An Adaptive System for Gait Recognition in Multi-View Environments

Yu Guan
Department of Computer Science
University of Warwick
Coventry, UK
+44 024 7657 3793
g.yu@warwick.ac.uk

Chang-Tsun Li
Department of Computer Science
University of Warwick
Coventry, UK
+44 024 7657 3794
c-t.li@warwick.ac.uk

Yongjian Hu
Department of Computer Science
University of Warwick
Coventry, UK
+44 024 7652 3987
yongjian.hu@dcs.warwick.ac.uk

ABSTRACT
Gait recognition systems often suffer from the challenges when query gaits are under the coupled effects of unknown view angles and large intra-class variations (e.g., wearing a coat). In this paper, we deem it as a two-stage classification problem, namely, view detection and fixed-view gait recognition. First, we propose two simple yet effective feature types (i.e., global features and local features) for view detection. By using the detected view information, the corresponding gallery (i.e., enrolled gait) for the detected view can be adaptively selected to perform the fixed-view gait recognition. For fixed-view gait recognition, since the inter-class variations for training are normally small, whereas the query gait usually has large intra-class variations, random subspace method are adopted. We evaluate our approach on the largest multi-view gait database CASIA-B dataset. The avoidance of searching whole multi-view database as well as the competitive performance indicate that our proposed method is practical for gait recognition in real world surveillance scenarios.

Categories and Subject Descriptors
I.5 [**Pattern Recognition**]: Application

Keywords
Multi-view, Adaptive, Gait Recognition, Biometrics

1. INTRODUCTION
Compared with other biometrics like face or iris recognition, the most significant advantage for gait recognition is that it can be applied at a distance. According to early medical and physiological studies, the human gait has 24 different components, which indicates that the gait pattern is unique for individuals [1]. Based on this premise, gait analysis has contributed to evidence for convictions in criminal cases [2]. However, for computer vision based gait recognition systems, covariate factors can affect the recognition performance. The most common factors are camera viewpoints, carrying condition, clothing etc. Given the fact that, except for the camera viewpoint factor, the gallery (i.e., enrolled gait) for each individual is normally obtained under a single walking condition (e.g., no bag carried or normal clothes type), whereas the walking condition of the probe (i.e., query gait) is unknown, how to design a robust system to address these problems is an acute challenge. There are two main approaches of gait recognition: appearance-based and model-based. The former utilizes the whole motion pattern of human body, whereas the latter uses the human body structure for recognition. In this paper, we propose a two-stage appearance-based method to cast gait recognition, with the query gait under coupled influences of camera viewpoint, carrying condition and clothing.

The average silhouette over one gait cycle, known as Gait Energy Image (GEI) is widely used in recent appearance-based gait recognition algorithms because of its simplicity and effectiveness[3][6][7][8][9][11][12]. Compared with the traditional frame-based methods (e.g., the baseline with direct frame matching [10]), the average operation encodes the static/dynamic information of a number of gait binary frames into a single grayscale image, which makes GEI less sensitive to segmentation errors [3] and walking speed [12]. However, directly applying GEI makes the classification process prone to errors [6][11], especially when the camera viewpoint is large (e.g., more than 30˚). For the viewpoints factor, there are three scenarios in practices [4]: *1)* Fixed-view gait recognition when probe and gallery are matched from the same view. *2)* Cross-view gait recognition when probe and gallery are matched from two different views. *3)* Multi-view gait recognition where multiple views gallery gaits are combined to recognize probe gait for an unknown view. Most existing view-invariant algorithms for scenarios *2* or *3* ignore either the effects of walking condition (e.g., [4][15]) or extreme view angles (e.g., [7][13][15]), while most carrying condition/clothing-invariant algorithms tested under scenario *1* do not take the view covariate into consideration (e.g., [14]). Only a few works attempt to tackle the covariates coupled effects in real world surveillance scenarios, e.g., query gaits from a wide range of unknown view angles under the influence of carrying a bag/wearing a coat (e.g., Fig. 1).

