11

Continuous Hidden Markov Models for Depth Map-Based Human Activity Recognition

Md. Zia Uddin and Tae-Seong Kim
Department of Biomedical Engineering
Kyung Hee University
Republic of Korea

1. Introduction

There is an enormous volume of literature on the applications of Hidden Markov Models (HMMs) to a broad range of pattern recognition tasks. The first practical application of HMMs is much based on the work of Rabiner et al (Lawrence & Rabiner, 1989) for speech recognition. Since then, HMMs have been extensively used in various scientific fields such as computational biology, biomedical signal interpretation, image classification and segmentation, etc.

An HMM can be described as a stochastic finite-state automation that can be used to model time sequential data. In general, there are four basic parts involved in the HMM: namely states, initial state distribution, state transition matrix, and state observation matrix. A state represents a property or condition that an HMM might have at a particular time. Initial state distribution indicates each state probability of an HMM at the time of starting the modeling procedure of an event. The state transition matrix represents the probabilities among the states. The observation matrix contains the observation probabilities from each state. Once the architecture of an HMM is defined with the four essential components, training of the HMM is required. To train, the first step is to classify features into a specific number of clusters, generating a codebook. Then from the codebook, symbol sequences are generated through vector quantization. These symbol sequences later are used to model spatiotemporal patterns in an HMM. The number of states and initial state distribution of HMM are empirically determined in general. The state transition and observation probabilities from each state are usually initialized with uniform distributions and later adapted according to the training symbol sequences. In practice, there are some well-established training algorithms available to automatically optimize the parameters of the HMM. The Baum–Welch (Baum et al., 1970) training procedure is a standard algorithm which uses the Maximum Likelihood Estimation (MLE) criterion. In this training algorithm, the training symbol sequences are used to estimate the HMM parameters. Finally, a testing sequence gets analyzed by the trained HMMs to be recognized.

In an HMM, the underlying processes are usually not observable, but they can be observed through another set of stochastic processes that produces continuous or discrete observations (Lawrence & Rabiner, 1989), which lead to discrete or continuous HMMs respectively. In the discrete HMMs, the observation sequences are vector-quantized using a codebook to select discrete symbols. Though the discrete symbols for the observations

simplify the modeling, they have limited representation power. As the discrete symbols are obtained from the codebook, which is generated using some unsupervised classification algorithm such as the K-means (Kanungu et al., 2000) or the Linde, Buzo, and Gray (LBG)'s clustering algorithm (Linde et al., 1980), the quantized vectors may not be accurate: the quantized vectors represent incorrect symbols sometimes. In fact, modeling of an event is very much dependent on the codebook generation process in the discrete HMM that may result in unsatisfactory results. To mitigate this problem, continuous probability distribution functions for the observations can be used to model the time-sequential information that enables to get a model as accurate as possible. In the continuous HMMs, the observation probability distribution functions are a mixture multivariate of Gaussians. The architecture of a continuous HMM, the number of Gaussian components per state, and the number of training iterations are usually empirically determined. Once the event is modeled by the continuous HMM, one can calculate the probabilities of the observation sequence and the probable underlying state sequences. The principal advantage of using the continuous HMM is the ability to model the event directly without involving vector quantization. However, the continuous HMM requires much longer training and recognition time, especially when a mixture of several Gaussian probability density components is used.

Among a wide range of application areas in which HMMs have been applied to recognize complex time-sequential events, human activity recognition (HAR) is one active area that utilizes spatio-temporal information from video to recognize various human activities. In this video-based HAR methodology, it is essential to model and recognize key features from time sequential activity images in which various activities are represented in time-sequential spatial silhouettes. Once the silhouettes from the activity video images are obtained, each activity is recognized by comparing with the trained activity features. Thus, feature extraction, learning, and recognition play vital roles in this regard. In the video-based HAR, binary silhouettes are most commonly employed where useful features are derived from activity videos to represent different human activities (Yamato et al., 1992; Cohen & Lim, 2003; Niu & Abdel-mottaleb, 2004; Niu & Abdel-mottaleb, 2005; Agarwal & Triggs, 2006; Uddin et al., 2008a). To extract the human activity silhouette features, the most popular feature extraction technique applied in the video-based HAR is Principal Component Analysis (PCA) (Niu & Abdel-Mottaleb, 2004; Niu & Abdel-Mottaleb, 2005). PCA is an unsupervised second order statistical approach to find useful basis for data representation. It finds PCs at the optimally reduced dimension of the input. For human activity recognition, it focuses on the global information of the binary silhouettes, which has been actively applied. However, PCA is only limited to second order statistical analysis, allowing up to decorrelation of data. Lately, a higher order statistical method called Independent Component Analysis (ICA) is being actively exploited in the face recognition area (Bartlett et al., 2002; Kwak & Pedrycz, 2007; Yang et al., 2005) and has shown superior performance over PCA. It has also been utilized successfully in other fields such as speech recognition (Kwon & Lee, 2004). In (Uddin et al., 2008a), we introduced local binary silhouette features through ICA to represent human body in different activities usefully. To extend the IC features, we applied Linear Discriminant Analysis (LDA) on them to build more robust features and applied for improved HAR. To model the time-sequential human activity features, HMMs have been used effectively in many works (Yamato et al., 1992; Sun et al., 2002; Nakata, 2006; Niu & Abdel-Mottaleb, 2004; Niu & Abdel-Mottaleb, 2005; Uddin et al., 2008a; Uddin et al., 2008b). In (Yamato et al., 1992), the binary silhouettes were employed to develop some distinct discrete HMMs for different activities. In (Uddin et al., 2008a) and

(Uddin et al., 2008b), we applied the discrete HMM to train and recognize different human activities from binary and depth silhouettes features respectively. Continuous HMMs have also been applied in numerous HAR works (Sun et al., 2002; Niu & Abdel-mottaleb, 2004; Niu & Abdel-mottaleb, 2005; Nakata, 2006). In (Sun et al., 2002) and (Nakata, 2006), the authors utilized optical flows to build continuous HMMs for recognition. In (Niu & Abdel-Mottaleb, 2004) and (Niu & Abdel-Mottaleb, 2005), the authors applied binary silhouette and optical flow motion features in combination with continuous HMM to recognize different human activities.

Although the binary silhouettes are very commonly employed to represent a wide variety of body configurations, it sometimes produces ambiguities by representing the same silhouette for different postures from different activities. For instance, if a person performs some hand movement activities in the direction toward the camera, different postures can correspond to the same silhouette due to its binary-level (i.e., white or black) pixel intensity distribution. One example is shown in Fig. 1, which shows the RGB, binary, and depth images of right hand-up-down, left hand-up-down, both hands-up-down, clapping, and boxing activity respectively. It is obvious that the binary silhouettes are a poor choice to separate these different postures. Besides, from the binary silhouettes, it is not possible to obtain the difference between the far and near parts of human body in the activity video. For better silhouette representation than binary, in (Uddin et al., 2008b), we proposed IC features from the time-sequential depth silhouettes to be used with the discrete HMMs for robust HAR. The depth silhouettes better represent the human body postures than the binary by differentiating the body parts by means of different depth values. Thus, depth silhouettes can be utilized to overcome the aforementioned limitations available in binary silhouettes.

In this chapter, with a brief introduction of HMMs, especially the continuous HMM, we present its application to model various human activities based on the spatio-temporal depth silhouette features. Then, we show how to recognize various human activities from time-series depth maps (i.e., depth videos) of human activities. We demonstrate that superior recognition can be achieved by means of the continuous HMM and depth silhouette features in recognizing various human activities, which are not easily discernible with binary silhouettes. One of the aims of our HAR system is to be used in smart homes to monitor and recognize important daily human activities. This should allow continuous daily, monthly, and yearly analysis of human activity patterns, habits, and needs. In addition, when any abnormal activity is recognized, the system can automatically generate an alarm to draw attention of the smart home inhabitants to draw their attention to avoid unexpected injury and to provide assistance if needed.

The remaining sections of this chapter are structured as follows. Section 2 describes the basics of continuous HMMs. Section 3 explains the methodology of the HAR system from video image acquisition and depth silhouette feature extraction to modeling and training activity continuous HMMs for recognition. Section 4 shows the experimental results utilizing different feature extraction approaches with HMMs. Finally, Section 5 draws the concluding remarks.

