THE UNIVERSITY OF WARWICK

Original citation:

Yu, G. et al. (2012). Gait recognition under carrying condition: a static dynamic fusion method. Proceedings of SPIE, 8436, pp. 84360W. Optics, Photonics, and Digital Technologies for Multimedia Applications II. Brussels, Belgium, 17 April.

http://dx.doi.org/10.1117/12.922492

Gait recognition under carrying condition: a static dynamic fusion method

Yu Guan*, Chang-Tsun Li and Yongjian Hu
Department of Computer Science, University of Warwick, Coventry CV4 7AL, UK

ABSTRACT

When an individual carries an object, such as a briefcase, conventional gait recognition algorithms based on average silhouette/Gait Energy Image (GEI) do not always perform well as the object carried may have the potential of being mistakenly regarded as a part of the human body. To solve such a problem, in this paper, instead of directly applying GEI to represent the gait information, we propose a novel dynamic feature template for classification. Based on this extracted dynamic information and some static feature templates (i.e., head part and trunk part), we cast gait recognition on the large USF (University of South Florida) database by adopting a static/dynamic fusion strategy. For the experiments involving carrying condition covariate, significant improvements are achieved when compared with other classic algorithms.

Keywords: dynamic feature template, fusion, carrying condition, gait recognition

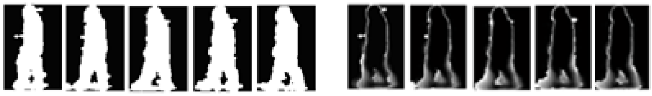
1. INTRODUCTION

Compared with other biometrics technologies like face or iris recognition, the most significant advantage for gait recognition is that it can be applied unobtrusively at a distance. According to the early medical and physiological studies, the human gait has 24 different components, which indicates that the gait pattern is unique for individuals [12]. However, covariate factors that can affect the recognition performance exist. These include, walking surface, shoe type, carrying condition, camera viewpoint, clothing, elapsed time, and so on [5]. How to address these problems is an acute challenge. There are 2 mainstreams in gait recognition, namely appearance-based gait recognition and model-based gait recognition. The former utilizes the whole motion pattern of the human body, whereas the latter uses the human body structure [3]. In this paper, we adopt the appearance-based methods to cast the gait recognition.

The average silhouette over one gait cycle, known as Gait Energy Image (GEI) is widely used in recent appearance-based gait recognition for its simplicity and efficiency [9]. Compared with the traditional frame-based methods (e.g., baseline algorithm with direct frame matching [5]), the averaging operation makes GEI insensitive to segmentation errors to some extent. However, the conventional GEI-based methods do not always perform well when an individual is in carrying status [9] [13], as the object he/she is carrying (e.g., a briefcase) may have the potential of being incorrectly regarded as a part of the human body. To solve such a problem, in this paper, dynamic and static features extracted from GEI are fused to enhance the performance. It is concluded that the static parts of the GEI (e.g., head part and trunk part) are the most important features against the dynamic parts (e.g. arms and legs) in gait recognition [1]. In [2], the GEI is divided into seven components and the head part and trunk part are assigned with higher weights against the other parts for human identification recognition and gender classification. Motivated by [1] and [2], in this paper, head part and trunk part from GEI are selected as static feature templates, with the corresponding scales fixed, based on anatomical information reported in [3]. The leg part is simply abandoned for its less significant contribution to static information representation, according to the findings in [1] [2]. To compensate such information loss, in this paper, we extract a novel dynamic feature template from GEI, and use it as an important part for the whole gait recognition system.

The rest of the paper is organized as follows: Section 2 describes the problem this work aims to solve. Section 3 covers the proposed dynamic feature template as well as the static feature templates used for classification. In Section 4, linear discriminant analysis (LDA) and score level fusion strategy are introduced. Experimental results are reported in Section 5 and conclusions are drawn in Section 6.

*G.Yu@warwick.ac.uk; phone +44 24 7657 3793



(a) (b)

Figure 1. (a) The pre-processed binary gait sequence. (b) ADE: The Absolute Difference from the Exemplar* for each frame. *****Exemplar is the average gait template over a whole gait sequence, whereas GEI is the average gait template over a gait cycle.

