A Dissertation on

Activity Recognition In A Video- A Real Time Approach

Submitted in partial fulfillment of the requirement

for the award of the degree of

MASTER of ENGINEERING

(Electronics & Communication Engineering)

Submitted by

Om Mishra

College Roll No: 11/E&C/2K9

University Roll No: 8519

Under the supervision and guidance of:

Prof. Rajiv Kapoor

Dept. of Electronics & Communication

Delhi College of Engineering, Delhi

DEPARTMENT OF ELECTRONICS & COMMUNICATION

DELHI COLLEGE OF ENGINEERING

UNIVERSITY OF DELHI

2009-2011

Certificate

This is to certify that the work contained in this major project entitled "**Activity Recognition In a Video- A Real Time Approach"** submitted by **Om Mishra (R. No.-8519)** of Delhi College of Engineering in partial fulfillment of the requirement for the degree of Master of Engineering in Electronics & Communication is a bonafide work carried out under my guidance and supervision in the academic year 2009-11.

The work embodied in this dissertation has not been submitted for the award of any other degree to the best of my knowledge.

Prof Rajiv Kapoor
HOD
ECE Department (DTU)

Guided by
Prof Rajiv Kapoor
ECE Department
DTU

Acknowledgement

It is a great pleasure to have the opportunity to extend my heartiest felt gratitude to everybody who helped me throughout the course of this project.

It is distinct pleasure to express my deep sense of gratitude and indebtedness to my learned supervisor **Dr. Rajiv Kapoor** for their invaluable guidance, encouragement and patient reviews. He kept on boosting me with time, to put an extra ounce of effort to realize this work. With their continuous inspiration only, it becomes possible to complete this dissertation.

I would also like to take this opportunity to present my sincere regards to all the faculty members of the Department for their support and encouragement.

I am grateful to my parents for their moral support all the time; they have been always around to cheer me up, in the odd times of this work. I am also thankful to my classmates for their unconditional support and motivation during this work. It would be injustice if I do not mention the names of **Ashish Chittora** and **Amit Kumar** –my two classmates. Both of these helped me a lot to complete this project.

Om Mishra

M.E. (Electronics & Communication Engineering)

College Roll No. 11/E&C/09

University Roll No.: 8519

Department of Electronics & Communication Engineering

Delhi College of Engineering, Delhi-42

TABLE OF CONTENTS

List of figures

List of tables

Abstract

The motivation behind this project is to develop software for tracking and recognizing the human activity major application in security, surveillance and vision analysis. The developed software must be capable of tracking the human body and recognizing its activity. The proposed method uses the approach for features extraction from the sequences of images. The method describes about the recognition of human activity with the help of change in energy produced by motion of the connected pixels in an image and then we used the support vector machine as the classifier. The proposed technique takes care of the real time implementation of the technique and in qualitative decision making both and shows better results. This technique is capable of understanding the activity. The statistical confidence is higher as compared to the previous techniques because the activity recognition is based upon the features of not just one organ but also on the dependent organs. This method works in real time and is inherently parallel.

Chapter 1

1.1 Introduction

Recognizing human activities for video surveillance is one of the most promising applications of computer vision. In recent years, this problem has caught the attention of researchers from industry, academia, security agencies, consumer agencies and the general populace too. Video surveillance is of increasing importance to many applications, such as elder-care, home-nursing, and unusual event alarming [1].

The task of recognizing human action poses several challenges. Human action is extremely diverse, and to build a system that can be used to successfully identify any type of action is a serious problem indeed. An interesting fact about human activity is the inherent similarity in the way actions are carried out. That is, people jump, stand, walk, bend down and get up in a more or less similar fashion, assuming, of course, there is no impediment in the performance of these actions.

Most systems that perform human motion analysis address general common tasks, such as: person detection & tracking, activity classification, behaviour interpretation and also person identification. Obviously, although some of these tasks can be considered independently, they must be solved in a common framework, where information can be communicated and exchanged between the different system modules. As the detection and tracking systems have progressed significantly in the past few years, [17, 22, 19], human motion and behaviour interpretation have naturally become the following step. In a surveillance scenario, tracking is the very first step and behaviour recognition the final goal. The task of activity recognition can be viewed as a bridge between the pixel measurements,

given by the tracker, and a more abstract behaviour description. In this paper, we focus on this intermediate level, that is essential to achieve the desired final large-scale interpretation. The need for such systems is increasing everyday with the number of surveillance cameras deployed in public spaces. Needless to say, the “traditional” job of the security operator, monitoring several video streams for extended periods of time, becomes impossible, as the number of cameras grows exponentially. Instead, we need systems able to detect, categorize and recognize human activity, calling for human attention only when necessary.

There are so many methods have been developed to recognize the activities and many other methods are in the process. So for this cause we are also introducing a new method for recognizing the human activities. We are basically applying our method for activity recognitions. We have applied this method to recognize the different activities for the video surveillance. We have taken the three types activity performed by the person walking, standing and jumping. The rest of the paper is structured as follows: Section 2 highlights the literature survey. Section 3 describes the research aims and objectives. Methodology for activity recognition for video surveillance is described in Section 4. Section 5 and 6 describe the classifiers for the activity recognition system and experimental results and conclusion respectively.

1.2 Background

Before discussing activity recognition, it is needed to know the basic terms used in this project. This work comes under pattern recognition and machine vision. So a brief overview of these terms has been given in the next section.

1.2.1 Pattern recognition

Automatic (machine) recognition, description, classification, and grouping of patterns are important problems in a variety of engineering and scientific disciplines such as biology, psychology, medicine, marketing, computer vision, artificial intelligence, and remote sensing. But what is a pattern? A pattern is defined as opposite of a chaos; it is an entity, vaguely defined, that could be given a name. For example, a pattern could be a fingerprint image, a handwritten cursive word, a human face, or a speech signal. Given a pattern, its recognition/classification may consist of one of the following two tasks: 1) supervised classification (e.g. discriminant analysis) in which the input pattern is identified as a member of a predefined class, 2) unsupervised classification (e.g., clustering) in which the pattern is assigned to a hitherto unknown class. Note that the recognition problem here is being posed as a classification or categorization task, where the classes are either defined by the system designer (in supervised classification) or are learned based on the similarity of patterns (in unsupervised classification). Interest in the area of pattern recognition has been renewed recently due to emerging applications which are not only challenging but also computationally more demanding.

The rapidly growing and available computing power, while enabling faster processing of huge data sets, has also facilitated the use of elaborate and diverse methods for data analysis and classification. At the same time, demands on automatic pattern recognition systems are rising enormously due to the availability of large databases and stringent performance requirements (speed, accuracy, and cost). In many of the emerging applications, it is clear that no single approach for classification is optimal and that multiple methods and approaches have to be used. Consequently, combining several sensing modalities and classifiers is now a commonly used practice in pattern recognition.

The design of a pattern recognition system essentially involves the following three aspects: making. The problem domain dictates the choice of sensor(s), pre-processing technique, representation scheme, and the decision making model. It is generally agreed that a well-defined and sufficiently constrained recognition problem (small intra-class variations and large interclass variations) will lead to a compact pattern representation and a simple decision making strategy. Learning from a set of examples (training set) is an important and desired attribute of most pattern recognition systems. The four best known approaches for pattern recognition are: 1) template matching, 2) statistical classification, 3) syntactic or structural matching, and 4) neural networks. These models are not necessarily independent and sometimes the same pattern recognition method exists with different interpretations.

1.2.2 Machine learning

Machine learning, a branch of artificial intelligence, is a scientific discipline concerned with the design and development of algorithms that allow computers to evolve behaviours based on empirical data, such as from sensor data or databases. A learner can take advantage of examples (data) to capture characteristics of interest of their unknown underlying probability distribution. Data can be seen as examples that illustrate relations between observed variables. A major focus of machine learning research is to automatically learn to recognize complex patterns and make intelligent decisions based on data; the difficulty lies in the fact that the set of all possible behaviours given all possible inputs is too large to be covered by the set of observed examples (training data). Hence the learner must generalize from the given examples, so as to be able to produce a useful output in new cases.

The computational analysis of machine learning algorithms and their performance is a branch of theoretical computer science known as computational learning theory. Because

training sets are finite and the future is uncertain, learning theory usually does not yield absolute guarantees of the performance of algorithms. Instead, probabilistic bounds on the performance are quite common. In addition to performance bounds, computational learning theorists study the time complexity and feasibility of learning. In computational learning theory, a computation is considered feasible if it can be done in polynomial time. There are two kinds of time complexity results. Positive results show that a certain class of functions can be learned in polynomial time. Negative results show that certain classes cannot be learned in polynomial time. There are many similarities between machine learning theory and statistics, although they use different terms.

Algorithms

Machine learning algorithms can be organized into a taxonomy based on the desired outcome of the algorithm.

A. Supervised learning

B. Unsupervised learning

C. Semi-supervised learning

D. Reinforcement learning

A. Supervised learning

Let us begin by considering the simplest machine learning task: supervised learning for classification.

Suppose we wish to develop a computer program that, when given a picture of a person, can determine whether the person is male or female. Such a program is called a classifier, because it assigns a class (i.e., male or female) to an object (i.e., a photograph). The task of supervised learning is to construct a classifier given a set of classified training examples—in

this case, example photographs along with the correct classes. The key challenge for supervised learning is the problem of generalization: After analyzing only a (usually small) sample of photographs, the learning system should output a classifier that works well on all possible photographs. A pair consisting of an object and its associated class is called a labelled example. The set of labelled examples provided to the learning algorithm is called the training set. Suppose we provide a training set to a learning algorithm and it outputs a classifier. How can we evaluate the quality of this classifier? The usual approach is to employ a second set of labelled examples called the test set. We measure the percentage of test examples correctly classified (called the classification rate) or the percentage of test examples misclassified (the misclassification rate). The reason we employ a separate test set is that most learned classifiers will be very accurate on the training examples. Indeed, a classifier that simply memorized the training examples would be able to classify them perfectly. We want to test the ability of the learned classifier to generalize to new data points. Note that this approach of measuring the classification rate assumes that each classification decision is independent and that each classification decision is equally important. These assumptions are often violated. The independence assumption could be violated if there is some temporal dependence in the data. Suppose for example, that the photographs were taken of students in classrooms. Some classes (e.g., early childhood development) primarily contain girls, other classes (e.g., car repair) primarily contain boys. If a classifier knew that the data consisted of batches, it could achieve higher accuracy by trying to identify the point at which one batch ends and another begins. Then within each batch of photographs, it could classify all of the objects into a single class (e.g., based on a majority vote of its guesses on the individual photographs). These kinds of temporal dependencies arise frequently. For example, a doctor seeing patients in a clinic knows that

contagious illnesses tend to come in waves. Hence, after seeing several consecutive patients with the flu, the doctor is more likely to classify the next patient as having the flu too, even if that patient's symptoms are not as clear cut as the symptoms of the previous patients. The assumption of equal importance could be violated if there are different costs or risks associated with different misclassification errors. Suppose the classifier must decide whether a patient has cancer based on some laboratory measurements. There are two kinds of errors. A false positive error occurs when the classifier classifies a healthy patient as having cancer. A false negative error occurs when the classifier classifies a person with cancer as being healthy. Typically false negatives are more costly than false positives, so we might want the learning algorithm to prefer classifiers that make fewer false negative errors, even if they make more false positives as a result.

