A Frontal View Gait Recognition Based on 3D Imaging Using a Time of Flight Camera

Tengku Afendi [1], F. Kurugollu[1], D. Crookes[1] and Ahmed Bouridane[2]

[1]The ECIT Institute,
Queen's University Belfast
Belfast, United Kingdom
{tzulcaffle01,f.kurugollu, d.crookes}@qub.ac.uk

[2]School of Computer Science,
Northumbria University,
Newcastle-upon-Tyne, United Kingdom
ahmed.bouridane@northhumbria.ac.uk

ABSTRACT

Studies have been carried out to recognize individuals from a frontal view using their gait patterns. In previous work, gait sequences were captured using either single or stereo RGB camera systems or the Kinect 1.0 camera system. In this research, we used a new frontal view gait recognition method using a laser based Time of Flight (ToF) camera. In addition to the new gait data set, other contributions include enhancement of the silhouette segmentation, gait cycle estimation and gait image representations. We propose four new gait image representations namely Gait Depth Energy Image (GDE), Partial GDE (PGDE), Discrete Cosine Transform GDE (DGDE) and Partial DGDE (PDGDE). The experimental results show that all the proposed gait image representations produce better accuracy than the previous methods. In addition, we have also developed Fusion GDEs (FGDEs) which achieve better overall accuracy and outperform the previous methods.

Index Terms— *Gait recognition, Gait data set, Time of Flight, Biometrics.*

1. INTRODUCTION

In recent years, biometrics research based on gait recognition has significantly increased. This is due to the nature of gait which is unique, unobtrusive, perceivable at a distance and only requires low camera resolution to accomplish the task [1]. The gait patterns of an individual can be captured either from the frontal or lateral views. Most previous work on gait recognition has been based on the lateral view because it provides much more temporal gait information. However, this approach requires a camera to be placed at a certain height and distance, so that it can capture enough information from the gait sequence. This imposes serious limitations on applications in environments such as narrow corridors. In order to alleviate these problems, frontal view based imaging can be considered for this kind of application. Furthermore, the frontal view gait patterns can be integrated with facial patterns to enhance biometric identification. Due to the limited gait information available in frontal view recognition, there are not many methods in the literature. Most attempts on frontal view gait recognition are based on a single RGB camera system. For example, Barnich and Droogenbroeck [2] proposed a gait feature based on a set of rectangles that will fit into any closed silhouette. However, the size of the rectangles will change if a subject wears bigger clothes or high heel shoes. Soriano et al. [3] and Balista et al. [4] applied Freeman Chain Code to the silhouette edge image obtained from the frontal view. The method depends on high precision of the silhouette segmentation which is very difficult to achieve in a complex background. The frontal view gait recognition algorithm in [5] employs the 3D gait volume by placing the edge points of the silhouettes in a 3D space. Silhouette alignment is obtained by stacking the normalized bounding boxes over time. The major drawbacks of this method are: the edge points and stacking methods are very much dependent on clothing, shoes, and carrying condition. Matovski et al. [6] applied Gait Energy Image (GEI) [7] and Gait Entropy Image (GEnI) [8] methods to frontal view gait recognition. Ryu and Kamata [9] proposed a frontal view gait recognition method using a stereo camera system which generates a human point cloud using Cartesian coordinates which are later converted into spherical coordinates. The main disadvantage of the single and stereo camera systems is their sensitivity to illumination that can reduce the accuracy of gait recognition as a result of silhouette segmentation failure. In order to overcome this problem Sivapalan et al. [10] and Chattopadhyay et al. [11] used the infrared based Kinect 1.0 camera, and proposed gait recognition methods based on binary voxel volume. However, the differential absorption of the infrared ray produces noise on different body parts[11]. Geisheimer et al. [12] proposed a micro Doppler radar to obtain gait signature from frontal view. However, it requires the subjects to wear infrared reflective markers which is not suitable for a real application. The main disadvantages of previous methods are that they do not consider the inner body shape and motion, and do not address the problem of carrying condition. In this study, we propose a new gait sequence captured using a laser-based Time of Flight (ToF) camera which can overcome some of the aforementioned problems encountered in methods using RGB and infrared cameras and radar by providing precise depth information for frontal view, and which may open up new avenues in gait recognition research. New methods to enhance the segmentation of human silhouette, and gait

cycle estimation as well as four new gait image representations, namely Gait Depth Energy Image (GDE), Partial GDE (PGDE), Discrete Cosine Transform GDE(DGDE), Partial DGDE (PDGDE), plus the combined Fusion GDEs (FGDEs) are proposed in this paper.

