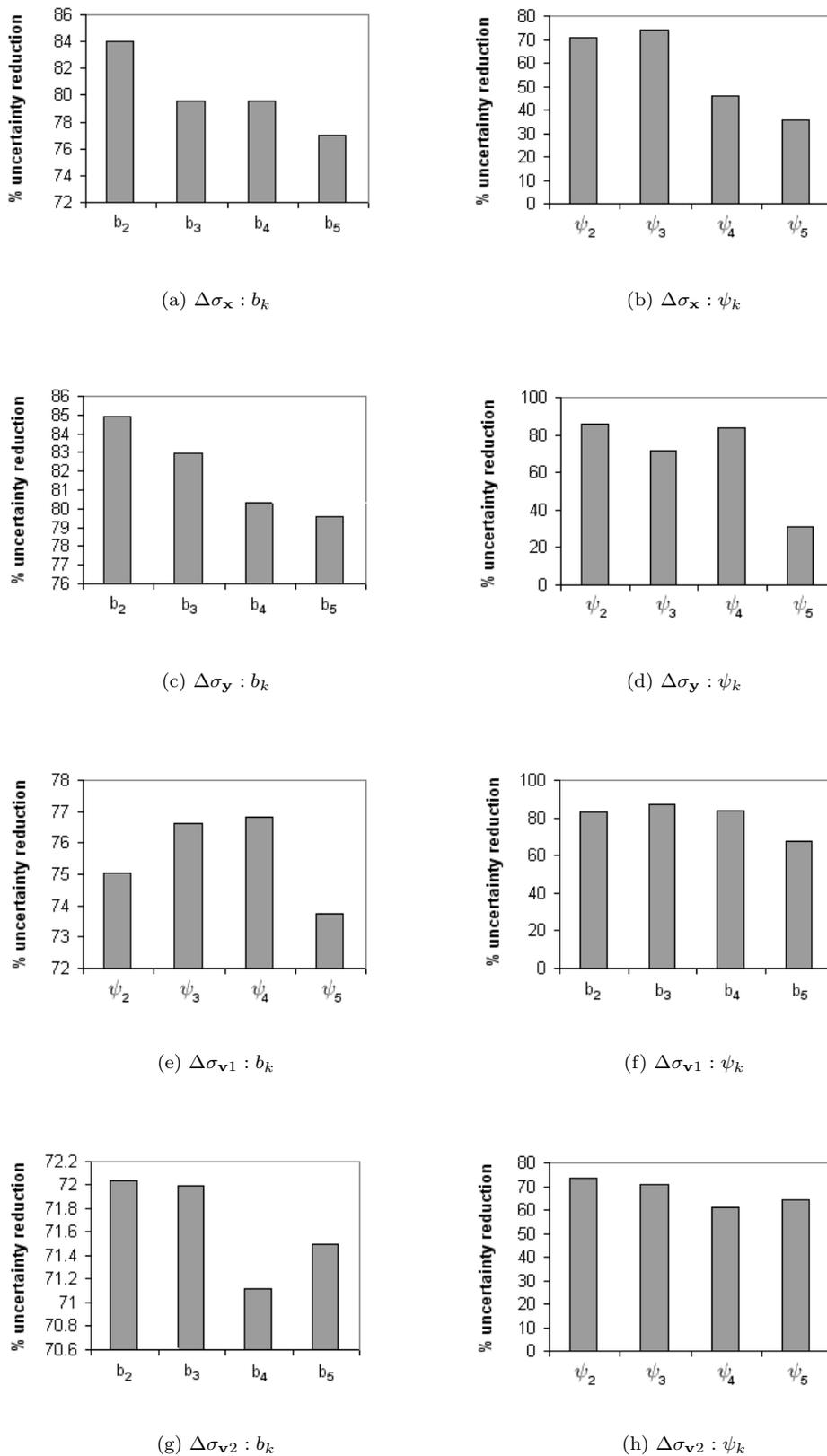


FIGURE 5.25: Estimated biometric parameter uncertainties computed at the 3σ confidence limits. The reduction in uncertainties between the independent set of pose projections \mathbf{M} and the uncertainties computed from the global optimization is quantified by the uncertainty difference as a percentage of the mean individual uncertainties $\Delta\sigma$.

5.11 Conclusions

We have demonstrated that generalized gait motion can be approximated by piecewise linear planar motion. The complete reconstruction algorithm requires no prior knowledge of the camera or worldspace, only that we can identify the limb joint landmarks for each subject. Reconstruction of generalized gait motion is a three step process: linear trajectory reconstruction, parameter fusion and finally global optimization. We have shown that the imaged motion can be decomposed into an accurate worldspace model, where the projection of worldspace structure includes modelling of radial lens distortion. We also gave details of how to model trajectory segments with non linear velocities, though without qualification of results. In the context of biometric gait analysis, this is a positive step towards making already established and future techniques more robust to changes in subject pose.

The set of limb swing plane homography mappings, computed from each reconstructed linear trajectory segment, form linear constraints on the camera intrinsic parameters. A minimum of two non parallel motion trajectories are required to find the intrinsic parameters of a camera, with the assumption of zero skew.

We assume that the camera intrinsic parameters and the ground plane rotation are fixed throughout the image sequence. Consequently, they can be removed from the set of planar homographies, such that the ground plane pose rotation aligns the canonical X axis with the first subject motion trajectory. The remaining set of reconstructed piecewise linear trajectory segments are then each parameterised by a single trajectory angle θ within the canonical ground plane. In the presence of image noise and effects of camera radial lens distortion, the set of reconstructed gait trajectory segments do not exactly satisfy the conditions for planar motion. We have discussed the accuracy of reconstruction through linear decomposition alone, as well as performing maximum likelihood estimation of both camera and subject pose parameters. The worldspace decomposition process accounts for the greatest significant proportion of the reprojection error within the initial fitting process. We model the camera by including components of radial lens distortion. Subsequently, the linear decomposition method provides a good set of initial estimates for the parameters of the camera and ground plane.

Each linear trajectory segment reconstructs the individual representations of the underlying biometric parameters. We described two methods to fuse the reconstructed parameters into a single representation. The linear method assumes each of the biometric coefficients are independent, thus fusion of these parameters can be thought of as no more than a statistical average over the complete set of coefficients. We also discussed a method that computes the maximum likelihood estimate of the fused parameters, by minimizing the image reprojection error. There is a significant improvement in residual fitting error when compared to using the linear method alone.

The parameter fusion step can be thought of as a half way stage between computing the set of reconstructed linear motion segments and the full worldspace model optimization. Its objective is twofold: i) Decompose the set of linear trajectory segment poses into a consistent planar motion parameterisation; and ii) Generate a consolidated set of underlying biometric parameters for each subject.

Maximum likelihood estimation of both sets of parameters are computed by sparse methods that minimize image reprojection error. We first consider the set of reconstructed limb plane homographies as fixed, thus minimization of the biometric parameters on the metric plane is enabled by direct reprojection of structure using these homographies. On the other hand, minimization of the camera and ground plane parameters is achieved by assuming that the set of reconstructed metric plane limb points are fixed, thus optimizing reprojection error through the parameterised worldspace projections.

Computation of the set of fused parameters through linear methods alone can give a poor conversion between piecewise linear segments and the worldspace model. Performing maximum likelihood estimations, on both biometric and system parameters within the fusion stage, generates good initial estimates of the global parameters and widens the region of convergence. The fusion step then removes some of the risk of encountering outlying local minima and potential slow convergence far from the true minimum, within the global parameter space.

The final global optimization step partitions the worldspace model into three sections: *system*, *subject* and *pose* parameters. The *system* parameters encode the camera intrinsic coefficients and ground plane pose rotation, the *subject* parameters encode the underlying dynamics of gait motion and the *pose* parameters determine the corresponding circumstantial values of initial gait phase and trajectory. We gave details of a sparse Levenberg-Marquardt minimization method that gives a true maximum likelihood estimate of the camera intrinsic, ground plane and each subject's underlying biometric parameters over the entire image sequence. The computed camera intrinsic parameters compare reasonably well with those obtained through an accurate calibration algorithm. An analysis of the parameter uncertainties was given, and showed that the global optimization method makes a significant improvement in both the level of uncertainty within the parameters and the root mean squared reprojection errors, when compared to the individual linear trajectory pose reconstructions.

While the algorithm can handle an arbitrary number of subjects, only a single subject has been used for testing. A large number of frames need to be manually marked for each subject, so the restriction mainly depends on the time required to mark up an image sequence. Similarly, the subject has been imaged walking with almost constant velocity around the test track. Further analysis needs to be done with more subjects and with walking at a number of different speeds.

Each subject, and their corresponding set of linear motion poses, influences the con-

straints on the camera intrinsic and ground plane rotation parameters. Predictions about changes in velocity and deviation from the path at turning points can only be validated by reconstructing the true 3D motion. In a slight oversight by the experimenter, no ground truth data was taken with the same landmark positions, that would have allowed us to compare reconstructions. Further work needs to be done to determine the accuracy within which we need to compute the landmark positions, in order to give a good reconstruction.

The gait function is able to reconstruct the subject motion fairly well, but in the presence of noise and with many data segments the reconstructed waveforms appear smoothed due to the nature of the Fourier representation. The detail within high energy phases of gait (*loading response* and *initial swing*) is missed. The higher the uncertainty within the reconstruction, the more like the basis function the gait waveform becomes. Since the first harmonic within the Fourier gait function of equation 5.13 contains no information about the underlying biometric, then it may be beneficial to replace this first order term with a function representing a statistically average gait motion. Instead of trying to measure the reconstructed gait signature itself, we then measure the departure from normal gait. Uncertainty within the reconstruction then shapes the reconstructed waveform more towards the statistically average gait function, thus preserving the detail within the high energy phases of gait.

There need not only be a single average gait motion function. A database of average gait functions can be generated to represent all the various gait modes. This range of modes can be wider than just the two normal walking and running types, but can also encompass the pathological conditions. From a reconstruction point of view, finding the function that best describes the gait mode gives us extra information about the circumstance of subject motion. How to best use the information reconstructed across different gait modes, in order to generate a unique biometric signature, is though a topic for further research.

5.12 Tables of parameter estimates

We present here the computed parameter and corresponding uncertainty estimates after global optimization for the set of subject biometric parameters \mathbf{B} . The maximum likelihood estimate (MLE) is found by optimizing all available parameters, including the camera intrinsic coefficients \mathbf{K}_c and \mathbf{K}_r . For comparative purposes we also show the parameter estimates and uncertainties for each individual linear trajectory pose segment, obtained through pose projection \mathbf{M} . We finally show the set of parameter estimates and uncertainties for each of the pose data blocks \mathbf{Z} and \mathbf{W} after optimization.

<i>Param</i>	a	b	c	d	e1	e2	MLE
d_2	0.79343	0.73767	0.72992	0.78166	0.79959	0.79557	0.76549

TABLE 5.15: Computed parameter estimates for parameters \mathbf{D} . Individual parameter estimates for each linear trajectory pose segment (a-e) and after global MLE are shown.

<i>Param</i>	a	b	c	d	e1	e2	MLE
d_2	0.03361	0.01501	0.02539	0.01814	0.02231	0.02004	0.00671

TABLE 5.16: The corresponding uncertainties of parameter estimates \mathbf{D} computed at the 3σ confidence limits.

<i>Param</i>	a	b	c	d	e1	e2	MLE
b_2	0.64409	0.56664	0.99231	0.88868	0.75239	0.79025	0.46713
ψ_2	-1.40415	-0.82466	-1.80979	-0.49576	-1.94606	-1.78927	-1.67389
b_3	0.0615	0.23226	0.35118	0.31576	0.40475	0.51466	0.19041
ψ_3	0.22714	-3.10169	-1.03506	1.60238	1.64414	2.02166	0.33561
b_4	0.22131	0.16079	0.27717	0.14345	0.35818	0.26729	0.05965
ψ_4	-0.91308	0.58736	1.43422	2.594	-2.79931	-2.63596	-0.43209
b_5	0.06317	0.11661	0.11865	0.20255	0.12165	0.08803	0.02993
ψ_5	-0.42605	2.38018	-2.15466	2.9069	-0.85679	-0.04464	-0.48506

TABLE 5.17: Computed parameter estimates for parameters \mathbf{X} . Individual parameter estimates for each linear trajectory pose segment (a-e) and after global MLE are shown.

<i>Param</i>	a	b	c	d	e1	e2	MLE
b_2	0.27762	0.17842	0.66027	1.06204	0.4764	0.4222	0.082072
ψ_2	0.76599	0.47638	1.0148	1.16425	0.86091	0.73876	0.24272
b_3	0.24768	0.16704	0.46178	0.6805	0.3965	0.36399	0.078802
ψ_3	5.04482	0.86791	1.85822	1.5928	1.50309	1.1851	0.52048
b_4	0.24569	0.16221	0.44212	0.38965	0.39137	0.33195	0.066878
ψ_4	1.78867	1.1707	2.35214	3.85267	1.83366	1.80178	1.15596
b_5	0.22848	0.14463	0.38763	0.23025	0.32785	0.27912	0.061301
ψ_5	3.66608	1.43854	4.02042	3.39437	3.31894	3.59887	2.08607

TABLE 5.18: The corresponding uncertainties of parameter estimates \mathbf{X} computed at the 3σ confidence limits.

<i>Param</i>	a	b	c	d	e1	e2	MLE
b_2	1.78687	2.778	1.375	1.4889	4.35266	3.72119	1.93923
ψ_2	0.1859	0.74311	-0.14958	1.22678	-1.87541	-1.93033	-0.72983
b_3	0.43535	0.74199	0.26348	0.18721	0.32759	0.30959	0.10605
ψ_3	-1.2744	-0.27206	-1.62704	0.24026	-1.9351	-1.94579	-1.45557
b_4	0.22428	0.27056	0.10224	0.09282	0.34549	0.23377	0.18312
ψ_4	-2.71894	-1.46198	3.13696	0.34272	-0.91272	-1.27855	2.20407
b_5	0.06535	0.13641	0.01124	0.05001	0.22536	0.26309	0.01634
ψ_5	1.5245	2.5714	2.57685	-0.11371	-0.46521	-0.35844	-1.41926

TABLE 5.19: Computed parameter estimates for parameters \mathbf{Y} . Individual parameter estimates for each linear trajectory pose segment (a-e) and after global MLE are shown.

<i>Param</i>	a	b	c	d	e1	e2	MLE
b_2	0.82122	0.89186	0.26502	1.09513	2.28349	1.40108	0.16992
ψ_2	0.88524	0.72058	0.32092	2.82343	0.90761	0.65415	0.14631
b_3	0.38193	0.35171	0.14888	0.57633	0.48014	0.34331	0.06469
ψ_3	1.43064	1.09075	0.71374	7.14951	1.8665	1.39638	0.64521
b_4	0.34058	0.2723	0.14133	0.61514	0.47837	0.33049	0.07141
ψ_4	2.11897	1.63676	1.50266	7.86026	2.30031	1.95434	0.47336
b_5	0.3018	0.22569	0.13585	0.27983	0.42509	0.30992	0.05703
ψ_5	4.84785	2.39385	11.9591	6.67012	2.90929	1.9895	3.54183

TABLE 5.20: The corresponding uncertainties of parameter estimates \mathbf{Y} computed at the 3σ confidence limits.

<i>Param</i>	a	b	c	d	e1	e2	MLE
b_2	0.21527	0.2533	0.11636	0.31704	0.30655	0.31474	0.27867
ψ_2	2.61757	2.70672	2.20133	2.34053	2.49635	2.51714	2.78381
b_3	0.08845	0.14989	0.02016	0.14992	0.16716	0.18676	0.12651
ψ_3	-2.83351	-2.51929	-1.32011	3.12348	-2.53854	-2.41671	-2.76678
b_4	0.01681	0.03281	0.01895	0.06527	0.00962	0.02885	0.02981
ψ_4	-0.34205	0.34917	2.96454	2.80206	0.78146	1.06585	2.34835
b_5	0.01089	0.02073	0.01607	0.06036	0.02211	0.03999	0.01615
ψ_5	-2.44617	2.48466	-1.13019	2.51263	2.52848	2.66093	-2.6584

TABLE 5.21: Computed parameter estimates for parameters \mathbf{V}_1 . Individual parameter estimates for each linear trajectory pose segment (a-e) and after global MLE are shown.

<i>Param</i>	a	b	c	d	e1	e2	MLE
b_2	0.09023	0.03952	0.07155	0.12095	0.05404	0.05416	0.01792
ψ_2	0.38331	0.16423	0.64288	0.75943	0.18077	0.16952	0.06501
b_3	0.08304	0.03735	0.06943	0.14042	0.04834	0.04878	0.01665
ψ_3	0.85554	0.26102	3.47868	1.06521	0.31401	0.29133	0.13625
b_4	0.07556	0.03534	0.06726	0.10456	0.04569	0.04446	0.01441
ψ_4	4.65207	1.12954	3.58584	1.95327	4.79756	1.59195	0.47727
b_5	0.06476	0.03183	0.05509	0.06501	0.03983	0.03827	0.01289
ψ_5	5.87016	1.53397	3.48077	1.06594	1.83809	1.00295	0.80295

TABLE 5.22: The corresponding uncertainties of parameter estimates \mathbf{V}_1 computed at the 3σ confidence limits.

<i>Param</i>	a	b	c	d	e1	e2	MLE
b_2	0.60971	0.62663	0.47776	0.59257	0.71095	0.75503	0.61337
ψ_2	1.9263	1.92781	1.91356	2.04861	1.84794	1.90877	1.99381
b_3	0.14521	0.11888	0.13487	0.14109	0.18495	0.20147	0.13791
ψ_3	-3.12982	-3.0294	-2.83897	-2.80517	3.11058	3.02179	-3.05652
b_4	0.03885	0.03033	0.03244	0.04031	0.04768	0.07688	0.03081
ψ_4	-2.60091	-3.00026	3.09014	-2.48211	2.81317	2.91047	-2.63747
b_5	0.05307	0.04865	0.04831	0.03599	0.04469	0.06095	0.03824
ψ_5	-1.89449	-1.73374	-1.51014	-1.46848	-2.06199	-1.89003	-1.41737

TABLE 5.23: Computed parameter estimates for parameters \mathbf{V}_2 . Individual parameter estimates for each linear trajectory pose segment (a-e) and after global MLE are shown.

<i>Param</i>	a	b	c	d	e1	e2	MLE
b_2	0.07412	0.03859	0.09177	0.03824	0.07683	0.07821	0.01854
ψ_2	0.18248	0.08831	0.28071	0.08914	0.18202	0.18523	0.04405
b_3	0.06214	0.03198	0.07475	0.03074	0.05976	0.05841	0.01483
ψ_3	0.48422	0.29208	0.65949	0.25025	0.40975	0.38269	0.12018
b_4	0.05882	0.03072	0.07332	0.03009	0.05631	0.05507	0.01465
ψ_4	1.54863	1.01938	2.25261	0.75578	1.15562	0.73408	0.48229
b_5	0.05784	0.02972	0.06973	0.02905	0.05141	0.04973	0.01365
ψ_5	1.11766	0.64189	1.48979	0.84864	1.1825	0.86968	0.36843

TABLE 5.24: The corresponding uncertainties of parameter estimates \mathbf{V}_2 computed at the 3σ confidence limits.

<i>Param</i>	a	b	c	d	e1	e2
θ	0	-2.954048	1.411748	-0.795410	-2.35076	-2.33378
α	-0.288967	-0.382725	-0.406140	-0.473025	-0.316038	-0.297291
τ	0.365178	0.326499	0.352361	0.221292	0.241182	0.234599
t_x	-9.099784	8.66166	-8.003123	-6.579745	3.39558	1.524352
t_y	0.196966	0.283015	0.037971	0.192821	0.389252	0.324117
t_z	19.448342	24.474308	30.076081	17.719765	15.729377	19.125916

TABLE 5.25: Computed parameter estimates for pose parameters \mathbf{Z} shown after global optimization.

<i>Param</i>	a	b	c	d	e1	e2
θ	0	0.013962	0.017697	0.025589	0.010290	0.009425
α	0.013655	0.015718	0.004088	0.020363	0.002799	0.00213
τ	0.014291	0.018608	0.009985	0.013675	0.007201	0.006409
t_x	0.208517	0.204597	0.255819	0.14359	0.119761	0.114534
t_y	0.092779	0.113465	0.143423	0.084625	0.073544	0.0891393
t_z	0.615131	0.764941	0.83771	0.535921	0.482151	0.594262

TABLE 5.26: The corresponding uncertainties of parameter estimates \mathbf{Z} computed at the 3σ confidence limits.

