Appl Intell (2010) 33: 193–206
DOI 10.1007/s10489-008-0159-2

Independent shape component-based human activity recognition via Hidden Markov Model

M. Zia Uddin · J.J. Lee · T.-S. Kim

Published online: 21 January 2009

Abstract In proactive computing, human activity recognition from image sequences is an active research area. In this paper, a novel human activity recognition method is proposed, which utilizes Independent Component Analysis (ICA) for activity shape information extraction from image sequences and Hidden Markov Model (HMM) for recognition. Various human activities are represented by shape feature vectors from the sequence of activity shape images via ICA. Based on these features, each HMM is trained and activity recognition is achieved by the trained HMMs of different activities. Our recognition performance has been compared to the conventional method where Principal Component Analysis (PCA) is typically used to derive activity shape features. Our results show that superior recognition is achieved with the proposed method especially for activities (e.g., skipping) that cannot be easily recognized by the conventional method. Furthermore, by employing Linear Discriminant Analysis (LDA) on IC features, the recognition results further improved significantly in the recognition performance.

Keywords PCA · ICA · LDA · K-means · LBG · HMM

M. Zia Uddin · J.J. Lee · T.-S. Kim (✉)
Department of Biomedical Engineering, Kyung Hee University,
Seocheon-dong, Giheung-gu, Yongin-si, Gyeonggi-do, 446-701,
Republic of Korea
e-mail: tskim@khu.ac.kr

M. Zia Uddin
e-mail: ziauddin@khu.ac.kr

J.J. Lee
e-mail: ljj@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. Proactive computing is a technology that proactively anticipates peoples' health-related needs and takes whatever action that is appropriate on their behalf. A system that can recognize various human activities has many important applications such as automated surveillance systems and smart home healthcare applications [1–3]. The common method for activity recognition is to extract some feature information including shape and motion following by the comparison with activity database. Thus, efficient feature extraction, learning, and classification of activity play a key role for human activity recognition. In general, human activity recognition is a challenging task as it has absence of rigid syntax like gesture or sign language recognition.

Many recognition methods have been proposed for human activity recognition. D. Gavrila presented a detailed survey on human action recognition in [4]. In general, human activities in images can be represented in two categories based on their features: one includes shape features [1, 2, 5–7] and another motion features such as optical flow, affine motion parameter and so on [1–3, 8–11]. Typically, binary shapes are commonly employed to represent different human activities [1, 2, 5]. In [1] and [2], PC-based shape features were used for recognition. In [5], 2D mesh features of binary shapes extracted from video frames were applied to recognize several tennis activities in time sequential images. In [6], the authors used a view independent approach in order to infer human postures utilizing 2D shapes captured by multiple cameras and 3D shape descriptions. In [7], Carlsson and Sullivan proposed shape matching key frame-based approach to recognize forehand and backhand strokes from tennis video clips where each shape was represented in the form of edge data from the Canny-edge detector. In [12–16], some shape deformation works are discussed. In addition, a lot of works using motion information as well, have been reported in the human activity recognition area. In [3], Robertson et al. presented a recognition method where actions were described by strong features based on trajectory information (position and velocity) and a set of local motion descriptors. In [8], Nakata applied the Burt-Anderson pyramid to extract useful features consist of multi-resolutional optical flows to recognize human activities. In [9], Sun et al. utilized affine motion parameters and optical flow features for distinguished human activity recognition. In [10], the authors applied Infinite Impulse Response (IIR) filter to measure motion features followed by PCA. In [11], Ben-Arie et al. used multidimensional indexing to recognize different actions that were represented by velocity vectors of major body parts. In some other works, the authors applied shape and motion features together to generate stronger features [1, 2]. In [1] and [2], in order to make the PC shape features more robust, the features were extended by augmenting motion features i.e. optical flow features to recognize view invariant human activities.

As surveyed, the most common shape feature extraction technique applied in video-based human activity recognition is PCA [1, 2]. The PCA approach, which is also known as the eigenface method, is a very common and popular unsupervised statistical approach to find global feature representation of the input. Usually, it shows optimality in the case of dimension reduction of the input feature space and pattern classification techniques are applied on that lower dimensional space for recognition. In the case of human activity recognition, PCA utilizes its second order statistical nature or characteristics to focus on the global representations of the shape images. There are a lot of extensions are available over the standard PCA [17–19]. However, PCA features in general do not lead to desirable recognition performance due to its extraction of global features. Lately, Independent Component Analysis (ICA), has been actively exploited especially in the face recognition area that utilizes the higher order statistics to find statistically independent local basis images for the face images. It generalizes PCA and has shown superior performance over PCA [20–22]. In addition to the face recognition area, it has been applied successfully in various fields such as speech recognition [23], bioinformatics [24], electroencephalogram (EEG) [25], and biomedical signal and image analysis [26]. However, ICA has not been applied to the human activity recognition field to the best of our knowledge.

As for the modeling and recognition of human activities, HMM has been utilized successfully in many papers such as [1–3, 5, 8, 9]. In [1] and [2], several HMMs were trained utilizing the shape and motion features. In [3], the authors used HMM to encode the scene rules and higher level human activity recognition. In [5], HMM was employed as recognition engine in combination with the shape features. In [8] and [9], the authors attempted to build HMMs utilizing motion features. In [6], Support Vector Machine (SVM) was implemented for view independent activity recognition.

In this work, we focus on the shape features and their improved representation via ICA. We propose the IC-based shape feature approach in combination with HMM for the first time in our best knowledge to recognize human activities. We employed ICA to focus on the local status of the activity shape features of different activities rather than global features like PCA. Moreover, we have further applied LDA to classify the IC features for better shape feature representation. For modeling and recognition of human activities, we applied discrete HMM where a sequence of discrete symbols obtained from a codebook is used. The recognition results using the IC shape feature-based approaches show significantly improved performance over the conventional PC-based approaches such as PCA and LDA on the PC features. Moreover, our proposed human activity recognition technique (i.e., LDA on the IC features) achieves superior recognition rate over other approaches. Especially, for skipping, it shows significant improvement recognition rate from 67.50% to 90%.

The overall structure of our paper is as follows. We begin with the methodology in Sect. 2 where the architecture of the proposed system is elaborated from video image preprocessing to activity data and experimental settings through feature extraction techniques and activity modeling, training, and recognition by HMM. We proceed to Sect. 3 where recognition results utilizing different shape feature-based approaches are presented and discussed. At last, we draw a conclusion of our work in Sect. 4.

2 Methodology

Our recognition system consists of video image preprocessing, feature extraction (i.e., PCA, ICA, LDA on the PC features, and LDA on the IC features), codebook generation, and recognition via HMM. Figure 1 shows the overall architecture of our proposed activity recognition system where H represents HMMs and L the likelihoods of HMMs.

