An Enhanced Strategy for Human Behavior Classification in Crowd

Submitted by

Robina Khatoon
Reg. # 586-FBAS/MSCS/F09

Project Supervisor

Syed Muhammad Saqlain
Assistant Professor
International Islamic University, Islamabad

Faculty of Basic and Applied Sciences
Department of Computer Science and Software Engineering
International Islamic University, Islamabad

1- Computer science

2. Automatic data processing

A dissertation submitted to the
Department of Computer Science and Software Engineering,
International Islamic University, Islamabad
as a partial fulfillment of the requirements
for the award of the MS (CS) degree

Declaration

I, Ms. Robina Khatoon solemnly declare that this thesis, neither as a whole nor as a part thereof has been copied out from any source. It is further declared that this dissertation is made entirely on the basis of my personal efforts under the sincere guidance and help of my supervisor. No portion of the work presented in this report has been submitted in support of any application for any other degree or qualification of this or any other university or institute of learning.

Robina Khatoon
Reg. # 586-FBAS/MSCS/F09

Abstract

Surveillance of specific places for recognizing abnormal events is under developed area of computer vision. This work presents a technique to recognize human behavior in crowd which is needed for surveillance. The techniques discussed in the literature mostly address problems like detection of human in crowd and overall behavior of the crowd is recognized by analyzing the crowd flow. Research has been done to recognize the behavior of individuals in the video sequences. A little work is done to recognize the individual behavior in a crowd so this research basically based on the area of recognizing individual behavior in a crowd. Human behavior modeling is a difficult task as it is based on human detection, tracking and analyzing normal and abnormal. The method presented in this research consist of sub-parts which are required to be followed in order to recognize human behavior. These sub-parts include segmentation, star skeletonization and classifier for recognizing human behavior.

Glossary

This section lists symbols and acronyms that frequently appear in this thesis.

Acronyms and Abbreviations

SRG	Seeded region growing
PCA	Principal Component Analysis
HMM	Hidden Markov Model
SVM	Support Vector Machine
S-LDA	Semi-Latent Dirichlet Allocation
S-CTM	Semi-latent Correlated Topic Model
RVM	Relevance Vector Machine
HOG	Histogram of Gradients
FAST	Features from Accelerated Segment Test
NWFE	Nonparametric Weighted Feature Extraction
LBP	Local Binary Pattern
S-E	Sampling-Expectation
KL	Kullback-Leibler
MOGs	Mixture of Gaussians
FCM	Fuzzy C-Mean
EM	Expectation Maximizing
TRF	Topic Random Field
MRF	Markov Random Field
DoG	Difference of Gaussian
RBF	Radial Basis Kernel

Acknowledgement

All praise to Almighty Allah, who gave us the understanding, courage and patience to complete this project.

Thanks to my parents and all well wishers, who helped me during my most difficult times and it is due to their untiring efforts that I am at this position today.

I express my gratitude to my teacher *Syed Muhammad Saqlain* for providing me opportunity to learn and enhance my knowledge. He had been ready to help and guide me throughout the project in any possible way and created dataset for experimentation. Second, my thanks go to *Shafina Shah* and *Umm-e-Zahoora* for the constructive discussions about my work and helping me during my most difficult times. Special thanks go to the teachers and students who act in videos and making it possible for me to generate required dataset for experimentation.

Last but not least thanks to my friends, class fellows especially *Ayesha Manzoor* and *Asmaah Jabeen* for their unending support throughout the course of project.

Department of Computer Science and Software Engineering
International Islamic University, Islamabad.

Dated: 05-07-2012

Final Approval

It is certified that we have read the thesis, titled **"An Enhanced Strategy for Human Behavior Classification in Crowd"** submitted by **Miss Robina** Reg. No. **586-FBAS/MSCS/F09**. It is our judgment that this thesis is of sufficient standard to warrant its acceptance by the International Islamic University, Islamabad, for the Degree of **Master of Science**.

Committee

External Examiner:
Dr. Abdus Sattar,
Former D.G. Pakistan Computer Baeuro

(Dr. Abdus Sattar)

Internal Examiner:
Dr. Ayyaz Hussain
Assistant Professor,
Department of Computer Science & Software Engineering,
International Islamic University Islamabad

(Dr. Ayyaz Hussain)

Supervisor:
Syed Muhammad Saqlain
Assistant Professor,
Department of Computer Science & Software Engineering,
International Islamic University Islamabad

(Syed Muhammad Saqlain)

Department of Computer Science & Software Engineering
International Islamic University Islamabad

Final Approval

Dated: ________

This is to certify that the Department of Computer Science, International Islamic University Islamabad, accepts this dissertation submitted, in its present form as satisfying the dissertation requirements for the degree of MS (CS).

Committee

Supervisor:
Syed Muhammad Saqlain
Assistant Professor,
Department of Computer Science & Software Engineering,
International Islamic University Islamabad

(Syed Muhammad Saqlain)

External Examiner:
Dr. Abdus Sattar,
Former D.G. Pakistan Computer Baeuro

(Dr. Abdus Sattar)

Internal Examiner:
Dr. Ayyaz Hussain
Assistant Professor,
Department of Computer Science & Software Engineering,
International Islamic University Islamabad

(Dr. Ayyaz Hussain)

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Surveillance is the monitoring of the behavior, activities, or other changing information in a secret manner. Surveillance can be carried out in different ways like biometric surveillance, camera surveillance, Ariel surveillance, computer surveillance etc [1]. Camera surveillance is under consideration in which camera is placed for observing an area. In past, human monitor the camera footage which requires great human effort and do not give promising results. A lot of work is done to automate such type of surveillance systems. Monitoring human behavior is under developed area.

In order to keep security measures most of the people pay attention to visual surveillance so they use different security camera and computer vision software which alert them in case of any abnormal event. Such type of surveillance systems are needed on different crowded places like airports, shopping malls, stations and private residential areas. It is useful for recognizing and monitoring threats, preventing criminal activity and the videos can be used for future investigation of any criminal activity.

The main tasks involved in visual surveillance include object detection, tracking and behavior recognition which are widely studied in computer vision. The main goal of object detection, tracking and behavior recognition is to give the vision to the machine like human, so it can track and recognize behavior of the target in image sequences. A lot of progress has been made in this field especially during the past ten years. However, some of the issues need to be solved before one can implement a robust video surveillance system for commercial applications.

1.1 Motivation

Most of the research done in computer vision focuses on problems such as image segmentation, object recognition and tracking, extracting objects features and so on. However, the most vital and difficult task of computer vision is to understand human activities. An advance visual surveillance system should be able to infer what is happening in the scene and raise warning if any abnormal event occurs. It is useful for recognizing and preventing criminal activities and the footage obtained from the surveillance camera can be used for future investigation of any criminal activity.

Many visual surveillance systems has been developed which can recognize abnormality in crowd by analyzing crowd flow. A little work is done in the area of recognizing individual behavior in the crowd. Such type of work may follow the steps like image segmentation, human detection, tracking and classification so following all these steps for single individual is a complex task to do. Some type of abnormal event may not change the behavior of overall crowd but may affect two or three persons in the crowd. The aim of this study is to recognize the individual's normal and abnormal behavior in the crowd and present a real-time surveillance system.

1.2 Objectives and Contributions

This thesis provides insights into some key issues in video surveillance; these issues include shadow removal human detection and human behavior classification. The main objective of this thesis is to propose a system which recognizes human normal and abnormal behavior in order to avoid abnormal activity in crowded places. The main contributions of this thesis is that it overcome the problem of detecting shadows as human by using shadow removal technique; it uses a better human detection method in order to overcome the problem of classifying vehicles as group of people which may increase accuracy of the proposed system. For instance, real-time processing is a desirable feature. More precisely, we require that the approach satisfies the constraints like single and fixed monitoring camera is used. A reasonable image resolution and frame rates are required since we have chosen to implement 2D plan information based method; a minimal frame-rate is required to capture most of the motion. A high density level of crowd is to be avoided due to the increasing probabilities of occlusions and mismatches in people detection.

1.3 Goals and Challenges

The main goal is to develop a real-time application which can be used in different public areas to find out the occurrence of any abnormal event and present a technique which out-performs the existing technique [29]. The main challenges are to find out such a human detection method that gives higher accuracy and incorporate it with other parts of the application. Secondly use a method of shadow removal to avoid a situation where shadows are classified as human.

1.4 Overview of Manuscript

In this section elaborates of the contents of this manuscript which is structured into five main chapters. The next chapter focuses on different techniques of segmentation, shadow removal and human detection. Chapter 3 contains previous work of researchers and the problems in the existing technique. The proposed method is over-viewed and detailed in chapters 4 and 5. The six and last chapter presents conclusions where a review of the applications and future work are also presented.

Chapter 2 discuss the key concepts of image segmentation, shadow removal, human detection, feature extraction and classification

Chapter 3 recalls the previous work related to image segmentation, shadow removal and human detection. The problems in the existing technique are also discussed in this chapter.

Chapter 4 provides detail design of the proposed system. It contains the description of proposed method for classification of human normal and abnormal behavior.

Chapter 5 discuss in detail about the datasets used for experimentation and which performance measures are used to evaluate the results. The evaluated results are also discussed in detail.

Chapter 6 consists of conclusions where a review of the applications and future work is also presented.

Chapter 2

Preliminary Knowledge

Computer Vision is field which includes methods for acquiring, processing, analyzing, and understanding of images in order to form decisions [2]. Computer vision is the broader research field which includes robotics, image retrieval, photo interpretation, computer graphics, video analysis and more. Algorithms in computer vision are generally categorized into three levels [3]:

1. Low-level vision algorithms in which image processing is directly related to pixels without deep analysis
2. Middle-level vision algorithms which includes pattern matching, object detection and tracking
3. High-level vision algorithms in which interpretation and semantics are extracted from images

Following are some applications of computer vision [2]:

- Controlling processes which focuses on industrial robot
- Detecting events which include visual surveillance
- Managing information which includes database indexing of image sequences
- Modeling objects or environments for example medical image analysis or topographical modeling
- Interaction for example work as the input to a device for computer-human interaction
- Automatic inspection, e.g. in industrialized application
- Geometric reconstruction: modeling, forensics and special effects
- Image and video editing
- Webcasting and indexing digital video

2.1 Image Segmentation

In computer vision, segmentation is the process of separating a digital image into multiple segments (sets of pixels, also known as super-pixels). [4] The purpose of segmentation is to represent the pixel into a meaningful form so that the analysis becomes easy. Image segmentation is done to detect objects and boundaries (edges) from the image which is later on used for recognizing behavior of the detected objects or for classification of the detected objects. It is used as a pre-processing step in medical imaging, surveillance system, vehicle tracking systems, human identification systems and so on. There are different methods of segmentation some of them are discussed below:

2.1.1 Thresholding

Thresholding basically involves transforming a color or grayscale image into a one bit binary image. [5] Each pixel in an image is assigned either black or white color depending on their color value. The assignment of black and white color basically depends on a value which is called as threshold. The pixels of the image values higher than the threshold are set to white otherwise black color. The following figure represent thresholding idea the best.

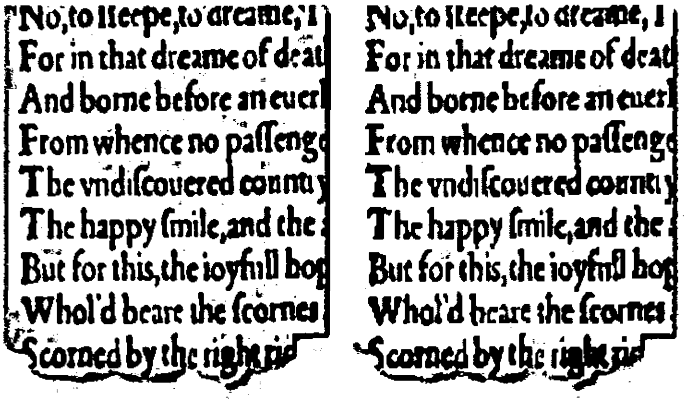


Figure 2.1: Thresholding example

In the figure 2.1, left image contain 24-bit information and after thresholding the image on the right is obtained which contains one bit information.

2.1.2 Clustering Methods

This method represents the image in compact form by a set of components. A set of pixels having same features are combined in the form of clusters.

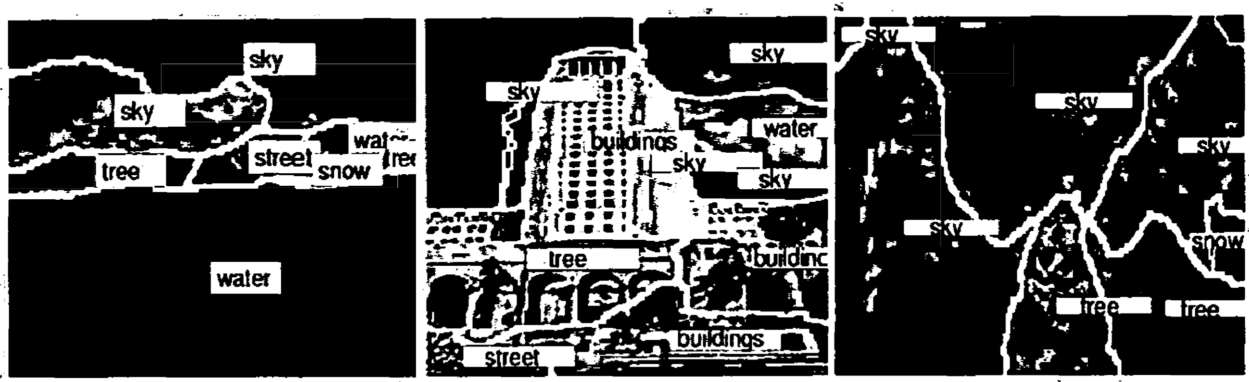


Figure 2.2: Example of clustering method

The K-means algorithm is an iterative technique that is used to divide an image into *K* clusters. [4] This non-hierarchical method at first takes the number of components of the population equal to the final required number of clusters in this step the final required number of clusters is chosen such that the points are equally farthest apart. [6] In the next step it creates clusters by examining each component in the population by comparing them to the minimum distance. When a new component is added to the cluster the centroid value is calculated and this is process is repeated until all the components are grouped into the required number of clusters.

2.1.3 Region Growing Methods

Seeded region growing (SRG) is a robust and simple method of segmentation which is fast and free of tuning parameters for segmentation introduced by [7]. For the selection of high level knowledge of image components is required and so, it makes implementation easy and can be applied on large datasets. The only drawback of the SRG algorithm is the difficulty in automating seed generation and dependency of output on order sorting of pixel as different order of processing pixels during region grow process leads to different final segmentation results. [8]

Seeded region growing approach divides an image into regions with respect to a set of seeds as presented in [9]. S1, S2... Sq are the set of seeds where in each step SRG add a pixel to the seed set. The initial seeds are further changed by the centroids of these generated homogeneous regions by involving the extra pixels step by step. The pixels are labeled by same symbol if they are in the same regions which are called as allocated pixel and the pixels are labeled by different symbols if they are in the variant regions which are called as unallocated pixels.

