MODEL-BASED HUMAN GAIT RECOGNITION USING FUSION OF FEATURES

Xiaxi Huang and Nikolaos V. Boulgouris

Department of Electronic Engineering
King's College London
United Kingdom

ABSTRACT

In this paper, we investigate the fusion of several features extracted from manually-labelled silhouettes. A novel approach for human gait recognition based on the combination of three discriminative features, i.e., the *area*, the *gravity centre*, and the *orientation* of each body component, is also proposed. Experimental results show that the proposed method exhibits considerably better performance, in comparison to all existing methods that use manually-extracted silhouettes.

Keywords: gait, recognition, fusion

1. INTRODUCTION

Gait recognition is a challenging signal processing technology for biometric identification [1] that has been developed rapidly during the past few years. As a biometric trait, gait has many unique advantages, such as unobtrusiveness, recognition at a distance, and operation using low-resolution images.

One of the earliest methods for gait recognition based on the separation of the human body into components was presented in [2]. In this method, an ellipse is fitted on each component of a binary silhouette and it is observed through time.

In [3], a set of silhouettes that were manually extracted and labelled was used. Using such silhouettes, however, was shown to have had an adverse impact on the performance of a gait recognition system, as low-quality, automatically-extracted silhouettes yielded superior results. This counter-intuitive finding was due to the existence of correlated segmentation errors in automatically extracted silhouettes. The above fact indicates that there is a need for further investigation of gait recognition using manual silhouettes in order to determine which types of gait information is important.

In [4], the importance of each body component of a walking subject was investigated using the manual silhouettes of [3]. A detailed analysis of the role and the contribution of each body component was presented and several ways were proposed for the efficient combination of the results obtained using the independent body components.

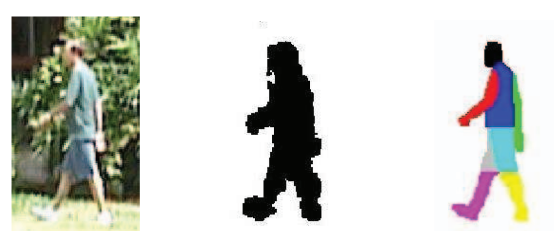
In [5], before the recognition process is applied, silhouettes are deformed using a Layered Deformable Model (LDM). By using this approach, detailed and modeled information of the manually silhouettes is obtained. However, the use of a large number of parameters increase the computational complexity of the algorithm.

The method proposed in the present paper considers the fusion of three discriminative component-based features (parameters): 1) the *area* of each body component 2) the *centre* of each body component, and 3) the *orientation* of each body component. The above features are fused based on component and temporal weighting. Experimental results show that the proposed method generally outperforms all other methods that use manual silhouettes.

This work was supported in part by the European Commission under contract FP7-215372 ACTIBIO. The authors would like to thank Prof. Sudeep Sarkar for providing the manual silhouettes of the database of the University of South Florida.

2. PROPOSED GAIT RECOGNITION SYSTEM

The NIST/USF HumanID gait challenge database includes a set of silhouettes in which the subjects were extracted and their body components were manually labelled [3]. An example of an original frame, its corresponding automatically-extracted silhouette, and manually-labelled silhouette is shown in Fig. 1. As seen in Fig. 1 (c), eight body components (i.e., head, torso, left / right arm, left / right thigh, and left / right leg) were labelled using different colours.



(a) (b) (c)

Fig. 1. Sample of the USF data set: (a) Original frame, (b)Automatically-extracted silhouette, (c) Manually-labelled silhouette.

Based on these manually labelled silhouettes, features of body components can be extracted and advanced approaches for gait recognition can be developed based on the extracted features. The block diagram of the proposed system is shown in Fig. 2. Initially, the frames are temporally aligned by re-sampling each sequence to a constant length. Subsequently, three features are extracted from each frame of each sequence; they include both shape and dynamic gait information. Then, component and temporal weights are calculated. This process takes temporal information into account during the recognition process. In the decision stage, distances for each feature between different subjects are calculated and combined. Finally, the combined distances are compared and a recognition decision is reached. In the following sections, each step of the algorithm will be described in detail.