2.1 Video image preprocessing

From every frame of each video clip, a Region of Interest (ROI) containing the binary activity shape is extracted. We assume that every video clip consists of single human activity. According to the algorithm given in [27], we extracted the ROI after the gray scale background subtraction. In some activities such as right and both hand waving, only the hands

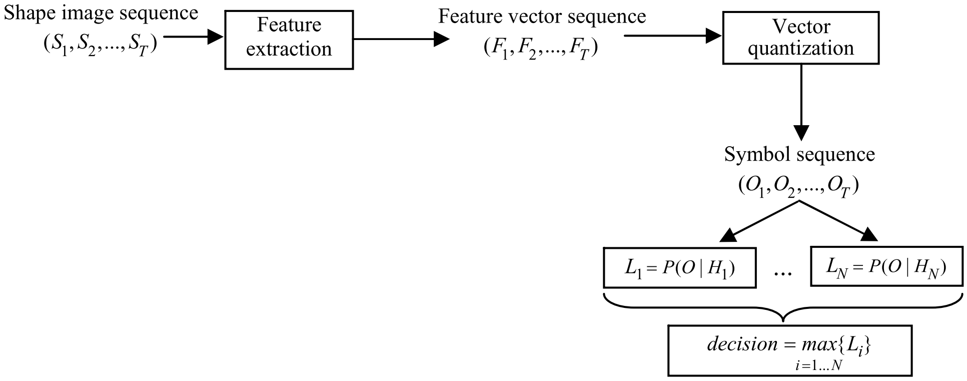


Fig. 1 Our proposed human activity recognition system

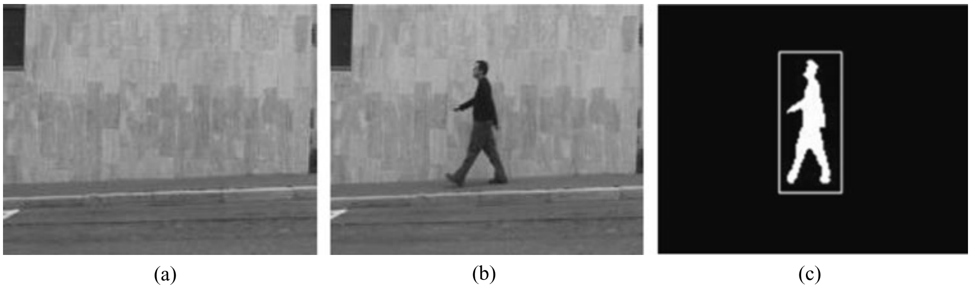


Fig. 2 (**a**) Background image, (**b**) Frame from a walking sequence, and (**c**) ROI within a rectangle

are the major moving components of human body except the whole body. Hence, instead of consecutive frames, we used static background images to subtract the background from the frames of each activity video clip. The probability of being a pixel to be background is given as

$$\mathrm{P}(R(x,y)) = \frac{1}{C}\sum_{i=1}^{C}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-(R(x,y)-B_i(x,y))^2}{2\sigma^2}\right) \quad (1)$$

where C is the number of background images. $\mathrm{P}(R(x,y))$ is the probability of being a pixel of the background in x and y position of the recent frame. $B_i(x,y)$ and $R(x,y)$ are the intensity values of the pixel in x and y position of the ith background image and recent frame respectively. To extract the ROI, the recent frame is converted to binary according to (2) using a threshold Th. A value of Th is experimentally determined on the basis of result of (1). Figure 2(c) shows the ROI of a sample frame and Fig. 3 a sequence of the generalized ROIs from the image sequences of walking and running. The general size of each ROI that we used for experiments in this work was 50 × 50.

$$BI(x,y) = \begin{cases} 1, & \mathrm{P}(R(x,y)) \le Th, \\ 0, & \mathrm{P}(R(x,y)) > Th. \end{cases} \quad (2)$$

To apply the activity feature extraction algorithms on these ROIs, every ROI is represented as a row vector where the dimension of the vector is equal to the number of pixels in the entire image. For instance, a ROI image of 50 × 50 size becomes a row vector of 2500 dimensions i.e., the size of the vector is 1 × 2500. Preprocessing steps are necessary before applying feature extraction algorithms on the images. The first step is to convert all the shape vectors to zero mean. Then feature extraction algorithms are applied on the zero mean input vectors. Having $(X_1, X_2, \ldots, X_T)$, T number of shape images and vectors, the mean shape vector $\bar{X} = (\frac{1}{T}\sum_{i=1}^{T} X_i)$ is calculated and subtracted from each

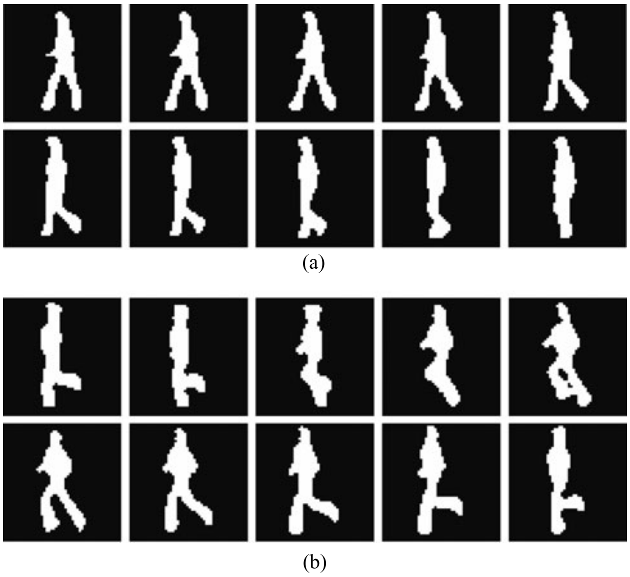


Fig. 3 Generalized ROIs from (**a**) a walking and (**b**) a running image sequence

shape vector to make it a zero mean vector $\tilde{X}_i = (X_i - \bar{X})$ where $1 \leq i \leq T$.

2.2 Feature extraction using PCA

The role of PCA is to approximate the original data with lower dimensional features. Its fundamental is to compute the eigenvectors of the covariance data matrix Q and then the approximation is done using a linear combination of a few top eigenvectors. The covariance matrix of the sample training image vectors and the principal components of the covariance matrix can be calculated respectively as

$$Q = \frac{1}{T}\sum_{i=1}^{T}(\tilde{X}_i \tilde{X}_i^T), \tag{3}$$

$$E^T Q E = \Lambda \tag{4}$$

where E represents the matrix of orthonormal eigenvectors and Λ the diagonal matrix of the eigenvalues. The size of the matrix E becomes $t \times m$ where t is the dimension of each shape image vector and m the number of principal components to be considered. Λ is a $m \times m$ diagonal matrix. Besides, E reflects the original coordinate system onto the eigenvectors where the eigenvector corresponding to the largest eigenvalue indicates the axis of largest variance and the next largest one is the orthogonal axis of the largest one indicating the second largest variance and so on. Usually, the eigenvalues that are close to zero carry negligible variance and hence can be excluded. So, the m eigenvectors corresponding some large eigenvalues can be used to define the subspace. Figure 4 shows the top 50 eigenvalues corresponding to the first 50 eigenvectors where a total of 750 shape image vectors are considered for PCA.

