An Efficient Human Action Recognition System Using ANFIS

S. Deepak Selvamani[1], G. Thavaseelan[2]

PG Scholar[1], Assistant Professor[2],
St.Peter's University, TN, India.

Abstract: - Action recognition is needed in the ground of visual examination arrangement to recognize the signal articles. This paper presents the adaptive neural fuzzy inference system for recognizing the various human activities from the indication video series. Human body situation patterns are accepted by self organizing map. It's been familiar with the constant human arrangements in the video series. Fuzzy inference system is suggested to detect the action organization. This technique maps, an established of response data into a set of desired. Bayesian framework to identify the various kinds of actions and appreciation results are formed for each camera. The suggested method is able to determine different occurrences of matching action completed by different people in different viewpoints truthfully than other standing methods in the modern literature.

Keywords: ANFIS, bayesian frameworks, human movement recognition, self organizing map, view invariance.

I. INTRODUCTION

Human movement acknowledgment is the most commonly used approach for video exploration. It is considered as a main problem for several presentations in the visual investigation in looseness techniques, human and mechanism edge, analysis of video events, theatre and sports etc. The term programme describes the human action dealings in a small percentage of time. Action is classify from the activity. A disorder is a constant event of small atomic appointments. For example the activity pushing contains the engagements like walk, run and jump etc. Be familiar with the human action is a very routine problem because the actions may perform in a different custom depending upon the event such as comparable actions with countless garbs, action may be performed by changed kinds of people in multifactorial vantage point or dissimilar people completed the same action but it may appear in several ways . Protest of mortal achievement is used to equal the communiqué of all mortal body approaches by a self organizing map (SOM) in a neural complex. In the planning phase SOM is used to train the data in the posture images and embodies the actions also.

ICGPC 2014
St.Peter's University, TN, India.

Adaptive neural unclear corollary system is mainly used for testing the data in the position images and harvests the accumulation results for human actions in the testing stage. It develops the prying. This method is very efficient to reduce the computational effect. Bayesian Background is used to identify the unknown actions and also produces combined gratitude results with high cataloguing accuracy. Unclear rules and the connection occupations parameters .For action arrangement Fuzzy reading system (FIS) is future. It robotically calculates the bound values without human direct Human action gratitude is a broadly studied area in computer vision. Its requests include investigation systems, video analysis, manufacturing and a variety of systems that involve exchanges between publics and automated devices such as human-computer borders. Its growth began in the early 1980s. To date, research has mainly attentive on knowledge and distinguishing actions from video categorizations taken by a single evident light camera. There is general nonfiction in action appreciation in a number of fields, counting computer vision, engine learning, decoration gratitude, signal dispensation, etc.. Between the changed types of landscapes for depiction, shapes and spatio-temporal curiosity points are most normally used. The methods planned in the past for shadow based action gratitude can be divided into two major groupings. One is to extract action descriptors from the arrangements of shadows. Predictable classifiers are normally used for gratitude . The other one is to extract features from each shadow and model the subtleties of the action obviously.

II. RELATED WORK

[1]Alexandros Iosifidis et al., proposed neural network method for recognizing the actions with multiple view points.[2] Song, E. Tuncel used video compression using geometric constraints.[3]Barr proposed video compression techniques . These techniques are mainly used in human and machine interface,analysis of video events and sports. [4]Seo and Milan far proposed regression kernel analysis. It arrests the data uniform in the occurrence of leaning of exploit and wrongness present-day in the data. It also finds the appraisal actions and does not requirement earlier familiarity overweight actions. [5]Chacon-Murguia and Sergio Gonzalez-Duarte proposed Adaptive neural fuzzy evaluation system (ANFIS) method. In this method Mamdani and sugeno ambiguous coordination are used. It determines the parameter value habitually according to the data and also falls the computation time.

[6]Lanz proposed Bayesian background technique. It is used to make out the abnormal appointments and produces the most collective recognition rates. [7]Ali and Shah proposed kinematic structures for achievement recognition. It represents the multipart human action in the videos. Kinematic features not view invariant because the same action viewed from different viewing angle. Blocking will also distress the presentation of the accomplishment. [8]Ahmad and Lee projected Hidden Markov Model. It be familiar with the actions from random view instead of any actual view. It is used to represent the actions from a number of viewing angles. HMM for association appreciation is used to create the time series data. It is the most widely used attitude for speech and expression recognition.

