Journal Of Al Azhar University Engineering Sector

Vol. 13, No. 46, January, 2018, 112-121

HUMAN ACTIVITIES RECOGNITION USING SHAPE MOMENTS AND HISTOGRAM OF NORMALIZED DISTANCES "HND"

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

[1]Electromechanical Institute, Housing and Building National Research Center

[2]Computer and Systems Engineering Department, Faculty of Engineering, Ain Shams University, Cairo 11517

ABSTRACT

this paper presents an algorithm for human activities recognition in videos based on a combination of two different feature types The first feature type concerns the shape and is called the Shape moments. The second feature type concerns the contour boundary coordinates and the feature is called Histogram of Normalized Distances from Center of gravity of the object Shape "COG" and it's Contour points "HND". Combining these features leads to the formation of a strong complementary feature vector that captures effective discriminate details of human action videos. We use two classifiers; the first is Multi-class Support Vector Machine and the second is Naïve Bayes classifier. The Recognition rate by using Multi-class SVM classifier is up to 95.6 % but by using Naive Bayes classifier is 97.2%.

Key Words: Suspicious human activities, Recognition, Contour Points.

I. Introduction

Human activity recognition is one of the growing areas of artificial intelligence and computer vision. The need for automated surveillance systems has 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 occurrence such as automatic reporting of a person with a bag loitering at 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. The rest of this paper is organized as follows. Section II shows the related work. Section III introduces the proposed System. Section IV Shows the Experiments. The Results in Section V and The Conclusion in Section IV.

II. Related Work

Suspicious activity recognition is very active research area in computer vision. Many papers were written for human activity recognition. In paper [1] the authors suggest a system use face profile, gait silhouette and using confident-frame based recognizing algorithm (CFR) to recognize human activities. The frame with high confidence are used to recognize the activity. In 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 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 Recognition System

The whole recognition process can be divided into stages: the first stage is to detect moving objects and trace them from frame to another frame over time by using blob analysis to detect groups of connected pixels which correspond to moving object, The challenge when the system tracks moving objects is an occlusion. The location of the objects is not available when they are occluded by others or things. So the proposed system used Kalman filter [5], [6] to overcome these missing measurements then we use multi threshold segmentation method to segment moving object from background since it's the effective tool. The second stage is to extract object contour by using the canny algorithm for edge detection [7] then the system determines the center of Gravity of segmented object 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 Shift Bounding Box from frame to frame by the same distance COG shifted so that Bounding Box center at the predicted location. Then the system determines the position of each point in the contour according to COG to cancel the effect of starting point variation. 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 distance on the maximum distance between COG and contour points ,Then the system draw the histogram of normalized distances between [0,1] and divided it into equal intervals step 0.1. We took the ratio of normalized distances in each interval and total density as a feature vector concatenated with other features vectors derived from shape moments [8]. The last stage of the algorithm is the classification and action recognition, we tried two methods of classification the first is multiclass Support vector machine (SVM) classifier and the second by using Naive Bayes classifier to classify the activity of each object in the video. Figure 1 shows tracked persons which the system detected two suspicious behaviors Kick and push.

Fig 1 Sample of suspicious Activities for tracked persons

A. Segmentation

Histogram thresholding and slicing techniques are used to segment the image. They may be applied directly to an image, but can also be combined with pre- and post-processing techniques. Thresholding is probably the most frequently used technique to segment an image. The thresholding operation is a grey value remapping operation defined by

$$g(v) = \begin{cases} 0 \; if v \; < t \\ 1 \; if \; v \; \geq t \end{cases} \tag{1}$$

Where v represents a grey value, and t is the threshold value.Thresholding maps a grey-valued image to a binary image. After the thresholding operation, the image has been segmented into two segments, identified by the pixel values 0 and 1 respectively. Thresholding is thus a simple but effective tool to isolate objects of interest; so in our work we use thresholding as a segmentation method. Thresholding selection techniques can be classified into two categories: bi-level and multi-level.In the former, one limit value is chosen to segment an image into two classes: one representing the object and the other one segmenting the background. When distinct objects are depicted within a given scene, multiple threshold values have to be selected for proper segmentation, which is commonly called multilevel thresholding. The way to extract the objects from the background is to select a threshold t that separates these modes, any point (x,y) for which g(x,y) > t is called an object point otherwise the point is called a background point .We use otsu’s method to determine the threshold, in this method we exhaustively search for the threshold that minimizes the intra class variance [9] (the variance within the class) .Intra class variance σ_{ω}^{2}(t) defined as a weighted sum of variances of the two classes.

