A STUDY ON VIEW-INSENSITIVE GAIT RECOGNITION

Ju Han, Bir Bhanu and Amit K. Roy-Chowdhury

Center for Research in Intelligent Systems
University of California, Riverside, California 92521, USA
{jhan,bhanu,amitrc}@cris.ucr.edu

ABSTRACT

Most gait recognition approaches only study human walking frontoparallel to the image plane which is not realistic in video surveillance applications. Human gait appearance depends on various factors including locations of the camera and the person, the camera axis and the walking direction. By analyzing these factors, we propose a statistical approach for view-insensitive gait recognition. The proposed approach recognizes human using a single camera, and avoids the difficulties of recovering the human body structure and camera calibration. Experimental results show that the proposed approach achieves good performance in recognizing individuals walking along different directions.

1. INTRODUCTION

Gait, which concerns recognizing individuals by the way they walk, has been an important biometric for recognizing non-cooperating individuals at a distance in real-world changing environmental conditions. Most existing gait recognition approaches only study human walking frontoparallel to the image plane [2, 4, 5, 7]. This makes them inapplicable for recognizing individuals walking along various directions in real video surveillance applications as illustrated in Figure 1.

To address this problem, Shakhnarovich et al. [6] estimate the sequence at any arbitrary view from multiple views from four fixed cameras. Their method is view invariant but the environmental requirement is very strict and inapplicable in real applications. Kale et al. [3] synthesize a side view silhouette sequence from a sequence of any arbitrary view if the person is far enough from the camera. Their method needs camera calibration that may not be available in real video surveillance applications. Wang et al. [8] include sequences of different walking derections in the database and evaluate the performance using the leave-one-out cross-validation rule. The results of using sequences of only one specific direction for training are not reported.

In this paper, we propose a statistical approach for view-insensitive gait recognition by analyzing the common properties of human walking along different directions. The proposed approach can recognize human walking along different directions according to the training sequence with a given walking direction, and avoids the difficulties of recovering the human body structure and camera calibration.

Fig. 1. Examples of human walking along different directions.

2. TECHNICAL APPROACH

2.1. Silhouette Preprocessing

We extract silhouette from original human walking sequences by a standard background subtraction approach, and begin with the extracted binary silhouette image sequences. The silhouette preprocessing includes size normalization (proportionally resizing each silhouette image so that all silhouettes have the same height) and horizontal alignment (centering the upper half silhouette part with respect to its horizontal centroid). Regular human walking can be considered as cyclic motion where human motion repeats at a stable frequency. Therefore, it is possible to divide the whole gait sequence into cycles and study them separately. In the preprocessed silhouette sequence, the time series signal of lower half silhouette part size from each frame indicates the gait frequency and phase information. We estimate the gait frequency and phase by a maximum entropy spectrum estimation method [4].

2.2. Gait Representation

In a cyclic human walking sequence, the order of poses in the cycles is the same, i.e., the limbs move forward and backward in a similar way among normal people. The difference exists in the phase of poses in a walking cycle, the extend of limbs, and the shape of the torso, etc. Therefore, it is possible to compose a spatio-temporal template in a single image instead of a ordered image sequences as usual. In our approach, we use gait energy image (GEI) as the gait template for individual recognition [1].

0-7803-9134-9/05/$20.00 ©2005 IEEE

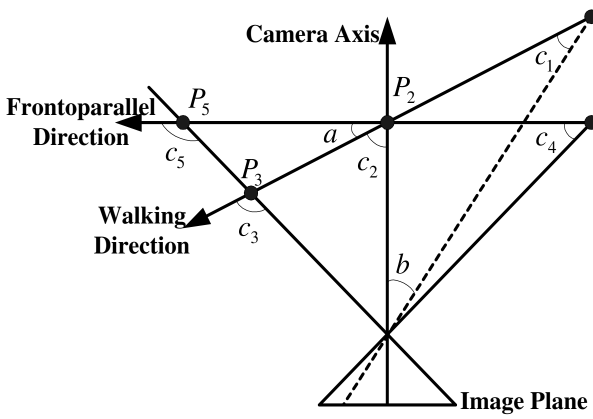

Fig. 2. The graph illustration (from the top view) of human walking along different directions.

