



A new spatio-temporal patch-based feature template for effective gait recognition

Mohammad H. Ghaemina¹  · Shahriar B. Shokouhi¹ · Ali Badiezadeh¹

Received: 26 August 2018 / Revised: 18 June 2019 / Accepted: 13 August 2019 /

Published online: 10 September 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Recognizing gait of people has been of great interest to the researchers of biometrics in the last decade. The robust features have been recently developed to identify human's gait under different conditions. But developing efficient gait template preserving spatio-temporal features of walking is still an open problem. To address this issue, we develop a patch-based feature that can describe rhythm of walking under covariate factors properly. In our method, a new gait signature (i.e. set of spatio-temporal features) is computed from distribution of local patches in a sequence. The given signature has been used to adjust the weights of spatio-temporal coordinates and the corresponding weights are concatenated with the Gabor features. As a result, a new augmented template called Patch Gait Feature (PGF) is derived accordingly. In addition, to verify how our feature template is efficient in gait recognition, we apply two common classification methods (PCA + LDA and Random Subspace Method (RSM)) separately and evaluate the results under different challenging conditions. The recognition rate on the USF dataset indicates Rank1/Rank5 accuracies of 61.59/80.67% with PCA + LDA and 76.01/86.59% with RSM and shows an improvement of about 5% with rational computational complexity compared with other related methods.

Keywords Biometrics · Gait recognition · Patch-based representation · Template features · Gait classification

✉ Mohammad H. Ghaemina
mhghaemina@elec.iust.ac.ir

Shahriar B. Shokouhi
bshokouhi@iust.ac.ir

Ali Badiezadeh
a_badiezadeh@alumni.iust.ac.ir

¹ School of Electrical Engineering, Iran University of Science and Technology, Tehran, Iran

1 Introduction

Recognizing human's gait pattern is a challenging issue in biometric systems and pattern recognition. Since the gait can be collected at a distance and in a hidden way or from low-resolution videos, it has many advantages in contrast to face and fingerprint biometrics [11, 21]. However, there are some challenges in the gait which affects the quality of a recognition system [6]. The most common variants are viewing angle, carriage condition, walking surface, type of shoes and aging [2]. To show the gait appearance in a video without a predefined model, two types of features are developed recently [6, 23]: "temporal" and "template". In the mentioned methods, the period of walking with the silhouettes are computed from the input video initially. In the first approach, the extracted silhouettes from all frames of walking in a gait period are computed and then for recognizing the gait, the silhouettes from gallery/probe sets are compared accordingly [6]. For this reason, the complexity of the search space increases exponentially to achieve a better recognition rate [23]. But in the latter methods, the matching process are simpler as they transform all silhouettes in a video sequence into a single template. Despite the simplicity and stability to the image noises, the main weakness of the template-based approaches is its limitation in precise gait representation.

Sarkar et al. [23] develop baseline algorithm (i.e. a temporal method) in which the correlation among the silhouettes in a period is a metric for recognition. For this purpose, three levels of correlations have been defined for correct matching which they are; 1) frame-to-frame; 2) gait period-to-period (including several frames), and 3) gait sequence-to-sequence (including several periods). Despite of the accuracy of recognition, the computationally of this method is too high on an outdoor dataset because of searching all frames of walking [6]. To simplify the recognition process and reduce the complexities, the template features have been extensively developed recently [1, 2, 6, 8, 11, 30]. The basic idea of template-based approach is presented by Han et al. [11]. They have developed Gait Energy Image (GEI) that is derived by averaging the silhouettes in a period of walking. As an alternative template feature, a set of Gabor filters has been applied to GEI that is called General Tensor Discriminant Analysis (GTDA) [26]. In addition, GTDA has been combined with Enhanced Gabor (EG) features for multi-view recognition [10]. Recently, augmented Gabor filter based on patch distribution feature (Gabor-PDF) has been developed to improve the gait representation under challenging conditions [7]. Although the template features have almost simple structure which can alleviate the complexities, its performance under different gait conditions has limitations. This is mainly due to removal of gait temporal ordering in the final template. This issue has been addressed by C. Wang et al. [29]. They provide Chrono-Gait Image (CGI), a colorful template in which the rhythm of walking in time domain is represented in color spectrum. It has been proved that such colorful template can represent gait in different conditions more accurately [29]. Moreover, Space-Time Interest Points (STIP) as a spatio-temporal model of walking is defined to handle viewing issues [14]. Atta et al. [1] recently propose 5/3 Gait Image (5/3GI) by extracting lifting 5/3 wavelet features from the contour of silhouettes. It is proved that 5/3GI template can preserve temporal information of walking in final template properly. Meanwhile, to handle different gait conditions such as clothing and carriage conditions, Gheble et al. [8] propose an adaptive outlier detection method which discards clothing area of the silhouettes in a period. They also use a similarity measurement to compare the most similar parts of silhouettes in gallery and probe sets. Furthermore, Fendri et al. [5] dynamically select different parts of

a silhouette that is more consistent to such covariates and apply a semantic classification to identify the relevant parts of human's body.

The mentioned approaches can accurately recognize the gait in clothing and carriage gait conditions. But, they suffer from an accurate motion model limiting the performance on other gait conditions. In other words, the nature of a gait is a spatio-temporal process and hence, the motion descriptors are required. To utilize the gait motion model, the spatio-temporal filtering methods can be used efficiently [6, 7, 27]. For example, a set of LoG/Sinc impulse responses has been found as an efficient spatial/temporal filtering for the gait modeling [7]. In addition, the output template can be computed by averaging the responses of filtering over a period. Following the ideas in the gait filtering methods [6, 7], there are some salient regions in walking space which provides meaningful information for classification. More precisely, a simple walking pattern includes local patches (i.e. STIP) in $2D + t$ space that can be used to describe its styles more accurately [2, 27]. The mentioned idea has been proposed earlier in action recognition studies [24, 25]. For example, S. Rahman et al. [22] describe an action with the local features using statistical information (i.e. $2D + t$ histograms) of the patches. The corresponding patches are easily extracted from a bank of spatio-temporal filtering. Recently, Ben et al. [2] propose a spatio-temporal Coupled Patch Alignment (CPA) feature to handle different viewing conditions. They extract a set of local patch features in a sequence of walking and apply nearest neighbour classifier to match the patches of different gait viewing angles.

Besides the gait multi-viewing problem, Xu et al. [31] develop a novel gait recognition approach based on Capsule Neural Network (CapsNet). They also apply a linear projection of low-level local features from gallery and probe sets to improve the performance of recognition. In this paper, we develop a new patch-based template to improve the recognition rate under different covariate factors. To denote the gait with the local patch features, each gait template, e.g. GEI, is represented by set of patch distribution features (PDF) which is derived by adding X-Y patch coordinates to the pixels of template [17–19, 30]. It is proved (in [30]) that using a set of 40D Gabor features together with X-Y coordinates of patches provides better representation of gait than 2D Discrete Cosine Transform (2D-DCT) and Grey-level features (developed in [17, 18]). The augmented templates can also be characterized with Universal Background Model (UBM) [13, 19, 30] that will exponentially increase the overhead of gait representation with respect to the size of dataset. In addition, temporal information of the patches is discarded in computing the augmented template. These issues have been addressed in this paper by proposing a new gait template based on spatio-temporal distribution of the patch features. The flowchart of our approach has been shown in Fig. 1.

As it is shown in Fig. 1, our method consists of three main steps: 1) Filtering the silhouettes and extracting the local patches, 2) Computing the patch probability distribution, and 3) Formatting the final template. In the first step, the local patches are extracted by looking for the extremums of local sliding window in $2D + t$ space. In the next step, three histograms of the patches coordinate in X-Y-T domain (or the gait signature) are computed and recorded for the weightings. In the final step, the weighted coordinates of template pixels along with the temporal ordering of patches are further concatenated with the template. We call our method as “Patch Gait Feature (PGF)”, in which the spatio-temporal local features of walking are embedded in an augmented template.