(a) (b) (c)

Figure 2. (a)-(b) Exemplar samples with/without a briefcase. (c) MADE: The Mean ADE for a gait sequence after denoising.

2. PROBLEM STATEMENT

After segmenting the walking people from the background, the human silhouette can be noisy, especially in the outdoor environments. For example, Fig.1 (a) shows a noisy sample gait sequence from the USF database [5]. GEI is a classic feature template which encodes both dynamic and static information of the human gait in a single image, and it can suppress the segmentation noise to some extent [9]. However, it can be heavily affected by the carrying condition, for instance, walking without a briefcase (similar to Fig.2 (a)) and walking with a briefcase (similar to Fig.2(b)) are significantly different in conventional GEI-based methods, although they come from the same person. From the classification viewpoint, it is a difficult classification task due to the large intra-class variations, whereas the inter-class variations can be small. If the Euclidean distance is adopted to reflect the dissimilarity between these two GEIs, without special processing, such distance may be enlarged even for the same individual when the briefcase area is taken into account. Dynamic features for gait recognition may solve this problem, since it less relies on the static parts of the GEI. However, the definition of the dynamic features vary [4][6][10]. In [4], the dynamic weight mask (DWM) is applied to GEI to capture the dynamic information, which assigns pixels with high variance greater weights based on the statistical analysis on corresponding pixel positions of the GEI training samples. However, the suppressed noise regions in GEI may be wrongly assigned with greater weights, since such regions normally have higher variance. Absolute differencing on raw consecutive frames is employed in [6], so the inner pixels of the carrying object are discarded in the first place. Yet for outdoor noisy data, this operation may introduce more noisy artifacts which affect the algorithm performance. Motivated by both [4] and [6], in this work, we propose a simple yet effective definition of dynamic feature template. To perform carrying condition invariant gait recognition, we also adopt a score level fusion strategy with the static parts (i.e., head part and trunk part) as suggested in [1]. Experimental results show that the proposed method outperforms most of the current algorithms.

3. FEATURE TEMPLATE SELECTION

To define the dynamic regions, we propose the Mean Absolute Difference from Exemplar (MADE) which encodes the relationship between each gait frame and the gait sequence it belongs to. Similar to the work in [4], the dynamic region mask is generated by the statistical analysis on all the MADEs in gallery, and pixels with greater variance are deemed as dynamic regions. It is worth mentioning that different with the enhanced GEI used in [4], we perform statistical analysis on the proposed MADE templates (e.g., Fig. 2(c)), which can avoid the potential of assigning higher weight to noisy pixels, since such noise pixels are suppressed according to the MADE definition. Apart from the dynamic feature template, areas covering the head and the body in GEI are also selected as static feature templates for classification.

3.1 Dynamic feature template definition

Given a pre-processed binary gait silhouette frame sequence (e.g., Fig.1(a)) from [7], $F = \{f_1, f_2, \dots f_N\}$, f_i represents the *i*th frame in the gait sequence, and *N* is the number of the frames. Different with GEI which is the average over a gait cycle, the Exemplar *E(x,y)* is the average over the whole gait sequence (e.g., Fig.2(a)-(b)).

$$E(x, y) = \frac{1}{N}\sum_{i=1}^{N} f_i(x, y) \tag{1}$$

Based on the Exemplar, for each gait sequence, to effectively encode the dynamic information, we define the *ADE* as the Absolute Difference from Exemplar for each frame (Fig.1 (b)) as follows:

$$ADE_i(x, y) = \left| f_i(x, y) - E(x, y) \right| \tag{2}$$

The following thresholding equations are applied to all the pixels of each *ADE* frame to reduce the noise.

$$ADE_i(x, y) = \begin{cases} 0 & if\,(ADE_i(x, y) \geq \theta_1) \\ 0 & if\,(ADE_i(x, y) \leq \theta_2) \\ ADE_i(x, y) & otherwise \end{cases} \tag{3}$$

