A COMPREHENSIVE FRAMEWORK FOR LEARNING EVENTS

Kalaivani P
Department of Electronics and Communication Engineering
Thiagarajar College of Engineering, Madurai, Tamil Nadu, India
kalaimathu.18@gmail.com,

Mohamed Mansoor Roomi S
Department of Electronics and Communication Engineering
Thiagarajar College of Engineering, Madurai, Tamil Nadu, India
smmroomi@tce.edu

Annalakshmi M
Department of Electronics and Communication Engineering
Sethu Institute of Technology, Virudhunagar (DT), Tamil Nadu, India
annam.baluss@gmail.com

ABSTRACT
Intelligent video surveillance system plays a vital role in learning the events remotely. In recent years, surveillance systems are widely used in all places starting from border security application to street monitoring systems. The surveillance system can also be used to monitor the activities of a student who learns the course through distance e-learning. The teacher can use a surveillance monitor to watch the behavior of the student from a remote place. In e-learning scenario, attending the course or learning with system are considered as usual activities and the other activities like walking, bending, paper passing and hand waving are referred to as unusual events. The intelligent surveillance system has to learn itself the events in the capturing video and make a decision about the event whether it is usual or unusual. This paper deals with an algorithm for a machine learning approach to learn the video events. It presents a detailed review of various techniques for abnormal event detection in video and it presents current scenario of research in this area. It detects the events using features of histogram of optical flow orientation, magnitude and entropy (HOFOME) combined with the histogram of oriented gradients (HOG). It classifies whether the event is usual or unusual with different machine learning classifiers namely Classification Tree (Ctree), Support Vector Machine (SVM). This paper presents the experimental results of the algorithm applied on benchmark dataset. The performance comparison shows that the proposed work outperforms the state of the art methods.
Keywords: Intelligent video surveillance; e-learning; abnormal event detection; HOFOME; HOG features; Ctree; SVM; classifiers.

INTRODUCTION

Internet-based education and e-learning have become an emerging technological trend as a result of advancements in network and information technology (Jian Yu 2009). In recent years, video surveillance system has been used as a third eye to monitor persons, places, events and more. Currently millions of surveillance cameras are used for several purposes (Wahyono 2016) like crime detection in country borders, airports, illegally parked vehicle detection, fire detection, human detection and tracking, smoke detection and unattended object detection in shopping centers, underground stations, sport stadiums, residential streets and more.

Video surveillance approach can also be used for an intelligent e-learning system to monitor the persons learning through distance education. It can be used as a virtual supervisor for online examinations, web-based online training. In e-learning system, teacher or course co-ordinator will be in one place who needs to monitor the events or activities of students in another place. Hence surveillance-based intelligent e-learning can play the role of learning or monitoring the events from remote distance. Whenever the student logs in e-learning course, the camera in front of the person starts capturing the happenings. A self-learning hardware system with software support shown in Figure 1 can be used for learning the events from distance which shows framework for the design of intelligent e-learning system. Hence, this paper presents a machine-learning algorithm for unusual event detection. In the context of outdoor pedestrian area monitoring, the non-pedestrian entities in walkway like skaters, bikers, small carts, wheel chair are abnormalities. In e-learning scenario, while attending the course in virtual class room, activities of persons like walking, bending, hand waving and paper passing are defined as abnormal events. It detects the events of video by using a hybrid model of extracting two features namely motion and shape. The motion feature is obtained with histogram of optical flow orientation, magnitude and entropy (HOFME) and the appearance or shape information is obtained with histogram of oriented gradients (HOG). It classifies the events as usual or unusual based on the training given to classifiers. Two different classifiers such as Ctree and SVM are used for comparing the performance of the proposed method. This approach is suitable for

monitoring the activities of person undertaking e-learning course while learning the course or attending the online examination.

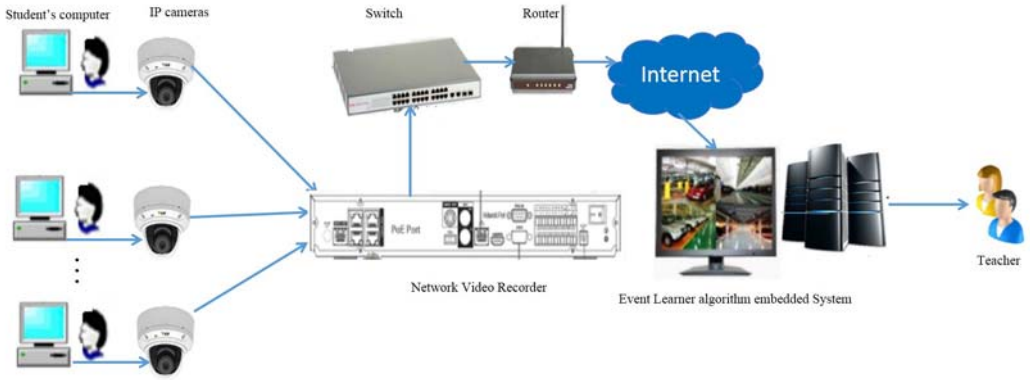


Figure 1: Framework for an Intelligent e-Learning system Design

The organization of this paper is as follows. Section II gives a detailed survey on unusual event detection and video summarization with comparison of various approaches used in literature. Section III describes the proposed approach in detail with the flow diagram. Section IV illuminates the experimental results and discussion with comparison table showing the performance comparison of the proposed approach with that of the existing methods. Section V gives the conclusion and future work.

LITERATURE REVIEW ON UNUSUAL EVENT DETECTION

Definition of Unusual event

The terminology event depends upon the scenario being considered in the application. It denotes what is happening in the area under coverage. In this paper abnormal event means that an event which happens unintentionally, abruptly and unexpectedly that needs an action to be taken. The generic flow diagram of unusual event detection is given in Figure 2.

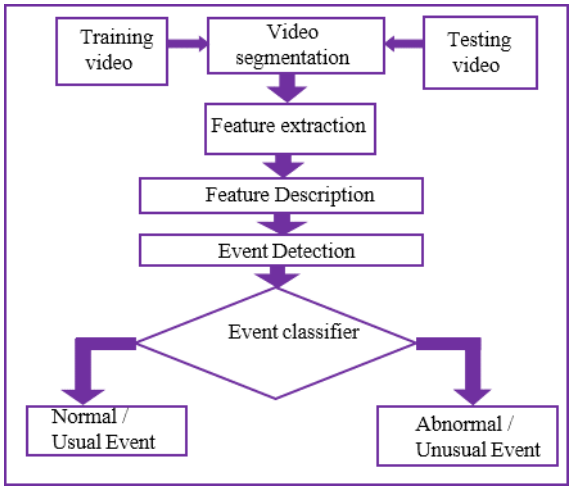


Figure 2: Flow diagram for Unusual Event Detection

Detailed survey on unusual event detection

An unsupervised approach has been developed by Hua Zhong (2004) to detect unusual activity in a large video. It detects the objects that are moving in the video and extract features using motion, color/ texture histograms. In the work proposed by Chen Change (2011), optical flow is calculated for every pixel in the region using Lucas-Kanade methods. Then a codebook is created and the Bayes classifier is used with a threshold to make a decision whether the event is abnormal or not. Gal Lavee (2005) uses the Nearest Neighborhood algorithm with Euclidean distance measure. A Neural network is trained using a back propagation algorithm and a decision tree is built with minimal entropy. Adam (2008) extracts information from regions and evaluates their normality. It uses Lucas-Kanade method for optical flow calculation and it considers both velocity and direction. Trajectory-