The term supervised learning includes not only learning classifiers but also learning functions that predict numerical values. For example, given a photograph of a person, we might want to predict the person's age, height, and weight. This task is usually called regression. In this case, each labelled training example is a pair of an object and the associated numerical value. The quality of a learned prediction function is usually measured as the square of the difference between the predicted value and the true value, although sometimes the absolute value of this difference is measured instead.

B. Unsupervised learning

In machine learning, unsupervised learning refers to the problem of trying to find hidden structure in unlabeled data. Since the examples given to the learner are unlabeled, there is no error or reward signal to evaluate a potential solution. This distinguishes unsupervised learning from supervised learning and reinforcement learning.

Unsupervised learning is closely related to the problem of density estimation in statistics. However unsupervised learning also encompasses many other techniques that seek to summarize and explain key features of the data. Many methods employed in unsupervised learning are based on data mining methods used to pre-process data.

Approaches to unsupervised learning include:

• clustering (e.g., k-means, mixture models, k-nearest neighbours, hierarchical clustering),

• Blind signal separation using feature extraction techniques for dimensionality reduction (e.g., Principal component analysis, Independent component analysis, Non-negative matrix factorization, Singular value decomposition).

Among neural network models, the self-organizing map (SOM) and adaptive resonance theory (ART) are commonly used unsupervised learning algorithms. The SOM is a topographic organization in which nearby locations in the map represent inputs with similar properties. The ART model allows the number of clusters to vary with problem size and lets the user control the degree of similarity between members of the same clusters by means of a user-defined constant called the vigilance parameter. ART networks are also used for many pattern recognition tasks, such as automatic target recognition and seismic signal processing.

C. Semi-supervised learning

In computer science, semi-supervised learning is a class of machine learning techniques that make use of both labelled and unlabeled data for training - typically a small amount of labelled data with a large amount of unlabeled data. Semi-supervised learning falls between unsupervised learning (without any labelled training data) and supervised learning (with completely labelled training data). Many machine-learning researchers have found that

unlabelled data, when used in conjunction with a small amount of labelled data, can produce considerable improvement in learning accuracy. The acquisition of labelled data for a learning problem often requires a skilled human agent to manually classify training examples. The cost associated with the labelling process thus may render a fully labelled training set infeasible, whereas acquisition of unlabeled data is relatively inexpensive. In such situations, semi-supervised learning can be of great practical value.

One example of a semi-supervised learning technique is co-training, in which two or possibly more learners are each trained on a set of examples, but with each learner using a different, and ideally independent, set of features for each example. An alternative approach is to model the joint probability distribution of the features and the labels. For the unlabelled data the labels can then be treated as 'missing data'. Techniques that handle missing data, such as Gibbs sampling or the EM algorithm, can then be used to estimate the parameters of the model.

D. Reinforcement learning

Reinforcement learning is an approach to artificial intelligence that emphasizes learning by the individual from its interaction with its environment. This contrasts with classical approaches to artificial intelligence and machine learning, which have downplayed learning from interaction, focusing instead on learning from a knowledgeable teacher, or on reasoning from a complete model of the environment. Modern reinforcement learning research is highly interdisciplinary; it includes researchers specializing in operations research, genetic algorithms, neural networks, psychology, and control engineering.

Reinforcement learning is learning what to do-how to map situations to actions-so as to maximize a scalar reward signal. The learner is not told which action to take, as in most forms of machine learning, but instead must discover which actions yield the most reward by trying them. In the most interesting and challenging cases, actions may affect not only the immediate reward, but also the next situation, and through that all subsequent rewards. These two characteristics-trial-and-error search and delayed reward-are the two most important distinguishing features of reinforcement learning. One of the challenges that arise in reinforcement learning and not in other kinds of learning is the trade off between exploration and exploitation. To obtain a lot of reward, a reinforcement learning agent must prefer actions that it has tried in the past and found to be effective in producing reward. But to discover which actions these are it has to select actions that it has not tried before. The agent has to exploit what it already knows in order to obtain reward, but it also has to explore in order to make better action selections in the future. The dilemma is that neither exploitation nor exploration can be pursued exclusively without failing at the task.

1.2.3 Computer vision

Activity recognition is a sub-part of computer vision. So it is needed to know the basics of computer vision. Computer vision (image understanding) is a discipline that studies how to reconstruct, interpret and understand a 3D scene from its 2D images in terms of the properties of the structures present in the scene. Computer vision is concerned with modelling and replicating human vision using computer software and hardware. It combines knowledge in computer science, electrical engineering, mathematics, physiology, biology,

and cognitive science. It needs knowledge from all these fields in order to understand and simulate the operation of the human vision system.

1.2.3.1 Computer Vision Hierarchy

Low-level vision: process image for feature extraction (edge, corner, or optical flow).

• Intermediate-level vision: object recognition and 3D scene interpretation using features obtained from the low-level vision.

• High-level vision: interpretation of the evolving information provided by the intermediate level vision as well as directing what intermediate and low level vision tasks should be performed. Interpretation may include conceptual description of a scene like activity, intention and behaviour.

1.2.3.2 Why study Computer Vision?

• Images and movies are everywhere

• Fast-growing collection of useful applications

– building representations of the 3D world from pictures

– Automated surveillance (who's doing what)

– Movie post-processing

– face recognition

• Various deep and attractive scientific mysteries

– How does object recognition work?

– Beautiful marriage of math, biology, physics, engineering

• Greater understanding of human vision.

Chapter 2

Literature survey

Much work has been done in activity recognition. Cai and Aggarwal [2] discuss the different approaches used in the recognition of human activities. They classify the approaches towards human activity recognition into state-space and template matching techniques. Liao et al [3] discuss methodologies which use motion in the recognition of human activity. Ayers and Shah [4] have developed a system that makes context-based decisions about the actions of people in a room. These actions include entering a room, using a computer terminal, opening a cabinet, picking up the phone, etc. Their system is able to recognize actions based on prior knowledge about the layout of the room. Davis, Intille and Bobick [10] have developed an algorithm that uses contextual information to simultaneously track multiple, non-rigid objects when erratic movements and object collisions are common. However, both of these algorithms require prior knowledge of the precise location of certain objects in the environment. In [4], the system is limited to actions like sitting and standing. Also, it is only able to recognize a picking action by knowledge of where the object is and tracking it after the person has come within a certain distance of it. In [8], Davis uses temporal plates for matching and recognition. The system computes history images (MHI's) of the persons in the scene. Davis [8] computes MHI's for 18 different images in 7 different orientations. These motion images are accumulated in time and form motion energy images (MEI's). Moment-based features are extracted from MEI's and MHI's and employed for recognition using template matching. Although template matching procedures have a lower

computational cost, they are usually more sensitive to the variance in the duration of the movement. A number of researchers have attempted the full three-dimensional reconstruction of the human form from image sequences, presuming that such information is necessary to understand the action taking place [11, 7, 15]. Others have proposed methods for recognizing action from the motion itself, as opposed to constructing a three-dimensional model of the person and then recognizing the action of the model [12, 5]. Rosario and Pentland [13], uses the Bayesian framework for modelling human actions. Given the correct probability density functions, Bayes theory is optimal in the sense of producing minimal classification errors. State space models have been widely used to detect, predict and estimate time series over a long period of time. Many state space systems use the hidden Markov model (HMM), a probabilistic model for the study of discrete time series. In [13, 16], HMMs have been applied to human activity recognition.

The proposed method uses the totally new approach for features extraction from the sequences of images .The method describes about the recognition of human activity with the help of change in kinetic energy produced by motion of the lattices which in turn based on connected pixels in an image. Then we used the support vector machine as the classifier. The various techniques which we explained above are either lack in the real time implementation of the technique or in qualitative decision making. The proposed technique takes care of both and shows better results.

Chapter 3

3.Research aims and Objectives

The main focus of this research was to develop an automated video surveillance system. Video surveillance has been an active research topic in these days. Recently our society has faced a lot of terrorist attacks in which a lot of people has been killed. On the border of the country, we have lost many soldiers due to manual monitoring. On the traffic road, road casualties have increased recently. These may be reduced by an effective surveillance system. These are some areas where video surveillance is demand of time.

For automated video surveillance an effective algorithm is needed for detecting any abnormal activity in the video in real time. We have developed a good algorithm for detecting any activities in the video. We have applied this algorithm on three different activities(Walking,jumping and standing position).

Suppose we want to monitor the activities happening in the restricted area.One option is that we appoint a person at the place where CCTV camera is placed. But it is not a good option as it is consuming a lot of manpower. Other option is that we take help from an automated video surveillance system. By analyzing the output of this system, we can find out what is happening at that place. But for effective operation of this option, video surveillance system should operate in real time.

We have tested our algorithm on similar situation. There may be three types of behaviour in the Walking, Jumping and Standing position. Our method is able to recognize these activities in real time.

Chapter 4

4.Methodology

We have divided our method in to three parts.

1. Feature extraction
2. Feature representation and Description
3. Classifier

The whole process of activity recognition, using proposed method, can be shown with the help of following block diagram .

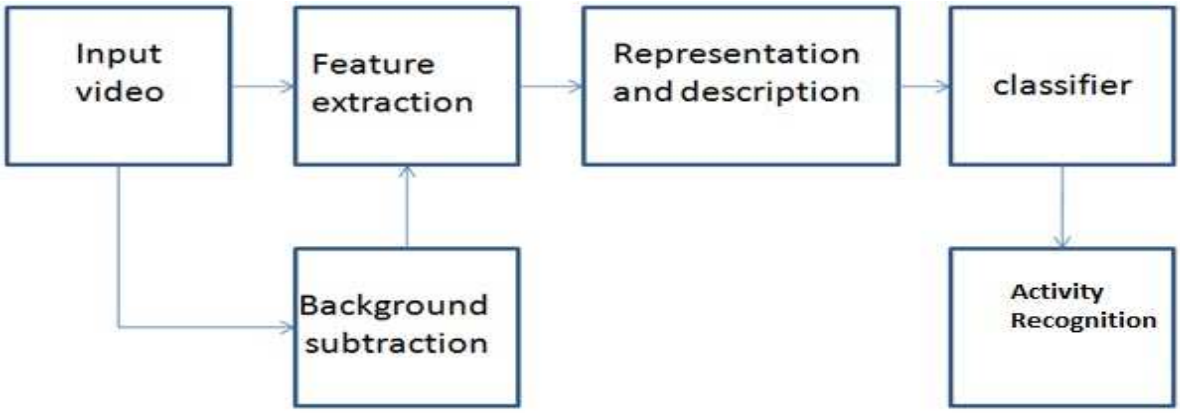


Figure 1. Block diagram of activity recognition system

The steps proposed in the block diagram, as shown in figure 1, are explained in following subsections.

4.1 Pre-processing:

There will be need of activity database for training. For this purpose videos of different activities (walking, jumping and stand position) have been taken shown in figures 2(a), 2(b) and 2(c) respectively.

