View Invariant Movement Recognition by using Adaptive Neural Fuzzy Inference System

V. Anitha[#1], K.S .Ravichandran[*2], B. Santhi[#3]

School of Computing, SASTRA University, Thanjavur-613402, India
[1] anithavijay28@gmail.com

ABSTRACT:- Action recognition is an essential in the field of visual surveillance system to identify the motion objects. This paper presents the adaptive neural fuzzy inference system for detecting the various human movements from the motion video sequences. Human body posture prototypes are identified by self organizing map. It also recognizes the continuous human movements in the video sequences. Fuzzy inference system is proposed to identify the action classification. This method maps, a set of input data into a set of desired output and it matches the data from the training and testing phase. Bayesian framework is used to recognize the various kinds of activities and produce the recognition results. The proposed work is able to detect continuous instances of similar action performed by several people in various view points accurately than other existing methods in the recent literature.

Keywords – ANFIS, bayesian frameworks, human movement recognition, multilayer perceptrons, self organizing map, view invariance.

I. INTRODUCTION

Human movement recognition is the most widely used approach for video analysis. It is considered as an important problem for many applications in the visual surveillance in compression techniques [2], human and machine interface [3], analysis of video events , entertainment and sports etc. The term movement describes the human action events in a small portion of time. Action is discriminate from the activity. An activity is a continuous event of small atomic actions. For example the activity jogging contains the movements like walk, run and jump etc. Recognizing the human action is a very challenging problem because the actions may appear in a different manner depending upon the event such as similar actions with various garbs, action may be performed by different kinds of people in multiple viewpoints or different people performed the same action but it may appear in various ways [1].

Representation of human action is used to match the similarity of all human body poses by a self organizing map (SOM) in a neural network. In the training phase SOM is used to train the data in the posture images and represents the actions also. Adaptive neural fuzzy inference system is mainly used for testing the data in the posture images and produces the aggregation results for human actions in the testing phase. It utilizes the fuzzy rules and the membership functions parameters [5]. For action classification Fuzzy inference system (FIS) is proposed. It automatically calculates the parameter values without human direct interference. This method is very efficient to reduce the computational effect. Bayesian Framework is used to recognize the unknown actions and also produces the combined recognition results with high classification accuracy [6].

II. RELATED WORK

Alexandros Iosifidis et al., proposed neural network method for recognizing the actions with multiple view points [1]. Barr proposed video compression techniques [2]. These techniques are mainly used in human and machine interface [3], analysis of video events [4] and sports. Ali and Shah proposed kinematic features for Action recognition. It represents the complex human action in the videos. Kinematic features not view invariant because the same action viewed from different viewing angle. Occlusion will also affect the performance of the action [7]. Seo and Milan far proposed regression kernel analysis. It captures the data even in the presence of misrepresentation of action and error present in the data. It also finds the similarity actions and does not need prior knowledge about actions [4]. Chacon-Murguia and Sergio Gonzalez-Duarte proposed Adaptive neural fuzzy inference system (ANFIS) mthod. In this method Mamdani and sugeno fuzzy system are used. It determines the parameter value automatically according to the data and also decreases the computation time [5]. Lanz proposed Bayesian framework technique. It is used to recognize the abnormal movements and produces the most combined recognition rates [6]. Ahmad and Lee proposed Hidden Markov Model. It recognizes the actions from random view instead of any particular view. It is used to represent the actions from various viewing angles. HMM for movement recognition is used to create the time series data. It is the most widely used approach for speech and word recognition [8]. Lena Gorelick et al., proposed Weizmann datasets. The motion video sequences of movement recognition are collected from Weizmann datasets [12].

Gkalelis et al., proposed linear discriminant analysis (LDA) and fuzzy vector quantization (FVQ). These methods have the capability to distinguish the similar movements. LDA reduces the dimensionality of the multiview movement video features. This method is powerful because low dimensionality features produce the recognition accuracy. It finds only the linear combination of features in a class of objects or events . FVQ is used to match the input image vector to basic movement pattern space. It also increases the quality of the input vector [9]. Lv and Nevatia proposed Pyramid Match Kernel algorithm. It provides the comparable rate between two same characteristics of the images. It achieves comparable result and lower computational cost. It also reduces the difficulty of movement recognition problem. But the single view action classification needs a large number of parameters to solve the ambiguity of the classification [10]. Yu et al., proposed appearance based gait recognition. It is valuable for robust gait recognition system. This method is not suitable to recognize the human action from the side view and also from various viewing angles [11]. Benezeth et al., proposed Background elimination method.It is the method for detecting the moving object. Moving objects are mainly used to create the binary images with black background [13]. Kohonen proposed self organizing map algorithm. It is used to identify the images with similar features and also grouping different portion of images. It is mainly used for training the data by using the unsupervised learning method [15]. Weinland et al., proposed principal component analysis (PCA). It is mostly used to decrease the high dimensional image features into low dimensional image features. It is useful for view invariant recognition for larger class of primitive actions. It does not perform linear separation and linear regression of classes and it does not perform the similar human actions also [17].

