# Real Time Extraction of Human Gait Features for Recognition

*Thesis submitted in partial fulfilment*

*of the requirements for the award of the degree of*

**Master of Technology**

*in*

**Communication & Signal Processing**

*by*

**Sonia Das**

**(211EC4321)**

*Under the supervision*

*of*

**Prof. Sukadev Meher**



**Department of Electronics & Communication Engineering**

**NATIONAL INSTITUTE OF TECHNOLOGY, ROURKELA**

राष्ट्रीय प्रौद्योगिकी संस्थान, राउरकेला

**May 2013**

# Real Time Extraction of Human Gait Features for Recognition

*Thesis submitted in partial fulfilment*

*of the requirements for the award of the degree of*

**Master of Technology**

*in*

**Communication & Signal Processing**

*by*

**Sonia Das**

**(211EC4321)**



**Department of Electronics & Communication Engineering**

**NATIONAL INSTITUTE OF TECHNOLOGY, ROURKELA**

राष्ट्रीय प्रौद्योगिकी संस्थान, राउरकेला

**May 2013**

# Abstract

Human motion analysis has received a great attention from researchers in the last decade due to its potential use in different applications such as automated visual surveillance. This field of research focuses on human activities, including people identification. Human gait is a new biometric indicator in visual surveillance system. It can recognize individuals as the way they walk. In the walking process, the human body shows regular periodic variation, such as upper and lower limbs, knee point, thigh point, stride parameters (stride length, Cadence, gait cycle), height, etc. This reflects the individual's unique movement pattern. In gait recognition, detection of moving people from a video is important for feature extraction. Height is one of the important features from the several gait features which is not influenced by the camera performance, distance and clothing style of the subject. Detection of people in video streams is the first relevant step of information and background subtraction is a very popular approach for foreground segmentation. In this thesis, different background subtraction methods have been simulated to overcome the problem of illumination variation, repetitive motions from background clutter, shadows, long term scene changes and camouflage. But background subtraction lacks capability to remove shadows. So different shadows detection methods have been tried out using RGB, YCbCr, and HSV color components to suppress shadows. These methods have been simulated and quantitative performance evaluated on different indoor video sequence. Then the research on shadow model has been extended to optimize the threshold values of HSV color space for shadow suppression with respect to the average intensity of local shadow region. A mathematical model is developed between the average intensity and the

threshold values.Further a new method is proposed here to calculate the variation of height during walking. The measurement of height of a person is not affected by his clothing style as well as the distance from the camera. At any distance the height can be measured, but for that camera calibration is essential. DLT method is used to find the height of a moving person for each frame using intrinsic as well as extrinsic parameters. Another parameter known as stride, function of height, is extracted using bounding box technique. As human walking style is periodic so the accumulation of height and stride parameter will give a periodic signal. Human identification is done by using theses parameters. The height variation and stride variation signals are sampled to get further analyzed using DCT (Discrete Cosine Transformation), DFT (Discrete Fourier Transformation), and DHT (Discrete Heartily Transformation) techniques. N - harmonics are selected from the transformation coefficients. These coefficients are known as feature vectors which are stored in the database. Euclidian distance and MSE are calculated on these feature vectors. When feature vectors of same subject are compared, then a maximum value of MSE is selected, known as Self-Recognition Threshold (SRT). Its value is different for different transformation techniques. It is used to identify individuals. Again we have discussed on Model based method to detect the thigh angle. But thigh angle of one leg can't be detected over a period of walking. Because one leg is occluded by the other leg. So stride parameter is used to estimate the thigh angle.

**Keywords:** *Gait Recognition, Background subtraction, Height estimation, Calibration, Stride, Silhouette.*

Department of Electronics and Communication Engg
**National Institute of Technology Rourkela**
ROURKELA-769 008, ODISHA, INDIA

May 29, 2013

# Certificate

This is to certify that the thesis titled as *"Real Time Extraction of Human Gait Features for Recognition"* by *"Sonia Das"* is a record of an original research work carried out under my supervision and guidance in partial fulfilment of the requirements for the award of the degree of *Master of Technology degree* in *Electronics and Communication Engineering* with specialization in *Communication and Signal Processing* during the session 2012-2013.

**Prof. Sukadev Meher**

# Acknowledgments

First and foremost, I am truly indebted to my supervisors Professor Sukadev Meher for this constant inspiration, excellent guidance   and valuable discussion leading to fruitful work is highly commendable. From finding a problem to solve it with careful observation by Professor Sukadev Meher  is unique who helped me a lot in my dissertation work and of course in due time. There are many people who are associated with this project directly or indirectly whose help, timely suggestion are highly appreciable for completion of this project. I would like to thankful to Deepak Kumar Panda, Aditya Acharya, Deepak Singh, Bodhisattwa Chakraborty, Lucky Kodwani, Saini Sikta, Vanita Devi and all friends, research members of Image processing lab  of  NIT Rourkela for their suggestions and good company I had  with.

I am very much indebted to Prof. Sarat Kumar Patra, Prof. Samit Ari, Prof. Kamala Kanta Mohapatra for providing insightful comments at different stages of thesis that were indeed thought provoking.

My special thanks go to Prof. Ajit Kumar Sahoo, Prof. Upendra Kumar Sahoo and Prof. Santosh Kumar Das for contributing towards enhancing the quality of the work in shaping this thesis.

My wholehearted gratitude to my parents Prasanta Kumar Das, Minati Das, and my friend Pradosh for their   love and support.

Sonia Das

Rourkela, May  2013

# Contents

# List of Figures

# List of Tables

# Chapter-1

# Gait Recognition

Gait is defined as —a manner of walking in Webster's New Collegiate Dictionary. Gait can be used as a biometric measure to recognize known persons and classify unknown subjects. The analysis of gait in real time finds considerable utility in applications ranging from the development of more intelligent human-computer inter-faces and visual surveillance systems to the video-based interpretation of mobility disorders.

## 1.1    Introduction

In video surveillance system human identification is an intriguing job. Human identification uses many biometric resources, for instance fingerprint, palm print, face, Iris, and hand geometry. Each individual resource requires a bound fundamental interaction between a person and a system. But gait as a behavioral biometric resource, which is non-invasive and arguably non-concealable nature. Gait recognition is one of the second generation biometrics which does not require subject cooperation. It is behavioral biometrics, can be captured from the distance and no-touching. Moreover, we extend our definition of gait to include certain aspects of the appearance of the person, such as: the aspect ratio of the torso, the clothing, the amount of arm swing, the period and phase of a walking cycle, etc.  Automatic capture and analysis of human motion is a highly active research area due to the number of potential applications and its inherent complexity. Gait can be detected and measured at low resolution, and therefore it can be used in such place where face or iris information is not available in high enough resolution for recognition.  For biometrics research, gait is usually

referred to include both body shape and dynamics, i.e. any information that can be extracted from the video of a walking person to robustly identify the person under various condition variations. The demand for automatic human identification system is strongly increasing and growing in many important applications. It has gained great interest from the pattern recognition and computer vision researchers for it is widely used in many security-sensitive environments such as banks, parks and airports [16]. Human identification uses many biometric resources, for instance fingerprint, palm print, face, Iris, hand geometry of an individual requires interaction between a person and a system. The biometric features of face will not give satisfactory result when there is a distance between the camera and person. In such case gait features will give an estimable result. There is an increased interest in gait as a biometric, mainly to its non-intrusive and arguably non-concealable nature [3]. Human gait recognition works from the observation that an individual's walking style is unique and can be used for human identification. The extraction and analysis of a pattern of human walking, or gait, has been an ongoing area of research since the advent of the still camera in 1896 [8]. There two areas dominate the field of gait research at the present. Clinical gait analysis focuses on collection of gait data in controlled environments using motion capture systems and biometric goals of human gait analysis analyze an individual's gait in a variety of different areas and scenarios. Gait analysis uses in biometric systems are largely based on visual data capture and analysis systems which process video of walking subjects in order to analyze gait.

Gait classification indicates walking, running, jumping. Gait recognition also called as gait based human identification. Recognizing people by gait depends on how the silhouette shape of an individual changes over time in an image.

## 1.2    Challenges

Although the performance of all three user authentication for biometric gait recognition approaches is encouraging, but there are several factors that may negatively influence the accuracy of such approaches. We can group the factors that influence a biometric gait system into two classes.

- *External factors*. Such factors mostly impose challenges to the recognition approach (or algorithm). For example, viewing angles (e.g. frontal view, side-view), lighting conditions (e.g. day/night), outdoor/indoor environments (e.g. sunny, rainy days), clothes (e.g. skirts in MV-based category), walking surface conditions (e.g. hard/soft, dry/wet grass/concrete, level/stairs, etc.), shoe types (e.g. mountain boots, sandals), object carrying (e.g. backpack, briefcase) and so on.

- *Internal factors*. Such factors cause changes of the natural gait due to sickness (e.g. foot injury, lower limb disorder, Parkinson disease etc.) or other physiological changes in body due to aging, drunkenness, pregnancy, gaining or losing weight and so on.

## 1.3    Related Works

In recent years, various techniques have been proposed for human recognition by gait. Little and Boyd [6] describe the shape of the human motion with scale-independent features from moments of the dense optical flow. Barron et.al. [11] describe the performance of optical flow techniques. Optical-flow-based methods cannot be applied to video streams in real time without specialized hardware. Sundaresan et al. [12] proposed a hidden Markovmodels (HMMs) based framework for individual recognition by gait. Sminchisescu et al.[13]describes Covariance scaled sampling for monocular 3D body tracking. The main disadvantage of 2-D model is that they require restrictions on the viewing angle. To overcome this many researchers use 3-D volumetric model such as elliptical, cylinders,

cones, spheres, etc. But Zhao et al. [14] sites volumetric models require more parameters than image based models and lead to more expensive computation and complexity due to use of linear method. It takes head point of the silhouette which can be easily detect at a distance.

## 1.3.1 Gait Analysis and Recognition

Human recognition based on gait can be grouped into three categories shown in fig1.1, namely machine vision (MV) based, floor sensor (FS) based and wearable sensor (WS) based.
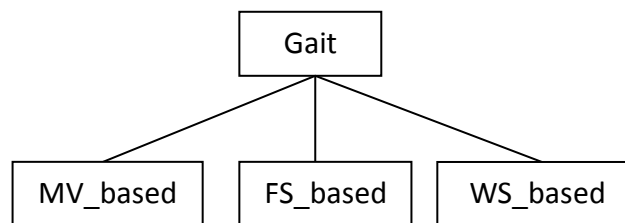


**Figure 1.1**: User authentication approaches

## 1.3.2 MV based Gait Recognition

In MV-based gait recognition technique, gait is captured using a video-camera from distance. Video and image processing techniques are applied to extract gait features for recognition purposes. Most of the MV-based gait recognition algorithms are based on the human silhouette [16]. That is the image background is removed and the silhouette of the person is extracted and analyzed for recognition.

## 1.3.3 FS based Gait Recognition

In FS-based approach, a set of sensors or force plates are installed on the floor [5, 6], which is shown in fig 1. 2. Such sensors enable to measure gait related features, when a person walks on them. Gait recognition approaches may be explicitly classified into two main classes. One of the main advantages of FS-based gait recognition is in its unobtrusive data collection. The FS-based gait recognition can be deployed in access control application and is usually installed in front of doors in the building. Such systems can find deployment as a standalone system or as a part of multimodal biometric system. This system can also indicate location information within a building [6].
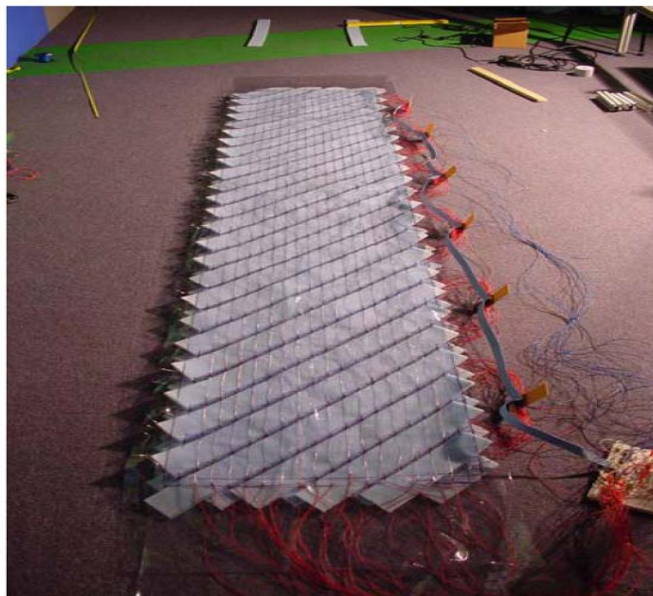


**Figure 1.2**: A prototype sensor mat from [6]

## 1.3.4 WS based Gait Recognition

In WS-based gait recognition, gait is collected using body worn motion recording (MR) sensors [16]. The MR sensors can be worn at different locations on the human body. In [24],

the MR sensor was attached to the lower part of the leg as shown in fig.1.3. The acceleration of gait, which is recorded by the MR sensor, is utilized for authentication. WS-based gait recognition was described by Morris [15]. However, the focus of this work was primarily on clinical aspects of the system [15]. Ailisto et al. [8] proposed WS-based gait recognition as a biometric authentication. In their approach, the MR sensor was attached to the waist of the subject. One of the main advantages of the WS-based gait recognition over several other biometric modalities is its unobtrusive data collection. The WS-based approach was proposed for protection and user authentication in mobile and portable electronic devices. With advances in miniaturization techniques it is feasible to integrate the MR sensor as one of the components in personal electronic devices.



**Figure 1.3** : The MR sensor attached to the lower leg [6].

Further gait recognition approaches may be explicitly classified into two main classes, namely model-based methods and motion-based methods.

- ➢ Model-Based method
- ➢ Motion-based method

## 1.3.5 Model based Method

A model-based approach is attempted to produce a biometric that has high fidelity to the original data [31]. The model based method is mainly used in the Medical studies. This is a novel approach to gait recognition by computer vision. The inherent advantage of a model-based approach is the potential ability to handle appearance transformations and practical effects, such as occlusion. Appearance transformations imply that an object's shape will be distorted by the camera's viewpoint. This can only be handled in area-based approaches by inclusion of marker points in each scene. A model-based method can also handle distorted scene without marker points since it relies on the presence of human motion in the sequence and as such an inherently model its time history/future. It approaches to feature extraction, use prior knowledge of the object. Model based methods are typically stick representations either surrounded by ribbons or blobs. But the disadvantage of implementing a model based approach is high computational cost, due to complex matching and is thus not suitable for real time systems.

## 1.3.6 Motion based Method

Motion based method is also called as Silhouette based method. In this method recognizing a person by gait intuitively depends on how the silhouette shape of an individual changes over time in an image sequence. Motion-based approaches can be further divided into two main classes [18]. The first class, state-space methods [18], considers gait motion to be composed of a sequence of static body poses and recognizes it by considering temporal variations of observations with respect to those static poses.

## 1.4    Gait Description

The following terms are used to describe the gait cycle, as given in [18]. In fig 1.4.the following terms are  described.
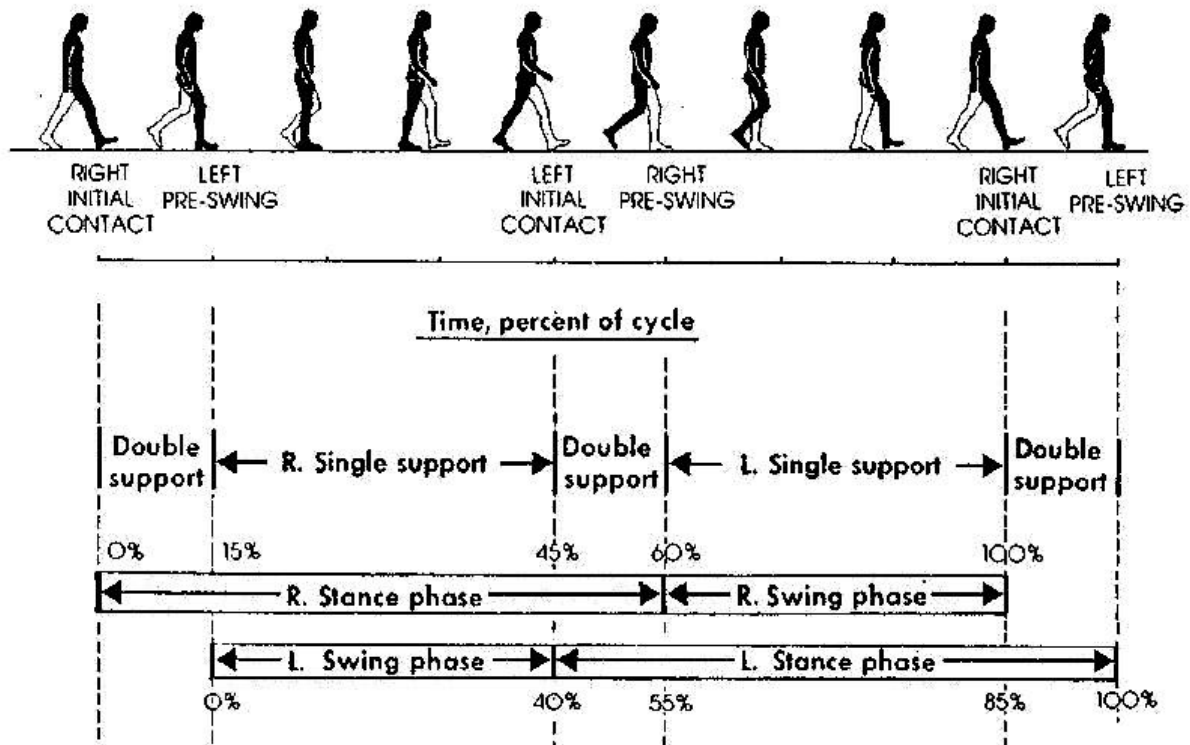


**Figure 1.4** : Relationship between gait cycle, step lengths and stride length [18].

## 1.4.1  Gait Cycle

A gait cycle is the time interval between successive in-stances of initial foot-to-floor contact ("heel strike") for the same foot. Each leg has two distinct periods. Period of time from one heel strike to the next heel strike of the same limb.

### 1.4.2 Step length

Distance between corresponding successive points of heel contact of the opposite feet. Step length is the distance between the points of initial contact of one feet and the point of opposite feet. In normal gait , right and left step lengths are similar.  Step length is the distance between the heel strike of one feet and the  heel strike of the opposite feet as you walk.

### 1.4.3 Stride Length

Linear distance between successive points of heel contact of the same feet.  It is the distance between successive points of initial contact points of initial contact of same feet. Right and left stride lengths are normally equal. Stride length is the distance between the heel strikes of the same foot. Stride length = 2× step length (in normal gait)

### 1.4.4 Stride Width

Side to side distance between the lines of two feet. It is also called as walking Base.

**Cadence:** Number of steps per unit time. During gait cycle each extremity passes through two passes.

**Stance Phase:** Stance Phase begins with heel strike and ends, when toe leaves the ground. Feet are in contact with the ground, heel-strike to toe-off 60%.

**Velocity:**  It is the distance covered by body in unit time. Velocity, the product of cadence and step length, is expressed in units of distance per time. Instantaneous velocity varies during the gait cycle. Average velocity (m/min) = step length (m) × cadence (steps/min).

There are two phases for gait cycle

(1) Stance phase: It references limb is contact with the floor.

(2) Swing phase: The limb is not contact either the floor.
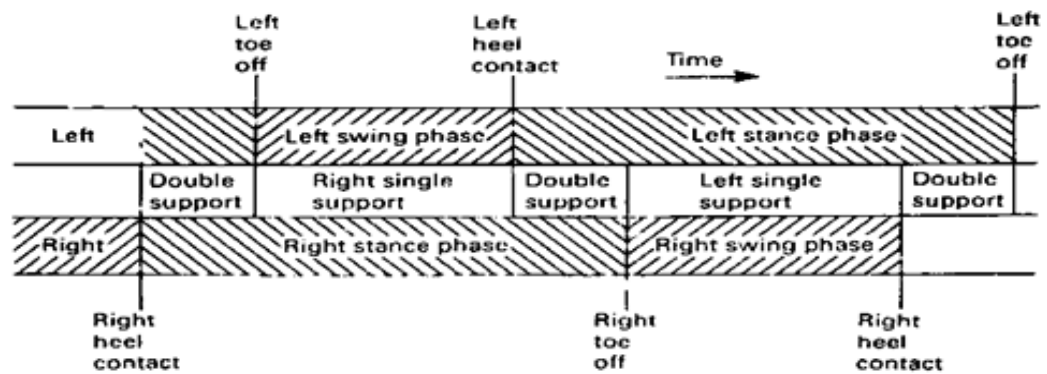


**Figure 1.5**: Phases for Gait cycle

► Time Frame:

A. Stance vs. Swing:

► Stance phase  =  60% of gait cycle

► Swing phase  =  40% of gait cycle

B. Single vs. Double support:

► Single support=  40% of gait cycle

► Double support=  20% of gait cycle

## 1.5    Problem statement

The main objective of gait recognition over other biometric recognition is

➢  Detect feature at low resolution i.e. perceivable at a distance

➢  Non contact.