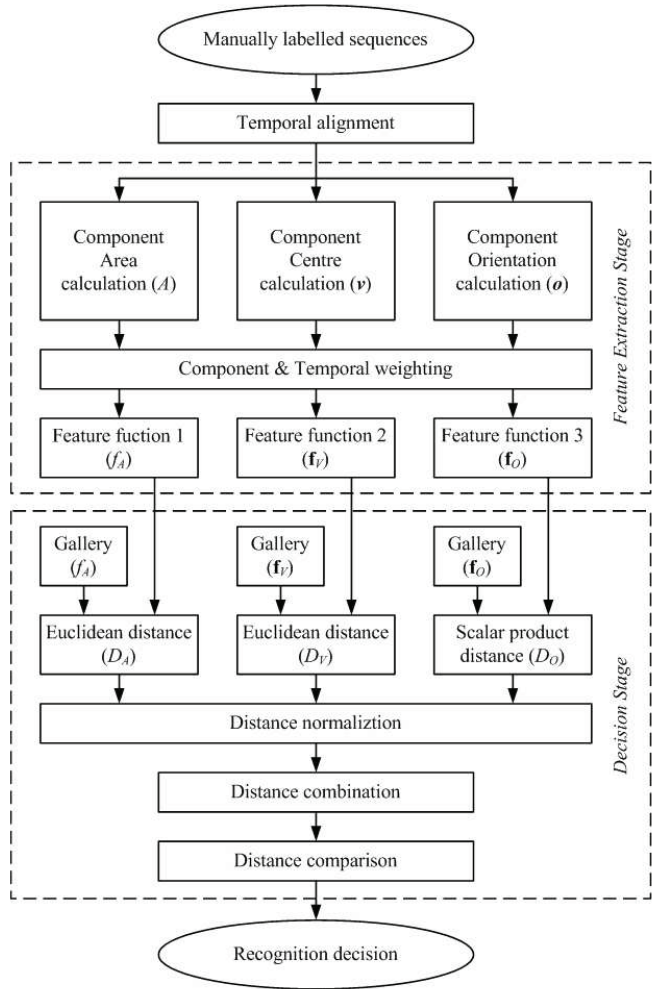


Fig. 2. Block diagram of the proposed algorithm

3. FEATURE EXTRACTION FROM MANUAL SILHOUETTES

Feature extraction is a very important process in a gait recognition system. In this work, we investigate three main features which seem suitable for the capturing of discriminatory gait information from sequences of manually extracted and labelled silhouettes.

3.1. Component area

The first feature that we consider is the area A_i of each body component $\mathcal{C}_i$, i.e., the number of pixels $\mathbf{x}$ in $\mathcal{C}_i$. These component areas are calculated for each frame of a temporally normalized sequence, in order to obtain a two-dimensional feature function, defined as $f_A(m,n) = A_{mn}$, where A_{mn} is the area for the mth component, $m = 1, \ldots, M$, in the nth frame, $n = 1, \ldots, N$.

3.2. Component centre

Due to the availability of labelling information for each pixel, it is possible to calculate the gravity centre $\mathbf{g}_m$ of each body component. After calculating the gravity centres, we calculate the vector distances $\mathbf{v}_m$ between each one of the body component centres $\mathbf{g}_m$ and the centre $\mathbf{g}$ of the entire silhouette, i.e., $\mathbf{v}_m = \mathbf{g}_m - \mathbf{g}$. Subsequently, a two-dimensional feature function for area information, $\mathbf{f}_V(m,n) = \mathbf{v}_{mn}$, is calculated.

3.3. Component orientation

The *component centre*, described above, reflects the position of the component with respect to the body centre, while the positions and distributions of the rest of the pixels remain unknown; the *area* gives size information, while shape information is disregarded. When the areas and gravity centres of two corresponding body components in two subjects are identical, they cannot be discriminated by those two features. In order to capture the structure of the silhouettes in a more accurate way, we extract a third feature - the *orientation* of each body component, which is defined as $\mathbf{f}_O(m,n) = \mathbf{o}_{mn}$, where $\mathbf{o}_{mn}$ is a vector that denotes the principle orientation of the mth body component in the nth frame.

4. COMPONENT AND TEMPORAL WEIGHTING

Since, as shown in [4], not all body components have the same importance, we apply component weighting separately for each feature. Based on the method in [4], we determine three sets of weights, arranged in three M-dimensional vectors $\mathbf{w}_{\mathcal{C}_A}$, $\mathbf{w}_{\mathcal{C}_V}$, $\mathbf{w}_{\mathcal{C}_O}$, where M is the number of body components.

In our system, all normalized gait sequences consist of one full gait cycle of length N. Within a gait cycle, however, the discriminating power of each frame is different, as some gait stances are more “revealing” than other gait stances. For this reason, weights are applied on the elements of the feature matrixes in the proposed algorithm.

Since the *component centre* feature reflects the arm's swing extent and the stride, the frames in which the subject is in a striding pose are more discriminative than the ones in which the subject is standing. Therefore, for $\mathbf{f}_V$, greater importance should be put on the temporal positions (i.e., frames) that correspond to a striding pose. Similarly, the *area* feature is more discriminative in the striding-subject frames, because in the standing-subject frames, a considerable amount of foreground pixels are overlapped, where errors are generated by using those incomplete areas.

