Gait based Person Recognition using GEI and Skeleton Data

Pola Qulta, Youssef Amr and Amr Abdelaziz
Alexandria University, es_Pola.Soliman2025, es_Youssif.Amr2025, es_Amr.Abdelaziz2025@alexu.edu.eg

***Abstract* - This project explores gait recognition for classifying individuals using both Gait Energy Images (GEI) and skeleton data. Deep learning is employed for human silhouette extraction from videos using a segmentation model, and OpenPose is utilized to obtain skeleton keypoints. Machine learning techniques, including Random Forest and Convolutional Neural Networks (CNNs), are implemented for classification. For GEI-based classification, a Random Forest model achieved an accuracy of 0.92, while a CNN achieved a test accuracy of 0.9231. Skeleton-based classification was performed with and without data augmentation; the model without augmentation achieved an accuracy of 0.91, whereas the augmented model achieved 0.98 accuracy. Ensemble models combining GEI and skeleton data achieved an accuracy of 0.98 without data augmentation and 0.97 with data augmentation. The results demonstrate the effectiveness of combining multiple modalities and data augmentation for gait recognition.**

Index Terms – Gait Energy, Pose Estimation, Person Recognition

INTRODUCTION

Human identification is a critical task in numerous applications, ranging from security systems to personalized human-computer interaction. Biometric methods, which rely on unique biological or behavioral traits, have become increasingly important for this purpose. Among these, gait recognition, the identification of individuals by their walking patterns, stands out as a particularly promising approach due to its non-intrusive nature and potential for use at a distance. This paper explores a gait recognition system that utilizes two distinct but complementary sources of information: Gait Energy Images (GEIs) and skeleton data extracted from video sequences. GEIs provide a holistic representation of motion patterns by capturing the average silhouette of a person during a gait cycle. In contrast, skeleton data, obtained through OpenPose, a pose estimation library, offers detailed information about the positions of joints and limbs throughout the gait cycle. To leverage the advantages of each approach, we use an ensemble method that combines the classification results from models trained on both GEI and skeleton data. The main contribution of this work is to present a detailed analysis of these two gait recognition methods, along with a combined ensemble approach, to achieve improved accuracy and robustness in person identification..

METHODOLOGY

I. Data Acquisition

The data used in this project was collected by a different group using a bimodal approach, capturing both visual and inertial data simultaneously. The visual data consists of videos of 45 subjects (32 males and 13 females) aged 18-23 walking a straight path within a squash playground. Each subject walked six times along a 7-meter path, three times in one direction and three times in the opposite direction. The subjects were filmed using two iPhone 7 cameras fixed in place at different angles: one at approximately 90 degrees to the subject and the other at approximately 60 degrees from the side, with both cameras positioned about 2.5 meters apart. All videos were recorded at a quality of 1080P at 30 frames per second. Simultaneously, each subject wore four inertial measurement units (IMUs), including three MetaMotionR (MMR) devices and one Apple Watch Series 1. The MMR devices were mounted on the right upper arm, right thigh, and right knee, while the Apple Watch was worn on the right wrist. To synchronize the visual and inertial data, subjects performed a distinct synchronization signal by quickly raising and lowering their right arm and leg before starting their normal walk. The subjects were also asked to wear darker clothing to avoid distortions when extracting silhouettes from the images. The data collection sessions occurred over two weeks, with approximately seven subjects recorded per day. This previously collected dataset provides visual and inertial gait data for each subject.

Key aspects of the data acquisition:

- Bimodal Data: Both visual (video) and inertial (IMU) data were collected.
- Participants: 45 subjects (32 males, 13 females) aged 18-23.
- Walking Path: 7-meter straight path walked six times.
- Visual Recording: Two iPhone 7 cameras at 90 and 60 degree angles.

- Inertial Sensors: Three MMR IMUs and one Apple Watch.
- Clothing: Subjects wore darker clothing for silhouette extraction.
- This setup ensured that a comprehensive dataset, including both visual and inertial modalities of gait, was available for analysis.

II. Initial Data Preprocessing

Before analysis, the collected data underwent several preprocessing steps to ensure quality and usability. Initial inspection revealed that some of the video recordings were corrupted or incomplete, and these were removed from the dataset. Additionally, some inertial data streams could not be synchronized with the corresponding visual data due to technical issues. To address the missing inertial data, a decision was made to prioritize the visual data, which was used in two ways: first, to generate stabilized and centered videos using a ResNet segmentation model, and then to utilize these to generate both Gait Energy Images (GEIs) and skeleton data as a replacement. The ResNet model was used to locate the person, center them and output a stabilized video. This process resulted in videos that focused solely on the subject's movement, eliminating unnecessary background information and stabilizing the gait cycle. Then OpenPose, a pose estimation library, was used to extract 3D skeleton keypoints from these stabilized videos.

