A Human Activity Recognition System Using ICA and HMM

MD. Zia Uddin, J. J. Lee, and T.-S. Kim

Abstract In this paper, a novel human activity recognition method is proposed which utilizes independent components of activity shape information from image sequences and Hidden Markov Model (HMM) for recognition. Activities are represented by feature vectors from Independent Component Analysis (ICA) on video images, and based on these features; recognition is achieved by trained HMMs of activities. Our recognition performance has been compared to the conventional method where Principle Component Analysis (PCA) is typically used to derive activity shape features. Our results show that superior recognition is achieved with our proposed method especially for activities (e.g., skipping) that cannot be easily recognized by the conventional method.

Keywords: Hidden Markov Model (HMM), Independent Component Analysis (ICA), Principal Component Analysis (PCA), K-means, LBG.

This work was supported by the MIC (Ministry of Information and Communication), Korea, under the ITRC (Information Technology and Communication Research Center) support program supervised by the IITA (IITA-2006-(C1090-0602-0002)).

* Lee, Jee Jun : MS, Department of Biomedical Engineering, Kyung Hee University e-mail: ljj@khu.ac.kr

** MD. Zia Uddin : MS, Department of Biomedical Engineering, Kyung Hee University e-mail: ziauddin@khu.ac.kr

*** Kim, Tae Sung : Professor, Department of Biomedical Engineering, Kyung Hee University e-mail: tskim@khu.ac.kr

1. Introduction

Recently human activity recognition is becoming an intensive field to study and research due to an interest in proactive computing. A system that can recognize various human activities has many important applications such as automated surveillance systems and smart home healthcare applications [1, 2]. In general, human activities in images can be represented in two categories based on their features: one includes shape features and another motion features such as optical flow. In this paper, we focus on shape features and their improved representation via ICA. Conventionally, most previous recognition works have used PCA to extract shape features of activities [1, 2]. Basically, PCA focuses on global representation of the shape features based on the second-order statistics. Recently another statistical approach called ICA exploiting higher order statistics for face recognition [3, 4]. It is proved that ICA features are very effective to describe local features. Thus our proposed approach is to use ICA to describe local information of activity shapes more effectively than PCA. As for recognition, due to successful applications of HMM in speech recognition [5, 6], recently HMM has been adopted in the human activity research field as well [1, 2, 7]. Once human activities are represented in features via PCA or ICA, HMM can be effectively used for human activity recognition, since it is a most suitable technique for recognizing time sequential feature information. Our results show that superior recognition is achieved with our proposed method, in which IC features are combined with HMM, especially for activities (e.g., skipping) that cannot be easily recognized by the conventional PC-based HMM method.

2. Related Works

In recent years, many recognition methods are proposed for human activity recognition. In [1] and [2], Niu and Abdel-Mottaleb used PC-based shape features to build HMM for recognition. They combined motion features with shape. In [7], Yamato et al. proposed 2D mesh feature and HMM based approach to recognize several tennis activities in time sequential images. In [8], Carlsson and Sullivan proposed another shape based approach to recognize forehand and backhand strokes from tennis video clips. For each person in each image sequence one key-frame per stroke was defined and using that one other frames in the sequence, depicting the same activity posture, were found. In [9], a view independent approach is proposed for classification and identification of human posture. The authors used 2D shapes captured by multiple cameras and shape description using a 3D shape descriptor. The descriptions are used for learning and recognition by SVM. In [10], Nakata proposed a multi-resolutional optical flow based method for recognition. He applied the Burt-Anderson pyramid approach to extract useful features consist of multi-resolutional optical flows. Sun et al. used affine motion parameters and optical flow features to build HMM for recognition in [11].

In our case, we propose a shape-based approach of using ICA in combination with HMM for the first time to recognize human activities.

3. Methodology

Video image preprocessing for shape feature extraction is discussed in Section 3.1. Sections 3.2 and 3.3 represent shape feature extraction via PCA and ICA respectively. ICA features extracted from sequential images are converted to a sequence of symbols, which are corresponding to the codewords of a codebook. In Section 3.4, we present codebook generation using vector quantization where two methods are described. Section 3.5 gives an overview of HMM models that are used for recognition using shape features. In learning HMM, the symbol sequences obtained from the training image sequences of each activity are used to optimize the corresponding HMM. In recognition, the symbol sequence is applied to all the HMMs and one is chosen that gives the maximum likelihood.

