

Adaptive Fusion of Gait and Face for Human Identification in Video

Xin Geng¹, Liang Wang², Ming Li¹, Qiang Wu³, and Kate Smith-Miles¹

¹School of Engineering and Information Technology,
Deakin University, Victoria 3125, Australia

{xge, ming, katesm}@deakin.edu.au

²Department of Computer Science and Software Engineering,
The University of Melbourne, VIC 3010, Australia

lwawang@csse.unimelb.edu.au

³Faculty of Information Technology,
University of Technology, Sydney, NSW 2007, Australia

wuq@it.uts.edu.au

Abstract

Most work on multi-biometric fusion is based on static fusion rules which cannot respond to the changes of the environment and the individual users. This paper proposes adaptive multi-biometric fusion, which dynamically adjusts the fusion rules to suit the real-time external conditions. As a typical example, the adaptive fusion of gait and face in video is studied. Two factors that may affect the relationship between gait and face in the fusion are considered, i.e., the view angle and the subject-to-camera distance. Together they determine the way gait and face are fused at an arbitrary time. Experimental results show that the adaptive fusion performs significantly better than not only single biometric traits, but also those widely adopted static fusion rules including SUM, PRODUCT, MIN, and MAX.

1. Introduction

Biometrics is a fast developing field for human recognition based upon one or more intrinsic physical or behavioral traits. Typical biometric traits include fingerprint, hand-geometry, face, iris, gait, ear, voice, etc. Single biometric systems are often affected by practical problems like noisy sensor data, non-universality and/or lack of distinctiveness of the biometric trait, unacceptable error rates, and spoof attacks [3]. To solve these problems, multi-biometric fusion is proposed by combining biometric traits from different sources [3]. Up to the present, most work on multimodal biometrics is based on *static fusion*, i.e., the fusion rules are predefined and remain fixed when the system is run-

ning. The most prominent problem of static fusion is that it cannot adapt to the changing environment and individual users. For human recognition in realistic environments, this paper proposes *adaptive multi-biometric fusion*, which can dynamically adapt the fusion rules to the external conditions.

As a typical example of adaptive multi-biometric fusion, the adaptive fusion of gait and face is investigated in this paper. Both gait and face are unobtrusive biometrics and can be simultaneously obtained by most video surveillance systems. There are some previous works on fusion of gait and face. For example, Shakhnarovich and Darrell [5] proposed to combine virtual gait and face cues generated by a 3D model derived from multiple camera views. Kale *et al.* [4] proposed the fusion of gait and face for a special ‘inverted Σ ’ walking pattern. Zhou and Bhanu [8] proposed a method to improve the side-view gait recognition by using the enhanced side-view face image generated from the video. Table 1 summarizes the main differences between this paper and the previous works. As can be seen, the fusion rules adopted by the previous works are all among the four static rules: SUM, PRODUCT, MIN, and MAX, i.e., combine the gait score and face score through the operators *sum*, *product*, *min* and *max*, respectively. Obviously none of them can respond to the changes of the external factors.

The most significant ‘external factors’ that might affect the relationship between gait and face in the fusion are the view angle and the distance from the subject to the camera. Fig. 1 gives the five representative walking patterns with suggested fusion schemes. Generally speaking, the performance of gait recognition is mainly affected by the view angle and is not sensitive to the distance. Usually the side

Table 1. Differences Between This Paper and Previous Works on Gait and Face Fusion

Work	Biometrics	No. of Cameras	Fusion Rules
Shakhnarovich and Darrell [5]	Virtual frontal face and side gait from a 3D model	4	SUM, PRODUCT, MIN, MAX
Kale <i>et al.</i> [4]	Frontal face and ‘inverted Σ ’ gait	1	SUM, PRODUCT
Zhou and Bhanu [8]	Side face and Side Gait	1	SUM, PRODUCT, MAX
This paper	Face and gait in 5 view angles	1	Adaptive fusion