In multi-pose/view face recognition, pose/view estimation can be deemed as a preprocessing step. For example, view-based pose estimation method [5] enrolls multiple faces under various poses or views, and matches the probe image with the gallery image which has the most similar pose condition. Bashir et al. [13] bring this idea to gait recognition by estimating the unknown probe view in the first place. However, there are some limitations in [13]: *1)* There are not sufficient supporting explanations for feature selection, i.e., the selected leg part from GEI is too suitable for the benchmark CASIA-B dataset, since feature template is not under the influence of wearing a coat or carrying a bag. *2)* The view detection algorithm is limited to the view range 36˚-144˚. To solve these problems, in this work, we propose two feature types (i.e., global features and local features) for view detection, which can be applied in a more general viewing condition. With the detected view, the multi-view gait recognition

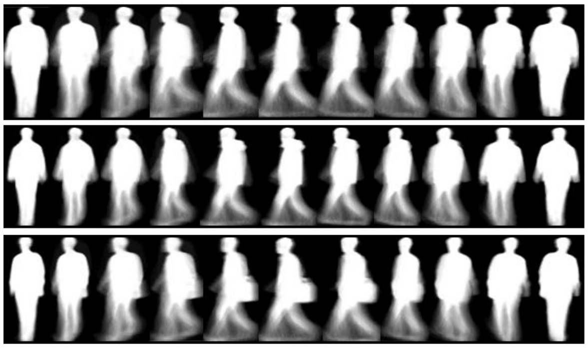

Figure 1. GEI examples from CASIA-B dataset [6] of an individual under different walking conditions from view 0˚ to 180˚, with an interval of 18˚. Top row: normal walking. Middle row: wearing a coat, Bottom row: carrying a bag.

is converted to a fixed-view gait recognition problem by adaptively selecting the gallery for the corresponding view. For the fixed-view gait recognition component, since the inter-class variations for training are normally small, whereas the query gait usually has large intra-class variations, random subspace method (RSM) are adopted. By applying the two-stage strategy, the adaptive gait recognition system can be used for query gait in unknown view angles under unknown walking conditions. It is also practical for multi-view gait recognition in real world surveillance scenarios, since with the detected view, exhausted searching through part or whole large-scale multi-view gallery can be avoided.

The remainder of this paper is organized as follows: Section 2 describes the details of the proposed approach. Experimental results and analysis are provided in Section 3, and Section 4 concludes the paper.

2. PROPOSED APPROACH

Compared with pose estimation in multi-pose/view face recognition, view detection is rarely explored in the context of multi-view gait recognition due to the lack of texture information. The only work in [13] directly employs the bottom 1/3 of the GEI (i.e., the leg part from GEI) as input feature template for gait view detection on the CASIA-B dataset (Fig. 1). However, the algorithm can only capture 7 views (i.e., from 36˚ to 164˚ with an interval of 18˚), because of the lack of discriminative view information, e.g., in Fig. 1, the leg part in 0˚ and 180˚ look almost identical. Besides, this method is based on the strong prior dataset knowledge, since the selected feature template (i.e., leg part) is barely affected by bag or coat (e.g., Fig. 1), so the robustness of the view detection method in [13] leaves unverified. To address these issues, we present a novel view detection approach in 2.1. For the whole adaptive gait recognition system, the proposed view detection method is served as a pre-processing component for the fixed-view gait recognition component, which is introduced in 2.2. By integrating both components, the whole adaptive system is shown in 2.3.

2.1 View Detection Component

For the view detection component, two types of features are proposed, namely, global features and local features. For a GEI, the former one is based on the entropy of row entries, whereas the latter is based on the standard deviation of the row entries whose values are neither *0* nor *1*.

Entropy is a statistical measure of randomness that can be used to characterize the texture of the input image. We borrow this concept to represent the global view information for each GEI row. For example, in Fig. 1, for any specific GEI rows in different views, the corresponding textures differ to some extent. For a GEI with m rows, the extracted global feature vector has a size of $m\times 1$, and the corresponding j^{th} entry is defined as:

$$H(j) = -\sum_{i=1}^{c} p_i \log_2 p_i \qquad (1)$$

Where c is the total number of the occurrence (i.e., the number of possible pixel intensity values for the j^{th} row), and p_i is the probability of the i^{th} occurrence.

