Chrono-Gait Image: A Novel Temporal Template for Gait Recognition

Chen Wang[1], Junping Zhang[1,⋆], Jian Pu[1],
Xiaoru Yuan[2], and Liang Wang[3,4]

[1] Shanghai Key Lab of Intelligent Information Processing
School of Computer Science, Fudan University, China
[2] Key Laboratory of Machine Perception (Ministry of Education)
School of Electronics Engineering and Computer Science
Peking University, Beijing 100871, China
[3] Department of Computer Science, University of Bath, BA2 7AY, United Kingdom
[4] National Laboratory of Pattern Recognition, Institute of Automation,
Chinese Academy of Sciences, China
chen.wang0517@gmail.com, jpzhang@fudan.edu.cn, mydaiyu@hotmail.com,
xiaoru.yuan@pku.edu.cn, lw356@cs.bath.ac.uk

Abstract. In this paper, we propose a novel temporal template, called Chrono-Gait Image (CGI), to describe the spatio-temporal walking pattern for human identification by gait. The CGI temporal template encodes the temporal information among gait frames via color mapping to improve the recognition performance. Our method starts with the extraction of the contour in each gait image, followed by utilizing a color mapping function to encode each of gait contour images in the same gait sequence and compositing them to a single CGI. We also obtain the CGI-based real templates by generating CGI for each period of one gait sequence and utilize contour distortion to generate the CGI-based synthetic templates. In addition to independent recognition using either of individual templates, we combine the real and synthetic temporal templates for refining the performance of human recognition. Extensive experiments on the USF HumanID database indicate that compared with the recently published gait recognition approaches, our CGI-based approach attains better performance in gait recognition with considerable robustness to gait period detection.

1 Introduction

Biometric authentication is useful in many applications such as social security, individual identification in law enforcement and access control in surveillance. Compared with other biometric features such as face, iris and fingerprint, the advantages of gait include: 1) the acquisition of gait data is non-contactable, non-invasive, and hidden; 2) gait is the only perceptible at a distance. However, the performance of gait recognition suffers from some exterior factors such as

⋆ Corresponding author.

K. Daniilidis, P. Maragos, N. Paragios (Eds.): ECCV 2010, Part I, LNCS 6311, pp. 257–270, 2010.

shoes, briefcases, clothing, and environmental context. Moreover, it depends on whether the spatio-temporal relationship between gait frames in a gait sequence can be effectively represented. Although it is a challenging task, the nature of gait indicates that it is a irreplaceable biometric [1] and can benefit the remote biometric authentication.

1.1 Related Work

The extraction of gait features plays a crucial role in improving the performance of gait recognition. There are two main extraction methods: model-based and model-free approaches. Model-based approaches recover the underlying mathematical construction of gait with a structure/motion model [2]. Wang et al. adopted procrustes analysis to capture the mean shapes of the gait silhouettes [3]. However, procrustes analysis is time consuming and vulnerable to noise. Veres et al. [4] and Guo and Nixon [5] 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 [6]. Wang et al. [7] combined structural-based and motion-based models by employing a condensation framework to refine the feature extraction. 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 [2,6]. Furthermore, it is necessary to understand the constraints of gait such as the dependency of neighboring joints and the limit of motion to develop an effective motion-based model [2].

As for the model-free approaches, we can divide them into two major categories based on the manners of preserving temporal information. The first strategy keeps temporal information in recognition (and training) stage [8,9,10,11]. Sundaresan et al. utilized a hidden Markov models (`HMMs`) based framework to achieve gait recognition [9]. Sarkar et al. [11] utilized the correlation of sequence pairs to preserve the spatio-temporal relationship between the galley and probe sequences. Wang et al. [8] applied principal component analysis (PCA) to extract statistical spatio-temporal features of gait frames. However, large-scale training samples are generally needed for probabilistic temporal modelling 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 of the dataset. The second strategy converts a sequence of images into a single template [1,12,13,14,15]. Liu et al. [12] proposed to represent the human gait by averaging all the silhouettes. Motivated by their work, Han and Bhanu [1] proposed the conception of gait energy image (`GEI`), and constructed the real and synthetic gait templates to improve the accuracy of gait recognition. Bashir et al. [16] also explored the invariant gait subspaces based on entropy. With a series of grayscale averaged gait images, Xu et al. employed discriminant analysis with tensor representation (DATER) for individual recognition [13]. Chen et al. proposed multilinear

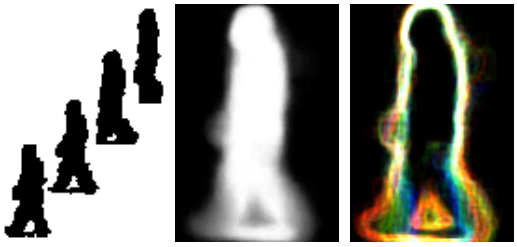

Fig. 1. From left to right: a gait sequence, gait energy image, and chrono-gait image

tensor-based non-parametric dimension reduction (MTP) [15] for gait recognition. However, the above template-based methods lose the temporal information of gait sequences more or less. 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 [13,14,15].

