Identification of Humans Using Gait

Amit Kale, Aravind Sundaresan, A. N. Rajagopalan, Naresh P. Cuntoor, Amit K. Roy-Chowdhury, Volker Krüger, and Rama Chellappa

***Abstract*—We propose a view-based approach to recognize humans from their gait. Two different image features have been considered: the width of the outer contour of the binarized silhouette of the walking person and the entire binary silhouette itself. To obtain the observation vector from the image features, we employ two different methods. In the first method, referred to as the indirect approach, the high-dimensional image feature is transformed to a lower dimensional space by generating what we call the frame to exemplar (FED) distance. The FED vector captures both structural and dynamic traits of each individual. For compact and effective gait representation and recognition, the gait information in the FED vector sequences is captured in a hidden Markov model (HMM). In the second method, referred to as the direct approach, we work with the feature vector directly (as opposed to computing the FED) and train an HMM. We estimate the HMM parameters (specifically the observation probability B) based on the distance between the exemplars and the image features. In this way, we avoid learning high-dimensional probability density functions. The statistical nature of the HMM lends overall robustness to representation and recognition. The performance of the methods is illustrated using several databases.**

I. Introduction

GAIT refers to the style of walking of an individual. Often, in surveillance applications, it is difficult to get face or iris information at the resolution required for recognition. Studies in psychophysics [1] indicate that humans have the capability of recognizing people from even impoverished displays of gait, indicating the presence of identity information in gait. From early medical studies [2], [3], it appears that there are 24 different components to human gait, and that, if all the measurements are considered, gait is unique. It is interesting, therefore, to study the utility of gait as a biometric.

A gait cycle corresponds to one complete cycle from rest (standing) position to-right-foot-forward-to-rest-to-left-foot-forward-to-rest position. The movements within a cycle consist of the motion of the different parts of the body such as head, hands, legs, etc. The characteristics of an individual are reflected not only in the dynamics and periodicity of a gait cycle but also in the height and width of that individual. Given the video of an unknown individual, we wish to use gait as a cue to find who among the N individuals in the database the person is. For a normal walk, gait sequences are repetitive and exhibit nearly periodic behavior. As gait databases continue to grow in size, it is conceivable that identifying a person only by gait may be difficult. However, gait can still serve as a useful filtering tool that allows us to narrow the search down to a considerably smaller set of potential candidates.

Approaches in computer vision to the gait recognition problem can be broadly classified as being either model-based or model-free. Both methodologies follow the general framework of feature extraction, feature correspondence and high-level processing. The major difference is with regard to feature correspondence between two consecutive frames. Methods which assume a priori models match the two-dimensional (2-D) image sequences to the model data. Feature correspondence is automatically achieved once matching between the images and the model data is established. Examples of this approach include the work of Lee *et al.* [4], where several ellipses are fitted to different parts of the binarized silhouette of the person and the parameters of these ellipses such as location of its centroid, eccentricity, etc. are used as a feature to represent the gait of a person. Recognition is achieved by template matching. In [5], Cunado *et al.* extract a gait signature by fitting the movement of the thighs to an articulated pendulum-like motion model. The idea is somewhat similar to an early work by Murray [2] who modeled the hip rotation angle as a simple pendulum, the motion of which was approximately described by simple harmonic motion. In [6] activity specific static parameters are extracted for gait recognition. Model-free methods establish correspondence between successive frames based upon the prediction or estimation of features related to position, velocity, shape, texture, and color. Alternatively, they assume some implicit notion of what is being observed. Examples of this approach include the work of Huang *et al.* [7], who use optical flow to derive a motion image sequence for a walk cycle. Principal components analysis is then applied to the binarized silhouette to derive what are called eigen gaits. Benabdelkader *et al.* [8] use image self-similarity plots as a gait feature. Little and Boyd [9] extract frequency and phase features from moments of the motion image derived from optical flow and use template matching to recognize different people by their gait.

A careful analysis of gait would reveal that it has two important components. The first is a structural component that captures the physical build of a person, e.g., body dimensions, length of limbs, etc. The second component is the motion dynamics of the body during a gait cycle. Our effort in this paper is directed toward deriving and fusing information from these two components. We

Manuscript received August 8, 2002; revised November 5, 2003. This work was supported by the DARPA/ONR under Grant N00014-00-1-0908. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Nasser Kehtarnavaz.

A. Kale is with the Department of Computer Science, University of Kentucky Lexington, KY 40506 USA.

A. Sundaresan, N. P. Cuntoor, and R. Chellappa are with the Department of Electrical and Computer Engineering and Center for Automation Research University of Maryland at College Park, College Park, MD 20740 USA.

A. N. Rajagopalan is with the Department of Electrical Engineering, Indian Institute of Technology, Madras Chennai 600 036, India.

A. K. Roy-Chowdhury is with the Department of Electrical Engineering, University of California at Riverside, Riverside CA 92521 USA.

V. Krüger is with the Aalborg University, Department of Computer Science, 6700 Esbjerg, Denmark.

Digital Object Identifier 10.1109/TIP.2004.832865

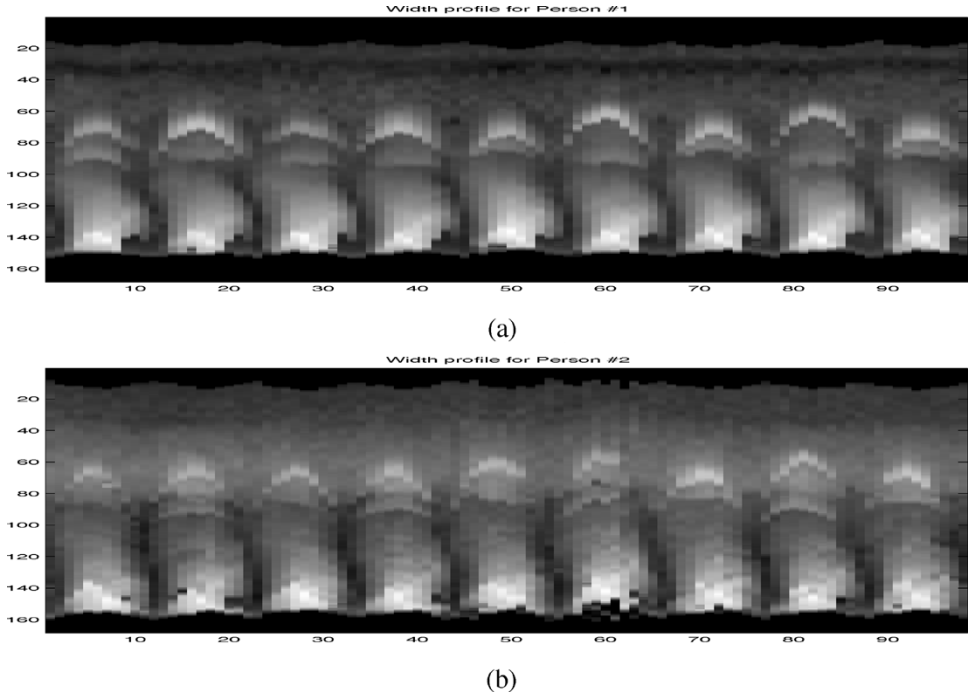


Fig. 1. Width vector profile for several gait cycles of two individuals. (a) Person 1. (b) Person 2.

propose a systematic approach to gait recognition by building representations for the structural and dynamic components of gait. The assumptions we use are: 1) the camera is static and the only motion within the field of view is that of the moving person and 2) the subject is monitored by multiple cameras so that the subject presents a side view to at least one of the cameras. This is because the gait of a person is best brought out in the side view. The image sequence of that camera which produces the best side view is used. Our experiments were set up in line with the above assumptions.

