Gait-Based Human Identification Using Appearance Matching

A. Kale[1], N. Cuntoor [1], B. Yegnanarayana [2], A.N. Rajagopalan[3] and R. Chellappa[1] *

[1] Center for Automation Research
University of Maryland at College Park
College Park, MD 20740

[2] Department of Computer Science and Engineering
[3] Department of Electrical Engineering
Indian Institute of Technology Madras
Chennai-600 036, India

June 11, 2003

EDICS 2-ANAL (Image/Video Processing:Analysis)

Abstract

In this paper, we present an appearance-based approach for recognizing human gait. Given the gait video of an individual, the images are binarized and the width of the outer contour of the silhouette of that individual is obtained for each image frame. Several gait features are derived from this basic width vector. Temporally-ordered sequences of the feature vectors are then used for representing the gait of a person. While matching the feature templates for recognition dynamic time-warping (DTW), which is a non-linear time-normalization technique, is used to deal with naturally occurring changes in the walking speeds of individuals. The performance of the proposed method is tested on indoor as well as outdoor gait databases, and the efficacy of different gait features and their noise resilience is studied. The experiments also demonstrate the effect of change in the viewing angle, characteristics of the floor surface, and frame-rate of data capture on the accuracy of gait recognition.

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

1 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 high enough resolution 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.

Approaches to gait recognition can be broadly classified as being model-based and model-free. Examples of the first kind include the work of Cunado et al [4] who extract 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 using simple harmonics. In [5], two sets of activity-specific static and stride parameters are extracted for different individuals. The expected confusion for each set is computed to guide the choice of parameters under different imaging conditions (viz. indoor vs outdoor, side-view vs angular-view etc). The set of stride parameters (four) which is smaller than the set of static parameters (two) is found to exhibit greater resilience to viewing direction. In a recent work, Lee et al [6] fit ellipses to different parts of the binarized silhouette of a person. The parameters of these ellipses such as the location of the centroid, eccentricity etc are used as features to represent the gait of that person.

Examples of model-free approach to gait include the work of Little and Boyd [7] who 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. Huang et al [8] also use optical flow to derive the motion image sequence corresponding to a gait cycle. Principal components analysis (PCA) is then carried out to derive what are called eigengaits for recognition. Benabdelkader et al [9] use image self-similarity as a gait feature. PCA is used to reduce dimensionality and the K-nearest neighbor rule is applied for classification.

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 size and shape 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 indexing tool that allows us to narrow the search down to a considerably smaller set of potential candidates.

In a pattern classification problem, choice of the feature as well as the classifier is important. We choose the width of the outer contour of the silhouette a person as the basic image feature since it contains structural as well as dynamical information of the person's gait. As will be shown, the outer contour contains sufficient information for recognizing gait. From the raw width vector, different low-dimensional features are derived. These include the smoothed and down-sampled width vector, the eigen-smoothed width vector, and the velocity feature vector. Temporally-ordered sequences of these feature vectors are used for compactly representing the person's gait. Typically, $5 - 10$ contiguous half-cycles of gait data per subject may be available and the number of frames per cycle ranges from 8 to 20. The amount of training data is inadequate for adopting statistical model-based approaches such as the Markov model, as it may not be possible to reliably estimate the parameters of the model. Hence a template-matching is adopted for comparing the probe and the reference sequences of the temporally-ordered feature vectors. Typically, gait cycles when taken at different times tend to be unequal in length due to changes in walking speeds of the individuals. Hence, a classifier based on direct template-matching is not appropriate. To deal with this issue, dynamic time-warping (DTW) is employed for matching gait sequences. The DTW concept has been quite successfully used by the speech community for text-dependent speaker recognition/verification [10]. DTW uses an optimum time expansion/compression function for producing non-linear time normalization so that it is able to

deal with misalignments and unequal sequence lengths of the probe and the reference gait sequences. Importantly, DTW can be used even with limited training data.

The performance of our approach is tested on five standard gait databases; namely, the University of Maryland (UMD) database, the MIT database, the Carnegie Mellon University (CMU) database, the University of Maryland angular-walk database (UMD3) and the University of South Florida (USF) database. These databases contain video sequences of individuals walking in a normal manner along certain pre-defined paths. The UMD and CMU databases have both frontal and side views, the MIT database has only side-view sequences, while the USF database is not restricted to side or frontal views. The proposed DTW-based approach is equally applicable to side as well as frontal views and the results are analyzed to determine the factors that contribute to accuracy of gait recognition. The idea is to study the efficacy of the features derived from the basic width vector and their resilience to noise in the estimate of the width vector. Our experiments also reveal the effect of difference in walking speeds, and the relevance of different body parts for gait recognition. Low frame-rate during data capture and methods to mitigate its effect are discussed. The effect of change in viewing angle on gait recognition is examined. It will be shown that the accuracy also depends on the surface characteristics of the floor.

The organization of the paper is as follows. In Section 2, the basic width vector, its extraction, and its relevance for the gait problem are explained. Section 3 discusses different low-dimensional features derivable from the basic width vector for gait representation. In Section 4, we describe how gait sequences can be matched using dynamic time-warping. Section 5 briefly reviews the gait databases used in our studies. Experimental results are discussed in Section 6. Conclusions and suggestions for further studies are given in Section 7.

2 The Width Vector

An important issue in gait-based human identification is the extraction of salient features that will effectively capture gait characteristics. In order to be independent of the color and texture of clothing and illumination, it is reasonable to consider the binarized silhouettes of a subject. Given the image

sequence of a subject, background subtraction [11] is applied to the image sequence. The resultant motion image is binarized to get the silhouettes (see Figure 1). A bounding box is placed around the part of the motion image that contains the outer contour of the moving person. The width along each row of the silhouette, computed as the difference in the locations of the right-most and the left-most boundary pixel in that row, is stored. The physical structure of the subject as well as the swing of the limbs and other details of the body are retained in the width vector thus derived but the pose information is lost. For the four individuals shown in Figure 1, an overlay of the width vectors derived from the silhouettes is given in Figure 2. The width-vector plots clearly bring out the fact that there is relatively more swing in the middle region (corresponding to hands) and in the end region (corresponding to feet), as compared to other parts of the body. However, these plots do not depict the temporal information. To bring out the temporal effects explicitly, the width vectors are plotted as a sequence of gray-level patterns in Figure 3 for the same four individuals. In this figure, about 5 full gait cycles are shown. The vertical axis corresponds to the index of the width vector, while the horizontal axis corresponds to the frame number. The total number of frames for each plot was fixed at 110 for uniformity. We call this the *temporal* plot of the width vector. It is clear that the width vector is approximately periodic and gives the extent of movement of different parts of the body. It varies with time in a quasi-periodic fashion, with the time taken for the completion of a half-cycle treated as the period.

Every component of the width vector contributes to the gait signature of a subject. The brighter a pixel, the larger is the value of the width vector in that position. The top part in each of the plots corresponds to the head-region, the middle part corresponds to the torso while the bottom part corresponds to the foot-region of the individual. Note that the intensity variations in the torso and leg region, (corresponding to the hand and leg-swing) are larger than in the head region. The extent of the dark region near the top and the bottom of the temporal width plots reflects the height differences between the individuals.

A comparison of the temporal width plots across individuals reveals interesting gait information. The structural or static differences among people is clearly revealed by the extent of the non-zero

portions in these plots. One can also decipher dynamic information. For instance in Figure 3, the hand motion in the case of Person 1 is more pronounced as compared to Person 2. Note that the intensities are brighter for Person 1 in the middle region compared to Person 2. There also exist visible differences in the brightness gradients or velocities for different people. i.e., the velocity profile for each individual is different. There are differences in the repetition frequency which may be useful to distinguish people at a gross level. For instance, Person 3 has 12 half-gait cycles for the same number of frames compared to 10 $\frac{1}{2}$ cycles for Person 1 and 10 cycles for Person 2. For most female walkers, the frequency was found to be higher compared to their male counterparts. Thus, considering a gait (half)-cycle as a unit of observation, we can capture both spatial and temporal characteristics of an individual's gait signature over several cycles.

3 Gait Representation

As discussed, the width vector captures important characteristics about an individual's gait. However, the dimension of the raw width vector can be quite high (168 in our experiments). In this section, we enumerate different low-dimensional but effective features that can be derived from the basic width vector. The idea is to arrive at a compact representation that exploits redundancy in the gait data for dimensionality reduction. Features that are directly obtained from the basic width vector are called direct features and these include the smoothed and down-sampled versions of the width vector. On the other hand, features derived using an eigen-analysis of the width vector across frames are called eigenfeatures. Corresponding to every frame of the gait cycle of an individual, one can extract any of the feature vectors discussed here. The gait of the individual is then represented by temporally ordering the feature vectors in accordance with the image frames in the gait cycle of that individual. We now describe each of the features in detail.

3.1 Direct Features

Examples of direct features are shown in Figure 4.

1. *The raw width vector*: This is the basic width vector discussed in Section 2, and is derived directly from the video sequence.

