Biometric Gait Identification Systems: From Spatio-Temporal Filtering to Local Patch-Based Techniques

Mohammad H. Ghaeminia, Shahriar B. Shokouhi, and Abdollah Amirkhani

1 Introduction

Human's gait or style of walking is considered to be an effective biometric measure for identifying the individuals in public places for surveillance programs [1]. This is mainly attributed to the fact that the images of the walking can be easily acquired from a distance. Nevertheless, any biometric system based on human's gait identification can be affected by certain external factors such as the type of clothing, bagging condition, camera viewpoint, surface, and the aging [1, 2]. Despite these challenges, it is important to note that the manner of walking can still be relied upon as a unique biometric benchmark for solving identification problems [3]. The key issue in handling a gait under covariate factors is developing an efficient biometric system. In this paper, we propose the criteria for ranking the gait biometric systems and evaluating them accordingly.

Basically, the most common way to describe the gait without a predefined model is using template feature [1, 3–5]. For example, gait energy image (GEI) [1], general tensor discriminant analysis (GTDA) [6], gait flow image (GFI) [7], chrono-gait image (CGI) [8], and 5/3 gait image (5/3GI) [3] are some of the template features that have been developed by the researchers. In such methods, the sequence of silhouettes (or features) is aggregated into a single image called "template." Since the nature of the gait is a spatio-temporal process, such conversion has three main limitations [9, 10]: (1) removing temporal ordering of gait, (2) utilizing not an efficient human's motion model, and (3) aggregating gait defects in the template feature.

M. H. Ghaeminia · S. B. Shokouhi
Department of Electrical Engineering, Iran University of Science and Technology, Tehran, Iran
e-mail: mhghaeminia@elec.iust.ac.ir; bshokouhi@iust.ac.ir

A. Amirkhani (✉)
Department of Electrical Engineering, Iran University of Science and Technology, Tehran, Iran
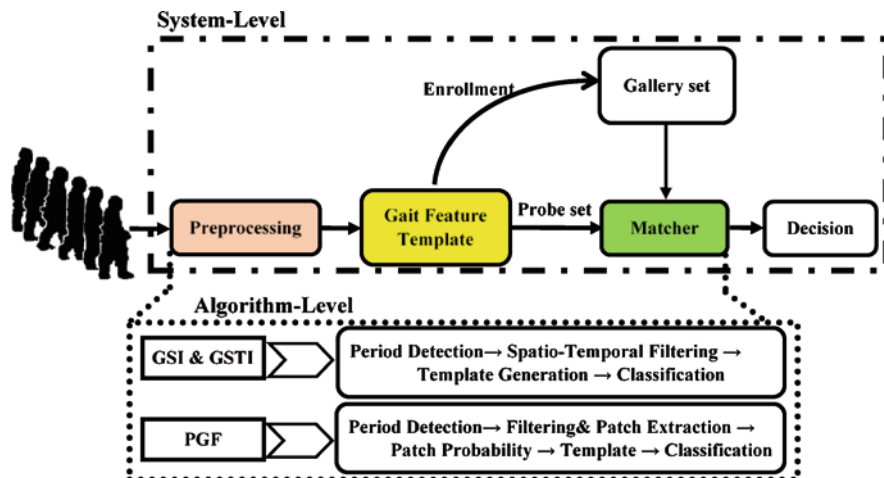e-mail: amirkhani@iust.ac.ir; amirkhani@ieee.org

Fig. 1 The main components of a gait biometric system

In order to resolve these issues, three separate algorithms have been proposed in this paper: (1) gait spatial image (GSI) features, (2) gait spatio-temporal image (GSTI) features, and (3) patch gait features (PGF). Each of these algorithms has a certain advantage in describing the rhythm of human walking; which will be examined in this paper. Besides the efficacy of the above methods in describing the gait, applying them in a real gait biometric system will be limited. Figure 1 shows different between the algorithm-level and system-level evaluations.

As shown in Fig. 1, a method of gait recognition is a subset of biometric system and hence, other benchmarks rather than recognition accuracy should be considered for evaluating the system. In this paper, the performance of gait biometric system is measured according to three proposed benchmarks: (1) recognition error rate, (2) computational and memory complexity, and (3) scalability.

Considering the recent state-of-the-art methods in gait recognition [3, 5, 9–11], the main contribution of the paper is as follows:

– Algorithm-level evaluation of the recent algorithms on three well-known datasets (USF [12], CASIA-B [13] and OU-ISIR-B [14]).
– Discussing on the benchmarks for validation of a gait biometric system.
– Evaluation of the proposed biometric systems according to given benchmarks.
– Providing an efficient tool for measuring the quality of gait system.

The rest of the paper is organized as follows: In next section we discuss on related gait recognition methods. The proposed algorithms (i.e., GSI, GSTI, and PGF) are explained shortly in Sect. 3. The proposed benchmarks for validating a gait biometric system are described in Sect. 4. Meanwhile, Sect. 5 provides the experimental results and discussion and the paper is concluded in Sect. 6.

2 Similar Works

The basis of efficient gait identification algorithms is gait template identification by means of template features [3, 5, 7, 15–18]. Therefore, the effective performance of a biometric system strongly depends on the type of algorithm used and the features selected. However, using appropriate features that are robust against the challenging covariates of human gait is itself a basic challenge. Moreover, some algorithms, despite being highly accurate in gait identification (e.g., [4, 19, 20]), have a large computational load and thus limited chance of development in a biometric system.

For appropriate gait identification, it is necessary to compute spatio-temporal features in one period and to compile them in a single image. This subject was investigated by Wang et al. [8] who presented a color template called chrono-gait image (CGI). Also, by extracting 5/3 wavelet features from silhouettes, Atta et al. [3] have recently proposed the 5/3 gait image (5/3GI) template, which not only has a small computational load but also a high gait identification accuracy. Fendri et al. [21] have dynamically selected the different sections of a silhouette which conform to gait covariates and which use the semantic classification scheme to identify different parts of human body. For detecting the gait template that matches human gait template, an effective spatio-temporal filter has been developed [9, 10, 22]. Recent research works have confirmed the existence of certain prominent regions in human gait (i.e., patches) that contain important information [11, 17]. By identifying these regions and describing human gait template on the basis of patches, gait identification accuracy can be increased.

