

A View Transformation Model Based on Sparse and Redundant Representation for Human Gait Recognition

Abstract

Background: Human gait as an effective behavioral biometric identifier has received much attention in recent years. However, there are challenges which reduce its performance. In this work we aim at improving performance of gait systems under variations in view angles, which present one of the major challenges to gait algorithms. **Methods:** We propose employment of a view transformation model based on sparse and redundant (SR) representation. More specifically, our proposed method trains a set of corresponding dictionaries for each viewing angle, which are then used in identification of a probe. In particular, the view transformation is performed by first obtaining the SR representation of the input image using the appropriate dictionary, then multiplying this representation by the dictionary of destination angle to obtain a corresponding image in the intended angle. **Results:** Experiments performed using CASIA Gait Database, Dataset B, support the satisfactory performance of our method. It is observed that in most tests, the proposed method outperforms the other methods in comparison. This is especially the case for large changes in the view angle, as well as the average recognition rate. **Conclusion:** A comparison with state-of-the-art methods in the literature showcases the superior performance of the proposed method, especially in the case of large variations in view angle.

Keywords: *Biometrics, gait analysis, human identification, sparse and redundant representation, view transformation model, view-invariant*

Submitted: 23-Oct-2019

Revision: 04-Dec-2019

Accepted: 14-Jan-2020

Published: 03-Jul-2020

Introduction

Human gait enjoys distinctive features such as recognition from distance and unobtrusiveness. Moreover, a typical gait recognition system does not require high quality video, which makes it inexpensive and easy to set up since one can use existing security cameras for gait recognition. These characteristics have helped the popularity among research community of human gait in recent years as a behavioral biometric identifier. Gait is particularly applicable in criminal cases. As a matter of fact, a number of criminal cases have already used gait, for example in identifying a bank robber^[1] and a burglar.^[2] However, human gait recognition suffers from some challenges and difficulties such as variations in viewpoint,^[3] walking speed,^[4] clothing,^[5] and carry conditions.^[6] Since gaits from the same person vary drastically from different viewpoints and camera view

is always unconstrained in real scenarios, this variation is one of the most critical challenges in human gait recognition. Therefore, gait recognition performance is dramatically dropped by changing in viewing angles.^[7]

This work proposes a robust scheme for gait recognition which is shown to be tolerant against variations in the viewing angle. The proposed scheme uses a cross-view approach to gait recognition, where the probe and gallery gaits are captured from two distinct viewpoints. Other approaches to this problem include fixed-view, where probe and gallery sequences are captured from the same viewpoint, and multi-view, where the probe sequence is recorded from single view and is processed against gallery gaits from multiple views. The most common approach among these in the literature is the cross-view approach.

The proposed scheme employs a view transformation model (VTM) based on sparse and redundant (SR) representation.

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

How to cite this article: Ghebleh A, Moghaddam ME. A view transformation model based on sparse and redundant representation for human gait recognition. *J Med Signals Sens* 2020;10:135-44.

**Abbas Ghebleh,
Mohsen Ebrahimi
Moghaddam**

*Department of Computer
Engineering and Science,
Shahid Beheshti University,
Tehran, Iran*

Address for correspondence:

*Dr. Mohsen Ebrahimi
Moghaddam,
Department of Computer
Engineering and Science,
Shahid Beheshti University,
Daneshjo Blv, Velenjak, Tehran,
Iran.
E-mail: m_moghadam@sbu.
ac.ir*

Access this article online

Website: www.jmssjournal.net

DOI: 10.4103/jmss.JMSS_59_19

Quick Response Code:



The VTM tries to learn an association between gait features from different viewing angles to map gait features from one view to another, and in turn facilitates comparison of probe gait from one viewpoint with gallery gaits of another viewpoint.

This paper is structured as follows: Section 2 gives a brief review of some existing schemes related to the present work. Section 3 provides the necessary background. The proposed scheme is recounted in Section 4. Section 5 presents empirical results showcasing the performance of the proposed scheme, as well as comparison with the existing work. Section 6 contains some concluding remarks.

Related Work

Human gait recognition schemes are typically categorized as model-based schemes and appearance-based schemes. Considering the model-based approaches like those in some studies,^[8-16] one aims at fitting each frame of the input gait sequence to a specific model of the human body. This is achieved via determination of the parameters (e.g., joint angles) of the model at hand. The obtained parameters of the model are then used as features for identification of a probe sequence against stored gallery data. In contrast, appearance-based approaches such as described in various studies^[17-31] focus on the shape of the subject's silhouettes in input frames and uses these to compute the desired feature. This usually leads to a single representation for a gait cycle. The most common of such representations is the gait energy image (GEI)^[17] which is simply the statistical mean, after alignment and normalization, of all silhouettes of a gait cycle.

The model-based approach seems to be more appropriate for cross-view gait recognition since given certain geometric assumptions, the calculated model have a view-invariant nature. However, these approaches generally suffer from model fitting errors due to typical low spatial resolution of the input frames.^[32] Appearance-based approach on the other hand, can recognize a subject even from relatively low spatial resolution images, while its performance suffers drastically under variations in the viewing angle. While the proposed scheme is based on the appearance-based approach, it aims at mitigating the challenge of viewpoint variations.

Appearance-based cross-view gait recognition schemes typically fall under one of the following descriptions:^[33] (1) those that use multiple cameras or camera calibrations to construct a three-dimensional (3-D) model; (2) those that aim at extracting view-invariant gait features; and (3) those that aim at learning cross-view mapping relationships of gait features.

Schemes following the first approach^[14,16,34-39] construct a 3-D model using cooperative multiple cameras or camera calibration and then project the obtained 3-D gallery into a 2-D silhouette. In theory, 2-D gaits for any desired view

can be obtained from the 3-D model, yet there are some practical limitations to this.^[3] This approach is suitable for a fully controlled and cooperative multi-camera setting, for example, a biometric tunnel^[40] which is expensive and complicated. Also, the processes of 3-D reconstruction and 2-D rendering are computationally demanding.