based anomalous event detection approach proposed by Claudio (2008) uses Support Vector Machines (SVMs). Although SVMs are used as tool for classification and clustering approaches, Claudio introduces novelty by using SVM to address the problem of anomalous event detection. In Reddy (2011), the anomaly detection is performed using region based approach which splits the scene into regions. It uses three feature descriptors namely average optical flow as a measure of speed, size and texture. Weilun Lao (2009) presents a framework of four processing levels for human behavior analysis. The levels are background modeling, object-based trajectory estimation, event-based semantic analysis and finally visualization which includes calibration of camera and reconstruction of 3-D scene. The work offered by Vijay Mahadevan (2010) combines both the spatial and temporal maps of anomaly detection. The method proposed by Zhigang Ma (2013) is named as Semantic Analysis via Intermediate Representation (SAIR). Ivanov (2009) proposes a method based on the acceleration and velocity of the objects in the scene for unusual event detection. A set containing macro-block motion vectors is used as feature for detection of abnormal event in compressed video streams (Nahum Kiryati 2008). Bin (2011) presents a sliding window technique to learn the initial dictionary of events. The occurrence of unexpected event in moving object environment can be detected early using statistical motion pattern formulated by Bayes rule as shown in equation (1).

The posterior probability,

$$P(\alpha_i \mid \beta) = \frac{P(\beta \mid \alpha_i)P(\alpha_i)}{\sum_{j=1}^{n} P(\beta \mid \alpha_j)P(\alpha_j)}; i = 1,2,...n \quad (1)$$

where, $P(\beta \mid \alpha_i)$ is the likelihood function, $P(\alpha_i)$ is priori probability and $\sum_{j=1}^{n} P(\beta \mid \alpha_j)P(\alpha_j)$ is the evidence.

A one-class classification method is used by Balakrishna Mandadi (2013) with an assumption that the training set consists only usual events. It models video by a bag-of-words method and uses a probabilistic approach for training data which uses Latent Dirichlet Allocation (LDA) framework. Kullback-Leibler divergence and Bhattacharya distance are used for the detection purpose.

Multi camera video surveillance system

Carter De Leo (2014) suggests anomaly detection in multi camera system by modeling the number of occurrence of activities as binomial distribution. Misrepresentation of anomalies are detected using PLCA mixture model. Than (2007) proposes a software system for tracking recurring events in multi-camera environment. It mainly focuses on people chasing in the manner of post-event analysis for the purpose of investigation of events. Hong Lin (2014) proposes a work for human activity identification system with multiple cameras. It can recognize activities even in cross-views. Wahyono (2016) handles real time processing of manifold data of four distinct cameras simultaneously using multi-threading approach and detects suspicious event automatically. In the work proposed by Hanning Zhou (2006), the video segments captured by different cameras are combined using a Coupled Hidden Markov Model (CHMM). The time dependency among the local activities are modeled with the formulation of Probabilistic Graphical Model (PGM). A Dynamic Time Warping (DTW) model is used and then optimization is done using Monte Carlo algorithm.

Comparison

Unusual event detection approaches used in the literature are compared and listed out in the Table 1. It lists out the various features and datasets being used. The literature publication frequency in IEEE for unusual event detection is shown in Figure 3.

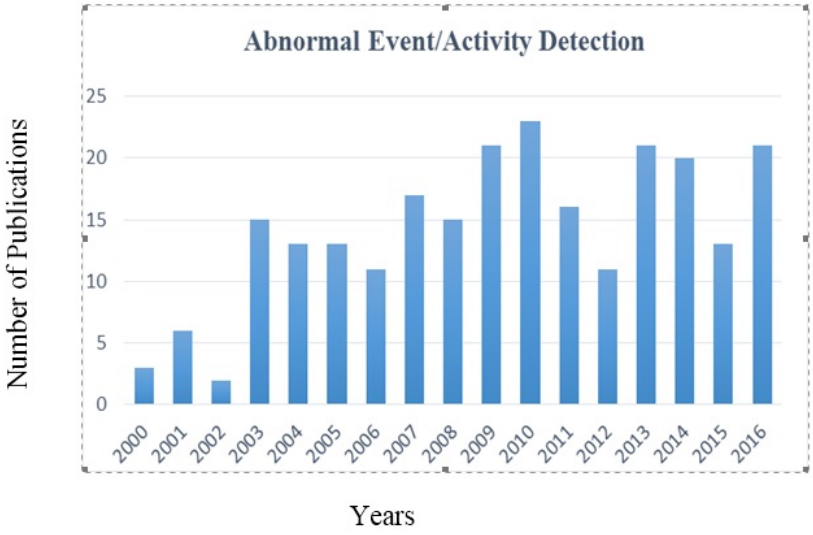


Figure 3: The frequency of publications in IEEE for abnormal event detection

Table 1: Comparison of state of art methods based on features and dataset

First author & Year	Methodology	Features	Dataset
Wahyono (2016)	Multi-threading strategy, GUI, Mixture of Gaussian model	--	i-LIDS
Hua Zhong (2004)	Histogram, Co-occurrence matrix, Bipartite graph co-clustering	Motion, color, texture	--
Chen Change (2011)	Lucas-Kanade method for optical flow, Bayes classifier	Motion	--
Gal Lavee (2005)	Nearest neighbor algorithm with Euclidean distance	Color, texture, shape	CanonZ100 camcorder Video
Adam A (2008)	Lucas-Kanade method for calculation of optical flow, pdf histogram.	Region based	--
Claudio (2008)	Support Vector Machine for trajectory based analysis	Spatial features	--
Reddy V (2011)	Region based segmentation, Cascaded model for classification.	Speed, size and texture	UCSD
Weilun Lao (2009)	Four level frame work for human motion analysis.	Shape	--
Vijay (2010)	Temporal and spatial anomaly maps are fusion and threshold detection	MDT, spatial saliency	UCSD
Zhigang (2013)	Semantic analysis via Intermediate Representation	Low level features	TRECVID MED11
Ivanov I (2009)	Acceleration and velocity of the objects are used for unusual detection.	Acceleration, velocity	PETS
Nahum (2008)	Probability density function of motion features.	Motion features	SONYTRV 900E PAL
Carter De (2014)	Probabilistic Latent component Analysis mixture model	Optical flow features	PETS 2001
Than V (2007)	Pixel-based background modeling, HOG for object classification,	--	MIT pedestrian dataset
Balakrishna (2013)	Latent Dirichlet Allocation (LDA) for Bag-of-words approach, Bhattacharya distance and Kullback-Leibler divergence for detection	Spatio-temporal features	Own video data captured

Hong Lin (2014)	Harris 3D and HOG/HOF descriptors, Bag-of-Words approach, view invariant feature representation.	Spatio-temporal interest points, global	IXMAS dataset
Hanning (2006)	Coupled Hidden Markov Model(CHMM)	Visual feature	Terrascope dataset

Video summarization:

The detected events are clustered together which can be used for summarizing the video covering the activity or behavior of candidates undertaking e-learning courses while learning and also attending the online examinations. Video summarization is helpful to quickly surf through the video instead of spending more time on it. Video summary is an abstract view of original video. It is constructed by concatenation of selected video segments called as key frames. The flow diagram for video summarization is shown in Figure 4.

