Detection of Abnormal Events in Surveillance Videos

K.Srinivas

A Thesis Submitted to
Indian Institute of Technology Hyderabad
In Partial Fulfillment of the Requirements for
The Degree of Master of Technology

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Department of Computer Science and Engineering

June 2015

Declaration

I declare that this written submission represents my ideas in my own words, and where ideas or words of others have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the Institute and can also evoke penal action from the sources that have thus not been properly cited, or from whom proper permission has not been taken when needed.

K. Srinivas

(Signature)

K. Srinivas

(K.Srinivas)

CS13M1005

(Roll No.)

Approval Sheet

This Thesis entitled Detection of Abnormal Events in Surveillance Videos by K.Srinivas is approved for the degree of Master of Technology from IIT Hyderabad

N.B.Vineeth

(———) Examiner N.B. Vineeth,

~~Dept. of Chem Eng~~ Dept. of CSE,

~~IITM~~ IITH

25.6.15

(———) Examiner Sumohana Channappayya

~~Dept. Math~~ Dept. of EE

~~IITH~~ IITH

C. Krishna Mohan

(Dr.C.Krishna Mohan) Adviser

Dept. of Computer Science and Engineering

IITH

~~(———) Co-Adviser~~

~~Dept. of Chem Eng~~

~~IITH~~

Naveen Sivadasan

(———) Chairman

~~Dept. of Mech Eng~~ Dept. of CSE

IITH

Abstract

The objective of this research work is to detect abnormal events in surveillance videos. It is one of the important and major challenging task in surveillance videos. Surveillance cameras are used for security, controlling and analyzing the crime, analysis of abnormal events in the places like Airports, Bus stands, Railway stations and other public places. The cameras capture huge data and store, which can be used to analyze the situations after their happening. Although this helps the operators to understand but no more use for controlling. Hence semi/fully automated intelligent systems are required to intimate the operators at the time of happening , thus hope of controlling huge damage. To illustrate, this can be used in finding terror attack, bomb placement, religion disputes, attacks, accidents, violation of traffic rules, etc. With this aim, researchers have focused on **Abnormal Detection in Surveillance Videos** where anomaly is detected and localized. Our aim here is to develop a more robust system to handle different anomaly and help the surveillance system.

We propose a novel algorithm for solving this problem. The algorithm contains two phases. First one is features extraction phase, second one is classification method phase. For feature extraction we use histograms of the orientation of optical flow (HOOF) method. Initially we split the video into sequence of frames. After that, by using HOOF method we extract the feature vector of each frame in that video stream. After feature extraction phase, we classify the frames as normal or abnormal by using two-class SVM. If the sequence of frames continuously detected as abnormal and crosses the contain threshold then the activity in that frame sequence is classified as an abnormal activity.

Key Words: Abnormal detection, Optical flow, HOOF, Two-class SVM.

Contents

Chapter 1

Introduction

In computer vision one of the major research areas is on surveillance videos. In surveillance videos,real time detection of abnormal events at the time of happening is one of the major challenging task. Abnormal means deviating from what is normal or uncommon behavior. Consider any public places like Airports, Bus stands, Railway stations, Temples , market, plaza and traffic places where unexpected events or abnormal events happens like fires, explosions, accidents, transportation disasters etc. At that time people scared and would trying to escape from those places. So that by taking crowds behavior before and after the event happened, we can detect abnormal events. For controlling the huge damage to the society and public safety, detection of abnormal events at the time of happening is one of the useful technologies. Hence semi/fully automated intelligent systems are required to intimate the man kind at the time of happening will lead to controlling the huge damage to the society. It can be used in finding the terror attack, Bomb placement, religion disputes, attacks, accidents, violation of traffic rules etc. Here our objective is to develop more robust system to handle various anomaly and help the surveillance system.

The key aspect of abnormal detection technique is the nature of the input data. The input is generally a collection of data instances like vector, record, event, point, sample, observation, entity etc. Each data instance is described using set of attributes like variable, feature, field, dimension etc. Each data instance contains only one attribute or multiple attributes. For multiple attribute data instance, all attributes might be of same type.

Anomalies are classified into two categories [1]. First one is point anomalies, if the test data instance is considered as abnormal with respect to rest of the data then the data instance is called as point anomaly. For example consider credit card fraud detection. Let the data set contains an individual person credit card transactions. Let us consider the data instance contains only one feature i.e amount spent. If any transaction in which the amount of spent is high compared to normal range of spent for that particular person then it is called point anomaly. Second one is contextual anomalies, if the test data instance is abnormal in particular context or time then it is called contextual anomaly. For example, temperature of weather 15 degrees is normal in winter. But same temperature value during summer is abnormal. Our abnormal detection method on crowded scenes is belongs to contextual anomaly category.

The algorithm of abnormal event detection contains two phases. First phase is feature extraction and second phase is classification. Transforming the input data instance into set of features is called feature extraction. Feature extraction is the process of selection of the variables or attributes which has some co relation with the class of that instance. Features are classified into two categories. First one is local features, it can also be referred as structural feature. These features contains structural elements like loop, branches, crossing point, joints, points, curve etc. Second one is global features, they can be computed on image pattern and they includes the features like histogram, projection, distance, etc. Our method is based on global features. In our experiment, for feature extraction process we take pixel values as input.

Based on the extent to which the labels are available,anomaly detection techniques can operate in one of the following methods. First one is supervised anomaly detection method, in this method the training data set has labeled instances for normal as well as abnormal class. Here,we build a predictive model for normal vs. abnormal class. Using this model, the test data is categorized which class it belongs to. Second one is semi-supervised anomaly detection method, in this method the training data set has labeled instance for only the normal class. Here we build a model for class corresponding to normal behavior. By using this model we can detect anomalies in test data.Third one is unsupervised anomaly detection method, in this method its do not required training data. In this category the normal instances are far more frequent than anomalies in the test data. If the assumption is false then the techniques in this category are suffered from high false alarm rate. In our experiment, we are using supervised learning method.

The data set used in our experiment are UMN and PETS2009. In this data sets we are taking video sequences, contains almost all possibilities of crowds behavior before and after abnormal event happened. For example in PETS2009 dataset, the normal sequences is people are walking in same direction. The abnormal sequence is people are running in same direction. In UMN data set, the normal sequence is people are walking in all directions. The abnormal sequence is people are running in all directions.