<i>Param</i>	a	b	c	d	e1	e2
v_x	0.935972	0.928409	0.926257	0.90898	0.928522	0.928393
f_0	4.09485	3.83101	3.886208	3.588989	3.676642	3.692307
p_1	0.084088	0.06132	0.213168	0.058756	0.281066	0.247376
p_2	-0.365631	-0.419974	-0.418252	-0.417963	-0.384351	-0.400053
q_1	0.106221	0.086503	0.093265	0.066537	0.062599	0.055458
q_2	0.045884	0.044089	0.025222	0.045167	0.044797	0.042783
q_3	0.470347	0.461969	0.49871	0.453301	0.426513	0.444543
q_4	0.592971	0.554872	0.470002	0.564661	0.587755	0.573584
r_1	0.032321	-0.395125	0.328061	0.032446	0.284525	0.308541
r_2	-0.528276	0.146931	0.28888	-0.529569	0.258062	0.281097
r_3	0.058979	-0.336322	0.334445	0.076322	0.329275	0.345886
r_4	-0.076076	-0.474677	0.20623	-0.071946	0.184042	0.198977

TABLE 5.27: Computed parameter estimates for pose parameters \mathbf{W} shown after global optimization.

<i>Param</i>	a	b	c	d	e1	e2
v_x	0.000253	0.000102	0.000214	0.000093	0.000163	0.00014
f_0	0.072801	0.060021	0.108952	0.048088	0.084872	0.086925
p_1	0.006187	0.005781	0.004353	0.006279	0.004645	0.00515
p_2	0.003998	0.003388	0.006353	0.003651	0.005305	0.006068
q_1	0.007769	0.006170	0.008491	0.006738	0.005398	0.005754
q_2	0.003564	0.003475	0.001911	0.003694	0.003147	0.003028
q_3	0.01122	0.009292	0.013654	0.009608	0.009537	0.010571
q_4	0.007784	0.006870	0.008464	0.006441	0.009079	0.009113
r_1	0.014429	0.014194	0.01367	0.016122	0.013513	0.014573
r_2	0.012129	0.012153	0.011704	0.012265	0.011363	0.011364
r_3	0.003555	0.002975	0.003144	0.002819	0.002683	0.002706
r_4	0.002555	0.002495	0.002589	0.00259	0.002766	0.002779

TABLE 5.28: The corresponding uncertainties of parameter estimates \mathbf{W} computed at the 3σ confidence limits.

Chapter 6

Conclusions

In conclusion to this thesis, we restate our hypotheses and describe what are believed to be the novel contributions achieved within this work. We then finalize the chapter with some suggestions for further development.

6.1 Restatement of hypotheses

We hypothesize that gait has the following features and properties.

- **Human locomotion can be modelled as a collection of dynamically moving, articulated limb segments.** Each limb is connected to the trunk and is composed of a number of inter-connected bone and joint structures. Each bone segment is rigid and of fixed length. These bone segments are allowed to freely pivot about the corresponding joint positions, although only within a constrained arc of motion.
- **Articulated leg motion is approximately planar.** While in reality the displacement of leg motion is within all three Euclidean directions, almost all the of the perceived motion is contained within a single plane. The variation of motion out of this plane is subtle and negligible in comparison to this major motion plane. Human motion can then be modelled by using a cardboard person assumption. A subject's body and leg segments are represented as a set of repeating spatio-temporal motion patterns within separate planes.
- **Normal gait is bilaterally symmetric with a half phase shift.** Walking uses a repetitious sequence of leg motion to move the body forward. This series of events is repeated by each leg with reciprocal timing. The stance period of one leg equals the swing of the other, thus motion on one leg swing plane is related to the motion of the other by a period of half the gait cycle.

- **Natural gait motion is piecewise linear.** In general, people tend to walk in straight lines with constant velocity. Deviation from this assumption infers inconsistent, non-repetitious leg motion and consequently suggests unnatural gait. Imaged gait can then be split piecewise into natural segments of gait motion.
- **Each individual has a set of possibly unique static features.** The static geometric features of gait that remain invariant over time are based on the fixed length measurements of limb segments. Similarly, there are static motion features that are derived from the representation of the dynamic leg motion function. These features are invariant to the circumstantial changes in subject motion such as stride length, cadence and consequently gait speed.
- **Parameterisation of each linear segment of gait motion can be split into two phases.** i) *Limb stance*, non-linear modelling of the articulated limb motion within the canonical motion plane; ii) *Pose projection*, linear projection of the worldspace subject motion structure into the image.

These two phases of subject motion projection are independent. The *pose projection* step is based on the linear projection of geometry into the image, thus we can employ a stratified approach, based on the geometric constraints of fixed limb segment lengths, to compute the inverse transformation that recovers the canonical motion structure of subject gait.

- **Gait motion commonly occurs within a fixed ground plane and is imaged by a static camera.** A further specialization of the epipolar geometry occurs when the cameras are related by motion within a plane. This is the dual situation to a person walking with unconstrained motion on the ground plane. In this case, the rotation axis is orthogonal to the translation direction. The imaged vanishing line of the ground plane is a fixed line and the vanishing point of the vertical direction a fixed point throughout the image sequence.

6.2 Contributions

Even though gait patterns are repeatable most of the time, changes in walking conditions affect these motion patterns. There are many factors, both physical and psychological that can influence the variations between our motion patterns such as walking speed, cadence, ground surface, load carrying and state of mind. The human skeletal structure is articulated but with fixed length limb segments. The geometric length properties of these limb segments provide one set of static parameters of gait motion that remain constant over the entire image sequence. Articulated leg motion can be modelled with a suitable periodic function, hence motion parameterisation may be robustly determined from all available leg data within the image sequence. A further set of static parameters

of gait motion can be determined from the coefficients of the articulated leg motion function.

- **Modelling of articulated leg motion.** Each articulated leg pose is represented by a Euclidean hip displacement $(x, y)^\top$ position followed by a series of connected rigid length bone segments defined by polar coordinates $(d, \theta)^\top$. The angular motion of each leg segment $\theta(t)$ can be represented by a Fourier series function. Good reconstruction results have been obtained with five Fourier harmonics.
- **Relationship between motion parameterisation and walking speed.** The set of reconstructed leg angle Fourier coefficients vary significantly over the range of walking speeds. Within each reconstruction, the amplitude harmonic coefficients have an exponential relationship, such that the first amplitude is the most significant, and encodes the gross angular variation. The higher amplitude harmonics are the least significant and are more responsive to measurement noise. The relationships between first harmonic amplitude and gait speed, and fundamental frequency and walking speed are both approximately linear.
- **Static measurements of parameterised gait motion.** The dynamics of articulated leg motion can be represented by a modified Fourier series that is pseudo-invariant to changes in stride length and cadence, within the range of customary walking speeds (3 - 6 km/h). We can align the Fourier signals by computing the time shift that zeros the first coefficient of phase. The signals may then be made invariant to scale by normalizing the first amplitude coefficients to unity. We represent the underlying biometric leg function through the set of normalized harmonic coefficients $(b_2, \psi_2, \dots, b_5, \psi_5)^\top$. The circumstantial parameters of gait motion $(f_0, a_0, a_1, t_s)^\top$ then allow us to distort this underlying biometric leg function by applying a series of linear deformations in order to better fit the waveform to the measured limb motion.
 - f_0 - scale waveform within the temporal axis.
 - a_1 - scale waveform within the θ axis.
 - t_s - offset waveform along the temporal axis.
 - a_0 - offset waveform along the θ axis.

The modified Fourier series function encodes the visual motion of the dynamics of articulated leg motion. However, there is no clear relationship between the underlying muscle motion and the resulting representation of leg angle function.

- **Parameterisation of articulated leg motion corresponding to overground walking.** Human motion is modelled by using a cardboard person assumption. A subject's body and limb parts are represented as a set of repeating spatio-temporal motion patterns within separate planes. The canonical representation of leg motion

is approximated by an articulated limb function $g(t)$ on two bilateral swing planes, which are inclined at an angle with the vertical. The dynamics of the gait function on both leg planes are related by a half phase shift, such that one leg undergoes exactly the same motion as the other, only half a gait cycle later. Corresponding left and right leg poses on the metric plane are then determined by evaluation of the biometric limb functions at $g(t)$ and $g(t + T/2)$ respectively, where T is the period of gait.

The canonical representation of gait assumes that a person walks from left to right with constant velocity. Since we are unable to gauge depth from monocular motion sequences, the scale ambiguity of a subject's height is resolved by normalizing the upper leg segment to unit length. We represent the non-linear articulated leg function by the modified Fourier series $g(t : f_0, \mathbf{D}, \mathbf{X}, \mathbf{Y}, \mathbf{V})$, where \mathbf{D} is the vector of normalized leg lengths, \mathbf{X} and \mathbf{Y} are the velocity and Fourier coefficients of the metric plane hip displacement functions, and \mathbf{V} the Fourier coefficients of the upper and lower leg angle functions.

We develop a stratified approach to linear trajectory gait reconstruction that uses the geometric constraints of articulated leg motion in order to recover the fronto-parallel view of gait dynamics. The stratification process for computing this rectification transformation is split into three stages: perspective, affine and similarity transformations of the imaged leg swing plane.

- **Projection of worldspace subject leg plane poses into the image.** Projection of the planarized human motion model into the image is achieved by a parameterised set of homography transformations \mathbf{M} that encode both the individual leg plane homography mappings. Each planar homography consists of a set of rigid motion transformations (scaling, rotation and translation). A subject's leg plane pose is computed by applying a rotation \mathbf{H}_α about the X axis to facilitate the leg swing plane inclination to the vertical, then applying the limb plane selection translation \mathbf{H}_β to map the required hip point to ∓ 1 . This is followed by a scaling \mathbf{H}_τ in the Z direction that generates the correct distance between both hip points for a subject. The worldspace orientation, subject displacement and subsequent projection into the image is then achieved via the linear pin-hole projection transformation $\mathbf{K}[\mathbf{R} | \mathbf{t}]$, where \mathbf{K} is the camera calibration matrix and \mathbf{R}, \mathbf{t} are the camera extrinsic pose matrices. Since the canonical spatio-temporal motion structure of gait is modelled on the metric $z = 0$ plane, projection of articulated leg points into the image is achieved by the homography matrices formed from the first, second and fourth columns of the pose projection transformation.
- **Stratified perspective transformation.** The first step in the stratified reconstruction process is to compute the perspective transformation \mathbf{H}_p that recovers

the affine properties of both leg swing planes. We initially assume that all of a subject's cardboard limb planes are parallel, thus they all share a common vanishing line within the image. Identification of landmark correspondences and subject periodicity is solved simultaneously by computing the self-similarity of structure over the image sequence. The imaged positions of repeated gait poses are related by a conjugate translation of the leg swing plane. We combine constraints from left and right leg planes in order to determine the coefficients of both transformations. We then extract the shared vanishing line of the leg swing planes from these coefficients and subsequently compute the required perspective transformation \mathbf{H}_p that restores the affine properties of the limb planes.

- **Stratified affine transformation.** Metric properties of the leg swing plane are then recovered by identifying the images of the circular points \mathbf{I}' and \mathbf{J}' from the fixed lengths of tracked leg segments throughout the image sequence. Linear constraints on the elements of the affine transformation \mathbf{H}_a , that maps the imaged circular points back to their canonical values, are computed from pairs of corresponding leg segments taken at different frame positions.
- **Stratified similarity transformation.** The scale ambiguity on both leg planes is removed by constraining the upper leg limb segment to be of unit length. The circular points \mathbf{I}, \mathbf{J} remain fixed under any similarity transformation \mathbf{H}_s , thus the remaining translational ambiguity is resolved by enforcing the bilateral symmetry constraint between both planes of recovered gait motion. We then compute a robust estimation of the gait motion function by fitting the articulated leg motion on both limb planes to a single modified Fourier series function.

In reality, people typically walk on a flat ground plane, hence the configuration and parameterisation of subject motion is specialized. Furthermore, people tend to walk in straight lines over a small number of gait cycles, thus an arbitrary length sequence of gait motion can be segmented piecewise into linear sections of gait. Reconstruction of a generalized gait motion is achieved in a three step process: i) Linear trajectory reconstruction; ii) Parameter fusion; and iii) Maximum likelihood estimation of the global motion parameterisation.

- **Parameter fusion.** The parameter fusion step can be thought of as a half way stage between computing the set of reconstructed linear motion segments and the full worldspace model optimization. Its objective is twofold: i) Decompose the set of linear trajectory segment poses into a consistent planar motion parameterisation; and ii) Generate a consolidated set of underlying biometric parameters for each subject.

Maximum likelihood estimation of both sets of parameters are computed by sparse methods that minimize image reprojection error. We first consider the set of

reconstructed leg plane homographies as fixed, thus minimization of the biometric parameters on the metric plane is enabled by direct reprojection of structure using these homographies. On the other hand, minimization of the camera and ground plane parameters is achieved by assuming that the set of reconstructed metric plane leg points are fixed, thus we optimize image reprojection error through the set of parameterised worldspace projections.

Since each subject pose projection is modelled by a set of homographies that map structure from the metric plane to the images of the leg swing planes, we can recover the camera intrinsic parameters by a method similar to Zhang's calibration algorithm [120]. At least two different trajectories of gait motion are required to form the required constraints on the image of the absolute conic ω .

- **Fusion of each piecewise segment of articulated leg motion into a single underlying gait motion function.** Each linear trajectory segment reconstructs the individual representations of the underlying biometric parameters. Fusion of these biometric parameters recovers an initial estimate of the unique properties of a subject's motion dynamics that are consistent over a range of gait modes.

We described two methods to fuse the reconstructed parameters into a single representation. The linear method assumes each of the biometric coefficients are independent, thus fusion of the parameters is achieved by a process of consolidation, via the statistical average over the complete set of coefficients. We also described a method that computes the maximum likelihood estimate of the fused biometric parameters, by minimizing image reprojection error. There is a significant improvement in residual fitting error when compared to using the linear method alone.

- **Fusion of the pose projections of each piecewise segment of gait motion into a consistent planar worldspace motion parameterisation.** Each subject pose is modelled by a linear motion trajectory with legs that swing within bilateral planes. The reconstruction phase described in chapter 4 essentially identifies these bilateral planes and the representation of leg motion within them. A consistent set of camera intrinsic parameters \mathbf{K}_c and \mathbf{K}_r , ground plane normal \mathbf{N} and the parameterised subject poses \mathbf{Z}_i are computed by fixing the reconstructed leg points within the metric plane and optimizing the set of bilateral limb planes.

We have described two methods to fuse the reconstructed parameters into a single representation. The linear method removes the identified camera calibration matrix from each of the leg plane homographies, then proceeds to decompose the elements of the extrinsic worldspace matrices into a product of the consistent ground plane rotation \mathbf{R}_N , and the set of pose trajectory rotations \mathbf{R}_θ within this plane. We also described a method that computes the maximum likelihood estimate of the fused system parameters by minimizing image reprojection error. We

showed that reprojection error is worst for oblique trajectories, with respect to the camera coordinate system, when performing linear decomposition alone.

- **Global optimization of parameters.** The final global optimization step partitions the worldspace model into three sections: *system*, *subject* and *pose* parameters. The *system* parameters encode the camera intrinsic coefficients and ground plane pose rotation, the *subject* parameters encode the underlying dynamics of gait motion and the *pose* parameters determine the corresponding circumstantial values of initial gait phase and trajectory. The form of the Jacobian matrix is sparse, and has a shape which can be exploited when solving for the potentially large number of subsidiary parameters. We gave details of a sparse Levenberg-Marquardt minimization method that gives a true maximum likelihood estimate of the camera intrinsic, ground plane and each subject's underlying biometric parameters over the entire image sequence.

6.3 Further work

- **Periodicity from self-similarity of pixel correspondences.** The task of determining periodicity and point correspondences over the image sequence has been made much simpler by the fact that we have manually marked the landmark interest features. There is a wide range of literature on periodicity detection and motion classification. The self-similarity based periodicity detection method outlined within chapter 4 is most closely related to the work of Cutler and Davis [21, 22]. Their work compares re-scaled image regions corresponding to a tracked subject, in order to determine the periodicity of self-similar pixel structures. Our method extends this work further by enforcing the geometric constraints of repeating planar motions, through identification of the imaged conjugate translations corresponding to subject motion. Our self-similarity method could easily be developed to enable periodicity detection and correspondence matching from pixel regions alone. Identification of the conjugate translation enables us to recover the affine properties of subject motion. To be of any practical interest, further investigation is required to enable segmentation of the required landmark features from the affinely recovered image regions of subject motion.
- **Lateral displacement.** Smoothing of the abrupt changes in the vertical displacement of the body is achieved by a series of limb motions. As a result, the head and body deviate laterally from the progression mid-line, thus a more realistic motion model can be employed to account for this type of displacement. We still assume that the motion dynamics of both legs is planar and inclined at an angle to the vertical. Each planar subject pose within the sequence is then laterally displaced in the Z direction by the simple harmonic lateral displacement function $z(t) = \gamma \sin(2\pi f_0 t + \psi)$. This parameterisation is non-linear and dependent

on the position of the subject within the image sequence, thus can only be modelled within the maximum likelihood estimation procedure. Initial estimates for the motion model parameters are first computed via the stratified reconstruction method.

- **Periodic gait series function.** The dynamics of gait have been captured by using a modified Fourier series function $g(t)$ to approximate the articulated leg motion. The unique features of gait are mostly apparent in the higher order Fourier components, where abrupt changes in the limb's muscles cause rapid changes in the leg function. The Fourier motion representation tends to capture the dynamics of the swing phases better, since the swinging motion from both legs accounts for 80% of the complete gait cycle. The apparent smoothing and subsequent misrepresentation of the gait function during the loading response and pre-swing phases of gait may possibly be avoided by changing the periodic gait reconstruction basis function.

Since the first harmonic within the Fourier gait function contains no information about the underlying biometric then it may be beneficial to replace this first order term with a function representing a statistically average gait motion. Instead of trying to measure the reconstructed gait signature itself, we then measure the departure from normal gait. Uncertainty within the reconstruction then shapes the reconstructed waveform more towards the statistically average gait function, thus preserving the detail within the high energy phases of gait.

- **Reconstruction error analysis for differing viewpoints.** While the work presented in this thesis demonstrates that subject motion can be recovered from many viewpoints, little has been done to analyse the major sources of reconstruction error. Further work needs to be performed to test the robustness of the method in the presence of noise and other sources of imaging error, such as camera radial lens distortion. In order for the method to have any practical application, we must be able to quantify the level of uncertainty within each of the reconstruction parameters. We must then determine what level of landmark measurement error is acceptable, in order for us to compute a reliable set of biometric motion features.
- **Reconstruction error analysis for differing gait speeds.** Further work needs to be done in order to validate the invariance of the reconstructed subject motion to changes in walking speed. The work presented within section 3.6 demonstrated that suitable biometric features could be extracted from the reconstructed motion of subjects walking on a treadmill, over a number of controlled walking speeds. We need to validate these same assumptions in light of the reconstructed subject motion sequences of overground walking. Subsequently, we should also compare the reconstructions corresponding to both treadmill and overground walking, over a suitable range of walking speeds.

- **Global parameterisation analysis for multiple subject motions.** While the gait reconstruction algorithm can handle an arbitrary number of subjects, only a single subject has been used for testing. A large number of frames need to be manually marked for each subject, so the restriction mainly depends on the time required to mark up an image sequence. Similarly, the subject has been imaged walking with almost constant velocity around the test track. Further analysis needs to be done with more subjects and with walking at a number of different speeds.

6.4 Publications associated with this thesis

- N. M. Spencer and J. N. Carter. Viewpoint invariance in automatic gait recognition. In Proc. Third IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'02), pages 1-6, 2002.
- N. M. Spencer and J. N. Carter. Towards pose invariant gait reconstruction. In Proc. International Conference on Image Processing (ICIP'05), Genova, Italy, volume III, pages 261-264, September 2005.

Appendix A

Projective Geometry

A.1 Classes of planar transformations

Geometry is the study of properties invariant under groups of transformations. From this point of view, 2D projective geometry is the study of properties of the projective plane \mathbb{P}^2 that are invariant under a group of transformations known as projectivities. A projectivity is an invertible mapping from points in \mathbb{P}^2 (homogeneous 3 vectors) to points in \mathbb{P}^2 and lines to lines. Projectivities form a group since the inverse of a projectivity is also a projectivity, as is the composition of two projectivities. A projectivity is called a *collineation*, a *projective transform* or a *homography*: the terms are synonymous.

