Non-temporal Multiple Silhouettes in Hidden Markov Model for View Independent Posture Recognition

Yunli Lee
Department of Media
Soongsil University
Seoul, South Korea
yunli@ssu.ac.kr

Keechul Jung
Department of Media
Soongsil University
Seoul, South Korea
kcjung@ssu.ac.kr

***Abstract*—This paper introduces a non-temporal multiple silhouettes in Hidden Markov Model (HMM) for offering view independent human posture recognition. The multiple silhouettes are used to reduce the ambiguity problem of posture recognition. A simple feature extraction of the 2D shape contour based histogram is used for image encoding and K-Means algorithm is applied for clustering and code-wording of eight simple postures from multiple views. Therefore, 3D volume reconstruction is not required, in return helps to reduce the complexity of modeling and computational power of feature extraction. HMM is trained to obtain view independent recognition model using multiple silhouettes. A combination of non-temporal multiple silhouettes, code-wording and HMM methods in this proposed approach make it possible to recognize human posture in view independent. The experimental results demonstrate the effectiveness of non-temporal multiple silhouettes in HMM for recognizing posture.**

Non-temporal multiple silhouettes; Hidden Markov Model; view independent; posture recognition

I. INTRODUCTION

Human gesture or action leads to the most active research in computer vision community by offering several important applications in our daily life, such as surveillance system, human-computer interfaces, games, virtual reality and others. These applications enable us to use hand or body gesture to convey information to the system or to control the device naturally without additional modal.

However, human body is not a rigid object and we could present various kinds of gesture that composed by a sequence of postures from each individual. Therefore, posture is a basic element of human gesture or action. The goal of our research is to recognize human posture in view independent without restricting the actor or observer position. Simple vision processing and feature extraction supply an intuitive 2D shape contour based histogram, and K-Means algorithm clusters the set of training posture silhouette images. Then, the code-wording of eight simple postures are trained by using the Hidden Markov Model (HMM) which is supporting view independent posture recognition.

In order to offer view independent posture recognition, several approaches have been proposed. Feiyue et al. [1] used stereo cameras of image pairs to propose a viewpoint insensitive representation for action recognition. Each action is represented in envelop shape, where this representation is viewpoint insensitive under assumption of affine camera projection model. Cen Rao [2] did a lot of related works on view invariant analysis for human activity in his Ph.D. study. He used trajectory of hand centroid to describe human action that are performed by hand. Both Feng Niu and Mohamed Abdel-Mottaleb [3], and Mohiuddin Ahmad and Seong-Whan Lee [4], used multiple view image sequences of combined features in Hidden Markov Model (HMM) for human action recognition. They built HMM model for each action in each viewing direction to characterize the variation from the change of viewing direction. However, this approach requires more HMM models of each action at each viewpoint for view independent characteristic. Liu Ren et al. [5] used extracted silhouettes from three video streams to match the body configuration of the user against a database of dance motion.

Different from previous works, we use synthetic models in virtual environment that relies on silhouettes extracted from seven virtual cameras to generate posture information in various view. These clean non-temporal multiple silhouettes are extracted automatically and clustered using K-Means algorithm for code-wording purpose. The order of cameras and code-wording are used to train the HMM for posture recognition model where it can handles view independent issue.

Section 2 describes the overview of our proposed HMM-based view independent posture recognition system using multiple silhouettes. We briefly explain the feature representation and extraction of silhouette images in Section 3. A simple K-Means classification used to create the code book is presented in Section 4. In Section 5, we introduce the concept of HMM and how it is adapted to offer view independent posture recognition. The experimental results of posture recognition are discussed in Section 6. Some thoughts about future work and conclusion are remarked in Section 7.

II. PROPOSED SYSTEM ARCHITECTURE

The system architecture of this proposed system consists of four parts. It contains a virtual studio of 3D model posture capturing module which automatically generates a sequence of non-temporal multiple silhouette images from seven virtual cameras; a 2D shape representation and extraction

978-0-7695-3521-0/09 $25.00 © 2009 IEEE
DOI 10.1109/ICCET.2009.113

module that extracts the contour points and computes the gradient point accumulation based on histogram bins template from silhouette image; a K-Means classification module that creates code-wording for non-temporal multiple silhouette images of defined postures for training the HMM models; and a posture recognition module which uses Baum Welch and forward algorithm [6-7] for training and evaluating the input sequence of non-temporal multiple silhouettes. The overview of the proposed system is illustrated in Figure 1.

