HUMAN ACTION RECOGNITION BASED ON NEURAL NETWORK

C.Seline Angel*

Abstract

In this paper based on neural network human actions are recognized. The action video is based on learning spacial related human body posture prototype using self organizing map. Self Organizing maps(SOM) is used to learn the human postures. SOM can be constructed based on three procedures they are competition, cooperation, adaption. In feature extraction the action recognition is expensive. So we use multilayer perceptron for action classification. fuzzy distance can produce a time inavariant action representation. Action are classified based on multilayer Perceptron. Inorder to recognize the action multi cameras are used. Action recognition is performed for each of N cameras by using MLP ie, feed forward neural network. Back propagation algorithm is trained in MLP. These actions are viewed from different angles, a view invariant action is recognized. The proposed method can be applied to videos depicting interaction between humans without any modification.

Index Terms—Multi-layer Perceptrons, Fuzzy Vector Quantization, Human Action Recognition, Bayesian frameworks.

* M.E(Computer Science and Engineering), Vins Christian college of Engineering, Chunkankadai, Nagercoil

I. INTRODUCTION

Action Recognition is one of the most challenging recognition problem in computer vision with many application such as intelligent surveillance. In the video-based surveillance application, even if the motion of persons is known, this is not sufficient to describe the posture of the person. The postures of the persons can provide important clues for understanding their activities. Therefore, accurate detection and recognition of various human postures both contribute to the scene understanding. The accuracy of the system is hampered by the use of a single camera, in case of complex situations and several people undertaking actions in the same scene. Often, the posture of people is occluded, so that the behavior cannot be realized in high accuracy [1],[2]. Actions are defined by using features based on motion information.[5],[6],[7]. A set of kinematic features that are derived from the optical flow for human action recognition in videos. The set of kinematic features includes divergence, vorticity, symmetric and antisymmetric flow fields, second and third principal invariants of flow gradient and rate of strain tensor, and third principal invariant of rate of rotation tensor. Each kinematic feature, when computed from the optical flow of a sequence of images, gives rise to a spatiotemporal pattern. set of kinematic features that are derived from the optical flow for human action recognition in videos. The visual recognition of complex movements and actions is crucial for the survival of many species. Movement recognition has been studied in psychophysical, neurophysiological and imaging experiments, and several cortical areas involved in it have been identified. We use a neurophysiologically plausible and quantitative model as a tool for organizing and making sense of the experimental data, despite their growing size and complexity. Thus, actions can be described as sequences of consecutive human body poses, in terms of human body silhouettes [10],[11],[12]. The camera viewing angle is not fixed and human actions are observed from arbitrary camera view points. Several researchers have highlighted the significant impact of the camera viewing angle variations on the action recognition performance ie,. extraction of motion descriptors from multiple cameras, and their classification into primitive actions such as raising and dropping hands and feet, sitting up and down, jumping, etc. To this new motion descriptors based on motion history volumes which fuse action cues, as seen from different viewpoints and over short time periods, into a single three dimensional representation. [19], [20]. This is the so-called viewing angle effect. To provide view-independent methods, the use of multi-camera setups has been adopted [21], [22], [23].

The main contributions of this paper are: *a*) the use of Self Organizing Maps (SOM) for identifying the basic posture prototypes of all the actions, *b*) the use of cumulative fuzzy distances from the SOM in order to achieve time-invariant action representations, *c*) the use of a Bayesian framework to combine the recognition results produced for each camera, *d*) the solution of the camera viewing angle identification problem using combined neural networks. The remainder of this paper is structured as follows. An overview of the recognition framework used in the proposed approach and a small discussion concerning the action recognition task is given in Section I-A. Section II presents details of the processing steps performed in the proposed method. Experiments for assessing the performance of the proposed method are described in Section III. Finally, conclusions are drawn in Section IV.

