A Literature Survey on Human Activity Recognition via Hidden Markov Model

Hemali S. Mojidra
M.E. [3rd sem] student, Computer Sci. & Eng. Dept.
Government Engineering College, Sector-28
Gandhinagar, India

Viral H. Borisagar
Assistant Professor, Computer Engineering Dept.
Government Engineering College, Sector-28
Gandhinagar, India

ABSTRACT
Human Activity Recognition (HAR) is popular research topic in computer vision and image processing area. Hidden Markov Models (HMMs) are used to recognize the pattern. In this paper, literature survey of different methodology and steps adapted to recognize human activities via trained Hidden Markov Model (HMM) is discussed. HMM is trained using parameters initialization of it. Parameters are initialized using feature extraction from sequence of images. Before Feature extraction image data are converted into binary or depth silhouettes. The conventional approach of features extraction from sequences of silhouetted images is using Principal Component Analysis (PCA) and novel approach is Independent Component Analysis (ICA) for HAR.

General Terms
Image Processing, Computer vision, Human Activity Recognition

Keywords
Human Activity Recognition (HAR), PCA, ICA, LDA, HMM.

1. INTRODUCTION
Numbers of literatures are available on the applications of Hidden Markov Models (HMMs) for the pattern recognition. The practical applications of HMMs are in speech recognition, computational biology, biomedical signal interpretation, image classification and segmentation, online world recognition, etc. 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. In practice, there are some well established training algorithms available to automatically optimize the parameters of the HMM. The Baum–Welch [15] training procedure is a standard algorithm which uses the Maximum Likelihood Estimation (MLE) criterion

Recently, human activity recognition (HAR) is becoming a concentrated field to study and research due to an interest in practical computing. Practical computing is a technology that helps to human to recognize the need by normal actions and activity reorganization. Such a system is used in health care center to recognize action and patient needs. Such Applications are described in [1-3]. The major task to recognize activity is extraction of feature information from activity database. Thus, efficient feature extraction, learning, and classification of activity play a key role for HAR. In general, HAR is a difficult task as it doesn't have any of rigid syntax like gesture or sign language recognition.

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 eigen face 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 HAR, PCA utilizes its second order statistical nature or characteristics to focus on the global representations of the shape images. In addition to the face recognition area, it has been applied successfully in various fields such as speech recognition [7], bioinformatics [8], electroencephalogram (EEG) [9], and biomedical signal and image analysis [10]. ICA to focus on the local status of the activity shape features of different activities rather while PCA focus on the local features. As for the modeling and recognition of human activities, HMM has been utilized successfully in many papers such as [1–6]. In [1] and [2], several HMMs were trained utilizing the shape and motion features. In [19] the IC-based shape feature approach is proposed in combination with HMM to recognize human activities from binary silhouettes. In [19] LDA is applied to classify the IC features for better shape feature representation.

The overall structure of the paper is as follows. It is begin with the basic steps for HAR in Section 2 where all the basic steps and approaches are described. In Section 3 comprises of the approaches are discussed using the surveyed paper results. Section 4 draws the concluding remarks.