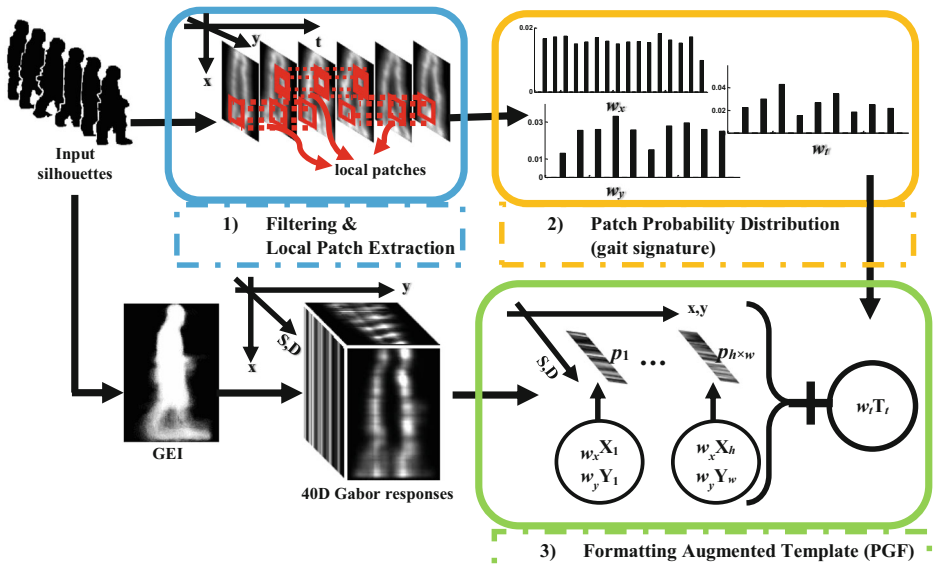


Fig. 1 The flowchart of our proposed method. The set of 40D Gabor features is augmented with the weighted coordinates which are derived from distribution of the local patches in a walking process. Here, “ w_x, w_y, w_t ” are the weights, “ p_i ” is the pixel coordinates and S, D are Gabor scales and directions

In comparisons with the recent gait filtering and patch-based features [6, 7, 19, 30], our major contributions are:

- A new patch distribution feature is proposed based on coordinates of the local patches in 2D + t domain.
- The simple walking pattern is characterized with proposed gait signature based on spatio-temporal coordinates of local patches over a sequence.
- The weighted X-Y-T coordinates are embedded with Gabor features and a new augmented feature template is formed accordingly. The weights of coordinates are derived according to the histograms of the patches.
- The complexity of the proposed method is comparable with other state-of-the-art methods.

To show how our method can describe the gait, we conduct comprehensive experiments on three well-known datasets (i.e. USF HumanID [23], CASIA-B [33] and OU-ISIR-B [20]). The structure of the paper is as follows. In the next section, local patch-based features are described. Then, the defined classifier is discussed in the third section that is used to recognize the proposed PGF. The experimental results are discussed in section 4 and the paper is concluded in the final section.

2 Gait feature extraction

In this section, the process of extracting patch features is described in detail. In the given approach, local patches along a sequence of walking are extracted and a new patch distribution

feature are calculated. Finally, the proposed template, i.e. PGF, will be derived based on the probability distribution of patches.

2.1 Preprocessing and silhouette extraction

The local patches are extracted from the sequence of silhouettes in a period of walking. However, some primary processing steps are applied to the raw input video. The major steps are background subtraction, foreground alignment and gait period detection. For background subtraction, Gaussian Mixture Model (GMM) is applied to model the background of the input gait video [23]. Once the background of the video is being computed, the foreground images including the sequence of human's walking are derived. These images are further aligned and resized according to the center of the foreground pixels and the normalized silhouettes with the same size and aspect ratio are generated [23].

Considering the sequence of aligned silhouettes, the period of a gait is easily derived by counting lower-half pixels in each frame [23]. It means that, the number of counted pixels have local extremums when two legs are near (or far) from each other. By computing the local extremums, the period is measured by median of distances between two consecutive minimums (or maximums) [6, 23]. Since each period includes of two analogous phases of walking, the starting phase for recording of the features is not important. But it is usual to start the phase of walking from the frames in which two legs are completely overlapped or near to each other [20, 23, 33]. Given a period of gait, the complete pattern of human's walking is consisting of several periods. In our method, the local patches and the corresponding PGF are computed in each period independently. In the following subsection, we explain this procedure in details.

2.2 Local spatio-temporal patches

Assuming a sequence of silhouettes in a period, the desired patches are extracted from the responses of gait filtering. Recently, the Gabor filtering has been known as an effective tool in human's motion modeling. Here, we apply it to highlight the salient regions during the walking. Let a set of Gabor filters be defined in a default pixel $z = (x, y)$ as follows:

$$\phi_{\tau, \nu}(z) = \frac{\|r_{\tau, \nu}\|^2}{\delta^2} e^{-\frac{\|r_{\tau, \nu}\|^2 \|z\|^2}{2\delta^2}} \left[e^{ir_{\tau, \nu} \cdot z} - e^{-\frac{\delta^2}{2}} \right], \quad (1)$$

where $r_{\tau, \nu} = \theta_{\tau} e^{i\phi_{\nu}}$ indicates the scales and directions of the Gabor function. Similar to conditions in [26, 30], we set $\sigma = 2\pi$, τ equals to $\{0, 1, 2, 3, 4\}$ and ν equals to $\{0, 1, 2, 3, 4, 5, 6, 7\}$ as well. As a result, the set of 40D Gabor impulse responses are obtained in 5 scales and 8 different directions. Also, $\exp(-(\delta^2/2))$ in eq. (1), is deducted to be the independent functions of DC value which are resistant to brightness variations.

In this paper, all of the gait frames (with the dimensions of $N_2 \times N_1$) are convoluted with 40D Gabor filters to extract the energy of walking within a sequence. Afterwards, the magnitude sum of 40D Gabor responses over different scales and orientations is computed for each silhouettes:

$$\begin{aligned}
 R_{SD}(z) &= \left| \sum_{\tau} \sum_{\nu} I(z) * \phi_{\tau,\nu}(z) \right| \\
 &= \left| I(z) * \sum_{\tau} \sum_{\nu} \phi_{\tau,\nu}(z) \right|,
 \end{aligned}
 \tag{2}$$

where $I(z)$ is the input silhouette and “*” is convolutional operator. To reduce complexity issues, the dimensions of the $R_{SD}(z)$ are reduced to $[N_2/2] \times [N_1/2]$ (noted as R_{SD-dn}), where $[N_2/2]$ and $[N_1/2]$ are the largest real numbers less than or equal to $N_2/2$ and $N_1/2$ respectively [1]. It has been proved that downsampling decreases the computational problems without reducing the accuracy of the identification [1].

In a conventional method such as [30], each GEI is represented as a set of Gabor features similar to the Eq. (2) and the existing augmented feature can be computed by adding spatial coordinates. The augmented Gabor features in spatial domain are shown as $p_h = [q_h^T, X_h, Y_h]^T \in \mathcal{R}^\rho$ ($\rho = 42$) by adding the pixels data. Here, q_h ($h = 1 \dots H$) is the amount of Gabor responses and $H = [N_1/2] \times [N_2/2]$ are total values of the pixels. It is clear that the augmented feature, q_h , has no data from the temporal information of gait. Moreover, all of the spatial coordinates have the same importance in q_h and hence, it cannot present a proper motion model. To extract spatio-temporal information that is consistent with human’s visual system, we first define patch window and extract local extremums from such the local window.

Given a sequence of Gabor responses, R_{SD-dn} , over a period, we define a sliding 2D + t window (or patch window) with dimensions of $w_x \times w_y \times w_t = w$ (e.g. $3 \times 3 \times 3$) along x , y and t . By moving the given window in the spatial and temporal spaces, the local maximums from the responses are recorded in each step accordingly. The resulting patches $z_{patch} = (X_i, Y_j, t_k)$ are characterized with location of extremums in which $0 \leq i < [N_1/2]$, $0 \leq j < [N_2/2]$ and $0 \leq k < T$ (T is the gait period). The gait signature is then extracted by computing distribution of the patches in 2D + t space which is discussed in following subsection.

2.3 Patch probability distribution

Given the recorded extremums in 2D + t walking space of an individual, we represent gait of individuals with the probability density function of patches over a sequence. It is mainly due to using statistical information of patches that can be applied to describe the gait. We formulate the probability distribution by set of m -bin histogram in spatial and temporal spaces. Suppose $b : \mathcal{R} \rightarrow \{1 \dots m\}$ as a function that assigns the input patches, z_{patch} , to the bin $b(z_{patch})$. The set of X-Y-T histograms, q_u , $u = 1 \dots m$, related to the locations of 2D + t patches are computed as follows:

$$\begin{aligned}
 q_{u,x} &= C_x \sum_{i=1}^{[N_1/2]} \delta[b(X_i) - u_x] & u_x &= 1 \dots [N_1/2] \\
 q_{u,y} &= C_y \sum_{j=1}^{[N_2/2]} \delta[b(Y_j) - u_y], & u_y &= 1 \dots [N_2/2] \\
 q_{u,t} &= C_t \sum_{k=1}^T \delta[b(t_k) - u_t] & u_t &= 1 \dots T
 \end{aligned}
 \tag{3}$$

where C_x , C_y and C_t are the normalization coefficients in which $\sum_{u=1}^m q_u = 1$. By calculating above histograms, three unique distributions corresponding to the horizontal, vertical, and temporal directions are computed accordingly. Figure 1a shows the process of computing spatio-temporal histograms. Each moving window in 2D + t space has a local maximum where its coordinate is noted to a bin and stacked in the corresponding histograms. The location of extremums over the full 2D + t walking space represent totally the distribution of spatio-

temporal patches for each people. Figure 2b, c and d show the set of X, Y and T histograms corresponding to three different people, each one with a unique color bar. Figure 3a shows locations of twenty most histogram values (in Fig. 2b, c and d) for visualizing how the local patches of different people are distributed in 2D + t space. The base template in Fig. 2 is mean of the 40D Gabor features ($R_{SD}(z)$ in eq. (2)) over a period and the circles denotes the patches in which the most values in histograms are shown with bigger radius. Meanwhile, the 2D projection of 2D + t patches in X-Y plane are shown in Fig. 3b.