2.1.4 Mixture of Gaussians Model

In this method each pixel is represented by a mixture of Gaussian functions that sum together to form a probability distribution functions. [10] The probability of observing a pixel c at some time instant is given by:

$$P(c) = \sum_{i=1}^{n} P(M_i)P(c \mid M_i) \qquad (2.1)$$

where M is the *i-th* mixture component, represented by a mean μ_i and a $p \times p$ covariance matrix $\sum_i$. The mean of each Gaussian function can be thought of as an educated guess of the pixel value in the next frame. The standard deviations and weight of each component are measures of confidence in that guess. Confidence is higher if higher weight and lower standard deviation. Mostly 3-5 Gaussian components are used per pixel. To determine if a pixel is part of the background or foreground we compare the pixel to the Gaussian components. If the pixel value is within a threshold value of a background component's standard deviation σ, it is considered part of the background otherwise pixel is part of foreground.

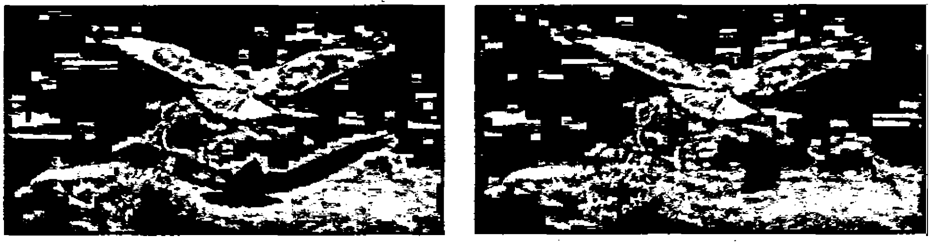
Mixture of Gaussians is the most complex method which gives us good performance but presents a tricky parameter optimization problem.

2.2 Shadow Removal

Shadow appears in images in which strong illumination variation exists. Existence of shadow reduces the accuracy of the computer vision application. It is used as a pre-processing step for different types of computer vision applications such as tracking, object detection, object recognition, human behavior classification. The reason for applying shadow removal is that shadows can cause object merging, object shape distortion or object loses and shadow pixel can become a part of foreground so can cause problem in application like object detection.

2.2.1 Ways of Applying Shadow Removal

Shadow removal can be used in two ways; one way is to segment out the shadows from the pre-segmented image and second way is to remove the shadow from image in the way that texture and color of image occupied by shadow can be restored. The former method is followed in [13] and the latter method is used in [11] and [12]. Results of the paper are given as example below:



(a) original image **(b) Image after shadow removal**

Figure 2.3: Removal of shadow from image preserving the texture and color of the original image [12]

(a) Original image **(b) Segmented image** **(c) Segmented image after shadow removal**

Figure 2.4: Removal of shadow from pre-segmented image [13]

There are two types of shadows which are as follows [14]:

1. Cast shadows
2. Form shadows

Form Shadows are created when an object is not directly facing the light source. For example a vehicle is parked where sun-light is falling on it from the backside so the shadow that appears on the car of its front side is a form shadow. **Cast shadows** are created when an object blocks the light source. For example, the shadow of a tree is created by the sunlight which falls on the ground. Cast shadows are darker than form shadows.

2.2.2 Taxonomy of Shadow Detection Algorithms

Various shadow detection approaches have been used for shadow detection. The taxonomy of shadow detection algorithms is given below:

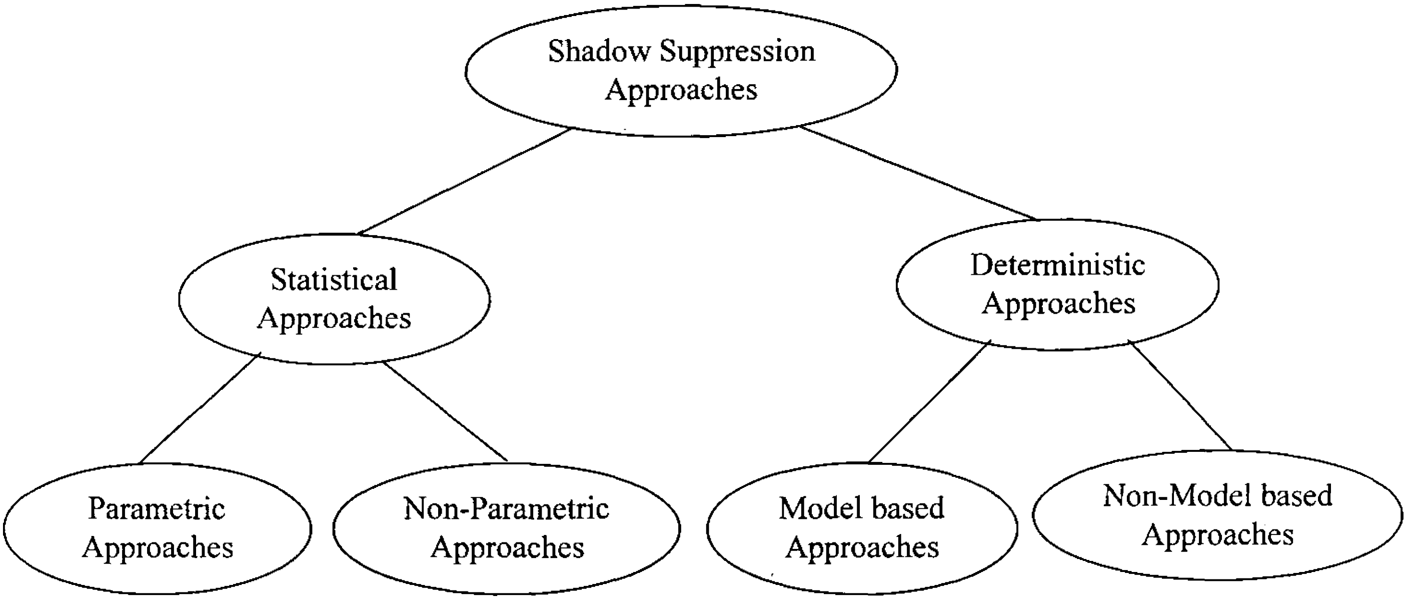


Figure 2.5: Taxonomy of Shadow Detection Algorithms [15]

In figure 2.5 various approaches of shadow detection are organized in two layers. The deterministic approaches use decision process for shadow pixel classification where the statistical approaches use probabilistic functions but the selection of parameter is difficult so it is further subdivided into parametric and non-parametric approach. Deterministic approach is future sub-divided due to the reason that whether on/off decision can be supported by model based knowledge or not. The results obtained from model based method are best but it sometimes becomes too complex and time consuming depending on the model selection as compared to non-model based methods. Different types of features are used for the shadow detection algorithms which are as follows:

1. **Spectral Features** include gray level and color features
2. **Spatial Features** include local and regions features
3. **Temporal Features** include static and dynamic features

Different approaches use different type of features depending on the performance of algorithm. For example an algorithm may give result using spectral features rather than temporal features.

2.3 Human Detection

Human identification in video sequences is the most important step for the surveillance applications. This is most difficult task as video sequences contain various objects other than human and some other features make it difficult which include variations in brightness, lightings and contrast level. Some of the techniques use segmented images for detection of human as in [16] where as some of the technique use images for human detection as in [17]. The features used for human identification vary for both approaches. The techniques which uses segmented images for human identification uses posture, size, speed etc as features whereas the techniques which uses real images for human identification uses features like texture, color, face detection, head tracking etc.

2.4 Summary

This chapter the reviews some of the key concepts of computer vision and their applications is presented. These concepts include image segmentation, shadow removal and human detection. Different approaches of segmentation such as thresholding, clustering method, region growing and mixture of gaussians are presented. Types of shadows and different ways of shadow removal are also part of the discussion. The ways that can be adopted for human identification are also presented.

Chapter 3

Literature Review

The classification of human behavior is done by passing the video sequences from different subsystems which may include image segmentation, shadow removal, human detection etc. A lot of work has been presented over the past 10 years which made enhancements by making use of different methods but some of the imperfections still exist in existing techniques. Some of the preceding efforts of the researchers in area of the sub-systems as well as for human behavior recognition are discussed below:

3.1 Human Behavior Classification

3.1.1 Detection of Emergency Events in Crowded Scenes

Human behavior classification can be applied in many real-time situations. *Andrade et al* presents in [18] an approach for detection of normal and abnormal events in crowd in case of emergency situations.

Firstly preprocessing is applied by background modeling and computation of the optical flow. Feature prototype is extracted by applying Principal Component Analysis (PCA) on optical flow fields of each frame that is obtained by calculating optical flow. The measure of similarity between the video segments is obtained by likelihood of the observations in the segments given by a Hidden Markov model. Similarity is evaluated for spectral clustering. The training is done on each cluster by using bank of HMM (Hidden Markov Model).

The results obtained from experiments show that bank of models is efficiently detecting simulated emergency situation in a dense crowd.

3.1.2 A Surveillance System for Human Behavior Analysis

Hsieh et al proposes in [19] a simple and rapid surveillance system which track human and differentiate between normal and abnormal behavior. Abnormal behaviors include climbing, falling, stopping, and disappearing.

Foreground regions are extracted by the following procedure: firstly frame differencing then holes filling, shadow removal, connected components labeling, and noise removal. Shadow removal method is done by the method proposed by *Cucchiara et al* in [20]. Noise removal is done by applying morphological operations i-e; erosion and dilation. Segmentation of multiple objects is done by analyzing the vertical projection of motion masks. Finally all the extracted foreground regions which are the moving objects are tracked by the integration of location estimation and weighted block-based similarity measurement.

Experimental results show that the system better deals with occlusion and efficiently detect and track the moving objects. Future work will deal with people and vehicle counting, suspicious behavior analysis, theft detection, and so on.

3.1.3 Human Aggressive Behavior Recognition using Binary Local Motion Descriptors

D. Chen et al presents in [21] an approach which is used to detect and monitor human aggressive behaviors. It uses local binary descriptor for detection. Point of interest is detected by the following steps:

- Find interest point using Harris corner detector (find high contrast points both in space and time, calculate the response function value for each pixel and extract local-maxima pixels as interest points)
- Extract cube at each interest point and convert it to binary by thresholding, where threshold is determined from the first frame
- A local binary feature is computed from the binary cube which consist of shape feature and motion feature

- Cluster the local motion features into a fixed size of feature codebook to reduce the feature space

Human behavior is then classified by the following steps:

- Codebook is generated on the basis of histogram, where code words are used as bins
- Each local binary feature is plot to its closest code word and added into the related bin
- A behavior descriptor is treated as a vector with the same size as the codebook
- Use a one-center SVM (Support Vector Machine) to train a model for all normal behaviors and detect aggressive behaviors as outliers

The system is accurate for retrieving aggressive behaviors from video records. This approach models the action of the arm, body, and the object together. The top 10 retrieval results include about 80% aggressive behaviors, which is much better than the random accuracy 36.2%.

3.1.4 Real-time Abnormal Motion Detection in Surveillance Video

N. Kiryati et al presents in [22] a novel approach for real time abnormal event detection. This approach is well suited for applications where limited computing power is required near the camera for compression and communication.

Firstly, the system converts the video to motion vectors by applying intra-frame compression technique. Intra-frames are generated at constant intervals, to allow random-access to the content, and to reduce accumulated errors. Secondly, generate motion features from motion vector by calculating magnitude, direction of the motion vector and total motion in the scene. Calculate area of dominant motion, motion homogeneity and motion direction. In order to evaluate abnormal motion detector; learning is done by the patterns of normal activity. So, in training phase estimate probability density function of feature vectors during normal conditions.

The experimental results show that this system is reliable for real-time operation. The abnormal action videos on which system is tested include running, jumping, and grass crossing actions.

3.1.5 Human Action Recognition by Semi-Latent Topic Models

Y. Wang and G. Mori proposed in [23] two new models for human action recognition which are based on topic models. The presented models are compared with previous topic models of pattern recognition. According to results these models are good in different aspects as:

1. Due to decoupling of model parameters training is much easier
2. The issue of choosing appropriate number of latent topics has been resolved
3. It gives better performance by sing information provided by class labels in training set

In this approach "frames" of video are "words" so each frame corresponds to a "word" and "video sequence" corresponds to a "document". Motion descriptor is used to represent the video frames. As the motion descriptor used is robust so any of the tracking or human detection method can be used. Normalized correlation is used to compare the motion descriptor of two frames. Codebook is constructed by computing affinity matrix where each entry in the affinity matrix is the frames and the similarity between the frames is calculated by using normalized correlation. Codewords of the codebook are the centers of clusters. Finally the replace each frame with its codeword and remove the temporal information. Semi-Latent Dirichlet Allocation (S-LDA) and Semi-latent Correlated Topic Model (S-CTM) are used for action classification.

The systems has its limitation also as it requires to improve the preprocessing stage of tracking and making human figures stable. Future work will be the testing the presented approach on more difficult datasets.

3.1.6 Human Action Classification using RVM

B. Yogameena et al presents in [24] a real-time video surveillance system which is competent of classifying normal and abnormal actions of persons in crowd. Abnormal actions include running, jumping, bending, walking, waving hand and fighting.

Firstly, blob detection subsystem detects the foreground pixels by subtracting a statistical background model. Each foreground blob is labeled as individual or group then each blob containing multiple people is divided into individuals using a projection method. Tracker is then initialized for each identified individual for separating vehicles from individuals. Skeleton points and the motion cues for each blob are selected as features. Individual's action is classified as normal or abnormal action using Relevance Vector Machine (RVM).

In SVM (Support Vector Machine) the number of support vectors can be much larger, so RVM (Relevance Vector Machine) is used to partially overcome that problem. Relevance Vector Machine (RVM) classification technique has been applied in many different areas of pattern recognition, including facial expressions recognition and pose estimation.

3.1.7. Crowd Event Recognition Using HOG Tracker

C. G'arate et al presents in [25] a new approach for crowd event recognition which uses HOG (Histogram of Gradients) tracker for recognizing crowd events such as crowd splitting, formation, walking, running, evacuation etc. This approach deals with overall behavior of the crowd for recognizing crowd events.

The system first detects the moving objects using moving segmentation and then feature points are calculated using FAST (Features from Accelerated Segment Test) approach. A feature vector is a collection of several elements which are the mean HOG descriptor, the start and end point of the trajectory of the tracked feature point, together with start and end time. It computes direction, speed, and crowd density using these elements. The system then calculates minimum distance which improves the feature point distribution for an object and prevents mixing of tracked points. Objects are tracked by tracking the feature points and build a 2D HOG descriptor for each detected feature point.

There are still some errors in the recognized events. This technique needs to improve the threshold computation at the level of scenario models. The computation of HOG motion in 3D is required. It also requires improving threshold computation so it remains independent from the scene. The approach has successfully validated on PETS dataset.

3.1.8 Unexpected Human Behavior Recognition using Multiple Features

This research [26] is related to the unexpected behavior recognition in the high density crowd scenes. The actions recognized by the system include running and fall detection.

The background model for each scene has been trained off line with the Average of Gaussian algorithm and shadow detection has been applied using the NRGB color model in order to have precise motion detection.

The features used by the system to recognize human action are:

i. **Accumulated Hitmap** counts consecutive foreground pixels for each location in an image and sets the corresponding point in the hitmap to the number of consecutive pixels. It is used to find the unexpected behavior.
ii. **Crowd pace** is estimated as ratio of sum of all non overlapping foreground pixels and the sum of foreground pixels.
iii. **Crowd density** counts the foreground pixels in an area.

The training of the system is done with image sequences of normal behavior. A limited numbers of features are used due to the reason of computational performance.