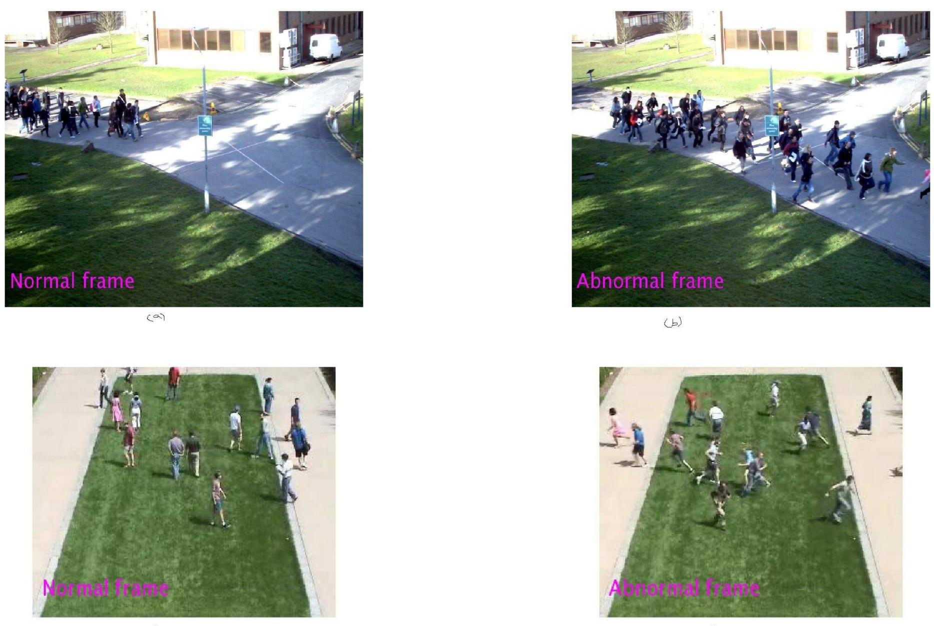


Fig:(a),(b) are normal and abnormal frames of PETS data set.(c),(d) are normal and abnormal frames of UMN data set.

Chapter 2

Review of previous work

Several approaches have been proposed to date for the abnormal event detection. In [2] a novel approach for abnormal event detection by using histogram of optical flow orientation method was proposed.This method i have taken reference for my proposed method. In [3] a novel approach for anomaly detection method was used. They have used local dictionaries from normal behavior descriptor for local regions. For finding the anomaly candidate behavior sparse reconstruction error can be used. They represent the given candidate data as sparse linear combinations of the optimal subset of usual features: dictionary. Descriptor extraction, Motion magnitude, Percentage foreground pixels for dense space time cube Background subtraction features are used in this approach. In [4] a novel approach for anomaly detection method was used. Here,they are finding normal crowd behavior model by using mixtures of dynamic textures. If the test data instance is out of this area then it labeled as anomalies. Here they are using temporal anomaly detection and spatial abnormal detection. In [5] a method was proposed for analysis of crowd scenes using holistic properties. In this firstly, they find crowd segmentation, after that they did perspective normalization. After that they did feature extraction by using segment features, internal edge features and texture features. In [6] the method introduce the sparse reconstruction cost for detecting both local abnormal event and global abnormal event. By updating the dictionary incrementally they further extend it to supports for online abnormal event detection. They took multi-scale HOF: 2 scales (smaller and larger) with 8 direction (16 bins) has the features. In [7] the method proposed for abnormal crowd behavior detection using social force model. This method can be used for both detecting and localizing the abnormal behaviors. Firstly grid of particles is placed over the image, after that finding the moving particles which are individuals. By using social force model they find the interaction forces between the particles. By mapping the interaction forces to image plane they find the force flow for every pixel in every frame. Force flow is used to find the model for normal crowd behavior. By using this model they can detect anomalies behavior of the crowds. In [8] the method proposed for anomalous crowd detection for both coherent and incoherent scenes. For modeling the crowded scenes, they are using particle trajectories. For complicated crowded scenes they are introducing chaotic dynamics chaotic invariant features. In [9] the method Bayesian framework can be proposed for detecting crowd escape events. Here,they are introducing potential destinations and divergent centers for characterizing crowd motion in both presence and absence of escape events.

Chapter 3

Abnormal Event Detection Method

3.1 Features Selection

A feature is defined as a descriptive parameter that is extracted from an image or a video sequence. Features can be classified into low-level and high-level features. Low-level features (also known as primitive features) such as color, texture, shape,object motion (for video) and spatial location of image elements (both for image and video). High-level features (also known as logical, semantic features) involve various degrees of semantics depicted in images and video. The features at this level can be objective or subjective. In our experiment,for computing feature extraction of a frame having 2 steps. First one is computing the optical flow, second one is finding the histogram of orientation optical flow.

3.1.1 Optical Flow

Consider any two sequences of frames, optical flow is the each pixel comes up with a vector that gives how much distance pixel was moving compared to previous frame. $\bar{u} = (a, b)$ is a velocity vector of each pixel in the image. It gives how quickly, the pixel was moving across the image and also it tells in which direction it is moving in the image plane. An abnormal action can be found by using direction and velocity of the movement, because of this we can use optical flow for scene description. The following diagram is the optical flow of two sequence of images.

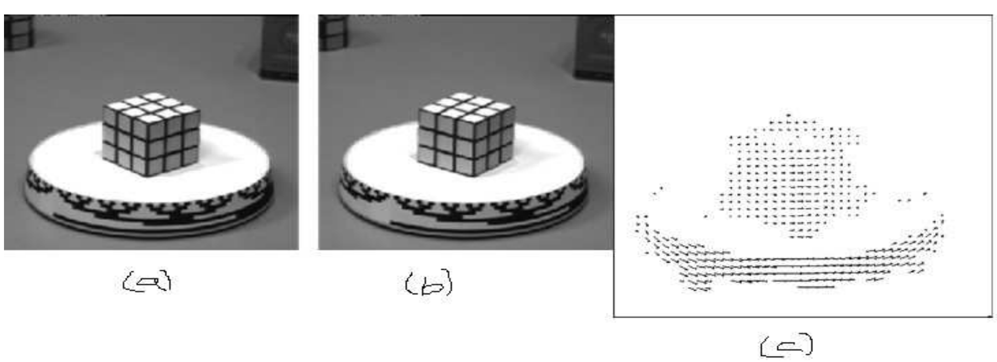


Fig:(a),(b) are sequence of frames.(c)optical flow of these frames.

