Gait analysis for human identification

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

[1] Center for Automation Research, University of Maryland, College Park, MD 20742
{kale, cuntoor, rama}@cfar.umd.edu
[2] Department of Electrical Engineering, IIT Madras
Chennai-600036, India
raju@ee.iitm.ernet.in
[3] Department of Computer Science and Engineering, IIT Madras,
Chennai-600036, India
yegna@cs.iitm.ernet.in

Abstract. Human gait is an attractive modality for recognizing people at a distance. In this paper we adopt an appearance-based approach to the problem of gait recognition. The width of the outer contour of the binarized silhouette of a walking person is chosen as the basic image feature. Different gait features are extracted from the width vector such as the dowsampled, smoothed width vectors, the velocity profile etc. and sequences of such temporally ordered feature vectors are used for representing a person's gait. We use the dynamic time-warping (DTW) approach for matching so that non-linear time normalization may be used to deal with the naturally-occuring changes in walking speed. The performance of the proposed method is tested using different gait databases.

1 Introduction

Gait refers to the style of walking of an individual. Often in surveillance applications, it is difficult to obtain face or iris information at a resolution that is sufficient for recognition. Studies in psychophysics [1] reveal that humans have the capability of recognizing people from even impoverished displays of gait, indicating the presence of identity information in the gait signature. From early medical studies [2], it appears that there are 24 different components to human gait and that if all gait movements are considered, gait is unique.

Approaches to gait recognition can be broadly classified as being model-based and model-free. Examples of the first kind include [3], [4] and [5]. In [3], the gait signature is extracted by fitting the movement of the thighs to an articulated pendulum-like motion model. In [4] several ellipses are fit to different parts of the binarized silhouette of a person. Statistical analysis of the parameters of these ellipses such as the location of centroid, eccentricity etc. are used to extract features for recognition. However, in the presence of noise, the estimates of these parameters may not be reliable. Examples of the model free approach include the work of Huang et al. [6] who use optical flow to derive the motion image sequence corresponding to a walk cycle. Principal components

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

analysis is then carried out to derive the so-called eigengaits for recognition. Other examples include the works of [7] where activity specific static parameters were used for gait recognition. The smaller set of parameters extracted is shown to have greater resilience to viewing direction. Representation using such a small set of parameters may however, ignore valuable structural cues about a person and may adversely affect performance in large datasets.

For normal walk, gait sequences are repetitive and exhibit nearly periodic behavior. Background subtraction is used to convert the video sequence into a sequence of binarized images. We choose the width of the outer contour of the silhouette as our basic image feature since it contains structural as well as dynamical aspects and compactly represents gait of an individual. From the basic width vector, direct (smoothed and downsampled) and eigen-based features are derived. Sequences of such temporally-ordered feature vectors are used for representing a person's gait. Typically, we have $5-10$ contiguous half cycles of walking data per subject and the number of frames per cycle ranges between 8 to 20. Since the amount of training data is rather limited for a statistical modeling approach, it may not be possible to reliably estimate the parameters of the model. Therefore, we use a more direct matching scheme for video comparison. Further, in the case of walking, unlike marching, different gait cycles tend to have unequal lengths. Hence, a classifier based on direct template-matching is not suitable. Dynamic time warping [8], on the other hand uses non-linear time normalization to compare test and reference gait patterns. The performance of different gait measurements is tested using the UMD, CMU and USF databases.

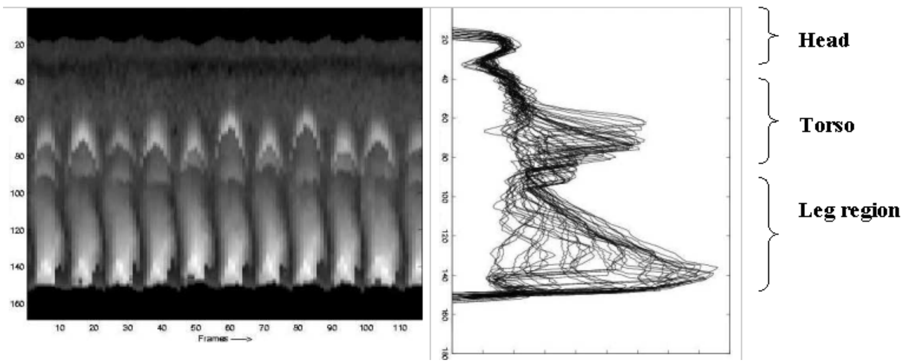
The organization of the paper is as follows. The basic width vector and its derivative features and the matching using DTW are explained in Section 2. . Experimental results are given in Section 3 while Section 4 concludes the paper.