A. Problem Statement

Let an arbitrary number of *Nc* cameras capturing a scene at a given time instance. These cameras form a *Nc*-view camera setup. This camera setup can be a converging one or not. In the first case, the space which can be captured by all the *Nc* cameras is referred as capture volume. In the later case, the cameras forming the camera setup are placed in such positions that there is not a space which can be simultaneously captured by all the cameras. A converging and a non-converging camera setup is illustrated in Figure 1. *Nt* video frames from a specific camera **f***i; i* = 1....,*Nt*, form a single-view video f=f_1^T, f_2^T,...., f_{Nt}^T

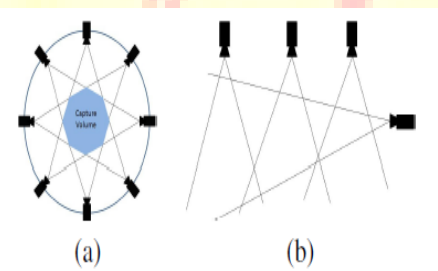

Fig. 1. a) A Converging and b) Non Converging camera setup

Actions can be periodic (e.g., walk, run) or not (e.g., bend, sit). The term elementary action refers to a single human action pattern. In the case of periodic actions, the term elementary action refers to a single period of the motion pattern, such as a walk step. In the case of non-periodic actions, the term elementary action refers to the whole motion pattern, i.e., a bend sequence.

II. PROPOSED METHOD

In this section, each step of the proposed method is described in detail. The extraction of posture data, used as input data in the remaining steps, is presented in subsection IIA. The use of a Self Organizing Map (SOM) to determine human body posture prototypes is described in subsection IIB. Action representation and classification are presented in subsections II-C and II-D, respectively. A variation of the original algorithm, that exploits the observation's viewing angle information is presented in subsections II-E and II-F. Finally, subsection II-G presents the procedure followed in the recognition phase.

A. Preprocessing Phase

An elementary action is captured by N cameras in elementary action videos consisting of Ntj ,$1 \leq j \leq NA$, video frames that depict one action period. The number Ntj may vary over action classes, as well as over elementary action videos coming from the same action class. Multi-period action videos are manually split in elementary action videos, which are subsequently used for training and testing in the elementary action recognition case.

Moving object segmentation techniques are applied to each action video frame to create binary images depicting person's body in white and the background in black. These images are centered at the person's center of mass. Bounding boxes of size equal to the maximum bounding box enclosing person's body are extracted and rescaled to $NH \times NW$ pixels to produce binary posture images of fixed size. Eight binary posture images of eight actions ('walk', 'run', 'jump in place', 'jump forward', 'bend', 'sit', 'fall' and 'wave one hand') taken from various viewing angles are shown in Figure 2.

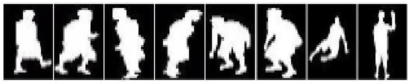

Fig 2. Posture images of eight action taken from various viewing angles.

B. Posture prototypes Identification

In the training phase, posture vectors Pi i = 1....Np *Np* being the total number of posture vectors consisting all the *NT* training action videos, having *Ntj* video frames each, are used to produce action independent posture prototypes without exploiting the known action labels. To produce spatially related posture prototypes, a SOM is used. The use of SOM leads to a topographic map (lattice) of the input data, in which the spatial locations of the resulting prototypes in the lattice are indicative of intrinsic statistical features of the input postures. The training procedure for constructing the SOM is based on three procedures:

1) *Competition:* For each of the training posture vectors **p***i*, its Euclidean distance from every SOM weight, **W**sj$\in R^D$; j = 1.....,N_S is calculated. The winning neuron is the one that gives the smallest distance:

$$j^* = \arg\min_{j} \| P_i - w_{sj} \|_2. \qquad (1)$$

2) *Cooperation:* The winning neuron j* indicates the center of a topological neighborhood h_{j*} Neurons are excited depending on their lateral distance,r_{j8} from this neuron. A typical choice of h_{j*} is the Gaussian function:

$$h_{j*}\, k(n) = \exp\left(-\frac{r_{j*k}^2}{2\sigma^2(n)}\right) \qquad (2)$$

where k corresponds to the neuron at hand, n is the iteration of the algorithm, r_{j*k}^2 is the Euclidean distance between neurons j*and k in the lattice space and is the ”effective width” of the topological neighborhood.

3) Adaptation: At this step, each neuron is adapted with respect to its lateral distance from the wining neuron as follows:

$$wsk\ (n+1) = wsk(n) + n\ (n) h_{j*k}(n)(P_i - (n))$$

where $\eta(n)$ is the learning-rate parameter: $\eta(n) = \eta(0)\ \exp(-\frac{n}{n_0})$ η(0)=0.1 in our experiments.