Considering a robust view detection algorithm should be human-independent, dynamic information is also employed to describe the local features. The definitions of dynamic features vary from previous literatures [16][17][18], while in this work, we simply define the dynamic features as the pixels in GEI with the values in the range of *(0,1)*. The pixels with value *0* and *1* are discarded, which have no positive contribution to the view detection algorithm, since they represent the background and the human static part, respectively. Standard deviation is then employed to define the local features for each GEI row to describe the variations of the corresponding dynamic features. Similarly, for a GEI with m rows, the extracted local feature vector has a size of $m\times 1$, and the corresponding j^{th} entry is defined as:

$$S(j) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2} \qquad X_i \in (0,1) \qquad (2)$$

Where n is the pixels number with the corresponding intensity values at the range of *(0, 1)* of the j^{th} row. X_i is the pixel intensity of the i^{th} pixel, and $\bar{X}$ is the average value of these n pixels. Compared with conventional feature template, e.g., GEI or leg part of GEI [13], the proposed global features (H) / local features (S) have lower dimensionality (i.e., the same as the image row). Initially, based on H/S, Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are employed for feature extraction. For a query gait in an unknown view, in terms of Euclidean distance, K nearest neighbors (KNN) are selected separately based on H and S to form a voting pool which has $2\times K$ candidates. A decision level fusion is then adopted to obtain the query gait view label by following the majority voting criterion out of the $2\times K$ candidates. Finally, the corresponding gallery for the detected view is adaptively selected from the multi-view enrolled gait database for fixed-view gait recognition.

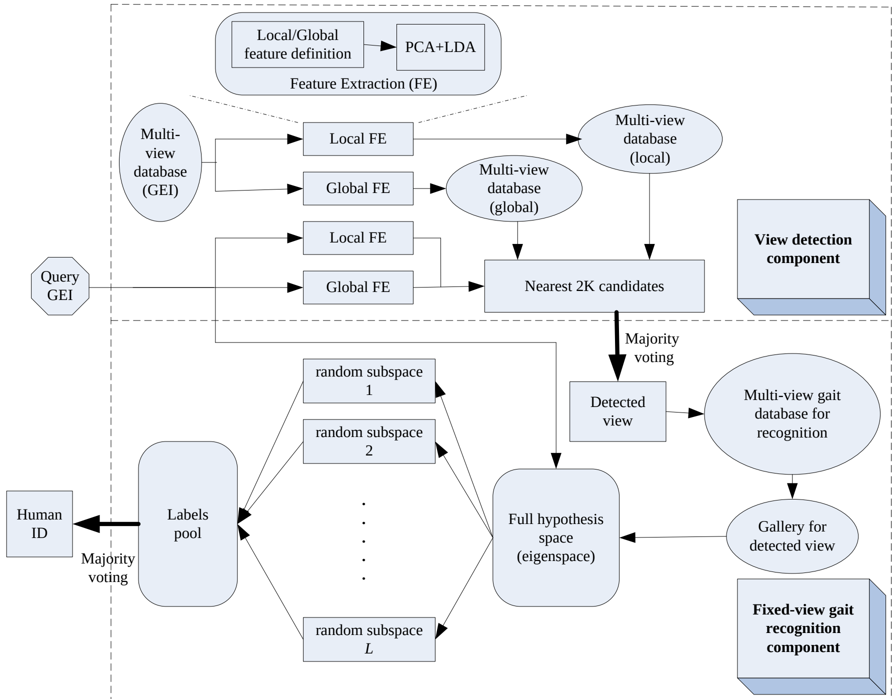

Figure 2. Overview of the multi-view adaptive gait recognition system.

2.2 Fixed-view Gait Recognition Component

We use the RSM framework as our fixed-view gait recognition component. In our previous work, we verify that when there are limited samples for training, overfitting avoidance can largely enhance the gait recognition performance and it is robust to the covariate factors like carrying condition, shoe type, small view change (i.e., less than 30˚) [8] and clothing [9]. Here we briefly review this RSM framework: Two dimensional PCA (2DPCA) is initially used for the full hypothesis space construction. The covariance matrix C then can be obtained by $C = \frac{1}{n}\sum_{j=1}^{n}(I_j - M)^T \times (I_j - M)$, where I_j is the j^{th} GEI sample, out of a total n samples in gallery. M is the average GEI over gallery, i.e., $M = \frac{1}{n}\sum_{i=1}^{n} I_i$. Then all the nonzero eigenvectors of C are preserved as the candidates to construct multiple inductive biases (i.e., random subspaces). For a certain inductive bias R^i, the corresponding eigenspace projection matrix can be constructed by randomly choosing N nonzero eigenvectors. To achieve optimal class separability, two dimensional LDA (2DLDA) is also adopted to further project eigenspace coefficients into canonical space, which maximizes the Fisher's criterion. Such two projection matrices can be used for feature extraction before nearest neighbor (NN) classifier is employed for labeling for a certain inductive bias. Since there are multiple inductive biases, majority voting is applied for the final decision.

