Human Identification Using Temporal Information Preserving Gait Template

Chen Wang, Junping Zhang, *Member*, *IEEE*, Liang Wang, *Senior Member*, *Senior IEEE*, Jian Pu, and Xiaoru Yuan, *Member*, *IEEE*

Abstract—Gait Energy Image (GEI) is an efficient template for human identification by gait. However, such a template loses temporal information in a gait sequence, which is critical to the performance of gait recognition. To address this issue, we develop a novel temporal template, named Chrono-Gait Image (CGI), in this paper. The proposed CGI template first extracts the contour in each gait frame, followed by encoding each of the gait contour images in the same gait sequence with a multichannel mapping function and compositing them to a single CGI. To make the templates robust to a complex surrounding environment, we also propose CGI-based real and synthetic temporal information preserving templates by using different gait periods and contour distortion techniques. Extensive experiments on three benchmark gait databases indicate that, compared with the recently published gait recognition approaches, our CGI-based temporal information preserving approach achieves competitive performance in gait recognition with robustness and efficiency.

Index Terms—Computer vision, gait recognition, biometric authentication, pattern recognition

1 Introduction

Biometric authentication has broad applications in social security, individual identification in law enforcement, and access control in surveillance. Unlike other biometric features such as iris, faces, palm, and fingerprint, the advantages of gait include: 1) Gait can be collected in a noncontactable, noninvasive, and hidden manner; 2) gait is the only perceptible biometric at a distance. However, the performance of gait recognition suffers from some exterior factors such as clothing, shoes, briefcases, and environmental context. Furthermore, whether or not the spatiotemporal relationship between gait frames in a gait sequence is effectively represented also influences the performance of gait recognition systems. Although it is a challenging task, the nature of gait indicates that it is an irreplaceable biometric [1] and can benefit the remote biometric authentication [2].

To build a successful gait recognition system, feature extraction plays a crucial role. Currently, gait feature extraction methods can be roughly divided into two major categories: model-based and model-free approaches. Model-based approaches assume that the gait can be modeled with a structure/motion model [3]. However, it is not easy to extract the underlying model from gait sequences [3], [4]. Model-free approaches either keep temporal information in the recognition (and training) stage [5], [6], [7], [8], or convert a sequence of images into a single template [1], [9], [10], [11], [12]. Although some model-free approaches such as Gait Energy Image (GEI) [1] have attractively low computational cost, such a conversion may lose the temporal information of gait sequences.

To preserve temporal information and display time-varying sequence in a single colored image, the visualization community has proposed some interesting strategies. For example, Woodring and Shen [13] displayed time-varying data by encoding the time varying information of the data into color spectrum. Jänicke et al. [14] proposed measuring local statistical complexity for multifield visualization. More recently, Wang et al. [15] claimed that critically important areas are the most essential aspect of time-varying data to be detected and highlighted. However, it is difficult and ineffective to directly employ such methods to generate a good temporal template for gait recognition since, unlike visualization data, gaits always have larger overlapped regions between frames [16].

Considering the pros and cons of gait recognition methods mentioned above, we pay more attention to the refinement of the single template method in this paper because of its simplicity and low computational complexity. We propose a multichannel temporal encoding technique, named Chrono-Gait Image (CGI), to encode a gait sequence to a multichannel image in order to preserve the temporal information of gait patterns well. When the number of multichannel is equal to 3, it can be regraded as a pseudocolor image. As an illustrative example, we show an example of a gait sequence, GEI and CGI, in Fig. 1. To enhance the discriminant ability of CGIs in complex environment, we also introduce several strategies to generate real and synthetic CGI templates. In comparison with the state-of-the-art methods, our major contributions are:

- *C. Wang, J. Zhang, and J. Pu are with the Shanghai Key Lab of Intelligent Information Processing, School of Computer Science, Fudan University, Handan Road 220, Shanghai 200433, China. E-mail: chen.wang0517@gmail.com, jpzhang@fudan.edu.cn, mydaiyu@hotmail.com.*
- *L. Wang is with the National Lab of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China. E-mail: wangliang@nlpr.ia.ac.cn.*
- *X. Yuan is with the Key Laboratory of Machine Perception, School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China. E-mail: xiaoru.yuan@pku.edu.cn.*

Manuscript received 30 Dec. 2010; revised 24 Sept. 2011; accepted 2 Dec. 2011; published online 20 Dec. 2011.
Recommended for acceptance by G. Mori.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2010-12-0995.
Digital Object Identifier no. 10.1109/TPAMI.2011.260.

0162-8828/12/$31.00 © 2012 IEEE Published by the IEEE Computer Society

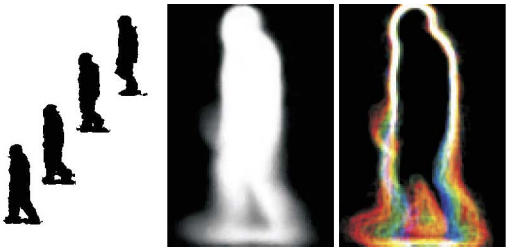


Fig. 1. From left to right: A gait sequence, gait energy image, and chrono-gait image.

1. Simple and easy to implement, CGI effectively preserves the temporal information in a gait sequence with a single template image.
2. Unlike intensity, multichannel technique, which has higher variance than grayscale, can enlarge the distance between gait sequences from different subjects and thus benefit gait recognition.
3. CGI is robust to different gait period detection methods which are usually a basis of constructing gait templates.
4. Our proposed method shows better robustness to surrounding environment according to our experiments.
5. Compared with most recently published gait recognition approaches, the training phase and recognition phase of CGI are quite efficient to make it competitive in some real-time scenarios.
6. To the best of our knowledge, multichannel encoding gait images as a temporal information preserving template for gait recognition has not yet been exploited in the biometric authentication community.

Experiments indicate that compared with several recently published approaches, the CGI temporal template attains competitive performance on three benchmark databases. It is worth noting that this paper is an extended version of our previous conference paper [16]. The main differences are that:

1. We generalize the previous pseudocolor gait template to a more general multichannel one. In this way, we can better understand the underlying mechanism that the proposed CGI has a good performance in gait recognition.
2. We improve the calculation of the average width of the leg region in the procedure of period detection and multichannel mapping to make it more reasonable, leading to better performance.
3. The whole technical details, including preprocessing, contour extraction, period detection, multichannel encoding, and representation construction for classification, are further clarified so that the work can be easily reproduced.
4. We evaluate the performance of our proposed CGI temporal information preserving template in three benchmark databases (i.e., USF, CASIA, and Soton) rather than the previous one (i.e., USF).
5. Furthermore, we conduct comprehensive experiments to study the influence of different parameters and variants in our algorithms to the performance of gait recognition, which also makes the properties of our proposed approaches clear.
6. We also discuss the limitation of our algorithms and some potential ways to avoid the issue.

The remainder of the paper is organized as follows: We give a brief survey on gait recognition in Section 2, and detail the proposed CGI temporal information preserving template in Section 3. We introduce the generation of real and synthetic CGI templates and the corresponding human recognition procedure in Section 4. Experiments are performed and analyzed in Section 5. We conclude the paper in Section 6.

2 Related Work

Gait features are very important in improving the performance of gait recognition. Generally speaking, there are two different gait feature extraction methods.

Model-based approaches are devoted to recovering the underlying mathematical construction of gait with a structure or motion model [3]. Wang et al. adopted procrustes analysis to capture the mean shapes of the gait silhouettes [17]. However, it is time consuming and vulnerable to noise. Veres et al. [18] and Guo and Nixon [19] employed the analysis of variance and mutual information, respectively, to discuss the effectiveness of features for gait recognition. Bouchrika and Nixon proposed a motion-based model by using the elliptic Fourier descriptors to extract crucial features from human joints [4]. Wang et al. [20] employed a condensation framework in which the structural-based and motion-based models are combined to refine the feature extraction. Chai et al. [21] divided the human body into three parts; then the variances of these parts are combined as the crucial gait features. Although the structure-based models can, to some degree, deal with occlusion and self-occlusion as well as rotation, the performance of the approaches suffers from the localization of the torso and it is not easy to extract the underlying model from gait sequences [3], [4]. Furthermore, it is necessary to understand the constraints of gait such as the dependency of neighboring joints and the limitation of motion to develop an effective motion-based model [3].

The model-free approaches can be divided into two major categories based on the manners of preserving temporal information. The first strategy keeps temporal information in the recognition (and training) stage [5], [6], [7], [8]. Sundaresan et al. utilized a hidden Markov models (HMMs) based framework to preserve such information [6]. By regarding gait data as a collection of cumulated frames, Kobayashi and Otsu [7] extracted the divergence between different gait states using the "Cubic Higher-order Local Auto-Correlation" (CHLAC) technique. Sarkar et al. [8] utilized the correlation of sequence pairs to preserve the spatiotemporal relationship between the galley and probe sequences. Wang et al. [5] applied principal component analysis (PCA) to extract statistical spatiotemporal features of gait frames. Liu and Sarkar [22] employed population HMM to model human walking and generated the dynamics-normalized stance-frames to recognize individuals. However, large-scale training samples are generally needed for probabilistic temporal modeling methods (such as HMMs) to obtain a good performance. A disadvantage for the direct sequence matching methods is the high

computational complexity of sequence matching during recognition and the high storage requirement.

