Activity Recognition and Abnormality Detection in Sports Crowd using Semi-supervised Learning

Undertaken by
Mehwish
345-FBAS/MSCS/F07

Supervisor
Mr. Asim Munir

Department of Computer Science and Software Engineering
Faculty of Basic and Applied Sciences
International Islamic University, H-10, Islamabad

(2011)

International Islamic University
CENTRAL LIBRARY

TH. 8849
Accession No.

MS
004
MEA

Computer Science

Computer Software

DATA ENTERED
28/05/13

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

Dated: 20-02-2012

Final Approval

It is certified that we have read the thesis, titled **"Activity Recognition and Abnormality Detection in Sports Crowd using Semi-supervised Learning"** submitted by **Miss Mehwish** Reg. No. **345-FAS/MSCS/F07**. It is our judgment that this thesis is of sufficient standard to warrant its acceptance by the International Islamic University, Islamabad, for the Degree of **Master of Science.**

Committee

External Examiner
Dr. Waseem Shahzad
Assistant Professor
Department of Computer Science
NU FAST, Islamabad

Internal Examiner
Ms. Fareeha Anwar
Lecturer
Department of Computer Science and Software Engineering,
Faculty of Basic and Applied Sciences,
International Islamic University, Islamabad.

Supervisor
Mr. Asim Munir
Assistant Professor,
Department of Computer Science and Software Engineering,
Faculty of Basic and Applied Sciences,
International Islamic University, Islamabad.

A dissertation submitted to the

Department of Computer Science and Software Engineering,

Faculty of Basic and Applied Sciences,

International Islamic University, Islamabad, Pakistan

as a partial fulfillment of the requirements

for the award of the degree of

Master of Science in Computer Science

بسم الله الرحمن الرحيم

In the name of Almighty Allah,
The most Beneficent, the most Merciful

Dedicated To

To My Parents

Declaration

I hereby declare that this research, neither in part nor in full, has been copied from any source, except where cited; hence, acknowledged. It is further declared that this research, in its entirety, is a product of my personal efforts, under the sincere guidance of my supervisor. No portion of the work being presented herein, has been submitted to any other university, institute, or seat of learning, in support to any piece of writing for bestowment of any other degree of qualification.

Mehwish
345-FBAS/MSCS/F07

Acknowledgements

All praise to the Almighty Allah, Who is the source of all knowledge and wisdom endowed to mankind.

I have deepest obligations to my loving parents, sister and brother for their encouragement and inspiration that have always been with me.

Many thanks to my project supervisor, **Mr. Asim Munir** whose constant motivation, unflagging efforts and uninvolved words of wisdom ever proved a lighthouse for me; it was earnestly felt whenever I swayed. Despite his never ending assignments of university management, student counseling, project supervision and teaching, they did never mind whenever I ask for an advice, within or without the time slot allocated for me.

Acknowledgement is also due to my teachers for dedicatedly instillimg and imparting enlightenment to me during the course of studies and afterwards for my project.

I am really grateful to Mr. Syed Ali Naqi Gilani for his continuous assistance. Without his motivation and help this research work would not be possible.

Mehwish
345-FBAS/MSCS/F07

Project in Brief

Project Title:	Activity Recognition and Abnormality Detection in Sports Crowd using Semi-supervised Learning
Undertaken By:	Mehwish
Supervised By:	Mr. Asim Munir
Tool Used:	MATLAB© 7.6
System Used:	HP Pavilion Entertainment PC
Operating System Used:	Windows 7
Starting Date:	3rd February, 2011
Completion Date:	18th August, 2011

Abstract

Now-a-days security has become a main concern for crowds where there is large number of persons like in playgrounds, railway stations, airports, roads etc. For groups of people this can be ensured by providing better in-depth supervision of situation through the use of multiple cameras. There are many types of movements and security task can be accomplished by detecting these types. Many techniques have been developed for this purpose. But most of them worked for small number of persons especially indoor, and those which were developed for crowds those also work in limited circumstances.

In our work large group of people is major concern particularly sports crowd where security is main issue. Motion of crowd is detected if it normal or abnormal and one factor of behavior that is peaceful and violent is also considered. Bayesian classifier is used which effortlessly classifies videos into normal and abnormal. Moreover videos' behavior is detected through root mean square technique. Results are evaluated through experimentation.

The aim of this research is to determine better technology for the detection of sports crowd's activity in terms of differentiating violent and peaceful motion along with classification of normal and abnormal motion.

TABLE OF CONTENTS

Chapter No	Contents	Page No

List of Figures

List of Tables

Chapter 1
Introduction

1 Introduction

Video surveillance systems are of great significance and increasingly being used in security systems and with strong motivation. Even the presence of video cameras at different places can help to reduce the number of crimes. Possible mishaps can be predicted and avoided or even if they occur then those can be stopped at an early stage. Appropriate and rapid actions may enable us to reduce damage. Given the ever-growing sources of danger and increasing losses, video surveillance gets an important place by providing the source of improving the security of people, buildings, goods, possessions and valuables.

In recent era many different techniques have been implemented to model and detect different types of human activities and interactions. Most of these techniques and models are designed for simple and usually indoor environments such as kitchens [1] , cargo bays [2], offices [3], and loading docks [4], old houses. In these situations activity recognition is mainly focusing on modeling the actions of only one person or small crowds. However, there are also few techniques to model larger crowds. To analysis the movements of crowd and behavior gains special interest when we are planning to build up a system for security purpose in any surveillance environment.

If we have to build an activity recognition or abnormality detection model in a crowd then we must keep in mind that it should be able to handle large variations, intensities and motion of real crowd. For this we require a huge amount of data for learning purpose. To classify different events of crowd, degree of similarity should be modeled between trained model and new video.

Different techniques were studied for motion detection and it is known that most of researches are dealing with small number of objects/persons. Only a limited work is carried out for large group of persons and huge crowds. Moreover those research efforts deal with only classification of normal and abnormal movements of crowd and no other factors are considered.

Hence presented work is divided into two main parts.

- Classification of normal and abnormal events
- Behavior Detection

In classification it is detected that crowd motion is normal or abnormal while in behavior detection it is detected whether that abnormal motion is peaceful or violent.

Bayesian classifier is used to classify normal and abnormal events. Root Mean Square Error (RMSE) is used for behavior detection of crowds which shows whether the crowd is peaceful or violent. Proposed approach not only deals with large number of objects but it also tells about their behavior, moreover it is fairly simple and robust. As in previous researches mostly Hidden Markov Model is used for abnormality detection which requires a huge amount if training data which is very hard to analyze and implement while it is not so in case of Bayesian classification.

1.2 Activity Recognition

Activity recognition comes under the identification of moving objects in some scene while considering environmental conditions. This is very important filed of research which has gained attention of many researchers since 1980 due to its significance regarding security purposes also in many other different fields.

Activity recognition can be used in many scenarios. Consider the following simple one in which there is an old man who lives alone in some apartment. There he performs simple daily tasks like sequence of tasks while making breakfast and then taking medicines. He is monitored continuously. At the end of the day his daughter or some other close person has access to secure website which has link with that person's apartment through sensor network. Through this access and scanning she finds that her father is having his food normally and taking medicines regularly which can put her at ease even while being away from him. Similarly an alarm can be generated if that man does something abnormal which is not there in his routine, so timely help can be provided.

Many different applications have been developed by researchers in field of activity recognition which are ensuring security and other facilities.

Types of Activity Recognition

Sensor-based, single-user activity recognition

This type combines the sensor networks area with data mining and machine learning techniques model large collection of human activities. [5][6] Mobile devices give enough data as well as calculation power to make physical activity recognition possible. This gives evaluation of energy consumption during daily life. The researchers of this field believe that if we allow computers and sensors for monitoring the behavior of agents then they can act well on our behalf.

Sensor-based, multi-user activity recognition

This type of activity recognition which recognizes activities of multiple users first appeared in work done by Olivetti Research Ltd (ORL). ORL used active badge systems in 90's. [7] for recognition of group activity patterns such as in office environment acceleration sensors were used.[8] Research on the recognition of multiple users in intelligent environment was done by Gu et al. [9] They worked in home environment to recognize activity of multiple users by use of sensor reading. Later they gave a pattern mining approach which gave results for both single and multiple user activities and combined them in a single solution. This can initiate many interesting research topics.

Vision-based activity recognition

This is a challenging type of activity recognition. In this behavior of agents is observed and understand which is taken from different cameras. This type has many applications in different environments and systems. Work done on vision based activity recognition appears in scientific conferences such as ICCV (international Conference on Computer Vision) and CVPR (Computer vision and Pattern Recognition).

1.3 Organization of Report

The rest of the thesis is organized as follows:

Chapter 2 is titled as "Motion Detection" in which different terms regarding motion detection are discussed in details. Also different techniques that are used for motion detection are also described.

Chapter 3 is titled as "Literature Survey" We include different research papers describing Motion Detection techniques. After summarizing these research papers we have identified the problems in the areas of Motion Detection.

Chapter 4 is named "Research Methodology" that we adopted for our work. This chapter explains our proposed technique in detail.

Chapter 5 is named as "Experiments and Results". Different sets of data videos are used for testing. And results of six of them are shown in this chapter.

Chapter 6 is titled as "Conclusion and Future Work" and References are given at the end.

Chapter 2
Motion Detection

2.1 Motion Detection

Motion detection is the identification of scene change. For motion detection it is checked how objects physically move from one state to another. For motions detection many techniques have been developed. Extraction of moving objects from scene is an important part of almost every application.

Video surveillance system gained an important place in security systems so many techniques have been developed for that, in a continuous video stream. Most of them are using very common approach of frame comparison in which current video frame is compared with previous frame. Frame comparison technique is helpful in video compression systems where only those changes are estimated and written which are not needed to be done for whole frames. But for motion detection applications it is not a desirable approach.

Another approach is to compare the current frame with the first frame in the video sequence. This comparison gives the moving objects in video stream if there were no object in first frame. But this approach does not give good results where background is not static. If there is an object in first frame but not in some others then this technique gives errors as motion is detected where ever there is some object. First frame can be renewed in some cases but still there are some situations in which we are not sure that first frame has no object and in these cases this technique fails.

There is another technique in which a frame is built which is called background of the scene and then this background is compared with each frame. Many approaches use this technique due to its efficiency but most of them are very complicated.