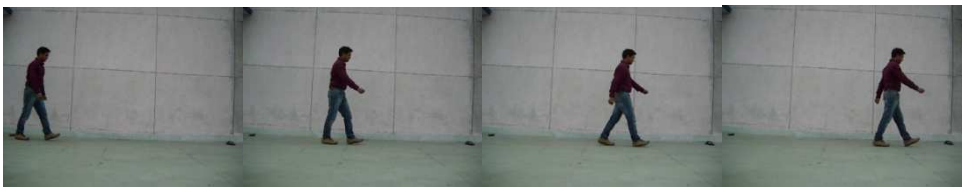


Figure 2(a). "Walking"

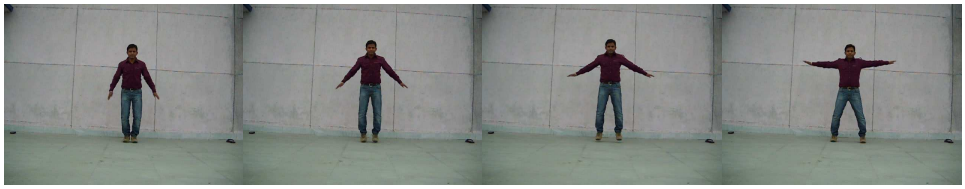


Figure 2(b) . "Jumping"

Figure 2(c). "Standing position"

To recognize any particular activity we have segmented the human body into three different parts i.e. lower part(leg),middle part(hand),upper part(head) as shown in figures 3(a), 3(b) and 3(c) respectively.

Figure 3(a). Lower body part (Leg) for different activities

Figure 3(b). Middle body part for different activities

Figure 3(c). Upper body part for different activities

The segmentation process can be done either by manual cropping of the images or by some existed automatic segmentation methods[26].Further more pre-processing is required because there is a large change in the energy due to gray level intensity of the pixel as compared to change in energy due to position.the energy due to intensity value will add those energy which is not of our interest because to recognize activity we are only concentrating on the motion of the connected pixels in sequences of frames. So to compensate this energy effect we removed the background from background subtraction.

4.2 Background subtraction

Background subtraction is a widely used method in Computer Vision for separating or segmenting out the foreground objects from the background of a video. The foreground objects are defined to be the parts of the image that changes and the background is made out of the pixels that stay relatively constant.

In the Computer Vision field, background subtraction is considered to be a low level processing task. It is usually performed as a pre-processing step before more high level tasks such as blob detection; tracking and object detection are performed.

Commonly used techniques for Background Subtraction Include

- Subtraction of reference background frame from each frame.
- Frame Differencing
- Gaussian Mixture Models (GMM)

In the first category, a reference frame is taken as background. Now this frame is subtracted from each frame of the video. Blobs found after background subtraction indicate foreground and remaining portion as background. But this method is not useful for moving background. To remove this problem, we use temporal frame differencing for background subtraction. In this approach consecutive frames are subtracted from each other. This approach is very adaptive to dynamic environments, but generally does a poor job of extracting all the relevant pixels, e.g., there may be holes left inside moving entities. Third approach is based on Gaussian Mixture Model. This is highly useful in case of modelling adaptive background. This tackles the problem of moving background and change in illumination of the scene. A brief overview of GMM has been given below.

Gaussian Mixture Model

GMM based method was first introduced by Stauffer and Grimson in 1999, and now it is the most widely used method for background subtraction due to its speed, simplicity and the ease of implementation. In this method, each pixel is modelled as a mixture of Gaussian distributions and any pixel intensity value that does not fit into one of the modelled Gaussian distributions is marked as a foreground pixel.

The background subtraction involves two different tasks, each of which needs to be performed real-time, with having only the video frames as the input.

1. Learning the background model
2. Classifying pixels as background or foreground

Learning the Background Model

Following parameters of each Gaussian component need to be learned dynamically

- The parameters of Gaussians
  - Mean
  - Variance and
  - Weight
- Number of Gaussians per pixel

The update equations for the Gaussian parameters are given below. These equations are executed for each Gaussian component for each pixel at the arrival of each video frame.

$$\begin{aligned}\hat{\pi}_m &\leftarrow \hat{\pi}_m + \alpha(o_m^{(t)} - \hat{\pi}_m) \\ \widehat{\vec{\mu}}_m &\leftarrow \widehat{\vec{\mu}}_m + o_m^{(t)}(\alpha/\hat{\pi}_m)\vec{\delta}_m \\ \widehat{\sigma}_m^2 &\leftarrow \widehat{\sigma}_m^2 + o_m^{(t)}(\alpha/\hat{\pi}_m)(\vec{\delta}_m^T\vec{\delta}_m - \widehat{\sigma}_m^2)\end{aligned} \quad (1)$$

Classifying Pixels

$\vec{x}^{(t)}$ = value of a pixel at time t in RGB colour space.

Bayesian decision R – if pixel is background (BG) or foreground (FG):

$$R = \frac{p(BG|\vec{x}^{(t)})}{p(FG|\vec{x}^{(t)})} = \frac{p(\vec{x}^{(t)}|BG)p(BG)}{p(\vec{x}^{(t)}|FG)p(FG)} \quad (2)$$

Initially set p(FG) = p(BG), therefore if the following condition is true we decide that the pixels is a background pixel

$$p(\vec{x}^{(t)}|BG) > c_{thr} \quad (3)$$

But we have used second approach 'frame differencing' for background subtraction because for real time application GMM based approach is not a better option. GMM based approach makes the method very slow as in this method each pixel is modelled by a group of Gaussians.

Now we converted these frames from gray to binary. So in this binary image object is represented by the white pixel as foreground and background by black pixels. Now we applied our method on binary image where we are only concentrating on white pixels which represent the object. So the processing time gets decreases and method becomes well applicable on real time application.

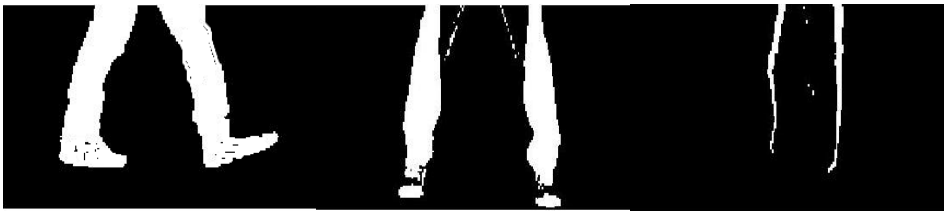


Figure 4(a). Binary image for lower body part for different activities

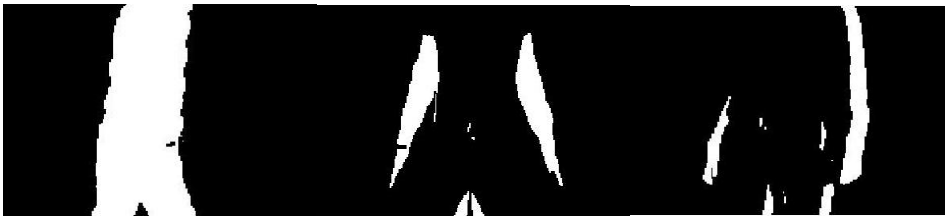


Figure 4(b). Binary image for middle body part for different activities

Figure 4(c). Binary image for upper body part for different activities

Figure 4(a), 4(b) and 4(c) shows the binary images of lower, middle and upper body parts respectively for different activities.

4.3 Features Extraction:

In the image processing, feature extraction from the images is very critical step for developing the method for any application. Extracted features vary from application to application. Basic image processing tools like morphology, histogram, filters etc. are used for feature extraction. Area, height, shape, diameter, centroid of the blob, perimeter of the shape are some examples of the features. Colour information in the image is also an

important feature. Efficiency of the method will improve as the number of features increases. But while making a real time system, we have to compromise with the efficiency. For real time application, redundant features are reduced using a dimension reduction technique. Principal Component analysis (PCA) is used for such an application. It is an unsupervised learning and is a standard technique commonly used for data reduction in statistical pattern recognition and signal processing. In this approach to perform dimensionality reduction on some input data, we compute the eigen values and eigen vectors of the correlation matrix of the input data vector, and then project the data orthogonally on to the subspace spanned by the eigen vectors belonging to the dominant eigen values.

Various approaches have been proposed for activity recognition in the video. They can be broadly categorized according to the type of feature extraction and representation adopted. It may be classified in to two ways.

4.3.1 Tracking based approach

One very popular category is based on trajectory modelling. It comprises tracking each object in the scene, and learning models for the resulting object tracks.

Tracking overview

In day to day life, there has been an increasing interest in image tracking and activity recognition systems. Due to the large amount of applications there those features can be used. Image tracking and activity recognition are receiving increasing attention among computer scientists due to the wide spectrum of applications where they can be used, ranging from athletic performance analysis to video surveillance. By image tracking we refer to the ability of a computer to recover the position and orientation of the object from a

sequence of images. There have been several different approaches to allow computers to derive automatically the kinematics pose and activity from image sequences. Video tracking is the process of locating a moving object in time using a camera. An algorithm analyses the video frames and outputs the location of moving targets within the video frame. The main difficulty in video tracking is to associate target locations in consecutive video frames, especially when the objects are moving fast relative to the frame rate. Here, video tracking systems usually employ a motion model which describes how the image of the target might change for different possible motions of the object to track.

After motion detection, surveillance systems generally track moving objects from one frame to another in an image sequence. The tracking algorithms usually have considerable intersection with motion detection during processing. Tracking over time typically involves matching objects in consecutive frames using features such as points, lines or blobs. Useful mathematical tools for tracking include the Kalman filter, the Condensation algorithm, the dynamic Bayesian network, the geodesic method, etc. Tracking methods are divided into four major categories: region-based tracking, active-contour-based tracking, feature based tracking, and model-based tracking. It should be pointed out that this classification is not absolute in that algorithms from different categories can be integrated together.

a. Region-Based Tracking

Region-based tracking algorithms track objects according to variations of the image regions corresponding to the moving objects. For these algorithms, the background image is maintained dynamically and motion regions are usually detected by subtracting the background from the current image. Wren *et al.* [39] explore the use of small blob features to track a single human in an indoor environment. In their work, a human body is

considered as a combination of some blobs respectively representing various body parts such as head, torso and the four limbs. Meanwhile, both human body and background scene are modelled with Gaussian distributions of pixel values. Finally, the pixels belonging to the human body are assigned to the different body part's blobs using the log-likelihood measure. Therefore, by tracking each small blob, the moving human is successfully tracked. Recently, McKenna *et al.* [40] propose an adaptive background subtraction method in which colour and gradient information are combined to cope with shadows and unreliable colour cues in motion segmentation. Tracking is then performed at three levels of abstraction: regions, people, and groups. Each region has a bounding box and regions can merge and split. A human is composed of one or more regions grouped together under the condition of geometric structure constraints on the human body, and a human group consists of one or more people grouped together. Therefore, using the region tracker and the individual colour appearance model, perfect tracking of multiple people is achieved, even during occlusion. Although they work well in scenes containing only a few objects (such as highways), region-based tracking algorithms cannot reliably handle occlusion between objects. Furthermore, as these algorithms only obtain the tracking results at the region level and are essentially procedures for motion detection, the outline or 3-D pose of objects cannot be acquired. (The 3-D pose of an object consists of the position and orientation of the object). Accordingly, these algorithms cannot satisfy the requirements for surveillance against a cluttered background or with multiple moving objects.

b. Active Contour-Based Tracking

Active contour-based tracking algorithms track objects by representing their outlines as bounding contours and updating these contours dynamically in successive frames. These

algorithms aim at directly extracting shapes of subjects and provide more effective descriptions of objects than region- based algorithms. Paragios *et al.* [46] detect and track multiple moving objects in image sequences using a geodesic active contour objective function and a level set formulation scheme. Peterfreund [47] explores a new active contour model based on a Kalman filter for tracking non rigid moving targets such as people in spatio-velocity space. Isard *et al.* [48] adopt stochastic differential equations to describe complex motion models, and combine this approach with deformable templates to cope with people tracking. Malik *et al.* [52] have successfully applied active contour-based methods to vehicle tracking. In contrast to region-based tracking algorithms, active contour-based algorithms describe objects more simply and more effectively and reduce computational complexity. Even under disturbance or partial occlusion, these algorithms may track objects continuously. However, the tracking precision is limited at the contour level. The recovery of the 3-D pose of an object from its contour on the image plane is a demanding problem. A further difficulty is that the active contour-based algorithms are highly sensitive to the initialization of tracking, making it difficult to start tracking automatically.

c. Feature-Based Tracking

Feature-based tracking algorithms perform recognition and tracking of objects by extracting elements, clustering them into higher level features and then matching the features between images. Feature-based tracking algorithms can further be classified into three subcategories according to the nature of selected features: global feature-based algorithms, local feature-based algorithms, and dependence-graph-based algorithms.