The most bigger radius in Fig. 3 corresponds to density of local patches in a walking space. It is clear from Figs. 2 and 3 that a set of X-Y-T histograms as gait signature can be used to describe style of walking properly. It means that for each walking conditions, there are some STIP that its probability in 2D + t space can be used to accurately characterize its motion. It should also be noted that most of the patches are localized on legs, hands and shoulder regions and hence, it signifies the salient regions for recognition. In the next subsection, the computed histograms are used to form augmented Gabor template.

2.4 Formatting augmented template

Once the histograms of the patches are being generated, the final template can be computed. To do this, we perform two processing steps: 1) applying proper filtering to GEI templates, 2) adding statistical distribution of patches to the template.

In the first step, Gabor filters found as an efficient filtering to represent the gait templates [30]. By applying the 40D Gabor functions to the GEI, we have a set of 40D Gabor responses to describe the GEI templates [9, 26]. Here, the magnitude part of 40D Gabor responses is downsampled and utilized for the proposed descriptor:

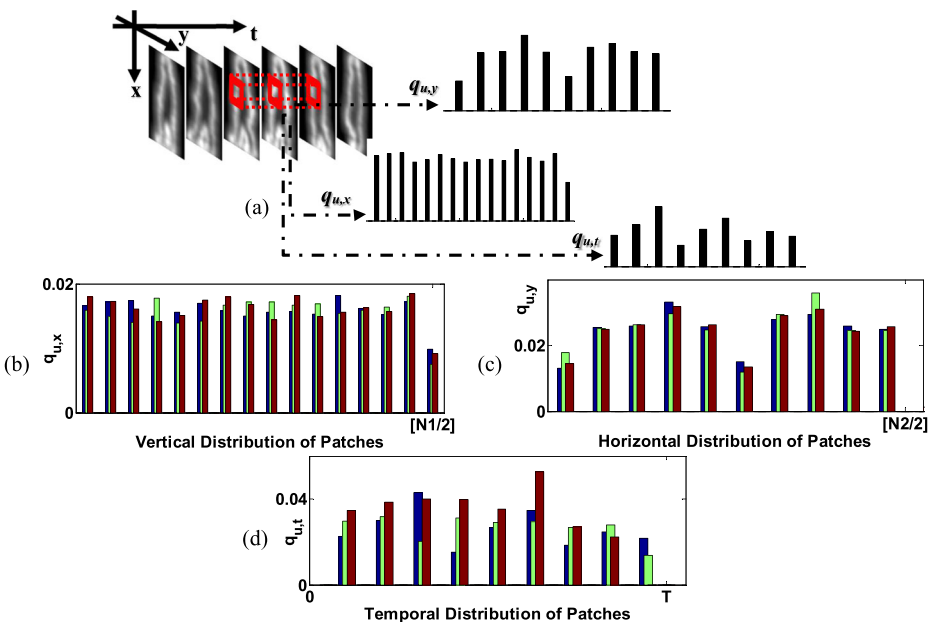


Fig. 2 (a) The process of computing X-Y-T histograms based on location of local extremums. (b), (c) and (d) The set of vertical, horizontal and temporal histograms for different individuals as shown in blue, green and red color bars

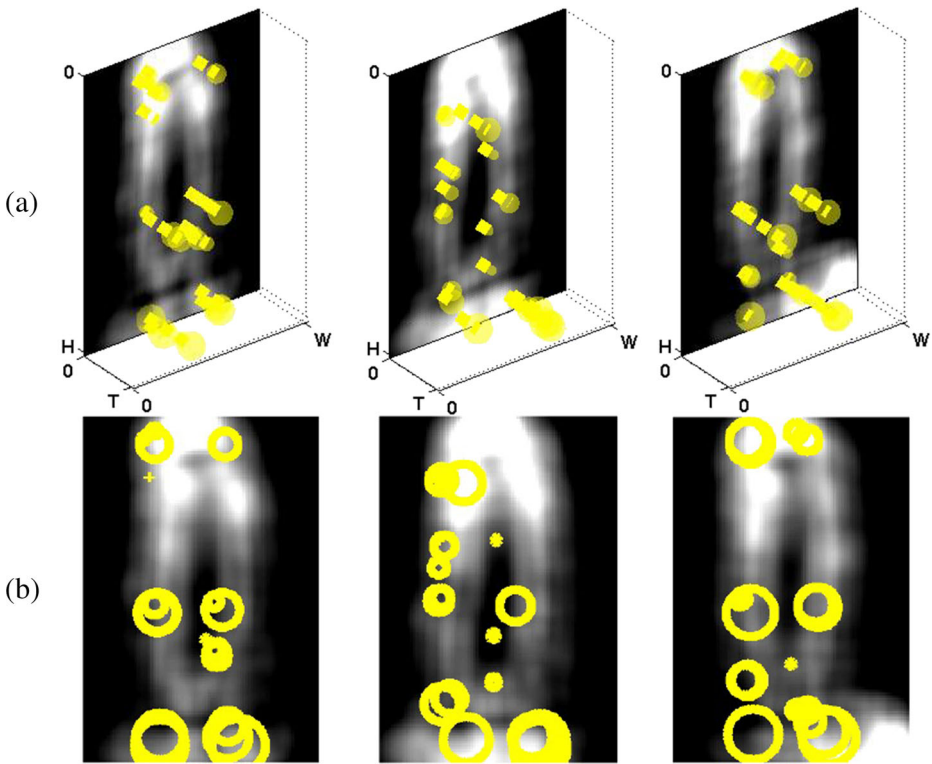


Fig. 3 The 2D+t locations of twenty top values of X-Y-T histograms for three different people. **a** 3D viewing direction and **b** 2D projection of locations

$$R_G(z) = |GEI(z) * \phi_{\tau,y}| \in \mathcal{R}^{[N_1/2] \times [N_2/2] \times 40} \tag{4}$$

Then, the augmented Gabor filter can be obtained by adding the probability distribution of 2D+t patches to $R_G(z)$. For this purpose, the values of histograms are concatenated to the responses of Gabor filters, $R_G(z)$. However, to properly consider distributions of the patches in final feature vector, we use the histogram values to weight spatial and temporal coordinates.

Let $\{z_h^* = (x_h^*, y_h^*)\}_{h=1 \dots H}$ be the normalized pixel locations, centered at center of silhouettes. Meanwhile, assume $\{t_k^*\}_{k=1 \dots T}$ be the normalized time stamp centered at mean of a period (i.e. $T/2$). An isotropic kernel, $k(\|z\|)$, is further defined to assign smaller weights to the pixels farther from the center. We also define the vectorized Gabor responses by $R_{G-c}^T = \{q_h^T\}_{h=1 \dots H} \in \mathcal{R}^{40 \times H}$ ($H = [N_1/2] \times [N_2/2]$). With the above assumptions, the spatial part of the proposed augmented feature vector is computed as follows:

$$G_{aug-s} = \left[R_{G-c}^T, q_{u,x} k(\|x_h^*\|), q_{u,y} k(\|y_h^*\|) \right]^T, \tag{5}$$

where $G_{aug-s} \in \mathcal{R}^{42 \times H}$ is the feature vector that is derived by adding weighted spatial information, $q_{u,x} k(\|x_h^*\|)$ and $q_{u,y} k(\|y_h^*\|)$, to Gabor features. The proposed PGF is computed by concatenating the temporal information to G_{aug-s} in eq. (5):

$$G_{aug_st} = \left[R_{G-c}^T, q_{u,x}k(\|x_h^*\|), q_{u,y}k(\|y_h^*\|), q_{u,t}k(\|t_h^*\|) \right]^T, \quad (6)$$

where $q_{u,t}k(\|t_k^*\|)$ is the weighted (and interpolated) values of the pixels in temporal domain within a period.

In this paper, the uniform kernels have been applied (in eq. (6)). Also, other types of kernels for spatial and temporal weighting can be used instead. The proposed PGF (i.e. $G_{aug_st} \in \mathcal{R}^{43 \times H}$) contains spatio-temporal information of walking in a period of time. To compute PGF, the weighted spatio-temporal coordinates in 2D + t walking space are concatenated to the set of 40D Gabor responses. As such, the weights are obtained from probability distribution of the 2D + t patches in a period of walking. In the next section, it will be discussed how to identify this pattern through the given features.

3 Classification

The augmented Gabor features are calculated for all individuals in given dataset. The conventional method to recognize the gait is assigning the features from probe sets to its closest feature to the gallery set by direct matching process. As an example, the 1-Nearest Neighbor (1-NN) has been used as a simple classification tool to assign a label, which has the minimum Euclidian distance to its neighbor class. In this paper, we briefly discuss on common classification tools that are developed for gait recognition.