θ_1 is the upper bound and θ_2 is the lower bound for each *ADE* template, which means pixels out of the range (θ_2, θ_1) are considered as noise. For example, the white areas out of the human body in Fig.1(b) are regarded as noise and will be removed through equation (3). In this paper, $\theta 1$ is assigned to 0.9 and $\theta 2$ is assigned to 0.1 for denoising. Then the Mean *ADE* template (MADE) is defined to represent the dynamic information of a gait sequence (e.g., Fig.2(c)).

$$MADE(x, y) = \frac{1}{N} \sum_{i=1}^{N} ADE_i(x, y) \tag{4}$$

Based on the MADEs in gallery, similar to the work in [4], we perform a statistical analysis to get a weighted mask, which assigns pixels with high variance greater weight. If we assume $MADE_i(x, y)$ is the *i*th class and there are *c* classes in the gallery, the standard deviation $std(x, y)$ for each pixel is as follows:

$$std(x, y) = \sqrt{\frac{1}{c} \sum_{i=1}^{c} \left[MADE_i(x, y) - \frac{1}{c} \sum_{i=1}^{c} MADE_i(x, y) \right]^2} \tag{5}$$

Using $std(x, y)$ to measure the richness of information is based on the concept of entropy, which is, pixels containing more information usually have more uncertainty, and in this paper we define the dynamic information based on such assumption. Considering the noisy pixels may also have large values of uncertainty, we perform the denosing operation by equation (3) before statistical analysis. The definition of MADE and the corresponding statistical analysis can be deemed as an improvement of the work in [4].

The MADE based dynamic weight mask (MDWM) is the normalized form of $std(x, y)$ at range [0, 1] for each pixel (e.g., Fig. 3(d)). Based on our definition, it can capture dynamic information for each gait sequences, through assigning weights to original GEI pixels with the corresponding values from MDWM. It can largely avoid the influence of the carrying condition, since for a query GEI, the MDWM can strengthen the dynamic information, whereas weaken the static information through weight assigning. The effect of carrying condition can be largely reduced in this process.

The dynamic feature template can then be extracted by an entry-wise multiplication between the MDWM and the GEI template [4]. However, in this work, instead of traditional one-gait-cycle GEI, we use half-gait-cycle GEI (e.g., Fig. 3(c)) to increase the size of the training samples for more dynamic information, according to the fact that most people walk symmetrically in the first and second half of a gait cycle [8].

$$Dy(x, y) = GEI(x, y) \times (MDWM(x, y)) \tag{6}$$

In equation (6), $\times$ denotes the entry-wise product, $Dy(x, y)$ is the dynamic features extracted from the corresponding pixels of GEI, and we denote this dynamic feature template as DyGEI (e.g., Fig. 3(a)).

3.2 Static feature template selection

For the fusion method, we also apply the static feature templates for classification, together with the proposed dynamic feature template. Instead of using the full GEI template, we only choose its corresponding head part and the trunk part, for their richness of discriminative static information [1]. As mentioned in Section 1, according to the anatomical information in [3], the scale of the head part (18.2% over the total height) and trunk part (33.8% over the total height) is fixed over the whole GEI.

(a) (b) (c) (d)

Figure 3. (a) DyGEI: the dynamic feature template for a sample GEI. (b) The head/trunk part being used as static feature templates. (c) sample half-gait-cycle GEI. (d) MDWM: The MADE based dynamic weight mask.

4. CLASSIFICATION

First, we perform the classification based on the static feature and dynamic feature separately, and then a fused classifier combining the static/dynamic classification scores is adopted for the final result.

4.1 Linear discriminant analysis

To project high-dimensional training data into lower space with optimal class separability, linear discriminant analysis (LDA) is applied, which seeks a transformation matrix W maximizing the ratio of the between-class scatter matrix S_B to the within-class scatter matrix S_W :

$$J(W) = \frac{\left|W^T S_B W\right|}{\left|W^T S_W W\right|} \tag{7}$$

However, to meet the input requirement of the traditional LDA, data have to be vectorized which may lead to high dimensionality. To avoid high dimensionality which may cause expensive computational cost and sometimes the singularity problem, in this work, the matrix-based 2DLDA is used. Mathematical details about 2DLDA can be found at literature [11].