Since PCA is a second order statistics-based analysis to represent the global information such as average faces or eigenfaces in the case of face recognition, after applying PCA on human shapes of different activities, it produces the global features representing frequently moving parts of human body in activities. Figure 5 depicts 10 basis images after PCA is applied on the images of 5 activities: namely walking, running, skipping, right hand waving, and both hand waving. The basis images are the resized 2D image forms of the eigenvectors.

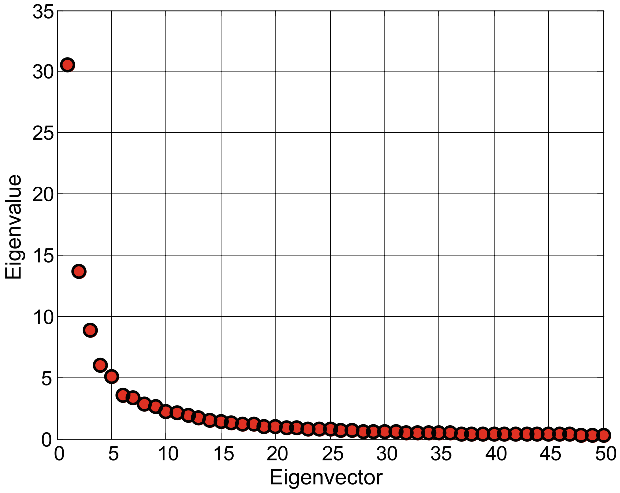


Fig. 4 Top 50 eigenvalues corresponding to the eigenvectors

Thus, the principal component representation of a shape image vector can be represented as follows.

$$P_i = \tilde{X}_i E_m \tag{5}$$

where P_i is the PCA projection of ith image. E_m is the leading m eigenvectors of Q corresponding to the top m eigenvalues. Suppose, there are k number of shape image vectors in the database where each vector is t dimensional and after applying PCA the top m eigenvectors are chosen, now, the size of the PCA representation of each image vector becomes $1 \times m$. For experiments, we considered 150 principal components after applying PCA over the training database of 750 images of five different activities and as a result, m becomes 150 and the size of the E_m is 2500×150 where each column vector represents a principal component. Hence, projecting each shape image vector with the size of 1×2500 onto the PCA feature space, it can be reduced to 1×150.

2.3 Feature extraction using ICA

ICA is recently introduced to solve a blind source separation problem where the objective is to decompose mixture of observed signals into a linear combination of some unknown independent signals and their mixing matrix [28]. The basic idea is to represent a set of random observed variables using some basis functions where the components are statistically independent. The ICA algorithm finds the statistically independent basis images. If S and X are the collection of the basis and input images respectively then the relation between X and S is modeled as

$$X = MS \tag{6}$$

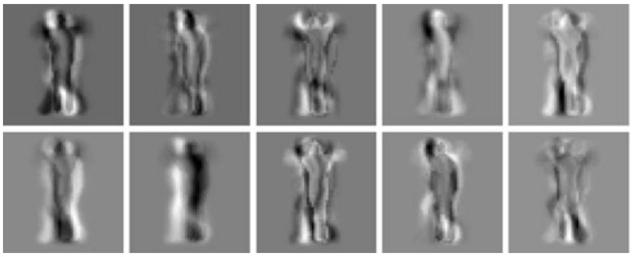


Fig. 5 Ten PCs from all activity shape images including walking, running, skipping, right hand, and both hand waving

where M represents an unknown linear mixing matrix of full rank. It is assumed that the sources are independent of each other and the mixing matrix is invertible. However, the ICA algorithm tries to find the mixing matrix M or the separating matrix so that

$$U = WX, \tag{7}$$

$$U = WMS \tag{8}$$

where U is an estimation of the independent sources [28]. The estimation problem of finding M can be helped by prewhitening of the observed vectors X. Therefore, X is linearly transformed into its prewhitened form Y such that

$$Y = RX, \tag{9}$$

$$R = \Lambda^{-\frac{1}{2}} E \tag{10}$$

where the correlation of Y is a unit matrix i.e., $E\{YY^T\} = I$. Λ substitutes the diagonal matrix of the eigenvalues and E orthonormal eigenvectors of the covariance matrix of X. After transforming the sample vectors X to Y, the ICA algorithm is applied on Y. Thus, the relation between Y and S becomes as

$$Y = BS \tag{11}$$

where B is the estimation of the unknown mixing matrix.

In brief, The ICA algorithm learns the weight matrix W, the inverse of mixing matrix M, is used to recover the set of independent basis images S. The shape images are denoted as variables and the pixel values of associated shapes are the observations of the variable. Before applying ICA, PCA is typically used to reduce the dimension of the total training image data. Thus, after applying the ICA algorithm on PCA subspace i.e., the top t dimensional m eigenvectors, the size of the weighting matrix W and the independent vector matrix S become $m \times m$ and $m \times t$ respectively. In general, the ICA basis images focus on the local feature information unlike the global information in the PC basis images. Figure 6 shows ten ICA basis images for all the activities where high contrast parts represent local human body components such as legs and hands used frequently in all activities. The ICA algorithm is performed on E_m^T and thus m independent basis images in the rows of S are produced:

$$S = WE_m^T, \tag{12}$$

$$E_m^T = W^{-1}S, \tag{13}$$

$$X_r = VW^{-1}S \tag{14}$$

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

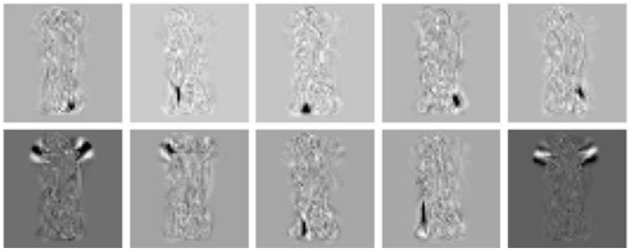


Fig. 6 Ten ICs from all activity shape images

Therefore, the independent component representation I_i of the ith shape vector $\tilde{X}_i$ from an activity image sequence can be expressed as

$$I_i = \tilde{X}_i E_m W^{-1}. \tag{15}$$

Since ICA is applied on the PC-features in this work, hence the size of the ICA representations of the shape image vectors are same as PCA. As we chose the top 150 principal components to apply ICA, therefore, the size of the IC feature of each shape vector and the weighting matrix W are 1×150 and 150×150 respectively.