We considered two image features, one being the width of the outer contour of the binarized silhouette, and the other being the binary silhouette itself. A set of exemplars that occur during a gait cycle is derived for each individual. To obtain the observation vector from the image features we employ two different methods. In the *indirect approach* the high-dimensional image feature is transformed to a lower dimensional space by generating the frame to exemplar (FED) distance. The FED vector captures both structural and dynamic traits of each individual. For compact and effective gait representation and recognition, the gait information in the FED vector sequences is captured using a hidden Markov model (HMM) for each individual. In the *direct approach*, we work with the feature vector directly and train an HMM for gait representation. The difference between the direct and indirect methods is that in the former the feature vector is directly used as the observation vector for the HMM whereas in the latter, the FED is used as the observation vector. In the direct method, we estimate the observation probability by an alternative approach based on the distance between the exemplars and the image features. In this way, we avoid learning high-dimensional probability density functions. The performance of the methods is tested on different databases.

The organization of the paper is as follows. Section II explores the issue of feature selection. Section III describes the two algorithms. In Section IV, we present experimental results and Section V concludes the paper.

II. Feature Selection

An important issue in gait is the extraction of appropriate salient features that will effectively capture the gait characteristics. The features must be reasonably robust to operating conditions and should yield good discriminability across individuals. As mentioned earlier, we assume that the side view of each individual is available. Intuitively, the silhouette appears to be a good feature to look at as it captures the motion of most of the body parts. It also supports night vision capability as it can be derived from IR imagery also. While extracting this feature we are faced with two options.

1) Use the entire silhouette.
2) Use only the outer contour of the silhouette.

The choice of using either of the above features depends upon the quality of the binarized silhouettes. If the silhouettes are of good quality, the outer contour retains all the information of the silhouette and allows a representation, the dimension of which is an order of magnitude lower than that of the binarized silhouette. However, for low quality, low resolution data, the extraction of the outer contour from the binarized silhouette may not be reliable. In such situations, direct use of the binarized silhouette may be more appropriate.

We choose the width of the outer contour of the silhouette as one of our feature vectors. In Fig. 1, we show plots of the width profiles of two different individuals for several gait cycles. Since we use only the distance between the left and right extremities of the silhouette, the two halves of the gait cycle are almost indistinguishable. From here on, we refer to half cycles as cycles, for

the sake of brevity. In Fig. 1, the x axis denotes the frame index, while the y axis denotes the index of the width vector (the row index). The ith horizontal line in the image shows the variations in the ith element of the width vector as a function of time. A brighter gray-scale indicates a higher value of the width. We observe that within each cycle, there is a systematic temporal progression of width vector magnitude, viz. the dynamics. A similar observation has been made in [10], where the gait patterns are analyzed as Frieze patterns. For the two width profile plots shown in Fig. 1, the differences are quite visible. For instance, by observing the bright patterns in the upper region of the two images, we see that the brightness is more pronounced in the first image as compared to the second. This area of the plot corresponds to the swings of the hand. Second, note that the brightness gradient (which translates to velocity in the video sequence) in the lower part of the images is more pronounced for Person 1 as compared to Person 2. This part of the plot corresponds to the swings of the extremities of the foot. Additionally, note that the height, as well as the visibility of the neck part of the two persons are different. It must be pointed out, however, that the differences need not be so pronounced for all individuals. Thus, the width profile contains structural and dynamic information peculiar to each individual. Besides this, the use of the width feature imparts uniformity to feature representation across different individuals. Also, by definition, the width vector is translation invariant. Hence, the width of the outer contour of the silhouette is indeed a potentially good candidate as a feature.

Given the image sequence of a subject, the width vectors are generated as follows.

1) Background subtraction as discussed in [11] is first applied to the image sequence. The resultant motion image is then binarized into foreground and background pixels.
2) A bounding box is then placed around the part of the motion image that contains the moving person. The size of the box is chosen to accommodate all the individuals in the database. These boxed binarized silhouettes can be used directly as image features or further processed to derive the width vector as in the next item.
3) Given the binarized silhouettes, the left and right boundaries of the body are traced. 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 the right-most and the left-most boundary pixels in that row.

In order to generate the binarized silhouette only, the first two steps of the above feature are used. One of the direct applications of the width feature is to parse the video into cycles in order to compute the exemplars. It is easy to see that the norm of the width vector show a periodic variation. Fig. 2 shows the norm of the width vector as a function of time for a given video sequence. The valleys of the resulting waveform correspond to the rest positions during the walk cycle while the peaks correspond to the part of the cycle where the hands and legs are maximally displaced.

III. Proposed Algorithms

Given a sequence of image features for person j, $\mathcal{X}^j = \{\mathbf{x}^j(1), \mathbf{x}^j(2), \ldots, \mathbf{x}^j(T)\}$, we wish to build a model for the gait of person j and use it to recognize this person from M different subjects in the database.

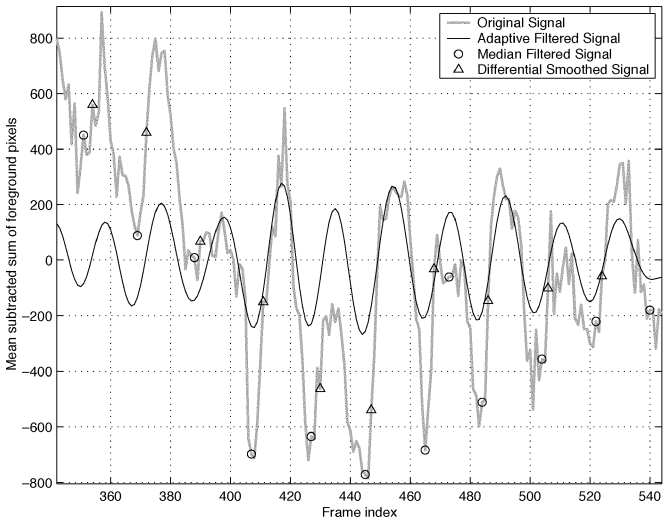


Fig. 2. Norm of the width vector as a function of time.

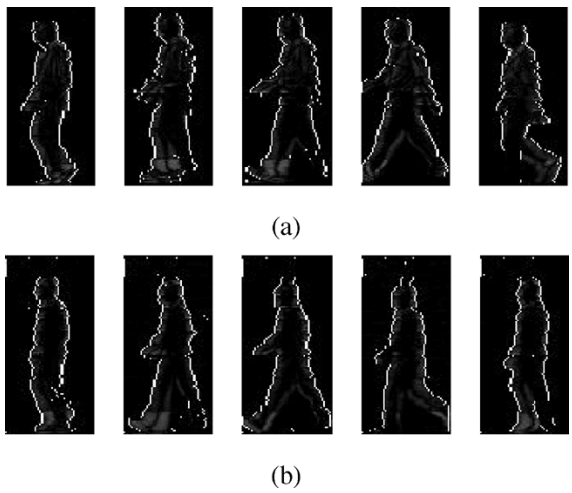


Fig. 3. Stances corresponding to the gait cycle of two individuals. (a) Person 1. (b) Person 2.

A. Overview

A closer examination of the physical process behind the generation of gait signature reveals that, during a gait cycle, it is possible to identify certain distinct phases or stances. In Fig. 3, we show five frames that we have picked from a gait cycle for two individuals. In the first stance, the person is at rest. In the second stance, he is just about to start and his hand is slightly raised. In the third stance, the hands and the feet are separated, while in the fourth stance, the hands and feet are displaced to a maximum. Finally, in the fifth stance, the person is returning to the rest state. Clearly, every person transits among these successive stances as he/she walks. Although these stances are generic, there exist differences not only in their image appearance based on the physical build of an individual, but also in the way an individual transits across these stances as he/she walks which represents the gait dynamics of the individual. A reasonable way to build a structural representation for a person is to pick N exemplars (or stances) $\mathcal{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_N\}$ from the pool of images that will minimize the error in representation of all the images of that person. The specifics of choice of exemplars may differ

for different approaches. Given the image sequence for an unknown person $\mathcal{Y} = \{\mathbf{y}(1), \ldots, \mathbf{y}(T)\}$, these exemplars can be directly used for recognition as

$$ID = \arg\min_j \sum_{t=1}^{T} \min_{n \in \{1,\ldots,N\}} d\left(\mathbf{y}(t), \mathbf{e}_n^j\right)$$