III. GEI EXTRACTION

Gait Energy Images (GEIs) were extracted from the preprocessed video data to create a holistic representation of gait patterns. First, each video was segmented into individual gait cycles, discarding the segments where subjects were turning. Silhouettes of the subjects were then extracted from each frame of the segmented videos using a combination of background subtraction, binarization, dilation, and contour finding. These silhouettes represent the subject's shape at each point in the gait cycle. To generate the GEI, the extracted silhouettes for each gait cycle were averaged using the formula:

$$G(i,j) = \frac{1}{N} \sum I(i,j,t) \qquad (1)$$

where N is the number of frames per clip, t represents the frame number, and I(i, j) is the silhouette image. This process results in a single GEI that represents the average motion pattern of the subject's gait for that cycle. These GEIs, with a dimension of 500x500, were then used as input features for the machine learning models.

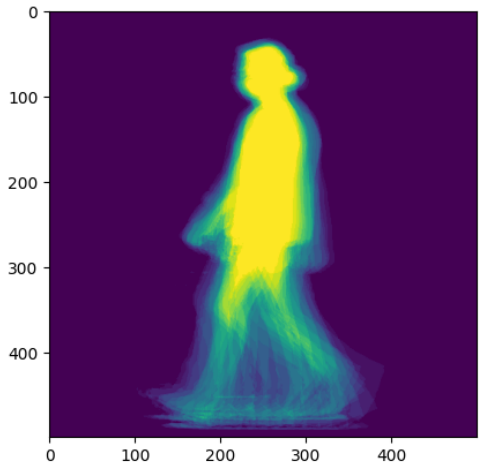

FIGURE I
EXAMPLE OF A GAIT ENERGY IMAGE

IV. Skeleton Data Extraction

Skeleton data was extracted from the preprocessed video data using OpenPose, a pose estimation library, to provide detailed information about the positions of joints and limbs throughout the gait cycle. The OpenPose BODY_25 model was utilized to detect 25 keypoints corresponding to different body joints in each frame of the video. These keypoints were extracted as 2D coordinates and stored in JSON format for each frame. This resulted in a time series of skeletal joint positions for each gait cycle, which were then used as input features for machine learning models. To create distinct temporal snippets of the skeleton data for the models, each skeleton sequence was divided into segments of 23 frames, to use as input to machine learning models. In addition, data augmentation was applied by using a sliding window of 23 frames with a stride of 5, to create new training samples. This process captured the temporal dynamics of the gait, providing a richer representation of motion patterns than static features alone.

FIGURE II
EXAMPLE OF SKELETON OUTPUT

V. Data Preprocessing

The creation of distinct and augmented datasets involved different approaches to preparing the skeleton and GEI data for model training. For the distinct datasets, each skeleton sequence was divided into non-overlapping temporal snippets of 23 frames. Similarly, a single GEI was created for each gait cycle by averaging the extracted silhouettes within that cycle. This resulted in a set of distinct, non-overlapping samples for both modalities. However, to enhance model robustness and increase the number of training samples, augmented datasets were also created. In the augmented approach, a sliding window of 23 frames with a stride of 5 was applied to both skeleton and GEI data. This means that overlapping samples were generated, with each sample consisting of a sequence of 23 frames taken every 5 frames. As a result, this process created more data points from the same raw data and allowed the models to learn from slightly different temporal perspectives of the gait cycle. In addition, to ensure sufficient sample sizes, labels with fewer than 10 instances were dropped from the distinct GEI dataset, labels with fewer than 20 samples were dropped from the distinct skeleton dataset, and labels with fewer than 230 samples were dropped from the augmented skeleton dataset.

VI. Classification Models

For gait-based person identification, two primary machine learning models were employed: a Random Forest classifier and a Convolutional Neural Network (CNN). The Random Forest model was used on flattened feature vectors of both GEI and skeleton data, treating each frame or GEI as an independent sample. In contrast, the CNN was used for the GEI data, leveraging its ability to capture spatial hierarchies and patterns within images. For skeleton data, the Random Forest model was applied to temporal snippets of skeleton keypoints. Specifically, the input to the Random Forest was created from segments of 23 frames, which were flattened into a one-dimensional vector. For the CNN, the input was the 500x500 GEI images. Data was split into training and testing sets, using a stratified split to ensure a balanced representation of different subjects in both sets. the model performance was evaluated using metrics such as precision, recall, F1-score, and accuracy. Furthermore, an ensemble model was created using a weighted average of predictions from the individual GEI and skeleton Random Forest models. The weights were applied to the probabilities from each model and then the final prediction was made from the maximum of those weighted probabilities. This method aimed to utilize the complementary information from both data modalities for more accurate classification. Finally, both a basic and data augmented methodology were tested for each model.

