ORIGINAL PAPER

On the selection of spatiotemporal filtering with classifier ensemble method for effective gait recognition

Mohammad H. Ghaeminia[1] · Shahriar B. Shokouhi[1]

Abstract
In this paper, we improve the performance of gait recognition by modeling human's motion with spatiotemporal gait features. Since existing methods often use average of silhouettes, i.e., gait energy image to model the gait, temporal information of walking may not be preserved under covariate factors. To handle such features in different conditions, we study the gait model from energy viewpoint. In the proposed method, energy of a gait, i.e., spatiotemporal feature, is derived from a newly designed filtering approach and the energies within a period will be aggregated into a single template that is called gait spatiotemporal image. The required features are truly extracted from spatial and temporal impulse responses that are redesigned and optimized for the gait. Moreover, to recognize the gait under covariate factors, a hybrid decision-level classifier based on random subspace method has been utilized for the given templates. Experimental results on well-known public datasets demonstrate the efficacy of our model. The proposed gait recognition system achieves the recognition rate of 72.25% for Rank1 and 85.64% for Rank5 on the USF dataset that is improved by at least 2% in Rank1 and 0.3% in Rank5 with respect to recent template-based methods.

Keywords Gait biometrics · Motion-based filtering · Spatiotemporal representation · Ensemble classification

1 Introduction

Gait recognition is one of the most challenging problems in the field of pattern recognition. Unlike the other biometrics such as face and fingerprint, gait has some advantages [1, 2]. It can be collected at a distance in an unobtrusive and hidden manner, and the low-resolution images can be used for representing gait traits. However, there are some covariate factors that can affect the performance of a gait recognition system [3]. To represent the appearance of gait in a video without any predefined model, robust features should be computed from the appearance (i.e., silhouette) [1, 4, 5]. Han et al. [2] introduce gait energy images (GEIs), or average of silhouettes, over one period of walking. Moreover, the augmented template features can be achieved with Gabor filter. For example, Xu et al. [1] introduce Gabor-based patch distribution feature (Gabor-PDF) combined with locality-constraint group

sparse representation (LGSR) to improve recognition rates. It is important to mention that the temporal ordering of gait will be lost during the conversion of silhouettes to a single gait template [4–6]. To preserve such information in the final template, a walking model based on spatiotemporal features should be considered within the averaging process. Hence, Wang et al. [5] develop a colored gait image named chronogait image (CGI), and Lam et al. [6] propose gait flow image (GFI). Atta et al. [4] recently apply lifting 5/3 wavelet filters to the contour of silhouettes to preserve temporal information of walking in final template. This template named 5/3 gait image (5/3GI) is able to handle different covariate factors under walking conditions.

All of the mentioned features have some privileges in gait modeling, but lack of efficient human's motion is noticeable in their features. To compute motion features within a timing duration, it is suggested to apply a filtering approach to the input video [7, 8]. Ghaeminia et al. [9] develop gait salient image (GSI) by applying opponent-based filtering and extracting salient gait features. They show that the quality of gait representation can be significantly improved by forming a template with a simple motion-based filtering. In this paper, we discuss on the method of action filtering and develop a filtering approach to compute required energy. Following the

✉ Mohammad H. Ghaeminia
  mhghaeminia@elec.iust.ac.ir

  Shahriar B. Shokouhi
  bshokouhi@iust.ac.ir

[1] School of Electrical Engineering, Iran University of Science and Technology, Tehran, Iran
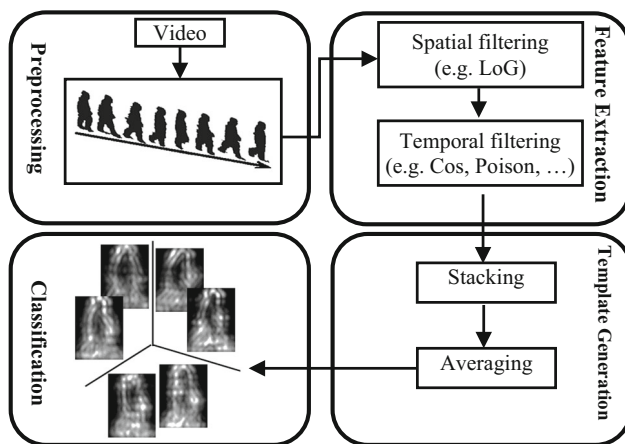
Fig. 1 The schematic of the proposed approach

idea in GSI method, we simplify the motion filtering process and propose two impulse responses to extract spatial and temporal energies from the gait motions. With consideration of a suitable filtering approach, the related templates can be computed by aggregating the responses and averaging them over a period [9, 10]. The block diagram of proposed approach to extract desired features is shown in Fig. 1.

Once the gait templates have been computed, it could be recognized with a robust classification method [3, 11].

In this regard, a classifier ensemble method, i.e., random subspace model (RSM), based on tensor representation of templates has been used as an efficient tool in face and gait recognition [3]. In RSM, the effect of covariate factors on final template can be significantly reduced by generating weak classifiers through random sampling from full feature space. In this paper, we apply RSM method to classify the proposed gait templates on publicly gait datasets such as USF [12] and OU-ISIR-B [13].