2.4 Linear discriminant analysis

LDA, a second order statistical approach, which is also known as the fisherface method, is a supervised classification approach that utilizes the class specific information maximizing the ratio of the within and between class scatter information. It looks for the vectors in the underlying space to create the best discrimination among different classes. It is well known for feature extraction and dimension reduction [29]. In order to obtain maximum discrimination, it projects data onto the lower dimensional space so that the ratio of the between and within class distance can be maximized. The within, S_W, and between, S_B, class scattering comparison is done by the following equations.

$$S_B = \sum_{i=1}^{C} J_i(\bar{m}_i - \bar{\bar{m}})(\bar{m}_i - \bar{\bar{m}})^T, \tag{16}$$

$$S_W = \sum_{i=1}^{C} \sum_{m_k \in C_i} (m_k - \bar{m}_i)(m_k - \bar{m}_i)^T \tag{17}$$

where J_i is the number of vectors in the ith class C_i. c is the number of classes and in our case, it represents the number of activities. $\bar{\bar{m}}$ represents the mean of all vectors, $\bar{m}_i$ the mean of the class C_i and m_k the vector of a specific class.

The optimal discrimination matrix is chosen by maximizing the ratio of the determinant of the between and within class scatter matrix as

$$D_{\text{opt}} = \arg\max_D \frac{|D^T S_B D|}{|D^T S_W D|} = [d_1, d_2, \ldots, d_t]^T \tag{18}$$

where D_{opt} is the set of discriminant vectors of S_W and S_B corresponding to the $(c-1)$ largest generalized eigenvalues λ and can be obtained via solving (19). The size of D_{opt} is $t \times r$ where $t \le r$ and r is the number of elements in a vector.

$$S_B d_i = \lambda_i S_W d_i, \quad i = 1, 2, \dots, (c-1) \tag{19}$$

where the rank of S_B is $(c-1)$ or less and hence the upper bound value of t is $(c-1)$ [30]. For instance, the LDA algorithm is applied on a database of 750 shape image vectors of five different classes where each vector is 150 dimensional. As a result, the size of the LDA subspace will be 4 × 150.

2.5 LDA on PC or IC features

LDA produces an optimal linear discriminant function which maps the input into the classification space on which the class identification of the samples is decided. Thus, to acquire a better feature space than PCA or ICA, the LDA algorithm can be applied on the extracted PC or IC features of binary shapes of different activities to classify. The feature vectors using LDA on the PC and IC features can be represented according to (20) and (21) respectively where Z_i and F_i indicate the LDA over PCA and ICA representation respectively for the ith shape image. Figure 7 shows the 3D representation of all the shape images after applying on 3 ICs that are chosen on the basis of the top kurtosis values. In this plot, using those ICs, the features do not seem to be well separable. Figure 8 demonstrates the 3D plot of the features after LDA on the ICA representations of the shape images of five classes where 150 ICs are taken. In the figure, the prototypes of the right hand and both hand waving are linearly separable. For walking, skipping, and running, the prototypes are close to each other due to the similarity of the activities, thus there are some overlaps in the prototypes of these classes. However, considering all the dimensions above three, the overlapped prototypes should be separable. Later on, in the experimental results section, we can see that the LDA on the IC features approach results a superior recognition rate over other approaches.

$$Z_i = P_i D_{\text{opt}}^T, \tag{20}$$

$$F_i = I_i D_{\text{opt}}^T. \tag{21}$$

Now, it is clear that through the LDA algorithm we can linearly classify the prototypes of different classes with the help of a low dimensional subspace. Thus, if we project 150 dimensional PCA or ICA representations of 750 shape image vectors of five different classes on the LDA subspace of them, we can end up with a much reduced dimensional LDA representation of each vector i.e., 1 × 4. Hence, the size of the training database containing 750 vectors becomes 750 × 4 from 750 × 2500.

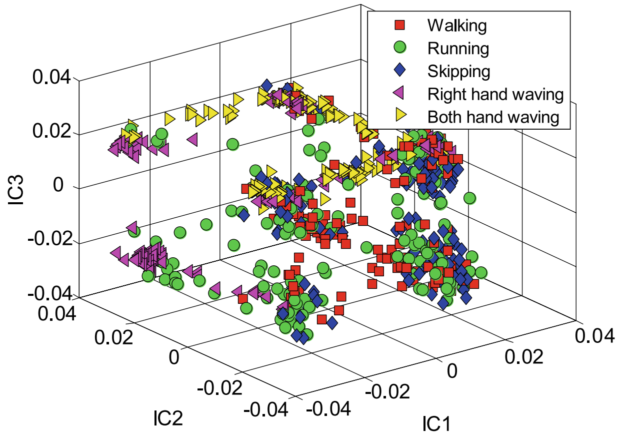


Fig. 7 3D plot of the IC features of 750 shapes of five activities

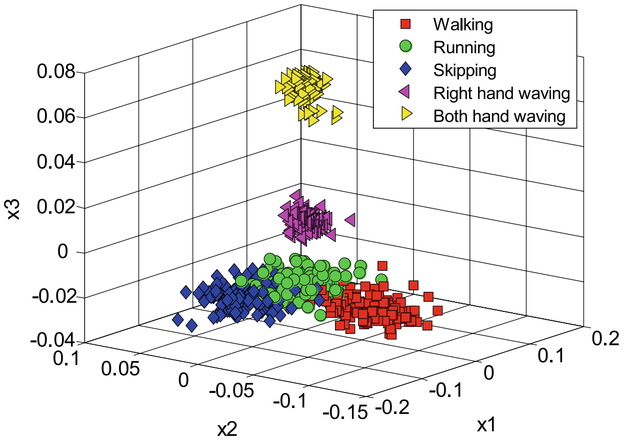


Fig. 8 3D plot of the features after LDA on the IC features of 750 shapes of five activities

2.6 Codebook generation

We need to symbolize the feature vectors before applying to train or recognize by HMMs. An efficient codebook of vectors should be generated using vector quantization from the training vectors. In our experiment, we have used two vector quantization algorithms: namely ordinary K-means clustering [31] and Linde, Buzo, and Gray (LBG)'s clustering algorithm [32]. In both of them, at first, the initial selection of the centroids is obtained. In the case of the K-means clustering, until a convergence criterion is met, it seeks the nearest centroid for every sample, assign the sample to the cluster, and compute the center of that cluster again. However, in the 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 the codebook size of one and recursively splits each codevector into two new codevectors. After splitting, optimization of the centroids is done to reduce the distortion. Since it follows the binary splitting technique [32], the size of the

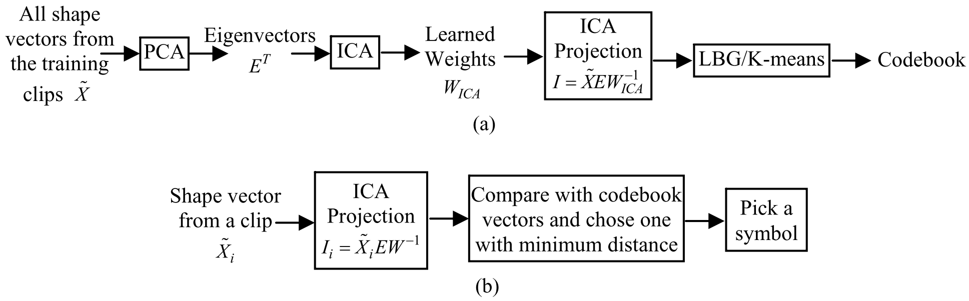

All shape vectors from the training clips $\tilde{X}$
PCA
Eigenvectors E^T
ICA
Learned Weights W_{ICA}
ICA Projection $I = \tilde{X}EW_{ICA}^{-1}$
LBG/K-means
Codebook
(a)
Shape vector from a clip $\tilde{X}_i$
ICA Projection $I_i = \tilde{X}_i EW^{-1}$
Compare with codebook vectors and chose one with minimum distance
Pick a symbol
(b)

