Reducing the Effect of Noise on Human Contour in Gait Recognition

Shiqi Yu, Daoliang Tan, Kaiqi Huang, and Tieniu Tan

National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences,
P.O. Box 2728, Beijing, 100080, China
{sqyu,dltan,kqhuang,tnt}@nlpr.ia.ac.cn

Abstract. Gait can be easily acquired at a distance, so it has become a popular biometric especially in intelligent visual surveillance. In gait-based human identification there are many factors that may degrade the performance, and noise on human contours is a significant one because to extract contours perfectly is a hard problem especially in a complex background. The contours extracted from video sequences are often polluted by noise. To improve the performance, we have to reduce the effect of noise. Different from the methods which use dynamic time warping (DTW) in previous work to match sequences in the time domain, a DTW-based contour similarity measure in the spatial domain is proposed to reduce the effect of noise. The experiments on a large gait database show the effectiveness of the proposed method.

1 Introduction

Gait has recently received an increasing interest from researchers. Gait is an attractive biometric feature for human identification at a distance because it is non-contact, non-invasive and can be easily acquired at a distance in contrast with other biometrics. Now gait has been considered as a suitable biometric for human identification in visual surveillance.

Many gait recognition methods have been developed in recent years, and most of them can be explicitly classified into two main categories: model-based approaches and appearance-based approaches. Model-based approaches [1,2,3] generally aim to recover gait features, such as stride dimensions, limb lengths and kinematics of joint angles, by model matching in each frame of a walking sequence. Appearance-based approaches [4,5] usually use the silhouettes as a whole to analyze dynamic features of motion bodies, so these methods are efficient and simple, and most of existing gait recognition approaches belong to this category.

The human contour, outline of the sihouette, is popular [5,6,7,8] in appearance-based gait recognition. A contour is more compact than its corresponding silhouette because only the points on the shape outline are considered. The shortcoming of contours is that they are more sensitive to noise than silhouettes. Dynamic time warping (DTW) is a common technique for matching time-varying signals. It has been used for matching sequences in the time domain [9,10,11,12]. We use DTW in the spatial domain and

proposed two contour similarity measures to make contour-based gait recognition algorithms robust to noise. Experiments were carried out on a large gait database, CASIA Gait Database (Dataset B) [13], which contains 124 subjects. The experimental results show that the proposed method can greatly improve the performance.

The remainder of this paper is organized as follows. Section 2 briefly introduces the contour extraction and representation, and Section 3 describes the effect of noise on contours. The proposed similarity measures are presented in Section 4. Experiments and results are shown in Section 5. Section 6 concludes the paper.

2 Human Contour Extraction

If video sequences are captured by fixed cameras, human silhouettes can be extracted by background subtraction and thresholding. A human contour, outline of a body, can be easily obtained from a human silhouette by a border following algorithm based on connectivity. The contour can be further viewed as being in the complex plane [7]. Each point on the contour can be represented by a complex number:

$$p_i = x_i + j \cdot y_i, (i = 1, 2, \cdots, N) \tag{1}$$

where N is the number of points on the contour. An equal number of points are used to express each contour. And each contour are counterclockwise unwrapped from the top point of the contour to turn it into a complex vector:

$$s = [p_1, p_2, \cdots, p_N]^T. \tag{2}$$

Here, walking direction is normalized based on gait symmetry in the side view.

Suppose a video sequence contains C gait cycles, then the sequence can be divided into C short sequences using the method mentioned in [14] and each one contains a gait cycle. To reduce the influence of the walking speed, contours in a gait cycle are interpolated to the same number T. The C gait cycles in a video sequence $\mathbb{S}$ are shown as follows:

$$
\begin{aligned}
S_1 &= [s_1, & s_2, & \cdots, s_T &] \\
S_2 &= [s_{T+1}, & s_{T+2}, & \cdots, s_{2T} &] \\
& \cdots \\
S_C &= [s_{(C-1)T+1}, s_{(C-1)T+2}, \cdots, s_{CT} &]
\end{aligned}
\tag{3}
$$

where s_i is a contour as illustrated in Equation 2, and S_c is a $N \times T$ matrix which represent a gait cycle.