1.2 Our Contribution

In the recent years, the visualization community has studied how to effectively represent a sequence of images with a single colored image. For displaying time-varying data, especially for volumetric data, Woodring and Shen [17] investigated several different color-mapping strategies by encoding the time varying information of the data into color spectrum. Jänicke et al. [18] measured local statistical complexity for multifield visualization. More recently, Wang et al. [19] claimed that critically important areas are the most essential aspect of time-varying data to be detected and highlighted. However, it is difficult to directly employ such methods to generate a good temporal template for gait recognition since these algorithms are inefficient to compress gait sequences (in which gaits always have large overlapped regions between frames) into a 2-dimensional gait image.

Considering the pros and cons of gait recognition methods mentioned before, we focus on the single-template method in this paper because of its simplicity and low computational complexity. But in order to well preserve temporal information of gait patterns, we borrow some ideas from color temporal encoding in the visualisation community. In brief, we propose a novel temporal template to encode a gait sequence to a color image, named as Chrono-Gait Image (`CGI`). To further improve the discriminant ability of `CGIs`, we also propose a simple strategy to generate real and synthetic templates. An example of a gait sequence, `GEI` and `CGI` is shown in Fig. 1. In comparison with the state-of-the-art methods, our major contributions are: 1) Simple and easy to implement, `CGI` effectively preserves the temporal information in a gait sequence with a single-template image. 2) Unlike intensity, color, 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) To the best of our knowledge, color encoding gait images as a temporal template for gait recognition is not yet exploited in the biometric authentication community. Experiments indicate that compared with several recently published approaches, the CGI temporal template attains competitive performance on USF HumanID benchmark database.

The remainder of the paper is organized as follows. The proposed CGI temporal template is detailed in Section 2. We discuss the generation of real and synthetic CGI templates and the corresponding human recognition procedure in Section 3. Experiments are provided and analyzed in Section 4. Section 5 concludes the paper.

2 Chrono-Gait Images

In this paper, we attempt to achieve individual recognition under a particular human motion. Note that the motion, regular human walking, is generally used in most of current approaches of individual recognition by gait.

2.1 Motivation and Pre-processing

Because of the basic structure of human body, regular human walking always has a fixed cycle with a particular frequency. However, some methods lose the gait cycle information when performing a individual recognition by gait. Meanwhile, other methods need high computational cost for preserving such information. To address the issue, several fundamental assumptions in this paper are: 1) most of 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) color can be used as a function of time. Under these assumptions, we can encode time-varying gait cycle information into a single chrono-gait image by color.

To obtain the CGIs, we directly employ the silhouette images that are extracted by the baseline algorithm proposed by Sarkar et al. [11]. Then we encode temporal information in the silhouette images with additional colors to generate a chrono-gait image. The goal of CGIs is to compress the silhouette images into a single image without losing too much temporal relationship between the images.

2.2 Period Detection

Regular human walking is a periodical motion. To preserve the temporal information, we need to detect the period in the gait sequence firstly. We propose to calculate average width W of the leg region in a gait silhouette image I as follows:

$$W = \frac{1}{\beta H - \alpha H} \sum_{i=\alpha H}^{\beta H} (R_i - L_i), 0 \leq \alpha \leq \beta \leq 1 \tag{1}$$

where H is the height of an image, L_i and R_i are the positions of the leftmost and rightmost foreground pixels in the ith line of the silhouette images, respectively.

To alleviate the influence of some exterior factors such as briefcase, shadow and surface that might be misclassified into the silhouette image, parameters α and β are used to constrain the computation of the gait period to the leg region.