2. Continuous Hidden Markov Model

HMM is considered to be one of the most suitable techniques for modeling and recognizing time sequential features (Lawrence & Rabiner, 1989; Kwon & Lee, 2004). Basically, a continuous HMM consists of two interrelated processes. The first process is related to an underlying, unobservable Markov chain with a finite number of states, a state transition

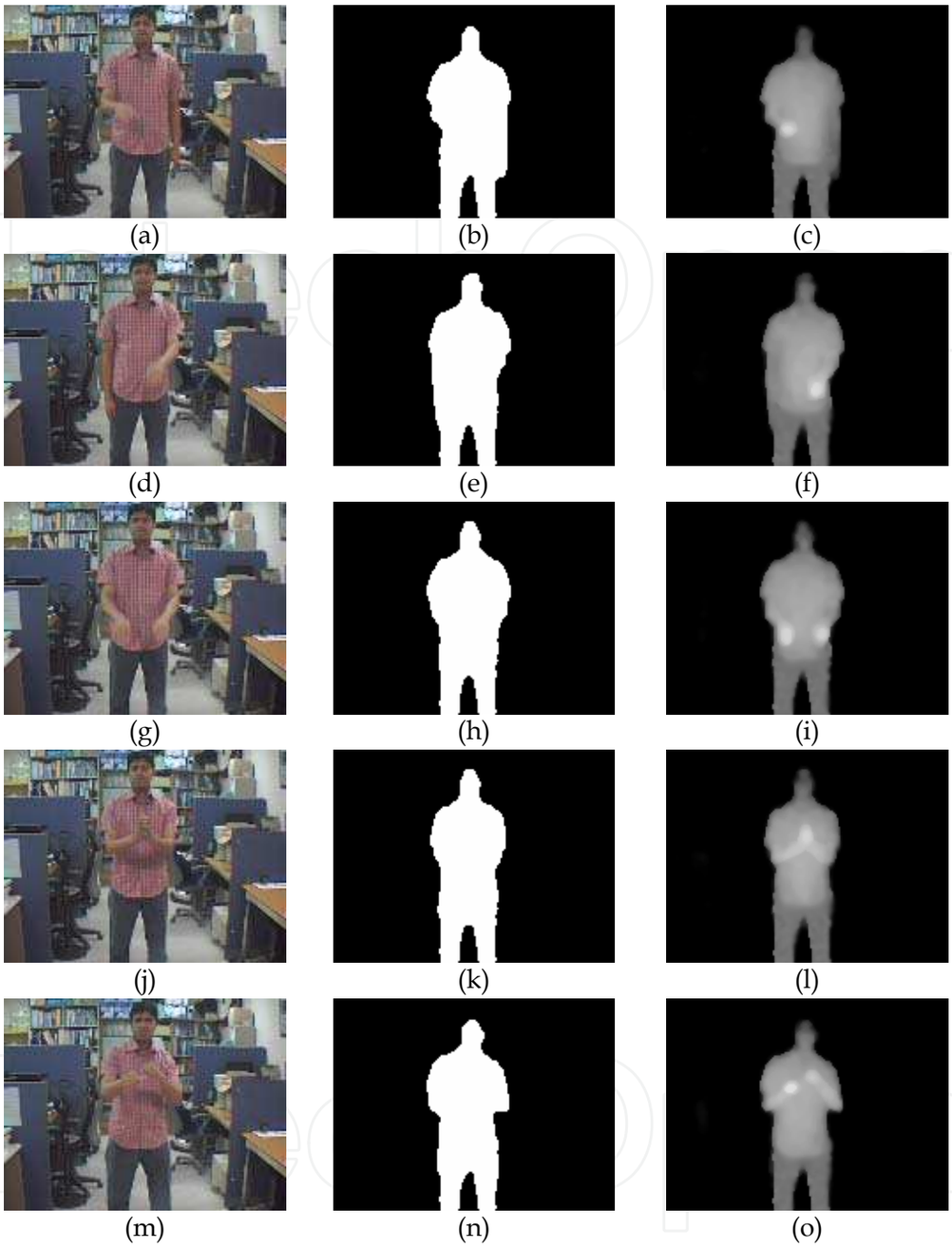


Fig. 1. RGB and their corresponding binary and depth images of (a)-(c) right hand-up-down, (d)-(f) left hand-up-down, (g)-(i) both hands-up-down, (j)-(l) clapping, and (m)-(o) boxing activity respectively.

probability, and an initial state probability distribution. The second consists of a set of density functions representing the observations associated with each state. As a continuous HMM can decode the time-sequential continuous information, it has been applied in many important applications such as speech recognition (Lawrence & Rabiner, 1989), human activity recognition (Sun et al., 2002), gesture recognition (Frolov et al., 2008) etc. A

continuous HMM denoted as $H = \{\Xi, \pi, A, B\}$ can be expressed as follows. $\Xi = \{\Xi_1, \Xi_2, ..., \Xi_q\}$ indicates the states where q is the number of states. The state of the model at time t can be expressed as $\Omega_t \in \Xi,\ 1 \le t \le T$ where T is the length of the observation sequence. The initial probability of the states π can be represented as

$$\pi = \{\pi_j\},\ \sum_{j=1}^{q} \pi_j = 1. \tag{1}$$

The state transition probability matrix is denoted as A where a_{ij} denotes the probability of a changing state from i to j i.e.,

$$a_{i,j} = \mathrm{P}(\Omega_{t+1} = \Xi_j \mid q_t = \Xi_i),\ \ 1 \le i,\ j \le q, \tag{2}$$

$$\sum_{j=1}^{q} a_{i,j} = 1,\ \ 1 \le i \le q. \tag{3}$$

The observation probability matrix is denoted as B where the probability $b_j(d)$ represents the probability of observing d from a state j that can be expressed as

$$\mathrm{b}_j(d) = \mathrm{P}(O_t = d \mid \Omega_t = \Xi_j),\ \ 1 \le j \le q. \tag{4}$$

Though there are various types of HMMs (Lawrence & Rabiner, 1989), one of the popular HMMs is the left-to-right model as shown in Fig. 2. In this model, the initial probability of the states π be initialized as {1,0,0,0} if there are four states and the modeling procedure is to be started from the first state. The transition matrix A can be initialized according to the transition between the states. The initial transitions can be uniformly distributed based on the connections between the states. Thus, the transition matrix can be represented as

$$A = \begin{bmatrix} 0.333 & 0.333 & 0.333 & 0 \\ 0 & 0.333 & 0.333 & 0.333 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5}$$

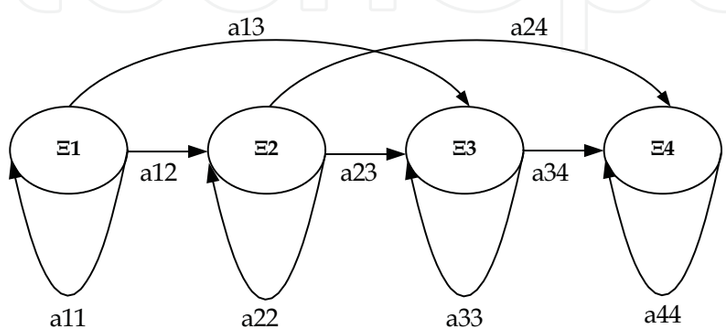


Fig. 2. A four state left-to-right HMM.

In the continuous observation probability density function matrix B, the commonly used distribution to describe the observation densities is the Gaussian one. To represent the continuous observation probability matrix, the mean and covariance are utilized. Weight coefficients are necessary to use during the mixture of the probability density function (pdf). Thus, the observation probability of O_t at time t from state j can be represented as

$$b_j(O_t) = \sum_{k=1}^{M} c_{j,k} b_{j,k}(O_t), \quad 1 \le j \le q, \tag{6}$$

$$\sum_{k=1}^{M} c_{j,k} = 1, \quad 1 \le j \le q \tag{7}$$

where c represents the weighting coefficients, M the number of mixtures, and O_t the observation feature vector at time t. The Baum-Welch algorithm (Baum et al., 1970) can be applied to estimate the HMM parameters as

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{q} \alpha_t(i)\beta_t(i)}, \tag{8}$$

$$\gamma_t(i,j) = \frac{\alpha_t(i)a_{i,j}b_j(O_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^{q}\sum_{j=1}^{q} \alpha_t(i)a_{i,j}b_j(O_{t+1})\beta_{t+1}(j)}, \tag{9}$$

$$\xi_t(j,k) = \frac{\sum_{i=1}^{q} \alpha_t(i)a_{i,j}c_{j,k}g_{j,k}(O_t)\beta_t(j)}{\sum_{i=1}^{q}\sum_{j=1}^{q} \alpha_t(i)a_{i,j}c_{j,k}g_{j,k}(O_t)\beta_t(j)}, \tag{10}$$

$$g_{j,k}(O_t) = \mathrm{N}(x \mid \mu_{j,k}, \Sigma_{j,k}), \tag{11}$$

$$g_{j,k}(O_t) = \sum_{k=1}^{M} c_{j,k} \frac{1}{(2\pi)^{P/2} \mid \Sigma_{j,k} \mid} \exp\left\{-\frac{1}{2}(O_t - \mu_{j,k})^T \Sigma_{j,k}^{-1}(O_t - \mu_{j,k})^T\right\} \tag{12}$$

where $\gamma_t(i)$ represents the probability of staying in the state i at time t. $\gamma_t(i,j)$ is the probability of staying in a state i at time t and a state j at time $t+1$. α and β are the forward and backward variables respectively. T contains the length of the observation sequence. $g_{j,k}(O_t)$ indicates k^{th} mixture component probability at time t, P dimension of the observation feature vector, and $\xi_t(j,k)$ probability of selecting k^{th} mixture component in state j at time t. The estimated HMM parameters can be represented as follows.

