Suspicious Human Activity Recognition Using Statistical Features

Hanan Samir[1], Hossam Abd El Munim[2], and Gamal Aly[2]

[1]Electromechanical Institute, Housing and Building National Research Center, Cairo 12611
[2]Computer and Systems Engineering Department, Faculty of Engineering, Ain Shams University, Cairo 11517

***Abstract*—This paper presents a new algorithm for suspicious human activity recognition in videos based on a combination of two different feature types. The first feature concerns the shape and is called shape moments. The second concerns the boundary coordinates and is called "Histogram of Normalized Distances (HND) from Center of gravity of the object shape (COG) and it's contour points" combining these features leads to the formation of a strong complementary feature vector that captures effective discriminate details of human action videos. The authors used two methods for classification, the Multi-class Support Vector Machine and Naive Bayes classifier. The classification by using the Multi-class SVM classifier verified recognition rate up to 95.6 %, but the Naive Bayes classifier verified 97.2%. The authors evaluated the suspicious activity recognition on 250 videos from HMDB data set. Five distinct suspicious human activities (e.g., Running, Punching, Kicking, Shooting guns and Falling floor, etc.) by 250 different persons. Experiments on HMDB show that the presented system can recognize suspicious activities effectively and accurately in surveillance videos.**

I. Introduction

Human activity recognition is one of the growing areas of artificial intelligence and computer vision. The need for automated surveillance systems have become urgent as reliance on the human factor gives inaccurate results in the recognition of suspicious activities. Public places like subway stations, airports, and government buildings require detection of abnormal and suspicious activities to prevent crime before an occurrence such as automatic reporting of a person with a bag loitering at the airport and to overcome acts of sabotage. Automated surveillance systems have other benefits than to identify suspected behavior such as monitoring of patients, children, and elderly persons. Where the elderly persons over the age of 80 will fall at least once a year. Most people are unable to get up on their own if medical care is not provided quickly the period of time spent on the floor and reduce their autonomy. Falls can cause fatal injuries if medical care is not given in time. Therefore, the need for an automated surveillance system is urgent to recognize the unusual activities like fainting and falling Floor to give elder people the ability to live more peacefully. Our contribution in this research that the system is able to track moving objects, identify activities and differentiate between normal and dangerous activities. The rest of this paper is organized as follows. Section II shows the related work. Section III introduces proposed system overview. Section IV shows the experiments. The results in section V and the conclusion, in section IV.

II. Related work

In recent years the field of suspicious human activity recognition draws the attention of many researchers, according to its importance to prevent crimes before the occurrence. Many papers were written for human activity recognition. In paper [1] the authors suggest a system use face profile, gait silhouette and using a confident-frame-based recognizing algorithm (CFR) to recognize human activities. The frame with high confidence is used to recognize the activity. In a paper [2] the authors used optical flows in object detection and they also explained the stages involved in video surveillance. Hidden Markov model was used in [3] to model human behavior as a stochastic sequence of actions. The authors explain actions using a feature vector containing a set of local motion descriptors and trajectory information. In another important paper [4] the authors use hand recognition, face recognition and body gestures to detect suspicious activities like object exchange, peeping into others answer sheet in examination halls.

III. Proposed System Overview

The entire recognition process can be divided into stages: the first stage is to detect and track moving objects from a frame to another frame over time by using blob parsing to detect sets of connected pixels that are compatible with moving objects, The challenge when the system tracks moving objects is the occlusions. The location of objects is not available when they are blocked by others or things. The proposed system, therefore used the Kalman candidate [5], [6] to overcome these missing measurements, then we use multi-threshold segmentation method for splitting the moving objects from the background since it's the most effective tool. Slicing techniques and histogram thresholding can also be combined with pre and post processing techniques to segment an image. The threshold sets the grayscaling images to a binary image, and the image is divided into two parts determined by the pixel values 0 and 1 that represent the background and the object. The second stage of the proposed system is to extract the perimeter of the object by using the canny algorithm for edge detection [7] then the system determines the Center of gravity of the segmented object (COG) and determines the bounding box for each object in the scene to label each object and recognize the activity of each object separately. The system shifts bounding box from frame to frame by the same distance the COG shifted so that bounding box center at the predicted location. The system then determines the position of each point in the ocean according to the COG. In the third stage, The system calculates the distances between COG and contour points of the object in each frame $D = \sqrt{[(x_i - x_{COG})]^2 + [(y_i - y_{COG})]^2}$ and normalize the distances by dividing the distance on the maximum distance between COG and contour points to cancel the effect of the scale variation from one object to another, and then draw the histogram of the normalized distances of each tracked object in each frame between [0,1] and divide it into equal intervals of step 0.1. We calculated the ratio of measured distances in each interval and total density as a sequential characteristic with other features vectors derived from moments of shape. Moments like, Mean, Skew, Kurtosis, etc. Provide useful descriptors for the shape. The system Combined moments to get invariant moments and those fixed on the gyro, translation, and scale. The last stage of the algorithm is the classification and action recognition. The recognition process is divided into two phases: the training phase and the testing phase. Figure 1 shows an overview of the proposed approach, including complete stages in the training and testing phase. We used two methods of classification the first is the multi-class Support Vector Machine (SVM) and the second by using Naive Bayes to make two classifications the first classification to differentiate between humans and other objects and the second classification to recognize the human activities. Figures 2, 3 and 4 illustrate four detected suspicious behaviors like punching, pushing, shooting a gun and falling floor.

