Journal of Computational Information Systems 7:4 (2011) 1155-1162
Available at http://www.Jofcis.com

Gait Modeling and Identifying Based on Dynamic Template Matching

Hao ZHANG†, Zhijing LIU, Haiyong ZHAO

School of Computer Science and Technology, Xidian University, Xi'an 710071, China

Abstract

We describe a new approach for gait modeling and identifying based on dynamic template matching, as it is suitable to video surveillance system requiring high accuracy and low computational complexity. This method reduces computational cost in gait recognition, significantly improves the recognition accuracy for gait and is applicable for video surveillance. Human appearance is extracted entirely from binary image of human silhouette in video sequences after Gaussian mixture models are implemented. First, we unfold the contour clockwise, and every video sequence is converted into a normalized distance in one dimensionality. Then we construct gait templates according to different subjects. Additionally, the distance based on dynamic time warping (DTW) is employed to measure the similarity between the probe sequences and gallery ones with predetermined parameters. Finally, the similarity is induced to identify human gait by DTW distance. In our experiments, we obtain about 92% in recognition accuracy. This approach is not only simple representation, but also applicable for video surveillance.

Keywords: Gait Modeling; Feature Extraction; Dynamic Template; Dynamic Time Warping (DTW); Video Surveillance

1. Introduction

Biometrics research has become a hot topic due to the demand of automatic human authentication and authorization in computer systems. Biometric resources such as iris, fingerprint, palmprint, and shoeprint have been thoroughly studied and employed in many applications. In vision community, human motion has been studied intensively [1], and a significant amount of progress has been achieved on human gait analysis and recognition.

The effective representation and measurement of gait are key issues. Currently, there are several successful representation models such as appearance-based models [2-5], stochastic statistical models [3], articulated biomechanics models [6], in which a set of parameters describes the gait, and other parameter-based models [7]. Several of these models can be combined to further improve the gait representation. Many appearance-based models have been developed for human gait recognition [4]. Some models use the silhouette of the entire body [3]; whereas others use the most discriminant parts [8] such as the torso and the thighs. Gait energy image (GEI) [9] is a spatio-temporal method which is proposed to characterize human walking properties for individual recognition by gait. Gait stances over one cycle is represented by population Hidden Markov Models (pHMM) [10] with their dynamic normalization. Zhang et al. [11] proposed a novel method for activity recognition by key frame detection. General tensor

† Corresponding author.
Email addresses: zhanghao@mail.xidian.edu.cn (Hao ZHANG).

April, 2011

discriminant analysis and Gabor feature [12] have been utilized to represent and analyze for gait recognition based on sequences from USF Human ID database. Similarly, Zhou et al. [13] proposed an object tracking algorithm based on dynamic template and motion prediction.

Video surveillance requires monitoring and analyzing many scenarios, and recognizing human gait simultaneously. Consequently, to design an efficient method with low spatial and temporal cost is a key issue. This paper represents human features in appearance-based model. An intact human silhouette in image is extracted, and converted into 1D (one-dimensionality) distance with normalized unfolding contour. The distance of Dynamic Time Warping (DTW) between galleries and probes is calculated and compared with the parameters to conclude gait similarity. Our main contributions are building eight states to represent gait cycle, using DTW-distance to measure the similarity and computing efficiently.

The paper is organized as follows. In Section 2, we introduce related theories, including contour unfolding and DTW. This is followed by a discussion of gait modeling process in Section 3, which consists of image acquisition and processing, feature extraction and representation, gait modeling. The experimental results are described and discussion is implemented in Section 4. In Section 5, the conclusions are discussed.

2. Contour Unfolding and Dynamic Time Warping

2.1. Contour Unfolding

Contour unfolding is a key issue for representation of human silhouette. As it is periodic, contour unfolding could describe states of human motion. To reduce silhouette redundancy and computational cost, 2D (two-dimensionality) shapes can be converted to 1D one to represent variable and spatiotemporal states, as shown in Figure 1.

