GAIT RECOGNITION BASED ON HUMAN BODY COMPONENTS

Nikolaos V. Boulgouris and Zhiwei X. Chi

Department of Electronic Engineering
King's College London
United Kingdom

ABSTRACT

This paper presents a novel approach for gait recognition based on the matching of body components. The human body components are studied separately and are shown to have unequal discrimination power. Several approaches are presented for the combination of the results obtained from different body components into a common distance metric for the evaluation of similarity between gait sequences. A method is also proposed for the determination of the weighting of the various body components based on their contribution to recognition performance. Using the best performing of the proposed methods, improved recognition performance is achieved.

Index Terms— human gait, recognition, verification.

1. INTRODUCTION

Gait recognition [1] is a fairly new technique for biometric identification based on the walking style of individuals. Recognition based on human gait has several advantages related to the unobtrusiveness and the ease with which gait information can be captured. Unlike other biometrics, such as face, iris, fingerprints, gait can be captured from a distant camera, without drawing the attention of the observed subject.

A few early attempts on gait recognition were presented in [2, 3, 4]. In [2], gait patterns were extracted from spatiotemporal image volumes in order to outline the contours of walking subjects. These contours were used for gait recognition. In [3], a template matching method based on an eigenspace representation of gait was proposed. In [4], phase information of gait features was extracted, based on optical flow, and was subsequently used for gait recognition.

In [5], a gait recognition method was proposed using recovered static body parameters. The static parameters that were used in [5] were the height, the distance between head and pelvis, the maximum distance between pelvis and feet and the distance between the feet. However, gait dynamics were not used and no attempt was made to quantify the contribution of each of these parameter in recognition performance.

One of the first methods that attempted to divide the human body into components that are treated separately was presented in [6]. In that work, a silhouette is considered to consist of seven ellipses. They study the behavior of each of these ellipses throughout the walking period in terms of the change of the magnitude component and the phase component obtained from Fourier Transform, as well as the discrimination power of each of these components. Feature vectors derived from the magnitude and phase components were used in person identification and gender classification tests, and produced good results based on a relatively small data set.

In [7], manually extracted and labelled silhouettes were used based on the USF HumanID Gait Challenge data set [8]. Although the manual silhouettes provide much more accurate representations of human shape, the results reported in [7] are inferior to those obtained using automatically extracted binary silhouettes. This is due to the fact that, in the automatically extracted binary silhouettes, there are correlated errors which contribute to the recognition performance. Consequently, more research needs to be conducted using the manually extracted and labelled silhouettes, which do not suffer from noise and background interference, in order to reliably determine what type of information carries discrimination power.

In this paper, we use the manual silhouettes of [7] in order to investigate the contribution of each body component of a walking person to the recognition performance of a gait recognition system. We provide a detailed analysis of the role and the contribution of each body component by reporting recognition results of systems based on each one of the body components. We also propose several ways to combine the results obtained using the independent body components, and we show that the best performing combination achieves better results than the system in [7] on the same manual silhouette data set.

The paper is organized as follows. Section 2 describes the component-wise comparison between two manual silhouettes. Section 3 presents the matching of gait styles for two silhouette sequences. Section 4 reports the detailed results using the proposed system. Finally, conclusions are drawn in section 5.

2. COMPONENT-WISE COMPARISON

The manual silhouettes consist of body components which can be clearly distinguished from one another, such as those presented in Fig. 1. On the other hand, the binary silhouettes contain limited information and are also plagued by error pixels due to the imperfect background subtraction. The availability of manually segmented and labelled silhouettes, for the the HumanID Gait Challenge data set, allows a more reliable investigation of issues related to the discrimination power of body components in gait recognition.

Fig. 1. Manual silhouettes.

The baseline algorithm in [8] uses the ratio $S(p_i, g_j)$ of the intersection over the union of two binary silhouettes as a measure of the similarity between them, i.e.,

$$S(p_i, g_j) = \frac{p_i \cap g_j}{p_i \cup g_j} \quad (1)$$

where g_j, p_i denote binary silhouettes from the gallery and probe set respectively, and j, i are silhouette indices running from 0 to the end of the walking cycle. The intersection and the union are calculated in terms of the number of pixels. The value of this ratio ranges within $[0, 1]$, with 0 representing no overlap between the two silhouettes, and 1 implying an identical match.