The frames in a gait cycle are aligned to make the stance with the greatest width be the first frame of the cycle. Some stances from a gait cycle are shown in Fig. 1.
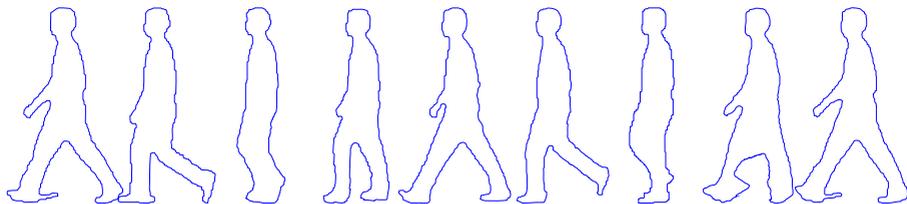

Fig. 1. Stances from a gait cycle

3   The Effect of Noise on Contours

To segment the human body is a hard problem especially in a complex background. The extracted silhouettes and contours may be affected by noise. Fig. 2.(a) shows a silhouette with noise where a portion is lost, and Fig. 2.(b) shows the corresponding ground-true silhouette. The following formula is used to measure the difference between a feature and its corresponding ground-truth feature.

$$\rho = \frac{|\mathbf{f} - \mathbf{f}_{gt}|}{|\mathbf{f}_{gt}|} \qquad (4)$$

where $\mathbf{f}$ is the feature vector and $\mathbf{f}_{gt}$ is the ground-truth feature vector. The ρ value between the two silhouettes (Fig. 2(a) and (b)) is 1.8%. This means that 1.8% of the human silhouette area is different.
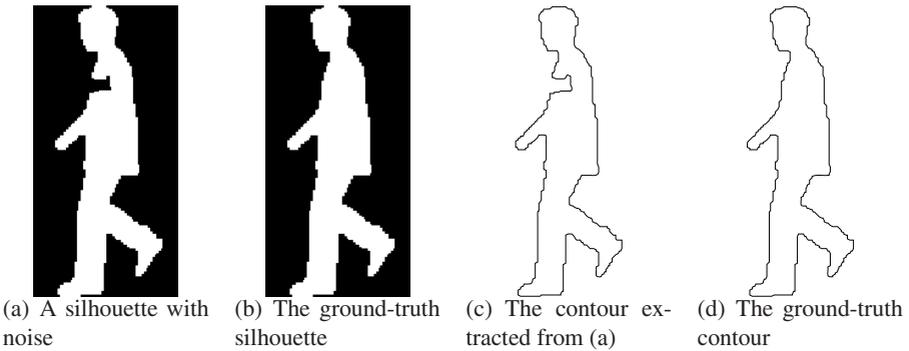

(a) A silhouette with noise (b) The ground-truth silhouette (c) The contour extracted from (a) (d) The ground-truth contour

Fig. 2. Silhouette with noise and the ground-truth one

Fig. 2(c) and (d) are the contours extracted from Fig. 2(a) and (b), respectively. We use the formula in Equation 2 to represent contours. The difference ratio, ρ, between the two contours is 7.6%, and it is over 4 times greater than the ρ value of silhouettes. It means that noise has a greater effect on contours than on silhouettes. This is consistent with one's intuition.

To improve contour-based methods, one solution is to improve human detection and segmentation algorithms. But it is not easy to segment human body perfectly from background especially in a complex outdoor scene. Another solution is to improve the similarity measure for contours. In the following section, we propose two DTW-based contour similarity measures.