3.1 Video Image Preprocessing

A threshold-based simple Gaussian probability distribution is used to subtract a background image from a recent frame to extract a Region of Interest (ROI). Figure 1 (c) shows a ROI of a sample frame and Figure 2 demonstrates a sequence of generalized ROIs from an image sequence of walking.

(a) (b) (c)

Figure 1. (a) Background image (b) Frame from a walking sequence (c) ROI with a rectangle.

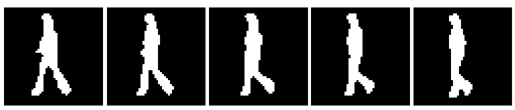

Figure 2. Generalized ROIs from a walking sequence.

To apply PCA or ICA to extract activity features from these ROIs, every normalized ROI is represented as a row vector where the dimension of the vector is equal to the number of pixels in the entire image. Preprocessing steps are necessary before applying PCA or ICA algorithm on the images. The first step is to make all the training vectors to zero mean. Then PCA or ICA algorithm is applied on the zero mean input vectors. Let we have T number of shape images and $X_1, X_2, ..., X_T$ are the sequential shape images. The mean shape image vector $\overline{X}$ is subtracted from each shape vector to make it a zero mean vector $\tilde{X}_i$ according to (1) where $1 \le i \le T$.

$$\tilde{X}_i = (X_i - \overline{X}) \tag{1}$$

3.2 Shape Feature Extraction Using PCA

PCA basis images for human shape images represent global information of activities. Figure 3 shows 10 basis images after PCA is applied on 750 images of 5 activities: namely walking, running, skipping, right hand, and both hand waving.

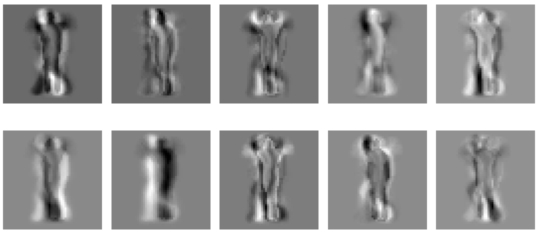

Figure 3. Ten PCs from all activity shape images including walking, running, skipping, right hand, and both hand waving.

The principle component representation of a shape image vector $\tilde{X}_i$ is as follows.

$$P_i = \tilde{X}_i E_m \tag{2}$$

where P_i is the PCA projection of i^{th} image. E_m is the top m eigenvectors corresponding to the first top eigenvalues of descending order.

3.3 Feature Extraction using ICA

The ICA algorithm finds the statistically independent basis images. The basic idea of ICA is to represent a set of random observed variables using basis function where the components are statistically independent. If S is collection of basis images and X is collection of input images then the relation between X and S is modeled as (3).

$$X = MS \tag{3}$$

where M represents an unknown linear mixing matrix of full rank.

The ICA algorithm learns the weight matrix W, which is inverse of mixing matrix M. W is used to recover set of independent basis images S. ICA basis image focuses on the local feature information rather than global information in PCA. ICA basis images show the local status of the movements in activity such as legs are open or closed for running. Figure 4 shows 10 ICA basis images for all activities. Before applying ICA, PCA is used to reduce dimension of total training image data. ICA algorithm is performed on E_m as follows.

$$S = WE_m^T \tag{4}$$

$$E_m^T = W^{-1}S \tag{5}$$

$$X_r = VW^{-1}S \tag{6}$$

where V is projection of images X on E_m and X_r the reconstructed original images.

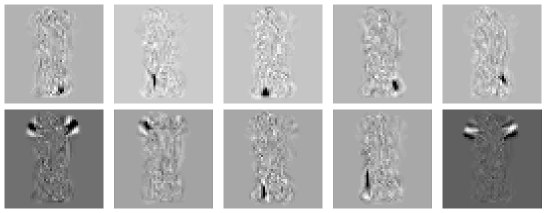

Figure 4. Ten ICs from all activity shape images.