Recently, in [7], manual silhouettes are used to construct silhouettes with higher resolution than the binary ones, and study the quality of silhouettes on some difficult probe sets. Meanwhile, the authors of [7] apply to the manual silhouettes the same approach that was taken in [8] using binary silhouettes, i.e., entire shape of silhouettes when calculating the similarity ratio.

In this paper, we take a different approach that exploits the availability of manually-segmented and labelled human body components. Specifically, first we calculate distances between silhouettes on a component by component basis and subsequently, we combine the component-wise distances in order to determine the final distance between two silhouettes. Using component-wise comparison, better recognition results can be achieved, since the similarity ratio of the entire body of two manual silhouettes does not reflect the difference among individual components. Suppose that in Fig 1, the first and the third silhouettes in the bottom row have identical area for the lower body, which includes thighs, legs, and feet. Although the similarity ratio measured on the entire body will consider the two silhouettes as identical, the ratio determined using each body component separately will be able to discriminate between the silhouettes.

In practice, we adopted a distance metric to describe the resemblance between two silhouettes with respect to a certain body component, namely,

$$d_\alpha(p_{\alpha i}, g_{\alpha j}) = 1 - \frac{p_{\alpha i} \cap g_{\alpha j}}{p_{\alpha i} \cup g_{\alpha j}},\ \alpha = 1, 2, \cdots, 8, \quad (2)$$

where $g_{\alpha j}$ is a body component of a manual silhouette from the gallery set and $p_{\alpha i}$ is a body component of a manual silhouette from the probe set. The index α refers to one of the body components. Therefore, the right-hand side of equation (2) is carried out component-wise. The correspondence is tabulated in Table 2. Also, the value of the distance ratio is between 0 and 1, and a smaller distance means a closer match for the relevant components.

α	body component
1	head
2	torso
3	left arm
4	right arm
5	left thigh
6	left leg
7	right thigh
8	right leg

Table 1. The correspondence of each body component with the index α.

3. GAIT STYLE MATCHING USING HUMAN BODY COMPONENT WEIGHTING

After calculating the distances between the corresponding body components, we formed a component distance vector

$$\mathbf{d}_{ij} = [d_1(p_{1i}, g_{1j}\ \ d_2(p_{2i}, g_{2j})\ \ \ldots\ \ d_8(p_{8i}, g_{8j})]^T \quad (3)$$

We tested several approaches for the combination of the individual component distances into a single distance that quantify the dissimilarity between two silhouettes. Specifically, the total distance between two silhouettes was taken to be equal to the *median*, *min*, *max*, and a *weighted sum* of the distances corresponding to each body part. Among all approaches, the weighted sum gave the best results. However, for reasons of completeness, we present the results for the rest of the combination approaches in the experimental results section. The total distance between a pair of manual silhouettes (p_i, g_j) was calculated as

$$\tilde{d}(p_i, g_j) = \mathbf{W}^T \mathbf{d}_{ij} \quad (4)$$

where

$$\mathbf{W} = [w_1,\ w_2,\ \cdots,\ w_8]^T \quad (5)$$

is a weight vector. In equation (4), the weighted sum of the distances is being taken over all body components, with each $d_\alpha(p_i, g_j)$, $\alpha = 1, \cdots, 8$, being weighted by a coefficient w_α. It is expected that some body parts are more reliable or carry more discrimination power than others. For instance, the right arm rarely appears, or does not appear at all, during one walking cycle for most subjects, and therefore it is reasonable to use a factor in order to appropriately weight its contribution.

So far, two types of distances have been introduced as shown in equations (2) and (4). The next step is to calculate the distance between gait sequences based on (4). Note that the component-based distance $\tilde{d}(p_i, g_j)$ in (4) is calculated on pairs of silhouettes (i, j) from two different gait sequences. However, in practice, the number of silhouettes from a gait sequence might be greater or smaller than that from another gait sequence depending on the subjects' walking pace. Therefore, before we can calculate the distance between these two gait sequences, we need to determine the correspondence between silhouettes from the two gait sequences that are to be compared.