By summarizing the state-of-the-art methods, the main contributions of the paper are highlighted here:

1. A gait energy model based on spatio-temporal filtering is being conceptualized to handle different gait conditions.
2. A robust template-based feature is computed based on the responses of filtering.
3. The ensemble technique is applied to classify new templates in the feature search space.
4. The effect of the defined parameters on the final template and corresponding recognition rate is also evaluated.

In the rest of the paper, we will discuss on the proposed template in Sect. 2. The classifier ensemble method is overviewed in Sect. 3, and the experimental results are presented in Sect. 4. The conclusion of paper is also summarized in Sect. 5.

2 The proposed energy model

In this section, we discuss thoroughly on the proposed method which encodes energy of gait into single template. The proposed filtering has two basic steps as shown in Fig. 1, computing spatial and making temporal impulse responses over a period. By convolving the gait sequence with both responses, the motion-based features (or motion energies) will be computed. Afterward, the local energies are being stacked over the sub-periods and averaged in a gait period to form a template-based feature.

2.1 Preprocessing and silhouette extraction

Since the gait is a periodic and rhythmic action, some preprocessing steps are needed in computing the period and silhouettes from the raw input video. After computing the background [with Gaussian mixture model (GMM)], the gait silhouettes can be derived with subtraction of input video from the background model [12]. Finally, the silhouette images have been aligned and resized according to the center of the image. Considering the periodical human's walking, we have to calculate the period to synchronize starting phase of the filtering. The period T could be easily derived by counting the number of foreground pixels (or leg regions) in the aligned-binary silhouettes [12]. The number of counted pixels has local extremums when two legs are nearest or farthest to each other. By computing such extremums, the period of a gait is the median of distances between two consecutive minimums (or maximums) [10, 12].

It is clear that the gait is a periodic activity, and hence in each period, there are sub-cycles with similar motion patterns. In our approach, we compute the features in each sub-cycles (i.e., 1/2 period) separately to achieve best performance. In the next subsection, we explain the filtering process in detail.

2.2 Spatiotemporal filtering

According to the Shabani's model [8], a video signal v_0 can be observed as a circuit network, in which each pixel $X = (x, y)$ connects to its spatially and temporally neighbors by a resistance [8]. The brightness v of each pixel is also considered as the charge of capacitor connected to each node, and the diffusion within the nodes is its flux at time t. Following the Kirchhoff's laws of assumed network and using the Fourier transform, we will have the set of spatiotemporal impulse responses to extract energy of input video signal as follows [8]:

$$v(x, y, t; s) = G * k * v_0, \quad \begin{cases} G(x, y; \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right), \\ k(t; \tau) = \text{sinc}(t - \tau) \cdot S(t) \end{cases}$$

$$(1)$$

where $S(t)$ denotes the Heaviside step function. Equation (3) means that 2D Gaussian and 1D Sinc functions are suitable to extract energy of a video signals. To make responses, R, invariant to the phase of motions, the video signal is further convolved with the even and odd parts of the temporal function [7, 8]:

$$R = (v * G * k_{ev})^2 + (v * G * k_{od})^2, \tag{2}$$

where k_{ev}/k_{od} are set of even/odd temporal Sinc filtering and "*" denotes the convolution process. Although the mentioned filters are suitable for the action, they are not applicable for modeling of gait for at least two reasons: (1) The spatial energy of a gait sequence is focused on the edge of silhouettes, and hence, a high-pass filter can detect the edges, respectively (e.g., like the features in [4, 5]). The Gaussian kernel from Eq. (3) as smoothing spatial filter is a kind of low-pass filter. (2) The walking process is a rhythmic and periodic activity [2, 12] in temporal domain in which single-tone filtering is needed for modeling. The Sinc kernel from Eq. (3) is an ideal low-pass filtering, i.e., rectangle filter in frequency domain, which contains different harmonics to handle all types of actions from walking to running. Therefore, to apply the Sinc filter to the gait, one may filter out different harmonics and keep just a few of them.

To design spatiotemporal filtering for the gait, some constraints should be added to the action model in Eq. (1). In the first step, the higher-order derivative of spatial signal can extract the edges of gait silhouettes properly. For this purpose, we suggest Laplacian of Gaussian (LoG) operator [5]:

$$\begin{aligned} \text{LoG} &\equiv \Delta(G(x, y; \sigma)) \\ &= \frac{\partial^2}{\partial x^2} G(x, y; \sigma) + \frac{\partial^2}{\partial y^2} G(x, y; \sigma), \end{aligned} \tag{3}$$