In the virtual studio of 3D model posture capturing module, seven virtual cameras are used to capture the 3D model posture. Seven units of camera are placed around the ceiling view. This module automatically generates seven clean silhouettes from each camera view. Each posture is represented with seven non-temporal views in clockwise order where the cameras' position is about 51 degree apart. The contour points are extracted from the silhouette image and the gradient of each point on the contour to the center point is computed and accumulated into 12 bins template. A symbol which corresponds to a code-word in the code book created by K-Means is assigned to each silhouette image. A sequence of code-word is formed by seven camera's view and used to train the HMM models. Without restricting the start position of the camera view and actor, six times permutation of the sequence training dataset is applied. Baum Welch algorithm [6-7] is used to obtain the HMM parameters from the training dataset.

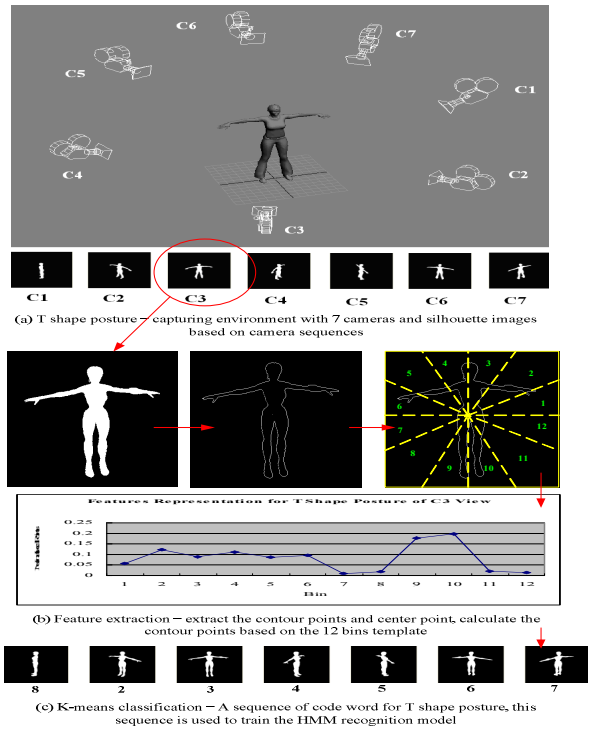

Figure 1. Overview of HMM-based view independent human posture recognition system.

A. *Multiple Views Posture Databases*

Our multiple views posture database comprises of eight postures as follows

1. T Shape,
2. Stand,
3. Both Hands Front,
4. Both Hands Up,
5. One Hand Up,
6. T Shape Foot Up,
7. T Shape Bend,
8. Crawl.

Each posture has been acquired from seven viewing position with 51 degree apart. Each posture was executed by seven synthetic 3D models for six times. The sequence of camera order is important to represent the posture information in view independent. However, we do not fix the starting position of the camera sequence. As long as the cameras sequence order is concerned, the starting position does not affect the learning performance. The database comprises 2352 images. As a whole, eight HMMs are created to model eight postures that support view independent using minimum number of cameras for training.

III. FEATURE EXTRACTION

We assume clean silhouettes are extracted from the system. A simple posture representation of silhouette image is used to extract the feature vectors. In this work, we extract the posture contour from each silhouette, while the center point of the posture contour is computed as a reference center point of 12 bins template. The gradient between each contour point and center is computed, and each bin from the defined template accumulates the number of contour points that overlaid on the corresponding bin. The feature named as gradient point accumulation is extracted from contour points of silhouette image. Let, $S_v = \{p_1, p_2, \ldots, p_n\}$ be the contour of silhouette image of a posture for each camera, where v is a camera index, $p_i = (px_i, py_i)$ is a contour point, and n is the total number of contour points, which is extracted using boundary following algorithm. The center point $c_{x,y}$, of the contour is obtained by the following equation from Yiğithan Dedeoğlu et al. [10]:

$$c_{x,y} = \left(\frac{\sum_{i=1}^{n} px_i}{n}, \frac{\sum_{i=1}^{n} py_i}{n} \right) \quad (1)$$

