

Gait Recognition using Time-of-Flight Sensor

Mohammad Omar Derawi¹, Hazem Ali², Faouzi Alaya Cheikh²

¹ Norwegian Information Security Laboratory, Gjøvik University College, Norway

² Media Technology Lab, Gjøvik University College, Norway

{mohammad.derawi,hazem.ali,faouzi.cheikh}@hig.no

Abstract: This paper develops a biometric gait recognition system based on 3D video acquired by a Time-of-Flight (ToF) sensor providing depth and intensity frames. A first step of the proposed gait analysis is the automatic extraction of the silhouette of the person via segmentation. The segmentation of the silhouette is performed on the depth frame which provide information which describes the distance from the camera of every pixel in the intensity frame. The range data is sensitive to noise thus we apply morphological filtering operations to enhance the segmented object and eliminate the background noise. The positions of the joint angles are estimated based on the splitting of the silhouette into several body segments, based on anatomical knowledge, and ellipse fitting. The resulting parameters from this analysis of the silhouette are used for feature extraction from each frame. The evolutions of these features in time are used to characterise the gait patterns of the test subjects. Finally, we do biometric performance evaluation for the whole system. To the best of our knowledge, this article is the first article that introduces biometric gait recognition based on ToF Sensor.

1 Introduction

The ability to use gait for people recognition and identification has been known for a long time. The earliest research started in the sixties of the twentieth century, where studies from medicine [MYN09] and psychology [Joh73] proved that human gait has discriminative patterns from which individuals can be identified. It is however just in the last decade that gait as a biometric feature has been introduced, and from a technical point of view gait recognition can be grouped in three different classes. Machine vision (MV) which uses video from one or more cameras, to capture gait data and video/image processing to extract its features. Floor sensors (FS), that use sensors installed in the floor, are able to measure gait features such as ground reaction forces and heel-to-toe ratio when a person walks on them. The third class uses wearable sensors (WS) where the gait data is collected using body-worn sensors.

MV based gait recognition is mainly used in surveillance and forensics applications [LSL08, HTWM04]. In MV image processing techniques are used to extract static like stride length which are determined by body geometry [BJ01], and dynamic features from body silhouettes. The MV based gait analysis techniques can be classified as model-based [BN07] and model free [HL10]. The main advantage of model based approaches is the direct extraction of gait signatures from model parameters, but it is computationally expensive. Model

free techniques characterize the body motion independently from body structure. MV gait analysis can also be categorized according to the technology used, as marker-based or marker-less. In marker based systems specific points in the subject's body are labelled by markers. By tracking these points in the video sequence the body motion can be tracked and analysed [Cut77, KN10]. MV based gait recognition provides wide range of gait features and many works utilized different sets of features and classification techniques. Benabdelkader et. al. [BCD02] used stride length and cadence as features extracted from 17 subjects' silhouettes walking in outdoor environment for 30 meters in a straight line at fixed speed to achieve EER of 11%, using linear regression for classification. Wang et. al. [WTNH03] utilized silhouette structure evolution over time to characterize gait, by calculating the silhouette centre and obtaining its contour they converted the 2D silhouette into 1D signal by calculating the distance between the centroid and every pixel on the contour. Principal component analysis (PCA) were used for dimensionality reduction of normalized distance signals using normalized Euclidean distance (NED) as similarity measure and nearest neighbour classifier with respect to class exemplars (ENN) classification approach. They achieved an EER of 20%, 13%, and 9% for 20 subjects filmed at 0, 45, and 90 degrees view respectively. The most related work to ours was done by He and Le [HL10], in which temporal leg angles was used as gait features for 4 walking styles slow, fast, incline and walking with a ball, on a running machine. They achieved wide range of CCR for the different walk styles using NN and ENN classification techniques. The best result for 9 subjects were in worst case 74,91% using NN for the shin parameters alone in fast walk and best case 100% using NN for merging thigh the shin parameters alone in ball walk, running the test over the whole CMU database 96.39 % was achieved for fast walk. Jensen et. al [JPL09] used ToF camera to analyse gait, in their work step and stride length, speed, cadence and angles of joints were extracted as extracted as gait features. They used model fitting technique to extract the joint angles. To the best of our knowledge, this article is the first article that introduces biometric gait recognition with the use of ToF Sensor.