In second step, the temporal part of the proposed filter is derived by applying a low-pass filter to remove different harmonics of the Sinc signal. Here, we represent a periodic Sinc signal by Fourier series:

$$\begin{aligned} \text{Sinc}_{ev}(t) &= \frac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n \cos(n\omega_\tau t) + \sum_{n=1}^{\infty} b_n \sin(n\omega_\tau t) \\ \text{Sinc}_{ev}(t + T) &= \text{Sinc}_{ev}(t) \quad t \geq 0, \end{aligned} \tag{4}$$

where ω_τ is π/T (T the gait period) and $b_n = 0$ for an even signal. Afterward, we keep two first harmonics, i.e., DC and first-order harmonic (related to simple walking motion), and filter out other harmonics in the frequency domain (related to other type of actions). The resulting impulse responses can be generalized as follows:

$$k_{ev}(t, \tau) = \frac{1}{2}a_0 + a_1 \cos(\omega_\tau(t - \tau)) \quad \text{for} \quad t \geq 0, \tag{5}$$

where a_0 and a_1 are the Fourier coefficients which can be easily computed by integral of Fourier cosine series. We called our temporal impulse responses as truncated cosine (Trunc. Cos.) in this paper. Once the suitable impulse responses are being computed, there are convolved with the input gait sequence [8] to extract energy of gait in a period [according to Eq. (2)]. It should be noted that the proposed kernels are sampled above the critical sampling rate (6σ in spatial and T in temporal domain) [9]. But to simplify the notations, we show the signals in continuous notation form.

2.3 Template generation

As discussed above, the proposed filtering is computed for each silhouette within a period. Now, the final template is easily computed by averaging the responses. Since there are similar patterns of walking within a period, it is better to aggregate the responses over the 1/2 [5, 10] sub-period in advance:

$$\text{GSTE} = \sum_{k=0}^{t_{max}-1} R_k, \tag{6}$$

where R_k is a response of spatiotemporal filtering, t_{max} is the number of silhouettes in a sub-period and gait spatiotemporal energy (or GSTE) is sum of such responses. The outputs of the filtering process should be normalized prior to computing the final template:

$$\text{GSTE}' = \frac{\text{GSTE} - \min(\text{GSTE})}{\max(\text{GSTE}) - \min(\text{GSTE})}. \tag{7}$$

The final template, named gait spatiotemporal image (or GSTI), can be computed by averaging the normalized GSTE over one period. Assume there are p sub-periods ($p = 2$ in this paper) in each period of gait, the proposed GSTI is expressed as:

$$\text{GSTI} = \frac{1}{p} \sum_{i=1}^{p} \text{GSTE}'_i. \tag{8}$$

Figure 2 shows an example of the primary steps in computing a GSTI template. The proposed filtering is applied to the silhouettes in Fig. 2a to generate the responses in Fig. 2b. The GSTE features in Fig. 2c is derived by stacking the responses in first and second half of period separately. The final GSTI template is then computed by averaging the GSTE features over a period (Fig. 2d).

3 Classifier ensemble method

The GSTI template, which is introduced in the previous section, is computed for each sequence in a given dataset. These
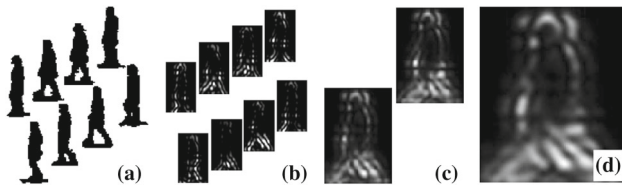
Fig. 2 The steps of generating GSTI templates: **a** input silhouettes, **b** responses of filtering, **c** GSTE features and **d** final GSTI template

templates will be further classified with RSM. In this section, we review the RSM which is consisting of three main steps [3]: random subspace sampling, enhancing dimensionality of generated features and making final decision.

3.1 Random subspace sampling

Assume there are n gait templates A_i $(i = 1 \ldots n)$ (or GSTI) in the training set (gallery). In the first step, we compute 2DPCA projection matrix leading to d eigenvectors with nonzero eigenvalues, $U = [u_1, u_2, \ldots, u_d]$ [3]. Afterward, the K random subspace T_{PCA}^k $(k = 1 \ldots K)$ can be computed by random selection of N $(N \leq d)$ unique eigenvectors from subsets U and repeating the process K-times. As a result, the random feature sets of ith template in kth subspace will be generated in lower dimension space as $X_i^k = A_i T_{\text{PCA}}^k$. It can be proved that random sampling of eigenvectors can preserve the covariate factors in lower dimension feature space efficiently [3]. However, some redundant information still remains in feature vector X_i^K that may affect the performance of final decision. To improve the recognition rate, another classification step will be applied in RSM.

3.2 Dimensionality enhancing

The randomly generated features have still redundant information that may affect quality of decision. To obtain more discriminant features for weak classifiers, an additional dimensionality reduction method should be performed. Here, two known techniques, i.e., 2D linear discriminant analysis (2DLDA) [14] and incremental dimension reduction algorithm via QR decomposition (IDR/QR) [14] can be used alternatively [3]. The features for final decision are then extracted from each method separately.