There are several important specializations of the projective transformations. We introduce these specializations starting from the most specialized and progressively generalize up to the most general projective transformation. This defines a hierarchy of transformations which any general projective transform can be broken down into. The distortion effects of each transformation group is described and a number of invariant properties mentioned about each. The more specialized transformations inherit the group of invariants of the more generalized transformations so to conserve space the invariants associated with more general transformation classes will be omitted from the discussions relating to the more specialized classes.

A.1.1 Euclidean Transformation

$$\mathbf{x}' = \mathbf{H}_e \mathbf{x} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} \mathbf{x} \quad (\text{A.1})$$

Where \mathbf{R} is a 2×2 rotation matrix (an orthogonal matrix whose transpose is the same as its inverse: $\mathbf{R}^\top \mathbf{R} = \mathbf{R}^{-1} \mathbf{R} = \mathcal{I}$), and \mathbf{t} is a translation 2-vector. A Euclidean

transformation has three degrees of freedom, 1 for rotation and 2 for translation. The transformation can be computed from two point correspondences. Figure A.1 shows the deformation of shape caused by applying the transformation to a geometric object.

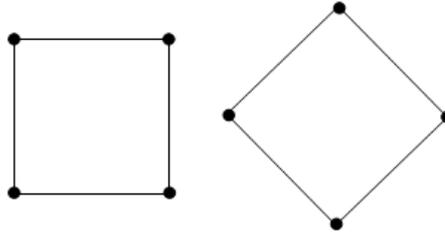


FIGURE A.1: Deformation of the Euclidean transformation is composed of rotation and translation only.

Lengths, areas and angles remain constant when the transformation is applied to any geometric object.

A.1.2 Similarity Transformation

$$\mathbf{x}' = \mathbf{H}_s \mathbf{x} = \begin{bmatrix} s\mathbf{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} \mathbf{x} \quad (\text{A.2})$$

Where \mathbf{R} is a 2×2 rotation matrix, \mathbf{t} is a translation 2-vector and s is a scalar that represents the isotropic scaling. Figure A.2 shows an example of the deformation caused by applying the transformation to a geometric object.

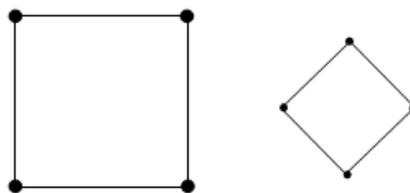


FIGURE A.2: Deformation of the similarity transformation is composed of scaling, rotation and translation only.

Angles, ratio of lengths, ratio of areas, and the circular points remain constant. One extra degree of freedom is added to the previous three of the Euclidean transformation, and as each point correspondence imposes two constraints on the transformation, two correspondences fully define the transformation. The term metric structure which is commonly used in reconstruction contexts implies that the rectification is defined up to a similarity.

Under any similarity transformation there are two points on the line at infinity \mathbf{l}_∞ which are fixed. These are the circular points \mathbf{I} and \mathbf{J} . The circular points get their name from the property that every circle on the image plane intersects \mathbf{l}_∞ at two fixed complex ideal points. Starting from the homogeneous representation of a conic and noting that a circle is a specialized conic with $a = c$ and $b = 0$.

$$ax_1^2 + bx_1x_2 + cx_2^2 + dx_1x_3 + ex_2x_3 + fx_3^2 = 0 \quad (\text{A.3})$$

$$a(x_1^2 + x_2^2) + dx_1x_3 + ex_2x_3 + fx_3^2 = 0 \quad (\text{A.4})$$

This conic intersects \mathbf{l}_∞ in the ideal points for which $x_3 = 0$, so by substitution of x_3 into the homogeneous equation for a circle.

$$x_1^2 + x_2^2 = 0 \quad (\text{A.5})$$

Setting $x_1 = 1$ then $x_2^2 = -1$, i.e. $x_2 = \pm i$, or in vector form.

$$\mathbf{I} = \begin{pmatrix} 1 \\ i \\ 0 \end{pmatrix} \quad \mathbf{J} = \begin{pmatrix} 1 \\ -i \\ 0 \end{pmatrix} \quad (\text{A.6})$$

The same result is obtained by setting $x_2 = 1$ as the vector is homogeneous and represents the same result up to scale. Identifying the circular points in an imaged plane allows the recovery of similarity properties (angles and ratios of lengths). Algebraically the circular points are the orthogonal directions of Euclidean geometry $(1, 0, 0)^\top$ and $(0, 1, 0)^\top$ packaged into a single complex conjugate entity.

$$\mathbf{I} = (1, 0, 0)^\top + i(0, 1, 0)^\top \quad (\text{A.7})$$

Consequently once the circular points are identified orthogonality and other metric properties are then determined. It can be easily shown that these points remain fixed under the similarity transformation.

$$\mathbf{I}' = \mathbf{H}_s \mathbf{I} \quad (\text{A.8})$$

$$= \begin{bmatrix} s \cos \theta & -s \sin \theta & t_x \\ s \sin \theta & s \cos \theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} 1 \\ i \\ 0 \end{pmatrix} \quad (\text{A.9})$$

$$= \begin{pmatrix} s(\cos \theta - i \sin \theta) \\ s(\sin \theta + i \cos \theta) \\ 0 \end{pmatrix} \quad (\text{A.10})$$

Since $e^{i\theta} = \cos \theta - i \sin \theta$ and $ie^{i\theta} = i \cos \theta + \sin \theta$ then

$$\mathbf{I}' = se^{i\theta} \begin{pmatrix} 1 \\ i \\ 0 \end{pmatrix} = \mathbf{I} \quad (\text{A.11})$$

with an analogous proof for \mathbf{J} . The invariance of the circular points to the similarity transformation suggests that once the image of the absolute conic or dual conic is defined metric structure can only be recovered up to the similarity transformation.

A.1.3 Affine Transformation

$$\mathbf{x}' = \mathbf{H}_a \mathbf{x} = \begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} \mathbf{x} \quad (\text{A.12})$$

Where \mathbf{A} is a non-singular 2×2 matrix and \mathbf{t} is a translation 2-vector. The affine matrix has six degrees of freedom: four for the \mathbf{A} matrix elements and a further two for the translation vector. The matrix can therefore be determined from three point correspondences.

The basic properties of affine transforms are:

- Maps straight lines to straight lines
- Maps parallel straight lines to parallel straight lines
- Preserves ratios of lengths along a given straight line

We investigate the properties of affine transformations geometrically by introducing a specialized type of affine transform namely the parallel projection.

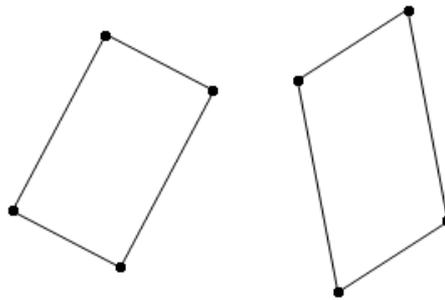


FIGURE A.3: Deformation of the affine transformation preserves parallelism, ratio of areas, ratio of lengths on collinear or parallel lines (midpoints), linear combinations of vectors (centroids) and the line at infinity \mathbf{l}_∞ .

A parallel projection maps straight lines to straight lines. Let \mathbf{l} be a line in the plane π_1 and let \mathbf{H}_a be a parallel projection mapping π_1 onto the plane π_2 . Now consider all the rays associated with \mathbf{H}_a that pass through \mathbf{l} . Since these rays are parallel, they must fill a plane. Call this plane π . The image of \mathbf{l} under \mathbf{H}_a consists of those points where the rays that pass through \mathbf{l} meet π_2 . These points are simply the points of intersection of π with π_2 . Since any two intersecting planes in \mathbb{R}^3 meet in a line, it follows that the image of \mathbf{l} under \mathbf{H}_a is a straight line. Figure A.4 shows geometrically the intersection of the plane π through \mathbf{l} on plane π_1 and the plane π_2 , and demonstrates that straight lines are mapped to straight lines between planes.

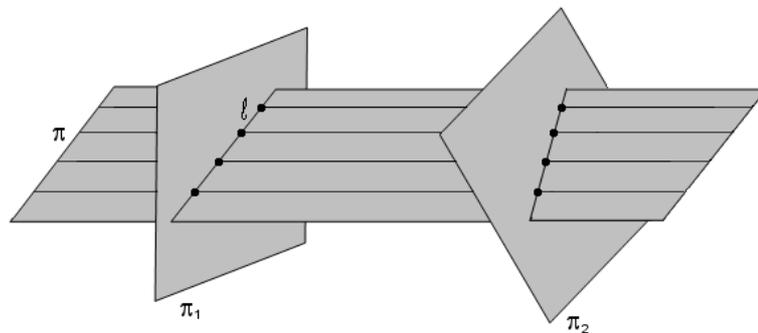


FIGURE A.4: Affine transformation maps straight lines to straight lines.

A parallel projection maps parallel straight lines to parallel straight lines. Let \mathbf{l}_1 and \mathbf{m}_1 be parallel lines in the plane π_1 , and let \mathbf{H}_a be a parallel projection mapping π_1 onto the plane π_2 . Let \mathbf{l}_2 and \mathbf{m}_2 be the lines in π_2 that are the images under \mathbf{H}_a of \mathbf{l}_1 and \mathbf{m}_1 . If \mathbf{l}_2 and \mathbf{m}_2 are not parallel, they must meet at some point, P_2 say. Let P_1 be the point on π_1 which maps to P_2 . Then P_1 must lie on both \mathbf{l}_1 and \mathbf{m}_1 . Since \mathbf{l}_1 and \mathbf{m}_1 are parallel, no such point of intersection can exist, which is a contradiction. It follows that \mathbf{l}_2 and \mathbf{m}_2 must indeed be parallel.

A parallel projection preserves ratios of lengths along a given straight line. Let A, B, C be three points on a line in the plane π_1 , and let \mathbf{H}_a be a parallel projection mapping π_1

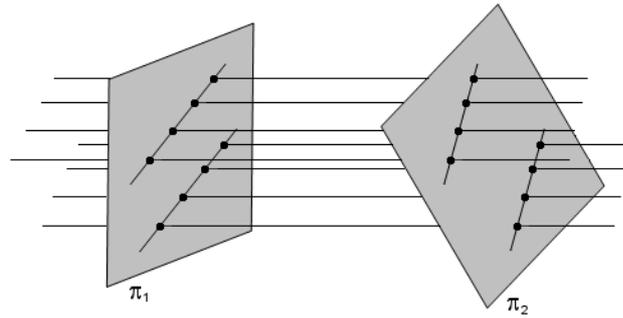


FIGURE A.5: Affine transformation maps straight parallel lines to straight parallel lines.

onto the plane π_2 . Let P, Q, R be the points in π_2 that are the images under \mathbf{H}_a of A, B, C . We know from property 1 that P, Q, R lie on a line; we have to show that the ratio $AB : AC$ is equal to the ratio $PQ : PR$. If the planes π_1 and π_2 are parallel then the parallel projection is an isometry, and so the ratios $AB : AC$ and $PQ : PR$ are equal, as required. On the other hand, if π_1 and π_2 are not parallel then we can construct a plane π'_1 through the point P which is parallel to π_1 . This plane intersects the ray through B and Q at some point B' , and the ray through C and R at some point C' . In this case the ratios $AB : AC$ and $PB' : PC'$ are equal. Now consider the triangle $PC'R$. The lines $B'Q$ and $C'R$ are parallel, since they are rays from the parallel projection. Hence $B'Q$ meets the sides PR and PC' in equal ratios. Thus $PQ : PR = PB' : PC'$. It follows that $PQ : PR = AB : AC$ as required.

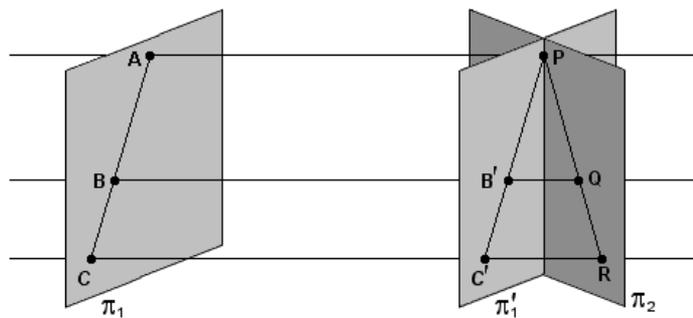


FIGURE A.6: Affine transformation preserves ratios of lengths along a given straight line.

In particular, if a point is the midpoint of a line segment then under parallel projection the image of the point is the midpoint of the image of the line segment.

Under the affine transformation the line at infinity \mathbf{l}_∞ is mapped onto itself. Using the transformation rule for lines, it is easily verified that the line at infinity remains invariant.

$$\mathbf{l}'_{\infty} = \mathbf{H}_{\mathbf{a}}^{-\top} \mathbf{l}_{\infty} = \begin{bmatrix} \mathbf{A}^{-\top} & \mathbf{0} \\ -\mathbf{t}^{\top} \mathbf{A}^{-\top} & 1 \end{bmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \mathbf{l}_{\infty} \quad (\text{A.13})$$

However \mathbf{l}_{∞} is not fixed pointwise under an affine transformation.

$$\begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{0}^{\top} & 1 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} = \begin{pmatrix} \mathbf{A} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ 0 \end{pmatrix} \quad (\text{A.14})$$

The ideal points are not fixed pointwise on the line at infinity \mathbf{l}_{∞} by the affine transformation unless $\mathbf{A}(x_1, x_2)^{\top} = \lambda(x_1, x_2)^{\top}$.

A.1.4 Perspective Transformation

$$\mathbf{x}' = \mathbf{H}_{\mathbf{p}} \mathbf{x} = \begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{v}^{\top} & v \end{bmatrix} \mathbf{x} \quad (\text{A.15})$$

Where \mathbf{A} is a non-singular 2×2 matrix, \mathbf{t} is a translation 2-vector and \mathbf{v} is a 2-vector of perspective coefficients. The matrix has nine elements with only their ratio significant, so the transformation is specified by eight parameters. Note also that it is not always possible to scale the matrix so that the parameter v is unity since it may be zero. The transformation may be computed from four point correspondences, with no three collinear on either plane.

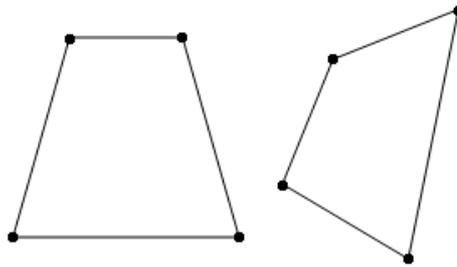


FIGURE A.7: Deformation of the perspective transformation preserves none of the previous properties, though the cross ratio (ratio of ratio of lengths) remains invariant.

Ideal points defined by the intersection of parallel scene lines depend only on the line direction. This was shown earlier through example and can be algebraically expressed by describing lines as a series of points with initial position vector and parameterised direction $\mathbf{l} = \mathbf{x}(\lambda) = U + \lambda D$.

$$\mathbf{x}(\lambda) = \begin{pmatrix} U_x + \lambda D_x \\ U_y + \lambda D_y \\ 1 \end{pmatrix} \quad (\text{A.16})$$

$$= \begin{pmatrix} \frac{U_x}{\lambda} + D_x \\ \frac{U_y}{\lambda} + D_y \\ \frac{1}{\lambda} \end{pmatrix} \quad (\text{A.17})$$

As $\lambda \rightarrow \infty$ the parameterised point on the line \mathbf{l} becomes ideal and is a function of direction only.

$$\mathbf{x}(\infty) = \begin{pmatrix} D_x \\ D_y \\ 0 \end{pmatrix} \quad (\text{A.18})$$

Ideal points when transformed by the affine matrix remain ideal, however ideal points transformed under the perspective transformation become finite.

$$\begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{v}^\top & v \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} = \begin{pmatrix} \mathbf{A} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ v_1 x_1 + v_2 x_2 \end{pmatrix} \quad (\text{A.19})$$

From this result we see that parallelism of lines is not preserved by the perspective transformation.

A.1.5 Cross Ratio

The ratio of distances is not preserved under a perspective transformation, however, the ratio of ratios of distances is invariant. The cross ratio of four points on a line is preserved under perspective transformations. There are many results in projective geometry which result in an interpretation in terms of the cross ratio. It seems likely that all invariant properties of a geometric configuration can ultimately be interpreted in terms of some number of cross ratio constructions.

The cross ratio is defined with respect to figure A.8 and is given by

$$\text{Cr} \{\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4\} = \frac{(X^3 - X^1)(X^4 - X^2)}{(X^3 - X^2)(X^4 - X^1)} \quad (\text{A.20})$$

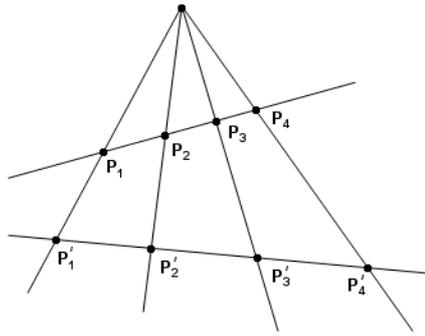


FIGURE A.8: The cross ratio for all lines cutting the pencil is the same i.e. $\text{Cr}\{\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4\} = \text{Cr}\{\mathbf{P}'_1, \mathbf{P}'_2, \mathbf{P}'_3, \mathbf{P}'_4\}$. This configuration corresponds to perspective projection onto a line.

where X^1, X^2, X^3, X^4 represent the corresponding positions of each point along the line, e.g. $(X^3 - X^1)$ is the distance between points P_3 and P_1 .

The pairs of points $\mathbf{P}_1, \mathbf{P}_2$ and $\mathbf{P}_3, \mathbf{P}_4$ are called *harmonic* if

$$\text{Cr}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4) = -1 \quad (\text{A.21})$$

The harmonic relation is associated with the orthogonality of directions in higher dimensional spaces. Note that it is permissible, with the cross ratio defined as above, to write the projective parameter of point $(1, 0)^\top$ as ∞ , and to use this in cross ratio computations.

Since points and lines are dual, there exists an equivalent cross ratio for lines. The dual relation to collinearity is incidence at a point. The cross ratio of the pencil can be defined in terms of the angles between the lines.

Other permutations of the points in the definition of the cross ratio will also lead to a scalar invariant. The four points can be permuted $4!$ different ways. There are only six distinct values of the cross ratio within the 24 permutations. If the cross ratio for the standard definition is defined as τ then the six distinct values are related by the set.

$$\left\{ \tau, \frac{1}{\tau}, 1 - \tau, \frac{1}{1 - \tau}, \frac{\tau - 1}{\tau}, \frac{\tau}{\tau - 1} \right\} \quad (\text{A.22})$$

There exists a rational function of the cross ratio value which is independent of the effects of permutation, namely the *j-invariant* which is defined by:

$$j(\tau) = \frac{(\tau^2 - \tau + 1)^3}{\tau^2 (\tau - 1)^2} \quad (\text{A.23})$$

Appendix B

Matrices and Linear Systems

B.1 Orthogonal regression

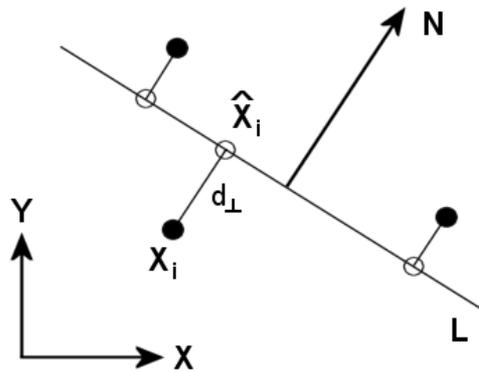


FIGURE B.1: Orthogonal regression of a set of points minimizes the sum of perpendicular distances between point distribution and fitted line.