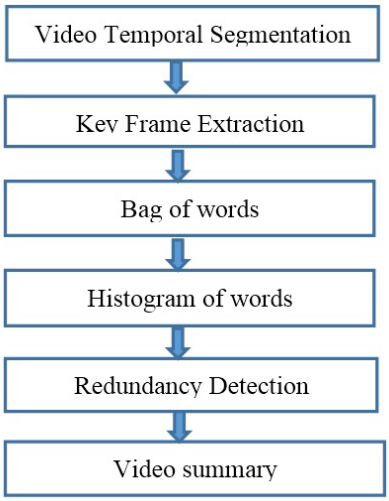


Figure 4: Flow diagram for video summarization

Video summarization types

Key frame based video summarization:

Key frames or representative frames are a set of significant images being extracted from video source. The summarized video is also known as static story board or still-image abstract.

Video skim based video summarization:

The original video is segmented in time as video clips of shorter duration. Each segment is then fused by a gradual or cut effect. This is also called as movie story board, moving image abstract or summary sequence. The best example for this video skimming is the trailer of the movie.

Video summarization Methodologies

A framework for video summarization starts from video temporal segmentation through key frame selection to redundancy detection. Shiyang Lu (2014) proposes a work to achieve statically summarized video by identifying key frames with significant local features. Yael Pritch (2009) suggests a new technique for organized object reading and for framing the ground truth to work with SVM classifier. Michael Gygli (2014) uses temporal super frame segmentation with estimation of low-level, mid-level, high-level features. Zheng Lu (2013) segments video into sub shots using static-transit grouping method which is useful for unstructured egocentric video analysis. The work proposed by Kadir (2001) employs subsampling technique based on motion for video summarization. The work proposed by Sandra (2008) extracts color and visual features from color histogram adaptation technique. Similar frames are clustered with K-means algorithm. Each cluster is represented by its key frames which represent video summary.

Multi camera surveillance video summarization

Carter De (2014) presents an approach for multi camera video summarization for handling intra activity redundancy and inter activity redundancy. It uses probabilistic latent component analysis (PLCA) algorithm for identifying higher level activities in video. In object based method proposed by Fatih Porikli (2004), object tracking is performed at each camera using background subtraction and mean-shift analysis. Then Bayesian belief network is employed to create a correspondence between various camera objects. The features and dataset used in the literature are listed in Table 2. The frequency of publications in IEEE under the topic of video summarization is shown as chart in Figure 5.

Table 2: Comparison of the state of the art methods based on features and dataset used

First author	Methodology	Features	Dataset
Carter De (2014)	Probabilistic Latent Component Analysis(PLCA)	Motion	Own dataset, PETS2001
Shiyang (2014)	Bag-of-Importance model with group sparse property.	Local visual features	Open video project OVP, VSUMM, Youtube.
Michael (2014)	Super frame temporal segmentation, selecting visual interestingness, combining subset of super frames	Temporal gradients, motion features	SumMe dataset, Berkeley Segmentation dataset.
Zheng Lu (2013)	Static-transit for segmenting video into sub shots, SVM classifiers	Optical flow, blur feature	UT Egocentric (UTE), ADL
Kadir A. (2001)	Temporal sub-sampling of motion	Motion	--
Sandra (2008)	Histogram adaptation, line profiles, K-means algorithm	Color , visual descriptors	Open video storyboard.
Fatih (2004)	Background subtraction, mean-shift analysis, Bayesian belief	Object based	ETRI dataset

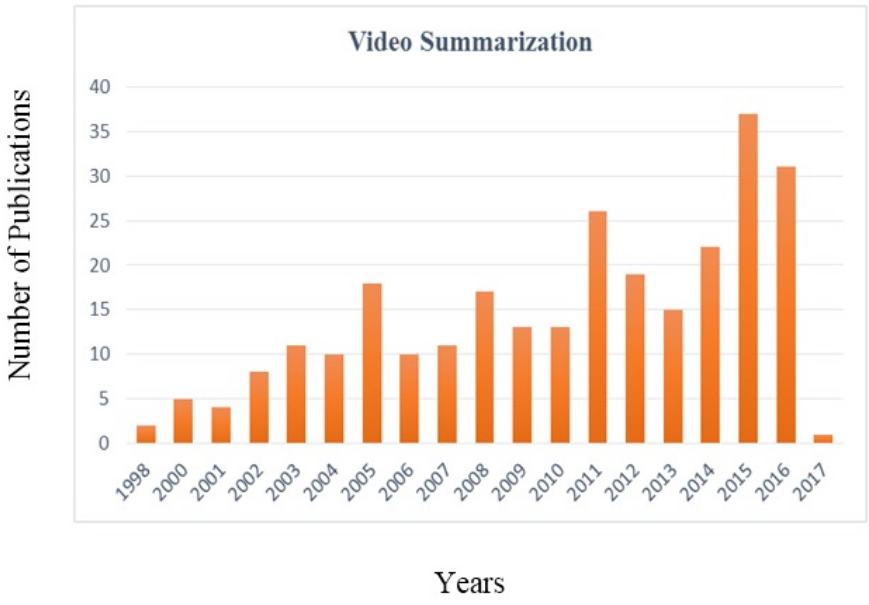


Figure 5: Frequency of publications in IEEE for video summarization

The literature survey gives an exhaustive idea for selection of methodology, dataset and features. The frequency of publications in unusual/abnormal event detection and video summarization depict the focus of researchers in those topics during the period of 2000 to 2016 and 1998 to 2017(January) respectively.

Limitations of existing works:

- The existing work of Colque (2016) results in moderate performance due to lack of shape and appearance information in extracted feature.
- In the existing work, classifier was trained with only normal patterns. During test phase, if the events differ from normal patterns then they were considered as anomalous which results in the equal error rate of is 32% for UCSD Peds 1 dataset .

Contributions:

- Exhaustive Literature survey on unusual event detection and video summarization
- A new hybrid model with motion and appearance information is being used to describe the event or activity in precise manner.
- In training phase, both normal and abnormal event features are used to improve the performance of the classification.

PROPOSED METHODOLOGY

The video covering the activity or event is preprocessed to make it suitable for further processing. The abnormal event / activity in video is detected by combining the motion information and appearance information. The motion feature is extracted using optical flow with histogram of optical flow orientation magnitude and entropy (HOFME). The appearance or shape feature is extracted with histogram of oriented gradient (HOG). The flow diagram for the proposed work is shown in Figure 6.