4   Algorithm Overview

To improve the robustness of contour to noise, one straightforward idea is to match contours first before computing the similarity. If each of the body components, such as head, hands and feet, can be correctly matched between two human body contours, the effect of noise should be greatly reduced. Dynamic time warping is used in our proposed method to match two contours.

4.1 Dynamic Time Warping

Dynamic time warping (DTW) [15] is an algorithm for measuring the similarity between two series, which may vary in time or speed, and has been widely used in speech processing, gesture recognition [16], etc. DTW can find an optimal match between two given series P and Q. Suppose the length of P and Q are n and m, respectively:

$$P = [p_1, p_2, \cdots, p_m]^T \tag{5}$$

$$Q = [q_1, q_2, \cdots, q_n]^T \tag{6}$$

To find the best alignment of the two series, an $m \times n$ matrix D, namely a cost matrix, is created. The element $d(i, j)$ of the matrix D is the distance between p_i and q_j:

$$d(i, j) = dist(p_i, q_j) \tag{7}$$

where $dist$ can be the Euclidean distance or other kinds of distances. In our experiments, the Euclidean distance was used.

A warp path W can be constructed as shown in Fig. 3 and expressed in the following way:

$$W = [w_1, w_2, \cdots, w_K] \qquad \max(m, n) \leq K \leq m + n \tag{8}$$

where K is the length of the warp path, and the kth element of W, w_k, is $\langle i, j \rangle$ which means the ith element of P corresponds with the jth one of Q.

To ensure every element of the two series to be used in the warp path, the warp path must start at the beginning of each series at $w_1 = \langle 1, 1 \rangle$ and finish at the end of both series at $w_K = \langle m, n \rangle$. Another constraint on the warp path is the two neighbor steps must be adjacent cells (including diagonally adjacent cells). Given $w_k = \langle i, j \rangle$, $w_{k+1} = \langle i', j' \rangle$ must satisfy $i' - i \leq 1$ and $j' - j \leq 1$.
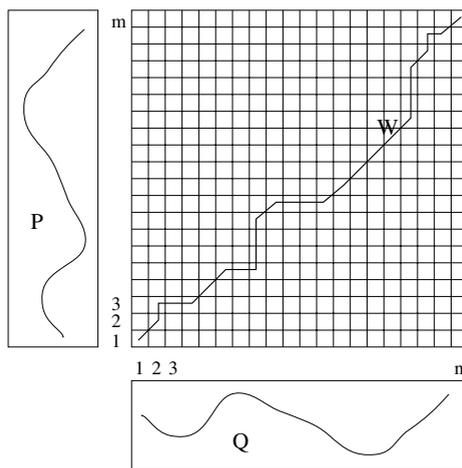
Fig. 3. An example of dynamic time warping

The optimal warp path is the minimum-distance one, where the distance of a warp path is

$$L = \sum_{k=1}^{K} d(i_k, j_k) = \sum_{k=1}^{K} d(w_k) \tag{9}$$

and $w_k = \langle i_k, j_k \rangle$ is the kth matched pair of the two series. This optimal path can be found very efficiently using dynamic programming [17].

4.2 Contour Similarity Measures

Since an optimal match can be found through DTW, using the average path distance of L to measure the similarity between two contours is a matter of course. The average distance between two contours P and Q is

$$\bar{L}(P, Q) = \frac{1}{K} \sum_{k=1}^{K} d(w_k) \tag{10}$$

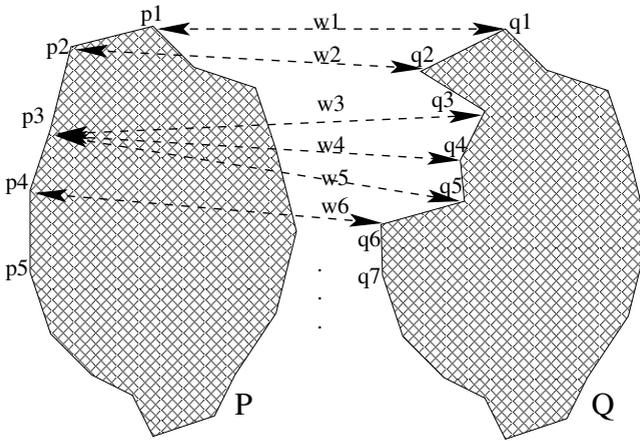where $W = [w_1, w_2, \cdots, w_K]$ is the optimal match between the contours P and Q.