[9]Gkalelis et al., proposed linear discriminant analysis (LDA) and fuzzy vector quantization (FVQ). These methods have the ability to decide the similar movements. LDA moderates the dimensionality of the multi view undertaking video structures. This method is powerful for the reason that low dimensionality features food the recognition accuracy. It finds only the linear combination of features in a class of objects or events. FVQ is used to tie the input appearance vector to plain movement form . It also increases the quality of the input vector. [10] Lv and Nevatia proposed Pyramid Match Kernel algorithm. It delivers the similar rate between two same physical characteristics of the images. It reaches analogous outcome and lesser computational cost. It also decreases the effort of movement recognition problem. But the single view action arrangement needs a bulky number of factors to solve the uncertainty of the classification.

[11]Yu et al., suggested appearance founded gait recognition. It is valuable for robust gait recognition system. This method is not fit to identify the human act from the side vision and also from various looking positions. [12]Lena Gorelick et al., recommended Weizmann datasets. The wave video arrangements of movement recognition are composed from Weizmann datasets . [13] Benezeth et al., suggested Background rejection method. It is the method for sensing the stirring object. Stirring objects are mostly used to make the binary metaphors with black background.

[14] Piccardi produces circumstantial edge detection. [15]Kohonen suggested self organizing map algorithm. It is used to detect the images with similar structures and also group diverse portion of images. It is mostly used for keeping fit the statistics by using the unsubstantiated learning method.[16]Maddalena methods background deduction for visual observation uses [17]Weinland et al., recommended principal component analysis (PCA). It is generally used to reduction the extraordinary dimensional matching features into little dimensional image features. It is beneficial for version invariant recognition for larger class of simple actions. It does not perform rectilinear separation and direct deterioration of modules and it does not accomplish the like human actions also.

III. EXPERIMENTAL SETUP

Human action recognition is preset finding of enduring events from video data. Action recognition is the definition of video divisions containing such actions. Video segment is used to show the assets of the actions. The video arrangements are together from Weizmann datasets. The video arrangements are transformed into borders and kept in the database. It holds the actions are Bend, jump in place, run, and walk movement hands etc. The recommended method can be made up of identification of bearing prototypes, challenging of data with ANFIS method, act classification and performance recognition. The outline of the suggested method is as shown in the Fig. 1.

The diagram signifies a SOM is used to sequence the data in the training phase. Primarily video structures are converted into frames. The contextual rejection method is used to generate the binary images. Then the binary images are trained. For removing the structures of image edge detection method is represented. It is mainly used for finding the withdrawal in the binary images. It also falls the quantity of data in the binary images. Subsequently grouping of SOM, Fuzzy consequence system automatically trials the data. Finally a Bayesian framework is used to differentiate the actions.

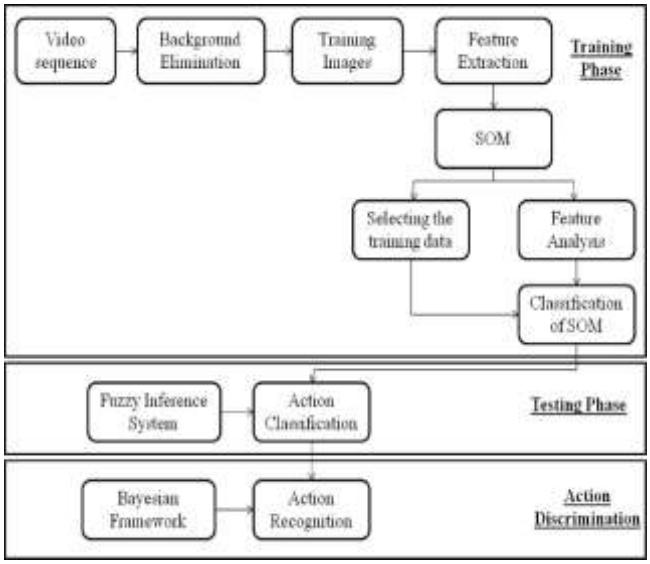

Fig1 overview of proposed method

A. PREPROCESSING PHASE

In action recognition, basic action video structures are transformed into video frames. Moving object partition techniques are used to generate binary images. Background rejection is a widely used methodology for detecting the stirring object. After the background elimination, the person's body is removed and yields the posture frames of binary images with the analogous size. Binary image frames of five activities are bend, jump in place, run, walk and movement of dualistic hands by means of the edge detection method is represented in Fig 2. It is the needed methods for detecting the edges in the binary images.

