Frontal View Gait Recognition with Fusion of Depth Features from a Time of Flight Camera

Tengku Mohd Afendi, Fatih Kurugollu, Senior Member, IEEE, Danny Crookes, Senior Member, IEEE, Ahmed Bouridane, Senior Member, IEEE, and Mohsen Farid, Senior Member, IEEE

Abstract

Frontal view gait recognition for people identification has been carried out using single RGB, stereo RGB, Kinect 1.0 and Doppler radar. However, existing methods based on these camera technologies suffer from several problems. Therefore, we propose a four-part method for frontal view gait recognition based on fusion of multiple features acquired from a Time of Flight (ToF) camera. We have developed a gait data set captured by a ToF camera. The data set includes two sessions recorded seven months apart, with 46 and 33 subjects respectively, each with six walks with five covariates. The four-part method includes: a new human silhouette extraction algorithm that reduces the multiple reflection problem experienced by ToF cameras; a frame selection method based on a new gait cycle detection algorithm; four new gait image representations; and a novel fusion classifier. Rigorous experiments are carried out to compare the proposed method with state-of-the-art methods. The results show distinct improvements over recognition rates for all covariates. The proposed method outperforms all major existing approaches for all covariates and results in 66.1% and 81.0% Rank 1 and Rank 5 recognition rates respectively in overall covariates, compared with a best state-of-the-art method performance of 35.7% and 57.7%.

Index Terms—Gait recognition, frontal view, Time of Flight camera, fusion of features, depth gait data set.

I. INTRODUCTION

Gait is the combination of posture and the way we move our whole body during the walking process [1]. It has been used as a discriminating feature in much recent research related to clinical analysis, gender classification, age estimation, forensics tools, and biometrics.

One interesting application in which gait features are used is biometrics. Among the earliest evidence for using gait as a biometric was the work of Murray *et al.* [2] and Johansson [3]. From a human anatomical point of view, Murray *et al.* suggested that gait is unique to an individual. Based on the experiments conducted by Johannson [3] and Stevenage *et al.* [4], they concluded that humans have the ability to identify individuals based on their gait. Unlike other biometrics such as fingerprint, finger veins, palmprint and palm veins, gait recognition can be used without direct contact with the sensing device. Unlike face and iris recognition, gait recognition does not require any specific postures or positions. It does not require the

cooperation or even awareness of the individual under observation. Also, the gait is hard to conceal and difficult to disguise [1]. Gait features are perceivable at a distance, and only low resolution is required [5] - [7].

Although several approaches have been presented for gait recognition, most limit their attention to the lateral view, since this is considered to provide much more spatial and temporal information [8], [9]. However, this approach requires the camera to be placed at a certain height and distance, to capture full gait sequences. However, this is only applicable in outdoor or wide indoor spaces, and not in applications such as a secure narrow corridor. In such situations, frontal view gait recognition can be applied. Frontal view gait patterns can also be integrated with facial patterns to enhance biometric identification.

Early attempts at using frontal view gait recognition used a single RGB camera. Barnich and Droogenbroeck [10] proposed gait features derived from a set of rectangles fitting any closed silhouette in RGB video frames. However, the size of the rectangles has to be changed if a subject wears bigger clothes or high heel shoes. They managed to produce good results but tests were not carried out on the clothing and shoes covariates. Soriano et al. [8] and Balista et al. [11] applied Freeman Chain Code to the silhouette edge image. The method depends on high quality silhouette segmentation which is very difficult to achieve in a complex background. The frontal view gait recognition algorithm in [12] 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 that the edge points and stacking methods are very dependent on clothing, shoes, and carrying conditions. Soriano et al. [8] achieved 100% accuracy but the experiment only involved normal walk, with only 4 subjects who had to wear a special suit. Balista et al. [11] performed analysis on the gait irregularities only, and no gait recognition results were presented. Matovski et al. [1] applied Gait Energy Image (GEI) [13] and Gait Entropy Image (GEnI) [14] methods to frontal view based gait recognition. The GEI is generated by averaging the binary silhouettes in one gait cycle. This reduces the silhouette noise, so GEI is less sensitive to noise. However, according to Bashir et al. [14], the presence of shadow (through lighting effects) can affect the accuracy of the GEI algorithm. Overall GEI produces good results in the experiments conducted in [1]; this is because the environment (background, lighting, walking surface and location) remains the same, eliminating the different types of shadow that would affect the accuracy of GEI. The GEnI, based on Shannon Entropy, produces high intensities in the dynamic areas such as legs and hands and low or zero intensities in the static areas. Unlike the GEI, GEnI is less affected by the presence of shadows. However, GEnI may only be effective if the subjects walk with constant or almost constant speed at all times, which may not be true in all conditions. If a test subject (or probe) walks slower than the subject in the gallery (the training set), the Shannon Entropy produces lower intensities especially in the dynamic areas. Likewise, when walking faster than normal speed, GEnI produces higher

intensities especially in the dynamic areas. The speed covariate experiment in [1] showed that GEnI experienced only a slight drop in performance; however, the speed variations were only 25% or less. Higher speed variations will decrease the performance of GEnI.

RGB cameras are widely used in lateral view gait recognition. Recently, Aggarwal and Vishwakarma [15] applied Zernike moment on a gait image representation called Average Energy Silhouette Image (AESI) to detect the presence of an object carried by an individual. Also, the features called Mean of Directional Pixels (MDP) and Spatial Distribution of Gradients (SDOG) are applied. MDP only considers the horizontal body motion, which is more suitable for lateral view gait recognition. It is not suitable for frontal view because the motion of the lower part of the body and hands are not horizontal. On the other hand, SDOG takes into account gradient information based on local orientation. Hence, it only considers the spatial features of gait. SDOG features are not suitable for gait recognition if there are changes of features caused by the temporal motion such as different walking speeds. Also, the experiments conducted did not involve speed covariates. The overall performance for this method was 91.47% for all CASIA datasets, 72.7% and 84.67% for OU-ISIR Treadmill Dataset B and USF datasets respectively. The method in [16] uses spatial-temporal and kinematic features from gait silhouettes and applied a deterministic learning method to produce dynamic gait features. For the spatio-temporal method, the silhouette is divided into several regions and the median of all widths is computed. However, the widths of the leg and hand regions change with different walking speeds. The widths of the head area may also change if a person moves his/her head position. The kinematic features are generated from moving body parts. If the positions of body parts are measured relative to the height of an individual, this will change if the individual uses different types of shoes. The accuracies of this methods are 94%, 99%, 90%, 98% and 94.4% on CASIA B, CASIA C, TUM-GAID, OU-ISIR, Treadmill Dataset A and USF-Human ID respectively. Castro et al. [17] combined the optical flow method and a Convolutional Neural Network (CNN) to produce new gait features. The optical flow method is sensitive to illumination changes. Another disadvantage of the GFI is that walking slower or faster than the gallery walking speed may produce different horizontal and vertical components of the optical flow, so this may affect the accuracy of the gait recognition. The method only achieved 59.4% average accuracy on the TUM-GAID dataset. Both methods in [18] and [19] combined GEI and a CNN to produce features for gait recognition. The problem with CNN is the computational complexity of the algorithm. Hence, the gait image size has to be small and in some cases the GEI image resolution needs to be reduced, thereby reducing the significant features in GEI and optical flow images [17]-[19]. The overall performance of the method in [18] and [19] on the CASIA B dataset was 86.70% and 95.88% respectively. Castro et al. [20] combined optical flow and a people detection algorithm that detects whether the moving object is human or not. This produces motion features called Tracklets. The people detection is based on a predetermined model of the human body. However, the

detection and the optical flow algorithms used in this work are not robust against different illumination conditions or similarity between the clothing colors of an individual and the background colors of the given image. These degradations generate incorrect Tracklet features. In addition, the Tracklets based optical flow is not robust to walking speed variation. Overall performance of the method on the lateral view CASIA B dataset is 95.2%.

A single RGB camera may not be able to provide enough information in a frontal view gait image sequence. Hence, Ryu and Kamata [9] used a stereo RGB camera system which generates a human point cloud using spherical coordinates. The method in [9] is scale invariant. However, it ignores the vertical axis change of direction which is important for features in the shoes covariate. The experiment involved 20 subjects with normal walk, fast walk, slow walk and walk with a bag. It achieved overall performance of 98.7%. However, the experiments were not conducted under rigorous gait recognition conditions. This was because the Curve Spread method [8] which uses Freeman Chain Code features which are also susceptible to noise, achieved only 82.5% overall accuracy.

The main disadvantage of the single and stereo RGB camera systems is that performance drops dramatically if the underlying human silhouette segmentation algorithm fails. RGB systems are sensitive to color differences between clothing or footwear (foreground color) and the environment (background color). Even if the actual foreground and background colors are different, illumination, shade and shadow may change the colors. Also, using a stereo RGB camera system is compute-intensive because of the stereo matching process in the post production of the 3D images.

In order to overcome this problem, Sivapalan et al. [21] and Chattopadhyay et al. [22] used the infrared based Kinect 1.0 camera system to produce depth measurement of the object in its scene. The human silhouette segmentation based depth is not affected by the illumination problems of RGB camera systems. They produced features known as Gait Energy Volume (GEV) and Pose Depth Volume (PDV). These two features are based on binary voxel. The construction of the binary voxel is highly sensitive to outliers which affect its accuracy [23]. Chattopadhyay et al. [23] proposed another method using front and back views from two Kinect 1.0 cameras. Due to the limitation of the Kinect 1.0 camera's range, the proposed method only captured an incomplete gait cycle. Therefore, the features in this method were based on only a few frames from the time interval, so the accuracy is affected by different walking speeds and different lengths of the first step. The lighting conditions were also controlled because this can affect the depth measurement of the Kinect 1.0 camera. This requirement was supported by research in [24] - [27] which found that the Kinect 1.0 camera is highly sensitive not only to lighting conditions but also to types of surfaces and colors. The Kinect 1.0 also produces noise on different body parts [22]. In Chattopadhyay et al. [22] the Kinect camera is able to capture full depth variation in limbs only but not the whole body over one gait cycle. Also, in [23],

the Kinect cameras were not able to record complete gait cycles. The good algorithm such as frontal view GEI [1] requires features from both lower and upper body parts in a complete gait cycle. In [1], the frontal view gave more than 90% accuracy. However, the rank 1 accuracy of the frontal view GEI in [22] and [23] were as low as 33.33% and 37.93% respectively. This is because the frontal view GEI features of the whole body over one gait cycle are not completely constructed. The problems experienced in [22] and [23] are caused by the sensor range limitation of the Kinect camera. The maximum sensor range for the Kinect 1.0 and 2.0 cameras is only 4.0m [28]. Recently, Zou et al. [29] combined features from a tri-axial accelerometer sensor in a smart phone and a Kinect camera to identify individuals based on their gait patterns. From the color and depth images of the Kinect camera, features called Eigengait and TrajGait are produced. The Eigengait is based on the EigenFace [30] features that are sensitive to lighting conditions [31]. Furthermore, Trajgait is sensitive to motion and can be affected by walking speed covariates. Also, for the accelerometer sensor, users may forget to bring their smart phones.

Geisheimer et al. [32] and Tahmoush and Silvious [33] proposed a method using both micro Doppler radar and infrared sensors to obtain a gait signature from a frontal view. Simultaneous infrared and radar measurements were taken with the goal of eventually correlating radar features to their biomechanical source. However, both methods in [32] and [33] are not suitable for a real application because the subjects need to wear infrared reflective markers.

Balazia and Sojka [34] use features extracted from the motion of joint angles through the Maximum Margin Criterion method. This method used the CMU-MoCap dataset that recorded the motion of joint angles with an optical marker-based Vicon system [35]. Similar to [32] and [33], this method is not suitable for real applications because the subjects are required to wear a black jumpsuit which has 41 markers taped to it.