Despite the simplicity of direct matching process, it has some limitations such as [6]: (1) The gait samples are commonly taken in similar conditions, and hence the features may overfitted. (2) The number of patterns in the training space is small and limited therefore the gait features cannot describe the topology of the main walking space truly. (3) The dimensions of the feature space are much higher than the training samples, since each pixel of the input pattern is considered as a dimension.

The above restrictions are known as “Under Sample Problem” (USP) [6, 9, 29]. Several methods have been proposed to address the training limitation problem. A simple strategy is to increase the amount of training data by producing extra templates. For example, the number of features in GEI and CGI methods have been increased by producing synthetic templates [11, 29]. But an efficient and useful solution for removing USP is to use dimensional reduction techniques [6]. For example, a two-step technique can be used to optimize search space. Here, we briefly reviewed the common classification methods that has been applied in this paper for proposed gait classification. In this regards, the principal component analysis (PCA) followed by linear discriminant analysis (LDA), known as PCA + LDA, has been found as a useful and efficient approach in face and gait classification [10, 29]. In this approach, the PCA is firstly applied to project the given template into a low-dimensional feature space. In the second step, the LDA technique is further applied to eliminate redundant information in feature space. By computing PCA + LDA transformation matrix in the gallery feature space, the features in the probe set can be projected accordingly. In other words, there are feature vectors with the reduced dimensions in both of the gallery and probe sets that can be used for the classification purpose [6, 11, 29]. To classify the gait patterns, the projected feature vector of probe sets that has minimum Euclidean distances to its gallery set is labelled as final decision [6].

Although the PCA + LDA method can efficiently enhance dimension of feature space, it has some limitations in discriminating the covariate factors. This is mainly due to removing structure

of gait templates within the projection process. To provide efficient tools that can discriminate the gait variants in feature space, the Random Subspace Method (RSM) has been developed recently [9]. It is proved that RSM provides better gait recognition rates in comparison with PCALDA under different gait covariates [7, 9]. More precisely, the RSM supersedes conventional methods due to reserving structure of the given template and generating weak classifiers through random sampling of the feature space [7]. The RSM has been consist of three main steps: creating random subspace from 2DPCA projection matrix, enhancing local dimensions and making hybrid-based decision. In first step, the 2DPCA from the input feature vector are computed and K (e.g. $K = 1000$ in this paper) random eigenvectors (projection matrix) are taken accordingly. Next, the dimensionality of randomly selected vectors is enhanced by applying another dimensionality reduction technique. In RSM, two different methods, i.e. 2DLDA and incremental dimension reduction algorithm via QR decomposition (IDR/QR) have been used alternatively. It should be note that in IDR/QR, the feature vectors are being vectorized within the process. As a result, two different classifiers are defined in RSM as 2DPCA + 2DLDA and 2DPCA + IDR/QR to label the features in K random subspaces. The hybrid decision-level fusion (HDF) within the K classifiers are derived by majority voting of the all classifiers in subspace for each individuals [9].

Assuming that there are n_G individuals in the gallery set each one has $g_i (i = 1 \cdots n_G)$ feature vectors. Also, suppose that $G_{aug,i}$ is a feature vector of i^{th} individual in the gallery set and $P_{aug,j} (j = 1 \cdots n_P)$ is the feature vector of the individuals in the probe set. For each feature set, there are K random samples with given mean of samples, $m_{G_i^k} (i = 1 \cdots c)$ in dataset. According to the Euclidean distance among the center of the feature vectors of gallery/probe sets in the low-dimensional subspace, the distance of the features between the input probe set ($P_{aug,j}$) and the given gallery ($G_{aug,i}$) in each subspace is measured as:

$$D(P_{aug,j}^k, G_{aug,i}^k) = \frac{1}{n_P} \sum_{j=1}^{n_P} \|P_{aug,j}^k - m_{G_i^k}\|, \quad i = 1 \cdots n_G, \tag{7}$$

In RSM for k^{th} subspace, the minimum distance of the probe set to individuals in the gallery set is considered as weak classifier:

$$l^k(P_{aug}^k) = \underset{i}{\operatorname{argmin}} D(P_{aug,j}^k, G_{aug,i}^k), \quad i = 1 \cdots n_G \tag{8}$$

where $l^k(P_{aug}^k)$ is the label of given individual in probe set in k^{th} subspace. Now, to make the hybrid decision for the input template, the mode of labels among all subspaces is computed. In RSM, two types of decisions have been taken based on 2DLDA and IDR/QR techniques. Let l_{LDA} and l_{QDR} be the computed labels derived from different methods of RSM. A hybrid classifier (HC) is defined according to fusion of final decisions in each methods as follows:

$$l_{HC}(P_{aug}) = \begin{cases} G_{aug,i}, & \text{if } l_{LDA}(P_{aug}) = G_{aug,i}, \\ G_{aug,i}, & \text{if } l_{QDR}(P_{aug}) = G_{aug,i}, \\ 0, & \text{Otherwise} \end{cases} \quad i = 1 \cdots n_G \tag{9}$$

In other words, the final classifier is assigned the label correctly if one of the given classifiers making final decision correctly [7, 9]. In the following section, we bring the experimental

results of the proposed gait recognition system which is known as PGF + PCALDA and PGF + RSM.

4 Results and discussion

In this section, PGF is evaluated by performing comprehensive experiments on three publicly available datasets including USF [23], CASIA (Dataset B) [33] and OU-ISIR (Dataset B) [20]. To fully assess the proposed method, we compare the results with the recent approaches based on different classifier techniques. In first step, the nearest neighbor is performed on the augmented Gabor features (PGF + 1NN). Secondly, to make a fair comparison, the PGF recognition system is evaluated with two types of algorithms: evaluating PGF + PCALDA with the different template features that has applied simple classifier, comparing the results of PGF + RSM with the methods which utilize advanced classifiers. The selected methods for evaluations are baseline algorithm [23], 2DLDA [16], GTDA [26], Gabor+GTDA [26], DATER [32], MTP [3], GEI [11], CGI [29], PGM [12], GEI + Sparse bilinear discriminant analysis (SBDA) [15], GSI [6], 5/3GI [1], STIP [14], locality-constrained group sparse representation (LGSR) [30], Gabor+RSM [9], GEI + RSM [9], view-invariant multiscale gait recognition (VI-MGR) [4], local patch-based subspace ensemble learning algorithm (LPSELA) [19], gait spatio-temporal image (GSTI) [7] and Two-Stream Generative Adversarial Network (TS-GAN) [28]. We also consider “Rank1” and “Rank5” Correct Classification Rate (CCR) as two common benchmarks for performance evaluation. In each experiments, the most values of Rank1 and Rank5 has been bolded.

4.1 Gait Conditions

As discussed earlier, we evaluate the proposed method on three well-known datasets: USF HumanID dataset (dataset version 2.1) [23], CASIA Dataset (Dataset B) [33] and OU-ISIR (Dataset B) [20]. Most of the results in this section have been assessed on the USF dataset, due to more popularity, challenging gait conditions and noisier quality of silhouettes provided in the dataset.

The USF dataset consists of 122 individuals walking in elliptical paths in front of the camera. Gait conditions include walking surface (S), shoes type (H), viewing angle (V), carrying condition (C) and elapsed time (T). Considering these conditions, the set of “Grass, shoe type A, without briefcase, time t_1 (in May)” was considered as a gallery set. Then 12 unique testing conditions have been developed for the probe sets.

Table 1 The specification of probe sets in USF gait dataset

Experiment	A	B	C	D	E	F	G	H	I	J	K	L
Covariates	V	H	VH	S	SH	SV	SHV	B	BH	BV	THC	STHC
Num. of People	122	54	54	121	60	121	60	120	60	120	33	33
Variance	Shoes/View		Surface+Shoes/View				Briefcase + Shoes/View			Time + Shoes +Surface +Clothing		

V View, H Shoe, S Surface, C Carriage, T Time, C Clothing

The list of conditions for probe sets with the number of individuals per experiment are shown in Table 1. In USF dataset, the sequence of silhouettes with period of walking are provided. To extract silhouettes, the background subtraction method, has been applied to the raw input video and then foreground images are aligned to the center of images and cut to 128×88 pixels.

From Table 1, there are different number of subjects in each probe. We define the weighted average identification rate for the results (W-AvgI) as:

$$W\text{-AvgI} = \frac{\sum_{i=1}^g w_i R_k}{\sum_{i=1}^g w_i}, \quad (10)$$

where w_i denotes the number of people in the probe sets, g denotes the number of experiment, i.e. $g = 12$ in Table 1, and R_k are Rank1/Rank5 recognition rates.

Another test criterion is CASIA dataset (Dataset B) [33]. The dataset consists of 124 individuals and data are given in 10 different indoor conditions as follows: six sets with natural conditions (NM01-NM06), two sets of walking with carrying bag (BG01-BG02), and two sets of walking with clothing condition (CL01-CL02).