The following is the method for computing optical flow proposed by B.Horn and B.Schunck [10]. Assume image intensity is constant for the corresponding pixel in the two sequence of images.

$$I_0(x, y, t) \approx I_1(x + dx, y + dy, t + dt) \tag{3.1}$$

After applying Taylor series ,we will get following equation.

$$I_x a + I_y b + I_t = 0 \tag{3.2}$$

By adding the global constraint of smoothness to brightness constancy, we can solve the problem.

$$E = \int\int \Big(\Big(I_x a + I_y b + I_t \Big)^2 + \alpha^2 \Big(|\bigtriangledown a|^2 + |\bigtriangledown b|^2 \Big) \Big) \mathrm{d}x\mathrm{d}y \tag{3.3}$$

Where I_x, I_y, I_t are the derivatives of the image intensity values along the x,y and time t dimensions receptively.a,b are the components of the optical flow and α is a regularization constant.

Minimize this function by using variational calculus to get following equations.

$$I_x(I_x a + I_y b + I_t) - \alpha^2 \bigtriangleup a = 0 \tag{3.4}$$

$$I_y(I_x a + I_y b + I_t) - \alpha^2 \bigtriangleup b = 0 \tag{3.5}$$

where $\bigtriangleup = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is the Laplace operator.

$$\bigtriangleup a(x, y) = \bar{a}(x, y) - a(x, y) \tag{3.6}$$

$$\bigtriangleup b(x, y) = \bar{b}(x, y) - b(x, y) \tag{3.7}$$

By using iterative scheme we can find a,b.

$$a^{k+1} = \bar{a}^k - \frac{I_x(I_x \bar{a}^k + I_y \bar{b}^k + I_t)}{\alpha^2 + I_x^2 + I_y^2} \tag{3.8}$$

$$b^{k+1} = \bar{b}^k - \frac{I_y(I_x \bar{a}^k + I_y \bar{b}^k + I_t)}{\alpha^2 + I_x^2 + I_y^2} \tag{3.9}$$

Where k denotes the algorithm iteration.Here the computations are based on sequence of two images, so that we can take single time step only.

3.1.2 Histogram of Oriented Optical Flow(HOOF)

It is a graphical representation of the frequency of pixels (vectors) in an optical flow as a function of their orientation. Firstly we can divide the image into blocks , after that each block is divided into cells where HOOF computed. A block contains $b_h \times b_w$ cells,where b_h, b_w are the number of cells in y and x direction. Each cell contains $c_h \times c_w$ pixels,where c_h, c_w are the height and width of the cell. In our experiment,we are changing the resolution of each frame into 240X320. After that each frame is divided into blocks of size 2X2 cells and each cell having size 8X8 pixels. Here we are taking overlapping element of two adjacent blocks is 50%. After dividing the frame like this,we get

39 blocks in horizontal direction and 29 blocks in vertical direction. After that we are calculated histogram of each cell having 9 bins.

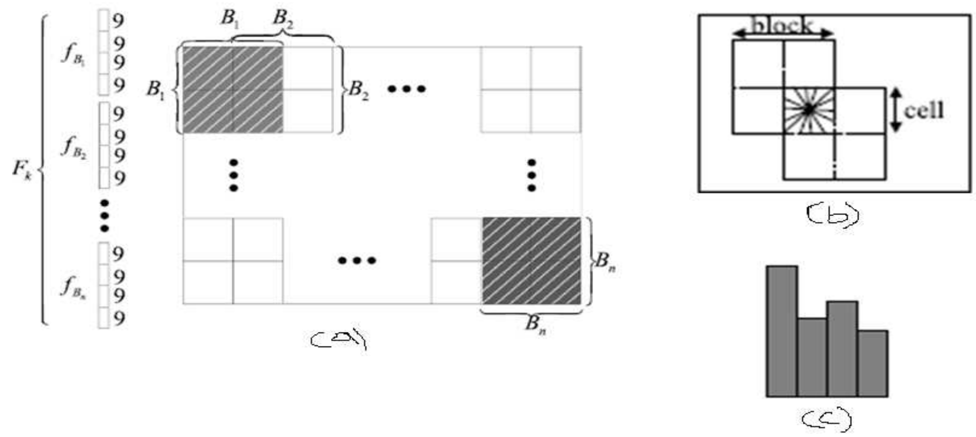


Fig: (a)HOOF computation in the kth frame.(b)Dividing block into cells(c)HOOF of the cell

We are explained calculation procedure of HOOF [11] with one example as follows. Here we are taking entire frame as one block and the block having only one cell.

1. Compute the optical flow of two sequence of frames in the video.

2. Each optical flow vector is binned according to its angle from the horizontal axis and also by using its magnitude it can be weighted.

3. Thus, all optical flow vectors $u = [a, b]^T$, with direction $\theta = \tan^{-1}(b/a)$ range.

$$-\frac{\pi}{2} + \pi\frac{k-1}{B} \leq \theta < -\frac{\pi}{2} + \pi\frac{k}{B} \tag{3.10}$$

   where $1 \leq k \leq B$ and B is the number of bins of the histogram.