$$\sigma_{\omega}^{2}(t) = \omega_0\sigma_0^2(t) + \omega_1\sigma_1^2(t) \quad (2)$$

Weights ω 0 , 1 {\displaystyle \omega _{0,1}} $\omega_0, \omega_1$are the probabilities of the two classes separated by a threshold t {\displaystyle t} t and σ 0 , 1 2 {\displaystyle \sigma _{0,1}^{2}} $\sigma_0, \sigma_1$are variances of these two classes .The class probability $\omega_0, \omega_1$are computed from the L histograms since

$$\omega_0 = \sum_{i=0}^{t-1} p(i), \omega_1 = \sum_{i=t}^{L-1} p(i) \quad (3)$$

B. Features Extraction

The System has many concatenated Features

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 between [0,1] and divided it into equal intervals step 0.1 to form the first feature vector.

2- Shape moments Features [8]

Shape moments can be used to represent global and invariant shape characteristics of image features. Central moments, Variance, Skew and Kurtosis are useful descriptors of shape. The central moments of order (p+q) of a shape f(x; y)is defined by

$$\mu_{pq} = \iint (x - \bar{x})^p (y - \bar{y})^q f(x, y)\, dx dy \quad (4)$$

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

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}}, \gamma = \frac{p+q}{2} + 1 \quad (5)$$

Moments have been shown to be a very useful set of descriptors for matching , The first few terms give the general shape and the later term fill in finer detail, We can completely reconstruct the shape if we have enough moments. There are seven moments which are invariant under translation since if we translated the object, we only change the mean not the variance or higher-order moments .so none of the central moments affected by translation, Also if we rotate the shape we can change the relative variances and higher-order moments but certain quantities such as Eigen values of the covariance matrix are invariant to rotation. Resizing the object by a factor of s is the same as scaling the x and y coordinate by s, the of Eigen values of the covariance matrix stay the same. By combining moments we can thus produce invariant moments, ones that are invariant to rotation, translation and scale. There are seven moments which are invariant under translation , Rotation and scale variation calculated as follows

$$h_1 = \eta_{20} + \eta_{02} \tag{6}$$

$$h_2 = (\eta_{20} - \eta_{02})^2 + (2\eta_{11})^2 \tag{7}$$

$$h_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{03} - \eta_{21})^2 \tag{8}$$

$$h_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{03} + \eta_{21})^2 \tag{9}$$

$$h_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{03} + \eta_{21})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} - \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{03} + \eta_{21})^2] \tag{10}$$

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

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

C. Classification Methods

We used two methods, the first was Multi-Class Support vector machine (SVM) as a multiple class classifier by constructing a net consisting of two- class and the second classification method was Naive Bayes classifier

1- Multi- Class Support Vector Machine

Multi-class SVMs are usually implemented by combining several two classes SVMs. There are many trails to implement multi-class Support vector machine, Many of the most prominent attempts are Hierarchical multiclass SVM [10] which we implemented in our research and Tree structured multiclass SVM [11], These algorithms consider the distribution of classes along with the distances of class centers.

i- Hierarchical multi-class SVM

The different ways to construct the binary trees divides the data set into two subsets from root to the leaf until every subset consists of only one class. There are some definitions:

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

$$m_i = \frac{1}{L_i} \sum_{s=1}^{L_i} \phi(x_s) \quad (13)$$

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

$$D = \|m_i - m_j\| \quad (14)$$

Where

$$\|m_i - m_j\| = \sqrt{\left(\frac{1}{L_i} \sum_{s=1}^{L_i} x_s - \frac{1}{L_j} \sum_{t=1}^{L_j} x_t\right)^2} \quad (15)$$

Definition 3: The radius of hyper-sphere in feature space is

$$R_i = max_{t=1,\dots,l_j} \|x_t - m_i\| \quad (16)$$

Hierarchical order is determined according to the class distribution. The bigger classification area is made the upper classes. In the hierarchical binary tree, the inclusion of classes is according to the distance of the class center. Class similarity considers the class distance and the class distribution both. The construction of the binary tree is from bottom to the top, the inverted binary tree. The classes with less similarity are the upper nodes and the classes with larger similarity are lower nodes. Figure 2 describes the construction order of binary tree for classifying the five classes Walking, Running, Waving, Punching, Shooting.