where $\mathbf{y}(t)$ represents the image of an unknown person at the tth time instant, while $\mathbf{e}_n^j$ represents the nth exemplar of the jth individual. Note, however, that such a simple discrimination criterion is susceptible to failures not only due to noise, but, more importantly, due to the presence of structural similarities among people in the database. To improve discriminability, the dynamics of the data must be exploited. A closer look at the gait cycle reveals that there is a temporal progression in the proximity of the observed silhouette to the different exemplars. Note that at the start of the gait cycle, a frame is closer to the first exemplar as compared to the other four. As time progresses, the frame will be closer to the second exemplar as compared to the others and so on. A similar behavior is reflected with regard to the remaining exemplars as well. Underlying the proximity of the silhouette to the exemplars is a probabilistic dependence across the exemplars. This encompasses information about how long a gait cycle persists in a particular exemplar as well as the way in which the gait cycle transits from one exemplar to the other. For two people who are similar in physical build, this dynamic knowledge can be used to improve the recognition performance. Because the transitions are systematic, it is possible to model this probabilistic dependence by a Markov matrix, as follows:

$$A = [P\left(\mathbf{e}_i(t) | \mathbf{e}_j(t-1)\right)] \tag{1}$$

for $i, j \in \{1, \ldots, N\}$. The matrix A encodes the dynamics in terms of state duration densities and transition probabilities. Often, in a practical situation, only a finite amount of training data is available and modeling can be difficult if the feature dimensionality is high. The dimension of the feature vectors described in the previous section is at least 100. Directly using the feature vectors to estimate the structure of the person and the dynamics of gait is clearly not advisable. We propose two different approaches to model the structure and dynamics of gait.

B. Approach 1: Indirect Approach

1) Gait Representation: In this approach, we pick N exemplars (or stances) $\mathcal{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_N\}$ from the pool of images that will minimize the error in representation of all the images of that person. If the overall average distortion is used as a criterion for codebook design, the selection of the N exemplars is said to be optimal if the overall average distortion is minimized for that choice. There are two conditions for ensuring optimality. The first condition is that the optimal quantizer is realized by using a nearest neighbor selection rule

$$q(\mathbf{x}) = \mathbf{e}_i, \Longleftrightarrow d(\mathbf{x}, \mathbf{e}_i) \leq d(\mathbf{x}, \mathbf{e}_j), \;\; j \neq i, \;\; 1 \leq i, \;\; j \leq N$$

where $\mathbf{x}$ represents an image in the training set, $d(\mathbf{x}, \mathbf{e}_i)$ is the distance between $\mathbf{x}$, and $\mathbf{e}_i$, while N is the number of exemplars.

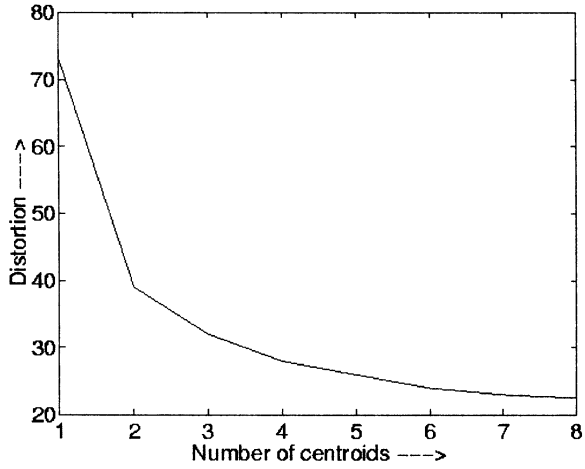


Fig. 4. Rate-distortion curve for number of exemplars vs distortion.

The second condition for optimality is that each codeword/exemplar $\mathbf{e}_i$ is chosen to minimize the average distortion in the cell C_i, i.e.

$$\mathbf{e}_i = \arg\min_{\mathbf{e}} E\left(d(\mathbf{x}, \mathbf{e}) | \mathbf{x} \in C_i\right)$$

where the C_is represent the Voronoi partitions [12] across the set of training images. To iteratively minimize the average distortion measure, the most widely used method is the k-means algorithm [12], [13]. However, implementing the k-means algorithm raises a number of issues. It is difficult to maintain a temporal order of the centroids (i.e., exemplars) automatically. Even if the order is maintained, there could be a cyclical shift in the centroids due to phase shifts in the gait cycle (i.e., different starting positions). In order to alleviate these problems, we divide each gait cycle into N equal segments. We pool the image features corresponding to the ith segment for all the cycles. The centroids (essentially the mean) of the features of each part were computed and denoted as the exemplar for that part. Doing this for all the N segments gives the optimal exemplar set $\mathcal{E} = \{\mathbf{e}_1^*, \ldots, \mathbf{e}_N^*\}$.

Of course, there is the issue of picking N. This is the classical problem of choosing the appropriate dimensionality of a model that will fit a given set of observations, e.g., choice of degree for a polynomial regression. The notion of "best fit" can be precisely defined by an objective function involving a penalty for the model complexity. Examples include minimum Bayes information criterion [14], minimum description length [15], etc. Similar problems exist for the case where there exists no parametric model for the data set, e.g., vector quantization. In problems like image compression, it is a common practice to look at the rate-distortion curves to examine the marginal reduction in the distortion as the bits per pixel are increased. We take a similar approach here. In Fig. 4, we plot the average distortion as a function of the number of centroids for the UMD database. It can be observed that beyond five centroids, the average distortion does not decrease as rapidly with the increase in the number of centroids.

In order to reliably estimate the gait dynamics, we propose a novel way of compactly encoding the observations, while retaining all the relevant information. Let $\mathbf{x}(t)$ denote the feature

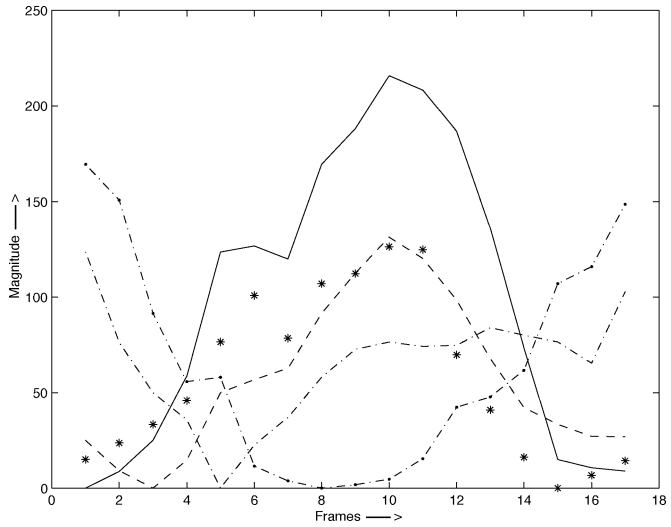


Fig. 5. FED vector components plotted as a function of time.

extracted from the image at time t. The distance of $\mathbf{x}(t)$ from the corresponding exemplars $\mathbf{e}_n \in \mathcal{E}$ can be computed to build an FED, $\mathbf{f}(t)$, which serves as a lower (N) dimensional representation of the image at time t. For instance, for the jth individual we compute the nth element of the FED vector as

$$\left[\mathbf{f}_j^{\mathcal{X}^j}(t)\right]_n = d\left(\mathbf{x}^j(t), \mathbf{e}_n^j\right) \tag{2}$$

where $t \in \{1, \ldots, T\}$, $\mathbf{e}_n^j$ denotes the nth exemplar of the jth person, and $n \in \{1, \ldots, N\}$. Thus, $\mathbf{f}_j^{\mathcal{X}^j}(t)$ constitutes an observation vector for person j. Similarly, $\mathbf{f}_j^{\mathcal{X}^i}(t)$ represents the observation sequence of the person i encoded in terms of the exemplars of person j. Note that the dimension of $\mathbf{f}_j^{\mathcal{X}^j}(t)$ is only N. These observations can be derived for several such gait cycles in the database.