This can be achieved using a rule, which makes sure that each silhouette p_i from the gait sequence of subject p is associated with an appropriate silhouette g_j from another gait sequence of subject g. The process for determining the correspondence of silhouettes from the compared gait sequences is based on linear time normalization [9]. Let T, R be the number of silhouettes in a gait cycle of the probe and gallery sequence respectively. Then if $R < T$, $i = j \cdot \frac{T}{R}$, whereas if $R > T$, $j = i \cdot \frac{R}{T}$.

Let $\mathcal{F}$ denote the set of pairs (i, j) determined as above. Once the correspondence between silhouettes from the probe and gallery sequences is established and the set $\mathcal{F}$ is constructed, the final distance between the nth probe subject and the rth gallery subject is calculated as

$$D(n, r) = \frac{1}{N_\mathcal{F}} \sum_{(i,j) \in \mathcal{F}} \tilde{d}(p_i, g_j) \quad (6)$$

and $N_\mathcal{F}$ is the cardinality of $\mathcal{F}$. A gait recognition system based on the above distance metric will be seen to yield considerable gains in terms of recognition performance.

4. OPTIMAL DETERMINATION OF WEIGHTING COEFFICIENTS

For the purpose of determination of the weights in (5), we use the gallery set as the training set, since the gallery set contains all the subjects that appear in the probe sets. We define a component-wise distance between the nth and the rth subject:

$$D_\alpha(n,r) = \frac{1}{N_{\mathcal{F}}} \sum_{(i,j)\in\mathcal{F}} d_\alpha(p_{\alpha i}, g_{\alpha j}) \tag{7}$$

Note that the index α in equation (7) is included in order to emphasize that only one body component is involved in the determination of the component-wise distance between two subjects. This distance, i.e. $D_\alpha(n,r)$, is subsequently used for the determination of the weighting coefficient of the corresponding body component. Specifically, for the nth subject in the gallery set, we calculate the distance with each one of the rest of the subjects in the gallery set using (7), and obtain a matrix $\mathbf{Q}_n$ as follows:

$$\mathbf{Q}_n = \{q_{n(\alpha,l)}\}$$

$$= \begin{bmatrix} D_1(n,1) & \cdots & D_1(n,l) & \cdots & D_1(n,N_g-1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ D_8(n,1) & \cdots & D_8(n,l) & \cdots & D_8(n,N_g-1) \end{bmatrix} \tag{8}$$

where N_g is the number of subjects in the gallery set, $n, l \in [1, N_g]$, and $l \neq n$. Then, an 8-dimensional vector $\mathbf{b}_n$ is derived from each $\mathbf{Q}_n$ such that

$$\mathbf{b}_n = \frac{1}{N_g - 1} \left[\sum_{l=1}^{N_g-1} q_{n(1,l)}, \ \cdots, \ \sum_{l=1}^{N_g-1} q_{n(8,l)} \right]^T \tag{9}$$

Essentially, the elements of the above vector are the average component-by-component distances of the gallery subject n to the rest of the subjects in the gallery. Apparently, the greater the average distance for a specific component, the more it should contribute to the determination of the final distance using (4). This is why the weighting coefficients $w_\alpha, \ \alpha = 1, \ \cdots, 8$, are calculated from the average of $\mathbf{b}_n$ over all n, i.e.,

$$w_\alpha = \frac{\sum_{n=1}^{N_g} b_{n\alpha}}{\sum_{\tau=1}^{8} \sum_{n=1}^{N_g} b_{n\tau}} \tag{10}$$

where

$$b_{n\tau} = \frac{1}{N_g - 1} \sum_{l=1}^{N_g-1} q_{n(\tau,l)} \tag{11}$$

The denominator in (10) is intended to normalize the weighting coefficients, so that they sum up to 1. Using (10) and (11), the weighting coefficients are

$$w_\alpha = \frac{\sum_{n=1}^{N_g} \sum_{l=1}^{N_g-1} q_{n(\alpha,l)}}{\sum_{\tau=1}^{8} \sum_{n=1}^{N_g} \sum_{l=1}^{N_g-1} q_{n(\tau,l)}} \tag{12}$$