2.3 Adaptive Gait Recognition System

We integrate these two components for an adaptive system. For a query gait from an unknown view angle, out of the whole multi-view database, the detected view gallery is selected for the matching process. It is worth mentioning that in the fixed-view gait recognition component, we randomly choose N eigenvectors, rather than the ones with top N largest energy. This is due to the small inter-class variations for training and the large intra-class variations for testing (e.g., wearing a coat). In this case, following least reconstruction error criterion by directly using PCA will lead to overfitting [9]. While in the view detection component, during PCA feature extraction, we choose the eigenvectors with the top N largest energy. This is because between different views, the inter-class variations are large while the inter-class variations are small. By integrating the view detection component and the fixed-view gait recognition component, such adaptive system is illustrated in Fig. 2.

3. EXPERIMENTS

3.1 Data and Parameters

The CASIA-B gait database [6] is the largest publicly available multi-view gait dataset, which consists of 124 subjects under 11 view angles (i.e., from 0˚ to 180˚ with an interval of 18˚). For each subject there are 10 walking sequences consisting of 6 normal walking sequences (i.e., Set A), 2 coat wearing gait sequences (i.e., Set B) and 2 bag carrying gait sequences (i.e., Set C). Out of the 6 normal walking sequences in Set A, the first 4 sequences are normally taken as gallery (i.e., Set A1), while the rest 2 sequences are taken as probe (i.e., Set A2) for the view-invariant algorithms evaluation. Besides, sequences in Set B (resp. Set C) are normally used as probe for the clothing-invariant (resp. carrying condition invariant) algorithms evaluation. Recently, Zheng et al. [7] made the GEIs available for this database, and it can be downloaded at http://www.cbsr.ia.ac.cn/users/szheng/. In this new GEI-based CASIA-B dataset, different with the original one, there is only one GEI corresponding to each gait sequence and the image is resized to 240×240. Considering the image size

Table 1. Correct View detection Rates (CVRs) for CASIA-B dataset (%)

-	Set A2			Set B			Set C		
-	H	S	F	H	S	F	H	S	F
0°	76	89	89	50	89	79	72	87	89
18°	98	88	98	85	77	84	88	84	92
36°	86	76	89	68	77	79	76	70	85
54°	93	87	98	90	78	95	93	89	95
72°	75	85	90	81	76	89	81	71	88
90°	74	77	86	53	74	69	50	77	73
108°	74	73	85	69	36	64	66	57	69
126°	91	73	93	80	47	83	88	53	89
144°	81	74	85	67	67	73	75	62	82
162°	93	77	88	89	78	88	85	66	85
180°	72	85	88	78	70	86	63	78	76
mean	83.0	80.4	**89.9**	73.6	69.9	**80.8**	76.1	72.2	**83.9**

* F is the decision level fusion based on H (i.e., global features) and S (i.e., local features).

should not be a factor that affects the view detection rate, in our experiments, we further subsample the GEI size to 128×88 for computational efficiency.

Out of the whole adaptive system, for individual independent view detection, subjects used in training and testing should be mutually exclusive, which means that the view detection system can be independently trained on an available dataset and then be tested on a different dataset. We choose 50% of the subjects from Set A1 for training, while the remaining 50% of the subjects from Set A2, Set B and Set C are used for testing. The adaptive gait recognition experiments are conducted for all the subjects in Set A2, Set B and Set C, respectively.

There are several parameters in our adaptive system. For the KNN in the view detection component, we set $K=10$ and so in the decision level fusion, there are $2\times K=20$ candidates in the voting pool. For the random subspace dimension N as well as the random subspace number L in the fixed-view gait recognition component, we use the default values in [9], i.e., $N=5$, and $L=500$.

