Motion–Based Gait Recognition for Recognizing People in Traditional Gulf Clothing

Mohammad Asif Towheed
Department of Computer Science and Engineering
American University of Sharjah
Sharjah, UAE
b00066375@aus.edu

Wasif Kiyani
Department of Computer Science and Engineering
American University of Sharjah
Sharjah, UAE
b00066791@aus.edu

Mumtaz Ummar
Department of Computer Science and Engineering
American University of Sharjah
Sharjah, UAE
g00067218@aus.edu

Tamer Shanableh
Department of Computer Science and Engineering
American University of Sharjah
Sharjah, UAE
tshanableh@aus.edu

Salam Dhou
Department of Computer Science and Engineering
American University of Sharjah
Sharjah, UAE
sdhou@aus.edu

Abstract— **Gait recognition is gaining popularity as it can recognize people in a non-intrusive and a non-contact manner. However, gait recognition is known for its susceptibility to clothing conditions. In this paper, we propose a solution specific to clothing conditions in the Gulf region where Abaya and Kandura are considered traditional clothing. The paper proposes a solution capable of training users based on traditional clothing and recognizing them in Western style clothing and vice-a-versa. The solution uses depth imaging, optical flow, accumulated motion and Discrete Cosine Transformation (DCT). Motion is calculated from consecutive images where the magnitudes and phases of motion vectors are accumulated into separate matrices. DCT and zonal coding is then applied to these matrices to form one concise feature vector that represents a walk. Experimental results, with 38 participants, showed that the proposed method is suitable for gait recognizing with such clothing constraints. The average classification accuracy is 88%. In comparison to an existing method, it is shown that the proposed method results in much more accurate recognition results yet at a higher computational cost.**

Keywords—Gait recognition, Computer Vision, Motion analysis, Image processing

I. INTRODUCTION

Every person has different biometrics and therefore biometrics are used to identify people. One category of biometrics is based on physical measures derived from fingerprint, iris and face. Another category includes behavioral characteristics like signature and gait. A desired feature of biometric is time-invariance, which means that the biometric representation does not change with time. Iris, fingerprint and to some extent gait, have this desired feature.

Identifying people by the way they walk, or gait recognition has other advantages. Users can be enrolled into a recognition system in a non-intrusive and non-contact manner. Additionally, full user collaboration is not required, and the gait can be captured from a short or a long distance. However, gait biometric on the other hand is susceptible to clothing, carrying conditions, camera viewpoint, walking speed and occlusion caused by foreground objects.

There are many applications of the gait biometric including authentication, which can be used in access-control, and recognition, which can be used in surveillance and security.

Gait recognition assumes that a full cycle of gait is available to the recognition system, however in real life this might not be the case because of occlusion and short walking distances. To solve this problem, authors in [1] proposed to create an incomplete Gait Energy Image (GEI) from fewer silhouettes of a person and use a deep auto-encoder to reconstruct a complete GEI from an incomplete GEI. Likewise, in [2] a solution was proposed that uses Constrained Fuzzy C-Means to divide a full gait cycle into a number of phases and using gait cycle analysis to join feature variables into one feature descriptor.

To address some of the shortfalls of gait recognition pertaining to clothing and carrying conditions, [3] proposed a gait recognition solution that uses multi-link gravity center track which is robust to dress and carrying conditions. The solution divides the person's silhouette into multiple links and calculates the gravity center track of each rank. Body regions are ranked prior to matching ranked silhouette regions. The view invariance of gait recognition was addressed by [4]. The solutions are based on robust skeleton points produced from a CNN network to produce a Skeleton Gait Energy Image (SGEI). Two different CNN- architectures are proposed for verification and identification of gait. To alleviate the walking speed problem in gait recognition, [5] proposed a gait feature extraction method that obtains accurate information even if the person is walking at different speeds. The solution is based on gait acceleration sampled from various body locations like the right side of pelvis, left thigh, wrist, left upper arm and right ankle. Other shortfalls of gait recognition are addressed by means of fusing it with other biometrics. For example, in [6], a multi biometric system is proposed which fuses the recognition results of gait, signature and iris.

Many research papers on gait recognition assume that silhouette images are readily available for the calculation of gait energy images. Such silhouette images are generated through segmentation, which if done automatically might not generate accurate results. This is particularly true if stereo vision is not used. To address this fact, a gait recognition system was proposed based on dense trajectories for feature extraction instead of action patterns [7].

