HMM-Based Gait Recognition with Human Profiles

Heung-Il Suk and Bong-Kee Sin

Computer Engineering, Pukyong National University
{daedalos, bkshin}@pknu.ac.kr

Abstract. Recently human gait has been considered as a useful biometric supporting high performance human identification systems. We propose a view-based pedestrian identification method using the dynamic silhouettes of a human body modeled with the hidden Markov model (HMM). Two types of gait models have been developed both with a cyclic architecture: one is a discrete HMM method using a self-organizing map-based VQ codebook and the other is a continuous HMM method using feature vectors transformed into a PCA space. Experimental results showed a consistent performance trend over a range of model's parameters and the recognition rate up to 88.1%. Compared with other methods, the proposed models and techniques are believed to have a sufficient potential for a successful application to gait recognition.

1 Introduction

Recognizing people by their gait, the style of walking of an individual, can be performed without asking them to take any specific actions and even without making them be aware whether they are being watched or not. From Johansson's studies in psychophysics with moving light displays (MLD) attached to body parts, it appeared that humans have the capability of recognizing their acquaintance only through their gait [1].

Recently human motion analysis has been receiving increasing attention from computer vision researchers and it is well explained in the review papers by J. K. Aggarwal et al. [2], C. Cedras et al. [3], and D. M. Gavrila et al. [4]. According to these papers two distinct methods for human motion analysis are distinguished: 'model-based method' using a priori shape models, and the other, called 'appearance-based' or 'view-based', without using explicit shape models. Both methods take a common sequential process of (1) feature extraction, (2) feature correspondence, and (3) high-level processing. The difference between them lies in the way of processing feature correspondence. 'Model-based' methods compare the input features taken from an input image with the parameters of the 2D or 3D body models prepared in advance, and make feature correspondence automatically. On the contrary, the 'appearance-based' methods carry out the feature correspondence by varying the values of position, velocity, shape, color, and so on from consecutive frames.

A brief review is in order. In the works of A. Kale et al. [5, 6], they used the width of the pedestrian's silhouette of the binarized images as the feature vector and developed

five basic stances from k-means clustering. Then a new feature vector composed of the Euclidean distance between an input feature vector and each of the 5 basic stances was the final feature vector for training HMM and recognizing people. On the other hand, J. J. Little et al. [7] extracted the frequency and phase of the gait derived from optical flow as the feature vector. In [8], C. BenAbdelkader et al. encoded the planar dynamics of a walking person in a 2D plot consisting of the pairwise image similarities of the sequence of images of the person and proposed a recognition method using k-nearest neighbor after reducing the dimension of the feature vector. Whereas R. Collins et al. [9] introduced a view-dependent method using template matching of body silhouettes. The key frames from a test sequence were extracted by performing cyclic gait analysis and those frames were compared to training frames using normalized correlation. The recognition was performed by nearest neighbor matching among correlation scores.

In this paper, we use people's silhouette as a profile vector and represent the characteristics of the gait with the sequences of the profile vectors. Since body shape alone is not enough for gait recognition, we need to take account of the gait dynamics which can be modeled by a 'Markov chain'. A hidden Markov model or HMM is a Markov chain variant that is very powerful for modeling highly variable and noisy patterns. We have chosen this model for gait recognition. In this work, the profile vector used for the recognition is composed of a fixed number of horizontal distance measurements of the left, right boundary of a person from the center and then normalized with respect to the height of the person. This feature can be directly compared with the features of A. Kale et al. [5, 6] and it has turned out that the features used in our work showed better performance. We have designed two separate recognizers, each using discrete hidden Markov models (DHMM) and continuous hidden Markov models (CHMM) respectively. For the training of DHMMs, each feature vector is quantized to a codeword by a self-organization map (SOM). For the training of CHMMs, feature vector is mapped to a point in the PCA space to reduce the dimension. In this way we kept the models from being overfitted due to the lack of training data.

The contribution of this paper is found in the way of representing the structural characteristics and the modeling dynamic characteristics of the human gait. The structural characteristics are represented by the left and right boundaries of the person and the dynamical characteristics are well models by an HMM. The HMM has one-way directed circular ring topology fitted to recognizing a person in a video sequence without external segmentation into a single gait cycle. This type of an HMM does not require a specific starting point in a gait cycle. Another benefit of this model we believe is that the longer the input sequences are, the better the recognition rate is.

