Gait Biometrics Coursework

Tom Evans
School of Electronics and Computer Science
University of Southampton
Email: tee1g21@soton.ac.uk
Word Count: 1981

***Abstract*—This project presents a gait biometric system which uses static images from the Southampton Gait Database. With each subject captured in two fixed standing poses, the study explores the effectiveness of silhouette and pose-based features for both classification and verification. Binary silhouettes were extracted using manual and machine learning-based approaches, followed by the computation of features such as width profiles, shape descriptors, and elliptical Fourier descriptors. Pose estimation was performed using MediaPipe to derive anatomical ratios and keypoint coordinates. A range of classifiers, including k-NN, Random Forest, and Logistic Regression, were evaluated, with the best classification accuracy (CCR 0.82) achieved using a combination of silhouette and pose features. Verification performance was more modest, with the lowest Equal Error Rate (EER) of 0.46. Silhouette features consistently outperformed pose-based features, demonstrating strong discriminative power even in static imagery. The report concludes with a discussion of the successes and limitations posed by the approach and proposes directions for future work.**

I. Introduction

This project investigates gait-based subject recognition using a constrained dataset provided by the coursework specification. The dataset consists of 88 training images and 22 test images, with the test set containing 11 individuals, all of whom also appear in the training set. Each subject is captured in two fixed standing poses from a single viewpoint. As a result, no true open-set verification is possible, and traditional motion-based or multi-view gait analysis techniques cannot be applied. Given these limitations, this project focuses on extracting silhouette and pose-based features from still images to evaluate their effectiveness for both classification and verification.

II. Background

The data provided for this coursework originates from the Southampton dataset presented by Shutler et al. [1]. By 2004, three recognition algorithms using binary silhouette descriptors had been applied to the dataset, achieving up to 97.3% accuracy.

Han and Bhanu [2] proposed Gait Energy Images (GEI), which average silhouettes across gait cycles. They extracted low-frequency components via discrete Fourier transform and used a nearest neighbour classifier, achieving 99.2% accuracy on the *CMU Mobo* dataset. Similarly, Wang et al. [3] employed PCA to reduce silhouette features such as width, height, and inter-body part distances, before classifying with nearest neighbour.

Castro et al. [4] proposed a fully pose-based gait recognition method using 3D joint coordinates from modern pose estimation systems. Features such as pairwise joint distances and angles were extracted per frame to represent movement. While results were competitive across standard gait datasets, pose features alone did not match the performance of silhouette-based methods.

Hussain et al. [5] evaluated five common classifiers: k-Nearest Neighbour (K-NN), Random Forest (RF), Decision Tree (DT), Logistic Regression (LR) and Stochastic Gradient Descent (SGD), in a gait analysis system for foot disease diagnosis. K-NN and RF achieved %99 accuracy, DT reached % 97, whilst LR and SGD achieved %82. PCA was also used to reduce feature dimensionality.

Yao et al. [6] employed data augmentation techniques such as horizontal flipping to increase training data size with CNN classification. Augmentation consistently improved accuracy across different walking scenarios, such as carrying bags and wearing coats.

III. Method

A. Data Labelling

Class labels were derived from the image filenames, where the second three-digit sequence (e.g., *016z050pf*) indicated the subject ID. This was used to label the training set. Test image labels were assigned manually through visual inspection and stored in a dictionary mapping to the original filenames.

B. Silhouette Feature Extraction

Silhouette-based features were implemented due to their widespread use in gait systems discussed in Section II. Binary silhouettes were initially generated using a manual method (see Figure 1):

- Converted images to HSV colour space.
- Define green screen using colour thresholding.
- Applied a binary thresholding to each image to isolate individuals from the green screen.
- Remove background outside of a hard-coded bounding box surrounding the individual.
- Extract the largest contour from each image and remove all others.
- Resize and save to output directory.

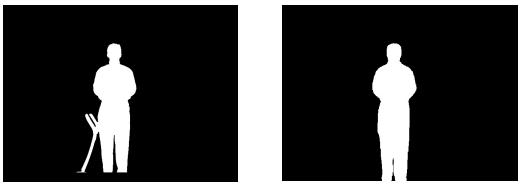
As Figure 1 shows, individuals standing close to the treadmill caused parts of it to be included in the silhouette, resulting in distortions that appeared as extensions of the body. To

(a) Successful silhouette extraction

(b) Unsuccessful silhouette extraction

Fig. 1. Comparison of manual silhouette extraction: (a) shows a successful result where the subject is cleanly isolated, while (b) shows a failure case with incomplete or noisy extraction.

overcome the limitations of the manual method, the DeepLab image segmentation model [7] was used to extract silhouettes automatically. This model performs pixel-wise classification to isolate the person, producing clean binary silhouettes without background elements like the treadmill. This yielded more reliable results across the dataset (see Figure 2).



