JAUES

Journal Of Al Azhar University Engineering Sector

Vol. 12, No. 44, July, 2017, 933-945

HUMAN ACTIVITY RECOGNITION SYSTEM (HARS) USING HISTOGRAM OF NORMALIZED FOURIER DESCRIPTOR

Hanan Samir[1], Hossam Abd Elmunim[2], and Gamal Aly[2]
[1]Housing and Building National Research Center
[2]Computer and Systems Engineering Dept., Faculty of Engineering, Ain Shams University

ABSTRACT
This paper presents a system, which is able to recognize different continuous human activities in real time from videos using a single stationary camera. The proposed system has a motion descriptor, Called histogram of normalized Fourier descriptor of object contour. After image segmentation the proposed system extract the object contour and estimate the FD ”Fourier descriptor” of that contour then normalize FD to cancel the effect of starting point variation ,Rotation and Scale. The histogram of the normalized Fourier descriptor is used as a feature vector.The authors used three methods for classification Support Vector Machine,One versus all support vector machine and Naive Bayes classifier. The classification by using SVM shows better performance than other methods. Experimental results on two data sets weizman and KTH validate the proposed system reliability and efficiency.

Keywords: HumanActivity,Recognition,Fourier Descriptors, Object contour

1. INTRODUCTION

Human activity recognition from video sequences or images is a difficult task, due to problems such as scale variations, background clutter, Occlusion and lighting variations and such activities are composed of different elementary actions so in order to analyze them, human elementary actions must be recognized at first. But elementary actions within the same class may be expressed by different persons with different body movements and actions between different classes may be difficult to be distinguished. Another difficulty is the way humans perform an activity, which depends on their physical mental or emotional state, and this makes the identification of the activity difficult to be performed [1]. In the past few years, automatic activity recognition has drawn a lot of attention in the field of video analysis technology due to its important role in many applications such as surveillance and healthcare systems. In a surveillance environment, the automatic detection of suspicious activities can be used to prevent crime or dangerous behaviors, such as automatic reporting of a person with a bag loitering at airport or station .So our paper contributes in recognizing human elementary action which has main role in recognizing suspicious behavior.

2. RELATED WORK

There are many categories for human activity recognition such as

2.1 **Space-time methods**

Involve activity recognition methods.These approaches consider the activities in three dimensions “Space-time volume” .The system construct three dimension model for each training video and when unlabeled video is provided, the system construct the three dimension model for this unlabeled video and compare it with each known activity model to measure the similarity and recognize the activity (Shabani, A. H., Clausi, D., and Zelek, J. S. (2011))[2] or trajectories (Vrigkas, M., Karavasilis, V., Nikou, C., and Kakadiaris, I. A. (2013) [3]).

2.2 **Stochastic methods**

Recognize activities by applying statistical models to represent human actions. There are many stochastic methods for human activity recognition, such as hidden Markov model (HMMs) (Lan, T., Wang, Y., and Mori, G. (2011)[4]; Iosifidis, A., Tefas, A., and Pitas, I. (2012a) [5) and hidden conditional random fields (HCRFs) (Quattoni, A., Wang, S., Morency, L. P.,Collins, M., and Darrell, T.)[6], these methods consider the structure to be recognized as a predictable sequence of states.

2.3 **Rule-based methods**

Rule-based methods use rules or sets of attributes to model an activity to describe an event. These approaches have the ability to recognize actions in complex scenes with multiple subjects (Morariu, V. I., and Davis, L. S. (2011) [7]).

2.4 **Shape-based methods**

Shape-based approaches efficiently represent activities with high-level reasoning by modeling the motion of human body parts. These approaches consider human silhouette as limbs jointly connected to each other and describe human body parts in 2D as rectangles and in 3D space as cylinders to obtain a model for human appearance. The human silhouette has an important role in recognizing human activity. (Sigal, L., Isard, M., Haussecker, H., and Black, M. J. (2012b)[8]; Tran, K. N., Kakadiaris, I. A., and Shah, S. K. (2012) [9]; Rocío Díaz de León, Luis Enrique Sucar(2002)[10]).

2.5 **Affective methods**

Affective methods are a combination of pattern recognition ,Computer vision and artificial intelligence (Picard, R. W. (1997)) [11].They represent human activities according to the affective state of a person (Liu, N., Dellandra, E., Tellez, B., and Chen, L. (2011b) [12]; Martinez, H. P., Yannakakis, G. N., and Hallam, J. (2014) [13]). Affective computing studies model the ability of a person to recognize, and control his/her affective states in terms of hand gestures, facial expressions, speech, and activity recognition (Pantic, M., and Rothkrantz, L. (2003)) [14].