➢ Non invasive.

➢ Give accurate result without user co-operation.

➢ Camouflage can avoidable.

➢ Dynamic in nature

The current areas of biometric research include automatic face recognition, eye (retina) identification, fingerprints, hand geometry, vein patterns, and voice patterns.

The face may be hidden or at low resolution, fails due to illumination changes, Pose variation, aging effects, and expression Variation; the palm or finger may be obscured, both require user cooperation to contact the palm or finger with the device; the ears may not be seen; Iris recognition fails due to eye lash occlusion, eyelid occlusion, Specular reflection. However, people need to walk so their gait is usually apparent.As gait is a dynamic property which is never being changed wittingly. This motivates using gait as a biometric.

## 1.6    Overview

The overview of the proposed model for feature extraction is shown in Fig.1.6. The algorithm consists of three main modules. The first module tracks the walking person and extracts the head and feet point from each frame. The detection of head and feet points consist of following steps-Input video frames extraction, silhouette detection, corner point's detection, head and feet point detection. The second module uses calibration process to get intrinsic and extrinsic camera parameters. DLT is used in the third module to give approximate height, stride length of the subject in each frame.
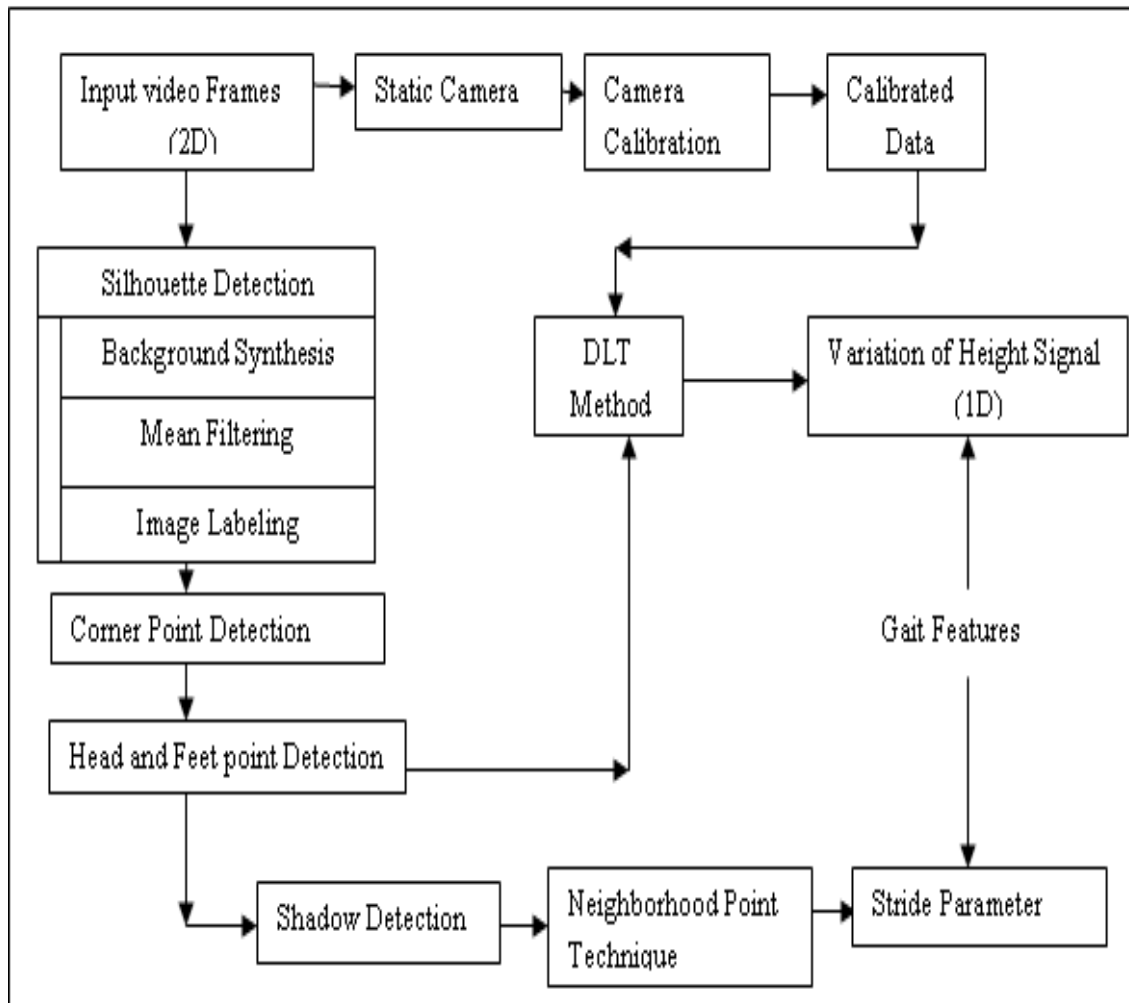
**Figure 1.6**: System overview of gait features extraction process

## 1.7     Organization of the Thesis

The remaining part of the thesis is organized as follows. Chapter 2 presents a brief survey of background subtraction methods for motion segmentation along with mean filtering, image labelling. In chapter 3 we have discussed shadow detection using different colour spaces such as RGB, YCbCr and HSV and optimizing the threshold levels for shadow detection. Gait feature extraction via silhouette based method and model-based method is described in chapter 4 and chapter 5 respectively. Human recognition Process is discussed in chapter 6. Finally, Chapter 7 concludes the thesis with the suggestions for future research.

## 1.8    Conclusion

Gait is an emergent biometric and recent study confirmed its potential use for surveillance applications. It is mostly used by computer vision researchers to approach the problem of gait analysis and recognition using different methodologies including model-based or model-free methods, most of their contributions and research studies were limited to the use of a silhouette-based approach or anatomical-based methods for gait recognition applied to walking subjects recorded from the side view, frontal view, oblique view without examining the effects of every-day factors including clothing, load carriage and high-heeled shoes.

# Chapter-2

# Background Subtraction Methods

In this chapter different background subtraction methods are discussed which the basic process for silhouette extraction is. Four background subtraction methods are discussed. one is simple frame differencing method. Others are running average method, sigma delta method and effective sigma delta methods, etc. These methods are verified using manually taking videos and database videos. Quantitative performance analysis is done and experimental results for tracking and classification of moving objects are drawn at the end of the chapter.

## 2.1    Introduction

In Automated Video Surveillance system aims to track an object in motion and classify it as a Human. It is used to recognize the region of interest i.e. the moving human in a video scene. Gait recognition is employed for person-specific identification in certain scenes for visual surveillance system. Motion detection and tracking, gait feature extraction are the different process for gait recognition. Motion detection includes background estimation, motion segmentation and human tracking. Motion detection is the process of detecting a change in position of an object relative to its surrounding, or the change in the surrounding relative to an object. A tracking algorithm measures and predicts the motion of a moving object over tie. Silhouette of moving objects are commonly used for feature for tracking.

Background estimation and motion segmentation are the important parts of human detection. It targets at a segmenting region corresponding to moving objects from the rest of an image. In this chapter, different background modeling techniques are analyzed and verified.

## 2.2    Motion Segmentation

Motion segmentation was described into three major classes of method as frame differencing, optical flow, and background subtraction by Hu et al. [1]. Motion detection targets the moving region such as human. Detecting moving regions provides a focus of attention on tracking, feature extraction and analysis. The segmentation methods use either temporal or spatial information in the image sequence.  The motion segmentation is adumbrated as:

1) Background subtraction: Background subtraction is a popular method for motion segmentation, but it requires a static background. It detects moving regions from video sequences by taking the difference between the current image and the reference background image in a pixel-by-pixel fashion. It is simple, but extremely sensitive to changes in dynamic scenes derived from lighting and extraneous events etc. Therefore, it is highly dependent on a good background model to reduce the influence of these changes [20], as part of environment modelling.

2) Temporal differencing: Temporal differencing makes use of the pixel-wise differences between two or three consecutive frames in an image sequence to extract moving regions. It is very adaptive to dynamic environments, but it can't extract all the relevant pixels, so there may be holes left inside moving entities. As an example of this method, Lipton et al. [28] detect moving targets in real video streams using temporal differencing. After the absolute difference

between the current and the previous frame is obtained, a threshold function is used to determine changes.

3) Optical flow. Optical flow based motion segmentation uses characteristics of flow vectors of moving objects overtime to detect moving regions in an image sequence. Optical-flow-based methods can be used to detect independently moving objects even in the presence of camera motion. However, most of these methods are computationally complex and very sensitive to noise, and cannot be applied to in real time without specialized hardware. More detailed discussion of optical flow can be found in Barron's work [11].

## 2.3    Related Work

The background subtraction [20], [21], and [22] is the most popular and common approach for motion detection. The idea is to take the difference between the current image and model image of background by using thresholding procedure. It gives silhouette region of an object. This approach is simple and computationally affordable for real-time systems, but is extremely sensitive to dynamic scene changes from lightning and extraneous event etc. Therefore it is highly dependent on a good background maintenance model. The problem with background subtraction [23] is to automatically update the background from the incoming video frame and it should be able to overcome the following problems:

- Motion in the background: Non-stationary background regions, such as branches and leaves of trees, a flag waving in the wind, or flowing water, should be identified as part of the background.

- Illumination changes: The background model should be able to adapt, to gradual changes in illumination over a period of time.

- Memory: The background module should not use much resource, in terms of computing power and memory.

- Shadows: Shadows cast by moving object should be identified as part of the background and not foreground.

- Camouflage: Moving object should be detected even if pixel characteristics are similar to those of the background.

- Bootstrapping: The background model should be able to maintain background even in the absence of training background (absence of foreground object).

For gait recognition proper silhouette extraction is important. So the idea is to simulate different background subtraction techniques which are available in the literature and compare experimental results for different gait videos.

## 2.4   Background Modelling

Moving object segmentation using background subtraction is very important for many visual applications: visual surveillance of both in outdoor and indoor environments, traffic control, and behavior detection during sport activities, and gait recognition. The Method of background extraction during training sequence and updating it during input frame sequence is called background modeling. The main challenges in motion segmentetion is extraction a clean background and its updating.

## 2.4.1  Simple Background Subtraction

In simple background subtraction an absolute difference is taken between every current image. $I_t$(x, y) and the reference background image B(*x, y*) to find out the motion detection mask D(*x, y*). The reference background is generally the first frame of a video, without containing foreground object.

$$D(x,y) = \begin{cases} 0 & if \ |I_t(x,y) - B_t(x,y)| \leq \tau \\ 1 & otherwise \end{cases} \qquad (2.1)$$

$I_t(x,y)$ = Current frame

$B(x,y)$ = Background frame

$\tau$          = Threshold value

Where $\tau$ is a threshold, which decides whether the pixel is foreground or back-ground. If the absolute difference is greater than or equal to $\tau$, the pixel is classified as foreground; otherwise the pixel is classified as background. If the background is not available in the video then background modeling method is used to construct the background.

## 2.4.2  Running Average

The commonly, fastest and the most memory compact background modeling is running average method. In this method, background extraction is done by arithmetic averaging on train sequence. After background extraction, background may change during detection of moving objects. Illumination changes are an important reason of background changes. Because of scene

illumination change and some other reasons, background image must be updated in each frame.

In running average method, background is updated as follow:

$$B_t(x, y) = (1 - \beta) B_{t-1}(x, y) + \beta I_t(x, y) \tag{2.2}$$

$$D(x, y) = \begin{cases} 0, & if \ |I_t(x, y) - B_t(x, y)| \leq \tau \\ 1, & if \ |I_t(x, y) - B_t(x, y)| \leq \tau \end{cases} \tag{2.3}$$

$\beta$ must be in range (0,1). As per the signal and system point of view, (2.2) is an Infinite Impulse Response (IIR) filter. Therefore, running average method is an IIR system.

## 2.4.3  Sigma-Delta Estimation

The $\Sigma$- $\Delta$ background estimation is a simple non-linear method of background subtraction [29]. It is a recursive computation of a valid background model of the scene. However, this model degrades quickly under slow or varying light conditions, due to the integration in the background model of pixel intensities belonging to the foreground objects.

➢ **Algorithm of Σ-Δ Estimation**

$B_0 = I_0$ // *initialize background model B*

$V_0 = 0$                    // Initialize variance $V$

**For** each frame t

$$\Delta_t(x, y) = |I_t(x, y) - B_t(x, y)|$$

**If** $\Delta_t \neq 0$ // Compute current difference

$$V_t(x, y) = V_{t-1}(x, y) + sgn(N \times \Delta t(x, y) - V_{t-1}(x, y)) // \text{Update variance } V$$

**End If**

$$D_t(x, y) = \begin{cases} 1 & if \Delta t(x, y) > V_t(x, y) \\ 0 & if \Delta t(x, y) < V_t(x, y) \end{cases} // \text{Compute detection image } D$$

**If** $D_t == 0$ // Update background model $B$

$$B_t(x, y) = B_{t-1}(x, y) + sgn(I_t(x, y) - B_{t-1}(x, y)) // \text{with relevance feedback}$$

**End If**

**End For**

$B_t$ represents the background-model image at frame *t*, $I_t$ represents the current input image, and $V_t$ represents the temporal *variance estimator image* (or *variance image*, for short), carrying information about the variability of the intensity values at each pixel. It is used as an adaptive threshold variable to be compared with the difference image. Pixels with higher intensity fluctuations will be less sensitive, whereas pixels with steadier intensities will signal detection upon lower differences. The only parameter to be adjusted is *N*, with typical values between 1

and 4. $D_t$ is the detection image or *detection mask*. This binary image highlights pixels belonging to the detected foreground objects.

## 2.4.4  Effective $\sum$-$\Delta$ Estimation

An M ×N resolution digital image was taken where x and y are spatial coordinates. the an original input image $F_f(x, y)$ is defined bellow.

$$F(x,y) = \begin{bmatrix} F_f(0,0) & F_f(0,1)\cdots & F_f(0,N-1) \\ F_f(1,0)\vdots & F_f(1,1)\ddots & F_f(1,N-1)\vdots \\ F_f(M-1,0) & \cdots F_f(M-1,1) & F_f(M-1,N-1) \end{bmatrix}$$

In McFarlane's $\Sigma$–$\Delta$ estimation algorithm [30], the new value $B_f(x,y)$ of background is determined by the previous background value $B_{f-1}(x,y)$ plus $sgn(F_f(x,y) - B_{f-1}(x,y))$. there new background values $B_f(x,y)$ do not consider the attribute of the original input image$F_f(x,y)$. Therefore when the moving objects are slowing down, stopping, or frequently appearing, the ghost effect are occurred in their built background images. In order to improve the ghost effect in the built background image, temporary input image $F_f^*(x,y)$. When $F_f^*(x,y)$ is not equal to$F_f(x,y)$, i.e., $F_f(x,y)$ belong to the attribute of moving object; the new background value$B_f(x,y)$ does not need to be adjusted in this frame. Otherwise, the new background value $B_f(x,y)$ must be adjusted with the $\Sigma$–$\Delta$ background estimation. Let C(x, y) be the counter for each pixel at the coordinate (x, y), α be the sampling interval of the frames and can be represented as $\mathrm{sgn}(\alpha) = 1\, if\ \alpha > 0, \mathrm{sgn}(\alpha) = -1\, if\ \alpha < 0$, $sgn(\alpha)\ = 1$ if $\alpha > 0$, $sgn(\alpha) = -1$ if α < 0, and $\mathrm{sgn}(\alpha) = 0$ *if* $\alpha = 0$ and T be the threshold of C(x, y). When C(x, y) is less than or equal to T, a new value of $(F_f^*(x, y) + B_{f-1}(x,y))$ divided by 2 is used to replace the new background

$B_f(x, y)$. It can adjust the background value $B_f(x, y)$ quickly to approach the real background value. Otherwise background video is adjusted by the sgn function with a multiple interval of α.

➢ **Algorithm of Effective Background Σ-Δ Estimation**

Input: $F_f(x, y)$

Output: $B_f(x, y)$

**For** each frame

*//Initialization*

**For** each pixel (x, y):

$F_0^*(x, y) \leftarrow F_0(x, y)$

$B_0(x, y) \leftarrow F_0(x, y)$

$C(x, y) \leftarrow 0$

  **For** each frame f and each pixel (x, y):

$$F_f^*(x, y) \leftarrow F_{f-1}^*(x, y) + sgn(F_f(x, y) - F_{f-1}^*(x, y))$$

**End For**

*//Median Adaptive Computing*

**If** f is a multiple α

**For** each pixel (x, y):

  **If** $F_f^*(x, y) == F_f(x, y)$

    **If** C(x, y) ≤ T

$$B_0(x, y) \leftarrow \frac{(F_f^*(x, y) + B_{f-1}(x, y))}{2}$$

C(x, y) ← C(x, y) +1

**End If**

  **End if**

 **End For**

     **Else** $B_f(x, y) \leftarrow B_{f-1}(x, y)+$ sgn $((F_f^*(x, y) + B_{f-1}(x, y))$

  **End if**

 **End if**

**End For**

## 2.5    Experimental Results and Discussion

For the performance of different background subtraction technique own made database and Southampton database have been used. Own made database consists of 174 frames of $720 \times 480$ spatial resolution, acquired at a frame rate of 29 fps. In this video lightning conditions are good but there is a strong shadow casted by moving object. The scene consists of static background. Southampton video consists of frame of $720 \times 576$ spatial resolution, acquired at a frame rate of 25 fps. The light variation is more. $1^{st}$ row $- 5^{th}$ row videos as shown in fig. 2.2 are naming as Database video1, Database video2, Database video3, Database video4 an Database video5respectively. In fig. 2.1 shows simple background subtraction result from own made database. Here the reference background is already available. But some open source database, reference background is not available. It demands background modeling. Table 2.1-2.4 gave a quantitative performance analysis of three different methods which are further represented in fig. 2.3-2.6.
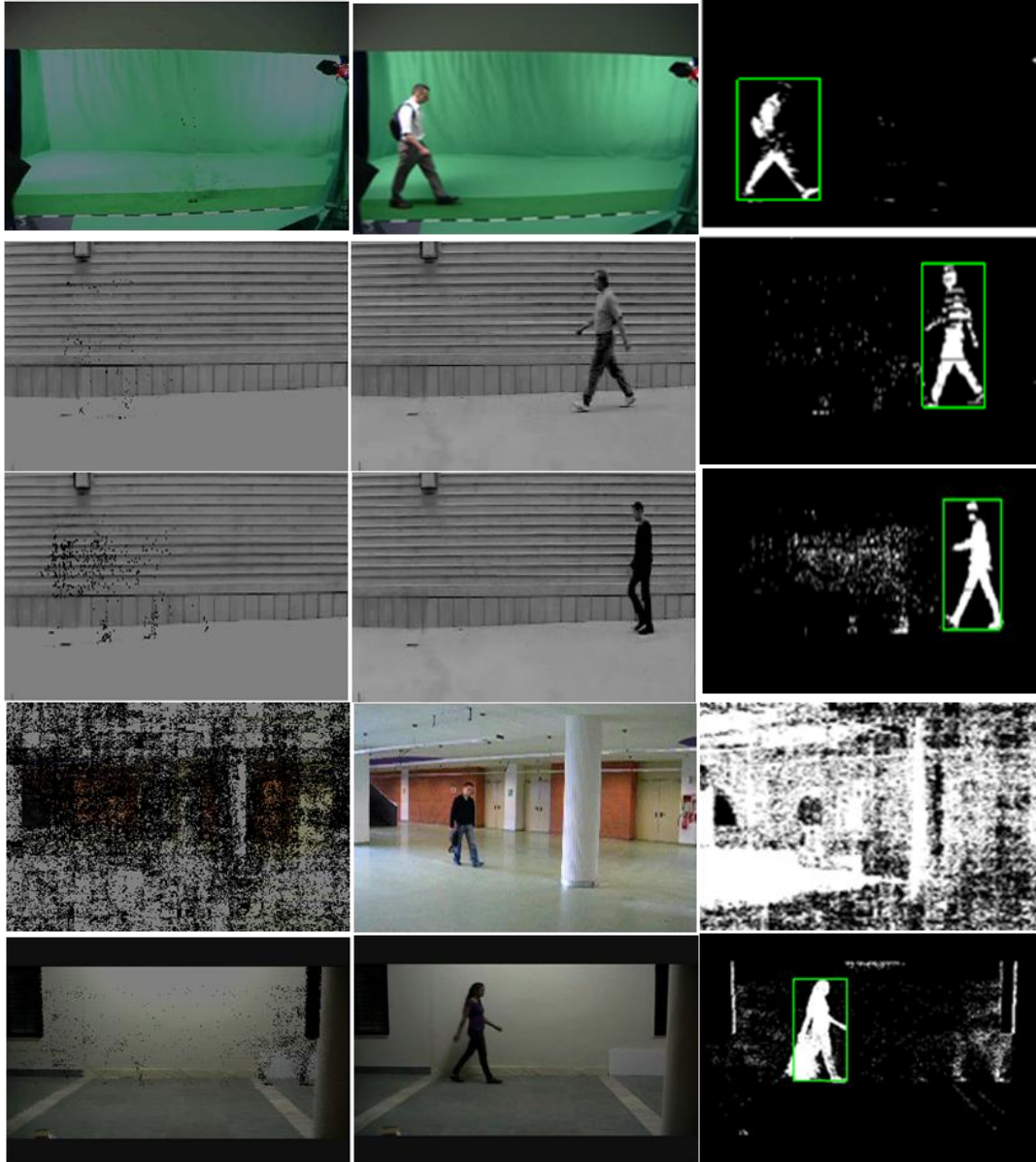