Given a preprocessed binary gait silhouette sequence $B(x, y, t)$, the grey-level gait energy image (GEI) is defined as follows

$$G(x, y) = \frac{1}{N} \sum_{t=1}^{N} B(x, y, t) \tag{1}$$

where N is the number of frames in the complete cycle(s) of a silhouette sequence, t is the frame number of the sequence (moment of time), x and y are values in the 2D image coordinate. GEI reflects major shapes of silhouettes and their changes over the gait cycle, and is not sensitive to incidental silhouette errors in individual frames.

2.3. View-Insensitive Gait Templates

In real video surveillance applications, training and testing data may contain people walking along various directions. Figure 2 shows the graph illustration of human walking along different directions from the top view. When a person walks, he/she is only visible within the camera view range. In this figure, there are three kind of angles: walking direction angle a, camera view b and walking view c. a is fixed in a walking sequence assuming that the person does not change the walking directions in the scene, b varies when the person is located at different positions, and c is determined by a and b. It is the walking view c that directly affects the shape of human in the image. That is, if the walking view c and human pose are the same for the same person at two moments of time, their shapes in the image are the same after size normalization. Even a person walks along the same direction in a scene, the walking view c is still different at different 3D locations. This makes the human shapes/silhouettes with the same pose different when their 3D locations are different. As show in Figure 2, if a person walks along the direction with angle a from P_1 through P_2 to P_3, the walking view c will change from c_1 through c_2 to c_3. If the same person walking along the frontaparallel direction from P_4 through P_2 to P_5, the walking view c will change from c_4 through $90°$ to c_5. The overlapping walking view of the two sequence ranges from c_4 to c_3 if $c_4 < c_3$. In this case, the parts within the overlapping walking view range can be used for view-insensitive gait recognition. When the difference of walking direction a is too large or the camera view range is too small, c_4 may be greater than c_3. Although there is no overlapping walking view range of the two sequences in this case, the beginning of the frontoparallel sequence and the end of the sequence with walking direction a may have similar walking views that still provide cues for recognizing human walking along different directions.

The walking views are generally unavailable in real applications. However, the above-mentioned fact indicates that the walking view (the angle between the walking direction and the camera view) ranges of human walking along different directions may be overlapping or adjacent if the difference of their walking directions is not too large. Therefore, we can generate a series of GEIs at a certain interval from each sequence, and some GEIs so obtained from the two sequences may have the similar appearance. If the GEIs of one sequence are used for training, the GEIs of the other sequence will be suitable for view-insensitive recognition. In our approach, we generate GEI set $\{G_j\}$ from the original sequence at the interval of $1/4$ walking cycle, i.e., two adjacent GEIs have the overlap of $3/4$ cycles. Figure 3 shows the GEIs generated from sequences of the same person walking along different directions. It is shown that the first several GEIs in the first row are similar to the last several GEIs in the second row, and the first several GEIs in the second row are similar to the last several GEIs in the third row. This is due to the overlap of the walking view range among sequences with different walking directions.

With the GEI templates obtained from each person (one or two sequences for each person), a statistical feature extraction method is used for learning efficient features. Features learned from the GEI templates characterize view insensitive human walking properties provided in training sequences.

2.4. Statistical Feature Extraction

Once we obtain a set of training gait templates, we use principal component analysis (PCA) to reduce their dimensionality followed by multiple discriminant analysis (MDA) to achieve the best class separability [2].