Fig. 4. The match between two contours

The contours can be affected by noise and have some distortions (q_3, q_4 and q_5 on the contour Q in Fig. 4). When this occurs, a point on a contour will be aligned to multiple points. That is, p_3 on the contour P will be aligned to q_3, q_4 and q_5 on the contour Q, and w_3, w_4 and w_5 have a common point p_3. To reduce the effect of noise, only the pair with the shortest distance (w_4) is kept and others pairs which have a common point are discarded in calculation of the contour similarity. So the distance between the contours P and Q can also be defined as

$$\bar{\mathbb{L}}(P, Q) = \frac{1}{K'} \sum_{k \in \mathbb{K}} d(w_k) \tag{11}$$

where K' is the number of elements in $\mathbb{K}$, $\mathbb{K}$ is a subset of $\{1, 2, \cdots, K\}$, and the pairs with a common point are removed and only the one with the shortest distance is kept.

Now we have two similarity measures: the traditional DTW average distance $\bar{L}$ and the improved DTW average distance $\bar{\mathbb{L}}$.

4.3   Video Sequence Similarity Measures

Suppose that there are two gait sequences $\boldsymbol{S}^G$ and $\boldsymbol{S}^P$, with sequence $\boldsymbol{S}^G$ from the gallery set (the training set), and sequence $\boldsymbol{S}^P$ from the probe set (the test set). The two sequences can be divided ino several gait cycles as illustrated in Equation 3. The distance between the two sequences is

$$SeqDist(\boldsymbol{S}^G, \boldsymbol{S}^P) = \min_{i,j}(CycleDist(\boldsymbol{S}_i^G, \boldsymbol{S}_j^P))$$
$$i \in \{1, 2, \cdots, C^G\}, j \in \{1, 2, \cdots, C^P\} \tag{12}$$

where $\boldsymbol{S}_i^G$ is the ith gait cycle in the sequence $\boldsymbol{S}^G$, and $\boldsymbol{S}_j^P$ is the jth gait cycle in the sequence $\boldsymbol{S}^P$. The distance between these two gait cycles can be:

$$CycleDist(\boldsymbol{S}_i^G, \boldsymbol{S}_j^P) = \frac{1}{T} \sum_{t=1}^{T} \bar{L}(\boldsymbol{s}_t^G, \boldsymbol{s}_t^P) \tag{13}$$

or

$$CycleDist(\boldsymbol{S}_i^G, \boldsymbol{S}_j^P) = \frac{1}{T} \sum_{t=1}^{T} \bar{\mathbb{L}}(\boldsymbol{s}_t^G, \boldsymbol{s}_t^P) \tag{14}$$

where $\boldsymbol{s}_t^G$ is the tth contour in the gait cycle $\boldsymbol{S}_i^G$, and $\boldsymbol{s}_t^P$ is the tth contour in the gait cycle $\boldsymbol{S}_j^P$.

5   Experimental Results

The proposed method was tested on CASIA Gait Database (Dataset B) [13] which contains 124 subjects (93 males and 31 females) captured from 11 views. One experiment was carried out for each view. There are six sequences for each subject at each view. The first four sequences were put into the gallery set, and the last two were put into the probe set as in [13]. Each sequence in the probe set was compared with all sequences in the gallery set using the DTW distance in Equation 12. Besides the DTW distance, the Euclidean distance was also involved for performance comparison.