2.1.1 Motion Alarm

Motion alarms can easily be added to any motion detection algorithm. Almost all of the approaches use most efficient technique of comparison of current frame with background. Each algorithm produces binary image which contains the difference of current frame and background. So, motion alarm is just a step away from this as the only thing that we need to add motion alarm in our technique is to calculate the total number of white pixels in difference frame.

It is an easy and simple task to add motion alarm to our approach but in some approaches it is achieved even in simpler way. An example of this is blob counting approach in which area of detected objects is also computed along with white pixel calculation. An alarm event is activated if this computed amount is greater than the predefined amount.

2.2 Video Camera [10]

A **video camera** is used to capture motion pictures. It was first developed for television industry but now a days these are used in other applications also. John Logie Baird was the developer of the earliest video cameras. These cameras were based on electromechanical Nipkow disk. Until 1980's all electronic designs were based on cathode ray tube technology which has common problem of image burn-in. Later on this tube was replaced with solid state image sensor such as CCDs and later with CMOS. These new technologies removed drawbacks of tube technology.

There are two ways to use video cameras.

In early broadcasting, cameras were used for live television transmission where cameras capture real time images and display them directly on screen, so observation was done immediately. Same cameras are still being used which give live production for television but most of them are used for security purposes.

In second way images are recorded and stored on storage devices. This saved record is used for further processing. There are different medias to store information. In early days video tapes were used for this purpose. Later on technology advanced and many other devices were introduced like optical discs, hard disks, flash memory etc. Recorded videos are used in television, film industry and also in surveillance system where recording is needed for further processing.

Modern video cameras have been modified in different designs and uses which are totally different from early television cameras. Here are the few types of video cameras.

- **Camcorders**: These are mobile cameras. Many devices are combined in this including VCR. These are widely used for personal stuff in homes, parties, electronic news gathering and related applications. Pocket video cameras is example of this type.
- **Closed-circuit television (CCTV):** These cameras are used for security purposes. As these are used for security so while designing certain things are kept in mind like its size which should b very small to hide it easily, secondly it is made to be operated automatically without need of any operator. Other ones are used for scientific and industrial purposes. These are usually not accessible to human as they are used in critical environment where human survival is impossible like radiation, heat, chemical exposure etc
- Webcams are video cameras used for online communication in computers. Larger video cameras can also be used in the same way but they need analog-to-digital converter which helps to store information on some storage media and then to send it to network.
- In recent century mostly digital cameras are used which converts signals directly from analog to digital. These cameras are usually smaller in size even smaller than security cameras, These can also be used as webcams or stationary cameras. Most of them are incorporated in some hardware like computers, mobile phones.
- There are some special purpose system also like in scientific research. For example in space or in artificial intelligence, in medical or robotics research. These are designed according to the requirements like for night vision and heat sensing these should work where there is non visible radiation for infrared photography and in medical for X-ray.

2.2.1 Moving vs Static Cameras

Cameras used for any kind of footage can be of two types in two different situations. One in which camera is moving and other in which camera is fixed and only scene in moving.

Static camera is fixed usually on some stand or tripod. These can also be hand held. The basic principal is that there is a fixed vision field and camera is stationary. And if there is zoom in some footage and camera is fixed then it is still considered static.

On the other hand moving cameras can also be hand held but mostly they are on carts, moving arms etc. these work on the principle of moving frame including panning.

2.3 Supervised Learning

Supervised learning comes under machine learning. The aim of this task is to deduce function from supervised training data which consists of set of training examples. Each example of supervised learning has an input object and desired output value. Input object is typically a vector and output value is supervisory signal.

A supervised learning algorithm works as follows

It analyzes training data.

Produces inferred function.

Later this function (also called classifier) produces correct output if we provide it with any valid input. So a training data is set which is used by learning algorithm to generalize. This generalization on the basis of training data works for any unseen situation in a reasonable way. Same task can be done in human as well as animal psychology. It is called concept learning in human and animal psychology.

2.4 Unsupervised Feature Selection:

In unsupervised learning unsupervised classification or clustering is done. The main aim of cluster analysis is to put same objects in one group based on some similarity measure. So natural grouping is done. [11]

In unsupervised learning a criteria is set and clustering is done according to this. So subset of features is selected which is main objective of unsupervised learning. [12]

Supervised vs. unsupervised learning [13]

If we see the difference between supervised and unsupervised learning theoretically then we would come to know that there is difference only in casual structure of model. The model of supervised learning defines the effect of inputs over outputs where inputs are on one end and outputs are on other end of casual chain.

In unsupervised learning it is assumed that variables are latent. It means that observations are at the end of causal chain. In supervised learning inputs are undefined. So if inputs are available then supervised model is not needed. On the other hand if inputs are missing then it is impossible to conclude anything about outputs so supervised model is trained. If some of the inputs are missing then we can model other available inputs, so missing inputs can b considered as latent variables so they cause no problem.

In supervised modeling output is produced by input observations and on the other hand in unsupervised learning it is assumed that set of latent are cause of all observations.

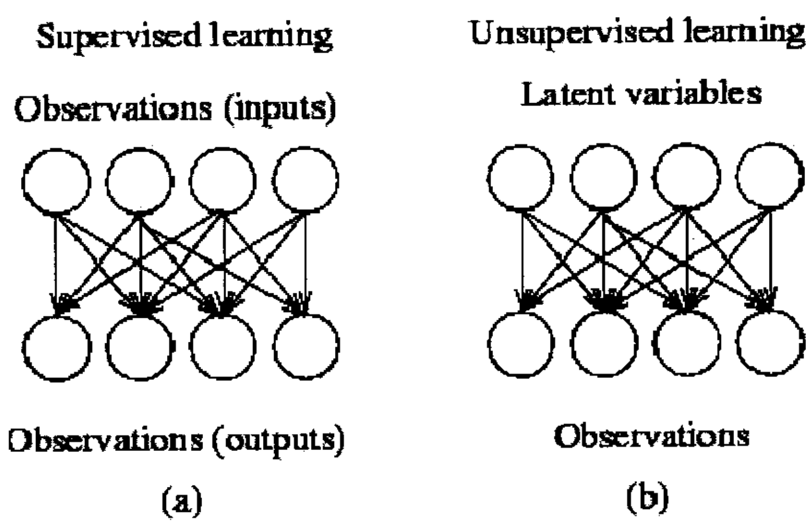


Figure 2-1: The causal structure of (a) supervised and (b) unsupervised learning

In supervised learning there are two sets of observations and we have to find out connection between them. So if there are large and complex models then learning is not easy with supervised modeling. If there are more number of steps then complexity increases. So if there are models with deep hierarchies then unsupervised learning is easy to implement.

2.5 Semi-supervised Learning

Semi-supervised learning falls between supervised and unsupervised learning. Here both types of data are used that is labeled as well as unlabeled. In semi supervised technique unlabeled data is used in large amount while labeled data is used in small amount which makes it better classifier. [14]

It is hard to obtain labeled data moreover it is expensive and time consuming to get it. Besides it need human effort. While unlabeled data is easy to obtain but it can't be used in every situation. These problems are removed in semi supervised learning as small amount of labeled data is used in it along with large amount of unlabeled data. So less human effort is required in it which gives high accuracy. [15] It is widely used in abnormality detection algorithms where unusual event can't be predicted in advance.

2.6 Optical Flow

Optical flow is noticeable motion of objects in an image frame with respect to previous frame. Optical flow is usually associated with motion field but not always. Example where motion filed and optical flow has no correspondence is barber's pole; these two are different in that which is shown in figure 2-2.

Generally these cases are unusual and mostly optical flow corresponds to the motion field.

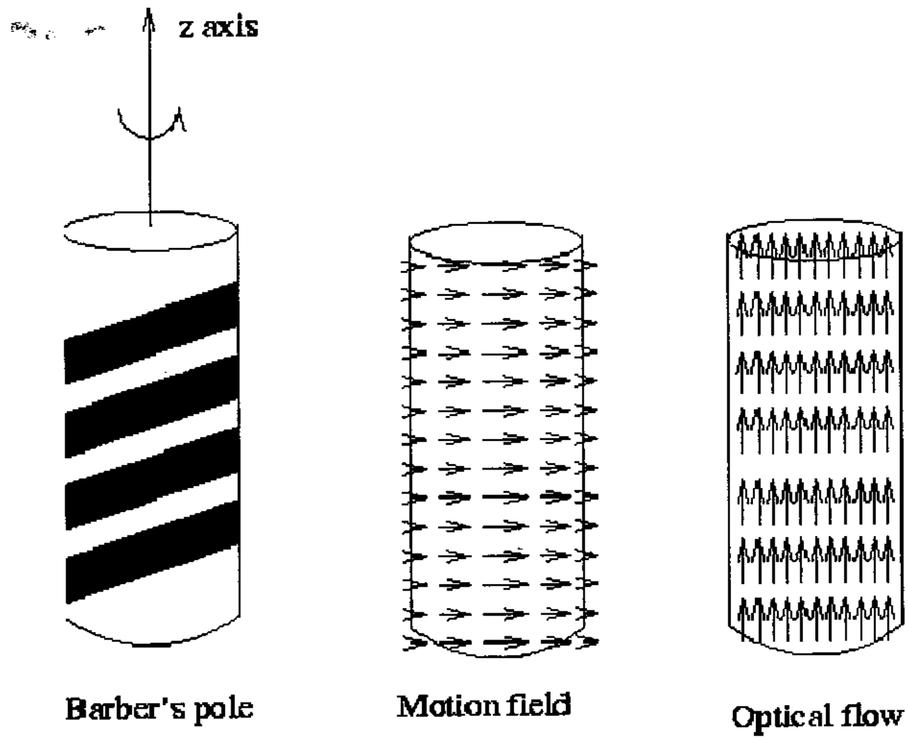


Figure 2-2: The motion field and optical flow of a barber's pole

While calculating optical flow a problem is faced called aperture problem. It is because of difficulty in calculating the component which is tangential to the intensity gradient. Only the optical flow component can be calculated which is in the direction of intensity gradient.

This problem is shown in figure 2-3, and further calculations are given below.

