GAIT RECOGNITION USING MULTIPLE VIEWS

Xiaxi Huang and Nikolaos V. Boulgouris

Department of Electronic Engineering
King's College London
United Kingdom

ABSTRACT

This paper proposes a gait recognition system in which several views are available. It is shown that each view has unequal discrimination power and, therefore, should have unequal contribution in the recognition process. In order to exploit the availability of multiple views, several methods for the combination of the results that are obtained from the individual views are tested and evaluated. A novel approach for the combination of the results from several views is also proposed based on the relative importance of each view. The proposed approach generates improved results, compared to those obtained by using individual views or by using multiple views that are combined using other combination methods.

Keywords: gait, surveillance, security

1. INTRODUCTION

Gait recognition [1] aims at the identification of individuals based on their walking style. 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, gait can be captured from a distant camera, without drawing the attention of the observed subject.

Although several approaches have been presented for the recognition of human gait, most of them limit their attention to the case in which only the side-view is available since this viewing angle is considered to provide the richest information of the gait of the waking person [2, 3, 4]. In [5], an experiment was carried out using two views, namely the frontal-parallel view and the side-view, from which the silhouettes of the subjects in two walking stances were extracted. This approach exhibited a higher recognition accuracy for the frontal-parallel view than that of the side-view. The side-view was also examined in [6] together with another view from a different angle, and the static parameters, such as the height of the walking person, as well as distances between body parts, were used in the template matching. Apart from the recognition rate, the results were also reported based on a small sample set using a confusion metric which reflects the effectiveness of the approach in the situation of a large population of subjects. The authors in [7] synthesize the side-view silhouettes from those captured by multiple cameras employing visual hull techniques, while in [8], the approach taken relied on the perspective projection and optical flow based structure of motion approach was taken instead. In [9], information from multiple cameras is gathered to construct a 3D gait model.

In this paper, we use the Motion of Body (MoBo) database from the Carnegie Mellon University (CMU) in order to investigate the contribution of each viewing direction to the recognition performance of a gait recognition system. In general, we try to answer the fundamental question: *if several views are available to a gait recognition system, what is the most appropriate way to combine them in order to enhance the performance and the reliability of the system?* We provide a detailed analysis of the role and the contribution of each viewing direction by reporting recognition results of systems based on each one of the available views. We also propose a novel way to combine the results obtained from independent views. In the proposed approach, we set a weight for each view, based on its importance as it is calculated using the statistical processing of the differences between views. The experimental results demonstrate the superior performance of the proposed weighted combination approach in comparison to the single-view approach and other combination methods for multiple views.

2. GAIT RECOGNITION USING MULTIPLE VIEWS

The CMU Motion of Body (MoBo) database does not contain explicitly the reference set and the test sets as in [2]. Therefore, we chose to use the "fast walk" sequences as the reference set, and the "slow walk" sequences as the test set. As mentioned in the introduction, our goal is to find out which viewing directions have the most significant contribution in a multiview gait recognition system. To this end, we adopt a simple and straightforward way in order to determine the similarity between gait sequences in the reference and test databases. Initially, from each gait sequence, taken from a specific viewpoint, we construct a simple template T by averaging all frames in the sequence ([10, 11])

$$T = \frac{1}{N_T} \sum_{a=1}^{N_T} t_a \quad (1)$$

where t_a, $a = 1, ..., N_T$ are the silhouettes in a gait sequence and N_T is the number of silhouettes.

Let T_i, R_i denote the templates corresponding to the ith and the jth subjects in the test database and the reference database respectively. Their distance is calculated using the following distance metric

$$d\left(T_i, R_j\right) = \|T_i - R_i\| = \|\frac{1}{N_{T_i}} \sum_{\alpha=1}^{N_{T_i}} t_{i\alpha} - \frac{1}{N_{R_j}} \sum_{\beta=1}^{N_{R_j}} r_{j\beta}\| \quad (2)$$

where $\|\cdot\|$ is the l^2-norm and $t_{i\alpha}$, $r_{j\beta}$ are the silhouettes belonging to the ith test subject and jth reference subject respectively. The associated frame indices α and β run from 1 to the total number of silhouettes in a sequence (N_{T_i} and N_{R_j} respectively). Essentially, a template is constructed for each subject by averaging all possible silhouettes, and then the Euclidean distance between two templates is calculated and used as their dissimilarity measure. This means that a smaller template distance corresponds to a closer match between two compared subjects.