3.1.9 Recognizing Human Actions using NWFE-Based Histogram Vectors

C. H. Lin et al in [27] recognizes human actions using NWFE (Nonparametric Weighted Feature Extraction) based histogram which classifies 10 actions including running, jumping, walking, and bending and so on.

The architecture of the system is given in figure 3.1. Body silhouette is extraction using background subtraction method [28] to extract foreground from the background. Features extracted for classification of human actions are distance signal and the width features but it increase the dimensionality of feature space which is reduced by PCA. The features are then mapped to the symbols in which *k*-means clustering to create a codebook where each codeword is the mean of one cluster. Human action is then recognized by following three steps:

Step 1: Compute the distance between two strings using their histograms, gives histogram vector
Step 2: Reduce the feature space using NWFE before training the classifier. The dimension reduction will enable the parameter estimation to be more accurate for classification purpose
Step 3: Classify human action using Bayes classifier

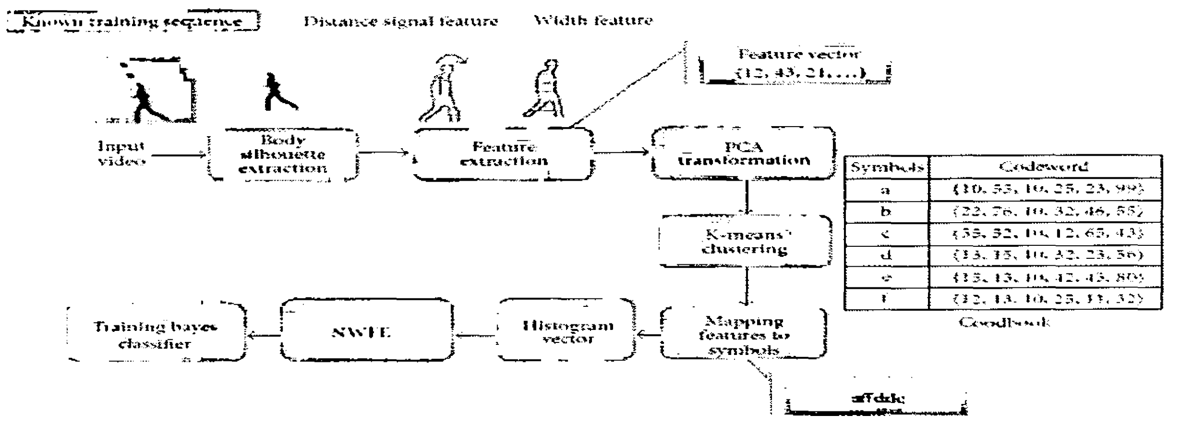


Figure 3.1: Architecture of the system

Time complexity of this method is not affected by size of dataset and because of dimensionality reduction through NWFE (Nonparametric Weighted Feature Extraction) the recognition time reduces from 0.5 to 0.1 msec. The system is computationally faster than the nearest-neighbour classifiers. The system is efficient as it reduces dimensionality of feature vectors using PCA.

3.1.10 Human Behavior Classification in Crowd using Projection and Star Skeletonization

B. Yogameena et al in [29] classifies actions which are person carrying a long bar, walking, bending and waving hand in crowd. This technique has some weaknesses in segmentation, human detection and behavior classification which are needed to be resolved.

In this paper foreground segmentation is accomplished in real-time using adaptive mixture of Gaussians method [30] where the number of gaussians used per pixel is four. Then projection is applied on head and ground plane to estimate the number of people in the group and to separate the individuals as mentioned in [31]. Following issues needed to be addressed while estimating pedestrian counts in groups:

1. Classifies slow-moving vehicles as group of people in certain cases
2. Do not deal with stationary people
3. When shadows appear larger than the people themselves, it classifies these shadows as human
4. The blobs used for estimating the number of people do not supply accurate estimate

Human features are extracted using star skeletonization [32] which gives five extreme points of human skeleton and motion cues. These features are then classified using SVM (Support Vector Machine) classifier, where the results obtained by SVM are not good.

In order to reduce the false alarm rate consider each minimum and maximum point as features, so do not use low-pass filter for smoothing. Experimental results are compared with P CA method.

3.2 Image Segmentation

3.2.1 Multi-Layer Background Subtraction Based on Color and Texture

J. Yao and J. M. Odobez in [33] proposes a background subtraction technique which uses multi-layers of the image and color and texture information of the objects for classification of pixels.

In the proposed method texture and the photometric invariant color measurements are combined for background and foreground detection. Local Binary Pattern (LBP) is used to model texture and the photometric invariant color measurements. The background modeling algorithm can be applied not only on the color images but can also work better for gray-scale images after little changes. Then the distance is calculated which is based on color and texture. The distance is matched with a threshold value; the smaller the distance the pixel have more chances of becoming background. After applying the update of the background model foreground detection is applied.

Local Binary Pattern (LBP) is useful for background modeling as it is robust to monotonic gray-scale changes, and accepts both global and local illumination changes. Local Binary Pattern (LBP) fails when both the background and foreground objects have the same texture. According to the results the proposed system performs better than the MOGs not only for classification of moving background pixel but also for the foreground objects.

3.2.2 Robust Foreground Detection in Video Using Pixel Layers

K.A. Patwardhan et al in [34] proposes foreground detection methods which performs better under dynamic background and moderately moving camera. The foreground clusters are obtained by decomposing the scene into layers.

Initially make a guess that which layer you want to extract from the scene. Then refine the extracted layer using Sampling-Expectation (SE). In order to extract all the layers from the scene the previous steps are performed iteratively. Kullback-Leibler (KL) divergence is used in order to validate that the extracted layer is meaningful and of significant size. Finally each incoming pixels are checked for adaptive models of the background if they do not hold this condition they are classified as foreground.

This techniques adapts to changes in the scene, give good results under moderately moving camera and for dynamic scenes. This technique is proposed to solve the problems including robust foreground and unusual region detection. According to quantitative comparison of the results this technique outperforms than MOGs (Mixture of Gaussians).

3.2.3 Edge-preserving Clustering Algorithms and MRI Image Segmentation

M.A. Balafar et al in [35] presents a technique which improves an already existing extension for Fuzzy C-Mean (FCM). In order to obtain better clustering results in the presence of noise the authors have introduced two extensions for Expectation Maximizing (EM).

FCM (Fuzzy C-Mean) is a clustering algorithm introduced by Bezdek. FCM is based on minimizing an object function. A new extension named FCM_EN was introduced by Szilágyi which speed up the clustering process for input image. In this extension a linearly-weighted sum image from the original image and its average image is used as input image for clustering. The edge-preserving mean of pixel is calculated in which a pixel i is considered as the average value of samples in the proposed neighborhood. In the second extension first mean filter is applied then histogram of input image is used as clustering data.

In this paper, extensions for FCM (Fuzzy C-Mean) and EM (Expectation-maximization) are introduced. The performance of FCM-EN, FCMFG is compared to the introduced algorithms. The results obtained from the experiments are better than the previous algorithms. Future work will include making the improvement to other clustering methods and also explore the effect of different clustering algorithms in segmentation of medical images.

3.2.4 Image Segmentation with Topic Random Field

B. Zhao et al in [36] proposes image segmentation technique which overcomes the two problems including the ignorance of spatial relationship which exist among local topic labels in an image and loss of information by representing image feature using the index of its closest match in the codebook.

When an image is given to the system initial over segmentation get started by the TRF (Topic Random Field). Over segmentation is done by partitioning the image into multiple homogeneous regions. In order to avoid the selection of regions which are greater than the size of the object we start over segmentation of image by spectral clustering. Four types of features are extracted for each of the over segmented region which includes shape, color, location and texture. The location information extracted from the image regions is represented in the form of a mask of size 8x8 and also the top and bottom most pixel in the region. The texture features are basically the average responses of filter bank in each region.

This paper presents TRF model for image segmentation. The results obtained are better than LDA-style models as TRF defines Markov Random Field (MRF) over hidden topic assignment of super-pixels in an image.

3.2.5 Light Refraction based Medical Image Segmentation

U. Güvenç et al in [37] proposes segmentation method for medical images through light refraction technique. This technique is based on region growing image segmentation method.

The similarity of the pixels is measured using light refraction law. Automatic segmentation is applied onto the input medical image which takes similarity measure as input as well. The similarity percent value is then compared with the similarity threshold value to decide whether the neighboring pixel belong to the same region or not. The automatic segmentation finally gives segmented image, region number and the number of pixels in the region as output.

The computational complexity of the proposed method is less than the previous methods as it does not require keeping in knowledge the number of region the image contain. Future work will include the calculation of the similarity measure using neural network.

3.3 Shadow Removal

3.3.1 Moving Shadow Detection with Low- and Mid-Level Reasoning

A. J. Joshi et al in [38] presents a method for detecting moving shadows by applying multi-level shadow identification scheme. The system is applicable without restrictions on the number of light sources, illumination conditions, surface orientations, and object sizes.

The first is to extract the foreground by Mixture of Gaussian technique. In the second step features are selected which differs the foreground from shadow. Four parameters are calculated for background model update which include edge magnitude error, edge gradient direction error, intensity ratio, and color error. The probability of shadow of being a pixel is calculated which is based on the four parameters. In order to estimate the shadow positions connected components are identified in an image. Along with labeling of blob some other important values are also calculated like blob area, perimeter, and the number of perimeter pixels that are neighbors to pixels of a blob of another type. In order to improve the accuracy misclassified blobs are removed.

The proposed method is applicable on all common cases having variant edge, color, and intensity cues. This method is independent of all the scenes so it can be used for all practical purposes.

3.3.2 Shadow detection using gradient based Background Subtraction

Shoaib et al in [39] proposes a new scheme of detecting cast shadow of human by gradient based background subtraction. The scheme do not use color information as it is not a reliable feature for shadow detection.

Foreground of the image sequences is extracted by Mixture of Gaussians technique. The foreground and the edge boundaries are given as input to the system. The system then applies contour closure to obtained closed contour of the foreground image. The disconnected regions of the blobs are obtained after applying contour removal. After extracting blobs, boundaries of the blobs are extracted. Neighborhood ratio is calculated by which is then matched with the threshold value. Finally shadow and quick illumination changes in the image are obtained.

This shadow detection technique can be incorporated in any vision based application or any surveillance system. The experimental results shows that the proposed system classifies 90% of the pixels of the objects and their shadow correctly

3.3.3 Removal of Shadow in Moving Object Detection in Video Sequence

A. Badola in [40] proposes an effective and robust shadow removal method for moving objects in the video surveillance system. It is relevant for the surveillance application where fast tracking is required in the areas where illumination changes occur. It is crucial challenge to deal with moving shadows in object detection, traffic controlling and object tracking applications. It is especially needed in automated surveillance applications or in vehicle monitoring where accurate tracking is very important even under different lightning conditions.

First read a video sequence then select a background reference image. Background reference image can be any frame of the video sequence. In this paper first frame is selected. Each frame is then normalized according to three colors of RGB. Each incoming frame is then subtracted from the background reference image and after applying threshold shadow is detected. Morphological operation like dilation and erosion are applied and finally object is classified and image is reconstructed.

Although this method removes shadow effectively but it models shadow in RGB color model so it may not give promising results for other color models.

3.3.4 Moving Cast Shadow Detection of Vehicle using Combined Color Models

B. Sum and S. Li in [41] a novel approach for detection of moving cast shadow is proposed which uses combined color models. This technique is good for the application which includes vehicle detection and tracking.

Firstly, convert the RGB color model to HIS color model because the intensity of shadow region is lower than the region of object. HSI can better deal with such situations than other models, such as RGB, YUV. Secondly, photometric color is employed in $c1c2c3$ color model to distinguish the dark and colorful object pixels from shadow pixels. As photometric color invariants are the functions which are invariant to changes in viewing direction and illumination condition. Finally, post processing is applied to improve accuracy by correcting failed shadow and object detection.

The results obtained from experiments prove that this method outperforms the two well known techniques and detect cast shadows accurately. Future work may include the use of edge information for shadow detection and investigating of other properties of shadow to obtain higher accuracy.

3.3.5 Human Shadow Removal with Unknown Light Source

C. C. Chen and J. K. Aggarwal [42] presents a method of shadow removal which removes shadow of human cast due to unknown light source. As in real world shadows have a wide spectrum of luminance values do we use multi-cue descriptor to detect human shadows.

The multi-cue descriptor is the concatenation of three features including color, log-polar coordinate and HOG (Histogram of Gradients). Experiments are made on six color spaces and according to results the mapping of RGB to HIS gives better accuracy than the other color spaces. Human shadow is connected to the bottom of the human figure so, log-polar coordinate feature is use to find the orientation of the shadow. The reason for using orientation transformed single-cell HOG as feature is that the dominant edge of the human is vertical and dominant orientation of the shadow can be in all directions so the strong edge direction is not available from the region of shadow. The descriptors from the labeled shadow images are used to train an RBF kernel Support Vector Machines (SVM) classifier.

The classifier is trained and tested on different set of videos so according to results the detectors gives higher accuracy. The risk of classifying human pixels as shadows is reduced by the 3 stage process. The average accuracy obtained is 96.37%.

3.4 Human Detection

3.4.1 Hierarchical Part-Template Matching for Human Detection and Segmentation

Z. Lin et al in [43] describes a Bayesian based approach to detect and segment human by combining local part-based and global template-based schemes. This approach is also capable of detecting human shapes and poses.

Different type of blob based and shape based approaches are presented in past. This paper presents a hierarchical part-template matching approach for human detection and segmentation. The method decomposes the global shape model and constructs a part-template tree for different shapes of human. Through synthesis of part detections; shape segmentation and poses are computed efficiently. Hypothesis is optimized for the set of detections using Bayesian MAP framework. In order to increase the accuracy the method is then combined with background subtraction.

This technique is tested on different datasets and compared with previous techniques. According to the results the presented technique performs well on images and video sequences with severe occlusion. Future work will be combining the presented approach is combined with appearance-based segmentation to improve the result of shape segmentations.

3.4.2 Human Detection Using Partial Least Squares Analysis

W.R. Schwartz et al in [44] presents a technique which detects human by using features like edges, color and texture. By these features we get a rich descriptor set producing a high dimensionality space. In such high dimensional space SVM performs well.

A strong set of features are selected for classification which made high discrimination; instead of using a complex classifier. These features may include the high vertical edges along with human body, texture of clothing, color of face and head of human. HOG descriptor is used to get these features. PLS is then used to reduce the dimensionality of the feature space. Finally a simple and efficient classifier is used to differentiate between human and non-human. This technique is tested on a number of datasets and results shows that the technique outperforms than the previous techniques.

3.4.3 Detecting and Counting People in Surveillance Applications

X. Liu et al in [45] presents a technique for segmenting people in crowd, counting people and tracking them over a time. The application of this method is counting people entering or leaving a particular site or event detection.

The videos are fed into the foreground detector. A tracker is initialized for each object. It maintains the trajectory of each person over the time. The calibration parameters are used to separates a group of people into individuals. The trajectories obtained from the tracker are used for people counting and event detection.

The presented technique is tested over 10 minute long video sequence and the results shows that it gives good results due the use of site geometry which extract relevant scene information and overcome the problem of human detection.

3.4.4 Human Detection in Crowded Scenes

Y. L. Hou and Pang in [46] detect human in the crowded video sequences. This is for the highly crowded area such as shopping malls.