III. MATERIALS AND METHODS

III.A. EXPERIMENTAL SETUP

Human action recognition is automated detection of ongoing events from video data. Action recognition is the finding of video segments containing such actions. Video segment is used to display the properties of the actions. The video sequences are collected from Weizmann datasets [12]. The video sequences are converted into frames and stored in the database. It contains the actions are Bend, jump in place, run, walk and wave two hands etc.

The proposed method consists of identification of posture prototypes, testing of data with ANFIS method, action classification and action recognition. The overview of the proposed method is as shown in the Fig. 1.

The diagram represents a SOM is used to train the data in the training phase. Initially video sequences are converted into frames. The background elimination method is used to create the binary images. Then the binary images are trained. For extracting the features of image edge detection method is represented. It is mainly used for finding the disconnection in the binary images. It also decreases the amount of data in the binary images. After classification of SOM, Fuzzy inference system automatically tests the data. Finally a Bayesian framework is used to recognize the actions.
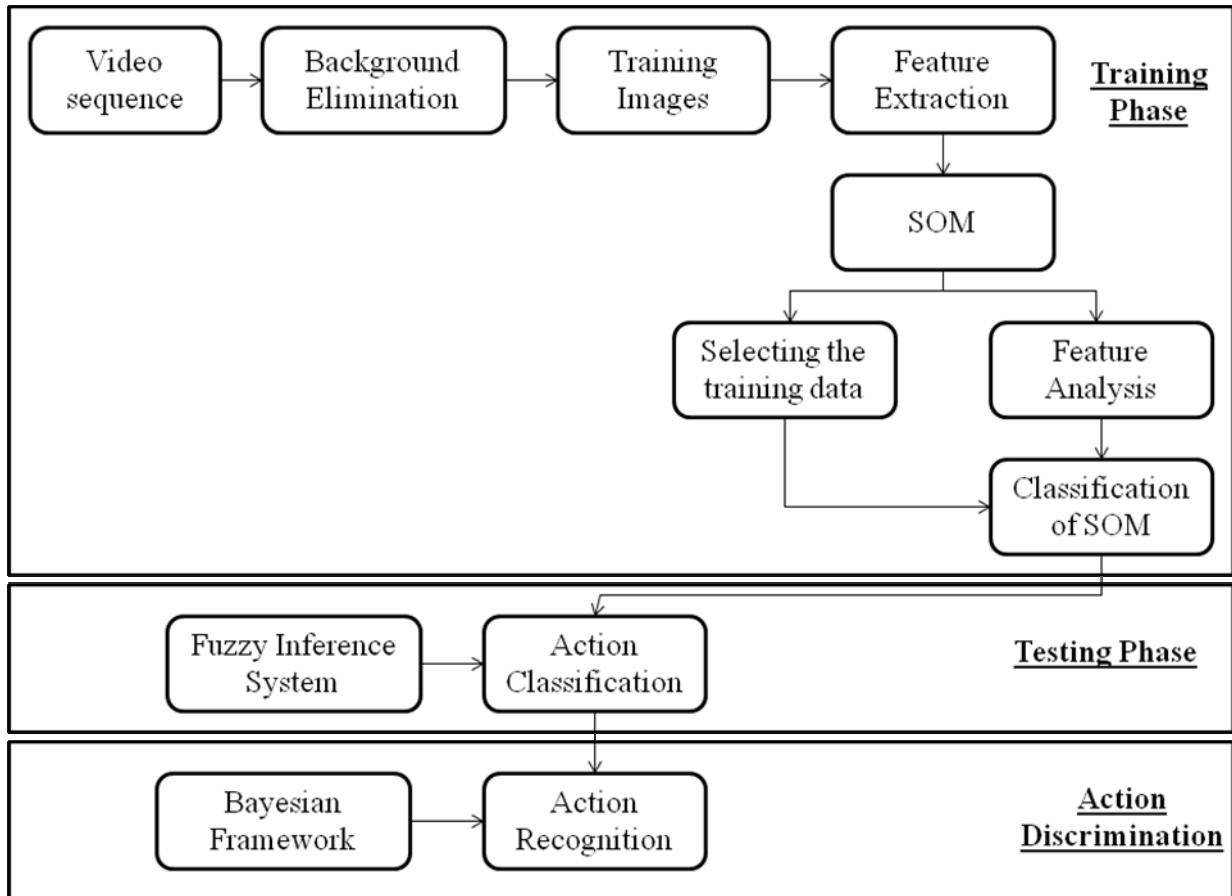
Fig. 1:A typical overview of the proposed method