(a) Manual silhouette (b) DeepLab silhouette

Fig. 2. Comparison of binary silhouette extraction of image *017z054pf.jpg* using Manual and DeepLab methods.

Once a dataset of binary silhouettes had been generated, features were extracted for classification. Given the limited size of the dataset, lightweight and well-established shape-based features were selected. These included:

- Width and Height profiles - width measurements at evenly spaced vertical (or horizontal) positions.
- Area, Perimeter, Compactness, dispersion and aspect ratio.
- Hu moments.
- Elliptical Fourier Descriptors.

C. Pose Estimation Feature Extraction

Given the promise of pose estimation, highlighted by Castro et al. [3], MediaPipe's pose detection model [8] was used to provide 3D keypoints for each individual in the train and test images, representing different parts of the body (Figure 3). From these keypoints, a combination of raw coordinates and derived ratios was used to represent body proportions. These included:

- Limb lengths.
- Ratios between leg and arm segments.
- Shoulder-to-hip width.
- Head-to-torso height.

These ratios were designed to capture consistent anatomical differences between individuals while being robust to image scale and camera perspective. This representation allowed the model to focus on distinctive structural characteristics rather than pixel-based details. Some common pose estimation features, such as limb angles, were not applicable as the dataset consisted of still images in fixed poses, making them non-discriminative across subjects.

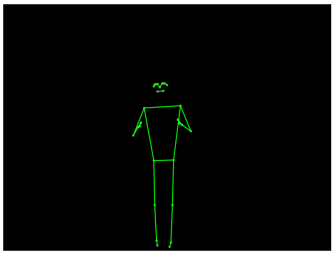


Fig. 3. Pose Estimation of image *016z050pf.jpg* over black background for visualisation.

D. Classification

Section II highlighted several commonly used classifiers in gait-based biometric systems. These lightweight classifiers are well-suited to this project, given the limited amount of training data available. The following classifiers were implemented using the `Scikit-learn` Python package:

- K Nearest Neighbour (k-NN)
- Random Forest (RF)
- Logistic Regression (LR)
- Decision Tree (DT)
- Support Vector Machine (SVM)
- Stochastic Gradient Descent (SGD)
- Gaussian Naive Bayes

As shown in section IV, SGD and Gaussian Naive Bayes were not pursued due to their consistently poor classification performance. Neural network approaches, including multi-layer perceptrons (MLPs) and custom CNNs implemented in PyTorch, were also explored but did not yield satisfactory results. This may be due to the limited size of the training dataset, which can hinder the ability of deep learning models to generalize effectively.

Classifier inputs (feature vectors) and parameters were defined using dictionaries, which allowed for automated scripting to optimise results via multiple executions of the pipeline. PCA dimension reduction and data augmentation, via horizontal flips and minor image shifts, were also employed after showing promise to improve gait performance in Section II.

E. Verification

To evaluate whether two samples belonged to the same subject, a pairwise verification approach was implemented. All possible pairs of feature vectors were generated, and each pair was labelled as either a genuine match (same subject) or an impostor pair (different subjects). Similarity scores were then computed for each pair using either Euclidean distance (inverted to act as a similarity score) or cosine similarity. By

comparing these scores against a threshold, the system could classify each pair as a match or a non-match.

F. Evaluation

The system was evaluated using the metrics specified in the coursework brief: Correct Classification Rate (CCR) for recognition, Equal Error Rate (EER) for verification, CCR at the EER threshold, and intra/inter-class variation histograms. Given the small dataset size, CCR was prioritised as it offered a simpler and more stable measure of performance. In contrast, threshold-based metrics like EER can be more sensitive to class imbalance and noise in similarity scores, which are more likely in limited data scenarios.

IV. Results

A. Classification

1) Silhouette Features

Silhouette features consistently outperformed pose estimation across most classifiers. The best result (17/22) was achieved using 1000 width measurements and 10 elliptical Fourier descriptors with LR (Table I). Width profile resolution had the greatest influence on performance: k-NN performed best with lower resolutions, while LR favoured higher ones. Additional features such as area, perimeter, dispersion, and compactness offered small gains for RF and SVM but tended to degrade performance with k-NN. EFD provided slight improvements, enabling the top result, while Hu moments and height profiles performed poorly overall, achieving a maximum CCR of just 0.136 and contributing little when combined with other features.