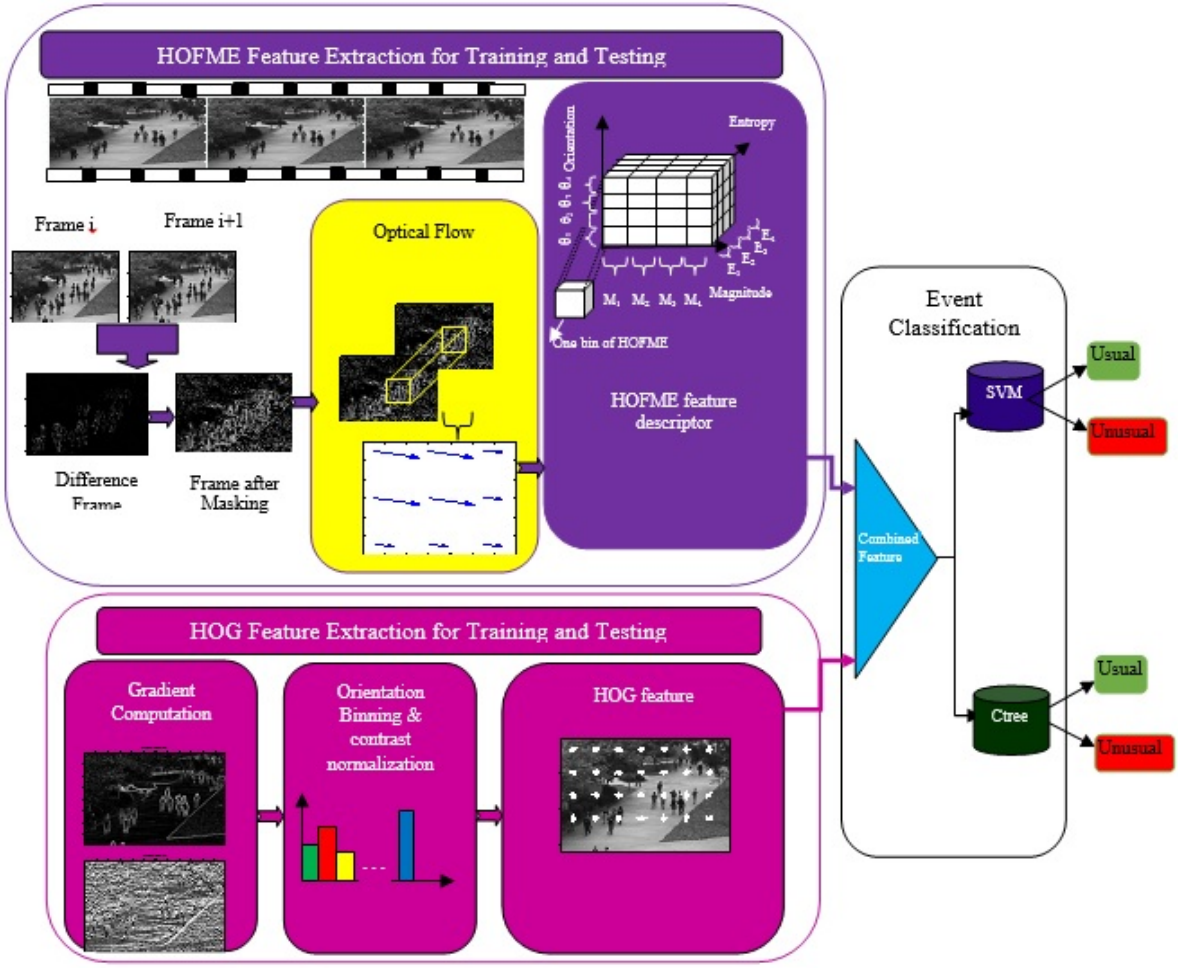


Figure 6: Proposed work flow

Histogram of Optical flow Orientation Magnitude and Entropy

The apparent motion of persons or objects from one frame to the next is represented by optical flow vector. In this paper optical flow is estimated using the pyramidal implementation of Lucas-Kanade algorithm.

Preprocessing:

The video is divided into non-overlapping spatio-temporal regions known as cuboids. As a first step, the video frames are converted to double precision images. Instead of extracting optical flow for a whole image, which is computationally expensive, each frame is compared with its next frame to find pixels having significant amount of motion. It is achieved with frame differencing between current frame f_c and next frame f_n. The pixel having the difference value greater than a threshold is considered and used for cuboid construction, otherwise the pixel is discarded (Colque 2016).Each cuboid consists of '*r*' number of rows '*c*' number of columns of image pixels with '*t*' number of frames ($r \times c \times t$).

Pyramidal Implementation of Lukas-Kanade Optical Flow:

The assumptions made in the Lucas-Kanade method are:

- Constant illumination assumption
- Spatial coherence between frames
- Motion in a small neighborhood is equivalent

For each and every pixel in the current frame C and next frame N, the optical flow in horizontal and vertical directions are determined (Jean-Yves Bouguet 2002) . Let a pixel intensity in current frame $C(x,y)=u=[u_x\ u_y]^T$ is displaced to $(x+d_x,y+d_y)$ in next frame N. The cost function or mean square error is given in equation (2).

$$f(d)=f(dx,dy)=\sum_{x=u_x-w_x}^{u_x+w_x}\sum_{y=u_y-w_y}^{u_y+w_y}(C(x,y)-N(x+d_x,y+d_y))^2 \qquad (2)$$

Where, w_x and w_y denote window size containing equivalent motion neighborhoods. The pseudo-code for pyramidal implementation of Lukas-Kanade optical flow algorithm is given in the algorithm 1.

Algorithm 1: Pyramidal implementation of Lukas-Kanade optical flow

Aim:For a point ***u*** on the image *C* its respective position on the image *N* is found using the following steps.

S-1: Construct pyramidal representations of *C* and *N*. $\{C^L\}_{L=0,1,\ldots,Lm}$ and $\{N^L\}_{L=0,1,\ldots,Lm.}$

S-2: Initialize the pyramid. $g^{Lm}=[\ g_x^{Lm}\ \ g_y^{Lm}\ ]^T=[0\ 0]^T.$

Loop1: for L=Lm down to 0 with step change of -1

Position of point u on image C^L: $u^L=[p_x\ p_y]^T=u/2^L.$

Derivative of C^L with respect to x: $C_x(x,y)=\frac{C^L(x+1,y)-C^L(x-1,y)}{2}$

Derivative of C^L with respect to y: $C_y(x,y)=\frac{C^L(x,y+1)-C^L(x,y-1)}{2}$

Spatial gradient matrix: $S=\sum_{x=p_x-w_x}^{p_x+w_x}\sum_{y=p_y-w_y}^{p_y+w_y}\begin{bmatrix}C_x^2(x,y) & C_x(x,y)C_y(x,y)\\ C_x(x,y)C_y(x,y) & C_y^2(x,y)\end{bmatrix}$

Initialize the iterative L-K: $\vec{v}^0=[0\ \ 0]^T$

Loop2: for k=1to K with step of 1

Image difference: $\delta C_k(x,y)=C^L(x,y)-N^L(x+g_x^L+v_x^{k-1},y+g_y^L+v_y^{k-1})$

Image mismatching vector: $b_k=\sum_{x=p_x-w_x}^{p_x+w_x}\sum_{y=p_y-w_y}^{p_y+w_y}\begin{bmatrix}\delta C_k(x,y)C_x(x,y)\\ \delta C_k(x,y)C_y(x,y)\end{bmatrix}$

Optical flow (Lucas-Kanade): $\vec{\eta}^k=G^{-1}b_k$

Next iteration assumption: $\vec{v}^k=\vec{v}^{k-1}+\vec{\eta}^k$

End of Loop 2 on k

Optical flow at level L: $d^L=\vec{v}^k$

Assumption for next level L-1: $g^{L-1}=\begin{bmatrix}g_x^{L-1} & g_y^{L-1}\end{bmatrix}^T=2\left(g^L+d^L\right)$

End of Loop1 on L

Final optical flow vector: $d=g^0+d^0$

Position of point on N: $v=u+d$

Solution: The respective point is located at v on image N.