From system perspective, as was stated earlier, there are certain limitations in using the above approaches in a biometric system. In other words, in system-level comparison, the measure of superiority is not just the accuracy of a particular method. For quantitative comparison between various systems, different methods have been developed; mostly on the basis of the performance of face and fingerprint biometrics [23–25]. For example, Grother et al. [23] and Olsen et al. [26] have developed standard ranking criteria for fingerprint biometric systems based on the quality of input sample. Also, Fernandez et al. [25] have examined and developed different evaluation measures, such as system error, in addition to the input sample quality. However, in gait identification, because of using noisy images recorded from a far distance, it is not a good idea to perform evaluations based on the quality of input sample. An appropriate criterion for evaluating the performances of different systems is to evaluate them based on system error [23, 25]. In addition to system error, two more criteria (i.e., computational load and scalability) will also be used in this paper to compare the performances of gait identification systems. The three selected approaches of GSI, GSTI, and PGF will be examined briefly in the next section and then the selection criteria will be thoroughly explained in Sect. 4.

3   Selected Approaches

Since human gait is essentially a spatio-temporal process, its proper description should be based on spatio-temporal features. For this purpose, a template feature based on spatio-temporal features has been developed in recent years. Of these methods, the three approaches of GSI [10], GSTI [9], and PGF [11] have been able to properly exploit the properties of corresponding features. In this section, the three cited methods will be introduced briefly and then the most effective classification technique for gait identification will be explained. It should be mentioned that the materials stated in this section pertain to algorithm-level development.

3.1   Preprocessing of Data

For the proper extraction of features, preliminary processing of information should be carried out on raw video recordings. The main preprocessing steps are as follows: resolving the background, extracting the foreground, and computing the gait period [10]. The Gaussian mixture model (GMM) has been proposed for modeling the background and subtracting the foreground from it [12]. Foreground images are obtained by subtracting the gait sequence from the background model. Following the extraction of foreground images, these images are normalized and made uniform by considering the obtained image centers [12]. In this way, silhouette images (foreground) with identical sizes and aspect ratios will be obtained. Then the gait period can be calculated by counting the pixels in the lower half of the silhouette image in each frame [10]. Because the values of corresponding pixels, when a walker's two feet move away from or get closer to each other, will have local maxima and minima. By computing the local extremums, the gait period (T) is easily obtained by considering the average distance between two successive maxima or minima.

3.2   Gait Spatio-Temporal Filtering

As we know, the sequence of human gait is a type of spatio-temporal signal in 3-D space, where x and y are two spatial coordinates and t is the temporal coordinate. Since there is no displacement in the y direction, in the GSI approach, the gait sequence has been considered in the x-t space. For this purpose, a directional filter has been developed for determining the direction of gait in this space [10]:

$$\begin{cases} Dir, L_1 = R_{A1} - R_{B2}; & Dir, L_2 = R_{A2} + R_{B1} \\ Dir, R_1 = R_{A1} + R_{B2}; & Dir, R_2 = R_{A2} - R_{B1} \end{cases}. \tag{1}$$

In the above equations, Dir,L_i and Dir,R_i ($i = 1, 2$), respectively, denote the main kernels for determining gait directions to the left or right in two different strengths. Also, R_{Ai} and R_{Bi} represent the impulse responses of spatio-temporal filter; whose spatial part is obtained by differentiating the Gaussian signal $G_m(x,\sigma) = \dfrac{\partial^m}{\partial x^m}\left\{\dfrac{1}{2\pi\sigma^2}\exp\left(-x^2\big/2\sigma^2\right)\right\}$, $x \in R$ (m and σ, respectively, denote the spatial order and scale) and whose temporal part is obtained from the exponential signal $F_n(t) = \left(\dfrac{1}{n!} - \dfrac{(kt)^2}{(n+2)!}\right)(kt)^n \exp(-kt)$, $0 \leq t \leq T$ (n and k, respectively, indicate the temporal order and scale). Now in each frame, the gait direction (or gait energy, EN_L and EN_R) is easily determined by the convolution of impulse responses in the silhouette image and the summation of directions:

$$\begin{cases} EN_L(t) = R_{L1}^2(t) + R_{L2}^2(t) \\ EN_R(t) = R_{R1}^2(t) + R_{R2}^2(t) \end{cases}, \tag{2}$$

that is, R_{Li} and R_{Ri} indicate the response of filtering obtained from convolution of silhouette with kernels in Eq. (1). The final gait direction in each frame is determined by subtracting the responses. Also, the GSI template is obtained by collecting the responses in every half-period and by averaging the responses of two successive half-periods:

$$GSI = \frac{1}{2}\sum_{k=1}^{2} EN_k, \tag{3}$$

in which, EN represents the sum of the responses in every half-period.

In the GSI approach, because of ignoring human gait in the vertical direction and also using a network of nonlinear filters, the computed responses will be rather unrealistic. So by employing a simple and realistic template in the GSTI method, the researchers have attempted to obtain a more precise human gait template. First, it has been assumed that an input video is a circuit network and that every pixel in it is equivalent to a resistor which is connected to adjacent pixels in the current frame and preceding frame. Also, in the network, the luminosity of each pixel has been considered as the load of a capacitor; and the goal is to calculate current flux in the circuit network. By solving the relevant equations and using Fourier transformation, the spatial impulse response (S_{IR}) and the temporal impulse response (T_{IR}) for human gait are obtained as follows [9]:

$$S_{IR} = LoG(x,y;\sigma) = \frac{\partial^2}{\partial x^2}G(x,y;\sigma) + \frac{\partial^2}{\partial y^2}G(x,y;\sigma), \tag{4}$$

$$T_{IR} = \frac{1}{2}a_0 + a_1\cos\left(\omega_\tau(t-\tau)\right) \quad t \geq 0. \tag{5}$$

In the above equations, $G(.)$ indicates the standard Gaussian signal, a_0 and a_1 are the coefficients of Fourier series, $\omega_\tau = \pi/T$ (T being the gait period), and $\tau = T/2$. It should be mentioned that in computing Eqs. (4) and (5), some limitations have been applied on Fourier transform solution to make it more compatible with human gait model [9]. Furthermore, the temporal response of Eq. (5) is an even signal; and after computing its corresponding odd signal, the squares of temporal filter responses are added together. After calculating the impulse responses of spatial and temporal filters, the shadow image is first convoluted with a spatial filter and then with a temporal filter in order to obtain the corresponding response. Eventually, similar to the GSI method, the responses in every half-period are collected and the average energy in a period is calculated as the GSTI template.