Fig. 9 (**a**) Codebook generation and (**b**) Symbol selection using ICA

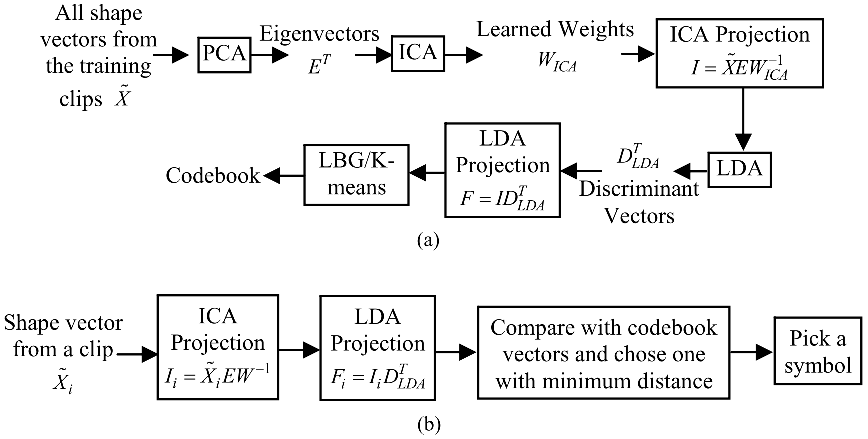

All shape vectors from the training clips $\tilde{X}$
PCA
Eigenvectors E^T
ICA
Learned Weights W_{ICA}
ICA Projection $I = \tilde{X}EW_{ICA}^{-1}$
LDA
D_{LDA}^T Discriminant Vectors
LDA Projection $F = ID_{LDA}^T$
LBG/K-means
Codebook
(a)
Shape vector from a clip $\tilde{X}_i$
ICA Projection $I_i = \tilde{X}_i EW^{-1}$
LDA Projection $F_i = I_i D_{LDA}^T$
Compare with codebook vectors and chose one with minimum distance
Pick a symbol
(b)

Fig. 10 (**a**) Codebook generation and (**b**) Symbol selection using LDA on the IC features

codebook must be power of two. In this work, we define the codebook size as the number of vectors in the codebook. In the 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 and there is less variation in its performance than K-means.

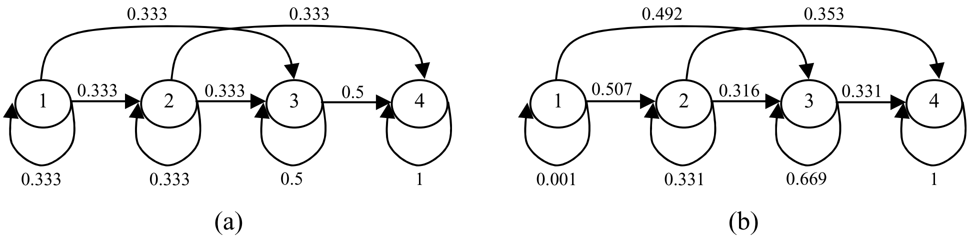
When a codebook is designed, the index numbers of the codevectors are used as symbols to apply on HMMs. As long as a feature vector is available, the index number of the closest codevector of the codebook from the feature is the symbol for that replace. Hence, every shape image is going to be assigned 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 as denoted as O. Figures 9 and 10 show the codebook generation and symbol selection from the codebook utilizing the IC features and the features after LDA on the IC features respectively.

2.7 Activity modeling and training using HMM

To learn and recognize of human activities, we have employed HMM that can deal with sequential image data with probabilistic learning capability for recognition. HMM is a stochastic process where an underlying stochastic process is usually not observable but it can be observed through another set of stochastic processes that produces observation symbols. The basic theory of HMM was developed by Baum et. al. [33, 34] and it has been applied extensively to solve a large number of problems such as speech recognition [35, 36] and handwritten character recognition [37]. Recently, it has been adopted to gesture recognition [38, 39] and human activity recognition as well [1–3, 8, 9]. Once human activities are represented as features then HMM can be applied effectively for human activity recognition as it is a suitable technique for recognizing time sequential feature information. Shape features are converted to a sequence of symbols that are corresponding to the codevectors of the codebook obtained by vector quantization. In learning HMM, the symbol sequences obtained from the training image sequences of distinct activity are used to optimize the corresponding HMM. Each activity is represented by a distinct HMM. In recognition, the symbol sequence is applied to all the HMMs and one is chosen that gives the highest likelihood.

An HMM is a collection of finite states connected by transitions. Every state of an HMM can be described by two types of probabilities: namely transition probability and symbol observation probability. A generic HMM can be expressed as $H = \{\Xi, \pi, A, B\}$ 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 a changing state from i to j and B observation symbols' probability from every state where the probability $b_j(d)$ indicates the probability of observing the symbol d from state j. If the number of activities is N then there will be a dictio-

Fig. 11 Walking HMM (**a**) before and (**b**) after training



nary $(H_1, H_2, \dots, H_N)$ of N trained models. We used the Baum-Welch algorithm [35] for HMM parameter estimation according to (22) to (30):

$$\alpha_1(i) = \pi_i b_i(O_1), \quad 1 \le i \le q, \tag{22}$$

$$\alpha_{t+1}(j) = \sum_{i=1}^{q} \alpha_t(i) a_{ij} b_j(O_{t+1}), \quad t = 1, 2, \dots, (T-1), \tag{23}$$

$$\beta_T(i) = 1, \quad 1 \le i \le q, \tag{24}$$

$$\beta_t(i) = \sum_{j=1}^{q} a_{ij} b_j(O_{t+1}) \beta_{t+1}(i), \quad t = (T-1), (T-2), \dots, 1, \tag{25}$$

$$\xi_t(i,j) = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^{q} \sum_{j=1}^{q} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}, \tag{26}$$

$$\hat{\pi}_i = \gamma_1(i), \quad 1 \le i \le q, \tag{27}$$

$$\gamma_t(i) = \sum_{j=1}^{q} \xi_t(i,j), \tag{28}$$

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}, \tag{29}$$

$$\hat{b}_j(d) = \frac{\sum_{t=1, O_t=d}^{T-1} \gamma_t(j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \tag{30}$$

