Gait Recognition Using Normalized Shadows

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

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

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

Abstract— **Surveillance of public spaces is often conducted with the help of cameras placed at elevated positions. Recently, drones with high resolution cameras have made it possible to perform overhead surveillance of critical spaces. However, images obtained in these conditions may not contain enough body features to allow conventional biometric recognition. This paper introduces a novel gait recognition system which uses the shadows cast by users, when available. It includes two main contributions: (i) a method for shadow segmentation, which analyzes the orientation of the silhouette contour to identify the feet position along time, in order to separate the body and shadow silhouettes connected at such positions; (ii) a method that normalizes the segmented shadow silhouettes, by applying a transformation derived from optimizing the low rank textures of a gait texture image, to compensate for changes in view and shadow orientation. The normalized shadow silhouettes can then undergo a gait recognition algorithm, which in this paper relies on the computation of a gait energy image, combined with linear discriminant analysis for user recognition. The proposed system outperforms the available state-of-the-art, being robust to changes in acquisition viewpoints.**

Keywords—Shadow Biometrics; Gait Recognition.

I. INTRODUCTION

Biometric traits such as fingerprint, iris, etc. are widely used in recognition systems, as they are unique to a user. However, they do not perform well in surveillance environments, as they require active user cooperation. A more limited set of biometric traits can be used to perform recognition over a distance and without user cooperation. A popular example among them is gait [1]. In the literature, gait recognition is performed following either a model-based, or an appearance-based approach.

Most model-based methods try to fit a model to the body or observed user motion. Examples include the construction of 3D models using information from multiple 2D video cameras [2] or from sensors capturing 2D video and depth information [3]. Other methods model the user's motion using key points such as the hip, knee and ankle positions [4], [5] or the head and feet positions [6]. These recognition methods rely on the use of additional information such as depth, external and internal camera parameters, floor position, etc. They also rely on an accurate detection of key points, which is not always possible, e.g., due to occlusions. These limitations often restrict the usage of model-based methods to controlled environments.

Appearance-based methods perform user recognition using spatiotemporal information obtained directly from observed gait sequences, thus being better suited for recognition in surveillance environments. Examples include methods that use gait representations, such as gait energy image (GEI) [7], Radon transform-based energy image [8] or feature vectors obtained by applying singular value decomposition to GEIs [9] [10]. If the viewing point changes, these methods either compute transformations by optimizing low rank textures of a gait texture image (GTI) to correct the view [11], or split the recognition process into two steps. The first step involves view identification by learning the leg region of a GEI [12], gait entropy image (GEnI) [13], perceptual hash of the GEI leg region [14] or feet positions in a GTI [15]. The second step applies user recognition for the identified view.

Gait-based recognition is effective in environments where the camera captures the entire body of the user, while it becomes harder or even impossible when the surveillance camera is carried by a drone or placed at a high position. For overhead cameras, body self-occlusions hide gait features, seriously hampering the performance of recognition systems. In such scenarios, however, there are often shadows cast from the user, which can provide an alternative for gait recognition, as Fig.1 illustrates. The shadow cast on the ground under a focused source of illumination, such as the sun (on a clear day) or a street lamp, also depicts the appearance of a user. Gait features can thus be obtained from the shadow, thus allowing successful user recognition. The use of shadows to obtain features to perform gait recognition was first suggested by Stoica, in 2008 [16].
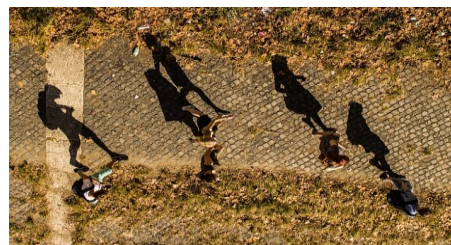

Fig. 1. Example of image captured by overhead camera (from *pixabay.com)*.