In this paper, gait recognition is addressed specific to the Gulf region where clothing can vary between traditional and western style. The traditional clothing includes Abaya for females and Kandura for men. These cover the whole body

978-1-7281-5052-9/19/$31.00 ©2019 IEEE https://doi.org/10.1109/AICCSA47632.2019.9035284

and cover the gap between legs which is typically used as a feature in gait recognition. The gait recognition for the Gulf region was partially addressed in [8]. The authors proposed an image accumulated differences solution that captures the motion of the whole body. However, the training and recognition was based on traditional clothing only.

In this work, we extend existing solutions by collecting a new data set that mixes traditional Gulf clothing with Western style clothing. Supervised machine learning is used for recognition. The training phase will therefore contain a mix of clothing styles and the gait models will be generated accordingly. As such, a person can be recognized by the way s/he walks regardless of the clothing style. Additionally, the proposed system uses a depth camera to assist in accurately segmenting the subjects prior to feature extraction which is mainly based on motion information and optical flow.

The rest of this paper is organized as follows. Section II introduces the proposed feature extraction approach. Section III introduces the developed software for gait recognition. Section IV describes the data collection process. Section V describes the training and classification algorithm used. Section VI presents the experimental results and Section VII concludes the paper.

II. Feature Extraction

In this section, we introduce two approaches for feature extraction based on accumulated image differences and accumulated optical flow motion vectors. In both solutions, captured images are pre-processed for segmentation and foreground subtraction using the depth camera. The segmented images are stored in a grayscale format with a pixel range of [0, 255].

A. *The accumulated absolute differences solution*

In previous work, feature extraction algorithm was proposed for gait recognition based on accumulated absolute differences [8]. In this work, this method is adopted by applying it to the segmented images using the depth camera. This would improve the method by removing the constraint of requiring a static background during training and testing. In this work, this solution is expanded to include optical flow as proposed in the next section.

The absolute differences between successive segmented images are computed and a threshold is applied to the result such that any difference less than a threshold is set to 0. This threshold is computed as the mean absolute difference between two consecutive images. Figure 1 shows an example of the thresholded absolute image difference.

Figure 1. Example of a thresholded absolute image difference

The second step of this algorithm is to accumulate all such thresholded image differences into one image, which is referred to as the Accumulated Differences image (AD image). The AD image is initially initialized to zeros and updated based on the values of each individual image pair difference as the one shown in Figure 1. If a pixel value is greater than zero in the image pair difference, then the corresponding pixel value in AD image is incremented by 1 and so forth for all image pair differences. Figure 2 shows an example of AD image.

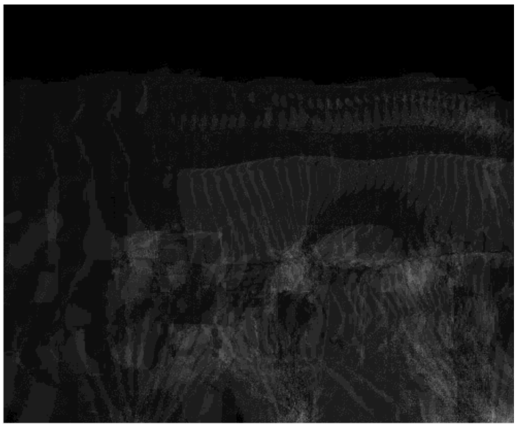


Figure 2. Example Accumulated Differences (AD) image.

The AD image is then converted to a feature vector by applying a 2-D Discrete Cosine Transformation to it using Equation 1 below.

$$F(u,v) = \frac{2}{\sqrt{MN}} C(u)C(v) * \sum_{i=0}^{M-1}\sum_{j=0}^{N-1} f(i,j)\cos\left(\frac{\pi u}{2M}\cdot(2i+1)\right)\cos\left(\frac{\pi v}{2N}\cdot(2j+1)\right) \quad (1)$$

Where the input image '*f*' has dimensions of $N \times M$ and $F(u,v)$ is the DCT coefficient at row u and column v of the DCT output matrix. $C(u)$ is a normalization factor equal to *1/(sqrt* (*2*)) for $u=0$ and 1 otherwise. In this work, we use DCT for its energy compaction property where most of the image information is concentrated in the top-left corner of the 2-D DCT matrix. As such, the top-left 10×10 DCT coefficients are selected and arranged into a 100^{th} dimensional feature vector.