3.3 Local Patch-Based Gait Features

The idea of patch-based approach is different from the previous methods. Here we are trying to seek the important regions in filter responses and to keep them in the final template. Therefore, there are three major steps in the PGF method [11]: (1) filtering the silhouette images and extracting the local patches, (2) calculating the probability distributions of patches, and (3) computing the PGF template. In the first step, the local patches, $p(x,y,t)$, are extracted by seeking the extremum points in a local sliding window in 3-D space. In this step, the response of Gabor filter for each silhouette image is calculated and becomes a basis for seeking the patches. In the second step, the histograms of patch distributions along the $x, y,$ and t directions are computed separately:

$$
\begin{aligned}
h_x &= hist\left[p\left(x_i, y, t \right) \right]_{i=1\ldots H} \\
h_y &= hist\left[p\left(x, y_i, t \right) \right]_{i=1\ldots W}, \\
h_t &= hist\left[p\left(x, y, t_i \right) \right]_{i=1\ldots T}
\end{aligned}
\tag{6}
$$

Here, H, W, and T are the image height and width and the gait period, respectively. The above triple histogram is unique to each individual and can be used as gait signature [11].

Now in the third step, by considering the above probability distributions as coordinate weights, the final template can be obtained. Here, the set of 40D Gabor responses (with five different scales and eight directions) is applied on the GEI, and the weighted coordinates of each pixel are added to the responses.

$$
PGF = \left[RG^T, h_i k\left(X_i \right), h_j k\left(Y_j \right), h_k k\left(t_k \right) \right]_{i=1\ldots H, j=1\ldots W, k=1\ldots T}^T,
\tag{7}
$$

in which, RG is response of the 40D Gabor filtering, $k(.)$ is the isotropic kernel for allocating more weight to the pixels near the image center or half-period, and h_l is

the weight resulting from the relevant histogram index. According to Eq. (7), the PGF template is an augmented Gabor template in which the spatio-temporal coordinates of patches have been added to filter response based on their probability distribution.

3.4 Classification

The introduced methods highlighted the manner of extracting powerful features from behavior template. However, the usefulness of recent approaches depends not only on feature extraction but also on the type of classification. The simplest method of classifying a template is to find the closest feature to a gallery determined by direct template matching. For example, the 1-nearest neighbor classification technique (1-NN) assigns a label to its nearest neighbor (with the least Euclidean distance) [8]. However, since feature dimensions are much more than the number of test samples and some features are overlooked, we have an undersample problem (UPS), which diminishes the recognition accuracy [6]. A common method for dealing with this problem is to use the two-step technique of PCA+LDA [10]. This technique is relatively simple and provides usual accuracy; because by applying the PCA matrix, the 2-D structure of data collapses and data turn into a one-dimensional vector. Despite the existence of various challenges, the random subspace method (RSM) has been able to identify gait templates with high accuracy [11, 27]. This classification method achieves weak classification by taking random samples from feature space and by making a decision based on a set of weak decisions. Considering the utility of the RMS classification technique, it has been employed in this paper. More details on this classification scheme can be found in the work of Guan et al. [27].

4 System-Level Evaluation Tools

So far, we explored the algorithms used for identifying behavior templates. However, quantitative measures are needed to evaluate the performances of these algorithms. By means of these quantitative criteria we can determine the utility and fidelity of different identification approaches. Every biometric system is evaluated based on different parameters so as to obtain its fidelity under different conditions. The existing approaches for evaluating the fidelity of biometric systems have been developed based on the performance of face and fingerprint biometrics [25, 28]; however, this procedure can also be extended to a gait biometric system.

In behavioral biometry systems the results are usually reported in terms of identification accuracy [28]; because the number of test sets is limited and the individuals in probe set are always a subset of the individuals in gallery. Moreover, one of the main challenges of behavioral biometry systems is the decline of their

performance in uncontrolled or unpredictable conditions [28]. Basically, the evaluation of every biometric system requires three parameters [28]: (1) evaluating the quality of input data (sensor quality), (2) guaranteeing the performance of algorithm, and (3) providing a utility benchmark. In this paper, the quality of input data is the same for all the evaluated algorithms; because the same standard datasets have been used for all of them. Also, the utility of the proposed algorithms in properly identifying gait templates has been confirmed. Therefore, the above three criteria [25], which are intra-system evaluation measures, are not sufficient for evaluating the performance of a gait biometric systems.

We define three important criteria for the extra-system evaluation of biometry systems: (1) system error, (2) execution speed and the amount of computer memory used, and (3) scalability. Each of these criteria will be subsequently explained.

4.1 False Match Rate (FMR) and False Non-match Rate (FNMR) Criteria

These two benchmarks are able to compute the amount of error for a system. The FMR criterion is an empirical estimation of probability (percent occurrence), by which a system falsely accepts an input template. Also, FNMR is an empirical estimation of probability, by which a system falsely rejects a matched input sample. FMR and FNMR criteria are equivalent to FAR and FRR, respectively; and for their computation procedures one can consult the available references [25]. Ordinarily, when comparing two biometry systems, the system which has a lower FRR at the same FAR is the better performing system.

The recognition error tradeoff (DET) graph has been used to graphically display system error; because it is not commonplace to use the ROC diagram in the evaluation of biometry systems [29]. The horizontal and the vertical axes of the DET graph, respectively, show the FAR and FRR values at different accuracies and Ranks. In DET diagram standards, the vertical axis is in percentage (%) and the horizontal axis has a logarithmic scale.

4.2 Computational and Memory Complexity

The time needed for identifying an individual and the amount of computer memory used for this purpose are determined by this criterion. In fact, time complexity and used memory constitute suitable measures for evaluating a system. When comparing two systems with the same error, the one with higher speed and less memory usage is considered a better system. A biometric system needs memory to compute the templates and to save the features.