4.2 Classification based on dynamic feature template

Assume W_{dynm} is the transformation matrix which can maximize equation (7). Given a query gait sequence P with nR GEIs, if we denote the corresponding extracted nR DyGEI as $\{D_1, D_2, \ldots D_{nR}\}$, the projected dynamic feature templates $\{PD_1, PD_2, \ldots PD_{nR}\}$ are:

$$PD_j = W_{dynm} D_j \qquad j = 1, \ldots nR \tag{8}$$

For the classifier based on dynamic feature templates, we define:

$$Dd(P, G_i) = \frac{1}{nR}\sum_{j=1}^{nR} PD_j - m_i \qquad i = 1, \ldots c \tag{9}$$

which is a set of distances reflecting the dissimilarity between the query gait sequence P and the ith class G_i in the gallery. For two gait sequences, the shorter the distance is, the more similar these two sequences are. In equation (9), m_i is the centroid of the projected dynamic feature of the ith class, out of c classes in gallery.

4.3 Score level fusion

The classification based on static feature templates (i.e., head part and trunk part) are similar to the dynamic feature template-based classification. Following equation (9), for a query gait sequence P, the sets of distances $Dh(P, G_i)$ or $Dt(P, G_i)$ are derived from the head part or trunk part for a score level fusion.

Table 1. The results(%) for Experiments H, I, J on USF database, with left (resp. right) H, I, J columns Rank-1(resp.Rank-5) performance.

Rank 1/5	H	I	J	H	I	J
Baseline [5]	61	57	36	85	78	62
Real GEI + LDA [9]	62	59	59	88	79	80
Fusion Real/Sync GEI + LDA [9]	64	60	60	90	83	82
GEnI + LDA [14]	82	63	66	-	-	-
EGEI + LDA [4]	52	52	58	-	-	-
EGEI + Gabor + DCV [4]	72	63	63	-	-	-
Fusion MSCT/SST [10]	49	43	30	78	75	61
Head feature template + LDA	61	57	36	81	84	63
Trunk feature template + LDA	62	53	30	84	86	57
Proposed dynamic feature template + LDA	70	64	42	89	83	71
Static (head+trunk) fusion + LDA	79	78	50	93	91	81
Proposed static dynamic Fusion + LDA	**91**	**90**	**66**	**97**	**93**	**87**

In General, if we only consider one feature type, e.g., dynamic feature template only, equation (9) can be directly used for classification. However, considering there are 3 feature types, similar to [9], such distances set has to be normalized in the following form:

$$normDd(P, G_i) = \frac{c(c-1)}{2\sum_{i=1}^{c}\sum_{j=1, j\neq i}^{c} Dd(G_i, G_j)} Dd(P, G_i) \tag{10}$$

where $2\sum_{i=1}^{c}\sum_{j=1, j\neq i}^{c} Dd(G_i, G_j) \,/\, c(c-1)$ is the average value of the distances between every two classes in gallery and it is used to normalize $Dd(P, G_i)$. Similarly, we can get $normDh(P, G_i)$ for head part and $normDt(P, G_i)$ for trunk part. Then final dissimilar distance can be defined as:

$$D(P, G_i) = normDd(P, G_i) + normDh(P, G_i) + normDt(P, G_i) \tag{11}$$

For a query gait sequence, after computing the normalized score using equation (11) with each gait sequence in the gallery, the one which holds a minimal score is regarded as identical with the query gait sequence.

5. EXPERIMENTS

We conduct gait recognition methods on the USF outdoor gait database [5][7]. It is a large database with 12 pre-designed experiments for algorithms comparison. In this work, we only choose 3 experiments which involve changes in carrying status. (i.e., Experiment H, I, J). In gallery, there are 122 gait sequences under a normal walking condition. For experiment H, there are 120 probe gait sequences and the covariate factor is carrying condition. For experiment I, there are 60 probe gait sequences under the combined influence of shoe type and carrying condition. There are 120 probe gait sequences for experiment J and the covariate factors are camera viewpoint and carrying condition. To evaluate the performance of the algorithms, we adopt the rank-1/rank-5 recognition accuracy. Rank-1 (resp. rank-5) shows the correct individual is ranked as the top 1 candidate (resp. top 5 candidates).