In the training phase local patches of human and their location information is given to the codebook. An interest point is detected for all the training images by scale-invariant DoG (Difference of Gaussian) detector. This detector gives scale and size of the patches. All the patches are divided into categories in order to limit the size of the codebook. The center of cluster is saved into the code book. In the testing phase the code book provides the location information of the patches of the test images. Average size of human is matched to the rectangular region for individual detection.

Future work will be the testing of technique on more complicated videos. In future the improvements will be made to the presented method.

3.4.5 Pedestrian Head Detection and Tracking Using Skeleton Graph

A new technique [47] of detecting human by detecting head and then tracking is presented in this paper. This technique uses skeleton graph for human detection in the crowded environment so this feature makes this technique different from other techniques.

The first step is background subtraction which is done to obtain foreground mask in the time T. Fourier Descriptor of outer contours is done for smoothing the silhouette. Skeleton of human body is determined after that head detection is done. In the extreme points of the skeleton degree inclination is considered. If the degree of inclination exist in the range of θ and $-\theta$ then the extreme point is detected as the head of the person. After estimating the pose of the head, it is tracked by using particle filter. Tracking is integrated in order to reduce error rate.

Experiments show that the proposed algorithm detected head reliably in the indoor/outdoor crowded environments. This technique gives promising counting results even in low or high resolution videos.

3.5 Problem Statement

Literature survey discussed above reviews previous methods of image segmentation, shadow removal, human detection and human behavior classification. Problems are identified in various techniques which are needed to be solved in future.

In the image segmentation techniques different methods are used like MoGs (Mixture of Gaussians), layered based approach, edge preserving clustering algorithms etc. The techniques based on different color models, HOG (Histogram of Gradients) descriptor, gradient based background subtraction and using multiple features are used for removal of shadow from video sequences. Human can be detection by head tracking, DoG (Difference of Gaussian) detector, PLS (Partial Least Square) and HOG (Histogram of Gradients) descriptor and hierarchical part-template matching. Though the results obtained all of the above mentioned methods are better than the previous techniques but some of the limitations exist.

A technique for human behavior classification in crowd through projection and star skeletonization is presented in [29]. It uses different sub-methods of different researchers to obtain an application for classification of human behavior. In this paper foreground segmentation is accomplished in real-time using adaptive mixture of Gaussians method [30] where the number of gaussians used per pixel is four which segment out the moving pixel of video sequences under different lightening conditions.

The projection is applied on head and ground plane to estimate the number of people in the group and to separate the individuals as mentioned in [31].

Following issues needed to be addressed while estimating pedestrian counts in groups:

1. Classifies slow-moving vehicles as group of people in certain cases
2. Do not deal with stationary people
3. When shadows appear larger than the people themselves, it classifies these shadows as human
4. The blobs used for estimating the number of people do not supply accurate estimate

In existing technique [29] human features are extracted using star skeletonization [32] which gives five extreme points of human skeleton and motion cues. These features are then classified using SVM (Support Vector Machine) classifier, where the results obtained by SVM are not good.

3.6 Summary

In this chapter different previous approaches of image segmentation, shadow removal, human behavior classification and human detection are discussed. In the image segmentation techniques different methods are used like MoGs (Mixture of Gaussians), layered based approach, edge preserving clustering algorithms etc. The techniques based on different color models, HOG (Histogram of Gradients) descriptor, gradient based background subtraction and using multiple features are used for removal of shadow from video sequences. Human can be detection by head tracking, DoG (Difference of Gaussian) detector, PLS (Partial Least Square) and HOG (Histogram of Gradients) descriptor and hierarchical part-template matching. The problems are mentioned which exist in [29] and are needed to be solved for getting a better human behavior classification application.

Chapter 4

Proposed Methodology

This research focuses on reducing the classification of shadows as human, false classification of human, and classifier problems to overcome. The proposed solution consists of different sub-parts which include segmentation, shadow removal, human detection, feature extraction and classification. The proposed research architecture is given below:

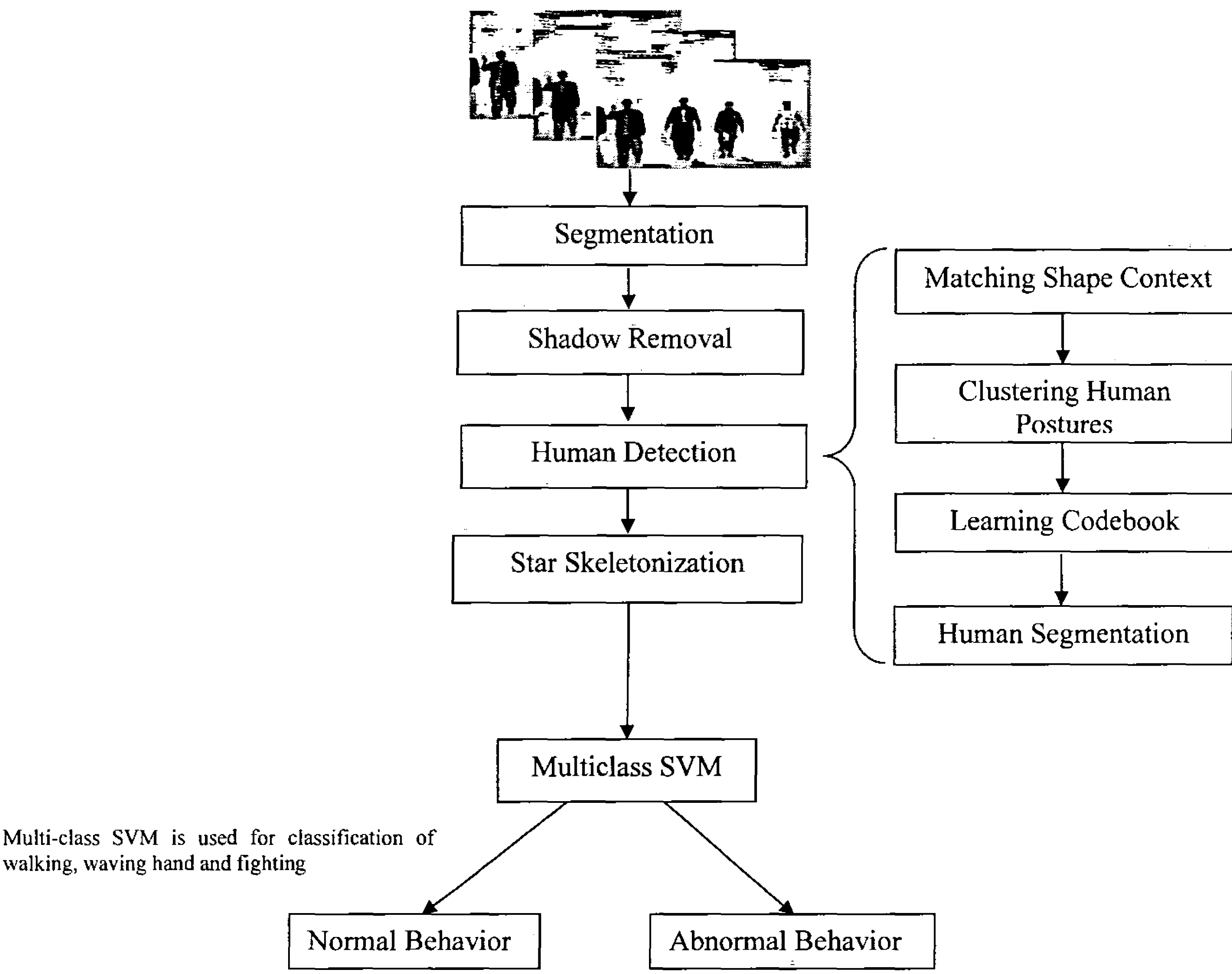


Figure 4.1: The block diagram of the proposed system

4.1 Segmentation

Segmentation is done by following the technique of mixture of gaussians. According to this technique [30] the probability of observing a pixel c at some time instant is given by:

$$P(c) = \sum_{i=1}^{n} P(M_i)P(c \mid M_i) \tag{4.1}$$

where M_i is the *i-th* mixture component, represented by a mean μ_i and a $p \times p$ covariance matrix. *Stauffer et al* in [48] used the simplifying assumption that different color channels are independent and have the same variance, thus $\sum_i = \sigma_i^2 I_{p\times p}$. It is possible to drop the assumption, as in [48], of uniform variance across channels and represent the covariance matrix as $\sum_i = diag(\sigma_{r,i}^2, \sigma_{g,i}^2, \sigma_{b,i}^2)$. Lastly, it is possible to assume nothing about the covariance matrix at all. $P(M_i)$ is the mixing coefficient corresponding to the *i-th* mixture component and $P(c \mid M_i)$ is the likelihood according to that mixture component. The individual mixture component is gaussians so the likelihood is given by:

$$P(c \mid M_i) = N_p(c, \mu_i, \textstyle\sum_i) = \frac{e^{-1/2(c-\mu_i)^T \sum_i^{-1}(c-\mu_i)}}{\sqrt{(2\pi)^p |\Sigma|}} \tag{4.2}$$

Every new pixel value c is checked against the existing K Gaussian distributions, until a match is found. A pixel c is said to match M_i if c is within δ standard deviations of μ_i. In other words, c matches M_i, if

$$\sqrt{(c-\mu_i)^T \textstyle\sum_i^{-1}(c-\mu_i)} < \delta \tag{4.3}$$

If no component matches the pixel, the last mixture component is replaced by a new distribution with initial mean equal to *c*, a small mixing component and a high variance.

All the unmatched components retain their mean and covariance matrix and their likelihood is updated as:

$$P(M_i) \leftarrow (1-\alpha)P(M_i) \tag{4.4}$$

If match is found for any of the K distribution, $P(M_k)$, μ_i and $\sum_k$ is updated as follows:

$$P(M_k) \leftarrow (1-\alpha)P(M_k) + \alpha \tag{4.5}$$

$$\mu_k = (1-\beta)\mu_k + \beta c \tag{4.6}$$

$$\sum_k \leftarrow (1-\beta)\sum_k + \beta(c-\mu_i)(c-\mu_i)^T \tag{4.7}$$

The parameters α and β are learning rate which are used as constant only because of performance concerns, but since in practice such a simplification has yielded better results. The mixture components are ordered by $P(M_k)/\sqrt{|\sum_i|}$ ratio so, that an observed pixel value c is matched against more probable components first. Once it has been determined that *c* matches a mixture component M_k, it must be determined whether M_k, is a part of the background process or not. Given a threshold $B \in [0...1]$, M_k is considered a part of the background if

$$\sum_{i=1}^{k-1} P(M_k) < B \tag{4.8}$$

4.1.1 Parameter Selection

a. Choice of number of mixture components

As the number of mixture components increases, the performance of the system decreases. The number of gaussians can be used from 3 to 5. In this paper 4 gaussians are used in order to increases the quality of the system.

b. Color Space

Different color spaces have been used with mixture of gaussians. Each color space has its own limitations such as HLS (Hue, Lightness, Saturation) is appropriate for the method but not when it is used with MPEG-4 compressed video since the saturation components are heavily quantized in decompressed video stream. RGB (Red, Green, Blue) color space gives better results when covariance matrix as diagonal is used.

c. Choice of α and β

Choosing a value of α close to 1 degenerates the method into a frame differencing approach. So, a small value of is used i-e; $\alpha = 0.001$. When using a constant value for β it is possible to set $\beta = \alpha$, thus updating the mixing coefficients and the mixture components at the same rate. But in this paper it is set to $\beta = C\alpha$, where C is a constant. In the particular examples $\beta = 1/4\alpha$ is used, which is chosen empirically.

d. Choice of B

If B is selected close to 0 a single mixture component will be considered a part of the background process and if close to 1 then multimodal background processes. Selecting a value close to 1 results in objects becoming a part of background. Select value between the two extremes depends upon nature of scene.

4.2 Shadow Removal

The extracted foreground is further tested for shadow detection [49] where the first frame of the video sequence is considered as the background image. A color / brightness difference value D is calculated as follows:

$$D = 18 \times D_{gb} + \left| \log\left(\frac{V_{BG}}{V_{curr}} \right) \right| \qquad (4.9)$$

where D_{gb} is the distance between the current pixel and the background pixel which is normalized by gb space; background brightness is represented by V_{BG} and V_{curr} is the brightness of current pixel. If D < 0.5 then the foreground is marked as lighting change pixel (that pixel can be shadow) and is discarded. Post-processing is then applied to the resultant image to remove noise and fill holes in the silhouette. The segmented image is reconstructed after marking the shadow pixels.

4.3 Human Detection

Human detection is the major step for human behavior classification. This is based on the technique of shape matching using shape context as in [50]. Firstly, the learning of human postures is done by finding the correspondence between the human contours which is done by matching shape context. Foreground is extracted from the input video sequences using edge detector. The contours are then represented as the set of points which basically contain the pixel locations of the edges. The correspondence between the two postures is found by finding the most similar shape context by mapping the sample points as in [51].

4.3.1 Matching Shape Context

The contours are the set of points $P = \{p_1, \ldots, p_n\}$ which contain set of n points. The rich local descriptor is used for finding the correspondence between the two shapes. A set of vectors originate from a point to all the other points in the contour. So a set of n-1 vectors contain the full description of the point so these vectors give the entire configuration of the shape relative to the reference point. The description based on the set of vectors is too much detailed so a histogram of relative points is computed. The shape context of a point p_i is obtained by calculating a coarse histogram of the remaining n-1 points. The bins of the histogram are uniform in log-polar space which makes the descriptor more sensitive to the position. The cost of matching the two points of two different human postures is computed as:

$$C_{ij} \equiv C(p_i, q_j) = \frac{1}{2}\sum_{k=1}^{K}\frac{\left[h_i(k) - h_j(k)\right]^2}{h_i(k) + h_j(k)} \qquad (4.10)$$

where p_i is the point of one contour and p_j is the point of another contour. $h_i(k)$ and $h_j(k)$ are the histograms of points p_i and q_j respectively. These histograms of each point are combined to form a single vector of every silhouette which is then used for matching the two human postures.

4.3.2 Clustering Human Postures

Next step is to create n clusters of the posture vectors using K-mean clustering technique. The similarity of the posture vectors is calculated as:

$$X^2 = \frac{1}{2}\sum_{k=1}^{K}\frac{[g(k)-h(k)]^2}{g(k)+h(k)} \qquad (4.11)$$

where $g(k)$ is the vector for one shape and $h(k)$ is the vector for another shape. The centers of clusters are stored as codebook entry. Each codebook entry contains two pieces of information:

i. Position of contour with respect to centroid
ii. Closest posture to which it belongs.

4.3.3 Learning Codebook

Learning of different postures is done by iterating through all foreground blobs, creating the posture vectors which are then matched against code book entries. If match found corresponding codebook entries cast votes for possible centroid of human in the scene and posture to which it belongs. Mean-Shift is used for finding maximums. Codebook activations cast votes for the possible locations and posture of human h_n in the scene. Each vote is then weighted by the distance of similarity found by comparing local posture and codebook instance C_i. This can be expressed as:

$$p(h_n, x \mid s, l) = \sum_i p(h_n, x \mid C_i, l)\, p(C_i \mid s) \qquad (4.12)$$

4.3.4 Segmenting human

A set of hypothesis is calculated for each search window W(x) using Mean-Shift. So, within a video frame contour is initialized with the posture h_n for which hypothesis is strong at location x.

$$p(h_i) = \sum_{x_j \in W(x)} p(h_n, x_j \mid s_k, l_k) \tag{4.13}$$

For segmentation of human from the video frames color and texture features are used as proposed in [52].