It is clear that, as we examine a gait cycle, the proximity of a frame from each of the stances changes with time. Correspondingly, the elements of the vector $\mathbf{f}_j^{\mathcal{X}^j}(t)$ would reflect this feature. To elaborate this further, note that for a frame at the start of the gait cycle, the first element of the observation vector will be smaller in magnitude as compared to the remaining four elements. As time progresses, the first element will increase in magnitude because the frame moves closer to the second stance. The magnitude of the second element will decrease as long as the frame is close to the second stance and then it will start to increase as well. A similar behavior is observed in the rest of the elements of the vector. The duration for which an element of this vector stays low encodes the stance duration density as also the probability of transition to another stance. Fig. 5 shows the evolution of the different components of the FED vector $\mathbf{f}_j^{\mathcal{X}^j}(t)$ for a half-gait cycle. As can be seen, there is a systematic succession of valleys for the different FED vector components across time. The FED vector representation is independent of the choice of features. The distance in (2) will change depending upon the feature. For example, for the case of the width feature, d corresponds to the Euclidean distance, whereas for the binary silhouette, d corresponds to the binary correlation. As described before, it is possible to model transition across exemplars by a Markov matrix. For the person j, it is possible to look upon the FED vector sequence $\mathbf{f}_j^{\mathcal{X}^j}(t)$ as the observed manifestation of the transition across exemplars (a hidden process). An HMM is appropriate for such a signal. HMMs use a Markov process to model the changing statistical characteristics that are manifested in the actual observations. The state sequence is *hidden*, and can only be observed through another set of observable stochastic processes. Each hidden state of the model is associated with a set of output probability distributions which can be either discrete probability mass functions or continuous probability density functions. Details on HMMs can be found in [16]. For the gait problem, the exemplars can be considered as analogues of states of the HMM, while the FED vector sequence can be considered as the observed process. Since the feature vectors are transformed to the intermediate FED vector representation, we refer to this approach as an indirect approach. In the proposed model for gait, the primary HMM parameters of interest are the number of states, the initial probability vector (π), the transition probability matrix (A), and the output probability distribution B, which we model as a continuous probability distribution. A brief explanation of each of these terms follows.

1) Initial probability (π): The initial probability vector is given by $\pi = \{\pi_i\}$, where π_i represents the probability of being in state i at the beginning of the experiment. For the gait problem, π_i can be thought of as the probability of starting in a particular stance.
2) Transition probability matrix (A): The entries of this matrix are given by a_{ij} where $a_{ij} = P(i_{t+1} = j | i_t = i)$. This represents the probability of being in state j at time $t+1$ given that the previous state was i. An HMM in which every state of the HMM can be reached from any other state, viz. every coefficient a_{ij} of A is positive, is referred to as ergodic HMM. When the coefficients have the property $a_{ij} = 0$ for $j < i$, i.e., if no transitions are allowed to states whose indices are lower than the current state, the HMM is referred to as a left-to-right model.
3) Probability of observation (B): The probability of observing symbol $\mathbf{x}$ while in state j is given by $b_j(\mathbf{x})$. Since the observations in our experiment are continuous valued, finding $b_j(\mathbf{x})$ turns out to be a problem of estimating the underlying probability density function of the observations. In the literature on HMMs, a Baum–Welch type of reestimation procedure has been formulated [16] for a mixture of any log concave or elliptically symmetric density function (such as the Gaussian).

In this paper, $\lambda = (A, B, \pi)$ will be used to compactly represent an HMM.

2) Gait Recognition: The HMM model parameters $\lambda = (A, B, \pi)$ serve as a means to represent the gait of different people. For robust recognition, it is reasonable that one must examine several gait cycles before taking a decision, i.e., instead of looking at a single walking cycle, it would be prudent to examine multiple cycles of a person to derive any conclusion about his gait. We assume that several gait cycles of an individual are given. The problem is to recognize this individual from a database of people whose models for gait are known *a priori*.

To begin with, the image sequence of the unknown person $\mathcal{Y} = \{\mathbf{y}(1), \ldots, \mathbf{y}(T)\}$ is subjected to the same image processing operations as the training image sequence, i.e., the

image feature $\mathbf{y}(t)$ of this person is generated for each frame and the FED vector $\mathbf{f}_j^{\mathcal{Y}}(t)$ is computed for all $j \in \{1, \ldots, M\}$ using (2). We wish to compute the likelihood that the observation sequence $\mathbf{f}_j^{\mathcal{Y}}$ was generated by the HMM corresponding to the jth person. This can be deciphered by using the forward algorithm [16] which computes this log probability as

$$P_j = \log\left(P\left(\mathbf{f}_j^{\mathcal{Y}} | \lambda_j\right)\right). \tag{3}$$

Here, λ_j is the HMM model corresponding to the person j. We repeat the above procedure for every person in the database thereby producing P_j, $j \in \{1, \ldots, M\}$. Suppose that the unknown person was actually person m. We would then expect P_m to be the largest among all P_js. A larger value of P_m will be the result of two factors.

1) The distance between $\mathcal{Y}$ and the stances of person m will be smaller than that between $\mathcal{Y}$ and any other person.
2) The pattern of transitions between stances/states for $\mathcal{Y}$ will be closest to that for person m.

Note that the observed image sequence must be in accordance with the transition probability matrix A as well as the observation probability B in order to yield a larger value for the log probability. If the values of $P_1, \ldots, P_M$ are observed for a sufficient number of gait cycles of the unknown person, one would expect that in a majority of cases P_m would be lower as compared to the rest of the P_is. For smaller databases, the performance can be easily examined in terms of a confusion matrix. For larger databases, a more convenient way of reporting recognition performance is to report the number of times the right person occurs in the top n matches where $n < M$, i.e., by way of cumulative match scores (CMS).

C. Approach 2: Direct Approach

1) Gait Representation: In this approach, we use the feature vector in its entirety to estimate the HMM $\lambda = (A, B, \pi)$ for each person. Hence, we refer to this approach as the direct approach. One of the important issues in training is learning the observation probability B. In general, if the underlying distribution of the data is non-Gaussian, it can be characterized by a mixture of Gaussians. As discussed before, the reliability of the estimated B depends on the number of training samples available and the dimension of the feature vector. In order to deal with the high dimensionality of the feature vector, we propose an alternative representation for B.

As discussed in the previous section it is possible, during a gait cycle, to identify certain distinct phases or stances. We build a structural representation for a person by picking N exemplars (or stances) from the training sequence, $\mathcal{X} = \{\mathbf{x}(1), \ldots, \mathbf{x}(T)\}$. We now define B in terms of the distance of this vector from the exemplars as follows:

$$b_n(\mathbf{x}(t)) = P(\mathbf{x}(t)|\mathbf{e}_n) = \beta e^{-\alpha D(\mathbf{x}(t), \mathbf{e}_n)}. \tag{4}$$

The probability $P(\mathbf{x}(t)|\mathbf{e}_n)$ is defined as a function of $D(\mathbf{x}(t), \mathbf{e}_n)$, the distance of the feature vector $\mathbf{x}(t)$ from the nth exemplar, $\mathbf{e}_n$. The motivation behind using an exemplar-based model in the above manner is that the recognition can be based on the *distance measure* between the observed feature vector and the exemplars. During the training phase, a model is built for all the subjects in the gallery. Note that B is completely defined by $\mathcal{E}$ if α and β are fixed beforehand. An initial estimate of $\mathcal{E}$ and λ is formed from $\mathcal{X}$, and these estimates are refined iteratively using expectation maximization [17]. We can iteratively refine the estimates of A and π by using the Baum–Welch algorithm [16] with $\mathcal{E}$ fixed. The algorithm to refine estimates of $\mathcal{E}$, while keeping A and π fixed, is determined by the choice of the distance metric. We describe in the following sections the methods used to obtain initial estimates of the HMM parameters, the training algorithm, and, finally, identification from a probe sequence.