• The features used in global feature-based algorithms include centroid, perimeters, areas, some orders of quadratures and colours etc. Polana *et al.* [49] provide a good example of global feature-based tracking. A person is bounded with a rectangular box whose centroid is selected as the feature for tracking. Even when occlusion happens between two persons during tracking, as long as the velocity of the centroid can be distinguished effectively, tracking is still successful.

• The features used in local feature-based algorithms include line segments, curve segments, and corner vertices etc.

• The features used in dependence-graph-based algorithms include a variety of distances and geometric relations between features.

The above three methods can be combined. In the recent work of Jang *et al.* [50], an active template that characterizes regional and Structural features of an object is built dynamically based on the information of shape, texture, colour, and edge features of the region. Using motion estimation based on a Kalman filter, the tracking of a non rigid moving object is successfully performed by minimizing a feature energy function during the matching process.

In general, as they operate on 2-D image planes, feature-based tracking algorithms can adapt successfully and rapidly to allow real-time processing and tracking of multiple objects which are required in heavy thruway scenes, etc. However, dependence graph- based algorithms cannot be used in real-time tracking because they need time-consuming searching and matching of graphs. Feature-based tracking algorithms can handle partial occlusion by using information on object motion, local features and dependence graphs. However, there are several serious deficiencies in feature-based tracking algorithms.

• The recognition rate of objects based on 2-D image features is low, because of the nonlinear distortion during perspective projection and the image variations with the viewpoint's movement.

• These algorithms are generally unable to recover 3-D pose of objects.

• The stability of dealing effectively with occlusion, overlapping and interference of unrelated structures is generally poor.

d. Model-Based Tracking

Model-based tracking algorithms track objects by matching projected object models, produced with prior knowledge, to image data. The models are usually constructed off-line with manual measurement, CAD tools or computer vision techniques. As model-based rigid object tracking and model-based non rigid object tracking are quite different, we review separately model-based human body tracking (non rigid object tracking) and model-based vehicle tracking (rigid object tracking).

1) Model-Based Human Body Tracking*:* The general approach for model-based human body tracing is known as analysis-by-synthesis, and it is used in a predict-match-update style. Firstly, the pose of the model for the next frame is predicted according to prior knowledge and tracking history. Then, the predicted model is synthesized and projected into the image plane for comparison with the image data. A specific pose evaluation function is needed to measure the similarity between the projected model and the image data. According to different search strategies, this is done either recursively or using sampling techniques until the correct pose is finally found and is used to update the model. Pose estimation in the first frame needs to be handled specially. Generally, model-based human body tracking involves three main issues:

• Construction of human body models;

• Representation of prior knowledge of motion models and motion constraints;

• Prediction and search strategies.

Previous work on these three issues is briefly and respectively reviewed as follows.

a) Human body models*:* Construction of human body models is the base of model-based human body tracking. Generally, the more complex a human body model, the more accurate the tracking results, but the more expensive the computation. Traditionally, the geometric structure of human body can be represented in the following four styles.

• **Stick figure**. The essence of human motion is typically contained in the movements of the torso, the head and the four limbs, so the stick-figure method is to represent the parts of a human body as sticks and link the sticks with joints. Karaulova *et al.* [42] use a stick figure representation to build a novel hierarchical model of human dynamics encoded using hidden Markov models (HMMs), and realize view-independent tracking of a human body in monocular image sequences.

• **2-D contour**. This kind of human body model is directly relevant to human body projections in an image plane. The human body segments are modelled by 2-D ribbons or blobs. For instance, Ju *et al.* [43] propose a cardboard human body model, in which the human limbs are represented by a set of jointed planar ribbons. The parameterized image motion of these patches is constrained to enforce the articulated movement of human limbs. Niyogi *et al.* [44] use the spatial-temporal pattern in XYT space to track, analyze and recognize walking figures. They examine the characteristic braided pattern produced by the lower limbs of a walking human, the projections of head movements are then located in the spatio-temporal domain, followed by the identification of the joint trajectories; The contour of a walking figure is outlined by utilizing these joint trajectories, and a more accurate gait

analysis is carried out using the outlined 2-D contour for the recognition of the specific human.

• **Volumetric models**. The main disadvantage of 2-D models is that they require restrictions on the viewing angle. To overcome this disadvantage, many researchers use 3-D volumetric models such as elliptical cylinders, cones, spheres, super-quadrics etc. Volumetric models require more parameters than image-based models and lead to more expensive computation during the matching process. Wachter *et al.* [45] establish a 3D body model using connected elliptical cones.

• **Hierarchical model**. Plankers *et al.* [53] present a hierarchical human model for achieving more accurate results. It includes four levels: skeleton, ellipsoid meatballs simulating tissues and fats, polygonal surface representing skin, and shaded rendering.

b) Motion models: Motion models of human limbs and joints are widely used in tracking. They are effective because the movements of the limbs are strongly constrained. These motion models serve as prior knowledge to predict motion parameters to interpret and recognize human behaviours, or to constrain the estimation of low-level image measurements.

A number of methods have been developed for learning two dimensional (2-D) motion paths [55], [56] resulting from tracking of objects or people [57]. Here, a large number of normal individuals or objects are tracked over time during the training phase. The resulting paths are then summarized by a set of motion trajectories, often translated into a symbolic representation of the background activity. In the detection phase, paths extracted from the monitored video are compared against those extracted in the training phase. Tracking is generally performed by means of graphical state-based representations, such as hidden Markov models or Bayesian networks [57]–[61]. Johnson and Hogg [62] consider human

trajectories in this context. The method begins by vector-quantizing tracks and clustering the result into a predetermined number of pdfs using a neural network. Based on the training data, the method predicts trajectory of a pedestrian and decides if it is anomalous or not. This approach was subsequently improved by simplifying the training step [63] and embedding it into a hierarchical structure based on co-occurrence statistics [64]. More recently, Saleemi et al. [65] proposed a stochastic, nonparametric method for modelling scene tracks. The authors claim that the use of predicted trajectories and tracking method robust to occlusions jointly permit the analysis of more general scenes, unlike other methods that are limited to roads and walkways. From a statistical perspective, the tracks amount to features, here and a nominal distribution of tracks, $g_o(l)$, is obtained in the training phase. For the anomalous tracks, the implicit assumption is that the tracks are uniformly distributed over the set of all tracks. In this statistical perspective, the optimal detection rule is a thresholding strategy, whereby outliers with respect to the nominal tracks are declared abnormal. Although there are advantages to using paths as motion features, there are clear disadvantages as well. First, tracking is a difficult task, especially in real time and in urban scenarios where often a large number of objects are present. Since the anomaly detection is directly related to the quality of tracking, a tracking error will inevitably bias the detection step. Second, since each individual or object monitored is related to a single path, it is hard to deal with people occluding each other.

4.3.2 Alternative approach avoiding tracking

Various authors have proposed alternative motion representations that avoid tracking. The most popular is dense optical flow, or some other form of spatio-temporal gradients. This approach focuses uniquely on motion information, ignoring abnormality information due to

variations of object appearance. This makes them impervious to abnormalities that do not involve motion outliers, e.g. a truck that crosses a bridge with weight restrictions. Furthermore, descriptors such as optical flow, pixel change histograms, or other traditional background subtraction operations, are difficult for crowded scenes, where the background is by definition dynamic, there are lots of clutter, and complicated occlusions. More complete representations that account for both appearance and motion have also been proposed. Overall, there is a great diversity of approaches to activity recognition. In general, it is quite difficult to compare two different solutions. Different representations of motion and appearance are combined with different graphical models for activity recognition, which are typically tailored to the type of video analyzed, or a specific scene domain.

A statistical framework is often used to describe activity recognition. In this setting, we are given features, l, in a suitably high-dimensional space. Feature instances are distributed with a probability density function (pdf), $g_0(.)$, if they come from a nominal distribution. Anomalous instances are distributed with pdf, $g_1(l)$. In the statistical framework, the anomaly detection problem amounts to predicting whether an instance l is distributed according to nominal or anomalous pdf. Thus, the detection problem can be stated as follows:

$$H_0: l \sim g_0(.) \ vs \ the \ alternative \ (anomaly) H_1: l \sim g_1(.)$$

If both pdfs are either known or can be estimated from training data this task reduces to the well-known likelihood ratio test (LRT) [38]. Nevertheless, video anomaly detection is generally difficult because of the following issues: The nominal and anomalous distributions are generally unknown and difficult to estimate even when the training data is given. This is because the boundary between normal and anomalous behaviour depends on the choice of

feature descriptors and distance metrics used that, in turn, significantly affect the performance. Anomalous behaviour is typically not apparent from raw video (pixel intensities) and the video data needs to be transformed to a feature space where anomalies may become apparent. Both normality and abnormality are diverse since a typical urban video contains a multitude of activities. Therefore, there is no single distribution that captures either the nominal or anomalous activity. A critical issue is the general lack of labelled data for training/ validation. This is particularly acute for anomalous data, since it is difficult to stage a rich set of anomalies, which is representative of a real-world scenario. Computational speed, in addition to statistical performance, is also an important performance metric for an anomaly detection system. For instance, there may be a need for real-time detection in surveillance situations. Finally, the nominal background activity is non-stationary over a long time scale, e.g., activities during the day are significantly different from those at night. Consequently, the non stationary has to be accounted for. There are many types of video anomalies, and it is generally difficult to group them. One possible classification is based on the fundamental dynamical nature of video data. In this perspective, video anomalies can be thought of as the presence/ absence of usual (or unusual) objects or motion attributes in unusual (or usual) locations or times. For instance, an abandoned- object anomaly occurs when an object is present at an unusual location for an extended period of time. A trespassing anomaly occurs when motion is present at unusual locations. An illegal U-turn anomaly occurs when an unusual motion attribute appears at the usual locations. Clearly, video anomalies can either be localized, as in the case of abandoned objects, or spatially distributed, as in the case of illegal U-turns. A significant effort has been devoted to video anomaly detection in surveillance applications over the last decade. The general problem, however, is still open because of the wide

variety of anomalies; most of the existing anomaly detection techniques solve the problem in a specific scenario. In the following, we review some major approaches to video anomaly detection. These techniques differ both in terms of what is known about the training data as well as the different transformations and metrics used for activity detection.