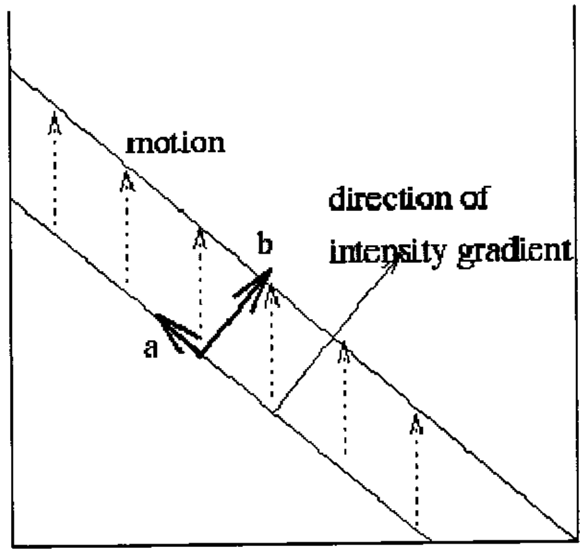


Figure 2-3: **The aperture problem. Only the component b can be measured.**

Intensity is denoted by $I(x,y,t)$. This is a function of three variables and first of all the change in I is noted. For this I is differentiated with respect to t:

$$\frac{dI}{dt} = \frac{\partial I}{\partial x}\frac{dx}{dt} + \frac{\partial I}{\partial y}\frac{dy}{dt} + \frac{\partial I}{\partial t} \qquad (1)$$

While calculating optical flow illumination tolerance is needed as it affects intensity. And it is hard to know whether the change in intensity is due to illumination or movement of objects. So we assume that there is no change in image intensity over time, i.e there is no change in illumination and shadows etc, so:

$$\frac{dI}{dt} = 0 \qquad (2)$$

which implies:

$$I_x u + I_y v + I_t = 0 \tag{3}$$

Where subscripts are the partial derivatives of I, and u and v are the x and y components of the optical flow vector.

This equation is called the *optical flow constraint equation.* As in this equation a constraint on the optical flow components is expressed.

This equation can be rewritten as:

$$(I_x, I_y) \cdot (\mathrm{u,v}) = - I_t \tag{4}$$

The component of the image velocity can be shown now in the direction of the image intensity gradient. This is shown in following equation:

$$(u, v) = \frac{-It}{\sqrt{Ix^2 + Iy^2}} \tag{5}$$

An ambiguity called Aperture problem can be there that is optical flow component at right angles to this direction cannot be determined.

Optical flow can be two dimensional as well as three dimensional. There are two main approaches compute 3D optical flow.

In first approach motion problem is converted into stereo problem. Then correspondence between number of points in image is found. And in second approach first optical flow is computed and then its geometrical properties are used to deduce 3D information.

2.7 Root Mean Square Error (RMSE)

Root Mean Square Error which is abbreviated as RMSE is basically the difference between two objects. Those are predicted by a model or an estimator and the ones actually observed from the thing being modeled or estimated.

RMSE is ideal when it is small as it is a good measure of accuracy.

RMSE is the other name of most familiar and popular term that is standard deviation, which can be calculated with help of following formula:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(xi - x)^2}{n-1}} \tag{6}$$

It is important to know the difference between RMSE and Range methods for finding the standard deviation which we talking about Statistical Process Control.

RMSE serves as a quadratic scoring rule which can be used to determine the average magnitude of error. Both the references do contain the equation of RMSE. If we want to say the formula in words we can go as it is the difference between the expected and corresponding observed values, both of them squared and then averaged over the whole sample. In the end just take the square root of the average. From the description it is clear that while calculating RMSE errors are first squared and then their average is taken, so it gives high weight to large values. This concludes that RMSE offers best results when large errors are not needed.

Both MAE (Mean Absolute Error) and RMSE (Root Mean Square Error) can be used at the same time to calculate the change in error in a set of forecast. MAE will always be lesser or equal to RMSE; the larger difference between them, individual error in sample will have larger variance. All the errors will be of same magnitude if both MAE and RMSE are equal.

Their value may range from zero to infinity. They are negatively oriented scores hence lower values are better.

2.7.1 Mean Absolute Error (MAE)

MAE is independent of the direction of errors as it measures the average magnitude of the error in a set of forecasts. Accuracy of continuous variables can be calculated by it. Its definition in words can be that MAE is the average over the verification sample of the absolute values of the differences between the expected and the corresponding observed values. All the differences are weighted equally in the average because MAE is a linear score.

Applications:

- It is used in the Protein nuclear magnetic resonance spectroscopy. RMSD (Root Mean Square Deviation) is now used for the estimation of the quality of the bundle of structures which are obtained.
- RMSD is used to judge submission of the Netflix prize from the test dataset's undisclosed "true" values.
- The RMSE and its extension coefficient of variation of the RMSD, CV(RMSE) are used to calibrate models to have an idea of building performance in simulation of energy consumption of buildings.
- RMSD is also used in meteorology to see the effectiveness of mathematical model prediction about the behavior of the atmosphere.
- Average distance between the atoms of superimposed proteins is called RMSD in bioinformatics.
- RMSD is the measure of the distance between a crystal structure conformation and a docking result in cheminformatics.

- It is also used to check whether an economic model fits economic indicators in economics.
- RMSD is used to check how good model or perception explains the abilities of the human senses in experimental psychology.
- Assessing the accuracy and precision of spatial analysis and remote sensing is also done by RMSD in GIS.
- RNSD and NRMSD is also used excessively to evaluate the calibration of a groundwater model in hydrogeology.
- Signal to Noise ratio is used to analyze how correctly an image can be reconstructed from the original image in imaging science. RMSD is part of the Peak Signal to Noise Ratio.
- It is also used in computational neuroscience to check how well a system learns a given model.

2.8 Statistical Classification

In statistical classification a training data is set which contains observations with known sub-population and on its basis an unknown sub-population is identified with new observations. Such type of classification shows variable behavior so we need statistics for its study.

Thus measurements are made new and individuals are grouped on these measurements and this grouping is based on quantitative information.

Grouping is done in cluster analysis too. Same objects are placed in same group. But it differs from statistical analysis as single data set is analyzed in cluster analysis and then it is observed that whether it can be divided into groups or not on the basis of similarity. Clustering comes under unsupervised learning while classification is supervised learning.

2.8.1 Naive Bayes Classifier

A Naive Bayes Classifier is probabilistic classifier. It is based on simple assumption that presence or absence of one feature of class is not related to other. For example if a fruit is orange on color, round and 4 inches in diameter then it is considered to be an orange. So if these features depend on each other or not or on some other features naïve bayes classifier considers all fruits with above three features as orange.

This classifier can easily be trained where there is supervised learning setting.

Assumptions used in this naïve design are very simple so this classifier works efficiently in complex situations.

Naïve Bayes classifier needs small amount of training data so it is easy to collect and manage that. Entire covariance matrix is not needed to determine in classification because independent variables are assumed so variance for each class is required only.

Chapter 3
Literature Survey

3 Literature Survey

A study over the topic was carried out focusing activity recognition and abnormality detection. Most relevant are discussed below.

3.1 Real-time Crowd Motion Analysis- 2008 [16]

In this paper abnormality detection of crowd is done by motion analysis rather tracking objects. Nacim Ihaddadene and Chabane Djeraba first find out regions of interest and then from those selected regions points of interest are extracted. Average motion direction for each region is computed and formed a direction map on its basis. If there is an abnormal situation like collapse then distance between average direction histogram and instant histogram corresponding to current frame increases. Entropy is measured in next step to analyze the optical flow vectors. In last step normal and abnormal video are categorized using a threshold.

Example of collapsing situation on an escalator exit is shown in following figure. Red part of curve shows the situation where collapse actually occurred.

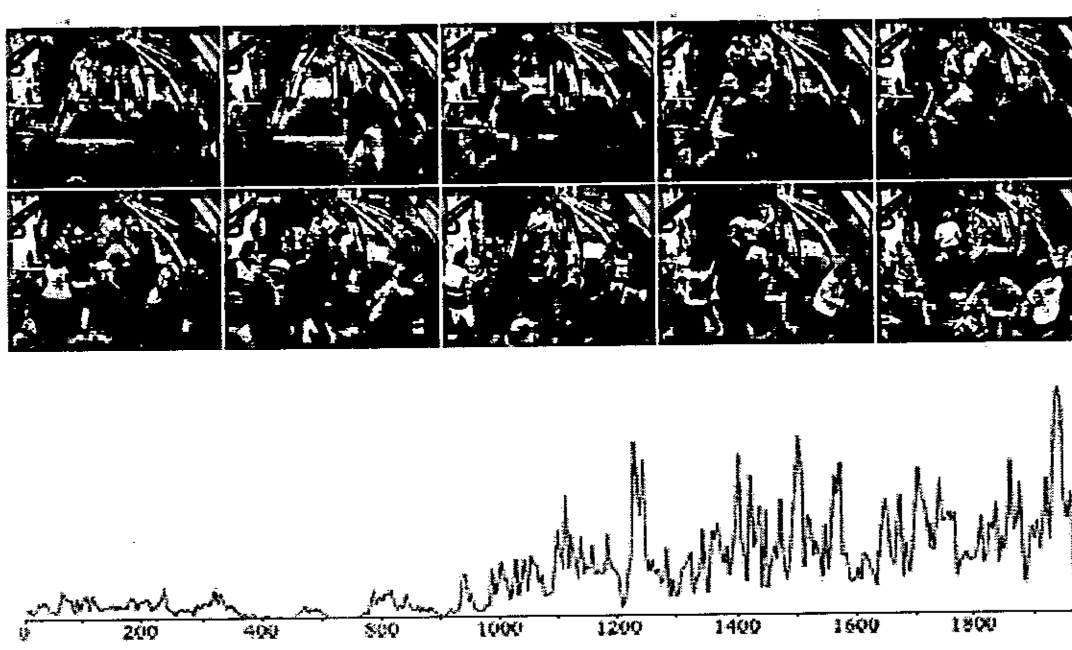


Figure 3-1: Measuring Variation

Advantages

- Large amount of data is not required as training is not done like previous approaches.
- Regions of interest are detected where variations are more important.
- Gives robust and promising results.

Limitations

- There is need to introduce context information to optimize system configuration.
- Point of interests need to be tracked over multiple frames to extend motion estimation.