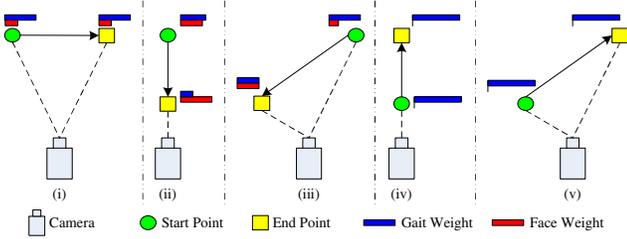


Figure 1. Adaptive fusion of gait and face in the five representative walking patterns.

view (i) will get the best result because more motion characteristics can be captured from this angle. As such, the oblique view (iii, v) is worse and the frontal/back view (ii, iv) is the worst. On the other hand, face recognition is affected by the view angle as well as image resolution. In contrast to gait, the frontal view (ii) is the best angle for face, the oblique frontal view (iii) is the next, the side view (i) is worse, the back view (iv) and the oblique back view (v) cannot be recognized at all. Moreover, the closer the face to the camera, the higher the resolution, and the more accurate the recognition. Accordingly, the weights of gait and face in the adaptive fusion are adjusted in real-time.

The rest of this paper is organized as follows. The framework of adaptive multi-biometric fusion is proposed in Section 2. Then the adaptive fusion of gait and face is investigated in Section 3. Experiments are reported in Section 4. Finally conclusions are drawn in Section 5.

2. Fusion Adaptable to the Context

What any biometric system confronts is an ever changing world. The reliability of the biometric models is affected by the variation of certain external conditions. The variability comes from the following two aspects:

- *User-induced variability.* The performance of biometric recognition is highly variable from user to user. Individual users differ in biometric distinctiveness, behavior habit, familiarity of the biometric system, and attitude toward the system. Typical user-induced vari-

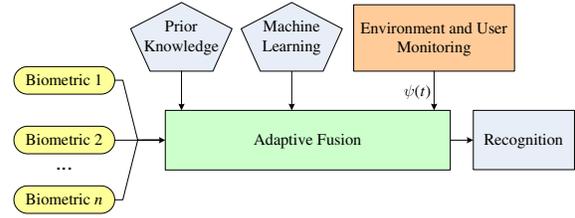


Figure 2. Framework of adaptive multi-biometric fusion.

able factors include intrinsic distinctiveness, pose, distance, expression, emotion, cooperativeness, *etc.*

- *Environment-induced variability.* The performance of biometric recognition can be also seriously affected by the environment where the system is working. For example, the lighting condition may affect all image/video based biometric traits, the background noise can affect the speech recognition, the extreme weather conditions, such as extremely high temperature and humidity, might make certain biometric sensors not work well, *etc.*

In order to deal with the variability, the multi-biometric system should be able to automatically adapt itself to the external changes. The framework of the adaptive multi-biometric fusion is shown in Fig. 2. The ‘Environment and User Monitoring’ component outputs a perceptual signal $\psi(t)$ indicating the changes of the outside world at time t . $\psi(t)$ might come from two different sources. The first is the additional sensors, such as the devices to detect the atmosphere temperature, humidity, *etc.* The second is the biometric data themselves, such as the brightness and contrast of the images, the emotion and pose of the user, *etc.* One of the most challenging issues is to determine which variation of the environment and the user might trigger the self-adjustment of the system and in which way the system can adapt itself to the variation. These problems could be solved through machine learning or by introducing prior knowledge from various areas, such as electronics, photography, psychology, *etc.* The learned/prior knowledge and the perceptual signal $\psi(t)$ together determine how the ‘Adaptive

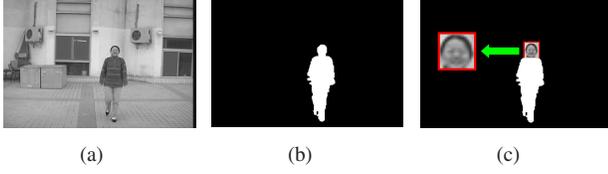


Figure 3. Silhouette and face extraction: (a) the original image, (b) the extracted silhouette, (c) extraction of the face image.

Fusion' component adjusts the system at time t . The adjustment might happen at two different levels. First, the fusion rule itself might be changed, such as drifting among the fusion rules of SUM, PRODUCT, MIN, and MAX. Second, in a particular fusion rule, the relationship of different biometric traits might be changed, such as adjusting the weight assigned to each biometric trait.