The independent component representation I_i of i^{th} shape vector $\tilde{X}_i$ from an activity image sequence can be expressed as

$$I_i = \tilde{X}_i E_m W^{-1}. \tag{7}$$

3.4 Codebook Generation

We symbolize feature vectors obtained using ICA or PCA before applying to train or recognize by HMM. An efficient codebook of vectors can be generated using vector quantization from training vectors. In our experiment, we have used two vector quantization algorithms: namely ordinary K-means clustering and Linde, Buzo, and Gray (LBG)'s clustering algorithm [12]. In both of them, first initial selection of centroids is obtained. In case of the K-means clustering, until a convergence criterion is met, for every sample it seeks the nearest centroid, assign the sample to the cluster, and compute the center of that cluster again. However, in case of LBG, recomputation is done after assigning all samples to new clusters. In LBG, initialization is done by splitting the centroid of whole dataset. It starts with codeword size of one and recursively splits into two codewords. After splitting, optimization of the centroids is done to reduce the distortion. Since it follows binary splitting methods, the size of the codebook must be power of two. In case of K-means, the overall performance varies due to the selection of the initial random centroids. On the contrary, LBG starts from splitting the centroid of entire dataset, thus there is less variation in its performance.

When a codebook is designed, index numbers of the codewords are used as symbols to apply on HMM. As long as a feature vector is available then index number of the closest codeword from the codebook is the symbol for that replace. Hence every shape image is going to be assigned a symbol. If there are K image sequences of T length then there will be K sequences of T length symbols. The symbols are the observations, O.

3.5 Training and Recognition using HMM

A generic HMM can be expressed as $\lambda = \{\Xi, \pi, A, B\}$ where Ξ denotes possible states, π the initial probability of the states, A the transition probability matrix between hidden states where state transition probability a_{ij} represents the probability of changing state from i to j, and B observation symbols' probability from every state where the probability $b_j(O)$ indicates the probability of observing the symbols O from state j. Before recognition, a distinct HMM is trained for every activity. If the number of activities is N then there will be a dictionary $(\lambda_1, \lambda_2, \ldots \lambda_N)$ of N models. We used the Baum-Welch algorithm for HMM parameter estimation [13]. The parameter estimation of observation sequence O is shown from (8) to (11).

$$\xi_t(i, j) = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^{q} \sum_{j=1}^{q} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)} \quad (8)$$

$$\gamma_t(i) = \sum_{j=1}^{q} \xi_t(i, j) \quad (9)$$

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T} \gamma_t(i)} \quad (10)$$

$$\hat{b}_j(d) = \frac{\sum_{t=1, O_t = d}^{T-1} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)} \quad (11)$$

where $\xi_t(i, j)$ represents the probability of staying in state i at time t and state j at time $t+1$. $\gamma_t(i)$ is the probability of staying in state i at time t. α and β are the forward and backward variables respectively that are calculated from transition and observation matrix. $\hat{a}_{ij}$ is the estimated transition probability from state i to state j and $\hat{b}_j(d)$ the estimated observation probability of symbol d from state j. q is the number of states used in the models.

Four-state left to right model is used for each activity. π is assigned as $\{1,0,0,0\}$. For B, the possible number of observations from every state is the number of vectors in the codebook. In each model, we applied uniform observation and transition probability from the states before training. The total transition and observation probability from any state is one. Figure 5 shows the structure and transition probabilities of an ICA-based running HMM before and after training with the codebook size of 32.

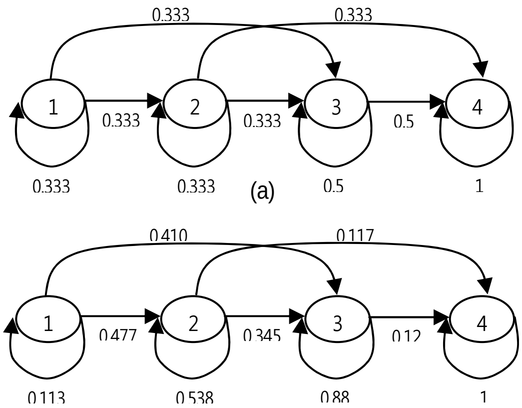

Figure 5. Running HMM structure and transition probabilities (a) before and (b) after training

To test observation sequence O, we have to find the appropriate HMM with the highest likelihood.

4. Experimental Result

Five activities have been recognized using our proposed approach: namely walk, run, skip, right hand wave, and both hand wave. Every sequence consists of 10 images. A total of 15 sequences from every activity are used to build feature space and for training corresponding HMM later on. So, the whole database consists of total 750 images. After applying ICA and PCA, 150 features are taken in feature space as it is known that the more number of features results in the better performance. We have tested its performance using four different sizes of codebook: namely 8, 16, 32 and 64. Two types of codebook generation schemes are used: K-means and LBG. A total 180 sequences are used for testing the models. Table 1 and Table 2 show the recognition results using ordinary K-means and LBG respectively