The second strategy converts a sequence of images into a single template [1], [9], [10], [11], [12]. Liu and Sarkar [9] proposed representing the human gait by averaging all the silhouettes. Motivated by their work, Han and Bhanu [1] proposed the concept of GEI, and constructed the real and synthetic gait templates to improve the accuracy of gait recognition. With a series of grayscale averaged gait images, Xu et al. employed discriminant analysis with tensor representation (DATER) for individual recognition [10]. Chen et al. proposed multilinear tensor-based nonparametric dimension reduction (MTP) [12] for gait recognition, and Zhang et al. generalized the MTP into low-resolution gait recognition [2]. Recently, Guo and Nixon proposed to utilize mutual information to select a subset of gait features to improve the performance of gait recognition [19]. However, the above template-based methods more or less lose the temporal information of gait sequences. For example, averaging template methods throw out all the temporal order information of the gait sequence. Moreover, the time and space computational complexities of those tensor-based approaches are too high to be employed in real applications [10], [11].

3 Chrono-Gait Images

Generally speaking, regular human walking always has a fixed cycle with a particular frequency because of the basic structure of human body. As a result, such walking is generally used in most of the current approaches of human identification by gait. However, some methods may neglect the influence of gait cycle information, e.g., GEI. Meanwhile, other methods require high computational cost to preserve such information. To address the issue, we propose to encode time-varying gait cycle information into a single chrono-gait image by using the multichannel technique. We also make several fundamental assumptions in this paper: 1) Most normal people have a similar gait gesture such as the stride length. 2) Each person has his/her unique gait behavior, such as the shape of the torso, the moving range of limbs, and so on. 3) Each channel of the multichannel method can be regarded as a function of time.

In the following sections, we will introduce some preprocessing techniques used for gait recognition. We also present a novel period detection method that extracts the gait period more accurately. Then, we detail the multichannel algorithm of generating the temporal information preserving gait template.

3.1 Preprocessing and Period Detection

To achieve a gait recognition system, some preprocesses, including background subtraction and foreground alignment, are required. Here, we assume that such preprocesses have been done to the original gait sequence. Concretely, we perform our gait recognition algorithm on the silhouette images. The silhouette images are generally obtained by using some well-known algorithms to subtract background and align foreground objects, e.g., the baseline algorithm proposed by Sarkar et al. [8]. Then, we employ different channels to encode spatial-temporal information in different phases of the gait period to generate the chrono-gait image. The goal of CGIs is to compress the silhouette images into a single multichannel image and preserve as much temporal relationship between continuous frames as possible.

Considering that regular human walking is a periodical motion, it is necessary to detect the period in the gait sequence for preserving the temporal information in the CGI template. We propose using the degree of the individual's two legs apart from each other to represent regular human walking, to detect the gait period in order to find some key frames, and to measure each gait frame's relative position in a single period. Since some exterior factors such as bag, briefcase, shadow, and surface that might be misclassified into the foreground can influence the performance of period detection, we calculate the average width W of the leg region in a gait silhouette image I as follows:

$$W = \frac{1}{\beta h - \alpha h + 1} \sum_{i=\alpha h}^{\beta h} (R_i - L_i), \quad 0 \le \alpha \le \beta \le 1, \qquad (1)$$

where h is the height of an individual (the foreground) in the image, L_i and R_i are the positions of the leftmost and rightmost foreground pixels in the ith line of the individual, respectively. Compared with our previous work [16], we use the height of the individual instead of the height of the whole image to calculate the average width of leg region. The reason is that some anatomical studies [23] show that the relative vertical positions (normalized by height) of some anatomical landmarks of an individual are similar to most people, e.g., the vertical positions of hip, knee, and ankle are 0.470, 0.715, 0.961 h, respectively. In the previous work, we assumed that the leg regions were located in a similar vertical position of the gait image. However, in some gait databases, the silhouette images are not normalized to make all the individuals the same height. Furthermore, some alignment technologies (e.g., alignment based on the centroid of the silhouette) may lead to the leg region of different individuals being located in different positions on the Y-axis of the image. Therefore, finding the leg region by the relative vertical position of the individual is more reasonable than by the relative position of the whole image. Here, the two parameters α and β are used to constrain the computation of the gait period to the leg region, and meanwhile decrease the influence of those external factors.

It is worth noting that Sarkar et al. [8] proposed detecting such key frames by counting the number of foreground pixels in the lower half of the silhouettes in their baseline algorithm. We show the difference between these two detection methods in Fig. 2, from which we can see that two detection methods pay attention to different parts of gait sequence. In the proposed period detection method, the average width W will have a local maximum when the two legs are farthest apart from each other and reach a local minimum when the two legs wholly overlap. Fig. 2 also indicates that our method produces sharper peaks and valleys, and thus preserves the correct temporal order well compared with the baseline algorithm [8].

3.2 Multichannel Mapping

To visualize time-varying information, several possible strategies can be considered in the visualization community.

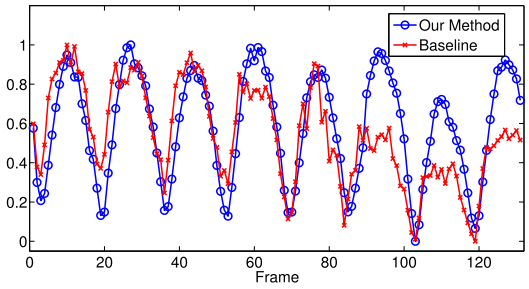


Fig. 2. Comparison between our method and the baseline algorithm on gait period detection. The X-axis denotes the order of gait frames. The Y-axis represents the average width of each frame for our method, and the number of foreground pixels in the lower half of the silhouette for baseline method. Both of them are normalized to $[0,1]$. Here, the values of α and β are 23/32 and 29/32, respectively.

A representative way is to employ pseudocolor to visualize such information for volume rendering [13]. In this method, Woodring and Shen proposed four integration functions: alpha compositing, first temporal hit, additive colors, and minimum/maximum intensity. Note that they assume that there is little overlapped foreground region between continuous frames, whereas in our case the overlap of foreground silhouettes is serious between gait frames. Consequently, we cannot directly use their methods to generate a good temporal information preserving template for gait recognition [16].

Since the outer contour of the silhouette images is an important feature [5], [17] and also preserves the spatial information with small degree of overlap, we attempt to extract the contours instead of silhouettes. To extract the contours of the silhouette images, there are various edge detection techniques such as gradient operator, LoG operator, and local information entropy [24]. We adopt local information entropy to obtain the gait contour from the silhouette image since it provides more abundant features than gradient and LoG operators. The local information entropy is defined as

$$h_t(x,y) = -\left(\frac{n_0}{|\omega_d(x,y)|}\ln\frac{n_0}{|\omega_d(x,y)|} + \frac{n_1}{|\omega_d(x,y)|}\ln\frac{n_1}{|\omega_d(x,y)|}\right), \tag{2}$$

where the d-neighborhood of point (x,y) based on D_8 distance (chessboard distance) is

$$\omega_d(x,y) = \{(u,v)|\max\{|u-x|,|v-y|\} \le d\},$$

and n_0 and n_1 are the numbers of foreground pixels and background pixels in $\omega_d(x,y)$, respectively. Term t represents the frame label, and x and y denote the horizontal and vertical values in the two-dimensional image coordinate, respectively. The neighborhood parameter d is set to 1 to emphasize the locality in our experiment. Then, we normalize the entropy by the following formula:

$$h'_t(x,y) = \frac{h_t(x,y) - \min_{x,y} h_t(x,y)}{\max_{x,y} h_t(x,y) - \min_{x,y} h_t(x,y)}. \tag{3}$$

Once the contours are extracted, we propose a liner interpolation function to encode the spatial-temporal information to k channels. First, we use a function to map each frame in a single $1/4$ gait period (change from the individual standing with two legs overlapping to taking a step with two legs apart from each other extremely or just the reverse process) into $[0,1]$ by computing the degree of two legs apart from each other, which is explained in (1), to represent each frame's position in the time domain of each period:

$$r_t = (W_t - W_{\min})/(W_{\max} - W_{\min}), \tag{4}$$

where W_t is the average width of the leg region of the tth frame. $W_{\max}$ and $W_{\min}$ are the extreme widths of the $1/4$ period which the tth frame belongs to.

Then, we give the tth frame different weights $C_i(r_t)$ in different channels according to their position in the time domain defined above. When the number of channels $k=1$, the strategy is similar to GEI; each frame will have the same weight in the only one channel without considering their position in the time domain, i.e., $C_1(r_t)=1$, and the temporal information will not be preserved here.