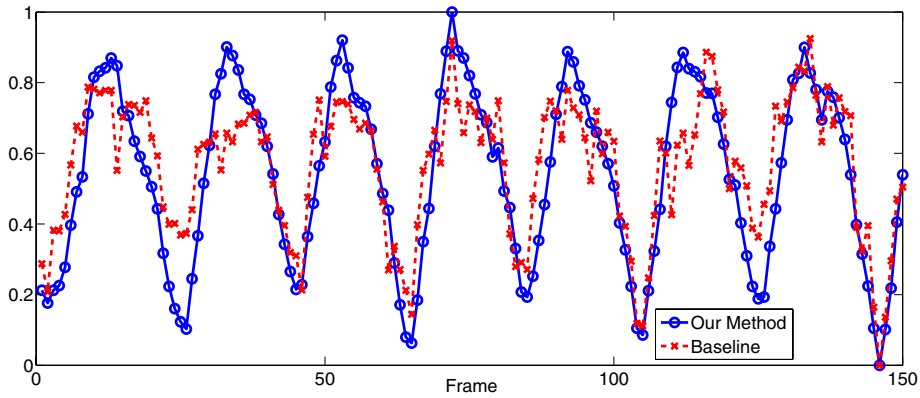

Fig. 2. Comparison between our method and baseline algorithm on gait period detection. The X-axis denotes the order of gait frames. The Y-axis represents the average width of each frame for our method, and the number of foreground pixels in the lower half of the silhouette for baseline method. Both of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

Sarkar et al. [11] proposed to detect such key frames by counting the number of foreground pixels in the lower half of the silhouettes in their baseline algorithm. A comparison of these two detection methods is illustrated as in Fig. 2, from which it can be seen 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 [11].

2.3 Color Mapping

It is worth mentioning that in the visualization community, Woodring and Shen [17] used pseudo-color to visualize time-varying information for volume rendering. They proposed four integration functions: 1) Alpha Compositing; 2) First Temporal Hit; 3) Additive Colors; 4) Minimum/Maximum Intensity. However, their methods can not directly be applied to generate a temporal template since they assume that there is little overlapped foreground region between continuous frames, while in our case, the overlap of foreground silhouettes is serious between gait frames. Some results using their recommended four functions are illustrated in Fig. 3. All these images are generated from the same gait sequence. Due to the serious overlap between gait frames, we can see that the resulting

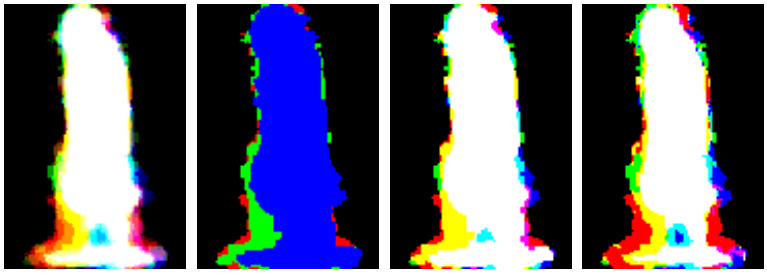

Fig. 3. Some results for visualizing gaits using Woodring and Shen's algorithm [17]. a) Alpha Compositing; b) First Temporal Hit; c) Additive Colors; d) Minimum/Maximum Intensity.

images lose many important features. Naturally, such results are not truly informative to represent gait patterns.

Since the outer contour of the silhouette images is an important feature [3,8], and preserve the spatial information with small degree of overlap, we thus attempt to extract the contours instead of silhouettes. There are various edge detection techniques such as gradient operator, LoG operator and local information entropy [20], to extract the contours of the silhouette images. We use local information entropy to obtain the gait contour 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 $\omega_d(x,y) = \big\{(u,v)|\max\{|u-x|,|v-y|\} \le d\big\}$, and n_0 and n_1 are the numbers of foreground pixels and background pixels, respectively. Term t represents the frame label, and x and y denote the values in the 2-d image coordinate.

Then we normalize the entropy by the following formula:

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

We also propose a liner interpolation function to encode the temporal information to three color components (R=Red, G=Green, B=Blue) as follows:

$$R(k_t) = \begin{cases} 0 & k_t \le 1/2, \\ (2k_t - 1)I & k_t > 1/2 \end{cases} \tag{4}$$

$$G(k_t) = \begin{cases} 2k_t I & k_t \le 1/2, \\ (2 - 2k_t)I & k_t > 1/2 \end{cases} \tag{5}$$

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

where $k_t = (W_t - W_{min})/(W_{max} - W_{min})$. W_i represents the degree of two legs apart from each other, which is explained in Equ. 1. W_{max} and W_{min} are