3. Adaptive Fusion of Gait and Face

As a typical example of adaptive multi-biometric fusion, the adaptive fusion of gait and face in video is studied in this section. The fusion happens at the score level, *i.e.* first obtain the scores from the *gait classifier* and the *face classifier* separately, then combine the two scores into one based on adaptive combination weights. In this case, as mentioned in Section 1, the fusion rule is adjusted according to the view angle and the distance from the subject to the camera.

3.1. Gait Classifier

Human motion can be regarded as temporal variation of human silhouettes. The gait classifier used in this paper is based on the silhouette images [6]. Assume the background to be steady, then the silhouette images can be generated through training a Gaussian model for each background pixel over a short period and comparing the background pixel probability to that of a uniform foreground model. One example of the silhouette image is shown in Fig. 3(b). After that, each silhouette image is centralized and normalized to the same size, and LPP (Locality Preserving Projection) [2] is used to get the corresponding low-dimensional embedding.

In detail, given the training data $\mathbf{G} = [\mathbf{x}_1; \mathbf{x}_2; \dots; \mathbf{x}_n]$, where \mathbf{x}_i is a vectorized silhouette image. Assume \mathbf{G} to be a graph with n nodes, an edge will connect nodes i and j if \mathbf{x}_i and \mathbf{x}_j are close. Here 'close' is defined by the K -nearest neighbors. The symmetric $n \times n$ edge matrix \mathbf{E} can be obtained with $g_{ij} = 1$ indicating an edge between nodes i and j , and $g_{ij} = 0$ otherwise. Then the transform matrix $\mathbf{W}_g = [\mathbf{w}_1 \ \mathbf{w}_2 \ \dots \ \mathbf{w}_l]$ can be calculated by solving the generalized eigenvector problem

$$\mathbf{GLG}^T \mathbf{w} = \lambda \mathbf{GBG}^T \mathbf{w}, \quad (1)$$

where \mathbf{B} is a diagonal matrix whose entries are column (or

row) sums of \mathbf{E} , $\mathbf{L} = \mathbf{B} - \mathbf{E}$ is the Laplacian matrix. The \mathbf{w}_i 's in \mathbf{W}_g are the solutions of Equation (1) sorted by the corresponding eigenvalues. l is the dimensionality of the subspace. Finally, the projection of a video \mathbf{X} is calculated by $\mathbf{Y} = \mathbf{XW}_g$. Suppose the gallery gait video of person j is \mathbf{X}_j , the probe gait video is \mathbf{X} , each row ($\mathbf{X}_j(i)$ or $\mathbf{X}(i)$) stores one frame. The mean Hausdorff distance d_H is used to measure the similarity between them, then the gait score $G(\mathbf{X}, j) = -d_H(\mathbf{XW}_g, \mathbf{X}_j\mathbf{W}_g)$, where

$$d_H(\mathbf{XW}_g, \mathbf{X}_j\mathbf{W}_g) = \Delta(\mathbf{XW}_g, \mathbf{X}_j\mathbf{W}_g) + \Delta(\mathbf{X}_j\mathbf{W}_g, \mathbf{XW}_g), \quad (2)$$

$$\Delta(\mathbf{XW}_g, \mathbf{X}_j\mathbf{W}_g) = \text{mean}_i(\min_j \|\mathbf{X}(i)\mathbf{W}_g - \mathbf{X}_j(j)\mathbf{W}_g\|). \quad (3)$$

3.2. Face Classifier

The first step of face recognition is face detection. Since the subject silhouette has already been extracted, face detection can be greatly simplified. Through some empirical experiments, the upper 1/6 of the silhouette is chosen as the face region. One example of the face image extraction is shown in Fig. 3(c). The extracted face images are first histogram equalized and then transformed into a vector of unit length to reduce the variation of illumination.