The rest of the paper is organized as follows. The ToF data set and the gait recognition algorithm are presented in Section 2 and Section 3 respectively. The experimental results for the proposed methods and comparison with their counterparts are provided in Section 4. The contributions are summarized in Section 5 along with some concluding remarks.

2. THE TOF DATA SET

The ToF data set was created by capturing gait sequences of 46 subjects from a frontal view. The subjects were asked to walk towards the camera six times and with five different covariates: two normal walks, one slow walk, one fast walk, one carrying two bags in both hands, and one carrying a ball with both hands. All the covariates are necessary for the evaluation on the effect of the proposed method over different gait patterns of an individual. A ToF camera [13] was used to capture the gait patterns. We set the capturing speed at 50fps, with the Multi Frequency Spatio Temporal Mode which makes four captures to disambiguate the distance of the objects from the sensor's nominal unambiguous range before producing one frame. The camera resolution is 160 by 120 in spatial coordinates and 16-bit in depth. Figure 1 shows the experimental setup; a subject was asked to walk towards the camera from line AB to CD and after reaching line CD the subject walked to the side of the camera (point E). The camera produces a brightness image (b-image), a horizontal distance image (x-image), a vertical distance image (y-image), and a depth image (z-image) which is the perpendicular distance between the camera and the target point. Therefore, a single point belonging to an object in the scene is represented by a three-tuple (x,y,z) which is found from the corresponding locations in the x-, y- and z- images.

3. PROPOSED METHOD

The methods proposed in [6], GEI and GEnI, used the average of binary silhouettes over one gait cycle. These image representations only contain the information on the 2D shape and 2D contour motion of the human body. On the other hand, our proposed image representations use frontal depth information as the feature for gait recognition. This produces the 3D shape and 3D contour motion which are important features for gait recognition especially when a person is walking perpendicular to the optical axis of the camera. The Kinect 1.0 camera only has 11-bit depth resolution which limits the range and accuracy [14] compared to the 16-bit depth resolution of our ToF camera. Also the optimal sensor range for the Kinect 1.0 camera is only 4 meters(m) while the ToF camera can sense up to 7m.

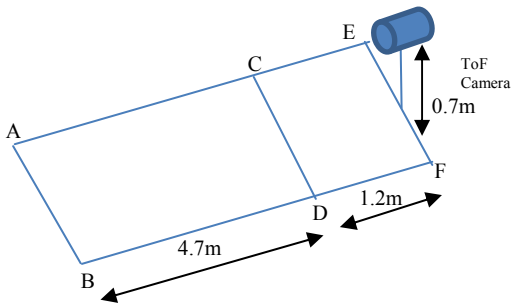


Fig. 1. The experimental setup with ToF camera.

In contrast to the ToF camera, the Kinect 1.0 camera depth sensing are affected by shade, dark blue color and shiny surface [15]. The proposed gait recognition framework consists of: human silhouette segmentation, gait cycle estimation, image representations and the recognition algorithm. The method starts with obtaining the foreground z-image by means of a simple frame difference, and thresholding with a value identified using Rosin's threshold method [16]. One of the main problems using the TOF-camera is that the emitted light from the camera is reflected in many directions by the objects and after multiple reflections this diffuse light illuminates all the objects in the scene. In this way, a fraction of the detected light signal is formed by the reflection of this diffuse light and this is not related to the distance [17]. In order to rectify this, a depth image enhancement algorithm, which is applied to the foreground z-image, is proposed. First, the pixels whose depth values corrupted by noise are removed based on the following equations:

$$U_{lim} = tp_{lim} + mean(Z_{fore}) \quad (1)$$

$$L_{lim} = mean(Z_{fore}) - tp_{lim} \quad (2)$$

where Z_{fore} shows the foreground z-image and tp_{lim} indicates the permitted fluctuation from the mean value of Z_{fore} and is set to 1500 experimentally. After this stage, the algorithm cleans up the separated blobs by keeping only the blob having the maximum area. This will speed up the removal of spurious pixels in the next stage. In the third stage, the spurious pixels in Z_{fore} due to the multiple reflection problem are tackled by using the x- and y-images which correspond to actual x- and y- coordinates of Z_{fore} image pixels in the 3D representation of the scene. Using the least squares fitting method with a first degree polynomial, the algorithm estimates the pixel value at each coordinate of each row in the x-image and each column in the y-image, as shown in Figure 2. In order for the algorithm to make the correct pixel value estimation, the algorithm only makes the estimation if the number of points in each row (x-image) or column (y-image) is greater than 10 and with 90% of nonzero elements. For the same reason, we group pixels into different

groups if they are greater than *diffgrp* which is measured by the consecutive difference between two pixels. The empirical value of *diffgrp* is 50. The pixel estimation is made based on the group with the highest number of pixels. Then, we remove the pixel in the z-image based on the distance between the actual pixel values and the estimated pixel value (*distval*) greater than 100 which is decided empirically. Finally, only the blob having maximum area is retained and other blobs are deleted. Figures 3(a), (b) and (c) show the raw z-image, the segmented image using Rosin's thresholding, and the enhanced image using our proposed algorithm, respectively. The proposed gait cycle algorithm starts by removing the body parts outside the torso area, which is defined according to the height of the subject. Based on the human anatomy [18], the torso is the area between 0.8 and 0.5 of a person height. From the torso area the middle column of coordinates is identified. The middle column of the torso is used to separate the left and right legs. Here, the leg area is the area below 0.35 of the height [18]. After dividing the leg into left and right areas, we find the mean of the pixel values in each area, which determines the average distance of the corresponding leg from the camera. The difference between the mean of the left and right legs in each frame is used as the feature to identify the gait cycle. In this study, 3 consecutive minimum values in a periodic function are used as one gait cycle. In order to reduce the fluctuation of local minima that are not caused by the minimum distance between the two legs, a threshold value is determined by the average of local minima. Any value in the local minima will be removed from the sequence if it is greater than the average value. The final frame in a gait sequence is decided by the mean of z distance of a silhouette $\leq$ 2400 which is adjusted according to the experimental setup. Since the image of a subject is bigger and more accurate if he or she is closer to the camera, it was decided to use images of the gait sequence within the last three local minima for the development of the image representations.

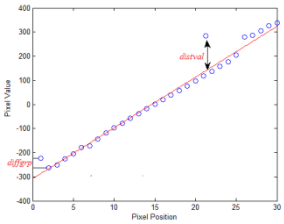


Fig. 2. The estimated pixel value sequence generated using Least Squares Fitting with pixel position in x-row or y-column.

(a) (b) (c)

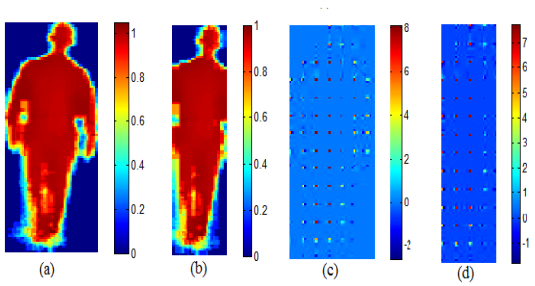
Fig. 3. (a) The raw z-image (b) The human silhouette produced by Rosin's threshold (c) The enhanced human silhouette