Given the problems experienced by the above methods and sensor technologies, we propose a frontal view gait recognition method based on using a 3D Time of Flight (ToF) camera, which can generate more accurate depth images. Unlike single or multiple RGB camera systems, a ToF camera produces gait images which are based on the depth, so it is not affected by color problems, or by illumination, shadows and shade. ToF technology does not require compute-intensive depth reconstruction. Also, unlike a RGB stereo based camera, a ToF camera delivers reliable depth information in low or repetitively texturized areas [25].

However, if we use ToF technology, a novel method is required because of the nature and interpretation of ToF images. New algorithms are required at all stages in the recognition process.

In comparison with existing studies in this area, the contributions of the research presented in this paper are:

- *New human silhouette extraction algorithm* – This new algorithm not only extracts the human silhouette but also reduces the multiple reflections problem experienced by a ToF camera.

- *Gait cycle frames selection algorithm* – To select the frames for one gait cycle, a new gait cycle detection algorithm based on depth information is developed.
- *Novel gait image representations* – Four gait image representations are developed. Each representation performs better than the others on certain covariates. This suggests that the gait image representations can be fused, to make the algorithm more robust overall.
- *Adaptive Multi-Stage Fusion Classifier* – Our algorithm is a hierarchical classifier that fuses the novel gait image representations. It identifies the covariates and applies a specialized classifier for that specific covariate.

These four algorithms are an extended version of our work in [36]. Compared to the previous paper, this paper explains the proposed algorithms thoroughly. Also, the gait image representations have been improved with removal of the area below the shin, using an α parameter. In addition, this paper introduces the new Adaptive Multi-Stage Fusion Classifier.

The remainder of this paper is organized as follows: Section II describes the development of the proposed data set. Section III introduces the proposed gait recognition method that includes: the new human silhouette extraction algorithm; the gait cycle frame selection algorithm; the development of the new gait image representations; and the novel fusion classifier. Experimental results are presented and analyzed in Section IV. Section V concludes the paper.

II. The Development Of The Proposed Data Set

In this research, a Fotonic B series ToF camera [37] is used to capture frontal view gait sequences. It measures the distances between the camera and objects based on the travel time of the emitted light from the camera to the objects and back again. The Kinect 1.0 and 2.0 cameras' sensor range is 4.0 meters [28], while the ToF camera can sense accurately the depth of objects up to 7m [37]. This difference is significant because, unlike the Kinect cameras, the proposed ToF camera is able to capture images of the whole body over one complete gait cycle as explained in the previous section. Both Kinect 1.0 and 2.0 cameras have depth sensors and an RGB camera. The cameras produce colored point clouds that suffer from a non-accurate association between depth and RGB data, due to a non-perfect alignment between the camera and the depth sensor. Moreover, depth images suffer from a geometric distortion; this requires calibrations that relate the 3D coordinates to 2D image coordinates [38], [39]. The proposed camera ToF does not need to be calibrated to produce 3D measurements

Our ToF camera has two disadvantages over the Kinect. The ToF camera has lower spatial resolution than the Kinect 1.0 [26]. The Kinect 1.0 and 2.0 have 320 x 480 and 512 x 424 spatial resolutions respectively [28], while our ToF camera has only 160 x 120 spatial resolution. However, it has been shown that gait recognition can be carried out using low resolution human silhouette images [5] - [7]. The second

disadvantage is that the cost of a ToF camera is greater than either Kinect camera. However, it is expected that the cost of ToF cameras will decrease significantly in the future [40], [41]. Therefore, this disadvantage is not critical in the long term for frontal gait recognition applications. We capture gait image sequences at 50 frames per second (fps). We set the predefined filter to Multi Frequency Spatio Temporal, which improves the sensing accuracy by taking four captures before producing one frame of the depth image. The ToF camera used in the tests generates four files which store horizontal distance, vertical distance, depth distance and brightness images. The 16-bit Portable Gray Map (PGM) file format is used to store all the images. In this research, only horizontal distance (X), vertical distance (Y) and depth distance (Z) images are used for gait recognition. The depth distance is the perpendicular distance from a target point to the origin of the coordinates. All the distance measurements are in millimeters (mm).

The aerial view of the experimental setup is illustrated in Fig. 1. The height of the ToF camera is 0.7m using a tripod in area C. Referring to Fig. 1, a subject was asked to stand in area A and walk towards the camera through Area B until he or she crossed line B1-B2. After crossing line B1-B2, the subject was asked to turn left and enter area D. Then the same procedure is repeated for different covariates. The distance between lines A1-A2 and B1-B2 is 4.7m and the distance from lines B1 and B2 to the camera is 1.2m. The two parallel dotted lines illustrate the walking direction. The subjects were not controlled to walk strictly with respect to the center/vertical axis of the camera. Hence the subjects were allowed to walk freely as they were approaching the camera. This produces frontal or nearly frontal view gait sequences. Based on the setup in Fig. 1, two sessions, repeated seven months apart, were conducted. The first and second sessions were conducted in May 2013 and December 2013 respectively. This is because the gait of a person can vary over time (time covariate). The first session involved 46 people, and each subject was asked to do 6 walks which involved 5 different covariates: 2 normal walks, 1 slow walk, 1 fast walk, 1 carrying two bags with one bag in each hand, and 1 carrying a ball with both hands. The walking speed was normal for both carrying cases. In the second session, only 33 subjects who were involved in the first session participated. In this session, the subjects were asked to do 5 walks, one for each of the 5 covariates. In the second session, we did not require the subjects to wear the same types of footwear and clothing as in the first session. This was to make the tests for time covariate more realistic. Before the start of each data collection session, a subject was also briefed about the covariates and most importantly about the walking speeds: normal walk, slow walk and fast walk. Since the exact walking speed is not controlled, briefing is vital, so that the subjects bore their natural variations of walking speeds in mind before capturing their gait image sequences with those covariates. The following metadata were also collected: gender (57 percent male), age (19 to 59 years old), height (1.50 m – 1.88m), and weight (42 - 114 kg). Of 46 subjects, 44 were right foot dominant and only one was left foot dominant. This information can be used for analyzing the performance of gait recognition in different

categories. This metadata can also be used in future research such as gender classification, age and height estimation, based on gait.

III. The Proposed Method

The proposed method consists of four stages: human silhouette extraction based on multilevel segmentation, frames selection based on the gait cycle detection, feature extraction through different gait image representations, and classification based on an Adaptive Multi-Stage Fusion Classifier (AMSFC).

A. Human Silhouette Extraction

The first stage is the extraction of the human silhouette from the depth image (Z-image). The algorithm starts by applying a simple background subtraction technique and then converting the subtracted image to a black and white image using Rosin's threshold method [42]. Experiments in [42] show that this method produces better results on the difference images than other thresholding methods such as Tsai's [43], Otsu's [44], Kapur's [45] and Ridler and Calvard's [46]. After that, the initial human silhouette filled with depth distances, Zinit, is obtained by multiplying the current foreground image with the black and white image.

One of the main problems when using a TOF-camera is that the emitted light from the camera is reflected in many directions by the objects. Thus, a fraction of the detected light signal is not related to the distance [47]. To reduce this problem, a depth image enhancement algorithm is proposed. First, the Zinit noisy pixels with values greater than upper and lower thresholds are removed. The thresholds are the permitted fluctuations from the average of Z_{init} (x,y) > 0. Next, the algorithm cleans up the image by removing small connected blobs below the maximum area. This will speed up the removal of remaining noisy pixels in the next step.

In the third step, the problem of remaining noisy pixels in Z_{init} is tackled by using the *X* and *Y* images (the actual horizontal and vertical coordinates of the human silhouette) using the linear least square fitting (LSF) method. LSF was chosen because both the horizontal and vertical coordinates have a linear relationship to their sequence positions, as shown in Fig. 2. The pixels in red circles are the noisy pixels which will be eliminated by the proposed algorithm. A horizontal vector H_U of any row, y, is produced using:

$$H_U(n)_y = X_f(\mathbf{x}_n, y) \quad \text{for } \mathbf{x}_1 < \mathbf{x}_2 < \cdots < \mathbf{x}_N \qquad (1)$$

where X_f denotes the horizontal image at current frame *f* and $\mathbf{x}_n$ is the column in X_f with $\mathbf{x}_1$ being the leftmost and $\mathbf{x}_N$ the rightmost columns of a row, *y* in Z_{edge}. Similarly, a vertical vector V can be generated. Before applying the Least Square Fitting (LSF) method in [46], the number of elements in H_{U_y} ($\#H_{U_y}$) and V ($\#V_x$) for each *y* and *x* are inspected using the following:

$$Decide = \begin{cases} Apply\ \text{LSF} & if\ \#\text{Ⱨ}_y \geq \text{Ɣ} \\ Apply\ \text{LSF} & if\ \#\text{Ⱶ}_x \geq \text{Ɣ} \\ Do\ nothing & Otherwise \end{cases} \quad (2)$$

Equation (2) avoids the incorrect generation of LSF lines due to noisy pixels in a short sequence of Ⱨ_y and Ⱶ_x with the empirical value for Ɣ is 10. Other criteria that need to be met before applying LSF to Ⱨ_y and Ⱶ_x are as follow:

$$Decide = \begin{cases} Apply\ \text{LSF} & if\ \left(\#\text{Ⱨ}_y \neq 0\right) > \left(\text{ƭ} \times \#\text{Ⱨ}_y\right) \\ Apply\ \text{LSF} & if\ \left(\#\text{Ⱶ}_x \neq 0\right) > \left(\text{ƭ} \times \#\text{Ⱶ}_x\right) \\ Do\ nothing & Otherwise \end{cases} \quad (3)$$

Equation (3) avoids the incorrect generation of LSF lines due to the presence of too many background pixels in Ⱨ_y and Ⱶ_x. The empirical value for ƭ is 0.9. LSF uses a grouping strategy to isolate the noisy pixels from the noise-free ones. The group is decided based on the difference between one pixel in one group and the next pixel in another group exceeding a certain threshold (in this case, 50). Next, the group which has the most members is used for plotting the LSF line. If the conditions in (2) and (3) are fulfilled, then the LSF method in [48] is applied by using the data in the group with the maximum number of elements. After this, the algorithm retains the blob with maximum area and deletes the smaller blobs. In certain cases, noisy pixels still exist near the feet and connect to the silhouette. Such noise is reduced by identifying the leftmost and rightmost columns of the upper body. Here the upper body is defined as silhouette above the knees position which is $0.715 \times h$ [49], and h is the height of the silhouette. Finally, all the parts of the columns below the knee positions that are outside the leftmost and rightmost columns of the upper body are deleted. Figs. 3(a), (b), (c) and (d) show examples of the background image, the foreground image, the image produced by Rosin's segmentation method and the image enhanced by the proposed algorithm respectively.

B. Gait Cycle Frames Selection

The gait cycle frames selection is the second stage of the proposed method. The video frames selection involves a gait cycle detection algorithm which uses the mean difference between two legs as the feature. To compute the mean difference between the two legs in the depth dimension of each frame, the center between the two legs needs to be determined. This is based on the midpoint of the abdomen, rather than the legs, because the image of the leg closer to the camera is bigger than the one further away. The midpoint of the abdomen is the area between $0.2 \times h$ and $0.5 \times h$. The algorithm then divides the legs area into left and right. The legs area is the area below $0.65 \times h$ which empirically is between knees and thigh. Then, the means of depth for both left and right are computed. After that, the difference between the means of the left and right areas is measured. An example of the mean depth difference between the non-zero-pixel values for the two legs in each frame of a gait sequence is shown in Fig. 4. The local minimum is detected at frame *d*, if the

mean difference between the two legs is less than at frames $d - 1$ and $d + 1$. After that the mean of all minima are computed and shown as the horizontal line in Fig. 4. All the local minima higher than the mean of all minima are removed. If a frame is too close to the frontal view camera, the camera may not be able to capture the whole-body silhouette. This is because the person is too close to the camera. Therefore, for the development of the gait image representations, only frames whose average of non-zero pixels $\geq \lfloor$ are selected. The $\lfloor$ value is set to 2400 which is identified experimentally. This value is identified based on the average of the last frame that contains the complete silhouette from top of the head to the feet of a subject. 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 gait image representations.