2 Gait Representation

2.1 The Width Vector

An important issue in gait is the extraction of salient features that will effectively capture gait characteristics. In order to be robust to changes of clothing and illumination it is reasonable to consider the binarized silhouettes of the subject. We choose the width of the outer contour of the silhouette as the feature vector. 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. The width along a given row is computed as the difference in the locations of the right-most and the left-most boundary pixels in that row and a width vector is generated for each frame. Note that the pose information is lost in the width-generation process. The overlay of the width vectors are given in Figure 1(b) for one individual. It should be noted that the width overlays does not capture the temporal aspect of the width vectors and the distinction between individuals is accentuated when the width vectors are plotted as a function of time as shown in Figure 1(a).

The variation of each component of the width vector can be regarded as the gait signature of that subject. From the temporal width plots, it is clear that the width vector is roughly periodic and gives the extent of movement of different parts of the body. The brighter a pixel, the larger is the value of the width vector in that position.



(a) (b)

Fig. 1. (a)Temporal plot of width vectors and (b) their overlay. The different regions of the body can be seen. In the temporal plot, the brighter a pixel, the larger is the magnitude of the width vector in that position.

2.2 Features derived from the basic width vector

In this section, we discuss different features derived from the basic width vector for gait recognition. The idea is to arrive at a compact representation that exploits redundancy in the gait data for dimensionality reduction.

The features we used include (i) smoothed and down-sampled versions of the width vector and (ii)differenced width vectors. 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 even at much lower dimensions. We are also interested in studying the effect of dynamics on gait identification. One way to extract the dynamics is to compute the velocity profile by taking the difference of successive frames in the walking sequence. Obviously, most of the structural information like girth of the person etc. is lost when we go the velocity domain. It is to be expected that neither dynamic nor structural information, by itself, will be sufficient to capture gait. Both are necessary and cannot be decoupled. Figure 2 (a) shows the raw width vector for an arbitrary frame while Figures 2 (b) and (c) show the case when the width vector is smoothed and down-sampled by a five point and 21 point filter respectively.

From the temporal width plot, we note that although the width vector changes with time as the person transits through a gait cycle, there is a high degree of correlation among the width vectors across frames. Most changes occur in the hand and leg regions. The rest of the body parts do not undergo any significant changes during a gait cycle. Hence, one would intuitively expect that the variations of the width vector may be restricted to a lower dimension subspace. Given the width vectors $\{W(1), \cdots, W(N)\}$,for N frames $W(.) \in R^{\boldsymbol{M}}$, 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

(a) (b) (c)

Fig. 2. (a) Raw width vector (b) Down-sampled and 5-pt smoothing width vector and (c) smoothing by a 21 point filter.

as

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

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

2.3 Matching Gait Sequences

A typical walking sequence has approximately 60 frames consisting of four half cycles of the subject walking. In the absence of large training data therefore, direct video-based matching is better suited than the use of statistical models. Direct frame-by-frame matching is not a realistic scheme since people may slightly alter their speed and style of walking. Instead of restricting the the frames of possible matches, it would be prudent to allow a search region at each time instant, during evaluation. Therefore, DTW [8] is chosen as the matching scheme. The key steps in the DTW algorithm are enforcing end point constraints, computing local and cumulative error computation followed by backtracking to obtain the warping path. To satisfy the end-point constraint, all the sequences are processed so that the first and the last frames are both rest stances. The Euclidean distance is used as the local distance measure when comparing two width vectors. The cumulative distance at the end of the warping path is recorded as the matching score between the reference and test patterns.

3 Experimental Results

In this section, we give a brief description of the databases that we have used in our experiments and demonstrate the performance of the proposed method for gait-based human identification.