The second approach employs view-invariant features to facilitate cross-view gait recognition. A brief description of some examples of this approach follows. The method of BenAbdelkader *et al.*^[41] uses a self-similarity plot to achieve robustness against limited view changes which achieved good performance with a limited range of view changes. Kale *et al.*^[42] make use of a perspective projection model to compute from an arbitrary view the side-view gaits. Jean *et al.*^[43,44] normalize all input data (from any view) onto a fixed plane, thus allowing direct comparison in that plane. Han *et al.*^[45] propose to select only parts of GEIs that overlap between various views to make a representation for the cross-view comparison. Finally, a joint subspace learning method is proposed by Liu *et al.*^[46] to mitigate the view variations challenge. This category works well when the angle between the sagittal plane of the person and the image plane is small; otherwise it fails.^[33] Furthermore, these methods are sensitive to noise which negatively affects recognition rates.^[3]

The third category of appearance-based cross-view gait recognition schemes maps gait features from one view to another by first training on the projection rule between the two views. Makihara *et al.*^[28] propose a VTM based on singular value decomposition (SVD). In addition to the SVD-based VTM, Kusakunniran *et al.*^[47] approaches optimization of the GEI feature through linear discriminant analysis (LDA). In a study by Zheng *et al.*,^[48] a method is proposed to obtain the VTM using robust principal component analysis. Other approaches to construction of appropriate VTMs include those that employ tools such as support vector regression,^[49] multilayer perceptron,^[50] and sparse regression.^[3] The method presented in Chen *et al.*^[51] constructs a VTM based on projection of gravity center trajectory and Kusakunniran *et al.*^[29] further improve the performance of this method. The method in study by Liu and Tan^[52] trains LDA-subspaces for constructing a VTM and Bashir *et al.*^[53] use canonical correlation analysis. The method in Kusakunniran *et al.*^[49] reformulates VTM construction as a regression problem. Hu *et al.*^[54] propose to apply a projection named view-invariant discriminative projection. Hu^[55] proposes enhanced Gabor gait which is a gait feature that includes a nonlinear mapping of statistical and structural characteristics of gait. Muramatsu *et al.*^[56,57] propose to use 3D training gait models to create a VTM.

Comparing to the first category, methods in this category are more feasible and less expensive since they do not use complicated multi-camera systems. Besides, they are more efficient and stable than the second category because

they are less sensitive to noise.^[3] Also, these methods can be applied for scenarios with no explicit interaction with the subjects, and can also be directly applied to views which are significantly different from the side view.^[58] One limitation of the third category is that these methods rely on supervised learning and it will be difficult for recognizing gait under untrained viewing angles.^[3] There are some works to address this challenge. For instance, Tian *et al.*^[59] proposes an innovative view-adaptive mapping approach. However, as mentioned in Yu *et al.*,^[60] small changes in view angles do not affect the recognition rates significantly, and if a sufficient number of cameras are employed, this challenge would be negligible. Another challenge in the third category is that most of the above mentioned methods train multiple mapping matrices, one for each pair of viewpoints.^[58] Also, performance of cross-view gait recognition drops when the change in viewing angle is large.^[3] The proposed method tries to mitigate these limitations. It creates a dictionary per each view angle and as the results show it performs well for big changes in view angle.

Background

VTM-based human gait recognition

The flowchart in Figure 1 presents a general outline of a VTM-based human gait recognition scheme. Such a scheme generally consists of three phases:^[33] Training, transformation and matching. In the training phase, the gait features of multiple training subjects are used to construct the appropriate VTM. This VTM is then used in the transformation phase to compute an input gait feature for a destination view from the source view. Finally, the matching phase calculates a score of recognition between the gait features of the probe and every gallery subject.

It should be noted that prior to these phases, a preprocessing module is needed which removes the background and extracts the silhouette from each frame. A simple yet effective method is to record the background in advance and subtract this recorded image from the input frame. The silhouette of the person is then obtained using some morphological operations. This method is easily applicable for security cameras in real applications. Some other methods for background detection and removal are described in study by Piccardi.^[61] The extracted silhouettes are then passed on to a feature extraction module which computes the desired features. These features are then passed to the training phase. Another important operation in preprocessing module is to detect the view angle of input sequence. In this work like most VTM methods, we assume that this angle is known. However, there are some methods like study by Chtourou *et al.*^[62] which can be used for walking direction estimation.

Sparse and redundant representation

The SR representation model has attracted great attention in past decades. This model is used to represent signals and images and yields great performance in many applications such as noise removal, image separation, and image compression. The main idea behind this model is that each signal can be obtained by a weighted sum of some basic atoms.^[63] More specifically,

$$x = D\alpha \quad (1)$$

Where x is the signal, D is a full-rank matrix called dictionary in which each column is an atom, and α is the SR representation of the signal x . In other words, α is a vector containing the weights of atoms. So, we can say that:

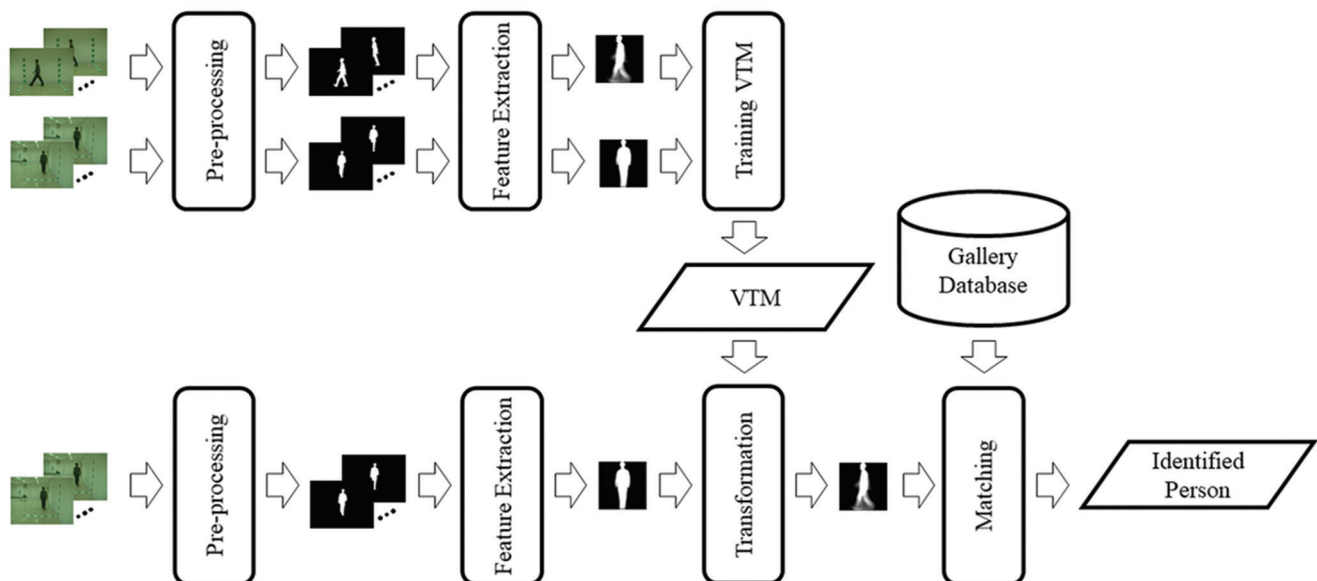


Figure 1: A general framework for VTM-based human gait recognition

$$\alpha = D^+ x \quad (2)$$

Where D^+ is a pseudoinverse of D . The number of atoms is typically more than the signal length ($m > n$), so the representation is called redundant. Furthermore, an important property of this model is sparsity, which means most values in α are zero. α is defined as the sparsest vector that can model x with at most ε errors:

$$\alpha = \underset{\alpha}{\operatorname{argmin}} \|\alpha\|_0 \quad \text{Subject to } \|D\alpha - x\|_2 < \varepsilon \quad (3)$$

Where $\|\cdot\|_0$ is the l^0 norm counting the nonzero entries of a vector. Solving the above equation is an NP-Hard problem but there are some estimation approaches that calculate α with good precision such as orthogonal matching pursuit which is used in this work.^[64]

Another challenge in this model is the dictionary. The dictionary must be rich enough to adequately describe the input signal in a sparse manner. There are some methods like K-SVD^[63] which are used to train and obtain the desired dictionary.