The work by Iwashita and Stoica [17] performs user recognition using features called gait stripes, obtained from shadow silhouettes under the presence of the sun. The method, however, implies the manual separation and normalization of shadow silhouettes, and considers a fixed camera view and a fixed time of the day. Further work by Iwashita et al. [18] introduces an automatic shadow segmentation method,

applying principle component analysis (PCA) to a GTI and computing the sum of intensities along the first principal component direction. The point of separation between the body and the shadow silhouettes corresponds to the highest sum value. They also proposed a method for shadow silhouette normalization [19], which requires manually setting a large number of parameters, but for producing results they instead used a fixed view and timestamps, to avoid the need for normalization. The same team used shadow contours as a feature for recognition [20], and further improved recognition results by combining contour and gait stripes features [21].

User recognition using shadows can also be performed in indoor environments as long as a focused source of illumination is used. The work of Iwashita et al. [22] uses two infrared illumination sources to cast two shadows, perpendicular to each other, captured by an overhead camera. A GEI, without any shadow separation, is used for recognition employing affine moment invariant features. The method is further improved, making it robust to appearance changes, by using a weighting technique [23]. Robustness to changes in view is addressed using a 3D model to synthesize the shadow of a user [24]. These methods use a fixed illumination source and capture shadows from a fixed viewpoint, requiring information about the illumination source position, thus limiting their use, e.g., in the presence of the sun.

The work discussed above introduced the use of shadows for gait recognition, but faces several limitations especially for outdoor operation, in the presence of an illumination source, such as the sun. Most of these methods either use entire silhouettes or separate the shadow silhouettes manually. When shadow separation is performed automatically [18], the results may not be accurate in some circumstances, as discussed in section II.A. Since the appearance of the shadow depends on the position of the sun and the camera, the shadow silhouettes should be normalized before attempting recognition. The orientation of the shadow is altered based on the position of the sun. However, a simple rotation of the shadow silhouette does not provide acceptable results, as shadows also suffer a perspective change caused by the varying position of the camera (i.e., change in viewpoint). Thus, to normalize the shadow silhouette, a homographic transformation is needed which rectifies the perspective and orientation of the shadow silhouette before considering it for user recognition. The methods considered in the literature typically require manually setting a large number of parameters to rectify the changes due to different camera viewpoints and shadow orientations.

This paper presents a system with two main contributions to address the above problems: (i) automatic shadow segmentation; and (ii) automatic shadow silhouette normalization. The system performs shadow segmentation by identifying the user's feet position, from the silhouette's contour orientation analysis. The feet positions obtained over the entire gait sequence are then used to identify a line that separates body and shadow silhouettes. Shadow silhouette normalization is then performed by applying a homographic transformation, obtained by optimizing the low rank textures of the shadow silhouettes' GTI. The optimization process transforms the GTI into a canonical view, providing the desired normalizing of the shadow silhouettes, correcting them

for orientation and viewpoint variations. For recognition purposes, GEIs are constructed from the shadow silhouettes, followed by linear discriminant analysis (LDA). The proposed system highlights the strength of gait recognition from shadows; also, it is robust to changes in viewpoint.

The rest of the paper is organized as follows. Section II presents the proposed system and section III discusses the experimental results. Section IV provides conclusions and suggests directions for future work.

II. THE PROPOSED SYSTEM

The proposed system for gait recognition using normalized shadows uses as input a set of binary foreground masks. They are obtained using robust principal component analysis [25] to identify the user and the shadow silhouettes as foreground. Automatic shadow segmentation is then applied to achieve the desired shadow separation, followed by shadow silhouette normalization.