2. *3-point smoothed width vector*: This is derived from the raw width vector by smoothing it with a 3-point mean filter.

3. *5-point, 11-point, and 21-point smoothing and down-sampling by a factor of 4, 8 and 16, respectively*: The motivation behind smoothing and down-sampling stems from the fact that the original width vector has redundancies. Hence, it should be possible to discriminate reasonably well using lower dimensional features. Also, the original raw width vector is derived by simple and quick pre-processing. Hence, it is usually noisy and smoothing helps to mitigate the effect of noise.

4. *Difference feature vector*: This is useful to study the effect of dynamics for gait-based human identification. There are many different ways to extract dynamic information from the width vector. One would be to simply take the difference of successive raw width vectors across frames thus preserving only the changes that occur between frames, during a gait cycle followed by smoothing and downsampling. There is a trade-off between smoothing and extracting dynamic information. Although some degree of smoothing is required to counter noise, too much smoothing can alter the dynamic information. Obviously, most of the structural information, like girth of the person, is lost when we go to the velocity domain. It is to be expected that neither dynamic nor structural information, in isolation, will be sufficient to capture gait. As will be shown later, both are necessary and cannot be decoupled.

3.2 Eigenfeatures

From the temporal width plots, we note that although the width vector changes with time within a gait cycle, there is a high degree of correlation among the width vectors across frames. Most changes occur in the hand and in the leg regions. The rest of the body parts do not undergo any significant changes during a gait cycle. Hence, it is reasonable to expect that gait information in the width vector can be derived with much fewer coefficients. This is what we attempt to do in the eigenanalysis of gait.

All the width vectors corresponding to several training gait cycles of an individual are used to construct the covariance or scatter matrix for that individual. Principal components analysis is then performed to derive the eigenvectors for this data.

Given the width vectors $\{W(1), \cdots, W(N)\}$,for N frames $W(.) \in R^M$, we compute the eigen vectors $\{V(1,) \cdots, V(M)\}$ corresponding to the eigen values of the scatter matrix arranged in the descending order and reconstruct the corresponding width vectors using $m(< M)$ most significant eigen vectors as

$$W_r(i) = (\sum_{j=1}^{m} w_j V(j)) + \bar{W}$$

where $w_j =< W(i), V(j) >$ and $\bar{W} = \frac{W(1)+\cdots+W(N)}{N}$.

Note that every individual has his/her own eigengait space. Reconstruction of the width vector using only the first and the second eigenvectors is shown in Figure 4(f). The consequence of approximating the width vector using only the first two eigenvectors is that the effect of noise is suppressed and the width vector is smooth. The first two eigenvectors capture the physical structure of the person as well as the typical arm and leg swings, and serve as a freeze-frame representation of the gait sequence.

4 Matching Gait Sequences using DTW

Since only limited training data is usually available for gait analysis, a template-matching approach is adopted in this paper for recognition. The input to the gait recognition algorithm is a sequence of image frames. Each image frame can be compactly represented by any of the feature vectors discussed in Section 3 to derive a template which consists of a temporally ordered sequence of feature vectors for representing gait. The number of frames in the reference gait sequence and the probe gait sequence depend on the number of gait cycles available. Larger the number of gait cycles used for matching, the better we can expect the performance to be. Typically, the number of frames in the reference and probe data will differ. Moreover, the reference and probe gait data are seldom synchronized. Therefore, direct matching of a probe gait sequence with a reference gait sequence should be avoided. In this work, a pattern-matching method based on dynamic programming paradigm is used for dealing with

this situation. Dynamic Time Warping (DTW) [10] can be used to compensate for the variability in the speed of walking which in turn reflects in the number of frames for each gait cycle. A distance metric (usually the Euclidean distance) defined as a function of time is computed between the two feature sets representing the gait data. A decision function is arrived at by integrating the metric over time. The DTW method was originally developed for isolated word recognition [12], and later adapted for text-dependent speaker verification [10]. The gait problem is analogous to text-based speaker identification/verification wherein different individuals utter the same text but differ only in the characteristics of their utterance [10].

Consider the $x-y$ plane shown in Figure 7 where the x and the y axes represent the frame numbers of the probe and the reference patterns, respectively. Assume that the first frame of the reference and probe sequence are both indexed as 1. Let the last frames of the reference and probe sequences be indexed as X and Y, respectively. The match between the two sets can be represented by a sequence of K points $C(1), C(2), ..., C(k), ..., C(K)$, where $C(k) = (x(k), y(k))$, and $x(k)$ is a frame index of the probe sequence and $y(k)$ is a frame index of the reference sequence. Here, $C(k)$ represents mapping of the time axis of probe sequence onto that of the reference sequence. The sequence

$$F = C(1), C(2), ...C(k), ..., C(K)$$

is called the warping path.

The process of time normalization uses certain constraints depending on the problem at hand. Some of the constraints are the following [13].

(a) Endpoint Constraints: The fixed endpoints of the patterns lead to a set of constraints on the warping function. They are of the form

$$x(1) = 1,\ y(1) = 1,\ x(k) = X, \text{ and } y(k) = Y \tag{1}$$

i.e., the first and the last frames of the probe sequence should be matched with the first and the last frames of the reference, respectively. Under these constraints, a typical warping would look like the one shown in Figure 7.

(b) Monotonicity Constraint: This constraint requires feature vectors belonging to a probe sequence (or the reference sequence) to be matched in monotonically increasing order. This maintains the temporal order of the sequence during time normalization. This is accomplished by ensuring that

$$x(k-1) \leq x(k), \text{ and } y(k-1) \leq y(k). \tag{2}$$

(c) Local Continuity Constraints: This constraint can be used to ensure use of each frame probe frame during normalization, and also to ensure that no more than one frame is skipped in the reference. This is achieved by setting the constraints

$$x(k) - x(k-1) = 1, \text{ and } y(k) - y(k-1) \leq 2 \tag{3}$$

(d) Global Path Constraints: This constraint restrains the extent of compression or expansion. The warping path can be constrained by defining a region around which the warping path is allowed to traverse. The region between the two parallel lines marked in the $x-y$ plane in Figure 6 defines the allowable grid or the region through which the warping path can traverse. The warping path can be limited to a band around the diagonal so that the search time for optimal path is significantly reduced. This is reasonable as it is unlikely that the probe sequence of the genuine case would deviate significantly from its reference.

After deciding the local and global constraints, the DTW algorithm is applied as follows.

1. Local Distance Computation: This involves computing the distance between the feature vectors representing the probe and reference frames. The distance between the probe and reference feature vectors $\mathbf{t_x}$ and $\mathbf{r_y}$ can be computed using

$$d(C(k)) = |\mathbf{t_x} - \mathbf{r_y}|$$

so that the total distance along the path F is given by

$$E(F) = \sum_{k=1}^{K} d(C(k)) \tag{4}$$

The distance $E(F)$ is the similarity score which can be used at the decision making stage.

2. Cumulative Distance Computation: The cumulative distance $D(x(k), y(k))$ is the minimum distance to reach $(x(k), y(k))$ at the k^{th} stage starting from $(1, 1)$. This distance is the sum of all the local distances of points through which the warping path passes to reach $(x(k), y(k))$ from $(1, 1)$. Initialize the cumulative distance $D(1, 1)$ to $L(1, 1)$, where $L(1, 1) = d(C(1))$. For all the other points lying within the global region of search, compute the cumulative distance using the local path constraints, i.e., the point $(x(k), y(k))$ can only be reached from the points $((x(k) - 1), y(k))$ or $((x(k) - 1), (y(k) - 1))$ or $((x(k) - 1), (y(k) - 2))$. If the distance at stage k is obtained as

$$D(x(k), y(k)) = \min \begin{cases} D((x(k) - 1), y(k)) + L((x(k), y(k)) \\ D((x(k) - 1), (y(k) - 1)) + L(x(k), y(k)) \\ D((x(k) - 1), (y(k) - 2)) + L(x(k), y(k)) \end{cases} \quad (5)$$

then $D(X, Y)$ gives the distance between the probe and the reference sequences.

3. Backtracking: Using the cumulative distance matrix and the local path constraints, the warping path F can be obtained by backtracking.

4.1 Implementation details

Use of dynamic time warping relies on similar start and end points for the probe and gallery. Hence prior to applying the DTW algorithm it is necessary to parse the video sequence into cycles. A simple way to achieve this is using the width feature explained in Section 2. From Figure 3 it is easy to see that $N(t) = (\sum_{j=1}^{M} I(j))(t)$ where $I(j)$ denotes the width at row j of the image, viz. the sum of widths will show a periodic variation (see Figure 4). The troughs of the resulting waveform correspond to the rest positions during the walk cycle while the crests correspond to the part of the cycle where the hands and legs are maximallly displaced. A half-cycle consists of frames between two successive troughs of sum of intensities plot. In general, the exact trough may be hard to determine in the presence of noise and picking a frame in its neighborhood is usually adequate. Given a video sequence, the width feature thus provides a natural parsing in terms of half-cycles. In our experiments, four half-cycles from the probe sequence are matched with four half-cycles from the gallery sequence. Euclidean distance was chosen as the local distance.