4.3   Scalability

This criterion expresses the performance of a system in huge set of data. To ensure a level of scalability that can support a very large dataset, it is necessary to use very strong support servers of high capacity. Assuming that the processing power of each server is one unit, the total processing power of ten servers with integrated processing capability is not 10 units. Even by increasing hardware and software resources, there will be a problem with system performance at large volumes of data; because with the rise in the number of samples, the degree of differentiability between samples decreases severely and system error goes up. In other words, a system's robustness against scalability indicates the reliability of that system under real conditions. Obviously, in comparing two systems with equal errors, the one with a lower error in a high-population dataset is a better system. Published papers do not present a specific criterion for assessing a system's robustness to scalability. In this paper, average Rank-1/Rank-5 accuracy achieved in three datasets will be used to evaluate the scalability criterion.

5   Results and Discussion

In this section, the results of algorithm- and system-level evaluations will be presented. Algorithm-level evaluation has been thoroughly discussed in the cited references [9–11]. In this paper, the results of algorithm-level evaluations are concisely presented and then the results of system-level evaluations are discussed. An algorithm-level comparison is carried out in order to assess the quality of the examined algorithms in gait template recognition; because an exact biometric system is dependent on the type of algorithm designed for it. The algorithms that have been selected for evaluation and comparison are as follows: GEI [1], CGI [8], 5/3GI [3], STIP [30], locality-constrained group sparse representation (LGSR) [20], view-invariant multiscale gait recognition (VI-MGR) [2], RSM [27], local patch-based subspace ensemble learning algorithm (LPSELA) [4], two-stream generative adversarial network (TS-GAN) [31], clothing-invariant gait recognition (CLIGR) [15], and local feature at the bottom layer based on capsule network (LBC) [18].

In this paper, three famous datasets (USF [12], CASIA-B [13], and OU-ISIR-B [14]) are used for algorithm evaluation. The USF dataset contains 122 individuals with five walking forms, CASIA-B has 124 individuals with ten walking forms, and the OU-ISIR-B dataset includes 48 individuals with 32 different forms of walking. With different test cases in the gallery and the probe sets, there are a total of 3176 unique tests for these three datasets, (1080 tests for USF, 1240 tests for CASIA-B, and 856 tests for OU-ISIR-B). For evaluating the examined algorithms, each of the above datasets has considered several gait challenges. The USF set has addressed the variations of walking manner with the change of shoes, viewing conditions, surface, bag, and time parameters. The CASIA-B set has considered gait variations with respect to "normal," "bag," and "cloth" conditions at different video recording

angles. And the OU-ISIR-B set has only dealt with the change of clothing, which is the most important and common parameter related to gait. In the following, the results of algorithm-level evaluations on the three mentioned datasets will be briefly discussed.

5.1 Algorithm-Level Evaluations

To verify the effective performance of the proposed algorithms, the first test is carried out on the CASIA-B dataset. Considering the high quality of silhouette images in this dataset, there is no need to use powerful classification techniques such as RSM; and satisfactory accuracy will be obtained by applying 1-NN. For a brief and accurate evaluation of the examined approaches, only Rank-1 values are presented. Table 1 shows the performance Rank-1 values achieved by the proposed algorithms.

Compared to the first group of algorithms, the average accuracy of Rank-1 is higher than that of GEI, CGI, and 5/3GI methods. The average accuracy of GSTI is close to that of CGI, which shows an acceptable performance. Also, in five tests (out of 6), the proposed approaches achieve a higher performance relative to GEI and CGI. The results of 5/3GI for certain experiments have not been revealed and their averages cannot be evaluated.

In the second part of Table 1, the satisfactory performance of the proposed algorithms can be revealed by comparing the results through multiview-based methods. Compared to STIP algorithm, the proposed algorithms are superior in three tests out of six; and in some other tests the results are very close. In comparison with the VI-MGR method, in two cases our algorithms have achieved greater performance. However, because of using synthetic information and relatively powerful classification technique, the VI-MGR algorithm has attained a higher accuracy. In summary, in the CASIA-B dataset, the proposed gait identification algorithms have achieved satisfactory or better recognition accuracies than other recent approaches.

In the second step, the performances of the proposed gait recognition algorithms in the OU-ISIR-B dataset are investigated. Here, the Rank-1 results obtained by

Table 1 Evaluating the Rank-1 recognition rates achieved by the proposed algorithms and other approaches in CASIA-B dataset

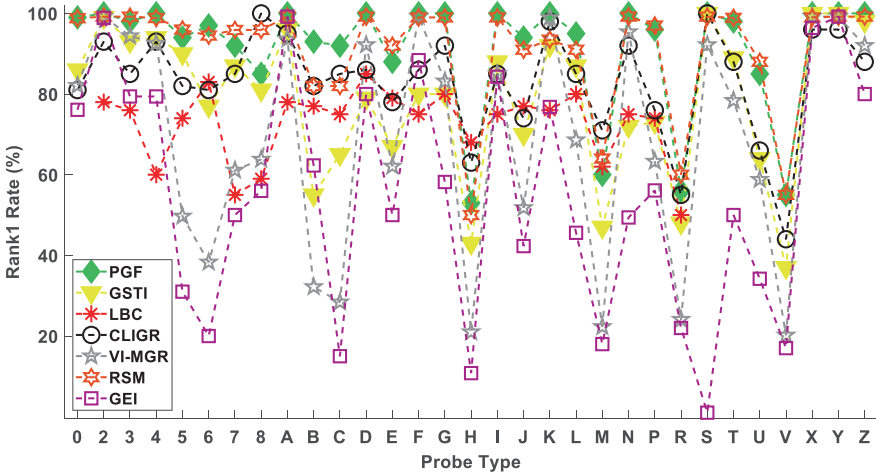| Gall-Prb	nm-nm	nm-bg	nm-cl	bg-bg	bg-cl	cl-cl	Avg.
GEI [1]	91.57	37.8	25.04	91.2	17.47	97.22	60.05
CGI [8]	88.06	51.93	46.88	89.81	32.52	95.37	67.43
5/3GI [3]	98	–	–	73	66	–	–
STIP [30]	95.4	60.9	52	73	29.8	70.6	63.62
VI-MGR [2]	100	89	76	79.03	–	71.77	–
TS-GAN [31]	–	–	–	–	–	–	63.1
GSI [10]	95.87	66.46	41.41	90.5	29.5	100	70.62
GSTI [9]	91.8	57.5	39.6	92.5	26.4	98	67.64
PGF [11]	92.4	65.7	50.8	91.5	33.5	97.5	71.9