2. BASIC STEPS AND APPROACHES FOR HAR
In Fig 1, all the basic steps and approaches are described.

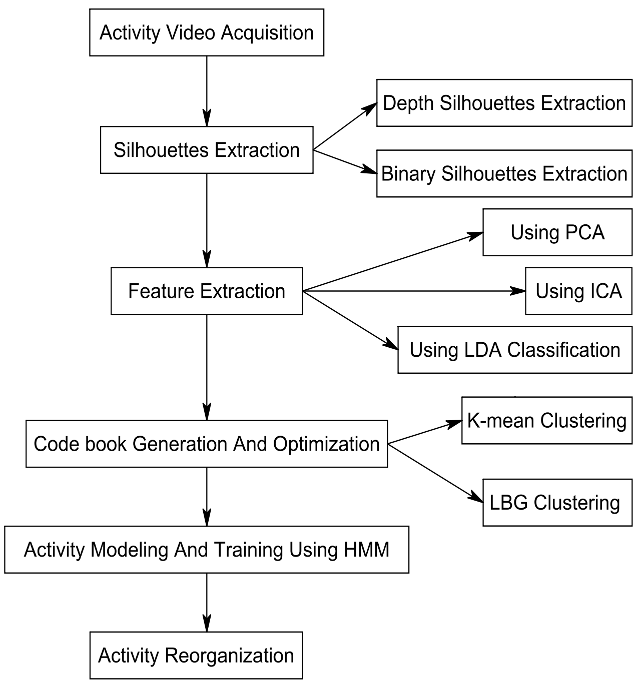

Fig 1 : Basic steps and approaches of HAR

2.1 Video Processing Silhouettes Extraction

Video of single activity is processed to generate the sequence of the images. Sequences of images are used for the HAR process. From sequence images silhouette image are generated. Binary silhouette image only contains 1 and 0. 3D video camera can be used to obtain the depth images, from which depth silhouette images are generated.

2.1.1 Binary and depth silhouettes extraction

From every frame of each video clip, a Region of Interest (ROI) containing the binary activity shape is extracted. Every video clip consists of single human activity. Algorithm to extract binary ROI is described in [19]. From the depth video, depth information of each frame is extracted to find the depth silhouettes. Algorithm to extract the depth silhouettes is described in [22]. According to the algorithm given in [11], ROI is extracted after the gray scale background subtraction for binary silhouettes extraction. To extract binary image, static background images are subtracted from the frames of each activity video clip[19].

In [23], 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. ZCAMTM, a commercial camera developed by the 3DV system, is used to acquire the RGB and depth images of different activities [23]. 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.

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. 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. 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. It is obvious that the binary silhouettes are a poor choice to separate these different postures. For better silhouette representation than binary, the time-sequential depth silhouettes to be used with the discrete HMMs for robust HAR [20, 21].

2.2 Feature Extraction

Feature extraction is a special form of dimensionality reduction. Transforming the input data into the set of features is called feature extraction. From the extracted features, human activities are identified. Principal Component Analysis (PCA) is the popular method for the feature extraction. Another method is Independent Component Analysis (ICA). Linear discriminant analysis (LDA) is the classification tool that can be applied on PC and IC feature in order to get better result.

2.2.1 Feature extraction using PCA

In [19], after preprocessing of the silhouette vectors, the the dimension reduction process is proceed to the training database contains the silhouette vectors with a high dimension. To extract the human activity silhouette features, the most popular feature extraction technique applied in the video-based HAR is Principal Component Analysis (PCA) [1-3]. 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 decor relation of data. 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 and then the approximation is done using a linear combination of a few top eigenvectors. Algorithm to extract feature using PCA is described in [19].

2.2.2 Feature extraction using ICA

Independent component analysis (ICA) is a computational method for separating a multivariate signal into additive subcomponents supposing the mutual statistical independence of the non-Gaussian source signals. It is a special case of blind source separation. When the independence assumption is correct, blind ICA separation of a mixed signal gives very good results. It is also used for signals that are not supposed to be generated by a mixing for analysis purposes [25].

ICA finds the independent components by maximizing the statistical independence of the estimated components. ICA can be define by “Minimization of Mutual Information” and “Maximization of non-Gaussianity” [25].

A higher order statistical method called Independent Component Analysis (ICA) is being actively exploited in the face recognition area [13] and has shown superior performance over PCA. The basic idea of ICA is to represent a set of random observed variables using basis functions where the components are statistically independent. It has also been utilized successfully in other fields such as speech recognition. In [19] local binary silhouette features through ICA to represent human body in different activities usefully is discussed. 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. ICA is applied in the depth silhouette-based HAR system to find out statistically independent local features for improved HAR.

2.2.3 Linear discriminant analysis (LDA)

Linear Discriminant Analysis (LDA) is an efficient classification tool that works based on grouping of similar classes of data. The LDA algorithm seeks the vectors in the underlying space to create the best discrimination among different classes. It finds the directions along which the classes are best separated by considering the within- class scatter but also the between-class scatter [12]. It has been used extensively in various applications such as facial expression recognition [12] and human activity recognition [18]. 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 .The within-class scatter matrix, SW and the between-class scatter matrix, SB are computed by the following equations: To extend the IC features, Linear Discriminant Analysis (LDA) is applied 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.