Proposed Method

We propose a VTM based on SR representation. Given the input from angle v_s , this model generates the corresponding output in another angle v_t .

Main idea

Assume that D_{v_t} and D_{v_s} are two dictionaries containing m atoms each. We call these dictionaries a *transform pair* if and only if for each $1 \leq i \leq m$, the i^{th} atom of them were transform pair. Two atoms are called transform pair if and only if they correspond to same regions in different view angles.

Assume that D_{v_t} and D_{v_s} are two dictionaries for view angles v_t and v_s respectively. If they are a transform pair then:

$$x_{v_t} = D_{v_t} \alpha = D_{v_t} D_{v_s}^+ x_{v_s} \quad (4)$$

Where x_{v_s} and x_{v_t} represents the image in source and target view angles and D_{v_s} and D_{v_t} denote the dictionaries corresponding to these angles. Loosely speaking, the input image (x_{v_s}) is coded using the corresponding dictionary (D_{v_s}) to obtain α which is then multiplied to the dictionary of target view (D_{v_t}) that produces the image in target view (x_{v_t}). The next section describes how to obtain a transform pair of dictionaries.

Training the VTM

As mentioned before, the K-SVD algorithm is used to train a dictionary using training data. Hence, we can use samples of each view to train a dictionary for that view. However, in our method, the corresponding atoms of the dictionaries must be correlated. For example, if the patch related to head

area is composed of atoms number 1, 3, 6, and 7 of D_{v_s} , then composing these exact atoms of D_{v_t} should make the head area in v_t . Training dictionaries independently loses these constraints. To mitigate this issue we propose the scheme presented in Figure 2. In this scheme, at first the corresponding patches from training samples in both view angles are extracted and linearized and then concatenated to form a bigger train sample. Samples obtained in this way are then used to train a dictionary (D_{all}). After that, D_{all} is split horizontally to make the desired dictionaries. Note that in this way, corresponding atoms in the dictionaries are correlated.

Transformation

Let p_{v_s} be a probe sample captured in v_s view angle. We first split p_{v_s} into n patches $p_{v_s}^1, p_{v_s}^2, \dots, p_{v_s}^n$. Each patch $p_{v_s}^i$ is encoded using $D_{v_s}^i$ which is trained using patches from the same location of train samples in v_s . The result would be the SR representation of the patch ($\alpha_{v_s}^i$) which is then decoded using $D_{v_d}^i$ to make $p_{v_d}^i$, the transformed patch in v_d . Finally, the transformed patches are merged to make the transformed probe (p_{v_d}). More formally:

$$p_{v_d}^i = d(e(p_{v_s}^i, D_{v_s}^i), D_{v_d}^i) \quad (5)$$

Where $e(\cdot)$ and $d(\cdot)$ denotes encoding and decoding respectively. The transformation process is presented in Figure 3.

Matching

After obtaining the transformed probe in the same view angle as gallery, we can compare them and find the similarity of the probe with each gallery sample using any criterion such as Euclidian distance. Then a classifier such as nearest neighbor is used to find the most similar sample. However, using the above method for transformation leads to some artifacts such as chessboard effect which affect the recognition process. To eliminate the chessboard effect we could use overlapping patches but this process increases the computational cost significantly. An alternate less costly approach is to use $\alpha_{v_s}^i$ as a feature instead of decoding to obtain the patch. We refer to this feature as SR feature in the following.

Instead of comparing two images we may compare their SR representation with respect to the same dictionary D . More formally, let x_1 and x_2 be two patches. Then:

$$x_1 = D\alpha_1 \quad (6)$$

$$x_2 = D\alpha_2 \quad (7)$$

Hence, the similarity of x_1 and x_2 can be estimated by similarity of α_1 and α_2 . In this way, the computation cost is reduced and the effect of artifacts is mitigated. In addition, using SR representation mitigates noises and less important data. Evaluation results show that using this representation leads to acceptable recognition rates.