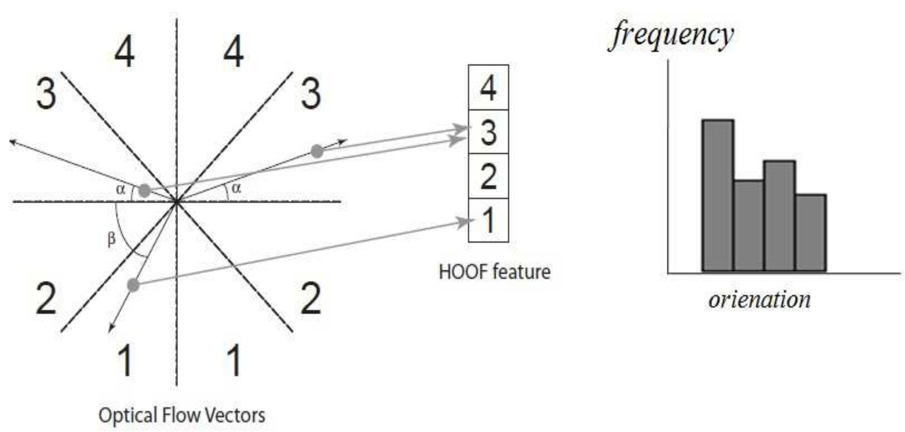


Fig:Histogram formation with four bins, B = 4

Like above procedure, we are calculating the HOOF of the cells. After that grouping the cell's HOOF together of the corresponding blocks and Normalization the group of histograms represent the block histogram. The set of these block histograms represents the feature descriptor(F_k)of the frame(number of blocks x number of bins dimension). In our experiment our feature vector dimension is 40716 X 1.

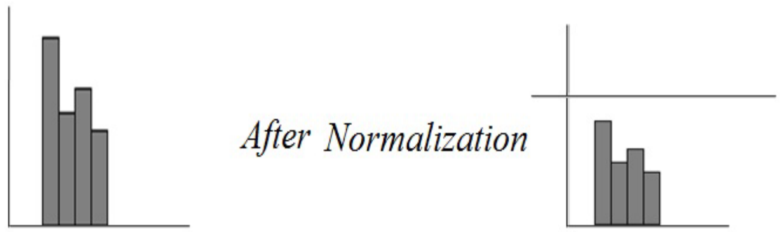


Fig:Normalization of the Histogram.

3.2 Classification

3.2.1 Theory Behind SVM

In our experiment we are used support vector machine(SVM) method for classification. Support vector machines (SVMs)provide a new approach to pattern classification problems with underlying basis in statistical learning theory, in particular the principle of structural risk minimization. The SVM models learn to separate the boundary regions between patterns belonging to two classes by mapping the input patterns onto a high dimensional space, and seeking a separating hyperplane in this space. The separating hyperplane is chosen in such a way as to maximize its distance (margin) from the closest training examples. Later, SVM has been extended to non-linear framework with the introduction of kernel methods. The theory behind non linear two class SVM explained [12] below.

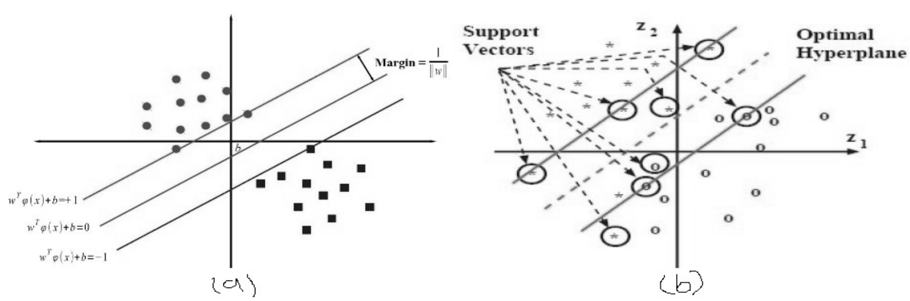


Fig:(a) Maximum-margin hyperplane for two class SVM (b)SVM for linearly non separable patterns.

Optimal hyperplane specified by (W,b) must satisfy the constraint:

$$y_i(W_i^T Z_i + b) \geq 1 - \beta_i \tag{3.11}$$

$$\beta_i \geq 0 \tag{3.12}$$

Here i= 0,1,2....N where N is the number of training examples. Z_i is the feature vector for X_i and y_i is the desired output.β_i is the measure of deviation for Z_i. The training examples for which they satisfied above constraints with equality sign are called support vectors. The optimum value of margin of separation is $\frac{2}{\|W\|}$

Given the training examples (X_i, Y_i) i=1,2.....N and find the values W and b such that they satisfied the above constraints [4.1] ,[4.2]. The weight vector W minimizes the cost function:

$$\frac{1}{2}W^T W + C\sum_{i=1}^{N}\beta_i \tag{3.13}$$

subject to satisfies the above constraints [4.1],[4.2].

where C is a user specified positive parameter. The optimal value of the weight vector satisfies following equation.

$$W = \sum_{i=1}^{N_s}\alpha_i y_i Z_i \tag{3.14}$$

where $0 \leq \alpha_i \leq C$ and the decision function is

$$f(x) = sgn\left(\sum_{i=1}^{N_s}\alpha_i y_i K(x, x_i) + b\right) \tag{3.15}$$

where N_s is the number of support vectors and the inner product kernel is defined as $K(x, x_i) = z^T z$. We have different type of kernel functions is there like polynomial, sigmoidal,Gaussian. In our experiment we are used gaussian kernel function.The following are the expressions of different kernels.

$$Gaussian\ kernel : K(x, x_i) = exp(\frac{-\|x - x_i\|^2}{\sigma^2}) \tag{3.16}$$

$$Polynomail\ kernel : K(x, x_i) = (x^T x_i + c)^d \tag{3.17}$$

$$sigmoidal\ kernel : K(x, x_i) = \tanh(ax^T x_i + c) \tag{3.18}$$

3.2.2 SVM Training

In this training phase, the features of the frames are trained to SVM. Mainly two parameters are supplied for SVM train function. First one is the array of feature vector of all the frames and second one is array of the labels of all frames. In addition ,some optional parameters are also used in this training phase like type of SVM(multi class,one class,regression) and type of kernel function(polynomial,sigmoidal,Gaussian),set gamma value etc. By taking these parameters the SVM train function returns a model which can be used as input parameter for SVM predict function.

3.2.3 SVM Prediction

In this prediction phase, the video is classified.Initially we extract the frames from entire video. By using feature extraction technique, For each frame we can compute the feature vector. For svm predict function, two parameters have to be supplied. First one is model returned from SVM train function. Second parameter is the array of feature vector of the video to be classified. By taking these parameters SVM predict function returns an array of labels assigned to each frame. In our experiment,for SVM classification we are used LIBSVM(library for support vector machines) software package [13].