2 Experiment Design

In order to verify the usefulness of the proposed system, we performed an individual gait identification experiment. In this section we will go through the different issues related to our experiment.

We used the Swiss ranger SR-4000 CW10 sensor by Mesa technologies [Mes] seen in Figure 1(a). The SR4000 is an optical imaging system housed in an anodized aluminium enclosure. The camera operates with 24 LED emitting infra-red in the 850nm range, it modulates the illumination light emitting diodes (LED) at modulation frequency of 15MHz. Range measurements are obtained at each pixel using the phase shift principle, with non-ambiguity range of 10 meters. The camera has USB port for data acquisition and supplied with software library for C and Matlab.

The subject's motion was filmed from the side by means of a ToF camera at 30 frames/sec, while the subject was walking on a track in front of the camera as shown in Figure 1(b).

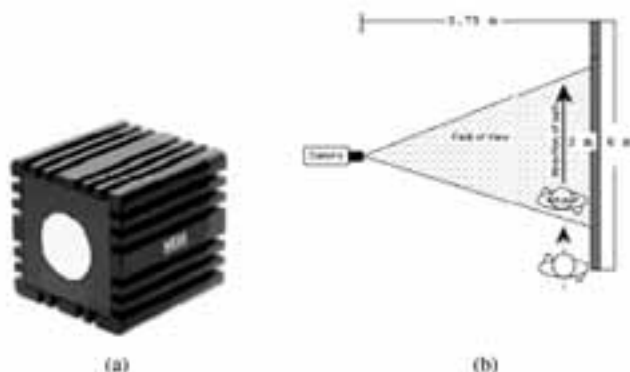


Figure 1: (a): SR-4000 ToF sensor, (b): set-up of the experiment

Due to the camera's narrow field of view, the length of the filmable track was limited to about 3 meters, therefore the subjects were asked to walk back and forth 10 times on a track longer than the camera's field of view's width. we used this longer track to allow recording the subject in full motion. To reduce the noise in the distance data, the camera was calibrated such that the image starts from the walking track in order to eliminate the reflection from the floor. The camera was put on a tripod at 0.7 meter from the floor and was tilted up about 5 degrees, as recommended by the camera manufacturer, see Figure 1(b).

The experiment was carried out on a solid surface in the lab. The subjects were asked to walk a fixed track in front of the camera. This fixed track allow the participants to walk for 1.5 to 2 gait cycles depending on the participants gait characteristics, and because some of the participants may not start precisely at the marker where the field of view of the camera starts. Each participant walks the track at least 5 times to extract one full gait cycle from each pass in front of the camera. The experiment procedure by the participant can be summarized in three steps to be repeated 5 times on average. First, the user *walks the track, turns around and walks the track back*.

The experiment was done in lab at Gjøvik University College. An invitation was sent to the students at the faculty to participate in the experiment, and 30 participants volunteered. They were of different age and height groups. The average age for the volunteers was 29.1 years, the average height was 176.9 cm. The participants were asked to wear the same type of shoes during the two sessions. In the experiment two sessions we collected data for the 30 subjects over a month, time gap between the two session varied from subject to subject, for a few subjects it was one week and for others about a month.

3 Feature Extraction

The image sequences of the subjects were acquired while walking in front of the camera. Followed by segmentation to extract the subjects body silhouette, morphological operations are applied to reduce background noise and fill holes in the extracted human silhouettes. Next, each of the enhanced human silhouettes is divided into six body segments based on human anatomical knowledge [NTT⁺09]. Ellipse fitting is applied to each of the six segments, and the orientation of each of the ellipses is used to calculate the orientation of each of the lower body parts for further analysis. The following steps are hereby described in more details:

Video segmentation is the process of partitioning a video spatially or temporally. It is an integral part of many video analysis and coding systems, including video indexing and retrieval, video coding, motion analysis and surveillance. In order to perform gait analysis of a person from image sequence, the subject needs to be extracted from the background of the video sequence. Image segmentation is used to separate foreground objects like people, from the background of the image sequence. Thresholding is the simplest image segmentation technique, in which each pixel of the original image is compared to a specified threshold, if the pixel's value is greater than the threshold value it is set as foreground pixel with value 1 if not it is set to zero as background pixel producing a binary image. In some complex images the operation can be iterated using two thresholds, in this case threshold works like band pass filtering. Histograms can be used to find the proper threshold values [Ots79], where peaks correspond to foreground objects are used to determine the threshold values. If the image's histogram shows no clear peaks, then, thresholding can not produce acceptable segmentation.

