Gait-based Recognition of Humans Using Continuous HMMs*

A. Kale[1], A.N. Rajagopalan[2], N. Cuntoor[1] and V. Krüger[1]

[1] Center for Automation Research
University of Maryland at College Park
College Park, MD 20742
[2] Department of Electrical Engineering
Indian Institute of Technology Madras
Chennai-600 036, India

Abstract

Gait is a spatio-temporal phenomenon that typifies the motion characteristics of an individual. In this paper, we propose a view based approach to recognize humans through gait. The width of the outer contour of the binarized silhouette of a walking person is chosen as the image feature. A set of stances or key frames that occur during the walk cycle of an individual is chosen. Euclidean distances of a given image from this stance set are computed and a lower dimensional observation vector is generated. A continuous HMM is trained using several such lower dimensional vector sequences extracted from the video. This methodology serves to compactly capture structural and transitional features that are unique to an individual. The statistical nature of the HMM renders overall robustness to gait representation and recognition. Human identification performance of the proposed scheme is found to be quite good when tested in natural walk conditions.

1 Introduction

The need for automated person identification is growing in many applications such as surveillance, access control and smart interfaces. It is well-known that biometrics are a powerful tool for reliable automated person identification. Established biometric-based identification techniques range from fingerprint and hand geometry methods to schemes like face recognition and iris identification. However, these methodologies are either intrusive or restricted to very controlled environments. For example, current face recognition technology is capable of recognizing only frontal or nearly frontal faces. When the problem of person identification is attempted in natural settings, such as those that occur in the automatic surveillance of people in strategic areas, it takes on a new dimension. Biometrics such as fingerprint or iris are then no longer applicable. Furthermore, night vision capability (an important component in surveillance) is usually not possible with these biometrics. A biometric that can address some of these shortcomings is the human 'gait'. The attraction of using gait as a biometric is that it is nonintrusive and typifies the motion characteristics specific to an individual. It is a well-known fact that people often recognize others by simply observing their gait which may justify using it as a cue for recognizing people from a small database. However, if the database is large, then gait information, by itself, may not be sufficient to discriminate each individual. But it still makes good sense to use gait as an indexing tool to greatly narrow down the search for potential targets. Early medical studies [7] suggest that if all movements are considered, gait is unique. In all, it appears that there are 24 different components to human gait. However, from a computational perspective, it is quite difficult to accurately extract these components. Precise extraction of body parts and joint angles in real visual imagery is a very cumbersome task and can be unreliable. Hence, the problem of representing and recognizing gait turns out to be a challenging one. A careful analysis of gait reveals that it has two important components, a structural component which captures the physical build of a person and a dynamic component which captures the transitions that the body undergoes during a walk cycle. If one could effectively capture these components, then it should be possible to recognize gait.

In this paper, we present a method that directly incorporates the structural and transitional knowledge about the identity of the person performing the activity. This knowledge is used to generate a lower dimensional observation vector sequence which is then used to design a continuous density HMM for each individual. In the next section we give an overview of the prior work in the area of activity and gait recognition. Section 3 motivates our approach. Section 4 describes our methodology for human recognition using gait. Section 5 describes our experimental results and sec-

*Supported by the DARPA/ONR grant N00014-00-1-0908.

tion 6 concludes the paper.

2 Prior Work

The task of recognizing people by the way they walk is an instance of the more general problem of recognition of humans from gesture or activity. We take a closer look at the relation between the problems of activity recognition and activity-specific person identification. A good review of the state of the art in activity recognition can be found in [1]. For human activity or behavior recognition most efforts have used HMM-based approaches [11, 12, 13] as opposed to template matching which is sensitive to noise and the variations in the movement duration. In [13], discrete HMMs are used to recognize different tennis strokes. In [11], continuous HMMs are used to recognize American sign language. In [12] a parametric continuous HMM has been applied for activity recognition. All these approaches involve picking a lower dimensional feature vector from an image and using these to train an HMM. Note that, if we choose a not too unreasonable set of features, the trajectories corresponding to distinct activities will be far apart in the vector space of the features. Hence, in principle, with a small degradation in performance, it is possible to replace the continuous approaches in [11, 12] by building a codeword set through kmeans clustering over the set of the lower dimensional observation vector space and using a discrete HMM approach as in [13]. The scenario is very different in the problem of recognition of humans from activity. Primarily, there is considerable similarity in the way people perform an activity. Hence, feature trajectories corresponding to different individuals performing the same activity tend to be much closer to one another as compared to feature trajectories corresponding to distinct activities. The aforementioned activity recognition approaches, if directly applied to human identification using gait will almost certainly fail in the presence of noise and structurally similar individuals in the database.