C. Development of Gait Image Representations

We propose four gait image representations, namely Gait Depth Energy (GDE), Partial Gait Depth Energy (PGDE), Discrete Cosine Transform GDE (DGDE) and Partial DGDE (PDGDE). GDE is similar to Depth Energy Image (DEI) [50]. The DEI is based on the average distances in one gait cycle. If DEI is applied directly, the absolute depth distances between the camera and a person in the gallery may differ from the absolute depth distances in a probe of the same individual. This would affect the performance. To overcome this, we normalize the DEI, giving GDE. Hence, the different sizes of silhouettes arising from different distances between the camera and the subject are allowed for.

The normalized depth image $\boldsymbol{\mathcal{Z}}_\eta$ is produced using the following equations:

$$\mathcal{Z}_{cr}^{nz}(k) = \mathcal{Z}_{cr}(x,y) > 0 \qquad (4)$$

$$\mathcal{Z}_\eta = \frac{\mathcal{Z}_{cr}(x,y)}{\overline{\mathcal{Z}_{cr}^{nz}}} \qquad (5)$$

$\overline{\mathcal{Z}_{\mathrm{cn}}^{\mathrm{nz}}}$ is the mean of the non-zero elements in $\boldsymbol{\mathcal{Z}}_{cn}$. From $\boldsymbol{\mathcal{Z}}_\eta$ the GDE is produced by averaging the frames which contain $\boldsymbol{\mathcal{Z}}_\eta$ in one gait cycle. The formula to produce the GDE image, $\boldsymbol{\mathcal{Z}}_{\mathrm{GDE}}$ is as follows:

$$\mathcal{Z}_{\mathrm{GDE}}(x,y) = \frac{1}{T_f}\sum_{j=1}^{T_f} \mathcal{Z}_\eta(j)(x,y) \qquad (6)$$

where T_f is the total number of frames in one gait cycle.

For the DGDE gait image representation, $\boldsymbol{\mathcal{Z}}_{\mathrm{DGDE}}$ is produced by applying Discrete Cosine Transform (DCT) [51] to K by L blocks of $\boldsymbol{\mathcal{Z}}_{\mathrm{GDE}}$. The top left corner of the K by L block is the zero-frequency (or DC) coefficient. The DC coefficient holds most of the image energy and represents the proportional average of the K by L block. The total energy remains the same in the K by L blocks but the energy distribution changes, with most energy concentrated in the DC and low frequency coefficients. In static areas of the gait, like the

abdomen, the DC and low frequency coefficients are more significant than the high frequency coefficients. However, in dynamic areas, like hands and lower legs, the high frequency coefficients contribute more to the gait signature based on the $\mathcal{Z}_{\mathrm{DGDE}}$. Hence, this makes DGDE more robust to noise, to variations in walking and to other inherent factors of gait. Figs. 5(a) and 5(b) show the $\mathcal{Z}_{\mathrm{GDE}}$ and $\mathcal{Z}_{\mathrm{DGDE}}$ respectively.

The PGDE gait image representation is produced by deleting the left and right sides of $\mathcal{Z}_{cr}$. This will eliminate the indiscriminate (or non-discriminating) features in the gait image representations due to the different swing of both hands. This is caused by different speeds of walking, carrying objects, mood and other inherent variations of gait. $\mathcal{Z}_{\mathrm{PGDE}}$ is produced by identifying the rightmost and leftmost columns of the silhouettes in $\mathcal{Z}_{cr}$ at the shoulder row (about $0.2 \times h$) [49]. Then all the columns outside these boundaries are deleted. After resizing $\mathcal{Z}_{nr}$ in all the selected frames in one gait cycle, Equation (6) is adapted to produce $\mathcal{Z}_{\mathrm{PGDE}}$. The $\mathcal{Z}_{\mathrm{PDGDE}}$ is produced by applying DCT to K by L blocks of $\mathcal{Z}_{\mathrm{PGDE}}$ (see Fig. 6(a) and (b)).

In addition to all these four gait image representations, we also enhance each gait image representation by removing the indiscriminative area below the shin. The indiscriminative features in this area are caused by the different heights of the feet lifted because of speed variations, types of shoe and other inherent factors of gait. The percentage height of the indiscriminative leg area with respect to the height of each gait image representation is named α. Each gait image representation has a different α which is empirically identified. Fig. 7 shows the GDE image representation after applying the removal based on α.

D. Adaptive Multi-Stage Fusion Classifier

Robustness is one of the most important aspects of a gait recognition method. The method developed must be robust against any motion of pixels or features due to walking speed variation. Another factor that needs to be taken into account is carrying objects. Due to the presence of carried objects, the structure of the body and limited swing of arms/hands would reduce the accuracy of the gait recognition. Therefore, a method is also required to reduce the impact of carrying objects in gait recognition.

Therefore, our adaptive multi-stage fusion classifier is divided into two main parts: an algorithm for the case when the subject is carrying an object, and another for when they are not. There are two cases of carrying an object: the upper body case (e.g. carrying a ball with both hands) and the lower body case (e.g. carrying a bag in each hand). The flow chart of the algorithm when carrying an object (the first part of the adaptive multi stage fusion classifier) is presented in Fig. 8.

The algorithm starts by detecting the presence of an object around the lower body. Using GDE, the algorithm identifies

whether the person is carrying objects around his/her lower body (*LC*) based on the following equations:

$$O_L = max\left(\left[\frac{1}{n}\sum_{x=1}^{n} \mathcal{Z}_{\mathrm{GDE}}(x,y)\right]_{y \in r_l}\right) \qquad (7)$$

$$LC\ is\ TRUE \text{ if } O_L + C_L > \left(\frac{1}{n}\sum_{x=1}^{n} \mathcal{Z}_{\mathrm{GDE}}(x,y)\right)_{y \in r_l} \quad (8)$$

where x and y are row and column pixel coordinates, $r_l = \{\frac{m}{2} + 1 \ldots m\}$. $\mathcal{Z}_{\mathrm{GDE}}(x,y)$ is GDE pixel value at (x, y), n and m are the width and height of the GDE and C_L is a constant value identified empirically as 0.1. C_L and O_L are identified by using the GDE in the gallery. If a person is not carrying an object around his/her lower body, the algorithm checks whether the person is carrying an object around the upper body. If a person is carrying an object around the upper body using both hands, the area which is normally occupied by the hands will have fewer pixels because both hands are nearer to the body center. Based on this, a person is identified as carrying an object around the upper body, U_C, using:

$$O_C = \arg min\left(\left[\frac{1}{n}\sum_{x=1}^{n} \mathcal{Z}_{\mathrm{GDE}}(x,y)\right]_{y \in r_u}\right) \quad (9)$$

$$U_C\ is \text{ TRUE if } O_C - C_U > \left(\frac{1}{n}\sum_{x=1}^{n} \mathcal{Z}_{\mathrm{GDE}}(x,y)\right)_{y \in r_u} \quad (10)$$

where $r_u = \{0.4 \times m \ldots 0.42 \times m\}$ are the estimated rows where the hands are absent because of carrying an object. The empirically determined value of C_u is 0.02. If *LC* and *UC* are true then the proposed algorithm divides the PDGDE into two halves – upper PDGDE and lower PDGDE. Then pixel by pixel matching is carried out for both halves. The matching score for each half is then multiplied by predetermined weights β_{u1} and β_{u2} (for upper and lower halves) if an object is detected around the upper body. If an object is detected around the lower body, the weights for upper and lower halves of PDGDE are β_{l1} and β_{l2} respectively. The empirical values for β_{u1}, β_{u2}, β_{l1} and β_{l2} for our dataset are 0.7, 0.3, 0.8 and 0.2 respectively. The weights for the upper halves are higher than the weights for the lower halves for both *LC* and *UC*.

In this work, we consider a secure corridor application. Therefore, only small objects are typically carried. For our dataset, a small object (a football) was used. β_{u1} is greater than β_{u2} because the ball used is small and the object does not have impact on the upper part of the PDGDE. Also, in a secure corridor application, small objects are typically carried around the upper body. For the lower body case, β_{l1} is greater than β_{l2} because the presence of the bags affects the shape and gait of the lower body when the bags are too close to the legs. The *K*-by-*L* DCT block size of the PDGDE is 10-by-10 for both *LC* and *UC*. Finally, a minimum distance classifier is employed to find the identity of a person in the gallery. If the algorithm identifies that no object is being carried, the subject's identity will be determined by the recognition algorithm for the non-carrying object case. The recognition algorithm for non-carrying object uses DGDE and PDGDE. Both DCT based gait image representations are used because of their robustness against noise and other gait invariant factors as discussed earlier. The difference between DGDE and PDGDE is DGDE includes the swing of hands but PDGDE removes them. The swing of both hands can sometimes be a useful feature, but it can also disturb

the accuracy of the gait recognition. Therefore, we decide to fuse both gait image representations for the non-carrying object recognition algorithm.

The five features applied for DGDE and PDGDE for the proposed non-carrying object recognition algorithm are: each pixel, mean of each row, mean of each column, standard deviation of each row and standard deviation of each column. For each feature, a minimum distance classifier is applied to identify the correct match. Therefore, ten matches of subjects in the gallery are generated using both DGDE and PDGDE gait image representations. Hence, the algorithm creates two sets, $\mathrm{M}_{\mathrm{DGDE}}$ and $\mathrm{M}_{\mathrm{PDGDE}}$, each consisting of five matches from the five features which generated from DGDE and PDGDE. Next the following equations are applied:

$$\begin{aligned} \tilde{\mathrm{m}}_{\mathrm{DGDE}} &= \arg mode[\mathrm{M}_{\mathrm{DGDE}}] \\ \tilde{\mathrm{m}}_{\mathrm{PDGDE}} &= \arg mode[\mathrm{M}_{\mathrm{PDGDE}}] \\ \mathrm{m}_{\mathrm{DGDE}} &= mode[\mathrm{M}_{\mathrm{DGDE}}] \\ \mathrm{m}_{\mathrm{PDGDE}} &= mode[\mathrm{M}_{\mathrm{PDGDE}}] \end{aligned} \tag{11}$$

The decision on which classifier to use (probability distribution (PD) or Hidden Markov Model (HMM)) is based on the following:

$$\begin{aligned} \acute{\mathrm{M}}_S = \tilde{\mathrm{m}}_{\mathrm{DGDE}}, if\ \tilde{\mathrm{m}}_{\mathrm{DGDE}} = \tilde{\mathrm{m}}_{\mathrm{PDGDE}}\ or\ \mathrm{m}_{\mathrm{DGDE}} = 5 \\ \acute{\mathrm{M}}_S = \tilde{\mathrm{m}}_{\mathrm{PDGDE}}, if\ \ \mathrm{m}_{\mathrm{PDGDE}} = 5 \\ Use\ \mathrm{PD}, if\ \mathrm{m}_{\mathrm{DGDE}} > \mathrm{m}_{\mathrm{PDGDE}} \\ Use\ \mathrm{HMM}, otherwise \end{aligned} \tag{12}$$

If $\mathrm{m}_{\mathrm{DGDE}} > \mathrm{m}_{\mathrm{PDGDE}}$, it shows that there is little motion of the body; otherwise it indicates large motion of body. The relative motion of the body is with respect to the gait image representation in the gallery. The reasons for selecting PD and HMM in (12) are: (i) PD is based on the similarity of the probability distribution between the respective columns in the gallery and probe; (ii) the HMM classifier observes the similarity of probability distribution not only in the respective columns, but also in the adjacent columns in the gallery and the probe.