After estimating the gait cycle, the GDE is generated by the following formula:

$$GDE(x,y) = \frac{1}{k}\sum_{j=1}^{k} Z_{nj}(x,y) \quad (3)$$

where k is the number of frames in one gait cycle, Z_{nj} (x, y) is the normalized z-image at frame j which carried out by dividing the z-image by the mean of the z-image. The PGDE is introduced to reduce the impact of carrying conditions. The PGDE is generated by cropping the left and right body parts. This is carried out by identifying the boundaries which are the start and end columns of the silhouette image above the shoulder line. The shoulder line is defined as 0.8 of a subject's height [18]. The cropped images in one gait cycle are then averaged to produce PGDE. The DGDE and PDGDE are generated by applying the two-dimensional Discrete Cosine Transform (DCT). The DCT has the ability to pack the energy of the highly correlated images which makes it suitable when a shift of the pixel value position occurs. In this experiment the DCT with 8-by-8 blocks is applied to GDE and PGDE to generate DGDE and PDGDE respectively. Figures 4(a), (b), (c) and (d) show the GDE, PGDE, DGDE and PDGDE images respectively. Although the PDGDE and the PGDE managed to overcome the impact of carrying conditions compared to DGDE, the left and right sides of the body parts contain little information that enables us to distinguish between certain individuals. However, while the DGDE and DPGDE managed to overcome the shift in position, the GDE and PGDE perform better than DGDE and DPGDE if the pixel differences between the gallery which contains the gait sequences that are known to the algorithm and the probe which is a gait sequence that is presented to the algorithm for recognition is very low. For these reasons we introduce the Fusion GDEs (FGDEs) which uses all four representations by dividing each representation into upper and lower body. The matching algorithm in the proposed method starts by identifying the weights for each upper and lower section of the all four image representations. The weights, w_k for the kth section of the image representations in the gallery, *GIgallery* are computed based on the following formula:

$$w_k = \frac{1}{(\sum_{i=1}^{T}\sum_{j=1}^{T}|GIgallery_i - GIgallery_j|_k \times totaldiff)^p} \quad (4)$$



(a) (b) (c) (d)

Fig. 4. The gait image representations: (a) GDE (b) PGDE (c) DGDE (d) PDGDE

where T is the total number of subjects in the gallery, p is a power factor that increases the discriminatory power of each of the sections, and *totaldiff* is given as follows:

$$totaldiff = \sum_{k=1}^{Q} \frac{1}{(\sum_{i=1}^{T}\sum_{j=1}^{T}|GIgallery_i - GIgallery_j|)_k} \quad (5)$$

where $Q = 8$, which is the total number of sections in all four image representations. Finally, the matching is implemented based on the following formula:

$$MatSub = \arg min \sum_{k=1}^{Q} w_k \times \left|GIprobe - GIgallery_j\right|_k \quad (6)$$

for j = {1,2,..T} where *MatSub* is the matched subject in the gallery for a test probe, *GIprobe*.

4. RESULTS AND DISCUSSION

In our experiments, one of the 'normal walk' gait sequences was used as the gallery and another 'normal walk' and the other four covariates were used as probe. The Cumulative Match Score (CMS) Rank 1 and Rank 5 were applied to evaluate the accuracies and robustness of the proposed methods, which are compared with two previous non-depth frontal view representations: GEI [6] and GEnI [6]. The silhouettes used to generate GEI and GEnI are produced by converting the depth silhouette to the binary silhouette. As shown in Table 1, all the proposed methods outperform the previous methods in overall accuracies and in almost all covariates.

Although the Rank 1 result of GEnI is better than GDE for 'carrying bags' covariate, both of them produce the same accuracy at Rank 5. This shows that GDE is competitive with GEnI on this covariate. It can also be observed that PDGDE and PGDE managed to overcome the impact of carrying conditions better than DGDE and GDE. However, GDE and DGDE perform slightly better than PGDE and PDGDE respectively on 'normal walk', 'slow walk' and 'fast walk' because little information is contained in the structure and the swings of hands. DGDE and PDGDE are better than GDE and PGDE because of their ability to overcome the pixel shift problem or misalignment due to the swing of the body during walking. FGDEs performed better for Rank 1 than the other methods on overall accuracies as well as in all five covariates. The discrimination power p was experimented with, and the result is presented in Figure 5. The overall Rank 1 accuracy is increasing from $p = 1$ until 4. At these p values the accuracy of the 'normal walk' and 'slow walk' are constant and start to drop when p reaches 3. However, the accuracy of the 'fast walk' covariate increases until it reaches 3 and drops when p is higher than 3. On the other hand, the accuracy of the carrying bags and ball improves as p increases. The reason for this is because the upper body for PDGDE weight has the highest value; hence the upper body for PDGDE image representation dominates the score level fusion matching. Based on the experiment, the recommended value for p is in the range of 2.5 to 4.5 with the overall Rank 1 accuracy more than 80%.