When $k>1$, we can separate the whole $1/4$ period into $k-1$ equal parts by k separating points $p_i = i/(k-1)$, $i=0,1,\ldots,k-1$, and employ the ith channel to describe the temporal information in the $(i-1)$th and ith parts; the weight $C_i(r_k)$ can be defined as

$$C_i(r_t) = \begin{cases} \left(\dfrac{r_t - p_{i-2}}{p_{i-1} - p_{i-2}}\right) I & p_{i-2} < r_t \le p_{i-1}, \\ \left(1 - \dfrac{r_t - p_{i-1}}{p_i - p_{i-1}}\right) I & p_{i-1} < r_t \le p_i, \\ 0 & \text{others}, \end{cases} \tag{5}$$

where I is the maximum of intensity value, e.g., 255. Note that we need to introduce the 0th and the kth parts and two virtual separating points p_{-1} and p_k to make the definition of C_1 and C_k uniform. However, they are not needed in the real calculation. For visualization, we take $k=3$ and give different weights to each frame in the Red, Green, and Blue channels. That means we map the human's motion into the continuous variation in the RGB space:

$$B(r_t) = C_1(r_t) = \begin{cases} (1-2r_t)I & 0 \le r_t \le 1/2, \\ 0 & 1/2 < r_t \le 1, \end{cases} \tag{6}$$

$$G(r_t) = C_2(r_t) = \begin{cases} 2r_t I & 0 \le r_t \le 1/2, \\ (2-2r_t)I & 1/2 < r_t \le 1, \end{cases} \tag{7}$$

$$R(r_t) = C_3(r_t) = \begin{cases} 0 & 0 \le r_t \le 1/2, \\ (2r_t - 1)I & 1/2 < r_t \le 1. \end{cases} \tag{8}$$

3.3 Representation Construction

We calculate the multichannel gait contour image $\mathbf{C}_t$ of the tth frame in the gait sequence as

$$\mathbf{C}_t(x,y) = \begin{pmatrix} h'_t(x,y) * C_1(r_t) \\ h'_t(x,y) * C_2(r_t) \\ \vdots \\ h'_t(x,y) * C_k(r_t) \end{pmatrix}. \tag{9}$$

The equation indicates that unlike the usual binary boundary image, we needn't select some values in $h'(x,y)$ as the signal of a boundary. Given the gait contour images $\mathbf{C}_t$, a CGI temporal template $\mathbf{CGI}(x,y)$ is defined as follows:

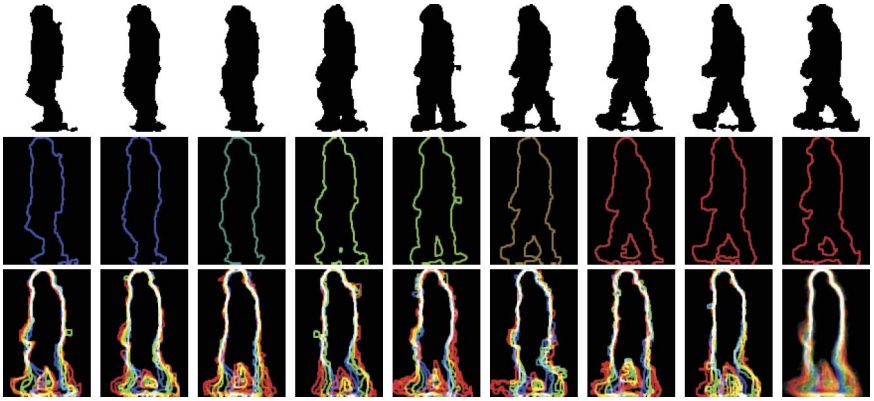


Fig. 3. An example of generating a CGI temporal template.

$$\mathbf{CGI}(x,y) = \frac{1}{p}\sum_{i=1}^{p}\mathbf{PGI}_i(x,y), \tag{10}$$

where p is the number of $1/4$ gait periods, and $\mathbf{PGI}_i(x,y) = \sum_{t=1}^{n_i}\mathbf{C}_t(x,y)$ is the sum of the total n_i multichannel contour images in the ith $1/4$ gait period. Note that we use the normal addition when we calculate the $\mathbf{CGI}(x,y)$, while we use the saturated addition (which mean the result will be I when the sum exceeds I) when we calculate the $\mathbf{PGI}(x,y)$.

The whole process to generate CGI is shown in Fig. 3. Here, we choose $k = 3$ because this setting maps the gait sequence into a RGB image which can be visualized well and thus help to illustrate the whole generating process better. The first row shows nine silhouettes in the first $1/4$ gait period. And the second row shows the corresponding colored gait contour images after edge detection and color mapping. Note that we have more than two values in each contour image. Then, we sum all nine of these images to obtain the first one $\mathbf{PGI}_1$ in the third row, representing this $1/4$ period. The second to the eighth images on the third row represent **PGI**s corresponding to other different $1/4$ periods in the same gait sequence. At last, we average all these frames to get the final CGI shown as the last one in the third row. It is not difficult to see that we obtain a better visualization result and a more informative gait template, which will be demonstrated in gait recognition experiments.

4 Human Recognition Using CGI

Now we can employ the proposed CGI temporal template for individual recognition by measuring the similarity between the gallery and probe templates. However, there are probably several disadvantages of doing so: 1) Since the gait sequences are sampled from similar physical conditions, the templates attained from such sequences may result in overfitting. 2) Due to the fact that the number of CGIs is small, it is a typical small sample size problem and thus cannot characterize the topology of essential gait space. 3) If we regard one pixel as one dimension, the dimensions of the original gait space are very high and the performance of gait recognition systems suffers from the problem of the curse of dimensionality. To solve these issues, we propose to generate CGI-based real templates and synthetic templates, projecting the templates into certain low-dimensional discrimination subspace with the dimension reduction method.

Specifically, we generate the real templates by referring to the multichannel image of each period as a temporal information preserving template. In other words, we average continuous four **PGI**s in one period. One advantage is that such a template keeps the similar gait temporal information as the CGI of the whole sequence owns. Furthermore, we generate synthetic templates to enhance the robustness to the exterior factors such as shadows. Similarly to Han and Bhanu [1], we cut the bottom $2 \times i$ rows from the CGI and resize to the original size using the nearest neighbor interpolation. If parameter i varies from 0 to $K-1$, then a total of K synthetic templates will be generated from each CGI template. Some examples of real and synthetic templates are shown in Fig. 4. For visualization, we also set $k = 3$ in these figures.

To address the curse of dimensionality issue without losing the computational efficiency, we employ Principal Component Analysis and Linear Discriminant Analysis (PCA+LDA) [25] to project the real and synthetic templates in the gallery set into a low-dimensional subspace. With the projection matrix calculated by PCA+LDA, the real/synthetic templates in the probe set will be projected into a low-dimensional subspace. Let $\hat{\mathcal{R}}_p$ and $\hat{\mathcal{S}}_p$ be the real and synthetic templates of the individual in probe sets, respectively, and let $\mathcal{R}_i$ and $\mathcal{S}_i$ be the real and synthetic templates of the ith individual in the gallery sets, respectively. In the subspace, the real/synthetic templates are recognized according to the minimal euclidean distances ($d(\hat{\mathcal{R}}_p, \mathcal{R}_j)$ or $d(\hat{\mathcal{S}}_p, \mathcal{S}_j)$) between the probe real/synthetic feature vectors

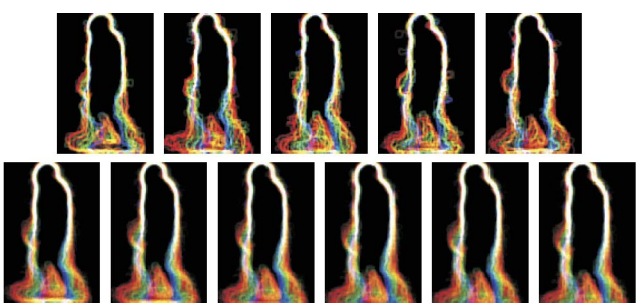


Fig. 4. Examples of real templates (top) and synthetic templates (bottom) for a gait sequence.

(a) USF HumanID Gait Database (b) CASIA Gait Database (c) Soton Large Gait Database

Fig. 5. Example images from the three benchmark databases used in our evaluation experiments [8], [26], [27].

to the class center of the gallery real/synthetic feature vectors. To further improve the performance, we propose to fuse the results of these two types of templates using the following equation:

$$d(\hat{\mathcal{R}}_p, \hat{\mathcal{S}}_p, \mathcal{R}_i, \mathcal{S}_i) = \frac{d(\hat{\mathcal{R}}_p, \mathcal{R}_i)}{\min_j d(\hat{\mathcal{R}}_p, \mathcal{R}_j)} + \frac{d(\hat{\mathcal{S}}_p, \mathcal{S}_i)}{\min_j d(\hat{\mathcal{S}}_p, \mathcal{S}_j)}, \quad i, j = 1, \ldots, C, \tag{11}$$

where C is the number of classes, i.e., the number of subjects here. We assign the probe template to the kth class if

$$k = \arg\min_i d(\hat{\mathcal{R}}_p, \hat{\mathcal{S}}_p, \mathcal{R}_i, \mathcal{S}_i), \quad i = 1, \ldots, C. \tag{12}$$

More details about real and synthetic templates can be referred in Han and Bhanu's work [1].