2) Pose Features

As Section II suggested, pose-based features underperformed compared to silhouette features. achieving a maximum CCR of 0.545 (12/22) using raw *xy* coordinates and body ratios as features and classifying with DT (Table II). K-NN, RF, and LR all performed poorly, though slightly better when using only raw coordinates. Adding body ratios improved performance for DT, while SVM consistently performed poorly across all feature sets. Pure ratio-based features were ineffective, likely due to limited pose variation and the small dataset.

3) Combined Features

Combining silhouette and pose features occasionally improved results. LR consistently performed well, reaching 16-17/22 across several feature combinations (Table III). Other classifiers showed limited benefit: K-NN performance decreased, while SVM and DT saw little change. Results with RF were inconsistent, though it achieved the highest CCR (0.82, 18/22) across all experiments using width profiles (2000), raw *xy* coordinates, and body ratios.

B. Verification

Table V shows verification performance across different features. Cosine similarity generally performed better for silhouette-based features, achieving the lowest EER of 0.46 using *WP:100* combined with 10 EFDs, with a CCR of 0.54 at that threshold. Pose-based features gave mixed results; in some cases, Euclidean distance performed slightly better. For example, height profiles achieved the lowest Euclidean EER despite poor classification performance.

C. Intra-Inter Class Variation

Figures 4 and 5 show the intra- and inter-class distance distributions for the best-performing silhouette and pose-based feature sets, respectively. Silhouette features produced a clearer separation between intra- and inter-class distances, with minimal overlap, indicating stronger discriminative power. In contrast, pose-based features showed substantial overlap, explaining their weaker verification performance. The distance distribution for the combined feature set closely mirrored the silhouette-based plot, suggesting that pose features had minimal influence when fused with silhouette data.

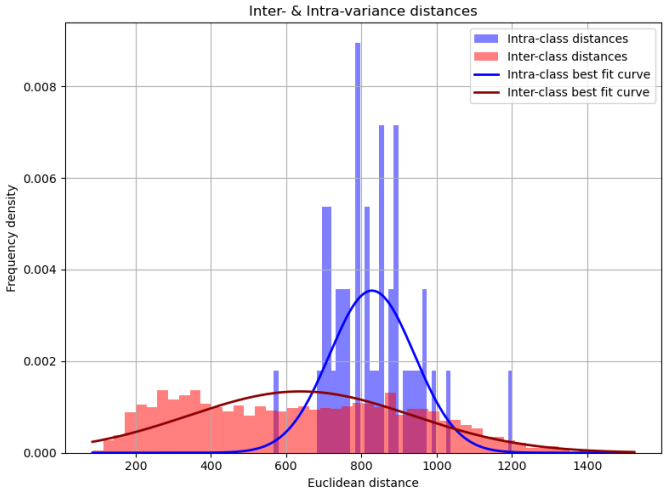


Fig. 4. Inter & Intra class variance distance histogram for best performing silhouette features

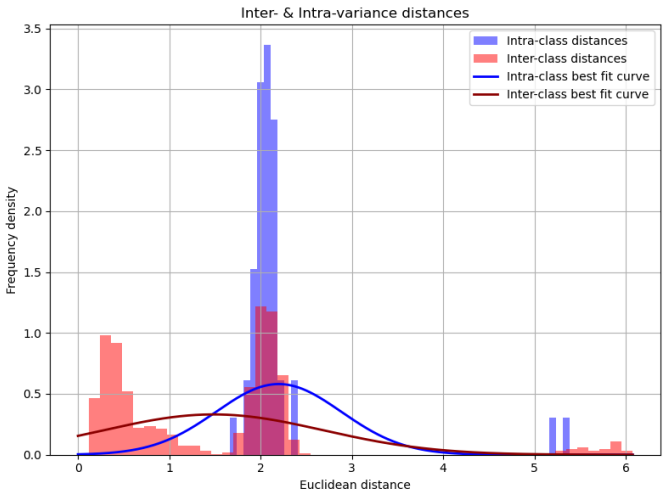


Fig. 5. Inter & Intra class variance distance histogram for best performing pose estimation features.

D. PCA and Data Augmentation

Both PCA and data augmentation were tested to enhance classification and verification performance but showed limited success. PCA yielded mixed results, often having little effect or slightly reducing accuracy in both tasks. Data augmentation occasionally improved classification (CCR) but was inconsistent, sometimes degrading performance due to noise or distortion from random transformations.