Following the 2DLDA algorithm [3], we obtain K transition matrix T_{LDA}^k $(k = 1 \ldots K)$ where each one has M selected eigenvectors. As an alternative solution, in IDR/QR technique, the T_{QRD}^k $(k = 1 \ldots K)$ will be derived based on QR decomposition of eigenvectors to extract discriminant features. It should be noted that in IDR/QR, the random features should be vectorized before training the model that may affect quality of generated features. Once the transition matrices are computed, two-staged dimensionality reduction techniques

as 2DPCA + 2DLDA (or 2DLDA) and 2DPCA + IDR/QR (or IDR/QR) will be derived for each subspace. The feature vector of ith template in training/testing sets in kth subspace for the methods is derived as $Y_{i,\text{LDA}}^k = \left(T_{\text{LDA}}^k\right)^T X_i^k$ and $Y_{i,\text{QRD}}^k = \left(T_{\text{QRD}}^k\right)^T X_{\text{vec},i}^k$ [3]. The hybrid decision level is then achieved based on the outputs of random classifiers which are discussed in the following subsection.

3.3 Classification

Now, each feature sets in kth subspace can make weak decision in according to the covariate factors. Suppose there are c classes in the training set (gallery) and each of them has n_i ($i = 1, \ldots, c$) samples. For the kth subspace, let $m_i^k (i = 1, \ldots, c)$ be the mean of the samples in each class and R^k (i.e., Y_{LDA}^k or Y_{QRD}^k) be the feature samples of probe set (including n_p gait samples). The Euclidean distance between R^k and the mean of ith class of the gallery m_i^k can be expressed as:

$$d\left(R^k, m_i^k\right) = \frac{1}{n_p} \sum_{j=1}^{n_p} \left\| R_j^k - m_i^k \right\|, \quad i = 1, \ldots, c. \quad (9)$$

The minimum distance of a given probe template to each class, $\{\omega_i\}_{i=1}^c$ in the gallery set is considered as weak decision:

$$\Omega^k(R^k) = \arg\min_{\omega_i} d(R^k, m_i^k), \quad i = 1, \ldots, c. \quad (10)$$

The distance criteria in Eqs. (9) and (10) are computed for both of 2DLDA and IDR/QR in each subspace to generate weak classifiers. A hybrid decision-level fusion (HDF) among K subspace in each set of weak classifiers can be achieved simply by majority voting of all K classifiers [3]. More precisely, for a probe gait query $R = \{R^k\}_{k=1}^K$, the mode of K labels in all of the subspaces is considered as the final decision. Let $\Omega_{\text{LDA}}(R)$ and $\Omega_{\text{QRD}}(R)$ be the final decision corresponding to the 2DLDA and IDR/QR-based features for a query gait R. The hybrid classifier (HC) can be performed by [3]:

$$\Omega_{\text{HC}}(R) = \begin{cases} \omega_i, & \text{if } \Omega_{\text{LDA}}(R) = \omega_i, \\ \omega_i, & \text{if } \Omega_{\text{QDR}}(R) = \omega_i, \quad i \in [1, c]. \\ 0, & \text{Otherwise} \end{cases} \quad (11)$$

It is inferred from Eq. (11) that HC decision is guaranteed if one of the corresponding classifiers recognizes given individual correctly.

4 Experimental results

In this section, we verify the efficiency of the proposed approach by doing comprehensive experiments on publi-

Table 1 The set of experiments for probe sets in USF gait dataset

Experiment	A	B	C	D	E	F
#People	122	54	54	121	60	121
Covariates	V	H	VH	S	SH	SV
Experiment	G	H	I	J	K	L
#People	60	120	60	120	33	33
Covariates	SHV	B	BH	BV	THC	STHC

V view, *H* shoe, *S* surface, *B* briefcase, *T* time, *C* clothing

Table 2 Clothing conditions in the OU-ISIR-B dataset

Type → clothes	Type → clothes	Type → clothes
0 → CP/CW	C → RP/DJ/Mf	N → SP/HS
2 → RP/HS	D → CP/HS	P → SP/Pk
3 → RP/HS/Ht	E → CP/LC	R → RC
4 → RP/HS/Cs	F → CP/FS	S → Sk/HS
5 → RP/LC	G → CP/Pk	T → Sk/FS
6 → RP/LC/Mf	H → CP/DJ	U → Sk/PK
7 → RP/LC/Ht	I → BP/HS	V → SK/DJ
8 → RP/LC/Cs	J → BP/LC	X → RP/FS/Cs
9 → RP/FS	K → BP/FS	Y → RP/FS/Cs
A → RP/Pk	L → BP/Pk	Z → SP/FS
B → RP/DJ	M → BP/DJ	–

RP regular pants, *BP* baggy pants, *SP* short pants, *HS* half shirt, *FS* full shirt, *LC* long coat, *CW* casual wear, *RC* rain coat, *Ht* hat, *CP* casual pants, *Sk* skirt, *Pk* parker, *DJ* down jacket, *Cs* casquette cap, *Mf* muffler

cally datasets. The parameters of our model and classification framework are discussed and the performance is compared with other state-of-the-art methods. The Rank1 and Rank5 correct classification rates (CCR) [12] are two metrics to evaluate the performance of our method. Moreover, due to random nature of the RSM, each of the experiments has been repeated 10 times and mean value of correct classification rates is considered as final performance (similar to [3]).