B. *Proposed Accumulated Optical Flow Algorithm*

In this work, the reviewed method [8] was further expanded to include optical flow motion estimation between consecutive images. The image capturing and segmentation is the same as described in the previous section. The Gunnar Farneback optical flow algorithm is used to track the motion from one frame to the other. This function returns a flow image for each pair of consecutive frames in a sequence of frames.

Applying optical flow to the set of frames results in a motion vector for each pixel. The motion vector is represented as two values; magnitude and phase. The resulting flow image can be stored as a RGB image, where the "Red" stores the phase values and the "Blue" stores the magnitude. The "Green" stores a constant value between 0 and 255. In our work, we used a constant value of 255. Figure 3 shows an example of a flow image of an image pair.

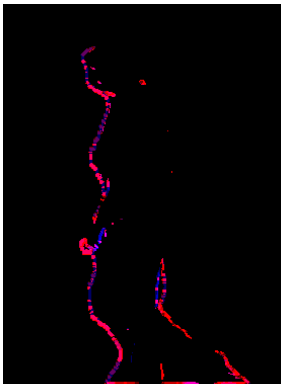


Figure 3. Example flow image of an image pair.

The second step of this algorithm is to accumulate all phases and magnitude of motion vectors into one RGB image, which is referred to as the Accumulated Motion Vectors Image (AMV) image. The AMV image is initialized initially to zeros and updated based on the values of each individual flow image as the one shown in Figure 3. If a phase value is greater than zero in the flow image, then the corresponding Red pixel value in AMV image is incremented by the same phase value. Likewise, if a magnitude value is greater than zero in the flow image then the corresponding Blue pixel value in AMV image is incremented by the same magnitude value. And so forth for all flow images in the rest of the video sequence. Figure 4 shows an example of an AMV image scaled to the range of [0, 255].

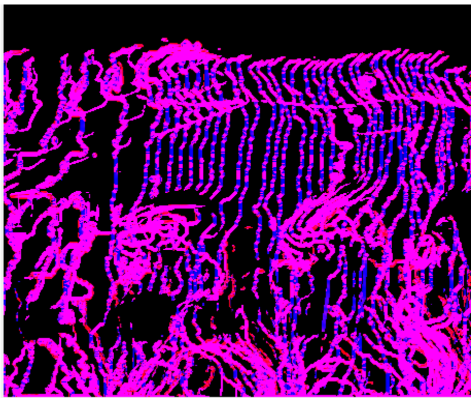


Figure 4. Example Accumulated Motion Vectors (AMV) image.

The AMV image is then converted to a feature vector by applying a 2-D Discrete Cosine Transformation to the "Red" and "Blue" components. Again, these components pertain to the phases and magnitude of motion vectors. The 2-D DCT equation is given in Equation 1 above. Since 2-D DCT transformations compacts the energy of the image in the top-left corner of the DCT image, we select the top-left 10×10 DCT coefficients of the "Red" and "Blue" components. This results in two 100^{th} dimensional feature vectors, which are concatenated into one vector of 200 elements. Figure 5 shows the flowchart of the proposed feature extraction method.

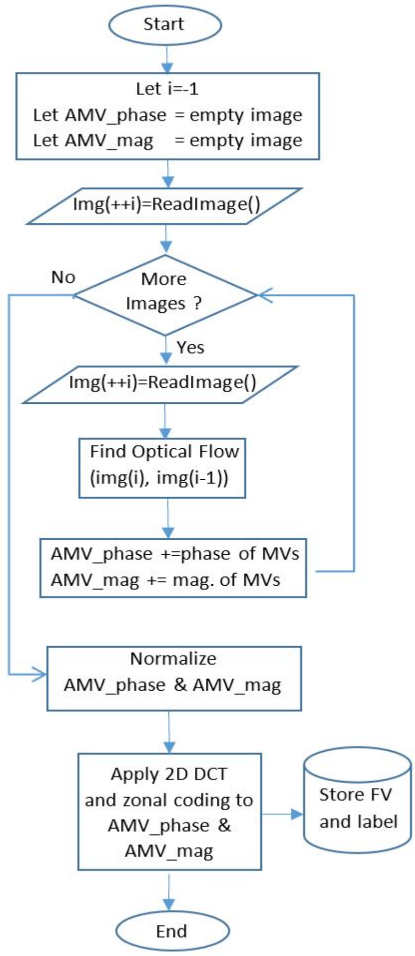


Figure 5. Flowchart of proposed feature extraction solution.