5 Evaluation Experiments

In this section, we will introduce our experiment setting, evaluate the performance of our CGI template by comparing with other state-of-the-art algorithms, study its robustness under different parameters and strategies, and discuss its pros and cons.

5.1 Experiment Settings

We evaluate the CGI algorithm on three benchmark databases, including the USF HumanID Gait Database (silhouette version 2.1) [8], CASIA Gait Database (Data set B) [26], and Soton Large Gait Database [27]. Some examples from these databases are illustrated in Fig. 5.

In the USF HumanID Gait Database (version 2.1), the gait sequences are sampled from 122 individuals walking in elliptical paths on concrete and grass surfaces, with/without a briefcase, wearing different shoes, and with different elapse time. By choosing the sequences with "Grass, Shoe Type A, Right Camera, No Briefcase, and Time t_1" for the gallery set, Sarkar et al. [8] developed 12 experiments, each of which is under a specific condition (Table 1). They also provide a manual silhouette version [28] in which some parts of body for each frame are labeled manually on a subset of the whole database.

The CASIA Gait Database (Data set B) consists of 124 individuals. For each individual, six gait sequences were captured under normal conditions, two sequences were captured when the people walking with a bag, and the other two sequences were captured when the people wearing a coat (named NM-01 to NM-06, BG-01, BG-02, CL-01, CL-02, respectively). Each gait sequence has 11 different view directions, from 0 to 180 degrees with 18 degrees between each two nearest view directions. Yu et al.'s work [26] give more details about this database. In our experiment, we only use the data captured from 90 degrees.

The Soton Large Gait Database consists of 115 individuals performing a normal walk. Each gait sequence was captured from the oblique view. This database includes gait sequences walking both from left to right and from right to left. For simplification, we flip some of them horizontally to make all the data have the same walking direction. More details about this database can be found in Shutler et al.'s work [27].

All three of these databases provide the silhouette benchmark images after background subtraction. Only the silhouette images of the USF HumanID Gait Database have been aligned already. Furthermore, we align these silhouettes by aligning their horizontal centroid and cut the silhouette images of CASIA Database and Soton Database into 160×100 and 140×91, respectively. All of our experiments are based on these aligned silhouette images.

In most of our experiments, we evaluate the performance of our algorithms based on the USF HumanID Gait Database. The reasons are that 1) this database is collected from an outdoor environment and thus is more challenging with respect to the number of subjects and the number of affecting factors, 2) it provides a preliminary experimental setting for gallery and probe sets, and most of the existing algorithms have used it for algorithm evaluation and comparison, and 3) the silhouette qualities in USF are of higher noise than those in SOTON and CASIA, and thus can be used to evaluate the algorithm robustness reasonably.

We evaluate the "Rank1" and "Rank5" performances of several recent approaches including baseline algorithm (based on silhouette shape matching) [8], GEI [1], HMM [29], IMED+LDA [11], 2DLDA [11], DATER [10], MTP [12], Tensor Locality Preserving Projections (TLPP) [30], and Dynamics-Normalized Gait Recognition Algorithm

TABLE 1
The Details of the Gallery and Probe Sets

Data Label	Data Set Size	Variances
Gallery	122	G, A, R, NB
Probe A	122	G, A, L, NB
Probe B	54	G, B, R, NB
Probe C	54	G, B, L, NB
Probe D	121	C, A, R, NB
Probe E	60	C, B, R, NB
Probe F	121	C, A, L, NB
Probe G	60	C, B, L, NB
Probe H	120	G, A, R, BF
Probe I	60	G, B, R, BF
Probe J	120	G, A, L, BF
Probe K	33	G, A/B, R, NB, T
Probe L	33	C, A/B, R, NB, T

Abbreviation note: G-Grass, C-Concrete, A-Shoe A, B-Shoe B, R-Right View, L-Left View, NB-No Briefcase, BF-Briefcase, T-Elapsed Time.

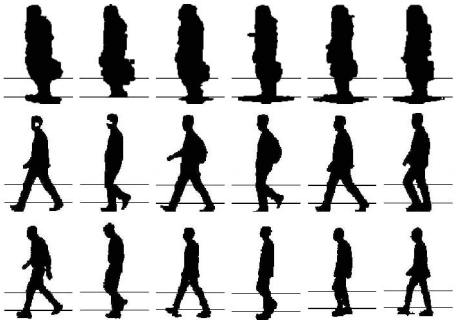


Fig. 6. Gait examples with two additional lines. The upper line is at αh and the lower one is at βh (h is the height of an individual). Figures in the first row, the second row, and the third row are collected from the USF HumanID Gait Database, CASIA Gait Database, and Soton Large Gait Database, respectively.

(DNGRA) [22]. The Rank1 performance means the percentage of the correct subjects ranked first, while the Rank5 performance means the percentage of the correct subjects appearing in any of the first five places in the rank list. We also report the average performance by computing the ratio of correctly recognized subjects to the total number of subjects.

In Section 3.1, we introduce two parameters α and β to find the leg region of an individual. From the experiment in Section 5.6, we can see that our proposed approach is robust to these two parameters. Thus, we only report the experimental results based on $\alpha = 23/32$ and $\beta = 29/32$ in other experiments. Note that we do not employ the anatomical results mentioned in Section 3.1 (the knee and ankle are 0.715 and 0.961 h, respectively) directly as we also need to use these two parameters to decrease the influence of some exterior factors such as briefcase and shadow. Some examples in these database are shown in Fig. 6. From the figure, we can see that most of the briefcase and bag is above the upper line and most of the shadow is under the lower line. Therefore, it means that the influence of briefcase, bag, and shadow can be effectively decreased.

In Section 3.2, we introduce a function to map a gait frame into different number of channels. In Section 5.5, we conduct an experiment to evaluate the influence of the number of channels k. We find that CGI achieves the best recognition performance with $k = 3$. Thus, in all the other experiments, we only report the results based on $k = 3$.

We also employ the fusion of real and synthetic templates introduced in Section 4 to further improve the performance. To make the experiment fair, we use the same strategy to generate real and synthetic templates of GEI and CGI, assigning the same parameters to PCA and LDA to reduce the data set into a subspace. The fusion results are obtained using the same formula (11).

5.2 The Effectiveness of CGI Template

To evaluate the performance of the proposed CGI temporal information preserving template, we employ a simple 1-nearest neighbor classifier (1-NN) on the original GEI and CGI without using real/synthetic templates and Principal Component Analysis/Linear Discriminant Analysis (PCA/LDA). We also provide the performance of the baseline algorithm [8] in the USF HumanID database for comparison. The results are summarized in Table 2. It can be seen from Table 2 that 1) CGI achieves the best average performance among all the algorithms. 2) CGI is very robust to the briefcase condition shown in Experiments H, I, and J. Specifically, the accuracy is improved by almost 20 percent compared with GEI. 3) Compared with GEI, CGI has better Rank5 performance than GEI in 9 out of 12 conditions, while in all the remaining three conditions, i.e., surface conditions, baseline algorithm provides better performance than both GEI and CGI. We can suggest that the gait templates are more sensitive to the surface condition than the baseline algorithm because of the shadows or some other factors.

TABLE 2
Comparison of Recognition Performance on the USF HumanID Database Using 1-NN

Exp.	Rank1 Performance (%)			Rank5 Performance (%)		
	baseline [8]	GEI	CGI	baseline	GEI	CGI
A	73	84	**87**	88	93	**96**
B	78	87	**94**	93	**94**	**94**
C	48	**72**	**72**	78	**93**	**93**
D	**32**	19	17	**66**	45	41
E	22	18	**25**	**55**	53	45
F	**17**	10	12	**42**	29	32
G	**17**	13	13	**38**	37	35
H	61	56	**78**	85	77	**91**
I	57	55	**80**	78	77	**97**
J	36	40	**54**	62	69	**82**
K	3	**9**	6	12	15	**30**
L	3	3	**9**	15	15	**27**
Avg.	40.96	41.13	**48.64**	64.54	61.38	**66.81**

To discover which components of the proposed CGI temporal templates are crucial to the performance of gait recognition, we compare the results of several variants of the contour-based temporal template with those of the silhouette-based template, which is employed by most of the gait recognition systems [10]. Here, GEI-contour means that we compute the GEI based on contour images, and CGI-gray means that we average each CGI into a grayscale image. Examples of GEI, GEI-contour, CGI-gray, and CGI are shown in Fig. 7.