Because the focus here is the contour similarity measure, a simple and most commonly used classifier, nearest neighbor classifier, was used in the experiments. The correct classification rates (CCRs) are listed in Table 1. We can find from Table 1 that the DTW-based measures greatly improve the recognition rate of gait recognition. The improve DTW-based measure, $\bar{\mathbb{L}}$, which achieves the highest CCR. The average CCR by the Euclidean distance is only 67.7%. The KFD method in [18] obtained better results than the Euclidean distance, but the average CCR is just 2.6% higher than that by the Euclidean distance. The improvement by the KFD method is not as great as the DTW-based methods achieved.

Table 1. CCRs by different methods(%)

Method	0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	Average
Improved DTW($\bar{\mathbb{L}}$)	93.5	84.7	85.9	80.2	83.9	83.5	73.0	80.6	89.9	90.7	92.7	85.3
Trad. DTW($\bar{L}$)	91.1	81.9	83.1	79.4	81.0	81.0	70.6	80.6	88.7	86.7	90.3	83.1
Euclidean Dist.	60.5	51.2	63.3	67.3	72.2	71.4	63.3	72.6	79.0	77.8	66.1	67.7
KFD in [18]	71.8	49.2	72.6	69.4	77.8	75.0	69.8	71.4	71.0	77.8	67.7	70.3
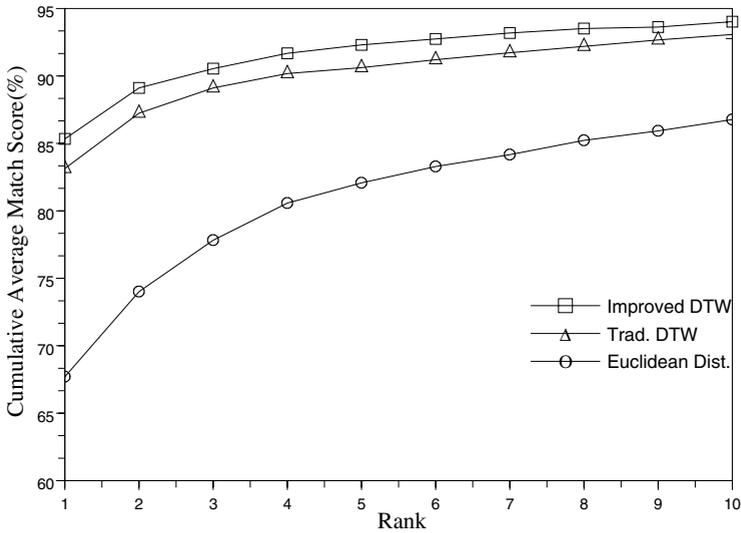
Fig. 5. Cumulative Average match scores of three different measures

Cumulative match scores (CMS) are also used to evaluate the performance of proposed methods. The match scores in Fig. 5 are the average match scores of 11 views from rank 1 to 10. From Fig. 5, we can draw similar conclusions to that from Table 1.

Although the DTW-based contour similarity measures are simple methods, they greatly improve the robustness to noise. The experiments discover their potential in gait recognition and show contour matching is effective in reducing the impact of noise.

6   Conclusions and Future Work

We have presented two novel contour similarity measures for gait recognition. DTW is used to match the points on two contours before computing their similarity. The similarity between two contours is computed based on their optimal match. Experiments on a large gait database show that the DTW-based measures are superior to direct contour comparison and even some other feature extraction methods. Although the proposed

method is simple and straightforward, it indicates a principle that contour matching can reduce the effect of noise.

In future we will design more sophisticated similarity measures to make the algorithm more robust to noise and even some other variations. Some other advanced classifiers, such as SVM and AdaBoost, will also be investigated.