The optimal number of update iterations is determined by performing a comparative study on the produced lattices. In a preliminary study, we have conducted experiments by using a variety of iteration numbers for the update procedure. Specifically, we trained the algorithm by using 20, 50, 60, 80, 100 and 150 update iterations. Comparing the produced lattices, we found that the quality of the produced posture prototypes does not change for update iterations number greater than 60. The optimal lattice topology is determined using the crossvalidation procedure, which is a procedure that determines the ability of a learning algorithm to generalize over data that was not trained on. That is, the learning algorithm is trained using all but some training data, which are subsequently used for testing. This procedure is applied multiple times (folds). The test action videos used to determine the optimal lattice topology were all the action videos of a specific person not included in the training set. A 12 × 12 lattice of posture prototypes produced using action videos of action classes ’walk’, ’run’, ’jump in place’, ’jump forward’, ’bend’, ’sit’, ’fall’ and ’wave one hand’ captured from eight viewing angles ’0^0’, ’ 45^0’, ’ 90^0’, ’ 135^0’, ’ 180’, ’ 225^0’, ’ 270^0’ and ’ 315^0’ (with respect to the person’s body) is depicted in Figure 3.

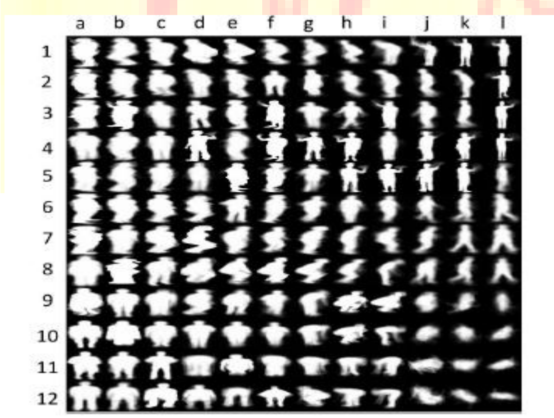

Fig. 3. A 12×12 SOM produced by posture frames of eight actions captured from eight viewing angles.

Figure 4 presents the winning neurons in the training set used to produce the 12× 12 lattice presented in Figure 3 for each of the action classes. In this Figure, only the winning neurons are shown, while the grayscale value of the enclosing square is a function of their wins number. That is, after determining the SOM, the similarity between all the posture vectors belonging to the training action videos and the SOM neurons was computed and the winning neurons corresponding to each posture vector was determined. The grayscale value of the enclosing squares is high for neurons having large number of wins and small for those having a small one. Thus, neurons enclosed in squares with high grayscale value correspond to human body poses appearing more often in each action type. As can be seen, posture prototypes representing each action are quite different and concentrate in neighboring parts of the lattice. Thus, one can expect that the more non-overlapping these maps are the more discriminant representation they offer for action recognition.

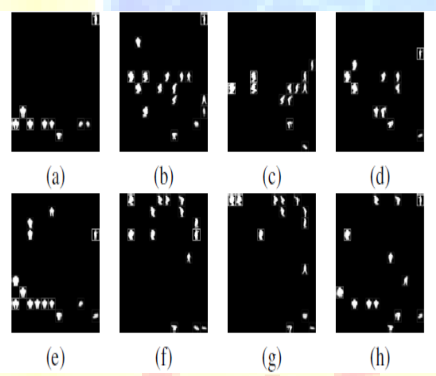

Fig. 5. Wining neurons for eight views: a) ’0^0’, b) ’ 45^0’, c) ’ 90^0’,d) ’ 135^0’, e) ’ 180’, f) 225^0’,g) ’ 270^0’ and h) ’ 315^0’

C. Action Representation

To find the similarity of every posture vector with every posture prototype fuzzy distance is calculated. Let posture vectors p*i; i* = 1*,…. , Ntj, j=1,….,* *NA* consists an action video. Fuzzy distances of every p*i* to all the SOM weights w*sk, k=1… NS* are calculated to determine the similarity of every posture vector with every posture prototype

$d_{ik} = (\|Pi - wsk||_2)^{-\frac{2}{m-1}}$ (4)

where *m* is the fuzzification parameter (*m* > 1). Its optimal value is determined by applying the cross-validation procedure. We have experimentally found that a value of *m* = 1:1 provides satisfactory action representation. Fuzzy distances allow for a smooth distance representation between posture vectors and posture prototypes.