2.6 **Behavioral methods**

Human behavior is affected by many factors according to human psychological state .These factors decomposed into many components including moods, actions, interactions with other people, so to recognize human behavior we need to determine the proper features and this happen when we get information about human psychological state (Candamo, J., Shreve, M., Goldgof, D. B., Sapper,D. B., and Kasturi, R. (2010).[15]).Behavioral approaches aim to recognize behavioral attributes such as facial expressions, gestures, and auditory cues (Song, Y., Morency, L. P., and Davis, R. (2012a) [16]; Vrigkas, M., Nikou, C., and Kakadiaris, I. A. (2014b) [17]).

2.7 **Social networking methods**

Social networks affect to a large extent on the behavior of humans ,Such as Twitter, YouTube and Facebook. Interaction with social networks can be considered as a type of activity. Social networking methods model the human behavior from gestures, body motion and speech (Patron-Perez, A., Marszalek, M., Reid, I., and Zisserman, A. (2012) [18] ; Marn-Jimnez, M. J., Noz Salinas, R. M., Yeguas-Bolivar, E., and de la Blanca, N. P. (2014) [19]).

By estimating the orientation and location of the faces of persons and computing a line of sight for each face we can model social interactions and we can get the location of the individuals Fathi, A., Hodgins, J. K., and Rehg, J. M. (2012) [20] and by modeling relationships between interacting person we can predict joint social interaction Park, H. S., and Shi, J. (2015) [21].

3. PROPOSED SYSTEM

The whole recognition process can be divided into two phases: the training phase and the testing phase. The system implementation consists of the five main parts shown in Fig.1.In this work, we make the assumption of a static background. After transforming the video into number of frames we segment the object in each frame using multi-threshold segmentation method since it is simple and effective tool to isolate object of interest. In the second stage we filter the segmented object and made edge detection using Canny algorithm, By this way we extracted the contour of the object and determined the position of each point on the contour. To cancel the effect of starting point variation on the position of each point in the object contour we determine the position of each point according to the center of the shape.Then we calculate Fourier descriptors for each point on the shape contour with removing scale and Rotation effects and these done by getting normalized Fourier descriptors by dividing absolute value of Fourier descriptor on the maximum absolute value of it (abs (Fourier descriptors))/(max(abs Fourier descriptors)).Then we get The histogram of normalized Fourier descriptor and divided it into equal intervals between [0,1]step 0.1. We took the ratio of normalized Fourier descriptor in each interval and total density as a feature vector. Last part of the implementation is the classification and action recognition (We tried many classification methods like two classes SVM,One vs all SVM and Naive Bayes classifier) .

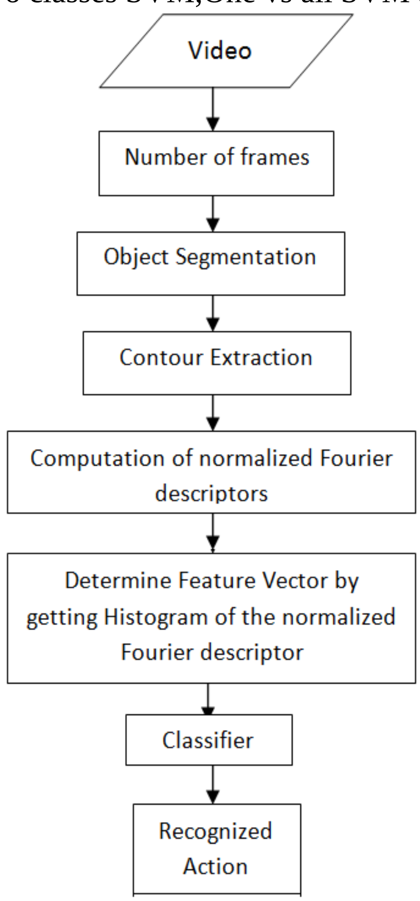


Fig. 1: proposed HARS System stages0

3.1 Image Segmentation

3.1.1 Threshold based segmentation Algorithm

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

$$g(v) = \begin{cases} 0 & if\, v < t \\ 1 & if\, v \geq t \end{cases} \qquad (1)$$

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 [22] (the variance within the class) .Intra class variance $\sigma_1 \omega^{\dagger 2}$ (t) defined as a weighted sum of variances of the two classes.

$$\sigma_\omega^2(t) = \omega_0 \sigma_0^2(t) + \omega_1 \sigma_1^2(t) \qquad (2)$$

Weights ω_0, ω_1 are the probabilities of the two classes separated by a threshold t and σ_0, σ_1 are variances of these two classes .The class probabilityare computed from the L histograms since

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

3.1.2 Canny edge detection algorithm

The Process of Canny edge detection algorithm **[23]** can be broken down to 5 different steps:

1. Apply Gaussian filter to smooth the image in order to remove the noise.
2. Find the intensity gradients of the image.
3. Apply non-maximum suppression to get rid of spurious response to edge detection.
4. Apply double threshold to determine potential edges.
5. Track edge by hysteresis: Finalize the detection of edges by suppressing all the other edges that are weak and not connected to strong edges.

Canny edge detection algorithm is one of the most strictly defined methods that provides good and reliable detection.