The PD uses Gaussian density distribution to estimate the similarity between gallery and probe of each column of GDE. First the following probabilities are calculated:

$$P\left(C_{k,x} \mid \mathcal{Z}_{\mathrm{GDE}}(x,y)_j\right) = \frac{P\left(\mathcal{Z}_{\mathrm{GDE}}(x,y)_j \mid \acute{\mathrm{C}}_{k,x}\right) \times \acute{\omega}_1}{P\left(\mathcal{Z}_{\mathrm{GDE}}(x,y)_j\right)} \tag{13}$$

$$P\left(\neg C_{k,x} \mid \mathcal{Z}_{\mathrm{GDE}}(x,y)_j\right) = \frac{P\left(\mathcal{Z}_{\mathrm{GDE}}(x,y)_j \mid \acute{\mathrm{C}}_{j,x}\right) \times \acute{\omega}_2}{P\left(\mathcal{Z}_{\mathrm{GDE}}(x,y)_j\right)} \tag{14}$$

where $y = (0.1905 \times h \ldots\ 0.3714 \times h)$. The range of y is the area approximately starting from the shoulders to the end of the chest or elbow. This area is chosen because it has been identified as the least dynamic area in gait motion. $P\left(C_{k,x} \mid \mathcal{Z}_{\mathrm{GDE}}(x,y)_j\right)$ is the probability of a class for column x in the k^{th} subject in the gallery

given the pixel value $\mathcal{Z}_{\mathrm{GDE}}(x,y)$ of the GDE image of the probe j. $P\left(\neg C_{k,x} \mid \mathcal{Z}_{\mathrm{GDE}}(x,y)_j\right)$ is the probability of $\mathcal{Z}_{\mathrm{GDE}}(x,y)_j$ not being in the $\acute{C}_{k,x}$ column class. Equations (13) and (14) are based on the Bayes Decision Theory, $\acute{\omega}_1$ and $\acute{\omega}_2$ are the prior probabilities which are empirically identified, and $\acute{\omega}_1 + \acute{\omega}_2 = 1$. Another condition is $\acute{\omega}_1 >> \acute{\omega}_2$; this condition is helpful when noise occurs on any pixel in a column of GDE. $P(\mathcal{Z}_{\mathrm{GDE}}(x,y)_j)$ is the sum of numerators in (13) and (14). Equations (13) and (14) are computed based on the Gaussian probability distribution. Next, the number of pixels belonging to each subject in the gallery is counted, determined by the following equation:

$$\begin{aligned} &\acute{\mathcal{C}}(x,y)_k = 1; \\ &if\ P\left(C_{k,x} \mid \mathcal{Z}_{\mathrm{GDE}}(x,y)_j\right) > P\left(\neg C_{k,x} \mid \mathcal{Z}_{\mathrm{GDE}}(x,y)_j\right) \\ &else\ \ \acute{\mathcal{C}}(x,y)_k = 0 \end{aligned} \qquad (15)$$

The matched subject $\acute{\mathrm{M}}_S$ in the gallery is based on the following formula:

$$\acute{\mathrm{M}}_S = \underset{k\in\{1\ldots\mathbb{N}\}}{\arg max}\left[\sum_{x=1}^{n}\sum_{y=0.1905\times h}^{0.3714\times h}\acute{\mathcal{C}}(x,y)_k\right] \qquad (16)$$

where $\mathbb{N}$ is the last subject in the gallery.

On the other hand, if $\mathfrak{y}_{\mathrm{DGDE}} \leq \mathfrak{y}_{\mathrm{PDGDE}}$, HMM is used to find the $\acute{\mathrm{M}}_S$ in the gallery. The HMM is characterized as the finite set of hidden states, $S = \{s_1, s_2 \ldots s_N\}$ and a set of parameters $\Theta = \{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi}\}$ [52]. The transition matrix $\boldsymbol{A} = \{a_{ij}, 1 \leq i, j \leq N_s\}$ represents the transition probability of going from state i to state j with $a_{ij} \geq 0$ and $\sum_{j=1}^{N_s} a_{ij} = 1$ where N_s is the number of states. The emission parameter $\boldsymbol{B} = \{b(o|s_j)\}$ indicates the probability of observation o, when the system state is s_j. In this paper the continuous HMM with Gaussian density is used. Hence $b(o|s_j)$ is represented as [52]:

$$b(o|s_j) = \mathcal{N}(o|\mu_j, \sigma_j) \qquad (17)$$

where $\mathcal{N}(o|\mu_j, \sigma_j)$ denotes the Gaussian density at o. $\boldsymbol{\pi} = \{\pi_i\}$, the initial state probability distribution, represents the probabilities of initial states with $\pi_i \geq 0$ and $\sum_{i=1}^{N} \pi_i = 1$.

In our problem, the HMM is implemented based on the idea that a depth pixel value in any position of a column can sometimes stray/shift into neighboring columns. This is due to misalignment of the gait image representation, noise, motion of the body and clothes, and other inherent factors of gait. Hence, the states are a column and its neighboring columns. Therefore, there are 2 states for the pixels at the first and last columns and 3 states for those at the columns between the first and last columns. Hence, the shift of a depth value between one column to the neighboring columns can occur horizontally within the same row or in different rows. The shift of a depth value may occur vertically within a column. In this case, it does not change the probability of the state of a state or column. Hence this does not affect the accuracy of the gait recognition.

Fig. 9 shows the proposed ergodic 2-state and 3-state HMM models applied in this work. Since we have

limited training data, the transition probabilities a_{ij} were identified using the two normal walks that produce the best accuracies.

The transition matrix *A* for both the 2-state model and 3-state model are as follows:

$$\boldsymbol{A_{2s}} = \begin{bmatrix} a_{11} = 0.97 & a_{12} = 0.03 \\ a_{21} = 0.97 & a_{22} = 0.03 \end{bmatrix} \quad (18)$$

$$\boldsymbol{A_{3s}} = \begin{bmatrix} a_{11} = 0.97 & a_{12} = 0.015 & a_{13} = 0.015 \\ a_{21} = 0.97 & a_{22} = 0.015 & a_{23} = 0.015 \\ a_{31} = 0.97 & a_{32} = 0.015 & a_{33} = 0.015 \end{bmatrix} \quad (19)$$

where $\boldsymbol{A_{2s}}$ and $\boldsymbol{A_{3s}}$ are the transition matrices of the 2- and 3- state models respectively. For $\boldsymbol{A_{2s}}$, the transition probabilities from state 1 are higher than in $\boldsymbol{A_{3s}}$ because of the dynamic attribute of the leftmost and rightmost columns of GDE. In this work, S_1 is always the column in which pixels are being observed. The initial state probabilities π_i are the elements of the vector $\boldsymbol{\pi}$ and the probabilities are identified empirically based on the two normal walks which produce the best gait recognition accuracy. The initial state probabilities are stated in the following equations:

$$\boldsymbol{\pi_{2s}} = \{\pi_1 = 0.97, \pi_2 = 0.03\} \quad (20)$$

$$\boldsymbol{\pi_{3s}} = \{\pi_1 = 0.97, \pi_2 = 0.015, \pi_3 = 0.015\} \quad (21)$$

where $\boldsymbol{\pi_{2s}}$ and $\boldsymbol{\pi_{3s}}$ are the $\boldsymbol{\pi}$ for 2- and 3-states respectively.

In this work, the recursive Viterbi algorithm is applied to find the optimal state sequence and its Viterbi probability score for each observed column. The total Viterbi probability score of the optimal state sequences in all columns, P_k^* is computed as follows:

$$P_k^* = \sum_{x=1}^{W} p_T^*(x) \quad (22)$$

where $p_T^*(x)$ is the Viterbi probability score of the optimal state sequence in a column *x*. Therefore, the matched subject $\mathfrak{M}$ is computed as follows:

$$\mathfrak{M} = \underset{k \in \{\mathfrak{m}_{\mathrm{EDGDE}},\ \mathfrak{m}_{\mathrm{EPDGDE}}\}}{\arg\max} (P_k^*) \quad (23)$$

IV. Experimental results & discussion

In this section, we first discuss the parameters used in the proposed methods and how they can be applied with different ToF camera settings. Then the experimental results for the proposed algorithm are presented and discussed.

A. Parameters Settings

The first stage of the proposed method involves silhouette extraction. The ϒ value is proportional to the size of the silhouette. Hence, bigger silhouettes require bigger ϒ values. Other parameters for this algorithm can be tuned based on the quality of images in the gallery.

In the gait cycle frames selection, the τ value is used to identify the last frame (where the subject is closest to the camera) so that the camera can capture the entire body silhouette. This value is identified based on the average of the last frame that contains the complete silhouette from top of the head to the feet of the tallest subject in the gallery.

The *K-by-L* block size for applying DCT to GDE and PGDE is 10×10. This is not application-dependent. The sizes of GDE and PGDE are 105×54 and 105×32 respectively. The *K* and *L* values are proportional to the sizes of GDE and PGDE.

Another parameter called α is used to identify the starting position of the indiscriminate features around the shin area. In this experiment, the α values applied are identified by using two normal walks. This is carried out because of the limited training data available. Hence at least two sets of galleries are required. The best *K*, *L* and α values are the ones producing the highest matching accuracy between the two galleries. A simple direct matching algorithm such as in (24) and (25) can be used [53].

$$D(k)(x,y) = \left|I^1_{G*}(x,y) - I^2_{G*}(k)(x,y)\right| \; for \; k = 1 \ldots ₦ \qquad (24)$$

$$R = \arg min\left(\sum_{x=1}^{w_{G1}} \sum_{y=1}^{G1} D(k)(x,y)\right) for \; k = 1 \ldots ₦ \qquad (25)$$

where I^1_{G*} is a gait depth image representation from the first gallery, $I^1_{G*}(k)$ is the gait depth image representation of the k^{th} subject in the second gallery, w_G and h_G are the width and height of the gait image representation and R is the matched subject in the first gallery.

The identification of carrying objects for the upper and lower body cases involves two experimental parameters, C_L and C_U. These are identified based on small objects carried by individuals. The small objects are selected for secure corridor applications. β_{u1}, β_{u2}, β_{l1} and β_{l2} can be identified based on the training data.

Similarly, $\boldsymbol{A}_{2s}$, $\boldsymbol{A}_{3s}$, $\boldsymbol{\pi}_{2s}$ and $\boldsymbol{\pi}_{3s}$ can be identified by means of HMM training with data related to the non-carrying object covariates. These parameters can be tuned based on subjects' walking speed. In the environments where people walk much faster or slower than their normal walk, the values of these parameters can be increased.

As discussed, in different environments and subjects, the values of the parameters may differ, but if the same aforementioned procedures are carried out based on the training data, it will produce similar results as presented in part B of Section IV

B. Experimental Results

In this work, ten experiments were carried on the proposed method and compared with four existing methods: Frontal View Gait Energy Image (FVGEI) [1], Frontal View Gait Entropy Image (FVGEnI) [1], Gait Energy Volume (GEV) [21] and Robust Frontal Gait Recognition (RFGR) [54]. All the methods are evaluated using Rank 1 and Rank 5 which are the key performance indicators that measure the accuracy and robustness of the algorithms. The gallery is one of the normal walks captured in the earlier of the two recording sessions. The silhouettes used to generate FVGEI and FVGEnI are produced by converting depth silhouettes to binary silhouettes.