Chapter 4

Overview of Abnormal Event Detection Method

Block Diagram for anomaly detection:

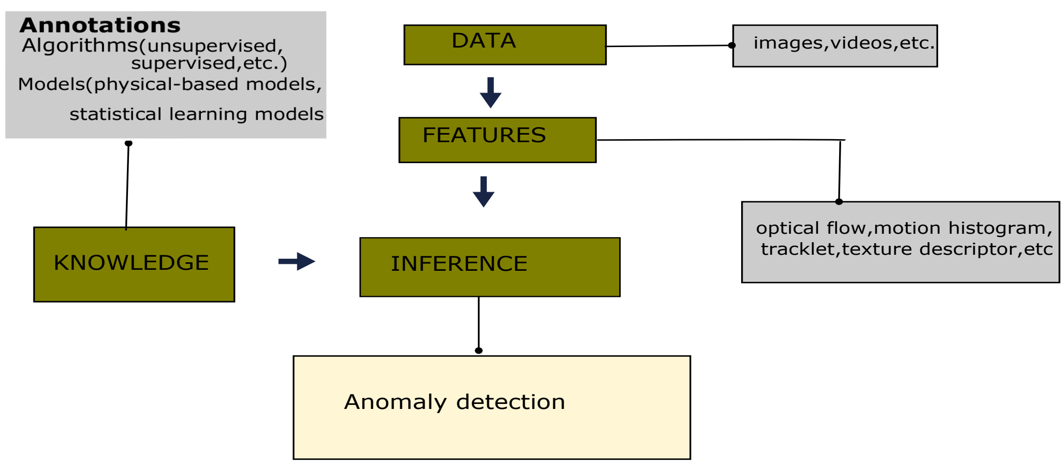


Fig:General structure for analysis of anomaly detection in crowded scenes.

Overview:

Let's consider $\{f_1, f_2.....f_n\}$ be the set of training frames having labels both normal and abnormal behavior.In training step,we are computing a model which can be used in testing step.

STEP1: Computing optical flow for each frame,It gives movement features of each pixel in the frame.

$$\{f_1, f_2.....f_n\} \rightarrow \{OF_1, OF_2.........OF_{n-1}\} \tag{4.1}$$

where $\{OF_1, OF_2.........OF_{n-1}\}$ are the optical flows of the corresponding training frames. Optical flow cannot be computed at the last frame in the sequence.

STEP2: computing feature vectors of the frames by using Histogram of oriented optical flow method(HOOF).

$$\{OF_1, OF_2........OF_{n-1}\}\overrightarrow{HOOF}\{FV_1, FV_2......FV_{n-1}\} \tag{4.2}$$

where $\{FV_1, FV_2......FV_{n-1}\}$ are the HOOF feature vectors of the corresponding training frames.

STEP3: By applying two-class SVM , we can obtain the support vectors for the HOOF feature vectors of the training frames.

$$\{FV_1, FV_2......FV_{n-1}\}\overrightarrow{SVM}\{SV_1, SV_2.........SV_m\} \tag{4.3}$$

where m $\leq n-1$ and $\{SV_1, SV_2.........SV_m\}$ are the support vectors.

STEP4: Based on the support vectors computed from training step, each testing frame $\{TF_1, TF_2, TF_3......TF_q\}$ can be classified.

$$f\big(FV_k\big) = sgn\bigg(\sum_{i=1}^{m} \alpha_i y_i K(SV_i, FV_k) + b\bigg) \tag{4.4}$$

$$= 1 \Rightarrow f(FV_k) \geq 0$$

$$= -1 \Rightarrow f(FV_k) < 0$$

where FV_k is the HOOF feature vector of the testing frame to be classified and SV_i is the support vector."1" represents normal frame and "-1" represents abnormal frame.

The following diagram is the flow chart of the Abnormal event detection method.

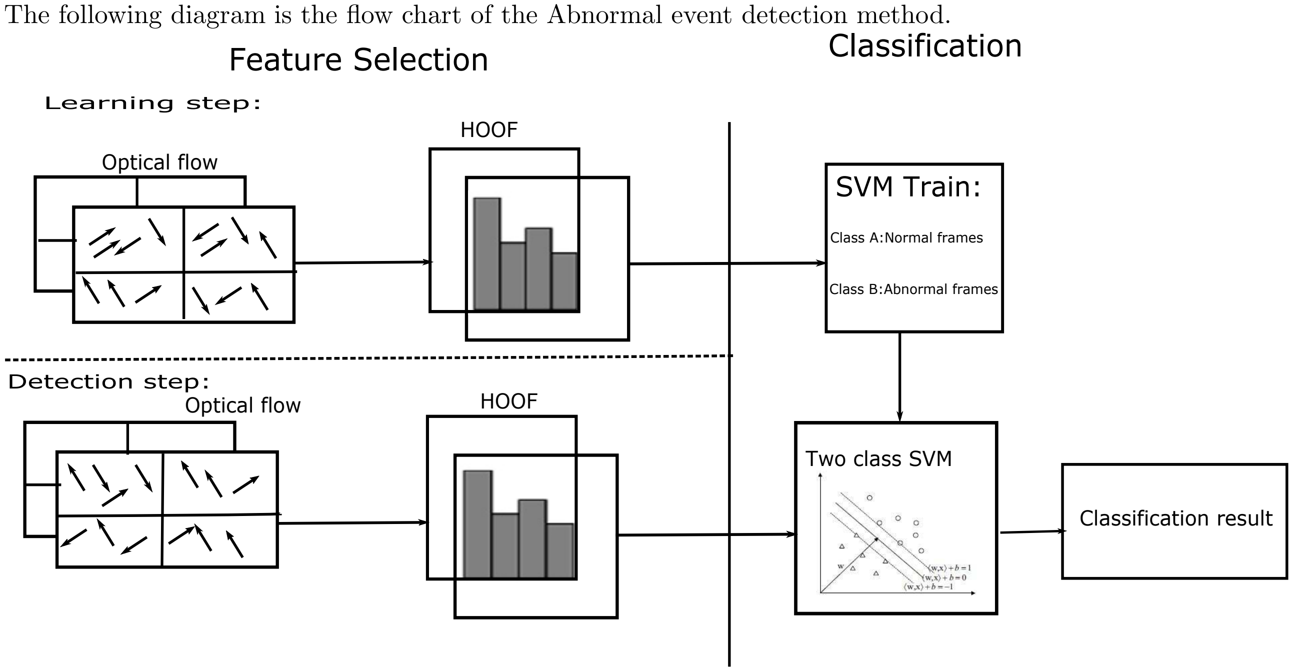


Fig:Flow chart of our abnormal event detection method.