Fig. 2 Evaluating the performances of various algorithms in the OU-ISIR-B dataset

Table 2 Comparing the average Rank-1 results obtained by the proposed and more recent algorithms in the OU-ISIR-B dataset

Algorithms	Avg. of Rank-1 (%)
GEI [1]	55.6
RSM [27]	90.7
VI-MGR [2]	69.03
CLIGR [15]	81.71
LBC [18]	73.33
GSTI [9]	77.25
PGF [11]	90.62

these algorithms in the probe set (which includes 31 different combinations of clothing) have been presented. However, due to the proximity of the results of GSI and GSTI algorithms, the GSI results have not been listed. Figure 2 graphically displays the fidelities of the examined algorithms when considering different clothing conditions. Also, Table 2 shows the average accuracies achieved by the proposed algorithms in the OU-ISIR-B dataset.

According to Fig. 2, in five tests (probes D, F, G, K, and N), the VI-MGR algorithm achieves a higher accuracy than GSTI. Also, in comparison with RSM, the results of GSTI can still be observed. Compared to the RSM approach, the GSTI algorithm uses a simpler filter structure; because instead of the 40-component Gabor functions, it exploits spatial and temporal impulse responses.

In addition, the PGF algorithm (which has a patch-based structure) enjoys a higher accuracy and fidelity relative to the GSTI algorithm. More precisely, in five tests out of 31 (probes 2, S, X, Y, and Z), the recognition accuracy of PGF is 100% and the algorithm has been able to identify all the individuals in the probe set. Also in 2 probes out of 31 (probes H and M), the patch-based methods exhibit weaker

Table 3 Rank-1 and Rank-5 results of different algorithms obtained for the USF dataset

Exp.	A	B	C	D	E	F	G	H	I	J	K	L	Avg.
Method	Rank-1 performance												
LGSR	95	93	89	51	50	29	36	85	83	68	18	24	70.07
RSM	100	95	94	73	73	55	54	97	99	94	41	42	81.15
VI-MGR	95	96	86	54	57	34	36	91	90	78	31	28	68.13
LPSELA	95	91	78	66	59	46	52	93	88	69	30	27	70.49
GSI	92	95	86	38	31	17	24	79	83	76	25	19	58.44
GSTI	97	95	93	53	49	41	46	96	97	92	33	21	72.25
PGF	100	96	98	62	59	43	46	100	99	94	28	30	76.01
	Rank-5 performance												
LGSR	99	94	96	89	91	64	64	99	98	92	39	45	85.31
RSM	100	98	98	85	84	79	73	98	99	98	55	58	88.59
VI-MGR	100	98	96	80	79	66	65	97	95	89	50	48	83.75
LPSELA	100	96	93	84	83	73	74	95	96	89	64	52	86.09
GSI	98	95	95	67	52	43	56	97	95	95	43	37	76.60
GSTI	100	96	97	78	76	72	74	99	99	99	42	36	85.64
PGF	100	98	98	80	77	77	60	100	100	99	48	45	86.59

performance relative to more recent approaches (CLIGR [15] and LBC [18]). Nevertheless, the performance of other methods under such conditions is very low. The results in the above figure indicate a performance below 60% for the PGF algorithm in probes H, M, R, and V. This is mainly due to the coverage of body parts during patch extraction. In these conditions, the proposed approaches cannot extract enough information from lower body parts. The results of Fig. 2 and Table 2 show the relative superiority of the proposed algorithms compared to the other algorithms.

In the third step of this section, the complete Rank-1 and Rank-5 results of different algorithms obtained for the USF dataset have been discussed.

According to Table 3, the average Rank-1 and Rank-5 results in the GSTI and PGF approaches have improved relative to all other methods (except RSM). Also, the results of PGF are very close to those of the RSM algorithm. The proposed methods are vulnerable to the change of "surface" and "time" (probes F, G, K, and L), but their performance is still comparable to all the other approaches. The Rank-5 recognition rates of the proposed algorithms have improved in most of the tests. More precisely, in three tests out of 12 (probes A, H, and I), the recognition accuracy of PGF is 100% and all the relevant subjects have been identified. Also, in the rest of the tests, the Rank-5 values are close to their maximum, and subjects can be identified with a high accuracy. The obtained results indicate that the average recognition success of PGF has increased by about 5% relative to other methods (e.g., LGSR [20], VI-MGR [2], and LPSELA [4]).

So the proposed algorithms are appropriate tools for human gait identification and provide a suitable basis for recognition of individuals under challenging and more complicated conditions.

5.2 System-Level Evaluations

In the previous section, the proposed algorithms for the recognition of behavioral templates were reviewed. However, quantitative measures are needed to evaluate the system-level performance of these algorithms. The utility and the fidelity of the proposed biometric systems can be obtained by means of these quantitative criteria. As it was stated in Sect. 4, the three criteria of system error, algorithm execution speed and the amount of required computer memory, and scalability have been used in this paper. The amount of system error is computed by using the FAR and FRR error parameters and evaluated by means of the detection error tradeoff (DET) diagram. In biometric systems for face recognition, (normally in the authentication mode), the above errors [33] are expressed as "FRR 1% @ FAR 1/10,000". However, since behavioral detection systems act in the identification mode, error calculation will be a bit different. Suppose the mth individual from a probe set is compared with all the individuals of a gallery (G individuals) and that these individuals are arranged descending. Now if the true (or corresponding) individual is located in the ith position, then in computing the Rank-r parameter, the values of FAR and FRR will be calculated as follows:

$$\begin{cases} \text{FAR} & = & \dfrac{i-1}{G}, & \text{FRR} & = & \dfrac{r-i}{G}. & \text{if } i \leq r \\[2mm] \text{FAR} & = & \dfrac{r}{G}, & \text{FRR} & = & \dfrac{i-r}{G}. & \text{if } i > r \end{cases} \tag{8}$$

Now by changing the value of r and considering the values of Rank-r and by recalculating the error values according to the above equation, the DET graph is obtained. In the DET diagram standards, the vertical axis shows the FRR percentage and the horizontal axis has a logarithmic scale. Due to its popularity and comprehensiveness, the USF dataset has been used here for error estimation; and the obtained error values are the weighted average errors in the whole probe sets and in different tests. Table 4 shows the weighted average error values with respect to Rank-1 and Rank-5 in the proposed systems. Also, Fig. 3 illustrates the DET diagrams for these systems in the USF dataset.