A. Shadow Segmentation

Iwashita and Stoica [18], performed user body and shadow silhouette separation by summing the intensity values along the first principal component direction (aligned with the x-axis in Fig.2) of a GTI, and selecting the highest sum of intensities as the point of separation. That method assumes the highest overlap between body and shadow silhouettes to occur at the feet position. However, as illustrated in Fig. 2 (b), this is not always the case, notably due to changes in shadow orientation resulting from the camera and illumination source positions.
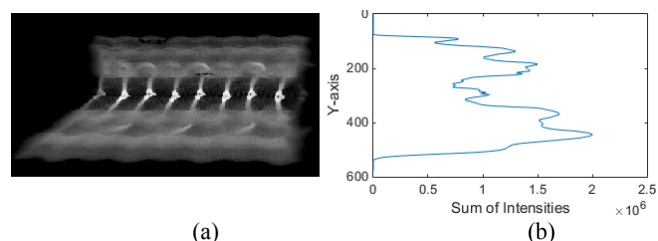

(a)   (b)

Fig. 2. *The shadow separation method in [18]: Example of a GTI (a) and sum of its intensity values along the first principal component (b).*

The proposed overcomes this limitation by splitting the shadow segmentation process into two steps. The first step identifies the feet position in individual foreground masks by analyzing the relative orientation of the shadow and the user body silhouettes – see example in Fig. 3(a). The second step then uses the feet positions obtained along the gait sequence to fit a line that separates the shadows and the user body silhouettes – see Fig. 3(b).

1) Individual feet position identification

Since the shadow silhouette is projected onto the ground, connecting with the body silhouette at the feet position, the main orientation of the body silhouette differs from that of the shadow silhouette (except when the illumination source is perfectly aligned with the user). Thus, the feet position can be identified as the foreground mask's contour point that exhibits a sudden orientation change. By considering only the lower part of the foreground mask contour, i.e., the bottom-most non-zero values of the contour image, as illustrated with blue "*" in

Fig. 3 (a), the proposed method selects the contour point with highest y coordinate value as the feet position. This assumption is true during the double support phase of the gait cycle. However, along the gait cycle, especially during mid stance, the arm or knee positions may in some cases be incorrectly identified as the desired feet position, as shown in Fig 3(a). To overcome this possible problem, the proposed method computes the difference between two consecutive y coordinate values among the selected part of the contour. This difference is represented by a red "." in lower part of Fig. 3 (a). The differences corresponding to the shadow part of the contour are typically low, while the difference corresponding to the knee or hand position of the contour is high and thus can be discarded from the selected part of the contour. Here, the threshold is empirically set to 50.
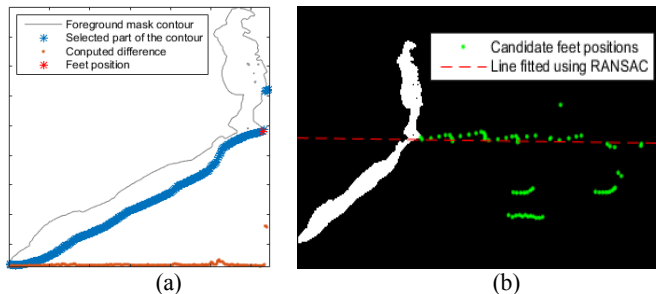

Fig. 3. *Proposed shadow separation method: Feet position selection (a), line fitting using RANSAC (b).*

2) Shadow separation

Once the feet positions for the entire gait sequence are obtained, the proposed method applies a line fitting algorithm to those feet positions obtained over the entire gait sequence. It then classifies the part of the silhouette below the line as the shadow silhouette.

Notice that for the cases when the illumination source is perfectly aligned with the user, there is no change in the orientation of the shadow with respect to the body of the user. In such cases, the previous step may incorrectly identify the feet position, selecting a random point along the contour. To make the proposed method robust against such errors, the proposed method uses the iterative random sample consensus (RANSAC) line fitting algorithm [26].