B. IDENTIFICATION OF POSTURE IMAGES

In the training segment sequences of the posture images are assembled to a fixed number of classes using a self organizing map (SOM) algorithm. It is a special class of neural networks. It uses the unsubstantiated learning method

which does not want any peripheral resources for getting the preferred output. The SOM is used to identify and combination of different portion of images with analogous features. An output neuron with the smallest value is resolute as the winner in a competition that unit is called best matching unit.

- *Initialization*: Weights are modified casually.
- *Sampling*: Produce the sample X and give it to the network.
- *Similarity Matching*: The winning neuron N is mapped with the weight of the input vector. It is considered as the best matching neuron.

$$N = \text{argmin}(j)\ (X - W_j) \quad (1)$$

- *Updating*: Adjust the parameter of the neighbourhood function.

$$\Delta W = \gamma .\ h_{ij}\ (X - W_j) \quad (2)$$

Where h_{ij} is the neighbourhood function, γ is the information rate reliant on time.The algorithm is skilled up to 100 iterations. This practice is applied multiple times for training the data which were not trained. to finish it represents the trial.SOM is also used to distinguish the continuous movement of human actions. Continuous progress of twist postures.

C. TESTING AND CLASSIFICATION OF DATA WITH ANFIS METHOD

In this segment, the client gives an input posture picture for which the resultant output image is hardened. Here the effort data is normalized and then checked with the ANFIS manner. It uses the sugeno type furry conclusion system for guidance routine. It utilizes the habitual credentials of fluffy rules and membership function parameters.

D. FIS CLASSIFIER

The wooly suggestion classification is used to work out the production value for the given enter value. Here, Sugeno type fuzzy inference system is to spot the bound value. It is the difficult method but it gives the apparent outcome which are more capable. In exploit classification, FIS classifier fulfilled the training of data up to 100 epochs. Once skilled, FIS is used to classify the each taxing data in the posture images and catalog the proceedings depending upon the images are already trained by SOM. as a final point it produces the most frequent occurrences of trial.

E. ACTION RECOGNITION

In the action acknowledgment part, video frames are segmented by using the surroundings eradication way and the skin are also extract. The input frame is compared with the posture retained in the record. If a identical posture is obtained, the posture is billed for the label name of the recent frame. Otherwise the new label name is assigned to the current frame of the posture which is retained in the database.

In the Bayesian structure case, the individual actions are fed to the FIS classifier to be familiar with the matching action that computes the most collective recognition rate depending on the Bayesian loom . It produces the mutual respect results with high sorting precision. A uncertainty medium represents the most feasible identification rate is shown in Table 1. lastly it recognizes the action such as bend, walk and run etc. A recognition rate obtained for Bayesian approach is 86.66%.
(a)Input image (b)Binary image (c)Segmented image
(d)Matched image

IV. RESULTS OF ACTION REGONITION

The results and discussions of the human action recognition is based on Bayesian approach. There are two phases during the papers method. In the coaching phase, SOM is skilled and matches the similitude of all human actions. In the testing phase, FIS is secondhand to test the in rank in the position images and produces the aggregation cost for human actions. In action recognition video sequences are poised from the Weizmann datasets. Here 20 videos from the Weizmann datasets are old for action recognition. Each video describes one mortal performing one action. The film sequences are altered into frames and stored in the database. It contains the action such as interweave, stroll and run, step in place and signal two hands etc.

The input image is in use from the database as publicized in Fig 3 (a). The grayscale image is converted into a dual image using edge detection method. It detects the spacious collection of boundaries 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 evaluate. It is also used for extracting fore from setting 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 retain in the database.