After the calculation of fuzzy distances, each posture vector is mapped to the following distance vector. Distance vectors di = $[d_{i1}, d_{i2}, \ldots . d_{ins}]$are normalized to produce membership vectors. The membership vector that correspond to the final representations of the posture vectors in the SOM posture. The use of the mean vector leads to a duration invariant action representation. That is expect the normalized cumulative membership of a specific action to be invariant to the duration of the action. The mean vector

$S=\frac{1}{Ntj\sum_{i=1}^{Ntj} ui}$, s $\in R^{N_s}$ (5)

Ntj membership vector comprising the action video is called action vector and represents the action video. The use of main mean vector leads to a duration invariant action representation

D. Single-view Action Classification

The neural network is mostly based on Artificial neural network. Artificial network are compose a_j=1,….,N_A of interconnecting artificial neurons. MLP is proposed for the action classification task consisting of *NS* inputs (equal to the dimensionality of action vectors **s**), N_A outputs (each corresponding to an action class) and using the hyperbolic tangent function

$f_{sigmoid}$(x)=αtan $h(bx)$ (6)

where the values α=1.7159 and b=$\frac{2}{3}$. For each of the action vectors **s***i*, MLP response $[\hat{O}_{i1}, \ldots, \hat{O}_i N_A]$ is calculated.

The modification of weight that connects neurons *i* and *j* follows the update rule. Each action video is classified to the action class *aj* , *j* = 1*,…,NA* that corresponds to the MLP maximum output:

$\hat{O}ik= f_{sigmoid} S_i^T w_{Ak}$ (7)

The training procedure is performed in an on-line form, i.e., adjustments of the MLP weights are performed for each training action vector. After the feed of a training action vector **s***i* and the calculation of the MLP response $\hat{O}i$ the modification of weight that connects neurons *i* and *j* follows the update rule:

$$\Delta W_{\mathrm{Aji}}(\mathrm{n+1}) = \mathrm{c}\Delta W_{\mathrm{Aji}}(n) + \Box\delta_{\mathrm{j}}(n) y_{\mathrm{i}}(n) \qquad (8)$$

This procedure is applied until the Mean Square Error(MSE) falls under an acceptable error rate ε

$$E\left[\left(\frac{1}{N_T}\left(\hat{o}_i - O_i\right)2\right)\frac{1}{2}\right] < \varepsilon. \qquad (9)$$

Finally, expecting that most of the recognized actions $\hat{a}_i$ will correspond to the actual action class of *S*, *S* is classified to an action class by performing majority voting over the action classes indicated in $\hat{a}_i$

$$\hat{a}_i = \underset{j}{\mathrm{argmax}}\ \hat{O}_{ij},\ i = 1, \dots, N, j = 1, \dots, N_A.$$

Thus a vector $\hat{a}_i = [\hat{a}_{i1}, \dots, \hat{a}_{NA}]$ containing all the recognized action classes is obtained. Using this approach view independent action is achieved.

E. Combination of single-view action classification results using a Bayesian Framework

The classification of an action vector set *S* consisting of $N \leq N_c$ action vectors i = 1…..N, each corresponding to an action video coming from a specific camera used for recognition,in one of the action classes of the action class set *A*, can be performed using a probabilistic framework. Each of the *N* action vectors **s***i* of *S* is Fed to the MLP and *N* vectors containing the responses are obtained. The problem to be solved is to classify *S* in one of the action classes a_j given these observations i.e., to estimate the probability MLP outputs $\hat{a}_i$ will be set to zero, as no recognition result is provided for these cameras.

Since MLP responses are real valued, estimation is very difficult. Let ˆ*ai* denote the action recognition result corresponding to the test action vector **s***i* representing the action video captured by the *i*-th camera, taking values in the action class set *A*.

Let $P(\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_{NC} | a_j)$ be the joint probabilities of all the NC cameras observing one of the *NA* action classes $\hat{a}_j$. Furthermore, the conditional probabilities $P(\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_{NC} | a_j)$ that camera 1 recognizes action class, camera 2 recognizes action class etc., given that the actual action class of S is $\hat{a}_j$, can be calculated. Using these probabilities, the probability $P(\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_N | a_j)$of action class $\hat{a}_j$; *j* = 1………*NA*, given the classification results $\hat{a}_j$ can be estimated using the Bayes formula:

$$P\left(\frac{a_j}{\hat{a}_1}, \hat{a}_2, \ldots, \hat{a}_N\right) = \frac{P(\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_N | a_j) \cdot P(a_j)}{\sum_{i=1}^{N_A} P(\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_N | a_l) \cdot P(a_i)} \qquad (10)$$

In the case of equiprobable action classes, *P*(*aj*) = 1*NA*. If this is not the case, *P*(*aj*) should be set to their real values andthe training data should be chosen accordingly. Expecting that training and evaluation data come from the same distributions, $P(\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_{NC} | a_j)$can be estimated during the training procedure. The conditional probabilities are estimated in the training phase. These test action vectors are fed to the action recognition MLP and *N* action recognition results $\hat{a}_i$ are obtained. In the case of majority voting, the action vector set *S* is classified to the action class a_j that has the most votes.