III.B. METHODS

1. PREPROCESSING PHASE

In action recognition, elementary action video sequences are converted into video frames. Moving object segmentation techniques [13, 14] are used to create binary images. Background elimination is a widely used approach for detecting the moving object. After the background elimination, the person's body is extracted and produces the posture frames of binary images with the similar size. Binary image frames of five actions are bend, jump in place, run, walk and wave two hands by using the edge detection method is represented in Fig 2. It is the essential methods for detecting the edges in the binary images.

Fig. 2: Binary image frames of five actions

2. IDENTIFICATION OF POSTURE IMAGES

In the training phase videos sequences of the posture images are clustered to a fixed number of classes using a self organizing map (SOM) algorithm [15]. It is a special class of neural networks. It uses the unsupervised learning method which does not need any external resources for getting the desired output. The SOM is used to identify and grouping different portion of images with similar features. An output neuron with the smallest value is determined as the winner in a competition that unit is called best matching unit.

The basic process for training the data based on SOM has the following steps.

1) Initialization: Weights are initialized randomly.

2) Sampling: Produce the sample X and give it to the network.

3) Similarity Matching: The winning neuron N is mapped with the weight of the input vector. It is considered as the best matching neuron.

$$N = \text{argmin (j) } (X - W_j)$$

4) Updating: Adjust the parameter of the neighbourhood function.

$$\Delta W = \gamma . \, h_{ij} \, (X - W_j)$$

Where h_{ij} is the neighbourhood function, γ is the learning rate dependent on time.

The algorithm is trained up to 100 iterations. This procedure is applied multiple times for training the data which were not trained. Finally it represents the actions.

SOM is also used to recognize the continuous movement of human actions. Continuous movement of bend postures is shown in Fig 3.

Fig. 3: Continuous movement of bend postures

3. TESTING AND CLASSIFICATION OF DATA WITH ANFIS METHOD

In this phase, the user gives an input posture image for which the corresponding output image is tested. Here the input data is normalized and then checked with the ANFIS method. It uses the sugeno type fuzzy inference system for training routine. It utilizes the automatic identification of fuzzy rules and membership function parameters [5].

4. FIS CLASSIFIER

The Fuzzy Inference System is used to compute the output value for the given input value. Here, Sugeno type fuzzy inference system is to identify the parameter value. It is the complicated method but it gives the probable results which are more efficient. In action classification, FIS classifier completed the training of data up to 100 epochs. Once trained, FIS is used to classify the each testing data in the posture images and classify the actions depending upon the images are already trained by SOM. Finally it produces the most repeated occurrences of actions.

5. ACTION RECOGNITION

In the action recognition phase, video frames are segmented by using the background elimination method and the features are also extracted. The input frame is compared with the posture retained in the database. If a same posture is obtained, the posture is allocated for the label name of the current frame. Otherwise the new label name is assigned to the current frame of the posture which is retained in the database.

In the Bayesian framework case, the human actions are fed to the FIS classifier to recognize the corresponding action that computes the most combined recognition rate depending on the Bayesian approach [6]. It produces the combined recognition results with high classification accuracy. A confusion matrix represents the most probable recognition rate is shown in Table 1. Finally it recognizes the action such as bend, walk and run etc. A recognition rate obtained for Bayesian approach is 86.66%.

IV. RESULTS AND ANALYSIS

The results and discussions of the human action recognition is based on Bayesian approach. There are two phases in the recognition method. In the training phase, SOM is trained and matches the similarity of all human actions. In the testing phase, FIS is used to test the data in the posture images and produces the aggregation results for human actions.

