06241 Abstracts Collection
Human Motion - Understanding, Modeling, Capture and Animation. 13th Workshop "Theoretical Foundations of Computer Vision"
— Dagstuhl Seminar —

Bodo Rosenhahn[1], Reinhard Klette[2], Dimitris Metaxas[3]

[1] MPI für Informatik, DE
rosenhahn@mpi-inf.mpg.de
[2] Univ. of Auckland, NZ
r.klette@auckland.ac.nz
[3] Rutgers Univ. - Piscataway, US
dnm@cs.rutgers.edu

Abstract. From 11.06.06 to 16.06.06, the Dagstuhl Seminar 06241 "Human Motion - Understanding, Modeling, Capture and Animation. 13th Workshop "Theoretical Foundations of Computer Vision"" was held in the International Conference and Research Center (IBFI), Schloss Dagstuhl. During the seminar, several participants presented their current research, and ongoing work and open problems were discussed. Abstracts of the presentations given during the seminar as well as abstracts of seminar results and ideas are put together in this paper. The first section describes the seminar topics and goals in general.

Keywords. Computer Vision, Computer Graphics, Biomechanics

06241 Summary – Human Motion - Understanding, Modeling, Capture and Animation. 13th Workshop

Reinhard Klette, Dimitris Metaxas and Bodo Rosenhahn

Modeling, tracking and understanding of human motion based on video sequences is a field of research of increasing importance, with applications in sports sciences, medicine, biomechanics, animation (avatars), surveillance, and so forth. Progress in human motion analysis depends on research in computer graphics, computer vision and biomechanics. Though these fields of research are often treated separately, human motion analysis requires an interaction of computer graphics with computer vision, which also benefits from an understanding of biomechanic constraints. This seminar brought together specialists and students from these disciplines, studying and contributing to the subject of human motion analysis from different perspectives. The interdisciplinary character of the

Dagstuhl Seminar Proceedings 06241
Human Motion - Understanding, Modeling, Capture and Animation. 13th Workshop "Theoretical Foundations of Computer Vision"
http://drops.dagstuhl.de/opus/volltexte/2006/721

seminar allowed to bring people together which normally would not have met at disciplinary conferences.

Eadweard Muybridge (1830-1904) is known as the pioneer in motion capturing with his famous experiments in 1887 called "Animal Locomotion" (Do all feet leave the ground during the gallop of a horse? He used photography to answer the question.) The field of animal or human motion analysis has developed into many directions since then. However, human-like animation and recovery of motion is still far from being satisfactory. Various groups are dealing with different aspects of modeling, estimation and animation of human motions. Motivations differ, and define directions of research. Examples of motivations are the analysis of movements for disease detection (hip dislocations, knee injuries etc.), sports movement optimization (ski or high jumping, golf playing, swimming, etc.), the animation of avatars in movies (e.g. Gollum in Lord of the Rings), or the realistic character animation in computer games.

New results and specific research strategies have been discussed at this seminar to approach this highly complex field. The seminar intention was to discuss theoretical fundamentals related to those issues and to specify open problems and major directions of further development in the field of human motion related to computer vision, computer graphics or biomechanics. The seminar schedule was characterised by flexibility, working groups, and sufficient time for focused discussions. The participants of this seminar enjoyed the atmosphere and the services at Dagstuhl very much. The quality of this center is unique.

There will be an edited book (within Springer's series on Computational Imaging) following the seminar, and all seminar participants have been invited to contribute with chapters. The deadline for those submissions is in September 2006 (allowing to incorporate results or ideas stimulated by the seminar), and submissions will be reviewed (as normal). Expected publication date is the end of 2007 or early 2008.

Imitation Learning of Human Movement and Hand Grabbing to Adapt to Environment Changes

Stephan Al-Zubi (Universität Kiel, D)

We propose a model for learning the articulated motion of human arm and hand grabbing. The goal is to generate plausible trajectories of joints that mimic the human movement using deformation information. The trajectories are then mapped to a constraint space. These constraints can be the space of start and end configuration of the human body and task-specific constraints such as avoiding an obstacle, picking up and putting down objects.

Such a model can be used to develop humanoid robots that move in a human-like way in reaction to diverse changes in their environment and as a priori model for motion tracking. The model proposed to accomplish this uses a combination of principal component analysis (PCA) and a special type of a topological map called the dynamic cell structure (DCS) network. Experiments on arm and hand movements show that this model is able to successfully generalize movement using a few training samples for free movement, obstacle avoidance and grabbing objects. We also introduce a method to map the learned human movement to a robot with different geometry using reinforcement learning and show some results.