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

We observed that large number of states in HMMs could not improve the recognition rate significantly on our initial activity data sets. Therefore, four-state left to right model is chosen for each activity. Since we start training from the first state, hence 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. Figure 11 shows the structure and transition probabilities of a walking HMM after training with the codebook size of 32. In order to test the sequence O, we find the appropriate HMM with the highest likelihood. The following equation represents the likelihood of an observation sequence O at time t using a forward variable α where a model H is given:

$$\mathrm{P}(O|H) = \sum_{i=1}^{q} \alpha_t(i). \tag{31}$$

2.8 Activity data and experimental settings

To assess the proposed approach, the activity video clips regarding our experiments were collected from the human activity databases used in [40] and [41]. In [40], the authors performed their approach to recognize ten activities: namely walking, running, skipping, jacking, jumping-forward-on-two-legs, jumping-in-place-on-two-legs, galloping sideways, waving-two-hands, waving one-hand, and bending. In [41], a local SVM approach was introduced to recognize six human activities: namely walking, jogging, running, boxing, waving-two-hands, and hand clapping. Thus, from the databases mentioned above, we collected the video clips related to five human activities (i.e., walking, running, skipping, right hand waving, and both hand waving) that we tried to recognize in this work. Furthermore, the collected video clips were split into several clips where each one contained 10 consecutive frames. After background subtraction from every frame, we calculated the binary ROI with the size of 50 × 50 pixels. A total of 15 sequences from every activity were used to build the feature space. So, the whole database consisted of a total of 750 images and after applying ICA and PCA on that, 150 features were taken into the feature space. In order to train and test each activity model, we applied 15 and 40 image sequences respectively.

As we applied discrete HMM because of its simplicity and efficiency to train and recognize the time-sequential features, hence a good codebook is supposed to be designed to obtain the discrete observation sequence for each training or testing video clip. So, it is necessary to study the effective codebook generation algorithms and the codebook sizes. As a result, we tested the K-means and LBG algorithms on both the IC and PC shape features with the different codebook sizes (i.e., 8, 16, 32, and 64) along with HMM in order to determine the optimal one.

Based on the study of the optimal codebook size and algorithm as mentioned above, the performance of LDA on the PC and IC features were studied since LDA is a good classifier that tries to find out the best linear discrimination among the prototype vectors of different classes. Hence, stronger features were obtained by applying the LDA classification algorithm on the PC and IC features. In this regard, in conjunction with the vector quantization and HMM, we studied four different feature extraction approaches: namely PCA, LDA on the PC features, ICA, and LDA on the IC features.

3 Experimental results and discussion

Basically, PCA is a common method for binary shape feature extraction. However, PCA, the global feature extraction as well as dimension reduction technique, is not sensitive enough to classify the shapes of different classes effectively. On the other hand, ICA, a higher order statistical approach, has already shown its superiority over the PC-based approaches. In general, vector quantization is a necessary step to quantize the feature vectors obtained from the activity video frames to generate a good codebook to obtain the observation symbols for HMM-based recognition. Hence, to determine the better codebook generation algorithm and the optimal codebook size, the K-means and LBG algorithms with different codebook sizes were applied. Tables 1 and 2 show the recognition results of the PC and IC-based approaches respectively where the codebook is designed by the K-means clustering algorithm. The recognition results of both the approaches using the LBG codebook is reflected in Tables 3 and 4. The graphical representation of the experimental performance is also shown in Fig. 12 where the IC feature-based approach along with the LBG codebook of different sizes indicates the superiority over PCA. Thus, in this study, the IC features are better than the PC features to represent the human activity shape features for five activities. Besides, the best recognition rate was obtained using the LBG algorithm with the codebook size of 32. Usually, video sensor-based human activity consists of time sequential information. Consequently, HMM, a powerful and effective probabilistic approach, was used to model and recognize different human activities from the time sequential video frames. Thus, complex human activities can be recognized effectively using ICA in combination with LBG and HMM.

As LDA is a strong tool to find out the underlying feature space to classify the feature vectors linearly, a better feature space can be created through classification of the shape features by LDA. Thus, we continued to focus on the classification of the PC and IC shape features by LDA to create better features. Hence, the LDA-based experiments over the

Table 1 Recognition result of the PC-based approach using the K-means codebook

Codebook size	Activity	Recognition rate (%)	Mean	Standard deviation
8	Walking	87.50	72.50	21.43
	Running	62.50		
	Skipping	40		
	RHW	80		
	BHW	92.50		
16	Walking	95	82.50	22.57
	Running	95		
	Skipping	42.50		
	RHW	92.50		
	BHW	87.50		
32	Walking	100	90.50	13.39
	Running	100		
	Skipping	67.50		
	RHW	92.50		
	BHW	92.50		
64	Walking	95	88.50	17.55
	Running	100		
	Skipping	57.50		
	RHW	92.50		
	BHW	97.50		

*RHW = Right Hand Waving

**BHW = Both Hand Waving

Table 2 Recognition result of the IC-based approach using the K-means codebook

Codebook size	Activity	Recognition rate (%)	Mean	Standard deviation
8	Walking	77.50	74	18.42
	Running	65		
	Skipping	47.50		
	RHW	85		
	BHW	95		
16	Walking	97.50	83.50	23.23
	Running	95		
	Skipping	42.50		
	RHW	87.50		
	BHW	95		
32	Walking	100	92	13.85
	Running	100		
	Skipping	67.50		
	RHW	95		
	BHW	97.50		
64	Walking	97.50	88.50	17.55
	Running	100		
	Skipping	57.50		
	RHW	92.50		
	BHW	95		