A 12 bins template is defined with 30 degree apart from each bin. Then, the gradient G, between the center point $c_{x,y}$ of contour with respect to each contour point on the contour is computed using below equation:

$$G(p[i], c_{x,y}) = \arctan\left(\frac{py_i - c_y}{px_i - c_x}\right), \forall i \in [1, \cdots, n] \quad (2)$$

Human varies in shape and size. As such, we propose a non-temporal multiple silhouettes to represent a posture because each silhouette may have a different contour shape.

Therefore, normalization is required to standardize the extracted features. The normalized accumulation of gradient point for 12 bins template, $B = \{b_1, b_2, \ldots, b_{12}\}$ is shown in Figure 2. B is bin with 12 dimensions feature vectors that are used to classify the contour shape of each silhouette image.

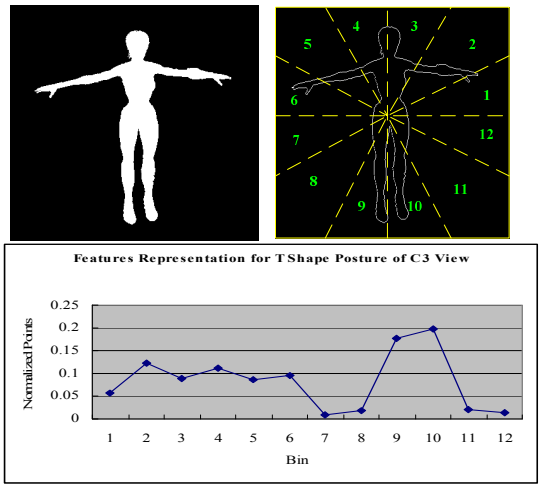

Figure 2. Bins distribution template and 2D shape representation graph of T Shape.

IV. MAPPING FEATURE TO SYMBOL

We propose seven multiple silhouettes to represent a posture, where each silhouette consists of 12 dimension feature vectors. We map each silhouette image into code-wording representation that forms a sequence of symbols for a posture in seven viewpoints using K-Means clustering. This result in a sequence of symbols represented as a posture and used to train the HMM model subsequently. Hence, a simple K-Means algorithm is applied to cluster the training feature vectors posture into K clusters, which equals to the size of code book. Consequently, the size of K clusters equivalent to the number of HMM output symbols, M.

A. K-Means and Code-wording

Simple K-Means algorithm [8] was introduced for clustering, where K is the number of cluster that cluster the data using means information from given dataset. For the silhouettes clustering, we initialized the means value of K corresponds to the defined postures. Since K is equivalent to M, these K symbols are used to generate code book for assigning a sequence of code-wording symbols for defining the input postures dataset to train the HMM models. Each silhouette is transformed into the symbol which is assigned to the cluster which is nearest to the means vector of the feature space.

For posture recognition, we select non-temporal multiple silhouettes of feature vectors for representative camera views from each posture as code-word in the code book. The extracted feature vectors would be mapped to a symbol, which code-word of the most similar (minimal error) feature vectors in the code book. Figure 3 shows the mapping result for the defined postures from the clustering result of K-Means, where K is defined as 56 by assuming the training dataset of eight postures has seven different views for each posture. The clustering result of simple K-Means is reasonable even though some postures are mistaken.