Contours and Optical Flow: Cues for Capturing Human Motion in Videos

Thomas Brox (Universität Bonn, D)

Capturing human motion from video data is a research topic with many applications. While the industrial systems applied in practice rely on tracking markers attached to a person, research meanwhile focuses on tracking features that naturally appear in videos. Besides local patches, contours and the optical flow are such features that allow for tracking the pose of objects and humans. Highlighting here the usage of contours and optical flow, it will be shown how these cues can be extracted reliably from the image. Avoiding strict assumptions like a static background, a region based active contour that includes the tracked object model as a shape prior is a good method to extract the object silhouette from the image. Supplementing the optic flow derived with a method that allows for large motions as a complementary feature yields a rather general and robust system for human tracking.

Keywords: Segmentation, optical flow, pose estimation

Visualisation of Athletes Torque Generation Ability during Handbike and Rowing Propulsion

Harald Böhm (TU München, D)

INTRODUCTION

Inverse dynamics computation of constraint forces and resultant muscle torques in dynamic movements are included in the most motion analysis software programs. Visualization of calculated kinematics and kinetics and their presentation by means of three dimensional human avatars are commonly used. Recently the knowledge about internal muscle properties has been increased and is very useful to evaluate sport equipment design as well as the ergonomics of different human movements. To our best knowledge there exists no implementation of muscular properties in commercial motion analysis software packages.

Therefore the aim of this study is to demonstrate the use of muscular properties studying two different propulsive movement techniques.

METHOD

Isometric muscular joint torques of the elbow and shoulder joints have been measured within their complete range of motion. A 3D motional analysis on a rowing ergometer and a handbike was carried out during maximal power output of two athletes.

Joint angles and angular velocities of the elbow and shoulder joints were calculated according to the ISB standard. To evaluate the maximum power output of rowing and handbiking propulsion their motion trajectories were compared regarding their torque generation ability.

RESULTS

In rowing the torque potential at the elbow joint is up to 20% higher than in the handbike movement. This is mainly caused by a greater elbow flexion. The shoulder did not show differences between both movements, this is caused by the plateau region of the isometric joint torque curve within the typical range of movement.

DISCUSSION

The demonstrated evaluation of motion analysis data shows that rowing is more effective concerning the maximum torque generation abilities of the athlete. However the results are limited to a two dimensional representation of the data, three dimensional problems based on three joint angles as well as additional dependence on discomfort parameters measured over the range of motion (Bubb 2000) leads to a multidimensional consideration, which requires sophisticated knowledge in computer graphics combined with creative ideas to visualize the acquired motion data in an understandable way.

See also: Bubb, H.; Estermann, S. Influence of Forces on Comfort Feeling in Vehicles, SAE Conference on Digital Human Modeling, Dearborn Ed.: SAE International. Warrendale, USA: SAE International, 2000

Joint work of: Böhm, Harald; Krämer, Christian

THE EUCLID CODE - Using Geometric Algebra for Motion Computations

Leo Dorst (University of Amsterdam, NL)

The geometric algebra of a metric vector space has *rotors*, which are manifestly structure preserving representations of orthogonal transformations. They can do what quaternions do for IR^3, but in a fully real manner. They are exponentials of bivectors, and having a logarithm permits straightforward interpolation of rotations.

The fairly new *conformal model* of Euclidean geometry gives a rotor representation of rigid body motions (and even of arbitrary conformal transformations). Now the logarithm permits interpolation between such motions, robust estimators, and the general use of linear techniques for general motion computations.

The talk is illustrated with interactive visualization software, also used to generate the figures in our upcoming book.

The role of Manifold learning in Human Motion Analysis

Ahmed Elgammal (Rutgers Univ. - Piscataway, USA)

Human body is an articulated object with high degrees of freedom.

Despite the high dimensionality of the configuration space, many human motion activities lie intrinsically on low dimensional manifolds.

Although the intrinsic body configuration manifolds might be very low in dimensionality, the resulting appearance manifold is challenging to model given various aspects that affects the appearance such as the shape and appearance of the person performing the motion, or variation in the view point, or illumination.

In this talk we present the results of a three year NSF funded project on learning nonlinear spatiotemporal models for human motion. Our objective is to learn representations for the shape and the appearance of moving (dynamic) objects that support tasks such as synthesis, pose recovery, reconstruction, and tracking. We studied various approaches for representing global deformation manifolds that preserve their geometric structure. Given such representations, we can learn generative models for dynamic shape and appearance. We also address the question of separating style and content on nonlinear manifolds representing dynamic objects. We learn decomposable generative models that explicitly decompose the intrinsic body configuration (content) as a function of time from the appearance/shape (style factors) of the person performing the action as time-invariant parameters. We show results on pose recovery, body tracking, gait recognition, as well as facial expression tracking and recognition.