Table 1. Recognition results using K-means codebook

Code - book Size	Activities	Recognition rate (PCA)	Recognition rate (ICA)
8	Walk	89.09 %	76.36 %
	Run	62.5	65
	Skip	38.09	1
	RHW	80	84
	BHW	93.33	95.56
16	Walk	96.36	96.36
	Run	95	95
	Skip	42.85	42.85

	RHW	92	88
	BHW	88.89	95.56
32	Walk	100	100
	Run	100	100
	Skip	66.67	66.67
	RHW	92	96
	BHW	93.33	97.78
64	Walk	94.54	98.18
	Run	100	100
	Skip	57.14	57.14
	RHW	92	92
	BHW	97.48	95.56

*RHW=Right Hand Wave
**BHW=Both Hand Wave

Table 2. Recognition results using LBG codebook

Code - book Size	Activities	Recognition rate (PCA)	Recognition rate (ICA)
8	Walk	94.54%	94.54%
	Run	67.5	87.5
	Skip	57.14	66.67
	RHW	84	84
	BHW	84.44	84.44
16	Walk	96.36	96.36
	Run	82.5	85
	Skip	47.61	66.67
	RHW	92	96
	BHW	77.78	77.78
32	Walk	100	100
	Run	100	100
	Skip	66.67	85.71
	RHW	92	96
	BHW	93.33	97.78
64	Walk	98.18	100
	Run	100	100
	Skip	66.67	85.71
	RHW	88	92
	BHW	93.33	95.56

5. Conclusion

We have presented an ICA and HMM-based approach for human activity recognition. The preliminary results show improved performance than other shape based approaches. For more robust human activity recognition, we plan to include motion information with proposed shape features.

6. Acknowledgement

This work was supported by the MIC (Ministry of Information and Communication), Korea, under the ITRC (Information Technology and Communication Research Center) support program supervised by the IITA (IITA-2006-(C1090-0602-0002)).

Reference

[1] Niu, Feng, Abdel-Mottaleb, Mohamed: View-Invariant Human Activity Recognition Based on Shape and Motion Features. In: Proceedings of the IEEE Sixth International Symposium on Multimedia Software Engineering, pp. 546-556 (2004)

[2] Niu, Feng, Abdel-Mottaleb, Mohamed: HMM-Based Segmentation and Recognition of Human Activities from Video Sequences. In: Proceedings of IEEE International Conference on Multimedia & Expo, pp. 804-807 (2005)

[3] Bartlett, M., Movellan, J., Sejnowski, T.: Face recognition by independent component analysis. In: IEEE Transactions on Neural Networks, vol. 13, pp. 1450-1464 (2002)

[4] Yang, J., Zhang, D., Yang, J. Y.: Is ICA Significantly Better than PCA for Face Recognition?. In: Proceedings of IEEE International Conference on Computer Vision, pp. 198-203 (2005)

[5] Lawrence, R., Rabiner, A.:Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. In: Proceedings of the IEEE, 77(2), pp. 257-286 (1989)

[6] Bregler, C., König, Y.: Eigenlips for Robust Speech Recognition. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal, Processing, pp. 669-672 (1994)

[7] Yamato, J., Ohya, J., Ishii, K.: Recognizing Human Action in Time-Sequential Images using Hidden Markov Model. In: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, pp. 379-385 (1992)

[8] Carlsson, S., Sullivan, J.: Action Recognition by Shape Matching to Key Frames. In: IEEE Computer Society Workshop on Models versus Exemplars in Computer Vision, pp. 263-270 (2002)

[9] Cohen, I., Li, H.: Inference of Human Postures by Classification of 3D Human Body Shape. In: IEEE International Workshop on Analysis and Modeling of Faces and Gestures, pp. 74-81 (2003)

[10] Nakata, Toru: Recognizing Human Activities in Video by Multi-resolutional Optical Flow. In: International Conference on Intelligent Robots and Systems, pp. 1793-1798 (2006)

[11] Sun, X., Chen, C., Manjunath, B. S.: Probabilistic Motion Parameter Models for Human Activity Recognition. In: 16th International Conference on Pattern recognition, pp. 443-450 (2002)

[12] Linde, Y., Buzo, A., Gray, R.: An Algorithm for Vector Quantizer Design. In: IEEE Transaction on Communications, vol. 28(1), pp. 84–94 (1980)

[13] IWAI, Yoshio, HATA, Tadashi, YACHIDA, Masahiko: Gesture Recognition based on Subspace Method and Hidden Markov Model. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 960-966 (1997)