4.1 Gait Datasets

We evaluate our method on two benchmark datasets: USF HumanID gait dataset [12] and OU-ISIR gait dataset (dataset B) [13]. The USF dataset is including 122 individuals that are captured in an outdoor environment under five different situations. Camera viewpoint (right or left), walking surface (concrete or grass), shoe type (A or B), carrying condition (with or without a briefcase) and the elapsed time are five covariate factors. Considering such gait conditions, the set of "grass, shoe type A, right viewpoint, no briefcase, time t_1 (May)" covariates is defined as gallery and 12 experiments are developed as probe sets. The number of individuals in each probe set and the differences of conditions to the gallery set are shown in Table 1.

Another benchmark is OU-ISIR gait dataset (dataset B). The OU-ISIR-B dataset [13] has been published recently to study effect of clothing conditions. This dataset includes of 48 individuals walking on a treadmill with 32 types of different clothing. Table 2 presents the clothing combinations provided in the database. The gallery set includes 48 individuals with standard clothes (i.e., type 9), while the probe set consists of 856 gait sequences with other 31 clothing conditions. As such, the silhouettes in the OU-ISIR-B dataset is aligned on their horizontal centroid and cut to 128×88 silhouette images.

4.2 Parameters selection

To recognize a GSTI template in given dataset, two sets of parameters are being defined: template parameters and RSM variables. The first set of parameters affect quality of final
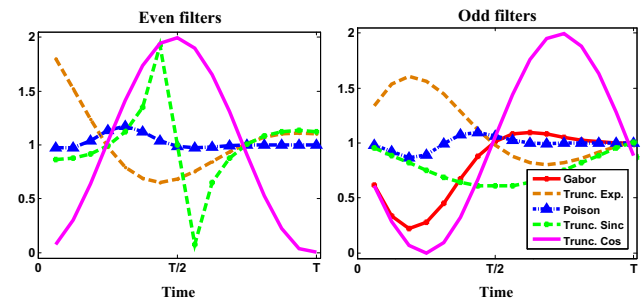

Fig. 3 Continuous profile of different truncated signals used as temporal impulse responses

template, while the others affect the performance of recognition.

In following subsections, we will review environmental settings and the optimal solution for better recognition is being discussed.

4.2.1 Template parameters

The temporal part of the proposed filtering plays an important role in computing the template. Here, to evaluate the efficacy of proposed temporal kernel, different signals such as Gabor [7], Trunc. Exp., Poison, Trunc. Sinc [8] and Trunc. Cos. signals are selected as different temporal kernels. The profiles of the even and odd parts of the mentioned kernels (in continuous form) over a gait period are shown in Fig. 3.

In our experiments, each silhouette in a gait period has been convolved with second-order derivate of Gaussian kernel (with $\sigma = 5$) and one of the temporal signals [according to Eq. (2)]. The final template is then computed and the GSTI templates are compared with 1-NN in the USF dataset. The Rank1 and Rank5 recognition rates across different temporal impulse responses are listed in Table 3. From Table 3,

Table 3 Comparisons of the recognition performance of final template using different temporal signals and 1-NN classifier on the USF dataset

Exp.	Rank1 performance (%)					Rank5 performance (%)				
	Gabor [7]	Trunc. Exp. [8]	Poison [8]	Trunc. Sinc [8]	Trunc. Cos.	Gabor	Trunc. Exp.	Poison	Trunc. Sinc	Trunc. Cos.
A	87	88	87	87	88	97	97	97	97	97
B	89	91	89	91	93	95	95	95	95	95
C	78	78	80	80	80	91	89	91	91	93
D	26	25	25	27	29	54	56	55	58	60
E	27	27	25	29	30	45	47	44	45	50
F	15	15	15	15	16	36	35	35	35	43
G	10	14	10	14	15	42	37	42	40	39
H	74	74	73	74	76	90	90	91	92	93
I	69	69	67	67	69	85	87	84	85	87
J	66	67	65	65	64	81	83	82	82	82
K	10	0	7	4	10	25	19	22	22	28
L	4	4	4	4	10	22	19	22	22	22
Avg.	50.33	50.47	49.71	50.6	52.01	68.02	67.91	68.04	68.54	70.63

in 11 out of 12 gait challenging conditions, the Rank1 and Rank5 performance of the proposed filtering has the highest one. Therefore, an optimum solution for gait modeling can be achieved by selecting a truncated cosine signal as temporal impulse response. We will also get similar performance in OU-ISIR-B dataset with the same setting of parameters.

4.2.2 Parameters of the RSM

Given GSTI templates from previous subsection, the main parameters of the RSM are as: (1) the number of random subspaces (i.e., K), (2) the dimension of the random subspaces (i.e., N) and (3) the dimension of second-stage classifier (i.e., M in 2DLDA or IDR/QR).

Generally, it is proved that the performance of recognition in not too sensitive to the M, unless it is too small and hence, the RSM is not depended to the M [3].