2) Initial Estimate of HMM Parameters: In order to obtain a good estimate of the exemplars and the transition matrix, we first obtain an initial estimate of an ordered set of exemplars from the sequence and the transition matrix and then iteratively refine the estimate. We observe that the gait sequence is quasiperiodic and we use this fact to obtain the initial estimate $\mathcal{E}^{(0)}$. We first divide the sequence into cycles. We can further divide each cycle into N temporally adjacent clusters of approximately equal size. We visualize the frames of the nth cluster of all cycles to be generated from the nth state. Thus, we can get an initial estimate of $\mathbf{e}_n$ from the feature vectors belonging to the nth cluster of all cycles. In order to get reliable initial estimates of the exemplars, we need to robustly estimate the cycle boundaries (see [18]). A corresponding initial estimate of the transition matrix, $A^{(0)}$ (with $A_{j,j}^{(0)} = A_{j,j \bmod N+1}^{(0)} = 0.5$, and all other $A_{j,k}^{(0)} = 0$) is also obtained. The initial probabilities $\pi_n^{(0)}$ are set to be equal to $1/N$.

3) Training the HMM Parameters: The iterative refinement of the estimates is performed in two steps. In the first step, a Viterbi evaluation [16] of the sequence is performed using the current values for the exemplars and the transition matrix. We can, thus, cluster feature vectors according to the most likely state they originated from. Using the current values of the exemplars $\mathcal{E}^{(i)}$ and the transition matrix $A^{(i)}$, Viterbi decoding on the sequence $\mathcal{X}$ yields the most probable path $\mathcal{Q} = \{q^{(i)}(1), q^{(i)}(2), \ldots, q^{(i)}(T)\}$, where $q^{(i)}(t)$ is the estimated state at time t and iteration i. Thus, the set of observation indices, whose corresponding observation is estimated to have been generated from state n is given by $\mathcal{T}_n^{(i)} = \{t : q^{(i)}(t) = n\}$. We now have a set of frames for each state and we would like to select the exemplars so as to maximize the probability in (5). If we use the definition in (4), (6) follows:

$$\mathbf{e}_n^{(i+1)} = \arg_{\mathbf{e}} \max \prod_{t \in \mathcal{T}_n^{(i)}} P(\mathbf{x}(t)|\mathbf{e}) \tag{5}$$

$$\mathbf{e}_n^{(i+1)} = \arg_{\mathbf{e}} \min \sum_{t \in \mathcal{T}_n^{(i)}} D(\mathbf{x}(t), \mathbf{e}). \tag{6}$$

The actual method for minimizing the distance in (6), however, depends on the distance metric used. We use the inner product (IP) distance (7). We have experimented with other distance measures, namely the Euclidean (EUCLID) distance and the sum of absolute difference (SAD) distance [18]

$$D_{IP}(\mathbf{x}, \mathbf{e}) = 1 - \frac{\mathbf{x}^\top \mathbf{e}}{\sqrt{\mathbf{x}^\top \mathbf{x} \mathbf{e}^\top \mathbf{e}}}. \tag{7}$$

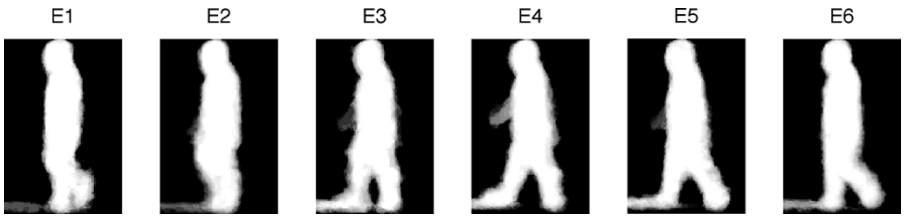


Fig. 6. Exemplars estimated using IP distance measure.

Note that though $\mathbf{x}$ and $\mathbf{e}$ are 2-D images, they are represented as vectors of dimension $D \times 1$ for ease of notation. $\mathbf{1}_{D\times 1}$ is a vector of D ones. The equations for updating the exemplars is given by (8). $\tilde{\mathbf{x}}$ denotes the normalized vector $\mathbf{x}$

$$\mathbf{e}_n^{(i+1)} = \sum_{t\in\mathcal{T}_n^{(i)}} \tilde{\mathbf{x}}(t). \tag{8}$$

The exemplars estimated for one observation sequence are displayed in Fig. 6. Given $\mathcal{E}^{(i+1)}$ and $A^{(i)}$, we can calculate $A^{(i+1)}$ using the Baum–Welch algorithm [16]. We set $\pi_n^{(i+1)} = 1/N$ at each iteration. Thus, we can iteratively refine our estimates of the HMM parameters. It usually takes only a few iterations to obtain an acceptable estimate.

4) Identifying From a Test Sequence: Given the sequence of the unknown person $\mathcal{Y}$ and the exemplars and HMM model parameters for the different people in the database, we wish to recognize the unknown person. As before, the given image sequence of the unknown person is subjected to the same image processing operations as the training image sequence to extract the relevant image features. As explained before, the likelihood that the observation sequence was produced by the jth individual in the database is computed using the forward algorithm as

$$P_j = \log\left(P(\mathcal{Y}|\lambda_j)\right). \tag{9}$$

Note that λ_j implicitly includes the exemplar set corresponding to person j.

IV. Experimental Results

In this section, we demonstrate the performance of the proposed algorithms on different databases. Our experiments are aimed at finding how well the two methods perform with respect to several different variations such as size of database, speed of walking, clothing, illumination, etc. We have considered normal walk as well as treadmill data for analysis. Our video sequences were taken from 1) the Carnegie Mellon University (CMU) database, 2) the University of Maryland (UMD) database, and 3) the University of South Florida (USF) database. For the sake of brevity, we present detailed results of the indirect approach using width vectors on the CMU and UMD databases while the results of the direct and indirect approaches with the binarized silhouette feature will be presented for the USF database.

Silhouettes and the feature vectors for the person are extracted using the procedure described in Section II. In general, the choice of N depends on the frame rate. For the UMD and CMU databases, we chose N to be five. However, for the USF database, which has a higher frame rate, we found that $N = 6$ is a better choice. The Viterbi algorithm was used to identify the probe sequence, since it is efficient and can operate in the logarithmic domain using only additions. For every gait cycle, we rank order the probabilities and the corresponding person indices in descending order. We then evaluate performance by letting each person in the database be the unknown u and plot the fraction of times that the right person is within the top n matches as a function of n. This curve known as the cumulative match score characteristic (CMS) was first used in the context of face recognition by Philips *et al.* [19].

A. CMU Database

This database[1] has 25 people walking at a fast pace and slow pace on a treadmill and a sequence of people walking while carrying a ball. There are about 16 cycles in each sequence. Half of the cycles were used for training and the other half for testing. The size of the image was 640 × 480. We did the following experiments on this database: 1) train on slow walk and test on slow walk, 2) train on fast walk and test on fast walk, 3) train on slow walk and test on fast walk, 4) train on fast walk and test on slow walk, and 5) train on walk carrying a ball and test on walk carrying a ball. In cases 1), 2), and 5), for each person, the sequences were divided into two halves, one half used for training and the other for testing, while in the cases 3) and 4), the entire slow/fast sequence was used for training and the other fast/slow sequence was used for evaluation.

The results obtained using the proposed method are given in Figs. 7 and 8. It can be seen that the right person in the top three matches 90% of the times for the cases where the training and testing sets correspond to the same walking styles. 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. In an effort to understand this, we ran an experiment whereby we artificially increased the number of frames per activity cycle using interpolation and observed the resulting HMM. It was seen that the A matrix tends toward diagonal dominance. This occurs on account of the fact that the HMM does not provide adequate representation of extreme temporal durations of activity. The probability of t consecutive observations in state i can be written as

$$d_i(t) = a_{ii}^t(1 - a_{ii})$$

[1]More details about the data are available at the URL http://hid.ri.cmu.edu/HidEval/evaluation.html

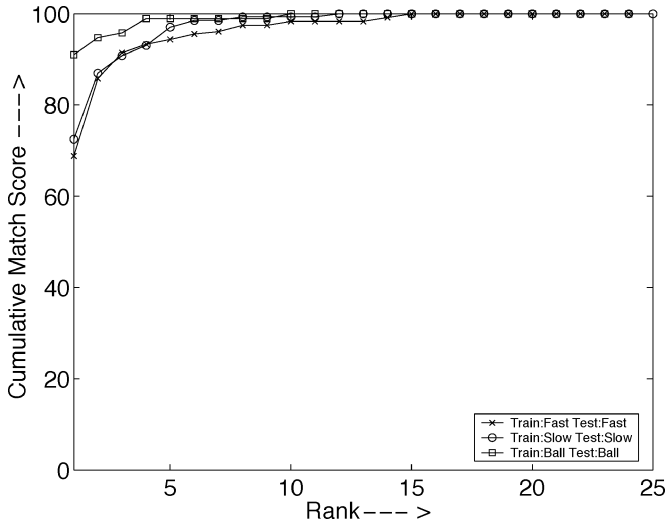


Fig. 7. Cumulative match characteristic for normal walk and walk when carrying an object for the CMU database.