As seen, the final weighting coefficients can be calculated directly using the matrix $\mathbf{Q}_n, \ n = 1, \ \cdots, \ N_g$, which include the average distance of a gallery subject to the rest of the subjects in the gallery. The resulting weighting coefficients are such that components that exhibit greater average distances (and thus carry more discriminant power) contribute more to the total distance. It should be noted that only one cycle for each subject is available in USF's database of manual silhouettes. This prohibits the use of other optimizations, e.g., based on LDA, that could potentially further improve the results by considering multiple instances of the same person. In the experimental results section we will see that our approach achieves superior results than the method presented in [7].

5. EXPERIMENTAL RESULTS

For the experimental evaluation of our method, we used the manual silhouettes[1] provided by the University of South Florida (USF). This database contains human gait sequences captured under different conditions, such as camera viewing angle, shoe type, walking surface, and walking with/without carrying a briefcase. The gallery (reference) set of gait sequences was used as the system database and the probe (test) sets B, D, H, K were considered to contain sequences of unknown subjects who are to be recognized by comparison of their gait sequences to the sequences in the gallery set.

For the performance evaluation, we report Cumulative Match Scores (as in [10]) at rank 1 and rank 5. Rank 1 results report the percentage of the subjects in a probe set that were identified exactly. Rank 5 results report the percentage of probe subjects whose true match in the gallery set was in the top five matches. Intuitively, we studied the relative importance of the human body components, in the gait recognition system outlined in sections 2 and 3 using one body component at a time. Fig 2 reports the recognition performance at rank 1, illustrated using vertical bars for the various body components. Figures 2 (a)–(d) correspond to probe sets B–D, respectively. It is interesting to see that each component exhibits unequal discrimination power, which varies from one probe set to another. Specifically, for probe B, the best performance is achieved using the torso and the right thigh. This behavior is roughly the same for probe D. For probe H, the torso remains the most powerful, but gives similar performance with the head and the left thigh. Surprisingly, the right arm which is occluded by the body in most of the frames, has good discrimination power. This is due to the fact that the pattern of its appearance, rather than its shape, is very useful for discriminating between individuals.

The average performance for all set is reported in Fig. 2(e). In general, the right thigh and right arm appear to be the most discriminative body components followed by the torso. This is an important conclusion which contradicts previous evidence, acquired using tests on automatically-extracted silhouettes, that the lower body carries almost all discrimination power [8]. It is now becoming clear that correlated noisy pixels in the lower part of automatically-extracted silhouettes led to the wrong conclusions about the discrimination power of the lower body in comparison to the upper body.

The results obtained using independent body components can be dramatically improved using a combination of their corresponding distances given by equation (2). As mentioned in the previous section, the final distance between two silhouettes was considered to be equal to the *minimum*, *maximum*, and the *median* of the distances corresponding to individual body components. We also considered the *weighted distance* given by equation (4). As seen in Table 2, the first three approaches in general do not yield good results, although the combination based on the median operator occasionally gives decent performance.

The best performing of our methods is based on the distance metric given by equation (4). The results of this approach, which

[1] We used version 6 of the silhouettes.

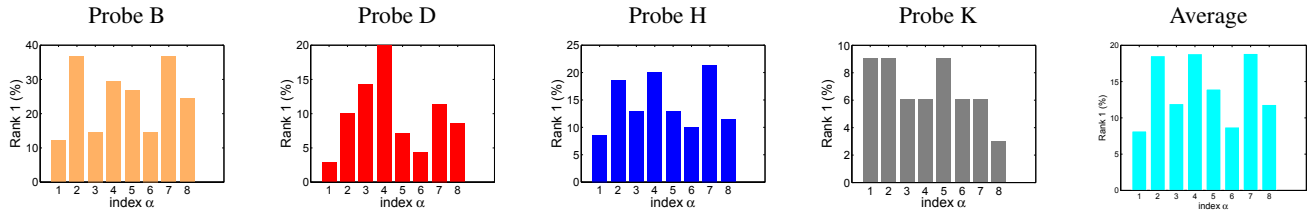

Fig. 2. Recognition percentage at Rank 1 for Probe sets B, D, H, K with respect to each body component and the average of all four Probe sets.