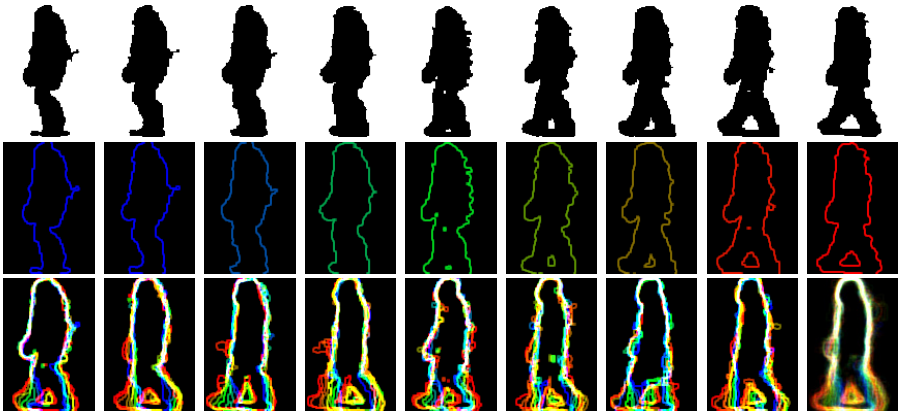

Fig. 4. An example of generating a `CGI` temporal template

the extreme widths of the period which the ith frame belongs to, and I is the maximum of intensity value, e.g., 255.

2.4 Representation Construction

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

$$\mathbf{C}_t(x,y) = \begin{pmatrix} h'_t(x,y) * R(k_t) \\ h'_t(x,y) * G(k_t) \\ h'_t(x,y) * B(k_t) \end{pmatrix} \tag{7}$$

Given the colored gait contour images $\mathbf{C}_t$, a `CGI` temporal template $\mathbf{CG}(x,y)$ is defined as follows:

$$\mathbf{CG}(x,y) = \frac{1}{p}\sum_{i=1}^{p} \mathbf{PG}_i(x,y), \tag{8}$$

where p is the number of 1/4 gait periods, and $\mathbf{PG}_i(x,y) = \sum_{t=1}^{n_i} \mathbf{C}_t(x,y)$ is the sum of the total n_i colored contour images in the ith 1/4 gait period.

The whole process to generate `CGI` is shown in Fig. 4. The first row shows 9 silhouettes in the αth 1/4 gait period. And the second row shows the corresponding colored gait contour images after edge detection and color mapping. Then we sum all these 9 images to obtain the first one $\mathbf{PG}_\alpha$ in the third row, representing this 1/4 period. The second to the eighth images on the third row represent $\mathbf{PG}$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 better visualization result and more informative gait template (which will be demonstrated in gait recognition experiments).

3 Human Recognition Using CGI Template

To employ the proposed CGI temporal template for individual recognition, one way is to directly measure the similarity between the gallery and probe templates. However, there are probably several disadvantages of doing so: 1) the templates obtained from the gait sequences may lead to overfit since such sequences are collected under similar physical conditions; 2) the number of CGIs is too small to characterize the topology of essential gait space; 3) when one pixel is viewed as one dimension, we will face to the problem of curse of dimensionality. Therefore, one solution to this is to generate two types of templates, namely real templates and synthetic templates. Meanwhile, we can project the templates into certain low-dimensional discrimination subspace with the dimension reduction method.

Specifically, we generate the real templates by referring to the colored image of each period (i.e., averaging continuous 4 **PG**s in one period) as a temporal template. One advantage is that such a template keeps the similar gait temporal information as the CGI of the whole sequence owns. Synthetic templates are used to enhance the robustness to the exterior factors such as the noise of shadow. Similar 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.

To address the curse of dimensionality issue, we employ Principal Component Analysis and Linear Discriminant Analysis (PCA+LDA) to project the real and synthetic templates in the gallery set into a low-dimensional subspace. And the real/synthetic templates in the probe set will be mapped into the low-dimensional subspace by using the projection matrix obtained by PCA+LDA. 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 identified based on 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 to the class center of the gallery real/synthetic feature vectors. To further improve the performance, we fuse the results of these two types of templates using the following equation:

$$d(\hat{\mathcal{R}}_p, \hat{\mathcal{S}}_p, \mathcal{R}_i, \mathcal{S}_i) = \frac{d(\hat{\mathcal{R}}_p, \mathcal{R}_i)}{min_j d(\hat{\mathcal{R}}_p, \mathcal{R}_j)} + \frac{d(\hat{\mathcal{S}}_p, \mathcal{S}_i)}{min_j d(\hat{\mathcal{S}}_p, \mathcal{S}_j)}, \quad i, j = 1, \cdots, C \tag{9}$$