Chapter 5

Experimental Results

In this section we are computed experimental results for analyzing the performance of the our abnormal event detection method. we are used PETS2009 [14] and UMN [15] data sets for our abnormal event detection method experiments. The UMN data set having the resolution 240X320 and PETS data set having the resolution 576X768. We are changing the resolution of PETS data set same as the UMN data set because of analyzing the performance of the our abnormal event detection method.

5.1 Generalized Training Dataset

Initially we are taking the training frames from both PETS2009 [14] and UMN [15] data sets having both normal and abnormal behavior. By taking these frames we can create generalized training data set containing all possibilities. Initially, By using our abnormal event detection method we compute a model for generalized training data set. Based on this model, we can detect abnormal events in the given video sequence.

The generalized training data set containing all possibilities of the normal and abnormal crowd behavior as explained below.

Normal behavior of the crowd:

- Frames in which People are walking in same direction -Taking 40 frames from PETS data set of *Time stamp 14-17(frames from 0 to 39).*
- Frames in which People are walking in different directions - Taking 350 frames from PETS data set of *Time stamp 14-55(frames from 0 to 349)* and taking 200 frames from UMN data set of lawn scene(frames from 51 to 150),plaza scene(frames from 101 to 200).
- Frames of cohesive crowd behavior - Taking 40 frames from PETS data set of *Time stamp 14-31(frames from 0 to 39).*
- Frames in which people are gathering(crowd formation) - Taking 120 frames from PETS data set of *Time stamp 14-33(frames from 101 to 220).*

Abnormal behavior of the crowd:

- Frames in which people are running in same direction - Taking 20 frames from PETS data set of *Time stamp 14-16(frames from 81 to 100).*

- Frames in which people are running in different direction - Taking 10 frames from PETS data set of *Time stamp 14-33(frames from 366 to 375).* and taking 60 frames from UMN data set of lawn scene(frames from 651 to 680),plaza scene(frames from 571 to 600).

- Frames of Crowd splitting behavior - Taking 40 frames from PETS data set of *Time stamp 14-31(frames from 71 to 110).*

- Frames of crowd Evacuation behavior - Taking 20 frames from PETS data set of *Time stamp 14-33(frames from 346 to 365).*

In our experiment we are taking 5 different abnormal video sequences from PETS2009 data set and 2 abnormal video sequences of 2 different scenes(lawn scene,plaza scene) from UMN data set. The detection results of these video sequences are explained next sections.

5.2 Evaluation Parameters

we are used the following parameters for analyzing the performance of our detection method as explained below.

−	Normal(predict)	Abnormal(predict)
normal(actual)	a	b
Abnormal(actual)	c	d

a − is number of correct predictions that an instance is normal.

b − is number of incorrect predictions that an instance is abnormal.

c − is number of incorrect predictions that an instance is normal.

d − is number of correct predictions that an instance is abnormal.

Accuracy(ACC): It is proportion of the total number of predictions that were correct.it is determined using following equation.

$$ACC = \frac{a+d}{a+b+c+d} \times 100 \tag{5.1}$$

True positive rate (or) True alarm rate(TPR): It is proportion of abnormal cases that were correctly identified.

$$TPR = \frac{d}{c+d} \times 100 \tag{5.2}$$

False positive rate (or) False alarm rate(FPR): It is proportion of normal cases that were incorrectly classified as abnormal, as calculated using the equation

$$FPR = \frac{b}{a+b} \times 100 \tag{5.3}$$

False negative rate (or) Missed alarm rate(FNR): It is the proportion of abnormal cases that were incorrectly classified as normal.

$$FNR = \frac{c}{c+d} \times 100 \tag{5.4}$$

5.3 PETS Dataset

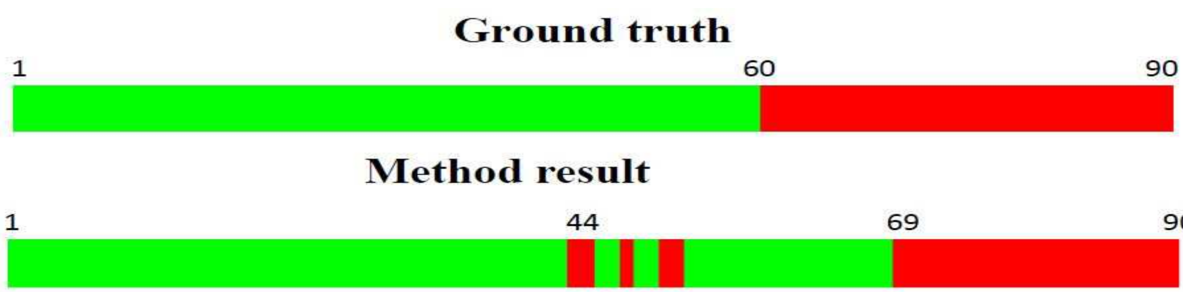
Video Sequence 1:

Normal scenes: People are walking in same direction.

Abnormal scenes: People are running in same direction.

Fig:Detected results of PETS Scene 1

Detection results: Here green represents Normal frames, Red represents abnormal frames. Here the following video sequence contains 90 frames. $1-59$ are normal testing frames, $60-90$ are abnormal testing frames. For ground truth, the abnormal event happened at 60th frame.



Our method detects abnormal event from 69th frame on wards(9 frames delay).And also at 44th, 48th,51th frames false alarms happened. The accuracy of the detection result is 83.33%.