HOFME Feature descriptor

The optical flow data of each cuboid ($r \times c \times t$) is used for feature vector construction. Hence total number of matrices of optical flow is one less than the number of frames in a video (nof-1). For each cuboid of the optical flow, its magnitude and orientation values are obtained. The entropy is computed from orientation matrix (Colque 2016). The orientation distribution is first obtained around a pixel *p* using its *m* values of neighborhoods (*m*=4) forming a patch. With the probabilities of the distribution p_i the entropy is calculated using the formula in equation (3). As there are three parameters namely orientation, magnitude and entropy, the histogram of these features is built as a cuboid with three co-ordinates as shown in Figure 4. Each feature value is quantized into four bins. The bin ranges of orientation (θ), magnitude (M) and entropy (E) are{(θ_1:0° to 90°), (θ_2:90° to 180°), (θ_3: 180° to 270°), (θ_4:270° to 360°)}, {(M_1:0 to 20), (M_2:20 to 40), (M_3:40 to 60), (M_4:60 to ∞)} and {(E_1:0 to $\frac{1}{2}$), (E_2:$\frac{1}{2}$ to 1), (E_3:1 to $\frac{3}{2}$), (E_4: $\frac{3}{2}$ to 2)} respectively. For example, a pixel with features (θ=110°, M=2, E=1) will fall under the bin of (θ_2, M_1, E_2). Hence each and every pixel with 3 features is grouped into one of the 64 bins in the feature cuboid.

The entropy for a pixel around a patch of *m* neighborhoods,

$$e(p_i) = -\sum_{i=1}^{m} \theta(p_i) \log[\theta(p_i)] \qquad (3)$$

HOG Feature

The histogram of oriented gradient (HOG) feature is extracted by following the steps shown in Figure 7. HOG features encode local shape and appearance information from regions within each frame. In HOG descriptor, the detector window is tiled with dense cell grids where each cell comprises histogram of gradient orientation bins each weighted by gradient magnitude. These features can be further utilized for classification of events in a video scene.

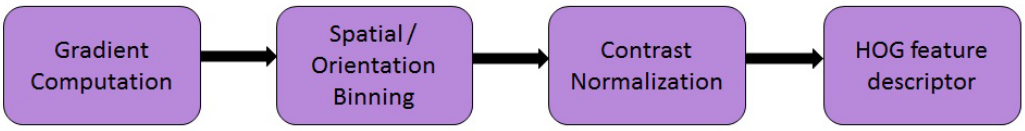


Figure 7: HOG feature extraction flow diagram

Gradient Computation

The input frame is passed to the central difference filter [-1 0 1] in order to compute the gradient. The forward difference is used to find the gradients at image borders. The gradient directions are determined in counterclockwise from positive x-axis and the measured angles are in the range of -180° to 180°.

Spatial/Orientation Binning

The subsequent step is to compute a weighted vote for an oriented histogram channel which is done according to the gradient component orientation centered on each pixel. Then, cells are formed by accumulating the votes into orientation bins. Cells may be in radial or rectangular form. Orientation bins can be spaced in the range of 0° to 180° for unsigned gradient or 0° to 360° for signed gradient. The bilinear interpolation is performed on votes between the neighboring bin centers in both location and direction to diminish the aliasing effect.

Contrast Normalization

Local contrast normalization is essential to achieve remarkable performance, since strengths of gradient differ over a wide range due to foreground-background disparity and local illumination variations. Cells are clustered into larger spatial blocks and then contrast normalization is performed on each block individually. An overlap stride of as a minimum half the block size is chosen to make sure adequate contrast normalization.

HOG feature descriptor

The HOG feature descriptor is a vector of all the elements of normalized cell responses from all of the blocks. Larger the block overlap values acquire more information at cost of increased feature vector size which improves

the performance. As HOG feature is computed for each frame but HOFME feature for pair of frames it is enough to calculate HOG one less than the number of frames in video (nof-1).

Combined Feature Descriptor

The motion information extracted using optical flow is combined with the appearance information extracted from HOG. Hence the histogram of oriented gradient (HOG) feature vector of size $(1 \times g)$ is appended with the histogram of optical flow orientation, magnitude and entropy (HOFME) having size of $(1 \times h)$. Hence the hybrid model results in the feature vector of size $(1 \times (g + h))$. The combined feature outperforms well in detecting the unusual event compared to the existing methods.

Classification

The combined feature vector is used as input for classifier. Large number of frames are used for training the classifier with frames having normal abnormal scenes along with training label. Less number of frames are chosen for testing phase compared to that used for training. The features extracted from training frames containing normal and abnormal patterns are used as test vectors with testing label. Two classifiers are used for performance comparison namely classification decision tree (Ctree) and support vector machine (SVM) classifier. Classification decision tree is modelled with the training data and their corresponding classification label using binary splits. The model is used to predict the test data with the known or trained input-output data history of the model. SVM classifies test data using a trained support vector machine. It splits the data into two classes with hyper plane selected by sequential minimal optimization (SMO) and do binary classification.

EXPERIMENTAL RESULTS AND DISCUSSION

The input video is converted into frames. The frames are preprocessed to have gray scale (if they are in RGB color space) and frame differencing is applied on two successive frames. If the absolute difference is smaller than a threshold then the pixel is discarded, otherwise the pixel is considered for cuboid construction. The binary mask is applied over those two frames under consideration in order to get the cuboid of moving pixels alone. The cuboid is constructed with spatial window size of 30x30 (r=c=30) and t=2 frames (the current and succeeding frame). The optical flow is computed for each cuboid using Lucas Kanade Thomasi pyramidal implementation. As the optical flow is operated pixel wise 900 magnitude values and orientation values are obtained. The orientation parameter is used for entropy calculation. The optical flow orientation, magnitude and entropy are used as three co-ordinates to build feature cuboid of size 4x4x4 as there are four bins in histogram of each parameter. Hence the optical flow feature vector size is of (1x64) which represents motion pattern. As the performance of motion feature alone in anomalous event detection is not remarkable (Colque 2016), appearance information is also included. The appearance or shape data is extracted as the histogram of oriented gradient HOG feature using Dalal and Triggs approach. It results in a feature vector having size of (1x648) for each frame in UCSD peds1 scenario. Both the features are combined, the resultant feature vector has size of (1x712) for each frame. Training video features and testing features are passed to classifiers. In training phase, frames with normal and abnormal events are chosen and their corresponding labels are prepared to train the classifiers. The trained model is then used to classify the test data.