Learning a Dynamic Independent Pose Distribution within a Bayesian Framework

Jürgen Gall (MPI für Informatik - Saarbrücken, D)

Capturing motion can be described by two steps in a Bayesian framework. Having an estimate of the pose in the current frame, the pose for the next frame is predicted by a dynamic model. Afterwards the estimate is updated according to the next frame. For human motion, simplified models of the dynamics are mostly used because of the complexity of the dynamics. These models do not integrate prior knowledge like anatomical constraints, self-intersection and unrealistic joint configurations in an appropriate way. We propose to integrate prior knowledge via an a priori pose distribution independent of the dynamic model. The pose distribution is learned from training samples using a Parzen-Rosenblatt estimator with a weighted Euclidean distance measure. Human motion capturing with an annealed particle filter demonstrates the benefits of the prior knowledge. The a priori distribution can be combined with various dynamic models and can be regarded as a soft constraint that penalizes unrealistic and improbable poses. Moreover, it is not restricted to motion patterns of the training data and thus more flexible than learning the dynamics.

Empowering Human-Computer Interactions with Segmented Motion

Richard Green (University of Canterbury - Christchurch, NZ)

One of the biggest hurdles in human-computer interaction is the current inability for computers to recognize human activities. Our studies suggest that such complex articulated motion differs significantly from what one might expect. We introduce hierarchical Bayesian models of concurrent movement structures for temporally segmenting this complex motion. These motion segments are then used for recognizing activities to enable applications to extend augmented reality and novel interactions with computers.

MoCap for Interaction Environments

Daniel Grest (Universität Kiel, D)

The accuracy marker-less motion capture systems is comparable to marker based systems, however the segmentation step makes strong restrictions to the capture environment, e.g. homogeneous clothing or background, constant lighting, etc. In interaction environments the background is non-static, cluttered and lighting changes often and rapidly. While the motion capture in such an environment has to be robust and must provide results in real-time, the accuracy of captured motion is not of great importance.

Stereo algorithms can provide data that is robust with respect to lighting and background and are available in real-time. The presented approach estimates arm movement with 5-6fps from up to 1000 correspondences by an optimized correspondence search using openGL and associative arrays.

Because speed is an issue, different optimization methods are compared, namely Gauss-Newton(Levenberg-Marquardt), Gradient Descent, and Stochastic Meta Descent. Experiments on human movement show the advantages and disadvantages of each method.

Additional cues like silhouette information and 2D corner tracking allow the estimation of complex movements even from a single view.

Keywords: Markerless Motion Capture, Pose Estimation, Real-Time Processing

The William Harvey Code: 3D Optical Flow computation and its Application to Cardiac Motion Analysis

Atsushi Imiya (Chiba University, J)

This talk focuses on the possibility to extract motion on the surface of beating heart using optical flow technique in computer vision.

Optical flow is a fundamental technique as a bases of motion tracking and motion segmentation. From 80?s methodologies of computer vision have been applied to medical image analysis and optical flow is applied to cardiac motion analysis first to 3D x-ray sine tomography, second 2D ultrasonic images, and then today motion analysis on gated cardiac MRI images is a challenging problem. Once a sequence of optical flow field is accurately and robustly computed from a sequence of images, it is possible to extract dynamic features form cardiac images as fundamental features both for biological and clinical data.

In the first part of the talk, I derived an accurate and robust optical flow computation algorithm employing the numerical computation of PDE derive as a method to solve optimization problem to compute optical flow.] In the second part, I derive a new linear image based anisotropic constrain for the optical flow computation, which allows us to compute the motion on the surface of the beating heart.

In the third part of the talk, I analyze incompressible condition of mechanics for the cardio optical flow computation. Then, I derive numerically a convergence condition for optical flow computation with a mechanical constrain.

Finally as a perspective, I propose some mechanical conditions which are suitable for the cardio optical flow computation.

Perspectives for Marker-less Tracking in Sports Applications

Uwe Kersting (University of Auckland, NZ)

Marker-based kinematic data collection is used for movement capture in laboratory settings. Inherent problems include adequate marker placement, skin movement artifacts or marker occlusion in certain body orientations. In outdoor settings the use of markers may be limited. A possible alternative are marker-less image-based motion tracking systems. We have proposed free-form surface patches to estimate segment orientations (1) as well as global and local morphing techniques (2). Further features include an advanced image segmentation method, dynamic occlusion handling and the inclusion of kinematic chains of higher complexity.