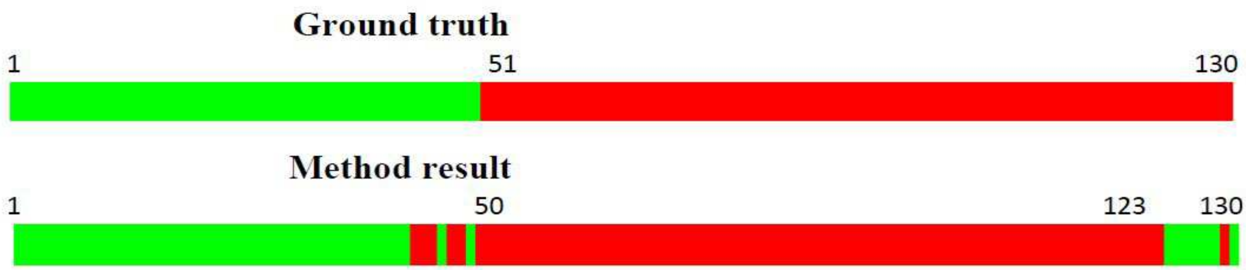
Video Sequence 2:

Normal scenes: Cohesive crowd

Abnormal scenes: Crowd splitting

Fig:Detected results of PETS Scene 2

Detection results: Here green represents Normal frames, Red represents abnormal frames. Here the following video sequence contains 130 frames. $1 - 50$ are normal testing frames, $51 - 130$ are abnormal testing frames. For ground truth, the abnormal event happened at 51th frame.



Our method detects abnormal event from 50th frame on wards(1 frames early).And also at 44th,45th, 48th frames false alarms happened,And also frames from 123 to 128,129th are considered as missed alarms. The accuracy of the detection result is 90%.

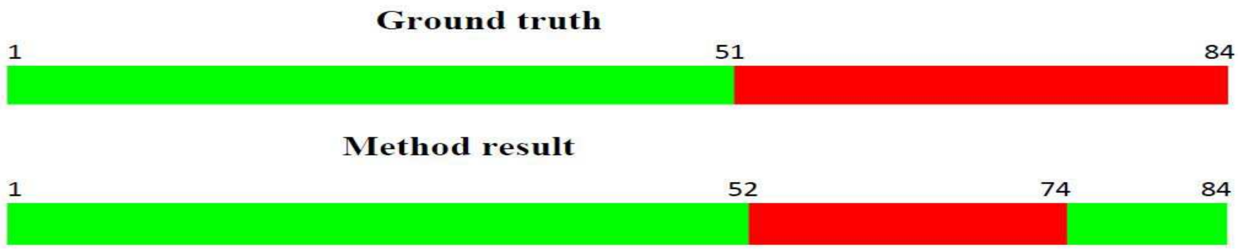
Video Sequence 3:

Normal scenes: People are walking in different directions.

Abnormal scenes: People are running in different directions.

Fig:Detected results of PETS Scene 3

Detection results: Here green represents Normal frames, Red represents abnormal frames. Here the following video sequence contains 84 frames. $1-50$ are normal testing frames, $51-84$ are abnormal testing frames. For ground truth, the abnormal event happened at 51th frame.



Our method detects abnormal event from 52th frame on wards(1 frames delay).Frames from 74 to 84 are considered as missed alarms. The accuracy of the detection result is 85.71%.

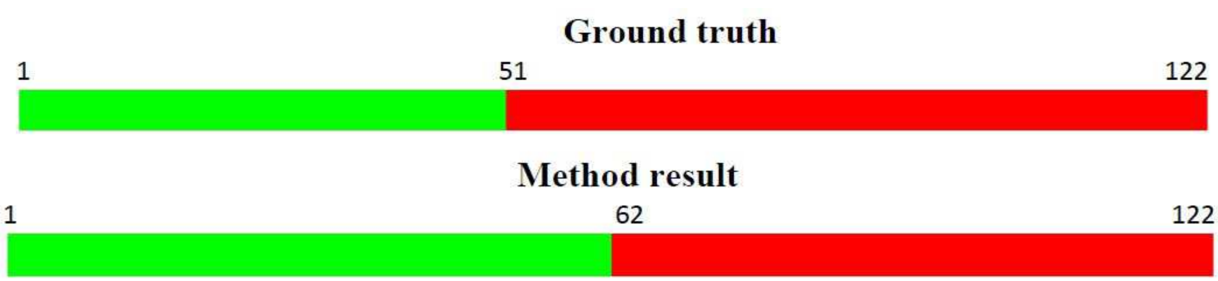
Video Sequence 4:

Normal scenes: People are walking in different directions.

Abnormal scenes: People are running in same direction.

Fig:Detected results of PETS Scene 4

Detection results: Here green represents Normal frames, Red represents abnormal frames. Here the following video sequence contains 122 frames. $1-50$ are normal testing frames, $51-122$ are abnormal testing frames. For ground truth, the abnormal event happened at 51th frame.



Our method detects abnormal event from 62th frame on wards(11 frames delay).Frames from 51 to 61 are considered as missed alarms. The accuracy of the detection result is 90.98%.

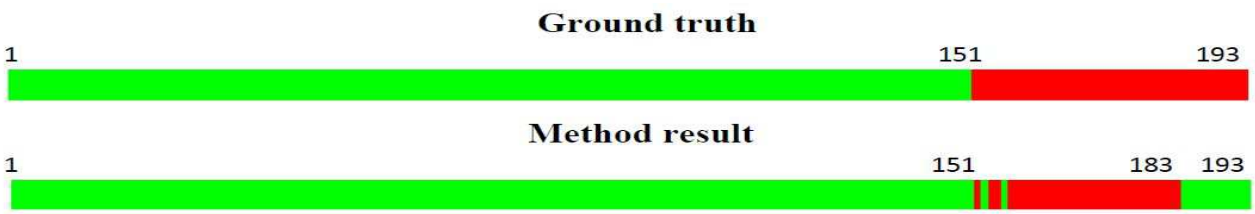
Video Sequence 5:

Normal scenes: People are gathering (or) crowd formation.

Abnormal scenes: Evacuation of the people.

Fig:Detected results of PETS Scene 5

Detection results: Here green represents Normal frames, Red represents abnormal frames. Here the following video sequence contains 193 frames. $1-150$ are normal testing frames, $151-193$ are abnormal testing frames. For ground truth, the abnormal event happened at 151th frame.



Our method detects abnormal event from 151th frame on wards.Frames from 183 to 193 and 152th,155th are considered as missed alarms.The accuracy of the detection result 93.26%.