In the figure, once the phases and magnitudes of the motion vectors are accumulated and stored into a new image, the resultant image is normalized by dividing its values by the total number of images in the walk sequence. This is important as different walks have different number of images.

III. PROPOSED SOFTWARE IMPLEMENTATION

The software development environment used in this work is Python-Spyder, and the software library used are NumPy, PIL, SciPy, Skimage, Matplotlib.pyplot, Sklearn, CV2, Pyrealsense2 and Pickle. The proposed system uses a depth camera from Intel known as Intel® RealSense™ Depth Camera D435[1].

The system architecture used is the "call and return architecture" which achieves modifiability and increasing cohesion. There is no need for a multi-tiered architecture since the bulk of the software are the training and testing algorithms done by a single operator. Figure 6 shows the interface of the system.

[1] https://www.intelrealsense.com/depth-camera-d435/

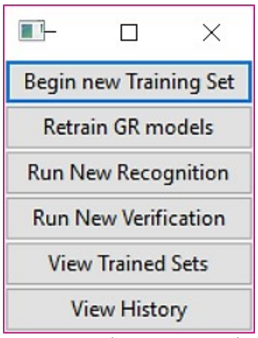


Figure 6. Main system's interface

The main available functions are training, where new subjects can enroll in the system. Figure 7 shows the interface of this function. New subjects are asked to walk in front of the camera 10 times. The software interface allows for gait recognition where the subject walks once in front of the camera and the system will recognize him/her accordingly.

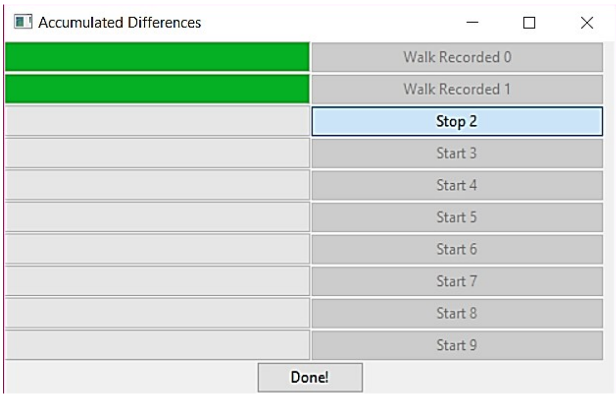


Figure 7. Enrolling a new subject into the gait system

The software also allows for automatic verification so it records the recognition results automatically. Figure 8 shows the interface for the automatic software verification.

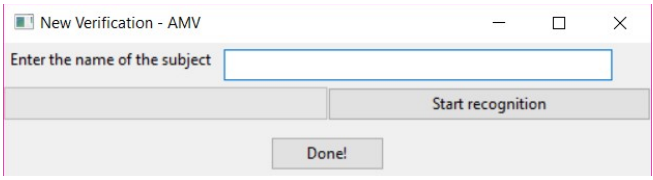


Figure 8. Interface for software verification

IV. DATA COLLECTION

In this section, data collection process for gait recognition is presented. Intel® RealSense™ Depth Camera D435 is used for data collection and processing. The camera has a wide field of view (FOV) and depth sensing capabilities. It comes with a cross-platform SDK which enables the use of many computing languages and wrappers. It contains an infra-red projector, right and left images and RGB module. The resolution of the camera is 1920 × 1080 at 30 frames per second. In this work, all images are spatially downsized to a resolution of 640 × 480.

A total of 38 undergraduate university students participated in the data collection from both genders. Each participant was asked to walk in front of the camera 5 times with and without Gulf clothing. So, the total number of walks is 10 per subject. The Gulf clothing used is an Abaya and Kandura for female and male participants respectively.