We seek a line \mathbf{L} which minimizes the sum of squared distances between the measured \mathbf{X}_i , and estimated $\hat{\mathbf{X}}_i$ points. The perpendicular distance of a point $\mathbf{X}_i = (x_i, y_i, 1)^\top$ from the line $\mathbf{L} = (a, b, c)^\top$ can be expressed by the function:

$$d_{\perp}(\mathbf{X}_i, \mathbf{L}) = \frac{ax_i + by_i + c}{\sqrt{a^2 + b^2}} \quad (\text{B.1})$$

The corresponding sum of squares cost function for the complete point distribution is given by:

$$\mathcal{C} = \sum_{i=1}^n d_{\perp}(\mathbf{X}_i, \mathbf{L})^2 \quad (\text{B.2})$$

$$\mathcal{C} = \sum_{i=1}^n (a'x_i + b'y_i + c')^2 \quad (\text{B.3})$$

Writing the line normal as $\mathbf{N} = (a', b')^{\top}$ where $\|\mathbf{N}\| = 1$ and \mathbf{x}_i as the inhomogeneous 2-vector representation of \mathbf{X}_i then

$$\mathcal{C} = \sum_{i=1}^n (\mathbf{N}^{\top} \mathbf{x}_i + c')^2 \quad (\text{B.4})$$

We then proceed to minimize the cost function \mathcal{C} over the parameter c' by first differentiating, then finding the minimum.

$$\frac{\partial \mathcal{C}}{\partial c'} = 2 \sum_{i=1}^n (\mathbf{N}^{\top} \mathbf{x}_i + c') = 0 \quad (\text{B.5})$$

$$\frac{\partial \mathcal{C}}{\partial c'} = \sum_{i=1}^n \mathbf{N}^{\top} \mathbf{x}_i + \sum_{i=1}^n c' = 0 \quad (\text{B.6})$$

$$c' = -\frac{1}{n} \sum_{i=1}^n \mathbf{N}^{\top} \mathbf{x}_i = -\mathbf{N}^{\top} \bar{\mathbf{x}} \quad (\text{B.7})$$

Here $\bar{\mathbf{x}}$ represents the centroid of the point set. We are then able to substitute back into the cost function of equation B.4 the minimized expression of c' to obtain a least squares expression in terms of the line normal and point set.

$$\mathcal{C} = \sum_{i=1}^n \left(\mathbf{N}^\top \mathbf{x}_i - \mathbf{N}^\top \bar{\mathbf{x}} \right)^2 \quad (\text{B.8})$$

$$= \sum_{i=1}^n \left(\mathbf{N}^\top \Delta \mathbf{x}_i \right)^2 \quad \text{where} \quad \Delta \mathbf{x}_i = \mathbf{x}_i - \bar{\mathbf{x}} \quad (\text{B.9})$$

$$= \sum_{i=1}^n (\mathbf{N}^\top \Delta \mathbf{x}_i) (\Delta \mathbf{x}_i^\top \mathbf{N}) \quad (\text{B.10})$$

$$= \mathbf{N}^\top \left(\sum_{i=1}^n \Delta \mathbf{x}_i \cdot \Delta \mathbf{x}_i^\top \right) \mathbf{N} \quad (\text{B.11})$$

$$= \mathbf{N}^\top \mathbf{M} \mathbf{N} \quad (\text{B.12})$$

We can then make a symmetric Eigen-decomposition of the moment matrix \mathbf{M} via the substitution $\mathbf{M}\mathbf{N} = \lambda\mathbf{N}$, and note that $\mathcal{C} = \mathbf{N}^\top \lambda \mathbf{N} = \lambda$ gives the cost error of the point distribution. The minimum of the least squares cost function occurs with least Eigenvalue λ , thus the Eigenvector corresponding to the smallest Eigenvalue is the solution we require for the line normal vector \mathbf{N} , and substitution of the minimized \mathbf{N} into equation B.7 gives the required value for parameter c' of the line \mathbf{L} . The moment matrix \mathbf{M} is symmetric, positive-definite hence all Eigenvalues are real and non-negative.

B.1.1 Constrained orthogonal regression

The minimization may also be constrained so that the line \mathbf{L} passes through a point $\tilde{\mathbf{X}}$, which may be ideal. The problem can then be stated as: minimize the fitting cost error $\|\mathbf{A}\mathbf{L}\|$ subject to $\|\mathbf{L}\| = 1$ and $\tilde{\mathbf{X}}^\top \mathbf{L} = 0$, where design matrix \mathbf{A} is formed by stacking each of the point constraints of the form $(u_i, v_i, 1) \cdot \mathbf{L} = 0$.

We first define the orthogonal complement of the constraint equation $\tilde{\mathbf{X}}^\top \mathbf{L} = 0$ by computing a rotation matrix \mathbf{R} that maps the constraint vector $\tilde{\mathbf{X}} = (\tilde{u}, \tilde{v}, \tilde{w})^\top$ onto the Z axis $(0, 0, 1)^\top$. The homogeneous vector that represents the constraint point is then equivalent to $\tilde{\mathbf{X}} \equiv \mathbf{R}^\top (0, 0, 1)^\top$ and consequently the two other vectors orthogonal to $\tilde{\mathbf{X}}$ are given by the transformations of both X and Y axes.

$$\mathbf{C}_\perp = \mathbf{R}^\top \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \quad (\text{B.13})$$

The constraint vector $\tilde{\mathbf{X}}$ and the complement matrix \mathbf{C}_\perp are orthogonal such that $\tilde{\mathbf{X}}^\top \mathbf{C}_\perp = \mathbf{0}$. Any 2-vector \mathbf{v} then satisfies the equation:

$$(\tilde{\mathbf{X}}^\top \mathbf{C}_\perp) \cdot \mathbf{v} = 0 \quad (\text{B.14})$$

Subsequently, the fitted line consistent with the constraint equation $\tilde{\mathbf{X}}^\top \mathbf{L} = 0$ is parameterised by the mapping $\mathbf{L} = \mathbf{C}_\perp \cdot \mathbf{v}$, for suitable \mathbf{v} . Since \mathbf{C}_\perp has orthogonal columns then $\|\mathbf{L}\| = \|\mathbf{C}_\perp \mathbf{v}\| = \|\mathbf{v}\|$. The problem then reduces to minimize $\|(\mathbf{A}\mathbf{C}_\perp) \cdot \mathbf{v}\|$ subject to $\|\mathbf{v}\| = 1$. We then compute the moment matrix $\mathbf{M} = (\mathbf{A}\mathbf{C}_\perp)^\top (\mathbf{A}\mathbf{C}_\perp)$ i.e. $\mathbf{M} = (\mathbf{C}_\perp)^\top (\mathbf{A}^\top \mathbf{A}) \mathbf{C}_\perp$. Writing each imaged constraint point as $\mathbf{X}_i = (u_i, v_i, 1)^\top$ then the symmetric matrix $\mathbf{A}^\top \mathbf{A}$ is defined as:

$$\mathbf{A}^\top \mathbf{A} = \begin{pmatrix} \sum_i u_i^2 & \sum_i u_i \cdot v_i & \sum_i u_i \\ \sum_i u_i \cdot v_i & \sum_i v_i^2 & \sum_i v_i \\ \sum_i u_i & \sum_i v_i & n \end{pmatrix} \quad (\text{B.15})$$

The Eigenvector with smallest Eigenvalue of the 2 by 2 symmetric matrix \mathbf{M} is the solution we require for the parameter vector \mathbf{v} . The fitted line consistent with the constraint point $\tilde{\mathbf{X}}^\top$ is then found by applying the mapping transform $\mathbf{L} = \mathbf{C}_\perp \cdot \mathbf{v}$.

B.1.2 Orthogonal projection of points onto a line

We seek to find the point $\hat{\mathbf{X}}$ that lies on the line $\mathbf{L} = (a, b, c)^\top$, closest to an arbitrary point $\mathbf{X} = (x, y, 1)^\top$.

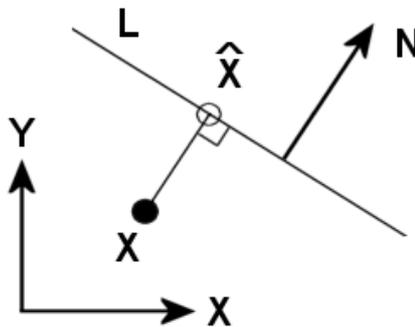


FIGURE B.2: Orthogonal projection of a point onto a line.

We first normalize the line \mathbf{L} such that its normal vector $\mathbf{N} = (a, b)^\top$ has unit norm $\|\mathbf{N}\| = 1$. Any orthogonal line to \mathbf{L} has the form $\mathbf{L}_\perp = (b, -a, c)^\top$. The orthogonal line that passes through the point \mathbf{X} must satisfy the condition $\mathbf{X}^\top \mathbf{L}_\perp = 0$, thus can be written.

$$\mathbf{L}_\perp = \begin{pmatrix} b \\ -a \\ -bx + ay \end{pmatrix} \quad (\text{B.16})$$

The intersection of both lines \mathbf{L} and \mathbf{L}_\perp then generates the required point $\hat{\mathbf{X}} = \mathbf{L}_\perp \times \mathbf{L}$.

$$\hat{\mathbf{X}} = \begin{pmatrix} b^2x - aby - ac \\ a^2y - abx - bc \\ 1 \end{pmatrix} \quad (\text{B.17})$$

B.2 Singular value decomposition

Given a matrix \mathbf{A} , the SVD is a factorization such that $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$, where \mathbf{U} is a matrix with orthogonal columns, \mathbf{V} a square orthogonal matrix, and \mathbf{D} is a diagonal matrix with non-negative entries, which are known as the singular values. The decomposition may be carried out in such a way that the diagonal entries of \mathbf{D} are in descending order, and we assume that this is always done, thus the column of \mathbf{V} corresponding to the smallest singular value is the last column.

The SVD exists for non-square matrices. Of most interest is the case where \mathbf{A} has more rows than columns i.e. \mathbf{A} is an $m \times n$ matrix where $m \geq n$. In this case \mathbf{A} may be factorized as $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$ where \mathbf{U} is an $m \times n$ matrix with orthogonal columns, \mathbf{D} is an $n \times n$ diagonal matrix and \mathbf{V} is an $n \times n$ orthogonal matrix. The fact that \mathbf{U} has orthogonal columns means that $\mathbf{U}^\top\mathbf{U} = \mathcal{I}_{n \times n}$. Furthermore \mathbf{U} has the norm-preserving property that $\|\mathbf{U}\mathbf{x}\| = \|\mathbf{x}\|$ for any vector \mathbf{x} .

A description of the singular value decomposition algorithm is not given here. A description of how it works is given in [38] and a practical implementation is given in [87].

B.2.1 Pseudo-inverse

Given a diagonal matrix \mathbf{D} we define its pseudo-inverse to be the diagonal matrix \mathbf{D}^+ such that

$$\mathbf{D}_{ii}^+ = \begin{cases} 0 & \text{if } \mathbf{D}_{ii} = 0 \\ \mathbf{D}_{ii}^{-1} & \text{otherwise} \end{cases} \quad (\text{B.18})$$

Given an $m \times n$ matrix \mathbf{A} with $m \geq n$, let the SVD of \mathbf{A} be $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$. The pseudo-inverse of matrix \mathbf{A} is then defined by the matrix.

$$\mathbf{A}^+ = \mathbf{V}\mathbf{D}^+\mathbf{U}^\top \quad (\text{B.19})$$

B.2.3 Least squares solution of homogeneous equations

If we consider the over determined system of equations of the form $\mathbf{Ax} = 0$. The obvious solution $\mathbf{x} = 0$ is not of interest, so we seek a non-zero solution to the set of equations. We note that if \mathbf{x} is the solution to this set of equations, then so is $k\mathbf{x}$ for any scalar k . A reasonable constraint would be to seek a solution for which $\|\mathbf{x}\| = 1$. Our problem statement then becomes: find the \mathbf{x} that minimizes $\|\mathbf{Ax}\|$ subject to $\|\mathbf{x}\| = 1$.

If we let the SVD of \mathbf{A} be $\mathbf{A} = \mathbf{UDV}^\top$, then we must minimize $\|\mathbf{UDV}^\top \mathbf{x}\|$. However $\|\mathbf{UDV}^\top \mathbf{x}\| = \|\mathbf{DV}^\top \mathbf{x}\|$ since \mathbf{U} has the norm preserving property $\|\mathbf{U}\tilde{\mathbf{x}}\| = \|\tilde{\mathbf{x}}\|$. Similarly $\|\mathbf{V}^\top \mathbf{x}\| = \|\mathbf{x}\|$ since \mathbf{V} is an orthogonal matrix, thus we need to minimize $\|\mathbf{DV}^\top \mathbf{x}\|$ subject to the condition $\|\mathbf{V}^\top \mathbf{x}\| = 1$. The problem can then be restated by using the substitution $\mathbf{y} = \mathbf{V}^\top \mathbf{x}$, hence we are required to minimize $\|\mathbf{Dy}\|$ subject to $\|\mathbf{y}\| = 1$.

Since the elements of \mathbf{D} are ordered in descending order the minimized solution of \mathbf{y} under the constraint $\|\mathbf{y}\| = 1$ is $\mathbf{y} = (0, 0, \dots, 1)^\top$, where \mathbf{y} has only one non zero entry, 1 in the last position. Finally the required solution $\mathbf{x} = \mathbf{Vy}$ is simply the last column of \mathbf{V} .

B.2.4 Approximating a 3×3 matrix by a rotation matrix

The problem considered here is to solve the best rotation matrix \mathbf{R} to approximate a given 3 matrix \mathbf{Q} such that the Frobenius norm of the difference between matrices $\mathbf{R} - \mathbf{Q}$ is minimized.

$$\min_{\mathbf{R}} \|\mathbf{R} - \mathbf{Q}\|_F^2 \quad \text{subject to} \quad \mathbf{R}^\top \mathbf{R} = \mathcal{I} \quad (\text{B.21})$$

If the trace of a square matrix is defined as the sum of all diagonal elements of the matrix.

$$\text{trace}(\mathbf{A}) = \sum_{i=1}^n a_{ii} \quad (\text{B.22})$$

The trace function has the properties:

$$\text{trace}(\mathbf{A}) = \text{trace}(\mathbf{A}^\top) \quad (\text{B.23})$$

$$\text{trace}(\mathbf{A} + \mathbf{B}) = \text{trace}(\mathbf{A}) + \text{trace}(\mathbf{B}) \quad (\text{B.24})$$

$$\text{trace}(\alpha \mathbf{A}) = \alpha \cdot \text{trace}(\mathbf{A}) \quad (\text{B.25})$$

The Frobenius matrix norm can be written as:

$$\|\mathbf{R} - \mathbf{Q}\|_F^2 = \text{trace}((\mathbf{R} - \mathbf{Q})^\top (\mathbf{R} - \mathbf{Q})) \quad (\text{B.26})$$

$$= \text{trace}(\mathbf{R}^\top \mathbf{R} - \mathbf{R}^\top \mathbf{Q} - \mathbf{Q}^\top \mathbf{R} + \mathbf{Q}^\top \mathbf{Q}) \quad (\text{B.27})$$

$$= 3 - 2 \cdot \text{trace}(\mathbf{R}^\top \mathbf{Q}) + \text{trace}(\mathbf{Q}^\top \mathbf{Q}) \quad (\text{B.28})$$

The problem of minimizing equation B.21 is equivalent to that of maximizing $\text{trace}(\mathbf{R}^\top \mathbf{Q})$. Let the Singular Value Decomposition of \mathbf{Q} be $\mathbf{Q} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$, where $\mathbf{D} = \text{diag}(d_1, d_2, d_3)$.

$$\text{trace}(\mathbf{R}^\top \mathbf{Q}) = \text{trace}(\mathbf{R}^\top \mathbf{U}\mathbf{D}\mathbf{V}^\top) \quad (\text{B.29})$$

$$= \text{trace}(\mathbf{V}^\top \mathbf{R}^\top \mathbf{U}\mathbf{D}) \quad (\text{B.30})$$

Now define an orthogonal matrix $\mathbf{Z} = \mathbf{V}^\top \mathbf{R}^\top \mathbf{U}$ then:

$$\text{trace}(\mathbf{R}^\top \mathbf{Q}) = \text{trace}(\mathbf{Z}\mathbf{D}) = \sum_{i=1}^3 z_{ii} \cdot d_i \quad (\text{B.31})$$

Since \mathbf{Z} is an orthogonal matrix then:

$$\sum_{i=1}^3 z_{ii} \cdot d_i \leq \sum_{i=1}^3 d_i \quad (\text{B.32})$$

Clearly the maximum is achieved when $\mathbf{Z} = \mathcal{I}$ hence matrix \mathbf{R} is best approximated by the transform.

$$\mathcal{I} = \mathbf{V}^\top \mathbf{R}^\top \mathbf{U} \quad (\text{B.33})$$

$$\mathbf{R} = \mathbf{U}\mathbf{V}^\top \quad (\text{B.34})$$

B.3 Parameterising a rotation matrix

A rotation matrix \mathbf{R} is a 3×3 matrix that is orthogonal $\mathbf{R}\mathbf{R}^\top = \mathbf{R}^\top \mathbf{R} = \mathcal{I}$ with unit determinant $\det(\mathbf{R}) = 1$. The inverse of a rotation matrix is the same as its transpose $\mathbf{R}^{-1} = \mathbf{R}^\top$.

Another way of representing a 3D rotation is to specify an axis of rotation of unit length \mathbf{r} and an angle of rotation about the axis α in radians. These two quantities can be jointly specified by a single 3-vector $\mathbf{v} = \alpha \cdot \mathbf{r}$. The angle of rotation is given by the magnitude of the vector $\alpha = \|\mathbf{v}\|$ and the axis of rotation is given by normalizing the vector to unit length $\mathbf{r} = \mathbf{v} / \|\mathbf{v}\|$.

B.3.1 Rodrigues Formula

If we define the matrix representation of the cross product $\mathbf{r} \times \mathbf{p} = [\mathbf{r}]_{\times} \mathbf{p}$, where $[\mathbf{r}]_{\times}$ is a skew-symmetric 3 by 3 matrix i.e. $([\mathbf{r}]_{\times})^{\top} = -[\mathbf{r}]_{\times}$.

$$[\mathbf{r}]_{\times} = \begin{pmatrix} 0 & -r_z & r_y \\ r_z & 0 & -r_x \\ -r_y & r_x & 0 \end{pmatrix} \quad (\text{B.35})$$

The conversion from vector to matrix representation is given by the Rodrigues formula.

$$\mathbf{R} = \cos \alpha \cdot \mathcal{I} + (1 - \cos \alpha) \mathbf{r} \cdot \mathbf{r}^{\top} + \sin \alpha \cdot [\mathbf{r}]_{\times} \quad (\text{B.36})$$

We first note that $\cos \alpha \cdot \mathcal{I} + (1 - \cos \alpha) \mathbf{r} \cdot \mathbf{r}^{\top}$ is symmetric, and $\sin \alpha \cdot [\mathbf{r}]_{\times}$ is skew-symmetric. To convert from the matrix to vector representation, the transpose of equation B.36 is then given by

$$\mathbf{R}^{\top} = \cos \alpha \cdot \mathcal{I} + (1 - \cos \alpha) \mathbf{r} \cdot \mathbf{r}^{\top} - \sin \alpha \cdot [\mathbf{r}]_{\times} \quad (\text{B.37})$$

Any matrix \mathbf{A} can be decomposed into the sum of symmetric $(\mathbf{A} + \mathbf{A}^{\top})/2$ and skew-symmetric $(\mathbf{A} - \mathbf{A}^{\top})/2$ parts. The skew-symmetric part of \mathbf{R} is then given by the equation.

$$(\mathbf{R} - \mathbf{R}^{\top})/2 = \sin \alpha \cdot [\mathbf{r}]_{\times} = [\sin \alpha \cdot \mathbf{r}]_{\times} \quad (\text{B.38})$$

The value of $\sin \alpha$ is given by the length of the vector $\sin \alpha \cdot \mathbf{r}$, from which we can recover the rotation angle α and subsequently compute the rotation axis vector \mathbf{r} and the Rodrigues rotation vector $\mathbf{v} = \alpha \cdot \mathbf{r}$.

Appendix C

Non-Linear Minimization

C.1 Newtonian iteration

An arbitrary function can be approximated at a local position u by computing a Taylor's series about u .

$$f(x) = f(u) + f'(u)(x - u) + \frac{1}{2}f''(u)(x - u)^2 \dots \quad (\text{C.1})$$

For any function of multiple variables $\mathbf{a} = (a_1, \dots, a_N)^\top$ we can make a second order Taylor's approximation to the function about a suitable local position \mathbf{u} .

$$f(\mathbf{a}) = f(\mathbf{u}) + \sum_i \left. \frac{\partial f}{\partial a_i} \right|_{\mathbf{u}} (a_i - u_i) + \frac{1}{2} \sum_{i,j} \left. \frac{\partial^2 f}{\partial a_i \partial a_j} \right|_{\mathbf{u}} (a_i - u_i)(a_j - u_j) \quad (\text{C.2})$$

$$f(\mathbf{a}) = \mathbf{c} + \mathbf{b}^\top (\mathbf{a} - \mathbf{u}) + \frac{1}{2} (\mathbf{a} - \mathbf{u})^\top \mathbf{D} (\mathbf{a} - \mathbf{u}) \quad (\text{C.3})$$

where

$$\begin{aligned} \mathbf{c} &= f(\mathbf{u}) && N - \text{vector} \\ \mathbf{b} &= \nabla f(\mathbf{u}) = \left. \frac{\partial f}{\partial a_i} \right|_{\mathbf{u}} && N - \text{vector} \\ \mathbf{D} &= \left. \frac{\partial^2 f}{\partial a_i \partial a_j} \right|_{\mathbf{u}} && N \times N - \text{matrix} \end{aligned}$$

The matrix \mathbf{D} , whose components are the second partial derivatives of the function, is called the Hessian matrix of the function at \mathbf{u} .