A global constraint of $max(0.1N_{gallery}, 0.1N_{probe})$ was chosen. Using the similarity matrix resulting from the matching algorithm, the cumulative match characteristic as explained in [14]. Essentially the cumulative match characteristic is a plot of the number of times the right person occurs in the top n matches where $n < G$ where G denotes the number of people in the gallery, as a function of n.

5 Data

The UMD Database: Gait data was captured with the out-door video research (OVR) facility at the University of Maryland [1]. Cameras located at right angles and at a height of 4.5 meters above the ground were used for data capture. The UMD dataset consists of 44 individuals. There are 38 male and 6 female subjects with different ethnicity, physical build etc. The individuals were asked to walk normally along a T-shaped path. Because the cameras are orthogonally placed and the path is T-shaped, a side-view and a frontal view could be captured simultaneously as the person walked. The videos for each individual were recorded with intervals ranging from about half-a-day to a maximum of one month. Each image frame was of size 170×138 pixels. Labeling was done to mark the gender, test/training sequences, direction of walk, number of gait cycles, and left/right foot forward.

The CMU Database: The CMU database [2] consists of different views of indoor gait sequences of 25 individuals walking on a treadmill. Different situations such as a person walking at a slow pace, fast pace, and walking while carrying a ball in his/her hands are considered. Even though it is an indoor database, it serves as a good testbed for evaluating performance under variations in walking speed and also to assess the relative importance of body parts. For example, in the sequence where the person is carrying the ball, there is very little upper-body movement. The size of each image frame in this database is 640×486 pixels.

The MIT Database: The MIT database [3] consists of side-views of gait sequences of 25 subjects

[1] http://degas.umiacs.umd.edu/hid
[2] http://hid.ri.cmu.edu
[3] http://www.ai.mit.edu/people/llee/HID/intro.htm/

captured indoors, and collected on four different days. There are 4-8 segments per subject per day with 2-5 contiguous half cycles per segment. The number of subjects on a particular day varies from 4 to 16. As a result, the degree of overlap across days is less. The frame rate of the camera is only 15 frames per second (as compared to 25 for the UMD database), and the image size is 128×104 pixels. A lower frame rate leads to sparse temporal sampling of the gait information and must be tackled appropriately. This database provides a good testbed for evaluating training and test databases captured on different days.

The UMD3 Database: The UMD3 database consists of outdoor gait sequences of 12 people walking along straight-line paths at azimuth angles of $0, 15, 30, 45$ and 60 degrees from the side view. The data was captured at the rate of 25 frames per second by a tripod-mounted camera. Two sequences of each person were captured with a time interval of at least 1 hr between them. One of the sequences was used as reference and the other was used as the probe. This database provides a testbed for assessing gait recognition performance as a function of the viewing angle. The image size for zero azimuth is 210×105 pixels.

The USF Database: In this database [15], each subject was required to walk in a counter-clockwise manner along two elliptical courses. The elliptical courses are located outdoors and approximately 15 meters along the major-axis and 5 meters along the minor-axis. To enable testing on different floor surfaces, the first course was laid out on flat concrete, while the second was laid out on a typical grass-lawn. Each course was viewed by two cameras. Their lines of sight were not parallel but verged so that the whole ellipse was just visible from the two cameras. When the persons walked along the rear portion of the ellipse, their view was approximately sideways. Subjects were also asked to bring a second pair of shoes, so that they could walk a second time in a different pair of shoes. A little over half of the subjects walked in two different shoe-types. The image size of the subjects is 100×50 pixels. The database has 70 individuals of which 75% are males.

6 Experimental Results

In this section, we present the results of our approach to gait recognition on all the databases described in Section 5. Each of the databases has its own interesting features, and our exercise reveals several interesting facts about the gait-based human identification problem. The UMD database has gait sequences for side as well as frontal views. Hence, the performance of all the features discussed in Section 3 were first studied for this database to get a feel for what might be a good feature vector for gait representation and recognition. The method described in Section 4 was used for recognition.

6.1 Results for the UMD Database

For this database there were approximately 40 frames corresponding to 4 half-cycles in the gallery and probe. The recognition results for the direct features are given in Table 1. From the table, we note that all the features have done well despite the presence of noise in the estimate of the basic width vector. The experiment also shows that dimensionality reduction is possible to a great extent by exploiting the redundancy in the gait data. In the experiments to come, we choose the 5-point smoothed 42-dimensional width vector, which is between the two extremities of no-smoothing and very heavy-smoothing, as the direct feature for performance analysis.

It must be mentioned here that similar to hand-dominance (right/left handedness), foot-dominance (right/left leggedness) also exists in individuals. In practice, it is difficult to align the reference and the probe sequences accurately with respect to heel-strike. The result of ignoring heel-strike information (left-foot forward or right-foot forward first) in gait analysis is reflected in Table 2. Suppose there are five half-cycles in both reference and probe sequences for a particular subject. The first four half-cycles of the two sequences are matched to generate a matrix of similarity scores. Then, the reference sequence is matched with a probe sequence shifted by half-cycle to generate another matrix of similarity scores. Of the two shifted probe sequences, only one can provide a better match if the subject exhibits foot-dominance. The two similarity scores are combined using the minimum-error criterion. The improvement in recognition when heel-strike is taken into account can be noted from Table 2.

For the eigenfeatures, the performance results obtained by combining different numbers of eigenvectors is given in Table 3. Again, four half-cycles of each subject were used for matching. Note that by using just the first two eigenvectors an accuracy of 80% is achievable. Other eigenvectors are noisy and, in fact, tend to lower the accuracy. Hence, we have used only the first two eigenvectors for computing the eigenfeatures in our experiments. It is also of interest to study the effect of the number of half cycles on matching. Intuitively, one would expect to obtain a better match with more number of half cycles, i.e., with a longer gait sequence. When the number of half-cycles is reduced, systematically from four to one, the corresponding results are given in Table 5. While reducing the number of half-cycles, the number of frames was kept approximately constant by linearly interpolating the components of the width vector. As expected, the performance degrades as the length of the sequence is reduced. **A comparison of the results using our method using two eigenvectors with the Lee and Grimson [6] and Bobick [5] is given in Table 4**

To assess the relative discriminability of the structural component of the width vector versus its dynamic component, we computed the difference of the width vectors corresponding to successive frames in the gait sequence. Eigen-decomposition of the velocity vector is carried out and the vector reconstructed using the first two eigenvectors was computed. Table 6 shows the results obtained by considering only the velocity profile as the feature of interest. Alternatively, the width vector could have been first subjected to eigendecomposition, and the successive differences of the projected values used for matching. Note that the accuracy drops significantly if only the velocity information is used. Clearly, for gait-based human identification both structural and dynamic information are important.

6.1.1 Frontal Gait

Here, the aim is to recognize humans using the frontal gait information as a cue. As a subject walks toward the camera, an apparent increase in the size of the subject and in the swing of the arms and legs can be observed. The outer contour of the subject carries information that is useful in discriminating among subjects. Figure 6(a) shows the variation of the width vector (that reflects in the outer contour) as a function of time. The size of the width vector grows indicating that the subject is approaching the

camera. We retain the last four half cycles of each sequence as shown in Figure 6(b), and the raw width vector itself is used for matching. The results are tabulated in Table 7. To account for the apparent change in the size of the subject, we normalize the width vectors by computing an appropriate scale factor. The position of the head and the feet (top and bottom pixels) are identified in each frame and smoothed using a median filter. To the resulting sinusoidal patterns shown in Figure 6(c), two straight lines, one to the top and one to the bottom, are fit. The distance between the two lines in each frame is used to compute the normalizing factor. The normalized plot is shown in Figure 6(d). The recognition rates obtained before and after normalization are summarized in Table 7. Note that the accuracy for the frontal gait sequences is lower as compared to the side-view case. This is because the swing is less pronounced now, and the dynamics are not as effectively captured as in the case of the side-view. For the UMD database, because we have two orthogonal cameras taking pictures simultaneously, we get both frontal and side-view gait sequences. The improved results obtained by combining the evidence from both the views at the decision level are given in Table 7.

6.2 Results for the CMU database

The following experiments were performed on this database: (i) Train on slow-walk and test on slow-walk, (ii) train on fast-walk and test on fast-walk, (iii) train on slow-walk and test on fast-walk and (iv) train on walk carrying a ball and test on walk carrying a ball.**For this database, the number of frames corresponding to 4 half-cycles varied between 55 for slow walk to about 75 for fast walk.** The results are given in Table 8.