(a)                                              (b)                                              (c)
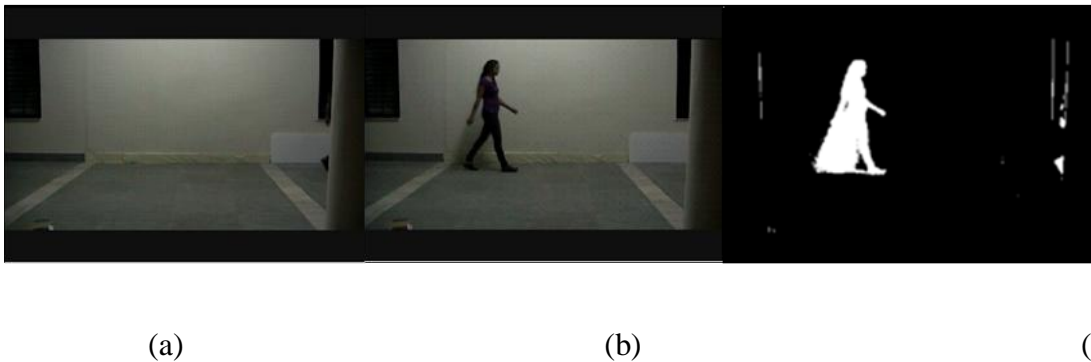
**Figure 2.1** : Simple background subtraction method: (a) Available reference back ground; (b) $48^{th}$ frame of own made database video; (c) Silhouette generated after simple background subtraction.
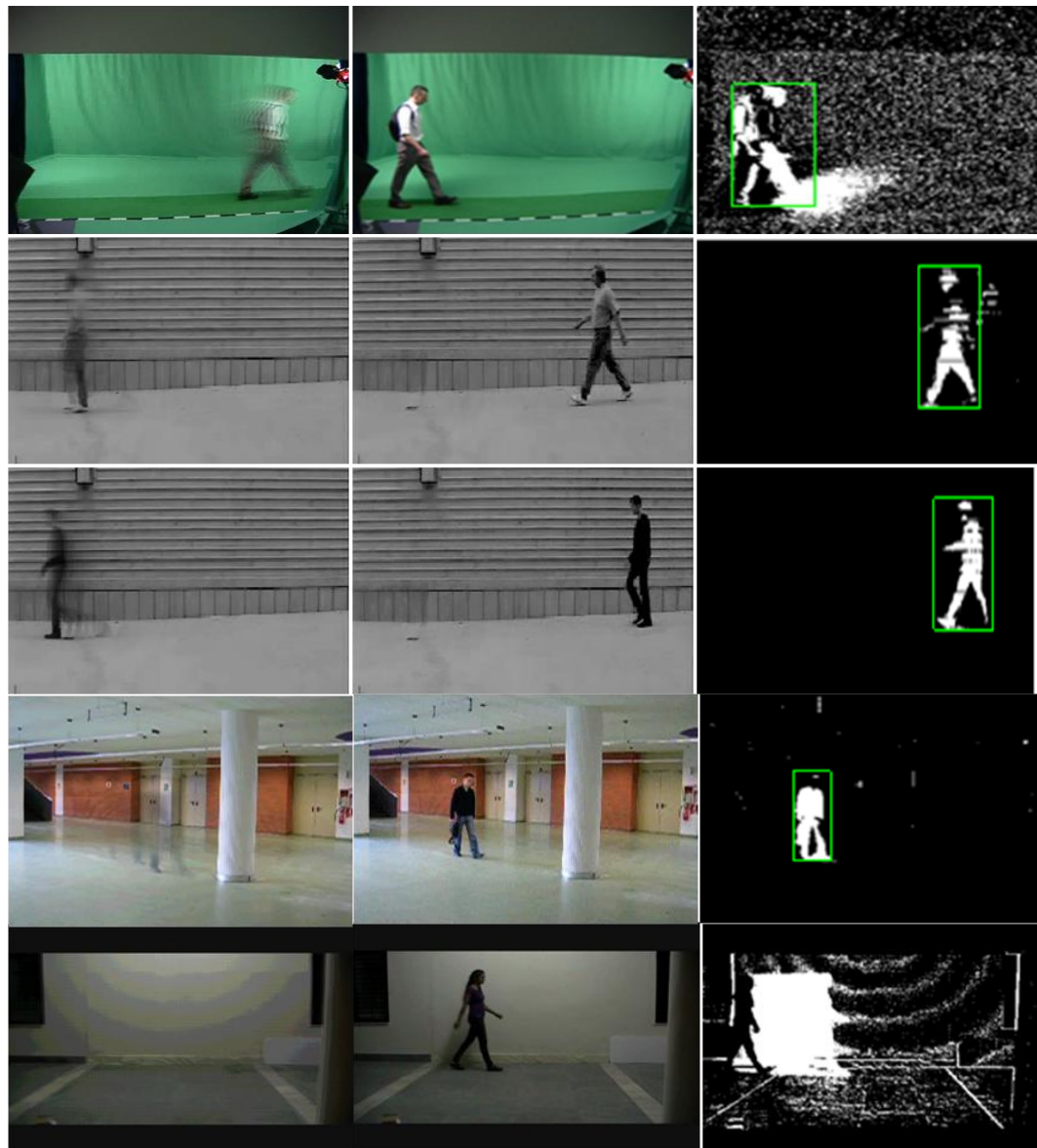
(a)
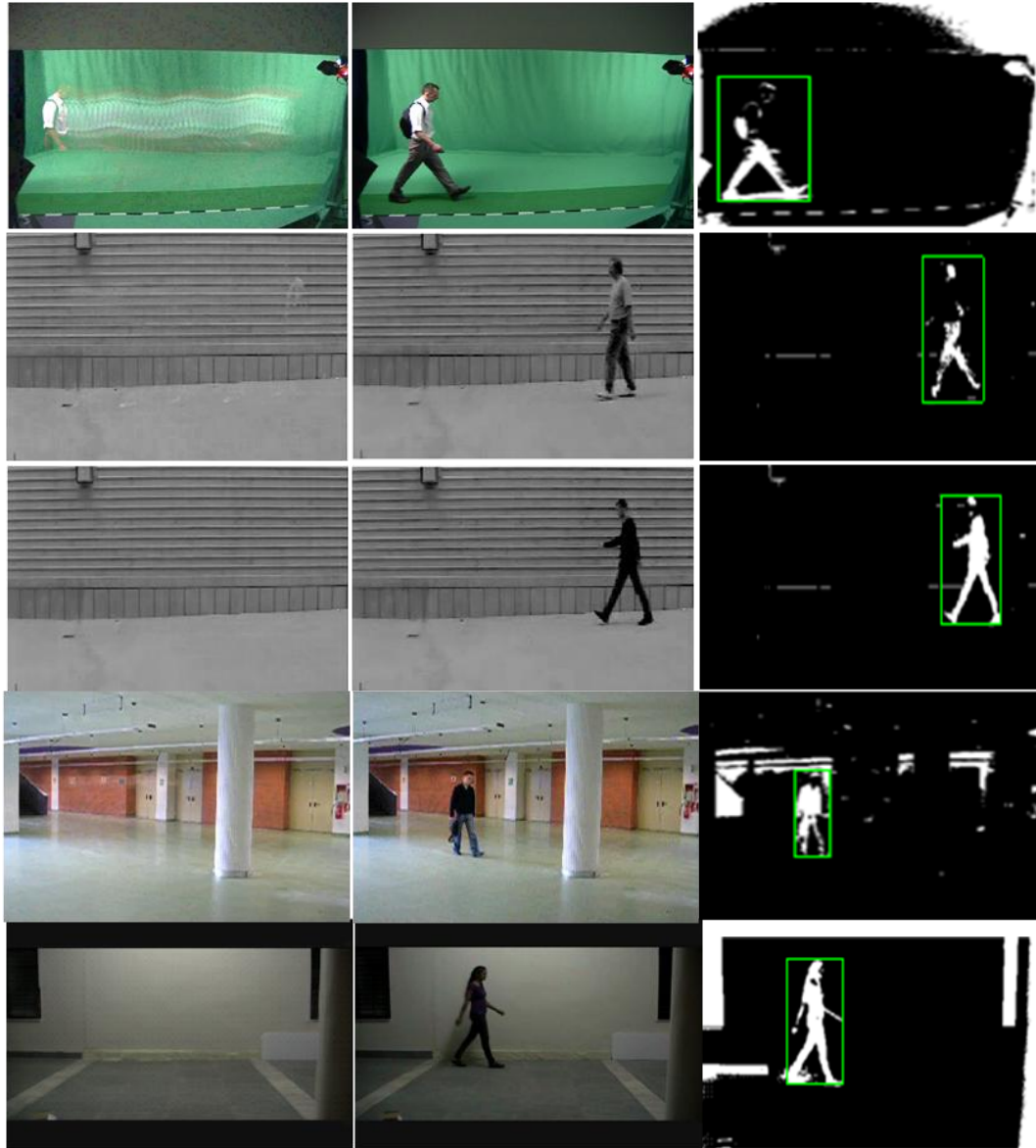
(b)

(c)

**Figure 2.2** : In first row database of Southampton, in 2$^{nd}$ and third row database gait video-1 and 5 are shown respectively. The first Column of (a) ,(b), (c) shows reconstructed background using Running average method, Sigma delta method and Effective sigma delta method respectively., 2$^{nd}$ column shows database video frames 5, 10 ,15 from top to down respectively. 3$^{rd}$ column shows silhouette detection and tracking the silhouette.

## 2.5.1 Quantitative Performance Analysis

There are different approaches to evaluate the performance of the background subtraction algorithms. In a binary decision problem, the classifier labels samples as either positive or negative. In our context, samples are pixel values, "positive" means foreground object pixel, and "negative" means background pixel. In order to quantify the classification performance, with respect to some ground truth classification, the following basic measures can be used:

• True positives (TP): correctly classified foreground pixels.

• True negatives (TN): correctly classified background pixels.

• False positives (FP): incorrectly classified foreground pixels.

• False negatives (FN): incorrectly classified background pixels.

Precision, recall, F-measure are the basic measures used in evaluating search strategies.

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$F - measure: S_F = \frac{2 \times Recall \times Precision}{(Recall + Precision)}$$

$$correct classification: S_{CC} = \frac{TP + TN}{TP + TN + FP + FN}$$

Table 2.1: Pixel based accuracy result for video1 database

|  | Recall | Precision | F- measure | Correct Classification |
|---|---|---|---|---|
| Running Average | 0.5782 | 0.4668 | 0.5165 | 0.635 |
| Sigma delta | 0.4589 | 0.8077 | 0.5852 | 0.78 |
| Effective Sigma delta | 0.7355 | 0.8150 | 0.7732 | 0.855 |

Table 2.2: Pixel based accuracy result for vodeo2 database

|  | Recall | Precision | F1 measure | Correct Classification |
|---|---|---|---|---|
| Running Average | 0.1863 | 0.0750 | 0.1069 | 0.532 |
| Sigma delta | 0.259 | 0.896 | 0.4018 | 0.713 |
| Effective Sigma delta | 0.5504 | 0.8546 | 0.6695 | 0.798 |

Table 2.3: Pixel based accuracy result for video3 database

|  | Recall | Precision | F1 measure | Correct Classification |
|---|---|---|---|---|
| Running Average | 0.7151 | 0.8417 | 0.7732 | 0.872 |
| Sigma delta | 0.7743 | 0.9704 | 0.86 | 0.924 |
| Effective Sigma delta | 0.8763 | 0.9608 | 0.9166 | 0.9514 |

Table 2.4: Pixel based accuracy result for video4 database

|  | Recall | Precision | F1 measure | Correct Classification |
|---|---|---|---|---|
| Running Average | 0.985 | 0.439 | 0.607 | 0.35 |
| Sigma delta | 0.72 | 0.959 | 0.85 | 0.89 |
| Effective Sigma delta | 0.94 | 0.67 | 0.78 | 0.824 |
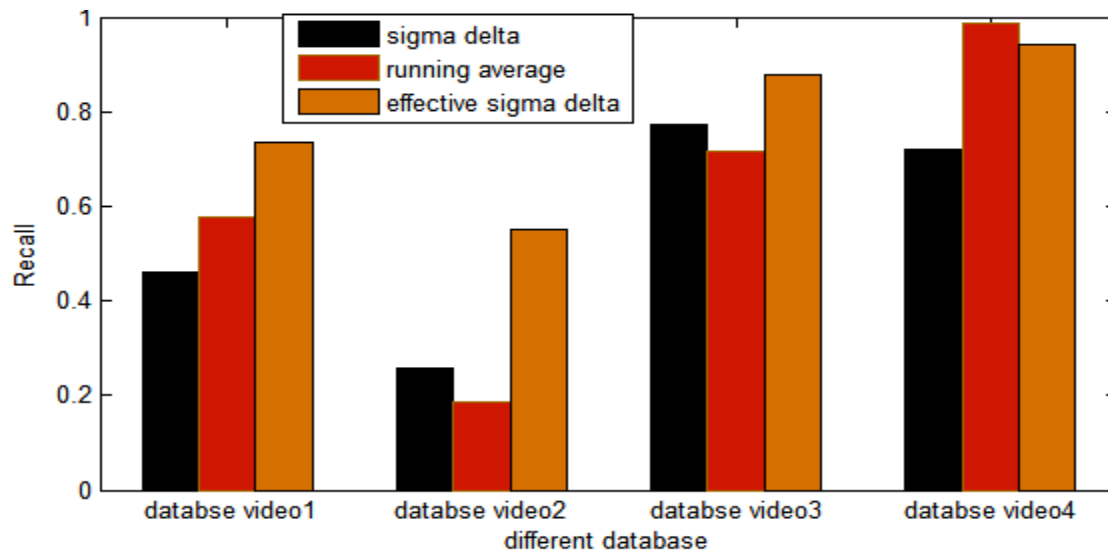


**Figure 2.3** : Recall bar for four different videos
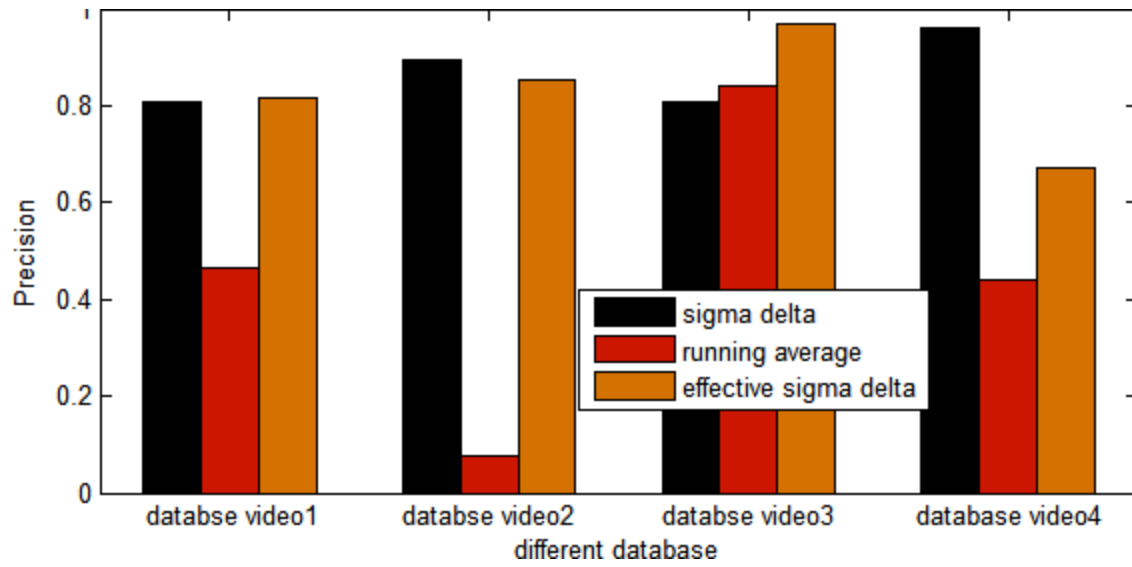
**Figure 2.4** : Precision bar for four different videos
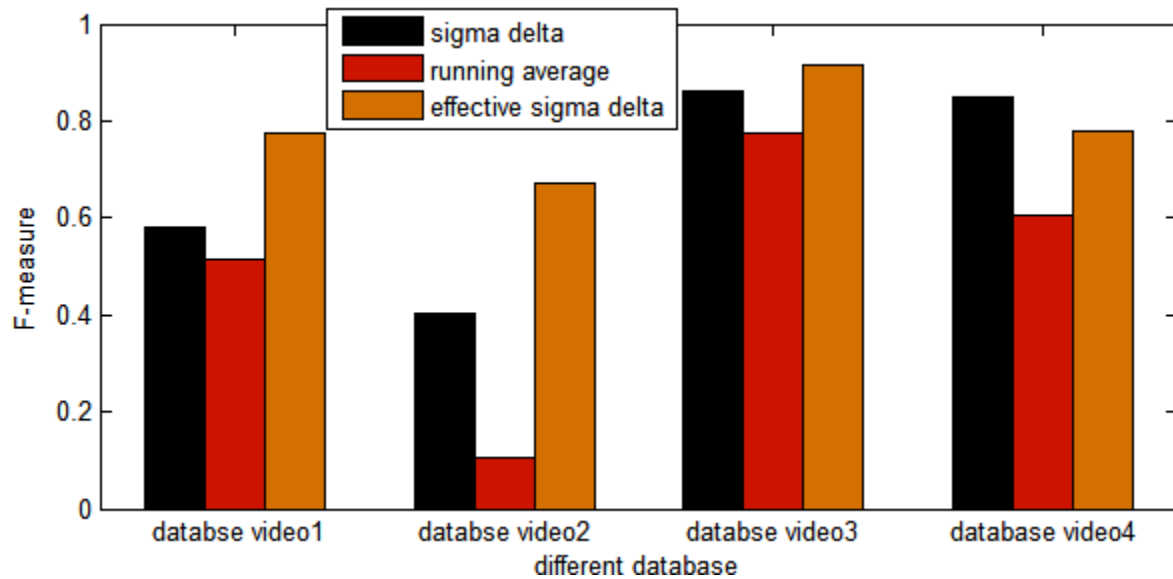


**Figure 2.5** : F-measure bar for four different videos
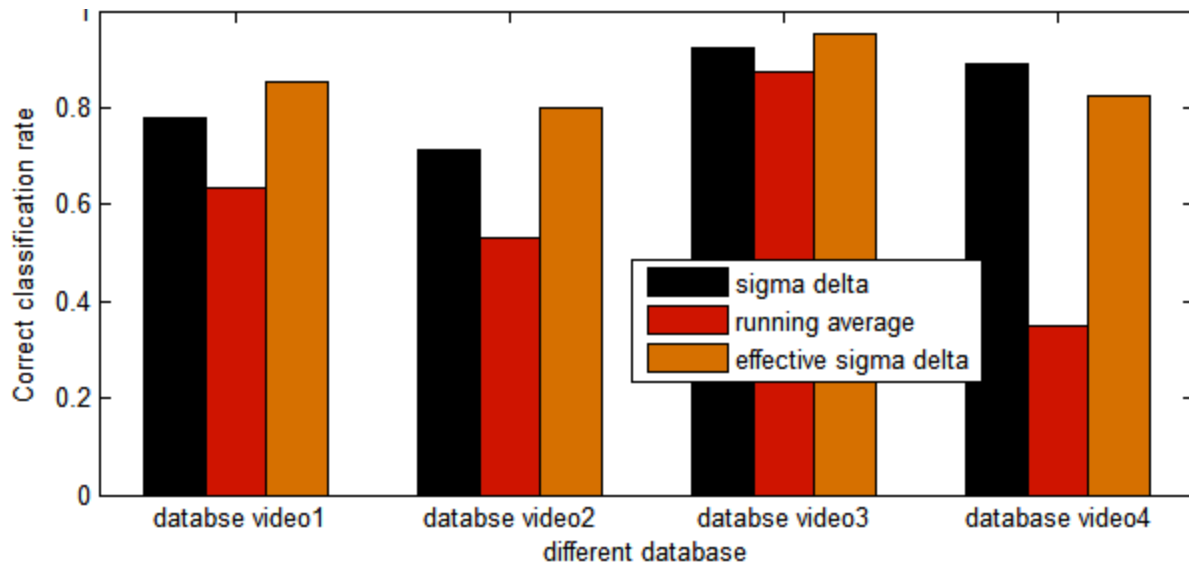
**Figure 2.6**: Correct classification bar for four different videos

## 2.6    Conclusion

In this chapter different background subtraction algorithms have been discussed. Sigma delta method gives better background model as compared to other methods for static background where the light variation does not occur when the intensity of light changes then effective sigma delta method gives better result. Quantitative performance analysis compared between different videos using different methods. The effective sigma delta method gives high recall, precision, F measure and better correct classification rate as compare to other methods.