In the Bayesian framework case, the *N* action vectors **s***i* are fed to the viewing angle identification MLP to recognize the corresponding viewing angle $\hat{v}_i$ and the action vector set *S* is classified to the action class a_j that provides the highest cumulative probability according to the Bayesian decision.

Figure 6 illustrates the probabilities $P(a_j \backslash \hat{a}_j, v_i)$ *j* = 1……,*NA; i* = 1….*NC*, the probabilities to correctly classify an action video belonging to action class from each of the viewing angles ˆ*vi* for an action class set A ={’walk’, ’run’, ’jump in place’, ’jump forward’, ’bend’, ’sit’, ’fall’, ’wave one hand’} produced using a 12 × 12 SOM and an 8-camera setup, in which each camera captures the person from one of the eight viewing angles ’0^0’, ’ 45^0’, ’ 90^0’, ’ 135^0’, ’ 180’, ’ 225^0’, ’ 270^0’ and ’ 315^0’ In this Figure, it can be seen that action vectors belonging to action classes ’jump in place’, ’bend’, and ’fall’ are almost correctly classified by every viewing angle. On the other hand, action vectors belonging to the remaining actions are more difficult to be correctly classified for some viewing angles. As was expected, the side views are the most capable in terms of classification for action classes ’walk’, ’run’, ’jump forward’ and ‘sit’, while in the case of action class ‘wave one hand’ the best views are the frontal and the back ones.

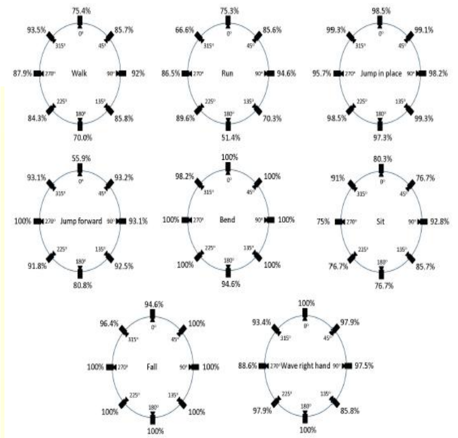

Fig. 6. Single view action classification results presented as input to the Bayesian framework for eight actions captured from eight viewing angles.

F. Action Recognition

Let a person performing an action captured from $N \leq NC$ cameras. In the case of elementary action recognition, this action is captured in N action videos, while in the case of continuous action recognition, a sliding window consisted video frames is used to create the N action videos used to perform action recognition at every window location. These videos are preprocessed as discussed in Section II-A to produce $N \times Nt$ posture vectors $\mathbf{p}i$ $i = 1....,N;$ $j = 1.....,Nt$, where Nt $Nt = Ntw$ in the elementary and the continuous action recognition tasks, respectively. Fuzzy distances from all the test posture vectors to every SOM posture prototype $\mathbf{w}Sk;$ $k = 1....,NS$, are calculated and a set S of N test action vectors, $\mathbf{s}i$, is obtained. These test action vectors are fed to the action recognition MLP and N action recognition results are obtained.

III. EXPERIMENTS

The experiments conducted in order to evaluate the performance of the proposed method are presented in this section. To demonstrate the ability of the proposed method to correctly classify actions performed by different persons, as variations in action execution speed and style may be observed, leave-one-person-out cross-validation procedure was applied in the i3DPost multi-view action recognition database [32] and the experiments are discussed in subsection III-C. Subsection III-D discusses the operation of the proposed method in case of multi-period action videos. Subsection III-E presents its robustness in the case of action recognition at different frame rates between training and test phases.