To save space, we only show the fusion results in Table 3. From Table 3 it can be seen that 1) compared with GEI, GEI-contour and CGI obtain a remarkable improvement on Experiments H, I, J. Furthermore, CGI is slightly better than GEI-contour. It means the key to the improvement on briefcase condition is contour. One possible reason is that contour weakens the influence from regions inside the briefcase's silhouettes. 2) We also notice that CGI and GEI perform much better than GEI-contour on Experiments D,

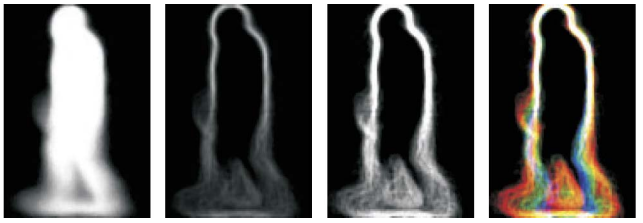


Fig. 7. From left to right: GEI, GEI-contour, CGI-gray, CGI obtained from the same gait sequence.

TABLE 3
Comparison of Recognition Performance of GEI, GEI-contour, CGI-gray, CGI on USF HumanID Database

Exp.	Rank1 Performance (%)				Rank5 Performance(%)			
	GEI	GEI-contour	CGI-gray	CGI	GEI	GEI-contour	CGI-gray	CGI
A	87	84	89	**91**	96	94	**99**	97
B	93	93	**94**	93	94	**96**	**96**	**96**
C	72	70	**81**	78	**94**	93	91	**94**
D	42	30	36	**51**	71	52	73	**77**
E	43	35	40	**53**	68	58	67	**77**
F	26	14	24	**35**	45	34	51	**56**
G	28	27	30	**38**	55	45	55	**58**
H	62	78	79	**84**	83	95	94	**98**
I	58	77	**78**	**78**	83	92	**97**	**97**
J	50	62	58	**64**	79	85	83	**86**
K	6	**12**	**12**	3	24	**33**	**33**	27
L	6	6	**9**	**9**	27	**33**	30	24
Avg.	51.57	52.30	55.95	**61.69**	72.55	70.56	76.93	**79.12**

E, F, G. It indicates that although contour instead of silhouette degenerates the recognition performances on surface condition, using the proposed CGI temporal information preserving template can make up for such loss and further improve the performance of gait recognition. 3) Compared with CGI-gray, CGI has better Rank1 performance in 9 out of 12 specific conditions and improves the average recognition ratio by about 5 percent. We can thus infer that with the proposed multichannel encoding technique, the temporal information of the gait sequence benefits gait recognition.

5.3 Environments and the Number of Training Data

We compare CGI and GEI on the CASIA database (Dataset B). As there are 10 gait sequences for each individual, we can adopt any one of them as the training data and generate one CGI template and one GEI template for each individual, and use the remaining nine gait sequences as testing data. Then, we employ a 1-NN classifier to identify each testing gait sequence based on the euclidean distance. Note that here we do not generate real and synthetic templates since: 1) There are only one or two complete gait periods in the CASIA gait sequence, thus generating real templates is meaningless. 2) The quality of background subtraction is quite good in the CASIA Gait Database and the shadows have been efficiently removed; therefore we do not need to generate synthetic templates here. Obviously, there are $10 \times 9 = 90$ different pairs of training data and testing data. To identify the influence of different environments, we categorize them into nine groups according to the sampling environments.

The experimental results are summarized in Table 4. The first column means different training environments and the first row means different testing environments.

TABLE 4
Under Different Training and Testing Environments, the Comparison of Rank1 Performance of GEI/CGI on the CASIA Database Using 1-NN (Percent)

	Normal	Bag	Coat
Normal	**91.57**/88.06	31.71/**43.67**	24.07/**42.98**
Bag	43.90/**60.19**	**91.20**/89.81	19.91/**37.27**
Coat	26.00/**50.77**	15.05/**27.78**	**97.22**/95.37

The recognition rate in each cell is the average of all the experiments belonging to this group. For example, there are 12 experiments belonging to the case where the training environment is normal condition and the testing environment is walking with a bag. It can be seen from Table 4 that 1) when we focus on the three numbers on the diagonal, we can find that GEI has better performance than CGI in all the three groups. It means that when the training and testing environments are the same, the performance of GEI is slightly better than that of CGI. 2) CGI wins in all six groups left and improves the accuracy by more than 15 percent on average. That means CGI performs better when the training environment and the testing environment are different. Therefore, we can further validate the conclusion drawn from the experiments on the USF HumanID database, i.e., CGI is more robust than GEI for the external environment.

We can observe one interesting phenomenon from the previous experiment, i.e., under a different training and test environment, CGI makes a significant improvement compared with GEI due to its robustness to external environment. However, under the same environment, CGI performed slightly worse than GEI. One possible reason is the insufficient number of training samples. Specifically, in the CASIA database there are only one or two complete gait periods in one gait sequence, while in the USF database there are more than five complete gait periods in one gait sequence. Considering the fact that the GEI template is the average of all the frames in one gait sequence while CGI template is the average of PGIs representing each $1/4$ gait period defined in Section 3.2. Therefore, CGI may need more training data than GEI. That may be one limitation of CGI.

To further validate this hypothesis, we evaluate it on the USF Database (Manual Silhouettes Version) [28], which contains only one gait period in each gait sequence. We found that CGI cannot perform as well as GEI in this situation. CGI losses 4.17 and 0.93 percent on Rank1 and Rank5 performance, respectively, compared with GEI. It justifies the limitation of CGI again when only one period exists in a gait sequence.

Furthermore, we carry out another experiment to study on the relationship between the number of training data and the recognition rate. We divide the CASIA database

TABLE 5
Comparison of Recognition Performance of GEI and CGI Using Different Numbers of Training Data on the CASIA Database

K	1	2	3	4	5
NM-06 Performance (%)					
GEI Rank1	**86.11**	**95.37**	**97.22**	97.22	98.15
CGI Rank1	84.26	**95.37**	96.30	**99.07**	**100.00**
GEI Rank5	**97.22**	97.22	98.15	98.15	**100.00**
CGI Rank5	**97.22**	**98.15**	**99.07**	**100.00**	**100.00**
BG-01 Performance (%)					
GEI Rank1	42.59	45.37	46.30	47.22	48.15
CGI Rank1	**49.07**	**62.96**	**62.04**	**63.89**	**68.52**
GEI Rank5	70.37	75.00	77.78	76.85	76.85
CGI Rank5	**75.93**	**85.19**	**90.74**	**90.74**	**90.74**
BG-02 Performance (%)					
GEI Rank1	47.22	46.20	52.78	50.93	51.85
CGI Rank1	**56.48**	**68.52**	**64.81**	**73.15**	**75.00**
GEI Rank5	75.00	79.63	76.85	76.85	75.93
CGI Rank5	**80.56**	**87.04**	**88.89**	**89.81**	**89.81**
CL-01 Performance (%)					
GEI Rank1	22.22	26.85	28.70	28.70	27.78
CGI Rank1	**43.52**	**46.30**	**48.15**	**45.37**	**49.07**
GEI Rank5	51.85	56.48	56.48	58.33	60.19
CGI Rank5	**61.11**	**66.67**	**70.37**	**71.30**	**72.22**
CL-02 Performance (%)					
GEI Rank1	20.37	25.93	25.93	25.93	25.93
CGI Rank1	**39.81**	**41.67**	**42.59**	**39.81**	**44.44**
GEI Rank5	51.85	55.56	61.11	57.41	61.11
CGI Rank5	**62.04**	**68.52**	**71.30**	**73.15**	**72.22**

Abbreviation note: NM-Normal, BG-Bag, CL-Cloth.

into two parts: the first five sequences captured under normal conditions and the other five gait sequences. To control the number of training data, we introduce a parameter $K = 1, 2, 3, 4, 5$, which means we use the first K gait sequences in the first part as training data. The gait template for one individual is the average of all the templates representing that the gait sequence belongs to this individual in the training data. And we use the second part as the testing data. The experimental results are illustrated in Table 5.

From Table 5, we can find that:

1. Both GEI and CGI perform better when they have more training data (in most situations).
2. With the increment of K, the recognition rate of CGI grows faster than that of GEI.
3. When the training and testing environments are the same (NM-06), the recognition rate of CGI can finally exceed that of GEI with the increment of K.
4. The performance on bag and cloth conditions can also benefit from the increment of training data under different conditions (normal condition here).

And the improvement of CGI is larger than that of GEI. Therefore, the robustness of CGI to external environment can cover the limitation to some degree since even if we cannot get enough data under one particular condition, we can use the data captured under other conditions to achieve a better CGI template.

In addition, we study the relationship between the number of training data and recognition rate on the Soton Large Gait Database. Making this experiment on this database has the following advantages: 1) Similarly to CASIA, most gait sequences in this database consist of one or two complete gait periods; 2) there are enough gait sequences for the same individual under the same environment, that is to say, there are 2,162 gait sequences under this condition so each individual has more than six gait sequences; 3) the silhouette qualities of the Soton Database are pretty high.

