

# Real Time Video Surveillance System

Sonali Vaidya  
ME-IT Student  
TCET, Mumbai

Kamal Shah, Ph.D  
Professor-IT  
TCET, Mumbai

## ABSTRACT

The role of Real Time video surveillance System has been elevated due to increased importance in safety and security. Traditional systems need human operators to understand the activities of objects of interest and take decisions. However, humans do make mistakes. The main reason is the nature of the task is passively watching multiple monitor screens where nothing special happens for a long period of time.

The Proposed Real Time video surveillance system is capable of detecting objects of interest, classify and track them. The Real time surveillance system with the capability of detecting and recognizing motion of a walking person in video which can lead to useful autonomous system. This can be done by using Gait Analysis for tracking scenarios and generating notification to authoritative person.

## General Terms

Video surveillance system

## Keywords

Gait analysis, GEI (Gait Energy Image), PCA (Principle Component Analysis), MDA (Multiple Discriminant Analysis)

## 1. INTRODUCTION

Due to increasing threats of terrorist, now a day the advanced video surveillance systems are more into use. In order to prevent the occurrence of potential danger, these systems need to analyse the behaviors of people.

One of these surveillance systems is the system proposed to identify an individual efficiently and accurately. Controlled environments such as school, colleges, offices, railway stations, airports etc. where these systems need to quickly detect threats.

Gait as a biometric has many advantages which makes it an attractive proposition as a method of identification. A main advantage of Gait is its unobtrusive identification. So even from far distance we can identify threat.

This facility gives the user enough time to identify the suspect before he could become a possible threat. Video footage of suspects are readily available with user, as surveillance cameras are comparatively low cost and installed in prime locations requiring security, in these cases the video needs to be checked against that particular suspect.

## 2. GAIT FOR VISUAL SURVEILLANCE

### 2.1 Introduction

The problem of “who is entering the area under surveillance” is of increasing importance for visual surveillance. Such type of personal identification can be treated as a special behavior-understanding problem. Human gait are now regarded as the main biometric features that can be used for personal identification in visual surveillance systems. In this case Gait Analysis is going to be used; Gait refers to the style of walking of an individual. It includes both the appearance and

the dynamics of human walking motion. Human gait is an identifying feature of a person that is determined by his/her weight, limb length, and habitual posture. And Gait recognition is the term typically used for the automatic extraction of visual clue that characterizes the motion of a walking person in video and is used for identification purposes in surveillance systems, which can then generate an alarm regarding the object of interest.

Gait is a behavioral biometric source that can be acquired at a distance. Gait recognition is typically used in the computer community to refer to the automatic extraction of visual cues that characterize the motion of a walking person in a video and is used for identification purposes in surveillance systems.

Automated surveillance system consists of three phases: detection, tracking and perception. In the perception phase, a high-level description is produced based on the features extracted during the previous phases from the temporal video stream. The main aim of automated surveillance system is to detect and track people in the scene as well as to perceive their behavior and report any suspicious activities to the authorized person.

Identification systems will undoubtedly play a key role in aiding law enforcement officers in their forensic investigations. More importantly, due to early recognition of suspicious individuals who may pose as threats, the system would be able to reduce future crimes. Human motion perception has been of interest to researchers from different disciplines due to the wide range of applications ranging from activity recognition to people identification. In fact, early studies by Johansson [1] on human motion perception using Moving Light Displays (MLD) have revealed that an observer can recognize different types of human motion based on joint motions. Moreover, the observer can make a judgment of the gender of the person [2], and even further identify the person if they are already familiar with their gait [3]. This leads to the conclusion that gait might be a potential biometric for surveillance systems. A biometric is a descriptive measure based on the human behavioral or physiological characteristics which distinguishes a person uniquely among other people; this unique description should be universal and permanent. Currently, as most biometric systems are still in their infancy, the use of biometrics is limited to identity verification and authentication [4]. Gait is an emergent biometric which is increasingly attracting the interests of researchers as well as the industry. Gait is nothing but the way of walking. Early studies by Murray revealed that gait might be a useful biometric for people identification, a total of 20 feature components including spatial displacement, ankle rotation and vertical tipping of the trunk have been identified to render uniquely the gait signature for every individual, while some of these features are difficult to extract using current computer vision systems, others are not consistent over time for the same person [5]. In one of the early experiments on gait recognition conducted by Cutting et al in 1978, it was demonstrated that people can recognize others just by gait cues [2].