A. The i3DPost multi-view database

The i3DPost multi-view database [32] contains 80 high resolution image sequences depicting eight persons performing eight actions and two person interactions. Eight cameras having a wide 45^0 viewing angle difference to provide 360^0 coverage of the capture volume were placed on a ring of 8m diameter at a height of 2m above the studio floor. The studio was covered by blue background. The actions performed in 64 video sequences are: ’walk’ (wk), ’run’ (rn), ’jump in place’ (jp), ’jump forward’ (jf), ’bend’ (bd), ’fall’ (fl), ’sit on a chair’ (st) and ’wave one hand’ (wo). The remaining 16 sequences depict two persons that interact. These interactions are: ’shake hand’ (sh) and ’pull down’ (pl).

B. The IXMAS multi-view database

The INRIA (Institut National de Recherche en Informatique et Automatique) Xmas Motion Acquisition Sequences database [22] contains 330 low resolution (291 × 390 pixels) image sequences depicting 10 persons performing 11 actions. Each sequence has been captured by five cameras. The persons freely change position and orientation. The actions performed are: ‘check watch’, ‘cross arm’, ‘scratch head’, ‘sit down’, ‘get up’, ‘turn around’, ‘walk in a circle’, ‘wave hand’, ‘punch’, ‘kick’, and
‘pick up’. Binary images denoting the person’s body are provided by the database.

C. Cross-validation in i3DPost multi-view database

The cross-validation procedure described in Subsection II-B was applied to the i3DPost eight-view database, using the action video sequences of the eight persons. Action videos were manually extracted and binary action videos were obtained by thresholding the blue color in the HSV color obtained for various SOM lattice topologies.

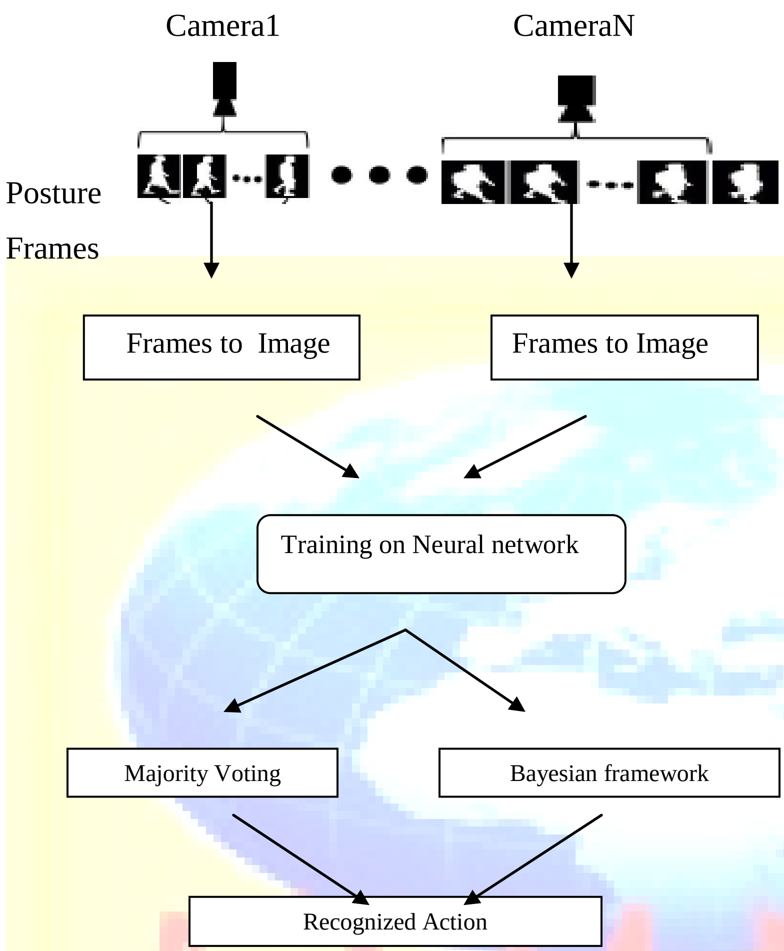

Fig. 7. Action Recognition system overview

majority voting and the Bayesian framework cases. It can be seen that high recognition rates were observed. The optimal topology was found to be a 12×12 lattice. A recognition rate equal to 93:9% was obtained for the majority voting case. The Bayesian approach outperforms the majority voting one, providing a recognition rate equal to 94:04% for the view-independent approach. As can be seen, the use of viewing angle information results to an

increase of the recognition ability. The best recognition rate was found to be equal to 94:87%, for the Bayesian approach incorporating the viewing angle recognition results. The confusion matrix corresponding to the best recognition result is presented in Table I