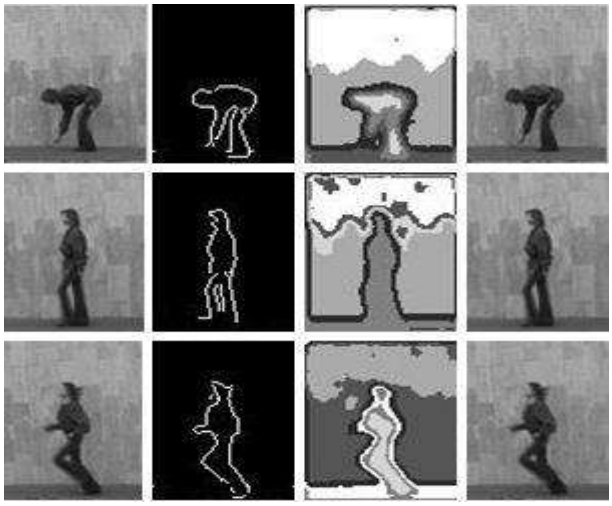

Fig 3 (a) Grayscale (b) Binary (c) Input (d) Matched

If an alike posture is obtained, the posture is owed for the label name of the existing frame. Otherwise the new label name is assigned to the current frame of the posture which is retained in the database. Finally matches the relationship of the action and recognize the actions such as curvature, hike and dart.

A. ANALYSIS

Bayesian style is used to diagnose the action and the result is presented in Table I by using the mistake matrix . It consists of information about known class and likely class. Here rows of the matrix style the known defined values and posts of the matrix describe the expected values. The transverse values are classified seamlessly and the off-diagonal values are falsely classified.

The overall true classification rate is equal to 86.66% for Bayesian style. An action which comprises changed body poses like bend is almost seamlessly classified. Parallel body poses like walk and run are tough to be suitably ordered.

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

B. PERFORMANCE METRICS

Presentation metrics equal the strong point and softness of different classifiers by computing the accuracy, recall and F1 metric. Presentation metrics and truth results are styled in the following.

- *ACCURACY:* It is the measure of the total number of predictions that were perfectly classified.

$$\text{Accuracy} = (T_P+T_N) / (T_P+ T_N + F_p +F_N) \quad (3)$$

- *PRECISION*: It is the measure of specific cases expected based on confident.

$$\text{Precision} = T_P / (T_P+F_P) \quad (4)$$

- *RECALL*: It is the measure of constructive cases that were acceptably designed. It is also called understanding. It is similar to the true positive rate.

$$\text{Recall} = T_P / (T_P+F_N) \quad (5)$$

- *F1 METRIC*: Figure of metric or F measure is the weighted mean of accuracy and recall. F1

$$\text{metric} = 2(\text{Recall}*\text{Precision}) (\text{Recall} + \text{Precision}) \quad (6)$$

- *SIMILARITY*: It is the measure of calculating the two or more different actions from the database.

$$\text{Similarity} = T_P / (T_p+F_N +F_P) \quad (7)$$

- *SPECIFICITY*: It is the measure of undesirable cases secret acceptably. It is same as the true

- Undesirable rate.

$$\text{Specificity} = T_N / (T_N+F_P) \quad (8)$$

The investigates were directed and the values are computed and organized .To begin an effective action appreciation method using analysis of spatiotemporal shadows measured during the activities, based on the idea that spatiotemporal disparities of human shadows encode not only spatial evidence about body poses at certain seconds, but also dynamic information about global body motion and the motions of local body parts. It appears to be achievable to use features that can be attained from space-time shapes for travelling the action properties. In contrast to feature tracking, eliminating space-time figures is also easier to instrument using current vision machineries, specifically in the case of motionless camera.

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

V. CONCLUSION AND FUTURE WORK

View invariant achievement acknowledgment method for an adaptive neural unclear reading organism to solve the standard action appreciativeness problem. ANFIS is the very appreciated implements to InterCity the images. It is a quick and upfront way of input selection. SOM is assembled from the dataset handing out and keeping fit the figures and the feedback interrogation is tested which is based on ANFIS. FIS classifier is used for organizing the given actions. It measures the connection amongst pictures and products the grouping results. Bayesian method is used to recognize the humanoid appointments using a only video illustration. This scheme also make out the constant human action. In future, this scheme can distinguish the human communication in the middle of societies and also calculate the unusual representation of hominid 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.8

[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.