(1) Centroid. Acquiring the human silhouette, we can obtain all points on the contour with algorithm of edge detection such as Canny, and calculate its centroid.

$$x_c = \frac{1}{N_b}\sum_{i=1}^{N_b} x_i \ , \quad y_c = \frac{1}{N_b}\sum_{i=1}^{N_b} y_i \tag{1}$$

Where (x_c, y_c) is the position of centriod, N_b is the number of points, and (x_i, y_i) is the point on the contour.

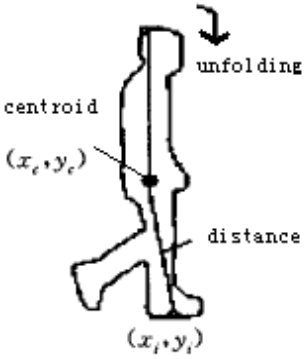

Fig. 1 Contour Unfolding

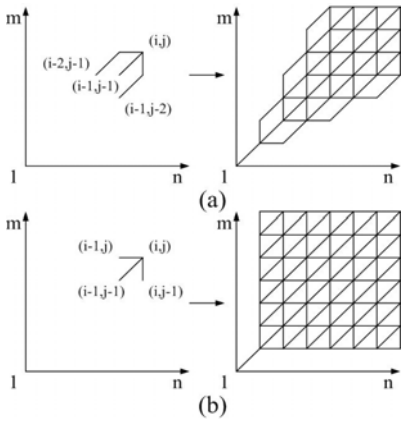

Fig. 2 Two Local Restrictions and Possible Paths

(2) Contour unfolding. As defining origin as left-bottom, we choose one of points which are the maximum in y-axis as starting point, and convert the contour to 1D distance which is the distance between centroid and each point of the contour.

$$d_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2} \tag{2}$$

(3) Normalization. To eliminate the errors in image size and distance length, we analyze statistical data and conclude the scope of normalized values.

2.2. Dynamic Time Warping (DTW)

Dynamic time warping is a flexible algorithm of pattern matching. It can match extensive, compact and deformable patterns in global and local areas to solve the similarity measurement of dynamic pattern and classification.

Assumed two time sequences are defined as follow:

$$A: \{a_1, a_2, \cdots, a_i, \cdots, a_n\}, B: \{b_1, b_2, \cdots, b_j, \cdots, b_m\}$$

The definition of warping path is

$$P=\{p_1, p_2, \cdots, p_k, \cdots, p_K\}$$

where $p_k=(i, j)_k$ and $\max(m, n) \leqslant K \leqslant m+n$.

DTW is denoted by objective functions as follow:

$$J = \frac{1}{\sum_{k=1}^{K} w_k} \min\{\sum_{k=1}^{K} w_k d(i_k, j_k)\} \tag{3}$$

$$P^* = \arg\min\{\sum_{k=1}^{K} w_k d(i_k, j_k)\} \tag{4}$$

where $d(i_k, j_k)$ is local distance between a_i and b_j, $d(i_k, j_k)=|a(i)-b(j)|$, w_k is an weighted factor which weights on local distance to distinguish different ones, $\sum_{k=1}^{K} w_k$ offsets the different warping paths to compare them on the same baseline.

DTW restricts the warping paths, in other words, they must meet global and local restriction. Local restriction consists of points, continuity and consistency restriction.

(1) Points restriction

$$p_1=(1, 1), p_K=(n, m)$$

Points restriction means that the starting and finishing points of two time sequences are consistent. We can lower the restriction of points actually and locate them into related regions.

(2) Continuity restriction

As $p_{k-1}=(i, j)$ and $p_k=(i', j')$ in local path, the continuity restriction is denoted by

$$i'-i \leqslant c, \ j'-j \leqslant c$$

where c is a positive integer, usually 1 or 2. Continuity restriction is used to restrict excessive extension in DTW. Figure 2 shows two typical continuities.

(3) Consistency restriction

As local path $p_{k-1}=(i, j)$ and $p_k=(i', j')$, it requires $i'-i \geqslant 0$ and $j'-j \geqslant 0$.

Global restriction defines a subset of n×m grids as searching area. Its restriction requires the warping path in particular area, such as strip or parallelogram shape.