3.2 Real-time abnormal motion detection in surveillance video - 2008 [17]

Following are the key ideas of this paper described by Nahum Kiryati et al:

- Avoid segmentation or tracking.
- Use the macro-block motion vectors that are generated anyway as part of standard video compression methods.
- Derive motion features from the motion vectors.
- Estimate the statistical distributions of the motion feature vectors that characterize normal activity during training.
- Unlikely feature vectors during online operation indicate abnormal motion.

Video is compressed eliminating spatial & temporal redundancy unlike common compression techniques. Video is converted to motion vectors. From these motion vectors motions features are derived.

Authors set a threshold to define the unlikely feature vectors (histogram cells). The feature vector associated with each of the incoming frames at the operational phase is computed. The system detects abnormal motion event after consecutive unlikely feature vectors.

Advantages

- Computationally efficient (75 CIF frames/sec on 2.8GHz PC)
- Its gives a modular approach as different features vectors and modeling methods can be used.
- Operates on the compressed video stream
- Abnormal motion is not associated with a particular object in the scene
- Possible extensions – relate the concepts of normal and abnormal motion with time and causality.

Limitation

- While using manually designed features it is impossible to train the system on a comprehensive set of abnormal motion patterns for every possible scene.

3.3 Modelling Crowd Scenes for Event Detection - 2006 [18]

In this paper the Ernesto L. Andrade et al presented a technique which detects abnormal events in crowds using unsupervised features extraction method.

For this first of all preprocessing is done in which background modeling is performed and optical flow is computed. Second step is calculation of feature prototypes in which principal component analysis is performed on optical flow field of all frames. In spectral clustering Multiple Observation Hidden Markov Model is trained. The measure of similarity between video segments is defined as:

$$S_{ij} = \tfrac{1}{2} \{\log P(W_j/B_i) + \log P(W_i/B_j)\} \text{ [18]} \qquad (1)$$

Similarity values form a similarity matrix S.

Then HMM training is done. Final model for video sequence has the form:

$$P(W/(M) = \sum_{k=1}^{K} \frac{N_k}{N} P(W/M_k) \qquad (2)$$

At the end normal and abnormal events classification is done. Video segment W^o_k is abnormal if

$$P(W^o_k|M) < Th_{Ab} \tag{3}$$

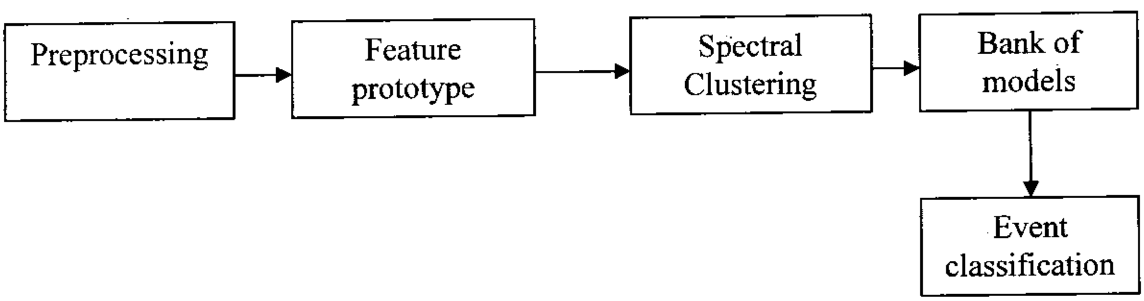


Figure 3-2: Block diagram for overall process

Advantages

- Spectral clustering is applied which automatically identifies the optimal number of models.
- The proposed technique not only detects events but also emergency situations in crowds.

Limitation

- In proposed approach analysis is applied to the whole frame that is why it is unable to detect small variations so local analysis should be done.

3.4 Semi-supervised Adapted HMMs for Unusual Event Detection [19]

In this paper Dong Zhang et al used semi-supervised adapted Hidden Markov Model framework for event detection. Dong Zhang et al used supervised learning for usual events and unsupervised learning for unusual events.

Framework for iterative adapted HMM is as follows:

1. First of all a usual event model is projected in which there is a large amount of training data.
2. Then segmentation is done and outlier is detected which is segment that has lowest likelihood.
3. Using this outlier adaptation technique i.e. Maximum a posteriori (MAP) is used, which makes training data sufficient.
4. Viterbi algorithm is used to decode test sequence.
5. Outlier detection is performed again.
6. Repeat step 3, 4, 5.
7. Stop when iterations complete.

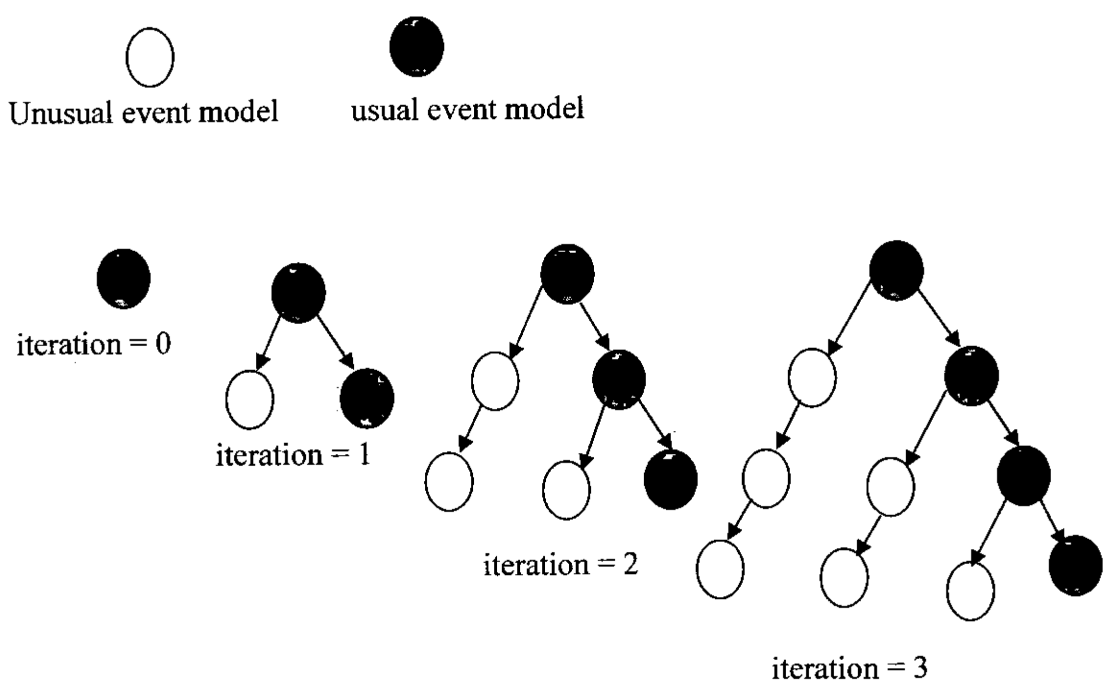


Figure 3-3: Illustration of the algorithm flow

Advantages

- Proposed technique deals with all types of unusual events.
- All types of data sets are used i.e. Audio events, Visual events and Audio-Visual events.

- Performance measures are used to check the effectiveness of framework.

Limitations

- There should be some criteria to optimize the number of iterations.
- While using semi-supervised approach feature selection is significant problem so feature selection process is needed to be improved.

3.5 Video Behaviour Profiling and Abnormality Detection without Manual Labeling – 2005 [20]

In this paper Tao Xiang and Shaogang Gong adopted discrete scene event based image feature extraction approach. Such type of technique claims to work well with noisy scenes.

First of all grouping of behavior pattern is done using eigenvectors. Then irrelevant eigenvectors are eliminated to do natural clustering. Video is divided into small segments having fixed time duration using online segmentation algorithm. Then foreground is extracted from background using Gaussian mixture background model and connected component labeling is performed to group foreground pixels. Each behavior pattern which is used in training set is modeled using Dynamic Bayesian Network (DBN). Affinity between two behavior patterns is computed as:

$$S_{ij} = \tfrac{1}{2} \{1/T_j \log P(P_j/B_i) + 1/T_j \log P(P_i/B_j)\} \; [20] \qquad (4)$$

Then multi-observation Hidden Markov Model is trained and on its basis final model is prepared. Then some threshold is defined and if value of final model is less than that threshold then that event is classified as abnormal event.

Experimental results showed that behavior model trained using unlabeled data is superior to labeled data training specially while working with unseen videos.

Advantages

- Works effectively even with noisy datasets.
- Unlabeled dataset is used which is easy to collect.
- The technique is simple and robust.

Limitations

- Needs large number of dataset for training purpose.
- Proposed technique cannot deal with unseen behavior patterns either normal or abnormal.

3.6 Activity Recognition and Abnormality Detection with the Switching Hidden Semi-Markov Model – 2005 [21]

In this paper Thi V. Duong et al dealt with an important research problem which is faced in some restrained environment that is human activity recognition of daily living. The main idea behind this application was to build a system for aged people which could do automatic monitoring and supporting for them. For this system needs to be able to do two things. First to learn the daily activities of its occupants and second to monitor them using this learned knowledge and to detect if there is some variation from usual activities and aware the caretaker when required..

Two layered model is introduced i.e Switching Hidden semi Markov Model (S-HSMM), its bottom layer is sequence of HSMM and top layer is Markov sequence of switching variables and these variables determine the parameters of HSMM so that it can be switched time to time. They considered sequence of six high-level activities which are performed in morning. The kitchen is divided into 28 square cells. There are some cells in which persons spend some time like taking something from fridge or cupboard, bottom layer captures these atomic activities. High-level activities such as making and eating breakfast are there in top layer.

The system is then provided with S-HSMM which consists of two models, one for duration and other for parameterization. This model is then converted into equivalent dynamic Bayesian network (DBN). Then this S-HSMM is applied to detect abnormalities in low-level activities' duration and to recognize high level activities. Its performance is compared with two existing HMM which are without activity hierarchy information and duration modeling. The performance of proposed model illustrates that both hierarchy and duration information are necessary for activity recognition of daily living.

Advantages

- Gives best results for simple indoor activities.
- The combination of duration and parameterization models surpass the results of hidden markov models without hierarchy information and duration modeling.
- Proposed model gives better computation time.

Limitation

- This technique is only for simple domains and does not work well for the complex ones.

3.7 Investigation into Optical Flow Super-Resolution for Surveillance Applications - 2005 [22]

This paper proposes a solution to overcome limitations of existing Super-resolution techniques for face identification. As proposed solution is based on optical flow so it is capable of resolving problems that occur due to occlusion.