According to Table 4, the PGF system has the least values of FRR and FAR; so it can be considered as a powerful and low-error system. However, a proper comparison cannot be made based on the values in the above table; because in comparing two biometric systems, the better system will have a lower FRR at the same FAR

Table 4 The specifications of the proposed biometric systems versus Rank-1 and Rank-5 errors

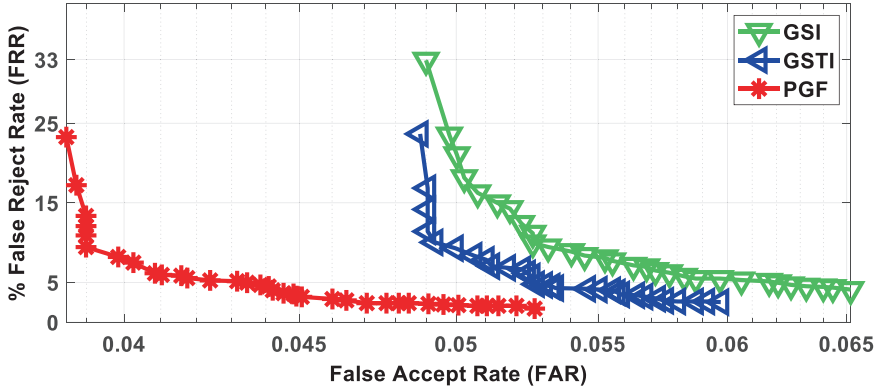| Biometric system	@Rank-1	@Rank-5
GSI	FRR 33% @ FAR 49/1000	FRR 16% @ FAR 51/1000
GSTI	FRR 24% @ FAR 48/1000	FRR 10% @ FAR 49/1000
PGF	FRR 23% @ FAR 38/1000	FRR 9% @ FAR 40/1000

Fig. 3 The DET diagrams resulting from the three proposed systems

value. For a more appropriate evaluation, the DET diagram has been used. As Fig. 3 shows, the PGF system has a much lower error than the other two systems. At a specific FAR (FAR = 10–1.3 ≈ 50/1000), the FRR value of PGF (2%) is less than that of GSTI (10%), which itself is less than that of GSI (18%). A 16% difference exists between the FRR values of PGF and GSI systems. So the PGF system has a lower gait recognition error.

Now the systems are evaluated in terms of processing time and required computer memory. The most important part of the proposed systems is their input image filter. The GSI system uses a quadruple directional filter (Eqs. (1) and (2)), the GSTI system uses a dual spatio-temporal filter (Eqs. (4) and (5)), and the PGF system employs a 40-component Gabor filter (Eq. (7)). Now if the image dimensions are *WxH* and the kernel dimensions are *wxh*, the approximate computation time of image filter will be of order $O(I_{filt}) \approx O(WHwh)$ and the memory used will be of order $O(I_{filt}) \approx O(WH + wh)$ [10]. Therefore, the time complexities of GSI, GSTI, and PGF systems in a dataset with the number of training and test individuals equal to $n_{te} + n_{tr}$ and with a gait period of T will, respectively, be $O(4(n_{te} + n_{tr})TWHwh)$, $O((n_{te} + n_{tr})TWHwh)$, and $O(10(T + 1)(n_{tr} + n_{te})WHwh)$ [9–11]. Also, by assuming $WH >> wh$, the amounts of memory required by these systems are $O(10TWH(n_{te}+n_{tr}))$, $O(3TWH(n_{te} + n_{tr}))$, and $O(10(T + 1)WH(n_{te} + n_{tr}))$, respectively [9–11]. Obviously, time complexity and required memory depend on three main parameters: (1) the size of input image (*WxH*), (2) the total number of individuals in the gallery and probe sets ($n = n_{tr} + n_{te}$), and (3) the average gait period (T). It should be mentioned that the dimensions of filter kernel in the examined systems have been considered as $w \times h = 39 \times 39 = 1521$ [9]. If the same dimensions are considered for input image ($W = H = N$), then the required execution time, t, for the proposed systems in the whole dataset will be as;

$$\text{GSI}: \quad t \approx O\left(4 \times 1521 \times nTN^2\right)$$
$$\text{GSTI}: \quad t \approx O\left(1521 \times nTN^2\right) \tag{9}$$
$$\text{PGF}: \quad t \approx O\left(10 \times 1521 \times (T+1)nN^2\right).$$

Also, the amounts of memory (M) used by the proposed systems are as follows:

$$\text{GSI}: \quad M \approx O\left(10 \times nTN^2\right)$$
$$\text{GSTI}: \quad M \approx O\left(3 \times nTN^2\right) \tag{10}$$
$$\text{PGF}: \quad M \approx O\left(10 \times (T+1)nN^2\right).$$