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

$$k = \arg\min_i d(\hat{\mathcal{R}}_p, \hat{\mathcal{S}}_p, \mathcal{R}_i, \mathcal{S}_i), \quad i = 1, \cdots, C \tag{10}$$

More details about real and synthetic templates can be referred in Han and Bhanu's work [1]. Note that although we use distance measurements and fusion functions different from those used in [1], extensive experiments indicate that

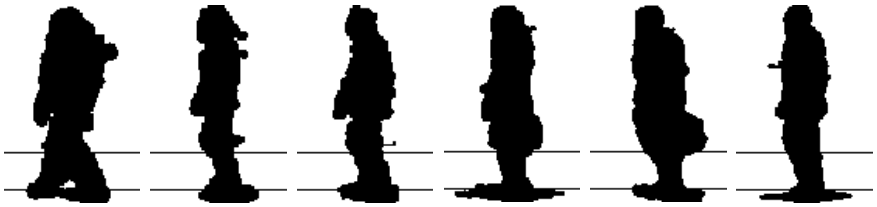

Fig. 5. Some gait examples with two additional lines. The upper line is at $3/4H$ and the lower one is at $15/16H$ (H is the height of a gait image). All of these cases are collected from different probe sets. The forth and fifth images are with briefcase and others are without briefcase.

the differences between the proposed fusion criterion and theirs are quite minor with respect to recognition accuray.

4 Experiments

We evaluate the `CGI` algorithm on the USF HumanID gait database (silhouette version 2.1) [11]. The gait database consists of 122 individuals walking in elliptical paths on concrete and grass surface, with/without a briefcase, wearing different shoes, and sampling in elapsed time. Sarkar et al. [11] selected the sequences with "Grass, Shoe Type A, Right Camera, No Briefcase, and Time t_1" for the gallery set, and developed 12 experiments, each of which is under some specific conditions. More details can be referred to Sarkar et al.'s work[11]. Because the USF database has provided the silhouette images after background subtraction and image alignment, all of our experiments are based on these silhouette images.

We evaluate the "Rank1" and "Rank5" performances of several recent approaches including baseline algorithm (based on silhouette shape matching) [11], `GEI` [1], HMM [21], IMED+LDA [14], 2DLDA [14], DATER [13], MTP [15] and Tensor Locality Preserving Projections (TLPP) [22]. The Rank1 performance means the percentage of the correct subjects ranked first while the Rank5 performance means the percentage of the correct subjects appeared 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 2.2, we introduce two parameters α and β. For the USF HumanID database, we choose $\alpha = 3/4$ and $\beta = 15/16$. Some examples in the database are shown in Fig. 5. From the figure, we can see that most of the briefcase is above the first $3/4H$ line and most of the shadow is under the $15/16H$ line. Therefore, it means that the the influence of briefcase and shadow can be effectively decreased.

To evaluate the performance of the proposed `CGI` temporal template, we employ a simple 1-nearest neighbor classifier (1-NN) on the original `GEI` and `CGI` without using real/synthetic templates and PCA/LDA. We also provide the performance of baseline algorithm [11] for comparison. The neighborhood parameter d introduced in Section 2.3, which is used to describe the size of locality, is set to

Table 1. Comparison of Recognition Performance on USF HumanID Database using 1-NN. Here, V–View, S–Shoe, U–Surface, B–Briefcase, T–Time, C–Clothing

Exp.	Gallery Size	Difference	Rank1 Performance (%)			Rank5 Performance (%)		
			baseline [11]	GEI	CGI	baseline	GEI	CGI
A	122	V	73	84	**89**	88	93	**98**
B	54	S	78	87	**91**	93	**94**	**94**
C	54	SV	48	72	**74**	78	93	**94**
D	121	U	**32**	19	20	**66**	45	40
E	60	US	**22**	18	18	**55**	53	43
F	121	UV	**17**	10	11	**42**	29	23
G	60	USV	**17**	13	13	**38**	37	32
H	120	B	61	56	**76**	85	77	**90**
I	60	SB	57	55	**75**	78	77	**93**
J	120	VB	36	40	**57**	62	69	**80**
K	33	TSC	3	**9**	6	12	15	**24**
L	33	TUSC	3	3	**9**	15	15	**24**
Avg.			40.96	41.13	**48.33**	64.54	61.38	**65.45**