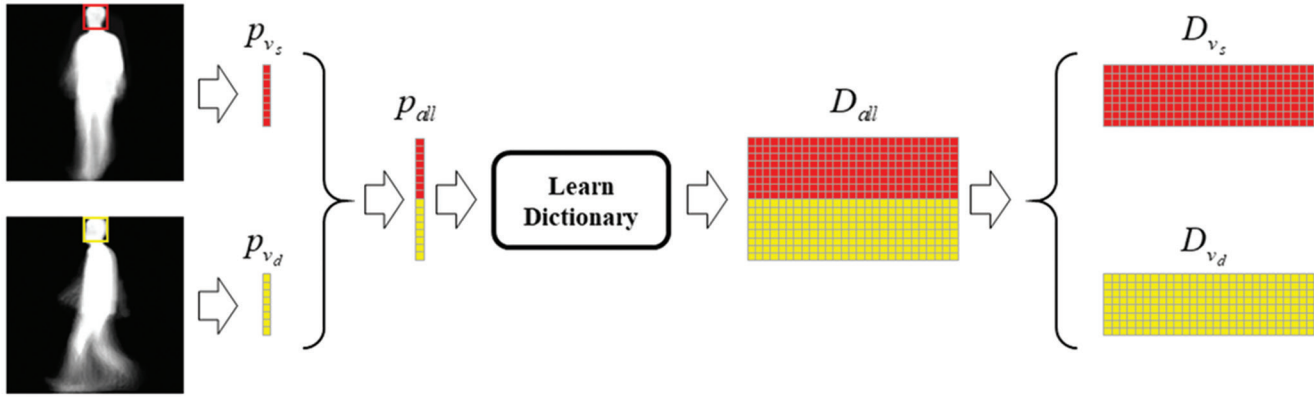


Figure 2: The scheme of training view transformation model

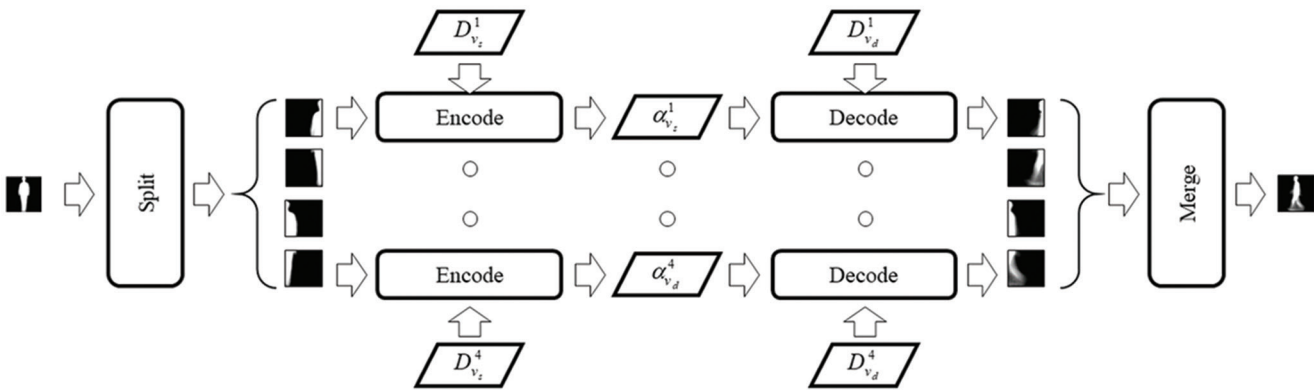


Figure 3: The view transformation process

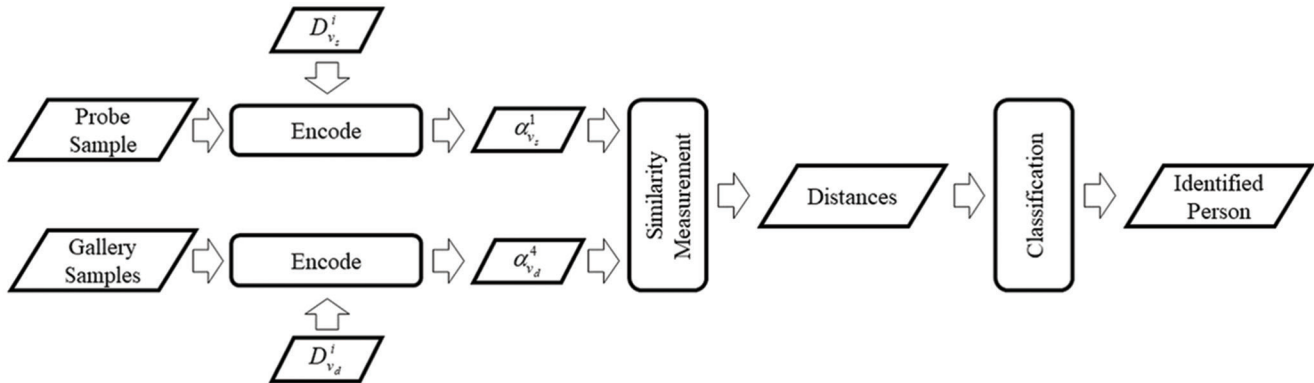


Figure 4: The transformation and matching processes

Figure 4 illustrates the transformation and matching processes. Probe sample is split and each patch $P_{v_s}^i$ is encoded using $D_{v_s}^i$. Similar process is performed for each patch at location i of j^{th} the gallery sample $g_{v_d}^{j,i}$ using $D_{v_d}^i$. After that, the dissimilarity of probe and gallery samples is calculated using their sparse and redundant representation. More formally:

$$D(p_{v_s}, g_{v_d}^j) = \sum_{i=1}^n f(\alpha_{v_s}^i, \alpha_{v_d}^i) \quad (8)$$

Where $f(\cdot)$ denotes the Euclidian distance. Finally, the gallery sample with minimum distance represents the subject.

Computational complexity

Considering the proposed method, the critical modules for analyzing the computational complexity are training VTM which is performed once and transformation that is executed for each probe. According to Rubinstein *et al.*,^[65] the computational

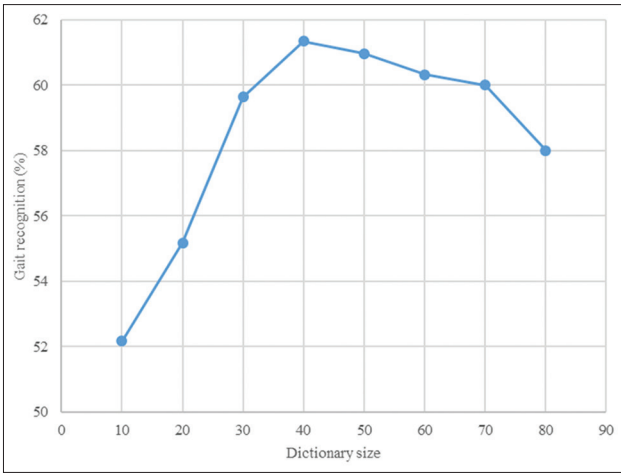


Figure 5: Average rank-1 recognition rates against dictionary size

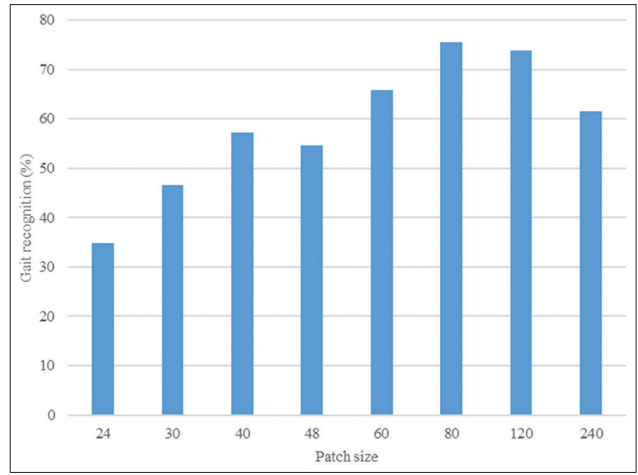


Figure 6: Average rank-1 recognition rates against patch size with 40 as dictionary size

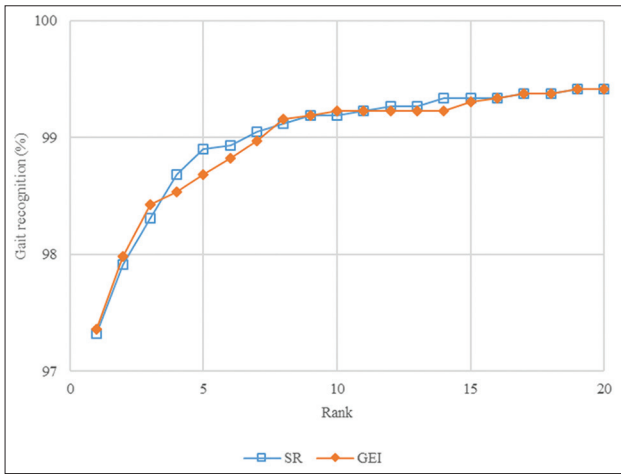


Figure 7: Average rank-1 recognition rates when probe and gallery are from the same angle using SR and GEI as feature