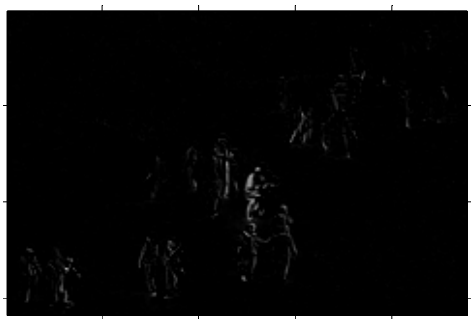
UCSD peds1 Dataset:

Matlab 2014a with image processing tool box is used for experimentation. The experimentation is performed on UCSD dataset which is out door scenario. The UCSD dataset is a publicly available annotated dataset for anomaly detection featuring pedestrian walkways (V. Mahadevan 2010). In this work the video sequences with only pedestrians are considered as normal event and the presence of non-pedestrian entities or unusual pedestrian motion pattern is marked as abnormal event. The UCSD ped1 dataset consists of 34 training video samples and 36 testing video samples. The unusual event detection accuracy is evaluated based on the criterion of frame-level, as the algorithm predicts the frame containing unusual event and compares with ground-truth annotations. Each video in peds1 has 200 frames hence the HOG and HOFME features are obtained for 199 frames of each. The frame showing normal event is displayed in Figure 8 (a). The difference between the successive frames is shown in 8 (b). Figure 8 (c) presents frame containing abnormal entity of a cycler and 8 (d) shows its difference from successive frame. The optical flow vector and HOG features are shown in 8 (e) and 8 (f) respectively.

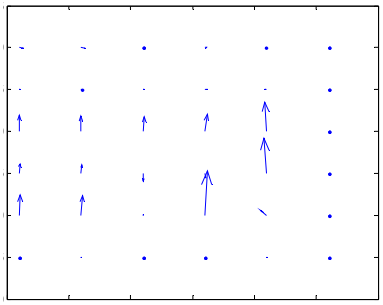
(a) Outdoor pedestrian path way

(b) Difference Frame

(c) Frame containing abnormal event



(d) Difference of successive frames



(e) Optical Flow for 2 successive masked frames

(f) HOG descriptor visualization

Figure 8: Frames starting from input stage to feature description stage of training and testing for UCSD peds 1 dataset

The algorithm predicts the events in UCSD Peds 1 dataset and classifies with accuracy values of 81.03% and 82.66% and using Ctree and SVM classifiers respectively. The equal error rate which is a measure of percentage of mis-classified frames is calculated as 18.97% and 17.34% using Ctree and SVM classifiers respectively. The performance of proposed work is shown by metrics of precision, recall and F1-measure in Figure 9. The

performance metric values of accuracy, precision, recall and F1-measure for Ctree and SVM are listed in Table 3. The receiver operating characteristics of the proposed approach of using the combined feature (HOG and HOFME) with three different classifiers are shown in Figure 10. The receiver operating characteristics of Ctree and SVM classifiers show that the classifiers perform remarkably well in classifying events since the curves far above the linear curve which separates the area diagonally.

Table 3: Performance comparison table of proposed work with different classifiers for UCSD peds1 dataset

Performance Metric	Classifier type	
	Ctree	SVM
Precision	0.6941	0.7051
Recall	0.9697	1
F1-measure	0.8091	0.8270

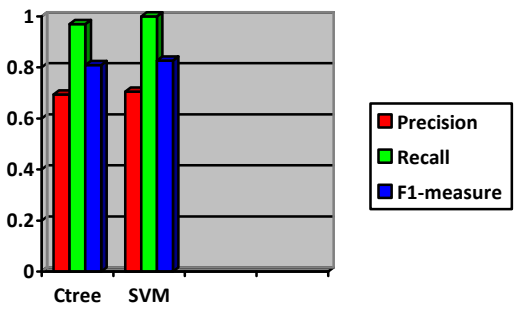
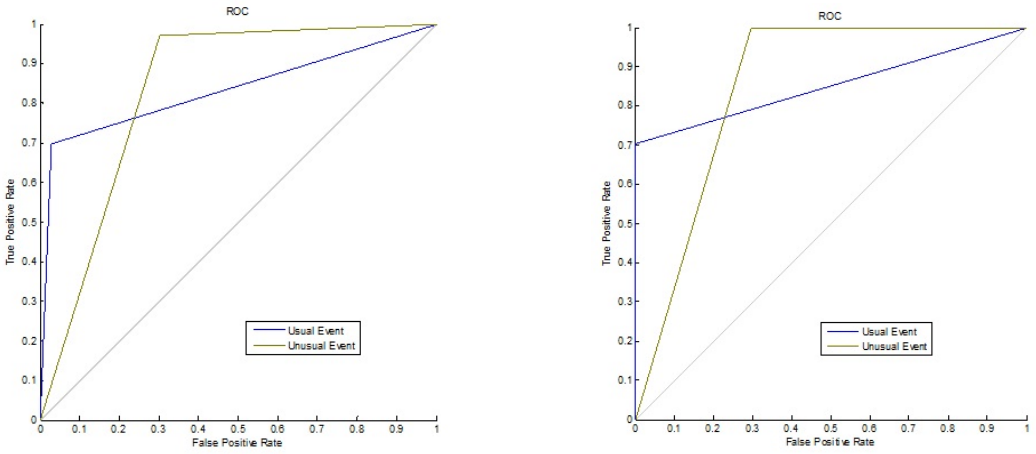


Figure 9: Different classifier performance comparison for UCSD peds 1 dataset



a) Ctree b) SVM

Figure 10: Receiver Operating Characteristics of proposed work with two classifiers for UCSD Peds 1 dataset

Comparison of various approaches

Quantitative results of proposed work in terms of Equal error rates (EERs), and Area Under Curve of ROC for Frame-level abnormality detection performed on UCSD Ped1 dataset are given in Table 4. The values of proposed work are compared with that of Colque (2016) and other existing works in Table 4. The method proposed by Colque (2016) extracts only the motion pattern with HOFME feature, it does not include any appearance information which is very essential in learning the events of a video. Hence the proposed work includes appearance or shape feature using HOG along with motion feature using HOFME descriptor. The proposed methodology achieves remarkable improvement in performance compared to published state of the art methods

Table 4: Quantitative comparison of proposed approach with existing methods for UCSD peds 1 dataset

Method	**EER (Equal Error Rate)**	**Area Under Curve (AUC) of ROC**
Social Force (V. Mahadevan 2010)	31%	67.5%
SF-MPPCA (V. Mahadevan 2010)	40%	59%
MDT (V. Mahadevan 2010)	25%	81.8%
MPPCA(Yang 2013)	40%	20.5%
Adam (Yang 2013)	38%	13.3%
Sparse (Yang 2013)	19%	46.1%
Yang (Yang 2013)	23%	47.1%
HOFME (Colque 2016)	33.1%	72.7%
Proposed HOG+HOFME		
-Ctree	18.97	94.25
-SVM	17.34	-

Indoor e-class room scenario

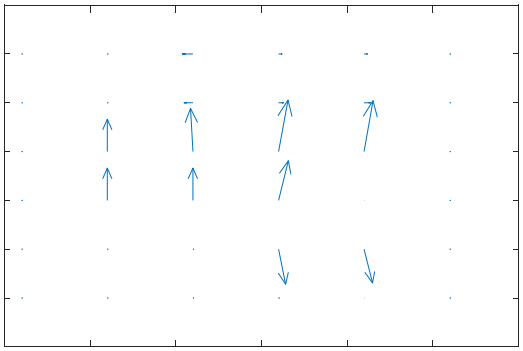
The video has been captured in e-learning class room with students listening an e-course. In this scenario, the activities of students like listening, writing are considered as normal events while walking, bending, passing a paper are considered as abnormal or unusual events. The video consists of frames having normal event and abnormal events with the frame size of 480x864x3 in RGB space. The frames are being converted to gray scale and resized to size of (300x600). Then the steps proposed in the flow diagram are followed to get HOFME motion feature with size of (1x64) and HOG appearance feature of size (1x4896). The hybrid model results in the feature size of (1x4960). The sample frame of events captured in an e-classroom video is shown in Figure 11(a), the gray scale version of the frame is shown in 11 (b) which does not have any abnormality as all students are listening the e-lecture. One of the abnormal activities mentioned above namely paper passing among students is shown in 11 (c) which is processed to extract region having significant motion as displayed in 11 (d).