RANSAC assumes that the input contains outliers, i.e., feet positions which do not vote consistently for a single line fit, and rather looks for a sufficient number of values that "agree" on the same line. To perform shadow separation, RANSAC randomly selects two feet positions and considers the line that passes through them. It then identifies other feet positions consistent with the model, classifying the rest as outliers. It iterates through the available feet positions until a sufficient number of samples agree with a single model. The resulting line is used to separate the shadow from the body silhouettes. As seen in Fig. 3(b), RANSAC successfully fits a line even in the presence of uncertainties in the identification of feet positions, allowing a successful shadow separation.

B. Shadow Silhouette Normalization

The obtained shadow silhouettes typically contain significant changes in orientation and perspective along the

gait sequence, due to the relative position of the illumination source, causing the shadows, and that of the camera, with respect to the moving user – see Fig. 4. As a consequence, shadow silhouettes obtained at the start and at the end of the same gait sequence cannot be matched even though belonging to the same user. Therefore, normalization of the shadow silhouettes is needed for user recognition.
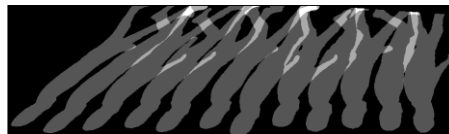

Fig. 4. *Shadow silhouettes at different instants for the same gait sequence.*

The shadow silhouette normalization method proposed in this paper uses a transformation obtained by optimizing the low rank textures of a GTI constructed from the shadow silhouettes. In this paper, the optimization of the low rank GTI texture is performed by applying the transform invariant low-rank textures (TILT) method [27]. TILT needs images containing regular symmetric patterns to obtain the desired transformation. This is the case for a gait cycle: silhouette shapes in the first half of the gait cycle roughly repeat in a symmetric way in the second half. Therefore, a GTI possesses the symmetry property expected from TILT inputs. When dealing with shadow silhouettes, the GTI can be obtained by vertically flipping the binary images containing the shadow silhouette and averaging all the resulting K binary images $I(x, y, t)$ belonging to the entire gait sequence, according to equation (1).

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

The model assumed by TILT is that a given GTI results from a transformation τ^{-1} applied to the sum of a low rank texture GTI^O with a sparse error E, according to (2).

$$GTI(x,y) = (GTI^O(x,y) + E).\tau^{-1} \qquad (2)$$

The representation in equation (2) can be used to construct the optimization problem described in equation (3).

$$\min_{GTI^O,E,\tau} rank(GTI^O) + \gamma\|E\|_0 , \ GTI.\tau = GTI^O + E \quad (3)$$

where $\|E\|_0$ denotes the number of nonzero entries in E. The weighting parameter γ is a tradeoff between the rank of the GTI and the sparsity of error E.

The aim of the optimization problem is to find the lowest possible rank of GTI^O, while at the same time having the fewest possible nonzero entries in E, that agree with the observed GTI up to the domain transformation τ. Since the optimization problem in (3) is nonconvex, TILT uses the convex relaxed form expressed in equation (4) and solves it via successive convex programming, as detailed in [26].

$$\min_{GTI^O,E,\tau} \|GTI^O\|_* + \lambda\|E\|_1 , \ GTI.\tau = GTI^O + E \quad (4)$$

where $\| \|_*$ denotes the nuclear norm, and $\| \|_1$ denotes the l^1-norm.

The transformation τ obtained represents the geometric projection to be applied to the GTI for normalization of the shadow silhouettes into a canonical view – see Fig. 5(b), (c).

Therefore, along with changing the orientation of the silhouettes, the domain transformation also rectifies the alterations caused by viewpoint changes.
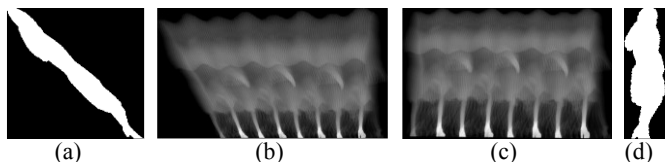

Fig. 5. *Proposed Shadow Silhouette Normalization method: Flipped shadow silhouette (a), GTI (b), GTI^o (c), Normalized shadow silhouette (d).*

C. User Recognition

The transformation of a shadow silhouette into the canonical view has some limitations. Notably, it cannot recover the self-occluded parts of the silhouette, with only the visible part being transformed into its canonical view.