3.2 Performance Evaluation

3.2.1 View detection evaluation

For view detection, with the mutually exclusive subjects for training and testing, the experimental results of the three probe sets (i.e., Set A2, Set B and Set C) are shown in Table 1. Based on H, the correct view detection rates (CVRs) for query gaits from range 18˚-54˚ and 126˚-144˚ are higher than other view angles. Based on S, the better CVRs can be achieved for query gaits in view 0˚ or 180˚. For the view range 72˚-108˚, although CVRs drop for both H and S, the decision level fusion of them (F) can give much better performance. Generally, for all the view angles of the three probe sets, view detection based on F tend to have better results, i.e., around 7% (resp. 10%) higher than H (resp. S) in terms of mean CVR. As expected, F successfully utilizes the compensation power between H and S, with the average CVRs 89.9%, 80.8% and 83.9% for the probe sets of normal walking, wearing a coat, and carrying a bag, respectively. For our view detection method, different walking conditions may affect the performance to some extent. Compared with the normal walking scenarios, the average CVR decreases by 9.1%

Table 2. Mutex Subjects Test for CASIA-B dataset (%)

-	Set A2	Set B	Set C
mutex	89.9	80.8	83.9
non-mutex	90.6	82.7	83.8

*Mutex denotes the subjects for training and testing are mutually exclusive. The CVRs are based on F.

Table 3. The average performance for the Fixed-view/Adaptive systems for CASIA-B dataset (%)

-	Set A2	Set B	Set C	mean
Fixed-view RSM	98.8	37.6	76.8	71.1
Adaptive RSM	96	33.9	70.5	66.8
Fix/Apt Diff (RSM)	2.8	3.7	6.3	**4.3**
Fixed-view Yu's method [6]	97.6	16.6	52.1	55.4
Adaptive Yu's method	93.3	14.8	47.1	51.7
Fix/Apt Diff (Yu's method)	4.3	1.8	5.0	**3.7**

* Fix/Apt Diff measures the degrading performance from fixed-view systems to the corresponding adaptive versions.

(resp. 6.0%) when walking with a coat (resp. with a bag). It should be pointing out that, unlike the work in [13], the input features are not pre-selected based on prior knowledge about the database. Although there are declines in terms of CVRs for different walking conditions, the performance is still competitive (i.e., at least more than 80%). So the proposed view detection method can be served as an effective component for the whole adaptive system, for various walking conditions.

Moreover, to verify our method is individual independent, we also run it on all the three probes without the constraint of mutually exclusive subjects for training and testing. The average CVRs based on F are reported in Table 2 (i.e., mutex and non-mutex). The close results indicate that our view detection algorithm has less potential in learning individual features over the view features.

3.2.2 Adaptive system evaluation

In the second component of the adaptive system, we choose the RSM framework from our previous work [8][9], which is verified to be robust to clothing[9], carrying condition, and small angle change[8]. Based on the detected view, the adaptive systems should have lower performance than the corresponding fixed-view systems which assume the query gait view is known, and the correct classification rates (CCRs) of the fixed-view gait recognition can be regarded as the upper bound of the adaptive system performance. We also implement Yu's method [6] (GEI direct matching), and incorporate it to this adaptive framework for comparison and discussion.

For fixed-view methods (i.e., RSM and Yu's method) and their corresponding adaptive versions, over the three different probe sets, the average performance and CCRs distribution over different views are reported in Table 3 and Fig. 3, respectively. Compared with RSM fixed-view method (resp. Yu's method), the average performance of the corresponding adaptive system only degrades 4.3% (resp. 3.7%), based on the detected view angles. Due to the fact that these fixed-view methods are invariant to small view change to some extent [6][8], so even in the misclassification cases, when a view angle is mistakenly assigned to a similar view label (i.e., by ±18˚), the corresponding adaptive systems still have potentials for the correct human identification. The insignificant performance differences between the fixed-view systems and the adaptive counterparts indicate the