EXPERIMENTS AND RESULTS

The table presents a comparison of different models used for gait-based person identification, showing their performance in terms of **precision, recall, F1-score, and accuracy.**

TABLE I
EXPERIMENT RESULTS

Model	Data Aug	Precision	Recall	F1	Accuracy
RF (GEI)	No	0.93	0.92	0.92	0.92
CNN (GEI)	No	0.93	0.92	0.92	0.9231
RF(Skeleton)	No	0.93	0.91	0.90	0.91
RF(Skeleton)	Yes	0.98	0.97	0.97	0.98
Ensemble RF	No	0.98	0.98	0.98	0.98
Ensemble RF	Yes	0.98	0.97	0.97	0.97

DISCUSSION

THE FINDINGS OF THIS STUDY UNDERSCORE THE EFFICACY OF BOTH GEI AND SKELETON DATA FOR GAIT-BASED PERSON IDENTIFICATION, WITH EACH MODALITY DEMONSTRATING STRONG PERFORMANCE. THE RANDOM FOREST MODEL USING FLATTENED GEI IMAGES YIELDED A MACRO AVERAGE F1-SCORE OF 0.92 AND AN ACCURACY OF 0.92, WHILE THE CNN TRAINED ON GEI ACHIEVED A TEST ACCURACY OF 0.9231, CONFIRMING GEIS AS A RELIABLE FEATURE FOR GAIT RECOGNITION.

FOR SKELETON-BASED ANALYSIS, A RANDOM FOREST MODEL TRAINED ON TEMPORAL SNIPPETS OF 23 FRAMES, WITHOUT DATA AUGMENTATION, ACHIEVED A MACRO AVERAGE F1-SCORE OF 0.90 AND AN ACCURACY OF 0.91, SHOWCASING THE DISCRIMINATIVE POWER OF RAW SKELETON DATA. DATA AUGMENTATION, USING A SLIDING WINDOW OF 23 FRAMES WITH A STRIDE OF 5, SIGNIFICANTLY ENHANCED PERFORMANCE, RESULTING IN A MACRO AVERAGE F1-SCORE OF 0.97 AND AN ACCURACY OF 0.98, EMPHASIZING THE VALUE OF TEMPORAL CONTEXT AND DATA AUGMENTATION FOR SKELETON-BASED METHODS.

THE STUDY FURTHER EXPLORED THE FUSION OF GEI AND SKELETON DATA WITH AN ENSEMBLE MODEL. THIS MODEL, CREATED BY AVERAGING PREDICTIONS FROM RANDOM FOREST MODELS TRAINED ON EACH MODALITY WITHOUT DATA AUGMENTATION, ACHIEVED A MACRO AVERAGE F1-SCORE OF 0.98 AND AN ACCURACY OF 0.98, INDICATING THE COMPLEMENTARY NATURE OF THE TWO DATA TYPES. HOWEVER, THE ENSEMBLE MODEL DID NOT IMPROVE RESULTS WHEN USING DATA AUGMENTATION, AS THE AUGMENTED SKELETON DATA SINGLE MODEL WAS ALREADY ABLE TO ACHIEVE A MACRO AVERAGE F1-SCORE OF 0.97 AND AN ACCURACY OF 0.98.

CONCLUSION

This study investigated the efficacy of gait-based person identification using both Gait Energy Images (GEI) and skeleton data, and explored various methods for optimizing performance. The experiments confirm that both modalities provide valuable information for identifying individuals based on their gait.

Data augmentation was shown to be highly effective in improving skeleton-based analysis. Using a sliding window of 23 frames with a stride of 5, the augmented Random Forest model achieved a macro average F1-score of 0.97 and an accuracy of 0.98.

These results indicate that while both GEI and skeleton data are useful for gait-based person identification, skeleton data combined with data augmentation provides superior results. Furthermore, the ensemble model can offer an additional increase in accuracy when using non-augmented data, by combining the information available in GEI data and skeleton data. The high performance of the Random Forest model on augmented skeleton data demonstrates a computationally efficient method for achieving high accuracy.

In conclusion, this study demonstrates the potential for robust and efficient gait-based person identification. It suggests that using a Random Forest model on augmented skeleton data is the most suitable method, due to its high accuracy and low computational complexity.

REFERENCES

[1] Z. Cao, G. Hidalgo, T. Simon, S. Wei, and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019.

[2] Z. Cao, T. Simon, S. Wei, and Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," in CVPR, 2017.

[3] T. Simon, H. Joo, I. Matthews, and Y. Sheikh, "Hand Keypoint Detection in Single Images using Multiview Bootstrapping," in CVPR, 2017.

[4] S. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional Pose Machines," in CVPR, 2016.

[5] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking Atrous Convolution for Semantic Image Segmentation," arXiv preprint arXiv:1706.05587, 2017.

[6] A. Madcor, "VSGD: A Bi-modal Dataset for Gait Analysis,", 2021.