3.1 UMD Database

The UMD dataset [4] contains outdoor gait sequences captured by two cameras placed at orthogonal to each other. 44 subjects are recorded in two sessions. We train with the

[4] http://degas.umiacs.umd.edu/hid

video data collected from the first session and test with that of the second session. For the UMD database the number of contiguous walk cycles varies from 4 to 6. To maintain uniformity, we use four half cycles for matching. The recognition results for the different gait measurements are presented in Table 1. We notice that even after considerable downsampling and smoothing, the recognition rates do not deteriorate rapidly. For the eigen features, we note that using just two eigenvectors an accuracy of 80 % is achievable. Increasing the number of eigen-vectors led to lower accuracy, since higher order eigen-vectors tend to be noisy. Henceforth, we use only the first two eigenvectors for computing eigen features. However, the accuracy drops significantly if only the velocity information is used. Thus, for gait recognition both structural as well as dynamic information are important.

		Rank				
Experiment	Feature Considered	1	2	3	4	5
Effect of smoothing and downsampling the raw width vector	Raw width vector	79.07	81.40	83.72	86.05	86.05
	5 point smoothed 42 dim feature	76.74	83.72	83.72	88.37	88.37
	11 point-smoothed 21 dim feature	76.74	83.72	83.72	83.72	90.70
	21 point-smoothed 11 dim feature	79.07	86.05	86.05	88.37	90.70
Eigen Decomposition of the width vector and reconstruction using different eigen vectors	Using eigen vector no.1	73.1	75.2	80.0	80.0	84.0
	Using eigen vector no.1,2	80	87	90	90	91
	Using eigen vector no.1,2,3	68	80	84	84	84
	Using eigen vector no.1,2,3,4	73	77	84	84	84
	Using eigen vector no.1,2,3,4,5	70	73	79	82	84
Velocity Profile	Smoothed and Differenced 168 dim	41.9	51.6	61.2	70.9	74.1
	Eigen decomp. of velocity profile	56	75	76	80	83

Table 1. UMD database: Analysis of Different features for gait recognition

3.2 CMU database

The CMU dataset [5] consists of 25 subjects walking on a treadmill, under different conditions such as slow walk, fast walk and walk when carrying a ball. Seven cameras are mounted at different angles. The first half of the gait sequence is used for training while the second half is used for testing. This dataset shows the effect of change in walking speed on recognition. The results are given in Table 2. It is seen that the eigensmoothed feature in general performs better than the direct smoothed feature since eigensmoothing exploits spatio-temporal redundancy rather than just the spatial sense. When the gallery is the slow walk sequence and the probe is the fast walk sequence the performance is found to be inferior than the case when the gallery and probe are both slow walk sequences. DTW is known to perform badly [9] when the ratio of gallery-length to probe-length is less than 0.5 or more than 2. In the CMU dataset, the ratio of cycle-length of the gallery to the cycle-length in the probe is atmost 1.36. The corresponding person was

[5] http://hid.ri.cmu.edu

Experiment	Feature Considered	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
Ball	Eigen Smoothed feature	95.4	100.0	100.0	100.0	100.0

Table 2. CMU database: Analysis of Different features for gait recognition

correctly identified as the top match. Thus the dynamic time warping method is robust to changes in walking speed. The value of the ratio for one of the mismatched cases was 1.15. To analyze this we consider a few frames in the gait cycles of the incorrectly recognized person under slow and fast-walk modes in Figures 3 (a) and (c). As is apparent from the figure, the posture as well as hand swings for the person are quite different in the cases of fast-walk and slow-walk. Thus the change in body dynamics and stride of the person rather than the length of the walk cycle are responsible for the bad recognition performance. Figures 3 (b) and (d), show the warping paths for the person with the highest ratio and the incorrectly recognized person, respectively.

Finally, the high accuracy in the case when the subjects are walking with a ball in his hand suggests that certain parts of the body may exhibit a more consistent pattern for recognition. This has also been noted in [10].

3.3 USF Database

The USF database [6] consists of outdoor gait sequences of 71 subjects walking along an elliptical path on two different surfaces (Grass and Concrete) wearing two different types of footwear (A and B). Two cameras, R and L capture that data. Seven experiments are set up as shown in Table 3. The USF database has the largest number of individuals among the databases that we have considered. The direct and eigen-smoothed width features were again considered. The eigen-smoothed feature gave better performance compared to the direct smoothed feature in this case as well. The CMC curve for the different experiments as described in Table 3 are shown in Figure 4. From the graph, several conclusions can be drawn. It is clear that difference in surface leads to the worst recognition performance while difference in viewing angle is affected least.