We choose K gait sequences randomly as training data and use the remaining sequences as testing data. The experimental results shown in Table 6 are the average of 20 repetitions. From Table 6 we can see that both CGI and GEI can benefit from the increment of training data. And CGI achieves better recognition performance. Since Soton Database provides high-quality gait frames, each of which is sampled under the same environment, it suggests that the limitation of CGI may be alleviated through improving the sampling qualities to reduce the noise.

5.4 Comparison with Other Published Algorithms

Finally, we compare our proposed algorithm with several recently published results on the USF HumanID database. The results are listed as in Table 7. We can see that CGI outperforms most of them on average Rank1/Rank5 performances, and is robust under most of the complex conditions. Note that we also achieve better results compared with our previous work [16] (61.69 percent versus 60.54 percent) because of the improvement on the CGI algorithm in this paper.

It is also worth mentioning that the time and space complexities of CGI are quite small. Furthermore, it can achieve competitive performance with some sophisticated approaches, e.g., pHMM employed by Liu and Sarkar [22] (DNGRA). Specifically, the time complexity of generating all CGI templates for each training and test data is $\Theta(N_{tr}TWHk + N_{te}TWHk)$, whereas pHMM takes $\Theta(N_{tr}TWHI_{Kmeans} + N_{tr}TWHN_s^2I_{pHMM} + N_{te}TWHN_s^2)$ to generate the dynamics-normalized stance-frames for each training and test data. Here, N_{tr} and N_{te} are the number of gait sequences in training data and test data, respectively. T means the average number of frames in each gait sequence. W and H are the width and height of each frame, respectively. k denotes the number of channels in CGI. N_s is the number of states in the pHMM model in DNGRA, while I_{Kmeans} and I_{pHMM} are the numbers of iteration for K-means clustering and pHMM training, respectively. Let $S_{tr} = N_{tr}TWH$ and

TABLE 6
Performance Comparison (Percent) of GEI and CGI Using Different Numbers of Training Data on the Soton Large Database

K	1	2	3	4	5	6
GEI Rank1	75.14±1.53	84.89±1.22	87.70±0.89	89.39±0.71	90.28±0.42	90.78±0.49
CGI Rank1	**80.90±1.24**	**89.37±1.11**	**91.99±0.88**	**93.23±0.60**	**93.68±0.46**	**94.00±0.47**
GEI Rank5	87.99±1.54	92.67±1.25	94.42±0.74	95.03±0.90	95.18±1.03	95.23±0.64
CGI Rank5	**91.88±0.70**	**95.03±1.05**	**96.25±0.78**	**96.72±0.88**	**96.76±1.01**	**96.78±0.73**

TABLE 7
Comparison of Recognition Performance (Percent) on the USF HumanID Database Using Different Methods

	A	B	C	D	E	F	G	H	I	J	K	L	Avg.
Rank1 Performance													
Baseline [8]	73	78	48	32	22	17	17	61	57	36	3	3	40.96
HMM [29]	89	88	68	35	28	15	21	85	80	58	17	15	53.54
IMED+LDA [11]	88	86	72	29	33	23	32	54	62	52	8	13	48.64
2DLDA [11]	89	93	80	28	33	17	19	74	71	49	16	16	50.98
GTDA (M&H) [11]	86	88	73	24	25	17	16	53	49	45	10	7	43.70
Gabor+GTDA (H) [11]	84	86	73	31	30	16	18	85	85	57	13	10	52.51
DATER [10]	87	93	78	42	42	23	28	80	79	59	18	21	56.99
TLPP [30]	87	93	72	25	35	17	18	62	62	43	12	15	46.95
MTP [12]	90	91	83	37	43	23	25	56	59	59	9	6	51.57
DNGRA [22]	85	89	72	57	66	46	41	83	79	52	15	24	62.94
GEI+Real [1]	89	87	78	36	38	20	28	62	59	59	3	6	51.04
GEI+Synthetic [1]	84	93	67	53	45	30	34	48	57	39	21	24	51.04
GEI+Fusion [1]	90	91	81	56	64	25	36	64	60	60	6	15	57.72
CGI+Real	90	91	80	33	33	18	22	84	80	61	3	6	54.49
CGI+Synthetic	86	89	67	54	62	33	33	57	58	52	0	9	54.28
CGI+Fusion (our method)	91	93	78	51	53	35	38	84	78	64	3	9	61.69
Rank5 Performance													
Baseline [8]	88	93	78	66	55	42	38	85	78	62	12	15	64.54
HMM [29]	–	–	–	–	–	–	–	–	–	–	–	–	–
IMED+LDA [11]	95	95	90	52	63	42	47	86	86	78	21	19	68.60
2DLDA [11]	97	93	93	57	59	39	47	91	94	75	37	34	70.95
GTDA (M&H) [11]	100	97	95	57	54	34	45	75	80	70	25	25	66.15
Gabor+GTDA (H) [11]	96	95	89	59	63	33	49	94	92	76	19	40	70.32
DATER [10]	96	96	93	69	69	51	52	92	90	83	40	36	75.68
TLPP [30]	94	94	87	52	55	35	42	85	78	68	24	33	65.18
MTP [12]	94	93	91	64	68	51	52	88	83	82	18	15	71.38
DNGRA [22]	96	94	89	85	81	68	69	96	95	79	46	39	82.05
GEI+Real [1]	93	93	89	65	60	42	45	88	79	80	6	9	68.68
GEI+Synthetic [1]	93	96	93	75	71	54	53	78	82	64	33	42	72.13
GEI+Fusion [1]	94	94	93	78	81	56	53	90	83	82	27	21	76.30
CGI+Real	95	94	93	66	65	48	52	96	97	87	24	27	75.05
CGI+Synthetic	94	96	83	76	75	53	57	92	85	78	27	21	74.95
CGI+Fusion (our method)	97	96	94	77	77	56	58	98	97	86	27	24	79.12

$S_{te} = N_{te}TWH$ be the sizes of training data and test data, respectively. We can rewrite the time complexity of CGI and DNGRA into

$$\Theta(kS_{tr} + kS_{te}), \Theta((I_{Kmeans} + N_s^2 I_{pHMM})S_{tr} + N_s^2 S_{te}).$$

Note that we often choose $k = 3$ in CGI, while $N_s = 20$, $I_{Kmeans} > 10$ in DNGRA [22] (note that I_{pHMM} is not mentioned in [22]). It is obvious that CGI is much faster than DNGRA, while the recognition performances provided by these two algorithms are competitive. In fact, CGI can process more than 50 frames each second,[1] making CGI quite competitive in real-time scenarios.

5.5 Effects of the Number of Channels

In Section 3.2, we introduced a function to map a gait frame into different number of channels. Here, we perform an experiment to evaluate the influence of the number of channels k on the USF Database. In this experiment, we generate real and synthetic CGI templates with the fusion strategy introduced in Section 4. The reported results are the average recognition performance on 12 probe sets here.

The experiment result is illustrated in Fig. 8. We can see that both Rank1 and Rank5 Performance rise in the beginning and then drop down rapidly with the increment of the number of channels k. Both Rank1 and Rank5 performance achieves the best performance when $k = 3$. One plausible interpretation is that when k is small, e.g., 1 or 2, the temporal information is less preserved in the CGI template and thus leads to a worse performance. On the other hand, considering that there are nine frames on average in a $1/4$ gait period in USF database and each channel will only employ a small part of these frames to encode temporal information (5). When we use four or more channels to generate a PGI, each channel may get two or less frames. Consequently, the performance is impaired by the insufficiency of training samples.

Similarly to the USF Database, CASIA Database and Soton Database also have nine frames in a $1/4$ gait period on average. Therefore, we perform our proposed multi-

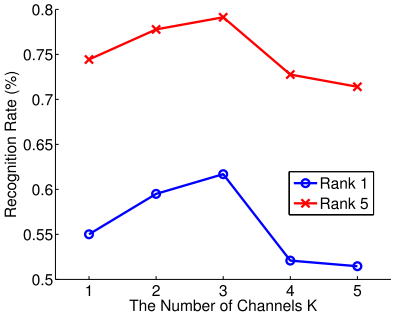


Fig. 8. Recognition performance using the CGI template with different numbers of channels.

1. We run a Matlab code on a machine with an Intel Core2 Duo CPU T9600 2.80 GHZ and 3 GB of DDR3 memory.

TABLE 8
Comparison of Average Rank1/Rank5 Performance of CGI Using Different α and β Values on the USF Database

α \ β	28/32	29/32	30/32
22/32	60.13/79.23	61.69/79.12	61.06/79.23
23/32	60.65/79.12	61.69/79.12	60.33/79.02
24/32	61.38/79.44	60.13/78.29	60.44/78.60
25/32	61.59/79.54	60.75/78.39	60.02/78.91

channel technique with the number of channels $k = 3$ in the other experiments since it makes a good tradeoff between preserving the temporal information in different phases of a gait period and demand on more training samples assigning to each channel.