Each walking sequence is captured under 11 different viewing angles, from 0 to 180 degrees, at intervals of 18 degrees between the two directions. Since the proposed augmented Gabor filter has been designed for 90° viewing conditions, the results from corresponding degrees have been presented. Both of the USF and CASIA datasets provide lateral viewing of walking people, but only in USF datasets, the aligned silhouettes have been provided. Similarly, the silhouettes are aligned in the CASIA based on their horizontal center, and then cuts into images of 160×100 (similar to the CGI conditions [29]).

Our third testing condition is OU-ISIR-B dataset [33] that has been published recently to study the effect of clothing conditions. This dataset includes of 48 individuals walking on a treadmill with 32 types of different clothing. Table 2 presents gallery and probe sets corresponding to different clothing combinations provided in the database. The gallery set includes 48 individuals with standard clothes (i.e. type 9) while the probe set consists

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

Type	clothes	Num. of people	Type	clothes	Num. of people	Type	Clothes	Num. of people
0	CP + CW	37	C	RP + DJ + Mf	48	N	SP + HS	14
2	RP + HS	14	D	CP + HS	14	P	SP + Pk	47
3	RP + HS + Ht	14	E	CP + LC	14	R	RC	46
4	RP + HS + Cs	14	F	CP + FS	14	S	Sk + HS	2
5	RP + LC	48	G	CP + Pk	14	T	Sk + FS	26
6	RP + LC + Mf	47	H	CP + DJ	14	U	Sk + PK	27
7	RP + LC + Ht	14	I	BP + HS	14	V	SK + DJ	27
8	RP + LC + Cs	14	J	BP + LC	35	X	RP + FS + Cs	33
9	RP + FS	48	K	BP + FS	35	Y	RP + FS + Cs	32
A	RP + Pk	48	L	BP + Pk	35	Z	SP + FS	32
B	RP + DJ	48	M	BP + DJ	35	–	–	–

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

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 (similar to the USF gait silhouette conditions).

4.2 Testing conditions

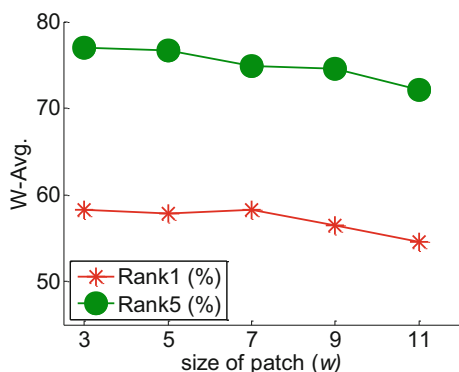
To illustrate how the proposed patch-based features can identify the type of walking, we first evaluate the effect of different settings on the performance of gait recognition. In order to compute a PGF, the size of patch window is being defined as $3 \times 3 \times 3$ (i.e. 3) pixels. However, other sizes can be used to describe the walking style in a sequence. Here, we repeat experiments with five different sizes of window as {3, 5, 7, 9, 11} and 1-NN classifier. The Rank1/Rank5 CCR on USF dataset is provided in Fig. 4. The performance of recognition shows that the results are almost invariant against size of window. However, the better performance is achieved by setting the size to 3. This is mainly due to preserving the local information that is provided with smaller window size.

Assuming the PGF template with patch size of 3 pixels, another setting condition in computing the PGF is using the weighted coordinates (step 3 in Fig. 1). To study effect of the weighting the coordinates, we compute the PGF using one of the four different weighting conditions: 1) weighting the coordinates in $2D + t$ domain (our approach), 2) weighting in the spatial domain (and not weighted the T coordinates), 3) weighting in the temporal domain (and not weighted the X-Y coordinates), and 4) no weighting in spatio-temporal domain (not weighted X-Y-T coordinates). The Rank1/Rank5 results of different PGF settings with RSM classifier on USF dataset are shown in Fig. 5.

The results of the mentioned four experiments are shown Fig. 5 with blue, red, black and green colors, respectfully. Table 3 shows the averaging results from different testing conditions which are demonstrated in Fig. 5.

From Fig. 5 and Table 3 we can find the following results: 1) In 5 out of 12 probes (A, B, H, I and J), the Rank1 and Rank5 of testing conditions are similar and hence the weighting of coordinates is not important in such probes. 2) In harder challenging conditions (C, D, E, F, G, K and L), the Rank1 and Rank5 of the proposed weighting are improved with the maximum of 17% (Rank 5 of probe F). In other words, our patch-based descriptor with proposed weighting has better performance under conditions of “walking surface” and “elapsed time” (in probes D, E, F, G, K and L). 3) The average of Rank1 and Rank5 for the proposed weighting are improved about 4%. The worse condition belongs to the second testing criteria when the pixels

Fig. 4 Effect of patch size for the proposed PGF on performance of recognition



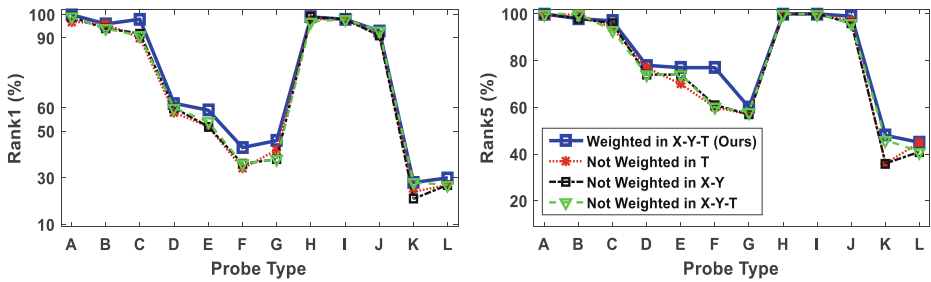


Fig. 5 Evaluation the effect of weighting coordinates on performance in USF dataset

just weighted in spatial domain (and not in temporal domain). It means that the proposed gait recognition system is very sensitive to temporal information. However, weighting the pixel only in temporal domain (and not in spatial domain) will also decrease the performance. As a result, the quality of features is improved when weighted pixels both in spatial and temporal domains are added to the augmented template.

4.3 Gait templates

To study the effect of proposed template on the performance of gait recognition, we first apply a simple classifier, i.e. 1-NN classifier. Using a simple classifier without applying PCA + LDA or RSM provides useful tools to analysis the robustness of input feature vectors against noises and other covariates.

Here, the 1-NN classifier on USF and CASIA dataset is applied to evaluate quality of the algorithm. In addition, the results of our PGF are compared with recently published methods (including baseline algorithm, GEI, CGI, 5/3GI, STIP, VI-MGR, GSI and GSTI). The evaluation results on USF dataset are provided in Table 4.

According to Rank1 CCR in Table 4, we have summarized the performance here. 1) The proposed template (PGF + 1NN) has the best weighted average of Rank1. 2) In comparison with the baseline algorithm and GEI, Rank1 accuracy has been improved in all conditions (except K and L). 3) The proposed patch-based feature improves the Rank1 CCR in 7 out of 12 experiments compared with other related methods. Although the suggested feature is vulnerable to “walking surfaces” and “elapsed time” conditions (in probes D, E, F, G, K, and L), it still has an acceptable performance. Likewise, the performance of other methods in Table 4 are weak in such conditions. 4) Compared to Rank5, the weighted average with 7 out of 12 conditions of our methods have the highest CCR. The performance of PGF + 1NN under “walking surface” and “elapsed time” conditions has been improved in Rank5 CCR and hence it is acceptable for recognition.

Table 3 The average of Rank1 and Rank5 for different testing conditions on USF dataset

Exp.	W-Avg. of Rank1 (%)	W-Avg. of Rank5 (%)
Weighted in X-Y-T (ours)	75.69	86.28
Not weighted in T	71.91	82.53
Not weighted in X-Y	72.44	82.39
Not weighted in X-Y-T	72.57	82.62

Table 4 Performance evaluation of different gait features on USF dataset using 1-NN classifier

Probes	Rank1 CCR (%)						Rank5 CCR (%)					
	Baseline [23]	GEI [11]	CGI [29]	GSTI [7]	5/3GI [1]	PGF (Ours)	Baseline	GEI	CGI	GSTI	5/3GI	PGF (Ours)
A	73	84	87	88	90	91	88	93	96	97	96	97
B	78	87	94	93	91	89	93	94	94	95	97	97
C	48	72	72	80	76	76	78	93	93	93	94	93
D	32	19	17	29	24	35	66	45	41	60	54	65
E	22	18	25	30	27	36	55	53	45	50	57	68
F	17	10	12	16	12	17	42	29	32	43	40	42
G	17	13	13	15	20	19	38	37	35	39	43	47
H	61	56	78	76	86	92	85	77	91	93	95	98
I	57	55	80	69	79	85	78	77	97	87	97	98
J	36	40	54	64	58	71	62	69	82	82	89	93
K	3	9	6	10	12	8	12	15	30	28	30	28
L	3	3	9	10	9	8	15	15	27	22	30	27
W-Avg1	41.0	41.1	48.6	52.0	51.9	57.19	64.5	61.4	66.8	70.63	72.2	75.74

The bold values indicate the values that outperform the previous works

In addition, the quality of images in CASIA are greatly better than silhouettes in USF dataset. For each individual, there is one or two gait periods in CASIA which reduces the number of PGF samples in each sequence and hence, the number of computed templates for correct classification are limited. [6, 26]. However, the CASIA dataset provides a good benchmark to evaluate the quality of the generated template using 1-NN classifier. To make a fair comparison with other methods [1, 4, 7, 11, 14, 28, 29] on CASIA, we define each condition as training set (gallery) and make all nine others as testing sets (probes). The resulting conditions are 90 (10 × 9) different combinations of the gallery and probe sets for evaluation using 1-NN classifier. To visualize the 90 unique tests, it has been divided into three main groups (i.e. *nm*, *bg*, *cl*), in which the corresponding results are averaged in each group.