4.3.3 Approach

The binary image obtained from pre-processing is used to extract useful information in terms of features. We have adopted second type of method for feature extraction that avoids tracking. For feature extraction related with any significant change in the video, each pixel of the image has been modelled by a Gaussian distribution function. To represent the whole greyscale of image by this function, Gaussian distribution functions [51, 54] have been represented either by varying variance value of Gaussian and keeping mean constant or by varying mean and keeping variance constant. Gaussian distribution function is represented by equation (1)

$$f_{ij}(z) = \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{(z-z_{mi})^2}{2\sigma_{zj}^2}} \quad (4)$$

Where z_{mi} and σ_{zj} are the Mean and Variance of the membership function $f_{ij}(z)$ respectively. *z* is the gray level value of the pixel at coordinate position *x* and *y* of the image. Taking pixel connectivity of order 8, every pixel in the image has 8 neighbourhoods corresponding to direction of 45, 90, 135, 180, 225, 270, 315, 360 degrees respectively except corner pixels. We have applied the Gaussian distribution function on each pixel.

Each Gaussian modelled pixel has been compared with the gray value of its neighbourhood pixels within some thresholding limit. Neighbourhood pixels, which satisfy this relation are termed as connected pixels of that Gaussian modelled pixel.

In gausssian membership function mean has been taken unity because object is represented by the white pixel in binary image.In binary images, Gaussian Membership Function of x and y, given by equation (4),play significant roles for activity recognition. After representing each pixel of the binary image by Gaussian Membership Function, given by equation (4), connected pixels for each pixel have been found..

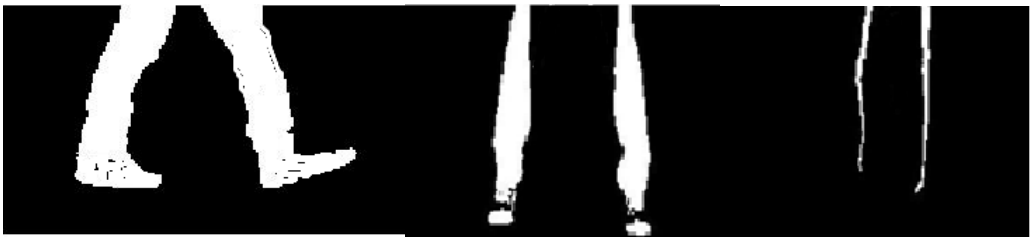


Figure 5(a). Connected pixel of lower part in binary image

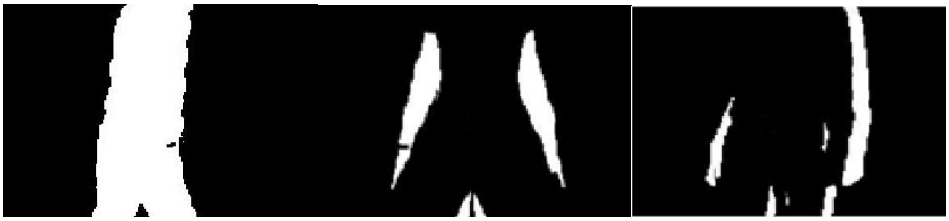


Figure 5(b). Connected pixel of middle part in binary image

Figure 5(c). Connected pixel of upper part in binary image

Figure 5(a), 5(b) and 5(c) shows the sequence of connected pixels of binary images shown in figure4(a), 4(b) and 4(c) respectively. The coordinates of all connected pixels are stored in the matrix. The data of matrix is considered as useful information features of the image. By this way useful information features are extracted from the image.

Algorithm:

Now we will find out the sequentially connected white pixel or lattices of the binary image.

1. A test image of same size of original binary image is taken.

2. Pixels present in the useful information matrix are made unity in the test image and remaining pixels are made zero.

3. Now test image is scanned for white pixels. Whenever a white pixel is found , it is taken as central pixel. As neighbourhoods of central pixel is named as testing pixels.

(a). We scan the testing pixels for connected components of the central pixel. Whenever it finds connected component , the central pixel ‘s and testing pixels ‘s coordinates are stored in a row of a matrix named as sequence connected pixels. Now central pixel in the test image is made zero and present testing pixel is taken as new central pixel. Now repeat this whole step . If it does not find any connected component, go to next step.

(b). Count the number of white pixels present in the test image. If these are greater than 1, go to step 3 ,otherwise terminate the loop.

(c). The white blobs represent the sequence of the connected pixels of the corresponding frame. These lattices changes with any dynamic change in the frame of the video.

4.4 Feature representation and description

After features have been extracted from the frames of the video, the resulting features representing any change in the video is represented and described in a form suitable for further computer processing.

4.4.1 Representation

Basically, representing a region involves two choices:

(1) We can represent the regions in terms of its external characteristics (its boundary).

(2) We can represent in terms of its internal characteristics (the pixels comprising the region).

An external representation is chosen when the primary focus is on the shape characteristics. An internal representation is selected when the primary focus is on the regional properties, such as colour and texture. Sometimes it may be necessary to use both types of representation. In either case, the features selected as descriptors should be as insensitive as possible to variations in size, translation and rotation.

There are various ways of representation of features like Chain codes, polygonal approximations, signatures, boundary segments skeletons.

4.4.1.1 Chain codes

Chain codes are used to represent a boundary by a connected sequence of straight- line segments of specified length and direction. Typically, this representation is based on 4-or 8-connectivity of the segments. The direction of each segment is coded by using a numbering

scheme. A boundary code formed as a sequence of such directional numbers is referred to as a Freeman chain code. The chain code of a boundary depends on the starting point. However, the code can be normalized with respect to the starting point by a straightforward procedure.

4.4.1.2 Polygonal approximations

A digital boundary can be approximated with arbitrary accuracy by a polygon. For a closed boundary, the approximation becomes exact when the number segments of the polygon is equal to the number of points in the boundary so that each pair of adjacent points defines a segment of the polygon. The goal of a polygonal approximation is to capture the essence of the shapes in a given boundary using the fewest possible number of segments. However, approximation techniques of modest complexity are well suited for image processing tasks.

4.4.1.3 Signatures

A signature is a 1-D functional representation of a boundary and may be generated in various ways. One of the simplest is to plot the distance from the centroid to the boundary as a function of angle. Regardless of how a signature is generated, however, the basic idea is to introduce the boundary representation to a 1-D function that presumably is easier to describe than the original 2-D boundary.

4.4.1.4 Boundary segments

Decomposing a boundary in to segments is often useful. Decomposition reduces the boundary's complexity and thus simplifies the description process. This approach is

particularly attractive when the boundary contains one or more significant concavities that carry shape information.

4.4.1.5 Skeletons

An important approach to representing the structural shape of a plane region is to reduce it to a graph. This reduction may be accomplished by obtaining the skeleton of the region via a thinning.

4.4.2 Description

Choosing a representation scheme, however, is only part of the task of making the data useful to a computer. The next task is to describe the region based on the chosen representation. For example, a region may be represented by its boundary and the boundary described by features such as its length, the orientation of the straight line joining its extreme points, and the number of concavities in the boundary.

Basically descriptors can be classified in two ways –boundary descriptors and regional descriptors.

4.4.2.1 Boundary descriptors

The length of a boundary is one of its simplest descriptors. The number of pixels along a boundary gives a rough approximation of its length. The value of the diameter and the orientation of a line segment connecting the two extremes points that comprise the diameter (this line is called the major axis of the boundary) are useful descriptors of a boundary. The minor axis of a boundary is defined as the line perpendicular to the major axis, and of such length that a box passing through the outer four points of intersection of

the boundary. The box just described is called the basic rectangle, and the ratio of the major to the minor axis is called the eccentricity of the boundary. This also is a useful descriptor. The shape of boundary segments can be described quantitatively by using statistical moments, such as the mean, variance, and the higher order moments.

4.4.2.2 Regional descriptors

Some simple descriptors are the area and the perimeter of a region. The area of a region is defined as the number of pixels in the region. The perimeter of a region is the length of its boundary. A more frequent use of these two descriptors is in measuring compactness of a region. A slightly different descriptor of compactness is the circularity ratio, defined as the ratio of a region to the area of circle the same perimeter. Other simple measures used as region descriptors include the mean and median of the intensity levels, the minimum and maximum intensity values, and the number of pixels with values above and below the mean. Topological properties are useful for global descriptions of regions in the image plane. Simply defined, topology is the study of properties of a figure that are unaffected by any deformation, as long as there is no tearing or joining of the figure. Another topological property useful for region description is the number of connected components.

4.4.3 Approach

As we have represented any change in the video in terms of sequence of the connected pixels with the help of Gaussian distribution functions. Every pixel has varying number of connected pixels. It means there are a number of directions of change corresponding to each connected pixel. To reduce this redundancy of change, we have taken most prominent direction of change corresponding to each connected pixel. Now resulting representation is

in terms of sequence connected pixels. These sequence connected pixels have been shown below in figure 5.

We have also derived sequence connected pixels corresponding to the background subtracted frames of the video.

These sequence connected pixels represent the motion vector in a frame. Motion vector is the most prominent change in the video. Now next task of the algorithm is to describe this motion vector in a suitable format so that it can be classified effectively according to the application. We have described these features represented in terms of motion vector using energy descriptors.

The crowd kinetic energy [41] of each frame is as follows:

$$E_i = \sum_{j=1}^{n} m_j V_j^2 \qquad (5)$$

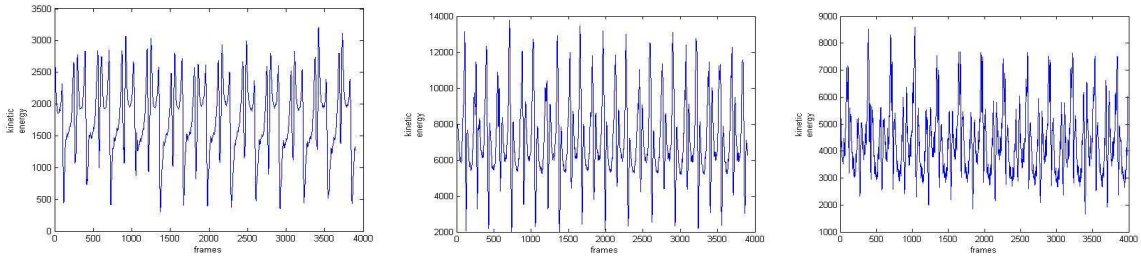
Where $\mathbf{E_i}$ is the crowd kinetic energy of the i^{th} frame in a video sequence, V_j is the magnitude of the j^{th} motion vector in the ith frame, and n is the total number of motion vectors in the i^{th} frame respectively. For a given scene with similar sizes of objects, we assume that $m_j = 1$. From (7) we can obtain the crowd kinetic energy for each video frame which can be used to indicate the global status of the crowd scene.