$$\hat{a}_{i,j} = \frac{\sum_{t=1}^{T-1} \gamma_t(i,j)}{\sum_{t=1}^{T} \gamma_t(i)}, \tag{13}$$

$$\hat{c}_{j,k} = \frac{\sum_{t=1}^{T} \xi_t(i,j)}{\sum_{t=1}^{T} \gamma_t(j)}, \tag{14}$$

$$\hat{\mu}_{j,k} = \frac{\sum_{t=1}^{T} \xi_t(j,k) x_t}{\sum_{t=1}^{T} \xi_t(j,k)}, \tag{15}$$

$$\hat{\Sigma}_{j,k} = \frac{\sum_{t=1}^{T} \xi_t(j,k) O_t O_t^T}{\sum_{t=1}^{T} \xi_t(j,k)} - \hat{\mu}_{j,k} \hat{\mu}_{j,k}^T \tag{16}$$

where $\hat{a}_{i,j}$ represents the estimated transition probability from the state i to the state j and $c_{j,k}$ the estimated k^{th} mixture weights in state j, $\mu_{j,k}$ the estimated mean of k^{th} mixture in state j, and $\hat{\Sigma}_{j,k}$ the estimated covariance of k^{th} mixture in state j.

3. Continuous HMM-based HAR system

In our HAR system, we apply the continuous HMM to model and recognize various human activities from time-sequential depth silhouette features. Our HAR system consists of depth silhouette extraction, feature extraction, modeling, and recognition via the continuous HMM. Fig. 3 shows the key processes of our depth silhouette feature-based activity recognition system.

3.1 Depth silhouette extraction

A Gaussian probability distribution function is used to remove background from the RGB frames and to extract the binary Region of Interest (ROI) based on which depth ROIs are extracted from the corresponding depth images acquired by a depth camera. ZCAM™, a commercial camera developed by the 3DV system, is used to acquire the RGB and depth images of different activities (Iddan & Yahav, 2001). The image sensor in the ZCAM produces the RGB and distance information for the object captured by the camera.

To capture the depth information, the image sensor first senses the surface boundaries of the object and arranges each object according to the distance information. The depth value indicates the range of each pixel in the scene to the camera as a grayscale value such that the shorter ranged pixels have brighter and longer ones contains darker values. The system provides both RGB and depth images simultaneously. Figs. 4(a) to (e) show a background

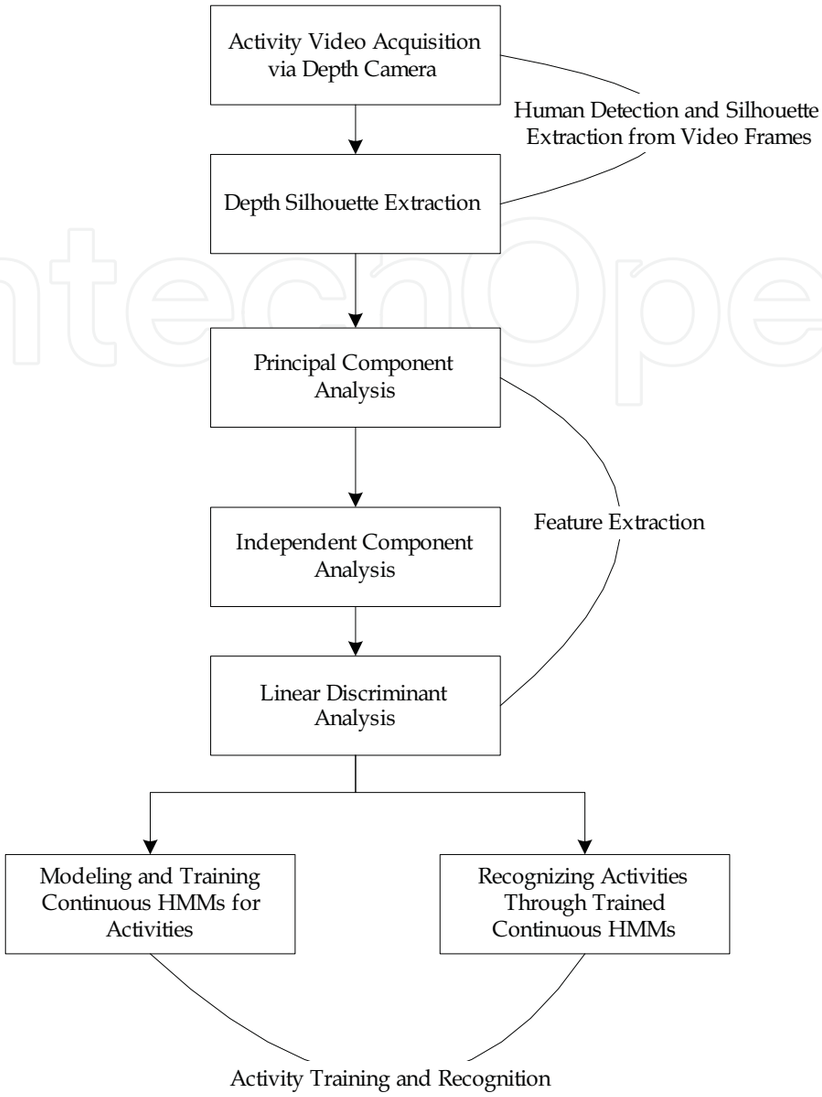


Fig. 3. Depth Silhouette-based human activity recognition system using continuous HMMs.

image, a RGB frame from a both hands up-down sequence, its corresponding binary, depth, and pseudo color image respectively. In the depth image, the higher pixel intensity indicates the near and the lower the far distance. For instance, in Fig. 4(d), the forearm regions are brighter than the body. Thus, different body components used in different activities can be represented effectively in the depth map or the depth information and hence can contribute effectively in the feature generation. On the contrary, the binary silhouettes contain a flat pixel value in the human body and hence cannot distinguish the body postures effectively. Consequently, by means of the infrared sensor-based camera, we obtained both the RGB and depth images of distinguished activities at 30fps to apply on our HAR system.

Since the raw depth images acquired by the camera consist of random noises, median filtering was applied on the depth images to make them smooth. Then, the depth silhouette was extracted from every depth image and resized to the size of 50x50. Fig. 5 shows sequences of depth silhouettes from right hand up-down, left hand up-down, both hands

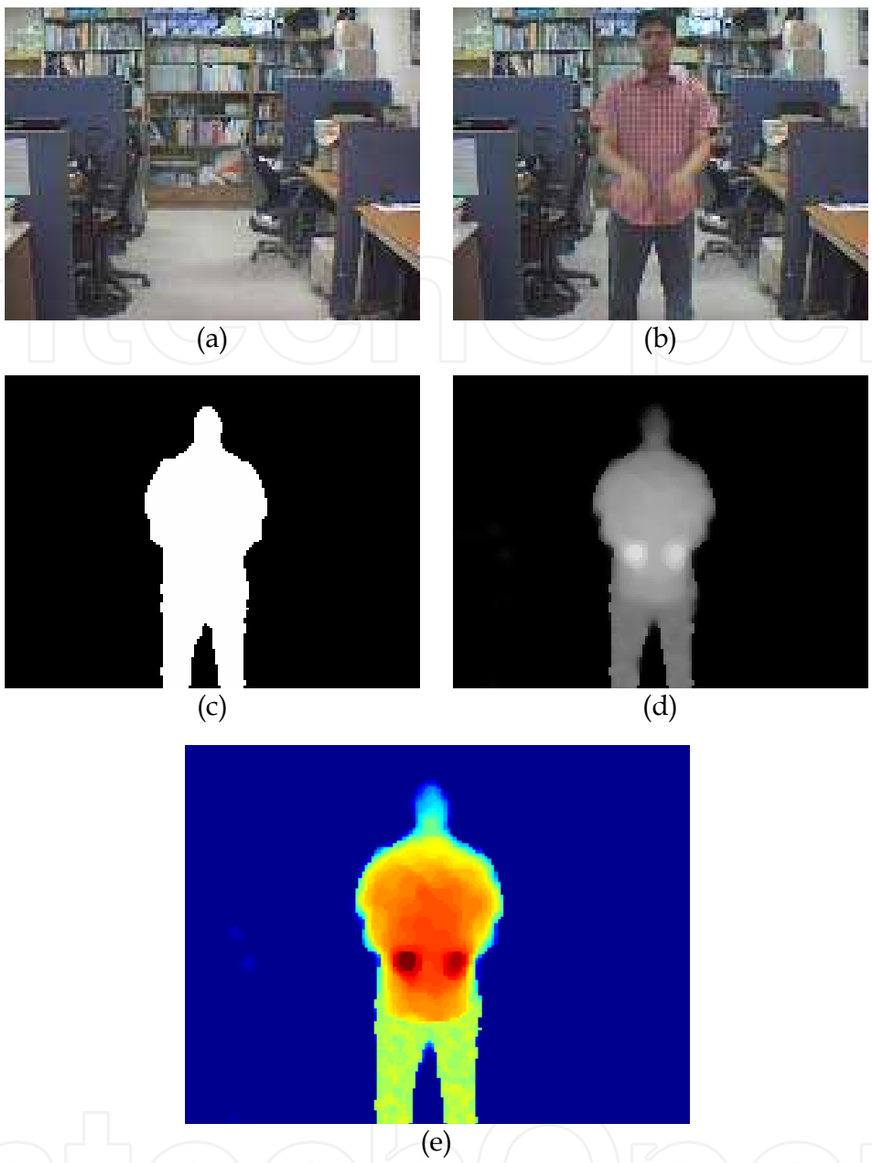


Fig. 4. Sample video images of (a) a background image, (b) a RGB frame from a both hands up-down sequence, (c) its corresponding binary, (d) depth, and (e) pseudo color image.