(a) Indoor e-classroom-RGB

(b) Indoor e-classroom-Gray scale

(c) Frame containing abnormal events

(d) Difference Frame containing major movement region in paper passing event



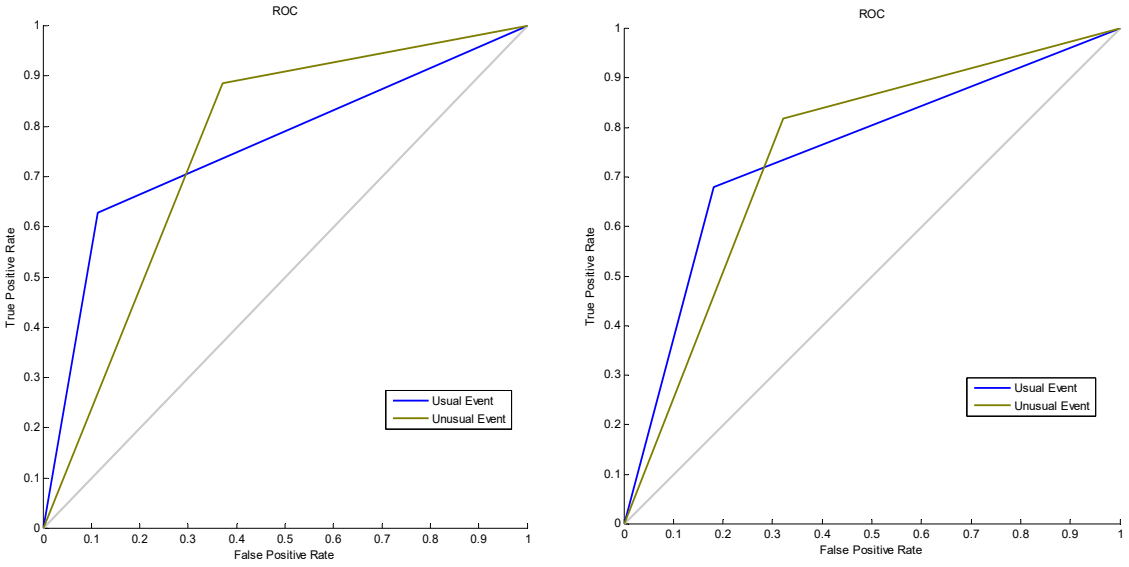
(e) Optical Flow for 2 successive masked frames

(f) HOG visualization

Figure 11: Frames starting from input stage to feature description stage of training and testing for e-learning classroom video

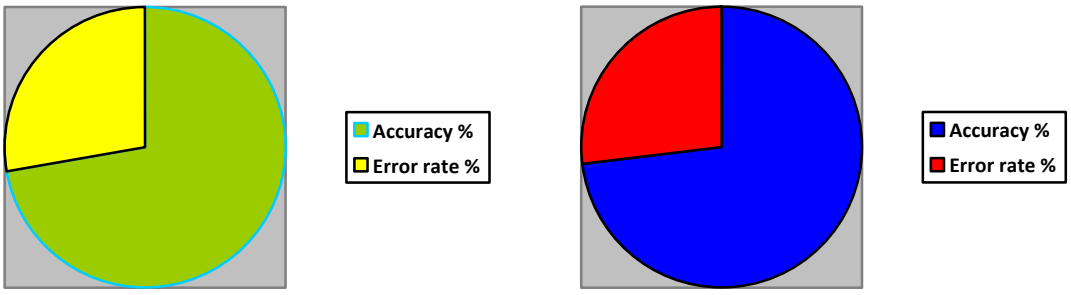
The optical flow is computed for the frame shown in 11 (c) and its consecutive frame after applying the mask over the frames. The computed optical flow vector has two parameters orientation and magnitude showing the velocity with which pixel move in consecutive frames which is displayed in 11 (e). The histogram of oriented gradients is plotted on the frame for visualization as shown in 11 (f).

The extracted features are combined to have motion and shape information with feature vector size of (1x4960). The classification tree (Ctree) and support vector machine (SVM) classifiers are trained with training frame features and an unknown frame is given for test to classify it as normal or abnormal case. The receiver operator characteristics of each classifier are shown in Figure 12. Both the classifiers yield good classification performance as the ROC far above the linear curve still the SVM classifier yields better results compared to Ctree. The area under the curve of ROC of Ctree is 80.96% and that of SVM is 77.99%. The Ctree classifier results in accuracy of 72.13% with error rate of 27.87% and SVM classifier predicts test samples with accuracy of 72.95% and error rate of 27.05% which is pictorially shown in Figure 13.The performance metrics like precision, recall and F1-measure of Ctree are 0.5735, 0.8864 and 0.6964 and that of SVM are 0.59, 0.8182 and 0.6857 respectively shown in Table 5 which are represented as bar chart in Figure 14.



(a) Ctree (b) SVM

Figure 12: Receiver operating characteristics of (a) Ctree and (b) SVM classifiers for e-learning classroom video



(a) (b)

Figure 13:Accuracy and error rate of (a) Ctree (b) SVM for e-learning classroom video

Table5: Performance comparison table of proposed work with different classifiers for e-learning classroom video

Performance Metric	Classifier type	
	Ctree	SVM
Precision	0.5735	0.59
Recall	0.8864	0.8182
F1-measure	0.6964	0.6857

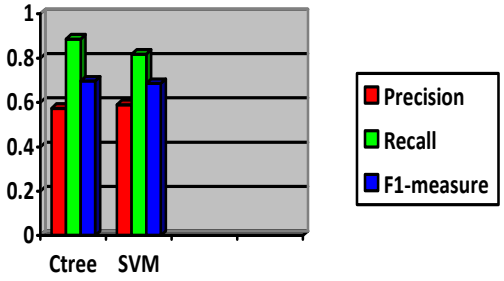


Figure 14: Comparison of performance metrics of Ctree and SVM classifiers for e-learning classroom video

CONCLUSION AND FUTURE WORK

This paper presents a methodology for learning the event remotely to monitor the behavior of students undertaking e-learning courses. The video surveillance system is applied for an intelligent e-learning system to monitor the activities of students while learning and attending online examinations. The proposed approach extracts the combined feature of motion and appearance from the video. The motion pattern is obtained from histogram of orientation, magnitude and entropy of optical flow (HOFME) and as an improvement over this approach (Colque 2016), appearance is included using histogram of oriented gradients (HOG). The combined feature is used for training and testing different classifiers Ctree and SVM to detect and classify whether the activity of the students is usual or unusual. The proposed approach is experimentally verified on a publicly available dataset for anomaly detection named UCSD peds1. The results show that the algorithm outperforms well compared to the existing published works. Hence abnormal event detection applicable for student monitoring goal in intelligent e-learning is possible using a video surveillance system with this algorithm embedded on it which is proved by the results obtained with indoor e-learning classroom video. The detected events can be combined to have a summarized video to quickly surf through the activities or behavior of e-learning candidates. The detailed survey for video summarization is also presented in this paper. It lists out techniques, datasets, features used in the literature for summarizing a video. An appropriate method can be derived from the survey and used for video summarization in future.