The aim of this study was to apply a marker-less tracking system to full body movements in sports. Marker data was recorded simultaneously to compare resulting kinematics to a commercially available marker-based tracking system.

A digital four camera system was used (Basler A602f, 180 Hz, SIMI motion). Tests were carried out on an outdoor sports field. To attain optimal contrast subjects had to wear tight white full-body suits. On top of the suite dark gray markers were placed on anatomical landmarks according to an existing upper body model (3). Marker data was tracked semi-automatically using SIMI software (Motion 7.0). Joint coordinate systems and joint motion were calculated using a customized Matlab script (The MathWorks 7.1). The marker-less system consists of three steps: segmentation, correspondence estimation and pose estimation as outlined in greater detail in (4). A series of experiments was conducted to compare marker-based and marker-less motion capture data In the lab-environment average differences between marker-based and silhouette-based joint angles of the upper extremity were less than 2.0 degrees for flexion-extension, while lower extremity values varied up to 3.4 degrees depending on the movements under investigation. However, in the outdoor environment larger deviations occurred. We think that improved body models and adjustments to the camera positioning will enable us to reduce discrepancies between marker-based and marker-less systems. It will be straight forward to apply simple centre of mass calculations to estimate energy exchange during jumps. Other perspectives for this are fast feedback sports technique evaluations.

Human Motion (Understanding, Modeling, Capture and Animation) - An Introduction

Reinhard Klette (University of Auckland, NZ)

The subject was first studied in arts (see, for example, static bodily poses of human bodies by ancient Greeks and Romans), leading to expressions of individual, collective (e.g., poses characterizing a married couple), or dynamic postures.

The 2D-3D pose estimation problem, as defined in computer vision (e.g., Grimson 1990) is based on the assumption of a (static) 3D model of an object.

Modeling of human bodies was again initiated in arts. For example, Leonardo da Vinci (1452-1519) discussed laws of static or kinematic chains (actually, kinematic trees) characterizing human motions such as walking up a stair. He introduced parameterized models for parts of the human body. Silhouettes are a way to match models with observed 2D projections, and the talk illustrated the incompleteness of information contained in a silhouette. Poses may also not only have no purpose, but even attempt to hide a purpose. Together with the diversity of human shape it is suggested to define particular directions of pose recognition. – Motion analysis is today mostly limited to the use of image sequences or video data. Further sensors are expected to emerge. Systematic studies of human locomotion date back to Muybridge, Marey and others at the end of the 19th century. Today we attempt to use motion analysis results also in medicine (e.g., clinical gait analysis), sports (e.g., enforcing rules about allowed bodily motions, modeling a multi-player game such as soccer, or analyzing motions of a golf player), technology (e.g., modeling 3D motion of a dummy during a crash test, or understanding traffic scenes), arts and human expressions (e.g., analyzing finger movements when playing a historic musical instrument, or when doing a string figure). Dynamic 3D shape reconstruction alone is not yet motion modeling or understanding. The diversity of potential applications defines different situations and prerequisites for used methods and technology (e.g., cameras with high frame rate), and it is expected that the field will develop within the next few years into a very vivid research area, combining results from biomechanics, computer vision, and computer graphics.

Recognition and Synthesis of Actions using a Dynamic Bayes Network

Volker Krüger (Aalborg University Ballerup, DK)

Our presentation consisted of two parts: In a first part we presented the new EU effort on Cognitive systems, PACO-PLUS.

PACP-PLUS is concerned with the goal of building a physically instantiated system that is able to perceive, learn and act in a real environment. PACO-PLUS aims at the design of cognitive robot system capable of developing perceptual, behavioral and cognitive categories in a measurable way and communicating and sharing these with humans and other artificial agents. The basic assumptions is that cognition is based on recurrent processes involving nested feedback loops operating on object-action complexes: Action defines the meaning (semantics) of an object and Objects suggest Actions the Affordance Principle. Main issues of PACO-PLUS concern the Sensor Representation, Action Representation, Quantification, Decision making and Planning, Novelty and Creativity, Language and Communication and the realization of the system in hardware.

The second part of the presentation concerned our present contribution concerning the representation and recognition of action.

There is biological evidence that human actions are composed out of action primitives, similarly to words and sentences being composed out of phonemes. Given a set of action primitives and an action composed out of these primitives we present a Hidden Markov Model-based approach that allows to recover the action primitives in that action. In our approach, the primitives may have different lengths, no clear “divider” between the primitives is necessary. The primitive detection is done online, no storing of past data is necessary. We verify our approach on a large database of MoCap and Silhouette data.

Hybrid Deformable Modeling Methods for Human Shape and Motion Analysis

Dimitris Metaxas (Rutgers Univ. - Piscataway, USA)

We will present novel methods for integrating discriminative and generative methods for human shape and motion analysis.