We now review some of the prior work done in the recognition of humans from gait. In [4], Huang et al. use optical flow to derive the motion image sequence corresponding to a gait cycle. The approach is sensitive to optical flow computation. Also, it does not address the issue of phase in a gait cycle. In another approach, Cunado et al. [3] extract gait signature by fitting the movement of the thighs to an articulated pendulum-like motion model. The idea is somewhat similar to the work by Murray [7] who modeled the hip rotation angle as a simple pendulum, the motion of which was approximately described by simple harmonic motion. Locating accurately the thigh in real image sequences can, however, be very difficult. Little and Boyd [5] extracted frequency and phase features from moments of the motion image derived from optical flow to recognize different people by their gait. As expected, the method is quite sensitive to the feature extraction process. Bobick and Johnson used static features for recognition of humans using gait [2]. Murase and Sakai [6] have also proposed a template matching method which is somewhat similar in spirit to the work reported in [4].

3 Our Approach

One of the issue that arises in the context of gait recognition is the viewing angle and invariance to thereof. It is reasonable to choose the viewing angle that yields maximum observable dynamics since the perceptible change in the structural gait information does not change significantly with viewing angle. We therefore analyze the side view of a person walking, allowing for minor angular variations.

A possible solution to gait representation/recognition lies in a closer examination of the physical process of gait generation. During a gait cycle, it is possible to identify certain distinct stances (Figure 1) that are generic, in the sense that every person transits between these successive stances as he/she walks. These stances partly encode identity information by virtue of the structural differences between people. In practice, an accurate time-stamping of these stances is impossible. A precise demarcation of when the image undergoes a transition from one stance to another is difficult. Hence, using structural information alone may not yield good discriminability.

There is a Markovian dependence from one stance to another. The gait cycle can be viewed as a doubly stochastic process in which the hidden process is represented by the transitions across the stances while the observable is the image generated when in a particular stance. The HMM is best suited for describing such a situation. Formally, a HMM is defined as a doubly stochastic process that is not directly observed but can only be studied through another set of stochastic processes that produce the given sequence of observations. Markovian transitions are assumed to occur between states and a random observation is output in a particular state. For a detailed discussion on HMMs and their applications see [9].

As described earlier, the gait cycle consists of distinct stances. It is our conjecture that these stances can be associated with the states of an HMM where the switch from one stance to another can be represented by transition probabilities between states.

4 Proposed Methodology

An important issue is the extraction of appropriate features that will capture the gait characteristics effectively. Intuitively, the silhouette of a person is a reasonable feature to look at as it captures the motion of most of the body parts and also encodes structural as well as transitional information. It is reasonably independent of the clothing worn by the person and it supports night vision capability as it can be easily derived from IR imagery. Successful training of the HMM depends largely on the dimension of the observation vector. Clearly, the silhouette information cannot be used as is due

to its large dimension. Compact encoding of the information contained in the silhouette is necessary for good performance. We now describe a procedure to efficiently encode this information. This is followed by a detailed description of training and evaluation modules.

4.1 Silhouette Extraction

In our experiments, the camera is assumed to be static and that only one person is within the field of view.

Given the image sequence of a subject, the silhouette is generated as follows:

1. Background subtraction is used to detect moving objects in each frame; subsequently a blob tracker tracks the fastest moving object in the scene, thereby reducing effects of minor disturbances in the background.
2. A standard 3×3 erosion filter is applied to the motion image to remove spurious noise.
3. Since we are interested in the outer contour of the body only, the left and right boundaries of the body are traced by examining the pixel intensities with a weighted low pass filter from leftmost and rightmost ends of the image.
4. The width of the silhouette along each row of the image is then stored. The width along a given row is simply the difference in the locations of rightmost and leftmost boundary pixels in that row.