REFERENCES

Adam, A., Rivlin, E., Shimshoni, I., & D. Reinitz (2008). Robust real-time unusual event detection using multiple fixed location monitors. *In IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.30, Issue 3 (pp.555-560).

Avila Sandra, E. F. de, Antonio da Luz Jr and Araujo Arnaldo de A. (2008). VSUMM: A Simple and Efficient Approach for Automatic Video Summarization. In *International Conference on Systems, Signals and Image Processing*.

Balakrishna Mandadi, Amit Sethi (2013). Unusual event detection using sparse spatio-temporal features and Bag-of-words model. *In IEEE International Conference on Image Information Processing.*

Cardillo G. (2008) ROC curve: compute a Receiver Operating Characteristics curve from http://www.mathworks.com/matlabcentral/fileexchange/19950.

Carter De Leo, B.S.Manjunath (2014). Multi camera video summarization and Anomaly Detection, *ACM Transaction on Sensor Networks*, Vol.10, No.2.

Chen Change Loy, Tao Xiang Shaogang Gong (2011). Stream-based active unusual event detection. *In Tenth Asian conference on Computer Vision*, Vol. part I, Springer-Verlag (pp.161-175).

Claudio Piciarelli, Christian Micheloni and Gian Luca Foresti (2008). Trajectory- based anomalous event detection. *In IEEE Transaction on Circuits and Systems for Video Technology*, Vol.18, No.11.

Fatih Porikli (2004). Multi – Camera Surveillance: Object-based Summarization Approach. *First Printing*, TR-2003-145.

Gal Lavee, Latifur Khan and Bhavani Thuraisingam (2005). A framework for a video analysis tool for suspicious event detection. *MDM/ KDD*, USA.

Hanning Zhou, Don Kimber (2006). Unusual event detection via multi-camera video mining. *In IEEE International Conference on Pattern Recognition*.

Hong Lin, Lekha Chaisorn, Yongkang Wong, An-An Liu, Yu-Ting Su, S. Mohan Kankanhalli (2014). View-invariant feature discovering for multi-camera human action recognition. *In IEEE International Workshop on Multimedia Signal Processing*.

Hua Zhong, Jianbo Shi and Mirkov Visontai (2004). Detecting unusual activity in video. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 2 (pp. 819-826).

Ivanov, Dufaux, T.M. Ha and Ebrahimi 2009. Towards generic detection of unusual events in video surveillance.

In sixth IEEE International Conference on Advanced Video and Signal based Surveillance (pp.61-66).

Jean-Yves Bouguet (2002). Pyramidal implementation of the Lucas kanade Feature Tracker description of the algorithm.

Jian Yu (2009). An Infrastructure for Real-time Interactive Distance e-Learning Environment. *In First international conference on Information Science and Engineering*.

Kadir Pekerm, A., Ajay Divakaran and Huifang Sun (2001). Constant pace skimming and temporal sub-sampling of video using motion activity. *In International Conference on Image Processing*.

Mahadevan, V., W. Li, Bhalodia V. & Vasconcelos N. (2010). Anomaly Detection in Crowded Scenes. *In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, San Francisco, CA.

Matteo Taiana, Athira Nambiar, Jacinto Nascimento, Dario Figueira and Alexandre Bernardino (2014). A Multi-camera Video Dataset For Research on High-Definition Surveillance. *In International Journal of Machine Intelligence and Sensory Signal Processing*, Vol.1, Issue 3.

Michael Gygli, Helmu Grabner, Hayko Riemenschneider and Luc Van Gool (2014). Creating summaries from user videos. *In European Conference on Computer Vision*.

Nahum Kiryati, Tammy Riklin Raviv, Yan Ivanchenko and Shay Rochel (2008). Real-time abnormal motion detection in surveillance video. *In International Conference on Pattern Recognition*.

Reddy, V., Sanderson, C., & Lovell B.C. (2011). Improved anomaly detection in crowded scenes via cell-based analysis of foreground speed, size and texture. *In IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp.55-61).

Rensso Mora Colque, et.al. (2016). Histogram of optical flow orientation and magnitude and entropy to detect anomalous events in videos. *In IEEE Transactions on Circuits and Systems for Video Technology*.

Shiyang Lu, Zhiyong Wang, Taomei, Genliang Guan and David Dagan Feng (2014). A Bag-of-Importance model with locality-constrained coding based feature learning for video summarization. *In IEEE Transaction on multimedia*, Vol.16, No.6.

Than V. Pham, Marcel Worring, Anrold W.M, Smeulders (2007). Multi-camera visual surveillance system for tracing of reoccurrences of people. *In IEEE International Conference on Distributed Smart Cameras*.

Vijay Mahadevan, Weixin Li, Viral Bhalodia and Nuno Vasconcelos (2010). Anomaly detection in crowded scenes. *In IEEE Conference on Computer Vision and Pattern Recognition* (pp.1975-1981).

Wahyono, Alexander Filonenko and Kang-Hyun Jo (2016). Designing interface and integration framework for multi-channel intelligent surveillance system. *In IEEE Conference on Human system interactions*.

Weilun Lao, Jungong Han and H.N. Peter (2009). Automatic video-based human motion analyzer for consumer surveillance system. *In IEEE Transactions on Consumer Electronics*, Vol.55, No.2.

Yang Cong, Junsong Yuan and Yandong Tang (2013). Video Anomaly Search in Crowded scenes via Spatio-Temporal motion context. *In IEEE Transactions on Information Forensic and Security.*

Yang Wu, Yasutomo Kawanishi, Michihiko Minoh, Masayuki Mukunoki and Shinpuhkan (2014). A Multi-Camera Pedestrian Dataset for Tracking People Across Multiple Cameras.

Yasmin Khan, S., Soudamini Pawar (2015). Video summarization: survey on event detection and summarization in soccer videos. *In International Journal of Advanced Computer Science and Applications*, vol.6, No.11.

Zaynab El Khattabi, Youness Tabii, Abdelhamid Benkaddour (2015). Video Summarization Techniques and Applications. *In International Journal of Computer, Electrical, Automation, Control and Information Engineering* Vol:9, No:4.

Zheng Lu and Kristen Grauman (2013). Story-driven summarization for egocentric video. *In IEEE Conference on Computer Vision and Pattern Recognition*.

Zhigang Ma, Yi Yang, Nicu Sebe, Kai Sheng and G.Alexander Hauptmann (2013). Multimedia event detection using a classifier-specific intermediate representation. *In IEEE Transaction on Multimedia*, Vol.15, No.7.