# Chapter-3

# Shadow Detection

As discussed in the previous chapter, Background subtraction lacks capability to remove shadows. In gait recognition background subtraction is not sufficient to track a human during walking but is more oriented to robust shape detection, even silhouette losses (due to the shadow cast over silhouette i.e. foreground). Shadow suppression models are helped to achieve this goal. This chapter represents moving shadow elimination methods using different colour spaces. It covers shadow detection method using RGB colour constancy within pixel; shadow eliminating operator using YCbCr colour space; shadow suppression using HSV colour information. Quantitative Performance analysis is done on own made and publicly available databases. Experimental results are shown at the end of the chapter.

## 3.1    Introduction

In realistic environments, the main problem in motion detection is how to distinguish between object and its moving shadow [24]. The moving shadows can affect the correct shape, position, measurements, and detection of moving objects. In particular, all the moving points of both objects and shadows are detected at the same time. Moreover, shadow points are usually next to object points and in most segmentation techniques shadows and objects are merged in a single blob. This cause two important drawbacks: the former is that the human shape is falsified by

shadows and all the measured geometrical properties are affected by an error (that varies during the day and when the luminance changes). This affects the feature extraction process of thigh angle, stride length, cadence, etc. Colour videos are the main format in most video application so moving shadow is detected and eliminated in colour space. In RGB colour space Kaew TraKul Pong and Bowden [25] found two properties of shadow, i.e. pixel value in the moving shadow region is darker than in background scene and other is the statistic property of the shadow region has a little variation in their attributes. In HSV colour space, Cucchiara et al. [26] eliminated vehicle's moving shadow by using invariance of chrominance. There are three basic facts with moving shadow detection and elimination in colour spaces. Firstly there are different classes of shadow due to various scenes and different properties of shadow. Secondly shadow in various colour spaces is different, which could bring on different detection results. Thirdly there are different set of threshold values for different mean of intensity of the shadow and background regions.

## 3.2    Classification of Shadow

A shadow is generally divided into two types, Static and dynamic shadow. Static shadow does not suffer motion detection because static shadow can be modelled as part of the back-ground. Dynamic shadow is taken by the moving vehicle, pedestrian and so on.

Human Vision System using moving shadow is be adapted the following conditions

- Shadow is the projection of moving object in background.
- Shadow always related with moving object. It reflects corresponding motion and behaviours of object.
- The shape of moving shadow could change with motion every time.

- The pixel values of shadow are darker than that of surrounding pixel or object.
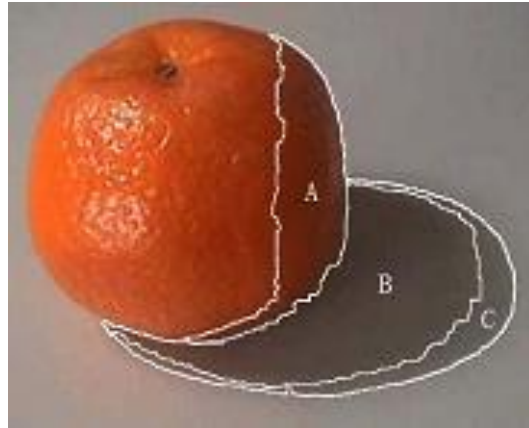


Figure 3.1 : Shadow Classification [29]

A shadow is an area where direct light from a light source cannot reach due to obstruction by an object. It is due to the occlusion of light source by an object in the scene.

Shadows are classified as two types

- Self shadow
- Cast shadow

## 3.2.1 Self Shadow

Self shadow occurs when the part of the object is not illuminated as shown in Fig 3.1 (region A). Penumbra. We are interested in cast shadow than self-shadow. For video surveillance it is not important that which region is umbra or penumbra, so shadow should be reclassified. In this paper, shadow is classified as follows.

- Invisible shadow
- Visible shadow

### 3.2.2 Cast shadow

The area projected on the scene by the object as shown in Fig.3 (region B and C) are called cast shadow. It can be dived into **Umbra** (dark shadow, Fig 3.1 region B) and **Penumbra**(soft shadow in Fig 3.1, region C).The part of the shadow where the direct light is only partly blocked by the object known as penumbra is shown in Fig 3.2.
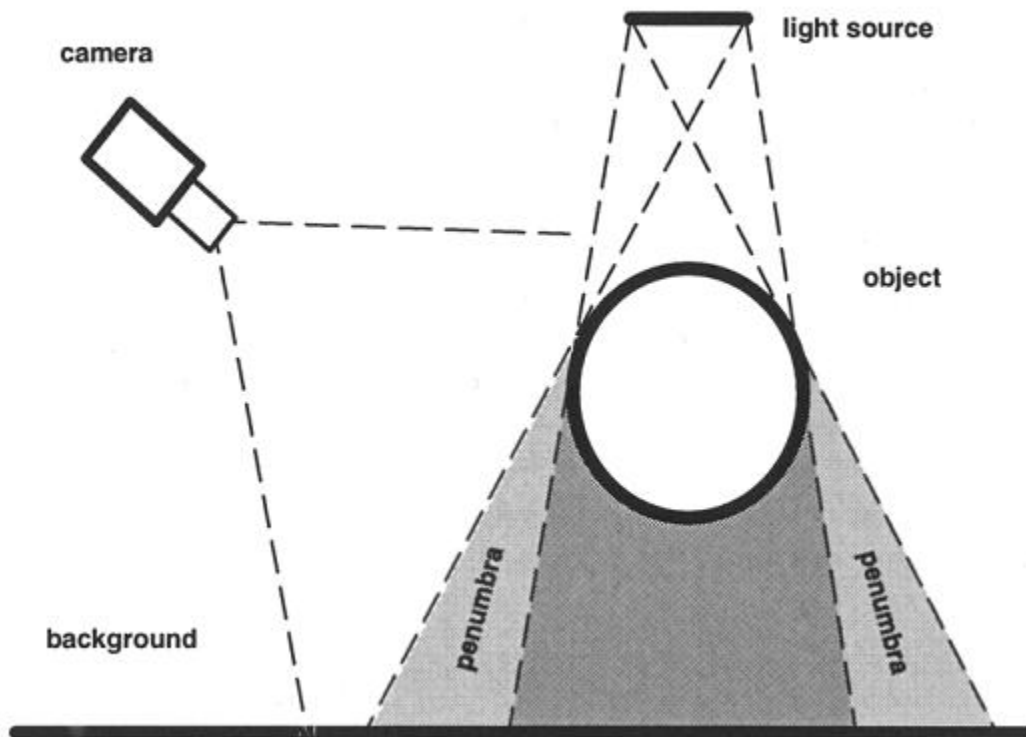


**Figure 3.2: Cast shadow generation: The scene grabbed by a camera consists of a moving object and a moving cast shadow on the background. The shadow is caused by a light source of certain extent and exhibits a penumbra.**

If light source is fixed, when objects move, not only self-shadow but also cast shadow is changed every time. Self-shadow is part of object, so it should not be removed during motion detecting. The shadow remove algorithms should eliminate the effects of cast shadow.

## 3.3  Shadow Analysis

A cast shadow can be described [23] as:

$$s_k(x, y) = E_k(x, y)\rho_k(x, y) \tag{3.1}$$

Where $s_k$ is the image luminance of the point of coordinate (x, y) at time instant k. $E_k(x, y)$ is the irradiance. It is computed as follows:

$$E_k(x, y) = \begin{cases} c_A + c_p \cos \angle(N(x, y), L) & illu\,\min ated \\ c_A & Shadowed \end{cases} \tag{3.2}$$

Where $c_A$ and $c_P$ are the intensity of the ambient light and that of the light intensity, respectively,

$\angle$ the direction of the light source and N(x, y) the object surface normal. $\rho_k(x, y)$ is the reflectance of the object surface.

1st step for shadow detection is the difference between the current frame and the reference image. The reference frame may be a previous frame or the reference image. Using e (3.1)the difference $D_k(x, y)$ can be written as

$$D_K(x, y) = S_{K+1}(x, y) - S_K(x, y) \tag{3.3}$$

k+1 is the previously illuminated frame which is covered by a cast shadow. According to the static background hypothesis, reflectance $\rho_k(x, y)$ of the background does not change with time, thus assume that

$$\rho_{k+1}(x, y) = \rho_k(x, y) = \rho(x, y) \tag{3.4}$$

Then, eq. 3.3 can be rewritten (using eqs. 3.1, 3.2 and 3.4) as [27]

$$D_k(x, y) = \rho(x, y)c_p \cos \angle(N(x, y), L) \qquad (3.5)$$

This implies (as assumed in many papers) that shadow points can be obtained by thresholding the frame difference image using eq.(3.5) .

Some Shadow hypotheses on the environment are outlined.

1. Strong light source

2. Static background (and camera)

3. Planer back ground

## 3.4  Useful Features for Shadow Detection

Most of the following features are useful for detecting shadows when the frame, which contains objects and their shadows, can be compared with an estimation of the background, which has no objects or moving cast shadows.

### 3.4.1  Intensity

The simplest assumption that can be used to detect cast shadows is that regions under shadow become darker as they are blocked from the illumination source. Furthermore, since there is also sudden change of illumination, there is a limit on how much darker they can become. These assumptions can be used to predict the range of intensity reduction of a region under shadow.

### 3.4.2  Chromacity

Most shadow detection methods based on spectral features use color information. They use the supposition that regions under shadow become darker but retain their chromacity. Chromacity is

a measure of color that is independent of intensity. The color transition model where the intensity is reduced but the chromacity remains the same is normally referred to as color constancy or linear attenuation [28]. Methods that use this model for detecting shadows often choose a color space with better separation between chromacity and intensity than the RGB color space.

## 3.5    Shadow Elimination Models

## 3.5.1  RGB Colour Constancy within Pixels Model

The spectral property is called color constancy within pixel [25]. A shadow casting on a background, pixel changes its brightness where as its color roughly remains same. A comparison of pixel-wise color information between current image and background image can help in shadow detection. The brightness and color information of a pixel can be divided by transferring the color space from RGB to a well-known normalized R-G using the following two equations.

$$
\begin{aligned}
C_r(x, y) &= \ln \frac{I_R(x, y)}{I_R(x, y) + I_G(x, y) + I_B(x, y)} \\
C_g(x, y) &= \ln \frac{I_G(x, y)}{I_R(x, y) + I_G(x, y) + I_B(x, y)}
\end{aligned}
\tag{3.6}
$$

Since the values of $C_r$ and $C_g$ remain roughly the same under different illumination conditions, the score of error for discriminating the pixel (x, y) as shadow is defined as

$$
\wedge(x, y) = \left| C_r(x, y) - C'_r(x, y) \right| + \left| C_g(x, y) - C'_g(x, y) \right|
\tag{3.7}
$$

Where $C$ contains the colour information of the current image and $C'$ contains the colour information of the background image. A smaller $\Lambda(x, y)$ represents that the colour of the pixel (x, y) does not change much, and it is more likely to be a shadow pixel.

### 3.5.2 Shadow Eliminating Operator using $YC_bC_r$

Shadow eliminating operator is designed here with grey and gradient properties of shadow. $YC_bC_r$ color space s is better for removing shadow. Because the luminance of shadow is lower than background and difference of gradient density between shadow and background is lower than that between the object and background.

The shadow elimination method is employed as follows

$$M_1 = \begin{cases} 1, \left| I^k_{C1}(x,y) - B^k_{C1}(x,y) \right| > D_{th} \\ 0, \textit{ otherwise} \end{cases}$$

$$M_2 = \begin{cases} 1, \left| G^k_{C2}(x,y) - G^k_{BC2}(x,y) \right| > D_{th1} \\ 0, \textit{ otherwise} \end{cases}$$

$$M_3 = \begin{cases} 1, \left| G^k_{C3}(x,y) - G^k_{BC3}(x,y) \right| > D_{th1} \\ 0, \textit{ otherwise} \end{cases}$$

(3.8)

$$M = M_1 \wedge (M_2 \vee M_3)$$

Where $M_1$ is the mask of moving region, $M_2$ and $M_3$ are the candidate edges, M is the edges of object,$I^k_{C1}$ and $B^k_{C1}$ are the *kth* frame and its background in $C_1$ channel, $G^k_{C2}$ and $G^k_{C3}$ are the gradient function in $C_2$ and $C_3$ channels. $G^k_{BC2}$ and $G^k_{BC3}$ are the gradient function of background in $C_2$ and $C_3$ channels, and $D_{th}$, $D_{th1}$ , $D_{th2}$ are predefined parameters. In $M_2$ and $M_3$ objects are not whole and in $M_1$ objects are whole including shadow. In $M$ only whole object is found. $C_1$ Channel is the moving region and $C_1$, $C_2$ and $C_3$ channels are the gradient differences. The absolute difference of the luminance between the background and the foreground which is greater than threshold $D_{th}$ and the pixel which gradient difference is lower than a threshold is considered as the candidate edge of shadow.

### 3.5.3  Shadow Suppression using HSV colour Information

The HSV color space corresponds closely to the human perception of color[26].This chapter used HSV color space to distinguish shadows than the RGB space. Thus, Sakbot [shadow] tries to estimate how the occlusion due to shadow changes the value of H, S and V. The moving objects points are only analysed, i.e. that are detected with a high difference according to eq (3.9). A cast shadow point darkens the background point, whereas an object point could darken it or not, depending on the object color texture.

A shadow mask $SP_K$ for each moving points (x, y) with three conditions as follows:

$$
SP_K = \begin{cases} 1, & if \ \alpha \leq \dfrac{I^V{}_K(x, y)}{B^V{}_k(x, y)} \leq \beta \\ \wedge (I^S{}_K(x, y) - B^S{}_K(x, y)) \leq \tau_s \\ \wedge (I^H{}_K(x, y) - B^H{}_K(x, y)) \leq \tau_H \\ 0 & otherwise \end{cases} \tag{3.9}
$$

The first condition works on the luminance (the V component). $I_K^V$ The intensity value for the component V of the HSV pixel at coordinates (x, y) in the frame k. $B_K^V$ The intensity value for the component V of the HSV pixel at coordinates (x, y) in the background frame k. $I_K^S$ The intensity value for the component S of the HSV pixel at coordinates (x, y) in the frame k. $B_K^S$ The intensity value for the component S of the HSV pixel at coordinates (x, y) in the background frame k. $I_K^H$ The intensity value for the component H of the HSV pixel at coordinates (x, y) in the frame k. $B_K^H$ The intensity value for the component H of the HSV pixel at coordinates (x, y) in the background frame k.

α: it gives how strong the light source is with regard to the reflectance and irradiance of objects.
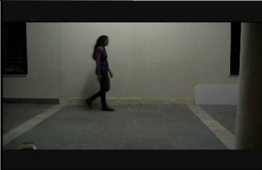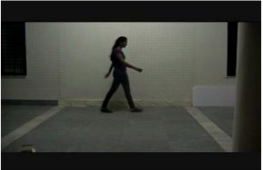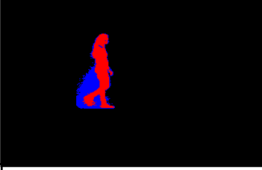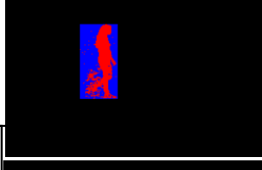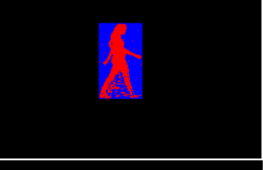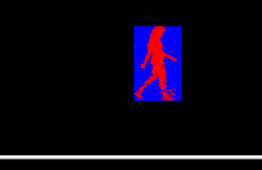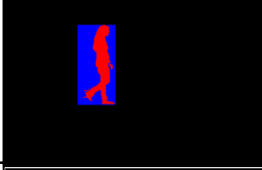
**β**: The use of **β** prevents the identification as s points where background was slightly changed by noise.

## 3.6    Experimental Results and Discussion
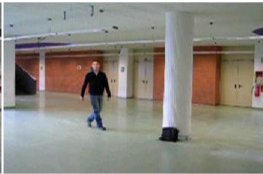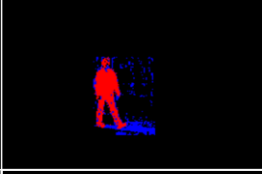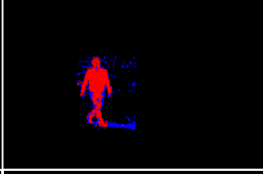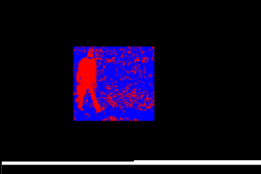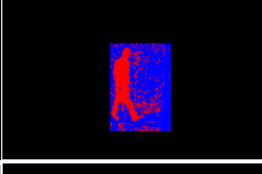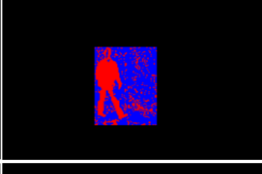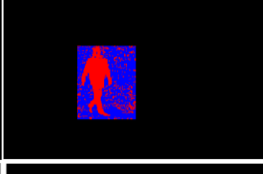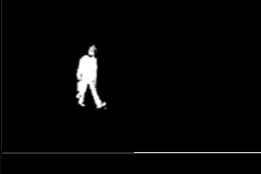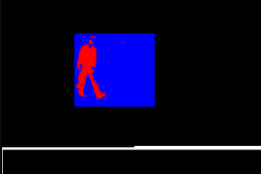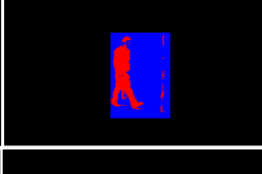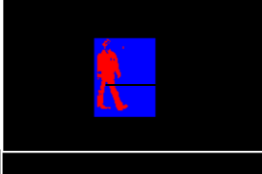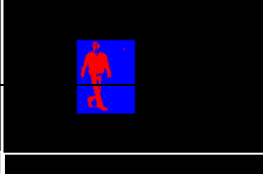
The three color space operators are implemented on three typical database sequences. The first sequence is own made database which are taken both day and night time on intensity variation conditions. MSA and Southampton sequences are taken in noise conditions and light fluctuation conditions respectively. All the above sequences are typical indoor environment. The main aim of the experiment is to separate moving cast shadow from human motion. So that good gait feature extraction is being extracted.

The original sequences are given together with the detection results in figs. (3.3-3.5),with the detected foreground objects and moving shadows being depicted in red and blue, respectively.