up-down, clapping, and boxing activities. To apply the feature extraction, each silhouette was converted to a vector with the size of the total pixels (i.e., 2500) in it. The first step before feature extraction is to make all the silhouette vectors zero mean.

3.2 Principal component analysis on depth silhouettes

After preprocessing of the silhouette vectors, we proceed to the dimension reduction process as the training database contains the silhouette vectors with a high dimension (i.e., 2500). In this regard, we applied PCA, one of the most popular methods to approximate original data in the lower dimensional feature space (Niu & Abdel-mottaleb, 2004; Niu & Abdel-mottaleb, 2005; Uddin et al., 2008a; Uddin et al., 2008b; Uddin et al., 2009). The main

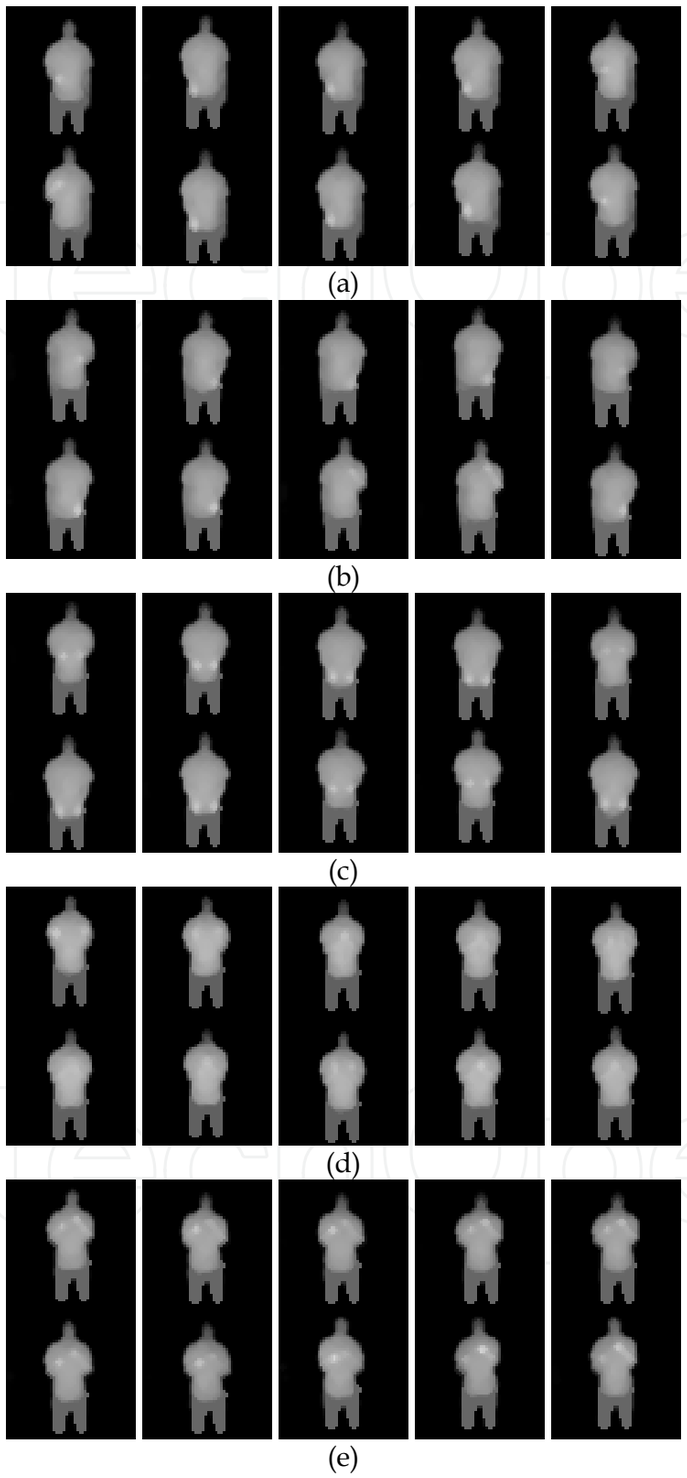


Fig. 5. Ten depth silhouettes from image sequences of (a) right hand up-down, (b) left hand up-down, (c) both hands up-down, (d) clapping, and (e) boxing activity.

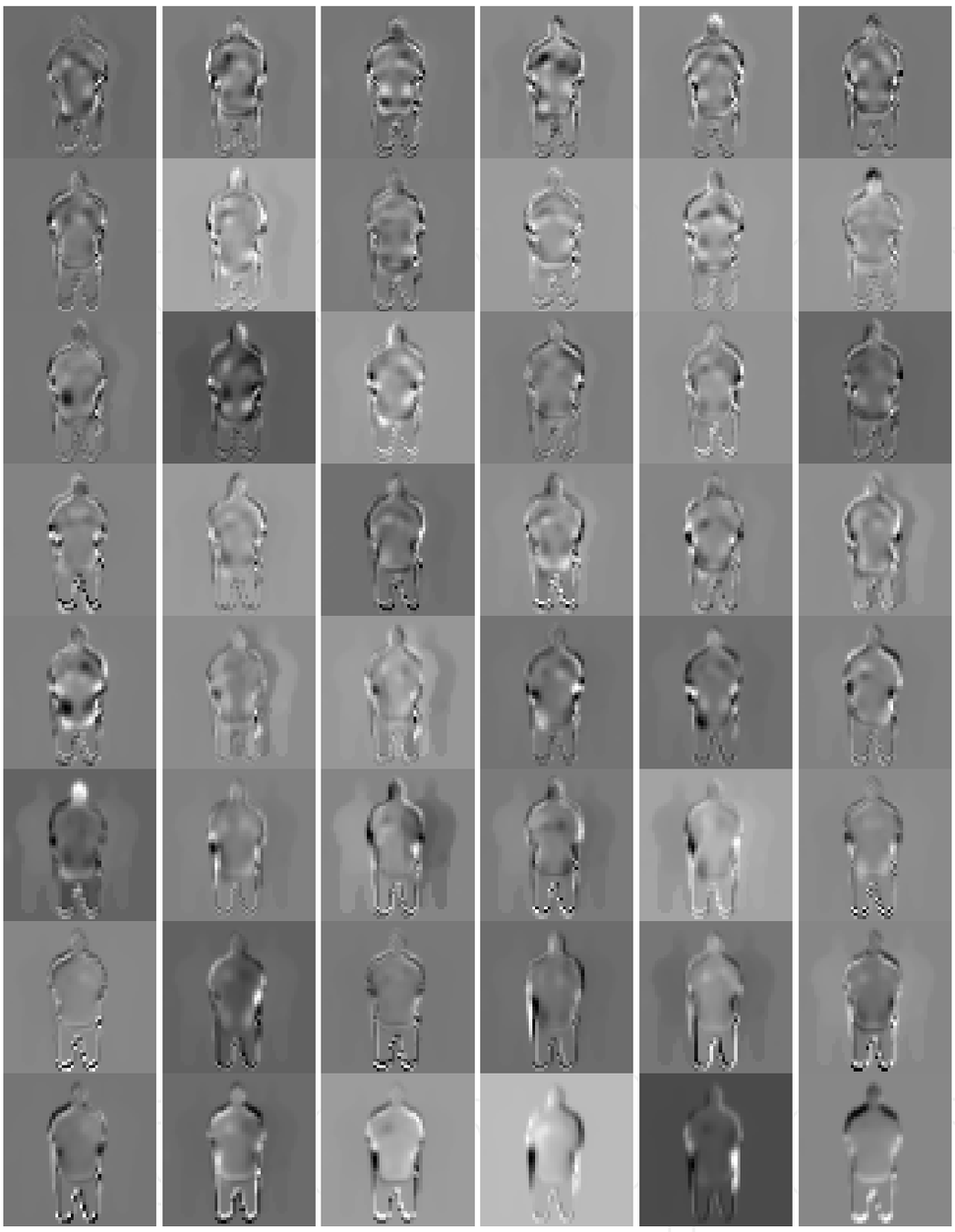


Fig. 6. Forty eight PCs of all the depth silhouettes of the five activities.

approach is to compute the eigenvectors of the covariance data matrix Q and then approximate using the linear combination of top eigenvectors. The covariance matrix of the sample training depth silhouette vectors and the PCs of the covariance matrix can be calculated as

$$Q = \frac{1}{T}\sum_{i=1}^{T}(\tilde{X}_i \tilde{X}_i^T), \tag{17}$$

$$\Lambda = E^T Q E \tag{18}$$

where E represents the matrix of eigenvectors and Λ diagonal matrix of the eigenvalues. 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.