2.3 Codebook Generation and Optimization

Feature vectors have to be symbolized before applying to train or recognize by HMMs. An efficient codebook of vectors should be generated using vector quantization from the training vectors. Generally two vector quantization algorithms are used: namely ordinary K-means clustering [13] and Linde, Buzo, and Gray (LBG)'s clustering algorithm [14]. 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. How-ever, in the case of LBG, recomputation is done after as-signing 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 code-vector into two new code vectors. After splitting, optimization of the centroids is done to reduce the distortion. It follows the binary splitting technique [14].

2.4 Activity Modeling and Training using HMM

HMM is a collection of finite states connected by transitions. Every state is characterized by transition and symbol observation probabilities. A generic HMM is expressed as H = {X, p , A,B} where X denotes possible states, p is the initial probability of the states, A is the transition probability matrix between hidden states and B is the observation symbols' probability from every state N number of activities require N no of trained HMM. A four-state left to right HMM is generally used to represent each activity. To estimate HMM parameters, Baum-Welch algorithm [15] is used generally.

3. COMPARISONS OF THE APPROACHES

In [19], walking, running, skipping, right hand, and both hand waving activities are recognized. In this, binary silhouettes extraction is used for the feature extraction process. To extract feature PCA and ICA approaches are applied. For code optimization K-mean and LBG clustering are applied. The mean result of activity recognition of IC base approach is higher than the PC base approach. From [19] all the average results of activity recognition rates of both the approaches are shown in Table 1. Graphical representation this result is illustrated in Fig 2. From the graph, LBG clustering technique gives better result compare to K-Mean clustering.

Table 1. Activity recognition rate of IC base feature using K-mean and LBG clustering. [19]

Codebook Size	Mean Activity Recognition Rate	
	K-means Clustering	LBG Clustering
8	74	84
16	83.50	84.50
32	92	96
64	88.50	94.50

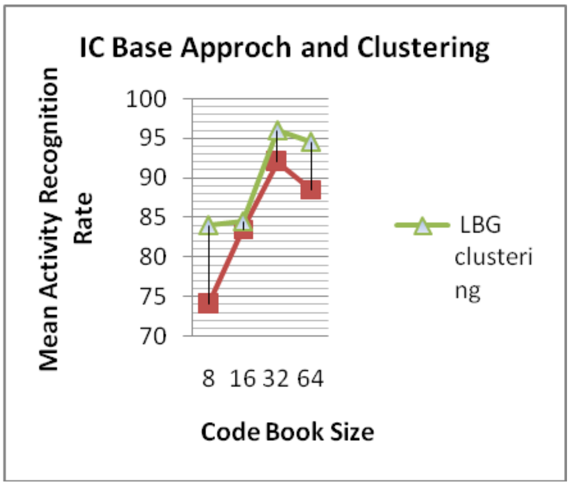

Fig 2: Comparison of K-Mean and LBG clustering [19]

From [19], mean activity recognition rates of K-mean and LBG clustering on IC base feature extraction are shown in Table 2. Graphical representation this result is illustrated in Fig 3. From graph, LBG clustering activity recognition rate is high compare to K-mean clustering. Activity recognition rate of LBG clustering gives higher result compare to K-mean clustering on PC base feature also.

Table 2. Mean Activity Recognition Rate of IC and PC base approach using LBG [19]

Codebook Size	Mean Activity Recognition Rate	
	PC Based Approach	IC Based Approach
8	78	84
16	79	84.50
32	90.50	96
64	89	94.50