A. *Features Extraction*

1- HND Feature

After the system had extracted the object contour in each frame, it determined the Center of gravity of object contour (COG) and the distances between COG and contour points $D = \sqrt{[(x_i - x_{COG})]^2 + [(y_i - y_{COG})]^2}$ then the system normalized the distances by dividing them on D_{max}. The last stage of the feature extraction is forming the feature vector, The system formed the feature vector by drawing the histogram of normalized distances of each tracked object in each frame between [0,1] and divided it into equal intervals step 0.1 to form the first feature vector.

2- Shape moments Features

moments describe the shape accurately, The central moments of order $(r + s)$ of a shape $f(x, y)$ is defined by

$$\mu_{rs} = \int\int (x - \bar{x})^r (y - \bar{y})^s f(x, y) dx dy. \tag{1}$$

where ($\bar{x}$,$\bar{y}$) is the shape centroid. Thus the normalized central moments are given by

$$\eta_{rs} = \frac{\mu_{rs}}{\mu_{00}^{\gamma}}, \gamma = \frac{r + s}{2} + 1 \tag{2}$$

The general shape represented by the first few terms and the finer detail represented in the later term. If we have enough moments we can reconstruct the shape. Shape moments can be used to represent characteristics of the global and invariant shape of image features. There are seven moments [8] like Central moments, Skew Variance and Kurtosis which are important descriptors of the shape and invariant under translation, rotation and resizing since if we translated the object we keep the variance and higher order moments but we only change the mean. Also, if we rotate the quantities like eigenvalues of the covariance matrix are invariant under rotation, but the relative variances and higher order moments changed. The eigenvalues of the covariance matrix stay without any changes when resizing the object by a factor of s the same as scaling the x coordinate and y coordinate by s. So we can produce invariant moments by combining moments. Thus the resultant moments are invariant to translation, rotation and scale.

$$h_1 = \eta_{20} + \eta_{02} \tag{3}$$

$$h_2 = (\eta_{20} - \eta_{02})^2 + (2\eta_{11})^2 \tag{4}$$

$$h_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{03} - \eta_{21})^2 \tag{5}$$

$$h_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{03} + \eta_{21})^2 \tag{6}$$

$$\begin{aligned} h_5 = {} & (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{03} + \eta_{21})^2] + \\ & (3\eta_{21} - \eta_{03})(\eta_{21} - \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{03} + \eta_{21})^2] \end{aligned} \tag{7}$$

$$\begin{aligned} h_6 = {} & (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{03} + \eta_{21})^2] + \\ & 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{03} + \eta_{21}) \end{aligned} \tag{8}$$

$$\begin{aligned} h_7 = {} & (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{03} + \eta_{21})^2] + \\ & (\eta_{30} - 3\eta_{12})(\eta_{03} + \eta_{21})[3(\eta_{30} + \eta_{12})^2 - (\eta_{03} + \eta_{21})^2] \end{aligned} \tag{9}$$

B. *Classification Methods*

There are many methods of classifications like k-nearest neighbors (KNN), Neural networks (NN), Naive Bayes classifier, Two classes support vector machine and Multi-class Support Vector Machine (SVM). We selected two of them for classification, the first was the Naive Bayes classifier and the second was the Multi-class Support Vector Machine Where they achieved the highest recognition rate.