Although gait recognition is still a new biometric and is not sufficiently established to be deployed in real world applications such as surveillance system, but it has the potential to overcome most of the limitations that other biometrics suffer from such as face, fingerprints and iris recognition. Face recognition in many cases has been proven to be unreliable for visual surveillance systems; this is due to the fact that people can disguise or hide their faces as well as that video data being captured can be inadequate at low resolution. Furthermore, another major drawback of face identification in security applications is its low recognition rates in poor illumination. Because most of the facial features cannot be recovered at large distances even using night vision capability [6]. Although fingerprint and iris recognition have proved to be robust for applications where authentication is required, such biometrics are inapplicable for situations where the subject's consent and cooperation are impossible to obtain.

## 2.2 Gait Analysis for Human Identification

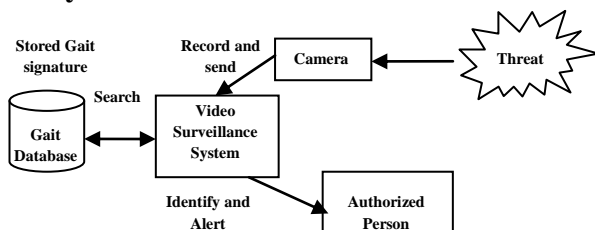
The term gait recognition signifies the identification of an individual from a video sequence of the subject walking. This does not mean that gait is restricted to walking; it can also be applied to running or various types of movement on foot.

Gait as a biometric is advantageous over other forms of biometric identification techniques for the following reasons:

- **Unobtrusive**- Gait of a walking person can be extracted without the user knowing they are being analyzed and without any cooperation with the user in the information gathering stage unlike fingerprinting or retina scans.
- **Distance recognition**- Gait of an individual can be captured at a distance, unlike other biometric techniques such as iris recognition
- **Reduced detail**- Gait recognition does not require images captured in very high resolution, unlike other biometric techniques such as iris recognition, which can be easily affected by low resolution images.
- **Difficult to conceal**- Gait of an individual is difficult to cover up, if they try to do so the individual will probably appear more suspicious. Unlike other biometric techniques such as face recognition, the individual's face can easily alter their looks.

## 3. REAL TIME VIDEO SURVEILLANCE SYSTEM

### 3.1 System Scenario



**Fig1: System Scenario**

The system scenario for a video surveillance system is depicted in Figure 1.

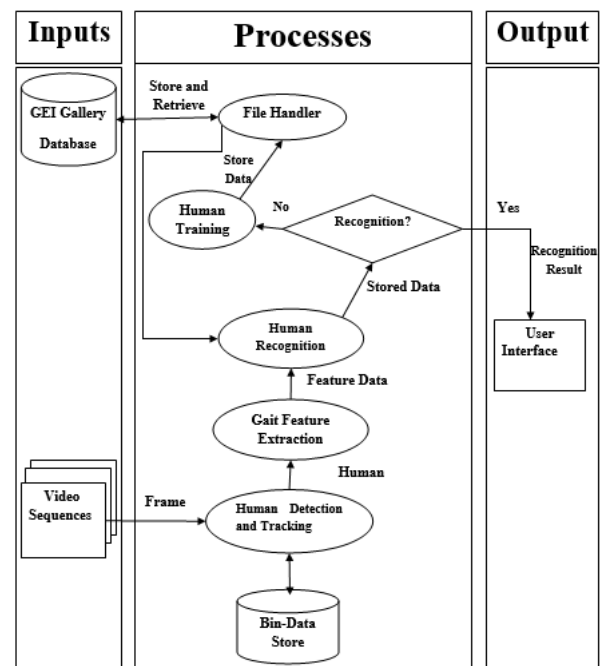
In this above system scenario, we are analyzing the video stream from surveillance cameras. If some individual walks by the camera whose gait signature has been previously recorded and they are a known threat, then the system will identify them and the appropriate authorities can be

automatically alerted and the person can be dealt with before they are allowed to become a threat. If the threat has been detected successfully from a distance, then it will give a time buffer for authorities to take an action.