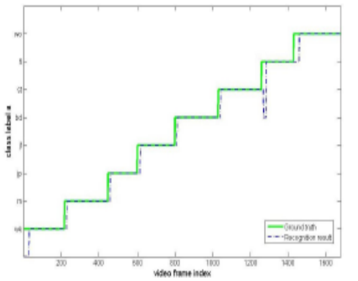

Fig. 8. a) Action recognition rates vs various lattice dimensions of the SOM

D. Continuous action recognition

This section presents the functionality of the proposed method in the case of continuous (multiple period) action recognition. Eight multiple period videos, each corresponding to one viewing angle, depicting one of the persons of the i3DPost eight-view action database were manually created by concatenating single period action videos. The algorithm was trained using the action videos depicting the remaining seven persons using a 12 × 12 lattice and combining the classification results corresponding to each camera with the Bayesian framework. In the test phase, a sliding window of *Ntw* = 21 video frames was used and recognition was performed at every sliding window position. A majority vote filter, of size equal to 11 video frames, was applied at every classification result. Figure 8b illustrates the results of this experiment.

TABLE I

CONFUSION MATRIX FOR EIGHT ACTION

	wk	rn	Jp	jf	Bd	st	Fl	wo
Wk	0.95	0.05						
Rn	0.05	0.95						
Jp			0.92	0.02		0.06		
Jf			0.05	0.9		0.05		
Bd					1			

St			0.13			0.87		
Fl							1	

E. Action recognition in different video frame rates

To simulate the situation of recognizing actions using cameras of different frame rates, between training and test phases, an experiment was set as follows. The cross-validation procedure using a 12×12 lattice and the Bayesian framework was applied for different camera frame rates in the test phase. That is, in the training phase the action videos depicting the training persons were used to train the algorithm using their actual number of frames. In the test phase, the number of frames consisting the action videos were fewer, in order to achieve recognition at lower frame rate.

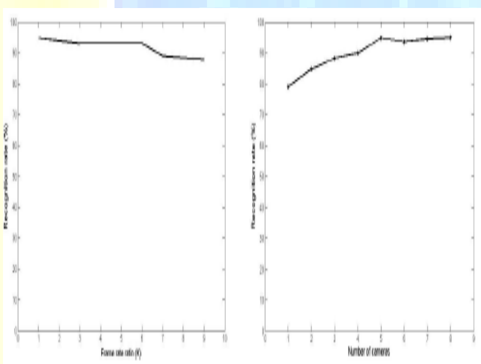

Fig. 9. a) Recognition results for different video frame rates between training and test phase. b) Action recognition rates vs various occlusion levels.

. TABLE II

CONFUSION MATRIX FOR EIGHT ACTIONS

	w	R	Jp	jf	b	H	p	st	f	W

	k	n			d	s	l		l	o
W k	. 9 5	0. 0 5								
r n	. 0 5	0. 9								
J p			0.8 1	0. 0 1		0. 0 2		0. 0 7		

In order to compare our method with other methods using the IXMAS action recognition datase we performed the LOOCV procedure by using the binary images provided in the database. In an off-line procedure, each image sequence was split in smaller segments, in order to produce action videos. Subsequently, the LOOCV procedure has been performed by using different SOM topologies and the Bayesian framework approach. By using a 13×13 SOM an action recognition rate equal to 89.8% has been obtained. As can be seen, the proposed method outperforms these methods providing up to 8.5% improvement on the action classification accuracy.

TABLE III

COMPARISON RESULT IN THE IXMAS MULTI-VIEW ACTION RECOGNITION DATABASE

Method[38]	Method[39]	Method[40]	Proposed method
81.3%	81%	80.6%	89.8%

IV. CONCLUSION

A very powerful framework based on Artificial Neural Networks has been proposed for action recognition. The proposed method highlights the strength of ANN in representing and classifying visual information. SOM human body posture representations is combined with Multilayer Perceptrons. Action and viewing angle classification is achieved independently for all cameras. A Bayesian framework is exploited in order to provide the optimal combination of the action classification results, coming from all available cameras. The effectiveness of the proposed method in challenging problem setups has been demonstrated by experimentation. According to authors knowledge, there is no other method in the literature that can deal with all the presented challenges in action recognition. Furthermore, it has been shown that the same framework can be applied for human interaction recognition between persons, without any modification.