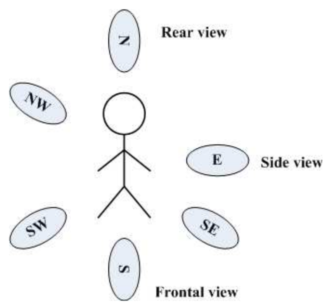
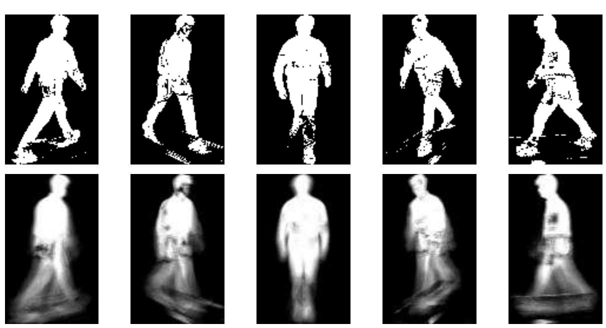


Fig. 1. Camera arrangement in the CMU Motion of Body (MoBo) database. Six cameras are oriented clockwise in the east, southeast, south, southwest, northwest, north, with the walking subject facing toward the south.

In order to evaluate the contribution of various viewing directions in the human gait recognition, we choose the Motion of Body (MoBo) database [12] from the Carnegie Mellon University (CMU) which contains walking subjects captured from six cameras located in positions as shown in Fig. 1. The database consists of walking sequences of 23 male plus 2 female subjects, who were recorded performing four kinds of activities, i.e., fast walk, slow walk, etc. We first take bounding boxes of silhouettes for the subjects, then align and normalize all silhouettes into uniform dimensions, i.e., 128 pixels tall and 80 pixels wide, in order to eliminate height variations among the walking subjects in the experiment. We use five (see Fig. 2) out of the available six viewing directions omitting the north view, since it is practically identical to the south view, i.e., the frontal view.



NW SW S SE E

Fig. 2. Available views for multiview gait recognition (upper row) and templates constructed using the five available views (bottom row).

3. COMBINATION OF DIFFERENT VIEWS USING A SINGLE DISTANCE METRIC

In this section, we propose a novel method for the combination of results from different views, in order to improve the performance of a gait recognition system based on a single view. In our approach, we use weights in order to reflect the importance of each view during the combination. This means that instead of using a single distance for the evaluation of similarity between walking persons i and j, we use multiple distances between the respective views and combine them in a total distance which is given by

$$D\left(T_i, R_j\right) = \sum_{v=1}^{V} w_v d_v\left(T_i, R_j\right) \tag{3}$$

where V is the total number of available views. Therefore, our task is to determine the weights w_v, which yield smaller total distance when $i = j$, and larger when $i \neq j$

Suppose that $d_{fv},\ v = 1, 2, ..., V$ are random variables representing the distances between a test subject and its corresponding reference subjects (i.e., "within class" distance), and $d_{bv},\ v = 1, 2, ...V$ are random variables representing the distances between a test subject and a reference subject other than its corresponding subject (i.e., "between class" distance).

In general, in order to maximize the efficiency of our system, we would like the weighed distance D_f between corresponding subjects in the reference and test databases

$$D_f = \sum_{v=1}^{V} w_v d_{fv} = \mathbf{w}^T \cdot \mathbf{d}_f \tag{4}$$

to be smaller than the weighed distance between the non-corresponding subjects

$$D_b = \sum_{v=1}^{V} w_v d_{bv} = \mathbf{w}^T \cdot \mathbf{d}_b \tag{5}$$

A recognition error takes place whenever $D_b < D_f$. Therefore, the probability of error is

$$P_e = P\left(D_b < D_f\right) = P\left(\mathbf{w}^T \cdot \left(\mathbf{d}_b - \mathbf{d}_f\right) < 0\right) \tag{6}$$

We define the random variable z as

$$z = \mathbf{w}^T \cdot \left(\mathbf{d}_b - \mathbf{d}_f\right) \tag{7}$$

if we assume that $\mathbf{d}_b$ and $\mathbf{d}_f$ are normal random vectors, then z is a normal random variable with Probability Density Distribution

$$P\left(z\right) = \frac{1}{\sqrt{2\pi}\sigma_z} e^{-\frac{1}{2}\frac{\left(z-m_z\right)^2}{\sigma_z^2}} \tag{8}$$

where m_z is the mean value of z, σ_z is the variance of z.