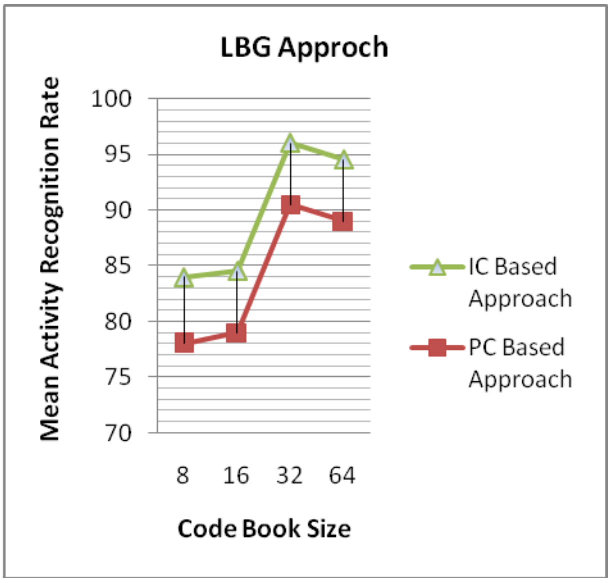

Fig 3: Comparison of IC and PC base approach [19]

In [20-22], depth camera is used for the activity recognition. PC and IC feature extraction approach give better result of activity recognition on depth silhouette image then binary silhouette image.

In [22], Activities such as Clapping, Boxing, Right hand-up-down, Left hand-up-down, both hands-up-down- are recognized. Both binary and depth silhouette image sequences are used for activity recognition. For all the approaches of feature extraction (PC and IC) and classification (LDA), depth map images give the better result. From [22], mean activity recognition rates of binary and depth silhouette approaches are shown in Table 3. Graphical representation this result is illustrated in Fig 4. From the graph, Depth silhouette images can give better result to recognize various human activities compared to binary silhouette images.

In [22], result of PC base, IC base, LDA on PC and LDA on IC feature extraction on binary and depth silhouette are shown in table. For all the activities binary images give less result compare to depth images.

Table 3. Activity recognition rate of binary and depth approaches. [22]

Approach	Mean Activity Recognition Rate	
	The Binary Silhouettes	The Depth Silhouettes
PCA	36.50	85.50
LDA on the PC features	36	86.50
ICA	43	93.50
LDA on the IC features	45.50	99

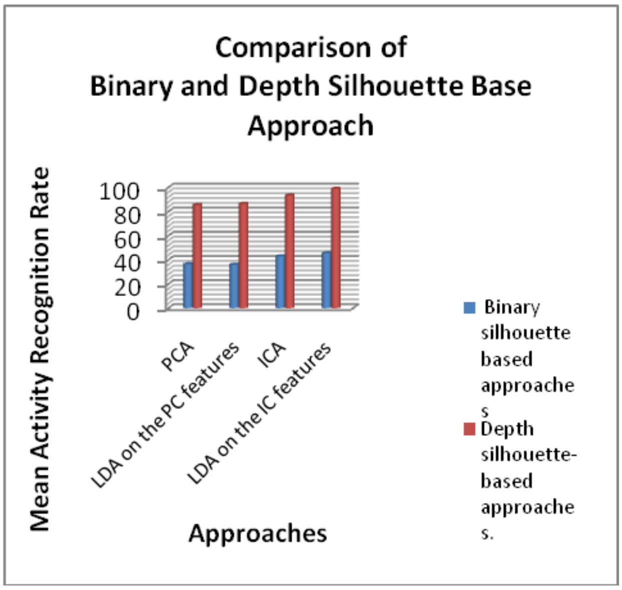

Fig 4: Comparison of binary and depth silhouette base approach [22]

4. CONCLUSION

In this Survey paper, the different approaches which are already proposed for human activity recognition using HMM are described. Different techniques are applied at each step of the approach for activity recognition. Depth silhouette images can give better result to recognize various human activities compared to binary silhouette images. IC based feature extraction gives better result compare to PC based extraction. By changing the parameter of the HMM, better result can be obtained.

5. REFERENCES

[1] Niu F, Abdel-Mottaleb M , "View-invariant human activity recognition based on shape and motion features," In: Proceedings of the IEEE sixth international symposium on multimedia soft-ware engineering, pp 546–556 , 2004.

[2] Niu F, Abdel-Mottaleb M, "HMM-based segmentation and recognition of human activities from video sequences," In: Pro-ceedings of IEEE international conference on multimedia & expo, 804–807, 2005.

[3] Robertson N, Reid I, "A general method for human activity recognition in video," Computer Vision Image Underst 104(2):232–248, 2006.