3.2 Fourier Descriptors (FD)

FD **[24]** can be used as a representation of 2D closed shapes. A point moving along the boundary generates the complex function

$$Z(m) = x(m) + jy(m) \quad (4)$$

Fourier Descriptor is defined as $Z(k) = \frac{1}{N}\sum_{m=0}^{N-1} Z(m)e^{\frac{-j2\pi mk}{N}}$ (k=0,…….,N-1) (5)

If the shape is scaled by factor S then $\dot{Z}(m) = S\,Z(m)$ (6)

If the shape is rotated by angle θ then $\dot{Z}(m) = Z(m)e^{j\theta}$ (7)

And if starting point is shifted then $\dot{Z}(m) = Z(m - m_0)$ (8)

$\dot{Z}(m) = SRZ(m) + T$ (9) **S:** Scale (R=$e^{\dagger j\theta}$) : rotation **T**: Translation

$$\dot{Z}(k) = \frac{1}{N}\sum_{m=0}^{N-1}[SRZ(m) + T]e^{\frac{-j2\pi mk}{N}} \quad (10)$$

$$\dot{Z}(K) = \frac{1}{N}\sum_{m=0}^{N-1} SRZ(m)e^{\frac{-j2\pi mK}{N}} + \frac{1}{N}\sum_{m=0}^{N-1} Te^{\frac{-j2\pi mK}{N}} \quad (11)$$

N: Number of boundary pixels So After obtaining the segmented object from the image we extracted the object contour and determined coordinates of each point on the contour then determined the center of the shape and the position of each point on the contour according to that center by this way we removed the effect of starting point variation then compute Fourier descriptors at each point on the contour.

3.2.1 Removing scale and rotation effects

We normalized Fourier descriptor by getting absolute value of it and divide this absolute value by maximum absolute value to get normalized Fourier descriptor = $\frac{absZ(K)}{\max(abs(Z(K)))}$ then we formed feature vector as the histogram of (normalized Fourier descriptors).

3.3 Classifier

We used Support vector machine (SVM) as a binary classification algorithm (We used it in the first time as two classes classifier and in the second time as one versus all classifier) and we used also Naïve Bayes classifier

3.3.1 Support Vector Machine based classifier

Support vector machine (SVM) **[25]** is a classifier method and it's goal is to design a hyperplane

$$g(x) = \omega^T x + \omega_0 \quad (12)$$

That classifies all training vectors in two classes,Let $x_i, i = 1,2, \dots N$ be the feature vectors of the training set x.These belong to either of two classes which are assumed to be linearly separable.

$$g(x) \geq 1 \;\forall\; x class\ 1 \ and \ g(x) \leq -1 \;\forall x \in class2 \quad . (13)$$

The best choice of hyperplane that leaves the maximum margin from both classes "The margin is this distance between the hyperplane and the closest elements from this hyperplane".the distance of a point from a hyper plane is given by $z = \frac{|g(x)|}{\|\omega\|}$ we can scale ω and ω_0 so that the value of $g(x)$ at the closest elements will be at least 1 So the total margin is computed by $\frac{1}{\|\omega\|} + \frac{1}{\|\omega\|} = \frac{2}{\|\omega\|}$ and the aim to minimize the $\|\omega\|$ and that will maximize the separability,when we minimize this weight we will have the biggest margin that will split two classes ,Minimizing ω is a nonlinear optimization task solved by the Karush –Kuhn-Tucker (KKT) conditions using lagrange multipliers λ_i ,x_i is the feature vectors of training set $\omega = \sum_{i=0}^{N} \lambda_i y_i x_i$ so when we solve this equation we try to minimize omega that will maximize the margin between the two classes .Support vector machine is a two-class classifier that can be used to as a multiple class classifier by constructing a net consisting of two- class .

3.3.1.1 Activity classification using Two classes SVM

We partition the data into Training and Validation Sets using the Standard Data Partition defaults of 60% of the data randomly allocated to the Training Set and 40% of the data randomly allocated to the Validation Set.