In the first row of  figs.( 3.3–3.5), the original sequences are listed;  HSV color information model and shadow remove using eq (3.9) are shown in $2^{nd}$ and $3^{rd}$ rows respectively; the detection results with shadow removal silhouette of eq.(3.6- 3.8) are shown in $4^{th}$ and $6^{th}$ rows respectively. In table (3.1-3.3) gives quantitative analysis and comparison on basis of shadow detection rate on three different databases. The graph of figs.3.6 (a-c) gives comparison analysis on three color spaces.

| | | | | |
|---|---|---|---|---|
| **Original images** | | | | |
| **Shadow is in blue using HSV space** | | | | |
| **Remove shadow Using HSV** | | | | |
| **Shadow is in blue using RGB space** | | | | |
| **Remove shadow Using RGB** | | | | |
| **Shadowis in blue using YCbCr space** | | | | |
| **Remove shadow Using YCbCr** | | | | |

**Figure 3.3** Experimental results on own made database

| | | | | |
|---|---|---|---|---|
| **Original Images** | | | | |
| **Shadow is in blue using HSV** | | | | |
| **Remove shadow Using HSV** | | | | |
| **Shadowis in blue using RGB** | | | | |
| **Remove shadow Using RGB** | | | | |
| **Shadow is in blue using YCbCr** | | | | |
| **Remove shadow UsingYCbCr** | | | | |

**Figure 3.4 :** Comparison of shadow detection rate of hsv, rgb, ycbcr color space on MSA database

| | | | | |
|---|---|---|---|---|
| **Original Images** | | | | |
| **HSV** | | | | |
| **Remove shadow Using HSV** | | | | |
| **Shadow is in blue using RGB** | | | | |
| **Remove shadow Using RGB** | | | | |
| **Shadow is in blue using YCbCr** | | | | |
| **Remove shadow Using YCbCr** | | | | |
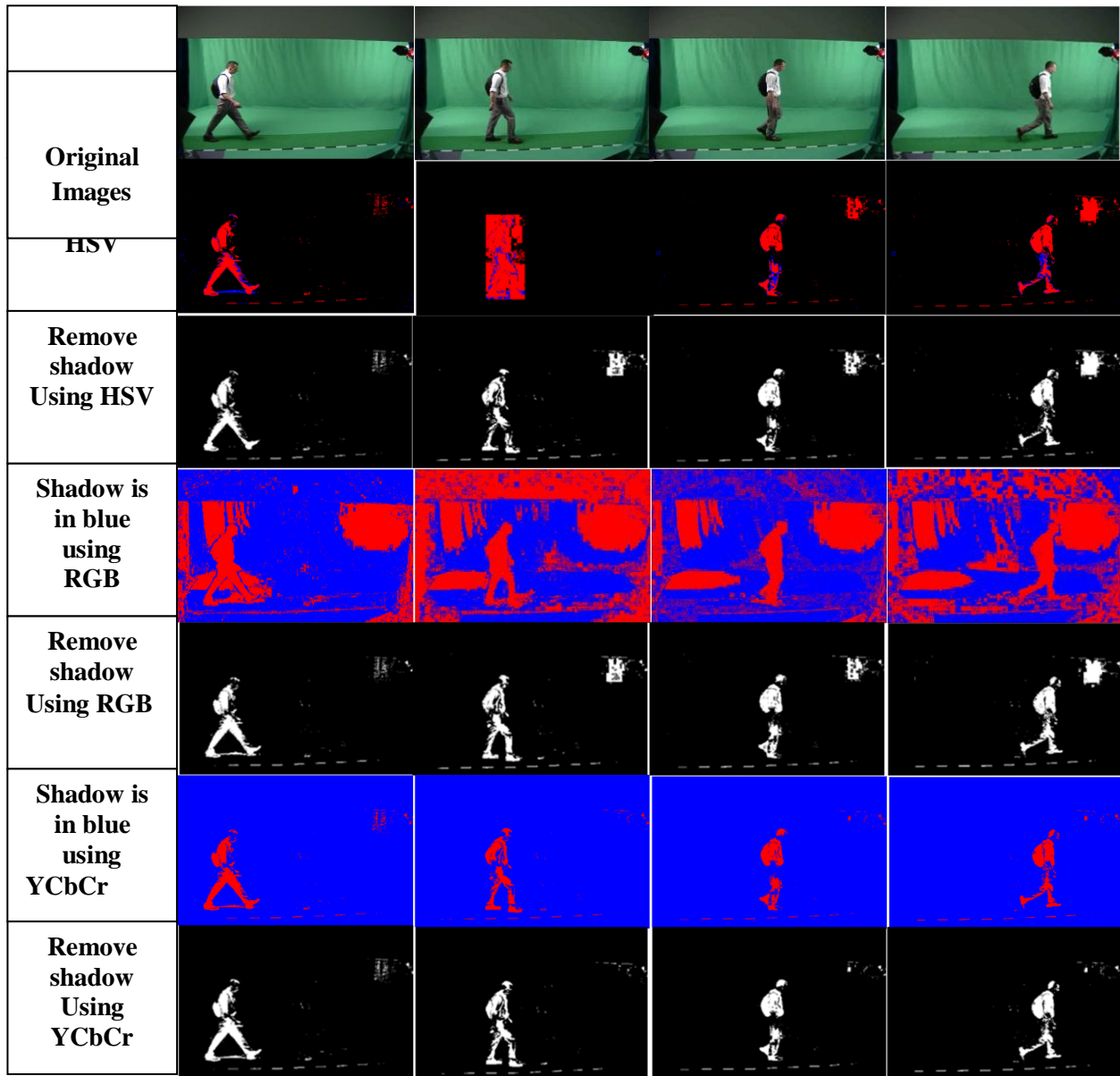
**Figure 3.5** : Comparison of shadow detection rate of hsv, rgb, ycbcr color space on Southampton database

**Table 3.1** Quantitative evaluations and Comparison on own made database

| Shadow Detection Rate (%) | Frame no: | 10 | 15 | 20 | 25 |
|---|---|---|---|---|---|
| RGB | | 54.54 | 55.1245 | 63.2620 | 65.3524 |
| YCbCr | | 81.42 | 81.1345 | 87.1232 | 86.41 |
| HSV | | 81.29 | 84.2519 | 88.2486 | 87.1 |

**Table 3.2** Quantitative evaluation and comparison on MSA database

| Shadow Detection Rate (%) | Frame no: | 10 | 15 | 20 | 25 |
|---|---|---|---|---|---|
| RGB | | 85.3302 | 84.3 | 86.4 | 85.432 |
| YCbCr | | 81.33 | 83.45 | 84.5 | 82.41 |
| HSV | | 74.31 | 77.43 | 79.12 | 76.4 |

**TABLE 3.3** Quantitative evaluation and comparison on Southampton database

| Shadow Detection Rate (%) | Frame no: | 3 | 15 | 25 | 35 |
|---|---|---|---|---|---|
| RGB | | 53.098 | 48.114 | 47.23 | 46.55 |
| YCbCr | | 58.154 | 53.23 | 46.119 | 47.421 |
| HSV | | 64.09 | 55.459 | 48.013 | 48.4521 |
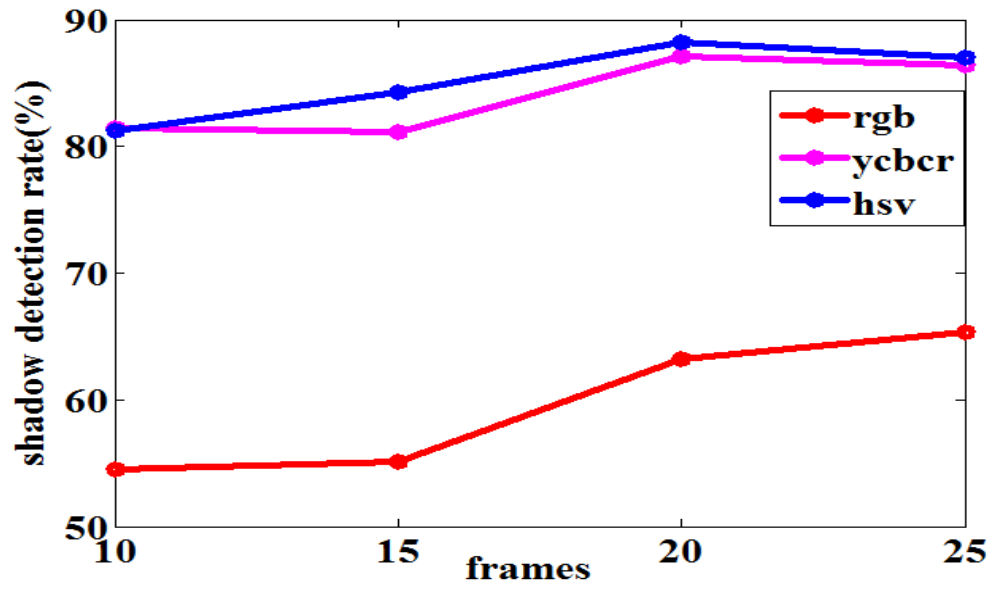
**(a)**
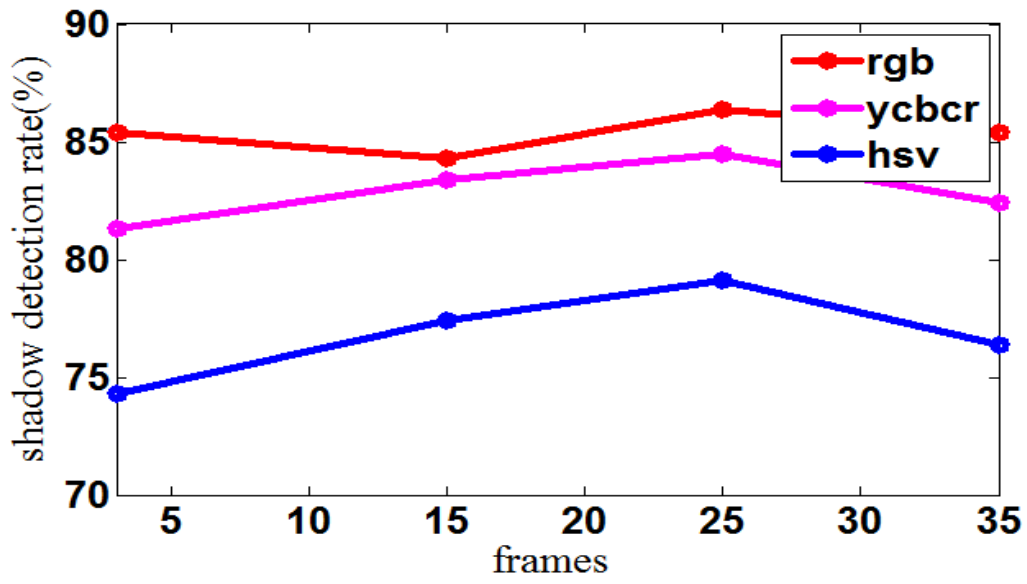


**(b)**
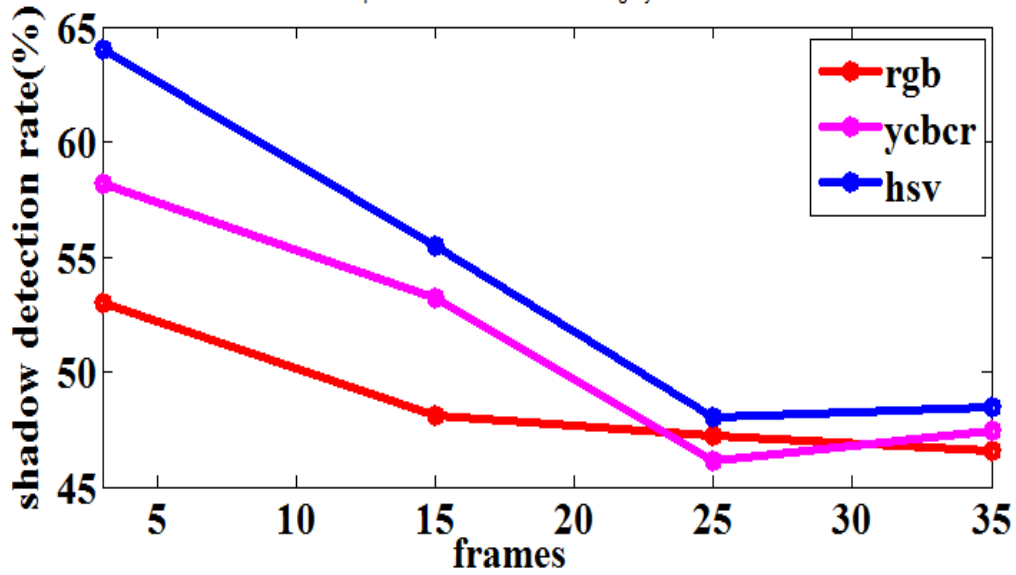
**(c)**

**Figure 3.6 :** Comparison of shadow detection rate of HSV, RGB, YCbCr color space on: (a) own made, (b) MSA and (c) Southampton database

## 3.7 Conclusion

In this chapter, RGB colour constancy pixel model, Shadow suppression operator using YCbCr model and shadow suppression using HSV colour information models have been discussed. The three models have been examined using own made and open source databases. The experimental results showed that HSV colour information model detects shadow region efficiently than other two colour models. The threshold values $(\alpha, \beta)$ of HSV model has been again analysed to optimize its range with respect to the intensity value of the blob to avoid arbitrarily taken its value in trial and error basis.

# Chapter- 4

# Feature Extraction via Silhouette based Method

Height is one of the important features from the several gait features which is not influenced by the camera performance, distance and clothing style of the subject. Height variation during walking process will give identity recognition. This chapter describes the proposed silhouette-based method deployed to derive the height variation process of walking subjects. Calibration Process is applied here to estimate the height at any distance. We present Direct Linear Transformation method (DLT) to transform 3-D information to 2-D information which makes the system computationally efficient for real time implementation. Experimental results of height variation and stride length variation of different subjects during gait are drawn at the end of the chapter.

## 4.1    Introduction

In gait feature extraction, silhouette based method is commonly adopted [18]. In silhouette based method, recognizing a person by gait intuitively depends on how the silhouette shape of an individual changes over time in an image sequence. A new method is proposed here to extract some gait features. This proposed method calculates the variation of height during walking. The measurement of height of a person is not affected by his clothing style as well as the distance from the camera. At any distance the height can be measured, but for that camera calibration is essential. Previously Lee et al. [27] used DLT method to estimate the height of a stationary person. So he only used intrinsic parameter of the camera. In the proposed method we used both intrinsic as well as extrinsic parameters of the camera because extrinsic parameter gives both translational and rotational matrix. Then DLT method is used to find the height of a moving person for each frame. The height is varying for each movement of subject. Another feature known as stride parameters are estimated using step length in terms of pixel. The step length is calculated in terms of width of bounding box of selected blobs. This gives stride length and gait cycle.

## 4.2    Camera Calibration

The goal of Camera Calibration is to determine the parameters of the camera. There are two types of camera parameters such as intrinsic and extrinsic [28]. Intrinsic or internal camera parameters are those, which describe the projection of objects onto the camera image. They establish the relationship between the points in the camera reference frame and the pixel coordinates of the points on the images got from the camera. Extrinsic or external camera parameters describe the location and the orientation of the camera. They establish the

relationship between the camera reference frame (coordinate system) and the world reference frame. Determining these parameters basically means finding the transformation (translation and rotation) which reconciles the 3-D coordinate system of the camera and that of the real world. Here both intrinsic as well as extrinsic parameters are used because although static camera is used here but the subjects are moving. In this paper, Camera Calibration Toolbox developed by Jean-Yves Bouguet [28] is used. A set of 9 monochrome test images was taken from the Cannon LEGRIA FS305 camera. Test images featured a planar checkerboard grid differently oriented in the each image. The images are loaded into the toolbox so that the grid corner cans can be extracted. Camera calibration Toolbox features an algorithm that uses the extracted corner points of the checkerboard pattern to compute a projective transformation between the image points of the n different images. It counts the number of squares in the grid. It is necessary to accurately mark the four corners, at most 5 pixels away from the corners. Otherwise some of the corners might be missed by the detector. Afterwards it is necessary to enter the dimensions of the each grid. In the bellow figure square size is 23×23 mm. The image corners are then automatically extracted to an accuracy of about 0.1 pixel. The process of corner extraction can be repeated as many times if the lens distortion is high. After the corner extraction calibration is performed in two steps: the initialization and the nonlinear optimization. The initialization step computes a closed-form solution for the calibration parameters based not including any lens distortion. The non-linear optimization step minimizes the total re-projection error shown in Fig 4.1 in the least squares sense over all the calibration parameters.

Afterwards, the camera intrinsic and extrinsic parameters are recovered using a closed-form solution. While the third and fifth order radial distortion terms are recovered within a linear least-

squares solution. A final nonlinear minimization of the re projection error, solve using a Liebenberg-Marquardt method, refines all the recovered parameters.
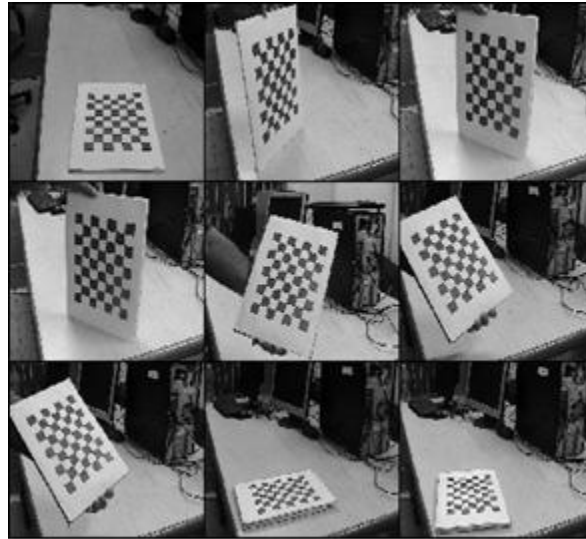


**Figure 4.1**: Images are used in calibration process

The camera matrix is formed using both the parameters [7]. The matrix converts 3D image to 2D image using (4.1).

$$\begin{pmatrix} u \\ v \end{pmatrix} = K\begin{bmatrix} R & t \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \qquad (4.1)$$

$$K = \text{Intrnisic matrix} = \begin{bmatrix} f_x & 0 & C_x \\ 0 & f_y & C_y \\ 0 & 0 & 1 \end{bmatrix} \qquad (4.2)$$

$f_x, f_y$ are the focal length in x and y direction

$C_x, C_y$ are the principal point in x and y direction

$X, Y$ and $Z$ are the object coordinate in the real world

$[R\ t]$ = joint rotation –translation matrix.

$u, v$ are the image coordinates

It is the combination of 3×3 rotation matrix and 1×3 Translation matrix.

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \quad t = \begin{bmatrix} x_t & y_t & z_t \end{bmatrix} \tag{4.3}$$

as in 2-D case, a homogeneous transformation matrix is defined, for 3-D case, a 3×4 matrix is

obtained that performs the rotation given by R ($\alpha$, $\beta$, $\Upsilon$), followed by a translation given by $x_t$, $y_t$,

$z_t$.

$$T = \begin{bmatrix} c\alpha c\beta & c\alpha s\beta s\gamma - s\alpha c\gamma & c\alpha s\beta c\gamma + s\alpha s\gamma & x_t \\ s\alpha c\beta & s\alpha s\beta s\gamma + c\alpha c\gamma & s\alpha s\beta c\gamma - c\alpha s\gamma & y_t \\ -s\beta & c\beta s\gamma & c\beta c\gamma & z_t \end{bmatrix} \tag{4.4}$$

where $c\beta = \cos(\beta)$, $c\alpha = \cos(\alpha)$, $c\gamma = \cos(\gamma)$

$s\alpha = \sin(\alpha),\ s\beta = \sin(\beta),\ s\gamma = \sin(\gamma)$

$\alpha$ =yaw angle, $\beta$ = pitch angle, $\gamma$ = Roll angle

T is the homogeneous transformation matrix

X, Y and Z are the object coordinates in the real world.