### 3.2 System Architecture

The most important objective of this proposed system is to:

- Develop a system that is capable of performing recognition of individuals derived from a video sequence of a walking person. The system should be able to store the derived gait signature for comparison at a later stage.
- In order to automate the classification process, automatic extraction of relevant gait features should be available from a video sequence.



**Fig2: System Architecture**

System Architecture contains:-

- **Inputs :**
  - **Video sequences:** Frames captured from a particular video.
  - **GEI Gallery Database:** contains GEI generated based on gait cycle detection.
- **Processes:**
  - **Bin-Data store:** Connected components data gets stored here.
  - **Human detection and tracking:** Detect human (object of interest) and then track him/her.
  - **Gait feature extraction:** Extracting the gait signature of an individual.
  - **Human training:** Use Principle Component Analysis for training extracted features.
  - **Human recognition:** Identify human.
  - **File handler:** It acts as an interface between bin-data store (input) and human recognition and human training.
- **Output:**

- **User interface:** Interface which will notify if the human is identified or not and display information about the identified person.

System Architecture in Fig. 2 works in following way:-

- Step 1 :-** Parse incoming video sequence into frames and extract images.
- Step 2 :-** Convert images into gray scale images.
- Step 3 :-** Perform Human Detection and Tracking (Refer Bin-Data for it.)
- Step 4 :-** Perform Feature Extraction
- Step 5 :-** From Extracted Feature perform Human Recognition, for that purpose system will refer GEI Gallery Database via File Handler for retrieving stored GEI of that particular human.
- Step 6 :-** If Human is recognized, then recognition result will be displayed on user interface
- Step 7 :-** Else Human Training will be perform to store GEI of that human. And that GEI will be stored in GEI Gallery Database via File Handler.

### 3.2 System Phases

There are 3 phases in this system:

- Human Detection and Tracking
- Feature Extraction
- Training and Recognition

## 4. PHASE 1: HUMAN DETECTION AND TRACKING

The advanced video surveillance system needs to analyze the behaviors of people in order to prevent the occurrence of the potential threat. Analysis of behavior of people requires the human detection and tracking system.

Aim of Human Detection and Tracking in this system is-

- To extract a good quality Human(s) silhouette image and
- To track this silhouette(s), from a video frame.

This is needed in order to perform the extraction of the gait feature from the walking sequence. The main aim of image segmentation in this system is to extract a good quality binary foreground image, from a video frame.

### 4.1 Background Subtraction

The background subtraction system is used to provide foreground image through the threshold of difference between the current image and reference image. If the reference image is the previous frame, this method is called temporal differencing.

The temporal differencing is very adaptive to dynamic environment, but generally does a poor job of extracting all relevant feature pixels. The mixture of Gaussian, Nonparametric Kernel, and codebook can behave better performance, but they need extra expensive computation and more memories. For realtime system, especially integrated with (Digital Video Recorder) DVR system, their cost is worthy, because the CPU usage and memories usage are very significant for system stability.

In proposed system, we can implement Gaussian model to be our method for background subtraction, model in moving object detection as this method can handle most of tough situations such as sudden light change, heavy shadow, the objects in background remove etc.

In this model, we use the running average as background model, and use the equation (1) and (2) to update in order to adapt to gradual light change.

$$\mu_{n+1} = \alpha \mu_n + (1 - \alpha) P_{n+1} \quad (1)$$

$$\sigma_{n+1} = \alpha \sigma_n + (1 - \alpha) |\mu_{n+1} - P_{n+1}| \quad (2)$$

Where  $\mu_n$  is a running average,  $\sigma_n$  is a standard deviation,  $P_n$  is pixel value,  $\alpha$  is updating rate in the  $n$ th frame.