PCA seeks a projection that best represents the data in a least square sense. Given the d-dimensional training template set $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{n_t}\}$, the d'-dimensional principal component vector $\mathbf{y}_k$ is obtained as:

$$\mathrm{y}_k = [a_1, ..., a_{d'}]^T = [\mathbf{e}_1, ..., \mathbf{e}_{d'}]^T \mathrm{x}_k, \quad k = 1, ..., n_t \tag{2}$$

where $\mathbf{e}_1, \mathbf{e}_2, ...,$ and $\mathbf{e}_{d'}$ are the d' eigenvectors of the scatter matrix S having the largest eigenvalues, and n_t is the number of the gait templates from all people in the training database.

MDA seeks a transformation matrix W that in some sense maximizes the ratio of the between-class scatter S_B to the within-class scatter S_W. The ratio is maximized when the

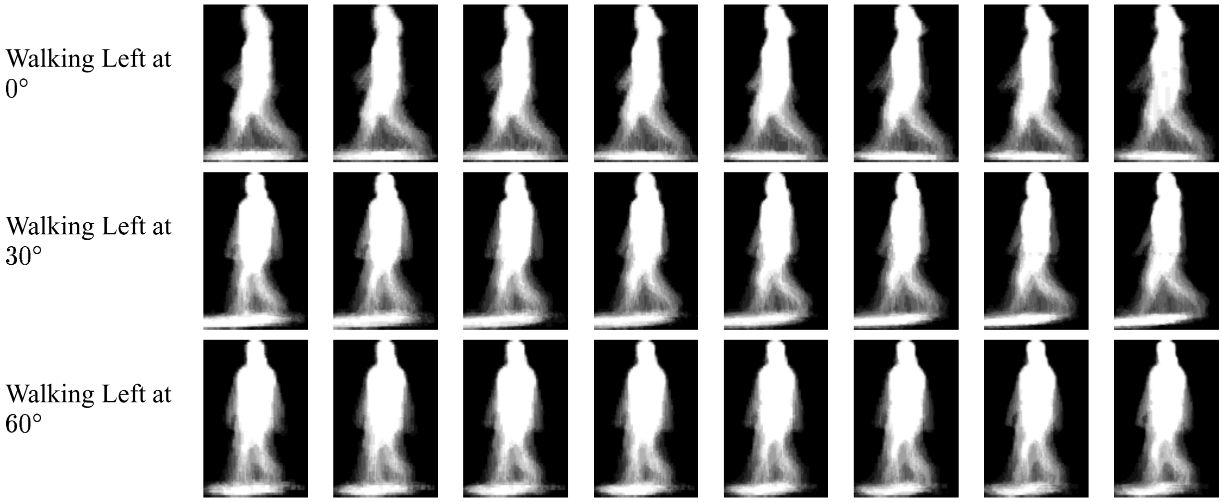

Fig. 3. GEIs generated from three sequences of the same person walking along different directions.

columns of W are the generalized eigenvectors that correspond to the largest eigenvalues in $S_B\mathbf{w}_i = \lambda_i S_W \mathbf{w}_i$. There are no more than $c-1$ nonzero eigenvalues, and the corresponding eigenvectors $\mathbf{v}_1,...,\mathbf{v}_{c-1}$ form transformation matrix. The $(c-1)$-dimensional multiple discriminant vector $\mathbf{z}_k$ is obtained as:

$$\mathrm{z}_k = [\mathbf{v}_1, ..., \mathbf{v}_{c-1}]^T \mathrm{y}_k, \quad k = 1, ..., n_t. \tag{3}$$