Morphological operations are shape based technique for processing of digital images [HSZ87]. Morphological operations are used to simplify image data preserving their main shape characteristics and eliminating irrelevant details. Morphological operations have two inputs the original image and structuring element to be applied to the input image, creating an output image of the same size. In a morphological operation, the value of each pixel in the output image is based on a comparison of the corresponding pixel in the input image with its neighbours. The shape and size of the structuring element constructs a morphological operation that is sensitive to specific shapes in the input image. The most basic morphological operations are dilation and erosion.

Ellipse fitting is used next to find the characteristics of the body parts. Having extracted body silhouette, the subject body are segmented into six parts [NTT⁺09] as illustrated in Figure 2. First, the centroid of the silhouette is determined by calculating its centre of mass. The area above the centroid is considered to be made of the upper body, head, neck and torso. The area below the centroid is considered made of the lower body, legs and feet. Next, one third of the upper body is divided into the head and neck. The remaining two thirds of the upper body are classified as the torso. The lower body is divided into two portions thighs and shins. Fitting an ellipse to each of the six body parts and finding their centres of mass, orientations, and major axes length we can characterize these body parts. Evolution of these parameters in the video sequence describes the human gait characteristics in time.



Figure 2: (a): body parts, (b): ellipse fitting model and (c) tracking of legs, blue ellipses for the leg closer to the camera

Leg tracking is performed next. Gait analysis requires reliable tracking of moving human body segments. As the human body is segmented into six parts, we need to track the lower limbs in the successive frames to construct meaningful output of the measured angles in the following step. To track the body parts along the image sequence we utilize the depth information acquired at each frame, we calculate the mean range values of each segment. The segment with higher mean range value belongs to the farthest away leg from the camera and vice versa, in Figure 2(c) a sample sequence of images with tracking results are shown.

Leg angles calculation is the last step of the features extraction. Human body is modelled as rigid segments connected by joints. The simplest model as 2D stick [GXT94], as in Figure (3-a). To extract the gait signatures we will mainly extract the thigh and shin angles from each frame of the video sequence to characterize the gait cycle. In this final step we extract the angles based on the data extracted from ellipse fitting to the body segments. The fitted ellipse parameters: orientation, major axis length (l_{major}) and centroid will be used to approximately calculate the start and end points coordinates for each of the body segments, such that each segment will be defined by two points in two dimensional space ((p_{i1}, p_{i2})) using equations ($x_1 = x_0 + l_{major} * \cos(\varphi)$), ($x_2 = x_0 - l_{major} * \cos(\varphi)$), ($y_1 = y_0 + l_{major} * \sin(\varphi)$) and ($y_2 = y_0 - l_{major} * \sin(\varphi)$) as shown in Figure (3-b). To calculate the angles of the leg segments, we reduced the impact of the non-precise

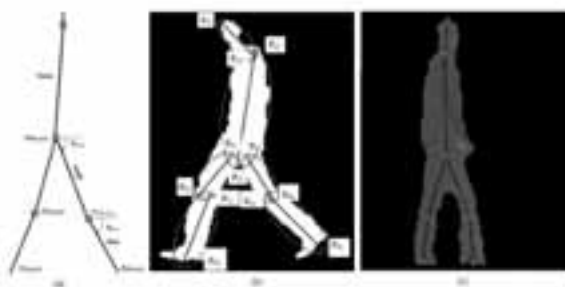


Figure 3: Illustration of joint locations, (a) 2D stick figure, (b) sample frame and (c) calculated angles.

fitting of the ellipses to the body segments, by estimating the location of the joints as the average location of the start point of one of the segments and the end point of the segment connected by the joint. The hip and knee joints' locations are calculated by the equation set ($p_1 = \frac{p_{22}+p_{21}+p_{21}}$), ($p_2 = \frac{p_{22}+p_{21}}$) and $p_3 = \frac{p_{21}+p_{21}}$ as shown in