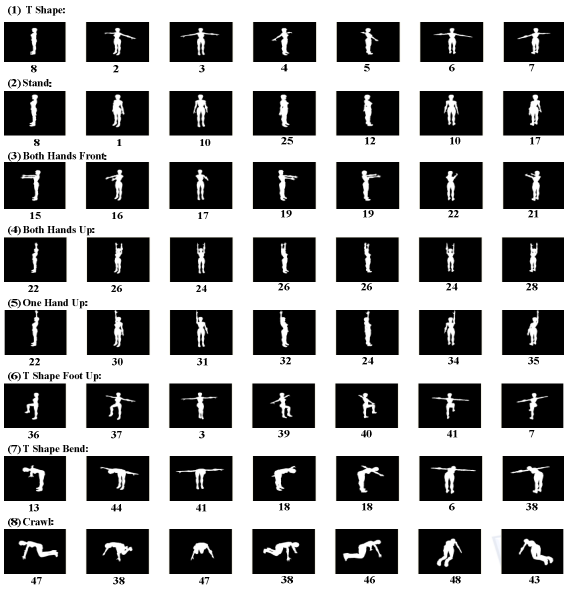

Figure 3. The defined postures and symbols of K-Means clustering.

V. HIDDEN MARKOV MODEL

Hidden Markov Models (HMM) is a probabilistic state transit model in which the system being modeled is assumed to be a Markov process where the states sequence could not be observed directly. There are widespread approaches to probabilistic sequence modeling [6-7]. HMM is well known as transition system which is able to attempt image understanding through model of patterns. It can deal with transitions in time, and also transition through another pattern. Thus, by taking this advantage of transition through another pattern, we learn a sequence of non-temporal multiple silhouettes from different views of multiple cameras for a defined posture that offers view independent human posture recognition system.

The basic parameters of HMM with N states with M observation symbols can be specified by $\lambda = (A, B, \pi)$ where λ represents the HMM model. The detailed explanation of HMM model parameters are as below:

- N is the number of states in the model where the states are hidden
- M is the number of observation symbols correspond to the physical output of the system being modeled
- A is a state transition probability matrix
- B is an observation probability for each state

- π is an initial state distribution

We propose only eight HMMs for recognizing the eight defined posture where each model can offer view independent by using a sequence of non-temporal silhouette images from multiple cameras. The set of parameters of HMM λ_j , where j is number of HMMs of the proposed recognition system, is learned by using Baum-Welch algorithm [6-7] from the training dataset. To recognize a given posture O, we evaluate $P(O \mid \lambda_j)$ with forward algorithm [6-7] for each j models and choose the maximum probability of model c to recognize the posture:

$$c = \arg\max_j P(O \mid \lambda_j) \quad (3)$$

A. View Independent HMM

View independent posture recognition enables user to perform posture command naturally without restricting the observation position. We have chosen seven cameras to be placed around the working space for capturing the posture. Non-temporal sequence of images is ordered in clockwise-order and six times permutation of view sequences has been applied to represent a posture where the user can perform the posture freely without any restriction.

These sequences are then used for training in HMM. The sequence of cameras is important for HMM to understand the posture in various views. HMM is able to learn the relation between each view image of each observed camera in order to support the view independent condition.

VI. Experimental Results

The proposed framework has been tested using synthetic dataset. The system has been trained using training data from seven 3D models with different size and gender. The models have been obtained from the 3D model database [9] and Poser software application. Eight postures as defined in Section 2.1 have been created using biped bone of 3D max studio application for 3D models.

We have used synthetic data to evaluate the accuracy of proposed system. The synthetic postures dataset are divided into two sets, one for training dataset, and another one is the testing dataset. Table 1 shows the confusion matrix obtained with five states of HMMs that learned from seven synthetic 3D models. The mean accuracy of the proposed recognition model is 97.7%.

TABLE I. Confusion Matrix for Postures Training Models

Posture	1	2	3	4	5	6	7	8
1	49	0	0	0	0	0	0	0
2	0	47	0	2	0	0	0	0
3	0	0	49	0	0	0	0	0
4	0	0	0	42	7	0	0	0
5	0	0	0	0	49	0	0	0
6	0	0	0	0	0	49	0	0
7	0	0	0	0	0	0	49	0
8	0	0	0	0	0	0	0	49

Another testing dataset has been created using two new 3D synthetic models. The postures were captured two times in same camera views and six times in new camera views. Figure 4 shows two sets of test data for T Shape posture which are generated from the new models in different camera views and classified to obtain input observation symbols. Table 2 shows the recognition results of testing dataset. The mean accuracy of this proposed recognition model is 89.8%.