Since we have obtained a Taylor's approximation of the function, we can compute the putative gradient of the function about the local point \mathbf{u} by differentiating equation C.3 with respect to the elements of \mathbf{a} .

$$\frac{\partial f(\mathbf{a})}{\partial \mathbf{a}} = \mathbf{b} + \mathbf{D}(\mathbf{a} - \mathbf{u}) \quad (\text{C.4})$$

$$\nabla f(\mathbf{a}) = \nabla f(\mathbf{u}) + \mathbf{D}(\mathbf{a} - \mathbf{u}) \quad (\text{C.5})$$

To determine the minimum of the function we must find the set of parameters \mathbf{a} that have zero gradient $\nabla f(\mathbf{a}) = 0$.

$$\nabla f(\mathbf{u}) + \mathbf{D}(\mathbf{a} - \mathbf{u}) = 0 \quad (\text{C.6})$$

$$\mathbf{a} = \mathbf{u} - \mathbf{D}^{-1}\nabla f(\mathbf{u}) \quad (\text{C.7})$$

This implies that at a current point in the parameter space $\mathbf{u} = \mathbf{a}_{cur}$ we can jump to the function minimum if we are able to compute the local gradient $\nabla f(\mathbf{a}_{cur})$ and Hessian matrix \mathbf{D} . This update step is known as a Newton iteration and approximates the local shape of the function by a quadratic in order to find a better local minimum. The basic idea is to iterate until convergence of the parameters \mathbf{a} , such that the parameter update step is smaller than some small tolerance ϵ i.e. $|\mathbf{a}_{next} - \mathbf{a}_{cur}| < \epsilon$.

$$\mathbf{a}_{next} = \mathbf{a}_{cur} - \mathbf{D}^{-1}\nabla f(\mathbf{a}_{cur}) \quad (\text{C.8})$$

If the functional approximation is good at the current local point then the update computes a set of parameters close to the actual minimum. On the other hand if we have a poor local approximation to the shape of the function at the current point then about all we can do is take a step down the local gradient direction $\nabla f(\mathbf{a}_{cur})$ toward the minimum by some constant factor μ .

$$\mathbf{a}_{next} = \mathbf{a}_{cur} - \mu \cdot \nabla f(\mathbf{a}_{cur}) \quad (\text{C.9})$$

C.2 Levenberg-Marquardt minimization

If we are able to model the data by some non-linear function with a set of parameters \mathbf{a} then we can compute a χ^2 merit function that describes a goodness of fit of the parameters to the measurement data points \mathbf{x} . The Levenberg-Marquardt method varies

smoothly between Newton iteration and gradient descent, in that where a Newton iteration step fails to decrease the χ^2 error, a step down the local function gradient is taken. The method varies continuously between the two depending on how close to the minimum we are and the shape of the function at the local point. The method has become the standard of non-linear least squares minimization routines.

The vector of data points are measured with a degree of uncertainty and correspondingly we define the measurement error deviation vector σ . The function $f(\mathbf{a})$ computes the set of data point estimates $\hat{\mathbf{x}} = f(\mathbf{a})$ that models the measurement data vector $\mathbf{x} = (x_1, \dots, x_M)^\top$. The corresponding χ^2 merit function is then given by equation C.10.

$$\chi^2(\mathbf{a}) = \sum_{i=1}^M \left(\frac{x_i - f(\mathbf{a} : i)}{\sigma_i} \right)^2 \quad (\text{C.10})$$

The gradient of the χ^2 function with respect to the model parameters \mathbf{a} , which will be zero at the χ^2 minimum, is given by the set of equations C.11.

$$\frac{\partial \chi^2}{\partial a_k} = -2 \sum_{i=1}^M \frac{1}{\sigma_i^2} (x_i - f(\mathbf{a} : i)) \cdot \frac{\partial f(\mathbf{a} : i)}{\partial a_k} \quad k = 1, \dots, N \quad (\text{C.11})$$

We must take another partial derivative in order to build up the Hessian matrix. We use the differentiation product rule $y = u(x) \cdot v(x)$ such that $y' = u'(x) \cdot v(x) + u(x) \cdot v'(x)$ defines the required set of partial derivatives. Equation C.11 is re-written as a product of the elements u_i and v_i .

$$\frac{\partial \chi^2}{\partial a_k} = -2 \sum_{i=1}^M u_i \cdot v_i \quad k = 1, \dots, N \quad (\text{C.12})$$

$$u_i = \frac{x_i - f(\mathbf{a} : i)}{\sigma_i^2} \quad v_i = \frac{\partial f(\mathbf{a} : i)}{\partial a_k} \quad (\text{C.13})$$

The set of partial derivatives of elements u_i and v_i obtained by differentiation with respect to the model parameters \mathbf{a} are given by:

$$\frac{\partial u_i}{\partial a_l} = -\frac{1}{\sigma_i^2} \frac{\partial f(\mathbf{a} : i)}{\partial a_l} \quad \frac{\partial v_i}{\partial a_l} = \frac{\partial^2 f(\mathbf{a} : i)}{\partial a_l \partial a_k} \quad l = 1, \dots, N \quad (\text{C.14})$$

Consequently the form of the second order partial derivatives is given by substitution of the required terms into the product rule $y' = u'(x) \cdot v(x) + u(x) \cdot v'(x)$.

$$\frac{\partial^2 \chi^2}{\partial a_l \partial a_k} = -2 \sum_{i=1}^M \left(\left(-\frac{1}{\sigma_i^2} \cdot \frac{\partial f(\mathbf{a} : i)}{\partial a_l} \frac{\partial f(\mathbf{a} : i)}{\partial a_k} \right) + \left(\frac{x_i - f(\mathbf{a} : i)}{\sigma_i^2} \cdot \frac{\partial^2 f(\mathbf{a} : i)}{\partial a_l \partial a_k} \right) \right) \quad (\text{C.15})$$

Since we know the model function then the $x_i - f(\mathbf{a} : i)$ terms are just the random measurement errors within the experiment, which tend to cancel out over the summation (zero mean). We can then remove the second derivative term from the product such that the elements of the Hessian matrix \mathbf{D} are given by the equation.

$$\frac{\partial^2 \chi^2}{\partial a_l \partial a_k} = 2 \sum_{i=1}^M \frac{1}{\sigma_i^2} \cdot \frac{\partial f(\mathbf{a} : i)}{\partial a_l} \frac{\partial f(\mathbf{a} : i)}{\partial a_k} \quad (\text{C.16})$$

We define the Jacobian \mathbf{J} as the $M \times N$ matrix of partial differentials obtained from the estimated points $\hat{\mathbf{x}} = f(\mathbf{a})$ with respect to the model parameters \mathbf{a} .

$$\mathbf{J} = \begin{bmatrix} \frac{\partial \hat{x}_1}{\partial a_1} & \cdots & \frac{\partial \hat{x}_1}{\partial a_N} \\ \frac{\partial \hat{x}_2}{\partial a_1} & \cdots & \frac{\partial \hat{x}_2}{\partial a_N} \\ \vdots & & \vdots \\ \frac{\partial \hat{x}_M}{\partial a_1} & \cdots & \frac{\partial \hat{x}_M}{\partial a_N} \end{bmatrix} \quad (\text{C.17})$$

The diagonal $M \times M$ covariance matrix of measurement data error deviations $\Sigma_{\mathbf{x}}$.

$$\Sigma_{\mathbf{x}} = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_M^2 \end{bmatrix} \quad (\text{C.18})$$

The residual fitting error M - vector \mathbf{r} .

$$\mathbf{r} = \begin{pmatrix} x_1 - \hat{x}_1 \\ x_2 - \hat{x}_2 \\ \vdots \\ x_M - \hat{x}_M \end{pmatrix} \quad (\text{C.19})$$

The Newton iteration step is then given by:

$$\mathbf{D} \cdot (\mathbf{a}_{next} - \mathbf{a}_{cur}) = -\nabla f(\mathbf{a}_{cur}) \quad (\text{C.20})$$

$$\mathbf{J}^\top \Sigma_{\mathbf{x}}^{-1} \mathbf{J} \cdot \delta \mathbf{a} = \mathbf{J}^\top \Sigma_{\mathbf{x}}^{-1} \mathbf{r} \quad (\text{C.21})$$

Solution of this set of equations of the form $\mathbf{A}\mathbf{x} = \mathbf{b}$ then gives the set of parameter updates $\delta \mathbf{a}$. The new parameters are then adjusted by the update.

$$\mathbf{a}_{next} = \mathbf{a}_{cur} + \delta \mathbf{a} \quad (\text{C.22})$$

We can then test whether the updated set of parameters \mathbf{a}_{next} decreases the χ^2 error. If the parameter update fails to decrease the χ^2 error then we must take a step down the local gradient direction $\nabla f(\mathbf{a}_{cur})$ toward the minimum by some constant factor μ . Consider the constant μ in the gradient descent equation C.9. We do not know what scale this parameter should be in order to facilitate a suitable step down the local gradient direction. Marquardt's first insight is that the components of the Hessian matrix, even if they are not usable in any precise fashion, give some information about the order of magnitude of the problem scale. The quantity χ^2 is non-dimensional, in that it is a pure number. On the other hand elements of $\nabla f(\mathbf{a}_{cur}) = \frac{\partial \chi^2}{\partial a_k}$ have dimensional units of $1/a_k$ and elements of $\delta \mathbf{a}$ units of a_k .

$$\begin{aligned} \mathbf{a}_{next} - \mathbf{a}_{cur} &= -\mu \cdot \nabla f(\mathbf{a}_{cur}) \\ \delta \mathbf{a} &= -\mu \cdot \nabla f(\mathbf{a}_{cur}) \end{aligned} \quad (\text{C.23})$$

The constants of proportionality c_k between elements of the vectors $\nabla f(\mathbf{a}_{cur})$ and $\delta \mathbf{a}$ must have dimensional units of a_k^2 , i.e.

$$\Delta a_k = -c_k \cdot \frac{\Delta \chi^2}{\Delta a_k} \quad (\text{C.24})$$

where units of $\Delta \chi^2$ are just pure numbers. The diagonal elements of the Hessian matrix $\mathbf{D}_{kk} = \frac{\partial^2 \chi^2}{\partial a_k^2}$ have similar units, in that they are reciprocally related $\Delta \chi^2 / \Delta a_k^2$. The scalar mapping between the two vectors is then given by:

$$\delta a_k = -\frac{1}{\lambda \cdot \mathbf{D}_{kk}} \cdot \nabla f(\mathbf{a}_{cur})_k \quad (\text{C.25})$$

where λ is a non-dimensional scaling factor that can be adjusted to cut down the gradient step. Setting $\lambda \gg 1$ decreases the update step while setting $\lambda \ll 1$ increases

it. Marquardt's second insight is that the two update equations for Newton iteration and gradient descent steps can be combined. We rearrange the gradient descent equation C.25 into a form similar to the Newton iteration equation C.20.

$$\lambda \cdot \mathbf{D}_{kk} \cdot \delta a_k = -\nabla f(\mathbf{a}_{cur})_k \quad (\text{C.26})$$

The two can be combined by defining a new matrix \mathbf{D}^* with augmented diagonal elements.

$$\begin{aligned} \mathbf{D}_{kk}^* &\equiv (1 + \lambda) \cdot \mathbf{D}_{kk} \\ \mathbf{D}_{jk}^* &\equiv \mathbf{D}_{jk} \quad (j \neq k) \end{aligned} \quad (\text{C.27})$$

The parameter update equation can then be written.

$$\mathbf{D}^* \cdot \delta \mathbf{a} = -\nabla f(\mathbf{a}_{cur}) \quad (\text{C.28})$$

When λ is very large the matrix \mathbf{D}^* is forced into being diagonally dominant, hence the update equation goes over to being identical to the gradient descent step in equation C.26. As λ approaches zero the update equation becomes identical to the Newton iteration equation C.20. The parameter λ can be varied to facilitate a smooth transition between Newton iteration and gradient descent. Given an initial guess for the set of parameters \mathbf{a} the Levenberg-Marquardt method can be implemented as follows:

1. Compute the initial $\chi^2(\mathbf{a})$ residual error.
2. Set an initial value for the Levenberg-Marquardt parameter $\lambda = 0.001$.
3. Compute the Jacobian matrix \mathbf{J} and the residual error vector \mathbf{r} for the model function with the set of parameters \mathbf{a} . Form the set of normal equations C.21, where $\mathbf{D} = \mathbf{J}^\top \Sigma_{\mathbf{x}}^{-1} \mathbf{J}$ and $-\nabla f(\mathbf{a}_{cur}) = \mathbf{J}^\top \Sigma_{\mathbf{x}}^{-1} \mathbf{r}$.
4. Augment the diagonal elements of \mathbf{D} with the scaling factor $(1 + \lambda)$ to form the matrix \mathbf{D}^* .
5. Solve the linear equations C.28 for the set of update parameters $\delta \mathbf{a}$ and evaluate the merit function $\chi^2(\mathbf{a} + \delta \mathbf{a})$.
6. If $\chi^2(\mathbf{a} + \delta \mathbf{a}) \geq \chi^2(\mathbf{a})$ fails to decrease the fitting cost then increase the Levenberg-Marquardt parameter λ by a factor of 10 and go back to step 4. If however the change in fitting cost is lower than a small tolerance ϵ_r such that $|\chi^2(\mathbf{a} + \delta \mathbf{a}) - \chi^2(\mathbf{a})| < \epsilon_r$ or the relative change in the update parameters is smaller than a small

tolerance ϵ_a such that $\|\delta\mathbf{a}\| / \|\mathbf{a} + \delta\mathbf{a}\| < \epsilon_a$ then return with the set of minimized parameters \mathbf{a} .

7. Since the update decreased the fitting cost, update the set of trial parameters $\mathbf{a} \leftarrow (\mathbf{a} + \delta\mathbf{a})$, decrease the Levenberg-Marquardt parameter λ by a factor of 10 and go back to step 3. If however the change in fitting cost is lower than a small tolerance ϵ_r such that $|\chi^2(\mathbf{a} + \delta\mathbf{a}) - \chi^2(\mathbf{a})| < \epsilon_r$ or the relative change in the update parameters is smaller than a small tolerance ϵ_a such that $\|\delta\mathbf{a}\| / \|\mathbf{a} + \delta\mathbf{a}\| < \epsilon_a$ then return with the new set of minimized parameters \mathbf{a} .

The Levenberg-Marquardt parameter λ is updated by applying multiplicative scaling factors 10 (increase), 1/10 (decrease). For some minimization problems you may find that the method spends a lot of time switching between two states i.e. $[i, d, i, d, i, \dots]$ where i denotes a λ parameter increase and d a decrease. This indicates that half the updates (and solution of the normal equations) failed to decrease the χ^2 error. We can keep a history of the past T update states and detect such a situation. It is advantageous to then apply an asymmetric update that decreases λ by a factor of 1/2 on successful decrease of the χ^2 merit function i.e. bias the method in favour of gradient descent to reduce the number of failed iterations.

Once the set of minimized parameters \mathbf{a} have been computed the covariance matrix $\Sigma_{\mathbf{a}}$ can be computed by inverting the Hessian matrix $\Sigma_{\mathbf{a}} = \mathbf{D}^{-1}$. For some minimization problems the vector \mathbf{a} over-parameterises the model. i.e. The nine parameters of a homography matrix encode the projective mapping that has only eight degrees of freedom. In this case the Jacobian matrix \mathbf{J} is rank degenerate, hence so is the Hessian matrix $\mathbf{D} = \mathbf{J}^T \Sigma_{\mathbf{x}}^{-1} \mathbf{J}$. In this case the pseudo-inverse must be used to compute the covariance matrix $\Sigma_{\mathbf{a}} = (\mathbf{J}^T \Sigma_{\mathbf{x}}^{-1} \mathbf{J})^+$. It is worth knowing in advance the theoretical rank of the system in order that we can enforce the rank constraint.

C.3 Confidence limits

We can summarize the probability distribution of errors within parameter estimation in the form of confidence limits. A confidence region is just a section of the N dimensional parameter space \mathbf{a} that contains a certain percentage of the total probability distribution. We can express a confidence by saying e.g. “there is a 99 percent chance that the true parameters fall within this region around the predicted value”. Certain percentages are customary within scientific usage, namely $\sigma - 68.3\%$, $2\sigma - 95.4\%$ and $3\sigma - 99.73\%$. The confidence and shape of the region are chosen by the experimenter, though obviously you want the shape of the region to be reasonably well centred about the estimated measurement \mathbf{a}_0 .

If we are able to model the measurement data through a set of parameters \mathbf{a} , then we can compute a $\chi^2(\mathbf{a})$ minimization between measured data points \mathbf{x} and the set of corresponding estimates $\hat{\mathbf{x}}$ from the fitting function $f(\mathbf{a})$.

$$\chi^2(\mathbf{a}) = \sum_{i=1}^M \left(\frac{x_i - f(\mathbf{a} : i)}{\sigma_i} \right)^2 \quad (\text{C.29})$$

The χ^2 fitting function error is minimum at the estimated parameters \mathbf{a}_0 . If the vector of parameters \mathbf{a} is perturbed away from \mathbf{a}_0 then χ^2 increases. The region within which χ^2 increases by no more than a set amount $\Delta\chi^2$ defines some N dimensional confidence region around \mathbf{a}_0 . There is then a suitable $\Delta\chi^2$ that causes the region to contain 68.3 percent, etc. of the probability distribution of parameter vectors \mathbf{a} . These regions are taken as the confidence limits for the estimated parameters \mathbf{a}_0 .

We are usually not interested in the full N dimensional confidence region but in individual regions of some smaller number of parameters ν . In this case the confidence regions in the ν dimensional subspace are the projections of the N dimensional regions, defined by the fixed $\Delta\chi^2$ boundaries, into the ν dimensional spaces of interest. We are frequently only interested in the confidence interval of each parameter taken separately, and as such $\nu = 1$. The χ^2 distribution with $\nu = 1$ degree of freedom has the same distribution as that of the square of a single normally distributed quantity, thus $\Delta\chi^2 < 1$ occurs 68.3 percent (1σ) of the time, $\Delta\chi^2 < 4$ occurs 99.73 percent (2σ) of the time, etc.

The covariance matrix that comes out of the χ^2 minimization has a clear quantitative interpretation only if the measurement errors are normally distributed. Let $\delta\mathbf{a}$ be a change in the parameters whose first component is arbitrary δa_1 , but the rest of whose components are chosen to minimize the χ^2 error. Since $\delta\mathbf{a}$ by hypothesis minimizes χ^2 in all but the first component the remaining components of the normal equations continue to hold and have the form:

$$\begin{aligned} \mathbf{J}^\top \Sigma_{\mathbf{x}}^{-1} \mathbf{J} \cdot \delta\mathbf{a} &= \mathbf{J}^\top \Sigma_{\mathbf{x}}^{-1} \mathbf{r} \\ \mathbf{D} \cdot \delta\mathbf{a} &= \mathbf{b} \end{aligned} \quad (\text{C.30})$$

$$\delta\mathbf{a} = \Sigma_{\mathbf{a}} \begin{pmatrix} b_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (\text{C.31})$$

Where $\Sigma_{\mathbf{a}}$ is the covariance matrix \mathbf{D}^{-1} of the estimated parameters \mathbf{a} .

$$\Sigma_{\mathbf{a}} = \begin{bmatrix} \Sigma_{a11} & \cdots & \Sigma_{a1N} \\ \vdots & \ddots & \\ \Sigma_{aN1} & & \Sigma_{aNN} \end{bmatrix} \quad (\text{C.32})$$

b_1 is an arbitrary constant we can adjust to give the desired left hand side.

$$\delta a_1 = \Sigma_{a11} \cdot b_1 \quad (\text{C.33})$$

$$b_1 = \delta a_1 / \Sigma_{a11} \quad (\text{C.34})$$

The value of $\Delta\chi^2$ is given in general by the equation.

$$\Delta\chi^2 = \delta\mathbf{a}^\top \mathbf{D} \delta\mathbf{a} \quad (\text{C.35})$$

This can be rewritten by substitution of $\delta\mathbf{a}$ from equation C.31. We also note that both \mathbf{D} and $\Sigma_{\mathbf{a}}$ are symmetric and are inverses of one another.

$$\Delta\chi^2 = \begin{pmatrix} b_1 & 0 & \cdots & 0 \end{pmatrix} \Sigma_{\mathbf{a}}^\top \mathbf{D} \Sigma_{\mathbf{a}} \begin{pmatrix} b_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (\text{C.36})$$

$$\Delta\chi^2 = \begin{pmatrix} b_1 & 0 & \cdots & 0 \end{pmatrix} \Sigma_{\mathbf{a}} \begin{pmatrix} b_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (\text{C.37})$$

$$\Delta\chi^2 = b_1^2 \cdot \Sigma_{a11} \quad (\text{C.38})$$

If we then substitute the value of b_1 from equation C.34 into expression C.38 we can compute a relationship between the confidence region $\pm\delta a_1$ and the formal standard error $\pm\sigma_1 = \sqrt{\Sigma_{a11}}$ defined from the covariance matrix.

$$\Delta\chi^2 = \left(\frac{\delta a_1}{\Sigma_{a11}} \right)^2 \cdot \Sigma_{a11} \quad (\text{C.39})$$

$$\Sigma_{a11} \cdot \Delta\chi^2 = \delta a_1^2 \quad (\text{C.40})$$

$$\delta a_1 = \pm \sqrt{\Delta\chi^2} \sqrt{\Sigma_{a11}} \quad (\text{C.41})$$

$$\delta a_1 = \pm k \cdot \sigma_1 \quad (\text{C.42})$$

We then find that the 68.3 percent confidence region corresponds to $\pm\sigma_1$, the 95.4 percent region corresponds to $\pm 2\sigma_1$, etc. We can use the same argument for each of the subsequent parameters of \mathbf{a} such that $\delta a_i = \pm k \cdot \sigma_i$ is the confidence region associated with each parameter a_i .