DTW is solved by optimization, and the solution is induced by iterative in Equation (5) as follow.

$$D(i, j) = d(i, j) + \min \begin{Bmatrix} D(i-2, j-1) + d(i-1, j) \\ D(i-1, j-1) \\ D(i-1, j-2) + d(i, j-1) \end{Bmatrix} \tag{5}$$

Different expressions are concluded by different local restrictions. Equation (5) is local restriction of Figure 2(a), and its initialization is denoted as follow:

$$D(2, 2)=d(1, 1)+d(2, 2)$$

By the local restriction of Figure 2(b), the iterative expression is denoted by

$$D(i, j) = d(i, j) + \min\{D(i-1, j), D(i-1, j-1), D(i, j-1)\}. \tag{6}$$

3. Our Approach

Building a convenient gait model is a critical issue in our study. Gait model is acquired from dataset, and the procedure is divided into three stages, i.e. image acquisition and processing, feature extraction and representation, gait modeling.

3.1. Image Acquisition and Processing

We adopt some gait videos of dataset as gallery sequences. Gaussian mixture modeling (GMM) is employed for background modeling and moving object detecting. Binary images of human silhouette are extracted by background subtraction. The connective silhouette is done by pre-processing of binary images, i.e. dilation and erosion.

3.2. Features Extraction and Representation

As a cycle of human gait is shown in Figure 3, its features can be detected and represented. Assumed the sequence number of frame is n and the ratio of human height to width is r. Therefore, their mapping is denoted by

$$r=f(n). \tag{7}$$

Where n is between 1 and N, N is total frame number of each sequence.

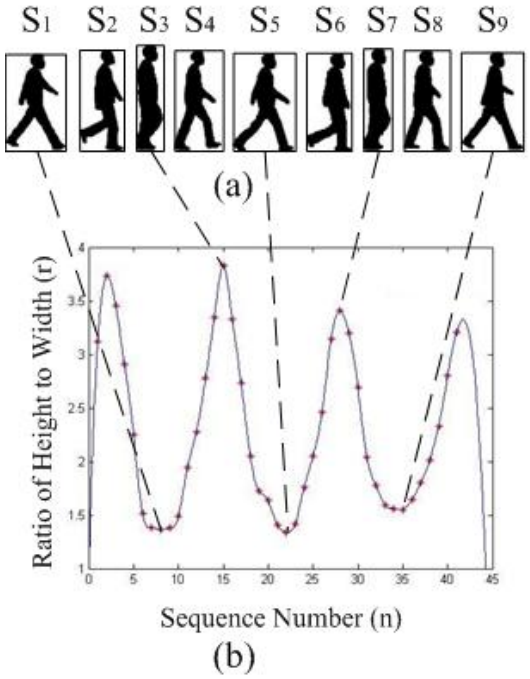

Fig. 3 (a) Some Images and Bounding Rectangles of Human Silhouettes in a Sequence