In action recognition video sequences are collected from the Weizmann datasets. Here 20 videos from the Weizmann datasets are used for action recognition. Each video describes one human performing one action. The video sequences are converted into frames and stored in the database. It contains the action such as bend, walk and run, jump in place and wave two hands etc.

The input image is taken from the database as shown in Fig 3 (a). The grayscale image is converted into a binary image using edge detection method. It detects the wide range of edges in the image. The binary image is as shown in Fig 3 (b). The binary image is segmented for clearly represent the action. By using segmentation techniques actions are easier to analyze. It is also used for extracting foreground from background model. The segmented image is as shown in fig 3 (c). The input image is matched with the actions in the database. Here the input image is matched with the posture retained in the database. If a same posture is obtained, the posture is allocated for the label name of the current frame. Otherwise the new label name is assigned to the current frame of the posture which is retained in the database. Finally matches the similarity of the action and recognize the actions such as bend, walk and run.
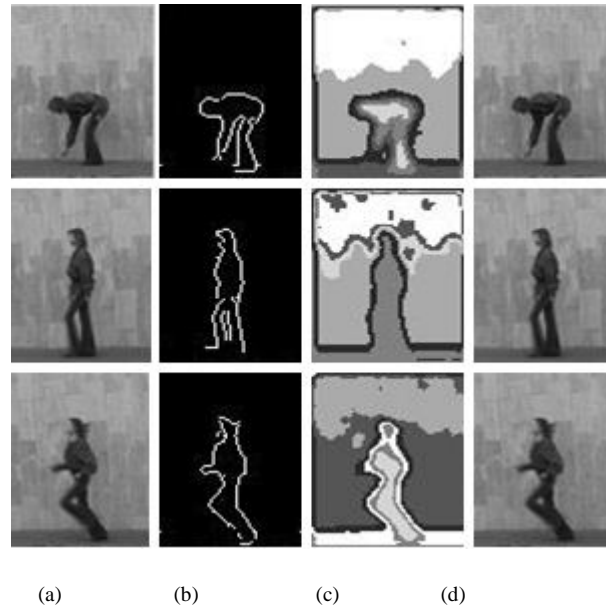
(a)   (b)   (c)   (d)
Fig. 3: Action recognition (a) Input image (b) Binary image
(c) Segmented image (d) Matched image

IV.A. ANALYSIS

Bayesian approach is used to recognize the action and the result is presented in Table I by using the confusion matrix [16]. It consists of information about known class and predicted class. Here rows of the matrix describe the known defined values and columns of the matrix describe the predicted values. The diagonal values are classified perfectly and the off-diagonal values are incorrectly classified.

The overall correct classification rate is equal to 86.66% for Bayesian approach. An action which contains different body poses like bend is almost perfectly classified. Similar body poses like walk and run are difficult to be correctly classified.

TABLE I
Confusion matrix for three actions

Predicted / Observed	Bend	Walk	Run
Bend	19	1	0
Walk	2	16	2
Run	1	2	17

IV.B. PERFORMANCE METRICS

Performance metrics compare the strength and weakness of different classifiers by computing the precision, recall and F1 metric [16]. Performance metrics and accuracy results are described in the following.

1. ACCURACY

It is the measure of the total number of predictions that were perfectly classified.

$$Accuracy = (T_P+T_N) / (T_P+ T_N + F_p +F_N)$$

The overall accuracy of the human action recognition is 86.66% as shown in Table II .

2. PRECISION

It is the measure of specific cases predicted based on positive classes.

$$Precision = T_P / (T_P+F_P)$$

3. RECALL

It is the measure of positive cases that were correctly calculated. It is also called sensitivity. It is similar to the true positive rate.

$$\text{Recall} = T_P / (T_P + F_N)$$

TABLE II
Performance Metrics

Metrics	Bend	Walk	Run
Precision	0.8636	0.8421	0.8947
Recall	0.9500	0.8000	0.8500
F1	0.9047	0.8205	0.8718
Similarity	0.8260	0.6956	0.7727
Specificity	0.9250	0.9250	0.9500

4. F1 METRIC

Figure of metric or F measure is the weighted mean of precision and recall.

$$\text{F1 metric} = 2 \, (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision})$$

5. SIMILARITY

It is the measure of calculating the two or more different actions from the database.

$$\text{Similarity} = T_P / (T_p + F_N + F_P)$$

6. SPECIFICITY

It is the measure of negative cases classified correctly. It is same as the true negative rate.

$$\text{Specificity} = T_N / (T_N + F_P)$$

The experiments were conducted and the values are computed and tabulated.