Table 5 shows Rank1 CCR for the proposed PGF + 1-NN under various conditions. The results in Table 5 are categorized into two different sections. 1) The results of the related template-based algorithms (including GEI [11], CGI [29], 5/3GI [1] and GSTI [7]) are provided in the first section. 2) The results of the recent multi-view methods (including STIP [14], VI-MGR [4] and TS-GAN [28]) are listed in the second section. Moreover, it has been presented only Rank1 results of CASIA conditions in Table 5.

Table 5 Evaluation Rank1 CCR of the proposed method on CASIA dataset using 1-NN classifier

Gallery-Probe	nm-nm	nm-bg / bg-nm	nm-cl / cl-nm	bg-bg	bg-cl / cl-bg	cl-cl	Avg.
GEI [11]	91.57	37.8	25.04	91.2	17.47	97.22	60.05
CGI [29]	88.06	51.93	46.88	89.81	32.52	95.37	67.43
5/3GI [1]	98	73	66	–	–	–	–
GSTI [7]	91.8	57.5	39.6	92.5	26.4	98	67.64
STIP [14]	95.4	60.9	52	73	29.8	70.6	63.62
VI-MGR [4]	100	89	76	79.03	--	71.77	--
TS-GAN [28]	--	--	--	--	--	--	63.10
PGF	92.63	69.33	51.0	93.0	34.5	100	73.41

The bold values indicate the values that outperform the previous works

In comparison with the methods in the first section of Table 5, the PGF method has a better average of Rank1. More precisely, in *bg-bg*, *bg-cl* and *cl-cl* conditions (columns 4, 5, and 6), the results show improvement in comparison with the template-based approaches. Furthermore, the results of our PGF are still comparable with respect to the methods in other parts of Table 5. The proposed method has better performance than STIP method in 4 out of 6 conditions. Meanwhile, the PGF results are competitive with the VI-MGR approach in *bg-bg* and *cl-cl* conditions (column 4 and 6). The VI-MGR provides higher accuracy due to the use of synthetic templates and a better classifier (i.e. subspace learning) [4, 9]. The total average of the Rank1 CCR in Table 5 confirms that the proposed feature has an acceptable performance in CASIA dataset (compared with TS-GAN [28]).

The performance of PGF on USF and CASIA datasets verifies effectiveness of our method in representing the gait under different conditions. Indeed, the patch-based features are able to improve the accuracy in most of the experiments.

4.4 Performance on clothing covariates

As discussed in subsection 4.1, the OU-ISIR-B dataset [33] is sampled from people walking on a treadmill and wearing different type of clothes. It should be noted that the most common covariate for rhythm of walking is use of different clothing. However, such variants are only presented in the probes K and L that have few number of people in each probe. As it is shown in Table 2, the OU-ISIR dataset consists of 32 combinations with clothing. To cover all possible testing conditions, it provides additional gallery set that can help for developing a clothing-invariant recognition system [8, 20]. Here, there is no need for pre-defined training set which may lead to unrealistic performance. To evaluate the performance of gait recognition system in OU-ISIR-B dataset, we compare the Rank1 results of 31 cloth types (i.e. the probe sets) of the proposed PGF with Clothing-Invariant Gait Recognition (CLIGR) [8], GSTI [7], VI-MGR [4] and Local Feature at the Bottom Layer Based on Capsule Network (LBC) [31]. The Rank1 results for different probes are shown in Fig. 6. To make fair comparisons, we apply RSM classifier to PGF and compare the results with other state-of-the-art methods.

From Fig. 6, it is clear that the proposed method outperforms the Rank1 in comparison with other methods in most challenging conditions. More precisely, in 2 out of 31 conditions (i.e. probes H and M), the PGF has lower Rank1 with respect to the CLIGR [8] and LBC [31]

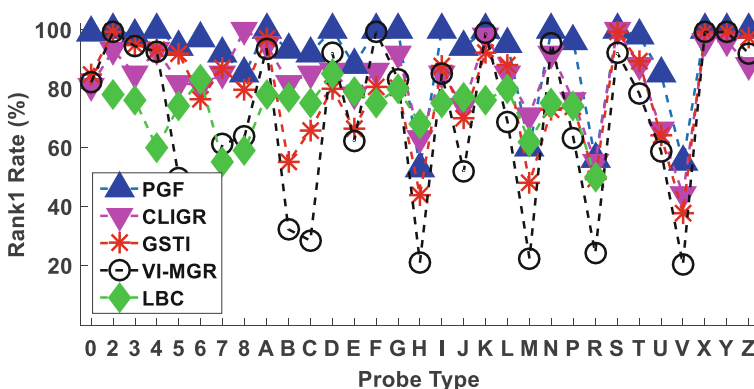


Fig. 6 Performance of the proposed gait recognition system for 31 probe types in OU-ISIR-B dataset

methods. However, the performance of other methods is pretty low in such conditions. Actually in probes H and M, the people wear “Down Jacket (DJ)” in which the parts of body for extraction of the patches are corrupted. As a result, local patches in PGF are failed to accurately locate the most important regions. From the results in Fig. 6, we can find that the performance of PGF has been lost below 60% in probes H, M, R and V. This is mainly due to covering of body parts in such probes, because of wearing “DJ” or “Rain Coat (RC)”. Furthermore, in probes B and C, the people are wearing “Regular Pans (RP)” with “DJ” and PGF has still acceptable accuracy compared with other methods. Truly in probes B and C, the upper body parts covered with unusual clothing while the lower body parts are covered with standard clothing. In such conditions, the proposed method extracts enough information from the lower body parts which can help to improve the accuracy of recognition. The performance of PGF in Fig. 6 verifies that our method is more robust when upper body clothing is changing (in probes B and C) and is weak to unusual clothing of lower body (in probes H, M, R and V). But in comparisons with other methods, our PGF has still better accuracy in most conditions.

The weighted average of Rank1 for the methods of Fig. 6 is listed in Table 6. Again we can conclude that our method outperforms the conventional methods in OU-ISIR-B dataset. It is mainly due to utilizing local information during the walking process that is provided in PGF.

4.5 Recognition system

In this subsection, the proposed gait feature recognition system has been compared to the recent methods on the USF HumanID dataset. To make fair comparisons, we evaluate the results with two sets of gait recognition systems: 1) comparing the PGF + PCALDA with the templates that utilizing simple or similar classifier, and 2) evaluating the PGF + RSM with the related methods that applying advanced classifier.

The bold values indicate the values that outperform the previous works.

The Rank1 and Rank5 CCR with the weighted average results (Eq. (10)) of the methods in first category (i.e. using simple classifier) have been summarized in Tables 7 and 8. It is clear that the weighted average results (W-AvgI) of Rank1 and Rank5 have acceptable rate in comparison with other related methods. Additionally, the proposed patch-based method is appropriate to represent the gait in most conditions due to the improvement of Rank1/Rank5 recognition rates. Specifically, in 4 out of the 12 conditions (probes A, C, D and E), Rank1 CCR is highest while in other testing conditions, the accuracies of our system is near to the highest. Our method is only vulnerable against walking surface and timing variants (probes D, E, F, G, K and L) in which other methods have also weak performance in such conditions.