Table 3 Recognition result of the PC-based approach using the LBG codebook

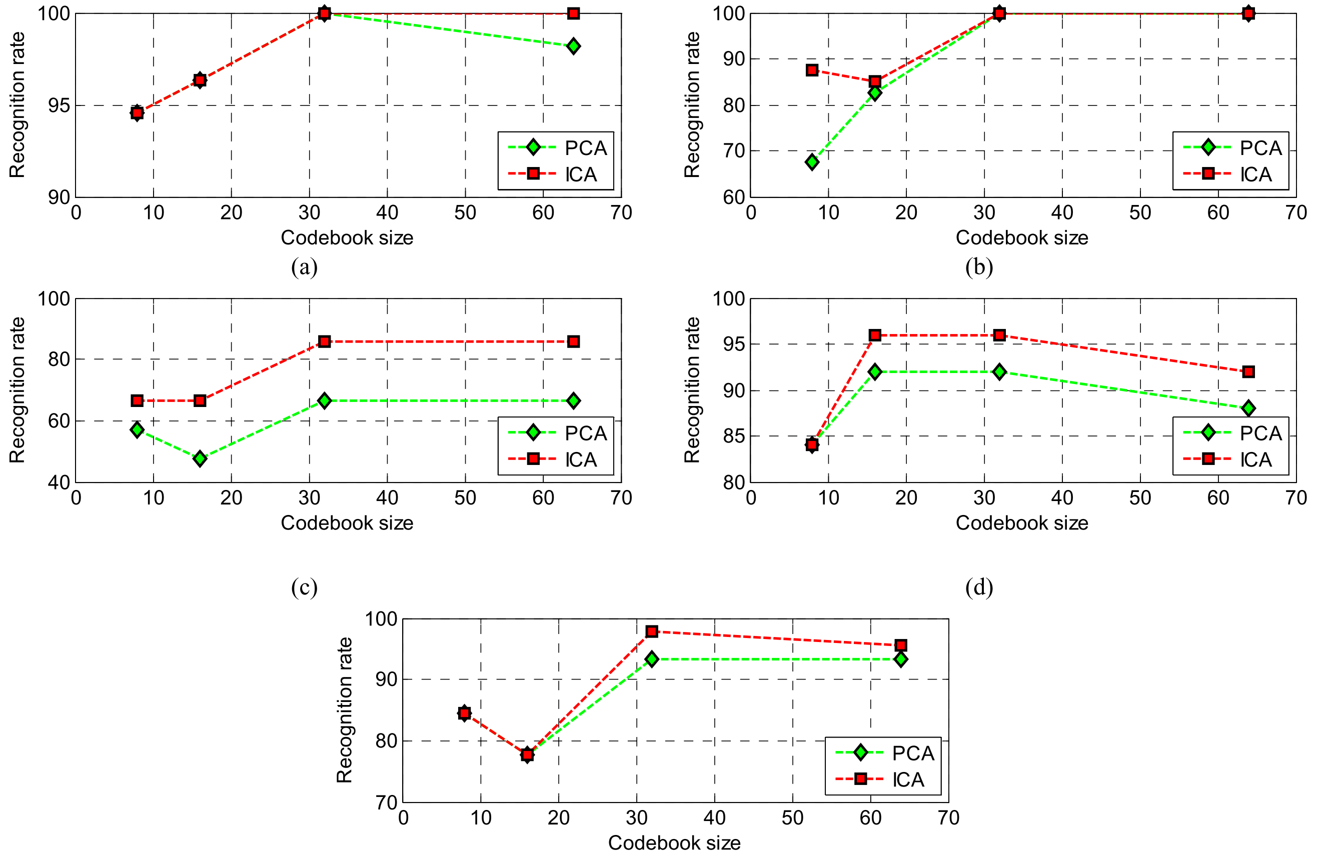
Codebook size	Activity	Recognition rate (%)	Mean	Standard deviation
8	Walking	95	78	15.15
	Running	67.50		
	Skipping	57.50		
	RHW	85		
	BHW	85		
16	Walking	95	79	19.01
	Running	82.50		
	Skipping	47.50		
	RHW	92.50		
	BHW	77.50		
32	Walking	100	90.50	13.39
	Running	100		
	Skipping	67.50		
	RHW	92.50		
	BHW	92.50		
64	Walking	97.50	89	12.94
	Running	100		
	Skipping	67.50		
	RHW	87.50		
	BHW	92.50		

Table 4 Recognition result of the IC-based approach using the LBG codebook

Codebook size	Activity	Recognition rate (%)	Mean	Standard deviation
8	Walking	95	84	10.09
	Running	87.50		
	Skipping	67.50		
	RHW	85		
	BHW	85		
16	Walking	97.50	84.50	12.42
	Running	85		
	Skipping	67.50		
	RHW	95		
	BHW	77.50		
32	Walking	100	96	6.28
	Running	100		
	Skipping	85		
	RHW	97.50		
	BHW	97.50		
64	Walking	100	94.50	6.22
	Running	100		
	Skipping	85		
	RHW	92.50		
	BHW	95		

Table 5 Recognition result of different approaches using the LBG codebook

Features	Activity	Recognition rate (%)	Mean	Standard deviation
PCA	Walking	100	90.50	13.39
	Running	100		
	Skipping	67.50		
	RHW	92.50		
	BHW	92.50		
LDA on the PC features	Walking	92.50	90	10.31
	Running	100		
	Skipping	72.50		
	RHW	92.50		
	BHW	92.50		
ICA	Walking	100	96	6.28
	Running	100		
	Skipping	85		
	RHW	97.50		
	BHW	97.50		
LDA on the IC features	Walking	100	97.50	4.33
	Running	100		
	Skipping	90		
	RHW	100		
	BHW	97.50		



(a) (b) (c) (d) (e)

Fig. 12 Recognition performance of (**a**) walking, (**b**) running, (**c**) skipping, (**d**) right hand waving, and (**e**) both hand waving of PCA and ICA using the LBG codebook

IC and PC-based representations of the shape features were accomplished along with the LBG codebook with the size of 32 and HMM. Table 5 lists the recognition results using different approaches: namely PCA, LDA on the PC features, ICA, and LDA on the IC features. The recognition results of the experiments reflect the highest recognition rate using LDA on the IC features. Our experimental results indicate the superiority of the IC feature-based approaches over the PC-based ones in this study.

In particular, using the proposed approach, it is possible to recognize the human activities with the help of binary shape-based features without segmenting the human body. However, this approach has limitations to describe some shapes in different activities. For instance, due to its binary representation, some body components (e.g., arms) are commonly hidden in binary shapes of different activities (e.g., clapping or boxing in the direction of the video sensor), which may cause ambiguities by assigning same binary shape to different activities. Hence, 3D representation of human body in different activities is an important issue in this regard and we are currently investigating this issue.

4 Conclusion

In this paper, we have presented a novel IC-based HMM approach for human activity recognition. Local shape features are focused here through ICA and our results show improved performance than other conventional approaches such as PCA for shape-based activity recognition. In our case, after several experiments utilizing different sizes and schemes of codebook on the human activity data of five activities, the optimal results were obtained where the codebook size was greater than 30. Besides, considering the outcomes of the previous experiments, to build stronger feature space, we have further classified the IC-based shape features using LDA and presented superior performance over other approaches. For more robust human activity recognition, we plan to include motion flow or 3D depth information with the proposed shape features.

Acknowledgements This work was supported by the MKE (Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the IITA (Institute of Information Technology Advancement) (IITA 2008-(C1090-0801-0002)).