Table 5.1: **Detection results of PETS scenes**

Different scenes	TPR	FPR	FNR	ACC
PETS Scene 1	70.96%	10.16%	29.03%	83.33%
PETS Scene 2	91.25%	12%	8.75%	90%
PETS Scene 3	64.70%	0%	35.29%	85.71%
PETS Scene 4	84.72%	0%	15.27%	90.98%
PETS Scene 5	69.76%	0%	30.23%	93.26%

<u>TABLE 6.1:</u> Detection results of PETS scenes.Here TPR='True positive rate'; FPR='False positive rate';FNR='False negative rate' or 'Missed alarm rate';ACC=accuracy.

5.4 UMN Dataset

Lawn scene:

Normal scenes: People are walking in different directions.

Abnormal scenes: People are running in different directions.

Fig:Detected results of UMN Scene 1

Detection results: Here green represents Normal frames, Red represents abnormal frames. Here the following video sequence contains 307 frames. $1 - 243$ are normal testing frames, $244 - 307$ are abnormal testing frames. For ground truth, the abnormal event happened at 244th frame.

Our method detects abnormal event from 247th frame on wards(3 frames delay).In between 1 to 244 frames some false alarms happened. The accuracy of the detection result is 83.06%.

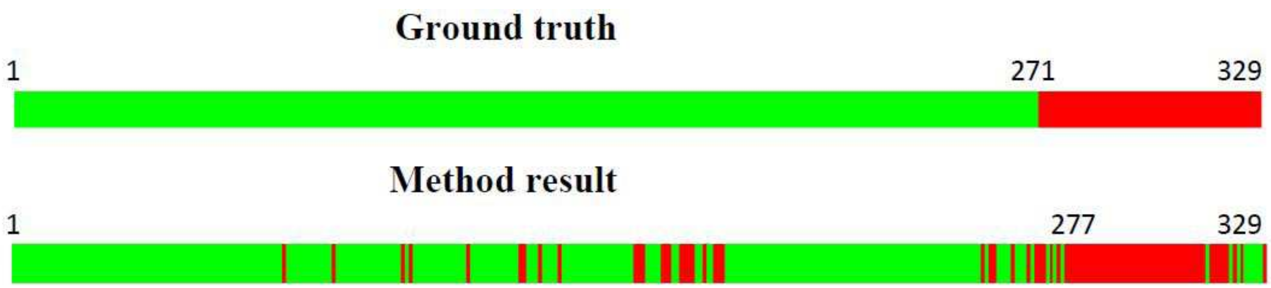
Plaza scene:

Normal scenes: People are walking in different directions.

Abnormal scenes: People are running in different directions.

Fig:Detected results of UMN Scene 2

Detection results: Here green represents Normal frames, Red represents abnormal frames. Here the following video sequence contains 329 frames. $1-270$ are normal testing frames, $271-329$ are abnormal testing frames. For ground truth, the abnormal event happened at 271th frame.



Our method detects abnormal event from 271th frame on wards.In between 1 to 271 frames some false alarms happened. The accuracy of the detection result is 87.53%.

Table 5.2: **Detection results of UMN scenes**

Different scenes	TPR	FPR	FNR	ACC
Lawn Scene	84.37%	17.28%	15.62%	83.06%
Plaza Scene	81.35%	11.11%	18.64%	87.53%

TABLE 6.2: Detection results of UMN scenes.Here TPR='True positive rate'; FPR='False positive rate';FNR='False negative rate' or 'Missed alarm rate';ACC=accuracy.

Chapter 6

Conclusion

A method for abnormal event detection is proposed. The method is based on two components − one is computing histograms of the orientation of optical flow(HOOFs), and second one is applying two class SVM for classification. We have tested our algorithm on several different video sequences in which it classified abnormal events. For reducing the wrong detections may be we can replace the optical flow with other approaches which can be capture more efficient features. From a machine learning perspective,deep learning theories can be used to increasing the learning and the classification performances.

References

[1] V. Chandola, A. Banerjee, and V. Kumar. Anomaly Detection: A Survey. *ACM Comput. Surv.* 41, (2009) 15:1–15:58.

[2] T. Wang and H. Snoussi. Histograms of optical flow orientation for visual abnormal events detection. In Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on. IEEE, 2012 13–18.

[3] S. Biswas and R. Babu. Sparse representation based anomaly detection with enhanced local dictionaries. In Image Processing (ICIP), 2014 IEEE International Conference on. 2014 5532–5536.

[4] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos. Anomaly detection in crowded scenes. In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. 2010 1975–1981.

[5] A. Chan, M. Morrow, and N. Vasconcelos. Analysis of crowded scenes using holistic properties. In Proceedings of the 11th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, PETS-09. 2009 .

[6] Y. Cong, J. Yuan, and J. Liu. Sparse reconstruction cost for abnormal event detection. In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011 3449–3456.

[7] R. Mehran, A. Oyama, and M. Shah. Abnormal crowd behavior detection using social force model. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. 2009 935–942.

[8] S. Wu, B. E. Moore, and M. Shah. Chaotic invariants of Lagrangian particle trajectories for anomaly detection in crowded scenes. In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. 2010 2054–2060.

[9] S. Wu, H.-S. Wong, and Z. Yu. A Bayesian Model for Crowd Escape Behavior Detection. volume 24. 2014 85–98.

[10] B. K. P. Horn and B. G. Schunck. Determining Optical Flow. *Artif. Intell.* 17, (1981) 185–203.

[11] N. Dalal, B. Triggs, and C. Schmid. Human detection using oriented histograms of flow and appearance. In Computer Vision–ECCV 2006, 428–441. Springer, 2006.

[12] N. Cristianini and J. Shawe-Taylor. An Introduction to Support Vector Machines: And Other Kernel-based Learning Methods. Cambridge University Press, New York, NY, USA, 2000.

[13] C.-C. Chang and C.-J. Lin. LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol.* 2, (2011) 27:1–27:27.

[14] I. PETS, Vellore. Performance Evaluation of Tracking and Surveillance (PETS) 2009 Benchmark Data. Multisensor Sequences Containing Different Crowd Activities,[link]:http://www.cvg.reading.ac.uk/PETS2009/a.html. 2009 .

[15] M. U. UMN, Minneapolis. Unusual Crowd Activity Dataset of University of Minnesota, Department of Computer Science and Engineering [Online]. Available: http://mha.cs.umn.edu/movies/crowdactivity- all.avi. 2006 .