Input = feature vector of one activity ,for Example the classification of Running and Jumping Activities ,Training Feature vectors of Running Activity for different videos and different person are [0.8768 0.0932 0.0147 0.0047 0.0030 0.0029 0.0016 0.0008 0.0017] &[0.8606 0.1093 0.0131 0.0058 0.0038 0.0026 0.0015 0.0010 0.0008 0.0017] &[0.9430 0.0380 0.0071 0.0036 0.0048 0.0012 0 0 0 0.0024]& [0.8527 0.1181 0.0136 0.0059 0.0040 0.0018 0.0012 0.0008 0.0005 0.0016]& [0.8421 0.1292 0.0144 0.0072 0.0036 0.0012 0 0.0012 0.0012] and Training Feature vectors for Jumping activities [0.8920 0.0797 0.0130 0.0053 0.0032 0.0023 0.0018 0.0008 0.0006 0.0015] & [0.8594 0.1099 0.0133 0.0054 0.0042 0.0025 0.0017 0.0007 0.0009 0.0018]& [0.8587 0.1112 0.0141 0.0062 0.0040 0.0020 0.0010 0.0008 0.0005 0.0015]& [0.8444 0.1228 0.0151 0.0065 0.0042 0.0024 0.0013 0.0012 0.0005 0.0015] &[0.8803 0.0900 0.0130 0.0070 0.0039 0.0021 0.0010 0.0007 0.0006 0.0016] The Two classes SVM tests the input feature vector to determine which class it belongs. When Fv [0.8421 0.1292 0.0144 0.0072 0.0036 0.0012 0 0.0012 0.0012] tested the recognized activity was Running activity and when Feature vector [0.8575 0.1109 0.0131 0.0066 0.0043 0.0022 0.0016 0.0014 0.0007 0.0016] tested the recognized activity was Jumping activity ,the Figures 2, 3 Show classification of some activities

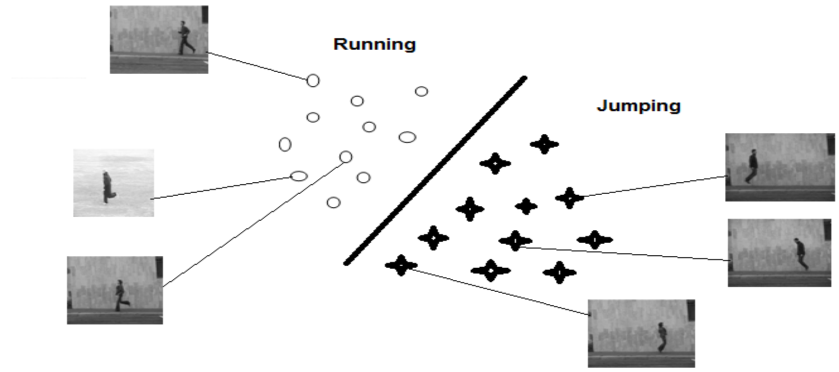


Fig 2 Example of SVM classification on two classes Running and Jumping

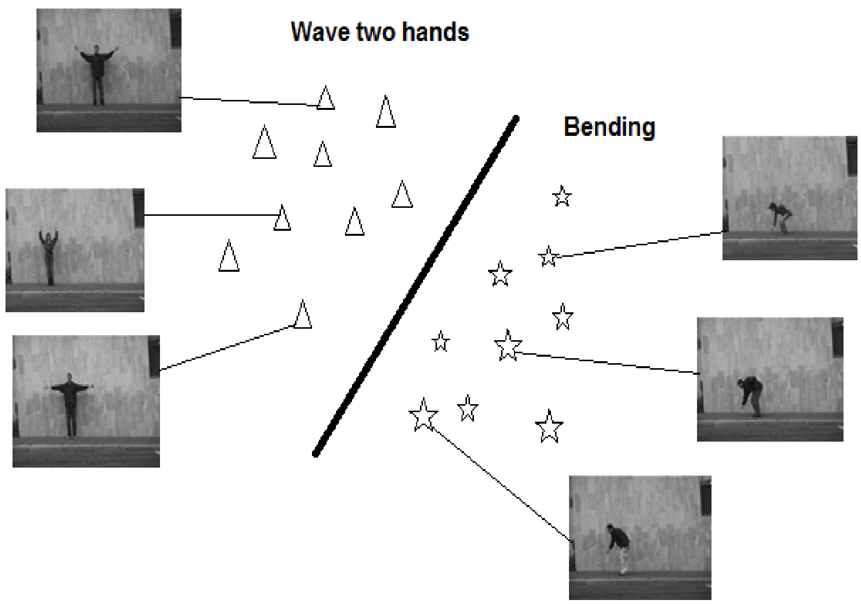


Fig 3 Example of SVM classification on two classes Wave two hands and Bending

3.3.2 Naïve Bayes based Classifier

The Bayesian Classification **[26]** represents a supervised learning method as well as a statistical method for classification. Assumes an underlying probabilistic model and it allows us to capture uncertainty about the model in a principled way by determining probabilities of the outcomes. It can solve diagnostic and predictive problems.This Classification is named after Thomas Bayes (1702-1761), who proposed the Bayes Theorem. Bayesian classification provides practical learning algorithms and prior knowledge and observed data can be combined. Bayesian Classification provides a useful perspective for understanding and evaluating many learning algorithms. It calculates explicit probabilities for hypothesis and it is robust to noise in input data.

Likelihood   Class Prior Probability

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)}$$

Posterior Probability   Predictor Prior Probability

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

- $P(c|x)$ is the posterior probability of *class* (c, *target*) given *predictor* (x, *attributes*).
- $P(c)$ is the prior probability of *class*.
- $P(x|c)$ is the likelihood which is the probability of *predictor* given *class*.
- $P(x)$ is the prior probability of *predictor*.