Figure 3(c) The extracted features for each of the segments were calculated by equation $\theta = \arctan \frac{y_2 - y_1}{x_2 - x_1}$. We calculate the inclination angle of thigh and shin for each of the subject's legs to characterize the gait by the evolution of these angles in time at each image of the video sequence. Since the gait is quasi-periodic movement we extract a single gait cycle from each video sequence. The extracted feature are filtered using a local median filter to filter out the outliers. The outliers can be due to losing track of the legs, or bad ellipse fitting, in Figure 4 plots for 5 different gait cycle for one subject before and after filtering.

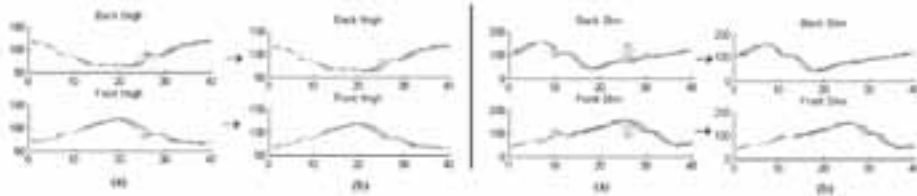


Figure 4: Extracted data for 5 different walking cycles,(a): original data, (b):filtered data.

4 Analysis and Results

As mentioned in earlier sections, each participant was filmed during two different sessions, and each session was downloaded as a separate file to the PC. Each of these files contained the data of 5 assessed gait data records, each again separated by leg types, i.e. front shin, back shin, front thigh and back thigh. In each file the data representing more than one gait cycle was manually aligned and cut such that one cycle started at a minimum and ended at a minimum value. This was done for all files. So the data from each collected file was split into 4 other files (for each leg type).

Using the above method the originally $30 \times 5 = 150$ collected files were split into $150 \times 4 = 600$ files, where each file contained the data of exactly one gait cycle for each leg type. Each file was labelled in such a way that participant, session number (1 or 2) and type of leg (front shin, back shin, front thigh, back thigh) are identifiable from the file name. Each file contained one column, representing the feature vector. The length of this feature vector varied indicating the length of gait cycle was varying from one participant to another. For the second session, 20 out of the 30 volunteers participated. This means that the performance evaluation over a certain time interval will only consist with 20 volunteers.

In our analysis in step 1, the noise was reduced by using the running median filter as described in the previous section. This resulted in having a filtered gait cycle representing our new feature vector. Since each file has dissimilar lengths we were unable to use distance metrics such as the Manhattan or Euclidean. Instead we applied a time series analysis named the Dynamic Time Warping (DTW) which is an algorithm for measuring similarity between two sequences which may vary in time or speed.

Several performance evaluations were calculated. Table 1 shows the performance of the first session with 30 subjects (second column) and the subset of 20 users (third column) who also participated at the second session. The first column indicates which template and test input were applied for performance testing. We notice that if we apply all the four types of legs as feature vector for one subject, we obtain a better EER than when applying them separately. This is due to the fact that more information is stored for a given subject. Since only a subset of users participated at the second session the EER has not changed significantly. With the performances for the second session we observe a significant change of the performance and the reason is that the users are more used to the walking in the second session and more comfortable with the experiment.

Template/Test	30 Participants	20 Participants - 1st	20 Participants - 2nd
Back thigh	8.42	7.48	4.72
Front thigh	7.39	6.62	6.02
Back shin	12.31	11.16	6.28
Front shin	11.32	10.24	9.41
All above	4.63	4.08	2.62

Table 1: EER Performance Results in % on the collected dataset. Second column is first session. Last column is session session

An interesting performance analysis is to investigate the change between the two session as can be seen in Table 2. We are analysing what will happen if we apply the first sessions data as training set and the second sessions data as test input. What we observe here is that the change over time becomes worse. Different shoe-type, clothes may have also an impact, and we realized that unfortunately not all participants came back for the second session.

Session 1	Session 2	Session 1 + Session 2
4.09	2.48	9.25

Table 2: EER Performance Results in % where session 1 as reference template and session 2 as test input (20 users).