Table I summarizes the results of the proposed methods and the four existing methods. As seen in Table I, our proposed method outperforms all the other methods in Rank 1 and Rank 5 for all covariates. The proposed method achieves perfect recognition (100%) for the normal walk experiment. All the other methods also produce good results on normal walk, except GEV and RFGR. GEV, which is based on binary voxel volume, also produces poor results on other covariates. This is because the construction of the binary voxel volume is highly sensitive to depth information, so noise causes severe misalignment of the voxel volume over one gait cycle. In [23] GEV achieved similar results (20% Rank 1 accuracy) for normal walk. On the other hand, RFGR which is based on Histogram of Oriented Gradient (HOG) produces slightly better result than GEV. However, the HOG reduces the depth features without considering whether they are discriminating or non-discriminating features, hence reducing the overall accuracy of RFGR.

The methods proposed in [1], FVGEI and FVGEnI, use the average of binary silhouettes over one gait cycle. These representations only contain information on the 2D shape and 2D contour motion of the body. However, our proposed representations use frontal depth information as the feature. 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 ToF camera. Another reason for the poorer performance of FVGEI and FVGEnI is the PCA-MDA classifier used by both methods. The problem with the PCA-MDA method is that the dimensionalities of the feature space of all the gait image representations in this experiment are much larger than the class or size of the gallery (which is 46). The feature space is the total number of pixels in the gait image (5670) for both FVGEI and FVGEnI. MDA fails when faced with this problem.
To overcome this problem, PCA is applied first before MDA, reducing the dimension to class-1 which is 45 for these experiments. This process removes the discriminant information in the feature space especially when the dimension reduction is enormous and/or the discriminant information is compacted within a small feature space. This fact is proved by replacing PCA-MDA with a pixel by pixel matching classifier in [53] (as shown in (24) and (25)) which does not reduce the feature space. The pixel by pixel matching classifier computes the sum of differences between each pixel in the probe and gallery images. Then a minimum

distance classifier is used to identify the match in the gallery. The overall Rank 1 results for FVGEI and FVGEnI using pixel matching are 40.0% and 38.6% respectively. These are slightly higher than the Rank 1 PCA-MDA versions of FVGEI and FVGEnI which are 35.7% and 33.8% respectively.

None of the gait image representations produce good results on the time-based covariate. This may be due to the change of clothing and footwear worn by the subjects between the two sessions. This suggestion is supported by the findings in [1] which shows that when the subjects wear consistent clothing and do not wear footwear, this produces significant improvement in the recognition results. However, in most cases people wear different types of clothes and shoes over a period of time (for example, in summer versus winter), and in most places people use footwear. Therefore, it is better not to restrict the clothing and footwear for the time covariate experiment. Other factors might be subjects' weight change or psychological state (eg. mood). We also note that the lateral view methods in [13], [14], [53], [55] - [62] yielded significant drops in the time covariate experiments.

A further evaluation of the impact of the covariates on our method uses the Cumulative Mean Score (CMS). CMS for the overall and each experiment for up to Rank 10 is presented in Fig. 10. The measurement is only made up to Rank 10 which reflects the not too difficult gait patterns where the proposed method can be improved in future. As can be observed, all the time covariate experiments produced lower accuracies than the overall accuracy. Based on the graph in Fig. 10, the bag & time covariates are the most challenging experiments. This may be caused by the weights of both bags that change the walking patterns. Another factor is the presence of both bags that closed or touched the legs of the subjects which can alter the shape of the silhouette and can also produce false depth information. For the non-carrying object covariates the normal walk has the least impact on the proposed algorithm followed by the fast and slow walks for both time and non-time-based covariate experiments.

Normal and fast walk covariates achieve almost the same accuracy with 100% score at Rank 1 and Rank 2 respectively. The slow walk covariate achieved 100% accuracy only at Rank 6. Similar trends are also shown by the time-based covariates.

For the time-based covariates, the overall patterns show that the normal walk is the best, followed by the fast and slow walks. The table also shows the matching accuracy of gait image representations after the indiscriminate area below the shin is removed by the α parameter. Different α values are shown in Table II. For almost all covariates, all the gait image representations show an improvement after the removal of disturbing features below the shin.

From Table II, it can be observed that PDGDE with features removed by α is the best gait image representation followed by

PGDE with features removed for all four carrying object covariates. It shows that removing the left and right

side of the silhouettes in one gait cycle can reduce the effect of carrying objects on gait recognition. Table II also shows that GDE with the area below the shin removed is the best feature for normal, slow and fast walks. PDGDE with features removed by α is the best in Rank 5 and achieved 100% accuracy. This is the reason for the selection of these 3 gait image representations for the fusion part of the non-carrying object algorithm.

In addition to the analyses that have been presented, we also analysed the impact of gender, age and parameters. Since the proposed method is more suitable and significant for non-time based covariates, the analysis focused on those five covariates only. Fig. 11 shows the comparison of accuracies between male and female individuals on the proposed method. In this study, walking sequences for 26 males and 20 females were recorded. Overall, the proposed algorithm identifies female individuals with 89% accuracy, which was slightly better than the recognition rate for males of 85.4%. The female individuals are easier to be identified than male individuals on slow walk, carrying bags and carrying ball covariates. On the fast covariate, the proposed method performs better on male individuals than female individuals. However, for normal walk the proposed method produces the same accuracy (100%) for both genders.

Table III shows the impact of age on the accuracy of the algorithm. Overall, the ranges of ages between 30-34 and 35-39 show the lowest gait recognition accuracy with both of them scoring only 80%. Hence, it shows that people in these age ranges are more difficult to identify. On the other hand, people in the younger age range (19-29 years) have more consistent gait patterns and are easier to be recognized.

Table IV presents the influence of weight on the proposed algorithm. Overall, the people with medium weight have the most reliable gait pattern followed by light and heavy people. From Table V, we can conclude that short individuals are least

affected by carrying object covariates. However, for non-carrying object covariate, the group of medium height people are easier to be recognized.

In order to examine the impact of a different gallery on the proposed algorithm, we swapped the normal walk gallery with normal walk probe. Fig. 12 shows the results. In this experiment, all the parameters remain the same. For all covariates, the proposed algorithm performs slightly better when using gallery 1 compared to gallery 2. However, for carrying ball covariate, the accuracy improves from 78.3% to 82.6%. Hence, the proposed algorithm can be improved in the future to overcome this problem. However, overall, the proposed algorithm achieved 87.0% and 82.2% when using galleries 1 and 2 respectively. This shows that using different gallery does have an impact on the proposed algorithm.

In this experiment, the gait image sequences were captured in two session, May 2013 (S1) and December 2013 (S2). Both sessions were conducted in the same room. However, the settings were not strictly the same: there were slight differences in the position and angle of the camera, the position of areas A, B, C and D, and the positions of items such as tables and chairs. Also, in S2, the subjects were not restricted to wear the same

types of clothes and shoes. Furthermore, some of the subjects had lost or gained weight and the mood of the subjects might not be the same. Hence, to test the effect of the α parameter on the slight change of environment and subjects' physiological and psychological factors, we compared the results of non-time covariate experiment in both sessions. The results are shown in Table VI. Although the α parameters were identified based on the dataset captured in S1, when they are used on non-time covariate experiment using S2 dataset, they produced better results on GDE, DGDE and PDGDE. Only the results for PDGDE for S2 is slightly lower (74.6%) than S1 (76.5%). Therefore, there is no significant impact of the α on the change of the environment and individuals' physiological and psychological factors.

In addition to the accuracies of the proposed and previous methods, a comparison of the computational cost of the proposed algorithm and the previous methods was also carried out. The results are presented in Table VII. The analysis was carried out using Matlab 2013a software with the following computer specifications:

- Computer System: Laptop Computer, Toshiba Satellite C640
- Microprocessor: Intel Pentium CPU 8940
- Microprocessor Clock Speed: 2.00GHz
- Random Access Memory (RAM): 4.00GB
- Operating System: Microsoft Windows 7 Ultimate (64-bit)

The computational cost analysis of the algorithms is based on the evaluation made by Guan *et al.* [62]. The algorithms are run 10 times and their maxima, minima, standard deviations and means are recorded. The running time of all the algorithms in Table VII are based on the Rank 1 accuracy. The proposed method has two main stages: the carrying objects and non-carrying object recognitions. The carrying objects, around both upper and lower, takes about 0.526s to recognise an individual if he/she carries an object. If the recognition process goes up to fusion of 10 features of DGDE and PDGDE, the proposed method takes about 0.82s. The next stage in the proposed method is either to apply PD or HMM to recognise individuals and this take about 13.47s and 28.01s respectively. Both of them are slow because they require probability-based computational methods. The average computational time for FVGEI and FVGEnI are 0.55s and 0.67s respectively. FVGEnI is a little slower than FVGEI because FVGEnI requires the mathematical operation of Shannon entropy. For RFGR, each silhouette needs to be segmented into three separate images before applying the HoG to these images. Hence, it takes longer than FVGEI and FVGEnI. GEV applies the binary voxel volume which requires high memory space. Hence, it takes more time for PCA-MDA to perform its operations. Therefore, the recognition process for GEV takes about 57.90s making it the slowest algorithm.

V. Conclusion

This paper presents a new framework for gait recognition using a 3D Time of Flight (ToF) camera. A new

data set was developed by capturing gait image sequences in two separate sessions with seven months between them. This enables experiments based on ten covariates: normal walk, slow walk, fast walk, carrying bag, carrying ball, normal walk & time, slow walk & time, fast walk & time, carrying bag & time, carrying ball & time. The paper also presents a four-part algorithm. The first part is a new human silhouette extraction algorithm which reduces the multiple reflection problem experienced by ToF cameras. The second part uses a new gait cycle algorithm to identify the gait cycle frames. For the third part, we developed four new 3D gait image representations: GDE, DGDE, PGDE and PDGDE. To improve performance, the features below the shin were removed. The final stage of the four-part algorithm is a novel Adaptive Multi-Stage Fusion Classifier. In experiments comparing the proposed method with four existing methods, the proposed method outperforms the previous methods overall and on all covariates for both Rank 1 and Rank 5 evaluation techniques. PDGDE contains the most suitable features for carrying objects covariates. This may be due to the removal of the left and right side of the silhouettes which reduces the impact of feature deviations caused by carrying an object. Although GDE is the best overall gait image representation for non-carrying object covariates, DGDE and PDGDE also produced similar or better accuracies than GDE on several non-carrying object covariates. This proves the significance of the fusion of features in the non-carrying object case. The time-based covariate affects the proposed algorithm significantly, just as it does existing methods. It is possible that the impact of changes over time may be more severe on 3D depth features than on 2D features. Therefore, future work could focus on combining both 2D and 3D features. However, our proposed method produced excellent results on the non-time based covariates.

Therefore, at this stage, we believe that the proposed approach is well suited to applications such as secure corridors in airport and train terminals. Finally, our gait data set may be suitable for research into gender classification and age and height estimation based on gait using a ToF camera.

REFERENCES

[1] 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, Apr. 2012.

[2] M. Murray, A. Drought and R. Kory, "Walking patterns of normal men," *Journal of Bones and Surgery*, vol. 46A. no. 2, pp. 335-360, Mar. 1964.

[3] G. Johansson, "Visual Perception of Biological Motion and a Model for its Analysis," *Perception and Psychophysics*, vol. 14, no. 2, pp. 201–211, 1973.

[4] S. V. Stevenage, M.S. Nixon, and K. Vince, "Visual Analysis of Gait as a Cue to Identity," *Applied Cognitive Psychology*, vol. 13, no. 6, pp. 513-26, Dec. 1999.

[5] M. S. Nixon, J. N. Carter, D. Cunado, P. S. Huang, and S.V. Stevenage, "Automatic gait recognition," in Unspecified Biometrics: Personal Identification in Networked Society, A. K. Jain, R. Bolle, and S. Pankanti, Eds., Boston:Kluwer Academic Publishers, pp. 231-250, 1999.

[6] S. D. Choudhury, and T. Tjahjadi, "Gait recognition based on shape and motion analysis of silhouette contours," *Computer Vision and Image Understanding*, vol. 117, no. 12, pp. 1770-1785, Dec. 2013.