Typical silhouette images extracted from a video sequence are shown in Figure 1. It may be noted that our silhouette extraction procedure is simple and straightforward. It is quite possible that the silhouette is sometimes not perfectly extracted. However, the advantage of using a statistical approach (such as the HMM) is that it is robust to such minor perturbations.

4.2 Training for Gait

Given an observation sequence, we seek a way to build a representation for the gait of every individual in the database. As discussed before we opt for the stochastic approach of using continuous HMMs. In this case, training involves learning the HMM parameters $\lambda = (A, B, \Pi)$ from the observation sequences. Here A denotes the transition probability matrix, B is the observation probability and Π is the initial probability vector. In order to capture the gait of an individual, we train the HMM using the width vectors derived from the silhouette for several gait cycles of the person. We express the pdf of the observation as

$$b_j(\mathbf{o}) = \mathcal{N}(\mathbf{o}; \mu_\mathbf{j}, \mathbf{U_j}), \quad 1 \leq j \leq N \tag{1}$$

where $\mathbf{o}$ is the observation vector and $\mu_\mathbf{j}$ N is the number of states in the HMM and $\mathbf{U_j}$ are the mean and covariance, respectively. The reliability of estimates of B depends on the number of training samples available and the dimension of the observation vector.

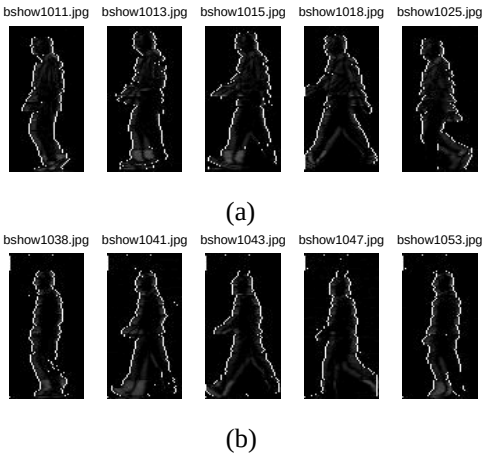

Figure 1. Five stances corresponding to the gait cycle of a) Person 1 and b) Person 2.

In a practical situation, only a finite amount of training data is available. Since the means and covariances in equation (1) have to be learnt from the training samples, the dimension of the observation vector becomes critical. The required number of training samples increases with the dimensionality of the observation vector. To be precise, assume for the moment that the data can be modeled by a single Gaussian distribution. Then, for a d-dimensional observation vector, we need at least d training samples to estimate the centroid and $\frac{d(d+1)}{2}$ training samples in order that the covariance matrix would have a well-defined inverse. In our experiments, the smallest dimension of the width vector of the silhouette is approximately 100. This implies that we require at least 100 observations to learn the mean value. To learn the covariance, we would need as many as 5150 vectors! For a mixture of m-Gaussian model, there would be a further m-fold increase in the number of training vectors. Clearly, the possibility of using the width vector directly is ruled out. A more compact way of encoding the observation, while retaining all relevant information, is needed. We propose the following methodology to tackle the dimensionality issue in the gait problem. To decide on the number of stances we plot the average rate distortion curve of the quantization error as a function of number of stances. From figure (2) we observe that the quantization error does not decrease appreciably beyond 5 stances. Let us denote the width vectors corresponding to the five stances for the jth person as $S_1^j, \cdots, S_5^j$. These stances are the ones that result from application of the k-means procedure to the training data available for that individual. The Euclidean distance between an observed width vector in frame k (denoted by $OW^j(k)$) and the lth stance is given by

$$||OW^j(k) - S_l^j|| \tag{2}$$

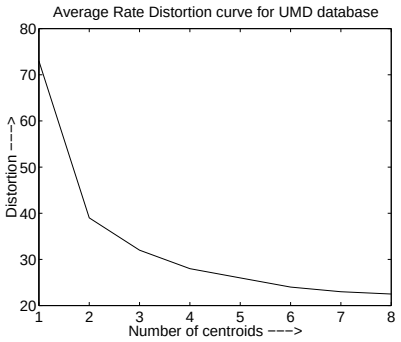

Figure 2. Average Rate Distortion Curve for UMD database

$O_j^i(k)$ represents the observation sequence of the person i encoded in terms of the stances of person j. Note that the dimension of $O_j^j(k)$ is only 5. The new 5-D vector, which is a measure of the similarity between the observed image and the five stances, has the following significance:

- Firstly, note that by virtue of self-similarity, the encoding of a width vector of person j in terms of the width vectors of the stances of person j will yield a lower Euclidean distance than when it is encoded in terms of the width vectors of person i. For instance, the five dimensional vector for a short person generated using the stances of a tall person will be large in magnitude.