Basically, the eigenvalues close to zero carry negligible variance and hence can be neglected. So, the several m eigenvectors corresponding to the largest eigenvalues can be used to define the subspace. Thus, the full dimensional depth silhouette vectors can be easily represented in the reduced dimension. After applying PCA on the depth silhouettes of various activities, it generates global features representing most frequently moving parts of human body in all activities. However, PCA being a second order statistics-based analysis can only extract global information (Niu & Abdel-mottaleb, 2004; Niu & Abdel-mottaleb, 2005; Uddin et al., 2008a; Uddin et al., 2008b; Uddin et al., 2009). Fig.6 shows 48 basis images after PCA is applied on 2250 images of the five activities. The basis images are the resized eigenvectors (i.e., 50x50) normalized in a grayscale. Fig. 7 shows the top 150 eigenvalues corresponding to the first 150 eigenvectors.

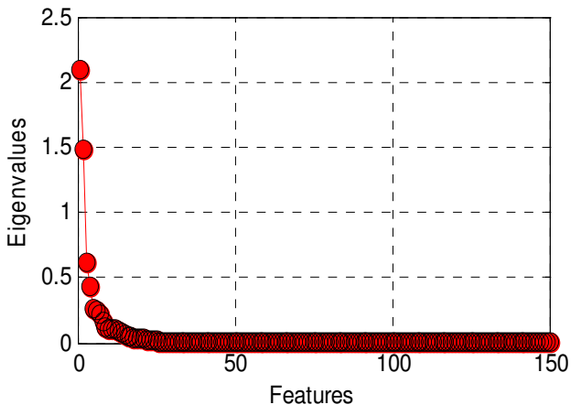


Fig. 7. Top 150 eigenvalues of the training depth silhouettes of the five activities.

If there are k number of depth silhouette vectors in the database where each vector is t dimensional and top m eigenvectors are chosen after applying PCA, the size of the PCA representations of all silhouette vectors becomes kxm where each silhouette vector represents the size of $1xm$. For our experiments, we considered 150 PCs after applying PCA over the training database of 2250 silhouettes of the five activities and as a result, m becomes 150 and the size of the E_m is 2500x150 where each column vector represents a PC. Thus, projecting each silhouette image vector with the size of 1x2500 onto the PCA feature space, it can be reduced to 1x150.

3.3 Independent component analysis on depth silhouettes

Independent Component Analysis (ICA) is a higher order statistical approach than PCA for separating a mixture of signals into its components. The most well-known applications of ICA are in the field of signal and image processing such as biomedical signals (Jung et. Al,

2001), speech (Lawrence & Rabiner, 1989; Kwon & Lee, 2004), face (Yang et al., 2005), etc. ICA has been recently applied in the field of facial expression analysis areas to focus on the local face features (Kwak & Pedrycz, 2007; Uddin et al., 2009).

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

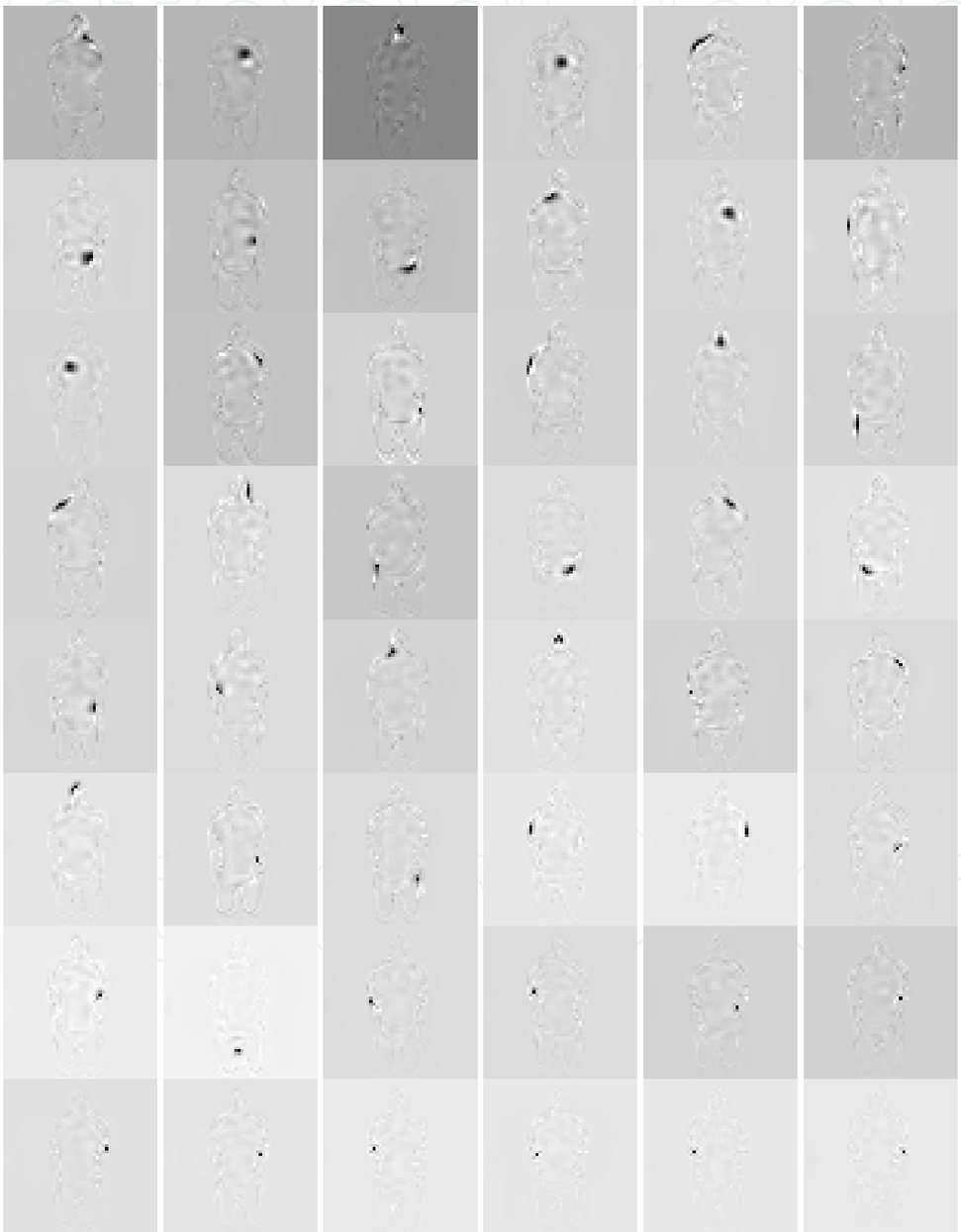


Fig. 8. Forty eight ICs of all the depth silhouettes of the five activities.

$$X = AS \tag{19}$$

where A represents an unknown linear mixing matrix of full rank.
An ICA algorithm learns the weight matrix W, which is inverse of mixing matrix A. W is used to recover a set of independent basis images S. The ICA basis images reflect local feature information rather than global information as in PCA. ICA basis images show the local features of the body parts in activity. Fig. 8 shows 48 ICA basis images for all activities. Before applying ICA, PCA is used to reduce the dimension of the image data. ICA is performed on E_m as follows.

$$S = WE_m^T, \tag{20}$$

$$E_m^T = W^{-1}S, \tag{21}$$

$$X_r = VW^{-1}S \tag{22}$$

where V is the projection of the images X on E_m and X_r the reconstructed original images. The IC representation I_i of i^{th} silhouette vector $\tilde{X}_i$ from an activity image sequence can be expressed as

$$I_i = \tilde{X}_i E_m W^{-1}. \tag{23}$$

Since ICA is applied on the PC features in our HAR work, the size of the ICA representations of the depth silhouette vectors are same as PCA. As the top 150 PCs are chosen to apply ICA, therefore, the size of the IC features of each depth silhouette vector and the weighting matrix W are 1x150 and 150x150 respectively.
However, as PCA considers the second order moments only, it lacks information on higher order statistics. On the contrary, ICA considers higher order statistics and it identifies the independent source components from their linear mixtures. Hence, ICA provides a more powerful data representation than PCA as it tries to provide an independent rather than uncorrelated feature representation. Thus, we applied ICA in our depth silhouette-based HAR system to find out statistically independent local features for improved HAR.