2   Profile Extraction

One of the simplest and most direct ways to represent the shape of a pedestrian is the silhouette against the background. The silhouette can be described as an ordered

sequence of boundary points and we define it as a profile vector that represents the shape characteristics of the person. A profile vector is composed of the horizontal distances from the horizontal center to the left and right boundary of the human blob B (Fig.1). The horizontal center of the person is calculated as

$$X_c = \frac{\sum_{(x,y)\in B} x}{|B|} \tag{1}$$

where, B is a set of foreground image pixels, X_c is the horizontal center of B, $|B|$ is the number of pixels and (x, y) is the coordinate of the pixel in the image. If we consider all the boundary pixels of the silhouette in composing a profile vector, its dimension of the vector will be too high and variable from one frame to another. So, we choose only 40 pixels of the silhouette of the human blob, 20 pixels from the left and 20 pixels from the right sampled at a fixed vertical interval from the bottom. The elements of the vector are then normalized with respect to the height of the human blob.
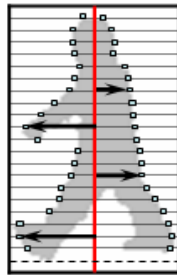

Fig. 1. Composition of a profile vector

3   Model of Human Gait

In this paper, we create two kinds of gait models using the discrete hidden Markov model (DHMM), and the continuous hidden Markov model (CHMM), to represent the dynamic characteristics of a human gait. The input data for the DHMM are the codeword sequences quantized from the profile vector sequences by a self-organization map (SOM) and those for the CHMM are the sequence vector transformed into a subspace of PCA. We first explain the vector quantization for the DHMM and then the PCA for the CHMM.

3.1   SOM

SOM is a topology preserving feature map. It being simple, we chose this for a vector quantizer of the feature vectors. Taking account of the fact that a human gait is composed of a cyclic sequence of stereotypic stances, we designed the output layer of the SOM to have a ring topology with each state corresponding to a typical stance in the

sequence. The SOM is trained with the profile vectors extracted from the input image sequences. The set of trained weight vectors are considered as a codebook for the gait pattern space. With this SOM we quantize each input profile vector into a codeword.

$$codeword = \arg \min_{k} \|\mathbf{x} - \mathbf{w}_k\|. \tag{2}$$

Here $\|\cdot\|$ is the Euclidean norm of the difference between the input profile vector and the code vector, $\mathbf{x}$ is an input vector and $\mathbf{w}_k$ is the codeword vector. We choose the codeword whose code vector has the minimum Euclidean distance from the input profile vector.

3.2 Principal Component Analysis

The principal component analysis is the method that can be used to reduce the dimension of any vectors by considering the variance of and the relationship among variables, while minimizing the loss of the information. It reduces the dimension of a vector by transforming it into the direction that has a large variance. This is originated from the fact that there is more information in the direction having a large variance compared to the direction having the small variance. We use this method to reduce the dimension of the profile vectors for the training of the CHMM.

It is known that the sum of the eigenvalues of the covariance matrix is equal to the total variance of the original variables. We can consider a mapping to a reduced dimension by specifying that the new components must account for at least a fraction of d of the total variance. Choosing a value of d, which is a fraction of the total variance, between 70% and 90% preserves most of the information in $\mathbf{x}$ [10]. We then choose k so that

$$\sum_{i=1}^{k} \lambda_i \geq d \sum_{i=1}^{p} \lambda_i \geq \sum_{i=1}^{k-1} \lambda_i \tag{3}$$

where, λ_i denotes the eigenvalues and p is the rank of the covariance matrix.

3.3 Gait Model

Recognizing people using structural features of body silhouette is not easy since simple structural characteristic can be shared by many individuals. We can exploit the observation that people have their own dynamic characteristics or temporal features, even though they have similar structural characteristics. These dynamic characteristics can be represented by the Markov chain in an HMM.

The transition of the states in an HMM is modeled by the 'Markov process'. The sequence of the transition of the states is not observable, i.e. hidden, and it is possible to estimate it through an observable data.

In this paper, we design the topology of an HMM as a circular ring (Fig. 2) considering that human gait is periodic. The parameters of the DHMM are trained using
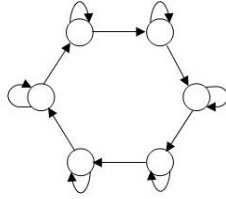
Fig. 2. Topology of a gait HMM

the codeword sequences quantized by the SOM and those of the CHMM are trained using the reduced vectors transformed into a PCA space.