The face recognition algorithm used in this approach is Fisherface [1], which tries to find a feature space that maximizes the ratio of the inter-personal difference and the intra-personal difference by applying Fisher's Linear Discriminant (FLD). In detail, suppose Ω_B is the between-class scatter matrix and Ω_W is the within-class scatter matrix, then the projection matrix $\mathbf{W}_f = [\mathbf{w}_1 \ \mathbf{w}_2 \ \dots \ \mathbf{w}_q]$ can be calculated by solving a generalized eigenvalue problem

$$\Omega_B \mathbf{w} = \lambda_i \Omega_W \mathbf{w}. \quad (4)$$

There are at most $c - 1$ nonzero eigenvalues, where c is the number of classes. So in this paper, q is set to $c - 1$. Suppose each video is represented as a matrix, each row stores the normalized face vector in one frame, the gallery video of person j is \mathbf{X}_j , the probe video is \mathbf{X} , then the face score of \mathbf{X} to person j is

$$F(\mathbf{X}, j) = -d_H(\mathbf{XW}_f, \mathbf{X}_j\mathbf{W}_f), \quad (5)$$

where d_H is the Hausdorff distance defined in Equation (2).

3.3. Fusion Adaptable to View Angle and Distance

As mentioned in Section 1, the main 'external factors' considered in the adaptive fusion of gait and face are the view angle and the distance from the subject to the camera. The way they affect the fusion depends on the walking patterns, which are illustrated in Fig. 1. Thus the first step of the adaptive fusion should be the judgement of the current

Table 2. Silhouette Analysis to Determine the Walking Patterns

Pattern	Conditions
(i)	$ h_e - h_s \leq \rho_1 f_h$
(ii)	$h_e - h_s > \rho_1 f_h, c_s - c_e \leq \rho_2 f_w$
(iii)	$h_e - h_s > \rho_1 f_h, c_s - c_e > \rho_2 f_w$
(iv)	$h_e - h_s < -\rho_1 f_h, c_s - c_e \leq \rho_2 f_w$
(v)	$h_e - h_s < -\rho_1 f_h, c_s - c_e > \rho_2 f_w$

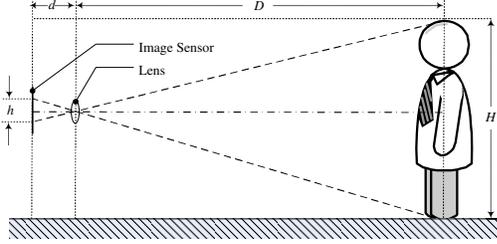


Figure 4. Estimation of the subject-to-camera distance.

walking pattern, which can be done through analyzing the silhouette images.

In the video clip of a walking person, suppose the first frame is s , the last is e , the width and height of each frame are f_w and f_h , the height of the silhouette in s is h_s , that in e is h_e , the horizontal position of the silhouette centroid in s is c_s , and that in e is c_e . Then the five walking patterns shown in Fig. 1 (i-v) can be determined through Table 2, where ρ_1 and ρ_2 are ratio threshold parameters. Note that each walking pattern corresponds to one view angle. There are more sophisticated ways to detect the view angle, but the simple rules in Table 2 are enough to distinguish the five walking patterns and can work very fast to ensure real-time response.

Another factor affecting the adaptive fusion is the distance from the subject to the camera. The distance can be roughly estimated for each frame in the video as shown in Fig. 4. Suppose the actual height of the subject is H , the height of the silhouette in the image is h , the distance from the subject to the camera lens is D , and the focal length of the lens is d , then

$$D = Hd/h = \alpha/h, \quad (6)$$

where $\alpha = Hd$. Without knowing α , it seems that D cannot be actually calculated. Fortunately, α can be removed as a common factor from the numerator and denominator when calculating the fusion weights (Equation (9)). Thus without loss of generality, $D = 1/h$.

Through the above analysis, the view angle and the subject-to-camera distance at any time t can be estimated. This corresponds to the output $\psi(t)$ of the ‘Environment and User Monitoring’ component in Fig. 2. Here $\psi(t)$ is generated from the biometric data themselves, and no addi-

tional sensors are required. The ‘Prior Knowledge’ that determines how the fusion component adapts itself to $\psi(t)$ is illustrated in Fig. 1. Basically, when the view angle changes from side to oblique then to frontal, the weight of gait in the fusion changes from higher to lower, while that of face changes from lower to higher. When face does not present in the image (pattern iv and v), the face weight is set to 0. When the distance becomes closer, the face weight increases, and vice versa. The rest of this section will discuss how this prior knowledge can be integrated into the adaptive fusion rules.