3.4 LDA on the independent depth silhouette component features

Linear Discriminant Analysis (LDA) is an efficient classification tool that works based on grouping of similar classes of data. It finds the directions along which the classes are best separated by considering the within-class scatter but also the between-class scatter (Kwak & Pedrycz, 2007; Uddin et al., 2009). It has been used extensively in various applications such as facial expression recognition (Kwak & Pedrycz, 2007; Uddin et al., 2009) and human activity recognition (Uddin et al., 2008). Basically, LDA projects data onto a lower-dimensional vector space such that the ratios of the between-class scatter and the within-class scatter is maximized, thus achieving maximum discrimination.
LDA generates an optimal linear discriminant function which maps the input into the classification space based on which the class identification of the samples can be decided (Kwak & Pedrycz, 2007). The within-class scatter matrix, S_W and the between-class scatter matrix, S_B are computed by the following equations:

$$S_B = \sum_{i=1}^{c} J_i (\overline{m_i} - \overline{\overline{m}})(\overline{m_i} - \overline{\overline{m}})^T, \tag{24}$$

$$S_W = \sum_{i=1}^{c} \sum_{m_k \in C_i} (m_k - \overline{m_i})(m_k - \overline{m_i})^T \tag{25}$$

where J_i is the number of vectors in i^{th} class C_i. c is the number of classes and in our case, it represents the number of activities. m represents the mean of all vectors, m_i the mean of the class C_i and m_k the vector of a specific class. The optimal discrimination matrix D_{opt} is chosen from the maximization of ratio of the determinant of the between and within-class scatter matrix as

$$D_{opt} = \arg \max_D \frac{\left| D^T S_B D \right|}{\left| D^T S_W D \right|} \tag{26}$$

where D_{opt} is the set of discriminant vectors corresponding to the $(c-1)$ largest generalized eigenvalues λ problem as

$$S_B d_i = \lambda_i S_W d_i. \tag{27}$$

The LDA algorithm seeks the vectors in the underlying space to create the best discrimination among different classes. Thus, the extracted local feature-based ICA representations of the binary silhouettes of different activities can be extended by LDA. LDA on the IC features for the depth silhouette vectors can be represented as

$$F_i = I_i D_{opt}^T. \tag{28}$$

Fig. 9 depicts the 3-D representation of the depth silhouette features after applying on three ICs that are chosen on the basis of top kurtosis values. Fig. 10 shows a 3-D plot of LDA on

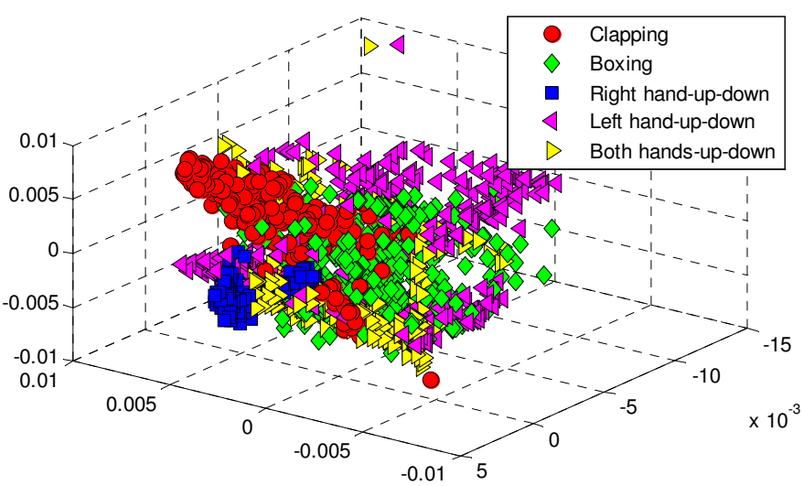


Fig. 9. A plot of the three IC features of 2250 depth silhouettes of all activities.

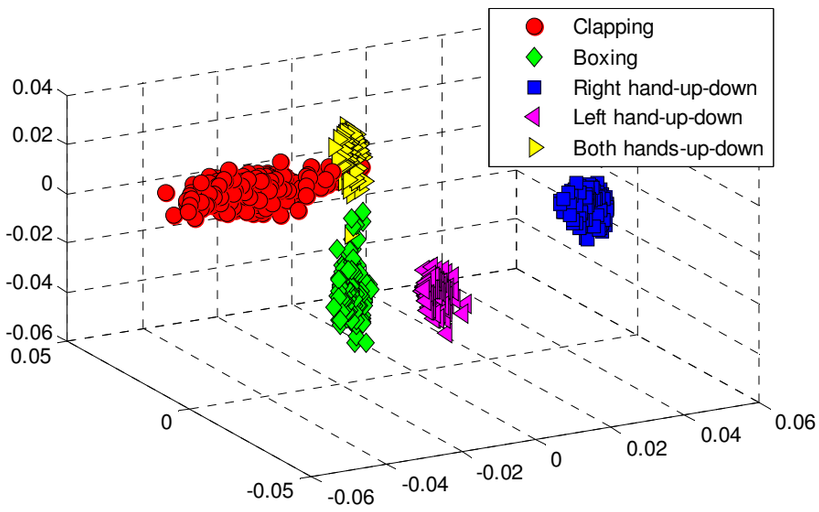


Fig. 10. A plot of the three LDA features of 2250 depth silhouettes of the activities.

the IC features of the silhouettes of the five activities where 150 ICs are taken into consideration. Fig. 10 demonstrates a good separation among the representations of the depth silhouettes of the five activities. Once the LDA algorithm is applied on a database of the depth silhouette vectors where each vector is 150 dimensional, the size of the LDA subspace becomes 4x150 as $(c - 1) = 4$. Hence, the LDA projection of the IC feature vector of each depth silhouette becomes 1x4.

3.5 Continuous HMM for depth silhouette-based activity training and recognition

Once human activities are represented in the time-sequential depth silhouette features, the continuous HMM can be applied effectively for HAR. In our system, we considered a four-state left-to-right HMM to model the human activities. The initial probability of the states π was initialized as {1,0,0,0}. The transition matrix A was uniformly initialized according to the transition between the states. Two mixtures per depth silhouette features were considered to model the activities by the continuous HMMs. Finally, the continuous HMMs were trained using the Baum-Welch parameter estimation algorithm. Each activity was represented by a distinct continuous HMM. Figs. 11 and 12 show the transition probabilities of a left hand up-down HMM before and after training respectively.

To recognize an activity, a feature vector sequence obtained from the activity image sequence was applied on all trained continuous HMMs to calculate the likelihood and the

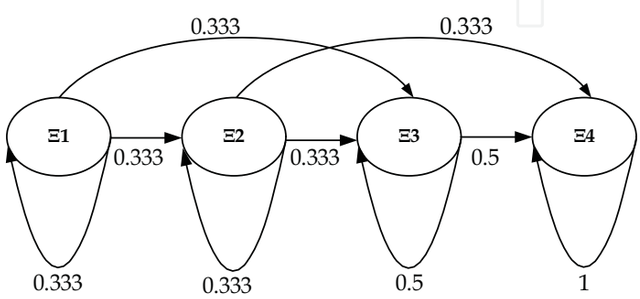


Fig. 11. A left hand up-down HMM before training.

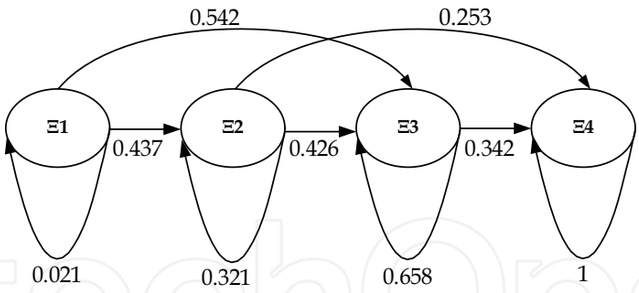


Fig. 12. A left hand up-down HMM after training.

one was chosen with the highest probability. Thus, to test a feature vector sequence O (i.e., $O_1, O_2, ..., O_T$), we found the appropriate HMM as

$$decision = \underset{i=1,2,...,N}{argmax}\{L_i\}, \tag{29}$$

$$L_i = \Pr(O \mid H_i) \tag{30}$$

where L represents the likelihood of O on corresponding trained activity HMM H.

4. Experimental setups and results

For the experiments, the binary (Uddin et al., 2008a) and depth silhouettes (Uddin et al., 2008b) were used as input to our HAR system. Five activities were recognized using different types of features with the continuous HMMs: namely right hand up-down, left hand up-down, both hands up-down, clapping, and boxing. Each activity sequence consisted of a time-series set of 30 silhouettes. A total of 15 sequences from each activity were used to build the feature space in training. Thus, the whole database consisted of a total of 2250 images. A total of 200 sequences (i.e., 40 sequences for each activity) were used in testing.