In particular we will present the coupling of deformable (generative) methods with learning-based discriminative approaches for facial and hand tracking methods with applications to American Sign Language analysis and recognition. In particular our coupling will be based on the error of fit between the deformable model and the image data. In addition, we will also present the integration of active shape(ASM) and deformable models for real time facial analysis. Finally, we will give a series of examples to demonstrate the effectiveness of our approach.

Automatic Classification and Retrieval of Motion Capture Data

Meinard Müller, Tido Röder (Universität Bonn, D)

The lifecycle of a motion clip in the conventional production of computer animations is typically very short: after some rehearsal, a motion clip is captured, incorporated in a single 3D scene, and then never used again. Both for efficiency and cost reasons, methods for further exploiting large collections of motion clips are gaining in importance. Here, an active field of research is the application of morphing and blending techniques for the creation of new, realistic motions from pre-recorded motion clips. This requires the identification and extraction of logically related motions scattered within some data set. Such content-based retrieval of motion capture data, which is one of the topics of this talk, constitutes a difficult and time-consuming problem due to significant spatio-temporal variations between logically related motions. Recent approaches to motion retrieval apply techniques such as dynamic time warping which, however, are not applicable to large data sets due to their quadratic space and time complexity. In our approach, we introduce various kinds of boolean features describing geometric relations between specified body points of a pose and show how these features induce a time segmentation of motion capture data streams. By incorporating spatio-temporal invariance into the relational features and adaptive segments, we are able to adopt efficient indexing methods allowing for flexible and efficient content-based retrieval and browsing in huge motion capture databases.

As a further application of relational motion features, a new method for automatic motion classification is presented. Here, we introduce the concept of weighted motion templates (WMT), by which the spatio-temporal characteristics of an entire motion class can be captured in one explicit, compact matrix representation. An efficient algorithm is described that automatically derives a WMT from a set of logically related training motions by some iterative averaging procedure. The resulting class WMT has a direct, semantic interpretation, and it can be manually edited, mixed, combined with other WMTs, extended, and restricted. Furthermore, a class WMT exhibits the characteristic as well as the variational aspects of the underlying motion class at a semantically high level. Classification is then performed by comparing a set of pre-computed class WMTs with unknown motion data and labeling matching portions with the respective motion class label. Here, the crucial point is that the variational (hence uncharacteristic) motion aspects encoded in the class WMT are automatically masked out in the comparison, which can be thought of as locally adaptive feature selection. In our extensive experiments, which were based on several hours of motion data, WMTs proved to be a powerful concept for motion classification.

Keywords: Motion capture, relational feature, retrieval, classification, indexing, motion template, adaptive segmentation, time alignment

Joint work of: Müller, Meinard; Röder, Tido

See also: Müller, M., Röder, T., Clausen, M. (2005): Efficient Content-Based Retrieval of Motion Capture Data. ACM Transactions on Graphics 24(3): 677-685 (Proceedings of ACM SIGGRAPH 2005).

Markerless Motion Capture for Biomechanical Applications

Lars Mündermann (Stanford University, USA)

Over the last several centuries our understanding of human locomotion has been a function of the methods to capture human movement that were available at the time. The Weber brothers (1836) reported one of the first quantitative studies of the temporal and distance parameters during human locomotion. The works of two contemporaries, Marey (1873) and Muybridge (1878), were among the first to quantify patterns of human movement using photographic techniques. Today most common methods for accurate capture of three-dimensional human movement require a laboratory environment and the attachment of markers or fixtures to the body Rs segments. These laboratory conditions can cause unknown experimental artifacts. Modern biomechanical and clinical applications require the accurate capture of normal and pathological human movement without the artifacts associated with standard marker-based motion capture techniques such as soft tissue artifacts and the risk of artificial stimulus of taped-on or strapped-on markers. The need for accurate information on the characteristics of normal and pathological human is motivated in part by the introduction of new clinical approaches for the treatment and prevention of diseases that are influenced by subtle changes in the patterns movement.

The need for markerless human motion capture methods is discussed and the advancement of markerless approaches is considered in view of accurate capture of three-dimensional human movement for biomechanical applications.

The role of choosing appropriate technical equipment and algorithms for accurate markerless motion capture is critical. The implementation of this new methodology offers the promise for simple, time-efficient, and potentially more meaningful assessments of human movement in research and clinical practice. The feasibility of accurately and precisely measuring 3D human body kinematics using a markerless motion capture system is demonstrated.