Before fusion, the gait/face classifier should be trained. Firstly the training set is divided into five subsets according to the five walking patterns. Then on each subset $\kappa = i, \dots, v$, a gait classifier (Section 3.1) G^κ and a face classifier (Section 3.2) F^κ are trained. During the test procedure, given a previously unseen video \mathbf{X} of n frames, the algorithm is to find out the identity of the person in it. Firstly, the walking pattern ω of \mathbf{X} is determined through the silhouette analysis shown in Table 2, and the corresponding gait classifier G^ω and face classifier F^ω are chosen. Then \mathbf{X} is divided into m subsets along the time axis with overlap v between the neighboring subsets. Each subset $\mathbf{X}^i, i = 1, \dots, m$ corresponds to a short period of time, and the fusion rule in it is assumed to be steady. The gait recognition algorithm usually works when the video sequence includes at least one walking cycle (two steps), thus the length of each subset p should include at least one walking cycle. For the similarity between \mathbf{X}^i and person j , a gait score $S_g^{ij} = G^\omega(\mathbf{X}^i, j)$, and a face score $S_f^{ij} = F^\omega(\mathbf{X}^i, j)$ can be calculated. These scores, with quite different ranges and distributions, must be normalized before fusion. The exponential transformation is used here for normalization. Given the original score to person j as S_j , the normalized score \tilde{S}_j is calculated by

$$\tilde{S}_j = \exp(S_j) / \sum_j \exp(S_j). \quad (7)$$

For each subset \mathbf{X}^i , the subject-to-camera distance in each frame is estimated by Equation (6), and the average distance \bar{D}^i in \mathbf{X}^i is calculated. Then the normalized gait score \tilde{S}_g^{ij} and face score \tilde{S}_f^{ij} are weighted combined as:

$$S_{ij} = \lambda \tilde{S}_g^{ij} + (1 - \lambda) \tilde{S}_f^{ij}, \quad (8)$$

$$\lambda = \frac{\bar{D}^i - \min_i(\bar{D}^i)}{\max_i(\bar{D}^i) - \min_i(\bar{D}^i)} \times (\max_\omega - \min_\omega) + \min_\omega, \quad (9)$$

where $[\min_\omega, \max_\omega]$ is the range of λ for the walking pattern ω . Obviously, that range for the walking pattern (iv) and (v) should be $[1, 1]$, i.e. $\lambda \equiv 1$. Note that the α in

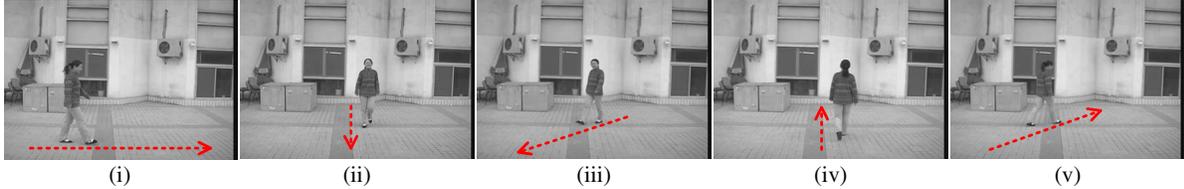


Figure 5. Typical images in the five walking patterns from the NLPR Gait Database.