1 in our experiment. The results are summarized in Tab. 1. It can be seen from the Tab. 1 that 1) `CGI` achieves the best average performance among all the algorithms. 2) As illustrated in Exp. H, I, J, the performance of `CGI` is very robust to the briefcase condition, and in such conditions, the accuracy is improved by almost 20% compared with `GEI`. 3) Compared with `GEI`, `CGI` has better Rank5 performance than `GEI` in 8 out of 12 specific conditions. 4) In all the remaining 4 conditions, both `GEI` and `CGI` perform worse than the baseline algorithm in the surface condition. We can infer that the gait templates are more sensitive to the surface condition than the baseline algorithm because of the shadows or some other factors.

To discover which components of the proposed `CGI` temporal templates are crucial to the performance of gait recognition, we compare several variants of the contour-based temporal template with the silhouette-based template, which is employed by most of the gait recognition systems [13]. Here `GEI-contour` means that we compute the `GEI` based on contour images, and `CGI-gray` means that we average each `CGI` into a graysclale image.

We also employ the fusion of real and synthetic templates introduced in Section 3 to further improve the performance. To make the experiment fair, we use the same strategy to generate real and synthetic templates, 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. More precisely, the reduced dimension is obtained when using 99% as the PCA cumulative contribution rate and the regularization parameter of LDA is set to 10^8 for real templates. And for synthetic templates, we set parameter K to 6, in other words, we cut the last 10 rows of each gait image to generate 6 synthetic templates. We use 99.5% as the PCA cumulative contribution rate and set the LDA parameter to 0. The size of locality d is also set to 1 here.

Table 2. Comparison of Recognition Performance of `GEI`, `GEI-contour`, `CGI-gray`, `CGI` using the Same Experiment Settings

Exp.	Rank1 Performance (%)				Rank5 Performance(%)			
	GEI	GEI-contour	CGI-gray	CGI	GEI	GEI-contour	CGI-gray	CGI
A	87	85	86	**92**	96	96	**97**	96
B	**93**	91	**93**	**93**	94	**96**	**96**	94
C	74	72	**81**	76	**94**	93	**94**	93
D	34	27	38	**47**	66	52	70	**75**
E	38	37	42	**48**	63	60	68	**70**
F	21	11	31	**34**	47	31	53	**54**
G	23	17	28	**40**	47	42	**58**	57
H	57	74	72	**82**	81	93	93	**94**
I	58	73	**75**	73	80	88	87	**93**
J	52	57	59	**62**	78	82	82	**83**
K	**9**	**9**	**9**	6	21	**30**	27	27
L	6	12	9	**15**	24	**36**	27	24
Avg.	49.06	49.90	55.74	**60.54**	70.46	69.52	75.89	**76.83**

Table 3. Comparison of Recognition Performance of `GEI`, `CGI` using two period detection method. We refer the period detection method proposed in Sarkar et al. [11] as "C" and our method proposed in section 2.2 as "W".

Exp.	Rank 1 Performance (%)				Rank 5 Performance (%)			
	GEI+C	GEI+W	CGI+C	CGI+W	GEI+C	GEI+W	CGI+C	CGI+W
A	87	88	**85**	**92**	96	95	93	96
B	93	91	**87**	**93**	94	96	94	94
C	74	76	78	76	94	93	93	93
D	**34**	**40**	47	47	66	66	72	75
E	38	38	53	48	63	63	70	70
F	21	25	30	34	47	45	55	54
G	23	28	37	40	47	50	55	57
H	57	58	**76**	**82**	81	78	93	94
I	**58**	**50**	68	73	80	80	88	93
J	**52**	**42**	58	62	78	76	84	83
K	9	12	3	6	**21**	**27**	24	27
L	6	6	18	15	24	27	24	24
Avg.	49.06	49.06	57.20	60.54	70.46	70.04	75.68	76.83

For saving space, we only report the fusion result in Tab. 2. From the Tab. 2 it can be seen that 1) `GEI-contour` and `CGI` obtain a remarkable improvement on Exp. H, I, J compared with `GEI`, and `CGI` is slightly better than `GEI-contour`. It means the key to the improvement on briefcase condition is the 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 more better than `GEI-contour` on Exp. D, E, F, G. It indicates that although contour instead of silhouette reduces the recognition rate on surface condition, using the proposed `CGI` temporal template can make up for such loss and further improve

Table 4. Comparison of Recognition Performance on USF HumanID Database using Different Methods, Rank1 Performance (%).