Since the total area of the body fluctuates at the same pace as the stride, it reflects the discriminating power (namely, weight) of the frame. Therefore, the total area of the body in each gallery sequence is calculated and described as:

$$\mathbf{A}_T = \left[\begin{array}{ccccc} A_{T1} & \ldots & A_{Tn} & \ldots & A_{TN} \end{array} \right] \tag{1}$$

where N is the number of frames in an aligned sequence.

Subsequently, the average body area for each aligned frame is calculated as:

$$\overline{\mathbf{A}}_T = \frac{1}{N_G} \sum_{q=1}^{N_G} \mathbf{A}_T^q = \left[\begin{array}{ccccc} \overline{A}_{T1} & \ldots & \overline{A}_{Tn} & \ldots & \overline{A}_{TN} \end{array} \right] \tag{2}$$

where N_G is the number of subjects in the reference database.

Finally, temporal weights for each aligned frame are defined as $\mathbf{w}_\tau = \left[\begin{array}{ccccc} w_{\tau 1} & \ldots & w_{\tau n} & \ldots & w_{\tau N} \end{array} \right]^T$, where

$$w_{\tau n} = \frac{\overline{A}_{Tn} - min(\overline{\mathbf{A}}_T)}{max(\overline{\mathbf{A}}_T)}, \quad n = 1, 2, ..., N$$

After having calculated the relative importance of body components for each one of the features, as well as the relative importance of gait poses in a gait cycle, we can determine the relative importance of each body component in each frame. For the *area* feature, this can be expressed as:

$$\mathbf{W}_A = \mathbf{w}_{\mathcal{C}_A} \cdot \mathbf{w}_\tau^T \tag{3}$$

where the weight for the mth component in the nth frame is calculated as $w_A(m, n) = w_{\mathcal{C}_A m} \cdot w_{\tau n}$.

In a similar way,

$$\mathbf{W}_V = \mathbf{w}_{\mathcal{C}_V} \cdot \mathbf{w}_\tau^T \tag{4}$$

While the *component orientation* feature reflects only the orientation information, its discriminating power should remain the same in all frames. Therefore, no temporal weighting, i.e., only component weighting, is applied on this feature.

5. DISTANCE CALCULATION AND FUSION

The decision stage of the proposed algorithm involves the measurement of the dissimilarity between subjects, i.e., the distances of their corresponding feature functions. We perform this calculation separately for each one of the features, and once the three feature distances are calculated, we combine them into a unique distance that expresses the dissimilarity between subjects.

The *area* feature function f_A is scalar, so the energy of the difference function is suitable for calculating the total area distance between two subjects:

$$D_A(p, q) = \sqrt{\frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} w_A^2(m, n) \left(f_A^p(m, n) - f_A^q(m, n) \right)^2} \tag{5}$$

where f_A^p and f_A^q are the two-dimensional *area* feature functions of the pth ($p = 1, 2, ..., N_P$) probe sequence and the qth ($q = 1, 2, ..., N_G$) gallery sequence respectively, and M is the number of body components that are considered.

The feature function $\mathbf{f}_V(m, n)$ is a vectors function, of which the magnitude contains important information. Therefore, the average Euclidean distance is used for this feature:

$$D_V(p, q) = \sqrt{\frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} w_V^2(m, n) \| \mathbf{f}_V^p(m, n) - \mathbf{f}_V^q(m, n) \|^2} \tag{6}$$

where $\| \cdot \|$ denotes the Euclidean distance operator.

In the *component orientation* feature function $\mathbf{f}_O(m, n)$, its elements are vectors but only contain orientation information, so the difference between two elements should be based on the angle between the two orientation vectors. In order to calculate this angle, the scalar product of two corresponding elements (in position (m, n)) is expressed as:

$$\mathbf{o}_{mn}^p \cdot \mathbf{o}_{mn}^q = |\mathbf{o}_{mn}^p| |\mathbf{o}_{mn}^q| \cos \theta_{mn}^{pq} \tag{7}$$

where $|\mathbf{o}_{mn}^p|$ and $|\mathbf{o}_{mn}^q|$ are the magnitudes of vector $\mathbf{o}_{mn}^p$ and $\mathbf{o}_{mn}^q$, and θ_{mn}^{pq} is the angle between them. As described in the previous section, $\mathbf{o}_{mn}$ is the eigenvector corresponding to the largest eigenvalue, and its magnitude is normalized to 1. Therefore, the distance for this feature is calculated as:

$$D_O(p, q) = \sum_{m=1}^{M} \sum_{n=1}^{N} w_O(m) \sqrt{1 - (\mathbf{o}_{mn}^p \cdot \mathbf{o}_{mn}^q)^2} \tag{8}$$

Before calculating the total distance, we normalize the distances for each feature using a min-max normalization method [6].