Therefore, using (7) and (8), the probability of error in (6) is expressed as

$$P_e = P\left(z < 0\right) = \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}\sigma_z} e^{-\frac{1}{2}\frac{\left(z-m_z\right)^2}{\sigma_z^2}} dz \tag{9}$$

Furthermore, if $q = \frac{z-m_z}{\sigma_z}$, then the above expression is equivalent to

$$P_e = \int_{-\infty}^{-\frac{m_z}{\sigma_z}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}q^2} dq \tag{10}$$

The probability of error can therefore be minimized by minimizing $-\frac{m_z}{\sigma_z}$, or equivalently by maximizing $\frac{m_z}{\sigma_z}$. To this end, we have to calculate m_z and σ_z. If $E\{\cdot\}$ denotes statistical expectation, then the mean value of z is

$$\begin{aligned} m_z &= E\{z\} = E\{\mathbf{w}^T\left(\mathbf{d}_b - \mathbf{d}_f\right)\} \\ &= \mathbf{w}^T\left(E\{\mathbf{d}_b\} - E\{\mathbf{d}_f\}\right) = \mathbf{w}^T\left(\mathbf{m}_{d_b} - \mathbf{m}_{d_f}\right) \end{aligned} \tag{11}$$

where $\mathbf{m}_{d_b}$ and $\mathbf{m}_{d_f}$ are the mean vectors of $\mathbf{d}_b$ and $\mathbf{d}_f$. The variance of z is

$$\begin{aligned}\sigma_z^2 &= E\{(z-m_z)^2\} \\ &= E\{\left(\mathbf{w}^T(\mathbf{d}_b-\mathbf{d}_f)-\mathbf{w}^T\left(\mathbf{m}_{d_b}-\mathbf{m}_{d_f}\right)\right)^2\} \\ &= E\{\mathbf{w}^T(\mathbf{d}_b-\mathbf{m}_{d_b})(\mathbf{d}_b-\mathbf{m}_{d_b})^T\mathbf{w} \\ &\quad -\mathbf{w}^T(\mathbf{d}_b-\mathbf{m}_{d_b})\left(\mathbf{d}_f-\mathbf{m}_{d_f}\right)^T\mathbf{w} \\ &\quad -\mathbf{w}^T\left(\mathbf{d}_f-\mathbf{m}_{d_f}\right)(\mathbf{d}_b-\mathbf{m}_{d_b})^T\mathbf{w} \\ &\quad +\mathbf{w}^T\left(\mathbf{d}_f-\mathbf{m}_{d_f}\right)\left(\mathbf{d}_f-\mathbf{m}_{d_f}\right)^T\mathbf{w}\} \qquad (12)\end{aligned}$$

If we assume that $\mathbf{d}_b$ and $\mathbf{d}_f$ are independent, then

$$\begin{aligned}\sigma_z^2 &= \mathbf{w}^T\cdot E\{(\mathbf{d}_b-\mathbf{m}_{d_b})(\mathbf{d}_b-\mathbf{m}_{d_b})^T\}\cdot\mathbf{w} \\ &\quad +\mathbf{w}^T\cdot E\{\left(\mathbf{d}_f-\mathbf{m}_{d_f}\right)\left(\mathbf{d}_f-\mathbf{m}_{d_f}\right)^T\}\cdot\mathbf{w} \\ &= \mathbf{w}^T\cdot\Sigma_{d_b}\cdot\mathbf{w}+\mathbf{w}^T\cdot\Sigma_{d_f}\cdot\mathbf{w} \qquad (13)\end{aligned}$$

Therefore, the optimization problem becomes equivalent to maximizing

$$\begin{aligned}\frac{m_z^2}{\sigma_z^2} &= \frac{\mathbf{w}^T\cdot\left(\mathbf{m}_{d_b}-\mathbf{m}_{d_f}\right)\cdot\left(\mathbf{m}_{d_b}-\mathbf{m}_{d_f}\right)^T\cdot\mathbf{w}}{\mathbf{w}^T\cdot\Sigma_{d_b}\cdot\mathbf{w}+\mathbf{w}^T\cdot\Sigma_{d_f}\cdot\mathbf{w}} \\ &= \frac{\mathbf{w}^T\cdot\Sigma_{d_c}\cdot\mathbf{w}}{\mathbf{w}^T\cdot\left(\Sigma_{d_b}+\Sigma_{d_f}\right)\cdot\mathbf{w}} \qquad (14)\end{aligned}$$

where

$$\Sigma_{d_c}=\left(\mathbf{m}_{d_b}-\mathbf{m}_{d_f}\right)\cdot\left(\mathbf{m}_{d_b}-\mathbf{m}_{d_f}\right)^T \qquad (15)$$