Following are the main steps of proposed solution:

1. First of all bilinear interpolation is used to interpolate the original image to twice of the input resolution.
2. Odd number of frames (N) are taken and optical flow is computed between current reference image and (N -1) / 2 previous and (N-1) / 2 following image.
3. Register the previous and following image to the reference image.
4. Fusion technique is computed across reference and registered images and by using this super-resolution image is estimated.
5. Final image is restored by applying Wiener deconvolution filter.

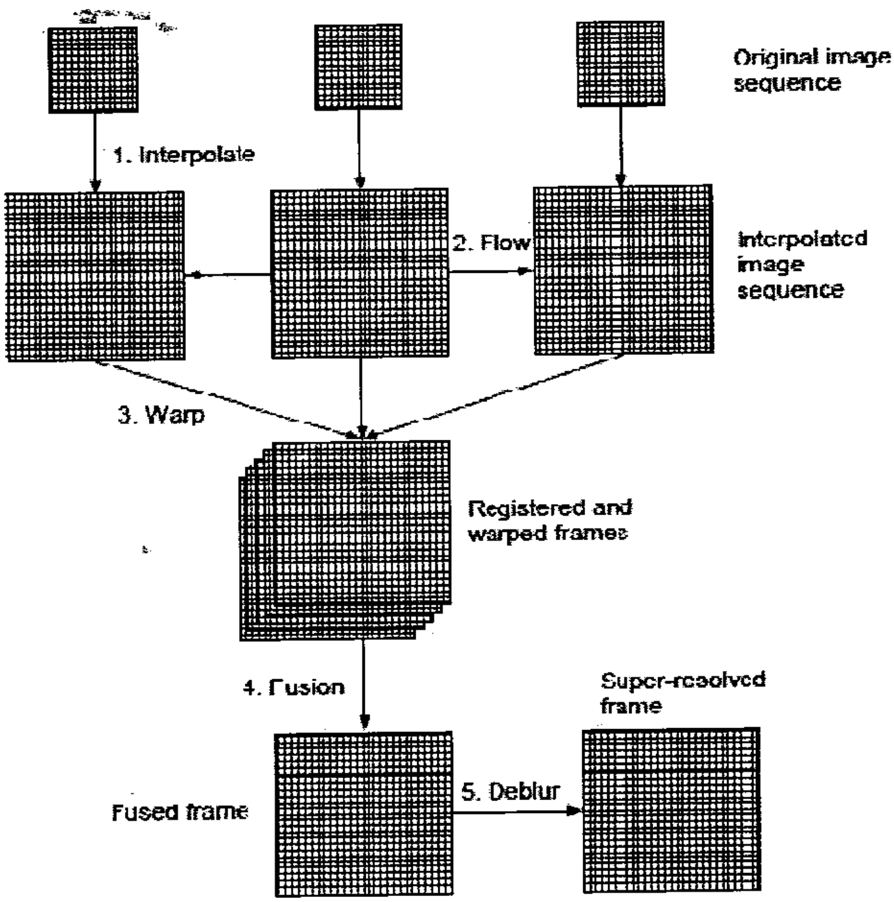


Figure 3-4: Super-Resolution system flow diagram

Advantages

- Proposed system is especially useful for surveillance applications
- Overcomes limitations of existing SR techniques regarding face images.
- Overcomes many problems of face images specially non-planarity and non-rigidity of the face by use of optical flow.
- Robust estimation methods are used to overcome occlusions and illumination variation problems.

Limitations

- The technique does not work well when motion between two input frames is too large.
- Super Resolution (SR) estimations are needed to be refined.

3.8 Learning Object Trajectory Patterns by Spectral Clustering - 2004 [23]

In this paper Fatih Porikli proposed a method for learning of trajectory patterns which is based on the disintegration of eigenvectors of affinity matrices.

Flow diagram of the process of event detection is below [23]

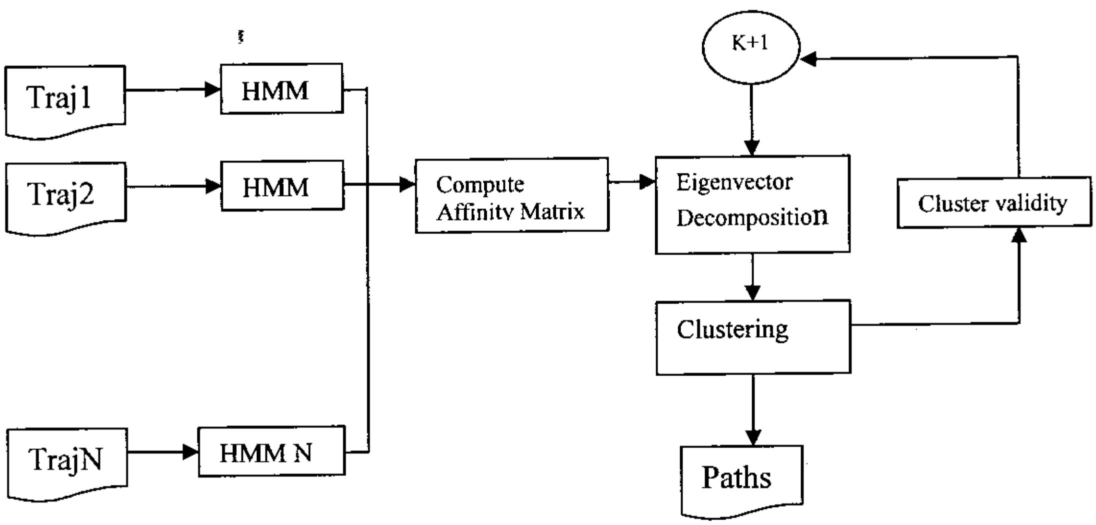


Figure 3-5: Flow diagram of spectral Clustering

Fatih Porikli first trained HMM for each trajectory .Then affinity matrix is computed. This affinity matrix is decomposed and largest eigenvectors are found. These eigenvectors help to find clusters. And thus cluster validity score is used to automatically compute optimal number of clusters.

Eigenvector decomposition is performed which is used in clustering. Author showed that number of eigenvectors and clusters is proportional to each other.

After the computation of eigenvalue a validity score is computed as:

$$\alpha_k = \sum_c^k \frac{1}{N_c} \sum_{i,j \in Z_c} \rho_{ij} \tag{5}$$

Advantages

- Normalization problems of previous methods are solved by using HMM for each trajectory.
- The proposed technique presents the relationship between the number of eigenvectors and clusters which helps in selecting the correct number of eigenvectors and also correct number of clusters is determined automatically.
- If the length of all trajectories is same then computational complexity of proposed technique is low.

Limitation

Relationship between different types of objects and trajectories is not shown.

3.9 Detecting Unusual Activity in Video - 2004 [24]

Hua Zhong et al used unsupervised approach i.e without any labeled training data to detect abnormal motion in large video sets. For this first of all feature selection is performed. Then the video is divided into equal small segments such that each segment contains an activity. The algorithm goes as follows:

- Moving objects are extracted from each frame.
- Histogram is quantized into prototypes.
- Co-occurrence matrix is computed between each video segment and prototype feature.
- Then pairwise similarity between all prototypes is computed.
- Smallest eigenvectors are selected, first rows of eigenvectors are coordinates for video segments and following rows are coordinates of prototypes in embedding space.
- In co-embedding space inter-cluster similarity is computed. And the clusters with small similarity are identified as unusual events.

As the goal was to detect abnormal events in long videos so the authors did not used model based approach which usually gives good results when normal activity is well defined.

Advantages

- Gives best results for long real life videos.
- Works well if normal activity is not well defined.
- Large dataset is used.

Limitation

- In unsupervised techniques data collection is easy but there are few ways to use them.

3.10 Computing 2D Optical Flow

The computation of differential optical flow is, essentially, a two-step procedure:

1. Measure the spatio-temporal intensity derivatives (which is equivalent to measuring the velocities normal to the local intensity structures) and

2. Integrate normal velocities into full velocities, for example, either locally via a least squares calculation [25, 27] or globally via a regularization [26, 27].

The basis of differential optical flow is the motion constraint equation. Assume I(x, y, t) is the center pixel in a n×n neighborhood and moves by δx, δy in time δt to I(x+ δx, y+ δy, t+ δt).

Since I(x; y; t) and I(x + δx, y + δy, t + δt) are the images of the same point (and therefore the same) we have:

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t) \qquad (6)$$

TH-8849

A 1st order Taylor series expansion about I(x; y; t) in equation (6) can be performed to obtain:

$$I(x + \delta x, y + \delta y, t + \delta t) = I(x, y, t) + \frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t + \mathrm{H.O.T.}; \qquad (7)$$

Above equation can be simplified as:

$$(Ix , Iy) \cdot (Vx , Vy) = - It \tag{8}$$

$$\nabla I . \underset{v}{\rightarrow} = - It \tag{9}$$

The velocity with smallest magnitude is Vn

Intensity derivatives Ix, Iy and It as:

$$\underset{v_n}{\rightarrow} = v_n\hat{n} = \frac{-I_t(I_x, I_y)}{\|\nabla I\|_2^2} \tag{10}$$

2D motion constraint equation can be rewritten as:

$$\underset{v}{\rightarrow} . \hat{n} = v_n \tag{11}$$

Which is equivalent to eq (9) [28]

After finding out the motion constraint equation any of the optical flow algorithms will be applied.

3.11 Problem Identification

Motion detection is a process of scene detection. Many approaches have been developed for this and main aim of all of them is to identify abnormal motion. Due to its importance regarding security purposes, it has become an important part of all video surveillance systems. Hence there is always a need to develop mechanism to ensure security of large crowds and also the identification of crowd behavior. In our work we found RMSE technique to be more effective to categorize videos' behavior and also Bayesian classifier effortlessly classifies videos into normal and abnormal. Experimental results show that our proposed technique gives efficient results. Also computation complexity is low in our method as compared to any other scheme in literature survey.

Chapter 4
Research Methodology

4 Research Methodology

Previously done research focused on the change in motion and abnormal events. Like in [19] an unusual event model is derived from the usual event model at each step of an iterative process, while [12] automatically detects the abnormal events in crowd. During study of previous research works it is found that those only detect abnormal motion but lack the detection of peaceful and violent motion of crowd, neither it has differentiated between them.