To have a better performance, it is suggested to increase the number of subspaces (or K) [3]. Moreover, small or large values of N may cause underlearning or overlearning problems [3]. We choose values of K and N empirically by fixing one parameter and changing another one and computing accuracies in each run. It is found that setting $K = 500$ and changing the value of N in the range 2–20 will lead to better accuracy rates. By running the experiments, average of Rank1 on the USF dataset for different values of N is shown in Fig. 4a.

By increasing the N, the accuracy is being increased slightly. However, by increasing N excessively, the computational overhead will be increased exponentially. Here, we choose $N = 10$ as an optimal value in our experiments. With setting the N to 10, we run the experiments again with different values of K in the range 100–1000. The averaged Rank1 for different values of K is shown in Fig. 4b. The accuracy is
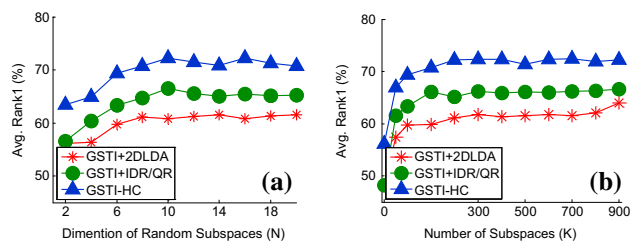

Fig. 4 Evaluating performance of the RSM with different parameters on the USF dataset: **a** fixing $K = 500$ and altering N, **b** setting $N = 10$ and changing the K

improved by increasing the number of subspaces. However, to preserve the computational resources, we choose 500 for K in our experiments.

4.3 GSTI templates

To demonstrate the effectiveness of the proposed spatiotemporal template, a simple 1-NN classifier is used to directly compare the GSTI templates. Direct matching of gait templates using 1-NN has some advantages such as: (1) The robustness of templates against noises can be examined, and (2) the performance of the recognition into a lower-dimensional feature space is guaranteed with considering redundant data in high-dimensional data. We compare the proposed template with recently published methods, namely baseline [12], GEI [2], CGI [5], GSI [9] and 5/3GI-Contour (or 5/3GI) [4] on USF dataset. The results of Rank1 and Rank5 identification rates are shown in Table 4. From Table 4, our feature template has the highest average of Rank1 and Rank5 with respect to the GEI, CGI and 5/3GI templates. More precisely, the results of Rank1 on GSTI are higher than

Table 4 Comparisons of the recognition performance on the USF dataset using 1-NN classifier

Exp.	Rank1 performance (%)						Rank5 performance (%)					
	Baseline [12]	GEI [2]	CGI [5]	GSI [9]	5/3GI [4]	Proposed model	Baseline	GEI	CGI	GSI	5/3GI	Proposed model
A	73	84	87	86	90	88	88	93	96	99	96	97
B	78	87	94	93	91	93	93	94	94	95	97	95
C	48	72	72	78	76	80	78	93	93	93	94	93
D	32	19	17	31	24	29	66	45	41	64	54	60
E	22	18	25	33	27	30	55	53	45	61	57	50
F	17	10	12	17	12	16	42	29	32	50	40	43
G	17	13	13	17	20	15	38	37	35	49	43	39
H	61	56	78	76	86	76	85	77	91	97	95	93
I	57	55	80	76	79	69	78	77	97	98	97	87
J	36	40	54	59	58	64	62	69	82	91	89	82
K	3	9	6	19	12	10	12	15	30	40	30	28
L	3	3	9	13	9	10	15	15	27	28	30	22
Avg.	41.0	41.1	48.6	52.6	51.9	52.01	64.5	61.4	66.8	76.5	72.2	70.63

GEI and CGI in hard situations such as surface (*D–G*) and elapsed time (*K–L*).

In comparisons with the GSI and 5/3GI, the GSTI template provides better Rank1/Rank5 results in some gait conditions while it has totally similar performance. Although the proposed method is sensitive to the surface conditions (probes *D–G*), it provides proper gait features with a simple filtering approach.

4.4 Performance on clothing covariates

The effects of clothing on the gait are the most common covariates in a human's walking. In USF dataset, such factor is only considered in probes *K* and *L* with a few number of individuals testing the cloth. To study effect of clothing on gait recognition, the OU-ISIR-B dataset is released recently [13] with 32 types of clothing as listed in Table 2. However, to cover all possible combinations of clothing, OU-ISIR-B dataset provides additional training set to make clothing-invariant recognition system [13]. In this paper, there is no need for predefined conditions which may lead to unrealistic performance. To investigate the performance of gait recognition system in OU-ISIR-B dataset, we compare the Rank1 results of 31 cloth types with GEI [2], Gabor+RSM-HDF [3] and VI-MGR [15] which is shown in Fig. 5.

It is clear from Fig. 5 that the proposed method outperforms the Rank1 in comparisons with GEI and VI-MGR in most of the conditions. More precisely, in 5 out of 31 conditions (i.e., probes *D*, *F*, *G*, *K* and *N*), the VI-MGR has better Rank1 while in others the GSTI-HC generates better result.