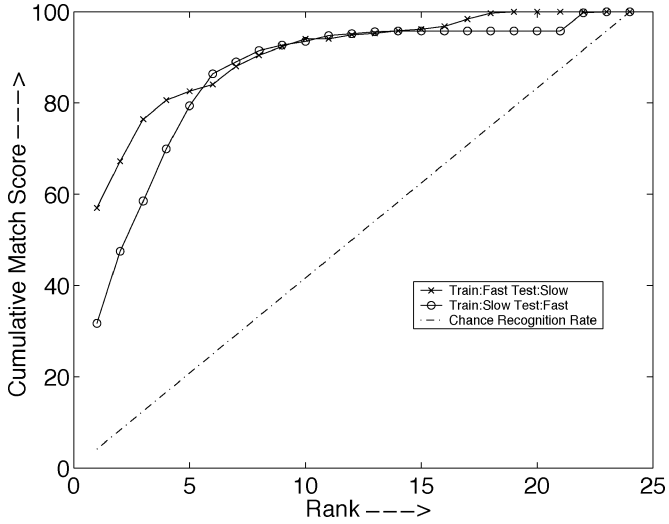


Fig. 8. Cumulative match characteristic for across speed testing for the CMU database.

where $d_i(t)$ is the probability of taking a self loop at state i for t times viz. a geometric distribution. As the duration of the activity $T \rightarrow \infty$ for a fixed N, this causes $a_{ii} \rightarrow 1$ and $a_{ij} \rightarrow 0$. Clearly, the geometric distribution does not represent a realistic description of the state duration density in our gait-modeling problem. Similar issues have been raised in the context of speech recognition and a solution is to explicitly model the distribution of state duration as has been done by Russell [20]. For the case of training with fast walk and testing on slow walk, the dip in performance is caused due to the fact that for some individuals as biomechanics suggests, there is a considerable change in body dynamics and stride length as a person changes his speed. For example, observe Fig. 9 which shows a few frames in the gait cycles of a person in the two scenarios. As is apparent from the figure, the posture as well as hand swings for the person are quite different for fast walk and slow walk.

When the subjects are walking with a ball in their hands, most of the gait dynamics are confined to the leg region. For this experiment, i.e., case 5), we observe from Fig. 7 that the top match is the correct match 90% of the time which is higher than the top match score (around 70%) in the normal walk cases. This suggests that for the purpose of recognition, certain parts of the body may be more effective than others. In particular, using the leg motion alone provides more discriminating evidence as compared to what might be obtained by weighting the evidences from the hand and leg motion coequally.

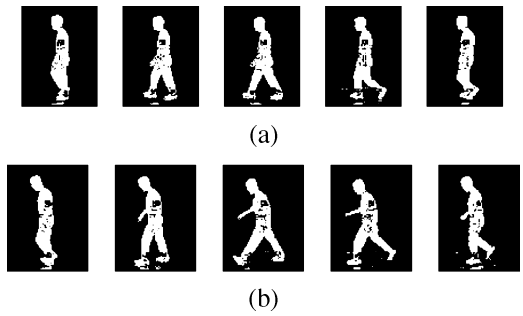


Fig. 9. Sample images of a person corresponding to different speeds. (a) Slow walk. (b) Fast walk.

B. UMD Database

It would be very useful to evaluate the utility of gait as a biometric in more realistic situations than those prevailing in the CMU database. To get a more realistic evaluation of gait, we designed our own experiment at the University of Maryland. We used outdoor surveillance cameras (Philips G3 EnviroDome camera system) at a height of 15 ft to capture data. The subjects were made to walk along a T-shaped path so that they present a side view to the surveillance cameras. This is in accordance with our basic assumption in Section I. We collected gait sequences of 44 individuals. For most individuals, the training and test video sequences were collected on different days. The database[2] is diverse in terms of gender, age, ethnicity, etc. Moreover, there was a change in clothing of the people across different days as well. This, and the fact that the data was collected outdoors under uncontrolled illumination, provides a realistic scenario for gait analysis. Each video sequence has approximately ten cycles. One sequence was used for training and the other for evaluation. The size of the image was 150 × 75. The result using the proposed method is shown in Fig. 10(a). It can be seen that the performance of the method does not degrade with an increase in the database size. The slight drop in performance can be attributed to changes in clothing conditions of some subjects and changes in illumination resulting in noisy binarized silhouettes. In Fig. 10(a), the dashed–dotted line represents the chance recognition rate which corresponds to the line connecting (1, $100/P$) and (P, 100), where P denotes the number of people in the gallery. In order to assess the confidence in our recognition capability, we computed the recognition performance by dropping one subject from the gallery and the probe when we compute the cumulative match characteristic. Essentially, this amounts to computing the CMS characteristics by eliminating row i and column i from the similarity matrix. By this procedure, we ensure that no individual in the gallery leads to a bias in

[2]More details about this data are available at the URL: http://degas.umiacs.umd.edu/hid/.

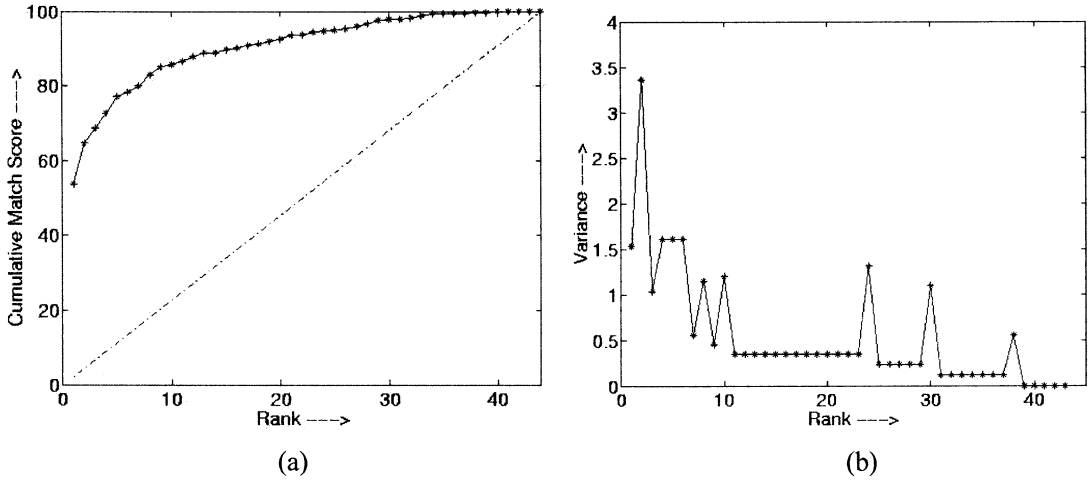


Fig. 10. Results for the UMD database using Algorithm 1 (44 people). (a) Cumulative match characteristic. (b) Recognition confidence.