To minimize the impact of having an incomplete shadow silhouette, this paper performs user recognition using features reflecting the dynamics of gait. For this, GEIs are obtained by averaging the N available cropped shadow silhouettes $Ic(x, y, t)$ belonging to a given gait cycle, according to (5).

$$GEI = \sum_{t=1}^{N} Ic(x, y, t)/N \qquad (5)$$

User recognition can then be performed by applying PCA for dimensionality reduction and data decorrelation, followed by LDA. LDA identifies a projection matrix $\emptyset$ onto a subspace that maximizes the ratio of intra- to inter-class scatter, using Fisher's criterion.

Given k classes with centroids $\bar{x}_k$ and a test GEI z, the system computes the Euclidean distance $d(,)$ in a transformed space, and selects the class with the lowest distance, according to equation (6).

$$arg \min_k d(z\emptyset, \bar{x}_k\emptyset) \qquad (6)$$

III. EXPERIMENTAL RESULTS

Since there are no publicly available gait shadow databases, the proposed system performance is tested using a database consisting of 12 users, created especially for this purpose. The database was captured outdoors, with users walking along a straight line, from point A to point B, as shown in Fig. 6 (a). For each user, 2 acquisition sessions were performed on different days, with a significant change in appearance in some cases, as illustrated in Fig. 6 (b). Acquisition was done between 12:30 and 14:30, acquiring 3 walking sequences per user in each session.

Fig. 6. Gait shadow sequences acqusition: acquisition location (a), example of appearance changes for the same user on different days (b).

Two tests were conducted to analyze the performance of the system in terms of recognition and robustness to viewpoint

changes. In these tests, two sequences of each session are used for training and one for testing, following the leave one out methodology. This gives a total of 4 training and 2 testing sequences for each user. For each sequence, two GEIs are computed, one from a complete gait cycle in the beginning and the other from a complete gait cycle at the end of the gait sequence. Each of these GEIs gathers silhouettes with distinct shadow orientations and viewpoint.

For the first test, complete gait sequences containing the two GEIs are used for training. Thus the system is trained for orientation and view changes, as illustrated in Fig. 4. Table I shows results obtained with the proposed method, without and with the proposed shadow normalization step. Also, results for the state-of-the-art gait stripes method are included; the method requires knowing the light source and user position for the suggested manual normalization step, absence of which significantly hampers its performance.

The first test is repeated for the proposed system by using the sequences obtained from one session for training and the other for testing. The correct recognition rate in this case reduces from 95% to 72%. However, this decrease in performance can be attributed to appearance changes as shown in Fig 6 (b), addressing which is beyond the scope of this work.

TABLE I. CORRECT RECOGNITION RATE

Methods		Correct Recognition Rate (%)
Gait Stripes [17] [18] [19]		70%
Proposed system	Non nornalized GEIs	86%
	Nornalized GEIs	95%

The second test highlights the importance of the proposed shadow normalization for robustness against viewpoint changes. Training is done using GEIs obtained from the first gait cycle, while testing is performed considering GEIs of the second gait cycle, i.e., this test considers significant changes in shadow orientations and viewpoint between the training and the testing sequences.

TABLE II. CORRECT RECOGNITION RATE WITH VIEWPOINT CHANGES

Methods		Correct Recognition Rate (%)
Gait Stripes [17] [18] [19]		53%
Proposed System	Non nornalized GEIs	31%
	Nornalized GEIs	89%

The results in Table II show that in face of shadow orientation and viewpoint changes, attempting recognition without shadow normalization is not a good option. In these conditions, also the gait stripes method performance is poor. Since the proposed normalization method transforms the shadow silhouettes to a canonical view, the proposed system becomes robust to such changes, highlighting the need for shadow silhouette normalization.