4.1 Motion Detection

Many approaches are used for motion detection in a continuous video stream. In most of the approaches motion detection is achieved through the comparison of video frames i.e. current and previous frame or of frames with background. Moreover a difference image between a captured image frame and an estimated background image is produced by background subtraction based on intensities. This works well if the illumination for the images to be processed is constant. Any change in the illumination will significantly affect the intensities. Under such conditions, it is difficult to identify, by motion detection algorithm, whether the changes of intensities is resulted from illumination or motion of object. Therefore illumination tolerance is needed to be considered for motion detection as well as segmentation. [21]

4.1.1 Abnormal Motion

Some temporal events are present in almost every video stream which are significant and rarely occur. These events are usually named as unusual, rare or abnormal events.

Unusual events occur infrequently and cannot be expected and are related to a task. [19] Unusual events constitute problems therefore they got attention in computer vision.

4.2 Proposed Method for Motion Detection

Process of activity recognition and abnormality detection is divided in two major tasks. The first task is to classify sports crowd into normal and abnormal. The other task is to detect peaceful and violent motion and differentiate crowd's behavior into peaceful and violent.

In order to detect crowd's behavior we have proposed following methodology.

Step 1: Data Collection and Preprocessing

Get data videos randomly and convert them into avi format. Set all frames size as 320 x 232 and frame rate as 12fps. Compute optical flow of every two consecutive frames of each video to find out zones of the scene where motion intensity is high. Apply Laplacian of Gaussian filter on each frame.

Step 2: Read Video and get relevant information

Read contents of avi file and get relevant frames and other details i.e. total number of frames in video, its frame rate and duration.

Step 3: Background Separation and Objects Calculation

Take first frame of video as background and subtract each element of frames from it using the following equation:

$$\text{Diff I} = f_1 - f_i \qquad \text{Where } i = 2 \text{ to } N \tag{1}$$

Compute global threshold using graythresh function to convert frames into binary image. The syntax of graythresh is below:

$$G = \text{graythresh(diff1)} \tag{2}$$

Find total number of objects in each frame and name it as SUM.

$$\text{SUM} = \sum_{i=1}^{N} \text{Ai} \tag{3}$$

Find out motion of these objects with respect to other frames.

Step 4: Perform Root Mean Square Error (RMSE) Technique

Calculate the RMSE between first frame and frame onward second frame to last and display it.

$$RMSE = \left[\frac{1}{MN}\sum_{n=1}^{N}\sum_{m=1}^{M}\left(F1(n,m) - F2(n,m)\right)^2\right]^{1/2}$$

Where F1 and F2 are video frames and M, N are the dimensions of the frames.

Compute RMSE of each frame and take their average. Also compute difference of RMSE of every two consecutive frames and take their average.

Step 5: Bayesian Classification

Classify video into normal and abnormal classes using Bayesian Classification. Use average consecutive RMSE as a variable in Bayesian classifier. Choose threshold through training data and classify videos as if average consecutive RMSE is greater than or equal to threshold then put that video in abnormal class otherwise in Normal class.

Step 6: Behavior Detection

Take three parameters i.e. SUM, Average RMSE and Average Consecutive RMSE and use them for behavior detection.

When SUM is greater than 35000, average RMSE is greater than 35 and average consecutive RMSE is greater than or equal to 0.80 then videos are violent otherwise peaceful. These parameters are set through repetitive experimentation.

The program goes as follows:

Video change to AVI

Frame size set to 320x232

Frame rat set to 12fps

Compute Optical Flow of video

Apply LOG filter on each frame

FOR Thirty frames of video

    Compute global threshold

    Find total number of objects

    Find motion of objects

```
    END FOR

    FOR Thirty video frames
        Compute RMSE
        AverageRMSE=avg(RMSE_i )
        Average consecutive RMSE = avg (RMSE_i-1 - RMSE_i )
    END FOR

IF  Average consecutive RMSE >= Threshold
    Declare Video Abnormal
ELSE
    Declare Video Normal

Also
If ( SUM > 3500 ) and ( average RMSE > 35 ) and ( average consecutive RMSE > 0.80)
    Violent behavior
    Else
    Peaceful behavior
```

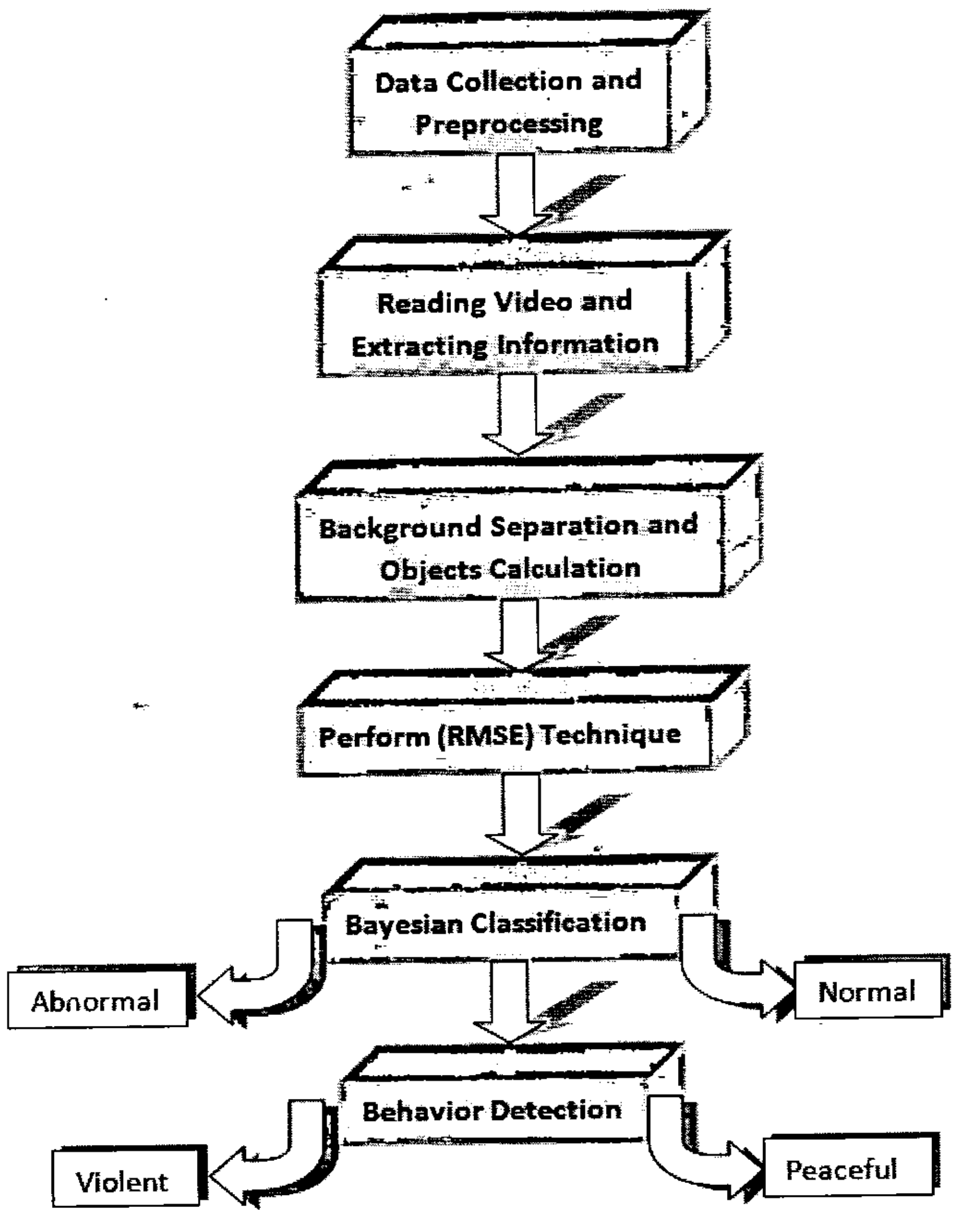


Figure 4-1: Block Diagram of Proposed Technique

Chapter 5
Experimental Results

5 Experimental Results

In our experiments we take different sets of normal and abnormal videos specifically of sports crowd that were obtained randomly.

5.1 Videos

We used several representative videos in our research including indoor, outdoor, road videos and specially sports crowd videos. Examples of few of them are as below

5.1.1 Normal videos

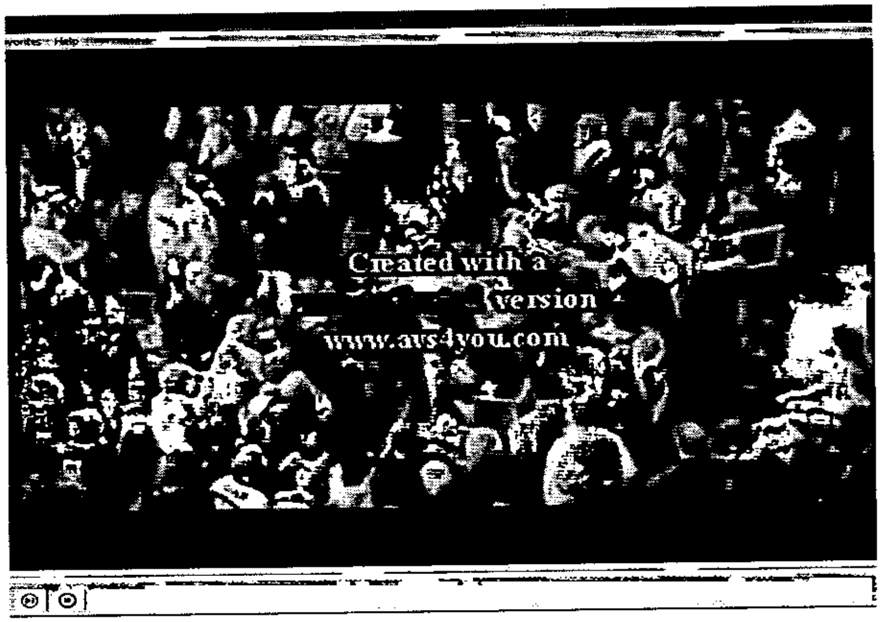
Following are examples of normal and peaceful videos.