[7] J. Zhang, J. Pu, C. Chen, and R. Fleischer, "Low-Resolution Gait Recognition," *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 40, no. 4, pp. 986-996, Aug. 2010.

[8] 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, Oct. 2004.

[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, Brussels Belgium, Sept. 11-14, 2011, pp. 3270-3273.
[10] 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, July 2009.
[11] 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, Jan. 2010.
[12] M. Goffredo, J. N. Carter and M. S. Nixon, "Front-view gait recognition," in IEEE 2nd International Conference on Biometrics: Theory, Applications and Systems, Arlington, Virginia, USA, Sep. 29 - Oct. 1, 2008, pp. 1-6.
[13] 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, Feb. 2006.
[14] K. Bashir, T. Xiang and S. Gong, "Gait recognition using gait entropy image," in 3rd International Conference on Crime Detection and Prevention, London, UK, Dec. 3, 2009, pp. 1-6.
[15] H. Aggarwal and D. Vishwakarma, "Covariate conscious approach for Gait recognition based upon Zernike moment invariants," *IEEE Trans. on Cognitive and Developmental Systems*, vol. PP. no. 99, pp.1-12, Jan. 2017.
[16] M. Deng, C. Wang, F. Cheng, and W. Zeng, "Fusion of spatial-temporal and kinematic features for gait recognition with deterministic learning," *Pattern Recognition*, vol. 67, pp. 186–200, July. 2017.
[17] F. M. Castro, M. J. Marín-Jiménez, N. Guil-Mata, and N. P. de la Blanca, "Automatic Learning of Gait Signatures for People Identification," in *Advances in Computational Intelligence*, vol. 10306, I. Rojas, G. Joya G., and A. Catala, Eds., Springer, Cham, 2017, pp. 257-270.
[18] M. Alotaibi and A. Mahmood, "Improved Gait recognition based on specialized deep convolutional neural networks," in IEEE Applied Imagery Pattern Recognition Workshop, Washington, DC, USA, Oct. 13-15, 2015, pp. 1-7.
[19] C. Yan, B. Zhang and F. Coenen, "Multi-attributes gait identification by convolutional neural networks," in 8th International Congress on Image and Signal Processing (CISP), Shenyang, China, Oct. 14-16, 2015, pp. 642-647.
[20] F.M. Castro, M.J. Marín-Jiménez, N.Guil-Mata, and R.Muñoz-Salinas, "Fisher Motion Descriptor for Multiview Gait Recognition," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 31, no.1, Jan. 2017.
[21] 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, Washington D.C. USA, Oct. 11-13, 2011, pp. 1-6.
[22] 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, Jan. 2014.
[23] P. Chattopadhyay, S. Sural, and J. Mukherjee, "Frontal Gait Recognition from Incomplete Sequences Using RGB-D Camera," *IEEE Trans. on Information Forensics and Security*, vol. 9, no. 11, pp. 1843-1856, Nov. 2014.
[24] C. D. Herrera, J. Kannala, and J. Heikkila, "Joint depth and color camera calibration with distortion correction," *IEEE Trans. on Pattern Analysis and Machine Intelligence*., vol. 34, no. 10, pp. 2058–2064, Oct. 2012.
[25] S. Schwarz, M. Sjostrom, and R. Olsson, "A Weighted Optimization Approach to Time-of-Flight Sensor Fusion", *IEEE Trans. on Image Processing*, vol. 23, no. 1, pp.214-225, Jan. 2014.
[26] B. Langmann B., K. Hartmann K. and O. Loffeld, "Depth camera technology comparison and performance evaluation," in 1st International Conference on Pattern Recognition Applications and Methods, Vilamoura, Algarve, Portugal, Feb. 6-8, 2012, pp. 438-444.
[27] M. A. Livingston, J. Sebastian, Z. Ai, and J. W. Decker, "Performance measurements for the Microsoft Kinect skeleton," in IEEE Virtual Reality Conference, Costa Mesa, California, USA, Mar. 4-8, 2012, pp.119-120.
[28] Q. Wang, G. Kurillo, F. Ofli and, R. Bajcsy, "Evaluation of Pose Tracking Accuracy in the First and Second Generations of Microsoft Kinect," in International Conference on Healthcare Informatics, Dallas, Texas, USA, Oct 21-23, 2015, pp. 380-389.
[29] Q. Zou, L. Ni, Q. Wang, Q. Li, S. Wang, "Robust Gait Recognition by Integrating Inertial and RGBD Sensors," *IEEE Trans. on Cybernetics*, vol. 48, no. 4, Apr.2018.
[30] M. Turk and A. Pentland, "Eigenfaces for Recognition," Journal of Cognitive Neuroscience, vol. 3, no. 1, 1991.
[31] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, July 1997.
[32] 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, Pacific Grove, CA, USA, Nov. 4-7, 2001, vol 1, pp. 834-838.
[33] D. Tahmoush and J. Silvious, "Radar micro-doppler for long range front-view gait recognition," in 3rd IEEE International Conference on Biometrics: Theory, Applications and Systems, Washington DC, USA, Sept 28-30,2009, pp. 1-6.
[34] M. Balazia and P. Sojka, "Learning robust features for gait recognition by Maximum Margin Criterion," in 23rd International Conference on Pattern Recognition, Cancun, Mexico, Dec. 4-6, 2016, pp. 901-906.
[35] CMU Graphics Lab, "Carnegie-Mellon Motion Capture (MoCap) Database," [Online]. Available: 2003, http://mocap.cs.cmu.edu. [Accessed: Aug. 7, 2017].
[36] T. Afendi, F.Kurugollu, D. Crookes and A. Bouridane, "A frontal view gait recognition based on 3D imaging using a time of flight camera," in Proceedings of 22nd European Signal Processing Conference (EUSIPCO), Lisbon, Portugal, Sept., 1-5, 2014, pp.2435-2439.
[37] Fotonic Inc. "Fotonic," [Online]. Available: http://www.fotonic.com/ content/Default.aspx. [Accessed: Apr. 25, 2013].

[38] F. Basso, A. Pretto and E. Menegatti, "Unsupervised intrinsic and extrinsic calibration of a camera-depth sensor couple," in Proceedings of IEEE International Conference on Robotics and Automation, Hong Kong, May 31 – June 7, 2014, pp. 6244-6249.
[39] A. Safaei and S. Fazli, "A new method in simultaneous estimation of Kinect-V2 sensor calibration using shuffled frog leaping algorithm," 2017 3in Prrd International Conference on Pattern Recognition and Image Analysis (IPRIA), Shahrekord, Iran, Apr. 19-20, 2017, pp. 211-215.
[40] M. Reynolds, J. Dobos, L. Peel, T. Weyrich, and G.J. Brostow, "Capturing Time-of-Flight data with confidence," in IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, USA, June 20-25, 2011, pp. 945-952.
[41] G. Diraco, A. Leone, and P. Siciliano, "Geodesic-based Human Posture Analysis by using a Single 3D TOF Camera," in IEEE Symposium on Industrial Electronics, Gdansk, Poland, June 27 – 30, 2011, pp. 1329-1334.
[42] P. L. Rosin, "Unimodal thresholding", *Pattern Recognition*, vol. 34, no.10, pp. 2083-2096, 2001.
[43] W.H. Tsai, "Moment-preserving thresholding: A new approach," *Computer Vision, Graphics and Image Processing*, vol. 29, no. 3, pp. 377-393, Mar. 1985.
[44] N. Otsu, "A threshold selection method from grey-level histograms," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 9, no.1, pp, 62-66, Jan. 1979.
[45] J.N. Kapur, P.K. Sahoo, and A.K.C. Wong, "A new method for grey-level picture thresholding using the entropy of the histogram," *Computer Vision, Graphics and Image Processing*, vol. 29, no. 3, pp. 273-285, Mar. 1985.
[46] T.W. Ridler, and S. Calvard, "Picture thresholding using an iterative selection method," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 8, no.8, pp. 629-632, Aug. 1978.
[47] 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, Iasi Romania, Jun 30-Jul 1, 2011, pp. 1-4.
[48] S. J. Miller. (2015, Feb.). The Method of Least Squares. Mathematics Department Brown University, Brown University, Providence, Rhode Island, USA, [Online]. Available: http:// web. williams. edu/ Mathematics/sjmiller/public_html/BrownClasses/54/handouts/MethodLeastSquares.pdf.
[49] J.H. Yoo, M.S. Nixon, and C.J. Harris, "Extracting human gait signatures based on body segment properties," in 5th IEEE Southwest Symposium on Image Analysis and Interpretation, Santa Fe, New Mexico, USA, April 7-9, 2002, pp. 35-39.
[50] C. Rougier, E. Auvinet, J. Meunier, M. Mignotte, and J. A. De Guise, "Depth Energy Image for gait symmetry quantification," in Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Boston, MA, USA, 30th Aug. - 3rd, Sept. 2011, pp. 5136–5139.
[51] A.K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, New Jersey, USA: Prentice Hall, 1989, pp.150-153.
[52] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," in Proccedings of IEEE, vol. 7, no.2, pp. 257-286, Feb. 1989.
[53] T.H.W. Lam, K.H. Cheung, J.H.K. Liu, "Gait flow image: A silhouette-based gait representation for human identification," *Pattern Recognition*, vol. 44, no. 4, pp. 973-987, Apr. 2011.
[54] B. R. Rowshan, C. N. Guerra, P. L. Correia and L. D. Soares, "Robust frontal gait recognition – merging viewpoints and depth ranges," in 3rd International Workshop on Biometrics and Forensics, Gjovik, Norway, Mar. 3-4, 2015, pp. 1-5.
[55] S. Sarkar, P.J. Phillips, Z. Liu, I.R. Vega, P. Grother, and K.W. Bowyer, "The Humanid Gait Challenge Problem: Data Sets, Performance, and Analysis," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 162- 177, Feb. 2005.
[56] N. V. Boulgouris, K. N. Plataniotis, and D. Hatzinakos, "Gait recognition using linear time normalization," *Pattern Recognition*, vol. 39, no. 5, pp. 969-979, May 2006.
[57] Z. Liu and S. Sarkar, "Improved gait recognition by gait dynamics normalization," *IEEE Trans. Pattern Analysis on Machine Intelligent*, vol. 28, no. 6, pp. 863–876, June 2006.
[58] T. Lam, R. Lee, D. Zhang, "Human gait recognition by the fusion of motion and static spatio-temporal templates," *Pattern Recognition*, vol. 40, no. 9, pp. 2563–2573, Sept. 2007.
[59] D. Tao, X. Li, X. Wu, and S. Maybank, "General tensor discriminant analysis and gabor features for gait recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.
[60] D. Xu, Y. Huang, Z. Zeng, and X. Xu, "Human gait recognition using patch distribution feature and locality-constrained group sparse representation," *IEEE Trans. Image Process*., vol. 21, no. 1, pp. 316–326, Jan. 2012.
[61] C. Wang, J. Zhang, L. Wang, J. Pu, and X. Yuan, "Human identification using temporal information preserving gait template," *IEEE Trans. on Pattern Analysis Machine Intelligence*, vol. 34, no. 11, pp. 2164–2176, Nov. 2012.
[62] Y. Guan, C. Li, and F. Roli, "On reducing the effect of covariate factors in gait recognition : a classifier ensemble method," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 37, no. 7, pp. 1521–1528, July, 2015.