The root mean square measurement error ϵ may be used in equation C.29 instead of the individual measurement errors σ_i . Where we do not know the individual measurement errors, we can assume that all data points have been specified with the same error $\sigma_i = 1$. This enables us to compute a χ^2 minimization of the parameters \mathbf{a} , then measure the root mean square fitting error ϵ between estimated and real data points. Using this value of ϵ we can then compute a credible estimate of the confidence limits associated with the set of model parameters \mathbf{a} .

C.4 Sparse Levenberg-Marquardt minimization

The LM algorithm described in section C.2 is suitable for a relatively small number of parameters. However when minimizing cost functions with respect to a large number of parameters the simple LM algorithm is not very suitable. The central step of the LM algorithm requires the solution of the normal equations which has a complexity $O(N^3)$. For large parameter vectors the computational cost of solving these equations is high, and susceptible to roundoff error. However many minimization problems have a certain block sparse structure that one can take advantage of in order to reduce the complexity of the minimization problem.

C.4.1 General sparse LM method

In the context of computer vision sparse techniques are useful in most reconstruction problems where the parameter vector can be split into two sections $\mathbf{P} = (\mathbf{a}^\top, \mathbf{b}^\top)^\top$. The first parameterises the set of system entities i.e. the set of homography / camera matrices and the second, the set of worldspace points. The Jacobian matrix $\mathbf{J} = [\partial\hat{\mathbf{x}}/\partial\mathbf{P}]$ has a block structure of the form $\mathbf{J} = [\mathbf{A} \mid \mathbf{B}]$, where $\mathbf{A} = [\partial\hat{\mathbf{x}}/\partial\mathbf{a}]$ and $\mathbf{B} = [\partial\hat{\mathbf{x}}/\partial\mathbf{b}]$.

The set of normal equations $\mathbf{J}^\top \Sigma_{\mathbf{x}}^{-1} \mathbf{J} \cdot \delta \mathbf{p} = \mathbf{J}^\top \Sigma_{\mathbf{x}}^{-1} \mathbf{r}$ that need to be solved in order to compute the parameter updates then have the form:

$$\left[\begin{array}{c|c} \mathbf{A}^\top \Sigma_{\mathbf{x}}^{-1} \mathbf{A} & \mathbf{A}^\top \Sigma_{\mathbf{x}}^{-1} \mathbf{B} \\ \hline \mathbf{B}^\top \Sigma_{\mathbf{x}}^{-1} \mathbf{A} & \mathbf{B}^\top \Sigma_{\mathbf{x}}^{-1} \mathbf{B} \end{array} \right] \begin{pmatrix} \delta_{\mathbf{a}} \\ \delta_{\mathbf{b}} \end{pmatrix} = \begin{pmatrix} \mathbf{A}^\top \Sigma_{\mathbf{x}}^{-1} \mathbf{r} \\ \mathbf{B}^\top \Sigma_{\mathbf{x}}^{-1} \mathbf{r} \end{pmatrix} \quad (\text{C.43})$$

$$\begin{bmatrix} \mathbf{U} & \mathbf{W} \\ \mathbf{W}^\top & \mathbf{V} \end{bmatrix} \begin{pmatrix} \delta_{\mathbf{a}} \\ \delta_{\mathbf{b}} \end{pmatrix} = \begin{pmatrix} \mathbf{e}_{\mathbf{a}} \\ \mathbf{e}_{\mathbf{b}} \end{pmatrix} \quad (\text{C.44})$$

We then augment the diagonal elements of the Hessian matrix with the Levenberg-Marquardt scale factor $(1 + \lambda)$. This augmentation alters the matrices \mathbf{U} and \mathbf{V} such that the normal equations can be written.

$$\begin{bmatrix} \mathbf{U}^* & \mathbf{W} \\ \mathbf{W}^\top & \mathbf{V}^* \end{bmatrix} \begin{pmatrix} \delta_{\mathbf{a}} \\ \delta_{\mathbf{b}} \end{pmatrix} = \begin{pmatrix} \mathbf{e}_{\mathbf{a}} \\ \mathbf{e}_{\mathbf{b}} \end{pmatrix} \quad (\text{C.45})$$

We aim to apply a set of transforms to both sides of the equation in order to make the parameter updates independent. We multiply each side of the normal equations by the matrix:

$$\begin{bmatrix} \mathcal{I} & -\mathbf{W}\mathbf{V}^{*-1} \\ \mathbf{0} & \mathcal{I} \end{bmatrix} \quad (\text{C.46})$$

This results in the elimination of the top right hand block of the Hessian matrix.

$$\begin{bmatrix} \mathbf{U}^* - \mathbf{W}\mathbf{V}^{*-1}\mathbf{W}^\top & \mathbf{0} \\ \mathbf{W}^\top & \mathbf{V}^* \end{bmatrix} \begin{pmatrix} \delta_{\mathbf{a}} \\ \delta_{\mathbf{b}} \end{pmatrix} = \begin{pmatrix} \mathbf{e}_{\mathbf{a}} - \mathbf{W}\mathbf{V}^{*-1}\mathbf{e}_{\mathbf{b}} \\ \mathbf{e}_{\mathbf{b}} \end{pmatrix} \quad (\text{C.47})$$

The first set of update parameters $\delta_{\mathbf{a}}$ may then be solved by solution of the set of equations of the form $\mathbf{A}\mathbf{x} = \mathbf{b}$.

$$(\mathbf{U}^* - \mathbf{W}\mathbf{V}^{*-1}\mathbf{W}^\top) \cdot \delta_{\mathbf{a}} = \mathbf{e}_{\mathbf{a}} - \mathbf{W}\mathbf{V}^{*-1}\mathbf{e}_{\mathbf{b}} \quad (\text{C.48})$$

Subsequently the second set of update parameters $\delta_{\mathbf{b}}$ can be found by back-substitution of $\delta_{\mathbf{a}}$ and solution of the corresponding equations of the form $\mathbf{A}\mathbf{x} = \mathbf{b}$.

$$\begin{aligned} \mathbf{W}^\top \delta_{\mathbf{a}} + \mathbf{V}^* \delta_{\mathbf{b}} &= \mathbf{e}_{\mathbf{b}} \\ \mathbf{V}^* \delta_{\mathbf{b}} &= \mathbf{e}_{\mathbf{b}} - \mathbf{W}^\top \delta_{\mathbf{a}} \end{aligned} \quad (\text{C.49})$$

C.4.2 Covariance matrix

The covariance matrix of the estimated parameters \mathbf{P} is given by $\Sigma_{\mathbf{p}} = (\mathbf{J}^T \Sigma_{\mathbf{x}}^{-1} \mathbf{J})^{-1}$ and in the over-parameterised case by the pseudo-inverse $\Sigma_{\mathbf{p}} = (\mathbf{J}^T \Sigma_{\mathbf{x}}^{-1} \mathbf{J})^+$. The covariance matrix $\Sigma_{\mathbf{p}}$ can be computed by a process of Gaussian elimination.

$$\begin{bmatrix} \mathbf{U} & \mathbf{W} \\ \mathbf{W}^T & \mathbf{V} \end{bmatrix} \begin{bmatrix} \Sigma_{\mathbf{a}} & \Sigma_{\mathbf{ab}} \\ \Sigma_{\mathbf{ab}}^T & \Sigma_{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} \mathcal{I} & \\ & \mathcal{I} \end{bmatrix} \quad (\text{C.50})$$

$$\begin{bmatrix} \mathbf{U} - \mathbf{WV}^{-1}\mathbf{W}^T & \mathbf{0} \\ \mathbf{W}^T & \mathbf{V} \end{bmatrix} \begin{bmatrix} \Sigma_{\mathbf{a}} & \Sigma_{\mathbf{ab}} \\ \Sigma_{\mathbf{ab}}^T & \Sigma_{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} \mathcal{I} & -\mathbf{WV}^{-1} \\ \mathbf{0} & \mathcal{I} \end{bmatrix} \quad (\text{C.51})$$

If we define the matrix $\mathbf{Y} = \mathbf{WV}^{-1}$, where we assume that matrix \mathbf{V} is invertible, then the individual covariance matrices of the parameter vector are given by:

$$\begin{aligned} (\mathbf{U} - \mathbf{WV}^{-1}\mathbf{W}^T)\Sigma_{\mathbf{a}} &= \mathcal{I} \\ \Sigma_{\mathbf{a}} &= (\mathbf{U} - \mathbf{WV}^{-1}\mathbf{W}^T)^+ \end{aligned} \quad (\text{C.52})$$

$$\begin{aligned} (\mathbf{U} - \mathbf{WV}^{-1}\mathbf{W}^T)\Sigma_{\mathbf{ab}} &= -\mathbf{WV}^{-1} \\ \Sigma_{\mathbf{ab}} &= -\Sigma_{\mathbf{a}}\mathbf{Y} \end{aligned} \quad (\text{C.53})$$

$$\begin{aligned} \mathbf{W}^T \Sigma_{\mathbf{ab}} + \mathbf{V} \Sigma_{\mathbf{b}} &= \mathcal{I} \\ \mathbf{V} \Sigma_{\mathbf{b}} &= \mathcal{I} + \mathbf{W}^T \Sigma_{\mathbf{a}} \mathbf{Y} \\ \Sigma_{\mathbf{b}} &= \mathbf{V}^{-1} + \mathbf{Y}^T \Sigma_{\mathbf{a}} \mathbf{Y} \end{aligned} \quad (\text{C.54})$$

C.4.3 Block sparse LM method

The sparse LM method gives a clear advantage when the Jacobian matrix obeys a certain sparseness condition. If the parameter vector can be divided up into segments $\mathbf{P} = (\mathbf{a}^T, \mathbf{b}_1^T, \mathbf{b}_2^T, \dots, \mathbf{b}_n^T)^T$, such that each of the parameters \mathbf{b}_i are independent then the Jacobian matrix has the form.

$$\mathbf{J} = \left[\begin{array}{c|ccc} \mathbf{A}_1 & \mathbf{B}_1 & & \\ \mathbf{A}_2 & & \mathbf{B}_2 & \\ \vdots & & & \ddots \\ \mathbf{A}_n & & & \mathbf{B}_n \end{array} \right] \quad (\text{C.55})$$

We suppose further that all the measurements \mathbf{X}_i are independent with covariance matrices $\Sigma_{\mathbf{x}_i}$ such that $\Sigma_{\mathbf{x}} = \text{diag}(\Sigma_{\mathbf{x}_1}, \dots, \Sigma_{\mathbf{x}_n})$. The corresponding set of normal equations can be written.

$$\left[\begin{array}{c|cccc} \sum_i \mathbf{A}_i^\top \Sigma_{\mathbf{x}_i}^{-1} \mathbf{A}_i & \mathbf{A}_1^\top \Sigma_{\mathbf{x}_1}^{-1} \mathbf{B}_1 & \mathbf{A}_2^\top \Sigma_{\mathbf{x}_2}^{-1} \mathbf{B}_2 & \cdots & \mathbf{A}_n^\top \Sigma_{\mathbf{x}_n}^{-1} \mathbf{B}_n \\ \mathbf{B}_1^\top \Sigma_{\mathbf{x}_1}^{-1} \mathbf{A}_1 & \mathbf{B}_1^\top \Sigma_{\mathbf{x}_1}^{-1} \mathbf{B}_1 & & & \\ \mathbf{B}_2^\top \Sigma_{\mathbf{x}_2}^{-1} \mathbf{A}_2 & & \mathbf{B}_2^\top \Sigma_{\mathbf{x}_2}^{-1} \mathbf{B}_2 & & \\ \vdots & & & \ddots & \\ \mathbf{B}_n^\top \Sigma_{\mathbf{x}_n}^{-1} \mathbf{A}_n & & & & \mathbf{B}_n^\top \Sigma_{\mathbf{x}_n}^{-1} \mathbf{B}_n \end{array} \right] \begin{pmatrix} \delta \mathbf{a} \\ \delta \mathbf{b}_1 \\ \delta \mathbf{b}_2 \\ \vdots \\ \delta \mathbf{b}_n \end{pmatrix} = \begin{pmatrix} \sum_i \mathbf{A}_i^\top \Sigma_{\mathbf{x}_i}^{-1} \mathbf{r}_i \\ \mathbf{B}_1^\top \Sigma_{\mathbf{x}_1}^{-1} \mathbf{r}_1 \\ \mathbf{B}_2^\top \Sigma_{\mathbf{x}_2}^{-1} \mathbf{r}_2 \\ \vdots \\ \mathbf{B}_n^\top \Sigma_{\mathbf{x}_n}^{-1} \mathbf{r}_n \end{pmatrix} \quad (\text{C.56})$$

$$\left[\begin{array}{c|cccc} \mathbf{U} & \mathbf{W}_1 & \mathbf{W}_2 & \cdots & \mathbf{W}_n \\ \mathbf{W}_1^\top & \mathbf{V}_1 & & & \\ \mathbf{W}_2^\top & & \mathbf{V}_2 & & \\ \vdots & & & \ddots & \\ \mathbf{W}_n^\top & & & & \mathbf{V}_n \end{array} \right] \begin{pmatrix} \delta \mathbf{a} \\ \delta \mathbf{b}_1 \\ \delta \mathbf{b}_2 \\ \vdots \\ \delta \mathbf{b}_n \end{pmatrix} = \begin{pmatrix} \mathbf{e}_a \\ \mathbf{e}_{b1} \\ \mathbf{e}_{b2} \\ \vdots \\ \mathbf{e}_{bn} \end{pmatrix} \quad (\text{C.57})$$

We augment the diagonal elements of the Hessian matrix with the Levenberg-Marquardt scaling factor $(1+\lambda)$, then apply the transformation matrix to both sides of the equations.

$$\left[\begin{array}{c|cccc} \mathcal{I} & -\mathbf{W}_1 \mathbf{V}_1^{*-1} & -\mathbf{W}_2 \mathbf{V}_2^{*-1} & \cdots & -\mathbf{W}_n \mathbf{V}_n^{*-1} \\ \mathbf{0} & \mathcal{I} & & & \\ \mathbf{0} & & \mathcal{I} & & \\ \vdots & & & \ddots & \\ \mathbf{0} & & & & \mathcal{I} \end{array} \right] \quad (\text{C.58})$$

The set of equations we need to solve in order to compute the parameter updates is given by.

$$\left[\begin{array}{c|ccc} \mathbf{U} - \sum_i \mathbf{W}_i \mathbf{V}_i^{*-1} \mathbf{W}_i^\top & & & \\ \hline \mathbf{W}_1^\top & \mathbf{V}_1^* & & \\ \mathbf{W}_2^\top & & \mathbf{V}_2^* & \\ \vdots & & & \ddots \\ \mathbf{W}_n^\top & & & & \mathbf{V}_n^* \end{array} \right] \begin{pmatrix} \delta \mathbf{a} \\ \delta \mathbf{b}_1 \\ \delta \mathbf{b}_2 \\ \vdots \\ \delta \mathbf{b}_n \end{pmatrix} = \begin{pmatrix} \mathbf{e}_a - \sum_i \mathbf{W}_i \mathbf{V}_i^{*-1} \mathbf{e}_{\mathbf{b}i} \\ \mathbf{e}_{\mathbf{b}1} \\ \mathbf{e}_{\mathbf{b}2} \\ \vdots \\ \mathbf{e}_{\mathbf{b}n} \end{pmatrix} \quad (\text{C.59})$$

We note that the transformation step only requires that we compute the inverse of each of the \mathbf{V}_i^* blocks and the corresponding set of matrix multiplications $\mathbf{W}_i \mathbf{V}_i^{*-1} \mathbf{W}_i^\top$ and $\mathbf{W}_i \mathbf{V}_i^{*-1} \mathbf{e}_{\mathbf{b}i}$. The first set of update parameters $\delta \mathbf{a}$ may then be solved by solution of the set of equations of the form $\mathbf{A} \mathbf{x} = \mathbf{b}$.

$$\left(\mathbf{U} - \sum_i \mathbf{W}_i \mathbf{V}_i^{*-1} \mathbf{W}_i^\top \right) \cdot \delta \mathbf{a} = \mathbf{e}_a - \sum_i \mathbf{W}_i \mathbf{V}_i^{*-1} \mathbf{e}_{\mathbf{b}i} \quad (\text{C.60})$$

Subsequently the second set of update parameters $\delta \mathbf{b}$ can be found by back-substitution of $\delta \mathbf{a}$ and solution of the corresponding equations of the form $\mathbf{A} \mathbf{x} = \mathbf{b}$. This substitution step can be performed for each of the individual parameter updates $\delta \mathbf{b}_i$.

$$\begin{aligned} \mathbf{W}_i^\top \delta \mathbf{a} + \mathbf{V}_i^* \delta \mathbf{b}_i &= \mathbf{e}_{\mathbf{b}i} \\ \delta \mathbf{b}_i &= \mathbf{V}_i^{*-1} \cdot (\mathbf{e}_{\mathbf{b}i} - \mathbf{W}_i^\top \delta \mathbf{a}) \end{aligned} \quad (\text{C.61})$$

Since we have already computed the inverses of each of the \mathbf{V}_i^* blocks then this parameter back substitution solution step amounts to no more than a few matrix multiplications. For parameter vectors with large numbers of $\mathbf{b}_1, \dots, \mathbf{b}_n$ blocks then solution of the normal equations using the basic LM method requires a computational complexity of the order $O(n^3)$ in inverting the Hessian matrix. On the other hand by applying the sparse LM method the computational complexity is of the order $O(n)$ required for inversion of each of the independent \mathbf{V}_i^* blocks of the Hessian matrix.

If we define the matrix $\mathbf{Y}_i = \mathbf{W}_i \mathbf{V}_i^{-1}$, where we assume that matrix \mathbf{V}_i is invertible, then the individual covariance matrices of the partitioned block parameter vector are given by:

$$\Sigma_{\mathbf{a}} = (\mathbf{U} - \sum_i \mathbf{W}_i \mathbf{V}_i^{-1} \mathbf{W}_i^{\top})^+ \quad (\text{C.62})$$

$$\Sigma_{\mathbf{a}b_i} = -\Sigma_{\mathbf{a}} \mathbf{Y}_i \quad (\text{C.63})$$

$$\Sigma_{\mathbf{b}_i, \mathbf{b}_i} = \mathbf{V}_i^{-1} + \mathbf{Y}_i^{\top} \Sigma_{\mathbf{a}} \mathbf{Y}_i \quad (\text{C.64})$$

$$\Sigma_{\mathbf{b}_i, \mathbf{b}_j} = \mathbf{Y}_i^{\top} \Sigma_{\mathbf{a}} \mathbf{Y}_j \quad (i \neq j) \quad (\text{C.65})$$

Appendix D

Implementation and Software

D.1 Introduction

There are a number of commercially available tools and libraries for use within computer vision, such as: Matlab, OpenCV, IUE, Targeted and VCL. Correspondingly, there are also many freely available numerical libraries, such as: Numerical Recipes in C++, and old Fortran code from LAPLACE and MAILSACK. Although these software libraries provide a large range of basic algorithms, due to the complexity of certain optimization problems within projective geometry, many of the larger sparse optimization methods are not freely available.