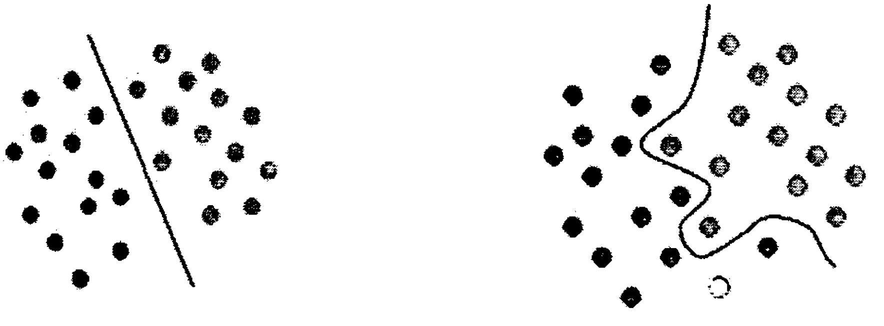
4.4 Feature Extraction

A large set of features can be used for analyzing human motion and classifying behavior on the basis of human motion. This research uses star skeleton for analyzing the behavior of human in the video streams as proposed in [53]. Boundaries of the contours are extracted which are then used for getting skeleton points. Extreme points are extracted from the boundaries and considered as skeleton points. Following steps are followed for creating human skeleton from silhouettes:

1. From the boundary of each contour centroids are evaluated
2. The distance of the centroid from each boundary pixel of each contour is calculated which gives a one dimensional discrete signal
3. The signal is then smoothed using low pass filter for noise reduction
4. Local maxima is taken as extreme points which is detected by finding zero-crossing difference function

4.5 Classification

The features obtained from star skeletonization are fed into SVM for classification. **Support Vector Machine** is a powerful tool for binary classification. Support Vector Machines are based on the idea of constructing decision boundaries which separating objects of different classes. Depending on the separating objects decision boundary can be linear or any complex structure. An example of separating red and green objects is presented in figure 4.2(a) in which objects are separated by linear classifier. Often most complex structures are needed for optimal separation to correctly classify test samples on the basis of learning that is done using train samples. Separating such hyperplanes is called as non-linear classification; an example of such type of classification is shown in figure 4.2 (b).



(a) Linear Classification **(b) Non-linear Classification**

Figure 4.2 Types of Classification [52]

SVM selects the best hyperplane out of many hyperplanes by calculating the distance of the nearest data point and the hyperplane. A hyperplane is selected for which the distance of data points from the hyperplane is maximum than the other hyperplanes. In order to deal with bulk of data and ignoring noise SVM introduces the concept of *soft margin.* If the training set is not linearly separable than fat decision margin is constructed to reduce error. *Kernel function* is a mathematical trick that allows SVM to perform two dimensional classifications on one dimensional feature space. Different kernel functions can be represented by the following examples [53]:

Linear Kernel: $k(x,y)=x^T y+c$ (4.14)

Polynomial Kernel: $k(x,y)=(\alpha x^T y+c)^d$ (4.15)

Gaussian Kernel: $$k(x,y)=\exp\left(\frac{\|x-y\|^2}{2\sigma^2}\right) \qquad (4.16)$$

Radial Basis Function: $$k(x,y)=\exp\left(-\gamma|x-y|^2\right) \qquad (4.17)$$

By using some non linear mapping SVM maps the input vector to high dimensional feature space. One should choose the kernel function that can separate the hyper plane with *maximum margins* and rather low dimensions. The hyperplanes can be modeled by the following equation [29]:

$$\omega_{st} X_i^T + b_{st} = 0 \qquad (4.18)$$

Where X_i is the ith row of matrix X and s and t are the two partitions separated by the hyperplanes ω_{st} is the weight factor and b_{st} is the bias term. So the classification function can be represented as follows:

$$f_{st}(X_i^T) = \omega_{st} X_i^T + b_{st} = 0 \qquad (4.19)$$

Multi-class Support Vector Machine is a mainly classifier that separates cases into different class labels and performs classification by constructing hyperplanes in multi-dimensional space. There are three approaches by which SVM can perform classification on three or more classes [54].

i. **Multi-class Ranking SVMs**

   All classes are classified using one SVM function.

ii. **One-against-all Classification**

   One binary SVM works for each class which separates members of one class from members of other classes.

iii. **Pairwise Classification**

   One binary SVM works for each pair of classes which separates members of one class from members of other classes.

In this research, classification is done by following *one-against-all classification* and *Radial Basis Function (RBF)* as kernel is used. Many kernel functions can be used for mapping test samples to the train samples but some of them can work well for large dataset. RBF Kernel is default and recommended kernel because it gives finite responses across the entire range of samples.

4.6 Pseudocode

Segmentation:

1. For each image of the video sequence follow step 2 to 5
2. For each pixel of the image and for each gaussian calculate equation 4.3
   if equation holds
      update mean variance and likelihood
   else
      update likelihood
3. Calculate mixture component rank by $P(M_k)/\sqrt{|\Sigma_i|}$ ratio
4. Sort rank values
5. Calculate equation 4.8
   if equation holds
      consider pixel as background
   else
      consider pixel as foreground

Shadow Removal:

1. For each segmented image follow steps 2 to 5
2. Consider first image as background image
3. Calculate the parameters of equation 4.9
4. Apply threshold to pixels of the image D< 0.5 to mark shadow pixels
5. Reconstruct the segmented image, remove noise and fill holes from the image

Human Detection:

1. For each shadowless segmented image follow steps 2 to 7
2. Apply background subtraction and edge detection to the images
3. Apply matching shape context and calculate similarity measure using equation 4.10

4. Cluster the human postures using K-mean clustering
5. Calculate similarity of posture vectors using equation 4.11
6. Create code book
7. Match foreground blobs against code book entries and cast votes for possible locations of human
8. Calculate maximum vote of possible human location using mean-shift
9. A set of hypothesis is calculated using equation 4.13. Select the location of human for which hypothesis is strong

Feature Extraction:

1. For each frame obtained from human detection follow steps 2 to 5
2. From the boundary of each contour within a frame centroids are evaluated
3. The distance of the centroid from each boundary pixel of each contour is calculated which gives a one dimensional discrete signal
4. The signal is then smoothed using low pass filter for noise reduction
5. Local maxima is taken as extreme points which is detected by finding zero-crossing difference function

Classification:

1. Load training file
2. Load testing file
3. Set parameters of SVM
4. Train SVM using RBF kernel
5. Classify using RBF kernel

4.7 Summary

In this chapter importance of different techniques are presented which are used in the existing technique [29]. The presented human detection and shadow removal techniques will reduce the false positives and may help in increasing the accuracy of the proposed approach. Binary and multi-class SVM classifier is discussed and specification of kernels and classification method are presented. SVM classifier is applied in various areas of computer vision like face recognition, human detection, finger print recognition etc.

Chapter 5

Experimental Results

This chapter discusses the quantitative measures that are used for evaluation of this research, the dataset used, different scenarios of dataset, parameter settings and the overall results obtained of classification. The results obtained from different set of modules like segmentation, human detection, feature extraction etc are also discussed in this chapter. The comparison of the results obtained from the existing technique [29] and the proposed solution on the basis of quantitative measures is also the part of discussion.

5.1 Quantitative Measure

When a classifier or model is applied to a dataset, the developer determines whether the results obtained are accurate or not with respect to the train data. Different measures can be obtained from the classified results like accuracy, precision, recall etc. Such measures are called as *Performance Measures*. These measures uses following terms:

- True positive = correctly identified actions
- False positive = incorrectly identified actions
- True negative = correctly rejected actions
- False negative = incorrectly rejected actions

Some of the performance measures that are evaluated for this research work are discussed below:

5.1.1 Accuracy

The overall correctness of the system is called as accuracy. It is calculated as ratio of sum of correct classification results and total number of classifications. Following formula can be used for calculating accuracy

$$Accuracy = \frac{No.of\ true\ positives + No.of\ true\ negatives}{No.of\ true\ positives + true\ negatives + false\ positives + false\ negatives} \quad (5.1)$$

5.1.2 Precision

The degree to which repeated measurements under unchanged conditions show the same results is called as *precision.* It is measure of correct classification among all positive classifications. Precision is the ratio of true positives and all positives.

$$Precision = \frac{number\ of\ true\ positives}{number\ of\ true\ positives+false\ positives} \qquad (5.2)$$

5.1.3 Negative Prediction

The measure of correct negative classifications against all negative classification is called as *negative prediction.*

$$Negative\ Predication = \frac{number\ of\ true\ negatives}{number\ of\ true\ negatives+false\ negatives} \qquad (5.3)$$

5.1.4 Recall

The fraction of relevant instances from the retrieved results is called as *recall.* It is also called as sensitivity and corresponds to the true positive rate. It is the ratio of true positives and false negatives.

$$Recall = \frac{number\ of\ true\ positives}{number\ of\ true\ positives+false\ negatives} \qquad (5.4)$$

5.1.5 F-measure

The performance of the test can be calculated using F-measure. It can be calculated by taking harmonic mean of precision and recall.

$$F = 2 \times \frac{Precision \times Recall}{Precision+Recall} \qquad (5.5)$$

5.2 Experimental Setup

The experimental setup includes the scenarios used in the dataset and the parameter setting for different modules of the proposed system like segmentation, shadow removal etc.

5.2.1 Dataset

The proposed method will be experimented on self-created video sequences and different available datasets to classify walking, waving hand and bending of people in crowd. Algorithms are tested by implementing in MATLAB R2008a on machine with Pentium IV 1.8 GHZ processor. Following datasets are used other than self-created dataset:

i. Weizmann (http://www.wisdom.weizmann.ac.il) dataset for the action of an individual walking
ii. CAVIAR (http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/) datasets for the action of an individual walking

Video sequences taken from above mentioned datasets of different scenarios will be considered while experimentation. Following are the scenarios:

Scenario 1: One person walking in corridor

Scenario 2: Two persons walking in corridor

Scenario 3: Three persons walking and waving hand in corridor

Scenario 4: One person walking in outdoor scene and shadow appears large

Scenario 5: One person walking with dog in outdoor scene

Scenario 6: Four person walking and waving hand in corridor

Scenario 7: Two persons fighting in corridor

5.2.2 Parameter Settings

Different settings of parameter are followed in each module. In segmentation parameter settings used for Gaussian mixture model are listed below:

Table 5.1: Parameter values used in segmentation

Parameters	Symbols	Values
Number of Gaussians	N	4
Learning Rate	α	10^{-3}
Learning Rate	β	$\alpha / 4$
Match Threshold	δ	2.5
Background Proportion	B	0.25
Initial weight		α
Initial variance	320.0	320.0
Low variance threshold		49.0

The parameter settings used in while classification using SVM classifier is given below:

Table 5.2: Parameter values used in SVM classifier

Parameters	Values
Kernel	Gaussian RBF Kernel
Type	Classification
Regularization Parameter (determines tradeoff between train error and smoothness)	20
Squared Bandwidth	0.4

5.3 Results and Discussion

Results of different modules like segmentation, shadow removal, human detection, feature extraction and classification for each scenario are presented in Figure 5.1, 5.2, 5.3 and 5.4.