5. CONCLUSION

In this paper, we have proposed a new methodology for frontal view gait recognition by using depth information obtained by means of a ToF camera. A new data set that consists of five walking covariates was created for research purposes. A new method to enhance the segmentation algorithm has also been presented which can be used not only in gait recognition but also in other applications using a ToF camera to obtain depth information of the scene. A novel technique for gait cycle estimation has also been developed for gait recognition using depth information. For this purpose, four new gait image representations and their fusion have also been proposed. The proposed methods outperform two major counterparts (frontal view GEI and frontal view GEnI). The fused representation (FGDEs) almost always gives the best overall performance.

Table 1. The comparative results of the proposed methods and two previous methods with parameter p set to 2.5.

	Normal Walk (%)		Slow Walk (%)		Fast Walk (%)		Carrying Bags (%)		Carrying Ball (%)		Overall (%)	
	Rank 1	Rank 5	Rank 1	Rank 5	Rank 1	Rank 5	Rank 1	Rank 5	Rank 1	Rank 5	Rank 1	Rank 5
Frontal GEI [6]	87.0	95.7	76.1	**93.5**	67.4	87.0	8.7	28.3	32.6	63.0	54.4	73.5
Frontal-GEnI [6]	87.0	95.7	69.6	84.8	60.9	82.6	13.0	30.4	32.6	58.7	52.6	70.4
GDE	91.3	95.7	76.1	**93.5**	69.6	87.0	8.7	30.4	34.8	65.2	56.1	74.3
PGDE	82.6	93.5	56.5	84.8	58.7	84.8	30.4	54.3	58.7	84.8	57.4	80.4
DGDE	**93.5**	95.7	78.3	**93.5**	76.1	89.1	8.7	34.8	32.6	69.6	57.8	76.5
PDGDE	87.0	95.7	71.7	93.5	73.9	89.1	54.4	84.8	78.3	91.3	73.0	90.9
FGDEs	**93.5**	**97.8**	**80.4**	91.3	**89.1**	**95.7**	**67.4**	**84.8**	**76.1**	**87.0**	**81.3**	**91.3**

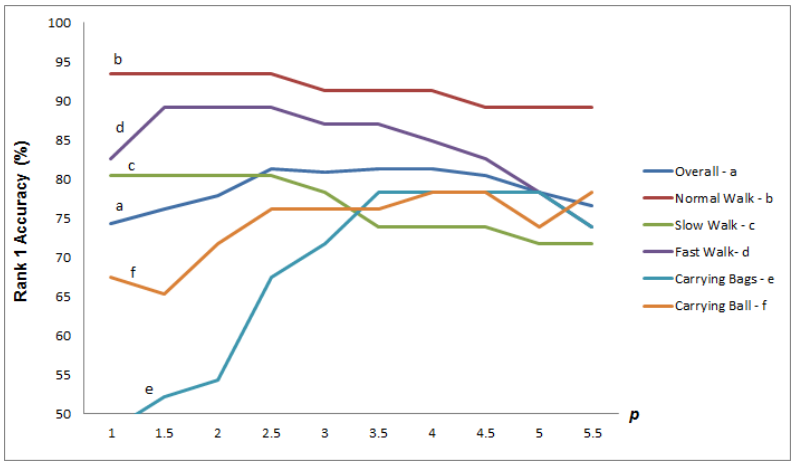


Fig. 5. The impact of the power factor *p* on the overall and walking covariates for Rank 1 accuracy.