Table 3. Recognition Rates (%) on the Test Set of the NLPR Gait Database

Walking Pattern	Gait-Only	Face-Only	Fusion of Gait and Face				
			SUM	PRODUCT	MIN	MAX	Adaptive
(i)	67.50	67.50	<u>80.00</u>	<u>72.50</u>	65.00	57.50	85.00
(ii)	55.00	75.00	60.00	45.00	50.00	60.00	85.00
(iii)	70.00	70.00	60.00	35.00	30.00	70.00	95.00
(iv)	50.00	45.00	<u>75.00</u>	<u>75.00</u>	80.00	<u>70.00</u>	80.00
(v)	60.00	35.00	55.00	45.00	30.00	<u>75.00</u>	90.00
All	61.67	60.00	<u>68.33</u>	57.50	53.33	<u>65.00</u>	86.67

Equation (6) can be removed from \bar{D}^i as a common factor of the numerator and denominator. It can be seen that the fusion weights are determined by both ω and \bar{D}^i . The walking pattern determines the possible range of the weight, and the subject-to-camera distance determines how the weight varies in that range. The values of min_ω and max_ω are chosen based on the relationship between the view angle and gait/face recognition, as has been shown in Fig. 1. After getting all the combined scores for the m subsets in \mathbf{X} , the average value $\bar{S}_j = \sum_i S_{ij}/m$ is calculated as the similarity between \mathbf{X} and person j . Finally the estimated ID for \mathbf{X} is determined by

$$ID(X) = \arg \max_j (\bar{S}_j). \quad (10)$$

4. Experiment

4.1. Methodology

The data used in the experiment is the NLPR Gait Database [7]. There are 20 different subjects in this database. Each subject walks along a straight-line path back and forth twice. There are three different angles between the path line and the image plane: lateral (0°), oblique (45°), and frontal (90°). In total, there are 240 gait image sequences ($20 \times 3 \times 2 \times 2$). These sequences are captured at a rate of 25 frames per second and the original resolution is 352×240 . The length of each sequence varies with the walking speed of the subjects, but the average length is about 90 frames. The typical examples from the NLPR Gait Database in the five walking patterns are shown in Fig. 5.

For gait score calculation, the silhouette images are normalized to 48×32 , $l = 25$ and $K = 15$. For face score

calculation, the face images are normalized to 25×25 , $q = c - 1 = 19$. For the adaptive fusion, the number of frames in each subset $p = 30$, and the overlap $v = 15$. The ratio threshold $\rho_1 = \rho_2 = 0.1$. The ranges of λ for the walking patterns (i)-(v) are set as $[0.7, 0.7]$, $[0.5, 0.8]$, $[0.6, 0.9]$, $[1, 1]$, and $[1, 1]$, respectively. When the upper bound equals to the lower bound, λ is a constant, *e.g.* $[0.7, 0.7]$ for pattern (i) means $\lambda \equiv 0.7$ because in the side view the subject-to-camera distance is relatively steady, and gait is more reliable than face (higher weight for gait in the fusion). The range for pattern (ii) is lower than that for pattern (iii) because gait recognition prefers pattern (iii) while face recognition performs better in pattern (ii). Finally, λ is set to the constant 1 for pattern (iv) and (v) because face does not present in these two walking patterns. We also tested several other configurations of the ranges, as long as the above prior knowledge about the relationship between gait/face recognition and the view angle is complied with, no significant difference was observed.

In the NLPR Gait Database, each subject walks along each path line back and forth *twice*. One of them is used as training data and the other is used as test data. In total, there are 120 training sequences and 120 test sequences. The recognition accuracy of gait-only, face-only, and the fusion of them are compared. The fusion methods include the adaptive fusion proposed in Section 3 and the fixed fusion rules commonly used by most previous work on multi-biometric fusion [5] [4] [8], namely SUM, PRODUCT, MIN, and MAX.

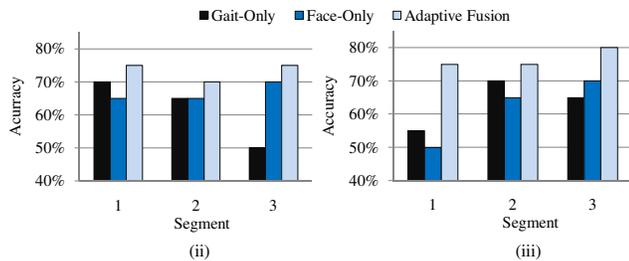


Figure 6. Recognition rates of the first three segments in pattern (ii) and (iii).