## 4.2.1  Camera Calibration Result

After  the first calibration was performed parameters are calculated in two following steps: After

the initialization and after the optimization .In both steps focal length ,principal point , skew,

distortion and pixel error are computed. The  focal length in pixels is stored in the 2×1 vector $f_c$

The  principal point is also stored in the  $2 \times 1$ vector  $c_c$ .The image distortion coefficients(the

radial and the tangential distortion ) are stored in the $5 \times 1$ vector  $k_c$ .
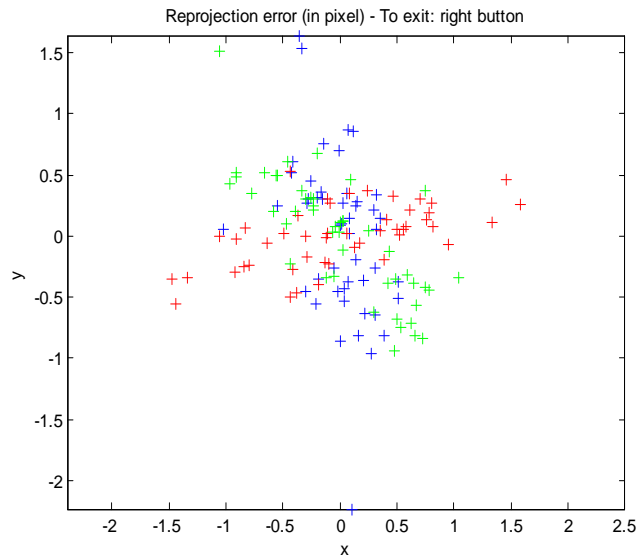


**Figure 4.2** : Reprojection Error
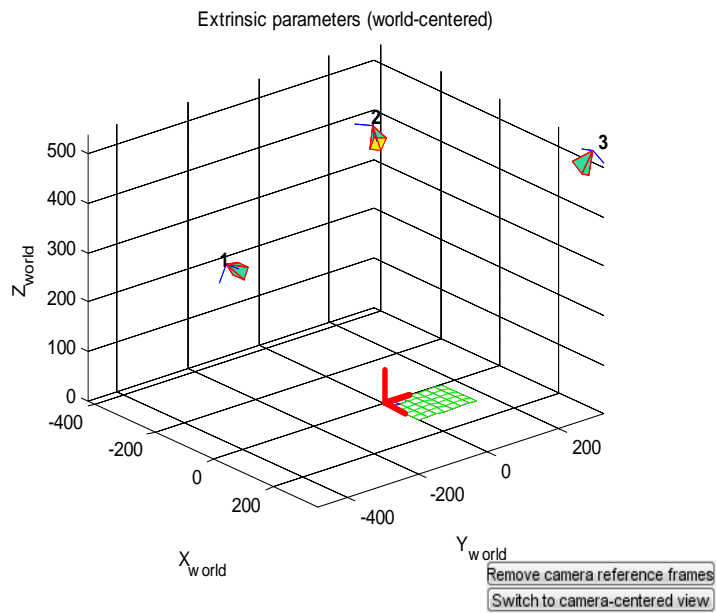


**Figure 4.3** : Extrinsic parameter (world centered)

Focal length    $f_c$   = [1229.17793   1229.57402]±[201.87944  207.26150]

Principal point $c_c$ = [383.50430 511.56501]±[137.20241 215.15207]

Distortion $k_c$ = [-0.16122  0.37309  0.02123  0.00278  0.00000]±[0.91615  14.96503 0.04703

0.02132   0.00000]

Pixel error        err =[0.17100  0.26601]

Translation vector Tc_ext = [-64. 593020 -546.708097 2891.118482 ]

Rotation matrix Rc_ext = [-0.075276 0.979786 0.185347

0.763052 -0.063060 0.643253 0.641938  0.189851 -0.742881]


## 4.3    Head and Feet Point Detection


Corner points are detected after detection of silhouette, using  Plessey  corner detector.  This operator considers a local window in the image and determines the  average change of intensity resulting from shifting the window by a small  amount in various directions.  This  operation  is repeated  for each pixel position  which  is  assigned  an interest value  equal to the minimum change produced by these shifts. Points of interest are the local   maximum of the interest values, since corners exhibit a large intensity variation in every direction.  Once the corner points  are detected,  then the top  max point and bottom  min point  of the silhouette  are selected. These points are called  head and feet point respectively. The whole process is shown in fig 4.5.
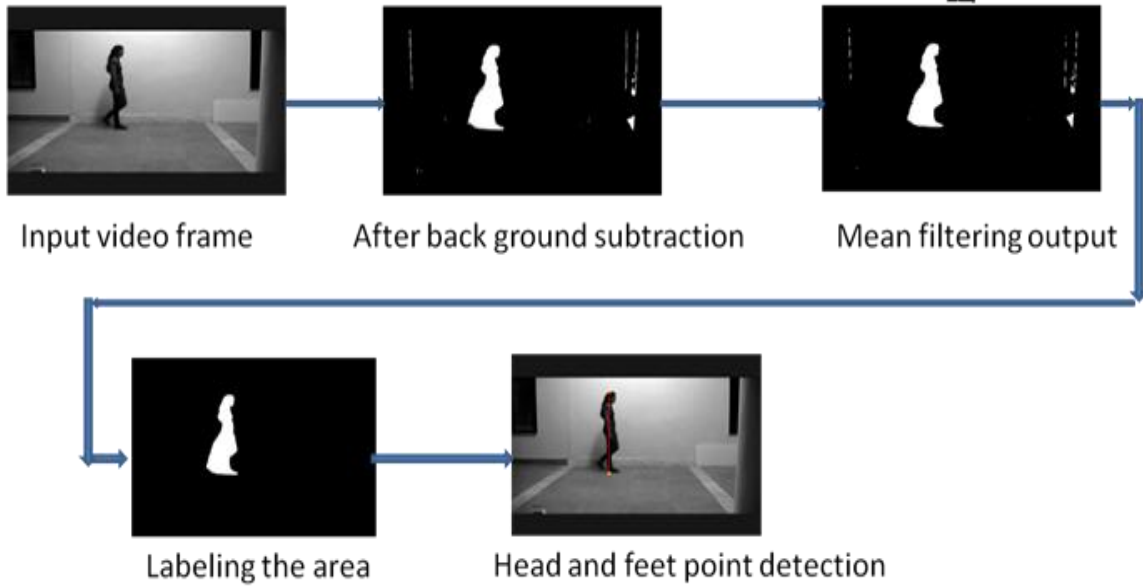
**Figure 4.5** : The process of getting head and feet point

## 4.4    Direct Linear Transformation

A point under the feet of the subject is chosen as the image coordinate $(u_1, v_1)$ and the coordinate

of a point over the head of the subject $(u_2, v_2)$ is arbitrarily chosen.

$$
\begin{aligned}
&c_{11}X_1 + c_{12}Y_1 + c_{13}Z_1 + c_{14} \\
&\quad - \left( u_1 c_{31}X_1 + u_1 c_{32}Y_1 + u_1 c_{32}Z_1 \right) = u_1 \\
&c_{21}X_1 + c_{22}Y_1 + c_{23}Z_1 + c_{24} \\
&\quad - \left( v_1 c_{31}X_1 + v_1 c_{32}Y_1 + v_1 c_{32}Z_1 \right) = v_1 \\
&c_{11}X_1 + c_{12}Y_1 + c_{13}Z_2 + c_{14} \\
&\quad - \left( u_2 c_{31}X_1 + u_2 c_{32}Y_1 + u_2 c_{33}Z_2 \right) = u_2 \\
&c_{21}X_1 + c_{22}Y_1 + c_{23}Z_2 + c_{24} \\
&\quad - \left( v_2 c_{31}X_1 + v_2 c_{32}Y_1 + v_2 c_{33}Z_2 \right) = v_2
\end{aligned}
\tag{4.5}
$$

$z_2$ can be obtained by solving the four equations in (2). Assuming that the image coordinate

under the feet $z_1$ is zero and the x– y coordinate of the points over the head and the feet are

equal. In case of height only Z-coordinate is different. The difference of $z_1$ and $z_2$ will give the

height of a subject.

## 4.5    Stride Parameters Detection

Stride length, Cadence, Gait cycle, and Stride length come under stride parameters [15]. This paper proposes neighborhood technique to find stride length of each subject in each frame.  The appropriate stride length is detected when the silhouette is shadow free, so HSV model-based shadow removal [23] technique is used. Boundary box technique is used to find the stride length. The width of the bounding box gives step length and once step length is detected, twice of that will give stride length. Gait cycle is estimated in terms of frame from periodic stride length signal.

## 4.6    Experimental Results and Discussion

In fig. 4 walking sequences are shown. The yellow marker indicates the head and feet point. In the sequences of the swing phase (one feet is in the ground and other feet toe-off) the vertical segment(red) between the head and feet  i.e. the height is maximum. In stance phase (when two feet contact in the ground and apart) the vertical segment that extends from the top of the head to the point halfway between the two feet. At that time height is the minimum. In fig.4.7 the height changing pattern is shown.

In Fig.4.9 periodic variation of stride length is shown. The boundary width of the blob gives the step length. The twice of step length gives stride length. The local maximum points of the curve in Fig. 4.6 are labelled with '+' in blue color, and their locations are 6, 23, 42, 59 respectively. Since the gait sequence is highly relative to the translation sequence after half gait cycle, the cycle length here is 36 frames.
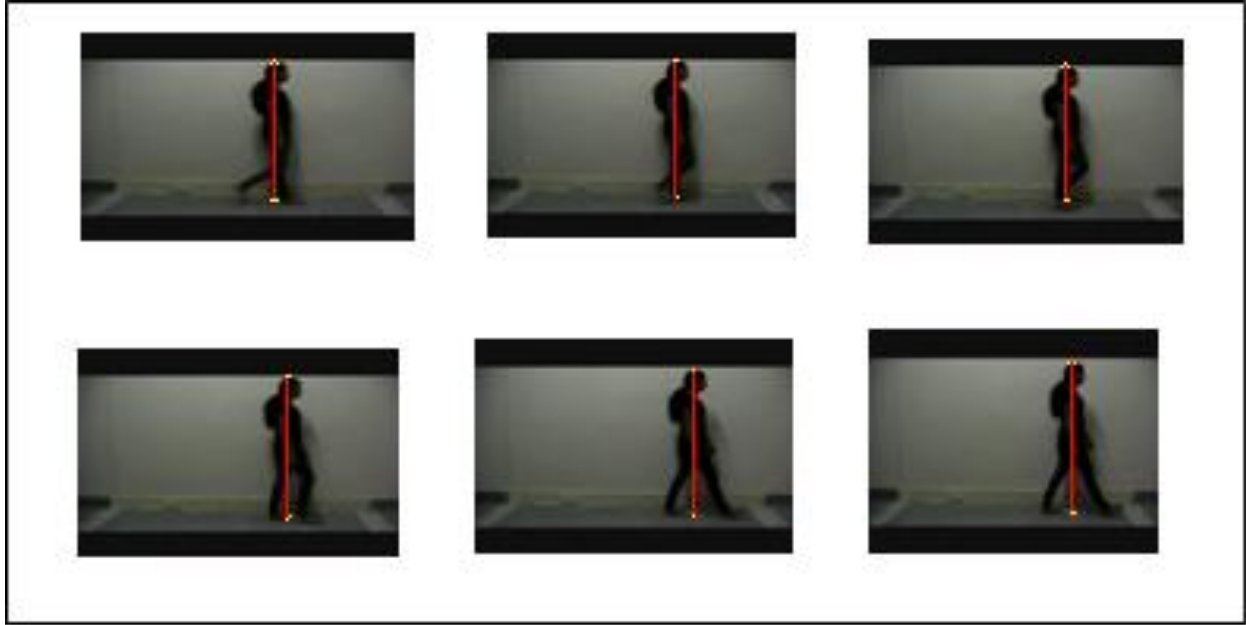
**Figure 4.6**: Walking sequence with extracted height model for subject. Frames run from left to right.In the swing phase (one feet is in the ground and other feet toe-off) the vertical segment between the head and feet and the height is maximum. In stance phase (when two feet contact in the ground and apart) the vertical segment that extends from the top of the head to the point halfway between the two feet. At that time height is the minimum.
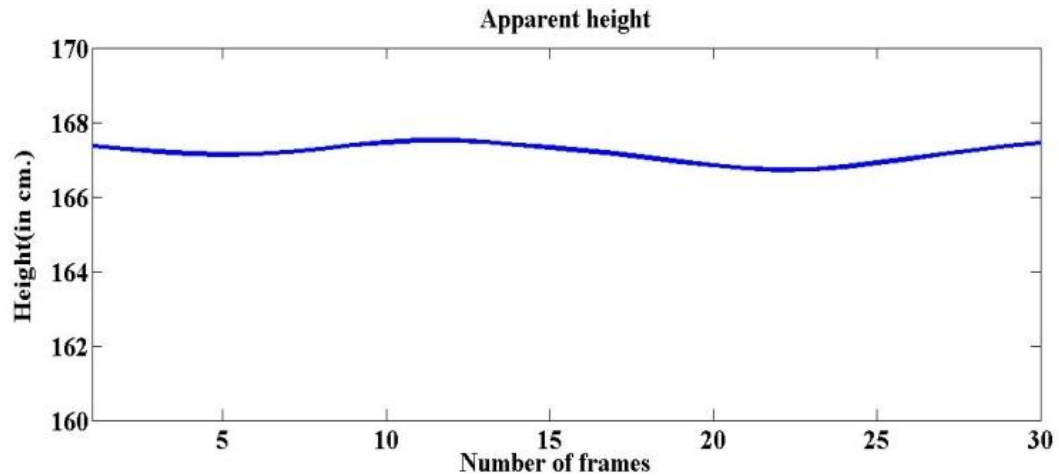


**Figure 4.7**: Height changing pattern extracted with the height model for the sequence in fig. 4.6
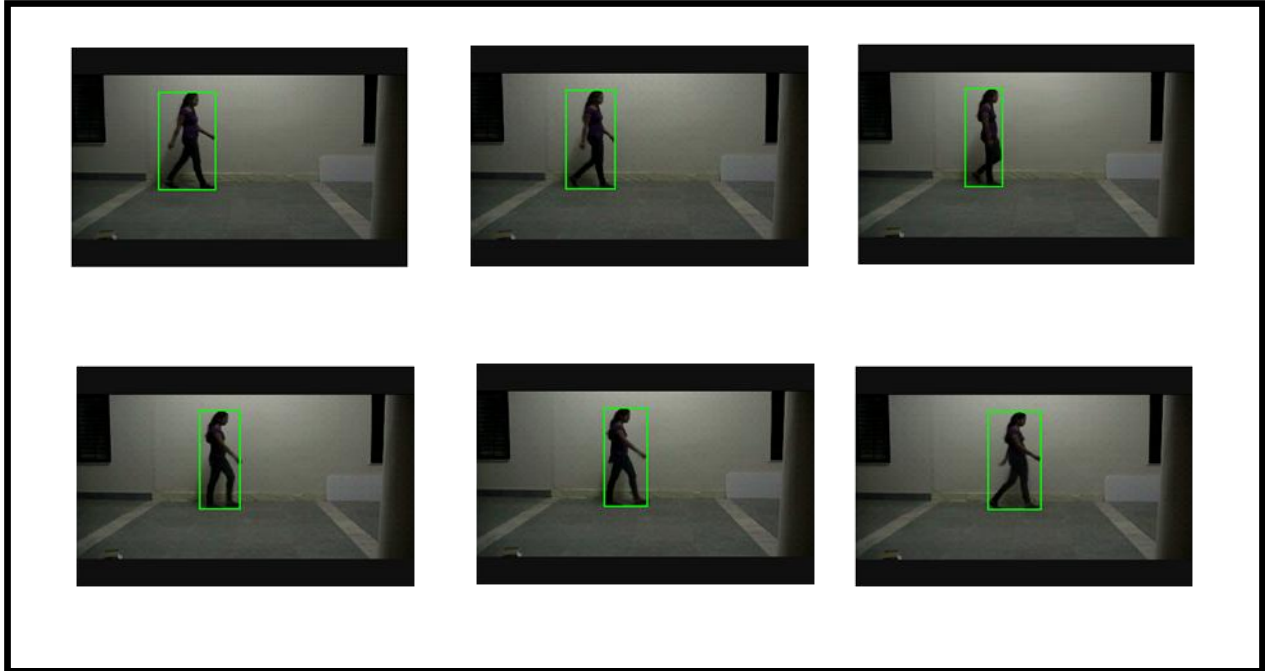
**Figure 4.8**: Variation of stride length is tracked using bounding box width.
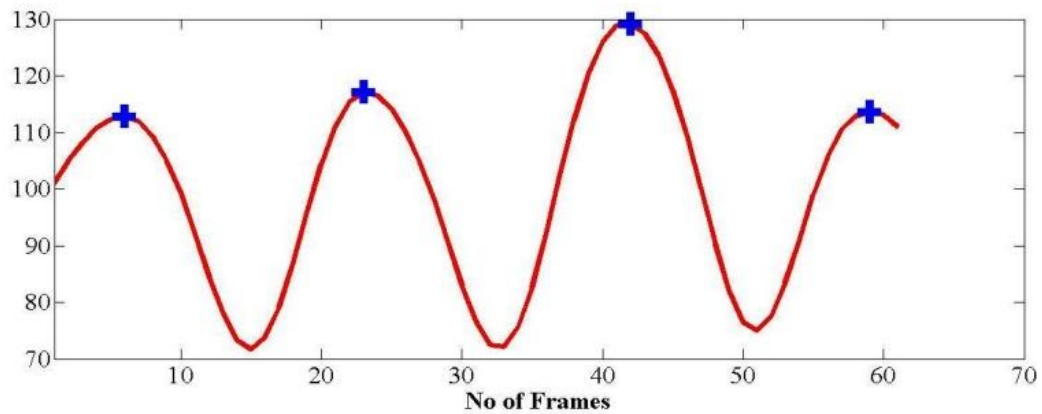


**Figure 4.9**: Changing of stride length during walking.

## 4.7    Conclusion

In this chapter gait features extraction process using silhouette based method has been discussed. Height variation is an important gait feature which can be measured at any distance. It gives a periodic sinusoidal wave form. The wave form has maxima in the swing phase and minima in the stance phase that means when step length is high then the height of the person is minimum. And when the step length is small then height is maximum. The stride length is useful for getting gait cycle. The frame difference between the first and the third maxima gives one gait cycle which is further very useful for gait feature extraction process.

# Chapter 5

# Feature Extraction via Model based Method

The Model based approaches use a model of either the person's shape (structure) or motion, in order to recover features of gait mechanics such as stride dimension, thigh angle, Kinematics of joint angles, knee point, etc. This chapter is pertained to calculate thigh angle with respect to horizontal line. As discussed in previous chapter gait cycle is an important parameter used to extract other features of gait. Here the gait cycle is used to detect the thigh angle when one leg is occluded by other leg.

## 5.1    Shape Model Estimation

A model-based approach is attempted to produce a biometric that has high fidelity to the original data. The model based method is mainly used in the Medical studies the process of model construction and fitting is divided into two phases. In the first phase, a shape model is estimated. The shape model consists of three major segments dedicated to head, torso and leg regions. which is shown in figure.5.1. This model has only three free parameters; the height of the body (H) and the center of mass coordinates; and the last static parameter to be estimated is the period of walking. In the chapter- 4 the height of the subject has already been estimated, uses that height as the reference height (H) and applied model proportion to get thigh region.
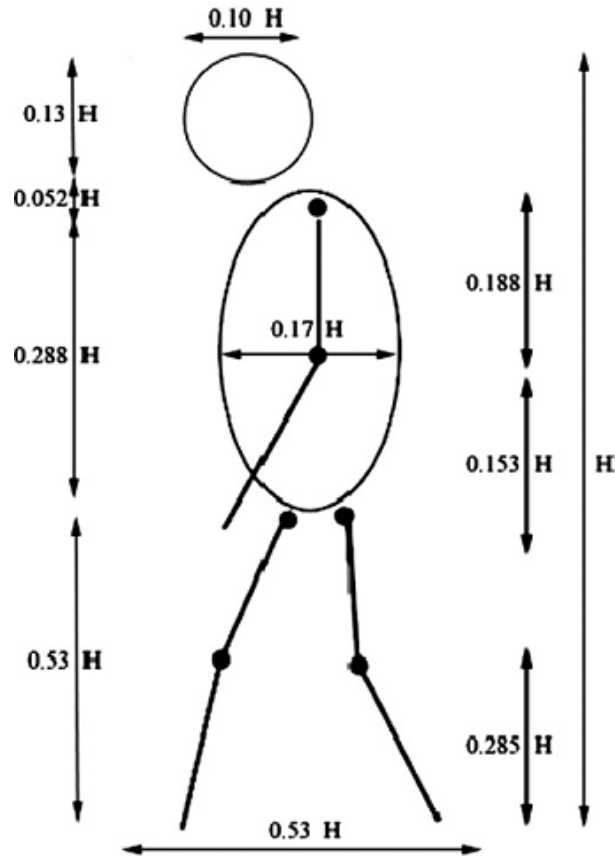
**Figure 5.1**: Model proportions [15]

## 5.2    Local Edge Linking Method

Thigh angle detection required the set of pixels that lie on the thigh. It requires edge linking technique Hough transform is used in edge linking technique to extract thighs in the sequence of images.    The detection of the proper thigh edge detection is required which is followed by linking algorithm. There are three fundamental approaches to edge linking

1.  Local processing

2.  Regional processing

3.  Global processing

## 5.2.1  Local Processing

It is the simplest approach for linking edge points to analyze the characteristics of pixels in a small neighborhood.