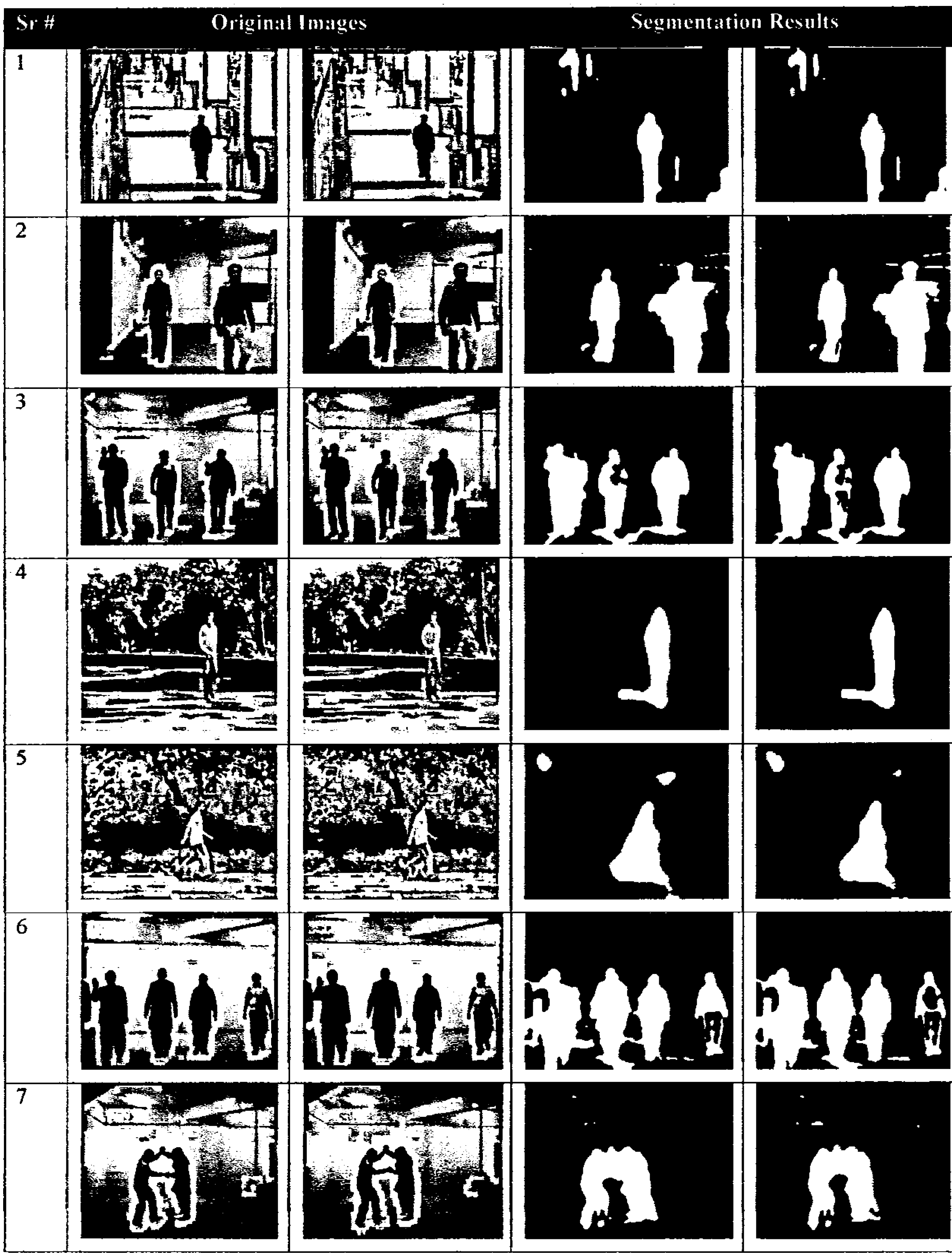


Figure 5.1 Results of segmentation for consecutive frames of each scenario

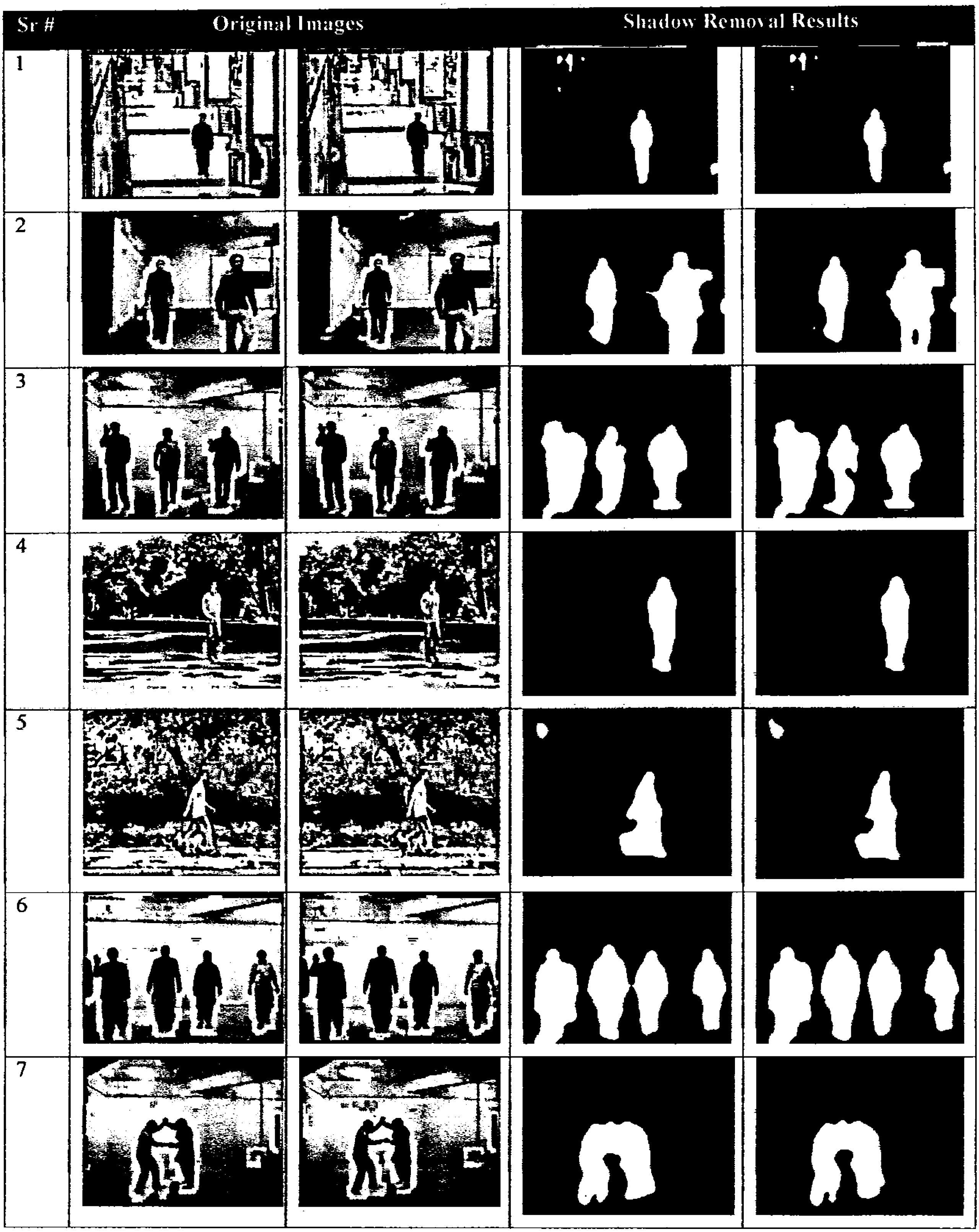


Figure 5.2 Results of shadow removal for consecutive frames of each scenario

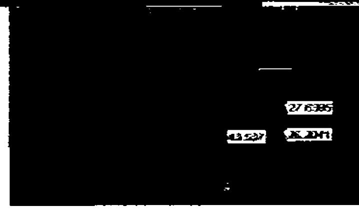
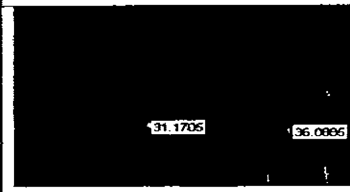
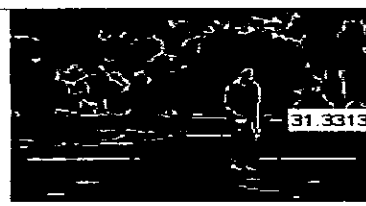
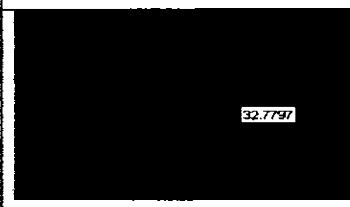
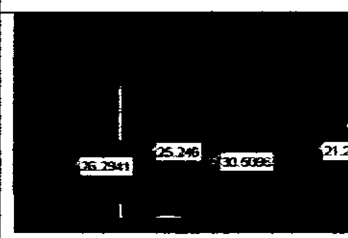
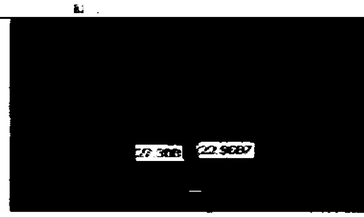
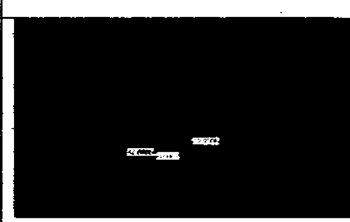
Sr #	Original Images	Technique [51] Results	Proposed Solution Results
1			
2			
3			
4			
5			
6			
7			

Figure 5.3 Comparative results of human detection of each scenario

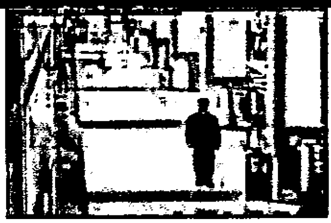
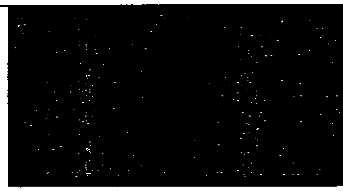
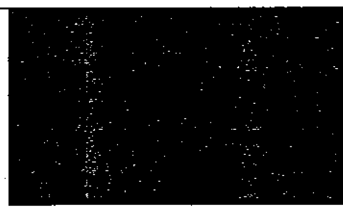
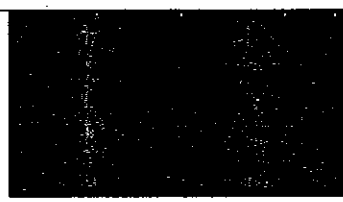
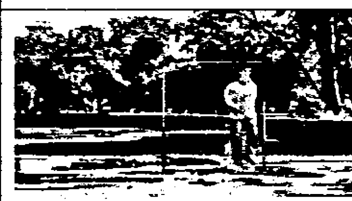
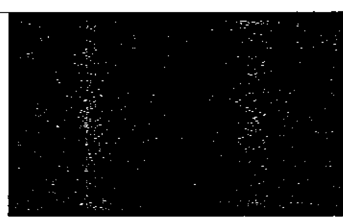
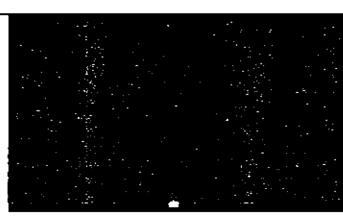
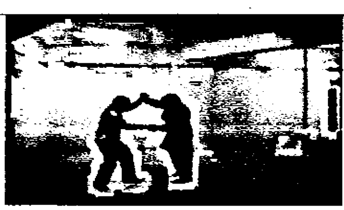
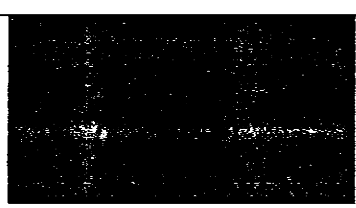
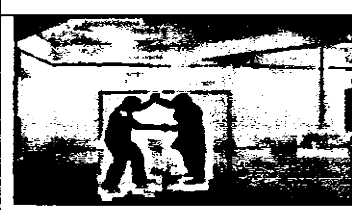
Sr #	Original Images	Feature Extraction Results	Classified Results
1			
2			
3			
4			
5			
6			
7			

Figure 5.4 Results of feature extraction and classification of each scenario, normal behaviors detected are shown by green color and abnormal by red color

The results of segmentation of two consecutive frames for each scenario are shown in figure 5.1. In the scenario 1 either there is different lightning condition in the scene but the results of segmentation are much better as mixture of gaussian deals with such conditions. In the scenario 4 the shadow of person appears large in outdoor scene and which is little bit handled by segmentation and will be totally removed after applying shadow removal technique. In scenario 5 there is flickering of leaves in the background which is well handled by mixture of gaussian as it eliminates misclassification due to cyclic motion in the background.

After applying shadow removal the results get much simplified and clear as the shadow is removed and preprocessing step made it much clear. Comparing the results of segmentation and shadow removal of scenario 1 and 4 it becomes clear that shadow removal is the required pre-processing step for overcoming the problem of existing technique [29] which says that when shadow of human gets larger than its size the shadow is also classified as human.

Existing technique [29] uses projection technique for classification of human which have many problems that are discussed in section 3.5. The technique for human detection presented in [51] have also some problems as it classifies non-human objects as human which can be removed by giving pre-processed image means image after segmentation and shadow removal. If actual image is passed and edge detection is applied and passed on to the system than it considers non-human objects as human. The quantitative measures for each scenario is evaluated using existing technique [51] and proposed technique given in table 5.3 and 5.4.

Table 5.3: Quantitative measures for each scenario using human detection technique [51]

	Precision	Recall	F-measure	Accuracy (in %)
Scenario 1	0.2755	1	0.2159	100
Scenario 2	1	0.7250	0.4202	72.50
Scenario 3	1	1	1	100
Scenario 4	0.0161	0.0333	0.0108	3.33
Scenario 5	0	0	0	0
Scenario 6	1	1	1	100
Scenario 7	0.6071	0.85	0.3541	85

Table 5.4: Quantitative measures for each scenario using proposed technique of human detection

	Precision	Recall	F-measure	Accuracy (in %)
Scenario 1	1	1	1	100
Scenario 2	0.95	1	0.4871	95
Scenario 3	1	1	1	100
Scenario 4	1	1	1	100
Scenario 5	1	1	1	100
Scenario 6	1	1	1	100
Scenario 7	1	1	1	100

The results of human detection are shown in figure 5.3 which contains original images, results of [51] and the results of proposed technique of each scenario. In scenario 1 [51] detected three humans in the scene whereas in original image one human exist which is accurately detection by the proposed solution. In the scenario 4 though the number of human detected by [51] is correct but the centroid of the person is false. In scenario 5 of outdoor scene two humans are detected by [51] whereas proposed method has correctly detected one person. So, from the results it becomes clear that the proposed methodology outperforms than existing technique [31] and the technique in [51] which makes the results more accurate.

The results of feature extraction and classification are shown in figure 5.4 for each scenario. From the results of feature extraction we come up with the conclusion that the extreme points are either not correctly identified or less number of extreme points are detected so because of this the accuracy of system is effected due to the false classification of human behaviors. So, the method of feature extraction is needed to be changed for the proposed system in order to reduce false classification and increasing accuracy of the system.

Different kernels of SVM classifiers are tested and the result shows that RBF kernel gives results consistently whereas linear kernel gives abrupt changes on each scenario. For normal behaviors the accuracy achieved by linear kernel is 100% in two scenarios but RBF kernel out-performs in some scenarios than linear kernel. For abnormal behavior RBF kernel out-performs as compare to linear kernel as given for scenario 7 shown in Figure 5.5. Table 5.5 and 5.6 shows different quantitative measures for each scenario on the basis of linear and RBF Kernel of SVM classifier. A comparison of both kernels on the basis of accuracy is shown in Table 5.7 and Figure 5.5.

Table 5.5: Performance Evaluation of each scenario using SVM classifier (RBF Kernel)

	Accuracy (% age)	Precision	Recall	F-measure	Training Time (in sec)
Scenario 1	85.19	1	0.8519	0.9200	4.22 e-01
Scenario 2	98.80	1	0.9880	0.9939	4.38 e-01
Scenario 3	96.53	1	0.9653	0.9823	4.22 e-01
Scenario 4	90.00	1	0.9000	0.9474	4.22 e-01
Scenario 5	94.23	1	0.9423	0.9703	4.22 e-01
Scenario 6	94.57	1	0.9457	0.9721	4.38 e-01
Scenario 7	93.10	0.9310	1	0.9643	4.22 e-01

Table 5.6: Performance Evaluation of each scenario using SVM classifier (Linear Kernel)

	Accuracy (% age)	Precision	Recall	F-measure	Training Time (in sec)
Scenario 1	100	1	1	1	3.59 e-01
Scenario 2	82.53	1	0.8253	0.9043	3.59 e-01
Scenario 3	89.58	1	0.8958	0.9451	3.28 e-01
Scenario 4	98.33	1	0.9833	0.9916	3.59 e-01
Scenario 5	82.69	1	0.8269	0.9053	3.59 e-01
Scenario 6	100	1	1	1	4.38 e-01
Scenario 7	43.10	0.4310	1	0.6024	3.13 e-01

Table 5.7: Comparison of accuracy for each scenario using different SVM kernels

	Accuracy using Linear Kernel (% age)	Accuracy using RBF Kernel (% age)
Scenario 1	100	85.19
Scenario 2	82.53	98.80
Scenario 3	89.58	96.53
Scenario 4	98.33	90.00
Scenario 5	82.69	94.23
Scenario 6	100	94.57
Scenario 7	43.10	93.10

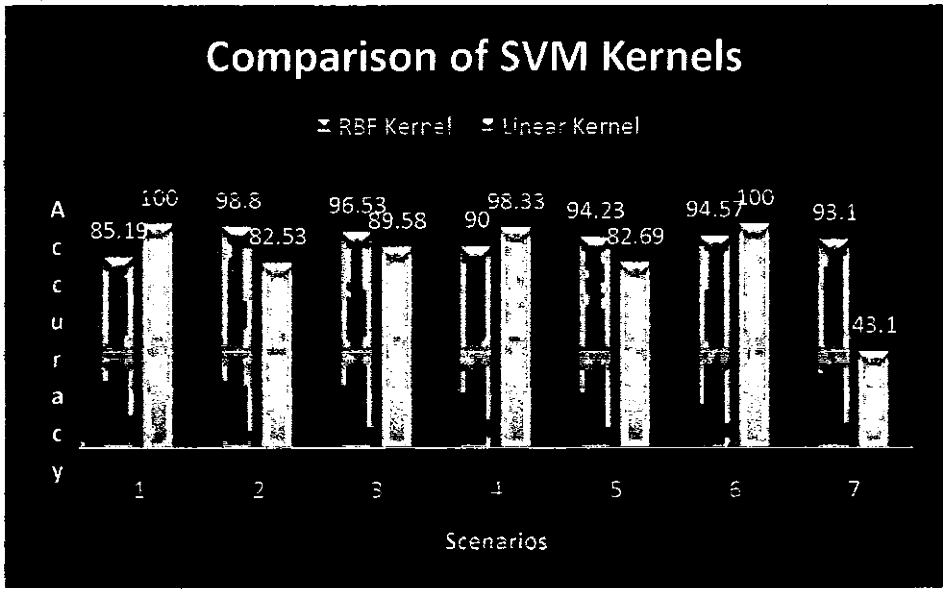


Figure 5.5 Comparison of RBF and Linear Kernel of SVM (in percentage) on the basis of accuracy for different scenarios

As the comparison of results shows that RBF kernel out-performs the Linear SVM Kernel in most of the scenarios. So, RBF kernel will be used for further performance evaluation. The overall performance of the system is given in Table 5.8 for normal and abnormal behaviors on the basis of accuracy, precision, recall, f-measure and training time.