(b) The Variable Ratios of Height to Width with Bounding Rectangles

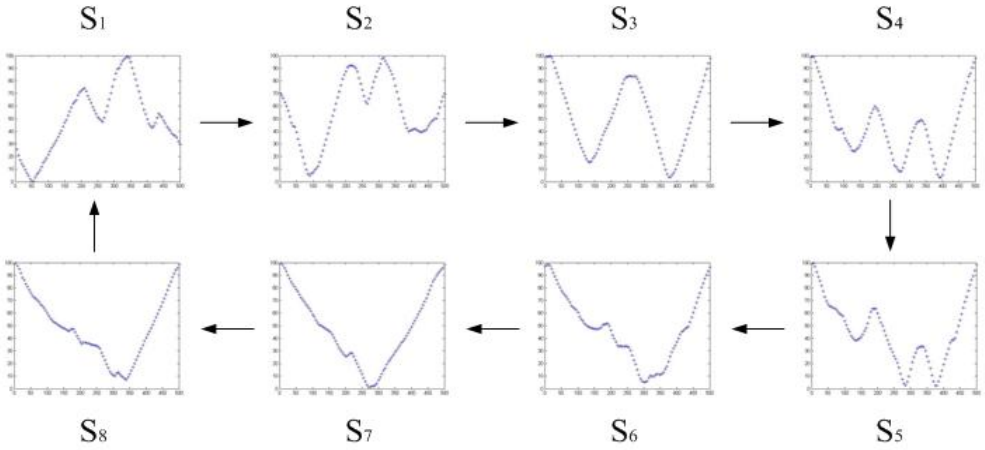

Fig. 4 Eight Transitions of Models States

Gait features are represented in contour unfolding. Assume N_b is the sum of points of contour, and the centroid (x_c, y_c) can be calculated by Equation (1). Assume d_i is the distance between centroid and any point of contour, 1D representations of distance are employed by Equation (2). As statistics of the number of points of contour in the dataset, the number of gallery sequences, which are Gaussian distributions between 400 and 600, is about 95%. Consequently, after normalization, N_b is equal to 500 and D is 100 assumed that D is normalized distance.

3.3. Gait Modeling

To calculate the distance of DTW, we interpolate the normalized data, in which x_i is positive integer and it is between 0 and 500, and the interval is σ, which is the margin of x_{i+1} and x_i. The values of points, $500/\sigma+1$, are induced by given data. A total cycle of gait is extracted as sample, and the motion state is represented by correlations of step length, such as S_i ($1 \leqslant i \leqslant 8$). S_3 and S_7 are maximums of the ratios, and S_1 and S_5 are minimums, while S_2, S_4, S_6 and S_8 are transitions. The sequence numbers corresponding S_1 to S_8 are induced by Equation (7). We select three consecutive frames to calibrate the error, in which the middle frame is n_i ($1 \leqslant i \leqslant 8$), i.e. n_i-1, n_i and n_i+1, and calculate the average to induce 8 related models shown in Figure 4.

We have utilized many gallery sequences in the database to train and calibrate errors. The value of every state is induced by steps, and every distance between gallery and probe sequences in states is calculated by DTW as Equation (6). Every array of data is interpolated by three different intervals when σ is equal to 20, 10 and 5 respectively, then the values of DTW distance are calculated statistically and shown in Table 1.

Table 1 Distributions of Variable Ratios with σ (%)

Ratios of σ	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	Average
20：10	83	77	73	96	79	58	81	67	76.8
10：5	83	69	89	81	67	55	64	67	71.9
20：5	54	53	65	68	53	32	52	45	52.8

Having observed the sequences of gait silhouettes, we are able to identify human walking more precisely from the states i.e. S_1, S_3, S_5 and S_7, compared with the transition states, S_2, S_4, S_6 and S_8. As twenty gallery sequences in the dataset are selected to train and calculate the distance related to states in every cycle, the DTW distance is variable with the transitions from S_1 to S_8. The DTW distances are shorter in S_1, S_3, S_5 and S_7 than S_2, S_4, S_6 and S_8. Therefore, to obtain precise recognition for gait, we can weight different states and sum DTW distance up, assumed that it weights 0.2 for S_1, S_3, S_5, S_7 and 0.05 for the others. The weighted equation is as follow:

$$DIST=(S_1+S_3+S_5+S_7)\times 0.2+(S_2+S_4+S_6+S_8)\times 0.05. \quad (8)$$

Table 2 Distributions of DTW Distance (%)

DIST	σ=20	σ=10	σ=5
[0,50)	15	5	0
[50,100)	35	20	5
[100,150)	35	35	5
[150,200)	5	30	25
[200,250)	5	5	25
[250,300)	5	5	30
[300,350)	0	0	5
[350,∞)	0	0	5

Table 3 Recognition Accuracy in Different Thresholds (%)

	σ=20 and DIST∈[0,150]	σ=10 and DIST∈[0,200]	σ=5 and DIST∈[0,300]	Average
Accuracy	85	90	90	88.3

We calculate DTW distance DIST with these interpolations in gallery sequences and describe their statistical distribution shown in Table 2. By their distribution analysis of DTW distance, we can induce the thresholds from statistical data, as shown in Table 3.