Probe set	Rank 1(%)				Rank 5(%)			
	min	max	med	$W^T d$	min	max	med	$W^T d$
B	12	29	29	49	39	46	61	78
D	10	20	24	26	29	37	44	53
H	3	20	14	16	14	43	36	46
K	6	6	12	15	15	30	24	39

Table 2. The results using several combinations of body component distance.

will be termed Component-Based Gait Recognition (CBGR), are reported in Table 3. The results are reported in comparison to the USF method [7], which relied on the entire body without considering individual body components. It can be clearly seen that there is considerable improvement on all probe sets at Rank 1 and Rank 5.

As mentioned previously, the performance of a gait recognition system based on high-quality, manually-extracted, and manually-labelled silhouettes seems to be inferior to that of a system based on automatically extracted silhouettes. This conclusion, which contradicts intuition, is due to correlated noise pixels which "boost" the performance of the automatic system and render the associated results inconclusive and possibly unreliable. However, as shown in the present paper, there is room for improvement of methods based on manual silhouettes. We believe that the study of systems based on manual silhouettes can provide more meaningful insights about the discrimination power of gait and the parameters that affect recognition performance.

Probe set	Rank 1(%)		Rank 5(%)	
	CBGR	USF	CBGR	USF
B	**49**	46	**78**	66
D	**26**	23	**53**	39
H	**16**	9	**46**	36
K	**15**	12	**39**	**39**

Table 3. Recognition percentage using the manual silhouettes with combined distance ratios of all body components in comparison to the USF method.

6. CONCLUSION

We presented a new approach for gait recognition using manually extracted and labelled silhouettes. The human body components were studied separately and were shown to carry different discrimination power. Several approaches were presented for the combination of the results of the different body components into a common distance metric for the evaluation of similarity between gait sequences. By combining the results from all body components, improved recognition performance was achieved.

7. ACKNOWLEDGEMENT

The authors would like to thank Professor Sudeep Sarkar at the University of South Florida for providing the manual silhouettes.

8. REFERENCES

[1] N. V. Boulgouris, D. Hatzinakos, K. N. Plataniotis, Gait recognition: a challenging signal processing technology for biometric identification, IEEE Signal Processing Magazine 22 (2005) 78–90.

[2] S. A. Niyogi, E. H. Adelson, Analyzing and recognizing walking figures in xyt, in: Proc. Computer Vision and Pattern Recognition, Seattle, WA, 1994, pp. 469–474.

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

[4] J. Little, J. Boyd, Recognizing people by their gait: The shape of motion, Videre: Journal of Computer Vision Research 14 (1998) 83–105.

[5] A. Y. Johnson, A. F. Bobick, A multi-view method for gait recognition using static body parameters, in: In 3rd Int'l Conf. on Audio- and Video-Based Biometric Person Authentication, Halmstad, Sweden, 2001, pp. 301–311.

[6] L. Lee, W. E. L. Grimson, Gait analysis for recognition and classification, in: Proc. Int'l Conf. Automatic Face and Gesture Recognition, 2002, pp. 155–162.

[7] Z. Liu, S. Sarkar, Effect of silhouette quality on hard problems in gait recognition, IEEE Trans. Man, and Cybernetics—Part B:Cybernetics 35 (2) (2005) 170–178.

[8] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, K. W. Bowyer, The humanid gait challenge problem:data sets, performance, and analysis, IEEE Trans. Pattern Analysis and Machine Intelligence 27 (2) (2005) 162–176.

[9] N. V. Boulgouris, K. N. Plataniotis, D. Hatzinakos, Gait recognition using linear time normalization, Pattern Recognition 39 (2006) 969–979.

[10] P. J. Phillips, H. Moon, S. A. Rizvi, P. J. Rauss, The feret evaluation methodology for face-recognition algorithms, IEEE Trans. Pattern Analysis and Machine Intelligence 22 (10) (2000) 1090–1104.