	A	B	C	D	E	F	G	H	I	J	K	L	Avg.
Baseline[11]	73	78	48	32	22	17	17	61	57	36	3	3	40.96
HMM[21]	89	88	68	35	28	15	21	**85**	**80**	58	17	15	53.54
IMED+LDA[14]	88	86	72	29	33	23	32	54	62	52	8	13	48.64
2DLDA[14]	89	**93**	80	28	33	17	19	74	71	49	16	16	50.98
DATER[13]	87	**93**	78	42	42	23	28	80	79	59	18	21	56.99
TLPP[22]	87	**93**	72	25	35	17	18	62	62	43	12	15	46.95
MTP[15]	90	91	**83**	37	43	23	25	56	59	59	9	6	51.57
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	89	81	28	30	15	13	82	75	60	3	3	51.98
CGI+Synthetic	89	89	67	53	53	27	33	59	60	56	3	15	54.49
CGI+Fusion	**92**	**93**	76	47	48	**34**	**40**	82	73	**62**	6	15	**60.54**

Table 5. Comparison of Recognition Performance on USF HumanID Database using Different Methods, Rank5 Performance (%).

	A	B	C	D	E	F	G	H	I	J	K	L	Avg.
Baseline[11]	88	93	78	66	55	42	38	85	78	62	12	15	64.54
HMM[21]	–	–	–	–	–	–	–	–	–	–	–	–	–
IMED+LDA[14]	95	95	90	52	63	42	47	86	86	78	21	19	68.60
2DLDA[14]	**97**	93	**93**	57	59	39	47	91	**94**	75	**37**	34	70.95
DATER[13]	96	**96**	**93**	69	69	51	52	92	90	83	40	36	75.68
TLPP[22]	94	94	87	52	55	35	42	85	78	68	24	33	65.18
MTP[15]	94	93	91	64	68	51	52	88	83	82	18	15	71.38
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	96	94	**93**	64	62	45	50	**94**	93	**85**	27	33	73.90
CGI+Synthetic	93	**96**	85	71	72	47	50	91	85	82	21	30	73.28
CGI+Fusion	96	94	**93**	75	70	54	**57**	**94**	93	83	27	24	**76.83**

the performance of gait recognition. 3) Compared with `CGI-gray`, `CGI` has better Rank1 performance in 9 out of 12 specific conditions and improve the average recognition ratio by about 5%. We can thus infer that with color encoding, the temporal information of the gait sequence benefits individual recognition by gait.

With the same parameter setting, we also investigate the influence of gait period detection as tabulated in Tab. 3. We observe from Tab. 3 that the divergence between the two detection methods is minor in almost all the experiments expect 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, it can work well using a basic period detection, and may work better if employing an advanced period detection method. Furthermore, `CGI` performs better than `GEI` when using both period detection methods.

Finally, we also compare the proposed algorithms with several state-of-the-art published results are illustrated as in Tab. 4 and Tab. 5. It is obvious that the proposed `CGI` outperforms others in both average Rank1 and Rank 5 performances, and is robust under most of the complex conditions. It is worth mentioning that since the time and space complexities of `CGI` are the same as those of `GEI`, the proposed `CGI` temporal template is very effective and competitive for real-world applications.

5 Conclusion

In this paper, we have proposed a simple and effective temporal template `CGI`. We extract a set of contour images from the corresponding silhouette images using local entropy principle, then color encoding the temporal information of gait sequence into the `CGI`. We also generate real and synthetic temporal templates and exploit fusion strategy to obtain better performance. Experiments in a benchmark database have demonstrated that compared with state-of-the-art, our `CGI` template can attain higher recognition accuracy.

In the future, we will study how to enhance `CGI`'s robustness in more complex conditions, and investigate how to select a more general color mapping function instead of the current linear mapping function. Furthermore, we will consider to generalize the proposed frameworks into other human-movement-related fields [23] such as gesture analysis and abnormal behavior detection.

Acknowledgements

This work was supported in part by the NFSC (No. 60635030, 60975044) and 973 program (No. 2010CB327900, 2009CB320903), Shanghai Leading Academic Discipline Project No. B114, Shanghai Key Laboratory of Intelligent Information Processing, China. Grant No. IIPL-09-016, and Hui-Chun Chin and Tsung-Dao Lee Chinese Undergraduate Research Endowment (CURE).