We note that the eigensmoothed feature performs better than the direct-smoothed feature. This can be attributed to the fact that eigensmoothing exploits spatio-temporal redundancy unlike direct-smoothing which uses only spatial smoothing. When the reference is the slow-walk sequence and the probe is the fast-walk sequence, the performance is found to be inferior than for the case when the reference and probe are both slow-walk sequences. DTW is known to perform poorly [16] when the ratio of reference sequence-length to probe sequence-length is less than 0.5 or more than 2. In the CMU dataset, this ratio for the worst case was 1.36. From Table 7, we note that the DTW-based method is

reasonably robust to changes in walking speed. In fact, in our experiments, the value of the ratio for one of the mismatched cases was 1.15. A few frames in the gait cycles of this incorrectly recognized person under slow and fast-walk modes are shown in Figures 8 (a) and (b). As is apparent from the figure, the posture as well as hand-swings for the person are quite different in cases of fast-walk and slow-walk. Figures 8 (c) and (d) show the warping paths for the person with the highest ratio and the incorrectly recognized person, respectively. Note that the warping path for the correctly recognized person is much more regular as compared to that of the incorrectly recognized individual. Hence, it is the change in the posture and body dynamics of the person rather than the mismatch in the length of the reference and probe sequences that was responsible for the mismatch.

Finally, we consider the case when the person is walking with a ball in his hand. In this situation, most of the gait dynamics is confined to the leg region. For this experiment, we observe from Table 9 that the recognition performance is very good (person identified correctly within the top 2 matches). This experiment suggests that for the purpose of recognition certain parts of the body may be more preferable than the others. Incidentally, this has been noted in kinesiology research as well [17].

6.3 Results for the MIT Database

The evaluation scheme for the MIT database is as follows. For training, data collected on days 2, 3 and 4 were used. Data collected on day 1 was used for testing. Overall, there are 8 subjects. We extract 2 half-cycles (which is the maximum number of half-cycles common to all the subjects in the database) and build the direct-smoothed width vector. Since the frame rate is 15 frames/second, on the average, there are only 7-8 frames per half-cycle **so that there were about 28 frames in four half-cycles**. We recognize that the length of the gait sequence is not adequate. As discussed earlier, the more the number of half-cycles, the better would be the possibility of obtaining a correct match. Nevertheless, 4 out of the 8 people were correctly recognized. Since the frame-rate is low for this database, it results in coarse-sampling in the temporal domain and presents a problem while matching. When we used linear interpolation between frames to partially compensate for the coarse sampling, the recognition rate increased to 5 matches out of 8. When the eigenfeatures were tried on this database, the results

were not very different.

6.4 Results for the UMD3 database

The UMD3 database contains gait data corresponding to different azimuth angles. The gait sequences were captured at different times at angles $\theta = 0$, 15, 30, 45 and 60 deg with respect to the side-view. To account for the foreshortening effect as a person walks at non-zero azimuths, the height of each individual was computed from the zero-azimuth width plots, and this was used to normalize the height corresponding to non-zero azimuth angles. Figure 9 shows the width plots for one of the individuals in the database for each of the above angles. **There were approximately 60 frames corresponding to 4 half-cycles for the UMD3 database.** It can be seen that for large azimuth angles the hand swings appear only in alternate half-cycles. This is because the body occludes the presence of hands in the alternate half-cycles. Another observation is that the range of variations in the width amplitude is small for large azimuth angles. Therefore, recognition performance can be expected to be more sensitive to the presence of noise in the gait sequence for higher values of θ.

Two full-cycles having a relative shift of one half-cycle were used as gallery sequence (similar to the first UMD database) and two full-cycles captured at a later time were chosen as the probe sequence. Both the 5-point direct-smoothed feature and the eigensmoothed feature were tested for recognition and the results are given in Table 10. In the table, L and R refer to two full gait-cycles with a relative shift of one half-cycle, while F refers to the case when we accept smaller of the two entries corresponding to L and R. It is clear that $\theta = 0$, which is the side-view, provides the best viewing direction for gait recognition. Since the variation in the amplitude of the width vector was small for large θs, eigensmoothing results in better performance. For smaller values of θs when the width vector components are relatively large, eigensmoothing does not seem to yield any significant improvement over direct smoothing.

6.5 Results for the USF Database

The USF database has the largest number of individuals among all the databases. It has variations with respect to floor surface (grass (G) or concrete(C)), shoe type (A or B), and camera viewing direction (left (L) or right (R)). The reference for all the experiments was chosen to be (G, A, R). **The number of frames corresponding to four half cycles varied from 65 to 90.** Different probe sequences for the experiments along with the cumulative match scores are given in Table 11 for the baseline algorithm [15] as well as our method using eigenfeatures. Note that recognition performance suffers most due to difference in surface characteristics, and least due to difference in viewing angle. However, unlike the UMD3 experiment, both the cameras see an 'almost' side-view of the subjects.

7 Conclusions

In this paper, we have proposed a new approach to gait recognition using the dynamic programming paradigm. The method was tested on five different databases using several features. The width of the outer contour of the binarized silhouette of a person was used as the gait feature. Different feature vectors were derived from the basic width vector using either direct smoothing and down-sampling or by way of eigenanalysis for gait representation and matching. During the matching process, a dynamic time warping algorithm was used to effectively handle unequal lengths of the reference and the probe gait sequences. Various situations such as changes in walking speed, sensitivity to viewing angle, different floor characteristics, low frame-rate etc were considered. The performance of our method was found to be quite satisfactory on all the databases. The eigensmoothed feature performed the best followed by the direct-smoothed feature vector. The velocity profile of the width vector is not sufficient to capture gait characteristics when used in isolation. Our experiments confirm that the side-view is the optimal one for capturing gait characteristics. As the azimuth angle increases, the recognition accuracy falls. The method is capable of recognizing even frontal gait but with lesser accuracy as compared to that of the side-view.

It would be interesting to analyze if the warping path characteristics of the DTW algorithm can

be used to determine the direction of motion. Also, a study that combines evidences from different component level features of the human body for the purpose of recognition deserves investigation.

References

[1] J. Cutting and L. Kozlowski, "Recognizing friends by their walk:gait perception without familiarity cues," *Bulletin of the Psychonomic Society*, vol. 9, pp. 353–356, 1977.

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

[3] M.P. Murray, "Gait as a total pattern of movement," *American Journal of Physical Medicine*, vol. 46, pp. 290–332, June 1967.

[4] D. Cunado, J.M. Nash, M.S. 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*, pp. 43–48, 1995.

[5] A.F. Bobick and A. Johnson, "Gait extraction and description by evidence-gathering," *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 423–430, 2001.

[6] L. Lee and W.E.L. Grimson, "Gait analysis for recognition and classification," *Proceedings of the IEEE Conference on Face and Gesture Recognition*, pp. 155–161, 2002.

[7] J. Little and J. Boyd, "Recognizing people by theirgait: the shape of motion," *Videre*, vol. 1, no. 2, pp. 1–32, 1998.

[8] P.S. Huang, C.J. Harris, and M.S. Nixon, "Recognizing humans by gait via parametric canonical space," *Artificial Intelligence in Engineering*, vol. 13, no. 4, pp. 359–366, October 1999.

[9] R. Cutler C. Benabdelkader and L.S. Davis, "Motion based recognition of people in eigengait space," *Proceedings of the IEEE Conference on Face and Gesture Recognition*, pp. 267–272, 2002.

[10] Sadaoki Furui, "Cepstral analysis technique for automatic speaker verification," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 29, no. 2, pp. 254–272, April 1981.

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

[12] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-26 no. 1, pp. 43–49, 1978.

[13] Jinu Mariam Zacharia, "Text-independent speaker verification using segmental, suprasegmental and source features," M.S. thesis, IIT Madras Chennai India, March 2002.

[14] 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.*, vol. 22, no. 10, pp. 1090–1100, October 2000.

[15] P. J. Phillips, S. Sarkar, I. Robledo, P. Grother, and K. W. Bowyer, "The gait identification challenge problem: Data sets and baseline algorithm," *Proc of the International Conference on Pattern Recognition*, 2002.

[16] L. Rabiner and H. Juang, *Fundamentals of speech recognition*, Prentice Hall, 1993.

[17] W. I Scholhorn, . Nigg B.M, D.J.Stephanshyn, and W. Liu, "Identification of individual walking patterns using time discrete and time continuous data sets," *Gait and Posture*, vol. 15, pp. 180–186, 2002.

Feature\Rank	1	2	3	4	5
Unsmoothed 168-dim feature	79.0	81.4	83.2	86.0	86.0
3-point smoothed 168-dim feature	79.0	81.4	83.7	86.0	86.0
5-point smoothed 42-dim feature	76.7	83.7	83.7	88.3	88.7
11-point-smoothed 21-dim feature	76.7	83.7	83.7	83.7	90.0
21-point-smoothed 11-dim feature	79.0	86.0	86.5	88.3	90.0

Table 1: UMD database: Cumulative match scores (CMS) for the first five ranks for different directly-derived feature vectors.

Feature\Rank	1	2	3	4	5
Set of 4 half cycles 1(a)	68.1	77.2	84.0	84.0	84.0
Set of 4 half cycles shifted by one 1(b)	70.4	79.5	81.8	86.3	86.3
Minimum of 1(a) and 1(b)	79.0	81.4	83.2	86.0	86.0

Table 2: UMD database: CMS for the first five ranks for the case of 2 full-cycles having a relative phase shift of 1 half-cycle.