But in normal gait conditions (probes A, B, and C), Rank1 results provide better accuracies than the recent spatio-temporal templates (e.g. CGI, GSI and 5/3GI). The Rank5 performance of the proposed PGF in Table 8 is still competitive in similar testing conditions. By summarizing the Rank1/Rank5 accuracies, our method supersedes conventional template-based

Table 6 Evaluation weighted average of Rank1 CCR (%) of the methods in OU-ISIR-B dataset

Methods	W-Avg. of Rank1 (%)
VI-MGR [4]	63.32
GSTI [7]	77.25
CLIGR [8]	81.73
LBC [31]	74.44
PGF (Ours)	90.62

Table 7 Comparisons of Rank1 CCR (%) of different templates using a simple classifier on USF dataset

Exp.	A	B	C	D	E	F	G	H	I	J	K	L	W–AvgI
<i>Method</i>	Rank1 Performance												
Baseline	73	78	48	32	22	17	17	61	57	36	3	3	40.96
2DLDA	89	93	80	28	33	17	19	74	71	49	16	16	50.98
GTDA	86	88	73	24	25	17	16	53	49	45	10	7	43.70
Gabor+GTDA	84	86	73	31	30	16	18	85	85	57	13	10	52.51
DATER	87	93	78	42	42	23	28	80	79	59	18	21	56.99
MTP	90	91	83	37	43	23	25	56	59	59	9	6	51.57
GEI (Real)	89	87	78	36	38	17	28	62	59	59	3	6	51.04
CGI (Real)	90	91	80	33	33	18	22	84	80	61	3	6	54.49
PGM	83	89	78	26	27	13	17	60	55	43	6	0	44.22
GEI + SBDA	93	94	80	44	45	25	34	80	83	62	18	21	60.83
GSI	92	95	86	38	31	17	24	79	83	76	25	19	58.44
5/3GI-Contour	92	91	81	42	35	21	25	87	80	63	9	6	61.59
PGF + PCALDA (Ours)	93	93	86	48	54	24	33	83	76	74	10	10	61.59

methods (such as CGI, GSI or 5/3GI) using spatio-temporal local patches in a given sequence. To evaluate the proposed PGF, a two-step classification method (i.e. PCA + LDA) has been utilized. However, the advanced classifiers such as ensemble methods [7, 9, 19] or sparse-based classifiers [15, 30] can be used to improve the performance. To verify effectiveness of our feature template, we compare the Rank1 and Rank5 results of PGF + RSM with other related methods in Tables 9 and 10. From Table 9, we can see that the proposed recognition system provides competitive result with respect to other methods. In other words, the average Rank1 of our method promoted other recognition rates except the Gabor+RSM [9]. More precisely, in 6 out of 12 conditions (in probes A, B, C, H, I and J), the Rank1 accuracy is the same or better than Gabor+RSM. In the remaining 6 probes (D, E, F, G, K and L), the proposed results are still competitive. The results of Rank1 in Table 9 indicate that our method is robust to find briefcase conditions but unable to locate the patches in surface and clothing of USF dataset. In such conditions, the main parts of body (including the legs) are covered while a person is walking on a “grass” or has a different clothing. The Rank5 recognition rates of our

Table 8 Comparisons of Rank5 CCR (%) of different templates using a simple classifier on USF dataset

Exp.	A	B	C	D	E	F	G	H	I	J	K	L	W–AvgI
<i>Method</i>	Rank5 Performance												
Baseline	88	93	78	66	55	42	38	85	78	62	12	15	64.54
2DLDA	97	93	93	57	59	39	47	91	94	75	37	34	70.95
GTDA	100	97	95	57	54	34	45	75	80	70	25	25	66.15
Gabor+GTDA	96	95	89	59	63	33	49	94	92	76	19	40	70.32
DATER	96	96	93	69	69	51	52	92	90	83	40	36	75.68
MTP	94	93	91	64	68	51	52	88	83	82	18	15	71.38
GEI (Real)	93	93	89	65	60	42	45	88	79	80	6	9	68.68
CGI (Real)	95	94	93	66	65	48	52	96	97	87	24	27	75.05
PGM	-	-	-	-	-	-	-	-	-	-	-	-	-
GEI + SBDA	96	96	93	74	69	53	55	92	93	83	40	36	77.15
GSI	98	95	95	67	52	43	56	97	95	95	43	37	76.60
5/3GI-Contour	96	96	94	74	72	57	59	94	97	87	27	30	78.39
PGF + PCALDA (Ours)	99	95	95	80	81	55	61	96	93	91	31	29	80.67

The bold values indicate the values that outperform the previous works

Table 9 Comparisons of Rank1 CCR (%) of the approaches using advanced classifier on USF dataset

Exp.	A	B	C	D	E	F	G	H	I	J	K	L	W–AvgI
<i>Method</i>	Rank5 Performance												
LGSR	95	93	89	51	50	29	36	85	83	68	18	24	70.07
GEI + RSM	98	95	88	54	60	37	44	90	93	83	33	21	70.16
Gabor+RSM	100	95	94	73	73	55	64	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
GSTI	97	95	93	53	49	41	46	96	97	92	33	21	72.25
PGF + RSM (Ours)	100	96	98	62	59	43	46	100	99	94	28	30	76.01

The bold values indicate the values that outperform the previous works

method listed in Table 7 also outperform the CCR in most of the conditions. Moreover, in 3 out of 12 conditions (in probes A, H and I), the PGF + RSM can completely recognize all of the given individuals in probe set. In other testing conditions, the Rank5 CCR of proposed system is near to the best value and can recognize the people accurately. Indeed, it is noted from Tables 9 and 10 that our patch-based feature can represent the gait more perfectly rather than conventional patch-based methods, i.e. LGSR [30] and LPSELA [19]. The experimental results in Tables 9 and 10 verifies that using advanced classifier can make a robust gait recognition system which improves the average of Rank1 by about 5% with respect to other methods (i.e. LGSR [30], GEI + RSM [9], VI-MGR [4] and LPSELA [19]). To calculate the PGF, a small and limited number of parameters (such as size of patch window) are defined, which the recognition rate is not quite dependent to such parameters. As a result, it provides a useful representation tools to describe the gait in more challenging and complex conditions.

4.6 Computational complexity

In this section, the computational complexity and required memory to calculate PGF are evaluated. The timing complexity is also compared with two recent approaches (i.e. CGI [29] and GSTI [7]).

From three main parts of the proposed approach (in Fig. 1), the first and third steps take the highest processing time. Because Gabor filters are calculated from the silhouettes. While in the second step, the local extremums and the corresponding probability distributions are computed. Since the filtering of the silhouettes needs high processing time for computing the extremums, the time/memory of filtering steps have dominant complexities in our method.

Table 10 Comparisons of Rank5 CCR (%) of the approaches using advanced classifier on USF dataset

Exp.	A	B	C	D	E	F	G	H	I	J	K	L	W–AvgI
<i>Method</i>	Rank5 Performance												
LGSR	99	94	96	89	91	64	64	99	98	92	39	45	85.31
GEI + RSM	99	99	97	71	68	49	56	98	97	91	40	38	79.01
Gabor+RSM	100	98	98	85	84	73	79	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
GSTI	100	96	97	78	76	72	74	99	99	99	42	36	85.64
PGF + RSM (Ours)	100	98	98	80	77	77	60	100	100	99	48	45	86.59

The bold values indicate the values that outperform the previous works

In the first step, a set of 40D Gabor filters within the T (gait period) is calculated. If the time complexity of each filter is represented by $O(I_{Gfilt})$, then the time complexity of the first step will be equal to $O(40TI_{Gfilt})$.

However, to reduce complexities, the responses of filtering have been further downsampled to $\frac{1}{4}$ of original size. Therefore, the memory consumption will be in the order of $O(10TI_{Gfilt})$. Following such assumptions, the computational complexity of the third step is approximately equal to $O(40I_{Gfilt})$. As a result, timing complexity of the entire method is in the order of $O(40I_{Gfilt}) + O(40TI_{Gfilt}) \approx O(40(T+1)I_{Gfilt})$.

Considering the timing complexities of computing a PGF, the complexity of entire gallery and probe datasets is $O(40(T+1)(n_{gl} + n_{pr})I_{Gfilt})$ where n_{gl} and n_{pr} are the number of datasets in the gallery and probe sequences, respectively. If the size of input image is $W \times H$ and size of Gabor kernel is $w \times h$, the computational overhead of a typical filtering is of order $O(I_{Gfilt}) \approx O(WHwh)$ [6]. Considering this assumption, the general complexity of all the PGF templates is in the order of $O(40(T+1)(n_{gl} + n_{pr})WHwh)$.

Assuming that, in the CGI method [29] with the k color channel ($k=3$), the computational overhead would be in the order of $O(k(n_{gl} + n_{pr})TWH)$ [6, 29]. Moreover, the GSTI computational overhead is also in the order of $O(4(n_{gl} + n_{pr})TWHwh)$ [6]. As a result, the complexity of the proposed algorithm increases to $O\left(\binom{40}{k} \left(\frac{T+1}{T}\right) wh\right)$ as compared to CGI. By assuming $k=3$, the increase rate is approximately equal to $O(13wh)$. Meanwhile, in comparison with the GSTI, the timing complexity increases approximately $O(40\left(\frac{T+1}{T}\right)) \approx O(40)$. This overhead is not too problematic since we can alleviate the complexity with fast filtering techniques. In other words, the time of computing a PGF for an individual is 820 ms using MATLAB 8.3.0 (2014a) running on an Intel (R) Core (TM) i7 processor with 8 GB RAM working at 2.39-GHz for USF dataset.