- In addition, the manner in which every component of this vector evolves with time encodes the transitional information unique to a person. This transitional information could be the key factor that can distinguish between two individuals that are structurally similar.

4.3 Gait Recognition

The stances together with the HMM represent the gait of an individual. For robust recognition, it is reasonable that one must examine several walk cycles before taking a decision i.e., instead of looking at a single half walking cycle, it is beneficial to examine multiple half-cycles of a person before any conclusion about his/her gait can be reached. We assume that several walk cycles of an individual are available. The problem is to recognize this individual from a database of people whose gait models are available.

Given the image sequence of the unknown person X, the width vector OW^X of this person is generated. Using the stances $S_1^i, \cdots, S_5^i$ for person i in the database, we compute the Euclidean distance of the width vector OW^X of the unknown person w.r.t. the stances of person i to yield $O_i^X(k)$ for the kth frame. The likelihood that the observation sequence O_i^X was generated by the HMM corresponding to the ith person can be computed using the forward algorithm

$$P_i = \log(P(O_i^X|\lambda_i)) \tag{3}$$

where λ_i is the HMM model corresponding to the person i. We repeat the above procedure for every person in the database thereby producing probabilities $P_j, 1 \leq j \leq N$. Suppose that the unknown person was actually person m. If the values of $P_1, \cdots, P_N$ are observed for a sufficient number of half cycles of the person X, we expect that in a majority of cases P_m would be higher compared to the rest of the P_is. We shall present our results in a format similar to the FERET protocol [8].

5 Experimental Results

For our experiments, the video sequences were taken from the following databases:

1. Little and Boyd's database [5]. This has 5 people with around 22 walk cycles for each subject. Half of the cycles were used for training and the other half for testing. The data was collected by a camera mounted on a tripod. The number of pixels on target was about 100.

2. University of Maryland (UMD) database: It has 43 people walking in a T-shaped path. The data was collected by a surveillance camera mounted at a height of 15ft. It has two sequences collected on different days for each person , each with 10 cycles. One sequence was used for training and the other for evaluation. The number of pixels on target was about 150.

3. Carnegie Mellon University (CMU) database: It has 25 people walking at a fast pace and slow pace on a treadmill. There are about 16 cycles in each sequence. Half of the cycles were used for training and the other half for testing. The data was collected by a camera mounted on a tripod. The number of pixels on target was about 630.

It should be pointed out here that a walk cycle consists of two strides(or half cycles) strides. Since we use only the extremities of the silhouette the two halves of the walk cycle are almost indistinguishable. The cycles we mention for the databases are really the half cycles.

Training:

Silhouettes corresponding to a walk cycle are extracted for each person in the database using the silhouette extraction procedure described in Section 3.1. The width vector is generated for each frame and encoded as a compact 5-D observation sequence using the stances of that person. This lower dimensional vector sequence (possibly of varying length) constitutes a training sequence. We train a 5-state, single Gaussian, ergodic HMM for each person. As expected, the transition probabilities and the observation probabilities

$T\backslash R$	1	2	3	4	5
1	10			1	
2		10	1	0	
3		2	9		
4	0	2		9	
5					11

Table 1. Confusion Matrix

$T\backslash R$	1	2	3	4	5
2		7	3	1	
4	1	2	2	6	

Table 2. Confusion Matrix for 3-state HMM

$T\backslash R$	1	2	3	4	5
2		7	3	1	
4	3	2		6	

Table 3. Confusion Matrix for 8-state HMM

turned out to be different for different people. We use the holdout method for error estimation.