Another option for gait recognition would be to use the body silhouettes. However, with the acquisition setup used, with a complex background, the quality of obtained body silhouettes was poor. In fact, for most of the body silhouettes, large parts of the legs or torso were missing. Nevertheless, for the 6 users for whom relatively complete body silhouettes could be obtained, the 2 tests were repeated, now with GEIs

computed from the body, instead of the shadow, silhouettes. Recognition rates of 84% and 56%, respectively for the first and second tests were obtained. For the same 6 users, the correct recognition rate using the shadow silhouettes was 98% and 94%, respectively. These results show the usefulness of the proposed system in situations where good quality user body silhouettes cannot be obtained.

IV. CONCLUSION

The paper presents a novel system that performs gait recognition using shadows obtained in the presence of the sun. The method includes a contribution for shadow silhouette segmentation, separating it from the user body silhouette by identifying the feet positions along time. A line is then fitted to these positions using RANSAC, in order to be robust against feet position identification errors that are prone to occur in some parts of the gait cycle. The shadow silhouettes then undergoes normalization, by optimizing the low rank textures of a GTI. The proposal made allows correcting the shadow silhouette orientation, while making the system robust to changes in camera observation viewpoints. The GEIs obtained from the normalized shadow silhouettes can now be used for recognition purposes. The results obtained show that the proposed system outperforms the state-of-the-art.

The system is currently tested on a database of only 12 users. To better evaluate the system, a bigger database is needed. Thus, future work includes testing the system on a larger database, also including more covariate factors, such as the time of the day, changes in walking direction and change in appearance caused due to use of coat, bags, etc.

REFERENCES

[1] Y. Makihara, , D.S. Matovski, M.S. Nixon, J.N. Carter and Y. Yagi, "Gait recognition: Databases, representations, and applications," Wiley Encyclopedia of Electrical and Electronics Engineering, 2015.

[2] G. Zhao, G. Liu and H. Li and M. Pietikainen, "3D gait recognition using multiple cameras," in 7th International Conference on Automatic Face and Gesture Recognition (FGR06), 2006.

[3] Y. Wang, J. Sun, J. Li and D. Zaho, "Gait recognition based on 3D skeleton joints captured by kinect," in IEEE International Conference on Image Processing (ICIP), 2016.

[4] I. Bouchrika, J.N. Carter and M.S. Nixon, "Towards automated visual surveillance using gait for identity recognition and tracking across multiple non-intersecting cameras," Multimedia Tools and Applications, 2nd ed., vol. 75, pp. 1201-1221, 2016.

[5] M. Goffredo, I. Bouchrika, J.N. Carter and M.S. Nixon, "Self-Calibrating View-Invariant Gait Biometrics," IEEE Transactions on Systems, Man, and Cybernetics, 4th ed., vol. 40, pp. 997 - 1008, 2010.

[6] F. Jeana, A.B. Albub and R. Bergevina, "Towards View-Invariant Gait Modeling:Computing View-Normalized Body," Pattern Recognition, 11th ed., vol. 42, pp. 2936-2949, 2009.

[7] J. Man and B. Bir, "Individual recognition using gait energy image," IEEE transactions on pattern analysis and machine intelligence,2nd ed., vol. 28, pp. 316-322, 2006.

[8] N. Boulgouris, V. Zhiwei and X. Chi, "Gait recognition using Radon transform and linear discriminant analysis," IEEE transactions on image processing, 3rd ed., vol 16, pp. 731-740, 2007.

[9] N. Liu, J. Lu and Y.P. Tan, "Joint subspace learning for view invariant gait recognition," IEEE Signal Processing Letters, Dallas, 7th ed., vol. 18, pp. 431-434, 2011.