3.3.2.1 Activity classification using Naïve Bayes classifier

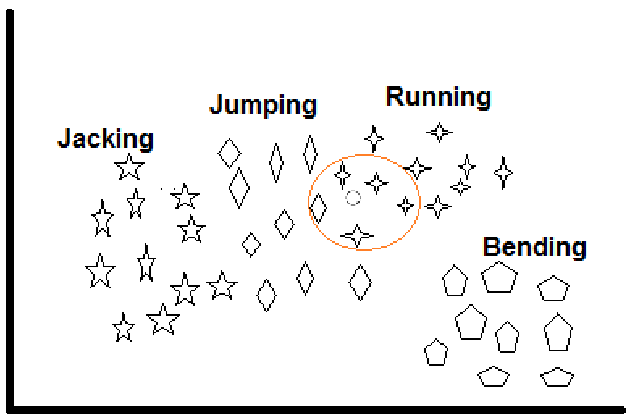


Fig 4 Example of Naive Bayes classification

The object can be classified as Jumping or Running or jacking or bending or etc..
The classifier computes the Prior probabilities for each class for example

$$\text{Prior Probability of Running Activity} = \frac{NumberofRunningObjects}{Totalnumberofobjects}$$

$$\textbf{Prior probability of Jumping Activity} = \frac{\textbf{Number of Jumping objects}}{\textbf{Total number of objects}}$$

$$\textbf{Prior probability of Jacking Activity} = \frac{\textbf{Number of Jacking objects}}{\textbf{Total number of objects}}$$

$$\textbf{Prior probability of Bending Activity} = \frac{\textbf{Number of Bending objects}}{\textbf{Total number of objects}}$$

In this Example 41 objects ,11 are running ,10 jumping , 11 Jacking and 9 Bending

$$\textbf{Prior probability of Running Activity}\,\alpha\, \frac{11}{41}$$

$$\textbf{Prior probability of Jumping Activity}\,\alpha\, \frac{10}{41}$$

$$\textbf{Prior probability of Jacking Activity} \, \alpha \, \frac{11}{41}$$

$$\textbf{Prior probability of Bending Activity} \, \alpha \, \frac{9}{41}$$

we are now ready to classify a new object (WHITE circle). Since the objects are well clustered. To measure likelihood, we draw a circle around X which encompasses a number (to be chosen a priori) of points irrespective of their class labels. Then we calculate the number of points in the circle belonging to each class label. From this we calculate the likelihood:

$$likelihod\ of\ x\ given\ Running\ \alpha\ \frac{number\ of\ Running\ in\ the\ vicinity\ of\ x}{total\ number\ of\ Running\ cases}$$

$$Likelihod\ of\ x\ given\ Running\,\alpha\ \frac{4}{11}$$

$$likelihod\ of\ x\ given\ Jumping\ \alpha\ \frac{number\ of\ Jumping\ in\ the\ vicinity\ of\ x}{total\ number\ of\ Jumping\ cases}$$

$$likelihod\ of\ x\ given\ Jumping\ \alpha\ \frac{1}{10}$$

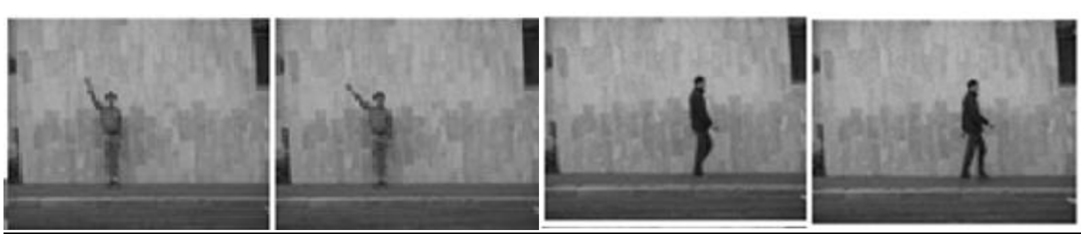
In the Bayesian analysis, the final classification is produced by combining both sources of information, i.e., the prior and the likelihood, to form a posterior probability using the so-called Bayes' rule

Posterior Probability of x being running α prior probability of Running x likelihood of x given Running $= \frac{11}{41} \times \frac{4}{11} = \frac{4}{41}$

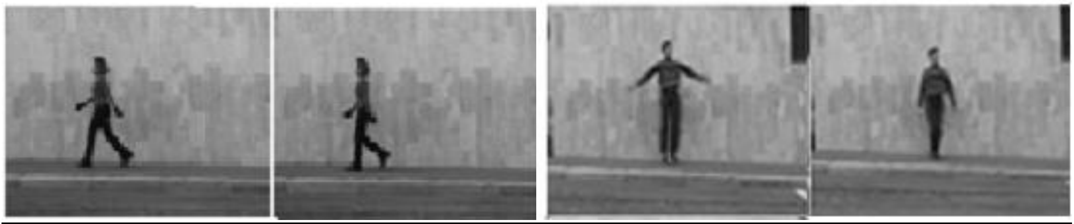
Posterior Probability of x being jumping α prior probability of jumping x likelihood of x given jumping $= \frac{10}{41} \times \frac{1}{10} = \frac{1}{41}$