1-Multi-Class Support Vector Machine

There are many types of Multi-class Support Vector Machines, The most popular of them are Hierarchical multi-class SVM [9] which we implemented in our research and tree structured multi-class SVM [10].

Hierarchical multi-class SVM

Different ways to build binary trees divide the data set into two subgroups of root to sheet so that each subgroup consists of only one category. There are some definitions:

Definition I: The center of class i in the feature space is given

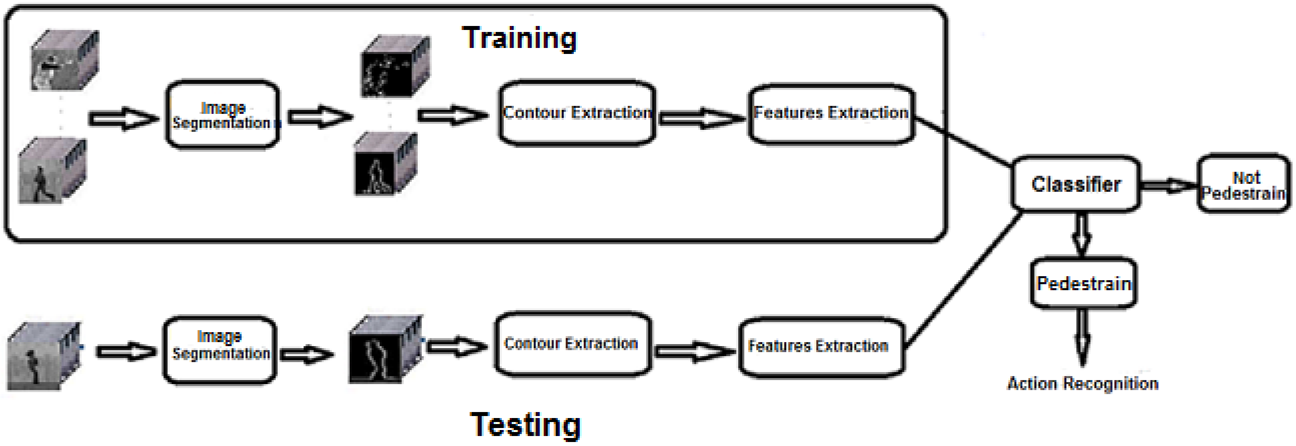

Fig. 1: An overview of our approach (i) Tracking the moving objects and converting the input videos into a number of images (ii) Image segmentation (iii) Contour extraction (iv) Features extraction (v) Classification.

Fig. 2: A sample of suspicious activities (punching and pushing) for tracked persons.

Fig. 3: A sample of suspicious activities (shooting a gun and punching).

Fig. 4: A sample of suspicious activities (In the right image a girl falling on the ground).

by:

$$m_i = \frac{1}{L_i}\sum_{s=1}^{L_i}\Phi(x_s) \tag{10}$$

Definition 2: The distance between class i and class j in the feature space is given by:

$$D = \parallel m_i - m_j \parallel \tag{11}$$

Where

$$\parallel m_i - m_j \parallel = \sqrt{(\frac{1}{L_i}\sum_{s=1}^{L_i} x_s - \frac{1}{L_j}\sum_{t=1}^{L_j} x_t)^2} \tag{12}$$

The hierarchy is determined according to the class distribution. The larger classification area is made in the upper layers. In the dual-row tree, categories are included according to the center of the class. The class similarity is the separation distance and class distribution. Build the binary tree from the bottom up, the binary inverted tree. Classes with Less similarity is the upper nodes and the most similar ones are the least nodes. figure 7 describes the construction order of the binary tree for classifying the twelve classes Running, Walking, Skipping, Kicking, Jumping, Pjumping, Waving, Punching, Shooting, Pushing, Bending, Jacking. The detected suspicious behaviors surrounded by red rectangles.

Tree structured multi-class SVM

The Decision Tree-based Support Vector Machine is the effective way of solving multi-class problems in intrusion detection systems and it can decrease the training and testing time of the intrusion detection system. The Tree structured multi-class SVM is a series of two- class SVMs. To determine the structure of the tree we have to compute the distance between the class patterns and the number of each class patterns, Let n denote the number of i^{th} class patterns x_i, Where i =

1,2,3,.....k. The Center point of i^{th} class patterns is calculated using the equation

$$C_i = \frac{\sum_{m=1}^{n_i} x_m^i}{n_i} \tag{13}$$

The Euclidian distance between i^{th} class and j^{th} class pattern is

$$Ed_{ij} = \parallel C_i - C_j \parallel \tag{14}$$

the distribution of two class patterns is given by a distance

$$d_{ij} = \frac{Ed_{ij}}{\gamma_i + \gamma_j} \tag{15}$$

Where

$$\gamma_i = \frac{\sum_{m=1}^{n_i} \parallel x_m^i - C_i \parallel}{n_i} \tag{16}$$

By calculating the distances of all pairwise classes we can divide the classes to two subsets, each subset has the classes with shorter separate distances and gradually we separate the classes of each subset depending on the separation distances between classes where the class that has a larger separation distance separated first from the subset.