In process to filter out some noises caused by camera movement, water wave and tree leaf shaking, we have to figure out the new way to create the difference image between the current image and the background image because the traditional way is not successful for dealing with such situations [7].

### 4.2 Connected Components Labeling

Connected component labeling can be applied on binary or gray level images and different measures of connectivity are possible. Connectivity checks are carried out by checking the labels of pixels that are North-East, North, North-West and West of the current pixel (assuming 8-connectivity) [11].

#### • Two-Pass Algorithm

Relatively simple to implement and understand, the two-pass algorithm iterates through 2-dimensional, binary data. The algorithm makes two passes over the image: In First pass it record equivalence and then assign temporary labels and In case of second pass it replace each temporary label by the label of its equivalence class. The input data can be modified or labeling information can be maintained in an additional data structure [11].

Here, the background classification is specific to the data, used to distinguish salient elements from the foreground. If the background variable is omitted, then the two-pass algorithm will treat the background as another region [10].

### 4.3 Object Tracking

In order to identify multiple human at the same time (move at same time in the video) we need to track the individuals and the human model has to be created for each individual. For that purpose we are using Appearance based tracking method. In this method we use the color histogram, velocity, the number of pixels and size as the human model to describe the humans. For tracking, we assume the human always moves in similar direction and similar velocity. During the process of tracking, we will check whether the people stop or change the direction. If the person doesn't move for period of time, we will check whether this person is false. Once the false person is found, system will learn this false alarm and adjust the background [7].

### 4.4 Object Classification

The aim of different smart visual surveillance applications is to extract semantics from video to be used in higher level analysis. Typical video scenes may contain a variety of objects such as people, vehicles, plants, animals, natural phenomenon (e.g. rain, snow). However, in surveillance applications, main target of interest are generally humans only.

Moving regions detected in video may correspond to different objects in real-world such as pedestrians, vehicles, etc. In order to track it reliability, it is very important to recognize the type of a detected object. Currently, there are two approaches for finding moving object classification which are motion-based and shape-based methods. Shape-based methods make

use of the object's 2D spatial information whereas motion-based methods use temporally tracked features of objects for the classification solution.

In this system, for human recognition we can use Shape-Based Approach to implement object classification using Jianpeng Zhou and Jack Hoang Algorithm's based on codebook theory which classify the human from other objects. The design of the codebook is critical for the classification. The partial distortion theorem for design codebook is that each partition region makes an equal contribution to the distortion for an optimal quantizer with sufficiently large  $N$  [18]. Based on this theorem, we used distortion sensitive competitive learning (DSCL) algorithm to design the codebook, which is explained in [7].

## 5. PHASE 2: FEATURE EXTRACTION

After detecting the subject the next step is to extract some useful features. For human motion analysis we use model-based approach. The Model-based, a shape model is established to match real images to this predefined model, and thus by extracting the corresponding features once the best match is obtained. Stick model is the most commonly used. Due to the advantages of the model-based approach it is being used for human motion analysis. It extracts detailed and accurate motion data.

### 5.1 Feature Extraction Types

There are 2 types of two types for feature extraction:

- Model-Based Feature Extraction
- Feature Selection on Gait Energy Image (GEI)

For the Model-Based Feature Extraction, we use a new method to extract human joints with better accuracy than blobs via incorporating prior knowledge to refine accuracy.

This technique has capability to extract moving joints of human body with high accuracy for indoor data as well as outdoor data filmed in an unconstrained environment. But this approach got some problems in the implementation such as it's highly complex, it takes long processing time. So, instead of Model-Based Feature Extraction we can use feature Selection on Gait Energy Image. We use a gait feature selection method to automatically select covariate condition invariant features for gait recognition.