However, the results of the proposed method are still comparable with a simple filtering scheme in comparison
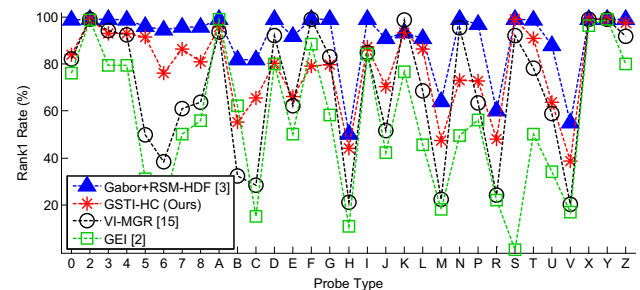

Fig. 5 Performance of gait recognition system for 31 probe types in OU-ISIR-B dataset

with Gabor+RSM-HDF. The Gabor+RSM-HDF has better Rank1 in 24 out of 31 clothing conditions due to using 40 Gabor kernel functions which needs more computational overhead.

The experimental results shown in Fig. 5 indicate that the proposed method can outperform the conventional methods on the OU-ISIR-B dataset. The results on the USF dataset (from Sect. 4.3) and OU-ISIR-B prove that our approach can extract energy of gait in a template-based feature more efficiently.

4.5 Experiments on USF

The performance of our method on USF dataset is being evaluated thoroughly in this section. To investigate efficacy of our method, we compare GSTI-HC with recently published methods including: Baseline [12], GEI [2], CGI [5], sparse bilinear discriminant analysis (SBDA) [16], Gabor-PDF+LGSR (LGSR) [1], VI-MGR [15], GEI+RSM-HDF

Table 5 Comparisons of the recognition performance on the USF dataset

Probe Set	A	B	C	D	E	F	G	H	I	J	K	L	Avg.
Probe Size	122	54	54	121	60	121	60	120	60	120	33	33	
Rank1 CCRs													
Baseline	73	78	48	32	22	17	17	61	57	36	3	3	40.96
GEI	90	91	81	56	64	25	36	64	60	60	6	15	57.66
CGI	91	93	78	51	53	35	38	84	78	64	3	9	61.69
SBDA	93	94	85	51	50	29	36	85	83	68	18	24	61.35
LGSR	95	93	89	62	62	39	38	94	91	78	21	21	70.07
VI-MGR	95	96	86	54	57	34	36	91	90	78	31	28	68.13
GeRSM	98	95	88	54	60	37	44	90	93	83	33	21	70.16
GbRSM	100	95	94	73	73	55	64	97	99	94	42	42	81.15
5/3GI	92	91	81	42	35	21	25	87	80	63	9	6	61.59
GSTI	97	95	93	53	49	41	46	96	97	92	33	21	72.25
Rank5 CCRs													
Baseline	88	93	78	66	55	42	38	85	78	62	12	15	64.54
GEI	94	94	93	78	81	56	53	90	83	82	27	21	76.23
CGI	97	96	94	77	77	56	58	98	97	86	27	24	79.12
SBDA	98	98	94	74	79	57	60	95	95	84	40	40	79.93
LGSR	99	94	96	89	91	64	64	99	98	92	39	45	85.31
VI-MGR	100	98	96	80	79	66	65	97	95	89	50	48	83.75
GeRSM	99	99	97	71	68	49	56	98	97	91	40	38	79.01
GbRSM	100	98	98	85	84	73	79	98	99	98	55	58	88.59
5/3GI	96	96	94	74	72	57	59	94	97	87	27	30	78.39
GSTI	100	96	97	78	76	72	74	99	99	99	42	36	85.64

(GeRSM) [3], Gabor+RSM-HDF (GbRSM) [3] and 5/3GI-Contour (5/3GI) [4].

The Rank1 and Rank5 accuracy rates are presented in Table 5. From Table 5, we can see that the average of Rank1 (or Rank5) in our method has been improved by 2% (or 0.3%) with respect to GEI+RSM-HDF [3] (or Gabor-PDF+ LGSR [1]). Meanwhile, the results of GSTI-HC are near to the Gabor+RSM-HDF [3].

In Gabor+RSM-HDF, Gabor-based templates are being used as gait features which the dimensionality of such templates is much higher than our feature template (according to different Gabor scales and orientations). Our method suffers from surface conditions where most of the related algorithms have weak performance too. To compute GSTI-HC, we need to define just a few parameters that indicate simplicity of the method.