5.6 The Robustness of α, β and Edge Detection

In this experiment, we investigate the influence of α and β on the USF Database. We manually tuned α and β around the relative vertical position of knee and ankle (0.715 and 0.961 h). Specifically, we tuned α in $\{22/32, 23/32, 24/32, 25/32\}$ and β in $\{28/32, 29/32, 30/32\}$ (considering the influence of shadow, the value of β will be less than 0.961). The result is reported in Table 8. We can see that the differences of recognition rate with different α and β values are quite subtle. The reason for the robustness is that the selection of α and β only affects the calculation of the average width of leg region W, but such an influence would not degenerate the performance of two crucial steps in CGI, i.e., period detection and calculating the mapping function shown in (4). First, we only use the local minimum and maximum of W to find some key frames and estimate the gait period. Second, we use the relative ratio of W in one period in the multichannel mapping (4). These two factors always depend on the relative variances of W rather than a specific value of W. Consequently, it leads to the conclusion that CGI is robust to α and β in some degree. With this good property, we can safely set α and β as some typical values when evaluating our method.

In our proposed approach, we employ local information entropy to extract the contour of the silhouette image. However, there exist many edge detection techniques, such as LoG operator, Sober operator, Canny operator, etc. Furthermore, for a binary image, there are several simple techniques, e.g., the morphology approach (the difference between the original image and the image after erosion by a 3×3 square mask) and extracting the perimeter pixels of the silhouette (function "bwperim" in matlab), etc. In this experiment, we compare the performance of CGI using different edge detection techniques. The experimental results are illustrated in Table 9.

From Table 9 we can see that the performance using local information entropy is significantly better than the others. One possible reason is that other edge detection techniques only provide a binary result of each pixel to be either edge or nonedge, while local information entropy provides more information about the degree of each pixel to be a contour pixel of the silhouette. It is also worth noting that local information entropy does not bring additional computing time complexity compared with the other edge detection techniques. Therefore, we employ local information entropy as the edge detection technique in CGI.

TABLE 9
Comparison of Average Rank1 and Rank5 Performances of CGI Using Different Edge Detection Techniques on USF Database

Edge Detection Technique	Rank1	Rank5
LoG Operator	54.18	73.38
Sobel Operator	52.92	73.17
Canny Operator	49.90	71.09
Morphology Approach	54.70	75.89
Bwperim	54.49	73.38
Local Information Entropy	**61.69**	**79.12**

5.7 Period Detection and Fusion Functions

With the same parameter settings, we also investigate the influence of gait period detection, as tabulated in Table 10. We observe from Table 10 that the divergence between the two detection methods is minor in almost all the experiments expect for a few groups of experiments highlighted in the table. One reason is that GEI, which uses the arithmetic average to generate the gait energy image, is insensitive to key frame selection and period detection. At the same time, this experiment indicates that our method is robust to the period detection, and it can work well using a basic period detection method. What's more, it may work better if employing an advanced period detection method. Furthermore, CGI performs better than GEI when using both period detection methods.

In Section 4, furthermore, we mentioned that we use different fusion functions from those used in Han and Bhanu's work [1]. The fusion function in [1] is

$$d(\hat{\mathcal{R}}_p, \hat{\mathcal{S}}_p, \mathcal{R}_i, \mathcal{S}_i) = \frac{c(c-1)d(\hat{\mathcal{R}}_p, \mathcal{R}_i)}{2\sum_{i=1}^{c}\sum_{j=1,j\neq i}^{c} d(\hat{\mathcal{R}}_p, \mathcal{R}_j)} + \frac{c(c-1)d(\hat{\mathcal{S}}_p, \mathcal{S}_i)}{2\sum_{i=1}^{c}\sum_{j=1,j\neq i}^{c} d(\hat{\mathcal{S}}_p, \mathcal{S}_j)},$$

where $d(\hat{\mathcal{R}}_p, \mathcal{R}_j)$ and $d(\hat{\mathcal{S}}_p, \mathcal{S}_j)$ are the distance between two templates, c is the number of classes, i.e., the number of subjects here.

In this experiment, we use two different fusion functions and keep the other experimental conditions the same. The results shown in Table 10 indicate that 1) the differences between the proposed fusion criterion and theirs are quite subtle with respect to recognition accuracy, and 2) when using F1, both the Rank1 and Rank5 average recognition rates are slightly better than using F2 for GEI and CGI.

6 Conclusion

In this paper, we have proposed a simple but effective temporal information preserving template CGI for gait recognition. We extract a set of contour images from the corresponding silhouette images using the local entropy principle, and encode the temporal information of gait sequence into the CGI using the multichannel technique. We also generate CGI-based real and synthetic temporal templates and exploit the fusion strategy to obtain better performance. Experiments on three benchmark databases have demonstrated that compared with state-of-the-art algorithms, our CGI template can attain higher or comparable recognition accuracy with good robustness and efficiency.

In the future, we will explore how to enhance CGI's robustness in more complex conditions, and investigate

TABLE 10
Comparison of the Recognition Performances of GEI, CGI Using Different Settings on the USF HumanID Database

Exp.	Rank 1 Performance (%)				Rank 5 Performance (%)			
	Two Period Detection Methods							
	GEI+C	GEI+W	CGI+C	CGI+W	GEI+C	GEI+W	CGI+C	CGI+W
A	87	87	92	91	96	95	97	97
B	93	91	93	93	94	96	94	96
C	72	74	78	78	94	94	94	94
D	42	43	46	51	71	70	77	77
E	43	42	52	53	68	68	78	77
F	26	26	**29**	**35**	45	46	55	56
G	28	28	37	38	55	53	57	58
H	62	60	79	84	83	82	96	98
I	58	58	**70**	**78**	83	83	97	97
J	50	48	58	64	79	78	88	86
K	**6**	**12**	0	3	24	21	21	27
L	6	6	9	9	27	27	18	24
Avg.	51.57	51.25	58.25	61.69	72.55	72.13	78.50	79.12
	Two Fusion Functions							
	GEI+F1	GEI+F2	CGI+F1	CGI+F2	GEI+F1	GEI+F2	CGI+F1	CGI+F2
A	87	86	91	91	96	96	97	96
B	93	94	93	93	94	96	96	96
C	72	74	78	76	94	93	94	94
D	42	36	51	50	71	70	77	76
E	43	40	53	50	68	65	77	77
F	26	21	35	32	45	40	56	55
G	28	27	38	32	55	52	58	57
H	62	62	84	83	83	87	98	98
I	58	62	78	80	83	85	97	98
J	50	51	64	62	79	78	86	88
K	6	6	3	3	24	24	27	24
L	6	3	9	9	27	15	24	27
Avg.	**51.57**	50.21	**61.69**	60.13	**72.55**	71.40	**79.12**	79.02

Abbreviation note: We refer our fusion function as "F1" and the function proposed by Han and Bhanu [1] as "F2," and refer the period detection method proposed in Sarkar et al. [8] as "C" and our method proposed in Section 3.1 as "W."

how to select a more general multichannel mapping function instead of the current linear mapping function. In addition, we will study how to make CGI effective when the gait sequence only contains few gait periods. We will also consider generalizing the proposed frameworks into other human-movement-related fields [31], [32] such as gesture recognition and abnormal behavior detection.

ACKNOWLEDGMENTS

This work was supported in part by the NSFC (No. 60975044, 60903062, 61175003), 973 program (No. 2010CB327900), Shanghai Leading Academic Discipline Project No. B114 and Hui-Chun Chin, and Tsung-Dao Lee Chinese Undergraduate Research Endowment (CURE).

REFERENCES

[1] J. Han and B. Bhanu, "Individual Recognition Using Gait Energy Image," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 28, no. 2, pp. 316-322, Feb. 2006.

[2] J. Zhang, J. Pu, C. Chen, and R. Fleischer, "Low-Resolution Gait Recognition," *IEEE Trans. Systems, Man, and Cybernetics, Part B,* vol. 40, no. 4, pp. 986-996, Aug. 2010.

[3] C.Y. Yam and M.S. Nixon, "Model-Based Gait Recognition," *Encyclopedia of Biometrics,* pp. 633-639, Springer, 2009.

[4] I. Bouchrika and M.S. Nixon, "Model-Based Feature Extraction for Gait Analysis and Recognition," *Proc. Third Int'l Conf. Computer Vision/Computer Graphics Collaboration Techniques and Applications,* pp. 150-160, 2007.

[5] L. Wang, T.N. Tan, H.Z. Ning, and W.M. Hu, "Silhouette Analysis Based Gait Recognition for Human Identification," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 12, pp. 1505-1518, Dec. 2003.

[6] A. Sundaresan, A. Roy-Chowdhury, and R. Chellappa, "A Hidden Markov Model Based Framework for Recognition of Humans from Gait Sequences," *Proc. Int'l Conf. Image Processing,* vol. 2, pp. 93-96, 2003.

[7] T. Kobayashi and N. Otsu, "Action and Simultaneous Multiple-Person Identification Using Cubic Higher-Order Local Auto-Correlation," *Proc. Int'l Conf. Pattern Recognition,* vol. 4, pp. 741-744, 2004.