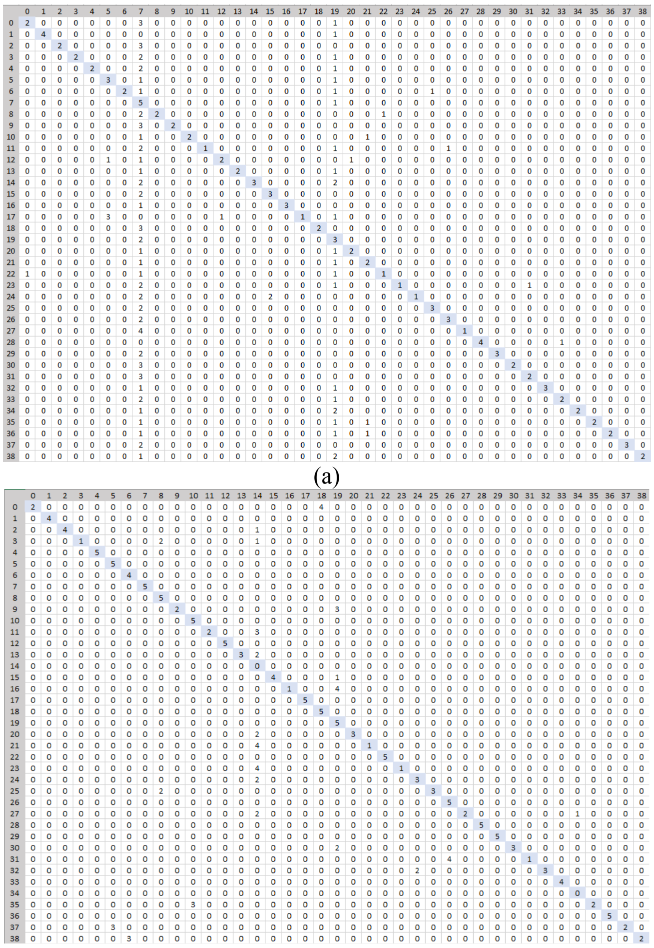
V. EXPERIMENTAL RESULTS

In this section, the results of the proposed method are presented and discussed.

An SVM classifier with a Linear kernel is used for model generation and classification. The collected data is split into 50% with traditional Abaya/Kandura clothing and 50% with western style clothing. Therefore, the training data percentage is 50%, while the rest of the data is used for testing.

The proposed method accuracy is reported using standard machine learning evaluation criteria: confusion matrices, precision, recall and f-scores.

Figure 9 plots the confusion matrices for AMV solution where training in part 'a' of the figure is based on traditional Abaya/Kandura clothing and training in part 'b' is based on Western style clothing.



(a)

(b)

Figure 9. Confusion matrices using the proposed Accumulated Motion Vectors solution (a) training on traditional Abaya/Kandura clothing (b) training on Western style clothing.

Table 1 shows the precision, recall, and f-scores results corresponding to Figure 9.

	Train traditional Test Western	Train western Test traditional
Precision	85%	91%
Recall	46%	68%
F-score	55%	71%

Table 1. Numerical summaries of precisions recall and f-scores using the proposed AMV solution.

The results reported in Figure 9 and Table 1 shows that the proposed AMV solution is capable of recognizing people by the way they walk even if the training is based on different style of clothing. The results also reveal that training on western style clothing and testing on traditional clothing results in higher recognition accuracy than case where the training and testing data sets are reversed. This is so as western style clothing reveal more information about the body shape and its movements, hence the higher recognition accuracy.

For comparison with the existing methods, the same experiments were performed using the AD solution [8] [9] summarized in Section 4.1. Table 2 shows a summary of precisions, recall, and f-score measures.

	Train traditional Test western	Train western Test traditional
Precision	30%	29%
Recall	32%	28%
F-score	29%	25%

Table 2. Numerical summaries of precisions recall and f-scores using the reviewed AD solution.

Comparing the results of the proposed solution in Table 1 against those of the reviewed solution in Table 2, it is shown that the proposed solution is much more suitable for gait recognition when training is based on a completely different style of clothing. This is expected, as the information retained the magnitudes and phases of motion vectors are more representative than thresholded absolute image differences.

The full processing time in terms of image capture, segmentation, feature extraction, training and testing for both the AD and AMV methods are shown in Figure 10. The processing time is performed on a Windows 10 computer with 16 GB RAM and an Intel® Core™ i7-7500U CPU @ 2.7GHz, 2 Cores with 4 logical processors.