Finally x classified as Running

Jacking Running



Wave one hand Walking from left to right



Walking from right to left Wave Two Hands

Fig 5 Example frames of Weizman database Activities

Running Waving

Fig 6 Examples frames of KTH database activities

4.HARS System Evaluation

This Section presents our evaluation on 86 videos from two data sets Weizman [27] and KTH[28]. 9 distinct Human activities(e.g., Walking, Running, Jumping, Handwaving,Bending,etc.) By ten different persons each person made about nine activities in varying orders and without any intentional pauses. Examples of the activities are presented in Fig 5 and Fig 6. In activity videos ,the person moves in front of a fairly uniform ,static background.

4.1 Segmentation Results of some activities

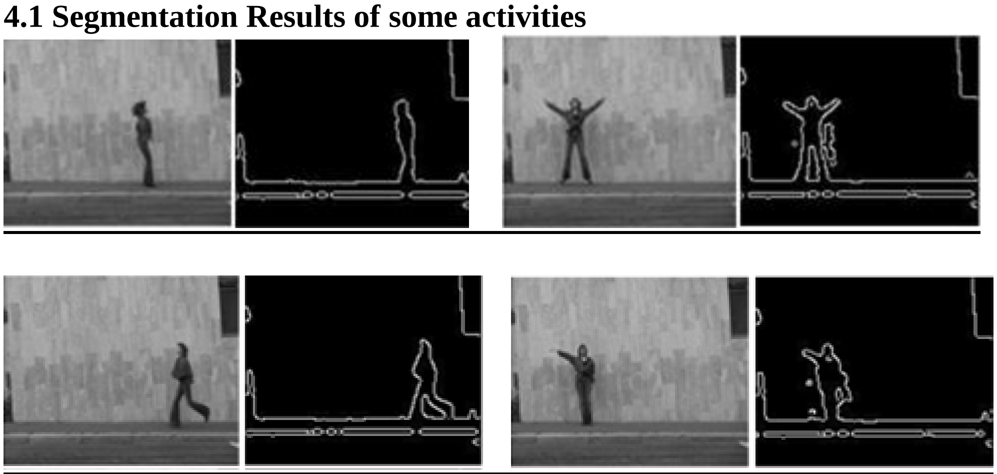
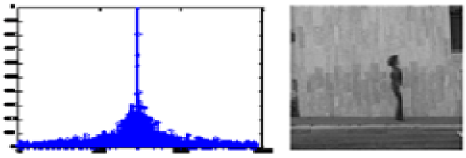
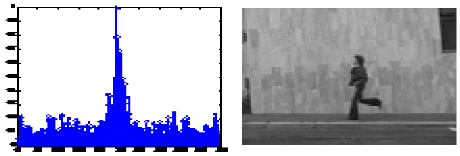


Fig 7 Segmentation by thresholding. On the left, an original image

4.2 Feature extraction of some activities

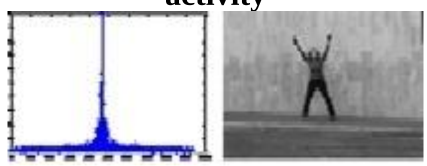


Jump Activity



Run Activity

Fig 8 An Example of Histograms of normalized Fourier descriptors for Jump and Run activity



Jack Activity

Fig 9 An Example of Histogram of normalized Fourier descriptors for Jack activity

5.RESULTS

We evaluate the performance of our proposed human activity recognition method first time by using the "two classes support vector machine"Table1 shows the confusion matrix and recognition rate .Table 2 shows the confusion matrix and recognition rate when using one versus all support vector machine classifier and table 3 shows the confusion matrix when using Naïve Bayes classifier.