2.5. Human Identification

For each training gait template, its gait feature vector is obtained as: $\mathrm{z}_k = T\mathrm{x}_k$, $k = 1, ..., n_t$, where $T = [\mathbf{v}_1, ..., \mathbf{v}_{c-1}]^T [\mathbf{e}_1, ..., \mathbf{e}_{d'}]^T$ is the complete transformation matrix based on combined PCA and MDA analysis. The set $\{\mathbf{z}\}$ is composed of the feature database for individual recognition. The class centers for $\{\mathbf{z}\}$ are given as follows

$$\mathrm{m}_i = \frac{1}{n_i} \sum_{\mathbf{z} \in \mathcal{D}_i} \mathbf{z}, \quad i = 1, ..., c \tag{4}$$

where c is the number of classes (individuals) in the database, $\mathcal{D}_i$ is the set of feature vectors belonging to the ith class, and n_i is the number of feature vectors in $\mathcal{D}_i$. Assuming that feature vectors for each class have Gaussian distribution with the same covariance matrix $\Sigma = \sigma I$, Bayesian classifier becomes minimum Euclidean distance classifier which will be used in the following individual recognition.

Given a probe gait silhouette sequence P, we follow the procedure in Section 2.3 to generate templates $\{G_j\}$, $j = 1, \ldots, n$. The corresponding feature vector set is obtained as follows

$$\{\hat{\mathcal{D}}_P\}: \quad \hat{\mathrm{z}}_j = TG_j, \quad j = 1, ..., n. \tag{5}$$

Let

$$D(\hat{\mathcal{D}}_P, \mathcal{D}_i) = \frac{1}{m} \sum_{j=1}^{m} ||\hat{\mathrm{z}}_j - \mathrm{m}_i||, \quad i = 1, ..., c \tag{6}$$

We assign $P \in \omega_k$ if

$$D(\hat{\mathcal{D}}_P, \mathcal{D}_k) = \min_{i=1}^{c} D(\hat{\mathcal{D}}_P, \mathcal{D}_i). \tag{7}$$

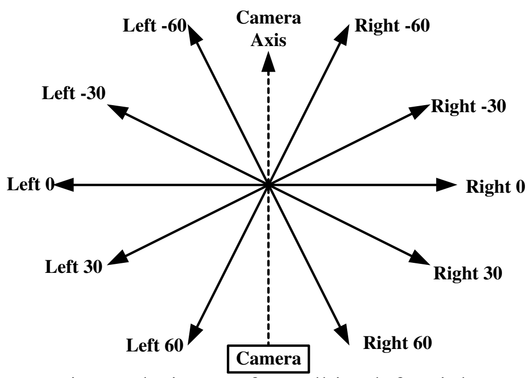

Fig. 4. Experiment design: Left - walking left; Right - walking right; positive angle - walking towards the camera; negative angle - walking away from the camera.

3. EXPERIMENTAL RESULTS

The video data used in our experiments are real human walking data recorded in outdoor environment. In these video data, there is only one walking person at the same time. Eight people walk along 10 different directions as shown in Figure 4. The beginning and end positions are different in the image sequences, and are not indicated in this figure. There is only one sequence per person per direction in the database. To obtain a uniform silhouette sequence representation for human walking left/right, we assume that people walk from right to left in the normalized silhouette sequence. If the person walks right in a sequence, we will reverse the silhouette along its vertical centroid axis, so that the person looks like walking left.

We first study the human recognition performance using sequences of one specific direction for training. In each experiment, we select sequences of one specific direction as the training data, and other sequences as the testing data. The experimental results are shown in Table 1. It can be seen that the average recognition performance is good. The performance of experiments using training data with $\pm 30°$ is better than those with $0°$ and $\pm 60°$ because the walking

Table 1. Recognition performance using sequences of one specific direction for training.

Training Data	Testing Data										
	L 0°	R 0°	L 30°	R −30°	R 30°	L −30°	L 60°	R −60°	R 60°	L −60°	Average
Left 0°		100%	100%	75%	75%	88%	63%	50%	50%	75%	75%
Right 0°	100%		88%	88%	88%	75%	38%	63%	50%	13%	67%
Sub-Average	100%		84%				50%				71%
Left 30°	88%	88%		88%	88%	100%	100%	50%	63%	75%	82%
Right −30°	75%	75%	75%		100%	88%	63%	88%	88%	50%	78%
Right 30°	75%	75%	88%	100%		88%	75%	88%	88%	63%	82%
Left −30°	75%	63%	100%	75%	88%		75%	63%	63%	75%	75%
Sub-Average	77%		90%				73%				79%
Left 60°	75%	38%	100%	63%	88%	88%		63%	63%	100%	75%
Right −60°	50%	50%	50%	88%	88%	75%	50%		88%	63%	67%
Right 60°	50%	63%	50%	88%	75%	50%	75%	88%		75%	68%
Left −60°	63%	50%	88%	63%	50%	88%	75%	63%	63%		67%
Sub-Average	55%		74%				72%				69%