The maximization of the above quality is reminiscent of the optimization problem that appears in two-class linear discriminant analysis. Trivially, the ratio can be maximized by determining a vector $\mathbf{w}$ that satisfies [13]

$$\Sigma_{d_c}\cdot\mathbf{w}=\Lambda\left(\Sigma_{d_b}+\Sigma_{d_f}\right)\cdot\mathbf{w} \qquad (16)$$

for some Λ. In the case that we are considering, the optimal $\mathbf{w}$ is given by

$$\mathbf{w}=\left(\Sigma_{d_b}+\Sigma_{d_f}\right)^{-1}\cdot\left(\mathbf{m}_{d_b}-\mathbf{m}_{d_f}\right) \qquad (17)$$

If we assume that the distances corresponding to different views are independent, then

$$\left(\Sigma_{d_b}+\Sigma_{d_f}\right)^{-1}= \left(\begin{array}{cccc}\frac{1}{\sigma^2_{d_{b1}}+\sigma^2_{d_{f1}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma^2_{d_{b2}}+\sigma^2_{d_{f2}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma^2_{d_{bV}}+\sigma^2_{d_{fV}}}\end{array}\right) \qquad (18)$$

where V is the total number of available views. Therefore, the optimal weight vector is

$$\mathbf{w}=\left(\frac{m_{d_{b1}}-m_{d_{f1}}}{\sigma^2_{d_{b1}}+\sigma^2_{d_{f1}}}\;\frac{m_{d_{b2}}-m_{d_{f2}}}{\sigma^2_{d_{b2}}+\sigma^2_{d_{f2}}}\cdots\frac{m_{d_{bV}}-m_{d_{fV}}}{\sigma^2_{d_{bV}}+\sigma^2_{d_{fV}}}\right)^T \qquad (19)$$

Of course, the practical application of the above theory requires the availability of database (other than the test database) which will be used in conjunction with the reference database for the determination of $m_{d_b}, m_{d_f}, \sigma_{d_b}, \sigma_{d_f}$. In our experiments we used the CMU database of individuals walking with a ball for this purpose.

In the sequel, we will use the weight vector in (19) for the combination of views in multiview gait recognition.

4. EXPERIMENTAL RESULTS

For the experimental evaluation of our methods, we used the Motion of Body (MoBo) database from the Carnegie Mellon University (CMU). The CMU database has 25 subjects walking on a treadmill. This is an artificial setting that might affect the results. However, using this database was essentially our only option since this is the only database that provides several views. We used the "fast walk" sequences as reference and the "slow walk" as test sequences. We also used the "with a ball" sequences in conjunction with the reference sequences for the determination of the weights in (19). The comparisons of recognition performance are based on Cumulative Match Scores at rank 1 and rank 5. Rank 1 results report the percentage of subjects in a test set that were identified exactly. Rank 5 results report the percentage of test subjects whose actual match in the reference database was in the top 5 matches. In this section, we present the results generated by the proposed view combination method. These results are compared to the results obtained using independent views and other combination methods.

Initially, we tried several simple methods for the combination of the results obtained using the available views. Specifically, the total distance between two subjects was taken to be equal to the *mean*, *max*, *min*, *median*, and *product rule* of the distances corresponding to each of the five viewing directions. Such combination approaches were originally explored in [14]. As shown in Figure 3 and Table 1, among all the above combination methods, the most satisfactory results were obtained by using the *Product* and *Min* combination rules.

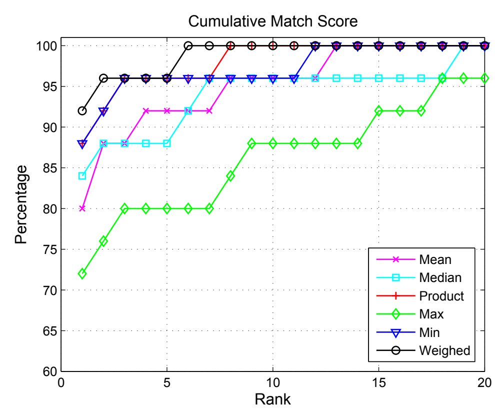


Fig. 3. Cumulative Match Score for the proposed and the other five combination methods

Subsequently, we applied the proposed methodology for the determination of the weights in Eq. (3). Based on Eq. (19), the weights for the combination of the distances of the available views were calculated and are tabulated in Table 2. As seen, the most suitable views seem to be the frontal (east) and the side (south) views since these views are given the heavier weights.