We started our experiments with the traditional binary silhouette-based HAR. After the background subtraction, the ROIs containing the binary silhouettes were extracted from the sequences. Since the binary silhouettes from the activities used in the experiments were similar to each other, the recognizer produced much low recognition rates for all the approaches (i.e., PCA, LDA on the PC features, ICA, and LDA on the IC features) as shown in Table 1.

In the following experiments, the binary silhouettes were replaced with the depth ones. The combination of PCA on the depth silhouettes with the continuous HMM were experimented first. PCA found the global depth silhouette features to be applied on the continuous HMMs, obtaining the mean recognition rate of 85.50%. By extending the PCA features by LDA, we obtained the mean recognition rate of 86.50%, improving the recognition marginally. We continued to apply ICA on the depth silhouettes to obtain improved local depth silhouette features, achieving the mean recognition rate of 93.50%. Finally, LDA on the IC features with the continuous HMMs produced the highest recognition rate of 99%. The recognition results using the various types of features from the depth silhouettes are shown in Table 2.

Approach	Activity	Recognition Rate	Mean	Standard Deviation
PCA	Right hand up-down	32.50%	36.50	6.02
	Left hand up-down	35		
	Both hands up-down	30		
	Clapping	40		
	Boxing	45		
LDA on the PC features	Right hand up-down	35	36	4.18
	Left hand up-down	32.50		
	Both hands up-down	37.50		
	Clapping	32.50		
	Boxing	42.50		
ICA	Right hand up-down	42.50	43	4.80
	Left hand up-down	37.50		
	Both hands up-down	50		
	Clapping	40		
	Boxing	45		
LDA on the IC features	Right hand up-down	45	45.50	4.11
	Left hand up-down	42.50		
	Both hands up-down	52.50		
	Clapping	42.50		
	Boxing	45		

Table 1. Recognition results using the binary silhouette-based approaches.

Approach	Activity	Recognition Rate	Mean	Standard Deviation
PCA	Right hand up-down	90%	85.50	4.47
	Left hand up-down	90		
	Both hands up-down	80		
	Clapping	85		
	Boxing	82.50		
LDA on the PC features	Right hand up-down	90	86.50	4.54
	Left hand up-down	92.50		
	Both hands up-down	82.50		
	Clapping	85		
	Boxing	82.50		
ICA	Right hand up-down	92.50	93.50	1.37
	Left hand up-down	95		
	Both hands up-down	92.50		
	Clapping	95		
	Boxing	92.50		
LDA on the IC features	Right hand up-down	100	99	1.36
	Left hand up-down	100		
	Both hands up-down	100		
	Clapping	97.50		
	Boxing	97.50		

Table 2. Recognition results using the depth silhouette-based approaches.

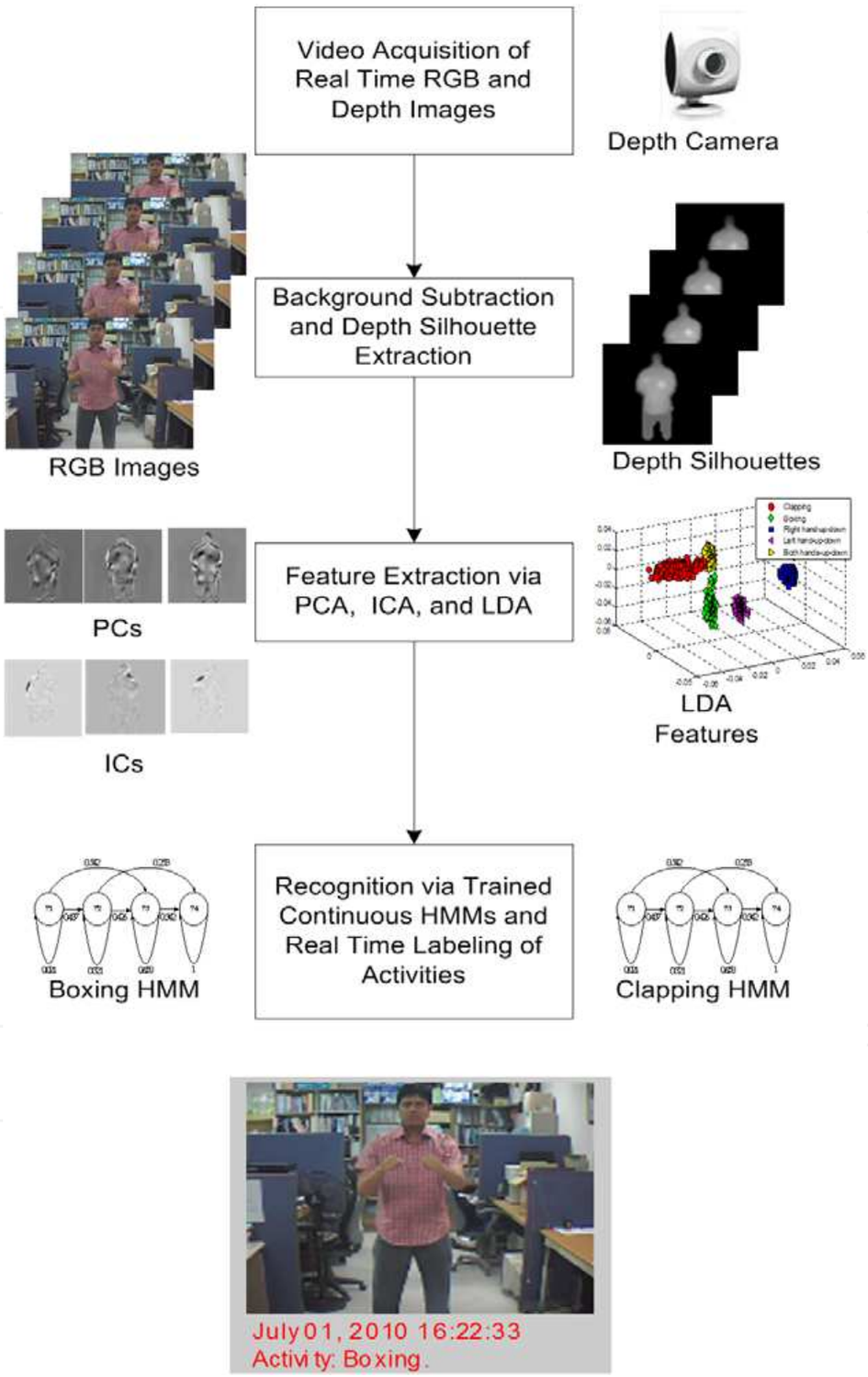


Fig. 13. Basic steps of our depth silhouette-based real-time HAR system.