4.6 Time complexity analysis

The complexity issues of the proposed method (GSTI-HC) can be further evaluated and compared with the recent methods such as (GEI and Gabor)+RSM-HDF [3]. Here, the timing costs of the gait recognition system compose of two

main parts, complexity of the filtering (or feature extraction) step and those of the classification step. To make a fair comparison with the methods in [3], we only compare the feature extraction step since the complexity of the classification step is the same. In the first step, the filtering of the input image takes the order of $O(I_{\text{spatial\_filt}} + 2\,I_{\text{temporal\_filt}})$ due to convolution with the spatial and oven/odd temporal impulse responses. Assuming the size of the input image as $W \times H$ and size of the proposed kernels as $w \times h$, the complexity of the filtering will be in the order of $O(I_{\text{filt}}) \approx O(WHwh)$ [9]. Moreover, the size of the temporal kernels ($w_t \times h_t = 1 \times T$) are smaller than the size of the spatial kernels ($w_s \times h_s = 6\sigma \times 6\sigma$) [9] and hence, $O(I_{\text{spatial\_filt}}) \gg O(2I_{\text{temporal\_filt}})$. With this assumption, the time complexity of the computing GEI, GSTI and Gabor templates is in the order of $O(1)$, $O(WHw_s h_s)$, $O(40WHw_s h_s)$ (due to using 40 set of Gabor kernel functions). It is clear that the computational overhead of the proposed filtering is 40 times faster than Gabor features. But the proposed filtering is slower than GEI feature while the performance of recognition has been improved (Table 5). More precisely, computational running time of a GSTI template is about 21.28 ms and hence 47 frames per

second[1] which is acceptable for real time gait recognition systems.

5 Conclusion

In this paper, a robust spatiotemporal template (i.e., GSTI) has been introduced for gait recognition. Template features are generated through an efficient filtering scheme in which the spatial and temporal information is extracted using appropriate impulse responses. To encode such information into final template, the computed features are aggregated and averaged over the gait cycles. In order to classify the proposed templates under challenging conditions, a hybrid decision level from the random classifiers has been merged to generate a robust classifier. The experimental results on two well-known datasets such as USF and OU-ISIR-B reveal that the proposed method improves the recognition rate in most challenging conditions in comparison with other state-of-the-art methods. The achieved recognition rate is 72.25% for Rank1 and 85.64% for Rank5 on the USF dataset which is promoted by at least 2% in Rank1 and 0.3% in Rank5 compared with the previous works. The available results demonstrate that the proposed method can be further improved to meet more real-life gait conditions.

References

1. Xu, D., Huang, Y., Zeng, Z., Xu, X.: Human gait recognition using patch distribution feature and locality-constrained group sparse representation. IEEE Trans. Image Process. **21**(1), 316–326 (2012)
2. Han, J., Bhanu, B.: Individual recognition using gait energy image. IEEE Trans. Pattern Anal. Mach. Intell. **28**(2), 316–322 (2006)
3. Guan, Y., Li, C.-T., Roli, F.: On reducing the effect of covariate factors in gait recognition: a classifier ensemble method. IEEE Trans. Pattern Anal. Mach. Intell. **37**(7), 1521–1528 (2015)
4. Atta, R., Shaheen, S., Ghanbari, M.: Human identification based on temporal lifting using 5/3 wavelet filters and radon transform. J. Pattern Recognit. **69**, 213–224 (2017)
5. Wang, C., Zhang, L.W.J., Pu, J., Yuan, X.: Human identification using temporal information preserving gait template. IEEE Trans. Pattern Anal. Mach. Intell. **34**(11), 2164–2176 (2012)
6. Lam, T.H.W., Cheung, K.H., Liu, J.N.K.: Gait flow image: a silhouette-based gait representation for human identification. J. Pattern Recognit. **44**(4), 973–987 (2011)
7. Dollar, P., Rabaud, V., Cottrell, G., Belongie, S.: Behavior recognition via sparse spatio-temporal filters. In: IEEE International Workshop VS-PETS, Beijing, China (2005)
8. Shabani, A.H., Clausi, D.A., Zelek, J.S.: Improved spatio-temporal salient feature detection for action recognition. In: Proceedings of the British Machine Vision Conference, Dundee (2011)
9. Ghaeminia, M.H., Shokouhi, S.B.: GSI: efficient spatio-temporal template for human gait recognition. Int. J. Biom. (IJBM) **10**, 29–51 (2018)
10. Ghaeminia, M.H., Badiezadeh, A., Shokouhi, S.B.: An efficient energy model for human gait recognition. In: DICTA, Gold Coast (2016)
11. Makihara, Y., Suzuki, A., Muramatsu, D., Li, X., Yagi, Y.: Joint intensity and spatial metric learning for robust gait recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI (2017)
12. 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. IEEE Trans. Pattern Anal. Mach. Intell. **27**(2), 162–177 (2005)
13. Hossain, M., Makihara, Y., Wang, J., Yagi, Y.: Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control. Pattern Recognit. **43**(6), 2281–2291 (2010)
14. Guan, Y., Li, C.-T., Hu, Y.: Random subspace method for gait recognition. In: IEEE International Conference on Multimedia Expo Workshops (2012)
15. Choudhury, S.D., Tjahjadi, T.: Robust view-invariant multiscale gait recognition. Pattern Recognit. **48**(3), 798–811 (2015)
16. Lai, Z., Xu, Y., Jin, Z., Zhang, D.: Human gait recognition via sparse discriminant projection learning. IEEE Trans. Circuits Syst. Video Technol. **24**(10), 1651–1662 (2014)

---

[1] The processing time has been measured with MATLAB 2013 running on a machine with Intel Core 2 Duo CPU P8400 2.20 GHz and 4 GB of DDR2 memory.