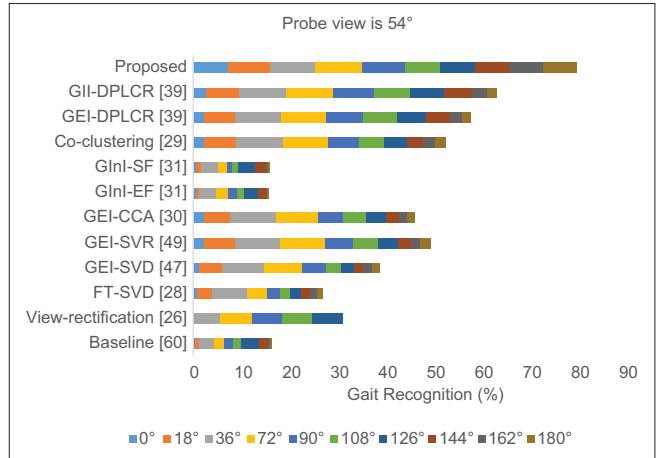


Figure 8: Performance comparison between rank-1 recognition rates of different methods for probe view 54° and various gallery views in the range from 0° to 180°

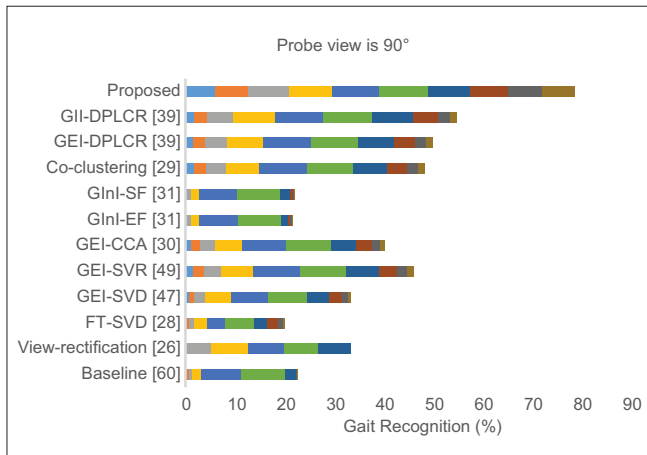


Figure 9: Performance comparison between rank-1 recognition rates of different methods for probe view 90° and various gallery views in the range from 0° to 180°

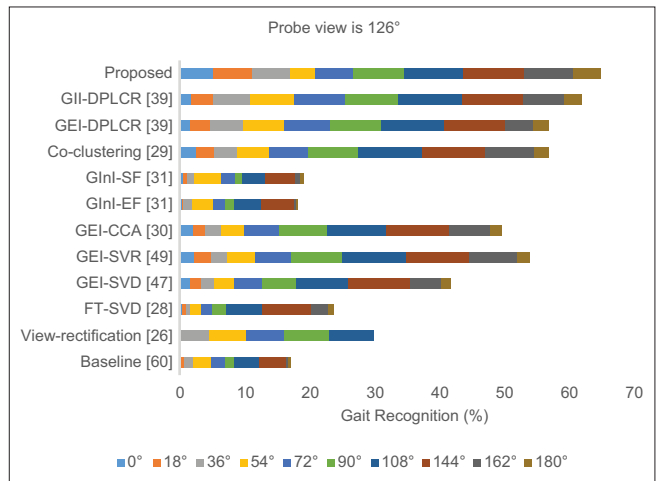


Figure 10: Performance comparison between rank-1 recognition rates of different methods for probe view 126° and various gallery views in the range from 0° to 180°

complexity for K-SVD algorithm which is used for training VTM is:

$$T_{K-SVD} \approx n.(K^2L + 2NL) \tag{9}$$

Where n is the number of training signals, k is target sparsity, L denotes number of atoms in dictionary and N is the signal length.

As mentioned before, we use OMP for transformation module. Its computational complexity is:

$$T_{OMP} = K^3 + 2KNL \tag{10}$$

Experimental Validation

Dataset

The CASIA gait database, Dataset B^[48,60] is utilized to assess the proposed method which contains sequences from 124 subjects. Eleven view angles (0°, 18°, 36°, 54°, 72°, 90°, 108°, 126°, 144°, 162°, and 180°) are considered and for each angle 6 sequences are recorded with normal clothing. Four sequences are used as gallery and others are used as probe samples.

Selection of parameter values

There are two main parameters which affect the performance of the proposed approach: Size of dictionary and patch size. To obtain the appropriate size of dictionaries and patch size, we have tested multiple values for these parameters. Dictionary size varies from 10 to 80, and seven different values for patch size are considered. The gallery view is 90° and all angles are used as probe view.

We use the average recognition rates with all angles and patch sizes per dictionary size to find the appropriate dictionary size as depicted in Figure 5. It is obvious that 40 has the best performance among others.

To find the appropriate patch size, using 40 as dictionary size the average recognition rates with all angles per each patch size is obtained that depicted in Figure 6. We can see that the 80 is the best choice for patch size.

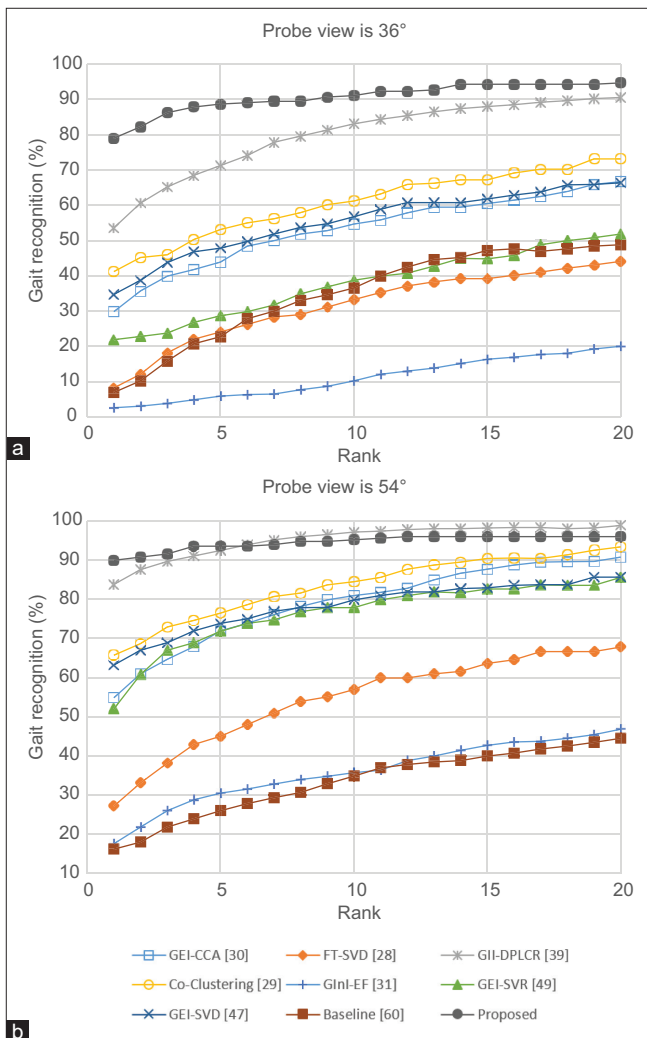


Figure 11: Plot of CMS curves of the methods in comparison for gallery view 90° and probe view (a) 36° and (b) 54°

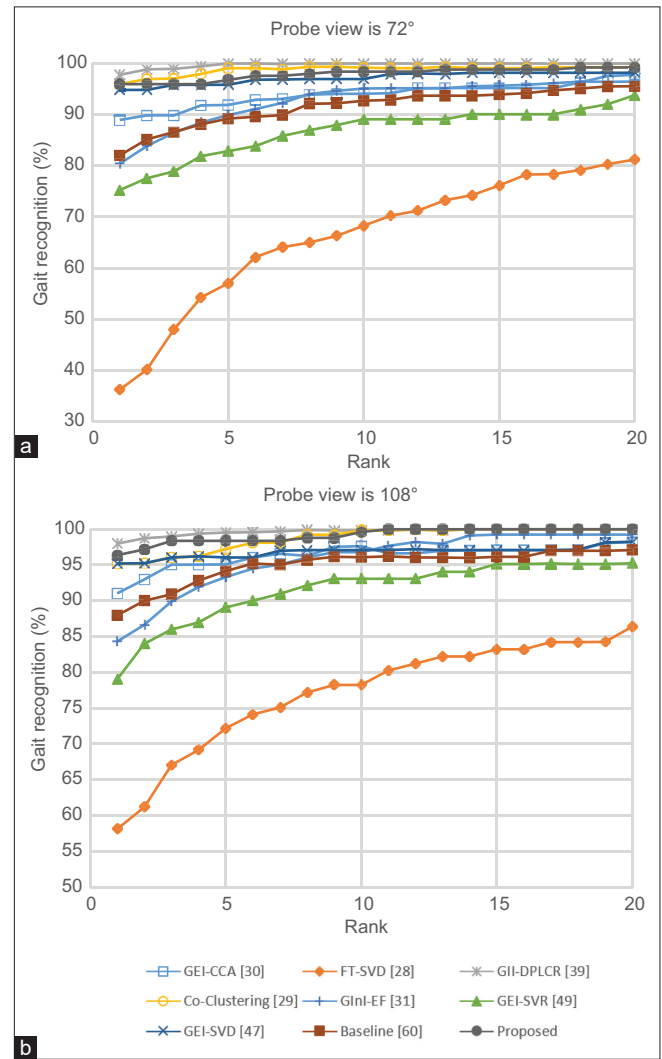


Figure 12: Plot of CMS curves of the methods in comparison for gallery view 90° and probe view (a) 72° and (b) 108°

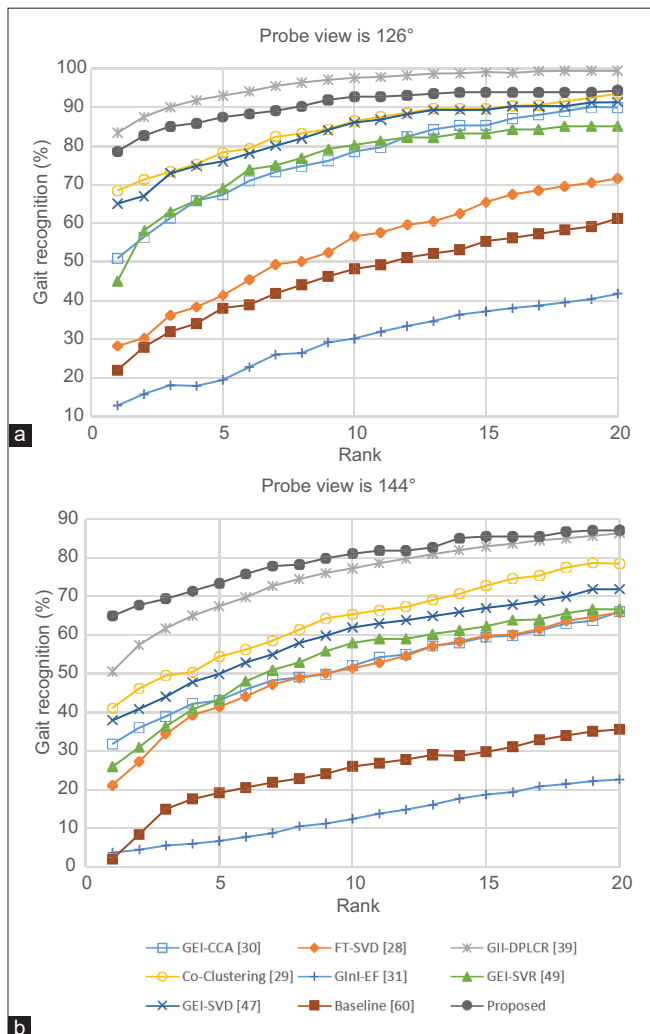


Figure 13: Plot of CMS curves of the methods in comparison for gallery view 90° and probe view (a) 126° and (b) 144°

Sparse and redundant feature

The performance of SR representation as a gait feature is investigated in this section. Towards this end, we have compared SR feature with the case which uses GEI as feature using Cumulative Matching Score (CMS) curves.^[6] The CMS or Cumulative Matching Characteristic is a rank based method of showing measured accuracy performance of a biometric system. The horizontal axis of the CMS graph is rank and the vertical axis is the recognition rate. The value for rank r shows the recognition rate within first r ranks.

Figure 7 shows the CMS curves of average recognition rates of GEI and SR features when gallery and probe are from the same angle. It can be observed that SR feature almost performs as well as GEI and both features have good performance when there is no view change. Therefore, using SR feature instead of GEI is acceptable.

Recognition rates

In study by Kusakunniran *et al.*,^[29] the angles 54°, 90°, and 126° of the dataset are used as probe views, while the remaining ten viewing angles are taken as gallery views for training the VTM. We follow these choices in testing the performance of the proposed method. We then use the obtained results to compare the performance of the proposed scheme with the algorithm of Zhaoxiang Zhang *et al.*^[39] and all the algorithm compared there-in. It is worth to mention that the baseline method as explained in Yu *et al.*^[60] is a simple method that does not do any action to mitigate the view angle challenge. It simply extracts GEI as feature, measures the distances using Euclidian distance and uses nearest neighbor as classifier. The reason of comparing this method is to highlight the effectiveness of other methods.

Comparisons of the rank-1 recognition rates according to each probe view 54°, 90°, and 126° are presented in Figures 8-10. It is observed that in most tests, the proposed method outperforms the other methods in comparison. This is especially the case for large changes in the view angle, as well as the average recognition rate. For example, when the probe view is 54°, the recognition rate of the proposed method is 46%, 40%, and 47% higher than the second best method with 0°, 162°, and 180° as gallery views, respectively. In addition, the proposed scheme performs approximately 16% better than GII-DPLCR which is the second best method. Similar results can be observed when the probe view is 90° and 126°.

Figures 11-13 report CMS curves for gallery view of 90° and various probe views in the range from 36° to 126° (except 90°). Investigating these curves, it is seen that for 36°, 54°, and 144° view angles, the proposed method outperforms others while it has competitive performance for the other view angles.

Concluding Remarks

Using SR representation, we propose in this work a VTM for cross-view gait recognition. We verify satisfactory performance of the proposed scheme using the CASIA Gait Database, Dataset B. These test results illustrate superiority of the proposed method in comparison with several state-of-the-art methods, especially in the case of large changes in the view angle. It is also observed that the average recognition rates for all angles are higher than these existing methods.

Financial support and sponsorship

None.

Conflicts of interest

There are no conflicts of interest.

References

- Larsen PK, Simonsen EB, Lynnerup N. Gait analysis in forensic medicine*. *J Forensic Sci* 2008;53:1149-53.
- BBC. How can you Identify a Criminal by the way they Walk? BBC Magazine; 2008. Available from: <http://news.bbc.co.uk/1/hi/magazine/7348164.stm>. [Last accessed on 2019 Oct 20].
- Kusakunniran W, Wu Q, Zhang J, Li H. Gait recognition under various viewing angles based on correlated motion regression. *IEEE Trans Circuits Syst Video Technol* 2012;22:966-80.
- Kovač J, Štruc V, Peer P. Frame-based classification for cross-speed gait recognition. *Multimed Tools Appl* 2019;78:5621-43.
- Ghebleh A, Moghaddam ME. Clothing-invariant human gait recognition using an adaptive outlier detection method. *Multimed Tools Appl* 2018;77:8237-57.
- Sarkar S, Phillips PJ, Liu Z, Vega IR, Grother P, Bowyer KW. The humanID gait challenge problem: Data sets, performance, and analysis. *IEEE Trans Pattern Anal Mach Intell* 2005;27:162-77.
- Yu S, Tan D, Tan T. Modelling the effect of view angle variation on appearance-based gait recognition. In: *Asian Conference on Computer Vision*. Springer, Berlin, Heidelberg; 2006. p. 807-816.
- Bobick AF, Johnson AY, editors. Gait recognition using static, activity-specific parameters. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE; 2001.
- Cunado D, Nixon MS, Carter JN. Automatic extraction and description of human gait models for recognition purposes. *Comput Vis Image Underst* 2003;90:1-41.
- Zhang R, Vogler C, Metaxas D. Human gait recognition at sagittal plane. *Image Vis Comput* 2007;25:321-30.
- Wang L, Ning H, Tan T, Hu W. Fusion of static and dynamic body biometrics for gait recognition. *IEEE Trans Circuits Syst Video Technol* 2004;14:149-58.
- Dockstader SL, Berg MJ, Tekalp AM. Stochastic kinematic modeling and feature extraction for gait analysis. *IEEE Trans Image Process* 2003;12:962-76.
- Haiping L, Konstantinos NP, Anastasios NV. A full-body layered deformable model for automatic model-based gait recognition. *EURASIP J Adv Signal Process* 2007;2008:1-13.
- Zhao G, Liu G, Li H, Pietikainen M, editors. 3D gait recognition using multiple cameras. In: *7th International Conference on Automatic Face and Gesture Recognition*. IEEE; 2006.
- Ariyanto G, Nixon MS, editors. Model-based 3D gait biometrics. *International Joint Conference on Biometrics*. IEEE; 2011.
- Bodor R, Drenner A, Fehr D, Masoud O, Papanikolopoulos N. View-independent human motion classification using image-based reconstruction. *Image Vis Comput* 2009;27:1194-206.
- Han J, Bhanu B. Individual recognition using gait energy image. *IEEE Trans Pattern Anal Mach Intell* 2006;28:316-22.
- Wang L, Tan T, Ning H, Hu W. Silhouette analysis-based gait recognition for human identification. *IEEE Trans Pattern Anal Mach Intell* 2003;25:1505-18.
- Bobick AF, Davis JW. The recognition of human movement using temporal templates. *IEEE Trans Pattern Anal Mach Intell* 2001;23:257-67.
- Liu J, Zheng N, editors. Gait history image: A novel temporal template for gait recognition. In: *IEEE International Conference on Multimedia and Expo*. IEEE; 2007.
- Lam TH, Cheung KH, Liu JN. Gait flow image: A silhouette-based gait representation for human identification. *Pattern Recognit* 2011;44:973-87.
- Yang X, Zhou Y, Zhang T, Shu G, Yang J. Gait recognition based on dynamic region analysis. *Signal Process* 2008;88:2350-6.
- Zhang E, Zhao Y, Xiong W. Active energy image plus 2DLPP for gait recognition. *Signal Process* 2010;90:2295-302.
- Bashir K, Xiang T, Gong S. *Gait Recognition Using Gait Entropy Image*; 2009.
- LI X, Chen Y. Gait recognition based on structural gait energy image. *J Comput Inf Syst* 2013;9:121-6.
- Goffredo M, Bouchrika I, Carter JN, Nixon MS. Self-calibrating view-invariant gait biometrics. *IEEE Trans Syst Man Cybern B Cybern* 2010;40:997-1008.
- Kusakunniran W, Wu Q, Zhang J, Ma Y, Li H. A new view-invariant feature for cross-view gait recognition. *IEEE Trans Inf Forensics Secur* 2013;8:1642-53.
- Makihara Y, Sagawa R, Mukaigawa Y, Echigo T, Yagi Y, editors. Gait recognition using a view transformation model in the frequency domain. In: *European Conference on Computer Vision*. Springer; 2006.
- Kusakunniran W, Wu Q, Zhang J, Li H, Wang L. Recognizing gaits across views through correlated motion co-clustering. *IEEE Trans Image Process* 2014;23:696-709.
- Bashir K, Xiang T, Gong S, Mary Q, editors. *Gait Representation Using Flow Fields*. BMVC; 2009.
- Arora P, Hanmandlu M, Srivastava S. Gait based authentication using gait information image features. *Pattern Recognit Lett* 2015;68:336-42.
- Chai Y, Ren J, Han W, Li H. *Human Gait Recognition: Approaches, Datasets and Challenges*; 2011.
- Muramatsu D, Makihara Y, Yagi Y. View transformation model incorporating quality measures for cross-view gait recognition. *IEEE Trans Cybern* 2016;46:1602-15.
- Shakhnarovich G, Lee L, Darrell T, editors. Integrated face and gait recognition from multiple views. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE; 2001.
- Zhang Z, Troje NF. View-independent person identification from human gait. *Neurocomputing* 2005;69:250-6.
- López-Fernández D, Madrid-Cuevas FJ, Carmona-Poyato A, Muñoz-Salinas R, Medina-Carnicer R. A new approach for multi-view gait recognition on unconstrained paths. *J Vis Commun Image Represent* 2016;38:396-406.
- Goffredo M, Seely RD, Carter JN, Nixon MS, editors. Markerless view independent gait analysis with self-camera calibration. In: *8th IEEE International Conference on Automatic Face & Gesture Recognition*. IEEE; 2008.
- Iwashita Y, Baba R, Ogawara K, Kurazume R, editors. Person identification from spatio-temporal 3D gait. In: *International Conference on Emerging Security Technologies (EST)*. IEEE; 2010.
- Zhaoxiang Zhang, Jiabin Chen, Qiang Wu, Ling Shao. GII representation-based cross-view gait recognition by discriminative projection with list-wise constraints. *IEEE Trans Cybern* 2018;48:2935-47.
- Seely RD, Samangoeei S, Lee M, Carter JN, Nixon MS, editors. The university of southampton multi-biometric tunnel and introducing a novel 3d gait dataset. In: *2nd IEEE International Conference on Biometrics: Theory, Applications and Systems*. IEEE; 2008.
- BenAbdelkader C, Cutler RG, Davis LS. Gait recognition using image self-similarity. *EURASIP J Adv Signal Process* 2004;2004:572-85.

42. Kale A, Chowdhury AR, Chellappa R, editors. Towards a view invariant gait recognition algorithm. In: Proceedings IEEE Conference on Advanced Video and Signal Based Surveillance. IEEE; 2003.
43. Jean F, Bergevin R, Albu AB, editors. Trajectories normalization for viewpoint invariant gait recognition. In: 19th International Conference on Pattern Recognition. IEEE; 2008.
44. Jean F, Bergevin R, Albu AB. Computing and evaluating view-normalized body part trajectories. *Image Vis Comput* 2009;27:1272-84.
45. Han J, Bhanu B, Roy-Chowdhury AK, editors. A study on view-insensitive gait recognition. In: IEEE International Conference on Image Processing. IEEE; 2005.
46. Liu N, Lu J, Tan YP. Joint subspace learning for view-invariant gait recognition. *IEEE Signal Process Lett* 2011;18:431-4.
47. Kusakunniran W, Wu Q, Li H, Zhang J, editors. Multiple views gait recognition using view transformation model based on optimized gait energy image. In: IEEE 12th International Conference on Computer Vision Workshops. IEEE; 2009.
48. Zheng S, Zhang J, Huang K, He R, Tan T, editors. Robust view transformation model for gait recognition. In: 18th IEEE International Conference on Image Processing; 2011: IEEE; 2011.
49. Kusakunniran W, Wu Q, Zhang J, Li H, editors. Support vector regression for multi-view gait recognition based on local motion feature selection. In: CVPR, 2010: IEEE Conference on Computer Vision and Pattern Recognition; IEEE; 2010.
50. Kusakunniran W, Wu Q, Zhang J, Li H. Cross-view and multi-view gait recognitions based on view transformation model using multi-layer perceptron. *Pattern Recognit Lett* 2012;33:882-9.
51. Chen X, Yang T, Xu J. Cross-view gait recognition based on human walking trajectory. *J Vis Commun Image Represent* 2014;25:1842-55.
52. Liu N, Tan YP, editors. View invariant gait recognition. In: IEEE International Conference on Acoustics, Speech and Signal Processing; 2010.
53. Bashir K, Xiang T, Gong S, editors. Cross View Gait Recognition Using Correlation Strength. *BMVC*; 2010.
54. Hu M, Wang Y, Zhang Z, Little JJ, Huang D. View-invariant discriminative projection for multi-view gait-based human identification. *IEEE Trans Inf Forensics Secur* 2013;8:2034-45.
55. Hu H. Enhanced Gabor feature based classification using a regularized locally tensor discriminant model for multiview gait recognition. *IEEE Transactions on Circuits and Systems Video Technol* 2013;23:1274-86.
56. Muramatsu D, Shiraishi A, Makihara Y, Uddin MZ, Yagi Y. Gait-based person recognition using arbitrary view transformation model. *IEEE Trans Image Process* 2015;24:140-54.
57. Muramatsu D, Shiraishi A, Makihara Y, Yagi Y, editors. Arbitrary view transformation model for gait person authentication. In: 2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS). IEEE; 2012.
58. Wu Z, Huang Y, Wang L, Wang X, Tan T. A Comprehensive Study on Cross-View Gait Based Human Identification with Deep CNNs. *IEEE Trans Pattern Anal Mach Intell* 2017;39:209-26.
59. Tian Y, Wei L, Lu S, Huang T. Free-view gait recognition. *PLoS One* 2019;14:e0214389.
60. Yu S, Tan D, Tan T, editors. A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition. 18th International Conference on Pattern Recognition. IEEE; 2006.
61. Piccardi M, editor. Background Subtraction Techniques: A Review. *IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat No 04CH37583)*. IEEE; 2004.
62. Chtourou I, Fendri E, Hammami M. Walking direction estimation for gait based applications. *Procedia Comput Sci* 2018;126:759-67.
63. Aharon M, Elad M, Bruckstein A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans Signal Process* 2006;54:4311-22.
64. Elad M. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Berlin: Springer Science & Business Media; 2010.
65. Rubinstein R, Zibulevsky M, Elad M. Efficient Implementation of the K-SVD Algorithm and the Batch-OMP Method. Department of Computer Science, Technion, Israel, Technology Report; 2008.
66. Phillips PJ, Grother PJ, Micheals RJ, Blackburn DM, Tabassi E, Bone M. *Face Recognition Vendor Test 2002: Evaluation Report*; 2003.

BIOGRAPHIES



Abbas Ghebleh is a PhD student of Software Engineering at Shahid Beheshti University, Tehran, Iran. He received his B.Sc. and M. Sc. in Software Engineering from Shahid Beheshti University, Tehran, Iran (2006 and 2010). His research interests are Digital Image Processing and Computer Vision including human gait recognition and

action recognition.

Email: a_ghebleh@sbu.ac.ir



Mohsen Ebrahimi Moghaddam is a professor in computer engineering and science department, Shahid Beheshti university of Iran since 2006. He has got his Ph.D., M.Sc., and B.Sc. from Sharif University in Iran. His research interests are image processing and pattern recognition specially using artificial intelligence techniques such as image security, watermarking, deblurring, and biometrics. He is the image processing lab chief in his department.

Email: m_moghaddam@sbu.ac.ir