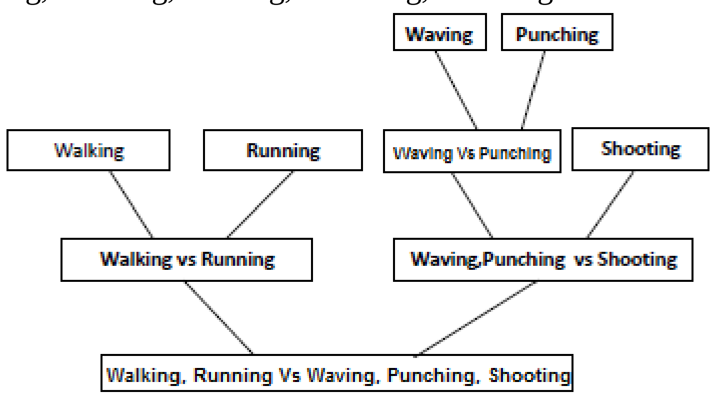

Fig 2: Inverted binary tree hierarchical multiciass SVM

ii- Tree structured multiclass SVM

It is a series of two-class SVMs. The distance between two class patterns, and the number of each class patterns are used to determine the structure of the tree.

2. NAIVE BAYES CLASSIFIER

Naive Bayes [12] is a conditional probability model. The Conditional probability means a measure of the probability of an event given that (by assumption, presumption, assertion or evidence) another event has occurred. If we want to solve classification problem and we have a number of features n represented by a vector $= (x_1; x_2; x_3; x_n)$, it assigns to this instance probabilities $p(c_k | x_1, \dots \dots, x_n)$ for each of k possible outcomes or classes. in the case of large features or if a feature can take on a large number of values, then basing such a model on probability tables is infeasible. We therefore reformulate the model to make it more tractable.

$$P(C_K|x) = \frac{P(C_K)P(X|C_K)}{P(x)} \quad (17)$$

Fig 3 Example frames of HMDB Database Activities

IV- Experiments

This Section presents our evaluation on 250 videos from HMDB data set [13] 5 distinct Suspicious Human activities (e.g., Running, Kicking, Punching, fall floor and shooting gun) By 250 different persons. In activity videos, the person moves in front of a fairly uniform, static background.

A. Recognition System Segmentation Results of some activities

Fig 4 shows original images on the left and the segmented images by using thresholding segmentation method on the right

Fig 4 Segmentation by thresholding. On the left, an original image

B. Recognition System feature extraction of some activities

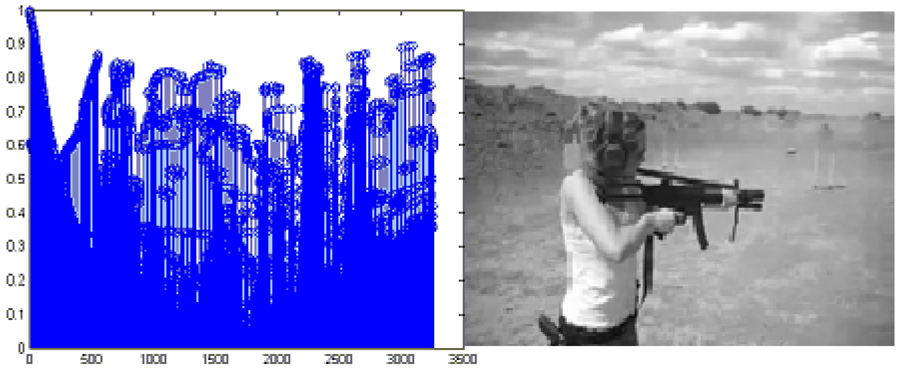

Fig 5 An Example of Histogram of Normalized distances from Center of gravity of the object and it's Contour points for Shoot gun activity

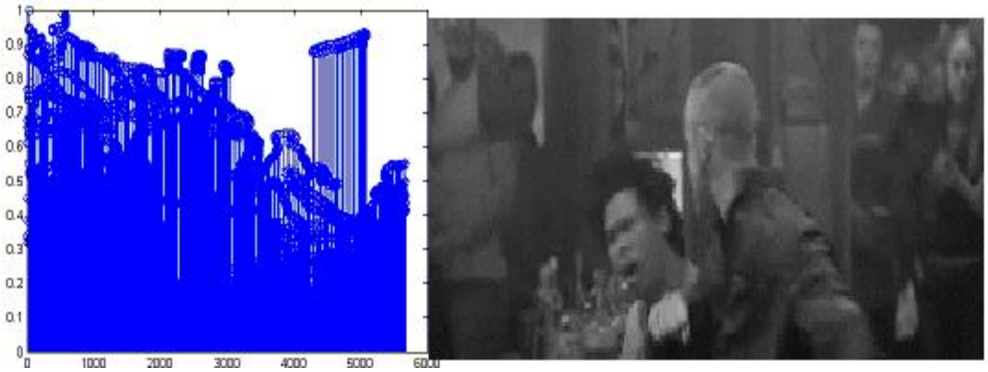

Fig 6 An Example of Histogram of Normalized distances from Center of gravity of the object and it's Contour points for punch activity