(a) Video 1 (b) Video 2

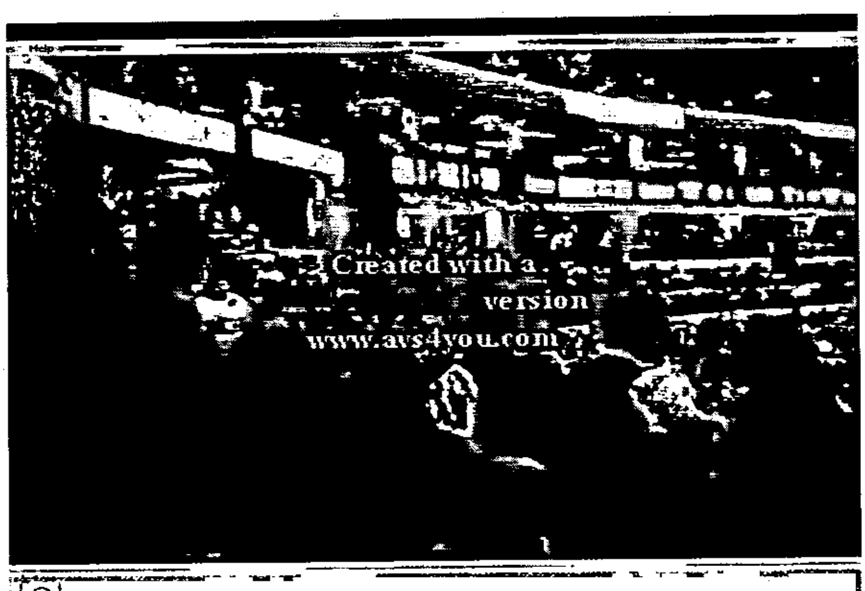
Figure 5-1: Examples of normal videos

5.1.2 Violent Videos

Following are examples of abnormal and violent videos.



(a) Video 1



(b) Video 2

Figure 5-2: Examples of violent videos

5.2 Video Frames

Examples of frames taken from two different videos presenting normal and peaceful behavior are shown in figure 5-2. A few frames detected as abnormal and violent are shown in figure 5-3.

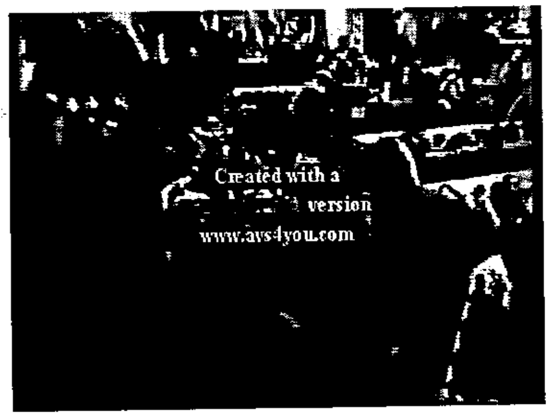
Frame 11

Frame 119

Frame 431

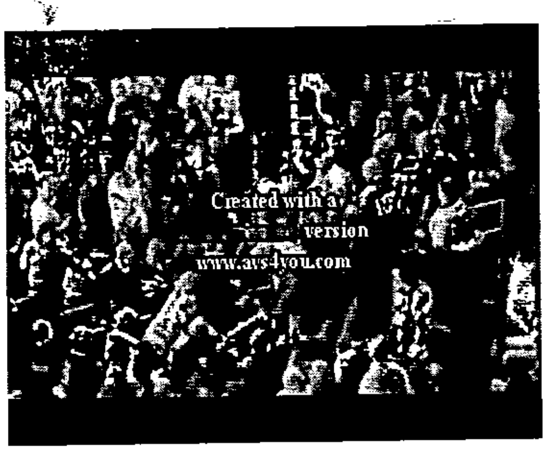
Frame 458

Figure 5-3: Examples of normal behavior

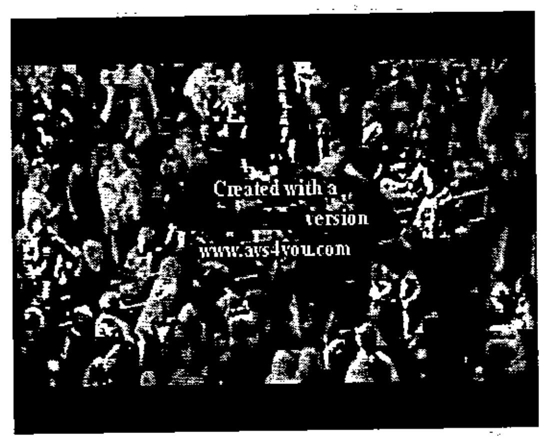


Frame 37

Frame 46



Frames 71



Frames 90

Figure 5-4: Frames taken from two crowd videos and detected as violent

5.3 Optical Flow

Optical flow of every two consecutive frames of each video is computed o determine motion flow. Results of optical flow of some frames along with relative frames is shown below

Frame 38

Frame 39

Optical Flow of Frame 38, 39

Figure 5-5: Optical Flow of two consecutive frames

5.4 Root Mean Square Error (RMSE)

Root mean square error indicates the error between frames.

Below is the illustration of the values of one video with 30 frames. It states 'X' n 'r' values whose depictions are as follows,

- "X" is RMSE of frames (30 values for 30 frames)
- "r" is difference of RMSE of every two consecutive frames

frame	1	2	3	4	5	6	7	8	9	10
x	0	12.5744	16.1254	18.7524	20.939	25.4368	26.294	28.73	31.8431	28.9003
r	0	12.5744	3.551	2.627	2.1866	4.4978	0.8572	2.436	3.1131	-2.9428

frame	11	12	13	14	15	16	17	18	19	20
x	28.3639	27.3286	25.3725	25.6669	24.9178	25.3086	25.6567	25.6351	25.3515	25.3416
r	-0.5364	-1.0353	-1.9561	0.2944	-0.7491	0.3908	0.3481	-0.0216	-0.2836	-0.0099

frame	21	22	23	24	25	26	27	28	29	30
x	25.3614	25.6049	24.1525	23.458	23.7279	24.6159	23.6994	23.9141	25.8674	24.8331
r	0.0198	0.2435	-1.4524	-0.6945	0.2699	0.888	-0.9165	0.2147	1.9533	-1.0343

Table 5-1: Illustration of RMSE values

The 'r' is calculated as the difference of 2 consecutive values and a related sample working is shown below:

frame	1	2	3	4	5	6
x	0	12.5744	16.1254	18.7524	20.939	25.4368
r		12.5744	3.551	2.627	2.1866	4.4978

Table 5-2: Calculations of RMSE values

In the above diagram, the arrows show the 2 'X' values taken to calculate a particular 'r' value of each video frame.

Proposed method produced three parameters and on the basis of these video behavior is determined as peaceful or violent. It was tested and results of six videos are shown in following table

Video	SUM	Average RMSE	Avg Consecutive RMSE	Standard Deviation	Number of Frames	Video Behavior	Status
Video1	505793	56	7.713	7.4464	1087	Violent	Truly accepted
Video2	579644	72	0.9092	8.2789	396	Violent	Truly accepted
Video3	143899	16	0.2243	3.25	602	Violent	Falsely Rejected
Video4	370195	31	0.0456	3.0125	476	Peaceful	Truly accepted
Video5	467551	50	1.7644	6.7929	1317	Violent	Truly accepted
Video6	436266	51	1.522	6.657	447	Peaceful	Falsely Rejected
Video7	515899	52	0.9632	6.34	1693	Violent	Truly accepted
Video8	130729	15	0.0101	1.1424	602	Peaceful	Truly accepted
Video9	143899	16	0.2243	1.2618	602	Peaceful	Truly accepted
Video10	333544	32	0.6641	6.9869	1317	Normal crowd	Truly accepted

Table 5-3: Results of video behavior

5.5 Bayesian Classification

Naïve Bayes Classifier is used for classification of videos into normal and abnormal as it is probabilistic model, robust to noise found in real data. Average consecutive RMSE is used as test data in model. It classified all videos as normal whose average consecutive RMSE value is less than threshold.

Results of classification for few videos are shown in the table below:

Videos	Avg Consecutive RMSE	Class	Status
Video 1	0.6641	Normal	Truly Accepted
Video 2	1.7644	Abnormal	Truly Accepted
Video 3	1.3937	Abnormal	Truly Accepted
Video 4	1.522	Abnormal	Falsely Rejected
Video 5	0.2243	Normal	Truly Accepted
Video 6	0.9632	Abnormal	Truly Accepted
Video 7	1.0343	Abnormal	Truly Accepted
Video 8	0.317	Normal	Falsely Rejected
Video 9	0.0101	Normal	Truly Accepted
Video 10	0.3025	Normal	Truly Accepted

Table 5-4: Bayesian Classification Results based on average consecutive RMSE

5.6 Accuracy Percentage

AP = (Number of accurate results/ Total) x 100

= (8/10) x 100

= 80%

Table 5.3 shows behavior detection of videos and table 5.4 shows abnormality detection. Ten videos are taken for experimentation and results show that proposed technique detects behavior of eight videos correctly and gives wrong results for only two videos.

For behavior detection it gives wrong results for video 3 and video 6. For video 3 it is satisfying our criteria of peaceful behavior while in reality this video is violent and reverse is the case of video 6.

For abnormality detection it gives wrong results for video 4 and video 8, rest of the results are truly accepted. Video 4 is abnormal while proposed technique detects it as normal and video 8 is abnormal whereas proposed method gives normal result for this.

So accuracy of proposed technique according to this set of data is found to be 80%.

5.7 GUI (Crowd Aggression Levels)

Final output is GUI figure given below which shows all of the results including details about videos, their behavior, RMSE values, video frames and classification.