4.2. Result

The recognition rates of gait-only, face-only and the fusion of them on the NLPR Gait Database are tabulated in Table 3. The best performance in each case is highlighted by boldface. As can be seen, the overall performance of single biometric recognition, either gait-only or face-only, is unsatisfying. The results of gait-only and face-only in different walking patterns support the prior knowledge about the relationship between gait/face recognition and the view angle, which is introduced into the adaptive fusion, *i.e.* gait recognition prefers the side view while face recognition performs better in the frontal view, and the back view of ‘face’ (actually the hair) in pattern (iv), (v) is not a reliable biometric trait (the corresponding accuracy is lower than 50%). When gait and face are combined, the recognition rates higher than both those of gait-only and face-only are underlined. It reveals that none of the *static fusion* rules can guarantee a better result over the single biometric trait in all cases. As for the overall performance, only SUM and MAX can get slightly better results than the single biometric recognition. This might be due to the usage of the fixed fusion rules without considering the reliability of different biometric traits under different conditions. An unreliable single biometric trait might deteriorate the performance of the other better one in the fusion. On the other hand, with the ability to dynamically adjust the fusion rule according to the view angle and subject-to-camera distance, the *adaptive fusion* performs significantly better than not only the single biometric recognition, but also all static fusion rules in all cases. Note that although in pattern (iv) and (v), only gait is used in the adaptive fusion, the results are different from those of gait-only because the gait classifier used in the adaptive fusion is trained on training sets separated by different walking patterns while that used in gait-only is trained on all training data. It is expectable to get better result of gait-only if trained and tested on the same walking pattern, but without the ability to detect the walking pattern, it cannot automatically choose the proper classifier for the current test video.

In the adaptive fusion, the fusion weight varies with the subject-to-camera distance in two (ii, iii) of the five walking patterns. Here the distance refers to the average dis-

tance of the segments obtained by dividing the image sequence along the time axis (the minimum number of segments in pattern (ii) and (iii) is 3). In order to discover the relationship between distance and recognition accuracy, gait-only, face-only and the adaptive fusion are tested segment by segment. The results on the first 3 segments are shown in Fig. 6, which support the prior knowledge about the subject-to-camera distance used in the adaptive fusion, *i.e.* the closer the subject to the camera (from segment 1 to 2 then to 3), the more accurate face-only relative to gait-only. Moreover, the adaptive fusion performs significantly better than both gait-only and face-only in all cases.

5. Conclusion

This paper proposes *adaptive multi-biometric fusion*. Up to the present, most existing work on multi-biometrics is based on *static fusion* rules. On the contrary, adaptive fusion can perceive the changes of the environment/users and dynamically adapt the fusion rule to those changes. To illustrate the advantages of adaptive fusion, the fusion of gait and face that is adaptable to the view angle and the subject-to-camera distance is investigated. Experimental results reveal that the adaptive fusion of gait and face performs significantly better than the conventional static fusion rules including SUM, PRODUCT, MIN and MAX.

References

- [1] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Machine Intell.*, 19(7):711–720, 1997.
- [2] X. He and P. Niyogi. Locality preserving projections. In *Proc. NIPS’03*, British Columbia, Canada, 2003.
- [3] A. K. Jain and A. Ross. Multibiometric systems. *Communications of the ACM, Special Issue on Multimodal Interfaces*, 47(1):34–40, 2004.
- [4] A. Kale, A. Roychowdhury, and R. Chellappa. Fusion of gait and face for human identification. In *Proc. ICASSP’04*, pages V–901–904, Montreal, Canada, 2004.
- [5] G. Shakhnarovich and T. Darrell. On probabilistic combination of face and gait cues for identification. In *Proc. FGR’02*, pages 169–174, Washington, DC, 2002.
- [6] L. Wang and D. Suter. Analysing human movements from silhouettes using manifold learning. In *Proc. AVSS’06*, page 7, Sydney, Australia, 2006.
- [7] L. Wang, T. Tan, H. Ning, and W. Hu. Silhouette analysis-based gait recognition for human identification. *IEEE Trans. Pattern Anal. Machine Intell.*, 25(12):1505–1518, 2003.
- [8] X. Zhou and B. Bhanu. Integrating face and gait for human recognition. In *Proc. CVPRW’06*, page 55, New York City, NY, 2006.