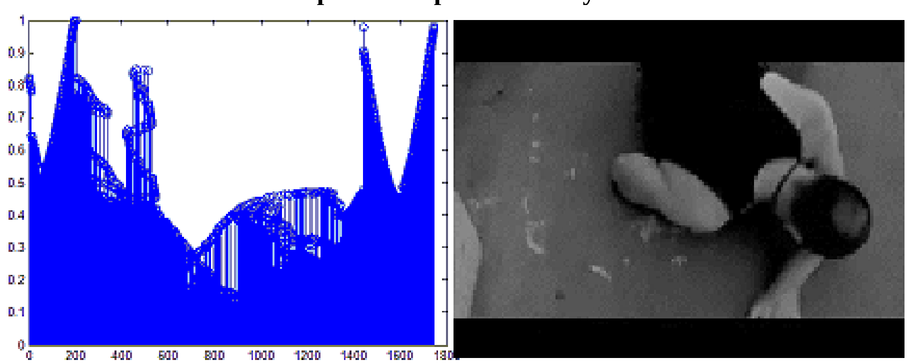

Fig 7An Example of Histogram of Normalized distances from Center of gravity of the object and it's Contour points for Fall Floor activity

V- Recognition System Results

Two different groups of experiments are presented. The first used the multiclass SVM and the Second used Naïve Bayes classifier. In the two groups, two experiments were conducted one used a feature vector of both (HND and Shape moments). The second employed HND feature only.

TABLE 1. Recognition System results of combined Shape moments and HND features using multiclass SVM classifier

Suspicious Human Actions	Recognition Results			
	Videos	Corrects	Wrongs	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 2. Confusion matrix results of combined Shape moments and HND features using multiclass 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 3. Recognition System results of combined Shape moments and HND features using Naïve Bayes classifier

Suspicious Human Actions	Recognition Results			
	Videos	Corrects	Wrongs	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 4. Confusion matrix results of combined Shape moments and HND features using Naïve 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

TABLE 5. Recognition System results of HND feature using multiclass SVM classifier

Suspicious Human Actions	Recognition Results			
	Videos	Corrects	Wrongs	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 6. Confusion matrix results of HND feature using multiclass SVM lassifier

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 7. Recognition System results of HND feature using Naïve Bayes classifier

Suspicious Human Actions	Recognition Results			
	Videos	Corrects	Wrongs	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 8. Confusion matrix results of HND feature using Naïve 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

IV- Conclusion

The research reported in this paper demonstrates optimal Suspicious human action recognition in terms of recognition rate accuracy by combining two different kinds of features. The first feature concerns the shape and is called Shape moments. The second concerns the boundary coordinates and is called HND. Combining these features leads to the formation of a strong complementary feature vector that captures effective discriminate details of human action videos. The SVM experimental results achieved a correct recognition rate of 95.6%. This result demonstrates that our algorithm promises excellent results in terms of accuracy for suspicious human action recognition. Moreover, the Naïve Bayes experimental result achieved was a correct recognition rate of 97.2% of the correct recognition rate. This result is very close to the optimal solution and indicates that these combined features can be applied in different classifiers successfully. In addition, the algorithm used in this research applied a new HND feature which is very useful for human recognition; it is also low time computation and complexity. It is proven based on the results, in this research, that these features types are very effective in terms of accuracy, especially when they are combined (e.g., the combination of HND and Shape moments features).

REFERENCES

[1] Weiyao Lin, Ming-Ting Sun, Radha Poovandran, and Zhengyou Zhang , Human Activity Recognition for Video Surveillance, IEEE International Symposium onircuit 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, pp.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 82,34– 45 (1960).

[6] Watada JunzoMusa Zalili,Jain Lakhmi C and Fulcher John,. Human Tracking A State of art Survey. KES 2010,Part2, LNAI 6277,pp. 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,179–187,1962.

[9] https://en.wikipedia.org/wiki/Otsu%27s method

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

[11] 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.

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

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