However, the crowd kinetic energy cannot provide any information about the crowd density, which is actually closely related to abnormal situations. So we define the modified crowd kinetic energy (MCKE) as follows:

$$\mathrm{ME_i} = \rho_i \sum_{j=1}^{n} V_j^2 \qquad (6)$$

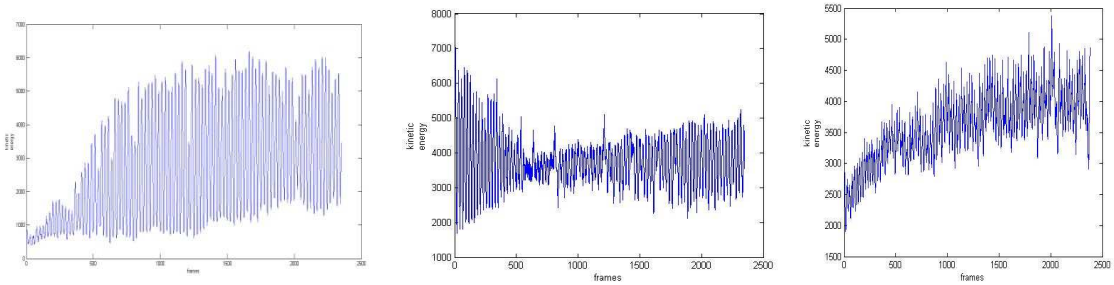
Where ρ_i is the ratio of foreground area to background area. In this application, we do not need accurate crowd density information. Thus the foreground to background ratio, which is a rough estimate of crowd density, is appropriate for abnormal events detection.

We have calculated this modified crowd kinetic energy for each frame of the video. This energy signal is shown in the figure.



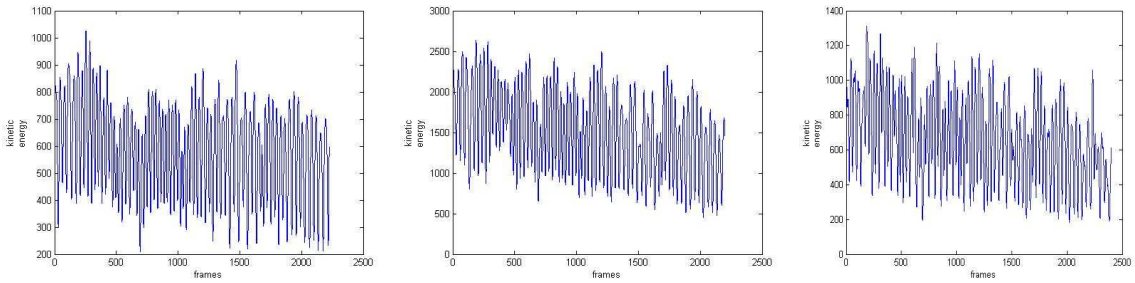
1. Upper body part 2. Middle body part 3. Lower body part

Figure 6(a). Energy graph of "walking" for different human body parts



1. Upper body part 2. Middle body part 3. Lower body part

Figure 6(b). Energy graph of "Jumping" for different human body parts



1. Upper body part 2. Middle body part 3. Lower body part

Figure 6(c). Energy graph of “Standing” for different human body parts

The energy of some frames of the video for activities (walking, jumping and standing) have been shown in figure 6(a), 6(b) and 6(c).

Now our main aim is to recognize activity which has been performed in the video.we have the feature data in the form of enregy of every frame which represets any activity.We have made the such automatic recognition system which will observe the test video for particular duration and then it will decide that which activity has been performed i.e.for training for our system we have taken the 150 frames as the duration and applied normal distribution on data(which is the kinetic energy) to find out the mean and variance and non parametric distribution [27,32,33] to find the bandwidth.so for every activity we have found the mean, variance and bandwidth.

These three values are used to plot unique points, corresponding to the each activity in three dimensions as shown in figure 7(a), 7(b) and 7(c).

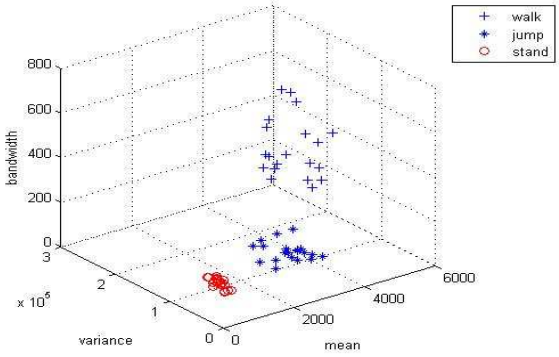


Figure 7(a). Lower body part

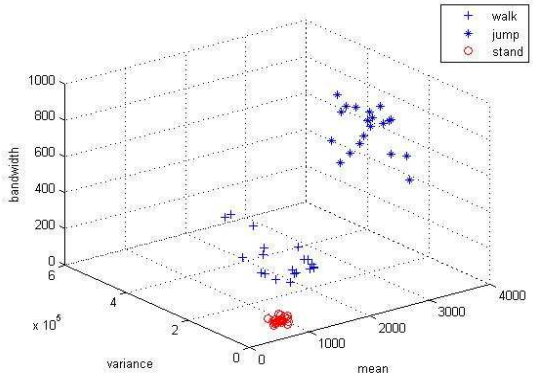


Figure 7(b). Upper body part

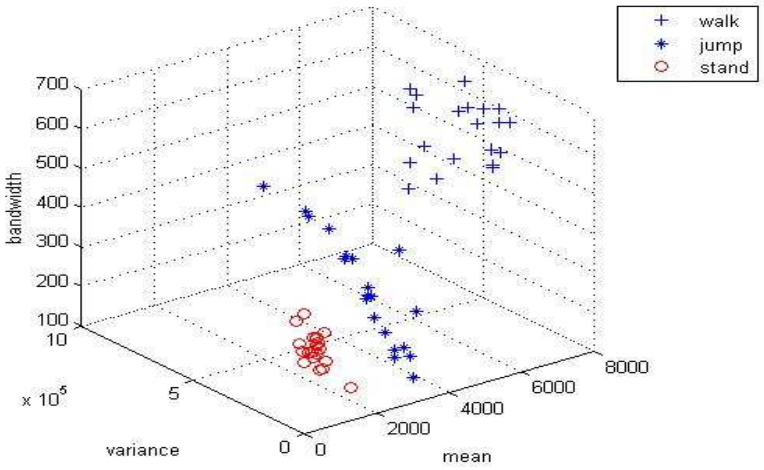


Figure 7(c). Middle body part

Figure 7. Representation of Three dimensional feature space for different activities

Chapter 5

5.Classifiers

5.1 Classifier –overview

Once a feature selection or classification procedure finds a proper representation, a classifier can be designed using a number of possible approaches. In practice, the choice of a classifier is a difficult problem and it is often based on which classifier(s) happen to be available, or best known, to the user. The simplest and the most intuitive approach to classifier design is based on the concept of similarity: patterns that are similar should be assigned to the same class. So, once a good metric has been established to define similarity, patterns can be classified by template matching or the minimum distance classifier using a few prototypes per class. The choice of the metric and the prototypes is crucial to the success of this approach.

We have given a brief introduction of some common classifiers used in pattern recognition and machine learning. These are described below:

5.1.1 k-Nearest Neighbour algorithm

In pattern recognition, the *k*-nearest neighbours algorithm (*k*-NN) is a method for classifying objects based on closest training examples in the feature space. *k-NN* is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification. The *k*-nearest neighbour algorithm is amongst the simplest of all machine learning algorithms: an object is classified by a majority vote of its neighbours, with the object being assigned to the class most common amongst its *k*

nearest neighbours (*k* is a positive integer, typically small). If *k* = 1, then the object is simply assigned to the class of its nearest neighbour.

The same method can be used for regression, by simply assigning the property value for the object to be the average of the values of its *k* nearest neighbours. It can be useful to weight the contributions of the neighbours, so that the nearer neighbours contribute more to the average than the more distant ones. (A common weighting scheme is to give each neighbour a weight of 1/*d*, where *d* is the distance to the neighbour. This scheme is a generalization of linear interpolation.)

The neighbours are taken from a set of objects for which the correct classification (or, in the case of regression, the value of the property) is known. This can be thought of as the training set for the algorithm, though no explicit training step is required. The *k*-nearest neighbour algorithm is sensitive to the local structure of the data.

Nearest neighbour rules in effect compute the decision boundary in an implicit manner. It is also possible to compute the decision boundary itself explicitly, and to do so in an efficient manner so that the computational complexity is a function of the boundary complexity.

5.1.1.1 Algorithm

The training examples are vectors in a multidimensional feature space, each with a class label. The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples. In the classification phase, *k* is a user-defined constant, and an unlabelled vector (a query or test point) is classified by assigning the label which is most frequent among the *k* training samples nearest to that query point. Usually Euclidean distance is used as the distance metric; however this is only applicable to continuous

variables. In cases such as text classification, another metric such as the overlap metric (or Hamming distance) can be used. Often, the classification accuracy of "k"-NN can be improved significantly if the distance metric is learned with specialized algorithms such as e.g. Large Margin Nearest Neighbour or Neighbourhood components analysis. A drawback to the basic "majority voting" classification is that the classes with the more frequent examples tend to dominate the prediction of the new vector, as they tend to come up in the *k* nearest neighbours when the neighbours are computed due to their large number. One way to overcome this problem is to weight the classification taking into account the distance from the test point to each of its *k* nearest neighbours.

5.1.1.2 Parameter selection

The best choice of *k* depends upon the data; generally, larger values of *k* reduce the effect of noise on the classification, but make boundaries between classes less distinct. A good *k* can be selected by various heuristic techniques, for example, cross-validation. The special case where the class is predicted to be the class of the closest training sample (i.e. when *k* = 1) is called the nearest neighbour algorithm. The accuracy of the *k*-NN algorithm can be severely degraded by the presence of noisy or irrelevant features, or if the feature scales are not consistent with their importance. Much research effort has been put into selecting or scaling features to improve classification. A particularly popular approach is the use of evolutionary algorithms to optimize feature scaling. Another popular approach is to scale features by the mutual information of the training data with the training classes. In binary (two class) classification problems, it is helpful to choose *k* to be an odd number as this avoids tied votes. One popular way of choosing the empirically optimal *k* in this setting is via bootstrap method.

5.1.2 Neural network

Neural networks can be viewed as massively parallel computing systems consisting of an extremely large number of simple processors with many interconnections. Neural network models attempt to use some organizational principles (such as learning, generalization, adaptivity, fault tolerance and distributed representation, and Pattern Recognition Models computation) in a network of weighted directed graphs in which the nodes are artificial neurons and directed edges (with weights) are connections between neuron outputs and neuron inputs. The main characteristics of neural networks are that they have the ability to learn complex nonlinear input-output relationships, use sequential training procedures, and adapt themselves to the data.

The most commonly used family of neural networks for pattern classification tasks is the single-layer perceptron, where the separating hyper plane is iteratively updated as a function of the distances of the misclassified patterns from the hyper plane. If the sigmoid function is used in combination with the MSE criterion, as in feed-forward neural nets (also called multilayer perceptrons), the perceptron may show a behaviour which is similar to other linear classifiers. It is important to note that neural networks themselves can lead to many different classifiers depending on how they are trained. While the hidden layers in multilayer perceptrons allow nonlinear decision boundaries, they also increase the danger of overtraining the classifier since the number of network parameters increases as more layers and more neurons per layer are added. Therefore, the regularization of neural networks may be necessary. Many regularization mechanisms are already built in, such as slow training in combination with early stopping.