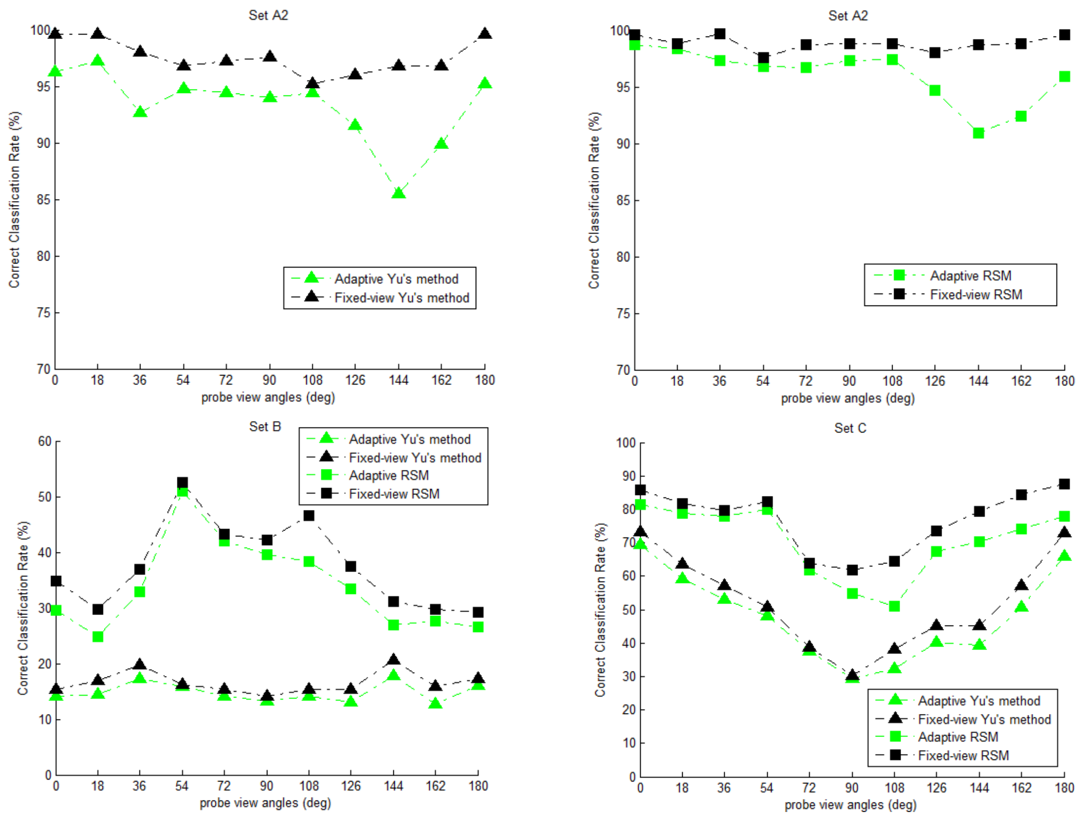

Figure 3. The correct classification rates (CCRs) distribution for CASIA-B Set A2 (i.e., normal walking), Set B (i.e., wearing a coat) and Set C (i.e., carrying a bag). The fixed-view methods assume probe with the view angles known, and the results can be deemed as the "ground truth" for the corresponding adaptive versions.

effectiveness of the adaptive framework. Compared with the adaptive Yu's method, as can be seen from Table 3, for multi-view Set A2 without the influences of other covariates, our method enhances the average CCR by 2.7%. In this case it is not significant, since both methods can achieve very good performance (i.e., more than 93% CCRs). However, under the multi-view environments, when coat (resp. bag) covariate is also taken into consideration in Set B (resp. Set C), our adaptive RSM method outperforms adaptive Yu's method by 19.1% (resp. 23.4%) in terms of average performance. It is worth mentioning that for Set B (resp. Set C), even compared with the fixed-view Yu's method, which assumes the probe view angles are known, our adaptive RSM system can still always have much better CCRs, as shown in Fig. 3.

4. CONCLUSIONS

Query gaits under the coupled influences of view angles and large intra-class variations are typical scenarios in uncontrolled environment. In this paper, by borrowing the concept of pose estimation in multi-pose/view face recognition [5], we present an adaptive gait recognition system to address these issues. The adaptive system includes two main components, i.e., view detection component and fixed-view gait recognition component. Compared with the only method [13], the proposed view detection component has a significant improvement in terms of angle coverage and robustness. For the fixed-view gait recognition component, we adopt RSM which can well solve the overfitting problem, especially when the inter-class variations for training are small and the intra-class variations for testing are large. By integrating these two components, the conventional process of exhausted searching through part or whole multi-view gallery can become self-adaptive, which is suitable for large-scale database. Based on these features as well as the competitive performance, the proposed adaptive system is practical for gait recognition tasks in the real world surveillance scenarios.