[4] Yamato J, Ohya J, Ishii K , "Recognizing human action in time-sequential images using hidden Markov model. In: Proceed-ings of IEEE international conference on computer vision and pat-tern recognition, pp 379–385, 1992.

[5] Nakata T, "Recognizing human activities in video by multi-resolution optical flow," In: Proceedings of international confer-ence on intelligent robots and systems, pp 1793–1798, 2006

[6] Sun X, Chen C, Manjunath BS , "Probabilistic motion para-meter models for human activity recognition," In: Proceedings of 16th international conference on pattern recognition, pp 443–450

[7] Kwon W, Lee TW, " Phoneme recognition using ICA-based feature extraction and transformation," Signal Process 84(6):1005– 1019, 2004

[8] Lee SI, Batzoglou S, "Application of independent compo-nent analysis to microarrays," Genome Biol 4(11):R76.1–21, 2003

[9] Makeig S, Bell AJ, Jung TP, Sejnowski TJ () Independent component analysis of electroencephalographic data. Adv Neural Inf Process Syst 8:145–151, 1996

[10] Jung T, Makeig S, Westerfield M, Townsend J, Courchesne E, Se-jnowski TJ, "Analysis and visualization of single-trial event-related potentials," Hum Brain Mapp 14:166–185, 2001

[11] Elgammal DH, Davis L, "Non-parametric model for back-ground subtraction," In: 6th European conference on computer vi-sion, Dublin, Ireland, 2000

[12] Kwak K-C, Pedrycz W, "Face recognition using an enhanced independent component analysis approach," IEEE Trans Neural Netw 18(2):530–541, 2007

[13] Kanungu T, Mount DM, Netanyahu N, Piatko C, Silverman R, Wu AY, "The analysis of a simple k-means clustering algo-rithm," In: Proceedings of 16th ACM symposium on computational geometry, pp 101–109, 2000

[14] Linde Y, Buzo A, Gray R, "An algorithm for vector quantizer design," IEEE Trans Commun 28(1):84–94, 1980

[15] Baum E, Petrie T, Soules G, Weiss N, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," Ann Math Stat 41:164–171, 1970

[16] Baum E, Eagon J, " An inequality with applications to sta-tistical estimation for probabilistic functions of Markov processes and to a model for ecology," Am Math Soc Bull 73:360–363, 1967

[17] Iwai Y, Hata T, Yachida M, "Gesture recognition based on subspace method and hidden Markov model," In: Proceedings of the IEEE/RSJ international conference on intelligent robots and systems, pp 960–966, 1997

[18] Uddin, M. Z., Truc, P. T. H., Lee, J. J., Kim, T.-S., "Human Activity Recognition Using Independent Component Features from Depth Images," In: Proc. of the 5th International Conference on Ubiquitous Healthcare, pp. 181—183, 2008

[19] Uddin, M. Z., Lee, J. J., Kim, T.-S., "Independent shape component-based human activity recognition via Hidden Markov Model," Journal of Applied Intelligence (Springer). DOI: 10.1007/s10489-008-0159-2., Netherlands,2009

[20] Md. Zia Uddin, Tae-Seong Kim, Jeong Tai Kim, "Video-based Human Gait Recognition Using Depth Imaging and Hidden Markov Model" A Smart System for Smart Home: 3rd International Symposium on Sustainable Healthy Buildings; Seoul, Korea, 27 May 2010

[21] Ahmad Jalal, Md. Zia Uddin, Jeong Tai Kim, Tae-Seong Kim, "Recognition of Human Home Activities via Depth Silhouettes and R-Transformation for Smart Homes," 5th International Symposium on Sustainable Healthy Buildings, Seoul, Korea, 10 February 2011

[22] Md. Zia Uddin and Tae-Seong Kim, "Continuous Hidden Markov Models for Depth Map-Based Human Activity Recognition" InTech, pp. 225-248.p, 19 April 2011

[23] http://www.3dvsystems.com

[24] Rabiner, L. R., "A tutorial on Hidden Markov Models and selected applications in speech recognition" In: Proc. of the IEEE, vol.77,no.2,pp.257—286,1989

[25] http://en.wikipedia.org/wiki/Independent_component_analysis