Recognition:

Given the gait cycles of an unknown person X and HMM models λ_i and stances for person i $1 \leq i \leq N$, we compute the 5-dimensional vector O_i^X. Using the Viterbi algorithm, we compute (3). The above procedure is repeated with respect to each person in the database. We rank order the person indices in descending order of the posterior probabilities. This procedure is repeated for several walk cycles. We present our results in terms of a cumulative match score (CMS) curve. It is also possible to give a confusion matrix as shown in table 1 for the Little and Boyd's database.

In the Little and Boyd database, Persons 2, 3 and 4 have similar structural characteristics and expectedly the false alarms are also somewhat predominant for these three subjects.The recognition results for the UMD and CMU databases are shown in figures 3a and 3b respectively. The result for the UMD database reveals that the performance of the method does not degrade significantly with an increase in the database size. However the slight drop in performance is due to drastic changes in clothing conditions of some subjects and changes in illumination(causing very noisy binarized silhouettes). It is natural for a person to change his speed of walking with time. The use of HMM enables us to deal with this variability without explicit time normalization. However for certain individuals and as biomechanics also suggests there is a considerable change in body dynamics as a person changes his speed which explains the slight drop in recognition rates. Observe that the results on CMU database when the HMM is trained using cycles from slow walk and tested using cycles from fast walk, the result is poor compared to the situation when the training and testing scenarios are reversed. This is because of the fact that with an increased number of frames per cycle the A matrix tends towards diagonal dominance on account of increased number of self loops. This suggests that explicit state duration modeling may be of interest [10].

The issue of the number of states deserves special attention. The choice of the number of states in an HMM model is always a tricky issue. The states of an HMM can be abstract quantities and it is not necessary that they must correspond to physical features of the underlying process. However, it would definitely be interesting, if the physical phenomenon can guide the choice of the number of states in an HMM. In Section 2, we had conjectured that the states and stances in a gait cycle are likely to be related.

We studied the performance of our gait recognizer as a function of the number of states. We experimented with 3, 5 and 8-state HMMs. The results obtained for the Little and Boyds database are reported. The worst case results for 3 and 8-state HMMs are given in Tables 2 and 3. From the tables we note that there is a considerable reduction in accuracy as compared to the 5-state case. The optimal state sequence obtained from the Viterbi algorithm revealed that the transitions in the states occur approximately at the same time instants that the shift in stances occurs in the observation sequence. On the other hand, the state sequence for 3-state and 8-state models did not have a corresponding physical interpretation . Thus, it appears that a 5-state HMM is best suited for our experiments, thereby confirming our conjecture relating the stances and the states.

6 Conclusion

In this paper, we have proposed an HMM-based approach to represent and recognize gait. A methodology is adopted to derive a low dimensional observation sequence from the silhouette of the body during a gait cycle. Learning is achieved by training an HMM for each person over several gait cycles. Gait recognition is performed by evaluating the log-probability that a given observation sequence was generated by an HMM model present in the database.

The method was tested on 3 different databases. In general, the recognition rates were found to be good. As anticipated, drastic changes in clothing adversely affects recognition performance. The method is sensitive to changes in viewing angle beyond ten degrees. The method is reasonably robust to changes in speed. In the case of human gait recognition we observed in some cases that the stride length changed appreciably with walking speed causing a slight drop in recognition performance. The method is however not robust to drastic changes in the silhouettes which might result due to changes in clothing or illumination.