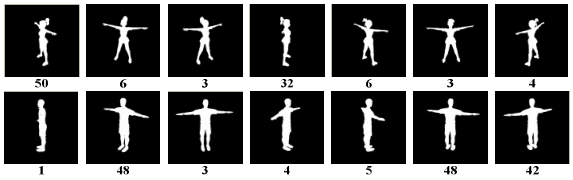

Figure 4. Two sets of new synthetic models with different camera views of T Shape posture.

TABLE II. Confusion Matrix for Posture Testing Models

Posture	1	2	3	4	5	6	7	8
1	14	0	0	0	0	2	0	0
2	0	15	1	0	0	0	0	0
3	2	1	13	0	0	0	0	0
4	2	0	0	14	0	0	0	0
5	0	0	0	0	16	0	0	0
6	4	0	0	0	0	12	0	0
7	0	0	0	0	0	1	15	0
8	0	0	0	0	0	0	0	16

VII. Conclusion and Future Work

This work proposes an efficient 3D human model posture recognition using non-temporal multiple silhouettes without any 3D volume reconstruction process. The contour points of silhouettes are extracted to represent the human posture. Each posture is composed from seven multiple contours of silhouette views which are represented in gradient point accumulation feature. The gradient point accumulation is allocated into 12 bins template. A set of training defined postures are clustered into K symbols that corresponds to a code-word of code book created by K-Means classification. Seven sequences symbol of code-wording expressing a defined posture. The recognition rate achieves average 97.7% of training models and 89.8% of new test models. The HMM-based non-temporal multiple silhouettes show a potential result for view independent human posture recognition.

For future work, we plan to apply the non-temporal multiple silhouettes of contour point features in HMM for view independent posture recognition in real environment. More postures definition and test data are required for evaluating the proposed features and HMM-based multiple silhouettes on the robustness of view independent human posture recognition system.

ACKNOWLEDGMENT

This work was supported by the 'Seoul R and BD Program (10581cooperateOrg93112)'.

REFERENCES

[1] Feiyue Huang, Huijun Di, Guangyou Xu, "Viewpoint Insensitive Posture Recognition for Action Recognition," Proceedings of Articulated Motion and Deformable Objects (AMDO 2006), LNCS, Vol. 4069, pp. 143-152, 2006.

[2] Cen Rao, "Invariance in Human Action Analysis," Ph.D. Dissertation of University of Central Florida, 2003.

[3] Feng Niu, Mohamed Abdel-Mottaleb, "View-Invariant Human Activity Recognition Based on Shape and Motion Features," IEEE Sixth International Symposium on Multimedia Software Engineering, pp. 546-556, 2004.

[4] Mohuiddin Ahmad, Seong-Whan Lee, "HMM-based Human Action Recognition Using Multiview Image Sequences," 18th International Conference on Pattern Recognition (ICPR 2006), Vol. 1, pp. 263-266, 2006

[5] Liu Ren, Gregory Shakhnarovich, Jessica K. Hodgins, Hanspeter Pfister, and Paul Viola, "Learning Silhouette Features for Control of Human Motion," ACM Transactions Graphics, Vol. 24, No. 4, pp. 1303-1331, October 2005.

[6] Lawrence R. Rabiner, "A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," Proceedings of the IEEE, vol. 77, no. 2, pp. 257-286, 1989.

[7] Lawrence R. Rabiner, B.H. Juang, "An Introduction to Hidden Markov Models," IEEE ASSP Magazine, vol. 3, pp. 4-16, 1986.

[8] Vitorino Ramos, Fernando Muge, "Map Segmentation by Colour Cube Genetic *K*-Mean Clustering," Proceedings of the ECDL, 2000.

[9] INRIA Gamma team research database, http://www-c.inria.fr/gamma/gamma.php

[10] Yiğithan Dedeoğlu, B. Uğur Töreyin, Uğur Güdükbay, A. Enis Çetin, "Silhouette-Based Method for Object Classification and Human Action Recognition in Video," 9th European Conference on Computer Vision (ECCV 2006), LNCS, vol. 3979, pp. 64-77, 2006.