Feature\Rank	1	2	3	4	5
Eigenvector 1	73	75	80	80	84
Eigenvectors 1,2	80	87	90	90	91
Eigenvectors 1,2,3	68	80	84	84	84
Eigenvectors 1,2,3,4	73	77	84	84	84
Eigenvectors 1,2,3,4,5	70	73	79	82	84
Eigenvector 2	58	63	67	72	74
Eigenvectors 2,3	61	64	67	82	76
Eigenvectors 2,3,4	61	65	68	68	72
Eigenvectors 2,3,4,5	68	68	72	74	74
Eigenvector 3	51	54	60	62	62
Eigenvectors 3,4	55	60	68	68	72
Eigenvector 4	22	28	35	42	42
Eigenvector 5	20	24	36	40	44

Table 3: CMS for the UMD database using different eigen-features.

Experiment	CMS Scores	
	Rank 1	**Rank 5**
Our approach(using 2 eigenvectors)	**80**	**91**
GaTech[5]	**17**	**42**
MIT[6]	**41**	**58**

Table 4: Comparison of different algorithms on UMD database.

Feature\Rank	1	2	3	4	5
Four half-cycles	80.0	87.0	90.0	90.0	91.0
Two half-cycles	71.0	80.0	82.0	84.0	87.0
One half-cycle	65.0	67.0	77.0	83.0	87.0

Table 5: Effect of length of gait sequence for the UMD database: CMS for the first five ranks.

Feature\Rank	1	2	3	4	5
Smoothed and differenced 168-dim feature	41.9	51.6	61.2	70.9	74.1
Eigen decomp. of velocity profile	56	75	76	80	83
Eigen decomp. of width vector and successive difference	32.0	40.0	51.0	61.0	65
of projected weights					
Eigen decomp. of width vector and using 2nd eigenvector	58.0	63.0	67.0	72.0	74.0

Table 6: UMD database: Cumulative match scores for the velocity profile.

Feature\Rank	1	2	3	4	5
Raw width vector (w/o normalization)	66.0	67.0	74.0	81.0	86.0
Normalization in the vertical direction	69.0	74.0	86.0	90.0	90.0
Normalization in both vertical and horizontal directions	71.0	84.0	87.0	87.0	87.0
Frontal + Side-view (decision based on fusion)	85.0	87.0	92.0	94.0	95.0

Table 7: CMS results for the frontal gait sequences of the UMD database.

Experiment	Feature	Rank				
		1	2	3	4	5
Slow vs Slow	Direct-smoothed feature	70.8	83.3	87.5	95.8	95.8
	Eigen-smoothed feature	95.8	95.8	95.8	95.8	100.0
Fast vs Fast	Direct-smoothed feature	83.3	83.3	83.3	83.3	87.5
	Eigen-smoothed feature	95.8	95.8	95.8	95.8	100.0
Fast vs. Slow	Direct-smoothed feature	54.1	75.0	87.5	87.5	87.5
	Eigen-smoothed feature	75.0	83.3	83.3	83.3	87.5

Table 8: CMU database: Recognition results for different experiments.

Experiment\Rank	1	2	3	4	5
Ball vs Ball	95.45	100.00	100.00	100.00	100.00

Table 9: CMS for the ball experiment in the CMU database using the eigen-smoothed feature.

Angle from side-view	Sequence chosen	Rank (direct)			Rank (eigen)		
		1	2	3	1	2	3
0 deg	L	75.00	91.67	100.00	91.67	91.67	100.00
	R	75.00	83.33	91.67	83.33	83.33	91.67
	F	91.67	91.67	100.00	91.67	91.67	100.00
15 deg	L	66.67	83.33	100.00	91.67	91.67	100.00
	R	58.33	75.00	83.33	91.67	91.67	100.00
	F	91.67	91.67	100.00	91.67	91.67	100.00
30 deg	L	66.67	83.33	91.67	83.33	83.33	91.67
	R	25.00	50.00	50.00	58.33	91.67	100.00
	F	83.33	83.33	91.67	83.33	83.33	100.00
45 deg	L	33.33	41.67	50.00	41.67	66.67	91.67
	R	66.67	75.00	83.33	83.33	91.67	100.0
	F	50.00	83.33	100.00	58.33	83.33	100.00
60 deg	L	33.33	50.00	58.33	58.33	75.00	91.67
	R	41.67	58.33	58.33	75.00	83.33	91.67
	F	58.33	83.33	91.67	75.00	83.33	91.67

Table 10: UMD3 database: CMS for the first three ranks for different azimuth angles using the direct and the eigen-feature vectors.

Experiment	Probe	Difference	Baseline		Our Approach	
			Rank 1	Rank 5	Rank 1	Rank 5
A	(G, A, L)	View	79	96	79	91
B	(G, B, R)	Shoe	66	81	67	79
C	(G, B, R)	Shoe, View	56	76	30	55
D	(G, B, L)	Surface	29	61	17	42
E	(C, A, R)	Surface, Shoe	24	55	15	39
F	(C, A, L)	Surface, View	30	46	16	30
G	(C, B, L)	Surface, Shoe, View	10	33	9	31

Table 11: Probe Sets and match scores for the USF database using the baseline algorithm and our approach.

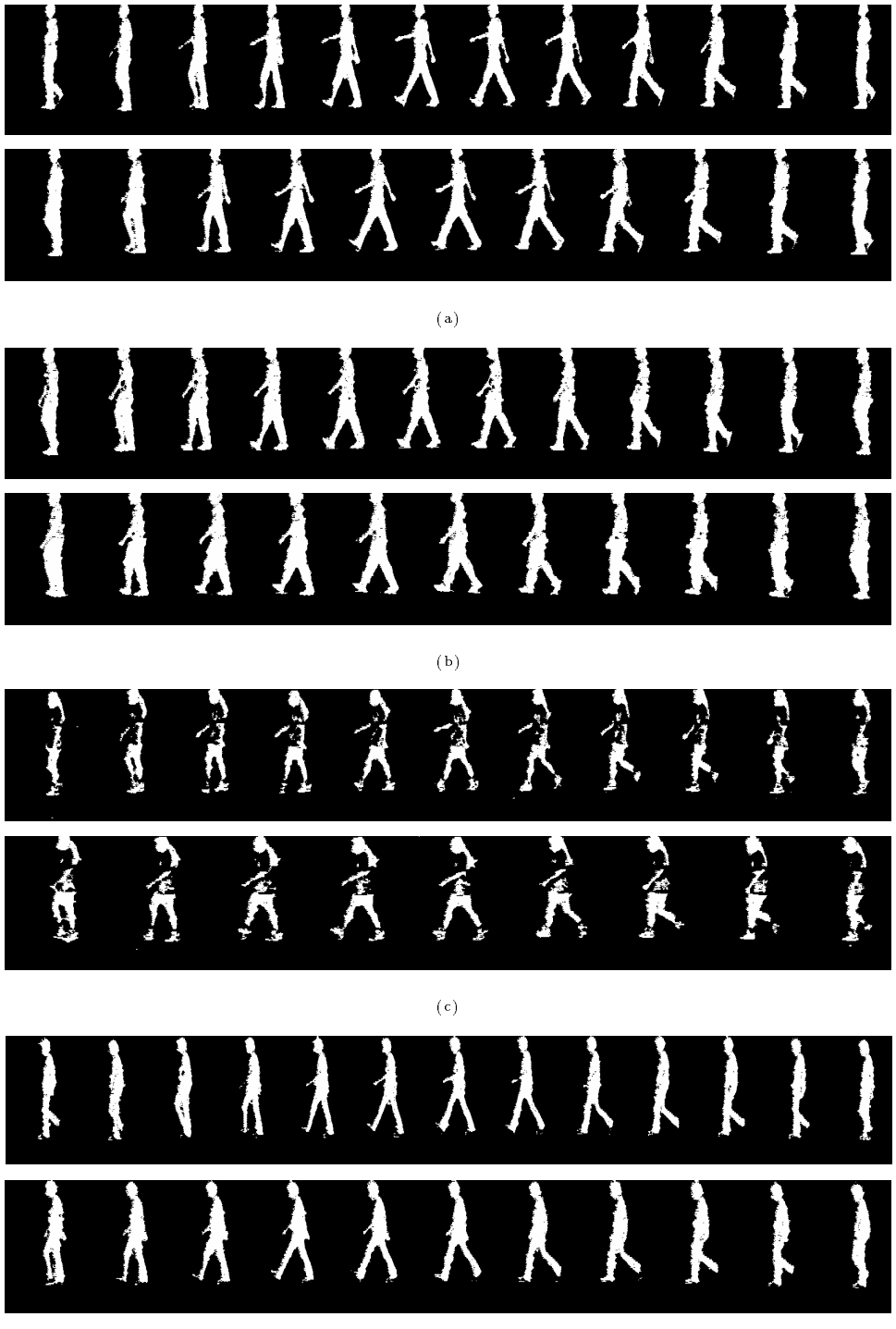

Figure 1: Silhouettes for a full gait-cycle corresponding to (a) Person 1 (b) Person 2 (c) Person 3 (d) Person 4.