Table1. Confusion matrix when using two classes SVM

	Run	Jump	Pjump	Jack	Skip	Bend	Wave one hand	Wave two hands	Walk
Run		Run:8/11 Jump:9/9	Run:11/11 Pjump:9/9	Run:11/11 Jack:9/9	Run:11/11 Skip:10/10	Run:11/11 Bend:8/9	Run:11/11 Wave one hand:10/10	Run:11/11 Wave two hands:10/10	Run:11/11 Walk:10/10
Jump	Jump:9/9 Run:8/11		Jump:9/9 Pjump:8/9	Jump:9/9 Jack:9/9	Jump:9/9 Skip:10/10	Jump:9/9 Bend:9/9	Jump:9/9 Wave one hand:10/10	Jump:9/9 Wave two hands:10/10	Jump:9/9 Walk:8/10
Pjump	Pjump:9/9 Run:11/11	Pjump:8/9 Jump:9/9		Pjump:7/9 Jack:9/9	Pjump:7/9 Skip:9/10	Pjump:8/9 Bend:8/9	Pjump:9/9 Wave one hand:10/10	Pjump:9/9 Wave two hands:10/10	Pjump:9/9 Walk:10/10
Jack	Jack:9/9 Run:11/11	Jack:9/9 Jump:9/9	Jack:9/9 Pjump:7/9		Jack:9/9 Skip:10/10	Jack:9/9 Bend:9/9	Jack:9/9 Wave one hand:10/10	Jack:9/9 Wave two hands:10/10	Jack:9/9 Walk:10/10
Skip	Skip:10/10 Run:11/11	Skip:10/10 Jump9/9	Skip:9/10 Pjump:7/9	Skip:10/10 Jack:9/9		Skip:8/10 Bend:7/9	Skip:10/10 Wave one hand:9/10	Skip:10/10 Wave two hands:10/10	Skip:9/10 Walk:8/10
Bend	Bend:8/9 Run:11/11	Bend:7/9 Jump:9/9	Bend:8/9 Pjump:8/9	Bend:9/9 Jack:9/9	Bend:7/9 Skip:8/10		Bend:2/9 Wave one hand:10/10	Bend:2/9 Wave two hands:10/10	Bend:8/9 Walk:8/10
Wave one hand	Wave one hand:10/10 Run:11/11	Wave one hand:10/10 Jump:9/9	Wave one hand:10/10 Pjump:9/9	Wave one hand:10/10 Jack:9/9	Wave one hand:9/10 Skip:10/10	Wave one hand:10/10 Bend:2/9		Wave one hand:10/10 Wave two hands:10/10	Wave one hand:9/10 Walk:9/10
Wave two hands	Wave two hands::10/10 Run:11/11	Wave two hands:10/10 Jump:9/9	Wave two hands:10l10 Pjump:9/9	Wave two hands:10/10 Jack:9/9	Wave two hands:10/10 Skip:10/10	Wave two hands:10/10 Bend:2/9	Wave two hands:10/10 Wave one hand:10/10		Wave two hands:9/10 Walk:9/10
Walk	Walk:10/10 Run:11/11	Walk:8/10 Jump:9/9	Walk:10/10 Pjump:9/9	Walk:10/10 Jack:9/9	Walk:8/10 Skip:9/10	Walk:8/10 Bend:8/9	Walk:9/10 Wave one hand:9/10	Walk:9/10 Wave two hands:9/10	
Recognition rate	97.56%	94.66%	94.66%	98.66%	93%	84.66%	93.63%	94.27%	92.36%

Jack vs not Jack	9/9
Bend vs not Bend	8/9
Jump vs not Jump	9/9
Pjump vs not Pjump	9/9
Wave one hand vs notWave one hand	9/10
Run vs not Run	8/11
Walk vs not Walk	8/10
Wave two hands vs not Wave two hands	9/10
Skip vs not Skip	6/10
Recognition rate	86.2%

Table2. Confusion matrix When using one vs all SVM

Run	8/11
Jump	8/9
Bend	7/9
Pjump	7/9
Skip	10/10
Wave one hand	8/10
Wavetwohands	7/9
Walk	10/10
Jack	9/9
Recognition rate	86.04%

Table3. Confusion matrix using Naive Bayes Classifier

6.CONCLUSION AND FUTURE WORK

In this paper we develop Human Activity Recognition system based on Fourier descriptors.The system used multi threshold segmentation method to segment moving objects from background since it's effective tool and then The system extracted object feature by using Fourier descriptors of object contour .The classification and action recognition are the last stage of the system. The system used three classification methods two classes SVM, One vs all SVM and Naive Bayes classifier .The two classes SVM achieved highest recognition rate up to 98.66% . The proposed system (HARS) opens the way to recognize human suspicious behavior since suspicious behavior is characterized as a sequence of elementary actions.

REFERENCES

[1] Gupta. A and Davis L. S, Objects in action: an approach for combining action understanding and object perception. Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Minneapolis, MN), 2007, 1-8

[2] Shabani, A. H., Clausi, D., and Zelek, J. S. Improved spatio-temporal salient feature detection for action recognition, in Proc. British Machine Vision Conference (Dundee); (2011), 1-12.

[3] Vrigkas, M., Karavasilis, V., Nikou, C., and Kakadiaris, I. A. Action recognition by matching clustered trajectories of motion vectors, in Proc.International Conference on Computer Vision Theory and Applications (Barcelona), (2013), 112-117.

[4] Lan, T., Wang, Y., and Mori, GDiscriminative figure-centric models forjoint action localization and recognition, in Proc. IEEE International Conference on Computer Vision (Barcelona); (2011), 2003-2010.