From Performance Theory to Character Animation Tools

Michael Neff (MPI für Informatik - Saarbrücken, D)

Generating expressive, character specific movement is a significant challenge that only a small number of people have truly mastered. My work focuses on creating tools that help with this task. Specifically, I try to leverage off the performing arts literature to better understand the nature of expressive movement and then use this understanding to build computational tools. I will show a range of results from this approach and also indicate how the approach can be used as a basis for both higher level, automatic motion generation and interactive motion performance.

Graphical Models for Human Motion Modeling

Vladimir Pavlovic (Rutgers Univ. - Piscataway, USA)

The human figure exhibits complex and rich dynamic behavior that is both nonlinear and time-varying. To automate the process of motion modeling we consider a class of learned dynamic models cast in the framework of dynamic Bayesian networks (DBNs) applied to analysis and tracking of the human figure. While direct learning of DBN parameter is possible, Bayesian learning formalism suggests that hyperparametric model description that considers all possible model dynamics may be preferred. Such integration over all possible models results in a subspace embedding of the original motion measurements. To this end, we propose a new family of Marginal Auto-Regressive (MAR) graphical models that describe the space of all stable auto-regressive sequences, regardless of their specific dynamics. We show that the use of dynamics and MAR models may lead to better estimates of sequence subspaces than the ones obtained by traditional non-sequential methods. We then propose a learning method for estimating general nonlinear dynamic system models that utilizes the new MAR models. The utility of the proposed methods is tested on the task of tracking 3D articulated figures in monocular image sequences. We demonstrate that the use of MAR can result in efficient and accurate tracking of the human figure from ambiguous visual inputs.

Qualitative and quantitative aspects of movement. The discrepancy between clinical gait analysis and activities of daily life

Dieter Rosenbaum (Universitätsklinikum Münster, D)

The first 3-dimensional movement analysis was already performed in the late 19th century and was based on the successful collaboration between the German anatomist Wilhelm Braune (1831-1892) and the mathematician Otto Fischer (1861-1917). While this first approach was extremely time consuming, the recent technical developments have helped to reduce the processing times so that clinical gait analysis nowadays can be considered as a standard tool in well-equipped research labs.

With an array of cameras and passive or active marker systems, the moving subject is being captured in the calibrated laboratory environment. Standard marker sets and biomechanical models are usually applied to extract the parameters of interest which will be used to describe the motion characteristics or identify and understand possible causes of movement disorders. The consideration of various body segments (i.e. arm; trunk; pelvis; hip, knee, and ankle joint) in the anatomical planes (i.e. frontal, sagittal and transverse plane) enables a detailed description of the individuals performance during walking. Comparisons between the clinically affected and contra-lateral side or between a patient and a control group are used to detect abnormal movement kinematics and kinetics.

It has to be realized, however, that clinical gait analysis is usually revealing the patient's best performance under ideal conditions. It remains unclear how indicative this is of his or her activities in daily life (ADL). Therefore, body-fixed sensors are increasingly used for ADL-monitoring, i.e. to evaluate patients in their natural environment. These small, light-weight and more or less non-obstructive sensors (e.g. accelerometers, gyroscopes, etc.) can record certain movement characteristics that are stored during long-term measurements and are read out for subsequent analysis. The systems are usually robust, easy to use and less expensive. It has to be noted, however, that they do not (yet?) provide the same level of detailed information as a 3-d movement analysis.

Therefore, two different approaches are available and have to be clearly distinguished: 1. Clinical gait analysis is laboratory-based, i.e. stationary, and describes how a patient is able to move. 2. ADL-monitoring is not confined to a lab environment and describes how much the patient is using his potential and which level of mobility is being used. Both approaches have their specific advantages and disadvantages which have to be considered before application and may be combined for a full description of an individuals movement characteristics. In the future, marker-less motion capturing systems might be a way to provide detailed of motion analysis in home-based environments

Tracking Clothed People

Bodo Rosenhahn (MPI für Informatik - Saarbrücken, D)

The talk presents an approach for motion capturing (MoCap) of dressed people.

A cloth draping method is embedded in a silhouette based MoCap system and an error functional is formalized to minimize image errors with respect to silhouettes, pose and kinematic chain parameters, the cloth draping components and external wind forces. We report on various experiments with two types of clothes, namely a skirt and a pair of shorts. Finally we compare the angles of the MoCap system with results from a commercially available marker based tracking system. The experiments show, that we are basically within the error range of marker based tracking systems, though body parts are occluded with cloth.

Non-convex manifold learning by robust flattening

Guy Rosman (Technion - Haifa, IL)

Presented is an algorithm for nonlinear dimensionality reduction that uses both local and global distances in order to learn the intrinsic geometry of a non-convex yet potentially flat, manifold.