References

1. Han, J., Bhanu, B.: Individual Recognition using Gait Energy Image. TPAMI 28, 316–322 (2006)
2. Yam, C.Y., Nixon, M.S.: Model-based Gait Recognition. In: Encyclopedia of Biometrics, pp. 633–639. Springer, Heidelberg (2009)
3. Wang, L., Tan, T.N., Hu, W.M., Ning, H.Z.: Automatic Gait Recognition Based on Statistical Shape Analysis. TIP 12, 1120–1131 (2003)

4. Veres, G.V., Gordon, L., Carter, J.N., Nixon, M.S.: What Image Information is Important in Silhouette-based Gait Recognition? In: CVPR, vol. 2, pp. 776–782 (2004)
5. Guo, B., Nixon, M.S.: Gait Feature Subset Selection by Mutual Information. TSMC, Part A 39, 36–46 (2009)
6. Bouchrika, I., Nixon, M.S.: Model-based Feature Extraction for Gait Analysis and Recognition. In: Gagalowicz, A., Philips, W. (eds.) MIRAGE 2007. LNCS, vol. 4418, pp. 150–160. Springer, Heidelberg (2007)
7. Wang, L., Tan, T., Ning, H., Hu, W.: Fusion of Static and Dynamic Body Biometrics for Gait Recognition. TCSVT 14, 149–158 (2004)
8. Wang, L., Tan, T.N., Ning, H.Z., Hu, W.M.: Silhouette Analysis Based Gait Recognition for Human Identification. TPAMI 12, 1505–1518 (2003)
9. Sundaresan, A., Roy-Chowdhury, A., Chellappa, R.: A Hidden Markov Model Based Framework for Recognition of Humans from Gait Sequences. In: ICIP, vol. 2, pp. 93–96 (2003)
10. Kobayashi, T., Otsu, N.: Action and Simultaneous Multiple-person Identification using Cubic Higher-order Local Auto-correlation. In: ICPR, vol. 4, pp. 741–744 (2004)
11. Sarkar, S., Phillips, P.J., Liu, Z., Vega, I.R., Grother, P., Bowyer, K.W.: The Humanid Gait Challenge Problem: Data Sets, Performance, and Analysis. TPAMI 27, 162–177 (2005)
12. Liu, Z., Sarkar, S.: Simplest Representation Yet for Gait Recognition: Averaged Silhouette. In: ICPR, vol. 4, pp. 211–214 (2004)
13. Xu, D., Yan, S., Tao, D., Zhang, L., Li, X., Zhang, H.J.: Human Gait Recognition With Matrix Representation. TCSVT 16, 896–903 (2006)
14. Tao, D., Li, X., Wu, X., Maybank, S.J.: General Tensor Discriminant Analysis and Gabor Features for Gait Recognition. TPAMI 29, 1700–1715 (2007)
15. Chen, C., Zhang, J., Fleischer, R.: Multilinear tensor-based non-parametric dimension reduction for gait recognition. In: Tistarelli, M., Nixon, M.S. (eds.) ICB 2009. LNCS, vol. 5558, pp. 1030–1039. Springer, Heidelberg (2009)
16. Bashir, K., Xiang, T., Gong, S.: Gait recognition using gait entropy image. IET Seminar Digests 2009 (2009)
17. Woodring, J., Shen, H.W.: Chronovolumes: A Direct Rendering Technique for Visualizing Time-varying Data. In: Eurographics, pp. 27–34 (2003)
18. Jänicke, H., Wiebel, A., Scheuermann, G., Kollmann, W.: Multifield Visualization using Local Statistical Complexity. TVCG 13, 1384–1391 (2007)
19. Wang, C., Yu, H., Ma, K.L.: Importance-driven Time-varying Data Visualization. TVCG 4, 1547–1554 (2008)
20. Yan, C., Sang, N., Zhang, T.: Local entropy-based transition region extraction and thresholding. Pattern Recognition Letters 24, 2935–2941 (2003)
21. Kale, A., Sundaresan, A., Rajagopalan, A.N., Cuntoor, N.P., RoyChowdhury, A.K., Kruger, V., Chellappa, R.: Identification of Humans using Gait. TIP 13, 1163–1173 (2004)
22. Chen, C., Zhang, J., Fleischer, R.: Distance Approximating Dimension Reduction of Riemannian Manifolds. TSMCB 40, 208–217 (2010)
23. Moeslund, T.B., Hilton, A., Krüger, V.: A Survey of Advances in Vision-based Human Motion Capture and Analysis. CVIU 103, 90–126 (2006)