Acknowledgements

References

1. Johnson, A.Y., Bobick, A.F.: A multi-view method for gait recognition using static body parameters. In: Proc. of 3rd International Conference on Audio and Video Based Biometric Person Authentication, Halmstad, Sweden, pp. 301–311 (June 2001)
2. Wang, L., Ning, H., Tan, T., Hu, W.: Fusion of static and dynamic body biometrics for gait recognition. IEEE Transactions on Circuits and Systems for Video Technology 14(2), 149–158 (2004)
3. Yam, C.Y., Nixon, M.S., Carter, J.N.: On the relationship of human walking and running: automatic person identification by gait. In: Proc. of International Conference on Pattern Recognition, Quebec,Canada, pp. 287–290 (2002)
4. Mowbray, S.D., Nixon, M.S.: Automatic gait recognition via fourier descriptors of deformable objects. In: Proc. of 4th International Conference on Audio- and Video-based Biometric Person Authentication, Guildford, UK, pp. 566–573 (June 2003)
5. Wang, L., Tan, T., Ning, H., Hu, W.: Silhouette analysis-based gait recognition for human identification. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(12), 1505–1518 (2003)
6. Niyogi, S.A., Adelson, E.H.: Analyzing and recognizing walking figures in xyt. In: CVPR, pp. 469–474. IEEE Computer Society Press, Los Alamitos (1994)
7. Yu, S., Wang, L., Hu, W., Tan, T.: Gait analysis for human identification in frequency domain. In: Proc. of the 3rd International Conference on Image and Graphics, Hong Kong, China, pp. 282–285 (December 2004)
8. Lam, T.H.W., Lee, R.S.T.: Human identification by using the motion and static characteristic of gait. In: ICPR, pp. 996–999. IEEE Computer Society Press, Los Alamitos (2006)
9. Cuntoor, N., Kale, A., Chellappa, R.: Combining multiple evidences for gait recognition. In: Proc. of the International Conference on Acoustics Speech and Signal Processing, Hong Kong (cancelled), April 6-10, 2003, vol. III, pp. 113–116 (2003)
10. Boulgouris, N.V., Plataniotis, K.N., Hatzinakos, D.: Gait recognition using dynamic time warping. In: Proc. of the 6'th IEEE workshop on multimedia signal processing, BC, Canada, pp. 263–266. IEEE Computer Society Press, Los Alamitos (2004)
11. Pandey, N., Abdulla, W., Salcic, Z.: Human identification using vector quantisation technique. In: Proceedings of the Eighth International Symposium on Signal Processing and Its Applications, Sydney, Australia, pp. 671–674 (August 2005)
12. Veeraraghavan, A., Member, S., Roy-Chowdhury, A.K., Chellappa, R.: Matching shape sequences in video with applications in human movement analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(12), 1896–1909 (2005)

13. Yu, S., Tan, D., Tan, T.: A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In: Proc. of the 18'th International Conference on Pattern Recognition (ICPR06), Hong Kong, China, pp. 441–444 (August 2006)
14. Sarkar, S., Phillips, P.J., Liu, Z., Vega, I.R., Grother, P., Bowyer, k.W.: The humanid gait challenge problem: Data sets, performance, and analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(2), 162–177 (2005)
15. Kruskal, L.: The symmetric time-warping problem: From continuous to discrete. In: Sankoff, D., Kruskal, J.B. (eds.) Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison. Addison-Wesley, Reading (1983)
16. Veeraraghavan, A., Chellappa, R., Roy-Chowdhury, A.K.: The function space of an activity. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), New York, USA, pp. 959–968. IEEE Computer Society Press, Los Alamitos (2006)
17. Keogh, E.J., Pazzani, M.J.: Derivative dynamic time warping. In: Proc. of the First SIAM International Conference on Data Mining, Chicago, USA, pp. 285–289 (April 2001)
18. Yu, S., Tan, D., Tan, T.: Modelling the effect of view angle variation on appearance-based gait recognition. In: Proc. of the 7'th Asian Conference on Computer Vision, Hyderabad, India, pp. 807–816 (January 2006)