FIGURES :

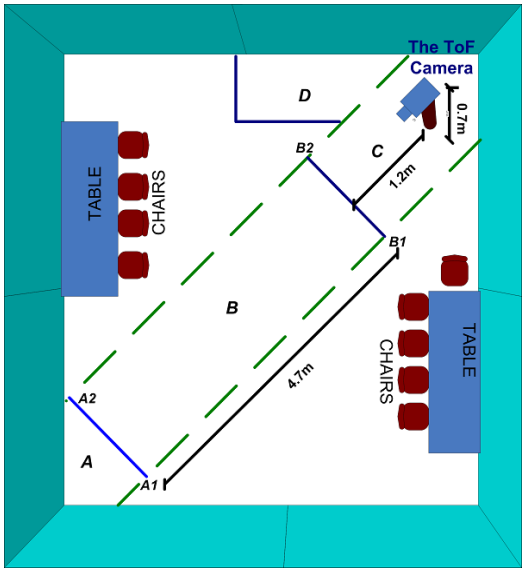


Fig. 1. Aerial view of the gait image sequence capture setup of the proposed frontal view gait data set using the ToF camera.

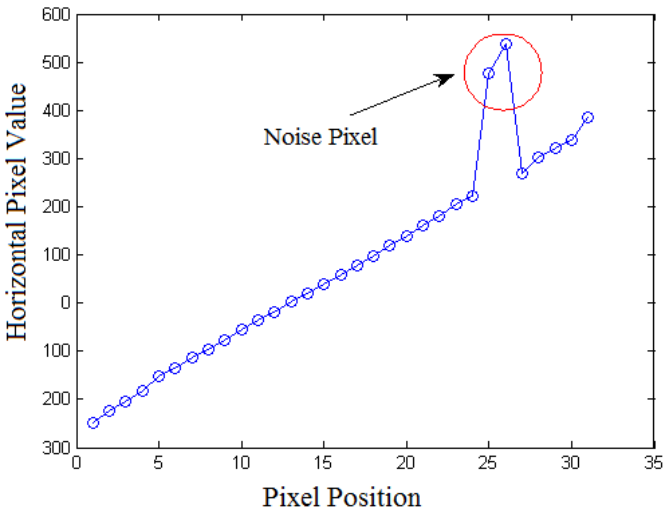


Fig 2. Example of one row of Human Silhouette (Horizontal Coordinates).

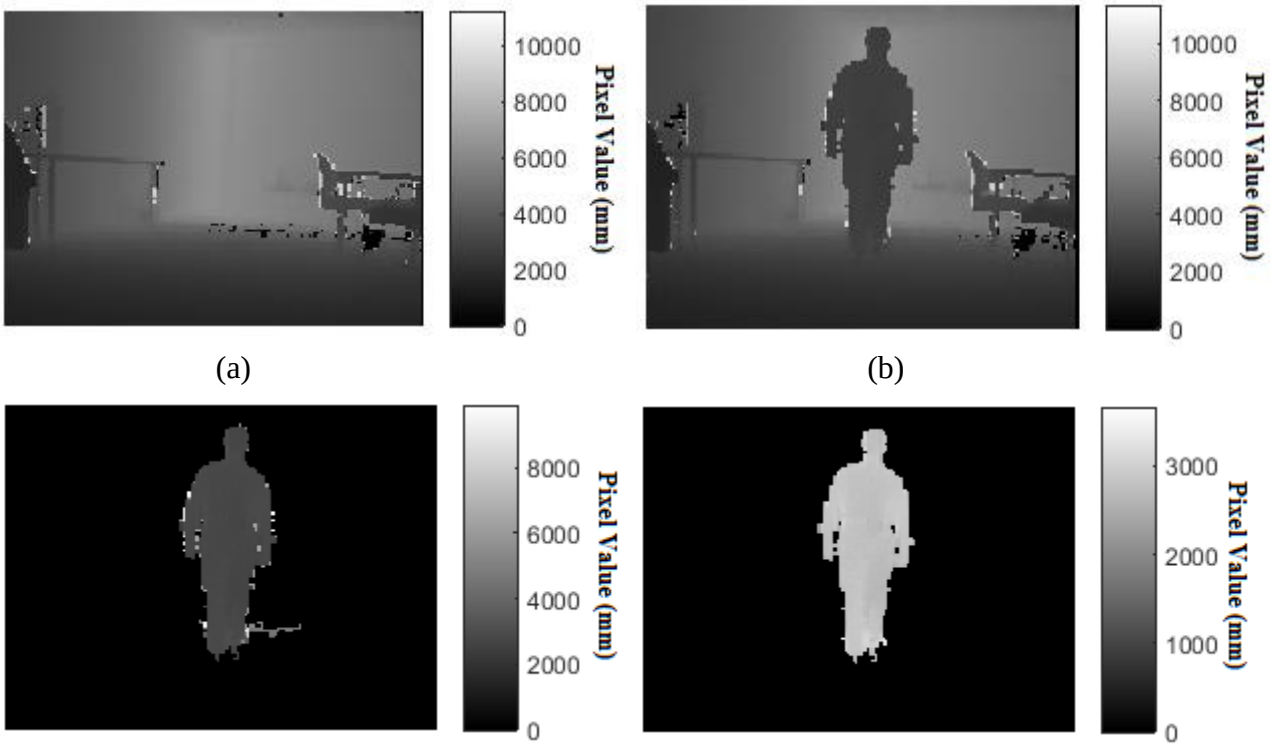


Fig. 3 (a) Background image (b) Foreground image (c) Image produced by Rosin's segmentation (d) Image enhanced by NRLSF.

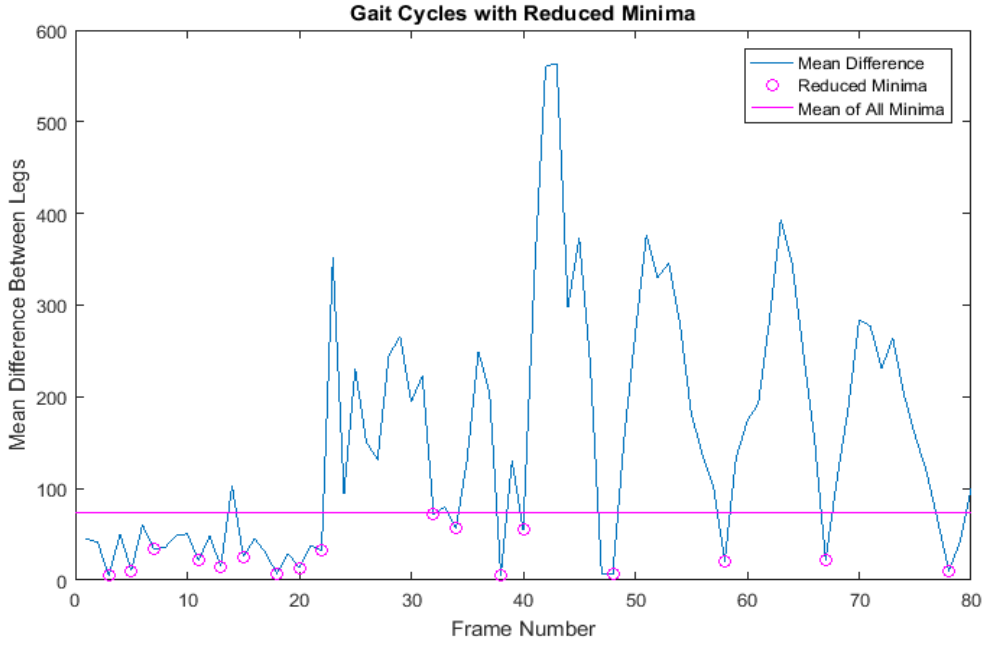
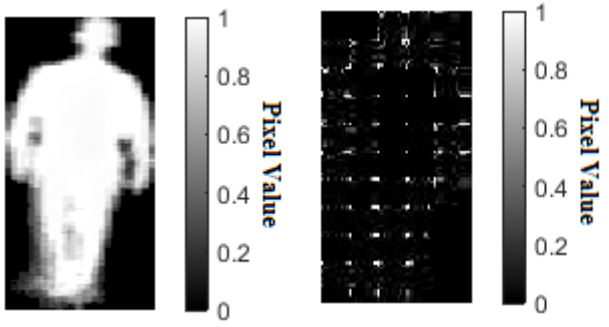
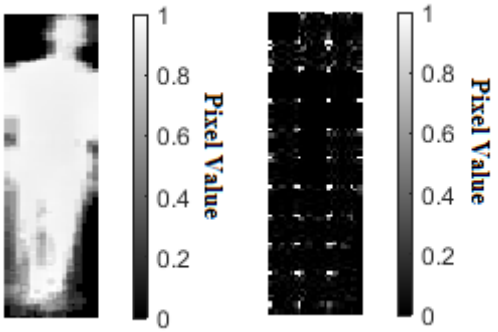


Fig 4. Gait cycle produced from the mean difference between the two legs.



(a) (b)

Fig. 5 Proposed gait full image representations: (a) GDE (b) DGDE.



(a) (b)

Fig. 6 Proposed gait partial image representations: (a) PGDE (b) PDGDE.

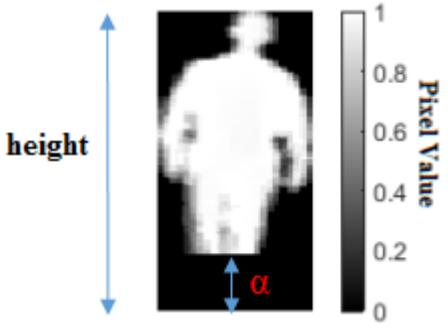


Fig. 7 The GDE removed by the α parameter.

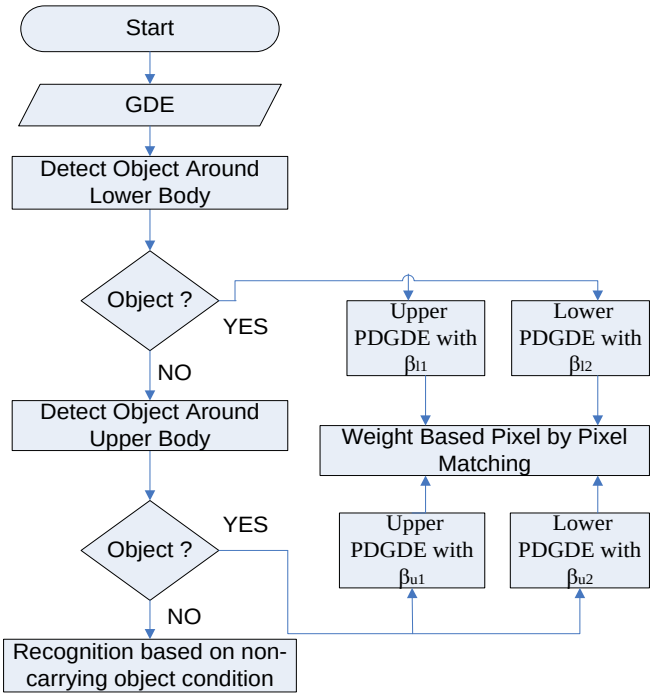
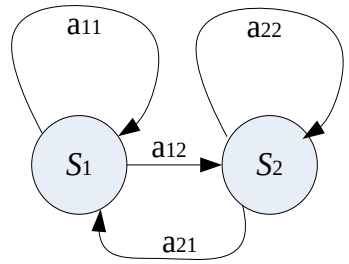
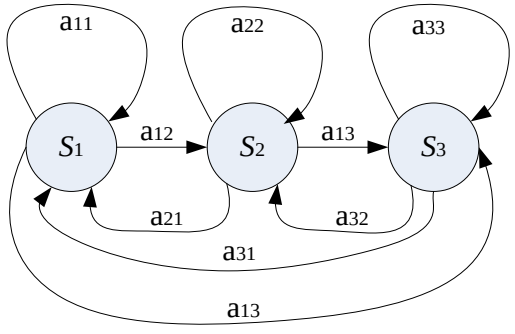


Fig. 8: Recognition algorithms for carrying objects.



(a)



(b)

Fig. 9 The proposed ergodic HMM model (a) 2-state (b) 3-state.