We give here the details of the software libraries and visualisation tools, that were implemented to provide the means required to label and analyse the captured image data. The software was built primarily out of necessity, though also provided a chance to further improve the author's understanding of projective geometry and computer vision. The algorithms employed to build a number of the tools used throughout the project are described within the text of Hartley and Zisserman [40]. The work within this chapter demonstrates the significant practical contribution made by the author in conjunction with the written theoretical material.

D.2 Testing

The software was built by using a modular design. Many of the minimization problems encountered during the project require that we evaluate a sufficiently complex model function. As a result, many of the components have been standardized and broken into smaller pieces, i.e. the homography transformation of one point set to another and the parameterisation of a rotation matrix by using a Rodrigues vector. This also enables us to test and verify each smaller component independently.

While the projection function that models the subject motion is already fairly complicated, computation of the analytical Jacobian matrix is an order of magnitude more complex. Splitting and testing each of the standard transformation functions removes many of the possibilities for introducing coding errors. Subsequently, the time it takes to track down and eliminate errors is also reduced. Due to the nature of a research project, many developers often use numerical differentiation methods (forward difference) to compute the Jacobian matrices. Since all the optimization problems within this project are similar and sparse, the large number of parameters involved requires some extremely long computation times to evaluate the Jacobian matrices. Analytical computation of the Jacobian matrices is therefore preferable and more accurate. Building the complete Jacobian matrix out of those derived from each individual standard transformation function is then just a question of performing a number of simple matrix-matrix multiplications.

For testing purposes, the computed analytical Jacobian matrices may be compared to their numerical forward difference counterparts by determining the Frobenius norm of the difference between both matrices. For the matrices to be deemed similar, this norm should differ by no more than some small empirical tolerance value.

D.3 Geometry Tools

The software tools built during the author's time at university arose and evolved as a result of the necessity to understand the basic principles of projective geometry. The geometry library is separate and written in C++, allowing it to be compiled on almost any platform. The visualisation tools make extensive use of graphical features, hence a degree of platform dependence is necessary. Software tools to perform manual marking of data and image visualisation were built using C++ and the Microsoft Foundation Classes (MFC) within Visual Studio 6.0. With the exception of a few important numerical routines, such as singular value decomposition, all software was implemented by the author during the period of study.

D.3.1 Primitive Visualisation

In order to enable interest features within an image to be manually marked to sub-pixel accuracy, a visualisation tool was built that allows users to zoom and move around to any part of the image plane. Points, lines and digital curves can be manually placed and manipulated within the image plane by appropriate interaction with the mouse. Primitives may be given numerical properties, such as an integer identification number that corresponds to a specific landmark index. The manual marking tool allows us to specify how we wish to handle applying new primitives to the image. New points may be

applied by clicking the mouse, with each point being added with increasing, decreasing or constant parameter values. We can page through the images of a sequence by clicking the appropriate button or corresponding keystroke on the keyboard. We have found that the fastest way to manually mark interest features, that track through an image sequence, is to pre-select a constant parameter value and insert a single point at a time onto each image frame within the sequence. Since a tracked interest point does not move very far between frames, applying consecutive point features requires that we only move the mouse a small distance, click the mouse once then press a single key on the keyboard to advance the frame. This process is repeated until all required landmark features are identified. The visualisation tool keeps track of any changes that have occurred to the geometry within the image sequence, hence saving all the modified geometry to a file is simple. Figure D.1 illustrates this simple point and click user interface that allows us to identify the low level 2D primitive features within the set of images.

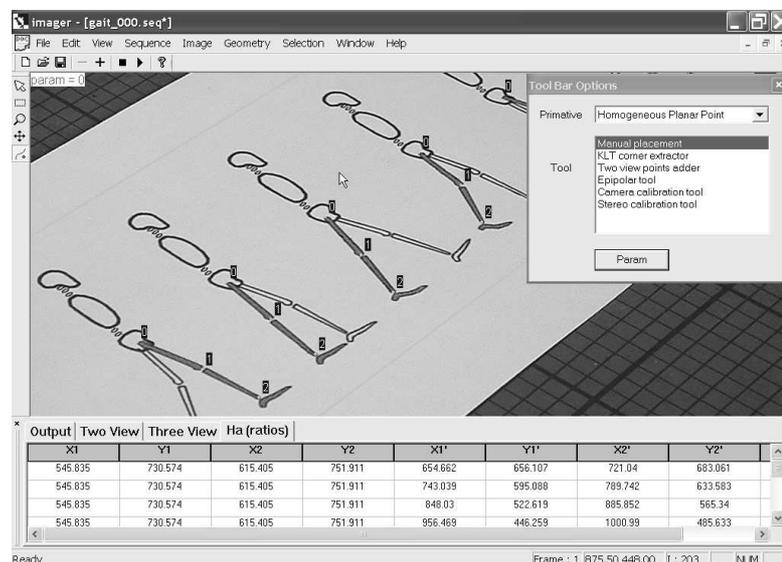


FIGURE D.1: Manual marking tool. The user interface allows a simple point and click interface to manually identify point features within an image.

The tool provides an easy way to display information about each primitive. Primitive attributes can be changed by right clicking close to the object to open the properties dialog. Primitives can be selected and grouped together. The selection mechanism allows us to identify the input primitives required for certain types of image operation. The tool provides an easy way to select, group, delete and parameterise primitives either by primitive type, area selection, group relationship, parameter or frame position.

D.3.2 Planar geometry

Planar transformation of structure is achieved by a 3×3 matrix mapping known as a homography. One of the most important features of the image tool is its ability to apply planar image distortions. The distorted output images are computed by the

method of bilinear interpolation. We can specify an arbitrary image transformation by manually entering the elements of the homography matrix or by loading them from a file. The image distortion dialog box also allows us to generate the homography mapping by applying a number of rigid motion transformations (rotation, translation and scale) to the current matrix, by entering the appropriate values into the custom controls. Additional controls allow us to set the identity matrix, compute the matrix transpose and matrix inverse. We can also specify the size of the output image should it differ from the input dimensions. Figure D.2 shows the image distortion dialog box.

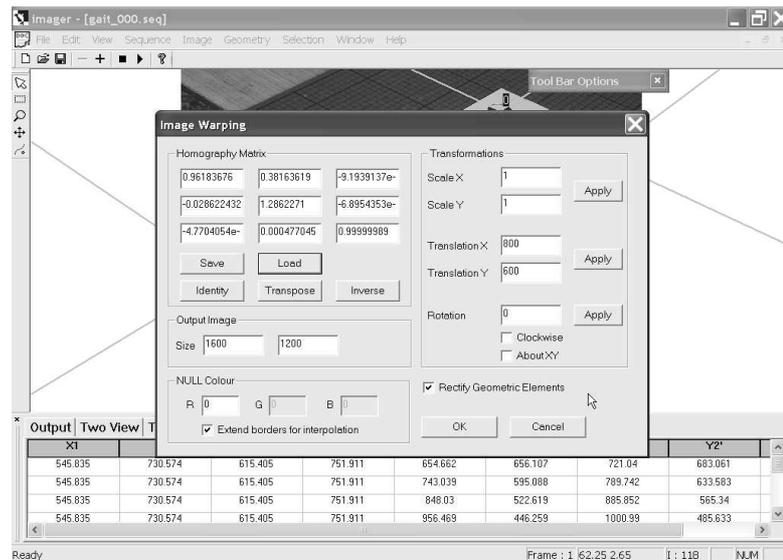


FIGURE D.2: Linear warping tool. Distortion of the image can be performed by specifying an arbitrary rectification homography.

Rectification of the geometric primitives within the image by the required planar transformation may also be performed. A number of sources can be used for the transformation homography. The two-view geometry toolbar maintains a number of special matrices that correspond to specific two-view mappings. Stereo point correspondences are determined by the epipolar geometry between views, which is encoded by the fundamental matrix \mathbf{F} . Similarly, point correspondences between the scene and image planes are determined by the homography mapping \mathbf{H} . These matrices can be computed from point correspondences alone, which can be manually picked within two selected views or from values entered directly into the correspondence table. We can then select this planar homography within the image distortion dialog box to perform the required image transformation.

The two-view dialog box also provides additional functionality to enable users to manipulate the set of stored transformation matrices. We can pre-multiply or post-multiply these matrices by an arbitrary 3×3 matrix. The toolbar also allows us to import and export point correspondences and sets of points from a variety of different text file formats. These points may then be added to any chosen image.

Primitives can also be used to form constraints on the elements of the rectification mapping. We can select geometric features, such as a line to compute the perspective transformation that maps it back to the ideal line, or a point in order to determine the stereopsis transformation that replaces it back to the ideal point. The stratified rectification toolbar also allows us to pick points from within the image that define a set of endpoint constraints. These endpoints are used to determine the affine transformation that restores the specified known length ratios. The computed rectification matrix can be used within the image distortion dialog to distort the image view. Figure D.3 shows the stratified rectification toolbar that allows us to enter the required algebraic constraints in order to compute the required rectification transformation.

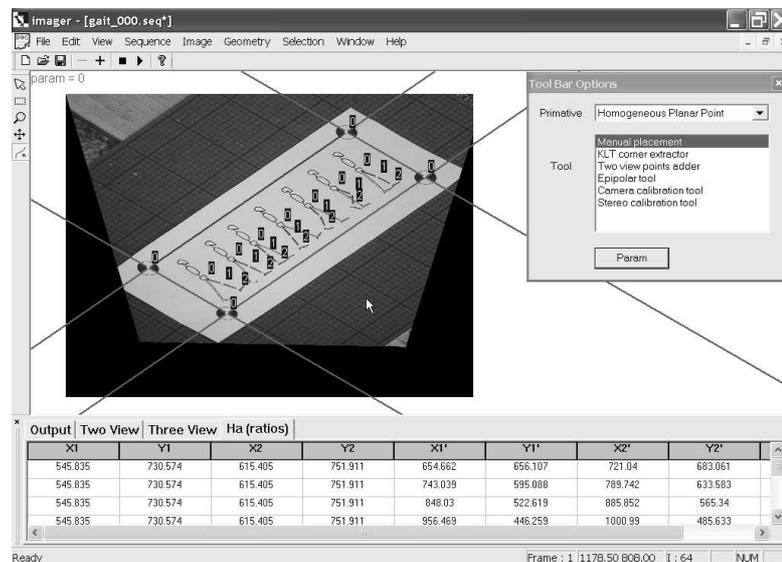


FIGURE D.3: Geometric stratified rectification constraints. Constraints formed from the known ratio of lengths between line segment endpoints can be identified by manually picking features from within the image.

D.3.3 Multiple view geometry

Throughout this project we need to be able to accurately measure a set of 3D points. In order to do this we must provide a method to calibrate an array of cameras. We have implemented Zhang's calibration technique [120], which requires that we mark imaged corner positions of a planar calibration target. Subsequently, we have built a tool that allows us to easily determine this set of points within each image, by clicking as close as possible the four corner points of the imaged calibration target. We have also automated this extraction process by using a technique based on the KLT corner detector to identify the calibration target point features. Consequently, the complete calibration algorithm can be automated and consists of four independent steps: i) Interest feature segmentation and detection; ii) Initial calibration; iii) Re-segmentation of interest features; iv) Final calibration. Figure D.4 shows a screen shot of the calibration toolbar and the corresponding output after performing the algorithm.

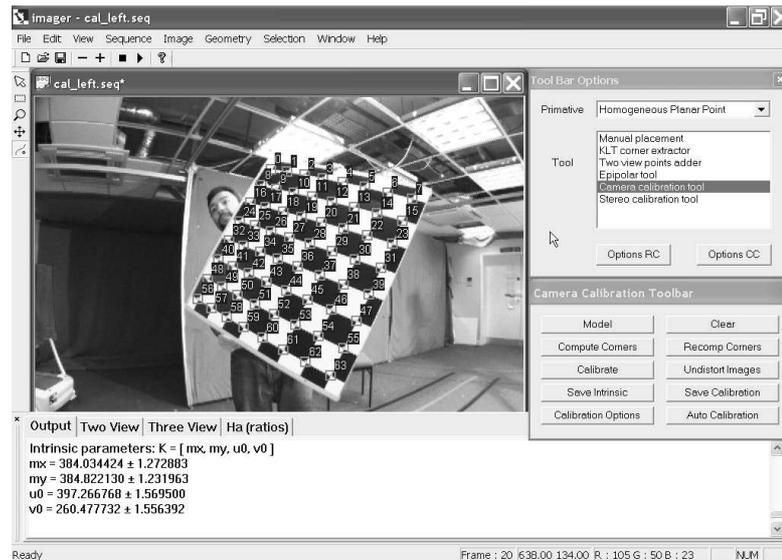


FIGURE D.4: Camera calibration tool. The tool allows the user to manually identify the image of the planar calibration target within each of the images of the sequence. The tool also fully automates the process of feature detection and camera calibration.

This calibration algorithm accurately determines the intrinsic parameters of the camera model. In order to determine workspace structure we must compute the extrinsic pose parameters between the cameras in the array. We have also implemented a stereo calibration algorithm that computes these extrinsic parameters between pairs of cameras. Larger numbers of cameras in the array can then be calibrated in a pairwise fashion in order to compute a consistent set of projection matrices. Further details of the calibration algorithm are outlined in section 2.5.3.

Once the set of camera intrinsic and extrinsic parameters are determined then image correspondences may be back projected to triangulate the set of workspace points. The three view toolbar allows us to manually pick point correspondences within three selected image views. These points are then added to the correspondence table and may be subsequently back projected. The toolbar also allows us to compute the correspondences within the camera views based on the set of integer identification numbers. This enables us to quickly triangulate the set of workspace points over the entire set of image frames. Figure D.5 illustrates the three view tool bar and the triangulation process.

After triangulation of the workspace structure we can view the set of workspace points by using the virtual viewing tool. This tool uses the OpenGL rendering interface to display a set of primitives that resemble the 3D points, planes and the set of cameras. Figure D.6 shows an image taken from the right camera view together with the corresponding virtual view, as seen from slightly behind the camera in order to show the placement of all the cameras in relation to the triangulated point set.

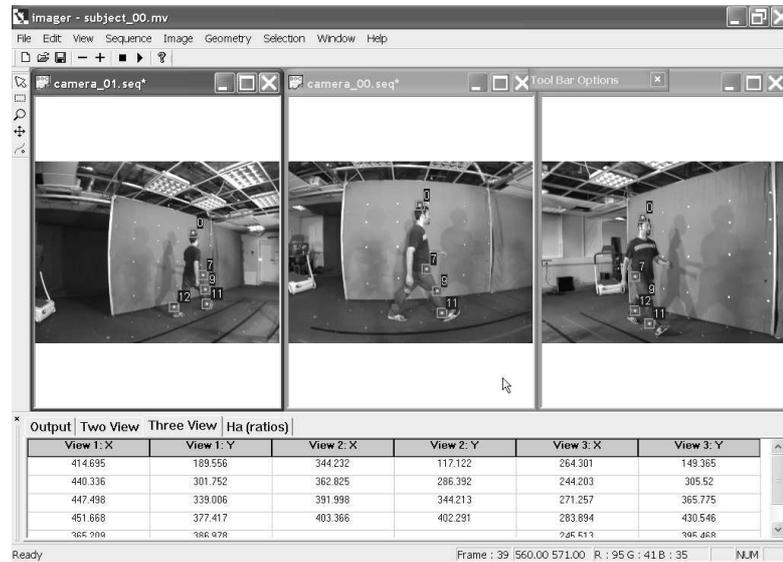


FIGURE D.5: Multiple view geometry. Point correspondences within two and three views can be manually selected in order to back project the set of rays that intersect in the worldspace points.

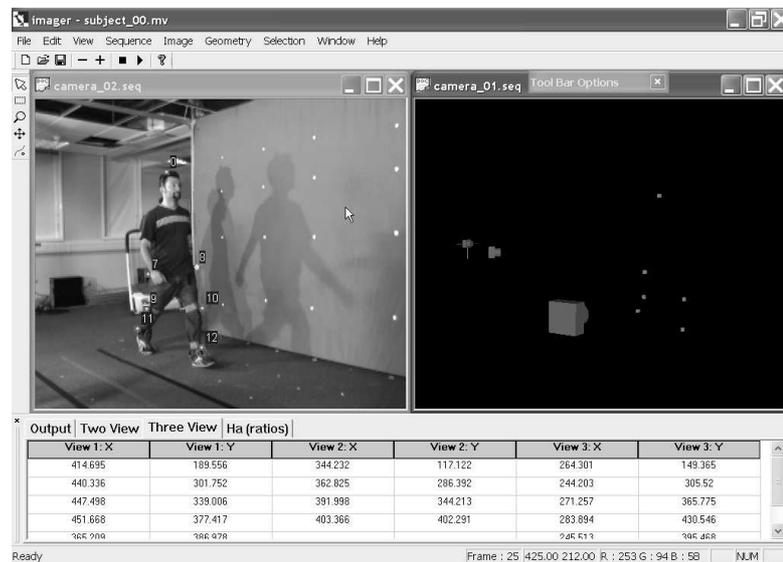


FIGURE D.6: Virtual view tool. The set of back projected worldspace points from the set of cameras can be viewed from a virtual camera. The position and orientation of the virtual camera can be controlled by using the mouse.

D.4 Gait system

The complete gait system requires that we input only the joint landmark features for each piecewise linear segment of subject motion. The set of outputs can be saved at each stage of the gait algorithm into simple text files. Consequently, no graphical interface is necessary and the gait system can be compiled as a simple command line program. Further analysis of the output data can be visualised later with the geometry tools. Figure D.7 shows: i) The single input text file to the system containing the set of subject image sequences, location of the output files and the system configuration parameters; ii) The set of computed output text files from the system.

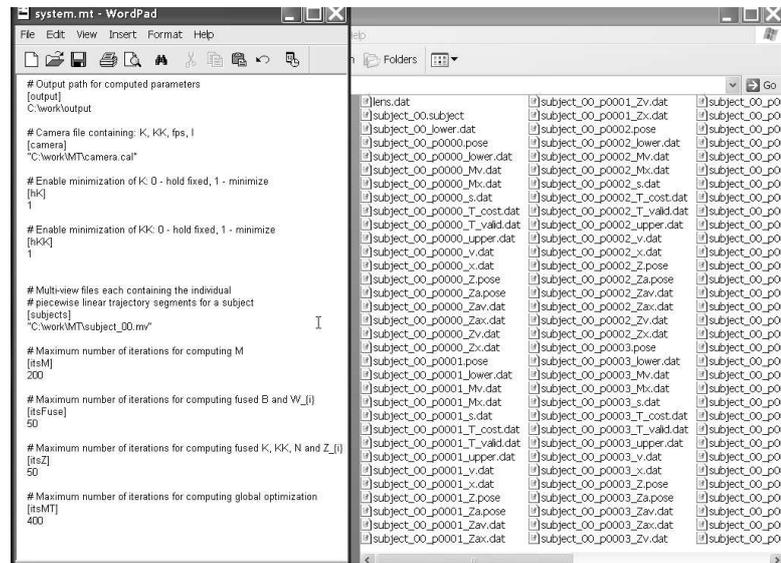


FIGURE D.7: Input (left) and output (right) from the gait algorithm.

The data and functionality of the complete gait system is best wrapped up within a single C++ class. The set of subject landmark position data matrices are stored within a tree structure in the class. The individual stages of the reconstruction algorithm are wrapped within a set of class methods. Each individual stage of the gait algorithm calls a number of virtual message handler routines: OnBeginXXX(), OnEndXXX() and OnFailXXX(). In the base implementation class these functions are empty, though they may be overridden in any further classes that inherit the base class. A number of experiments within this project are implemented by inheriting the behaviour of the base class and overriding the behaviour of the message handlers in order to output the required information to the text files. The gait system is then configured through a number of options in the configuration file and the complete reconstruction algorithm performed by a single call to the Run() method.