2- Naive Bayes Classifier

Naive Bayes classifier [11] is the suitable method when input dimensions are high, it's simple method and can outperform more sophisticated methods. We can specify it as a conditional probability model and that means a measure of the probability of an event given another event has occurred. Naive Bayes algorithms have many applications such as real time prediction, Multi-class prediction, Spam filtering, Text classification and Sentiment Analysis. So we can use it for making predictions in real time, also we can predict the probability of multiple classes of target variable and by using Naive Bayes classifier with collaborative filtering we can build a recommendation system to filter unseen information.

Shooting a gun

Punching

Fig. 5. Example frames of HMDB database activities.

IV. Experiments

This section presents our evaluation of 250 videos from HMDB dataset [12] 5 distinct suspicious human activities (e.g., Running, Kicking, Punching, Falling floor, and Shooting a gun) By 250 different persons. Fig 5 shows example frames of HMDB database activities, shooting and punching. In activity videos, the person moves in front of a fairly uniform, static background. HMDB that collected from various sources, the small proportion of public databases such as the Prelinger archive and mostly from movies, YouTube and Google videos. The dataset contains 51 action categories each containing a minimum of 101 clips.

A. Segmentation Results

Fig 6 illustrates the segmented images by using multi-threshold segmentation method, original images on the left.

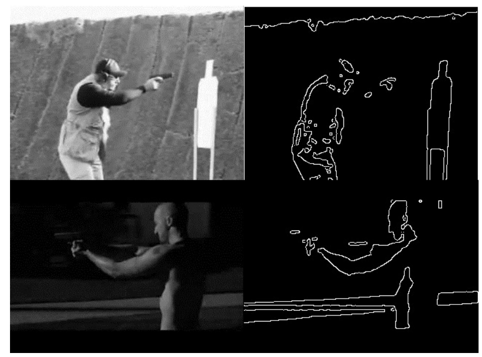

Fig. 6. Segmentation results of shooting gun activity images.

V. Results

Two different groups of experiments are presented. The first used the multi-class SVM and the Second used Naive Bayes classifier. In the two groups, two experiments were conducted one employed HND feature only and the second used a feature vector of both (HND and Shape moments). Table I presents recognition results of HND feature using multi-class SVM classifier, we can see that the proposed algorithm achieves 100% recognition rate in running activity and achieves total recognition results about 88.8% on HMDB data set, Table II presents the confusion matrix results. The experiments repeated in table III and table IV by using the Naive Bayes classifier, but achieved a higher recognition rate than multi-class SVM classifier about 90.4%. When the algorithm combined HND feature and Shape moments features, The recognition rate increased to 95.6% when using Multi-classes SVM and 97.2% when using Naive Bayes classifier, The recognition results and the confusion matrices are shown in tables V, VI, VII and VIII. Fig 8,9 show the accuracy of classification for experimental evaluation of learned features.