Gait Energy Image (GEI) is selected for gait representation, which is a spatio-temporal gait representation, constructed using silhouettes. Conventional silhouette based gait representations treat gait as a sequence of templates. In contrast, GEI represents gait using a single image which contains information about both body shape and human walking dynamics. GEI is thus a compact representation which makes it an ideal starting point for feature selection since it is computationally expensive if the number of features to select is high. In spite of its compactness, it has been demonstrated that GEI is less sensitive to noise and able to achieve highly competitive results compared to alternative representations [15].

### 5.2 Gait Energy Image (GEI) Feature Extraction

#### 5.2.1 Gait Cycle Detection

Gait Periodicity can be estimated by a simple strategy. We count the number of foreground pixels in the silhouette in each frame over time. The number will reach the maximum when the two legs are farthest apart (i.e. full stride stance),

and drop to a minimum when the legs overlap (i.e. heels together stance) [17]. But it is difficult to get the minimum or maximum number as the frames intensity change frequently. So we calculate the Average intensity of  $k$  consecutive frames.

#### 5.2.2 Size Normalization & Horizontal Alignment

Before extracting features, we should normalize all silhouette images to be the fixed size, then centroids of an image is calculated.

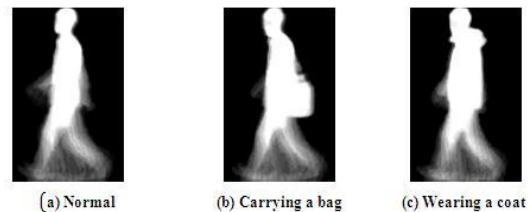
#### 5.2.3 Gait Representation

Given a human walking sequence, a human silhouette is extracted from each frame using the method in [14][17]. After applying size normalization and horizontal alignment to each extracted silhouette image, gait cycles are segmented by estimating gait frequency using a maximum entropy estimation technique presented in [14][17].

Gait Energy Image (GEI) is then computed as

$$G(x, y) = \frac{1}{N} \sum_{t=1}^N I(x, y, t) \quad (3)$$

Where  $N$  is the number of frames in a complete gait cycle,  $x$  and  $y$  are the image coordinates, and  $t$  is the frame number in the gait cycle [15].



**Fig 3: Gait Energy Images of a person under different carrying and clothing conditions, a subject from CASIA database [21].**

Examples of GEIs are shown in Fig. (3) Note that pixels with high intensity values in a GEI correspond to body parts that move little during a walking cycle (e.g. head, torso), while pixels with low intensity values correspond to body parts that move constantly (e.g. lower parts of legs and arms). The former mainly contain information about body shape and posture, whilst the latter tells us more about how people move during walking. We call the former static areas of a GEI and the latter dynamic areas of a GEI [14]. The dynamic areas are invariant to human appearance changes; they seem to be the most informative part of the GEI representation for human identification. The static areas of a GEI also contain useful information for identification (e.g. one's hair style). However, since they mainly contain body shape information, they can introduce variations that hinder human identification through gait [15].

## 6. PHASE 3 - TRAINING AND RECOGNITION

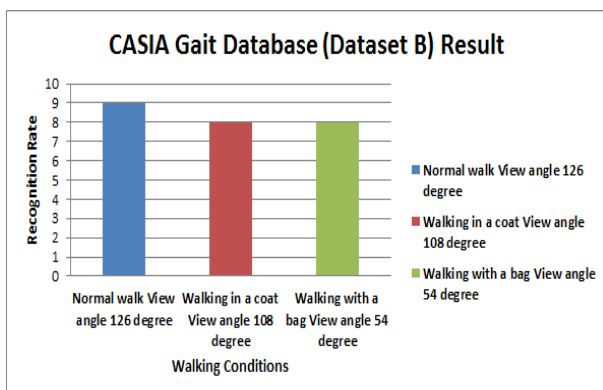
**Training** - The process of storing the extracted features (i.e. probe GEI) and the information needed about the trained humans (i.e. label, name, address etc.) in the gallery database to be used later for the recognition of walking humans. Training should be performed in a special environment with special conditions to get the best motion patterns.