Since the HMMs are trained for each person in the data set, it can be used as a way to represent the characteristics of the specific person's gait. In other words, when there is an input vector sequence taken from the video sequences of a person's gait, the HMM which best represents the characteristics of the person's gait will produce the highest likelihood for the input vector sequence. The likelihood of each HMM is computed by the forward algorithm.

$$\hat{i} = \arg \max_i P(\mathbf{X} | \lambda_i). \tag{3}$$

Here, $\mathbf{X}$ is the input vector sequence, λ_i is the HMM model for the ith person.

4   Experiment and Analysis

The proposed method was tested on a video database consisting of seven sequences for each of the six subjects, taken from the web site of the University of Calgary in Canada [11]. Those video sequences were captured at 15 frame rate, 24 bit colors and an image size of 320×160. Each sequence includes 85 frames and 6 gait cycles on average.

4.1   Data Coding

Fig. 3 shows the example of the vector quantization for a half cycle of a person's gait using an SOM with seven output nodes. The quantized vector sequences are used for the DHMM training.

A profile vector contains the 40 boundary points as marked in Fig. 4(a). Each of the elements is the horizontal distance from the vertical center to the left and right boundary of a human blob and is numbered in clockwise starting from the left bottom (Fig. 4(a)). The covariance matrix of all the profile vectors used for training motivates us to reduce the dimension. As shown in Fig. 4(b), only a small fraction of elements (covariances) are significant implying great variation and correlation. Those numbers in each axis indicate the nth element of a profile vector. It is conjectured that these bright colored parts correspond to key features to be used to recognize people. Those components are feet/legs (1-5, 35-40) and hands/arms (11-15) regions.
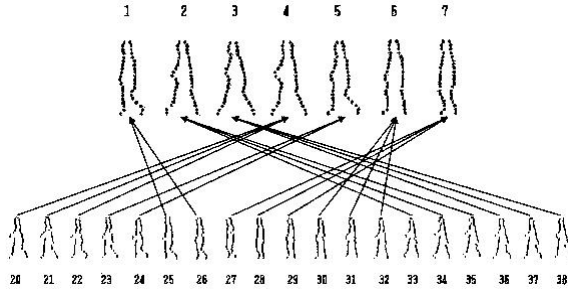
Fig. 3. Vector quantization example for half cycle

(a)   (b)
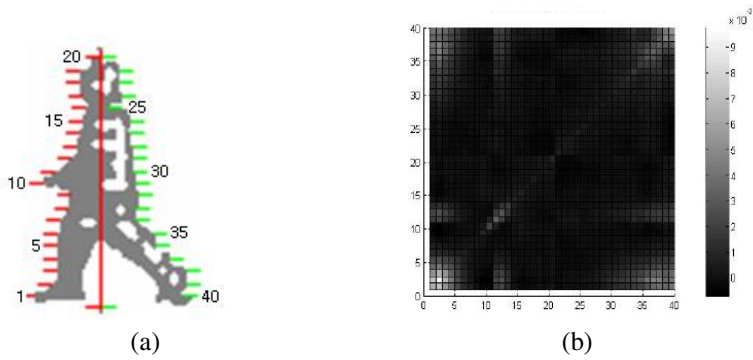
Fig. 4. (a) The elements of the profile vector (b) Visualization of the covariance matrix of the profile vectors

4.2 Recognition

Due to the lack of the test data, cross validation was used. In the DHMM test by varying the number of codeword and states, it showed the highest performance, 88.10%, with seven states and seven code vectors. The detailed results over a number of states with fixed codebook size are shown in Table 1.

Table 1. Results of DHMM (codebook size = 7)

| DHMM test					
# States	5	6	7	8	9
Hits(%)	40.48	47.30	88.10	69.05	0.0

For the test of the CHMM, we varied the reduced-dimension of each profile vectors. The highest performance was obtained with seven states and eight dimensions. The result with seven states is shown over a number of dimensions in Table 2.