TABLE I
PROBE SETS AND MATCH SCORES FOR THE USF DATABASE USING THE BASELINE ALGORITHM AND OUR INDIRECT AND DIRECT APPROACHES

Experiment (Probe)	Baseline		Indirect Approach		Direct Approach	
	Rank 1	Rank 5	Rank 1	Rank 5	Rank 1	Rank 5
A (G, A, L)	79	96	91	100	99	100
B (G, B, R)	66	81	76	81	89	90
C (G, B, L)	56	76	65	76	78	90
D (C, A, R)	29	61	25	61	35	65
E (C, B, R)	24	55	29	39	29	65
F (C, A, L)	30	46	24	46	18	60
G (C, B, L)	10	33	15	33	24	50

the performance computation. Having computed the CMS characteristics, thus, we find the variance in the cumulative match score at every rank. Smaller the value of this variance, the more reliable is the gait recognition performance. This is shown for the UMD database in Fig. 10(b).

C. USF Database

Finally, we consider the USF database[3] which has been identified as the gait challenge database [21]. The database has variations as regards viewing direction, shoe type, and surface type. Also, the subjects were asked to carry a briefcase for one testing condition.

We present the results of both our methods and a comparative analysis on this dataset. Different probe sequences for the experiments along with the cumulative match scores are given in Table I for the baseline algorithm [22] and our direct and indirect approaches. The image quality for the USF database was worse than the previous two databases in terms of resolution and amount of noise. We experimented with both the width feature as well as the binarized silhouette for the USF dataset. However, the extraction of the outer contour in this case is not reliable and the width vectors were found to be noisy. In Table I, we report only the results of our methods using the silhouettes as the image feature. G and C indicate grass and concrete surfaces, A and B indicate shoe types, and L and R indicate left and right cameras, respectively. We observe that the direct method is more robust to the presence of noise than the indirect method. We also note that recognition performance suffers most due to differences in surface and background characteristics, and least due to difference in viewing angle. Results from other research groups using this data can be found in [23] and websites (http://degas.umiacs.umd.edu/links.html). Recently, the gallery in the USF database was extended by adding subjects who walked with only one shoe type on grass, which happened to be labeled as Shoe B. Since the shoe type labeling is arbitrary, they were put in the gallery to increase the gallery size to 122. Also, the baseline algorithm itself was modified using a new background processing module. In Table II, we present results for the new baseline algorithm and our direct approach. In this table, $t2$ refers to the probe sequences captured in November and BF represents the people walking with a briefcase. The corresponding CMS curves are shown in Fig. 11(a).

TABLE II
RECOGNITION SCORES FOR NEW BASELINE AND HMM (122 PEOPLE)

Experiment (Probe)	Baseline	HMM
A (G, A, L)	73	89
B (G, B, R)	78	88
C (G, B, L)	48	68
D (C, A, R)	32	35
E (C, B, R)	22	28
F (C, A, L)	17	15
G (C, B, L)	17	21
H (G, A, R, BF)	61	85
I (G, B, R, BF)	57	80
J (G, A, L, BF)	36	58
K $(G, A, R, t2)$	3	17
L $(C, A, R, t2)$	3	15

[3]More details about this database can be found at http://figment.csee.usf.edu/GaitBaseline/.

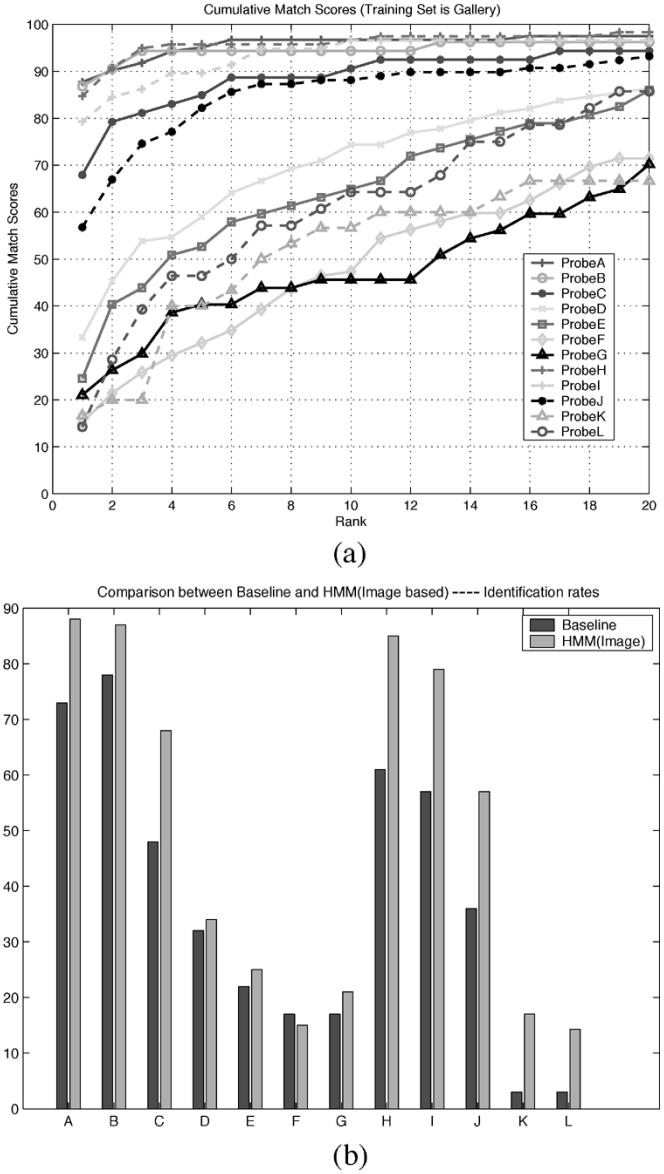


Fig. 11. Results for the USF database (122 people). (a) CMS plots. (b) Comparison with baseline.

V. Conclusion

In this paper, we have presented two approaches to represent and recognize people by their gait. The width of the outer contour of the binarized silhouette as well as the silhouette itself were used as features to represent gait. In one approach, a low-dimensional observation sequence is derived from the silhouettes during a gait cycle and an HMM is trained for each person. Gait identification is performed by evaluating the probability that a given observation sequence was generated by a particular HMM model. In the second approach, the distance between an image feature and exemplar was used to estimate the observation probability B. The performance of the methods was illustrated using different gait databases.

Acknowledgment

The authors would like to thank B. Yegnanarayana, Indian Institute of Technology, Madras, for helpful discussions on hidden Markov models.

References

[1] J. Cutting and L. Kozlowski, "Recognizing friends by their walk: gait perception without familiarity cues," *Bull. Psychonom. Soc.*, vol. 9, pp. 353–356, 1977.

[2] M. P. Murray, A. B. Drought, and R. C. Kory, "Walking patterns of normal men," *J. Bone and Joint Surgery*, vol. 46-A, no. 2, pp. 335–360, 1964.

[3] M. P. Murray, "Gait as a total pattern of movement," *Amer. J. Phys. Med.*, vol. 46, pp. 290–332, June 1967.

[4] L. Lee and W. E. L. Grimson, "Gait analysis for recognition and classification," in *Proc. IEEE Conf. Face and Gesture Recognition*, 2002, pp. 155–161.

[5] D. Cunado, J. M. Nash, M. S. Nixon, and J. N. Carter, "Gait extraction and description by evidence-gathering," in *Proc. Int. Conf. Audio and Video Based Biometric Person Authentication*, 1995, pp. 43–48.

[6] A. F. Bobick and A. Johnson, "Gait recognition using static, activity-specific parameters," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, 2001, pp. 423–430.

[7] P. S. Huang, C. J. Harris, and M. S. Nixon, "Recognizing humans by gait via parametric canonical space," *Artif. Intell. Eng.*, vol. 13, no. 4, pp. 359–366, Oct. 1999.

[8] R. Cutler, C. Benabdelkader, and L. S. Davis, "Motion based recognition of people in eigengait space," in *Proc. IEEE Conf. Face and Gesture Recognition*, 2002, pp. 267–272.

[9] J. Little and J. Boyd, "Recognizing people by their gait: The shape of motion," *Videre*, vol. 1, no. 2, pp. 1–32, 1998.

[10] Y. Liu, R. T. Collins, and Y. Tsin, "Gait sequence analysis using frieze patterns," in *Proc. ECCV*, vol. 2, 2002, pp. 657–671.

[11] A. Elgammal, D. Harwood, and L. Davis, "Non-parametric model for background subtraction," in *Proc. IEEE FRAME-RATE Workshop*, 1999.