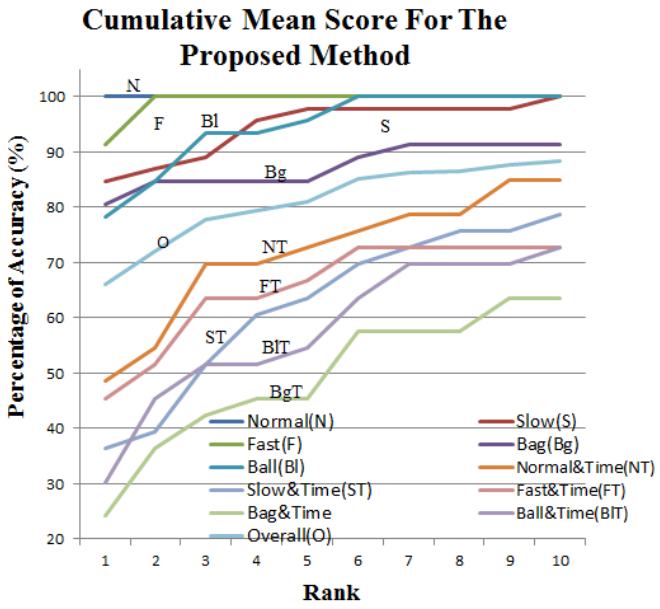


Fig. 10: The recognition rate of overall and each experiment based on Cumulative Mean Score (CMS) from Rank 1 to Rank 10.

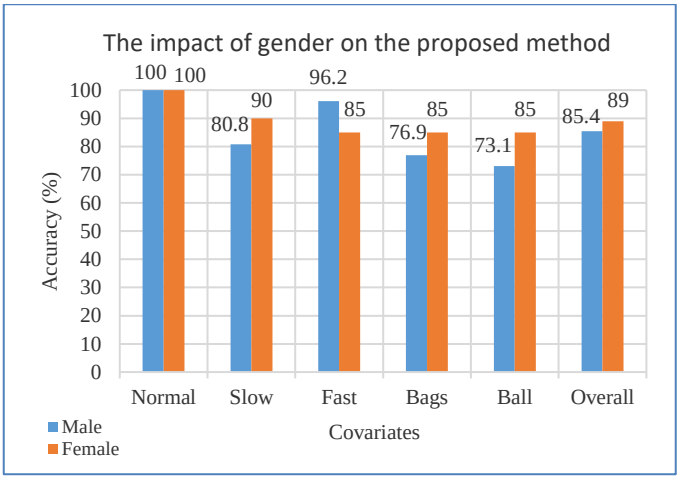


Fig. 11: The impact of gender on the proposed method.

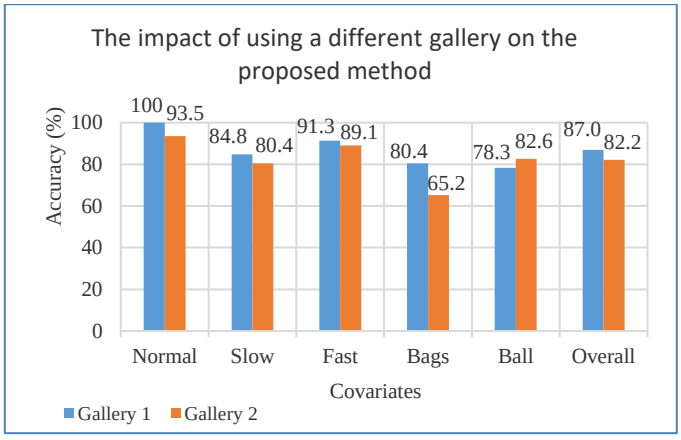


Fig. 12 The impact of using a different gallery on the proposed method.

TABLES:

TABLE I COMPARISON OF PERCENTAGE OF RECOGNITION PERFORMANCES OF THE PROPOSED METHOD AGAINST THE PREVIOUS METHODS

	FVGEI [1]		**FVGEnI [1]**		**GEV[21]**		**RFGR[54]**		**Proposed Method**	
	Rank 1 (%)	Rank 5 (%)	Rank 1 (%)	Rank 5 (%)	Rank 1 (%)	Rank 5 (%)	Rank 1 (%)	Rank 5 (%)	Rank 1 (%)	Rank 5 (%)
Normal	87.0	95.7	78.3	95.7	19.6	30.4	43.5	54.3	**100**	**100**
Slow	71.7	93.5	65.2	87.0	19.6	32.6	43.5	54.3	**84.8**	**97.8**
Fast	65.2	80.4	60.9	78.3	17.4	30.4	43.5	50.0	**91.3**	**100**
Carrying Bags	2.2	19.6	8.7	26.1	2.2	10.9	4.3	15.2	**80.4**	**84.8**
Carrying Ball	21.7	43.5	33.3	54.3	4.3	13.0	2.2	17.4	**78.3**	**95.7**
Normal & Time	24.2	51.5	33.3	51.5	12.1	32.6	21.2	45.5	**48.5**	**72.7**
Slow & Time	18.2	57.6	33.3	60.6	9.1	32.6	18.2	30.3	**36.4**	**63.6**
Fast & Time	27.3	48.5	18.2	39.4	12.1	34.8	24.2	42.4	**45.5**	**66.7**
Carrying Bags & Time	3.0	24.2	3.0	18.2	3.0	15.2	9.1	18.2	**24.2**	**45.5**
Carrying Ball & Time	9.1	45.5	18.2	30.3	3.0	17.4	3.0	21.2	**30.3**	**54.6**
Overall	35.7	57.7	33.8	56.5	10.7	24.7	22.3	35.4	**66.1**	**81.0**

TABLE II COMPARISON OF % RECOGNITION PERFORMANCE OF THE PROPOSED GAIT 3D DEPTH IMAGE REPRESENTATIONS USING PIXEL BY PIXEL MATCHING

	GDE		GDE with $\alpha = 0.19 \times h$		DGDE		DGDE with $\alpha = 0.11 \times h$		PGDE		PGDE with $\alpha = 0.2 \times h$		PDGDE		PDGDE with $\alpha = 0.22 \times h$	
	Rank 1	Rank 5	Rank 1	Rank 5	Rank 1	Rank 5	Rank 1	Rank 5	Rank 1	Rank 5	Rank 1	Rank 5	Rank 1	Rank 5	Rank 1	Rank 5
Normal	91.3	95.7	**93.5**	95.7	**93.5**	93.5	**93.5**	93.5	82.6	93.5	89.1	95.7	87.0	95.7	89.1	**100**
Slow	76.1	93.5	80.4	**97.8**	**82.6**	93.5	**82.6**	93.5	56.5	84.8	60.9	89.1	69.6	93.5	73.9	95.7
Fast	69.6	87.0	76.1	**95.7**	71.7	89.1	76.1	89.1	58.7	84.8	69.6	89.1	76.1	91.3	**87.0**	93.5
Carrying Bags	8.7	30.4	8.7	34.8	15.2	34.8	8.7	34.8	30.4	54.3	30.4	60.9	54.3	82.6	**58.7**	**89.1**
Carrying Ball	34.8	65.2	32.6	58.7	39.1	65.2	32.6	65.2	58.7	84.8	63.0	84.8	**73.9**	91.3	**73.9**	**95.7**
Normal & Time	27.3	60.6	36.4	**69.7**	33.3	63.6	**39.4**	63.6	24.2	54.5	24.2	63.6	27.3	66.7	27.3	63.6
Slow & Time	30.3	51.5	**33.3**	60.6	24.2	63.6	30.3	**63.6**	9.1	36.4	15.2	33.3	18.2	42.4	21.2	54.6
Fast & Time	27.3	51.5	33.3	**60.6**	33.3	57.6	**39.4**	57.6	18.2	48.5	27.3	54.5	15.2	57.6	24.2	54.6
Carrying Bags & Time	6.1	33.3	12.1	36.4	6.1	21.2	6.1	21.2	3.0	36.4	9.1	45.5	**24.2**	**51.5**	**24.2**	48.5
Carrying Ball & Time	9.1	27.3	15.2	27.3	15.2	30.3	12.1	30.3	12.1	45.5	24.2	51.5	**27.3**	54.5	18.2	**57.6**
Overall	41.0	62.0	44.8	65.8	44.6	63.8	44.8	63.5	39.0	65.3	44.8	69.6	51.4	75.7	**54.1**	**78.5**

TABLE III THE IMPACT OF AGE ON THE PROPOSED METHOD

	Rank 1 Accuracy (%)						
Age (Years)	**19-24**	**25-29**	**30-34**	**35-39**	**40-44**	**45-49**	**50-59**
Normal	100	100	100	100	100	100	100
Slow	100	88.9	80	80	71.4	75	100
Fast	87.5	100	80.0	100	85.7	100	10
Bags	87.5	88.9	70.0	60	100	75	66.7
Ball	100	77.8	70.0	60	85.7	75	66.7
Overall	**95.0**	**91.1**	**80.0**	**80.0**	**88.6**	**85**	**86.7**

TABLE IV THE IMPACT OF WEIGHT ON THE PROPOSED METHOD

	Rank 1 Accuracy (%)		
Weight (kg)	**40-59**	**60-79**	**80-114**
Normal	100	100	100
Slow	88.9	88.5	72.7
Fast	77.8	96.2	90.9
Bags	88.9	84.6	63.6
Ball	88.9	80.8	63.6
Overall	88.9	90.0	78.2

TABLE V THE IMPACT OF HEIGHT ON THE PROPOSED METHOD

	Rank 1 Accuracy (%)		
Height (m)	**1.50-1.64**	**1.65-1.74**	**1.75-1.89**
Normal	100	100	100
Slow	92.3	93.3	72.2
Fast	76.9	100.0	94.4
Bags	92.3	80.0	72.2
Ball	92.3	80.0	66.7
Overall	90.8	90.7	81.1

TABLE VI COMPARISON ON THE EFFECT OF α PARAMETER FOR SESSION 1 (S1) AND SESSION 2 (S2)

	GDE with $\alpha = 0.19 \times h$		DGDE with $\alpha = 0.11 \times h$		PGDE with $\alpha = 0.2 \times h$		PDGDE with $\alpha = 0.22 \times h$	
	S1	S2	S1	S2	S1	S2	S1	S2
Normal	93.5	93.9	93.5	97/0	89.1	87.9	89.1	90.9
Slow	80.4	75.8	82.6	75.8	60.9	66.7	73.9	66.7
Fast	76.1	81.8	76.1	87.9	69.6	69.7	87.0	72.7
Bags	8.7	6.1	8.7	12.1	30.4	48.5	58.7	66.7
Ball	32.6	60.6	32.6	57.6	63.0	69.7	73.9	75.8
Mean	**58.3**	**63.6**	**58.7**	**66.1**	**62.6**	**68.5**	**76.5**	**74.6**

TABLE VII COMPUTATIONAL COST ANALYSIS OF THE PROPOSED METHOD VERSUS THE PREVIOUS METHODS

	Algorithms			Maximum (s)	Minimum (s)	Standard Deviation (s)	Average (s)
Previous Methods	FVGEI [1]			0.61	0.53	0.023	0.55
	FVGEnI [1]			0.69	0.65	0.013	0.67
	GEV [21]			60.48	52.47	2.328	57.90
	RFGR [54]			9.61	9.47	0.042	9.50
Proposed Method	Carrying Objects			0.53[1]	0.52	0.003	0.53[2]
	Non-Carrying Object	Fusion of 10 features of DGDE & PDGDE		0.88	0.79	0.020	0.82
		Apply PD		13.54	13.39	0.050	13.47
		Apply HMM		28.21	27.90	0.070	28.01

[1] The actual measurement is 0.5321s before rounded up to the nearest hundredth
[2] The actual measurement is 0.5261s before rounded up to the nearest hundredth