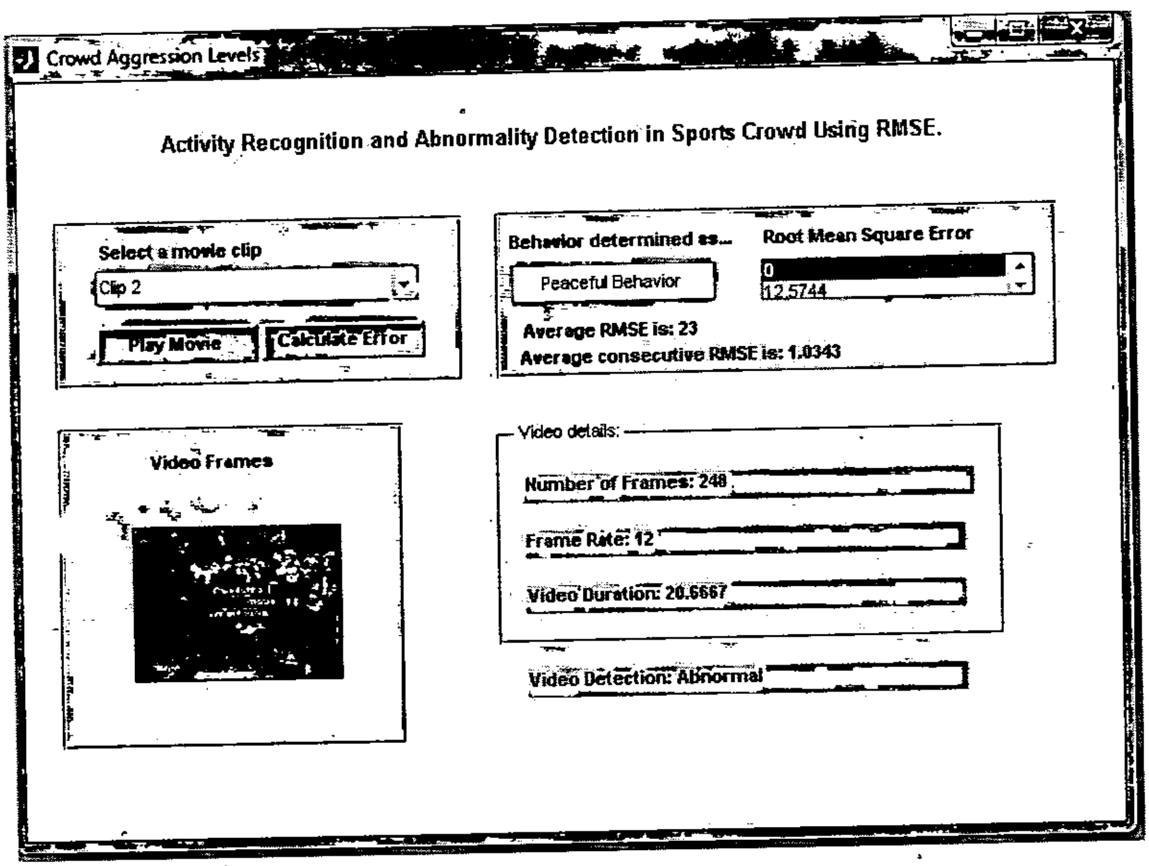


Figure 5-6: GUI (Crowd Aggression Levels)

Chapter 6
Conclusion and Future Work

6 Conclusion and Future Work

Based upon the experimental results the conclusions drawn.

6.1 Conclusion

Many techniques have been developed for video surveillance systems keeping in mind its importance regarding security purposes. Most of them work for the environment where number of persons is small usually indoor. Very few of them work for crowds and they also work in limited scenario.

Proposed technique not only detects abnormal events in crowd but also identifies crowd behavior and differentiates behavior between peaceful and violent. RMSE technique provides parameters which helps in behavior detection also as one of its parameter that is used to train Bayesian classifier which helps in classification of normal and abnormal events. Different crowd videos are taken for tests specifically sports crowd. The experiments show that our model is effective in detecting abnormal events in crowds. Also the relationship between different parameters efficiently detects crowd's behavior.

6.2 Future Work

There always remain spaces for improvement. For specific type of environments supervised learning may be used. Our work focuses mainly on sports crowds and usually moving cameras are used for their footage. In future this technique can also be applied for stationary footage considering dimensions and zoom factors. Moreover it may be applied on wide range of data to set up a threshold value with more conviction keeping in mind factors like background noise and illumination change.

Appendices

References:

[1]. T. V. Duong, H H. Bui, D. Q. Phung, and S. Venkatesh, "Activity recognition and abnormality detection with the switching hidden semi-markov model", IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), 1:838–845, 2005.

[2]. R. Hamid, A. Johnson, S. Batta, A. Bobick, C. Isbell, and G. Colema, "Detection and explanation of nomalous activities: representing activities as bags of event n-grams", IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), 1:1031–1038, 2005.

[3]. N. Oliver, A. Garg, and E. Horvitz, " Layered representations for learning and inferring office activity from multiple sensory channels", Computer Vision and Image Understanding, 96:163–180, 2004.

[4]. S. Gong and T. Xiang, "Recognition of group activities using a dynamic probabilistic network", Proceedings of the IEEE International Conference on Computer Vision, pages 742–749, 2003.

[5] Tanzeem Choudhury, Gaetano Borriello, et al. "The Mobile Sensing Platform: An Embedded System for Activity Recognition", Appears in the IEEE Pervasive Magazine - Special Issue on Activity-Based Computing, April 2008.

[6] Nishkam Ravi, Nikhil Dandekar, Preetham Mysore, Michael Littman, "Activity Recognition from Accelerometer Data", Proceedings of the Seventeenth Conference on Innovative Applications of Artificial Intelligence (IAAI/AAAI 2005).

[7] Want R., Hopper A., Falcao V., Gibbons J.: "The Active Badge Location System", ACM Transactions on Information, Systems, Vol. 40, No. 1, Pages. 91-102, January 1992

[8] Bieber G., Kirste T., Untersuchung des gruppendynamischen Aktivitaetsverhaltes im Office-Umfeld, 7. Berliner Werkstatt Mensch-Maschine-Systeme, Berlin, Germany, 2007

[9] Tao Gu, Zhanqing Wu, Liang Wang, Xianping Tao, and Jian Lu. "Mining Emerging Patterns for Recognizing Activities of Multiple Users in Pervasive Computing", In Proc. of the 6th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous '09), Toronto, Canada, July 13–16, 2009.

[10]. http://en.wikipedia.org/wiki/Video_camera

[11] Jennifer G. Dy, Carla E. Brodley: "Feature Selection for Unsupervised Learning", Journal of Machine Learning Research 5 (2004) 845–88, Submitted 11/2000; Published 8/2004

[12] M. Morita1,2, R. Sabourin1−3, F. Bortolozzi3 and C. Y. Suen2: "Unsupervised Feature Selection Using Multi-Objective Genetic Algorithms for Handwritten Word Recognition", 2003

[13]. http://users.ics.tkk.fi/harri/thesis/valpola_thesis/node34.html

[14]. www.wikipedia .com

[15] Xiaojin Zhu: "Semi-Supervised Learning Literature Survey", Last modified on July 19, 2008

[16] Nacim Ihaddadene and Chabane Djeraba, "Real-time Crowd Motion Analysis", 978-1-4244-2175-6/08/$25.00 ©2008 IEEE

[17] Nahum Kiryati, Tammy Riklin Raviv†, Yan Ivanchenko, Shay Rochel, " Real-time Abnormal Motion Detection in Surveillance Video", 978-1-4244-2175-6/08/$25.00 ©2008 IEEE

[18]. Ernesto L. Andrade1, Scott Blunsden2 and Robert B. Fisher, "Modelling Crowd Scenes for Event Detection", IEEE Computer Society Washington, DC, USA, Volume 01, 2006

[19]. D. Zhang, D. Gatica-Perez, S. Bengio, and I McCowan, "Semi-supervised adapted hmms for unusual event detection", IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), 1:611–618, 2005.

[20]. T. Xiang and S. Gong,"Video behaviour profiling and abnormality detection without manual labeling", Proceedings, IEEE International Conference on Computer Vision, 2005.
[21]. W. Li, Z. Zhang, and Z. Liu, "Expandable Data-Driven Graphical Modeling of Human Actions Based on Salient Postures", IEEE Transaction on Circuits and Systems for Video Technology, *Vol.18 No.11, pages 1499-1510, 2008.*
[22] F. Lin, C. Fookes, V. Chandran and S. Sridhara," Investigation into Optical Flow Super-Resolution for Surveillance Applications", In Lovell, Brian C. & Maeder, Anthony J. (Eds.) APRS Workshop on Digital Image Computing: Pattern Recognition and Imaging for Medical Applications, *21 February, 2005, Brisbane.*
[23]. F. Porikli, "Learning Object Trajectory Patterns by Spectral Clustering",IEEE International Conference on Multimedia and Expo (ICME), 2004
[24] H. Zhong, J. Shi, and M. Visontai, "Detecting unusual activity in video", IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004), 2:819–826, 2004.
[25] B.D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision", DARPA Image Understanding Workshop, 1981, Pages121{130 (see also IJCAI'81, Pages674{679).
[26] B.K.P. Horn and B.G. Schunck, "Determining Optical Flow", Arti_cial Intelligence, 17, 1981, Pages185{204.
[27] J. L. Barron, "Experience with 3D optical ow on gated MRI cardiac datasets" In 1st Canadian Conference on Computer and Robot Vision, pages 370{377, 2004.
[28] J.L.Barron and N.A.Thacker Tutorial: "Computing 2D and 3D Optical Flow", Tina Memo No. 2004-012, 2005.
[29]. W. Li, Z. Zhang, and Z. Liu, "Action Recognition Based on A Bag of 3D Points", in Proc. IEEE International Workshop on CVPR for Human Communicative Behavior Analysis *(CVPR4HB), pages 9-14, San Francisco, CA, USA, June 18, 2010.*
[30]. W. Lin, M-T. Sun, R. Poovandran, and Z. Zhang, "Activity Recognition Using A Combination of Category Components And Local Models for Video Surveillance", IEEE Transaction on Circuits and Systems for Video Technology, *Vol.18, No.8, pages 1128-1139, 2008.*
[31]. W. Lin, M.-T, Sun, R. Poovendran, Z. Zhang, "Group event detection with a varying number of group members for video surveillance", IEEE Trans. Circuits and Systems for Video Technology, vol. 20, issue. 8, Pages.1057--1067, 2010.
[32] M J. Black and P. Anandan, "A framework for the robust estimation of optical flow", 4th International onference on Computer Vision, pages 231–236, 1993.
[33] http://www.micquality.com/six_sigma_glossary/rmse.htm

*[34]*http://www.eumetcal.org/resources/ukmeteocal/verification/www/english/msg/ver_cont_var/uos3/uos3_ko1.htm