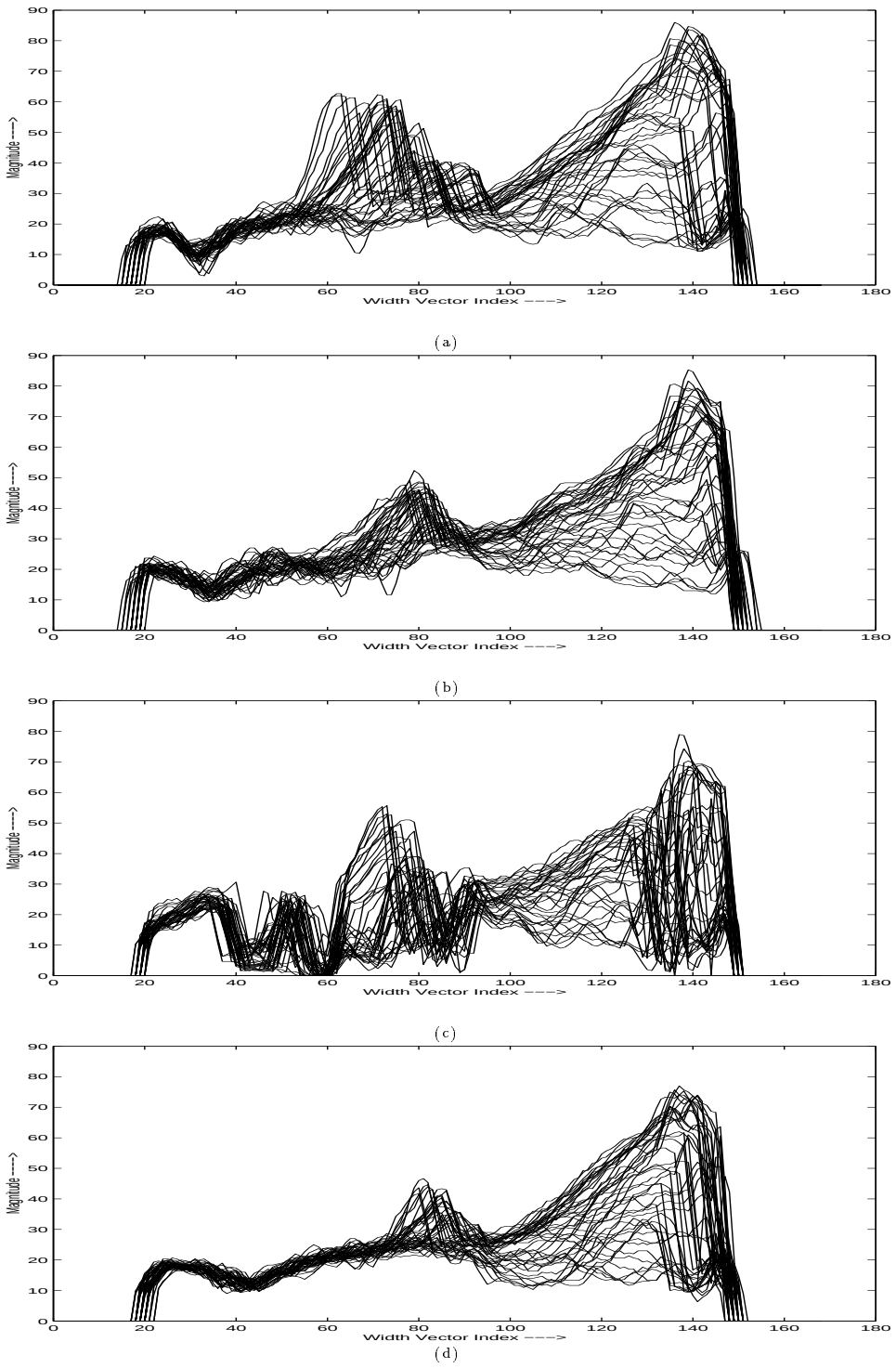

Figure 2: Width overlay plots for (a) Person 1 (b) Person 2 (c) Person 3 (d) Person 4.

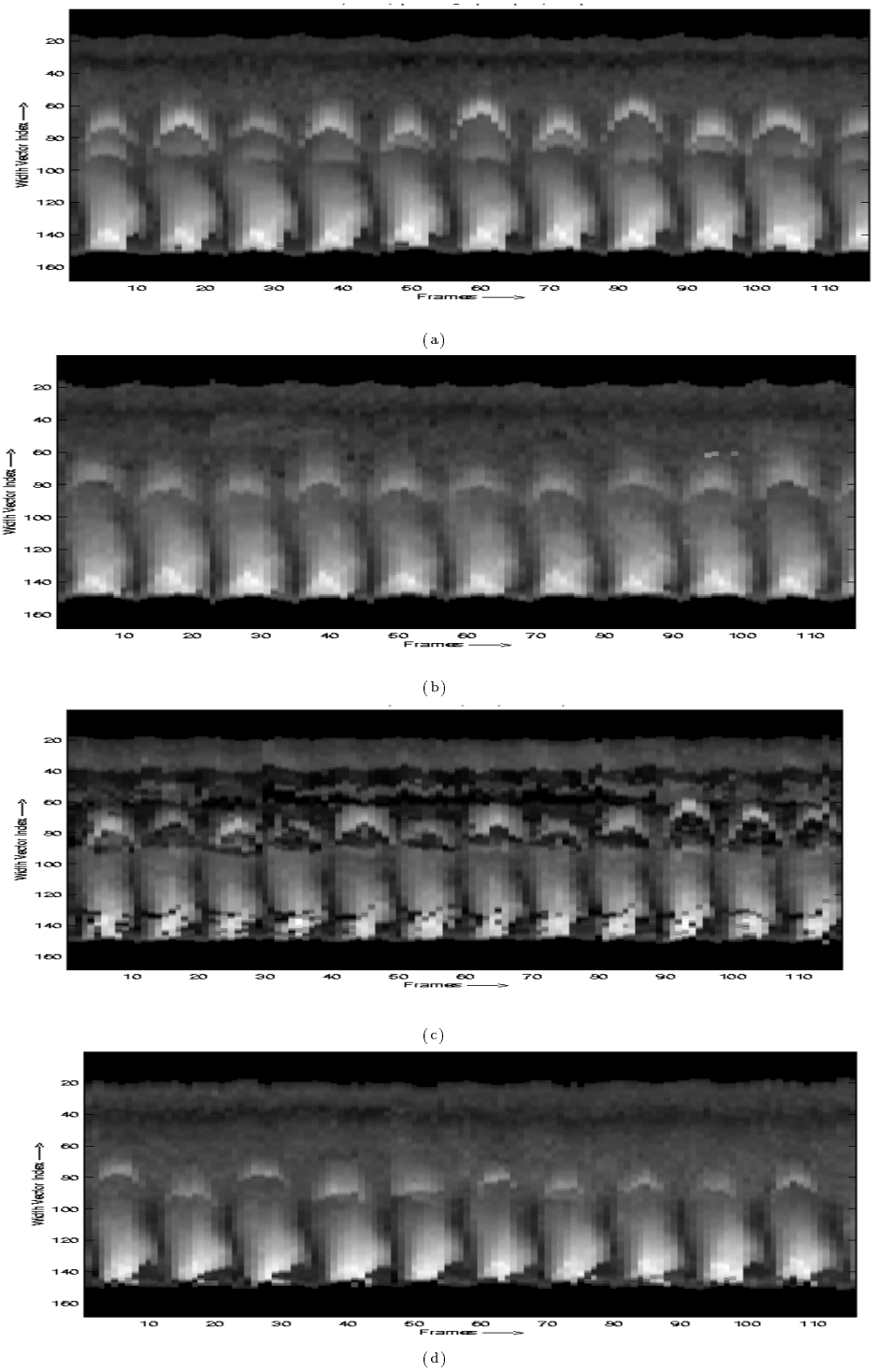

Figure 3: Temporal plot of the width vector for several gait cycles of (a) Person 1 (b) Person 2 (c) Person 3 (d) Person 4.