The other most commonly used family of neural networks for pattern classification tasks is the feed-forward network, which includes multilayer perceptron and Radial-Basis Function (RBF) networks. These networks are organized into layers and have unidirectional connections between the layers. Another popular network is the Self-Organizing Map (SOM), or Kohonen-Network, which is mainly used for data clustering and feature mapping. The learning process involves updating network architecture and connection weights so that a network can efficiently perform a specific classification/clustering task.

The increasing popularity of neural network models to solve pattern recognition problems has been primarily due to their seemingly low dependence on domain-specific knowledge (relative to model-based and rule-based approaches) and due to the availability of efficient learning algorithms for practitioners to use. Neural networks provide a new suite of nonlinear algorithms for feature extraction (using hidden layers) and classification (e.g., multilayer perceptrons). In addition, existing feature extraction and classification algorithms can also be mapped on neural network architectures for efficient (hardware) implementation. In spite of the seemingly different underlying principles, most of the well known neural network models are implicitly equivalent or similar to classical statistical pattern recognition methods. Most NNs conceal the statistics from the user. Despite these similarities, neural networks do offer several advantages such as, unified approaches for feature extraction and classification and flexible procedures for finding good, moderately non linear solutions.

5.1.3 Bayes classifier

The second main concept used for designing pattern classifiers is based on the probabilistic approach. The optimal Bayes decision rule (with the 0/1 loss function) assigns a pattern to

the class with the maximum posterior probability. This rule can be modified to take into account costs associated with different types of misclassifications. For known class conditional densities, the Bayes decision rule gives the optimum classifier, in the sense that, for given prior probabilities, loss function and class-conditional densities, no other decision rule will have a lower risk (i.e., expected value of the loss function, for example, probability of error). If the prior class probabilities are equal and a 0/1 loss function is adopted, the Bayes decision rule and the maximum likelihood decision rule exactly coincide.

5.1.4 Hidden markov model

Hidden markov model is a commonly used classifier in pattern recognition and machine learning. It is commonly used for speech processing. Before discussing it, it is needed to know about markov processes:

Consider a system which may be described at any time as being in one of a set of N distinct states S_1, $S_{2,\ldots,}$ S_N as illustrated in figure 8.

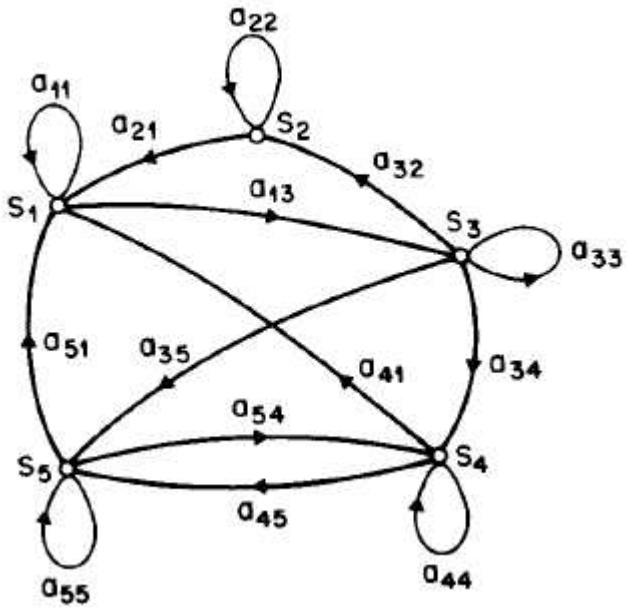


Figure 8, A Markov chain with 5 states (labelled S_1, to S_5) with selected state transitions.

At regularly spaced discrete times, the system undergoes a change of state (possibly back to the same state) according to a set of probabilities associated with the state. We denote the time instants associated with state changes as $t = 1, 2, \ldots$, and we denote the actual state at time t as qr. A full probabilistic description of the above system would, in general, require specification of the current state (at time t), as well as all the predecessor states. For the special case of a discrete, first order, Markov chain, this probabilistic description is truncated to just the current and the predecessor state, i.e.

$$P[q_t = S_j | q_{t-1} = S_i, q_{t-2} = S_k, \cdots] = P[q_t = S_j | q_{t-1} = S_i]. \quad (7)$$

The above stochastic process could be called an observable Markov model since the output of the process is the set of states at each instant of time, where each state corresponds to a physical (observable) event.

5.2 Support vector machine

Support vector machines (SVMs) are a set of related supervised learning methods that analyze data and recognize patterns, used for classification and regression analysis. The standard SVM takes a set of input data, and predicts, for each given input, which of two possible classes the input is a member of, which makes the SVM a non-probabilistic binary linear classifier. Since an SVM is a classifier, then given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that predicts whether a new example falls into one category or the other. Intuitively, an SVM model is a representation of the examples as points in space, mapped so that the

examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

More formally, a support vector machine constructs a hyper plane or set of hyper planes in a high or infinite dimensional space, which can be used for classification, regression or other tasks. Intuitively, a good separation is achieved by the hyper plane that has the largest distance to the nearest training data points of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier.

Whereas the original problem may be stated in a finite dimensional space, it often happens that in that space the sets to be discriminated are not linearly separable. For this reason it was proposed that the original finite dimensional space be mapped into a much higher dimensional space presumably making the separation easier in that space. SVM schemes use a mapping into a larger space so that cross products may be computed easily in terms of the variables in the original space making the computational load reasonable. The cross products in the larger space are defined in terms of a kernel function which can be selected to suit the problem. The hyper planes in the large space are defined as the set of points whose cross product with a vector in that space is constant. The vectors defining the hyper planes can be chosen to be linear combinations with parameters of images of feature vectors which occur in the data base. With this choice of a hyper plane the points x in the feature space which are mapped into the hyper plane are defined by the relation: Note that if becomes small as grows further from , each element in the sum measures the degree of closeness of the test point to the corresponding data base point . In this way the sum of kernels above can be used to measure the relative nearness of each test point to the data points originating in one or the other of the sets to be discriminated.

Multiclass SVM aims to assign labels to instances by using support vector machines, where the labels are drawn from a finite set of several elements. The dominating approach for doing so is to reduce the single multiclass problem into multiple binary classification problems. Each of the problems yields a binary classifier, which is assumed to produce an output function that gives relatively large values for examples from the positive class and relatively small values for examples belonging to the negative class. Two common methods to build such binary classifiers are where each classifier distinguishes between (i) one of the labels to the rest (one-versus-all) or (ii) between every pair of classes (one-versus-one). Classification of new instances for one-versus-all case is done by a winner-takes-all strategy, in which the classifier with the highest output function assigns the class (it is important that the output functions be calibrated to produce comparable scores). For the one-versus-one approach, classification is done by max-wins voting strategy, in which every classifier assigns the instance to one of the two classes, and then the vote for the assigned class is increased by one vote, and finally the class with most votes determines the instance classification.

Now we have the clusters points corresponding to every activity in three dimensions. The best way to classify such data is by Support Vector Machine. At first we need to train the SVM. SVMs can be trained by supervised learning.

As SVM is dedicated to binary classification problems, three popular strategies have been proposed to apply it to multi-class problems. Suppose we are dealing with a *K*-class problem. Thus *K* binary SVMs need to be trained. The scheme is the *one-against- one* method [28, 29], which trains $K(K-1)/2$ binary SVMs, each of which discriminate two of the *K* classes. Other newest schemes are Binary Decision Tree (BDT) SVM [30], Directed Acyclic Graph (DAG) SVM [31], which are more complex than other methods. The binary SVM separates the clusters in classes by hyper planes.

We used one against all SVM classification method, it gives satisfactory results. There is one more important issue is selection of the kernel function for SVM. Linear kernel may give erroneous results, instead of that polynomial or radial basis function kernel gives better results.

5.3 Decision:

In the previous section we classified the activity performed by different body parts of a person .the classifier had classified the activity into three different class (walk, jump and stand) for lower, middle and upper body parts; now we need to take a decision for the overall activity performed by a person. For a particular activity like walking, jumping or standing can be recognized if it comes under the class of that activity.

Chapter 6

6.1 Experimental Result

In this section experiments have been performed on the test videos. In these videos same activities have been performed for which our system is trained. The mean, variance and bandwidth are found for the test video using the above explained procedure. Finally the data is tested with classifier for its respective class and the overall performance of the system for these videos is obtained.

Figure 9(a) Figure9(b) Figure 9(c)

Figure 9 Test video#1

Figure 10(a) Figure 10(b) Figure 10(c)

Figure 10 Test video#2

The above figures i.e. figure 9 and figure 10. are the test video frames in which (a) represents walking , (b) represents standing and (c) represents jumping .

Database	Test Video#1	Test Video#2	Total
Recognition Rate	95.3 %	96.7 %	97.0 %

Table 1. Experimental results of Test videos.

We tested the proposed method with different type of kernel function for SVM. The kernel function may be either of linear, quadratic, polynomial or radial basis function type. Recognition rate varies for different kernel functions. As shown in the following table, RBF kernel gives best recognition rate.

SVM type	Linear	Quadratic	Polynomial	RBF
Recognition Rate	97.2 %	97.9 %	98.4 %	99.2 %

Table 2. Recognition rates (in percentage) taking different type of SVM kernel functions

The comparison of recognition rate of proposed method with other conventional methods is given in the following table:

Method	Recognition rate on ORL database
Naive Bayes Network (NBN)	89.5
Bayes Network (BN)	90.5
Multilayer	89.5

Perceptron (MLP)	
Radial Basis Function Network (RBF)	91.5
Support Vector Machine (SVM)	96.3

Table 3. Comparison of recognition rate (in percentage) with other techniques

Here naive bayes network(NBN), bayes network (BN),multilayer perceptron(MLP),radial basis function(RBF) and support vector machine(SVM) are the methods of behaviour recognition where support vector machine has given best recognition rate (96.3%) among them.

6.2 Conclusion

In this thesis, we have combined the concept of mathematics, physics and computer science along with signal processing to model a new method for recognizing the activities in the video in real time. Non-linear gaussian member-ship functions (GMF) have been applied on video for finding the dynamic features.Then we used crowd kinetic energy for modelling these features. Then SVM is used as the classifier.Experimental results demonstrate the effectiveness of our proposed method.

Chapter 7

References

[1] P. Harmo, T. Taipalus, J. Knuuttila, J. Wallet and A. Halme, “Needs and Solutions- Home Automation and Service Robots for the Elderly and Disabled,”*Int’l Conf. Intelligent Robot, Systems*, pp. 3201- 3206, 2005.

[2] J. K. Aggarwal and Qin Cai. Human motion analysis: A review. Computer Vision and Image Understanding, pages 428{440, 1999.

[3] J.K. Aggarwal, Qin Cai, W. Liao, and B. Sabata. Nonrigid motion analysis: Articulated and elastic motion. Computer Vision and Image Understanding, pages 142{156, 1997.

[5] Douglas Ayers and Mubarak Shah. Recognizing human action in a static room. In Proceedings Computer Vision and Pattern Recognition, pages 42{46, 1998.

[5] A. Bobick and J. Davis. Appearance-based motion recognition of human actions. Master's thesis, Massachusetts Institute of Technology, 1996.

[6] Qin Cai and J.K. Aggarwal. Automatic tracking of human motion in indoor scene across multiple synchronized video streams. International Conference on Computer Vision, 1998.