4 Conclusion

Appearance-based features were derived from the video sequence of subjects walking, captured both indoors and outdoors. The width of the outer contour of the binary silhouette was used as the basic feature. Different features were extracted from the width

[6] http://marathon.csee.usf.edu/GaitBaseline/

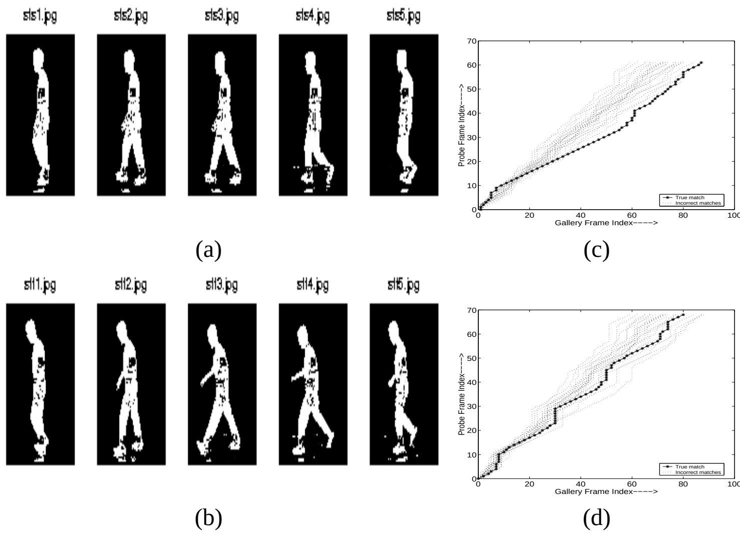


Fig. 3. Sample images the same subject corresponding to (a) slow-walk and (b) fast-walk(Notice the change in posture and body dynamics) (c)-Warping path for person with largest training to testing ratio (d) Warping path for person in (a) and (b).

Experiment	Probe	Difference
A	G,A,L (71)	View
B	G,B,R (41)	Shoe
C	G,B,L (41)	Shoe, View
D	C,A,R (70)	Surface
E	C,B,R (44)	Surface, Shoe
F	C,A,L (70)	Surface, View
G	C,B,L (44)	Surface, Shoe, View

Table 3. USF Dataset: 7 probe sets with the common gallery being G,A,R consisting 71 subjects. The numbers in the brackets are the number of subjects in each probe set.

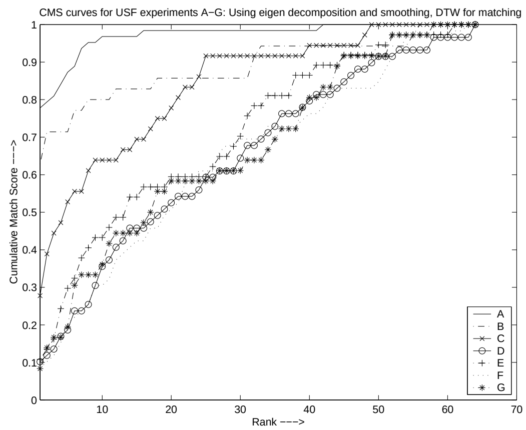


Fig. 4. CMS curves for USF database.

vector and dynamic time warping was used to match gait sequences. Eigenanalysis of the width vector shows that the gait signal evolves on a lower dimensional subspace and that gait possesses discriminative information. The method was found to be reasonably robust to changes in speed. The contribution of dynamic information for gait recognition was also studied. It was also found that the leg region by itself gave better recognition performance for one of the databases. One of our future areas of our research involves a systematic study of component level features extracted from the basic feature and combining the evidences and to study effects of viewing angle changes on recognition.

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, "Gait as a total pattern of movement," *American Journal of Physical Medicine*, vol. 46, pp. 290–332, June 1967.
3. 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.
4. 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.
5. A. Kale, N. Cuntoor, and R. Chellappa, "A framework for activity-specific human recognition," *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (Orlando, FL)*, May 2002.
6. 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.
7. A.F. Bobick and A. Johnson, "Gait extraction and description by evidence-gathering," *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2001.
8. Jinu Mariam Zacharia, "Text-independent speaker verification using segmental, suprasegmental and source features," M.S. thesis, IIT Madras Chennai India, March 2002.
9. L. Rabiner and H. Juang, *Fundamentals of speech recognition*, Prentice Hall, 1993.
10. 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.