From the memory consumption, and following similar discussion mentioned above, the most memory usage of the proposed approach is related to the first and third steps. At the first and third steps, the required memory is about $O(10TI_{Gfilt})$ and $O(10I_{Gfilt})$ respectively. However, the memory usage for the convolution of each Gabor filter is in the order of $O(I_{Gfilt}) \approx O(WH + wh)$ [6]. With this assumption, the required memory of the proposed approach is in the order of $O(10(T+1)(WH + wh))$. The memory usage of the approach for total silhouettes in the given dataset is $O(10(T+1)(WH + wh)(n_{gl} + n_{pr}))$. If we assume that the size of input silhouette in the USF dataset is 128×88 , the size of the Gabor function is 39×39 , the average of period in the dataset is $T \approx 30$, and $n_{gl} + n_{pr} = 1080$, the total amount of memory needed to calculate the proposed patch feature will be nearly 3.6 Gigabytes.

It should be noted that simple classifiers (e.g. 1-NN and PCA + LDA) have been proposed for evaluating the PGF templates. Obviously, using an advanced classifier such as ensemble method [7, 9] consumes extra computational cost and memory allocation for the recognition process.

5 Conclusion

In this paper, a new feature template has been proposed for human gait modeling that extracts spatio-temporal information of walking. The local patches are extracted in a walking sequence

and the probability distribution of patches in spatial and temporal domains are considered as gait signature. The corresponding coordinates in $2D + t$ domain are then weighted according to probabilities of local patch. The final augmented template, which is called PGF, is derived by concatenating a set of Gabor-based templates together with the weighted coordinates in the spatial and temporal domains. The proposed method provides an efficient gait representation template because local patches are described with their probability distributions (i.e. histogram) in full $2D + t$ domain. To make a gait recognition system based on PGF, two well-known classifiers (i.e. PCA + LDA and RSM) has been applied to classify the patterns. Extensive simulations and evaluations on the popular datasets demonstrate that our proposed method provide an enhanced performance in comparison with other state-of-the-art approaches. Moreover, our method has an optimal computational overhead that makes it proper for different gait conditions. To calculate PGF, there should be enough space for generating and storing the responses of Gabor filters. The proposed PGF template has efficient patch-based features to handle more challenging conditions in real life scenarios.

References

1. Atta R, Shaheen S, Ghanbari M (2017) Human identification based on temporal lifting using 5/3 wavelet filters and radon transform. *J Pattern Recognit* 69:213–224
2. Ben X, Gong C, Zhang P, Jia X, Meng QW a W (2019) Coupled patch alignment for matching cross-view gaits. *IEEE Trans Image Process* 28(6):3142–3157
3. Chen C, Zhang J, Fleischer R (2009) Multilinear tensor-based non-parametric dimension reduction for gait recognition. *Proceedings of 3rd International Conference on Biometrics*
4. Choudhury SD, Tjahjadi T (2015) Robust view-invariant multiscale gait recognition. *J Pattern Recognit* 48(3):798–811
5. Fendri E, Chtourou I, Hammami M (2019) Gait-based person re-identification under covariate factors. *Journal of Pattern Analysis and Application (PAA)*. <https://doi.org/10.1007/s10044-019-00793-4>
6. Ghaemina MH, Shokouhi SB (2018) GSI: Efficient Spatio-Temporal Template for Human Gait Recognition. *International Journal of Biometrics (IJBM)* 10(1):29–51
7. Ghaemina MH, Shokouhi SB (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
8. Ghebleh A, Moghaddam ME (2018) Clothing-invariant human gait recognition using an adaptive outlier detection method. *Journal of Multimedia Tools and Applications (MTAP)* 77(7):8237–8257
9. Guan Y, Li C-T, Roli F (2015) On reducing the effect of covariate factors in gait recognition: A classifier ensemble method. *IEEE Trans Pattern Anal Mach Intell* 37(7):1521–1528
10. Haifeng H (2013) Enhanced Gabor feature based classification using a regularized locally tensor discriminant model for multiview gait recognition. *IEEE Transactions on Circuits and Systems for Video Technology* 23(7):1274–1286
11. Han J, Bhanu B (2006) Individual recognition using gait energy image. *IEEE Trans Pattern Anal Mach Intell* 28(2):316–322
12. Hong S, Lee H, Kim E (2013) Probabilistic gait modelling and recognition. *IET Comput Vis* 7(1):56–70
13. 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
14. Kusakunniran W (2014) Recognizing gaits on spatio-temporal feature domain. *IEEE Transactions on Information Forensics and Security* 9(9):1416–1423
15. Lai Z, Xu Y, Jin Z, Zhang D (2014) Human gait recognition via sparse discriminant projection learning. *IEEE Trans Circuits Syst Video Technol* 24(10):1651–1662
16. Li M, Yuan B (2005) 2D-LDA: A statistical linear discriminant analysis for image matrix. *Pattern Recogn Lett* 26:527–532
17. Liu M, Yan S (2008) Y. Fu and and T. Huang, “Flexible X-Y patches for face recognition,” in *IEEE Int. Conf. Acoust., Speech, Signal Process*
18. Lucey S, Chen T (2004) A GMM parts based face representation for improved verification through relevance adaptation. *Proc IEEE Int Conf Comput Vis Pattern Recog*

19. 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 Recogn* 66:280–294
20. Makihara Y, Mannami H, Tsuji A, Hossain M, Sugiura K, Yag AM a Y (2012) The OU-ISIR Gait Database Comprising the Treadmill Dataset. *IPSJ Trans on Computer Vision and Applications* 4:53–62
21. More SA, Deore PJ (2012) A servey on gait biometrics. *World Journal of Science and Technology* 2(4): 146–151
22. Rahman S, See J (2016) Spatio-Temporal Mid-Level Feature Bank for Action Recognition in Low Quality Video. *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Shanghai
23. Sarkar S, Phillips PJ, Liu Z, Vega IR, Grother P, Bowyer KW (2005) The humanID gait challenge problem; data sets, performance, and analysis. *IEEE Trans Pattern Anal Mach Intell* 27(2):162–177
24. Shabani AH, Clausi DA, Zelek JS (2011) In: *Proceedings of the British Machine Vision Conference* (ed) Improved spatio-temporal salient feature detection for action recognition, Dundee
25. Shabani AH, Zelek JS, Clausi DA (2013) Multiple scale-specific representations for improved human action recognition. *Pattern Recogn Lett* 1(1):1771–1779
26. Tao D, Li X, Wu X, Maybank SJ (2007) General tensor discriminant analysis and gabor features for gait recognition. *IEEE Trans Pattern Anal Mach Intell* 29(10):1700–1715
27. Tong S, Fu Y, Ling XY a H (2018) Multi-View Gait Recognition Based on a Spatial-Temporal Deep Neural Network. *IEEE Access* 6:57583–57596
28. 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
29. Wang C, Zhang LWJ, Pu J, Yuan X (2012) Human identification using temporal information preserving gait template. *IEEE Trans Pattern Anal Mach Intell* 34(11):2164–2176
30. Xu D, Huang Y, Zeng Z, Xu X (2012) Human gait recognition using patch distribution feature and locality-constrained group sparse representation. *IEEE Trans Image Process* 21(1):316–326
31. Xu Z, Lu W, Zhang Q, Yeung Y, Chen X (2019) Gait recognition based on capsule network. *J Vis Commun Image Represent* 59:159–167
32. Xu D, Yan S, Tao D, Zhang L, Li X, Zhang H-J (2006) Human gait recognition with matrix representation. *IEEE Transactions on Circuits and Systems for Video Technology* 16(7):896–903
33. Yu S, Tan D, Tan T (2006) A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. *Proceeding of 18th International Conference on Pattern Recognition (ICPR)*, Hong Kong

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Mohammad H. Ghaemina received the BS degree in electronic engineering from University of Isfahan, Isfahan, Iran, in 2007, the MS degree in electronic engineering from Iran University of Science and Technology (IUST), Tehran, Iran, in 2010. He is currently finishing his PhD from IUST. His primary research interests fall in the field of computer vision, video analytics, biometric authentication, and pattern recognition for real world applications. He wrote a book with Dr. Shokouhi entitled *Fundamentals of Video-Based Object Tracking* in Persian in 2013 and some papers in well-known journals and conferences such as SIVP, IJBM, ICPR, CRV, MVIP and DICTA within 2010–2018.



Shahriar B. Shokouhi received his B.S. and M.S. degrees in Electronics in 1986 and 1989, both from the Department of Electrical Engineering of Iran University of Science & Technology (IUST). He received his Ph.D. on Electronics and Image Processing in 1999 from School of Electronic & Electrical Engineering, University of Bath, UK. He is currently an associate professor and director of undergraduate and graduate studies at the electrical engineering department of IUST. He is a member of IEEE. His research interests include Machine Vision, Trusted Hardware Design, and Intelligent Systems Design.



Ali Badiezhadeh received the B.S. degree from Shahid Chamran University, Ahvaz, Iran, in 2014 and the M.S. degree from Iran University of Science and Technology (IUST), Tehran, Iran, in 2017 both in electrical engineering. His research interests include image and video processing, machine learning, biometrics and deep learning.