Suppose the number of individuals in a dataset (considering the three evaluated datasets) has three default values of $n = \{150, 500, 1000\}$, which correspond to small, medium, and large datasets, respectively; and suppose the gait period has two default values of $T = \{30, 70\}$, which, respectively, specify small and large sequences of recorded human gait. With these assumptions, Figs. 4 and 5 illustrate the systems' computational loads and required memories versus the dimensions of input images and in terms of predefined parameters.
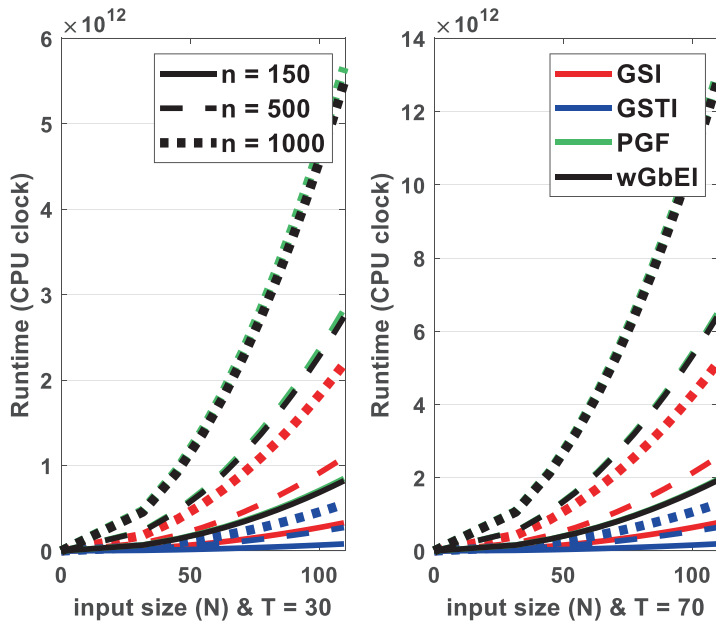

Fig. 4 Computation times required by the proposed systems versus image dimensions
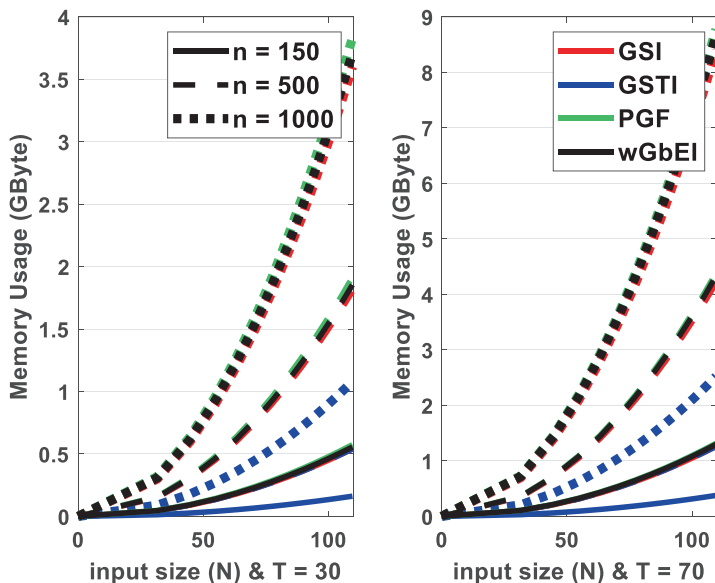
Fig. 5 The amounts of memory (in *GB*) required by the proposed systems at various settings

Table 5 Average system accuracies for all the tested datasets (3176 data in total)

Biometric system	Rank-1 (%)	Rank-5 (%)
GSI	70.89	84.32
GSTI	68.96	82.47
PGF	74.16	85.28

According to Figs. 4 and 5, GSTI is the fastest human identification system that requires less memory. Also, the slowest system is PGF, which uses a memory capacity which is almost equal to that of the GSI system. So in situations where system error is less important, the GSTI system is recommended; and in applications in which system error should be minimized, the PGF system will be the best choice. In these conditions, there should be a tradeoff between system error (Fig. 3), computation speed (Fig. 4), and the amount of memory used (Fig. 5). By executing the algorithms in MATLAB software (ver. 8.3.0, published in 2014) using a PC with Intel (R) Core i7 processor, 8 GB of RAM, and 2.39 GHz operating frequency, the computation times and the required memories for the GSTI, GSI, and PGF algorithms were obtained as 56, 14, and 5.5 frames/s and 3.6, 1.2, and 3.75 GB, respectively [9–11].

The last parameter to evaluate is system's robustness to dataset scalability. In evaluating biometric gait identification systems, the criterion of scalability has not been usually discussed [5]. So in order to measure the degree of robustness against scalability, the average results of the examined systems for all the tested datasets are

Table 6 Evaluating the fidelity of the proposed biometric systems

System benchmark	Superior system	Second superior system	Weakest system
FAR & FRR	PGF	GSTI	GSI
Time/memory	GSTI	GSI	PGF
Scalability	PGF	GSI	GSTI

computed in this section. A total of 3176 tests have been carried out (1080 tests on the USF dataset, 1240 tests on the CASIA-B dataset, and 856 tests on the OU-ISIR-B dataset). The average results have been tabulated in Table 5.

According to Table 5, the PGF system achieves a higher Rank-1/Rank-5 accuracy and performs better when a large dataset is involved. Also, the GSTI approach proves to be the weakest in this case. So the PGF system will have a better performance in a large-volume dataset.

5.3 System-Level Performance Evaluation Criteria

In the preceding section, the proposed systems were thoroughly evaluated. From the perspective of FAR and FRR errors, a system may produce minimal error but have a large computational load and require extensive memory. Also, error optimization is equivalent to system's robustness against the increase of dataset volume, and vice versa. In evaluating the proposed systems, Table 6 ranks the proposed biometric gait identification systems.

According to the above table, if the amount of computer memory used and the computation time are not important factors, the PGF system will be the most suitable, because of its lower error. Also, when the execution speed of algorithm and the memory it uses are of utmost importance, the GSTI system is recommended. In this case, the GSI system can also be used as a substitute system for GSTI; both of these systems have a greater computation speed and require less memory compared to the PGF system.

6 Conclusion

In this paper, effective gait template detection algorithms were reviewed and the evaluation criteria for the biometric systems related to these algorithms were presented. Today, there are three major challenges in properly identifying gait templates: (1) using temporal information, (2) making the algorithms compatible with human gait templates, and (3) eliminating extra information and noise from final features. As was previously mentioned, the three algorithms of GSI, GSTI, and PGF have been able to deal with the above three problems and to improve the fidelity of gait template

recognition in most cases. However, for a more exact evaluation of performance, these algorithms have been compared from the standpoint of biometric systems. For this purpose, three criteria of system error, computational load, and scalability have been defined and used to evaluate system performances. The fastest and the most accurate biometric systems are GSTI and PGF, respectively. But when the computational load is limited and a relatively low system error is expected, the GSI system could be reliably used as an alternative system. In general, the criteria presented in this paper can be used as a standard measure to evaluate biometric gait identification systems and to help us choose the right system suitable for real conditions.