[10] K. Shirag, Y. Makihara, D. Muramatsu and T. Echigo and Y. Yagi, "GEINet: View-invariant gait recognition using a convolutional neural network," in International Conference on Biometrics (ICB), 2016.

[11] W. Kusakunniran, W. Wu, Q. Zhang, Y. Ma and H. Li, "A new view-invariant feature for cross-view gait recognition," Information Forensics and Security, IEEE Transactions, 10th ed., vol. 8, pp. 1642-1653, 2013.

[12] K. Bashir, T. Xiang, S. Gong, "Cross View Gait Recognition using Correlation strength," in British Machine Vision Conference, 2010.

[13] S. Choudhury and T. Tjahjadi, "Robust View Invariant Multiscale Gait Recognition," Pattern Recognition, 3rd ed., vol. 48, pp. 798-811, 2015.

[14] T. Verlekar and P. Correia. "Walking direction identification using perceptual hashing." in IEEE 4th International Workshop on Biometrics and Forensics (IWBF), 2016.

[15] T. Verlekar, P. Correia and L. Soares, "View-Invariant Gait Recognition Exploiting Spatio-Temporal Information and a Dissimilarity Metric," in Biometrics Special Interest Group (BIOSIG), 2016 International Conference of the IEEE, 2016.

[16] A. Stoica, "Towards Recognition of Humans and their Behaviors from Space and Airborne Platforms: Extracting the Information in the Dynamics of Human Shadows," in Bio-inspired, Learning and Intelligent Systems for Security, Edinburgh, 2008 .

[17] Y. Iwashita and A. Stoica, "Gait Recognition Using Shadow Analysis," in Symposium on Bio-inspired Learning and Intelligent Systems for Security, Edinburgh, 2009.

[18] Y. Iwashita, A. Stoica and R. Kurazume, "Person Identification using Shadow Analysis.," in British Machine Vision Conference, BMVC 2010.

[19] Y. Iwashita, A. Stoica and R. Kurazume, "People identification using shadow dynamics," in 17th IEEE International Conference on Image Processing (ICIP), Hong Kong, 2010.

[20] Y. Iwashita, A. Stoica and R. Kurazume, "Finding people by their shadows: Aerial surveillance using body biometrics extracted from ground video," in Third International Conference on Emerging Security Technologies (EST), 2012 .

[21] Y. Iwashita, A. Stoica and R. Kurazume, "Gait identification using shadow biometrics," Pattern Recognition Letters, 16th ed., vol. 33, pp. 2148–2155, 2012.

[22] Y. Iwashita, K. Uchino, R. Kurazume and A. Stoica, "Gait identification from invisible shadows," in Sensing Technologies for Global Health, Military Medicine, Disaster Response, and Environmental Monitoring II and Biometric Technology for Human Identification IX, Baltimore, Maryland, 2012.

[23] Y. Iwashita, R. Kurazume and A. Stoica, "Gait Identification Using Invisible Shadows: Robustness to Appearance Changes," in Emerging Security Technologies (EST), 2014 Fifth International Conference on, Alcala de Henares, 2014.

[24] M. Shinzaki, Y. Iwashita, R. Kurazume and K. Ogawara, "Gait-Based Person Identification Method Using Shadow Biometrics for Robustness to Changes in the Walking Direction," in Conference on Applications of Computer Vision (WACV), 2015 IEEE Winter , Waikoloa, HI, 2015.

[25] P. Rodriguez and B. Wohlberg, "Fast principal component pursuit via alternating minimization," in IEEE International Conference on Image Processing, 2013.

[26] M.A. Fischler, and R.C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," Communications of the ACM,6th ed., vol. 24, pp. 381-395, 1981.

[27] Z. Zhengdong, X. Liang, A. Ganesh and Y. Ma, "Tilt: Transform invariant low-rank textures," International journal of computer vision, 1st ed., vol. 99, pp. 1-24, 2012.