**Classification-** The process in which individual items are placed into groups based on quantitative information on one or more characteristics inherent in the items (referred to as traits, variables, characters, etc) and based on a training set of previously labeled items. In this phase all GEIs stored in the Gallery will be retrieved and grouped into classes. Then the new features (i.e. probe GEI) will be assigned to one of the classes that has the minimal distance. Gait recognition can be performed by matching a probe GEI to the gallery GEI that has the minimal distance between them. However, direct template matching has been shown to be sensitive to noise and small silhouette distortions [14]. This is because the features extracted from silhouettes are usually high-dimensional. Working with huge vectors and comparing them and storing them is a computationally expensive, time consuming and needs a lot of storage space. Consequently, almost always a dimensionality reduction method is used to find most dominant features and remove redundant or less important once. When we obtain series of training GEI templates for each individual, the problem of their excessive dimensionality occurs. To reduce their dimensionality, there are two classical linear approaches for finding transformations for dimensionality reduction—Principal Component Analysis (PCA) and its variants Multiple Discriminant Analysis (MDA). PCA seeks a projection that best represents the data in a least square sense, while MDA seeks a projection that best separates the data in a least-square sense [20]. Huang et al. [19] combine PCA and MDA which seeks to project the original features to a subspace of lower dimensionality so that the best data representation and class separability can be achieved simultaneously [20], where in we can obtain multiple discriminant vectors of each gallery GEI and the probe GEI. This information in turn will help us to recognize the particular human.

## 7. EXPERIMENT SETUP AND RESULT

### 7.1 CASIA Gait Database (Dataset B)

In this experiment, we use the CASIA Gait Database (Dataset B). The database consists of 124 persons. Every person has 3 walking conditions- normal walk, walking in coat, and walking with bag. In this dataset view angles are ranging from 0° to 180°. For training purpose we took all 3 mentioned condition with view angle 90°. We analyse/test them on 3 condition ,a) Normal walk – View angle 126° b) Walking in coat- View angle 108° c) Walking with bag- View angle 54°



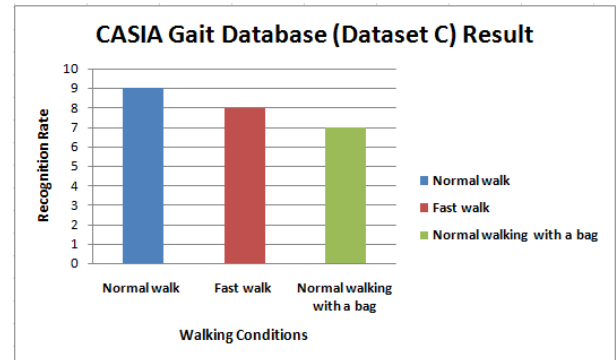
**Fig.4:Result of various walking conditions and view angles on Dataset B**

Fig.5, depicts that, for normal walk condition, out of 10 persons system recognized 9 persons, on the other hand for walking in coat and walking with bag out of 10 persons

system recognized 8 persons. This in turn gives us 83.33% of efficiency.

### 7.2 CASIA Gait Database (Dataset C)

In this case, we use the CASIA Gait Database (Dataset C), where images are collected by an infrared (thermal) camera. This database consists of 153 persons and takes into account 3 walking conditions: normal walking, fast walking, and normal walking with a bag. These all videos are captured at night.



**Fig.5:Result of various walking conditions on Dataset C**

Fig.7, depicts that, For normal walk condition, out of 10 persons system recognized 9 persons, for fast walk condition, out of 10 persons system recognized 8 persons and for normal walking with bag out of 10 persons system recognized 7 persons. This in turn gives us 80% of efficiency.

However the efficiency achieved in above two scenarios cannot be generalized as it is performed on less number of test cases and conditions under which they are tested may be changed on other time.

## 8. CONCLUSION

Here for real time video surveillance system, we used new spatio-temporal technique known as Gait Energy Image (GEI), and for getting maximum efficiency we used Principal Component Analysis (PCA) and Multiple Discriminant Analysis (MDA), where PCA maximizes the variance in all the variables accounted for by the factor whereas, MDA maximizes the difference between values of the dependent. Basically PCA and MDA are used for learning features from the trained GEI templates. And then based on the learned features recognition is then carried out.