Since our algorithm matches non-local structures, it is robust even to strong noise. We show experimental results demonstrating the advantages of our approach over state of the art methods.

Joint work of: Rosman, Guy; Bronstein, M. Alex; Bronstein, M. Michael; Kimmel, Ron

Motion Analysis using Scale-space Tree

Tomoya Sakai (Chiba University, J)

We elucidate the hierarchical structure of an image in a Gaussian scale space, and observe transition of the structure in image sequences. Local maxima and local minima of the image intensity are feature points of bright parts and dark cavities in the image, and saddles imply their connections. A point at infinity is a hidden local minimum, which represents a dark background of a bright image. These stationary points have hierarchical relationships indicated by their trajectories (stationary curves) in the scale space, and by specific gradient field curves (antidirectional figure-flow curves) at fixed scales. Therefore, tree representation for the hierarchical structure is derived from the differential geometric features in the scale space. For a sequence of images, we yield a sequence of the trees. The image sequence is segmented into subsequences according to structural changes of the trees. We show experimental results on image sequences of beating heart, rotating box, kicking horse, and so on. Our approach is non-model-based, and it achieves temporal segmentation without any prior knowledge about the image sequences.

Keywords: Deep structure, pseudograph, temporal segmentation, skeleton extraction

Joint work of: Sakai, Tomoya; Imiya, Atsushi

High Resolution Acquisition, Learning and Transfer of Dynamic 3D Facial Expressions

Dimitris Samaras (SUNY at Stony Brook, USA)

We present data-driven face models that accurately describe the appearance of faces under unknown pose and illumination conditions as well as the subtle geometry changes that happen during expressions.

I will focus on issues in 3D shape matching, a crucial step in the analysis of multiple 3D scans. In order to overcome challenges posed by noise and occlusion, we explore a family of conformal geometric maps including harmonic maps, conformal maps and least squares conformal maps. This allows us to simplify the 3D surface-matching problem to a 2D image-matching problem, by comparing the resulting 2D conformal geometric maps. These methods are applied to tracking of dense point clouds, that guarantees that nonrigid deformations will be accurately tracked as well. Once tracking is achieved, by reducing the dimensionality of our data to a lower dimensional space manifold and decomposing into style and content parameters, we transfer subtle expression information (in the form of a style vector) between individuals, as well as smoothly morph geometry and motion. In the second part of the talk I will present a statistical model of shape and spherical harmonic appearance information, which allows us to convert any single image of a face into a different pose and illumination.

Keywords: Facial motion tracking facial animation

See also: High Resolution Acquisition, Learning and Transfer of Dynamic 3-D Facial Expressions, Yang Wang, Xiaolei Huang, Chan-Su Lee, Song Zhang, Zhiguo Li, Dimitris Samaras, Dimitris Metaxas, Ahmed Elgammal, Peisen Huang In Computer Graphics Forum (EuroGraphics 2004) pp. III: 677-686.

Acquisition, Pose Estimation and Modeling of Human Body Shape and Motion

James Sherman (University of Maryland - College Park, USA)

There are different approaches to pose estimation and registration of different body parts using voxel data. We propose a general bottom-up approach in order to segment the voxels into different body parts (such as arms, legs, head and trunk) that are 1-D in nature. The voxels are first transformed into a high dimensional space which is the eigenspace of the Laplacian of the neighborhood graph. We exploit the properties of this transformation and fit splines to the voxels belonging to different body segments in eigenspace. The boundary of the splines is determined by examination of the error in spline fitting. We then use a probabilistic approach in order to register the segmented body segments by utilizing their connectivity and prior knowledge of the general structure of the subjects. We present results on real data, containing both simple and complex poses. We also present results on human body model estimation using this method for different human subjects. The method is fairly general and can be applied to voxel-based registration of any articulated or non-articulated object which is composed of primarily 1-D parts.

Keywords: Pose estimation, voxel based, body part registration

Joint work of: Sundaresan, Aravind; Chellappa, Rama; Sherman, James

Human Motion Analysis in Sports Footage Detection, Recognition and Tracking

Josephine Sullivan (KTH - Stockholm, S)

This talk gives an overview of recent work carried out at KTH in the field of human motion analysis. The focus will be on 3d reconstruction of human action from monocular video sequences and also on more recent results which focus on multi-target tracking.