Table 1 shows the experimental results of the proposed method and other conventional algorithms. The average recognition accuracy of the proposed static dynamic fusion method over the 3 experiments is promising compared with others. We also conduct experiments based on head feature template only, trunk feature template only, static feature templates fusion and proposed dynamic feature template only for further discussion. In Experiments H and I which involve the changes in carrying condition and shoe type, the performance of the proposed static dynamic fusion method is superb (i.e., more than 90% accuracy in terms of rank-1). Although it only achieves a 66% recognition rate in Experiment J which is under the combined influence of carrying condition and camera viewpoint, it still outperforms other algorithms. It should be pointed out that different with the work in [4], the proposed dynamic features template based on MADE can achieve a better performance, since instead of being assigned with higher weight, the noise areas are suppressed before the statistical analysis, according to the MADE definition.

Moreover, as can be seen from Table 1, recognition accuracy based on only one static feature template (i.e., head part only or trunk part only) is not enough. Intuitively, that is the lack of information, since head/trunk part is only a subset of the whole GEI. However, one interest observation is that fusing these two subsets may bring an even better performance than algorithms based on the whole set (i.e., Real GEI + LDA [9]). One of the explanations is that leg part may negatively affect the performance, since they can be easily corrupted by other covariate factors (e.g., by shadows, or briefcase). We fuse the static feature templates together with the proposed dynamic feature template, and the result shows that they are good compensation with each other.

6. CONCLUSION AND FURTHER WORK

This paper demonstrates an effective static/dynamic fusion method for gait recognition under carrying condition. Compared with other algorithms, the general performance of this work is promising based on the carrying condition experiments from the USF database. The dynamic feature template we defined in this paper is an improvement of the work in [4], and it also serves as an important part of gait recognition fusion system. The static/dynamic fusion method shows its effectiveness in the experiments under the influences of carrying condition. However, the static features we have chosen are sensitive to static parts shape distortion (e.g., the shape of head/trunk may change under a different camera viewpoint). More robust dynamic/static features (e.g., view-invariant features) will be explored in the future.

REFERENCES

[1] Veres, G., Gordon, L., Carter, J. and Nixon, M., “What image information is important in silhouette-based gait recognition,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 2 , 776-782 (2004).

[2] Li, X., Maybank, S.J., Yan, S. Tao, D. and Xu ,D., "Gait Components and Their Application to Gender Recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews,* 38 , 145-155 (2008).

[3] Winter, D.A., [The Biomechanics and Motor Control of Human Movement], John Wiley & Sons, (1990).

[4] Yang, X., Zhou Y., Zhang,T., Shu, G. and Yang, Jie., “Gait recognition based on dynamic region analysis,”. Signal Processing, 2350-2356 (2008).

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

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

[7] http://figment.csee.usf.edu/GaitBaseline/

[8] Huang, X. and Boulgouris, N.V., "Gait recognition for random walking patterns and variable body postures," *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP),* 1726-1729 (2010).

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

[10] Lam,T. H. W., Raymond, S. T. L.,and Zhang, D., “Human gait recognition by the fusion of motion and static spatio-temporal templates”, Pattern Recognition, 40 , 2563-2573 (2007).

[11] Li,M. and Yuan, B., “2D-LDA: a statistical linear discriminative analysis for image matrix,” Pattern Recognition Letters, 26, 527-532 (2005).

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

[13] Liu, Z. and Sarkar,S., “*Simplest* representation *yet* for *gait* recognition: averaged silhouette,” in Proc. IEEE Conf. Pattern Recognition, 211–214 (2004).

[14] Bashir, K., Xiang, T., and Gong, S., "Gait recognition using Gait Entropy Image," *International Conference on Crime Detection and Prevention (ICDP),* 1-6 (2009).