References

1. Han, J., & Bhanu, B. (2006). Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 28*(2), 316–322.
2. Choudhury, S. D., & Tjahjadi, T. (2015). Robust view-invariant multiscale gait recognition. *Journal of Pattern Recognition, 48*(3), 798–811.
3. Atta, R., Shaheen, S., & Ghanbari, M. (2017). Human identification based on temporal lifting using 5/3 wavelet filters and radon transform. *Journal of Pattern Recognition, 69*, 213–224.
4. Ma, G., Wua, L., & Wang, Y. (2017). A general subspace ensemble learning framework via totally-corrective boosting and tensor-based and local patch-based extensions for gait recognition. *Pattern Recognition, 66*, 280–294.
5. Rida, I., Almaadeed, N., & Almaadeed, S. (2019). Robust gait recognition: A comprehensive survey. *IET Biometrics, 8*(1), 14–28.
6. Tao, D., Li, X., Wu, X., & Maybank, S. J. (2007). General tensor discriminant analysis and Gabor features for gait recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*(10), 1700–1715.
7. Lam, T. H., Cheung, K. H., & Liu, J. N. (2011). Gait flow image: A silhouette-based gait representation for human identification. *Journal of Pattern Recognition, 44*(4), 973–987.
8. Wang, C., Zhang, J., Wang, L., Pu, J., & Yuan, X. (2012). Human identification using temporal information preserving gait template. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 34*(11), 2164–2176.
9. Ghaeminia, M. H., & Shokouhi, S. B. (Feb. 2019). On the selection of spatio-temporal filtering with classifier ensemble method for effective gait recognition. *Journal of Signal, Image and Video Processing, 13*(1), 43–51.
10. Ghaeminia, M. H., & Shokouhi, S. B. (2018). GSI: efficient spatio-temporal template for human gait recognition. *International Journal of Biometrics (IJBM), 10*(1), 29–51.
11. Ghaeminia, M. H., Shokouhi, S. B., & Badiezadeh, A. (2019). A new spatiotemporal patch-based feature template for effective gait recognition. *Multimedia Tools and Applications (MTAP), 79*(1-2), 713–736. https://doi.org/10.1007/s11042-019-08106-x.
12. Sarkar, S., Phillips, P. J., Liu, Z., Vega, I. R., Grother, P., & Bowyer, K. W. (2005). The humanID gait challenge problem; data sets, performance, and analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(2), 162–177.
13. Yu, S., Tan, D., & Tan, T. (2006). A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *Proceeding of 18th International Conference on Pattern Recognition (ICPR)*. Hong Kong: IEEE Computer Society.
14. Iwama, H., Okumura, M., Makihara, Y., & Yagi, Y. (2012). The OU-ISIR gait dataset comprising the large population dataset and performance evaluation of gait recognition. *IEEE Transactions on Information Forensics and Security, 7*(5), 1511–1521.

15. Ghebleh, A., & Moghaddam, M. E. (2018). Clothing-invariant human gait recognition using an adaptive outlier detection method. *Journal of Multimedia Tools and Applications (MTAP), 77*(7), 8237–8257.
16. Hong, S., Lee, H., & Kim, E. (2013). Probabilistic gait modelling and recognition. *IET Computer Vision, 7*(1), 56–70.
17. Ben, X., Gong, C., Zhang, P., Jia, X., & Meng, Q. W. (2019). Coupled patch alignment for matching cross-view gaits. *IEEE Transactions on Image Processing, 28*(6), 3142–3157.
18. Xu, Z., Lu, W., Zhang, Q., Yeung, Y., & Chen, X. (Feb. 2019). Gait recognition based on capsule network. *Journal of Visual Communications and Image Representation, 59*, 159–167.
19. Huang, Y., Xu, D., & Nie, F. (2012). Patch distribution compatible semisupervised dimension reduction for face and human gait recognition. *IEEE Transactions on Circuits and Systems for Video Technology, 22*(3), 479–488.
20. Xu, D., Huang, Y., Zeng, Z., & Xu, X. (2012). Human gait recognition using patch distribution feature and locality-constrained group sparse representation. *IEEE Transactions on Image Processing, 21*(1), 316–326.
21. Fendri, E., Chtourou, I., & Hammami, M. (2019). Gait-based person re-identification under covariate factors. *Journal of Pattern Analysis and Application (PAA), 22*(4), 1629–1642. https://doi.org/10.1007/s10044-019-00793-4.
22. Tong, S., Fu, Y., & Ling, X. Y. (2018). Multi-view gait recognition based on a spatial-temporal deep neural network. *IEEE Access, 6*, 57583–57596.
23. Grother, P., & Tabassi, E. (April 2007). Performance of biometric quality measures. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*(4), 531–543.
24. Teixeira, R. F., & Leite, N. J. (2017). A new framework for quality assessment of high-resolution fingerprint images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(10), 1905–1917.
25. Alonso-Fernandez, F., Fierrez, J., & Ortega-Garcia, J. (2012). Quality measures in biometric systems. *IEEE Security & Privacy, 10*(6), 52–62.
26. Olsen, M. A., Smida, M., & Busch, C. (2016). Finger image quality assessment features – Definitions and evaluation. *IET Biometrics, 5*(2), 47–64.
27. Guan, Y., Li, C.-T., & Roli, F. (2015). On reducing the effect of covariate factors in gait recognition: A classifier ensemble method. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 37*(7), 1521–1528.
28. Ward, J. A., Lukowicz, P., & Gellersen, H. (2011). Performance metrics for activity recognition. *ACM Transactions on Intelligent Systems and Technology, 2*(1), 1.
29. Bolle, R. M., Connell, J. H., Pankanti, S., & Senior, N. K. (2005). The relation between the ROC curve and the CMC. In *Fourth IEEE workshop on automatic identification advanced technologies (AutoID'05). Buffalo, NY, USA*.
30. Kusakunniran, W. (2014). Recognizing gaits on spatio-temporal feature domain. *IEEE Transactions on Information Forensics and Security, 9*(9), 1416–1423.
31. Wang, Y., Song, C., Huang, Y., Wang, Z., & Wang, L. (2019). Learning view invariant gait features with two-stream GAN. *Journal of Neurocomputing, 33*, 245–254.