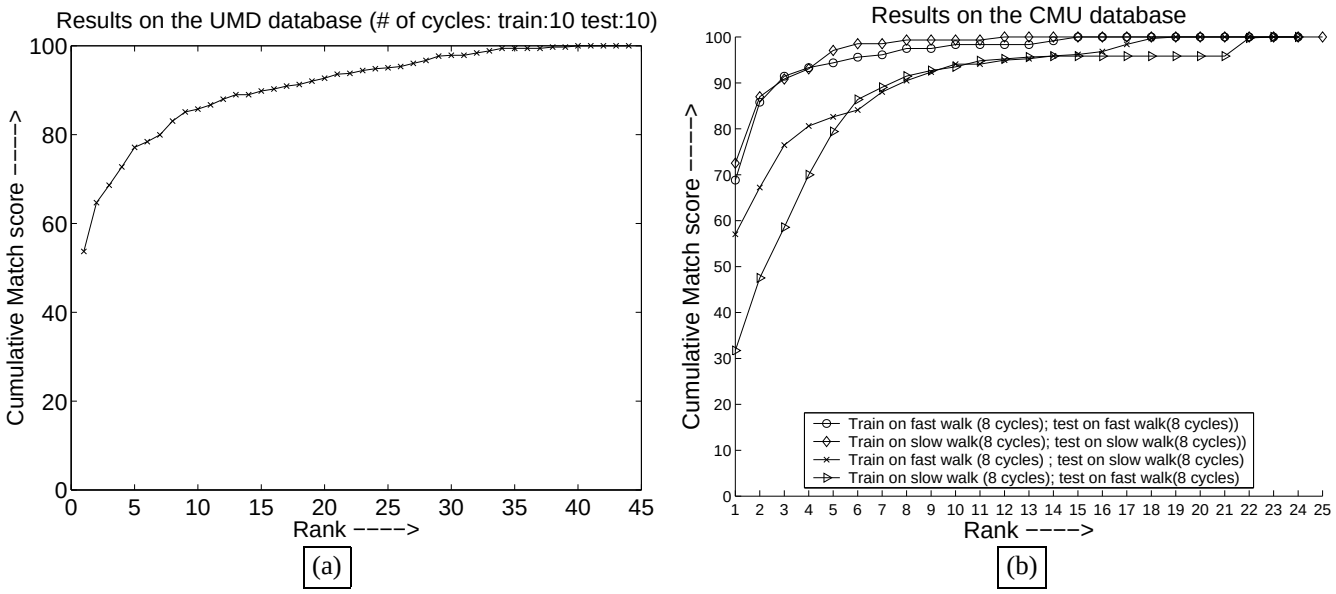

Figure 3. Identification Performance for (a) UMD database (b) CMU database

Presently we are looking at ways to make the scheme invariant to viewing angle and scale which might occur due to the use of multiple cameras. We are also exploring the use of better image metrics to make the 5-D vector more informative. It should be stressed here that the scheme has the potential to distinguish between humans and non-humans. It can also be extended to classify different activities such as walking and running. We are exploring the possibility of activity independent person identification.

References

[1] J. Aggarwal and Q. Cai. Human motion analysis:a review. *Computer Vision and Image Understanding*, 73(3):428–440, March 1999.

[2] A. Bobick and A. Johnson. Gait recognition using static activity-specific parameters. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (Lihue, HI)*, December 2001.

[3] D. Cunado, J. Nash, M. Nixon, and J. N. Carter. Gait extraction and description by evidence-gathering. *Proc. of the International Conference on Audio and Video Based Biometric Person Authentication*, pages 43–48, 1995.

[4] P. Huang, C. Harris, and M. Nixon. Recognizing humans by gait via parametric canonical space. *Artificial Intelligence in Engineering*, 13(4):359–366, October 1999.

[5] J. Little and J. Boyd. Recognizing people by theirgait: the shape of motion. *Videre*, 1(2):1–32, 1998.

[6] H. Murase and R. Sakai. Moving object recognition in eigenspace representation:gait analysis and lip reading. *Pattern Recognition Letters*, 17:155–162, 1996.

[7] M. Murray, A. Drought, and R. Kory. Walking patterns of normal men. *Journal of Bone and Joint surgery*, 46-A(2):335–360, 1964.

[8] P. J. Philips, H. Moon, and S. A. Rizvi. The feret evaluation methodology for face-recognition algorithms. *IEEE Trans. on Pattern Anal. and Machine Intell.*, 22(10):1090–1100, October 2000.

[9] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–285, February 1989.

[10] M. Russell and R. K. Moore. Explicit modelling of state occupancy in hidden markov models for automatic speech recognition. *Proceedings of IEEE Conference on Acoustics Speech and Signal Processing*, June 1985.

[11] T. Starner, J. Weaver, and A. Pentland. Real-time american sign language recognition from video using hmms. *IEEE Trans. on Pattern Anal. and Machine Intell.*, 12(8):1371–1375, December 1998.

[12] D. Wilson and A. Bobick. Nonlinear phmms for the interpretation of parameterized gesture. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (Santa Barbara, CA)*, June 1998.

[13] J. Yamato, J. Ohya, and L. Ishii. Recognizing human action in time-sequential images using hidden markov model. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 624–630, 1995.