Table 2. Recognition performance using sequences of two specific directions for training.

Train Data	L 0° R 0°	L 30° R −30°	R 30° L −30°	L 60° R −60°	R 60° L −60°
Perf.	72%	92%	89%	86%	87%

view overlap between the training data and testing data is larger in the former case. The recognition performance is not satisfactory in some extreme cases because the walking view difference between the training data and testing data is too large.

Some other interesting results are obtained from our experiments. If the person in the training sequence walks left (right), the average recognition rate for the testing sequences where the person walks left (right) is 82%, and that for the testing sequences where the person walks right (left) is 66%. This observation indicates that our method to unify people walking left/right introduces errors. On the other hand, if the person in the training sequence walks towards (away from) the camera, the average recognition rate for the testing sequences where the person walks towards (away from) the camera is 81%, and that for the testing sequences where the person walks away from (towards) the camera is 82%. These results do not include the sequences with walking direction of 0°. This observation indicates that walking towards or away from the camera has little effect on gait recognition.

We also study the human recognition performance using sequences of two specific directions for training. Experimental results in Table 2 shows that this combination dramatically improves the human recognition performance. Notice that the combination of left 0° and right 0° does not improve the performance because sequences with the two directions contain the same walking view ranges.

4. CONCLUSIONS

In this paper, we propose a statistical feature extraction approach for view-insensitive gait recognition based on the fact that the walking view (the angle between the walking direction and the camera view) ranges of human walking along different directions may be overlapping or adjacent if the difference of their walking directions is not too large. The proposed approach avoids the difficulties of recovering the human body structure and camera calibration. Experimental results show that the proposed approach achieves good performance in recognizing people walking along different directions from the training sequence with a given walking direction. It is also shown that the performance is further improved when more sequences with different walking directions are available for training.

5. REFERENCES

[1] J. Han and B. Bhanu. Statistical feature fusion for gait-based human recognition. in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2:842–847, 2004.

[2] P. Huang, C. Harris, and M. Nixon. Recognizing humans by gait via parametric canonical space. *Artificial Intelligence in Engineering*, 13:359–366, 1999.

[3] A. Kale, A. Chowdhury, and R. Chellappa. Towards a view invariant gait recognition algorithm. in *Proc. IEEE Conference on Advanced Video and Signal Based Surveillance*, pages 143–150, July 2003.

[4] J. Little and J. Boyd. Recognizing people by their gait: the shape of motion. *Videre: Journal of Computer Vision Research*, 1(2):1–32, 1998.

[5] P. Phillips, S. Sarkar, I. Robledo, P. Grother, and K. Bowyer. The gait identification challenge problem: data sets and baseline algorithm. in *Proc. International Conference on Pattern Recognition*, 1:385–388, 2002.

[6] G. Shakhnarovich, L. Lee, and T. Darrell. Integrated face and gait recognition from multiple views. in *Proc. IEEE Workshop on Visual Motion*, 1:439–446, 2001.

[7] D. Tolliver and R. Collins. Gait shape estimation for identification. in *Proc. 4th International Conference on Audio- and Video-Based Biometric Person Authentication, LNCS 2688*, pages 734–742, 2003.

[8] L. Wang, T. Tan, W. Hu, and H. Ning. Automatic gait recognition based on statistical shape analysis. *IEEE Transactions on Image Processing*, 12(9):1120–1131, 2003.