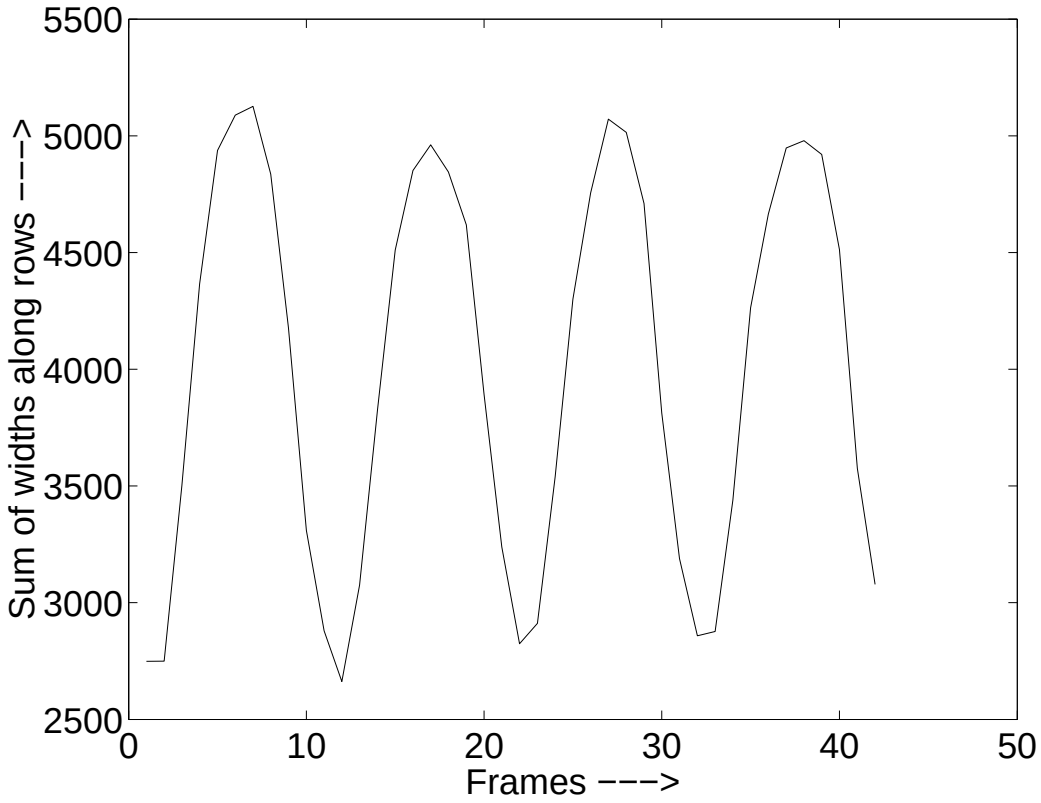

Figure 4: Sum of widths as a function of time.

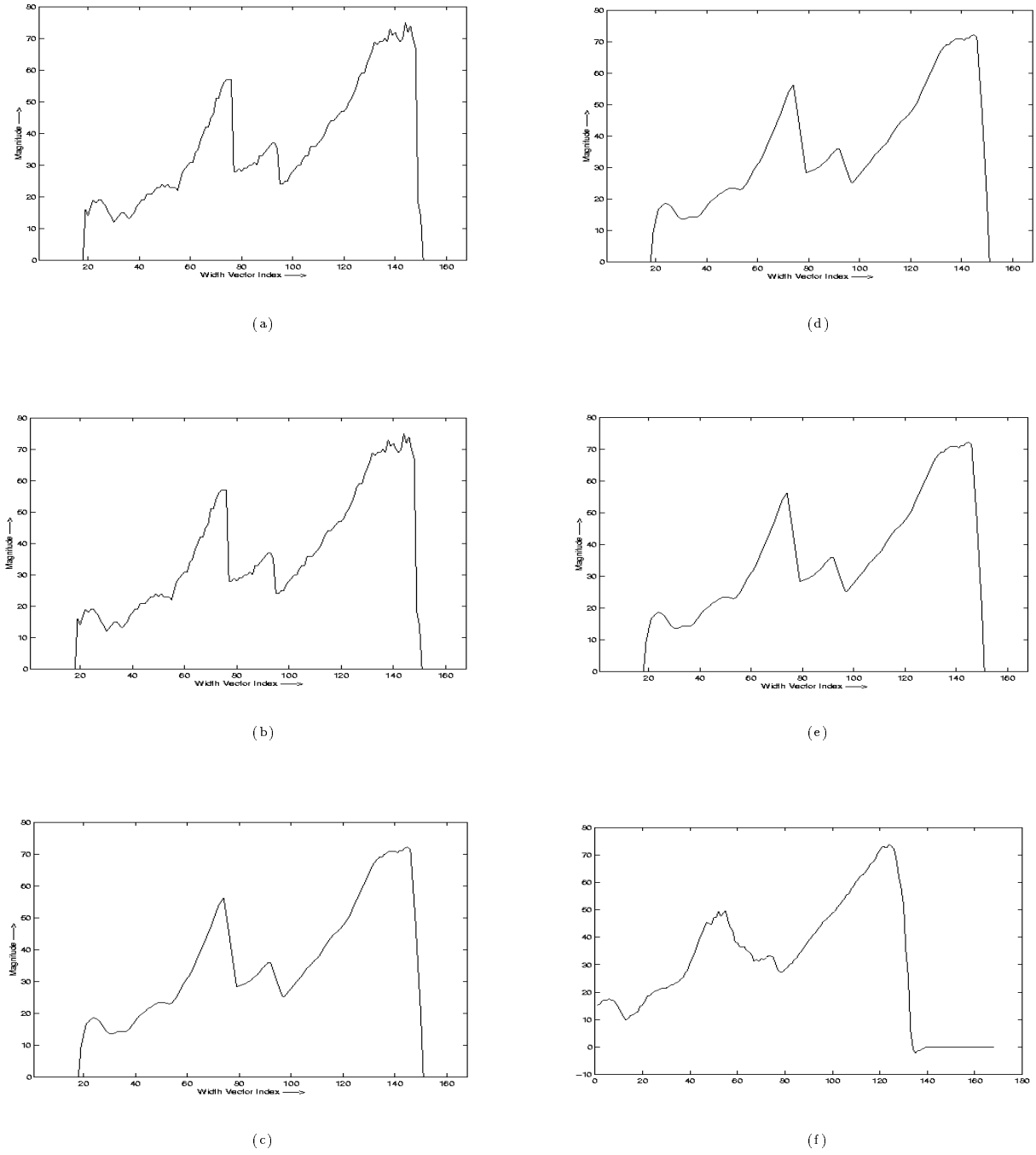

Figure 5: Width plots for an arbitrary frame. (a) Unsmoothed raw feature. (b) 3-point smoothed feature. (c) 5-point smoothed feature. (d) 11-point smoothed feature. (e) 21-point smoothed feature. (f) Reconstructed width vector using the first two eigenvectors.

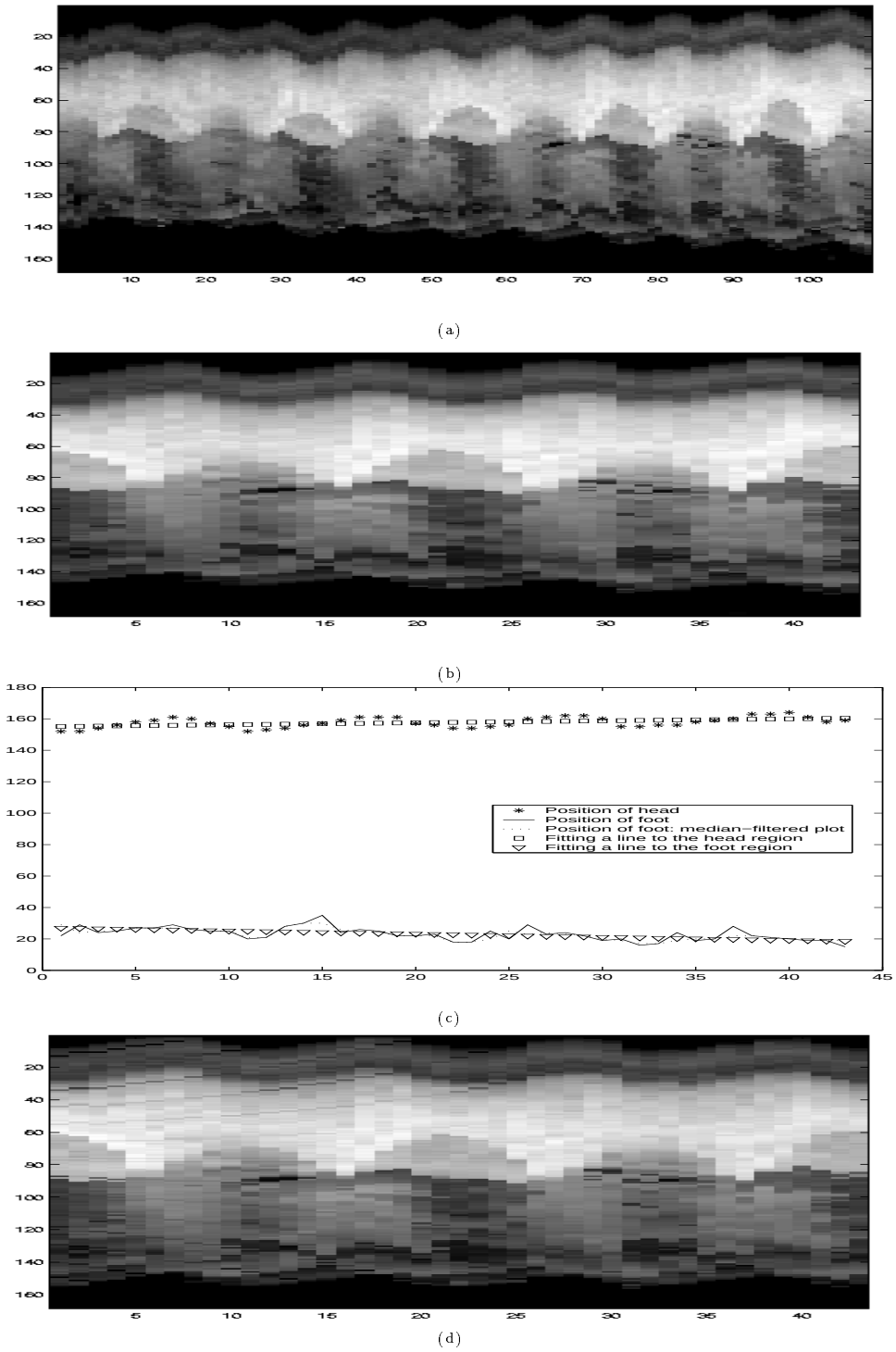

Figure 6: Outer contour plot for a frontal gait sequence in the UMD datasbase. (a) Subject walking towards the camera (b) 4 half-cycles chosen for recognition (c) The position of head and foot in each frame is located to compute the normalizing factor. Two straight lines are then fit to the trajectories of the head and foot. (d) The 4 half-cycles normalized.