4. Experimental Results and Discussion

We performed the experiments with probe sequences in Dataset B of CASIA and SOTON, where subjects motions are parallel to image plane. The statistical data and results of gait recognition are shown in Table 4.

As the interpolations are different, the weighted DTW distances are different, so are thresholds. It is a key issue for recognition to determine the threshold. In Table 4, compared with the thresholds in gait modeling, average accuracy for gait recognition improves by 3% to 5%, while the total improves by 4.1%. It concludes that recognition accuracy is improved, therefore, this method is effective and the thresholds are reasonable.

Table 4 Experimental Data and Results for Gait Recognition

Database	Dataset	Number of probe sequences	Recognition accuracy (%)				
			σ=20 and DIST∈[0,150]	σ=10 and DIST∈[0,200]	σ=5 and DIST∈[0,300]	Average	Total average
CASIA	Dataset B	100	91.0	92.0	94.0	92.3	91.9
SOTON	Small dataset	20	90.0	95.0	95.0	93.3	
	Large dataset	100	88.0	91.0	91.0	90.0	

5. Conclusions

We have proposed a novel and effective method for gait-based human recognition by DTW. The experimental results and analysis indicate that our method is effective in automated gait recognition. The key advantages include low computation cost and unnecessary learning by large data, both of which suggest it to be real-time and efficient for video surveillance. Whereas it requires clear-cut images, pre-processed images could be implemented for intact human silhouette before feature extraction and representation. Then recognition accuracy depending on the pre-processing stage also suffers from feature extraction, representation and computation.

Acknowledgements

The authors would like to thank CASIA and SOTON to provide activity databases and the anonymous reviewers for their constructive comments.

References

[1] R. Poppe, A survey on vision-based human action recognition, Image and Vision Computing, 28 (2010) 976-990.
[2] S. Sarkar, P. Phillips, Z. Liu, I. Vega, P. Grother, K. Bowyer, The humanID gait challenge problem: Data sets, performance and analysis, IEEE Trans. Pattern Anal. Mach. Intell., 27 (2005) 162-177.
[3] A. Kale, A. Sundaresan, A.N. Rajagopalan, N.P. Cuntoor, A.K. Roy-Chowdhury, V. Kruger, R. Chellappa, Identification of humans using gait, IEEE Trans. Image Processing, 13 (2004) 1163-1173.
[4] Z. Liu, S. Sarkar, Simplest representation yet for gait recognition: Averaged silhouette, Proc. IEEE Int'l Conf. Pattern Recognition, 4 (2004) 211-214.
[5] L. Wang, T. Tan, H. Ning, W. Hu, Silhouette analysis-based gait recognition for human identification, IEEE Trans. Pattern Anal. Mach. Intell., 25 (2003) 1505-1518.
[6] D. Cunado, M. Nixon, J. Carter, Automatic extraction and description of human gait models for recognition purposes, Computer Vision and Image Understanding, 90 (2003) 1-41.
[7] R. Cutler, L. Davis, Robust periodic motion and motion symmetry detection, Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, (2000) 615-622.
[8] L. Lee, G. Dalley, K. Tieu, Learning pedestrian models for silhouette refinement, Proc. IEEE Int'l Conf. Computer Vision, 1 (2003) 663-670.
[9] J. Han, B. Bhanu, Individual recognition using gait energy image, IEEE Trans. Pattern Anal. Mach. Intell., 28 (2006) 316-322.
[10] Z. Liu, S. Sarkar, Improved gait recognition by gait dynamics normalization, IEEE Trans. Pattern Anal. Mach. Intell., 28 (2006) 863-876.
[11] H. Zhang, Z. Liu, H. Zhao, G. Cheng, Recognizing human activities by key frame in video sequences, Journal of Software, 5 (2010) 818-825.
[12] D. Tao, X. Li, X. Wu, S.J. Maybank, General tensor discriminant analysis and gabor features for gait recognition, IEEE Trans. Pattern Anal. Mach. Intell., 29 (2007) 1700-1715.
[13] Z. Zhou, J. Zhang, L. Fang, Object tracking based on dynamic template and motion prediction, Journal of Computational Information Systems, 5 (2009) 1841-1846.