Bibliography

- [1] Pathokinesiology department, physical therapy department: Observational gait analysis handbook. The professional staff association of Rancho Los Amigos Medical center, Downey CA, 1989.
- [2] T. P. Andriacchi, J. A. Ogle, and J. O. Galante. Walking speed as a basis for normal and abnormal gait measurements. *Journal of biomechanics*, 10(4):261–268, 1977.
- [3] C. Angeloni, P. O. Riley, and E.D. Krebs. Frequency content of whole body gait kinematic data. *IEEE Trans. on Rehabilitation Engineering*, 2(1):40–46, 1994.
- [4] M. Armstrong, A. Zisserman, and R. Hartley. Self calibration from image triplets. In *Proc. European Conference on Computer Vision, LNCS 1064/5*, pages 3–16. Springer-Verlag, 1996.
- [5] C. BenAbdelkader. *Gait as a biometric for person identification in video sequences*. PhD thesis, University of Maryland, June 2002.
- [6] C. BenAbdelkader, R. Cutler, and L. Davis. Stride and cadence as a biometric in automatic person identification and verification. *5th International Conference on Automatic Face and Gesture Recognition*, 2002.
- [7] C. BenAbdelkader, R. Cutler, and L. Davis. View-invariant estimation of height and stride for gait recognition. *Workshop on Biometric Authentication ECCV*, 2002.
- [8] A. F. Bobick and A.Y. Johnson. Gait recognition using static activity-specific parameters. In *In Proceedings of Computer Vision and Pattern Recognition Conference (CVPR 2001)*, Kauai, Hawaii, December 2001.
- [9] D. A. Brannan, M. F. Esplen, and J. J. Gray. *Geometry*. Cambridge University Press, Cambridge CB2 2RU, UK, 1999.
- [10] B. Caprile and V. Torre. Using vanishing points for camera calibration. *IJCV*, 4(2):127–140, March 1990.

-
- [11] J. N. Carter and M. S. Nixon. On measuring gait signatures which are invariant to their trajectory. *Measurement and Control*, 32:265–269, 1999.
- [12] J. N. Carter and M. S. Nixon. On measuring trajectory invariant gait signatures. In *IAPRS*, volume XXXIII, Amsterdam, 2000.
- [13] R. Cipolla and P. Giblin. *Visual motion of curves and surfaces*. Cambridge University Press, Cambridge CB2 2RU, UK, 2000.
- [14] T. A. Clarke and J. G. Fryer. The development of camera calibration methods and models. *Photogrammetric Record*, 16(91):51–66, April 1998.
- [15] G. V. B. Cochran. *A primer of orthopaedic biomechanics*. Churchill Livingstone, New York, 1982.
- [16] A. Criminisi, I. Reid, and A. Zisserman. Single view metrology. In *Proc. 7th International Conference on Computer Vision*, pages 434–442, Kerkyra, Greece, September 1999.
- [17] D. Cunado, J. M. Nash, M. S. Nixon, and J.N. Carter. Gait extraction and description by evidence-gathering. In *2nd Proc. AVBPA*, pages 43–48, March 1999.
- [18] D. Cunado, M. S. Nixon, and J. N. Carter. Using gait as a biometric, via phase-weighted magnitude spectra. In *Lecture Notes in Computer Science (Proceedings of AVBPA97)*, pages 95–102, March 1997.
- [19] D. Cunado, M. S. Nixon, and J. N. Carter. Automatic gait recognition via model-based evidence gathering. In *AutoID '99: IEEE Workshop on Identification Advanced Technologies*, pages 27–30, 1999.
- [20] R. Cutler and L. Davis. View based detection and analysis of periodic motion. *Proc. International Conference on Pattern Recognition*, page P. Session A14, August 1998.
- [21] R. Cutler and L. Davis. Real-time periodic motion detection, analysis, and applications. *Computer Vision and Pattern Recognition*, pages 326–332, June 1999.
- [22] R. Cutler and L. Davis. Robust real-time periodic motion detection, analysis and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(2):129–155, 2000.
- [23] J. Davis. Visual categorization of children and adult walking styles. *International Conference on Audio and Video based Biometric Person Authentication*, 2001.
- [24] J. Davis and S. Taylor. Analysis and recognition of walking movements. In *In International Conference on Pattern Recognition (ICPR 2002)*, pages 315–318, Quebec, Canada, August 2002.

- [25] E. R. C. Draper. A treadmill-based system for measuring symmetry of gait. *Medical Engineering and Physics*, 22:215–222, 2000.
- [26] H. B. Falls and L. A. Humphrey. Energy cost of running and walking in young men. *Med sci sports*, 8:9–13, 1976.
- [27] O. Faugeras and Q. Luong. *The geometry of multiple images*. MIT press, Cambridge, Massachusetts, 2001.
- [28] O. D. Faugeras. *Three-dimensional computer vision: A geometric viewpoint*. MIT Press, 1993.
- [29] O. D. Faugeras. Stratification of three-dimensional vision: projective, affine, and metric representation. *J. Opt. Soc. Am*, A12:465–484, 1995.
- [30] O. D. Faugeras and S. J. Maybank. Motion from point matches: Multiplicity of solutions. *International Journal of Computer Vision*, 4:255–246, 1990.
- [31] O. D. Faugeras, L. Quan, and P. Sturm. Self calibration of a 1D projective camera and its application to the self calibration of a 2D projective camera. In *Proc. European Conference on Computer Vision*, pages 36–52, 1998.
- [32] F. R. Finley and K.A. Cody. Locomotive characteristics of urban pedestrians. *Arch Phys med rehabil*, 51:423–426, 1970.
- [33] J.D. Foley, A. Van Dam, S.K. Feiner, and J.F. Hughes. *Computer Graphics: Principles and practice*. Addison-Wesley, 1990.
- [34] J. P. Foster, M. S. Nixon, and A. Prugel-Bennett. New area based metrics for gait recognition. In *Proc. 3rd International Conference on Audio and Video Based Biometric Person Authentication*, pages 312–317, 2001.
- [35] H. Fujiyoshi and A. Lipton. Real-time human motion analysis by image skeletonization. In *Proc. IEEE Workshop Applications of Computer Vision*, page P. Session 1A, October 1998.
- [36] D. M. Gavrila. The visual analysis of human movement: A survey. *Computer Vision and Image Understanding: CVIU*, 73(1):82–98, 1999.
- [37] N. Goddard. The interpretation of visual motion: Recognizing moving light displays. In *Proc. IEEE Workshop motion*, pages 212–220, March 1989.
- [38] G. H. Golub and C. F. Van Loan. *Matrix computations*. The John Hopkins University Press, Baltimore, MD, second edition, 1989.
- [39] M. Han and T. Kanade. Multiple motion scene reconstruction with uncalibrated cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(7):884–894, July 2003.

- [40] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, Cambridge CB2 2RU, UK, 2000.
- [41] R. I. Hartley. Euclidean reconstruction from uncalibrated views. In J. L. Mundy, A. Zisserman, and D. Forsyth, editors, *Proc. 2nd European-US Workshop on Invariance*, pages 187–202, Azores, 1993.
- [42] R. I. Hartley. An algorithm for self calibration from several views. In *Proc. IEEE conference on computer vision and pattern recognition*, pages 908–912, Seattle, WA, June 1994.
- [43] J. B. Hayfron-Acquah, M. S. Nixon, and J. N. Carter. Automatic gait recognition by symmetry analysis. In *Proc. 3rd International Conference on Audio and Video Based Biometric Person Authentication*, pages 272–277, June 2001.
- [44] J. B. Hayfron-Acquah, M. S. Nixon, and J. N. Carter. Recognising human and animal movement by symmetry. In *International Conference on Image Processing, ICIP*, pages 290–293, October 2001.
- [45] J. Heikkila and O. Silven. A four-step camera calibration procedure with implicit image correction. In *in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico*, page 11061112, 1997.
- [46] B. Heisele and C. Wohler. Motion based recognition of pedestrians. In *Proc. International Conference on Pattern Recognition*, August 1998.
- [47] S. M. Hsiang and C. Chang. The effect of gait speed and load carrying on the reliability of ground reaction forces. *Safety Science*, 20(7-8):639–657, 2002.
- [48] P. S. Huang, C. J. Harris, and M. S. Nixon. Human gait recognition in canonical space using temporal templates. *IEE Proceedings Vision Image and Signal Processing*, 146(2):93–100, 1999.
- [49] P. S. Huang, C. J. Harris, and M. S. Nixon. Recognizing humans by gait via parametric canonical space. *Journal of Artificial Intelligence in Engineering*, 13:359–366, 1999.
- [50] V. T. Inman, H. J. Ralston, and F. Todd. *Human Walking*. Williams and Wilkins Company, Baltimore, MD, 1981.
- [51] A. Jain, R. Bolle, and S. Pankanti, editors. *Biometrics: Personal identification in a networked society*. Kluwer Academic Publishing, January 1999.
- [52] G. Johansson. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, 14(2):201–211, 1973.
- [53] G. Johansson. Visual motion perception. *Scientific Am*, 232:75–88, June 1976.

- [54] A. Y. Johnson and A. F. Bobick. A multi-view method for gait recognition using static body parameters. In *The 3rd International Conference on Audio and Video Based Biometric Person Authentication*, 2001.
- [55] A. Kale, R. Chowdhury, and R. Chellappa. Towards a view invariant gait recognition algorithm. In *Proc. IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS'03)*, pages 143–150, 2003.
- [56] C. Kirtley, M. W. Whittle, and R. J. Jefferson. Influence of walking speed on gait parameters. *Journal of Biomedical Engineering*, 7(4):282–288, 1985.
- [57] L. T. Kozlowski and J. E. Cutting. Recognizing the sex of a walker from a dynamic point-light display. *Perception and Psychophysics*, 21(6):575–580, 1977.
- [58] J. L. Lelas, G. J. Merriman, P. O. Riley, and D. C. Kerrigan. Predicting peak kinematic and kinetic parameters from gait speed. *Gait and Posture*, June 2002.
- [59] L. Li, E. C. H. van den Bogert, G. E. Caldwell, R. E. A. van Emmerik, and J. Hamill. Coordination patterns of walking and running at similar speed and stride frequency. *Movement Science*, 18:67–85, 1999.
- [60] D. Liebowitz. *Camera calibration and reconstruction of geometry from images*. PhD thesis, University of Oxford, Dept. Engineering Science, June 2001. D.Phil. thesis.
- [61] D. Liebowitz and S. Carlsson. Uncalibrated motion capture exploiting articulated structure constraints. In *Proc. 8th International Conference on Computer Vision*, Vancouver, Canada, 2001.
- [62] D. Liebowitz and A. Zisserman. Metric rectification for perspective images of planes. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 482–488, Santa Barbara, CA, U.S.A., June 1998.
- [63] D. Liebowitz and A. Zisserman. Combining scene and auto-calibration constraints. In *Proc. 7th International Conference on Computer Vision*, pages 434–442, Kerkyra, Greece, September 1999.
- [64] J. Little and J. Boyd. Describing motion for recognition. In *Proceedings of International Symposium on Computer Vision*, pages 235–240, Coral Gables, FL USA, November 1995.
- [65] J. Little and J. Boyd. Recognising people by their gait: The shape of motion. In *Videre*, volume 1, pages 1–32, 1998.
- [66] F. Liu and R. Picard. Detecting and segmenting periodic motion. Technical Report 400, M.I.T Media Laboratory Perceptual Computing Section, 1996.

- [67] F. Liu and R. Picard. Finding periodicity in space and time. In *International Conference on Computer Vision*, pages 376–383, January 1998.
- [68] L. Ma, Y. Chen, and K. Moore. An analytical piecewise radial distortion model for precision camera calibration. *CoRR: Computer vision and pattern recognition*, 2003.
- [69] G. Mather and L. Murdoch. Gender discrimination in biological motion displays based on dynamic cues. In *Proc. Royal Society*, volume 258, pages 273–279, 1994.
- [70] A. Matsas, N. Taylor, and H. McBurney. Knee joint kinematics from familiarised treadmill walking can be generalised to overground walking in young unimpaired subjects. *Gait and Posture*, 11:46–53, 2000.
- [71] S. J. Maybank. Theory of reconstruction from image motion. *Springer-Verlag, Berlin*, 1993.
- [72] S. J. Maybank and O. D. Faugeras. A theory of self-calibration of a moving camera. *The international journal of computer vision*, 8(2):123–152, August 1992.
- [73] L. Middleton, S. C. Wong, M. O. Jewell, J. N. Carter, and M. S. Nixon. A middleware for a large array of cameras. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3682–3687, Edmonton, Canada, 2005.
- [74] J. L. Mundy and A. Zisserman. *Geometric invariance in computer vision*. MIT press, Cambridge, Massachusetts, 1992.
- [75] M. P. Murray. Gait as a total pattern of movement. *American Journal of Physical Medicine*, 46(1):290–332, 1967.
- [76] M. P. Murray, A. B. Drought, and R. C. Kory. Walking patterns of normal men. *Journal of Bone Joint Surgery*, 46-A(2):335–360, 1964.
- [77] B. M. Nigg and W. Herzog, editors. *Biomechanics of the musculo-skeletal system*. John Wiley and Sons, University of Calgary, Calgary, Alberta, Canada, second edition, 1999.
- [78] M. S. Nixon, J. N. Carter, D. Cunado, P. S. Huang, and S. V. Stevenage. Automatic gait recognition. In A. Jain, R. Bolle, and S. Pankanti, editors, *Biometrics: Personal identification in a networked society*, pages 231–250. Kluwer Academic Publishing, January 1999.
- [79] S. A. Niyogi and E. H. Adelson. Analyzing and recognizing walking figures in XYT. In *CVPR*, pages 469–474, 1994.
- [80] S. A. Niyogi and E. H. Adelson. Analyzing gait with spatiotemporal surfaces. In *Proc. IEEE Workshop Motion of Non-Rigid and Articulated Objects*, pages 64–69, 1994.

- [81] Atos Origin. UK Passport Service biometrics enrolment trial, May 2005. <http://www.passport.gov.uk>.
- [82] J. C. Otis and A. H. Burstein. Evaluation of the VA-Rancho gait analyzer, Mark I. *Bull Prosthetics Research*, 18(1):21–25, 1981.
- [83] C. Peham, T. Licka, A. Mayr, M. Scheidl, and D. Girtler. Speed dependency of motion pattern consistency. *Journal of Biomechanics*, 31(9):769–772, 1998.
- [84] J. Perry. *Gait Analysis: Normal and pathological function*. SLACK Incorporated, Rancho Los Amigos medical center, Downey, CA, 1992.
- [85] R. Polana and R. Nelson. Detecting activities. In *Proc. CVPR*, pages 2–7, New York, NY, June 1993.
- [86] R. Polana and R. Nelson. Detection and recognition of periodic non-rigid motion. In *International Conference on Computer Vision*, volume 23, pages 261–282, June 1997.
- [87] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling. *Numerical recipes in C*. Cambridge University Press, 1988.
- [88] JBdeCM Saunders, V. T. Inman, and H. D. Eberhart. The major determinants in normal and pathological gait. *Journal of Bone Joint Surgery*, 35-A(3):543–557, 1953.
- [89] F. Schaffalitzky and A. Zisserman. Geometric grouping of repeated elements within images. In *Proc. BMVC*, pages 13–22, 1998.
- [90] F. Schaffalitzky and A. Zisserman. Planar grouping for automatic detection of vanishing lines and points. *Image and Vision Computing*, 18:647–658, 2000.
- [91] S. M. Seitz and C. R. Dyer. View invariant analysis of cyclic motion. *International Journal of Computer Vision*, 25(3):1–23, 1997.
- [92] J.G. Semple and G.T. Kneebone. *Algebraic Projective Geometry*. Oxford University Press, 1979.
- [93] N. M. Spencer and J. N. Carter. Viewpoint invariance in automatic gait recognition. In *Proc. Third IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'02)*, pages 1–6, 2002.
- [94] G. Stein. Accurate internal camera calibration using rotation, with analysis of sources of error. In *Proc. Fifth international conference on computer vision*, pages 230–236, Cambridge, Massachusetts, June 1995.
- [95] G. M. Strathy, E. Y. Chao, and R. K. Laughman. Changes in knee function associated with treadmill ambulation. *Journal of Biomechanics*, 16(7):517–522, 1983.

- [96] P. Sturm. *Vision 3D non calibrée: Contributions à la reconstuction projective et étude des mouvements critiques pour l'auto calibrage*. PhD thesis, INRIA Rhone-Alpes, 1997.
- [97] P. Sturm and S. Maybank. On plane based camera calibration: A general algorithm, singularities, applications. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 432–437, June 1999.
- [98] I.E. Sutherland. Sketchpad: A man-machine graphical communications system. Technical report 296. Technical report, MIT Lincoln Laboratories, 1963.
- [99] R. Szeliski. Image mosaicing for tele-reality applications. Technical report, Digital Equipment Corporation, Cambridge, USA, 1994.
- [100] R. Szeliski and S. Heung-Yeung. Creating full view panoramic image maosaics and environment maps. In *Proc. ACM SIGGRAPH Conference on computer graphics*, 1997.
- [101] R. Tanawongsuwan and A. F. Bobick. Gait recognition from time normalized joint angle trajectories in the walking plane. In *In Proceedings of Computer Vision and Pattern Recognition Conference (CVPR 2001)*, Kauai, Hawaii, December 2001.
- [102] R. Tanawongsuwan and A. F. Bobick. Characteristics of time-distance gait parameters across speeds. Technical report, Graphics, Visualization, and Usability Center, Georgia Institute of Technology, Atlanta, GA, 2003. GIT-GVU-03-01.
- [103] R. Tanawongsuwan and A. F. Bobick. Performance analysis of time-distance gait parameters under different speeds. In *In 4th International Conference on Audio and Video Based Biometric Person Authentication (AVBPA 2003)*, University of Surrey, Guildford, UK, June 2003.
- [104] C. J. Taylor. Reconstruction of articulated objects from point correspondences in a single uncalibrated image. *Computer Vision and Image Understanding*, 80(3):349–363, 2000.
- [105] A. Thorstensson, J. Nilsson, H. Carlson, and M. R. Zomlefer. Trunk movements in human locomotion. *Acta Physiol Scand*, 121:9–22, 1984.
- [106] A. Thorstensson and H. R. Roberthson. Adaptations to changing speed in human locomotion: speed of transition between walking and running. *Acta Physiol scand*, 131:211–214, 1987.
- [107] W. Triggs. Autocalibration from planar scenes. In *Proc. 5th European Conference on Computer Vision*, pages 89–105, Freiburg, Germany, June 1998.
- [108] P. Tsai, M. Shah, K. Keiter, and T. Kasparis. Cyclic motion detection for motion based recognition. *Pattern recognition*, 27(12):1591–1603, 1994.

- [109] R. Y. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off the shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):323–344, 1987.
- [110] R. L. Waters, B. R. Lunsford, J. Perry, and R. Byrd. Energy-speed relationship of walking: standard tables. *Journal of orthopedic research*, 6:215–222, 1988.
- [111] R. L. Waters and J. M. Morris. Electrical activity of muscles of the trunk during walking. *Journal of Anatomy*, 111(2):191–199, 1972.
- [112] G. Wei and S. Ma. A complete two-plane camera calibration method and experimental comparisons. In *Proc. Fourth international conference on computer vision*, pages 439–446, Berlin, May 1993.
- [113] G. Wei and S. Ma. Implicit and explicit camera calibration: Theory and experiments. *IEEE Transactions on pattern analysis and machine intelligence*, 16(5):469–480, 1994.
- [114] J. Weng, P. Cohen, and M. Herniou. Camera calibration with distortion models and accuracy evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(10):965–980, October 1992.
- [115] S. C. White, H. J. Yack, C. A. Tucker, and H. Y. Lin. Comparison of vertical ground reaction forces during overground and treadmill walking. *Med Sci Sports Exerc*, 30:1537–1542, 1998.
- [116] C. Y. Yam. *Model based approaches for recognising people by the way they walk or run*. PhD thesis, University of Southampton, Dept. Electronics and computer science, November 2002. D.Phil. thesis.
- [117] C. Y. Yam, M. S. Nixon, and J. N. Carter. Extended model based automatic gait recognition of walking and running. In *3rd Proc. AVBPA 2001*, pages 278–283, June 2001.
- [118] C. Y. Yam, M. S. Nixon, and J. N. Carter. Gait recognition by walking and running: A model based approach. In *Proc. Asian Conference on Computer Vision, ACCV*, pages 1–6, January 2002.
- [119] C. Y. Yam, M. S. Nixon, and J. N. Carter. On the relationship of human walking and running: Automatic person identification by gait. In *IEEE Proc. 16th International Conference on Pattern Recognition*, volume 1, August 2002.
- [120] Z. Zhang. A flexible new technique for camera calibration. In *Proc. 7th International Conference on Computer Vision*, Kerkyra, Greece, September 1999.
- [121] A. Zisserman, D. Liebowitz, and M. Armstrong. Resolving ambiguities in auto-calibration. *Philosophical Transactions of the Royal Society of London, Series A*, 356(1740):1193–1211, 1998.