Firstly, algorithm is presented for the 3D reconstruction of human action in relatively long ($>$ 30 second) monocular image sequences. A sequence is represented by a small set of automatically found representative keyframes. The skeletal joint positions are manually located in each keyframe and mapped to all other frames in the sequence. For each keyframe a 3D key pose is created, and interpolation between these 3D body poses, together with the incorporation of limb length and symmetry constraints, provides a smooth initial approximation of the 3D motion. This is then fitted to the image data to generate a realistic 3D reconstruction. The degree of manual input required is controlled by the diversity of the sequence's content. Sports footage is ideally suited to this approach as it frequently contains a limited number of repeated actions. Our method is demonstrated on a long (36 second) sequence of a woman playing tennis filmed with a non-stationary camera. This sequence required manual initialisation on $< 1.5\%$ of the frames, and demonstrates that the system can deal with very rapid motion, severe self occlusions, motion blur and clutter occurring over several concurrent frames. The result is seen to provide a qualitatively accurate 3D reconstruction of the motion.

Successful multi-target tracking requires solving two problems - localize the targets and label their identity. An isolated target's identity can be unambiguously preserved from one frame to the next. However, for long sequences of many moving targets, like a football game, grouping scenarios will occur in which identity labellings cannot be maintained reliably by using continuity of motion or appearance. This talk describes how to match targets' identities despite these interactions. Trajectories of when a target is isolated are found. These trajectories end when targets interact and their labellings cannot be maintained. The interactions (merges and splits) of these trajectories form a graph structure. Appropriate feature vectors summarizing particular qualities of each trajectory are extracted. A clustering procedure based on these feature vectors allows the identities of temporally separated trajectories to be matched. Results are shown from a football match captured by a wide screen system giving a full stationary view of the pitch.

Keywords: Human Motion, multi-target tracking

Multiview 3D Tracking with an Incrementally Constructed 3D Model

Tomas Svoboda (Czech Technical University, CZ)

We propose a multiview tracking method for rigid objects. Assuming that a part of the object is visible in at least two cameras, a partial 3D model is reconstructed in terms of a collection of small 3D planar patches of arbitrary topology. The 3D representation, recovered fully automatically, allows to formulate tracking as gradient minimization in pose (translation, rotation) space. As the object moves, the 3D model is incrementally updated. A virtuous circle emerges: tracking enables composition of the partial 3D model; the 3D model facilitates and robustifies the multiview tracking.

We demonstrate experimentally that the interleaved track-and-reconstruct approach successfully tracks a 360 degrees turn-around and a wide range of motions. Monocular tracking is also possible after the model is constructed. Using more cameras, however, significantly increases stability in critical poses and moves. We demonstrate how to exploit the 3D model to increases stability in the presence of uneven and/or changing illumination.

Joint work of: Zimmermann, Karel; Matas, Jiri

See also: Zimmermann, Karel and Svoboda, Tomas and Matas, Jiri; Multiview 3D Tracking with an Incrementally Constructed 3D Model; Third International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT), 2006

Advances and Challenges in Interactive Deformable Modeling

Matthias Teschner (Universität Freiburg, D)

The realistic simulation of complex deformable objects at interactive rates comprises a number of challenging problems, including deformable modeling, collision detection, collision response, and geometric constraints.

1. The deformable modeling approach has to provide interactive update rates, while guaranteeing a stable simulation. Furthermore, the approach has to represent objects with varying elasto-mechanical properties.
2. The collision detection algorithm has to handle large numbers of geometrically complex objects. In particular, the algorithm has to detect self-collisions of deforming objects.
3. The collision response method has to handle colliding and resting contacts among multiple deformable objects in a robust and consistent way. The method has to consider the fact that only sampled collision information is available due to the discretized object representations and the discrete-time simulation.

4. Geometric constraints provide powerful mechanisms to broaden the versatility of simulations in terms of scene modeling. Therefore, efficient constraint approaches have to be investigated and integrated into interactive frameworks.

The presentation discusses solutions to the aforementioned simulation aspects. Interactive software demonstrations illustrate all models, algorithms, and their application to human motion, surgery simulation and games. Finally, ongoing research and potential challenges will be discussed.

Video-based Capturing and Rendering of People

Christian Theobalt (MPI für Informatik - Saarbrücken, D)

By means of passive optical motion capture real people can be authentically animated and photo-realistically textured. To import real-world characters into virtual environments, however, also surface reflectance properties must be known. We describe a video-based modeling approach that captures human shape and motion as well as reflectance characteristics from a handful of synchronized video recordings. The presented method is able to recover spatially varying surface reflectance properties of clothes from multi-view video footage. The resulting model description enables us to realistically reproduce the appearance of animated virtual actors under different lighting conditions, as well as to interchange surface attributes among different people, e.g. for virtual dressing. Our contribution can be used to create 3D renditions of real-world people under arbitrary novel lighting conditions on standard graphics hardware.