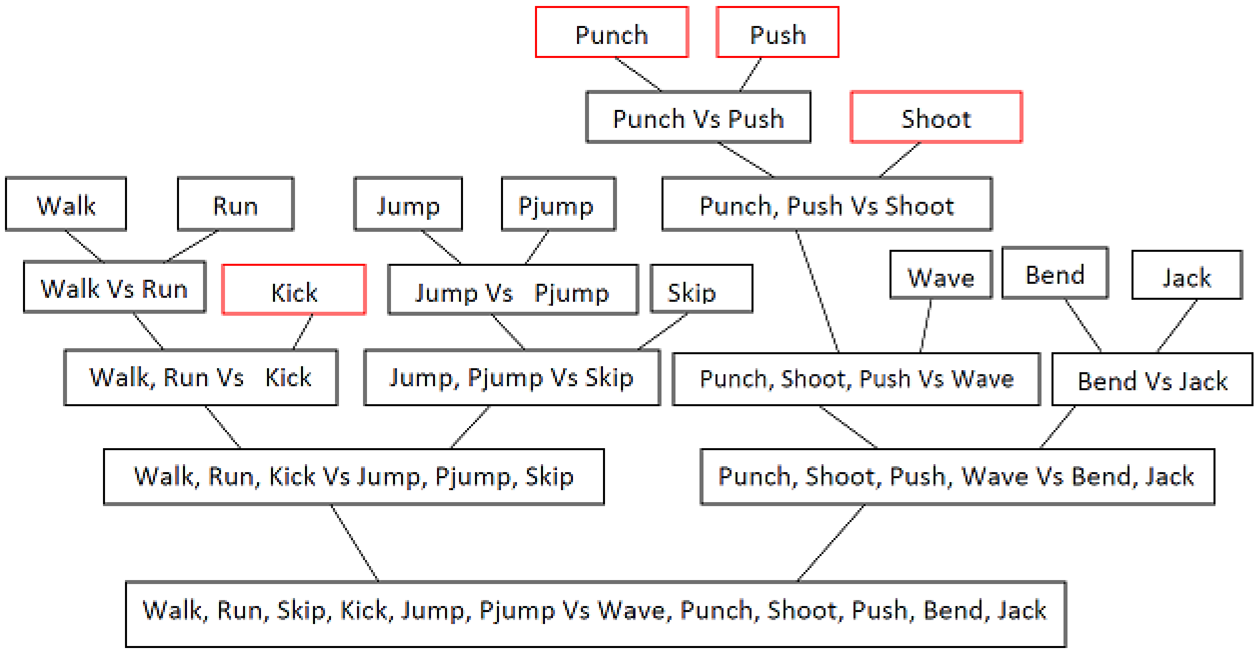

Fig. 7. Inverted binary tree hierarchical multiclass SVM.

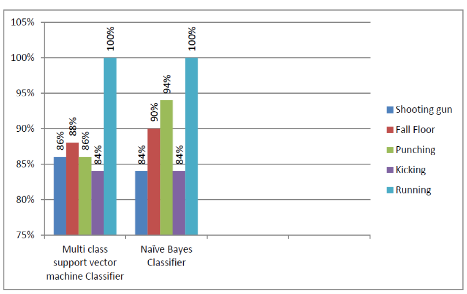

Fig. 8. Classification accuracy for experimental evaluation of HND feature.

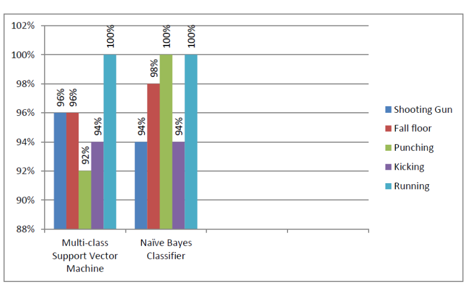

Fig. 9. Classification accuracy for experimental evaluation of combined shape moments and HND features.

TABLE I: Recognition results of HND feature using Multi-class SVM classifier

Suspicious Human Actions	Recognition Results			
	Videos	Correct	Wrong	Correct Rate
Shooting Gun	50	43	7	86%
Fall Floor	50	44	6	88%
Punching	50	43	7	86%
Kicking	50	42	8	84%
Running	50	50	0	100%
Total result	250	222	28	88.8%

TABLE II: Confusion matrix results of HND feature using multiclass SVM classifier

Suspicious Human Actions	Shooting Gun	Fall Floor	Punching	Kicking	Running
Shooting Gun	43		7		
Fall Floor		44	2	4	
Punching	3	4	43		
Kicking		5	3	42	
Running					50

TABLE III: Recognition results of HND feature using Naive Bayes classifier

Suspicious Human Actions	Recognition Results			
	Videos	Correct	Wrong	Correct Rate
Shooting Gun	50	42	8	84%
Fall Floor	50	45	5	90%
Punching	50	47	3	94%
Kicking	50	42	8	84%
Running	50	50	0	100%
Total result	250	226	24	90.4%

TABLE IV: Confusion matrix results of HND feature using Naive Bayes classifier

Suspicious Human Actions	Shooting Gun	Fall Floor	Punching	Kicking	Running
Shooting Gun	42		8		
Fall Floor		45	5		
Punching	1	2	47		
Kicking	3	5		42	
Running					50