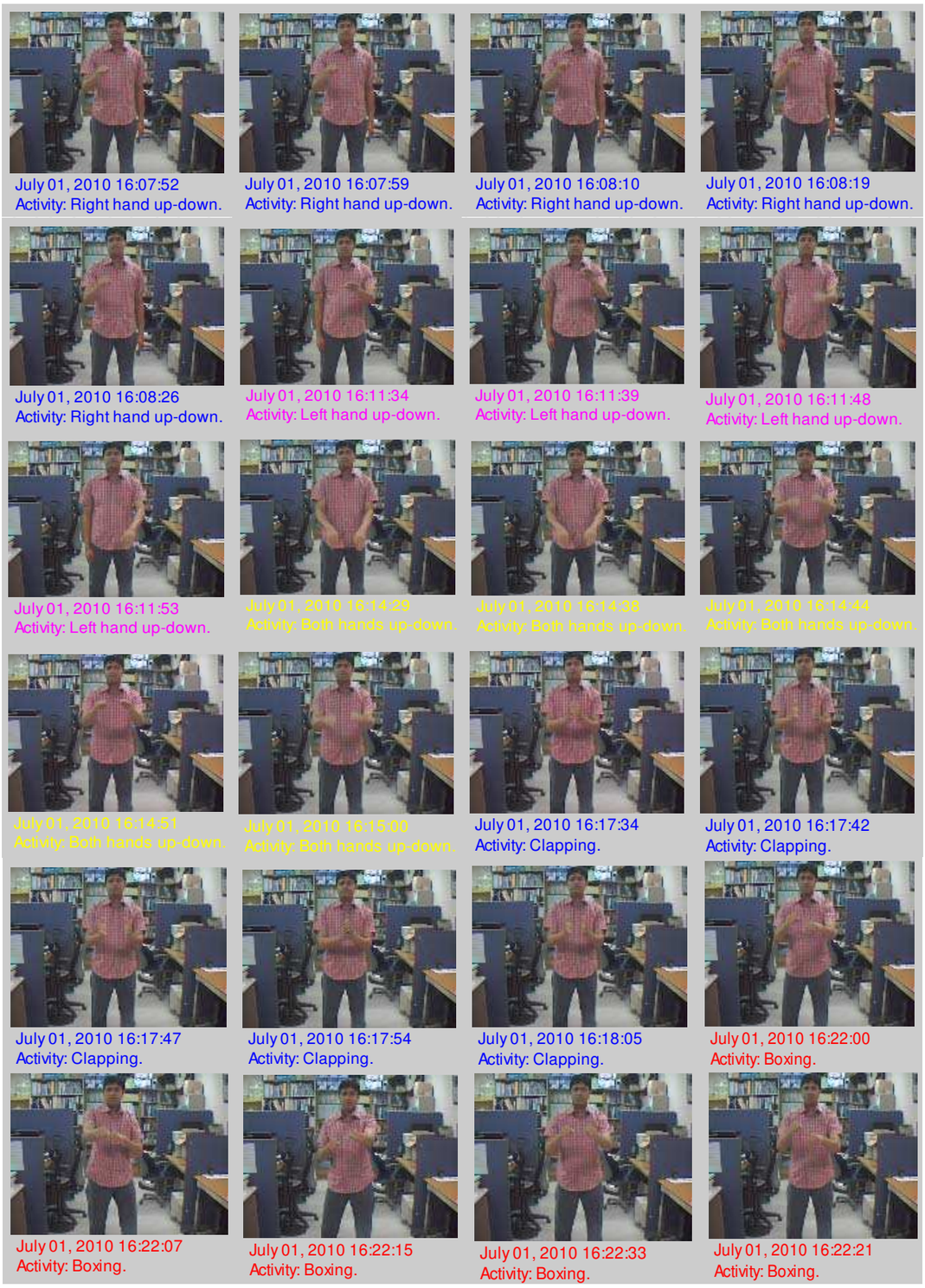


Fig. 14. Real-time human activity recognition results.

4.1 Real-time depth silhouette-based human activity recognition

After testing our system offline, we continued our study to recognize the five activities in real-time. We implemented a HAR system in a programming language, Matlab 7.4.0 to acquire the RGB as well as depth images of various activities in real-time and applied LDA on the IC features on the depth silhouettes with the continuous HMMs. We used a non overlapping window of 30 sequential depth frames from the real time video data and extracted the IC-based features to apply on the trained HMMs for recognition. Fig. 13 shows the architecture of our real-time HAR system. Fig. 14 shows some of our sample real-time recognition results with date and time where the result is shown at the bellow of the RGB images automatically. Here, the labeled RGB image is shown for the clarity of the activity though we used the depth silhouettes for activity recognition.

5. Conclusion

In this chapter, we have presented the basics of the continuous HMM and its application to human activity recognition. Our depth silhouette-based HAR methodology successfully incorporates the continuous HMM to recognize various human activities: we have shown that the depth silhouettes outperform the traditional binary silhouettes significantly, although the same HMM has been incorporated for recognition. We have also demonstrated real-time working of our presented HAR system. As presented, the HMMs can be effective in the field of HAR where pattern recognition of spatiotemporally changing information is critical.

6. Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science, and Technology (No. 2010-0001860).

7. References

Agarwal, A.; & Triggs B. (2006). Recovering 3D human pose from monocular images, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 28, pp. 44-58.

Bartlett, M.; Movellan, J. & Sejnowski, T. (2002). Face Recognition by Independent Component Analysis, *IEEE Transactions on Neural Networks*, Vol., 13, pp. 1450-1464.

Baum, E.; Petrie, T.; Soules, G.; & Weiss, N. (1970). A Maximization Technique Occurring in The Statistical Analysis of Probabilistic Functions of Markov chains. *Annals of Mathematical Statistics,* Vol., 41, pp. 164-171

Ben-Arie, J.; Wang, Z.; Pandit, P. & Rajaram, S. (2002). Human Activity Recognition Using Multidimensional Indexing, *IEEE Transactions on Pattern Analysis and Machine Intelligence Archive*, Vol., 24(8), pp. 1091-1104.

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

Cohen, I. & Lim, H. (2003). Inference of Human Postures by Classification of 3D Human Body Shape, *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pp. 74-81.

Frolov V.; Deml B.; & Hannig G. (2008) Gesture Recognition with Hidden Markov Models to Enable Multi-modal Haptic Feedback, Lecture Notes in Computer Science, Vol. 5024/2008, pp. 786-795.

Iddan , G. J. & Yahav, G. (2001). 3D imaging in the studio (and elsewhere…). *Proceedings of SPIE*, Vol., 4298, pp 48-55.

Iwai, Y.; Hata, T. & Yachida, M. (1997) Gesture Recognition Based on Subspace Method and Hidden Markov Model, *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 960-966.

Jung, T.; Makeig, S.; Westerfield, M.; Townsend, J.; Courchesne, E.; & Sejnowski, T. J. (2001). Analysis and Visualization of Single-Trial Event-Related Potentials, *Human Brain Mapping*, Vol., 14, pp. 166-185.

Kanungu, T.; Mount, D. M.; Netanyahu, N.; Piatko, C.; Silverman, R. & Wu, A. Y. (2000). The analysis of a simple k-means clustering algorithm, Proceedings of 16th ACM Symposium On Computational Geometry, pp. 101-109.

Kwak, K.-C. & Pedrycz, W. (2007). Face Recognition Using an Enhanced Independent Component Analysis Approach, *IEEE Transactions on Neural Networks*, Vol., 18(2), pp. 530-541.

Kwon, O. W. & Lee, T. W. (2004). Phoneme recognition using ICA-based feature extraction and transformation, *Signal Processing*, Vol., 84(6), pp. 1005–1019.

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

Linde, Y.; Buzo, A. & Gray, R. (1980). An Algorithm for Vector Quantizer Design, *IEEE Transaction on Communications*, Vol., 28(1), pp. 84–94.

Mckeown, M. J.; Makeig, S.; Brown, G. G.; Jung, T. P.; Kindermann, S. S.; Bell, A. J. & Sejnowski, T. J. (1998) Analysis of fMRI by decomposition into independent spatial components, *Human Brain Mapping*, Vol., 6(3), pp. 160–188.

Nakata, T. (2006). Recognizing Human Activities in Video by Multi-resolutional Optical Flow, *Proceedings of International Conference on Intelligent Robots and Systems*, pp. 1793-1798.

Niu, F. & Abdel-Mottaleb M. (2004).View-Invariant Human Activity Recognition Based on Shape and Motion Features, *Proceedings of the IEEE Sixth International Symposium on Multimedia Software Engineering*, pp. 546-556.

Niu, F. & Abdel-Mottaleb, M. (2005). HMM-Based Segmentation and Recognition of Human Activities from Video Sequences, *Proceedings of IEEE International Conference on Multimedia & Expo*, pp. 804-807.

Robertson, N. & Reid, I. (2006). A General Method for Human Activity Recognition in Video, *Computer Vision and Image Understanding*, Vol., 104(2), pp. 232 - 248.

Sun, X.; Chen, C. & Manjunath, B. S. (2002). Probabilistic Motion Parameter Models for Human Activity Recognition, *Proceedings of 16th International Conference on Pattern recognition*, pp. 443-450.

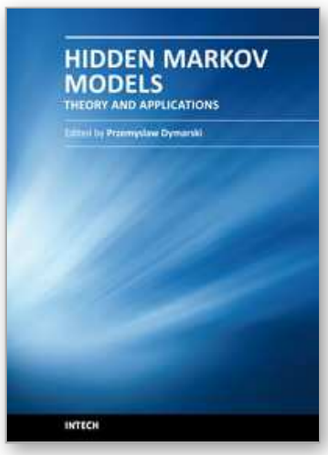
Yamato, J.; Ohya, J. & Ishii, K. (1992). Recognizing Human Action in Time-Sequential Images using Hidden Markov Model, *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 379-385.p

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

Uddin, M. Z.; Lee, J. J. & Kim T.-S. (2008a) Shape-Based Human Activity Recognition Using Independent Component Analysis and Hidden Markov Model, *Proceedings of The s21st International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, pp. 245-254.

Uddin, M. Z.; Lee, J. J. & Kim, T.-S. (2008b). Human Activity Recognition Using Independent Component Features from Depth Images, *Proceedings of the 5th International Conference on Ubiquitous Healthcare*, pp. 181-183.

Uddin, M. Z.; Lee, J. J. & Kim, T.-S. (2009). An Enhanced Independent Component-Based Human Facial Expression Recognition from Video, *IEEE Transactions on Consumer Electronics*, Vol., 55(4), pp. 2216-2224.



Hidden Markov Models, Theory and Applications

Edited by Dr. Przemyslaw Dymarski

ISBN 978-953-307-208-1
Hard cover, 314 pages
Publisher InTech
Published online 19, April, 2011
Published in print edition April, 2011

Hidden Markov Models (HMMs), although known for decades, have made a big career nowadays and are still in state of development. This book presents theoretical issues and a variety of HMMs applications in speech recognition and synthesis, medicine, neurosciences, computational biology, bioinformatics, seismology, environment protection and engineering. I hope that the reader will find this book useful and helpful for their own research.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen