Repositório ISCTE-IUL

View-Invariant Gait Recognition exploiting Spatio-Temporal Information and a Dissimilarity Metric

Tanmay Tulsidas Verlekar
Instituto de Telecomunicações,
Instituto Superior Técnico, Universidade
de Lisboa,
Lisboa, Portugal.
ttv@lx.it.pt.

Paulo Lobato Correia
Instituto de Telecomunicações,
Instituto Superior Técnico, Universidade
de Lisboa,
Lisboa, Portugal.
plc@lx.it.pt.

Luís Ducla Soares
Instituto de Telecomunicações,
Instituto Universitário de Lisboa
(ISCTE-IUL),
Lisboa, Portugal.
lds@lx.it.pt.

Abstract— **In gait recognition, when subjects do not follow a known walking trajectory, the comparison against a database may be rendered impossible. Some proposed solutions rely on learning and mapping the appearance of silhouettes along various views, with some limitations caused for instance by appearance changes (e.g. coats or bags). The present paper discusses this problem and proposes a novel solution for automatic viewing angle identification, using minimal information computed from the walking person silhouettes, while being robust against appearance changes. The proposed method is more efficient and provides improved results when compared to the available alternatives. Moreover, unlike most state-of-the-art methods, it does not require a training stage. The paper also discusses the use of a dissimilarity metric for the recognition stage. Dissimilarity metrics have shown interesting results in several recognition systems. This paper also attests the strength of a dissimilarity-based approach for gait recognition.**

Keywords— *View Invariance, Gait Recognition, Dissimilarity Space, Biometrics.*

I. INTRODUCTION

Biometric recognition can be used for a wide range of applications, from user authentication in controlled conditions, such as the access to a building or the login into a computer or smartphone, to more unconstrained scenarios with challenging recognition conditions, as the identification of a walking person from a surveillance camera footage.

In the particular case of surveillance scenarios, the user does not actively cooperate in the recognition procedure, the distance to the camera can be large and the image acquisition angle is typically unknown. This prevents traditional biometric traits, such as face, iris or fingerprint, from being used with acceptable results. Here, however, recognition based on gait (i.e., the way a person walks) has shown promising results.

Although a lot of work has already been done on gait recognition, most of it applies to fixed viewing angle image acquisition, typically when the walking direction is perpendicular to the camera axis [1], [2], [3], [4]. As the viewing angle changes, the problem of view-independent gait

recognition is raised, since the features that can be computed for recognition purposes also change. Combined with the fact that data collected at different moments from the same person can also exhibit significant differences, e.g., due to changes in clothing or carried items, view-independent gait recognition can hardly be considered a fully solved problem [5].

In the literature, various methods have been proposed to tackle the problem of camera viewpoint change when attempting gait recognition. These methods can be broadly classified into two groups: (i) model-based, and (ii) appearance-based methods [6].

The model-based approaches consist of methods that either use 3D models or try to model view-invariant gait features to be used for recognition purposes. Examples that use 3D models include [7], where parameters such as feet coordinates are used to set the model view and synthesize a new image for matching. In [8], a 3D model gait sequence is used for training and, therefore, 2D images from any arbitrary view can be re-projected using the associated projection matrices, which makes it possible to generate training data associated with any pair of target views that are to be matched. The methods described in [9] and [10] use multiple cameras to fit 3D models to the obtained sequences. Features like the trajectories of key joints are then used for user recognition. Requiring multiple cameras along the walking path, additional information about the external and internal camera parameters, or the position of the floor, make these systems effective only in controlled environments.

Other model-based methods, such as the one in [11], use the hip, knee and ankle positions to compute a set of anatomy-based ratios, and project the relevant points onto the lateral plane, for performing gait recognition. The method proposed in [12] computes the head and feet trajectories, converting them into a fronto-parallel view for matching. These methods are applicable to a limited set of views and are not robust to the typical occlusions caused by clothing and/or carried items. Also, being dependent on the detection of fixed points/ratios over a person's silhouette, any artifacts caused, for instance by shadows or self-occlusion, may lead to significant model errors.

Appearance-based methods, on the other hand, use spatio-temporal features computed directly from the observed sequences for recognition purposes. Most of these methods address view-invariant gait recognition by learning and mapping changes in appearance caused by changes in viewing angle. These methods are affected by the quality and the number of samples available for training. They are also adversely affected by changes in appearance caused by clothing and/or carried items, with some of them being applicable only over a limited range of views.

Examples of these techniques include the works presented in [13] and [14], where recognition is performed by exploring inter- and intra-user correlation along various views using principal component analysis (PCA) and linear discriminant analysis (LDA). In [15] and [16], a view transformation model (VTM) is constructed from the GEIs and recognition is performed by singular value decomposition of the VTM. In [17], a gait texture image (GTI) from any viewpoint is transformed into a lateral view, using the invariant low-rank texture transform (TILT). The obtained transformation matrix is used to alter the silhouettes into lateral view and then match them against the database by analyzing their contours.

To make appearance-based methods robust to changes caused by clothing or item carrying conditions, some methods split the recognition process into two steps. The first step detects the viewing angle, and the second step performs user recognition along the detected view. One such example is described in [5], where the leg region of gait energy images (GEI) is cropped and its entropy is computed and compared against the database to identify the view. A Gaussian filter is then applied to the GEI at different scales, generating a multiscale gait image, explored for recognition using random subspace learning (RSL). A similar approach is used in [18], with a Gaussian process classifier being applied to the cropped GEI for viewing angle identification, followed by canonical correlation analysis (CCA) for user recognition. The work in [6] obtained improved viewing angle identification results by computing a perceptual hash over the leg region of the user. In [19], viewing angle identification is performed based on the positions and heights of the person at the beginning and end of a gait cycle. This is followed by contour analysis for user recognition.

Although this group of methods is effective and robust against appearance change, it only works for views for which the systems are trained for. Hence, insufficient training data may hamper the performance of such methods.

This paper proposes a system that performs viewing angle identification that, unlike many existing methods, is not computationally expensive, is not limited in the range of walking directions considered and can also handle appearance changes. Additionally, the paper also proposes a dissimilarity-based approach [20], which has shown great promise in pattern recognition literature, and is here applied for gait recognition.

The novel viewing angle identification method proposed here first builds a GTI, and estimates the viewing angle by analysing the spatio-temporal evolution of the user's feet position. The proposed method does not require any previous training and is robust to appearance changes resulting from clothing variations or items being carried. Moreover, the method is computationally efficient and provides improved viewing angle identification results when compared to the state-of-the-art.

To perform user recognition, a dissimilarity-based approach is proposed. The method computes a feature vector of dissimilarity values based on the distances of each user's GEI to the selected prototypes. PCA is then applied to the dissimilarity vectors to reduce the problem of dimensionality, followed by LDA to get the features to be used for performing user recognition. Combined with the proposed viewing angle identification solution, the resulting view-invariant gait recognition system outperforms the state-of-the-art.

The remainder of the paper is organized as follows. The proposed system is presented in section II, with the corresponding experimental results being reported in section III. Finally, section IV provides some conclusions and directions for future work.

II. Ease Proposed View-Invariant Gait Recognition System

This paper proposes an appearance-based view-invariant gait recognition solution, composed of three main modules: (i) gait representation; (ii) viewing angle identification and (iii) user recognition, as illustrated in Fig. 1.
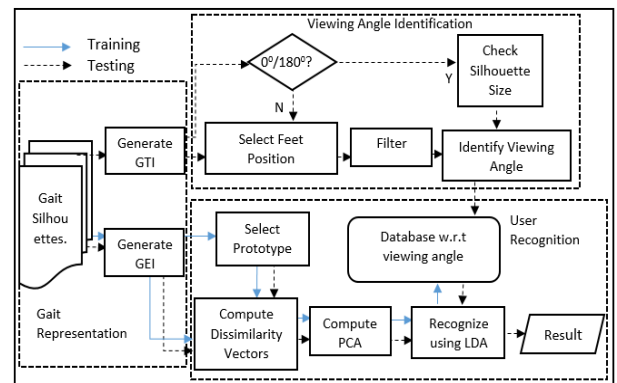

Fig. 1. *Proposed gait recognition system.*

The first module uses the silhouettes of the walking person to compute both the gait energy image (GEI) and the gait texture image (GTI). The second module consists of the novel viewing angle identification method being proposed. The third module performs the user recognition itself using a dissimilarity-based approach. A more detailed explanation of each of these modules is included in the following sections.

A. Gait Representation

The considered solution relies on two different gait information representations, as illustrated in Fig. 1, to perform user recognition. The first representation, denoted as GTI in [17], is here used for viewing angle identification, as it captures the spatio-temporal information associated with the person movement. The second is the well-known GEI, which is used for the actual user recognition. Both representations are computed from walking person's silhouettes.

Given a sequence of images containing silhouettes, the GTI is computed according to (1):

$$GTI(x, y) = \sum_{t=1}^{T} I(x, y, t)/T \qquad (1)$$

where $I(x,y,t)$ is the binary silhouette's pixel value considered for position (x,y) of the image, at time t. Notice that instead of considering only the silhouettes captured during one gait cycle, as done in [17], here T is the total number of available images. In fact, the goal of the proposed method is to capture the motion direction of a given person, and obtaining a larger set of feet positions to improve the desired viewing angle identification results.
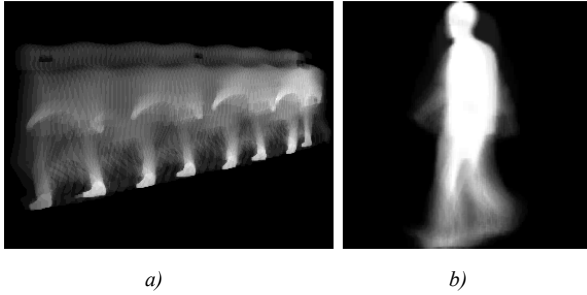
The GEI is generated as specified in [15]. The process can be described by (2), now providing as input, $Ic(x,y,t)$, a set of cropped silhouettes previously normalized in size to 240x240 pixels and conveniently aligned.

$$GEI(x, y) = \sum_{t=1}^{N} Ic(x, y, t)/N \qquad (2)$$

In this case, the average is performed over a complete gait cycle, assumed of duration N.

B. Viewing Angle Identification

Most of the appearance-based methods that are robust to appearance changes follow the general framework described in [18], [5] and [6], where only the bottom third of the GEI, corresponding to the leg region, is used since typically it is unaffected by variations in clothing and/or carrying conditions. The leg region is used for the subsequent analysis, keeping most information associated with the direction of walking, as illustrated in Fig. 3. However, as discussed in the introduction, these methods are sensitive to the GEI quality and the samples available for the training stage. In case the silhouettes' leg region is covered, e.g., by a long skirt, then the walking direction information cannot be accurately estimated. The proposed method overcomes these drawbacks, by relying only on the feet position information. Consequently, the proposed method can tackle a wider range of appearance changes than most state-of-the-art solutions.

The architecture of the proposed method, presented in Fig. 1, relies on a spatio-temporal analysis, using the GTI, to identify the feet positions. These positions correspond to the bottom-most non-zero values of the GTI contour, as illustrated

with a "+" in Fig. 4. The left- and right-most sides of the GTI do not correspond to feet positions, but rather to leg and waist regions. A filter is thus applied to remove these areas from the list of feet positions, simply by thresholding the difference between two consecutive positions, as that value increases for the tails of the sequence of acquired coordinates. The value of consecutive differences is represented by a "." in Fig. 4 b), and the horizontal line represents the threshold separating the feet position from the rest of the contour coordinates, which is empirically set to best separate the feet positions.
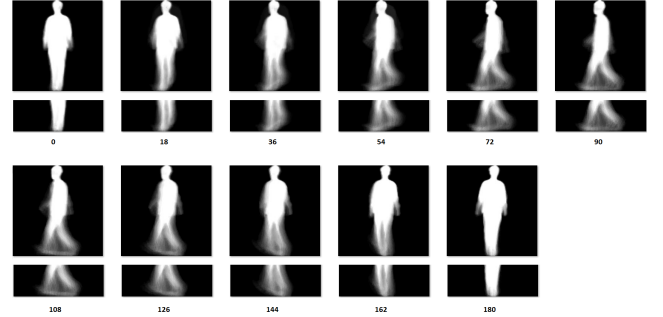
To detect the viewing angle, the walking direction is identified by applying PCA to the filtered feet positions. PCA transforms the feet positions onto a new coordinate system, such that the first principal component captures the largest coordinate's variance, and therefore its direction represents the dominant walking trajectory, from which the corresponding view angle can be estimated.

There are two cases for which the method presented above cannot be used. In fact, when the user walks towards the camera, or away from it, all the silhouettes will be overlapped in the GTI, not allowing to trace the feet positions - see example in Fig. 5 b) and d). The proposed method detects such situations by analyzing the summation of intensities along the y-axis, and checking if a single peak appears. For all other views, multiple peaks will be identified, as can be seen in Fig. 5 a) and c). Finally, the distinction between a user walking towards or away from the camera is simply done by comparing the sizes of the first and last silhouettes of the sequence.

Fig. 5. Distinguishing 0º or 180º from other views: a) and b) represent GTIs; c) and d) represent the vertical sums of intensities in the GTIs; b) and d) correspond to a 0º or 180º view, while a) and c) corresponds to other viewing angles.

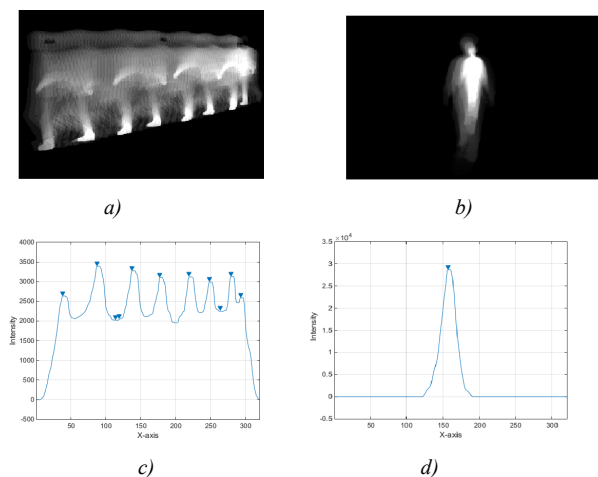## C. User Recognition

To perform user recognition, the use of a dissimilarity metric is explored, as illustrated in Fig. 1. To generate the dissimilarity measurements, a set of GEIs has first to be selected as prototypes. The prototypes can correspond to the entire database, to a random selection from the database, or to a set of carefully selected samples from the database [20]. Prototype selection is an important step, being the equivalent of defining good features in a feature-based classification strategy. Once GEI prototypes are selected, dissimilarity vectors can be created by computing the pairwise Euclidean distances between the user's GEI and each of the prototypes. Therefore, the dissimilarity representation can be viewed as a mapping from the initial GEI representation to the dissimilarity space defined by the prototypes. In the new representation, each element of a dissimilarity vector expresses the degree of difference to each individual prototype.

For the proposed implementation, the available training sequences are split into two groups. The first group is used to generate the prototypes, while the second group is used to create the dissimilarity vectors representing each of the users enrolled in the database, by computing the distance to each of the prototypes. The test sequences are also represented by dissimilarity vectors, again by computing the distances to the same prototypes, and are compared to the database for performing user recognition.

One commonly adopted classification strategy uses a K-Nearest Neighbour (K-NN) classifier. K-NN performs classification by computing a distance metric, such as the Euclidean distance, between the testing and the database dissimilarity vectors, to assign the test sequences to the class with the smallest distance.

This paper adopts another low complexity classification strategy, which achieves better results than K-NN, as depicted in Fig. 1. The proposed method applies PCA to obtain Eigen vectors using the training dissimilarity vectors. Eigen vectors

with the largest variance are retained and the data is projected onto it for dimensionality reduction. Then, LDA is employed to identify a projection matrix that maximizes inter- and intra-class scatter ratio using Fisher's criterion. Recognition is then performed using the Euclidean distance, where the test sequences are classified into the class with the smallest distance.

III. EXPERIMENTAL RESULTS

The proposed recognition system is tested on dataset B of the CASIA gait database, which gathers walking sequences with a large variation with respect to viewing angles. It was collected by the Institute of Automation of the Chinese Academy of Sciences [21]. This database contains gait silhouettes of 124 users, captured for 11 different walking directions, with angles of acquisition from 0º to 180º, with a step of 18º between adjacent views. For each angle, there are 10 sequences per user, out of which 6 are normal walking sequences (N), 2 wearing a large coat (C) and 2 carrying a bag (B).

A. Viewing Angle Identification Results

The proposed method requires no training and is tested using the entire database, to also evaluate the algorithm's robustness to appearance changes. Viewing angle identification evaluation is performed by quantizing the estimated direction to nearest multiple of 18º, as considered in the ground-truth.

Results are compared against those of state-of-the-art methods that are also robust to appearance changes, with the results being summarized in TABLE I. The methods reported in [5] and [6] use the first four normal sequences for training and the rest are used for testing. The method described in [18] also uses the first 4 normal sequences for training, but only from 60% of the subjects in the dataset; the remaining 40% are used for testing over a subset of the available viewing angles (36º to 144º). Since the proposed method does not need training, it uses 6 normal, 2 bag and 2 coat sequences for testing. The obtained results show the superior performance of the proposed method, evidencing its robustness against appearance changes.

TABLE I. CORRECT VIEWING ANGLE IDENTIFICATION RATE (%).

View(º)	GP Classifier [18]			Entropy Method [5]			PHash Method [6]			Proposed Method		
	N	C	B	N	C	B	N	C	B	N	C	B
0	-	-	-	83	80	79	91	86	86	98	96	98
18	-	-	-	94	87	85	90	81	76	99	99	99
36	84	83	84	88	85	80	70	66	56	98	98	98
54	91	88	91	92	90	89	92	87	72	99	99	98
72	85	84	85	81	80	78	96	90	88	98	99	98
90	74	68	74	89	79	72	95	88	86	97	97	97
108	86	83	86	79	75	70	94	92	87	95	94	95
126	91	92	91	90	88	85	96	92	91	98	99	99
144	93	93	93	83	81	79	95	91	83	96	96	96
162	-	-	-	89	86	84	95	89	88	97	96	96
180	-	-	-	82	80	75	95	90	88	99	97	99
Mean	86	84	86	86	82	78	92	87	82	97	97	97

B. User Recognition

Once the viewing angle is detected, user recognition can be performed. As mentioned in section 2, in this paper user recognition is performed using a dissimilarity space. Given a large database, the prototype selection would be done by selecting a random subset. However, since the current database is small, a representative from each class is included as prototype, to get a sufficiently descriptive set; in this example, the first normal walking sequence of each user is taken as a prototype. Then, the next three normal walking sequences of each registered user are considered for computation of the dissimilarity vectors to be used for registration of the user, and the remaining two normal sequences are used for testing. The test sequences are also represented in the dissimilarity space, i.e., using the dissimilarity vectors obtained by comparison with the prototypes, to perform user recognition.

Even though the viewing angle identification algorithm is robust to appearance changes, like the usage of a coat or carrying of items, as discussed in the previous section, the user recognition tests only consider the normal walking sequences. Extending the recognition module to handle appearance changes is left as future work.

Here, the goal is to understand how using the dissimilarity space for user recognition may affect the recognition results in comparison to state-of-the-art methods that work under similar conditions. Thus, results are compared against those reported in the following papers: (i) in [16] 22 user sequences are considered to construct a view transformation model; it then splits the remaining sequences into two halves, the first three are used for training and the rest for testing; (ii) the multiscale method in [5] uses four normal sequences for training and the remaining two for testing.

Gait recognition results are included in TABLE II. It can be seen that the dissimilarity space using PCA followed by LDA provides results comparable to the best reported by state-of-the-art methods. It is worth noting that a correct viewing angle identification rate of 97% leads to a correct user recognition rate of 98.6%, meaning that even for some cases where the view direction was incorrectly detected, it was labelled with an adjacent direction that still allows performing a correct user recognition, as the proposed method is robust to slight changes in viewing angles. To fully understand the effectiveness of the user recognition method using dissimilarity space, a second set of observations are made where the test sequences are 'ideally' sorted with respect to view. It can be seen that, in this setting, the proposed method provides even better results as shown in TABLE II under the column labelled "Ideal View Case". Notice that results of the multiscale method are provided rounded up to the nearest integer value in [5].

It should also be noted that an additional advantage of the proposed method lies in its simplicity. The method in [5] although providing slightly better results, when compared to the proposal for user recognition, is computationally more expensive. The method constructs a multiscale gait image by applying 5 different Gaussian filter scales to the GEI and uses subspace learning to recognize users, making it more complex than the proposed dissimilarity vector computation. The RSL method in [5] randomly selects 300 PCA components to construct a subspace and uses LDA for user recognition, repeating the process ten times with different subspaces. The final classification is then done by a maximum voting policy. In comparison, the proposed method selects a limited number of principal components, that covers 95% of the total variation in the data, and through a single application of LDA provides comparable results, thus making the system computationally more efficient.

TABLE II.   CORRECT USER RECOGNITION RATE (%).

View(º)	Proposed Dissimilarity Space Method			ARA Method [16]	Multiscale Method [5]
	View identification+ K-NN	*View identification+ PCA+LDA*	*Ideal View Case + PCA+LDA*		
0	89.5	99.6	100.0	98.1	100
18	91.5	99.6	100.0	97.6	99
36	86.1	99.2	100.0	97.3	100
54	85.6	98.8	99.2	98.1	99
72	82.1	99.2	100.0	98.4	100
90	80.2	98.0	99.6	94.9	100
108	80.9	97.1	100.0	95.1	99
126	80.5	99.6	100.0	98.4	99
144	81.9	96.7	99.2	94.9	100
162	86.5	97.1	99.6	96.0	100
180	93.5	100.0	100.0	95.1	99
Mean	85.3	98.6	99.8	96.7	99.5

IV. CONCLUSION

The paper presents a new method to perform user recognition independent of the viewing angle. It does this by splitting the problem into two steps. The first step involves viewing angle identification. This step is crucial for most appearance-based methods. Although most methods are robust to small changes in viewing angle, in situations where appearance changes are caused due to coats or bags, this step gains higher significance, as the method is able to cope with both changes in appearance and in viewing angle. The proposed work uses the spatio-temporal evolution of feet position, extracted from the GTI, to detect the viewing angle. The results obtained are better than those achieved by other state-of-the-art methods. The second step involves user recognition, where the user is matched against the database with respect to the detected viewing angle. The proposed work explores a dissimilarity space, using PCA followed by LDA, to perform recognition providing results equivalent to the state-of-the-art methods.

Future work includes exploring the dissimilarity metric in situations where appearance is modified due to bags or coats. It will result in a complete system that can be implemented in a surveillance environment. Further improving the viewing angle identification method will also lead to better recognition results as shown in TABLE II.

REFERENCES

[1] J. Fazenda, D. Santos, P. Correia, "Using gait to recognize people," In The International Conference on Computer as a Tool, EUROCON, Belgrade, 2005.

[2] N. Liu, Y. Tan, "View invariant gait recognition," In IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), Dallas, TX, pp. 1410-1413, 2010.

[3] Sruti, D. Choudhury, "Silhouette-based gait recognition using procrustes shape analysis and elliptic fourier descriptors," Pattern Recognition, , 9th ed., vol. 45, pp. 3414-3426, 2012.

[4] Y. Iwashita, K. Uchino, R. Kurazume, "Gait-based person identification robust to changes in appearance," Sensors - Open Access Journal, 6th ed., vol. 13, pp. 7884-7901, 2013.

[5] Sruti, D. Choudhury, T. Tardi, "Robust view invariant multiscale gait recognition," Pattern Recognition, 3rd ed., vol. 48, pp. 798-811, 2015.

[6] T. Verlekar, P. Correia, "Walking direction identification using perceptual hashing," In International Workshop on Biometrics and Forensics - IWBF, Limassol, Cyprus, 2016.

[7] Y. Iwashita, R. Kurazume, R. Baba, K. Ogawara, "Method for gait-based biometric identification robust to changes in observation angle," In 26th International Conference of Image and Vision Computing, New Zealand, Auckland, 2011.

[8] D. Muramatsu, Y. Makihara, Y. Yagi, "Cross-view gait recognition by fusion of multiple transformation consistency measures," IET Biometrics, 2nd ed., vol. 4, pp. 62-73, 2015.

[9] G. Zhao, G. Liu, H. Li, M. Pietikainen, "3D gait recognition using multiple cameras. in ieee 7th international conference on automatic face and gesture recognition", Southampton, pp. 529 - 534, 2006.

[10] T. Krzeszowski, B. Kwolek, A. Michalczuk, "View independent human gait recognition using markerless 3d human motion capture," In Computer Vision and Graphics, Springer. Berlin Heidelberg, 2012.

[11] M. Goffredo, I. Bouchrika, J. Carter, M. Nixon, "Self-calibrating view-invariant gait biometrics," IEEE Transactions on Systems, Man, and Cybernetics, 4th ed., vol. 40, pp. 997-1008, 2010.

[12] F. Jeana, A. Albub, R. Bergevina, "Towards view-invariant gait modeling: computing view-normalized body," Pattern Recognition, 11th ed., vol. 42, pp. 2936-2949, 2009.

[13] N. Liu, Y. Tan, "View invariant gait recognition," In IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), Dallas, TX, pp. 1410-1413, 2010

[14] L. Tan, "View recognition of human gait sequences in videos," In 17th IEEE International Conference on Image Processing (ICIP), 2010.

[15] W. Kusakunniran, Q. Wu, H. Li, J. Zhang, "Multiple views gait recognition using view transformation model based on optimized gait energy image", In IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops), Kyoto, pp. 1058-1064, 2009.

[16] H. Chaubey, H. Madasu, S. Vasikarla, "Enhanced view invariant gait recognition using feature level fusion", In Applied Imagery Pattern Recognition Workshop (AIPR), Washington, DC, pp. 1-5, 2014.

[17] W. Kusakunniran; Q. Wu, J. Zhang, "A new view-invariant feature for cross-view gait recognition," IEEE Transactions on Information Forensics and Security, 10th ed., vol. 8, pp. 1642-1653, 2013.

[18] B. Khalid, X. Tao, G. Shaogang, "Cross view gait recognition using correlation strength," In British Machine Vision Conference, London, 2010.

[19] S. Jia, L. Wang, X. Li, "View-invariant gait authentication based on silhouette contours analysis and view estimation," Journal of Automatica Sinica, IEEE/CAA, 2nd ed., vol. 2, pp. 226 - 232, 2015.

[20] R. Duin, E. Pękalska, "The dissimilarity space: Bridging structural and statistical pattern recognition," Pattern Recognition Letters, 7th ed., vol. 33, pp. 826 - 832, 2012.

[21] CASIA Gait Database. http://www.sinobiometrics.com. January 2005.