TABLE V: Recognition results of combined Shape moments and HND features using multi-class SVM classifier

Suspicious Human Actions	Recognition Results			
	Videos	Correct	Wrong	Correct Rate
Shooting Gun	50	48	2	96%
Fall Floor	50	48	2	96%
Punching	50	46	4	92%
Kicking	50	47	3	94%
Running	50	50	0	100%
Total result	250	239	11	95.6%

TABLE VI: Confusion matrix results of combined Shape moments and HND features using multi-class SVM classifier

Suspicious Human Actions	Shooting Gun	Fall Floor	Punching	Kicking	Running
Shooting Gun	48		2		
Fall Floor		48		2	
Punching	2	2	46		
Kicking		3		47	
Running					50

TABLE VII: Recognition results of combined shape moments and HND features using Naive Bayes classifier

Suspicious Human Actions	Recognition Results			
	Videos	Correct	Wrong	Correct Rate
Shooting Gun	50	47	3	94%
Fall Floor	50	49	1	98%
Punching	50	50	0	100%
Kicking	50	47	3	94%
Running	50	50	0	100%
Total result	250	243	7	97.2%

TABLE VIII: Confusion matrix results of combined shape moments and HND features using Naive Bayes classifier

Suspicious Human Actions	Shooting Gun	Fall Floor	Punching	Kicking	Running
Shooting Gun	47		3		
Fall Floor		49		1	
Punching			50		
Kicking		3		47	
Running					50

VI. Conclusion and Future Work

Our research presented suspicious human action recognition by using two types of features. The first features relate to the shape and is called shape moments. The second relates to the boundary coordinates and is called Histogram of Normalized Distances (HND). Combining these features creates a powerful, distinctive vector that includes powerful details to differentiate human activities. The multi-class Support Vector Machine achieved a recognition rate of 95.6%. Furthermore, the Naive Bayes classifier achieved a recognition rate of 97.2%. From the results, we found that the combination of HND feature with the shape moments features achieves better results. So these types of features are extremely effective in terms of accuracy, especially when they are combined. These combined features can be applied to various classifiers successfully.

References

[1] Weiyao Lin, Ming-Ting Sun, Radha Poovandran, and Zhengyou Zhang , *“Human Activity Recognition for Video Surveillance”*, IEEE International Symposium on Circuit System (ISCAS), Pages: 2737-2740, 2008.

[2] Teddy Co, *“A Survey on Behavior Analysis in Video Surveillance for Homeland Security Applications”*, IEEE Conference on Applied Imagery Pattern Recognition Workshop (AIPR), Pages: 1-8, 2008.

[3] Neil Robertson, Ian Reid, *“A General Method for Human Activity Recognition in Video”*, Journal of Computer Vision and Image Understanding, Vol. 104, No. 2, Pages: 232-248, 2006.

[4] Gowsikhaa D., Manjunath, Abirami S., *“Suspicious Human Activity Detection from Surveillance Videos”*, (IJIDCS) International Journal on Internet and Distributed Computing Systems”. Vol: 2 No: 2, pages: 141-148, 2012.

[5] R.E Kalman,*A new approach to linear filtering and prediction problems.* Transactions of the ASME,SER.D.,Journal of Basic Engineering 82, Pages 34-45, 1960.

[6] Watada JunzoMusa Zalili,Jain Lakhmi C and Fulcher John, *Human Tracking A State of art Survey.* KES 2010, Part2, LNAI 6277, Pages 454–463, 2010.

[7] https://en.wikipedia.org/wiki/Canny_edge_detector

[8] Hu,M,.*Visual pattern recognition by moment invariants.* IRE Tr.on.Inf.Theory, pages 179-187, 1962.

[9] Lili Cheng, Jianpei Zhang, Jing Yang, Jun Ma.*An Improved Hierarchical Multi-Class Support Vector Machine with Binary Tree Architecture* 978-0-7695-3112- 0/08 2008 IEEE DOl 10.1109I1CICSE.2008

[10] Jun GUO Computer CenterEast, *An Efficient Algorithm for Multiclass Support Vector Machines* 978-0-7695-3489-3/08 2008 fEEE DOl 10.1109/ ICACfE. 2008.48

[11] Naive Bayes text classification @http://nlp.stanford.edu/IR-book/html/htmledition/naive-bayes-text-classification-1.html

[12] HMDB Data set @http://serre-lab.clps.brown.edu/resource/hmdb-a-large-human-motion-database