REFERENCES

[1] Nixon, M. S., Carter, J. N., Cunado, D., Huang, P. S. and Stevenage, S. V., "Automatic Gait Recognition" in Jain, A. K., Bolle, R. and Pankanti, S. (eds.) Unspecified Biometrics: Personal Identification in Networked Society, Kluwer Academic Publishers, pp. 231-250, 1999.

[2] O. Barnich and M. Van Droogenbroeck, "Frontal-view gait recognition by intra- and inter-frame rectangle size distribution", Pattern Recognition Letters, vol. 30, no 10, pp. 893-901, 2009.

[3] M. Soriano, A. Araullo and C. Saloma, "Curve spreads – a biometric from front-view gait video", Pattern Recognition Letters, vol. 25, no. 14, pp. 1595-1602, 2004.

[4] J. A. Balista, M. N. Soriano and C. A. Saloma, "Compact time-independent pattern representation of entire human gait cycle for tracking of gait irregularities", Pattern Recognition Letters, vol. 31, no. 1, pp. 20-27, 2010.

[5] M. Goffredo, J. N. Carter and M. S. Nixon, "Front-view gait recognition" in IEEE 2nd International Conference on Biometrics: Theory, Applications and Systems, pp. 1-6, Arlington Virginia USA, Sep 29-Oct 1, 2008.

[6] D. S. Matovski, M. S. Nixon, S. Mahmoodi and J. N. Carter, "The Effect of Time on Gait Recognition Performance", IEEE Trans. on Information Forensics and Security, vol. 7, no. 2, pp. 543-552, 2012.

[7] J. Han and B. Bhanu, "Individual recognition using gait energy image", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 28, no. 2, pp. 316-322, 2006.

[8] K. Bashir, T. Xiang and S. Gong, "Gait recognition using gait entropy image" in 3rd International Conference on Crime Detection and Prevention, pp. 1-6, London UK, Dec 3, 2009.

[9] J. Ryu and S. Kamata, "Front view gait recognition using spherical space model with human point clouds" in IEEE International Conference on Image Processing, pp. 3270-3273, Brussels Belgium, Sep 11-14, 2011.

[10] S. Sivapalan, D. Chen, S. Denman, S. Sridharan and C. Fookes, "Gait Energy Volume and Frontal Gait Recognition using Depth Images" in International Joint Conference on Biometrics, pp. 1-6, Washington D.C. USA, Oct 11-13, 2011.

[11] P. Chattopadhyay, A. Roy, S. Sural and J. Mukhopadhyay, "Pose Depth Volume extraction from RGB-D streams for frontal gait recognition", Journal of Visual Communication and Image Representation, vol 25, no.1, pp. 53-63, 2014.

[12] J.L. Geisheimer, W.S. Marshall, and E. Greneker, "A Continuous-Wave (CW) Radar for Gait Analysis," in 35th Asilomar Conference on Signals, Systems and Computers, vol 1, pp. 834-838, Pacific Grove California USA, Nov 4-7, 2001.

[13] Fotonic Inc. "Fotonic", http://www.fotonic.com/content/Default.aspx, [April 25, 2013].

[14] M. A. Livingston, J. Sebastian, Z. Ai, and J. W. Decker, "Performance measurements for the Microsoft Kinect skeleton" in IEEE Virtual Reality Conference, pp.119-120, Costa Mesa California USA, Mar. 4-8 2012.

[15] C. D. Mutto, P. Zanuttigh, and G. M. Cortelazzo, "Time-of-Flight Cameras and Microsoft Kinect™: A user perspective on technology and applications", Springer US, 2013.

[16] P. L. Rosin, "Unimodal thresholding", Pattern Recognition, vol. 34, no.10, pp. 2083-2096, 2001.

[17] D. Falie and L. C. Ciobotaru, "A simple enhancement algorithm for time-of-flight camera range images" in International Symposium on Signals, Circuits and Systems, pp. 1-4, Iasi Romania, Jun 30-Jul 1, 2011.

[18] D. A. Winter, "The Biomechanics and Motor Control of Human Movement." Hoboken, NJ, Wiley, 1990.