Two principal properties are used for establishing similarity of edge pixels.

➢ The strength (magnitude)

➢ The direction of the gradient vector

The first property is based on this

$$M(x, y) = mag(\nabla f) = \sqrt{g_x^2 + g_y^2} \qquad (5.1)$$

$\nabla f$ =gradient of the image $f$

$M(x, y) = mag(\nabla f) =$ It is the magnitude (length)of the vector $\nabla f$ at the loacation $(x, y)$.

Let is the set of coordinates of a neighbourhood cantered at point $(x, y)$ in an image. An edge pixel with coordinates (s, t) in $S_{xy}$ is similar in magnitude to the pixel at (x, y) if

$$M(s,t) - M(x, y) \leq E \qquad (5.2)$$

Where $E$ is a positive threshold.

The second property is the direction angle of the gradient vector is given by (5.2)

An edge pixel with coordinates (s, t) in $S_{xy}$ has an angle similar to the pixel at (x, y) if

$$|(s,t) - \alpha(x, y)| \leq A \qquad (5.3)$$

Where  *A* is a positive angle threshold. The direction of the edge at pixel (x, y) is perpendicular to the direction of the gradient vector at that point.

A pixel with coordinates (s, t) in  $S_{xy}$  is linked to the pixel at (x, y) if both the magnitude and direction equations are satisfied. This process is repeated at every location in the image. A record must be kept of linked points as the centre of the neighborhood is moved from pixel to pixel. The preceding formulation is computationally expensive for scanning total image because all the neighbors of every point have to be examined.

**STEPS FOR LOCAL PROCESSING**

> Compute the gradient magnitude M(x, y) and angle arrays $\alpha(x, y)$, of the input image, *f*(x, y).

> From a binary image, g, whose value at any pair of coordinate (x, y)is given by:

$$\begin{cases} 1, if \ m(x, y) > T_M \ \ and \ \ \alpha(x, y) = A \pm T_A \\ 0, \qquad\qquad\qquad\qquad Otherwise \end{cases} \tag{5.4}$$

Where $T_M$ is a threshold, *A* is a specified angle direction, and $\pm T_A$ defines a "bands" of acceptable directions about *A*.

> Scan the rows of g and fill (set to 1) all gaps (sets to 0s) in each row that don't exceed a specific length, K. A gap is bounded at both ends by one or more 1s.  The rows are processed individually with no memory between them.

> To detect gaps in any other direction, θ, rotate g by this angle and horizontal scanning procedure is applied which is in step 3. Rotate the result back by − θ.

## 5.2.2  Regional Processing

In regional processing a piece of knowledge is required for the regional membership of pixels in the corresponding edge image. In such situations techniques are used for linking pixels on a regional basis. In the figure bellow shows a set of points representing an open curve in

Then perpendicular distance from all other points in the curve to this line and select the point that yield the maximum distance.
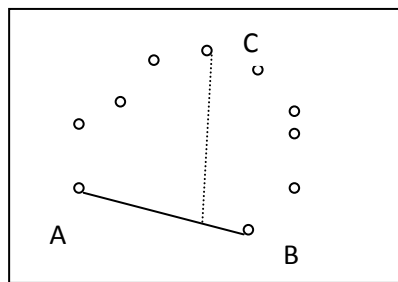


**Figure 5.2:** Regional processing curve

## 5.2.3  Hough Transformation

Hough transformation is a global processing technique. It looks at the global relationship between pixels. It is used to automatically detect simple features from an image. It transforms an image from the feature space into parameter space for line detection.

It is not a one to one relationship between pixels in the mage and cells in the parameter space matrix. Each cell in the parameter space represents a line that spans across the entire image. Transformation between feature space and parameter space (I) Project a line through each edge pixel at every possible angle. (ii) For each line, calculate the minimum distance between the line and the origin, (iii) Increment the appropriate parameter space accumulator by one.

The resulting matrix: The x-axis of the parameter space ranges from 1 to the square root of the sum of the squares of rows and columns from feature space. The number corresponds the farthest possible minimum distance from the origin to a line passing through the image. The y-axis represents the angle of the line. The axes could be switched, transforming from parameter space back to feature space is slightly more trouble for lines than circles. Suppose that, for n points in an image, we want to find all subsets of these points that lie on $(x_i, y_i)$ and the equation for a straight line $y_i = ax_i + b$. Infinitely many lines pass through the point $(x_i, y_i)$ all satisfying the equation for varying of a and b. The equation $b = -x_i a + y_i$ and considering the a, b plane yields the equation of a single line for a fixed point $(x_i, y_i)$ at $(a', b')$. $a'$ is the slope and $b'$ is the intercept of the line containing both $(x_i, y_i)$ and $(x_j, y_j)$ in the x y plane. In fact, all points that lie on this line have corresponding lines in the parameter space that intersect at $(a', b')$.
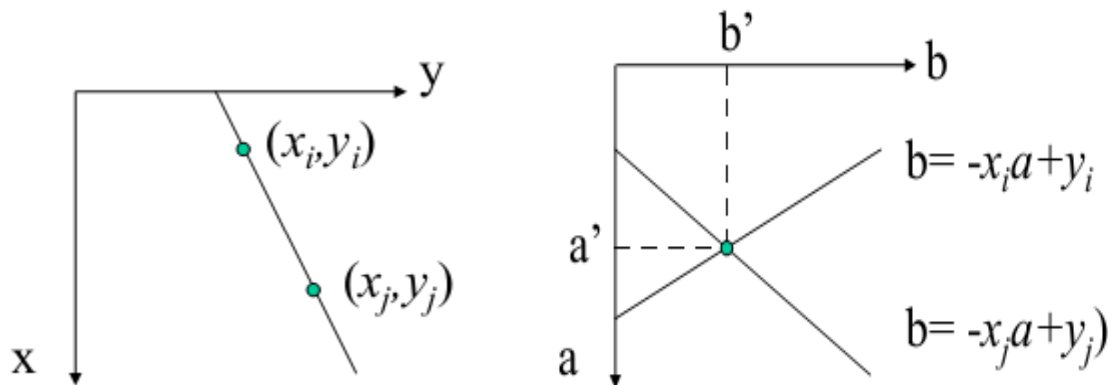


**Figure 5.3**: Line equation in terms of slope 'a' and y-intercept 'b'

A problem with Normal equation is given by

$$x \times \cos\theta + y \times \sin\theta = \rho \tag{5.5}$$

Solution: use the normal representation of a line given by
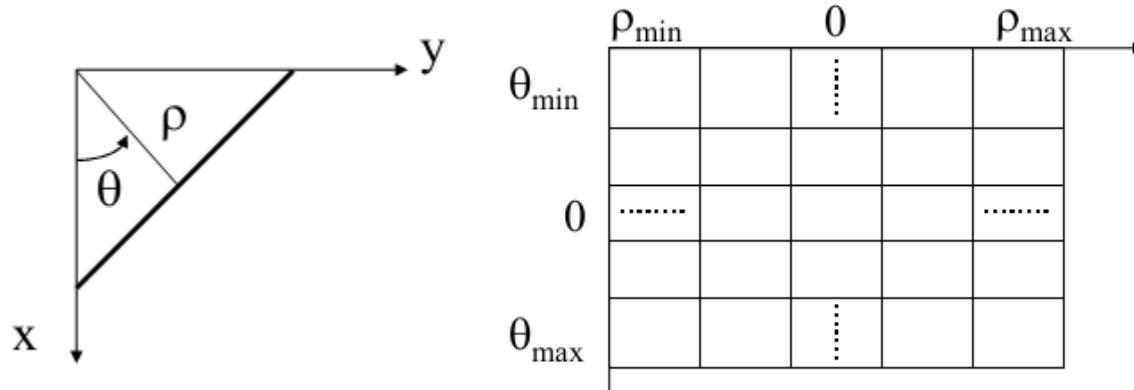


**Figure 5.4**: Accumulator to store $\rho$ and $\theta$

Instead of straight lines in the ab plane, we now have sinusoidal curves in the $\rho\theta$ plane. M collinear points lying on the line

$$x \times cos\theta_j + y \times sin\theta_j = \rho_i \qquad (5.6)$$

It yields M sinusoidal curves that intersect at ($\rho_i$ ,$\theta_j$) in the parameter space

• The range of $\theta$ is ±90°, measured with respect to the x axis

A horizontal line has $\theta=0°$, with $\rho$ equal to the positive x intercept

– A vertical line has $\theta=+90°$, with $\rho$ equal to the positive y intercept or

$\theta=-90°$, with $\rho$ equal to the negative y intercept

• The range of ρis ±(2)½D. Where D is the distance between corners in the image
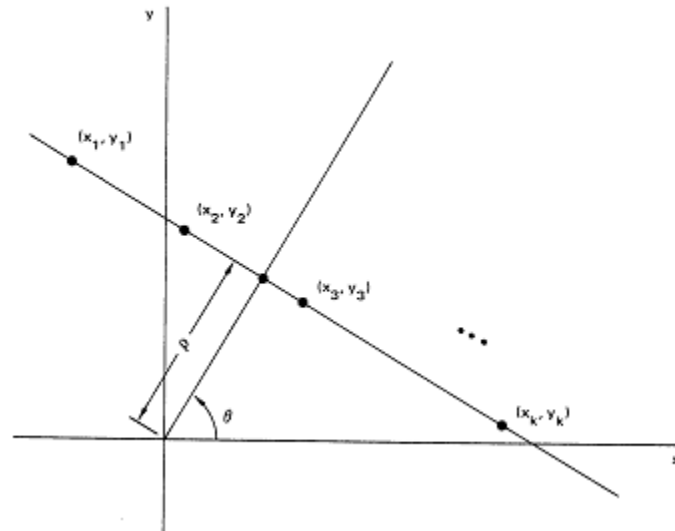


**Fig 5.5:** Projection of collinear points onto a line

must be kept of linked points as the center of the neighborhood is moved from pixel to pixel.The

preceding formulation is computationally expensive for scanning total image because all the

neighbors of every point have to be examined.

➢ Scan the rows of g and fill (set to 1) all gaps (sets to 0s) in each row that don't exceed a

   specific length, K. A gap is bounded at both ends by one or more 1s.  The rows are

   processed individually with no memory between them.

➢ To detect gaps in any other direction, θ, rotate g by this angle and horizontal scanning

   procedure is applied which is in step 3. Rotate the result back by − θ.

## 5.3    Procedure for getting thigh points

• Frame is extracted from video

- Background subtraction method and thresholding technique are used to get foreground object.

- Labeling the image to detect only human from other objects.

- Shadow removal technique using HSV color space is used to remove shadow and getting proper silhouette.

- Morphological operation is used to fill the holes in the silhouette.

- Lower point (feet point) and upper point (head point) is detected using corner detection technique.

- Height detection using calibration process.

- Canny edge detector is used to detect the edges.

- Thigh point  = calibrated height * 0.53

- Check the condition weather a person is coming from left to right or from right to left.

- Edge linking method is applied using local processing and region processing.First requires knowledge about edge points in the local region (3×3 or 5×5 neighborhood).The second requires that points on the boundary of a region be known.

- The second point in the thigh is known by using the (3×3 or 5×5 neighborhood) whose weight is equal to '1'.

  All the points which are passing through above two pints are the thigh points  and can be calculated using straight line equation.

- Calculate the thigh angle with respect to. horizontal line.

## 5.4    Experimental Results
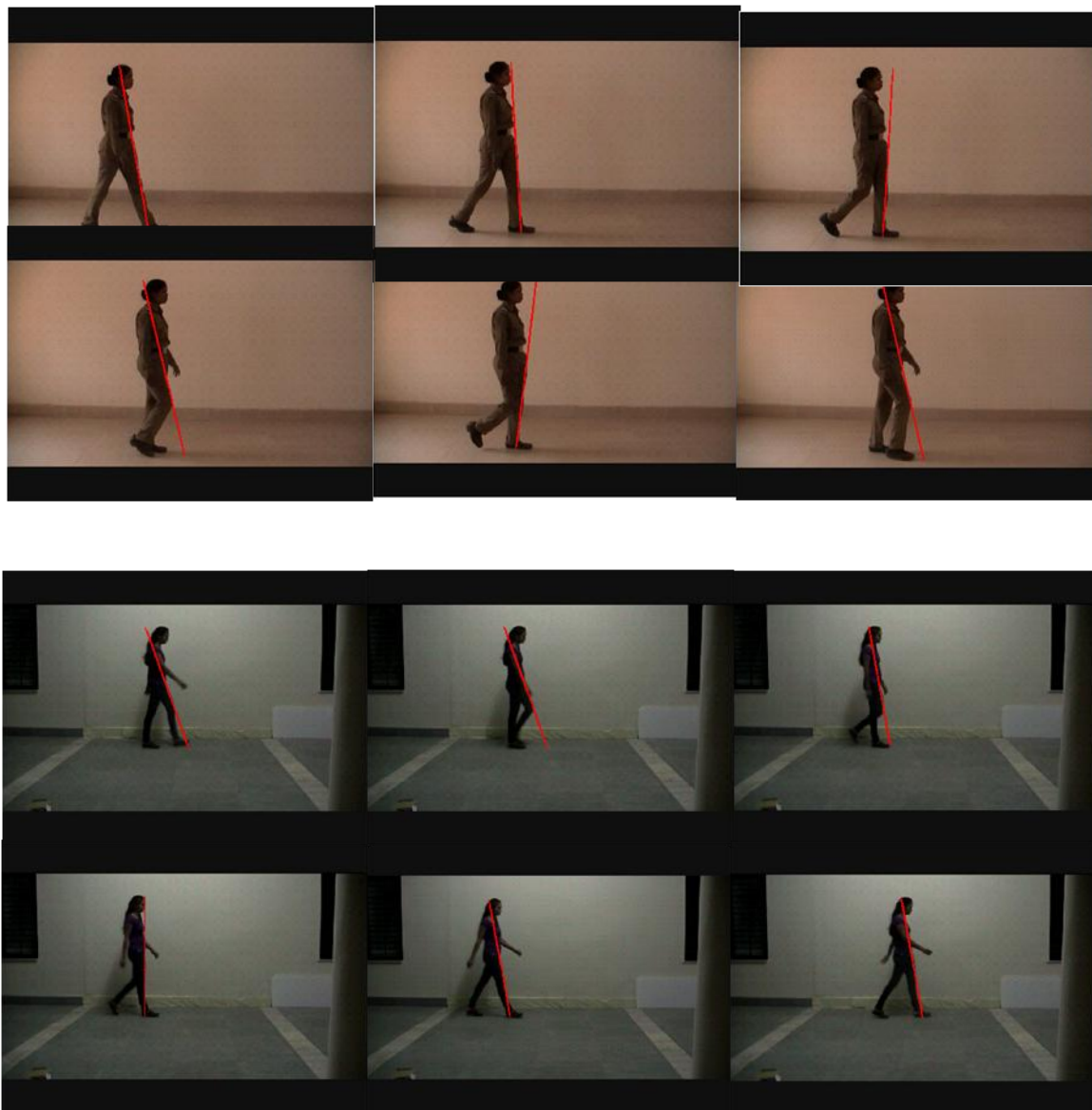


**Figure 5.6**: Thigh angle is shown on the sequence of databases

In the fig.5.6 the red line segments are representing thigh angle on the above sequence of frames.

In fig. 5.7 shoes angle variation curve with respect to horizontal line.
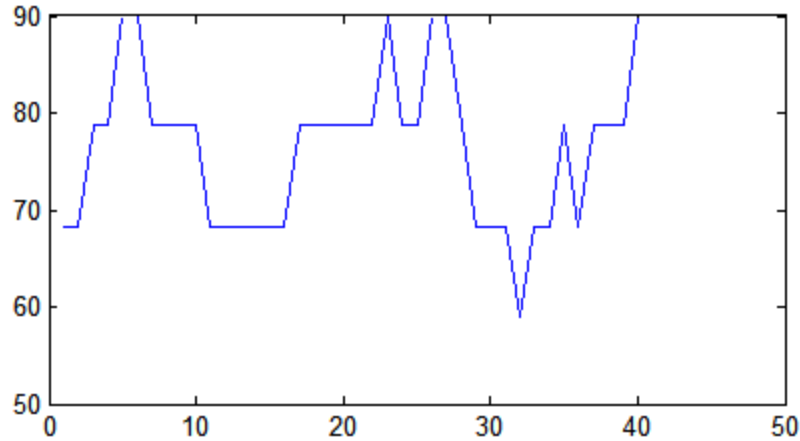
**Figure 5.7**: Variation of thigh angle with respect to horizontal line. X-axis represents number of frames and y-axis represents angles.

## 5.5    Conclusion

This chapter put to use region based and local processing method as compared to Hough transformation method for edge linking detection. Edge linking process using Hough transform has problems such as particular edge of thigh, limb, hand may not be always detected correctly. So local based processing and region based processing methods are used to improve accuracy of detecting thigh points .Gait cycle is used to estimate the thigh angle when one leg is occluded by other leg. Periodic thigh angle variation is a gait feature which can be further process for recognition.

# Chapter 6

# Human Recognition

This chapter describes the process of human recognition using different transformation techniques. The height variation and stride length variation features which are extracted from previous chapter-4 used separately and combined for human recognition. Different windowing techniques such as Blackman window and Rectangular window are applied on periodic variation signal with respect to frame. Then DFT, DHT, and DCT are used on the samples to extract the coefficients such as feature because frequency domain contains more information than time domain. Then self recognizing thresholding (SRT) is obtained to recognize individual.

## 6.1    Feature Identification Process

The process of human feature identification is shown in fig 6.1. In the proposed model 1-D height signal is generated by combining the height of each frame of the silhouette. Different windowing techniques such as Blackman window and Rectangular window are applied to get finite samples from this continuous signal. Windowing techniques are applied to avoid leakage outside the finite interval. DFT, DHT, and DCT are used on the samples to extract the feature. DFT is a powerful tool for analyzing and measuring the continuous signal. It produces the average frequency content of the signal over the entire time that the signal was acquired. It maps a length-N complex sequence to a length-N complex spectrum. DHT is a real-valued transform

closely related to DFT of a real-valued sequence [9]. It directly maps a real-valued sequence to a

real-valued spectrum while preserving some of the useful properties of DFT [12]. DCT is a real

and orthogonal transformation technique. It expresses a sequence of finitely many data points in

terms of a sum of cosine functions oscillating at different frequencies. In DCT most of the signal

components are stored in the lower harmonic components. These transformations are applied on

the apparent height signals. Then N- harmonics from the transformation coefficients are selected.

These are the feature vectors stored in the database. If any subject comes then his features are

extracted, then compared with the database features by using Euclidian distance and MSE

computation. If MSE value is greater than SRT then different subjects, otherwise same subject.
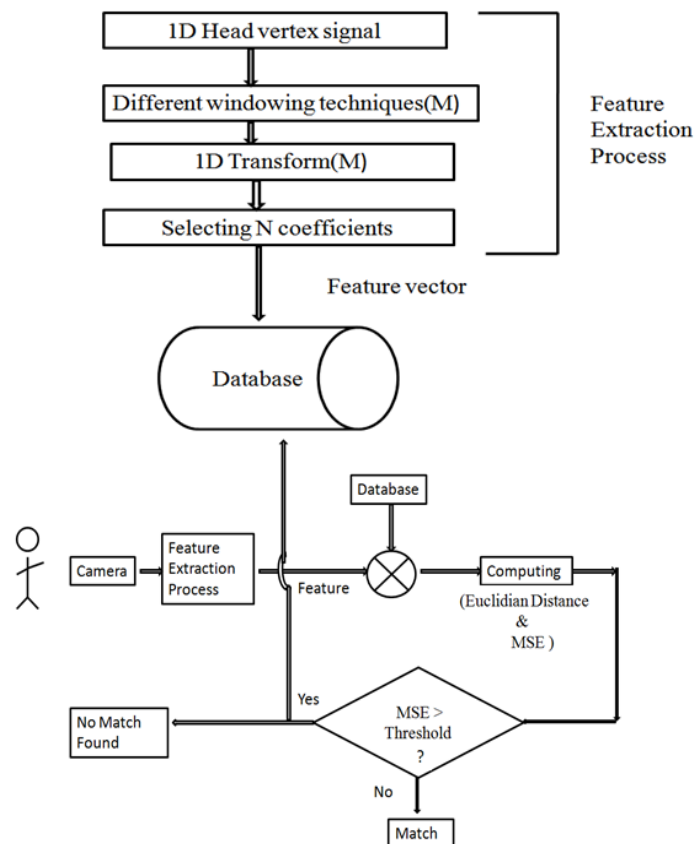


**Figure 6.1** : The process of features identification.

## 6.2     Simulation Results and Discussion

A video camera is placed with a plane normal to the subject's path in indoor environment with controlled illumination. Subjects are asked to walk in front of a plain, static background. A static video camera is used to capture video sequences at a distance of 500 cm, 250 cm, and 200cm.