[8] 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 Analysis and Machine Intelligence,* vol. 27, no. 2, pp. 162-177, Feb. 2005.

[9] Z. Liu and S. Sarkar, "Simplest Representation Yet for Gait Recognition: Averaged Silhouette," *Proc. Int'l Conf. Pattern Recognition,* vol. 4, pp. 211-214, 2004.

[10] D. Xu, S. Yan, D. Tao, L. Zhang, X. Li, and H.-J. Zhang, "Human Gait Recognition with Matrix Representation," *IEEE Trans. Circuits and Systems for Video Technology,* vol. 16, no. 7, pp. 896-903, July 2006.

[11] D. Tao, X. Li, X. Wu, and S.J. Maybank, "General Tensor Discriminant Analysis and Gabor Features for Gait Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 29, no. 10, pp. 1700-1715, Oct. 2007.

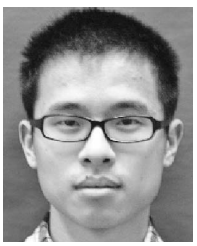
[12] C. Chen, J. Zhang, and R. Fleischer, "Multilinear Tensor-Based Non-Parametric Dimension Reduction for Gait Recognition," *Proc. Int'l Conf. Biometrics,* pp. 1037-1046, 2009.

[13] J. Woodring and H.W. Shen, "Chronovolumes: A Direct Rendering Technique for Visualizing Time-Varying Data," *Proc. Eurographics/IEEE TVCG Workshop Vol. Graphics,* pp. 27-34, 2003.

[14] H. Jänicke, A. Wiebel, G. Scheuermann, and W. Kollmann, "Multifield Visualization Using Local Statistical Complexity," *IEEE Trans. Visualization and Computer Graphics,* vol. 13, no. 6, pp. 1384-1391, Nov./Dec. 2007.

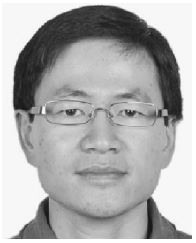
[15] C. Wang, H. Yu, and K.L. Ma, "Importance-Driven Time-Varying Data Visualization," *IEEE Trans. Visualization and Computer Graphics,* vol. 4, no. 6, pp. 1547-1554, Nov./Dec. 2008.

[16] C. Wang, J. Zhang, J. Pu, X. Yuan, and L. Wang, "Chrono-Gait Image: A Novel Temporal Template for Gait Recognition," *Proc. European Conf. Computer Vision,* pp. 257-270, 2010.
[17] L. Wang, T.N. Tan, W.M. Hu, and H.Z. Ning, "Automatic Gait Recognition Based on Statistical Shape Analysis," *IEEE Trans. Image Processing,* vol. 12, no. 9, pp. 1120-1131, Sept. 2003.
[18] G.V. Veres, L. Gordon, J.N. Carter, and M.S. Nixon, "What Image Information Is Important in Silhouette-Based Gait Recognition?" *Computer Vision and Pattern Recognition,* vol. 2, pp. 776-782, 2004.
[19] B. Guo and M.S. Nixon, "Gait Feature Subset Selection by Mutual Information," *IEEE Trans. Systems, Man and Cybernetics, Part A,* vol. 39, no. 1, pp. 36-46, Jan. 2009.
[20] L. Wang, T. Tan, H. Ning, and W. Hu, "Fusion of Static and Dynamic Body Biometrics for Gait Recognition," *IEEE Trans. Circuits and Systems for Video Technology,* vol. 14, no. 2, pp. 149-158, Feb. 2004.
[21] Y. Chai, Q. Wang, J.P. Jia, and R. Zhao, "A Novel Human Gait Recognition Method by Segmenting and Extracting the Region Variance Feature," *Proc. Int'l Conf. Pattern Recognition,* vol. 4, pp. 425-428, 2006.
[22] Z. Liu and S. Sarkar, "Improved Gait Recognition by Gait Dynamics Normalization," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 28, no. 6, pp. 863-876, June 2006.
[23] D. Winter, *Biomechanics and Motor Control of Human Movement.* John Wiley & Sons, Inc., 2009.
[24] C. Yan, N. Sang, and T. Zhang, "Local Entropy-Based Transition Region Extraction and Thresholding," *Pattern Recognition Letters,* vol. 24, no. 16, pp. 2935-2941, 2003.
[25] D.L. Swets and J. Weng, "Using Discriminant Eigenfeatures for Image Retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, no. 8, pp. 831-836, Aug. 1996.
[26] S. Yu, D. Tan, and T. Tan, "A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition," *Proc. 18th Int'l Conf. Pattern Recognition,* vol. 4, pp. 441-444, 2006.
[27] J. Shutler, M. Grant, M. Nixon, and J. Carter, "On a Large Sequence-Based Human Gait Database," *Proc. Fourth Int'l Conf. Recent Advances in Soft Computing,* pp. 66-72, 2002.
[28] Z. Liu, L. Malave, A. Osuntogun, P. Sudhakar, and S. Sarkar, "Toward Understanding the Limits of Gait Recognition," *Soc. of Photo-Optical Instrumentation Engineers,* vol. 5404, pp. 195-205, 2004.
[29] A. Kale, A. Sundaresan, A.N. Rajagopalan, N.P. Cuntoor, A.K. RoyChowdhury, V. Kruger, and R. Chellappa, "Identification of Humans Using Gait," *IEEE Trans. Image Processing,* vol. 13, no. 9, pp. 1163-1173, Sept. 2004.
[30] C. Chen, J. Zhang, and R. Fleischer, "Distance Approximating Dimension Reduction of Riemannian Manifolds," *IEEE Trans. Systems, Man, and Cybernetics, Part B,* vol. 40, no. 1, pp. 208-217, Feb. 2010.
[31] A.F. Bobick and J.W. Davis, "The Recognition of Human Movement Using Temporal Templates," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 3, pp. 257-267, Mar. 2001.
[32] T.B. Moeslund, A. Hilton, and V. Krüger, "A Survey of Advances in Vision-Based Human Motion Capture and Analysis," *Computer Vision and Image Understanding,* vol. 103, nos. 2/3, pp. 90-126, 2006.



Chen Wang is currently an undergraduate student in the School of Computer Science, Fudan University, Shanghai, China and will receive the BS degree in 2012. His current research interests include statistical learning and its applications in computer vision, web search, and web data mining.

Junping Zhang received the BS degree in automation from Xiangtan University, Xiangtan, China, in 1992, the MS degree in control theory and control engineering from Hunan University, Changsha, China, in 2000, and the PhD degree in intelligent systems and pattern recognition from the Institution of Automation, Chinese Academy of Sciences, in 2003. He has been an associate professor in the School of Computer Science, Fudan University since 2006. His research interests include machine learning, image processing, biometric authentication, and intelligent transportation systems. He has been an associate editor of *IEEE Intelligent Systems* since 2009. He has been an associate editor of the *IEEE Transactions on Intelligent Transportation Systems* since 2010. He is a member of the IEEE and the IEEE Computer Society.



Liang Wang received both the BEng and MEng degrees from Anhui University in 1997 and 2000, respectively, and the PhD degree from the Institute of Automation, Chinese Academy of Sciences (CAS) in 2004. From 2004 to 2010, he worked as a research assistant at Imperial College London, United Kingdom, and Monash University, Australia, a research fellow at the University of Melbourne, Australia, and a lecturer at the University of Bath, United Kingdom, respectively. Currently, he is a professor in the Hundred Talents Program at the National Lab of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, P.R. China. His major research interests include machine learning, pattern recognition, and computer vision. He has been widely published in highly ranked international journals such as the *IEEE Transactions on Pattern Analysis and Machine Intelligence and IEEE Transactions on Information Processing*, and leading international conferences such as CVPR, ICCV, and ICDM. He has received several honors and awards such as the Special Prize of the Presidential Scholarship of Chinese Academy of Sciences. He is an associate editor of the *IEEE Transactions on Systems, Man, and Cybernetics-Part B*, the *International Journal of Image and Graphics*, *Signal Processing*, *Neurocomputing*, and the *International Journal of Cognitive Biometrics*. He was a guest editor of four special issues, is a coeditor of five edited books, and was a cochair of six international workshops. He is currently a senior member of the IEEE, as well as a member of BMVA.

Jian Pu received the bachelor's degree from Beijing University of Chemical Technology, Beijing, China, in 2005 and is currently working toward the PhD degree in the School of Computer Science, Fudan University, Shanghai, China. His research interests include sparse coding and gait recognition.

Xiaoru Yuan received the PhD degree in computer science in 2006, from the University of Minnesota at Twin Cities. He is a faculty member in the School of Electronics Engineering and Computer Science at Peking University from January 2008. His primary research interests fall in the field of visualization with emphasis on information visualization, visual analytics, high performance rendering and visualization for massive data sets, and novel visualization user interface. He is a member of the IEEE.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.