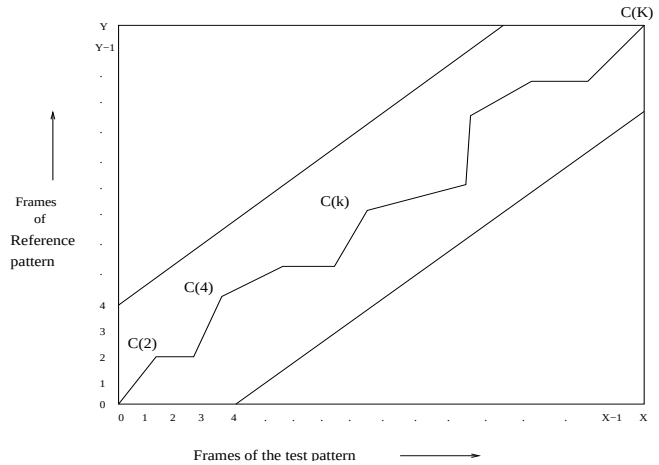

Figure 7: A typical dynamic time-warping path.

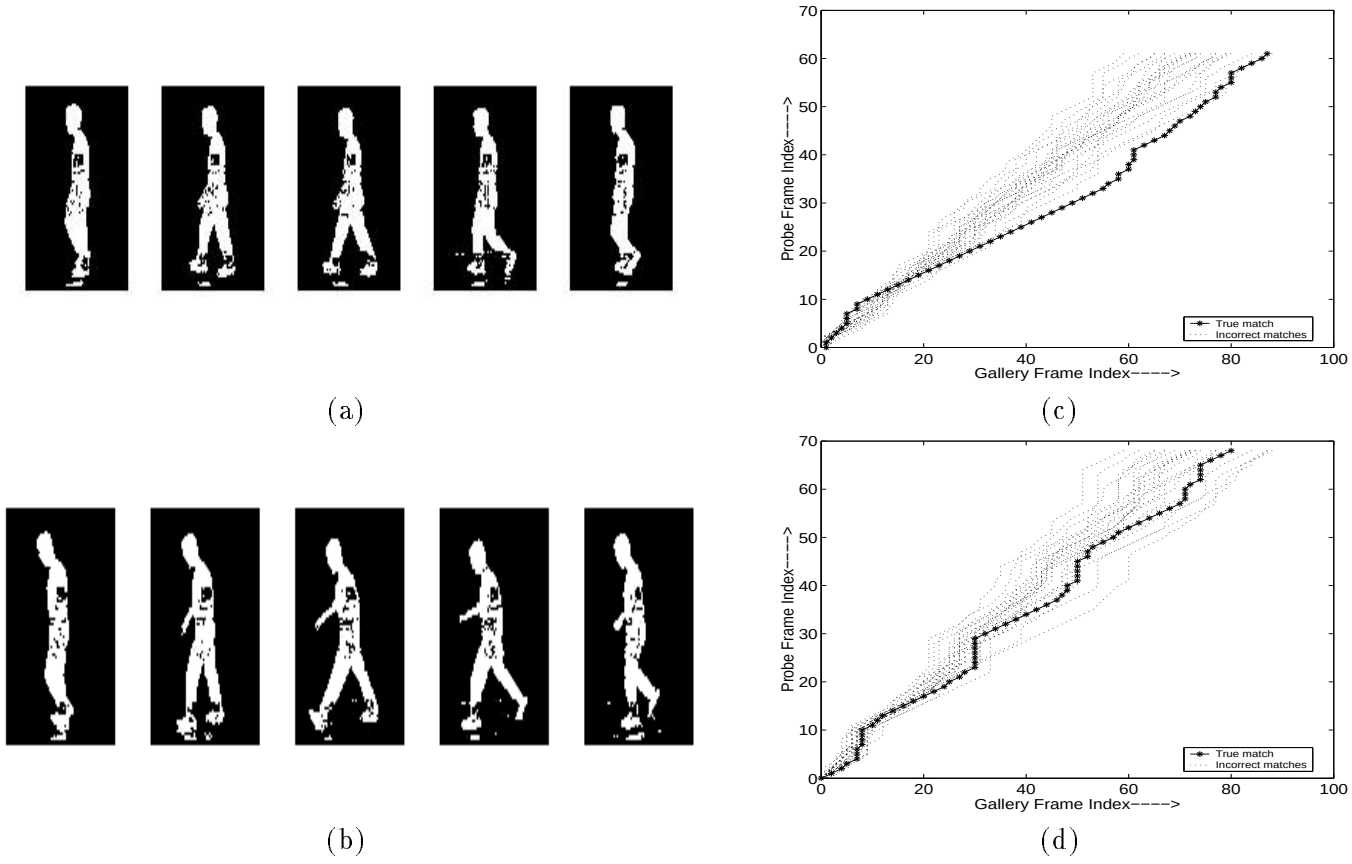

Figure 8: Sample images of the same subject corresponding to (a) slow-walk and (b) fast-walk (Notice the change in posture and body dynamics) (c) Warping path for person with largest reference sequence-length to probe sequence-length ratio (d) Warping path for the person in (a) and (b).

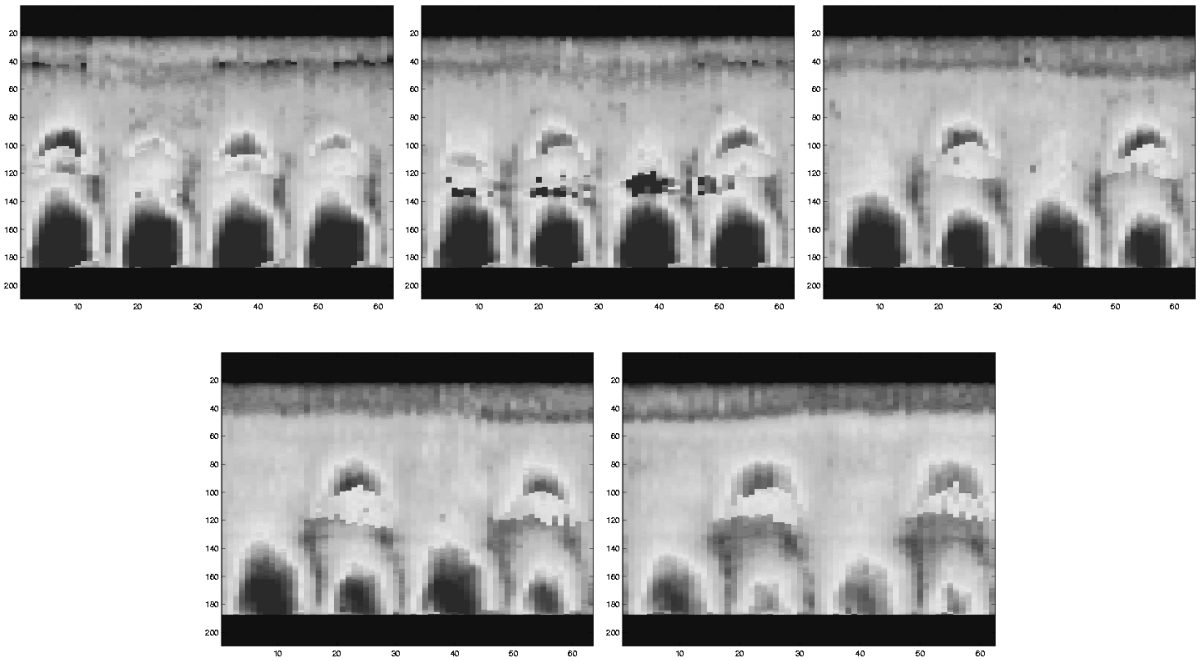

Figure 9: Width profile of an individual for different azimuth angles (a) 0 deg (b) 15 deg (c) 30 deg (d) 45 deg and (e) 60 deg.