5 Conclusion

In this paper we presented the first known results on gait recognition using the 3D ToF Sensor. When analysing the data we could already visually see that gait cycles were detectable for each subject which are dissimilar from others'. The experiment was performed over two different days (sessions) where each of the subjects (first session 30 subjects, second session 20 subjects) walked a track within the camera field of view. The best equal error rate obtained was 2.66 % for a separate session where the change over time we retrieve an equal error rate of about 9.25 %. Although the last mentioned result is not so low, this paper presents a first step towards a better performance in the future. Future work includes to work with multiple session over multiple days, and more cycles person to see the stability over time.

6 Acknowledgments

The authors would like to thank all the (anonymous) participants in this experiment. The writing of this article would not have been possible without their effort in the data collection phase.

References

- [BCD02] C. BenAbdelkader, R. Cutler, and L. Davis. Stride and cadence as a biometric in automatic person identification and verification. In *FG'02*, pages 372–377, 2002.
- [BJ01] A. F. Bobick and A. Y. Johnson. Gait recognition using static, activity-specific parameters. In *CVPR 2001*, volume 1, 2001.
- [BN07] I Bouchrika and M S Nixon. Model-Based Feature Extraction for Gait Analysis and Recognition. In *Mirage: Computer Vision ? Computer Graphics Collaboration Techniques and Applications*, volume 4418, pages 150–160, March 2007.
- [Cut77] & Kozlowski L. Cutting, J. Recognizing friends by their walk: Gait perception without familiarity cues. *Bull. Psychon. Soc.*, Vol. 9:pp. 353–356, 1977.
- [GXT94] Y. Guo, G. Xu, and S. Tsuji. Tracking Human Body Motion Based on a Stick Figure Model. 5:1–9, 1994.
- [HL10] Weihua He and Ping Li. Gait recognition using the temporal information of leg angles. In *ICCSIT 2010*, volume 5, pages 78–83, 2010.
- [HSZ87] Robert M. Haralick, Stanley R. Sternberg, and Xinhua Zhuang. Image Analysis Using Mathematical Morphology. *IEEE PAMI*, 9(4):532–550, july 1987.
- [HTWM04] Weiming Hu, Tieniu Tan, Liang Wang, and S. Maybank. A survey on visual surveillance of object motion and behaviors. *IEEE-J-SMCC*, 34(3):334–352, 2004.
- [Joh73] Gunnar Johansson. Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, 14:201–211, 1973.
- [JPL09] Rasmus Jensen, Rasmus Paulsen, and Rasmus Larsen. Analyzing Gait Using a Time-of-Flight Camera. In *Image Analysis*, volume 5575 of *LNCS*, pages 21–30. Springer Berlin / Heidelberg, 2009.
- [KN10] Kumar Amod Sohi B S Kumar Neelesh, Kunju Nissan. Active marker based kinematic and spatio-temporal gait measurement system using LabVIEW vision. *JSIR*, Vol.69(08):600–605, 2010.
- [LSL08] Peter K. Larsen, Erik B. Simonsen, and Niels Lynnerup. Gait Analysis in Forensic Medicine. *Journal of Forensic Sciences*, 53(5):1149–1153, 2008.
- [Mes] Mesa. <http://www.mesa-imaging.ch/>.
- [MYN09] A. Murai, K. Yamane, and Y. Nakamura. Characterization of motor skill based on musculoskeletal model. In *EMBC 2009*, pages 6542–6545, 2009.
- [NTT⁺09] Hu Ng, Wooi-Haw Tan, Hau-Lee Tong, Junaidi Abdullah, and Ryoichi Komiya. Extraction and Classification of Human Gait Features. In *Visual Informatics: Bridging Research and Practice*, volume 5857 of *LNCS*, pages 596–606. Springer Berlin Heidelberg, 2009.
- [Ots79] N. Otsu. A Threshold Selection Method from Gray-Level Histograms. *Systems, Man and Cybernetics, IEEE Transactions on*, 9(1):62–66, jan. 1979.
- [WTNH03] Liang Wang, Tieniu Tan, Huazhong Ning, and Weiming Hu. Silhouette analysis-based gait recognition for human identification. 25(12):1505–1518, 2003.