The length of each Video is 3-4 sec. and the frame size is 720×480 pixel resolution. 50 consecutive frames from each video are extracted. This is a very important step as the total result depends on the quality of the gait captured.

The method is tested on walking style of 10 different subjects and each subject is advised to walk minimum 4-5 times. First we compare the features of walking style of each subject with himself and then with others. Three transformation techniques are applied and their coefficients are compared to get MSE. In the Table I we put some experimental results of MSE of same subjects and different subjects to get threshold values for each transformation using two different windowing techniques such as Blackman and rectangular window.

Table 6.1 The Results of MSE when same subject and different subjects are compared on the basis of height parameter

| Method | Blackman Window Technique | | | | | |
|---|---|---|---|---|---|---|
| | $P1 \sim P2$ | $P1 \sim P2^{'}$ | $P1 \sim P3$ | $P1 \sim P1^{'}$ | $P1 \sim P1^{''}$ | $P1 \sim P1^{'''}$ |
| DCT | 57.042 | 40.51 | 27.69 | 4.641 | 1.8354 | 1.7046 |
| DFT | 1361.2 | 966.8 | 659.1 | 111.3 | 41.82 | 37.88 |
| DHT | 1308.7 | 995.6 | 592.7 | 79.6 | 32.56 | 24.38 |
| | Rectangular Window Technique | | | | | |
| DCT | 158.34 | 92.15 | 56.7 | 15.85 | 8.76 | 5.78 |
| DFT | 6594 | 4681 | 2731 | 421.1 | 182.4 | 113.6 |
| DHT | 6562 | 2358 | 1358 | 208.5 | 85.7 | 54.99 |

Here  $P1, P2$ and $P3$ are the coefficients of different subjects. $P1', P1'', P1'''$ are the coefficients of same subject's at different time of walking. Blackman window and Rectangular window techniques are used to get M number of samples from the 1-D signal. Then DCT, DFT, DHT transformations are applied on them. Then N no. of coefficients is taken from M no. of sample

coefficients which makes a N×1 feature vector. A tradeoff between computational burden and for

better result only first 4 coefficients are selected among 30 samples. In Fig. 8.threshold values

for different transformation techniques are plotted using Rectangular and Blackman window.

The Same experimental work is also done to find the SRT value of stride parameter. The SRT

value is 1.5e+05 in case of DHT using rectangular window technique. Comparison of average

recognition rate of subjects on the basis of height feature for different transformations using

rectangular window is shown in fig. 6.2 (a). (b) Separately compares the average recognition rate

of height and stride length (one of the stride parameter). Variation of height gives better

recognition rate than variation of stride length.
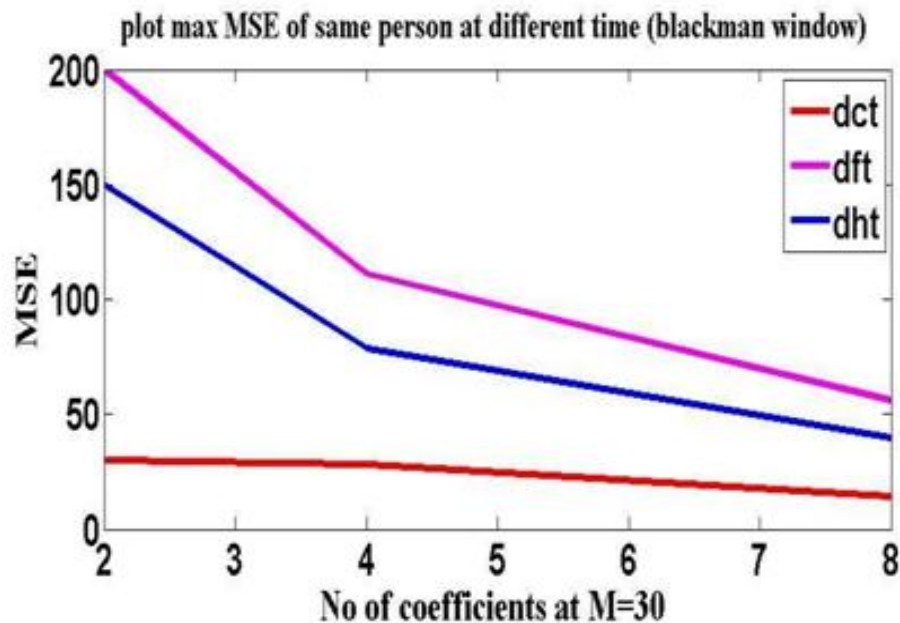


(a)
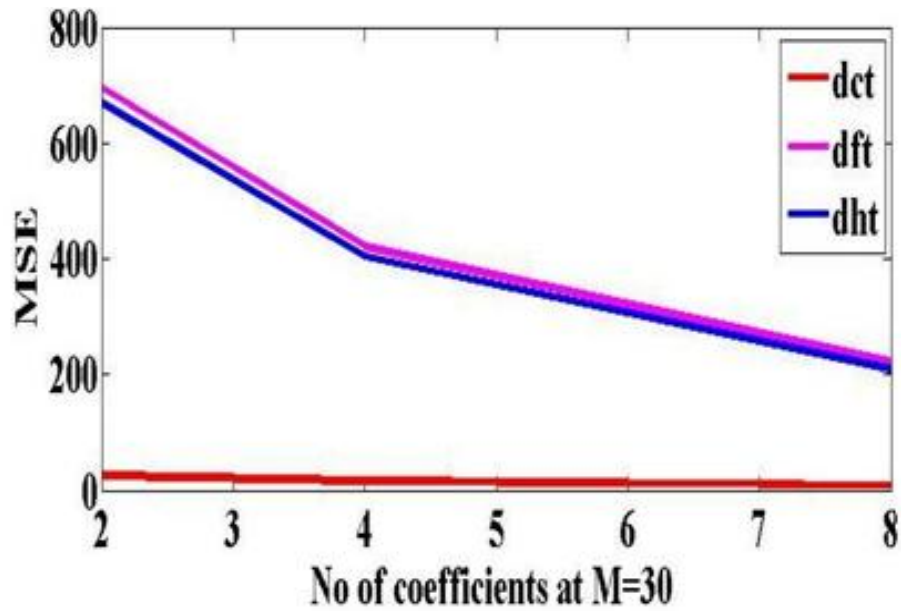
(b)

**Figure 6.2**: The above graphs show the threshold values by using different transformation techniques. Mean Square Error in y axis and no of coefficients in x-axis: (a, b) Compare MSE for same subject walking style using Blackman and Rectangular window techniques respectively.
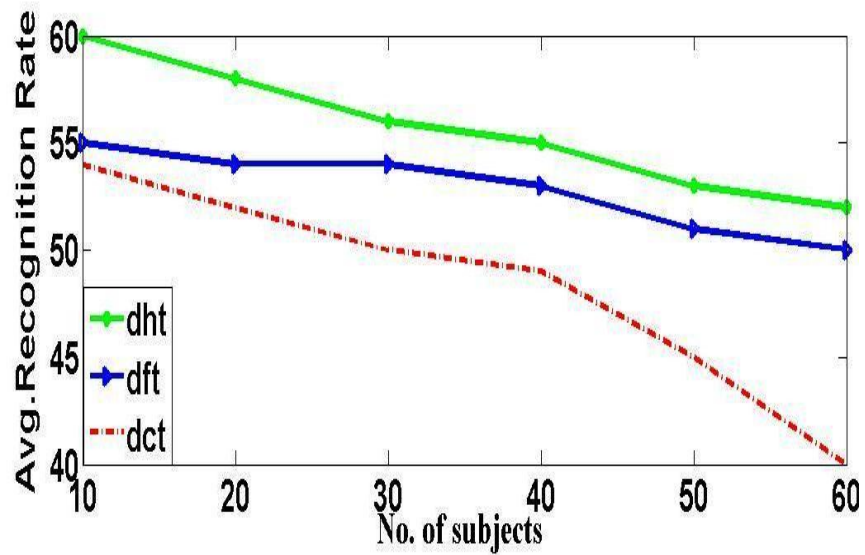


(a)
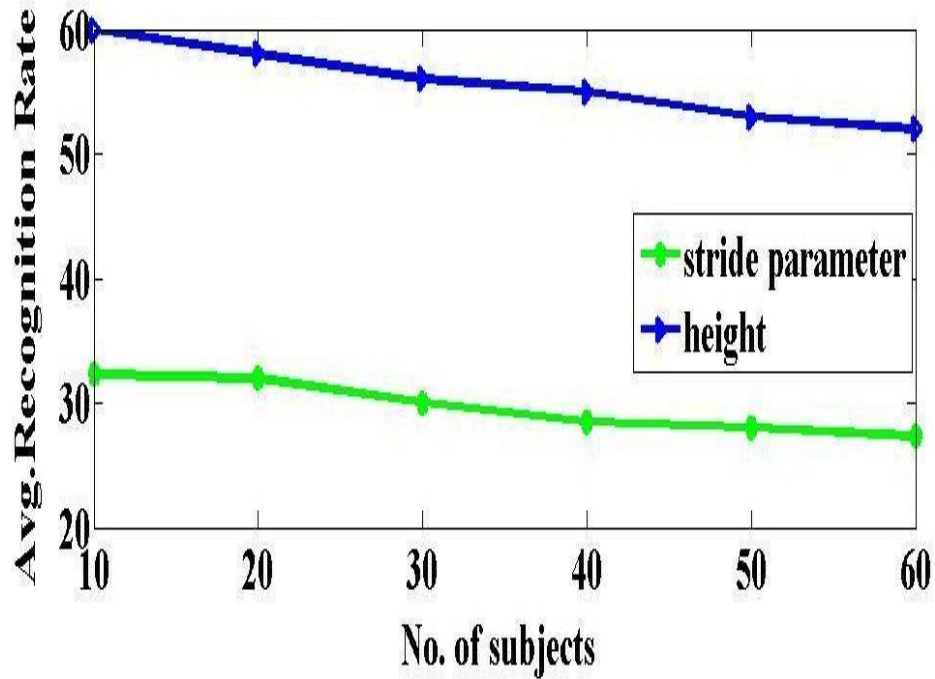
(b)

**Figure 6.3:** Comparison of Average recognition rate w.r.t number of subjects.: (a) Compare average recognition rate between different transformations using rectangular window for height parameter.. (b) Compare average recognition rate between two gait parameters: height and one of the stride parameter: Stride length

We carried out these SRT values for recognition of other 15subjects. These subjects also walked 4 times at different time. Table II compares recognition rate of different transformation methods used in the proposed model.

$$\text{Re cogination Rate} = \frac{\text{Number of Subjects Correctly recognized}}{\text{Total number of Subjects tests}} \times 100$$

Table 6.2 Recognition rates of different transformation techniques for height parameter.

| Method | Window Techniques | Recognition Rate (%) |
|--------|-------------------|----------------------|
| DFT | | 46 |
| DCT | Blackman Window | 56 |
| DHT | | 53 |
| DFT | | 57 |
| DCT | Rectangular Window | 55 |
| DHT | | 60 |

## 6.3   Conclusion

This chapter implemented DFT, DCT and DHT transformation techniques on height and stride variation signals for feature identification. Self recognition threshold (SRT) value is chosen for human recognition. On comparison between three transformations DHT gave better recognition rate than others. Another comparison was held between two gait features such as stride and height parameters. The average recognition rate of two parameter showed height gives better recognition rate than stride.

# Chapter 7

# Conclusion & Scope for Future Work

## 7.1    Conclusion

This thesis has investigated the use of height and stride related gait features for identifying a person. But gait feature extraction process is only possible when it works on a good silhouette. Here in this thesis, different recent background subtractions available in the literature have been studied and their performance tested on the different video test sequences. It should be noted that robust motion detection is a critical task and its performance is affected by the presence of varying illumination, background motion, camouflage, shadow, and etc. Thus only background subtraction is not enough to suppress the shadow from the foreground and make a perfect binary silhouette. Different shadow detection methods are tried out using RGB, YCbCr, and HSV color operators to make a shadow mask to suppress the shadow. The quantitative performance evaluation is done on different indoor videos and it is observed that HSV color model gives better performance. Further HSV color space has been used for shadow suppression model to optimize the threshold parameters with respect to the average intensity of local shadow region. A mathematical model is developed for the threshold parameters $\alpha$ $and$ $\beta$ .

Then we have discussed on height variation, and stride length parameter during walking which is periodic in nature. Height variation of a person during walking is radically distinctive from person to person. Our experimental result shows analysis of height signals

and stride signals using different windowing techniques and different transformation techniques. Our experimental result shows rectangular window using DHT method gives better recognition rate in compared to other transformation techniques.  Further result shows height feature gives better recognition rate (52%) as compared to (27%) by stride feature.  As height is a gait feature which is not influenced by the clothing style, camera performance, distance, etc. so its performance is better than the stride feature.

Again we have discussed on Model based method to detect the thigh angle. But thigh angle of one leg is not detected over a period of walking. Because one leg is occluded by the other leg. So gait cycle is used to estimate the thigh angle.

## 7.2    Scope for Future Work

➢ Although height and stride parameters are extracted and recognition rate is determined by considering each feature separately, so in future combining these two features may give better gait recognition rate.

➢ All the experiments are performed on the indoor sequences with lateral view of subjects. In future different view angle like oblique view, frontal view, etc. are considered for feature extraction.

➢ In addition, as the importance of the dynamic gait features for people identification is confirmed in this study, further research should be carried out to investigate the discriminatory power and analyze the kinematic characteristics of gait motion using more advanced statistical methods in order to derive more discriminative and efficient features from the gait dynamics.

# References

[1] H. Weiming, T. Tieniu , W. Liang, and S.Maybank, "A survey on visual surveillance of object motion and behaviours," *IEEE Transactions on Systems, Man, and Cybernetics, Applications and Reviews.,* vol.34, no. 3, pp. 334-352, 2004.

[2] L. Jiwen, and Z. Erhu, "Gait recognition for human identification based on ICA and fuzzy SVM through multiple views fusion," in *Pattern Recognition Letters.,* vol.28, no.16, pp. 2401-2411, 2007.

[3] C. B. Abdelkader, R. Cutler, H. Nanda, and L. Davis, "Eigen gait: Motion-based recognition of people using image self-similarity," *Audio-and Video-Based Biometric Person Authentication*, *Springer Berlin Heidelberg.,* pp. 284-294, 2001.

[4] L. Zongyi, and S. Sudeep, "Simplest representation yet for gait recognition: Averaged silhouette," in *Proc. IEEE Conf. Pattern Recognition,* vol. 4, no. 2,  pp. 211-214, 2004.

[5] J. Robert, and G. D. Abowd. "The smart floor: a mechanism for natural user identification and tracking,"*CHl'00 extended abstracts on human factors in computing systems., pp. 275–276, 2000.

[6] L. Middleton, B. A. Alex, B. Alex, and N. S. Mark, "A floor sensor system for gait recognition.*"Automatic identification advanced Technologies.,* pp. 171-176, 2005.

[7] D.Gafurov, H. Kirsi, and S.Torkjelx. "Gait recognition using acceleration from MEMS." in *Availability, Reliability and Security. IEEE Conf.*, pp. 6-10, 2006.

[8] J. Mantyjarvi, M. Lindholm, E. Vildjiounaite, S. M. Makela, and H. Ailisto. "Identifying users of portable devices from gait pattern with accelerometers." *Acoustics, speech and signal Processing, 2005, IEEE Conf.,* vol. 2, pp. 973-976, 2005.

[9] M. Pat, A.D. Bernardt, and R. C. Kory, "Walking patterns of normal men." *The Journal of Bone & Joint Surgery.*, vol.46, no. 2, pp. 335-360, 1964.

[10] N. S. Mark,  J. N. Carter, D. Cunado, P. S. Huang, and S. V. Stevenage, "Automatic gait recognition," in *Biometrics*, Springer US, pp. 231-249, 2002.

[11] B. John, L. David, J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *International Journal of Computer Vision.,* vol. 12, no. 1, pp. 43-77, 1994.

[12] S. Aravind, C. AmitRoy, and C. Rama, "A hidden markov model based framework for recognition of humans from gait sequences," *Image Processing, ICIP  IEEE conf.*, vol. 2, pp. 89-93, 2003.

[13] S. Cristian, and B. Triggs, "Covariance scaled sampling for monocular 3D body tracking," *Computer vision and Pattern Recognition, IEEE conf.,* vol. 1, pp. 432-447, 2003.

[14] Z. Tao, T. Wang, and H. Shum," Learning a highly structured motion model for 3D human tracking," *Proc. Asian Conf. Computer Vision*, vol. 1, pp. 144-149. 2002.

[15] Moore, Keith L. *clinically oriented anatomy*. Lippincott Williams & Wilkins, 2013.

[16] D. Gafurov, "A survey of biometric gait recognition: Approaches, security and challenges." In *NIK-2007 conference*. 2007.

[17] P. Remagnino, T. Tan, and K. Baker," Multi-agent visual surveillance of dynamic scenes," *Image and Vision Computing*., vol. 16, no.8, pp. 529–532, 1998.

[18] L.Wang,  T. Tan, H. Ning, and W. Hu. "Silhouette analysis-based gait recognition for human identification." *Pattern Analysis and Machine Intelligence, IEEE Trans.*, vol.  25, no. 12, pp. 1505-1518, 2003.

[19] S. Ching, S. Cheung, "Robust techniques for background subtraction in urban traffic video," *Electronic Imaging, International society for Optics and Photonics*., pp. 881–892, 2004.

[20] S. Toral, M. Vargas, F. Barrero, and M. G. Ortega. " Improved sigma-delta background estimation for vehicle detection." *Electronics letters*., vol. 45, no. 1 pp. 32-34, 2009.

[21] A. Prati, I. Mikic, M. M. Trivedi, and R. Cucchiara. "Detecting moving        shadows: algorithms and evaluation." *Pattern Analysis and Machine Intelligence, IEEE Transactions.,* vol.  25, no. 7, pp. 918-923, 2003.

[22] P. K. T. Pong, and R. Bowden, "An improved adaptive background mixture model for real-time tracking with shadow detection," *Video-Based Surveillance Systems, Springer US*. pp. 135-144, 2002.

[23] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, and S. Sirotti. "Improving shadow suppression in moving object detection with HSV color information," *Intelligent Transportation Systems, IEEE.*, pp.334-339, 2001.

[24] E. Salvador, A. Cavallaro, and T. Ebrahimi. "Cast shadow segmentation using invariant color features." *Computer vision and image understanding,* vol. 95, no. 2 pp. 238-259, 2004.

[25] L.T. Maloney, and B. A.Wandell, "Color constancy: a method for recovering surface spectral reflectance." *JOSA A.,* vol. 3, no. 1, pp.29-33, 1998.

[26] M. S. Nixon, "Model-based gait recognition." pp: 633-639, 2009.

[27] J. Lee, E.D. Lee, H. Tark, J. Hwang, D. Young Yoon, " Efficient height measurement method of surveillance camera image," Forensic science International., vol. 177, no. 1, pp. 17-23, 2008.

[28] A. Fetic, D. J. D. Osmankovic, "The Procedure of a camera calibration using Camera Calibration Toolbox for Matlab," *MIPRO, IEEE Conf.,* pp. 1752- 1757, 2012.

[29] A. Manzanera, and J. C. Richefeu, "A new motion detection algorithm based on $\Sigma - \Delta$ background estimation," Pattern Recognition Letters., vol. 28, no. 3, 2007.

[30] F. Cheng, and Y. Chen, "Effective $\Sigma - \Delta$ background estimation for video background generation," *Asia-Pacific Services Computing Conference*, pp. 1315-1321, 2008.

[31] D. K. Panda, " Motion detection, object classification and tracking for visual surveillance application," *Phd diss.*, 2012.