[5] Iosifidis, A., Tefas, A., and Pitas, I. Activity-based person identification using fuzzy representation and discriminant learning. IEEE Trans. Inform. Forensics Secur.7;(2012a),530-542
doi:.10.1109/TIFS.2011.2175921

[6] Quattoni, A., Wang, S., Morency, L. P., Collins, M., and Darrell, T. Hidden conditional random fields. IEEE Trans. Pattern Anal. Mach. Intell. 29, (2007). 1848-1852. doi:10.1109/TPAMI.2007.1124.

[7] Morariu, V. I., and Davis, L. S. Multi-agent event recognition in structured scenarios, in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition(Colorado Springs, CO);(2011),3289-3296.

[8] Sigal, L., Isard, M., Haussecker, H., and Black, M. J. Loose-limbed people: estimating 3D human pose and motion using non-parametric belief propagation. Int. J. Comput. Vis. 98; (2012b), 1548. doi:10.1007/s11263-011-0493-4.
[9] Tran, K. N., Kakadiaris, I. A., and Shah, S. K. Part-based motion descriptor image for human action recognition. Pattern Recognit. 45;(2012).2562-2572.doi:10.1016/j.patcog.2011.12.028.
[10] Rocio Diaz de Leon,Luis Enrique Sucar.Human silhouette recognition with Fourier descriptors (2000) doi :10.1109/ICPR.2000.903643.
[11] Picard, R. W. Affective Computing. Cambridge, MA: MIT Press; (1997).
[12] Liu, N., Dellandra, E., Tellez, B., and Chen, L. Associating textual features with visual ones to improve affective image classification, in Proc. International Conference on Affective Computing and Intelligent Interaction, Lecture Notes in Computer Science; (2011b). Vol. 6974 (Memphis, TN), 195204.
[13] Martinez, H. P., Yannakakis, G. N., and Hallam, J., Don't classify ratings of affect; rank them! IEEE Trans. Affective Comput. 5, 314326.(2014) doi:10.1109/TAFFC.2014.2352268. 7
[14] Pantic, M., and Rothkrantz, L. Towards an affect-sensitive multimodal human-computer interaction, in Proc. IEEE, Special Issue on Multimodal Human-Computer Interaction, Invited Paper; (2003). Vol. 91 (IEEE), 1370-1390.
[15] Candamo, J., Shreve, M., Goldgof, D. B., Sapper, D. B., and Kasturi, R. Understanding transit scenes: a survey on human behavior-recognition algorithms. IEEE Trans. Intell. Transp. Syst. 11;(2010), 206-224doi:. 10.1109/TITS.2009.2030963
[16] Song, Y., Morency, L. P., and Davis, R. Multimodal human behavior analysis: learning correlation and interaction across modalities, in Proc. ACM International Conference on Multimodal Interaction (Santa Monica, CA); (2012a). 27-30.
[17] Vrigkas, M., Nikou, C., and Kakadiaris, I. A. Classifying behavioral attributes using conditional random fields, in Proc. 8th Hellenic Conference on Artificial Intelligence, Lecture Notes in Computer Science; (2014b). Vol. 8445 (Ioannina), 95-104.
[18] Patron-Perez, A., Marszalek, M., Reid, I., and Zisserman, A. Structured learning of human interactions in TV shows. IEEE Trans. Pattern Anal. Mach. Intell. 34; (2012). 2441-2453.doi:10.1109/TPAMI.2012.24.
[19] Marn-Jimnez, M. J., Noz Salinas, R. M., Yeguas-Bolivar, E., and de la Blanca, N. P. Human interaction categorization by using audio-visual cues. Mach. Vis. Appl. 25; (2014). 71-84. doi:10.1007/s00138–013–0521–1.
[20]Fathi, A., Hodgins, J. K., and Rehg, J. MSocial interactions: a first-person perspective;(2012). in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Providence, RI), 1226-1233.
[21] Park, H. S., and Shi, J.. Social saliency prediction, in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Boston, MA);(2015), 4777-4785.
[22] https://en.wikipedia.org/wiki/Otsu%27s method
[23] https://en.wikipedia.org/wiki/Canny edge detector
[24] Cyrille Beaudry, Renaud Peteri, and Laurent Mascarilla . Action recognition in videos using frequency analysis of critical point trajectories. In Proc.IEEE International Conference on image processing ;paris, France,2014.,pages 1445–1449.
[25]SergiosTheodoridis,KonstantinosKoutroumbas,Patternrecognition Second Edition 5.
[26] Naive Bayes text classification@http://nlp.stanford.edu/IR book/html/htmledition/naive-bayes-text-classification-1.html
[27] Lena Gorelick,Moshe Blank,Eli Shechtman,Michal Irani,and Ronen Basri.Actions as space-time shapes. IEEE Trans.Pattern Anal. Machine Intell,29 ;December 2007 (12):2247–2253,.
[28]C.Schuldt,I.Laptev,andB.Caputo.Recognizing human actions:a local svm approach, In Proc.Int.confPatternRecognition,2004;volume 3,pages 32–36 vol.3.