## 9. REFERENCES

- [1] G. Johansson, "Visual Perception of Biological Motion and a Model for its Analysis", *Perception and Psychophysics*, vol.14(2):201-211, 1973.
- [2] L. T. Kozlowski and J. E. Cutting, "Recognizing the Gender of Walkers from Point-Lights Mounted on Ankles: Some Second Thoughts", *Perception & Psychophysics*, 23(5):459, 1978.
- [3] N. H. Goddard, "The Perception of Articulated Motion: Recognizing Moving Light Displays". PhD thesis, University of Rochester, 1992.
- [4] A. K. Jain, R. Bolle, and S. Pankanti, editors, "Biometrics: Personal Identification in Networked Society", Kluwer academic publishers, 1999.

- [5] M. P. Murray, "Gait as a Total Pattern of Movement. American Journal of Physical Medicine", 46(1):290-333, 1967.
- [6] A. Kale, A. N. Rajagopalan, N. Cuntoor, and V. Kruger, "Gait-Based Recognition of Humans using Continuous HMMs", In Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, pages 321-326, 2002.
- [7] Jianpeng Zhou and Jack Hoang, "Real Time Robust Human Detection and Tracking System", In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops - Volume 03, page 149, June 2005.
- [8] S.J. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler, "Tracking Groups of People", Computer Vision and Image Understanding, 80:42-56, 2000.
- [9] P.L. Rosin and T. Ellis, "Image Difference Threshold Strategies and Shadow Detection", In Proceedings of 6<sup>th</sup> British Machine Vision Conference, pp. 347-356, 1995.
- [10] Lifeng He; Yuyan Chao, K. Suzuki, "A Run-Based Two-Scan Labeling Algorithm", IEEE Transactions on Image Processing, 17(5):749-756, May 2008.
- [11] Sunny Kumar, Pratibha Sharma, "Offline Handwritten & Typewritten Character Recognition using Template Matching", IJCSET, 4(6):818-825, 6 June 2013.
- [12] Imed Bouchrika and Mark S. Nixon, "Model-Based Feature Extraction for Gait Analysis and Recognition", In Proceedings of 3rd international conference on Computer vision/computer graphics collaboration techniques, pp. 150-160, 2007.
- [13] Yigithan Dedeoglu, "Moving Object Detection, Tracking And Classification For Smart Video Surveillance", Bilkent University, 2004.
- [14] Khalid Bashir, Tao Xiang and Shaogang Gong, "Feature Selection for Gait Recognition without Subject Cooperation", In Journal of Pattern Recognition Letters, 31(13):2052-2060, 2010.
- [15] Khalid Bashir, Tao Xiang, Shaogang Gong, "Feature Selection On Gait Energy Image For Human Identification", IEEE transaction, PP 985-988, 2008.
- [16] A. Aguado, M.S. Nixon, M.E. Montiel, "Parameterising arbitrary shapes via Fourier descriptors for evidence-gathering extraction", computer vision and image understanding, 69(2):202-221, 1998.
- [17] S. Sarkar, P. Phillips, Z. Liu, I. Vega, P. Grother, and K. Bowyer, "The humanID gait challenge problem: Data sets, performance, and analysis", IEEE Transactions on PAMI, 27(2):162-177, 2005.
- [18] A. Gersho, "Asymptotically optimal block quantization" IEEE Transaction on Information Theory, vol. 25, pp. 373-380, July 1979.
- [19] P.S. Huang, C.J. Harris, and M.S. Nixon, "Recognizing Humans by Gait via Parametric Canonical Space," Journal of Artificial Intelligence in Eng., vol. 13, pp. 359-366, 1999.
- [20] J. Han and B. Bhanu, "Individual recognition using gait energy image", IEEE Transactions on PAMI, 28(2):316-322, Feb 2006.
- [21] Yu, S., Tan, D., Tan, T., "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition", In 18th Internat. Conference on Pattern Recognition, pp. 441-444, 2006.