The figure shows that by increasing the number of processed images per second, the processing time of the proposed AMV solution increases noticeably. This is an expected result, as the proposed solution relies on dense optical flow for the computation of motion vectors. The reviewed AD solution on the other hand, relies on thresholding absolute image differences which require relatively less processing time.

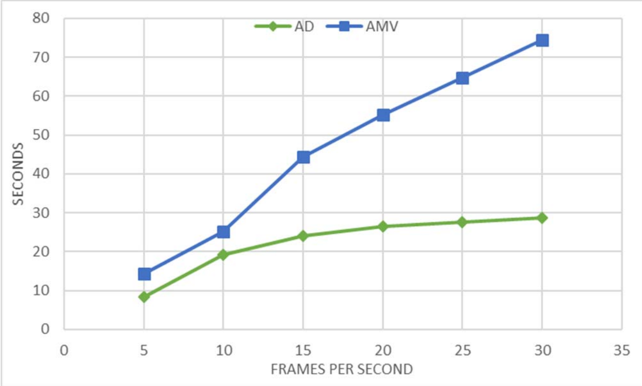


Figure 10. Total processing time per subject.

VI. CONCLUSIONS

This paper proposed a gait recognition solution specific to clothing conditions in the Gulf region where Abaya and Kandura are considered traditional clothing. The paper proposed a solution capable of recognizing users in traditional clothing although they are enrolled in the system using western style clothing and vice-a-versa. The solution uses Intel® RealSense™ Depth Camera D435, for image acquisition and foreground extraction. Once the users are segmented, the proposed solution used optical flow, accumulated motion and Discrete Cosine Transformation (DCT) for feature extraction. The motion was calculated from consecutive images where the magnitudes and phases of motion vectors are accumulated into separate matrices. This was followed by DCT and zonal coding where the top left 100 coefficients of each matrix was retained. The coefficients were concatenated to form one concise feature vector that represents a walk.

Experimental results showed that the proposed solution is suitable for recognizing 38 participants with such clothing constraints. The precision of the recognition was 85% and 91% when training and traditional and western clothing respectively. The proposed method outperformed the existing solutions in gait recognition given the clothing constraints used in this work.

REFERENCES

[1] M. Babaee, L. Li and G. Rigoll, “Gait Recognition from Incomplete Gait Cycle,” IEEE International Conference on Image Processing (ICIP), October 2018

[2] W. Zou and S. Kamata, “Frontal Gait Recognition from Incomplete RGB-D Streams Using Gait Cycle Analysis,” International Conference on Informatics, Electronics & Vision (ICIEV), June 2018

[3] S. Choudhury and T. Tjahjadi, “Gait recognition based on shape and motion analysis of silhouette contours,” Computer Vision and Image Understanding 117(12):1770–1785, December 2013

[4] L. Yao, W. Kusakunniran, Q. Wu, J. Zhang, Z. Tang, Robust CNN-based Gait Verification and Identification using Skeleton Gait Energy Image, pages 297 - 303, Australia, December 2018, Digital Image Computing: Techniques and Applications (DICTA)

[5] G. Pan, Y. Zhang and Z. Wu, “Accelerometer-based gait recognition via voting by signature points,” Electronics Letters 45(22):1116 – 1118, November 2009

[6] S. Gandhe and T. Jawale, “Human identilication using fusion of iris, signature and gait recognition,” International Conference on Global Trends in Signal

Processing, Information Computing and Communication (ICGTSPICC), December 2016
[7] T. Sattrupai and W. Kusakunniran, “Deep Trajectory Based Gait Recognition for Human Re-identification,” IEEE TENCON Region 10 Conference, October 2018
[8] T. Shanableh, K. Assaleh, L. Hajjaj and A. Kabani, “Gait Recognition System Tailored For Arab Costume of the Gulf Region,” IEEE International Symposium on Signal Processing and Information Technology , December 2009
[9] A. Al-Tayyan, K. Assaleh and T. Shanableh, “Decision-Level Fusion for Single-View Gait Recognition with Various Carrying and Clothing Conditions,” Image and Vision Computing, vol. 61, Elsevier, May 2017