The above conclusion is experimentally verified by studying the recognition performance that corresponds to each of the views independently. The Cumulative Match Scores and the recognition rates that are achieved using each view as well as those achieved by the proposed method are shown in Table 1. As it can be seen, the south and the east views have the highest recognition rates, as well as the highest weights, which means that the weights calculated by the proposed method correctly reflect the importance of the views. The re-

Single- / Combined-views Method	Rank 1(%)	Rank 5(%)
East	84	92
Southeast	64	76
South	88	96
Northwest	76	92
Southwest	72	76
Mean	80	92
Median	84	88
Product rule	88	96
Max	72	80
Min	88	96
Weighed (Proposed)	**92**	**96**

Table 1. The recognition rates for the single-view and combined-views methods.

View	E	SE	S	NW	S
Weight	**0.3332**	0.0603	**0.4036**	0.1188	0.0842

Table 2. The weights calculated by the proposed method.

sults obtained by the proposed combination method are generally better than those obtained from single views.

The proposed system was also evaluated in terms of verification performance. In an access control scenario, this means calculating the probability of positive recognition of an authorized subject versus the probability of granting access to an unauthorized subject. In Table 3, verification results are presented at 5%, 10% and 20% false alarm rate for the proposed method and the existing methods. As seen, within the five viewing directions, the frontal (east) and side (south) views have the best performances; and among the five existing combination methods, the *Min* method obtains the best results. As expected, the proposed method has improved verification performance, in comparison to any of the single-view methods as well as in comparison to the other methods for multiview gait recognition.

Single-view / Combined-views Method	Verification Rate (%)		
	FAR 5%	FAR 10%	FAR 20%
East	88	96	96
Southeast	68	72	76
South	92	96	100
Northwest	80	92	92
Southwest	76	76	84
Mean	88	92	96
Median	92	94	96
Product Rule	92	96	96
Max	72	76	84
Min	92	96	100
Weighed (Proposed)	**96**	**100**	**100**

Table 3. The verification rates for the single-view and combined-views methods.

5. CONCLUSION

In this paper, we investigated the availability of multiple views in the gait recognition task using the Motion of Body (MoBo) database from the Carnegie Mellon University (CMU). We showed that each view has unequal discrimination power and therefore has unequal contribution to the task of gait recognition. A novel approach was proposed for the combination of the results from different views into a common distance metric for the evaluation of similarity between gait sequences. By using the proposed method, which uses different weights in order to exploit the different importance of the views, improved recognition performance was achieved in comparison to the results obtained from individual views or by using other combination methods.

6. 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] S. Sarkar, P.J. Phillips, Z. Liu, I.R. Vega, P. Grother, and K.W. Bowyer, "The humanid gait challenge problem: data sets, performance, and analysis," *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 27, no. 2, pp. 162–176, Feb. 2005.

[3] N.V. Boulgouris, K.N. Plataniotis, and D. Hatzinakos, "Gait recognition using linear time normalization," *Pattern Recognition*, vol. 39, pp. 969–979, May 2006.

[4] M. Ekinci, "Gait recognition using multiple projections," in *Proc. Int. Conf. Automatic Face and Gesture Recognition*, Apr. 2006, pp. 517–522.

[5] R.T. Collins, R. Gross, and J. Shi, "Silhouette-based human identification from body shape and gait," in *Proc. Int. Conf. Automatic Face and Gesture Recognition*, Washington, DC, May 2002, pp. 351–356.

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

[7] G. Shakhnarovich, L. Lee, and T. Darrell, "Integrated face and gait recognition from multiple views," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Kauai, HI, Dec. 2001, vol. I, pp. 439–476.

[8] A. Kale, A. Chowdhury, and R. Chellappa, "Towards a view invariant gait recognition algorithm," in *Proc. IEEE Conf. on Advanced Video and Signal Based Surveillance*, 2003.

[9] G. Zhao, G. Liu, H. Li, and M. Pietikainen, "3d gait recognition using multiple cameras," in *Proc. Int. Conf. Automatic Face and Gesture Recognition*, Apr. 2006, pp. 529–534.

[10] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 28, no. 2, pp. 316–322, 2006.

[11] Z. Liu and S. Sarkar, "Simplest representation yet for gait recognition: Averaged silhouette," in *Int. Conf. on Pattern Recognition*, Aug. 2004.

[12] R. Gross and J. Shi, "The cmu motion of body (mobo) database," Tech. Rep. CMU-RI-TR-01-18, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, Jun. 2001.

[13] R.O. Duda, P.E. Hart, and D.G. Stork., *Pattern Classification*, John Wiley & Sons, Inc., 2001.

[14] J. Kittler, R.P.W. Duin M. Hatef, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.