[12] M. R. Anderberg, *Cluster Analysis for Applications*. New York: Academic, 1973.

[13] R. Duda and P. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.

[14] G. Schwartz, "Estimating the dimension of a model," *Ann. Stat.*, vol. 6, no. 2, pp. 497–511, 1978.

[15] J. Rissanen, "Universal coding, information, prediction, and estimation," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 629–636, July 1984.

[16] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, pp. 257–285, February 1989.

[17] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *J. Roy. Stat. Soc.*, vol. 39, no. 1, pp. 1–38, 1977.

[18] A. Sundaresan, A. RoyChowdhury, and R. Chellappa, "A hidden markov model based framework for recognition of humans from gait sequences," in *Proc. Int. Conf. Image Processing*, 2003.

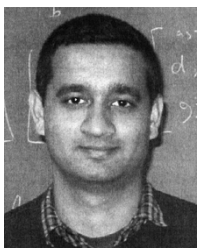
[19] P. J. Phillips, H. Moon, and S. A. Rizvi, "The feret evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, pp. 1090–1100, Oct. 2000.

[20] M. Russell and R. K. Moore, "Explicit modeling of state occupancy in hidden markov models for automatic speech recognition," in *Proc. IEEE Conf. Acoustics Speech and Signal Processing*, June 1985, pp. 5–8.

[21] P. J. Phillips, S. Sarkar, I. Robledo, P. Grother, and K. W. Bowyer, "The gait identification challenge problem: data sets and baseline algorithm," in *Proc. Int. Conf. Pattern Recognition*, 2002.

[22] ——, "Baseline results for the challenge problem of human id using gait analysis," in *Proc. 5th IEEE Int. Conf. Automatic Face and Gesture Recognition*, 2002.

[23] D. Tolliver and R. Collins, "Gait shape estimation for identification," in *Proc. AVBPA*, 2003, pp. 734–742.

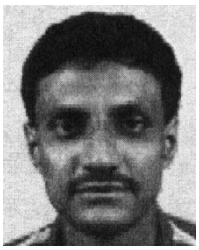


Amit Kale received the B.E. (Hons.) degree from Victoria Jubilee Technical Institute, affiliated with the University of Bombay, Bombay, India, in 1996, the M.Tech. degree from the Indian Institute of Technology, Bombay, in 1998, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Maryland, College Park, in 2003, where he worked on developing algorithms for human identification using gait.

He is presently a Postdoctoral Researcher in the Department of Computer Science, University of Kentucky, Lexington. His research interests are in image and video processing, computer vision, and pattern recognition.

Aravind Sundaresan received the B.E. (Hons.) degree from the Birla Institute of Technology and Science, Pilani, India, in 2000. He is pursuing the Ph.D degree in the Department of Electrical and Computer Engineering, University of Maryland, College Park.

He was a Software Engineer in the Multimedia Group, Sasken Communication Technologies, Bangalore, India, between July 2000 and July 2001. His research interests are in signal, image, and video processing, and computer vision.



A. N. Rajagopalan received the B.E. and M.Tech. degrees in electronics engineering from Nagpur University, Nagpur, India, and the Ph.D. degree in electrical engineering from the Indian Institute of Technology (IIT), Bombay, in July 1998.

During the summer of 1998, he was a Visiting Scientist at the Image Communication Laboratory, University of Erlangen, Erlangen, Germany. He joined the Center for Automation Research, University of Maryland, College Park, in October 1998 and was on the research faculty as Assistant Research Scientist until September 2000. Since October 2000, he has served as Assistant Professor in the Department of Electrical Engineering, IIT, Madras. He is a coauthor of the book *Depth from defocus: A real aperture imaging approach* (New York: Springer-Verlag, 1999). His research interests include depth recovery from defocus, image restoration, superresolution, face detection and recognition, and higher-order statistical learning.

Naresh P. Cuntoor received the B.E. degree from Karnataka Regional Engineering College, Surathkal, India, in 2000 and the M.S. degree from the University of Maryland, College Park, in 2003, where he is currently pursuing the Ph.D. degree in the Department of Electrical and Computer Engineering.

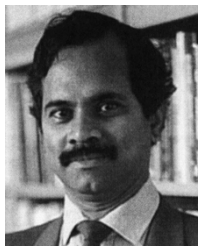
His areas of interest include computer vision and statistical pattern recognition.

Amit K. Roy-Chowdhury received the B.S. degree in electrical Eengineering from Jadavpur University, Calcutta, India, in 1985, the M.S. degree in engineering in systems science and automation from the Indian Institute of Science, Bangalore, in 1997, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Maryland, College Park, in 2002, where he worked on statistical error characterization of 3-D modeling from monocular video sequences.

He is an Assistant Professor in the Electrical Engineering Department, University of California, Riverside. He was previously with the Center for Automation Research, University of Maryland, as a Research Associate. He was involved in projects related to face, gait, and activity modeling and recognition. His research interests are in signal, image and video processing, computer vision, and pattern recognition.

Volker Krüger received the Diploma degree in computer science and math and the Ph.D. degree in engineering from the Christian Albrecht Universität zu Kiel, Kiel, Germany, in 1996 and 2000, respectively. His Ph.D. research was focused on face processing using wavelet networks.

He has done research in wavelet-based image processing. From 2000 to 2002, Dr. Krüger was a Research Assistant and Leading Scientist for the Human ID project at the University of Maryland, College Park. Since 2002, he has been an Associate Professor at Aalborg University, Esbjerg, Denmark. His major research is presently focused on the recognition of identity and actions of humans.



Rama Chellappa received the B.E. (Hons.) degree from the University of Madras, Madras, India, in 1975 and the M.E. (Distinction) degree from the Indian Institute of Science, Bangalore, in 1977. He received the M.S.E.E. and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, in 1978 and 1981, respectively.

Since 1991, he has been a Professor of electrical engineering and an affiliate Professor of Computer Science at the University of Maryland, College Park. He is also affiliated with the Center for Automation Research (Director) and the Institute for Advanced Computer Studies (permanent member). Prior to joining the University of Maryland, he was an Assistant Professor (1981 to 1986) and an Associate Professor (1986 to 1991) and Director of the Signal and Image Processing Institute (1988 to 1990) with the University of Southern California (USC), Los Angeles. Over the last 22 years, he has published numerous book chapters and peer-reviewed journal and conference papers. He has edited a collection of Papers on Digital Image Processing (Los Alamitos, CA: IEEE Computer Society Press, 1992), coauthored a research monograph on *Artificial Neural Networks for Computer Vision* (with Y.T. Zhou) (New York: Springer-Verlag, 1990), and co-edited a book on *Markov Random Fields: Theory and Applications* (with A.K. Jain) (New York: Academic, 1993). His current research interests are face and gait analysis, 3-D modeling from video, automatic target recognition from stationary and moving platforms, surveillance and monitoring, hyperspectral processing, image understanding, and commercial applications of image processing and understanding.

Dr. Chellappa has served as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON IMAGE PROCESSING, and IEEE TRANSACTIONS ON NEURAL NETWORKS. He was Co-Editor-in-Chief of *Graphical models and Image Processing*. He is now serving as the Editor-in-Chief of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE. He served as a member of the IEEE Signal Processing Society Board of Governors from 1996 to 1999. Currently, he is serving as the Vice President of Awards and Membership for the IEEE Signal Processing Society. He has served as a General the Technical Program Chair for several IEEE international and national conferences and workshops. He received several awards, including the National Science Foundation (NSF) Presidential Young Investigator Award, an IBM Faculty Development Award, the 1990 Excellence in Teaching Award from School of Engineering at USC, the 1992 Best Industry Related Paper Award from the International Association of Pattern Recognition (with Q. Zheng), and the 2000 Technical Achievement Award from the IEEE Signal Processing Society. He was elected as a Distinguished Faculty Research Fellow (1996 to 1998) at the University of Maryland, he is a Fellow of the International Association for Pattern Recognition, and he received a Distinguished Scholar-Teacher award from the University of Maryland in 2003.