[7] L. Campbell and A. Bobick. Recognition of human body motion using phase space constraints. IEEE International Conference on Computer Vision Proceedings of the 5th International Conference on Computer Vision, pages 624{630, 1995.

[8] James Davis and Aaron Bobick. The representation and recognition of action using temporal plates. In Proceedings Computer Vision and Pattern Recognition, pages 928{934, 1997.

[9] R. O. Duda and P. E. Hart. Pattern Classification and Scene Analysis. New York Wiley, 1973.

[10] Stephen S. Intille, James Davis, and Aaron Bobick. Real time closed world tracking. In Proceedings IEEE International Conference on Computer Vision and Pattern Recognition, pages 697{703, 1997.

[11] J. Rehg and T. Kanade. Model based tracking of self-occluding articulated objects. IEEE International Conference on Computer Vision Proceedings of the 5th International Conference on Computer Vision, pages 612{617, 1995.

[12] K Rohr. Towards model-based recognition of human movements in image sequences. CVGIP, Image Understanding, 59 n:94{115, 1994.

[13] Nuria Oliver Barbara Rosario and Alex Pentland. A bayesian computer vision system for modeling human interactions. Proceedings of ICVS99, 1999.

[14] Saad Ahmed Sirohey. Human face segmentation and identification. Master's thesis, University of Maryland, 1993.

[15] J. Yamato, J. Ohya, and K. Ishii. Recognizing human action in time sequential images using hidden markov models. In Proceedings of IEEE Interantional Conference on Computer Vision and Pattern Recognition, pages 379{385, 1992.

[16] Jie Yang, Yangsheng Xu, and Chiou S. Chen. Human action learning via hidden markov model. IEEE Transactions on Systems, Man and Cybernetics, A:34{ 44, 1997

[17] I. Hariataoglu, D. Harwood, L. Davis,“W4: A Real-Time Surveillance of People and Their Activities”, *IEEE Transactions on Pattern Analysis and Machine Intelligence* , Vol. 22, No. 8, August 2000.

[18] J.W. Davis, “Sequential Reliable-Inference for Rapid Detection of human actions”, IEEE Conf. on Advance Video and Signal Based Surveillance, pp. 169-176, July 21-22, 2003, Miami, FL.

[19] Somboon Hongeng, Ram Nevatia, Francois Bremond, “Video-based event recognition: activity representation and probabilistic recognition methods”, *Computer Vision and Image Understanding* 96, (2004) 129-162 .

[20] K. Kira and L. Rendell, “A practical approach to feature selection”, *Pro. 9th Int. Workshop on Machine Learning*, 1992, (pp. 249-256).

[21] Schrodinger, E. (1926). *"An Undulatory Theory of the Mechanics of Atoms and Molecules"* (http:/ / home. tiscali. nl/ physis/ HistoricPaper/ Schroedinger/ Schroedinger1926c. pdf). *Physical Review* 28 (6): 1049–1070. doi:10.1103/PhysRev.28.1049.

[22] Tao Zhao, Ram Nevatia, “Tracking Multiple Humans in Complex Situations”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 9, September 2004.

[23] R. Nevatia, F. Lv and T. Zhao, “Self-calibration of a camera from video of a walking human”, International Conference on Pattern Recognition, pp. I: 562-567, Quebec City, Canada, August 2002.

[24] Rmer Rosales, Stan Sclaroff, “3D Trajectory Recovery for Tracking Multiple Objects and Trajectory Guided Recognition of Actions”, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. 2, June 1999.

[25] N. Vlassis and A. Likas, “The kurtosis-EM algorithm for Gaussian mixture modelling,” IEEE Trans. SMC ,1999.

[26] Werghi.N,Yijun Xiao,Siebrt,J.P., “A functional based segmentation of human body scans in arbitrary postures, IEEE transections on Cybernatics, Part:B, Volume. 36, Issue. 1, Pub: 2006, PP. 153-165.

[27] Altman, Naomi and Christian Leger (1995): “Bandwidth selection for kernel distribution function,estimation,” Journal of Statistical Planning and Inference, 46, 195-214.

[28] S. Knerr, L. Personnaz, and G. Dreyfus, *Nurocosingle-layer learning revisited: A stepwise procedure for building and training a neural network*, Springer, 1990.

[29] U. Kressel, *Pairwise classification and support vector machines*, in Advances in Kernel Methods- Support Vector Learning (1999).

[30]Hossam Osman.*Novel multiclass SVM-based binary decision tree classifier*. IEEE International Symposium on Signal Processing and Information Technology, pages 880-883, 2007.

[31]Platt, J., Cristanini, N., Shawe-Taylor, J.: *Large margin DAGs for multiclass classification*. Advances in Neural Information Processing Systems 12. MIT Press. Pages 543–557, 2000.

[32] Azzalini, A. (1981): “A note on the estimation of a distribution function and quantiles by a kernel method,” Biometrika, 68, 326-328.

[33] Jin, Zhezhen and Yongzhao Shao (1999): “On kernel estimation of a multivariate distribution function,”Statistics and Probability Letters, 41, 163-168.

[34] Fengjun Lv, Jinman Kang, Ram Nevatia, Isaac Cohen, Gerard Medioni “Automatic tracking and labeling of human activities in a video sequence”,IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, Prague, May 11-14, 2004.

[35] O. Masoud, N. Papanikolopoulos, “Recognizing human activities”, IEEE Conf. on Advanced Video and Signal Surveillance, 2003.

[36] Aaron F. Bobick, James W. Davis, “The Recognition of Human Movement Using Temporal Templates”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 3, March 2001.

[37] Alexei Efros, Alexander Berg, Greg Mori, Jitendra Malik, “Recognizing Actions at a Distance”, *IEEE International Conference on Computer Vision*, Nice, France, October 2003.

EMRS DTC Technical Conference – Edinburgh 2006

[38] Tao Xiang and Shaogang Gong, “Video Behaviour Profiling and Abnormality Detection without Manual Labelling”, Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV’05)

[39] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, “Pfinder: real-time tracking of the human body,” IEEE Trans. Pattern Anal. Machine Intell., vol. 19, pp. 780–785, July 1997.

[40] S. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler, “Tracking groups of people,” Comput. Vis. Image Understanding, vol. 80, no. 1, pp. 42–56, 2000.

[42] I. A. Karaulova, P. M. Hall, and A. D. Marshall, “A hierarchical model of dynamics for tracking people with a single video camera,” in Proc. British Machine Vision Conf., 2000, pp. 262–352.

[43] S. Ju, M. Black, and Y. Yaccob, “Cardboard people: a parameterized model of articulated image motion,” in Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition, 1996, pp. 38–44.

[44] S. A. Niyogi and E. H. Adelson, “Analyzing and recognizing walking figures in XYT,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 1994, pp. 469–474.

[45] S. Wachter and H.-H. Nagel, “Tracking persons in monocular image sequences,” Comput. Vis. Image Understanding, vol. 74, no. 3, pp. 174–192, 1999.

[46] N. Paragios and R. Deriche, “Geodesic active contours and level sets for the detection and tracking of moving objects,” IEEE Trans. Pattern Anal. Machine Intell., vol. 22, pp. 266–280, Mar. 2000.

[47] N. Peterfreund, “Robust tracking of position and velocity with Kalman snakes,” IEEE Trans. Pattern Anal. Machine Intell., vol. 22, pp. 564–569, June 2000.

[48] M. Isard and A. Blake, “Contour tracking by stochastic propagation of conditional density,” in Proc. European Conf. Computer Vision, 1996, pp. 343–356.

[49] R. Polana and R. Nelson, “Low level recognition of human motion,” in Proc. IEEE Workshop Motion of Non-Rigid and Articulated Objects, Austin, TX, 1994, pp. 77–82.

[50] D.-S. Jang and H.-I. Choi, “Active models for tracking moving objects,” Pattern Recognit., vol. 33, no. 7, pp. 1135–1146, 2000.

[51] V. K. Madasu and S. Vasikarla, “Fuzzy edge detection in biometric systems”, in 36th Applied Imagery Pattern Recognition Workshop, IEEE, pp. 139-144, 2007.

[52] D. Koller, J. Weber, T. Huang, J. Malik, G. Ogasawara, B. Rao, and S. Russel, “Toward robust automatic traffic scene analysis in real-time,” in Proc. Int. Conf. Pattern Recognition, Israel, 1994, pp. 126–131.

[53] R. Plankers and P. Fua, “Articulated soft objects for video-based body modelling,” in Proc. Int. Conf. Computer Vision, Vancouver, BC, Canada, 2001, pp. 394–401.

[54] D.O. Aborisade, “Fuzzy Logic Based Digital Image Edge Detection”, in Global journal of Computer Science and Technology, pp. 78-84, 2010.

[55] C. Piciarelli, C. Micheloni, and G. Foresti, “Trajectory-based anomalous event detection,” IEEE Trans. Circuits Syst. Video Technol., vol. 18, no. 11, pp. 1544–1554, 2008.

[56] T. Chen, H. Haussecker, A. Bovyrin, R. Belenov, K. Rodyushkin, A. Kuranov, and V. Eruhimov, “Computer vision workload analysis: Case study of video surveillance systems,” Intell. Technol. J., vol. 9, no. 2, pp. 109–118, 2005.

[57] W. Hu, T. Tab, L. Wang, and S. Maybank, “A survey on visual surveillance of object motion and behaviors,” IEEE Trans. Syst. Man Cybern., vol. 34, no. 3, pp. 334–352, 2004.

[58] P. Kumar, S. Ranganath, H. Weimin, and K. Sengupta, “Framework for real time behavior interpretation from traffic video,” IEEE Trans. Intell. Transp. Syst., vol. 6, no. 1, pp. 43–53, 2005.

[59] M. Bennewitz, G. Cielniak, and W. Burgard, “Utilizing learned motion patterns to robustly track persons,” in Proc. IEEE Int. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, Nice, France, 2003, pp. 102–109.

[60] N. M. Oliver, B. Rosario, and A. P. Pentland, “A Bayesian computer vision system for modeling human interactions,” IEEE Trans. Pattern Anal. Machine Intell., vol. 22, no. 8, pp. 831–843, 2000.

[61] N. Vaswani, A. Roy-Chowdhury, and R. Chellappa, “Shape activity: A continuous state hmm for moving/deforming shapes with application to abnormal activity activity detection,” IEEE Trans. Image Processing, vol. 14, no. 10, pp. 1603–1616, 2005.

[62] N. Johnson and D. Hogg, “Learning the distribution of object trajectories for event recognition,” Image Vis. Comput., vol. 14, no. 8, pp. 609–615, 1996.

[63] N. Sumpter and A. Bulpitt, “Learning spatio-temporal patterns for predicting object behavior,” Image Vis. Comput., vol. 18, no. 9, pp. 697–704, 2000.

[64] C. Stauffer and E. Grimson, “Learning patterns of activity using real-time tracking,” IEEE Trans. Pattern Anal. Machine Intell., vol. 22, no. 8, pp. 747–757, 2000.

[65] I. Saleemi, K. Shafique, and M. Shah, “Probabilistic modeling of scene dynamics for applications in visual surveillance,” IEEE Trans. Pattern Anal. Machine Intell., vol. 31, no. 8, pp. 1472–1485, 2009.