V. CONCLUSION AND FUTURE WORK

View invariant action recognition method for an adaptive neural fuzzy inference system to solve the generic action recognition problem. ANFIS is the very useful tools to train the images. It is a quick and straightforward way of input selection. SOM is constructed from the dataset processing and training the data and the input query is tested which is based on ANFIS. FIS classifier is used for classifying the given actions. It measures the similarity between images and produces the classification results. Bayesian approach is used to recognize the human actions using a single video sample. This method also recognizes the continuous human action.

In future, this method can detect the human interaction between persons and also calculate the abnormal representation of human movements.

REFERENCES

[1] Alexandros Iosifidis, Anastasios Tefas, and Ioannis Pitas," View-invariant action recognition based on artificial neural networks," *IEEE Trans. Neural Netw.*, vol. 23, no. 3, pp. 412–424, Mar. 2012.
[2] B. Song, E. Tuncel, and A. Chowdhury, "Toward a multi-terminal video compression algorithm by integrating distributed source coding with geometrical constraints," *J. Multimedia*, vol. 2, no. 3, pp. 9–16, 2007.
[3] P. Barr, J. Noble, and R. Biddle, "Video game values: Human-computer interaction and games," *Interact. Comput,* vol. 19, no. 2, pp. 180–195, Mar. 2007.
[4] H. J. Seo and P. Milanfar, "Action recognition from one example," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 867–882, May 2011.
[5] Mario I. Chacon-Murguia and Sergio Gonzalez-Duarte," An adaptive neural-fuzzy approach for object detection in dynamic backgrounds for surveillance systems," *IEEE Trans. Ind. Electron.*, vol. 59, no. 8, pp.3286–3298, Aug. 2012.
[6] O. Lanz, "Approximate Bayesian multibody tracking," *IEEE Trans.Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1436–1449, Sep. 2006.
[7] S. Ali and M. Shah, "Human action recognition in videos using kinematic features and multiple instance learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 2, pp. 288–303, Feb. 2010.
[8] M. Ahmad and S. Lee, "HMM-based human action recognition using multiview image sequences," in *Proc IEEE Int. Conf. Pattern Recognit.*, vol. 1. Sep. 2006, pp. 263–266.
[9] N. Gkalelis, N. Nikolaidis, and I. Pitas, "View independent human movement recognition from multi-view video exploiting a circular invariant posture representation," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jun. – Jul. 2009, pp. 394–397.
[10] F. Lv and R. Nevatia, "Single view human action recognition using key pose matching and viterbi path searching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
[11] S. Yu, D. Tan, and T. Tan, "Modeling the effect of view angle variation on appearance-based gait recognition," in *Proc. Asian Conf.*

Comput. Vis., vol. 1. Jan. 2006, pp. 807–816.

[12] Lena Gorelick and Moshe Blank and Eli Shechtman and Michal Irani and Ronen Basri," Actions as Space - Time Shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2247-2253, Dec. 2007.

[13] Y. Benezeth, P. Jodoin, B. Emile, H. Laurent, and C. Rosenberger, "Review   and evaluation of commonly- pp.implemented background subtraction algorithms," in *Proc. 19th Int. Conf. Pattern Recognit.*, Tampa, FL,Dec. 2009, pp. 1–4.

[14] M. Piccardi, "Background subtraction techniques: A review," in *Proc. IEEE Int. Conf. Syst., Man Cybern,* vol. 4. Oct. 2004, pp. 3099– 3104.

[15] T. Kohonen, "The self-organizing map," *Proc. IEEE*, vol. 78, no. 9, pp. 1464–1480, Sep. 2002.

[16] L. Maddalena and A. Petrosino, "A self-organization approach to background subtraction for visual surveillance applications," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1168–1177, Jul. 2008.

[17] D. Weinland, R. Ronfard, and E. Boyer, "Free viewpoint action recognition using motion history volumes," *Comput. Vis. Image Understand.*, vol. 104, nos. 2–3, pp. 249–257, Nov. –Dec. 2006.