REFERENCES

[1] L. Weilun, H. Jungong, and P. With, “Flexible human behavior analysis framework for video surveillance applications,” International Journal of Digital Multimedia Broadcasting, vol. 2010, Article ID 920121, 9 pages, 2010.

[2] P. Barr, J. Noble, and R. Biddle, “Video game values: Human-computer interaction and games,” Interacting with Computers, vol. 19, no. 2, pp. 180–195, Mar. 2007.

[3] B. Song, E. Tuncel, and A. Chowdhury, “Towards a multi-terminal video compression algorithm by integrating distributed source coding with geometrical constraints,” Journal of Multimedia, vol. 2, no. 3, pp. 9– 16, 2007.

[4] T. Hfllerer, S. Feiner, D. Hallaway, B. Bell, M. Lanzagorta, D. Brown, S. Julier, and Y. B. nad L. Rosenblum, “User interface management techniques for collaborative mobile augmented reality,” Computers and Graphics, vol. 25, no. 5, pp. 799–810, Oct. 2001.

[5] S. Ali and M. Shah, “Human action recognition in videos using kinematic features and multiple instance learning,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 32, no. 2, pp. 288–303, feb. 2010.

[6] J. Hoey and J. Little, “Representation and recognition of complex human motion,” in Proceedings of IEEE Conference on Computer Vision, vol. 1. IEEE, 2000, pp. 752–759.

[7] Y. Wang and G. Mori, "Hidden part models for human action recognition: Probabilistic versus max margin," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 33, no. 7, pp. 1310–1323, july 2011. [8] H. J. Seo and P. Milanfar, "Action recognition from one example," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 33, no. 5, pp. 867–882, May 2011.

[9] M. Giese and T. Poggio, "Neural mechanisms for the recognition of biological movements," Nature Reviews Neuroscience, vol. 4, no. 3, pp. 179–192, Mar. 2003.

[10] N. Gkalelis, A. Tefas, and I. Pitas, "Combining fuzzy vector quantization with linear discriminant analysis for continuous human movement recognition," IEEE Transactions on Circuits Systems Video Technology, vol. 18, no. 11, pp. 1511–1521, Nov. 2008.

[11] L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 2247–2253, 2007.

[12] A. Iosifidis, A. Tefas, and I. Pitas, "Activity based person identification using fuzzy representation and discriminant learning," IEEE Transactions on Information Forensics and Security, no. 99.

[13] M. Ursino, E. Magosso, and C. Cuppini, "Recognition of abstract objects via neural oscillators: interaction among topological organization, associative memory and gamma band synchronization," IEEE Transactions on Neural Networks, vol. 20, no. 2, pp. 316–335, 2009.

[14] M. Schmitt, "On the sample complexity of learning for networks of spiking neurons with nonlinear synaptic interactions," IEEE Transactions on Neural Networks, vol. 15, no. 5, pp. 995–1001, 2004.

[15] B. Ruf and M. Schmitt, "Self-organization of spiking neurons using action potential timing," IEEE Transactions on Neural Networks, vol. 9, no. 3, pp. 575–578, 1998.

[16] C. Schuldt, I. Laptev, and B. Caputo, "Recognizing human actions: A local svm approach," in Proceedings of International Conference on Pattern Recognition, vol. 3. IEEE, pp. 32–36. [17] H. Jhuang, T. Serre, L. Wolf, and T. Poggio, "A biologically inspired system for action recognition," Proceedings of IEEE International Conference on Computer Vision, vol. 1, 2007.

[18] A. Iosifidis, A. Tefas, N. Nikolaidis, and I. Pitas, "Multi-view human movement recognition based on fuzzy distances and linear discriminant analysis," Computer Vision and Image Understanding, 2011.

[19] S. Yu, D. Tan, and T. Tan, “Modeling the effect of view angle variation on appearance-based gait recognition,” in Proceedings Asian Conf.Computer Vision, vol. 1, Jan. 2006, pp. 807–816.

[20] O. Masoud and N. Papanikolopoulos, “A method for human action recognition,” Image and Vision Computing, vol. 21, no. 8, pp. 729–743, Aug. 2003.

BIBILOGRAPHY

Name:Mrs C.Seline Angel,

Participated in International conference at Maria Engineering college, Atoor and National Conference at Bharathiyar Institute of Engineering for Women, Salem, India