Table 5.8: Overall performance evaluation of proposed system using SVM classifier

	Accuracy (% age)	Precision	Recall	F-measure	Training Time (in sec)
Normal	93.22	1	0.9322	0.9643	4.27 e-01
Abnormal	93.10	0.9310	1	0.9643	4.22 e-01

Multi-class SVM is also applied on the feature set using one-vs-all strategy and the results obtained after applying multi-class SVM are much better than SVM for normal behaviors whereas for abnormal behaviors SVM and multi-class SVM give same results. The details of performance evaluation using multi-class SVM classifier for each scenario are given in Table 5.9 and the overall performance for normal and abnormal behaviors using multi-class SVM classifier is given in Table 5.11. The comparative graph of SVM and multi-class SVM classification on the basis of accuracy is given in Figure 5.6.

Table 5.9: Performance Evaluation of each scenario using Multi-class SVM classifier

	Accuracy (% age)	Precision	Recall	F-measure	Training Time (in sec)
Scenario 1	100	1	1	1	1.14 e+02
Scenario 2	98.80	1	0.9880	0.9939	8.55 e+01
Scenario 3	94.44	1	0.9444	0.9714	1.18 e+02
Scenario 4	100	1	1	1	1.46 e+02
Scenario 5	94.23	1	0.9423	0.9703	8.03 e+01
Scenario 6	100	1	1	1	7.37 e+01
Scenario 7	94.83	0.9483	1	0.9735	2.14 e+02

Table 5.10: Performance Evaluation of each behavior using Multi-class SVM

	Accuracy (% age)	Precision	Recall	F-measure	Training Time (in sec)
Walking	98.2	1	0.9825	0.9910	4.79 e+02
Waving	97.22	1	0.9722	0.9857	4.27 e+01
Fighting	94.83	0.9483	1	0.9735	2.14 e+02

Table 5.11: Overall performance evaluation of proposed system using Multi-class SVM classifier

	Accuracy (%age)	Precision	Recall	F-measure	Training Time (in sec)
Normal	97.91	1	0.9791	0.9892	1.03 e+02
Abnormal	94.83	0.9483	1	0.9735	2.14 e+02

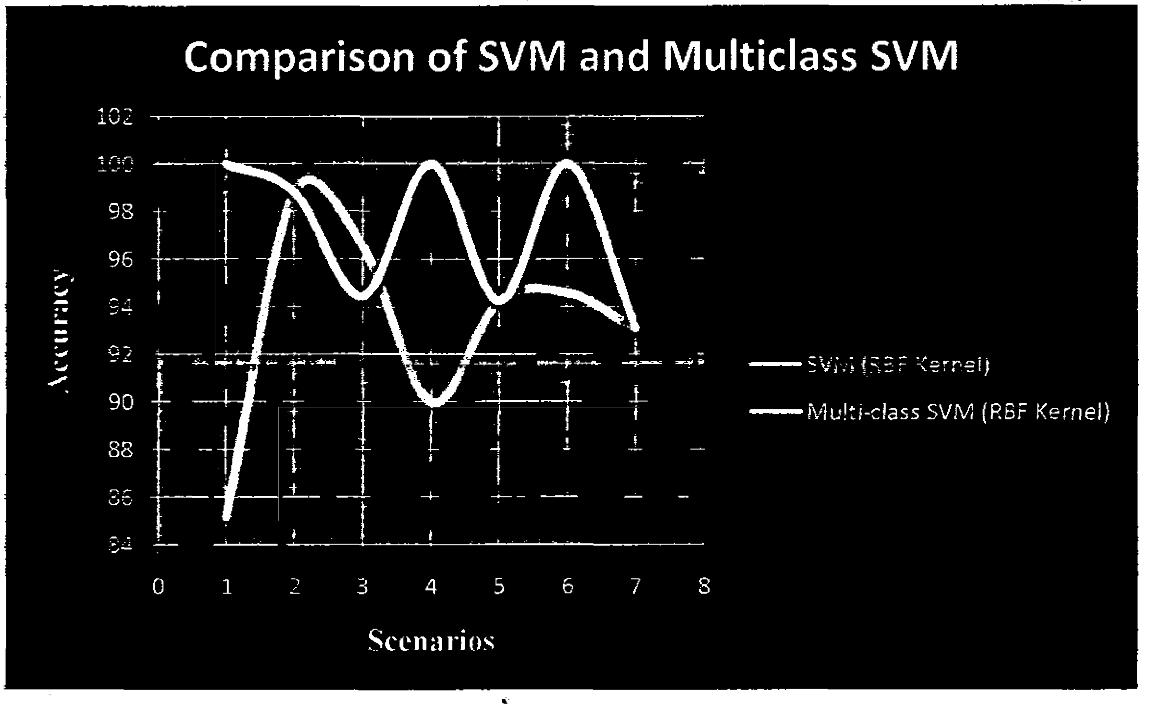


Figure 5.6 Comparison of SVM and Multi-class SVM (in percentage) on the basis of accuracy for different scenarios

Table 5.12 presents the comparison of proposed technique with existing technique [29] on the basis of accuracy of the system. The results shows that proposed system out-performs than the existing technique [29] for normal behaviors and abnormal behaviors when SVM classifier is applied using RBF kernel whereas when multi-class SVM is applied the proposed system out-performs than existing technique [29] for normal behaviors but for abnormal behavior [29] performs better with minor difference than proposed system. The reason for this difference is that the proposed system is trained on less dataset for abnormal behavior and secondly the abnormal behavior of proposed system is complex than the existing technique [29]. The existing technique [29] is tested on abnormal running and the proposed system is tested on fighting of two persons.

Table 5.12: Comparison of existing technique [29] and proposed technique based on accuracy (in percentage)

	Existing Technique [29] using SVM	Proposed Technique using SVM	Existing Technique [29] using Multi-class SVM	Proposed Technique using Multi-class SVM
Normal	89.85	93.22	96.45	97.91
Abnormal	82.50	93.10	96.70	94.83

5.4 Summary

In this chapter the complete experimental setup is presented including dataset details, different scenarios of video sequences and parameter settings. The results obtained after experimentation shows that proposed system give more accuracy than the existing technique [29] and proposed system give less false predictions. Due to using shadow removal technique and better human detection method the results of the proposed system are better than the existing technique [29]. This work can be further extended for classification of more human behaviors other than walking, bending and waving hand.

Chapter 6

Conclusions and Future Work

This chapter concludes the whole report on the basis of problems in the existing technique, proposed solution and results obtained. Future work that can be done using the proposed technique is also summarized. The importance of proposed technique in the area of computer vision is also discussed.

6.1 Conclusions

In this research work an enhanced strategy of human behavior classification is presented which outperforms the previous methods. We have validated this approach on the basis of quantitative measures giving higher accuracy on most on the videos which is not achieved in the existing technique [29]. The proposed methodology comprised on five modules including segmentation, shadow removal, human detection, feature extraction and classification whereas in the existing technique shadow removal is not included and human detection has many limitations because of which there was high rate of misclassification or false detection. The presented human detection technique is robust as applied in both indoor and outdoor scenes and give good results.

6.2 Future Work

The proposed solution can be further enhanced to be useful in many other scenarios like falling down, abnormal running and many other abnormal events that may based on such behaviors like target killing detection, snatch theft detection and suspicious activities detection. The future work that can be applied in the methodology may involve the use of HMM (Hidden Markov Model) for classification of events that are mentioned above and updating the system with more strong feature extraction method which reduce false classification of human behaviors and increasing the accuracy of the system.

6.3 Applications

There are many applications of the proposed technique which can be achieved by changing the flow and by introducing some of the other techniques.

6.3.1 Human detection and Tracking

The proposed solution can be applied to human detection and tracking systems. By following the steps of segmentation, shadow removal, human detection of the proposed technique and finally applying tracking to the detected human in the video. So, we finally get the trajectories of every individual in the video sequence which can be analyzed for identifying human motion.

6.3.2 Distributed Surveillance System

This area of surveillance system comprises of cameras fixed at different locations within a building or public places. The video sequences obtained from different cameras can also processed at the same time for analyzing the behavior of human that can be either normal or abnormal. The proposed system deals with less number of actions which can be further extended for adding many other actions like running, pulling, jumping etc.

6.3.3 Crowd Counting

Crowd counting applications can be applied on public as well as in private areas. This can done using proposed technique by applying segmentation, shadow removal, human detection and finally apply counting of human in each frame of the video sequence. Using this method we can also analyze the duration of presence of a human in the scene.

6.4 Summary

This chapter summarizes the overall discussed of the presented research work. It also explores the application areas of this research work and discussion related to the enhancement of work in future is also presented in it. Finally, this approach can be further extended for many other abnormal event detection that are based on individuals in the crowd not on the overall flow of the crowd.

References

[1] Wikipedia, http://en.wikipedia.org/wiki/Surveillance

[2] Wikipedia, http://en.wikipedia.org/wiki/Computer_vision

[3] M. Becha Ka ′ aniche, "Human Gesture Recognition", Ph.D. dissertation, Univ. of Nice Sophia Antipolis, Nov 2009

[4] Wikipedia, http://en.wikipedia.org/wiki/Segmentation_(image_processing)

[5] Generation5, http://www.generation5.org/content/2003/segmentation.asp

[6] George Mason University, Department of Computer Science, http://cs.gmu.edu/cne/modules/dau/stat/clustgalgs/clust5_bdy.html

[7] R. Adams and L. Bischof, "Seeded region growing", IEEE Trans. Pattern Analysis Machine Intelligence, pp. 641–647, Jun 1994.

[8] H. Rai, T.R.Gopalakrishnan Nair, "Gradient Based Seeded Region Grow method for CT Angiographic Image Segmentation", pp.1-6, Jan 2009.

[9] J. Fan, Guihua Zeng , M. Body, M.S. Hacid, "Seeded region growing: an extensive and comparative study", Pattern Recognition Letters, pp. 1139–1156, June 2005

[10] EE Times Design Article, http://www.eetimes.com/design/militaryaerospacedesign/4017685/Background-subtraction-part-1-MATLAB-models

[11] D. Miyazaki, Y. Matsushita, and K. Ikeuchi, "Interactive shadow removal from a single image using hierarchical graph cut" 9th Asian Conference on Computer Vision, pp. 234-245, 2009

[12] F. Liu and M. Gleicher, "Texture-Consistent Shadow Removal", 10th European Conf. on Computer Vision, pp. 437–450, 2008

[13] L. Xu, J.L. Landabaso, M. Pardàs, "Shadow Removal with Blob-Based Morphological Reconstruction for Error Correction" IEEE International Conf. on Acoustics, Speech and Signal Processing, pp. 729- 732, March 2005

[14] http://painting.about.com/cs/paintingknowhow/a/shadows.htm

[15] A. Prati, I. Miki, R. Cucchiara, M. Trivedi, "Comparative Evaluation of Moving Shadow Detection Algorithms", IEEE CVPR workshop on Empirical Evaluation Methods in Computer Vision, pp. 1-8, 2001

[16] J. Zhou and J. Hoang, "Real Time Robust Human Detection and Tracking System", IEEE conference on Computer Vision and Pattern Recognition, pp. 149- 149, June 2005

[17] Q. Zhu, Shai Avidan, M. C. Yeh and K. T. Cheng, "Fast Human Detection Using a Cascade of Histograms of Oriented Gradients", IEEE conference on Computer Vision and Pattern Recognition, pp. 1491- 1498, 2006

[18] E. L. Andrade, R. B. Fisher, and S. Blunsden, "Detection of emergency events in crowded scenes," in Proc. Inst. Eng. Technol. Conf. Crime Security, pp. 528–533, 2006

[19] Hsieh, Hsu, "A simple and fast surveillance system for human tracking and behavior analysis", IEEE Conference on Signal-Image Technologies and Internet-Based System, pp. 812-828, Dec. 2007

[20] R. Cucchiara, C. Grana, M.Piccardi, A. Prati, and S. Sirotti , "Improving Shadow Suppression in Moving Object Detection with HSV Color Information", IEEE Intelligent Transportation System Conference, pp.334-339, Aug. 2001

[21] D. Chen, M.Y. Chen, C. Gao, A. Hauptmann, H. Wactlar and A. Bharucha, "Recognition of Aggressive Human Behavior Using Binary Local Motion Descriptors" , 30th Annual International IEEE Engineering in Medicine and Biology (EMBC) Conference, pp. 20-24, Aug. 2008

[22] N. Kiryati, T. R. Raviv, Y. Ivanchenko, S. Rochel, "Real-time Abnormal Motion Detection in Surveillance Video", 19th International Conference on Pattern Recognition (ICPR 2008), pp. 8-11, Dec. 2008

[23] Y. Wang, G. Mori, "Human Action Recognition by Semi-Latent Topic Models" IEEE transactions on pattern analysis and machine intelligence, pp. 1762-1774, Oct. 2009

[24] B. Yogameena, S. Veeralakshmi, E. Komagal, S. Raju, and V. Abhaikumar, "RVM-Based Human Action Classification in Crowd through Projection and Star Skeletonization", Hindawi Publishing Corporation EURASIP Journal on Image and Video Processing, Jan. 2009

[25] C. G´arate, P. Bilinski, F. Bremond, "Crowd Event Recognition Using HOG Tracker", 12th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, 2009

[26] A. Zweng, M. Kampel, "Unexpected Human Behavior Recognition in Image Sequences using Multiple Features", International Conference on Pattern Recognition, 2010

[27] C. H. Lin, F. S. Hsu, and W. Y. Lin, "Recognizing Human Actions Using NWFE-Based Histogram Vectors", Hindawi Publishing Corporation, EURASIP Journal on Advances in Signal Processing, 2010

[28] J. Zhou and J. Hoang, "Real time robust human detection and tracking system," IEEE Conference on Computer Vision and Pattern Recognition Workshop, pp. 149–149, 2005.

[29] B. Yogameena, E. Komagal, M. Archana and S. Raju Abhaikumar, "Support Vector Machine-Based Human Behavior Classification in Crowd through Projection and Star Skeletonization", Journal of Computer Science, pp. 1008-1013, 2010

[30] S. Atev, O.Masoud, and N. Papanikolopoulos, "Practical mixtures of gaussians with brightness monitoring," IEEE Conference on Intelligent Transportation Systems, pp. 423–428, Oct. 2004.

[31] P. Kilambi, E. Ribnick, A. J. Joshi, O. Masoud, and N. Papanikolopoulos, "Estimating pedestrian counts in groups", Computer Vision and Image Understanding, pp.43–59, 2008.

[32] H. Fujiyoshi, A. J. Lipton, and T. Kanade, "Real-time human motion analysis by image skeletonization", IEICE Transactions on Information and Systems, pp. 113–120, 2004.