The three normalized distances, $\widetilde{D}_A$, $\widetilde{D}_V$, and $\widetilde{D}_O$,are combined together to obtain a total distance. Since the importance of those three distances are different, weights need to be put on them as well:

$$D = w_A \widetilde{D}_A + w_V \widetilde{D}_V + w_O \widetilde{D}_O \tag{9}$$

In this work, the above weight values were selected based on experimentation. However, the performance of the algorithm is generally insensitive to small weight variations.

Probe	Rank 1 Recognition Rate (%)				Rank 5 Recognition Rate (%)			
	BL [3]	CBGR [4]	LDM [5]	FGF	BL [3]	CBGR [4]	LDM [5]	FGF
B	46	49	51	**56**	66	78	73	**78**
D	23	26	21	**29**	39	53	43	**62**
H	9	16	20	**34**	36	46	44	**52**
K	12	13	6	**15**	39	39	39	**39**
Average	23	27	25	**34**	45	54	50	**58**

Table 1. The recognition rates for the proposed and three other existing methods.

6. EXPERIMENTAL RESULTS

For the experimental evaluation of our method, we use the manually-labelled sequences from the USF database [7]. In this database, each of the sequences includes one full gait cycle and each manual silhouette is segmented in eight areas. In the USF data sets, the right arm and right thigh of the subjects are partially occluded in most frames and even entirely missed in some frames, therefore, in the proposed algorithm, only the remaining six components are considered. In order to make a fair comparison, we followed the same experimental setup as in [3, 4, 5], in which five key data sets are used in the experiments: Gallery, Probe B, D, H and K, and the numbers of subjects in them are 71, 41, 70, 70 and 33 respectively. In particular, we present results obtained by using each of the features separately as well as by combining all three features. These results will be compared with the results of three other existing methods.

In order to evaluate the performance of the compared methods, Cumulative Match Scores (CMS) are calculated. In CMS, a Rank k result denotes the percentage of probe subjects whose corresponding gallery subject is within the top k matches. In Table 2, average Rank 1 and Rank 5 scores of all four probe sets obtained using only one feature and using the combination of the features are shown. As seen, among the three features, the *component orientation* yields better results than the *component area* and the *component centre*. The last row of the table presents the performance of a system based on the combination of the three features. As seen, the Fusion of Gait Features (FGF) outperforms any single feature for any probe set and at any rank.

The best performing of our methods above was also compared with three other methods. Specifically, our FGF method was compared to the Baseline algorithm (BL) [3], the Component Based Gait Recognition (CBGR) method [4], and the method based on Layered Deformable Model (LDM) [5]. In Table 1, the proposed and the other three methods' recognition rates at Rank 1 and 5 are tabulated. As we can see, at Rank 1, the proposed method achieves much better results for all probe sets and a considerably higher average score, compared to all other existing methods. At rank 5, the proposed method yields equally good or better results for all probe sets, and the average score is also much higher than that of the other methods.

Feature	Rank 1 (%)	Rank 5 (%)
Component Area	16	34
Component Centre	19	46
Component Orientation	26	50
FGF	**34**	**58**

Table 2. The average recognition rates of four probe sets for single and combined features.

7. CONCLUSIONS

This paper proposed a novel gait recognition method that was applied on manually labelled sequences. The proposed method uses three discriminative component-based features, namely, the *area* and the *orientation* of body components, as well as the vector distance between *centre*s of body components and the whole body. By combining these three features, improved performance is achieved in comparison to other existing methods that use manually extracted silhouettes.

8. REFERENCES

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

[2] L. Lee and W.E.L. Grimson, "Gait analysis for recognition and classification," in *Proc. Int. Conf. Automatic Face and Gesture Recognition*, Washington, DC, May 2002, pp. 148–155.

[3] Z. Liu and S. Sarkar, "Effect of silhouette quality on hard problems in gait recognition," *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 35, pp. 170–183, 2005.

[4] N.V. Boulgouris and Z.X. Chi, "Human gait recognition based on matching of body components," *Pattern Recognition*, vol. 40, no. 6, pp. 1763–1770, 2007.

[5] H. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos, "A full-body layered deformable model for automatic model-based gait recognition," *EURASIP Journal on Advances in Signal Processing*, 2008.

[6] A. Ross, K. Nandakumar, and A.K. Jain, *Handbook of Multibiometrics*, Springer, 2006.

[7] P.J. Phillips, S. Sarkar, I. Robledo, P. Grother, and K.W. Bowyer, "The gait identification challenge problem: Data sets and baseline algorithm," in *Proc. Int. Conf. Pattern Recognition*, Quebec City, Canada, Aug. 2002, pp. 385–388.