Table 2. Results of CHMM (# of states = 7)

CHMM test					
Dimension	5	6	7	8	9
Hits(%)	76.19	73.81	69.05	88.10	76.19

4.3   Performance

In the third experiment, we compared the performance with those of the two other methods by J. J. Little et al. [7] and A. Kale et al. [5] using the same data. In the J. J. Little et al. research, they achieved a recognition rate of 90.5% by composing feature vectors with the frequency and phase taken from optical flow using the video sequences of twice as big as the frame size and the frame rate of our data. Direct comparison shows that J. J. Little et al reported higher performance than ours, but it was on different (or more detailed) data set. On the other hand, A. Kale et al's method achieved 64.3% on our data set with five HMM states. The comparison between our method and that of A. Kale et al. is shown in Fig. 5. Although we can not strongly argue for the statistical significance of the result, our system's performance was higher in all cases with a single candidate.
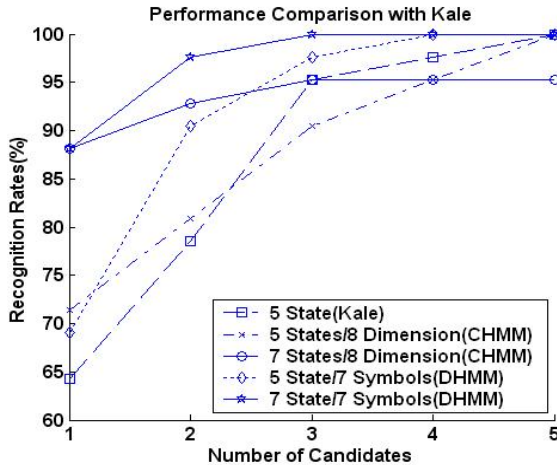
Fig. 5. Performance comparison with Kale's

5   Conclusion

We proposed an improved gait recognition method using HMMs. It added new features including SOM-based codebook and a profile vector representation for each frame, and the cyclic HMM topology for the gait dynamics. When tested on the video data from the web site of the University of Calgary in Canada by varying the number

of states, codebook sizes and the number of dimensions, we achieved the highest recognition rates with seven states in both the DHMM and the CHMM.

In this work, we used a background subtraction technique to extract a silhouette without considering clothes, illumination, camera angle and walking angle. And we used a small data set for test. These are the direction of immediate future work. One advantage of the model-based approach is that we can estimate the inner structure of the pedestrian from the SOM and HMM as shown in Fig. 6. This result can be used to confirm the silhouette shape and/or to enable further understanding of human motion of activity.

Fig. 6. Pre-estimated stick figure

References

1. G. Johansson.: Visual perception of biological motion and a model for its analysis, *Perception and Psychophysics*, vol. 14. no. 2 (1973) 201-211
2. J. Aggarwal and Q. Cai.: Human motion analysis - a review, *Computer Vision and Image Understanding*, vol. 73, no. 3 (1999) 428-440
3. C. Cedras and M. Shah.: Motion-based recognition - a survey, *Image and Vision Computing*, vol. 13, no. 2. (1995) 129-155
4. D.M. Gavrila.: The visual analysis of human movement - a survey, *Computer Vision and Image Understanding*, vol. 73. (1999) 82-98
5. A. Kale, A. Rajagopalan, N. Cuntoor, and V. Kruger.: Gait-based recognition of humans using continuous HMMs, *Proceedings of the IEEE Conference on Automatic Face and Gesture Recognition*, (2002) 321-326
6. A. Kale, A. N. Rajagopalan, A. Sundaresan, N. Cuntoor, A. RoyChowdhury, V. Krueger, R. Chellappa.: Identification of humans using gait, *IEEE Transactions on Image Processing*, September (2004)
7. J.J. Little and J.E. Boyd.: Recognizing people by their gait: the shape of motion, *Videre*, vol. 1, no. 2. (1998) 1-32
8. C. BenAbdelkader, R.Cutler, and L.Davis.: Motion-based recognition of people in eigen-gait space. *IEEE Conf Automatic Face and Gesture Recognition*, (2002) 254-259
9. R. Collins, R. Gross, and J. Shi.: Silhouette-based human identification from body shape and gait, *IEEE Conf Automatic Face and Gesture Recognition*, (2002) 351-356
10. Andrew R. Webb, *Statistical Pattern Recognition Second Edition,* John Wiley and Sons, 2002.
11. http://pages.cpsc.ucalgary.ca/~boyd/gait/experiment.html