References

1. Niu F, Abdel-Mottaleb M (2004) 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
2. Niu F, Abdel-Mottaleb M (2005) HMM-based segmentation and recognition of human activities from video sequences. In: Proceedings of IEEE international conference on multimedia & expo, pp 804–807
3. Robertson N, Reid I (2006) A general method for human activity recognition in video. Comput Vis Image Underst 104(2):232–248
4. Gavrila D (1999) The visual analysis of human movement: a survey. Comput Vis Image Underst 73:82–98
5. Yamato J, Ohya J, Ishii K (1992) 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
6. Cohen I, Lim H (2003) 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
7. Carlsson S, Sullivan J (2002) Action recognition by shape matching to key frames. In: IEEE computer society workshop on models versus exemplars in computer vision, pp 263–270
8. Nakata T (2006) Recognizing human activities in video by multiresolutional optical flow. In: Proceedings of international conference on intelligent robots and systems, pp 1793–1798
9. Sun X, Chen C, Manjunath BS (2002) Probabilistic motion parameter models for human activity recognition. In: Proceedings of 16th international conference on pattern recognition, pp 443–450
10. Masound O, Papanikolopoulos N (2003) Recognizing human activities. In: IEEE conference on advanced video and signal based surveillance, Miami, Florida, pp 157–162
11. Ben-Arie ZW, Pandit P, Rajaram S (2002) Human activity recognition using multidimensional indexing. IEEE Trans Pattern Anal Mach Intell 24(8):1091–1104
12. Belongie S, Malik J (2000) Matching with shape contexts. In: IEEE workshop on content-based access of image and video libraries
13. Bookstein FL (1978) The measurement of biological shape and shape change. Lecture notes in biomathematics
14. Bremermann HJ (1971) Cybernetic functionals and fuzzy sets. In: IEEE systems, man and cybernetics group annual symposium, pp 248–254
15. Carlsson S (1999) Order structure, correspondence and shape based categories. Shape contour and grouping in computer vision. LNCS, vol 1681. Springer, Berlin, pp 58–71
16. Sclaroff S (1996) Deformable prototypes for encoding shape categories in image databases. Pattern Recogn 30(4):627–640
17. Pujol JV, Lumbreras F, Villanueva JJ, (2001) Topological principal component analysis for face encoding and recognition. Pattern Recogn Lett 22:769–776
18. Kim HC, Kim D, Bang SY (2002) Face recognition using the mixture-of-eigenfaces method. Pattern Recogn Lett 23:1549–1558
19. Gottumukkal R, Asari VK (2004) An improved face recognition technique based on modular PCA approach. Pattern Recogn Lett 24:429–436
20. Bartlett MS, Lades HM, Sejnowski TJ (1998) Independent component representations recognition. In: SPIE symposium on electronic imaging: science and technology, human vision and electronic imaging III, San Jose, CA
21. Bartlett MS, Movellan JR, Sejnowski TJ (2002) Face recognition by independent component analysis. IEEE Trans Neural Netw 13:1450–1464
22. Liu C, Wechsler H (1999) Comparative assessment of independent component analysis (ICA) for face recognition. In: International conference on audio and video based biometric person authentication, Washington, DC
23. Kwon W, Lee TW (2004) Phoneme recognition using ICA-based feature extraction and transformation. Signal Process 84(6):1005–1019
24. Lee SI, Batzoglou S (2003) Application of independent component analysis to microarrays. Genome Biol 4(11):R76.1–21
25. Makeig S, Bell AJ, Jung TP, Sejnowski TJ (1996) Independent component analysis of electroencephalographic data. Adv Neural Inf Process Syst 8:145–151
26. Jung T, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ (2001) Analysis and visualization of single-trial event-related potentials. Hum Brain Mapp 14:166–185
27. Elgammal DH, Davis L (2000) Non-parametric model for background subtraction. In: 6th European conference on computer vision, Dublin, Ireland
28. Cardoso J-F (1997) Infomax and maximum likelihood for source separation. IEEE Lett Signal Process 4:112–114
29. Belhumeur PN, Hespanha JP, Kriegman DJ (1997) Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans Pattern Anal Mach Intell 19(7):711–720
30. Kwak K-C, Pedrycz W (2007) Face recognition using an enhanced independent component analysis approach. IEEE Trans Neural Netw 18(2):530–541
31. Kanungu T, Mount DM, Netanyahu N, Piatko C, Silverman R, Wu AY (2000) The analysis of a simple k-means clustering algorithm. In: Proceedings of 16th ACM symposium on computational geometry, pp 101–109
32. Linde Y, Buzo A, Gray R (1980) An algorithm for vector quantizer design. IEEE Trans Commun 28(1):84–94
33. Baum E, Petrie T, Soules G, Weiss N (1970) A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. Ann Math Stat 41:164–171
34. Baum E, Eagon J (1967) An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology. Am Math Soc Bull 73:360–363
35. Lawrence R, Rabiner A (1989) Tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE 77(2):257–286
36. Bregler C, König Y (1994) Eigenlips for robust speech recognition. In: Proceedings of the IEEE international conference on acoustics, speech, and signal processing, Adelaide, Australia, pp 669–672
37. Hu J, Brown MK, Turin W (1996) HMM based on-line handwriting recognition. IEEE Trans Pattern Anal Mach Intell 18(10):1039–1045
38. Eickeler S, Kosmala A, Rigoll G (1998) Hidden Markov model based online gesture recognition. In: Proceedings of international conference on pattern recognition (ICPR), pp 1755–1757
39. Iwai Y, Hata T, Yachida M (1997) 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
40. Gorelick L, Blank M, Shechtman E, Irani M, Basri R (2007) Actions as space-time shapes. IEEE Trans Patt Anal Mach Intell 29(12):2247–2253
41. Schuldt C, Laptev I, Caputo B (2004) Recognizing human actions: a local SVM approach. In: Proceedings of the 17th international conference on pattern recognition (ICPR'04), pp 32–36

M. Zia Uddin was born in Chittagong, Bangladesh in 1981. He received his B.Sc. degree in Computer Science and Engineering from International Islamic University Chittagong, Bangladesh. He is currently working toward his M.S. leading to Ph.D. degree in the Department of Biomedical Engineering at Kyung Hee University, Republic of Korea. His research interest includes neural network, pattern recognition, image processing, computer vision, and machine learning.

J.J. Lee received the B.S. degrees in Biomedical Engineering and Electrical Engineering from Kyung Hee University (KHU) in 2007 and the M.S. degree in Biomedical Engineering from KHU in 2009. During his studies at KHU, he worked as a Research Assistant in the Bioimaging Laboratory in the Dept. of Biomedical Engineering and the u-Lifecare Research Center at KHU. He is now with LG Electronics as a Research Engineer.

His research interests include biosignal and image processing, multisensor signal analysis, and pattern classification and recognition for advanced bio-imaging and machine learning applications.

T.-S. Kim received the B.S. degree in Biomedical Engineering from the University of Southern California (USC) in 1991, M.S. degrees in Biomedical and Electrical Engineering from USC in 1993 and 1998 respectively, and Ph.D. in Biomedical Engineering from USC in 1999. After his postdoctoral work in cognitive sciences at the University of California, Irvine in 2000, he joined the Alfred E. Mann Institute for Biomedical Engineering and Dept. of Biomedical Engineering at USC as a Research Scientist and Research Assistant Professor. In 2004, he moved to Kyung Hee University in Korea where he is currently an Associate Professor in the Biomedical Engineering Department.

His research interests have spanned various areas of biomedical imaging including Magnetic Resonance Imaging (MRI), functional MRI, E/MEG imaging, DT-MRI, transmission ultrasonic CT, and Magnetic Resonance Electrical Impedance Imaging. Lately he has started research work in proactive computing at the u-Lifecare Research Center where he serves as Vice Director.

Dr. Kim has been developing advanced signal and image processing methods, pattern classification and machine learning methods, and novel medical imaging and rehabilitation instruments and technologies. Dr. Kim has published more than 50 peer reviewed papers and 100 proceedings, and holds 3 international patents. He is a member of IEEE, KOSOMBE, and Tau Beta Pi, and listed in Who's Who in the World '09.