[33] J. Yao, J. M. Odobez, "Multi-Layer Background Subtraction Based on Color and Texture", IEEE the CVPR visual surveillance workshop, minneapolis, 2007

[34] K.A. Patwardhan, G. Sapiro, V. Morellas, "Robust foreground detection in video using pixel layers", IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 746–751, 2008

[35] M.A. Balafar, A. R. Ramli and S. Mashohor, "Edge-preserving Clustering Algorithms and Their Application for MRI Image Segmentation", International Multi-Conference of Engineers and Computer Scientists, 2010

[36] B. Zhao, L.F. Fei, and E. P. Xing, "Image Segmentation with Topic Random Field", ECCV, pp. 785–798, 2010

[37] U. Güvenç, R. Demirci and T. Karagül, "Light refraction based medical image segmentation", Scientific Research and Essays, pp. 1127-1132, May 2010

[38] Joshi, Atev, Masoud, Papanikolopoulos, "Moving Shadow Detection with Low- and Mid-Level Reasoning", IEEE International Conference of Robotics and Automation, 2007

[39] Shoaib, M. Dragon, R. Ostermann, "Shadow detection for moving humans using gradient-based background subtraction", IEEE International Conference on Acoustics, Speech and Signal Processing, 2009

[40] A. Badola, "Removal of shadow in moving object detection in video sequence", 2005

[41] B.Sun; S. Li, "Moving Cast Shadow Detection of Vehicle Using Combined Color Models", Chinese Conference of Pattern Recognition (CCPR), 2010

[42] C. C. Chen and Aggarwal, "Human Shadow Removal with Unknown Light Source", 20th International Conference of Pattern Recognition (ICPR), 2010

[43] Z. Lin, Davis, Doermann and D eMenthon, "Hierarchical Part-Template Matching for Human Detection and Segmentation", IEEE 11th International Conference of Computer Vision, 2007

[44] W.R. Schwartz, A. Kembhavi, D. Harwood, L. S. Davis, "Human detection using partial least squares analysis", IEEE 12th International Conference of Computer Vision, 2009

[45] X. Liu, P. H. Tu, J. Rittscher, A. Perera and N. Krahnstoever, "Detecting and counting people in surveillance applications", IEEE Conference of Advanced Video and Signal Based Surveillance, 2005

[46] Y. L. Hou, Pang, "Human detection in crowded scenes", 17th IEEE International Conference of Image Processing (ICIP), 2010

[47] K.E. Aziz, D. Merad, B. Fertil, N. Thome, "Pedestrian head detection and tracking using skeleton graph for people counting in crowded environments", 2011

[48] C. Stauffer, W.E.L Grimson, "Adaptive Background Mixture Models for Real time Tracking." Conference on Computer vision and Pattern Recognition, 1999.

[49] P. Kelly, C. O. Conaire, D. Monaghan, J. Kuklyte, D. Connaghan, N.E.O. Connor, J. Diego and P. Daras, "A low-cost performance analysis and coaching system for tennis", ACM MultiMedia Grand Challenge, 2010

[50] S. Belongie, J. Malik and J. Puzicha, "Shape Matchine and Object Recognition using Shape Context", IEEE transactions on pattern analysis and machine intelligence, 2002

[51] M. R. Sullivan and M. Shah, "Detecting and Segmenting Humans in Crowded Scenes", 15th International Conference on Multimedia, 2007

[52] StatSoft Electronic Statistic Textbook, http://www.statsoft.com/textbook/support-vector-machines/

[53] http://crsouza.blogspot.com/2010/04/kernel-support-vector-machines-for.html

[54] http://courses.media.mit.edu/2006fall/mas622j/Projects/aisen-project/index.html

Chapter 6

Conclusions and Future Work

This chapter concludes the whole report on the basis of problems in the existing technique, proposed solution and results obtained. Future work that can be done using the proposed technique is also summarized. The importance of proposed technique in the area of computer vision is also discussed.

6.1 Conclusions

In this research work an enhanced strategy of human behavior classification is presented which outperforms the previous methods. We have validated this approach on the basis of quantitative measures giving higher accuracy on most on the videos which is not achieved in the existing technique [29]. The proposed methodology comprised on five modules including segmentation, shadow removal, human detection, feature extraction and classification whereas in the existing technique shadow removal is not included and human detection has many limitations because of which there was high rate of misclassification or false detection. The presented human detection technique is robust as applied in both indoor and outdoor scenes and give good results.

6.2 Future Work

The proposed solution can be further enhanced to be useful in many other scenarios like falling down, abnormal running and many other abnormal events that may based on such behaviors like target killing detection, snatch theft detection and suspicious activities detection. The future work that can be applied in the methodology may involve the use of HMM (Hidden Markov Model) for classification of events that are mentioned above and updating the system with more strong feature extraction method which reduce false classification of human behaviors and increasing the accuracy of the system.

6.3 Applications

There are many applications of the proposed technique which can be achieved by changing the flow and by introducing some of the other techniques.

6.3.1 Human detection and Tracking

The proposed solution can be applied to human detection and tracking systems. By following the steps of segmentation, shadow removal, human detection of the proposed technique and finally applying tracking to the detected human in the video. So, we finally get the trajectories of every individual in the video sequence which can be analyzed for identifying human motion.

6.3.2 Distributed Surveillance System

This area of surveillance system comprises of cameras fixed at different locations within a building or public places. The video sequences obtained from different cameras can also processed at the same time for analyzing the behavior of human that can be either normal or abnormal. The proposed system deals with less number of actions which can be further extended for adding many other actions like running, pulling, jumping etc.

6.3.3 Crowd Counting

Crowd counting applications can be applied on public as well as in private areas. This can done using proposed technique by applying segmentation, shadow removal, human detection and finally apply counting of human in each frame of the video sequence. Using this method we can also analyze the duration of presence of a human in the scene.

6.4 Summary

This chapter summarizes the overall discussed of the presented research work. It also explores the application areas of this research work and discussion related to the enhancement of work in future is also presented in it. Finally, this approach can be further extended for many other abnormal event detection that are based on individuals in the crowd not on the overall flow of the crowd.

References

[1] Wikipedia, http://en.wikipedia.org/wiki/Surveillance

[2] Wikipedia, http://en.wikipedia.org/wiki/Computer_vision

[3] M. Becha Ka ´ aniche, "Human Gesture Recognition", Ph.D. dissertation, Univ. of Nice Sophia Antipolis, Nov 2009

[4] Wikipedia, http://en.wikipedia.org/wiki/Segmentation_(image_processing)

[5] Generation5, http://www.generation5.org/content/2003/segmentation.asp

[6] George Mason University, Department of Computer Science, http://cs.gmu.edu/cne/modules/dau/stat/clustgalgs/clust5_bdy.html

[7] R. Adams and L. Bischof, "Seeded region growing", IEEE Trans. Pattern Analysis Machine Intelligence, pp. 641–647, Jun 1994.

[8] H. Rai, T.R.Gopalakrishnan Nair, "Gradient Based Seeded Region Grow method for CT Angiographic Image Segmentation", pp.1-6, Jan 2009.

[9] J. Fan, Guihua Zeng , M. Body, M.S. Hacid, "Seeded region growing: an extensive and comparative study", Pattern Recognition Letters, pp. 1139–1156, June 2005

[10] EE Times Design Article, http://www.eetimes.com/design/militaryaerospacedesign/4017685/Background-subtraction-part-1-MATLAB-models

[11] D. Miyazaki, Y. Matsushita, and K. Ikeuchi, "Interactive shadow removal from a single image using hierarchical graph cut" 9th Asian Conference on Computer Vision, pp. 234-245, 2009

[12] F. Liu and M. Gleicher, "Texture-Consistent Shadow Removal", 10th European Conf. on Computer Vision, pp. 437–450, 2008

[13] L. Xu, J.L. Landabaso, M. Pardàs, "Shadow Removal with Blob-Based Morphological Reconstruction for Error Correction" IEEE International Conf. on Acoustics, Speech and Signal Processing, pp. 729- 732, March 2005

[14] http://painting.about.com/cs/paintingknowhow/a/shadows.htm

[15] A. Prati, I. Miki, R. Cucchiara, M. Trivedi, "Comparative Evaluation of Moving Shadow Detection Algorithms", IEEE CVPR workshop on Empirical Evaluation Methods in Computer Vision, pp. 1-8, 2001

[16] J. Zhou and J. Hoang, "Real Time Robust Human Detection and Tracking System", IEEE conference on Computer Vision and Pattern Recognition, pp. 149- 149, June 2005

[17] Q. Zhu, Shai Avidan, M. C. Yeh and K. T. Cheng, "Fast Human Detection Using a Cascade of Histograms of Oriented Gradients", IEEE conference on Computer Vision and Pattern Recognition, pp. 1491- 1498, 2006

[18] E. L. Andrade, R. B. Fisher, and S. Blunsden, "Detection of emergency events in crowded scenes," in Proc. Inst. Eng. Technol. Conf. Crime Security, pp. 528–533, 2006

[19] Hsieh, Hsu, "A simple and fast surveillance system for human tracking and behavior analysis", IEEE Conference on Signal-Image Technologies and Internet-Based System, pp. 812-828, Dec. 2007

[20] R. Cucchiara, C. Grana, M.Piccardi, A. Prati, and S. Sirotti , "Improving Shadow Suppression in Moving Object Detection with HSV Color Information", IEEE Intelligent Transportation System Conference, pp.334-339, Aug. 2001

[21] D. Chen, M.Y. Chen, C. Gao, A. Hauptmann, H. Wactlar and A. Bharucha, "Recognition of Aggressive Human Behavior Using Binary Local Motion Descriptors" , 30th Annual International IEEE Engineering in Medicine and Biology (EMBC) Conference, pp. 20-24, Aug. 2008

[22] N. Kiryati, T. R. Raviv, Y. Ivanchenko, S. Rochel, "Real-time Abnormal Motion Detection in Surveillance Video", 19th International Conference on Pattern Recognition (ICPR 2008), pp. 8-11, Dec. 2008

[23] Y. Wang, G. Mori, "Human Action Recognition by Semi-Latent Topic Models" IEEE transactions on pattern analysis and machine intelligence, pp. 1762-1774, Oct. 2009

[24] B. Yogameena, S. Veeralakshmi, E. Komagal, S. Raju, and V. Abhaikumar, "RVM-Based Human Action Classification in Crowd through Projection and Star Skeletonization", Hindawi Publishing Corporation EURASIP Journal on Image and Video Processing, Jan. 2009

[25] C. G´arate, P. Bilinski, F. Bremond, "Crowd Event Recognition Using HOG Tracker", 12th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, 2009

[26] A. Zweng, M. Kampel, "Unexpected Human Behavior Recognition in Image Sequences using Multiple Features", International Conference on Pattern Recognition, 2010

[27] C. H. Lin, F. S. Hsu, and W. Y. Lin, "Recognizing Human Actions Using NWFE-Based Histogram Vectors", Hindawi Publishing Corporation, EURASIP Journal on Advances in Signal Processing, 2010

[28] J. Zhou and J. Hoang, "Real time robust human detection and tracking system," IEEE Conference on Computer Vision and Pattern Recognition Workshop, pp. 149–149, 2005.

[29] B. Yogameena, E. Komagal, M. Archana and S. Raju Abhaikumar, "Support Vector Machine-Based Human Behavior Classification in Crowd through Projection and Star Skeletonization", Journal of Computer Science, pp. 1008-1013, 2010

[30] S. Atev, O.Masoud, and N. Papanikolopoulos, "Practical mixtures of gaussians with brightness monitoring," IEEE Conference on Intelligent Transportation Systems, pp. 423–428, Oct. 2004.

[31] P. Kilambi, E. Ribnick, A. J. Joshi, O. Masoud, and N. Papanikolopoulos, "Estimating pedestrian counts in groups", Computer Vision and Image Understanding, pp.43–59, 2008.

[32] H. Fujiyoshi, A. J. Lipton, and T. Kanade, "Real-time human motion analysis by image skeletonization", IEICE Transactions on Information and Systems, pp. 113–120, 2004.

[33] J. Yao, J. M. Odobez, "Multi-Layer Background Subtraction Based on Color and Texture", IEEE the CVPR visual surveillance workshop, minneapolis, 2007

[34] K.A. Patwardhan, G. Sapiro, V. Morellas, "Robust foreground detection in video using pixel layers", IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 746–751, 2008

[35] M.A. Balafar, A. R. Ramli and S. Mashohor, "Edge-preserving Clustering Algorithms and Their Application for MRI Image Segmentation", International Multi-Conference of Engineers and Computer Scientists, 2010

[36] B. Zhao, L.F. Fei, and E. P. Xing, "Image Segmentation with Topic Random Field", ECCV, pp. 785–798, 2010

[37] U. Güvenç, R. Demirci and T. Karagül, "Light refraction based medical image segmentation", Scientific Research and Essays, pp. 1127-1132, May 2010

[38] Joshi, Atev, Masoud, Papanikolopoulos, "Moving Shadow Detection with Low- and Mid-Level Reasoning", IEEE International Conference of Robotics and Automation, 2007

[39] Shoaib, M. Dragon, R. Ostermann, "Shadow detection for moving humans using gradient-based background subtraction", IEEE International Conference on Acoustics, Speech and Signal Processing, 2009

[40] A. Badola, "Removal of shadow in moving object detection in video sequence", 2005

[41] B.Sun; S. Li, "Moving Cast Shadow Detection of Vehicle Using Combined Color Models", Chinese Conference of Pattern Recognition (CCPR), 2010

[42] C. C. Chen and Aggarwal, "Human Shadow Removal with Unknown Light Source", 20th International Conference of Pattern Recognition (ICPR), 2010

[43] Z. Lin, Davis, Doermann and D eMenthon, "Hierarchical Part-Template Matching for Human Detection and Segmentation", IEEE 11th International Conference of Computer Vision, 2007

[44] W.R. Schwartz, A. Kembhavi, D. Harwood, L. S. Davis, "Human detection using partial least squares analysis", IEEE 12th International Conference of Computer Vision, 2009

[45] X. Liu, P. H. Tu, J. Rittscher, A. Perera and N. Krahnstoever, "Detecting and counting people in surveillance applications", IEEE Conference of Advanced Video and Signal Based Surveillance, 2005

[46] Y. L. Hou, Pang, "Human detection in crowded scenes", 17th IEEE International Conference of Image Processing (ICIP), 2010

[47] K.E. Aziz, D. Merad, B. Fertil, N. Thome, "Pedestrian head detection and tracking using skeleton graph for people counting in crowded environments", 2011

[48] C. Stauffer, W.E.L Grimson, "Adaptive Background Mixture Models for Real time Tracking." Conference on Computer vision and Pattern Recognition, 1999.

[49] P. Kelly, C. O. Conaire, D. Monaghan, J. Kuklyte, D. Connaghan, N.E.O. Connor, J. Diego and P. Daras, "A low-cost performance analysis and coaching system for tennis", ACM MultiMedia Grand Challenge, 2010

[50] S. Belongie, J. Malik and J. Puzicha, "Shape Matchine and Object Recognition using Shape Context", IEEE transactions on pattern analysis and machine intelligence, 2002

[51] M. R. Sullivan and M. Shah, "Detecting and Segmenting Humans in Crowded Scenes", 15th International Conference on Multimedia, 2007

[52] StatSoft Electronic Statistic Textbook, http://www.statsoft.com/textbook/support-vector-machines/

[53] http://crsouza.blogspot.com/2010/04/kernel-support-vector-machines-for.html

[54] http://courses.media.mit.edu/2006fall/mas622j/Projects/aisen-project/index.html