5. REFERENCES

[1] Murry, M. P. 1967. Gait as a Total Pattern of Movement. *American Journal of Physical Medicine, 46* (1). 290-332.

[2] Larsen, P.K., Simonsen, E.B. and Lynnerup, N. 2008. Gait Analysis in Forensic Medicine. *Journal of Forensic Sciences, 53*(5).1149-1153.

[3] Han, J. and Bhanu, B. 2006. Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 28*(2). 316-322.

[4] Kusakunniran, W., Wu, Q., Zhang, J. and Li, H. 2011. Cross-view and multi-view gait recognitions based on view transformation model using multi-layer perceptron. *Pattern Recognition Letters, 33*(7). 882-889.

[5] Pentland, A., Moghaddam, B. and Starner, T. 1994. View-based and modular eigenspaces for face recognition. *In*

Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR'94). (Seattle, WA, June 21-23, 1994), 84-91.

[6] Yu, S., Tan, D. and Tan, T. 2006. A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition. *In Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)*. (Hong Kong, China, September, 2006), 441-444.

[7] Zheng, S., Zhang, J., Huang, K., He, R. and Tan, T. 2011. Robust view transformation model for gait recognition. *In Proceedings of the 18th IEEE International Conference on Image Processing (ICIP'11)*. (Brussels, Belgium, September 11-14, 2011), 2073-2076.

[8] Guan, Y., Li, C.-T. and Hu, Y. 2012. Random subspace method for gait recognition. *In Proceedings of IEEE International Conference on Multimedia and Expo workshop (ICMEW'12).* (Melbourne, Australia, July 9 – 13, 2012). 284 – 289.

[9] Guan, Y., Li, C.-T. and Hu, Y. 2012. Robust Clothing-Invariant Gait Recognition. *In Proceedings of the 8th International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP'12).* (Piraeus-Athens, Greece, July 18-20, 2012). 321 – 324.

[10] Sarkar, S., Phillips, P.J., Liu, Z., Vega. I.R., Grother, P. and Bowyer, K.W. 2005. The humanID gait challenge problem: data sets, performance, and analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(2).162-177.

[11] Liu, Z. and Sarkar, S. 2004. Simplest representation yet for gait recognition: averaged silhouette. *In Proceedings of International Conference on Pattern Recognition (ICPR'04)*. (Cambridge, UK, August 23-26 2004), 211-214.

[12] Matovski, D., Nixon, M.S., Mahmoodi, S. and Carter, J. 2012. The Effect of Time on Gait Recognition Performance. *IEEE Transactions on Information Forensics and Security*, *7*(2). 543-552.

[13] Bashir, K., Xiang, T. and Gong, S. 2010. Cross-View Gait Recognition Using Correlation Strength. In *Proceedings of the 21st British Machine Vision Conference (BMVC'10).* (Aberystwyth, UK, August 30 - September 2, 2010), 109.1-109.11.

[14] Zhang, E., Zhao, Y. and Xiong, W. 2010. Active energy image plus 2DLPP for gait recognition. *Signal Processing, 90*(7). 2295-2302.

[15] Goffredo, M., Bouchrika, I., Carter, J.N. and Nixon, M.S. 2010. Self-Calibrating View-Invariant Gait Biometrics. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 40*(4). 997-1008.

[16] Lam, T. H. W., Raymond, S. T. L. and Zhang, D. 2007. Human gait recognition by the fusion of motion and static spatiotemporal templates. *Pattern Recognition*, *40*(9). 2563-2573.

[17] Wang, L., Ning, H., Tan, T. and Hu, W. 2004. Fusion of static and dynamic body biometrics for gait recognition. *IEEE Transactions on Circuits and Systems for Video Technology, 14*(2). 149- 158.

[18] Yang, X., Zhou, Y., Zhang, T., Shu, G. and Yang, J. 2008. Gait recognition based on dynamic region analysis. *Signal Processing, 88*(9). 2350-2356.