V. Discussion

A. Approach

1) Classifaction

Silhouette features consistently outperformed pose-based features, likely due to the limited variation in subject pose and the static nature of the dataset. With only two poses per individual, pose-based descriptors lacked the variability needed for reliable identification. In contrast, silhouette features, particularly width profiles, captured individual-specific body shape characteristics that remained distinguishable in still images. EFDs contributed minor improvements by encoding contour shape, while Hu moments and basic shape statistics performed poorly, possibly due to their sensitivity to noise and limited discriminative power in this context.

The impact of combining features varied by classifier. LR benefited from combined features, likely due to the increased data size; it consistently performed better with more data present, favouring high-resolution features. In contrast, k-NN saw no benefit, even within silhouette or pose feature sets. Its best performance came from individual features:(*WP:50*) for silhouette and (*raw xy*) for pose, suggesting sensitivity to feature scale and format. Combining diverse features likely disrupted its distance-based calculations, which could be mitigated by feature weighting or normalisation.

RF achieved the highest CCR overall and performed well with large and combined feature sets. However, its performance was unstable. Small changes in features sometimes caused noticeable variation, likely due to its sensitivity to small variations in high-dimensional data, exacerbated by the limited dataset. SVMs and decision trees performed poorly overall, likely due to their sensitivity to limited training data.

2) Verification

Overall, this approach achieved strong classification performance, with CCR scores exceeding 0.8 on certain feature sets. However, verification results were more modest, with the lowest EER at 0.46 and CCR at EER peaking at 0.54. Intra and inter-class distance distributions supported these findings: silhouette features showed moderate separation, while pose-based features had significant overlap. This suggests that although the features were sufficient for identifying the most likely class, they lacked the consistency and discriminative power required for reliable pairwise verification for such sparse training data.

The minimal impact of PCA and data augmentation was also likely due to the small dataset size, where reduced variance and limited diversity restricted their ability to enhance feature quality or generalisation.

B. Future Work

Future work could explore datasets containing walking sequences or varied pose angles to enable true gait analysis and improve the effectiveness of pose-based features and temporal silhouette methods, such as Gait Energy images proposed by Han and Bhanu [2]. This would allow models to capture temporal and view-dependent information, which static images cannot provide. Additionally, future evaluation could include open-set identification or verification scenarios, where the system must detect whether a subject is unknown, providing a more realistic assessment of generalisation and robustness.

VI. Conclusion

This project explored silhouette and pose-based features for subject recognition using static gait data. Silhouette features, particularly width profiles, consistently outperformed pose-based alternatives, achieving high classification accuracy. However, verification performance was modest across all feature sets. Overall, simple feature representations and classifiers proved most effective given the constraints of the dataset, and several limitations highlight promising directions for future work.

References

[1] J. D. Shutler, M. G. Grant, M. S. Nixon, and J. N. Carter, "On a large sequence-based human gait database," in *Applications and Science in Soft Computing*, ser. Advances in Soft Computing, vol. 24, Springer, 2002, pp. 339–346. DOI: 10 . 1007 / 978 - 3 - 540 - 45240 - 9_46. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-540-45240-9_46.

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

[3] L. Wang, T. Tan, H. Ning, and W. Hu, "Silhouette analysis-based gait recognition for human identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1505–1518, 2003.

[4] F. Castro, L. Sigal, R. Maranon, M. Castro, P. Arias, and A. Ortega, "Pose-based gait recognition," *arXiv preprint arXiv:2006.12399*, 2020.

[5] A. N. Hussain, S. A. Abboud, B. A. Jumaa, and M. N. Abdullah, "Gait classification using machine learning for foot diseases diagnosis," *Technium Romanian Journal of Applied Sciences and Technology*, vol. 4, no. 4, pp. 37–49, 2022, ISSN: 2668-778X. [Online]. Available: https://techniumscience.com/index.php/technium/article/view/6528.

[6] G. S. Carlos, R. O. França, W. G. dos Santos, and J. P. Papa, "Gait recognition using 3d convolutional neural networks," *PeerJ Computer Science*, vol. 7, e382, 2021. DOI: 10.7717/peerj-cs.382. [Online]. Available: https://peerj.com/articles/cs-382/.

[7] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, *Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs*, 2017. arXiv: 1606.00915 [cs.CV]. [Online]. Available: https://arxiv.org/abs/1606.00915.

[8] C. Lugaresi, J. Tang, H. Nash, *et al.*, "Mediapipe: A framework for perceiving and processing reality," in *Third Workshop on Computer Vision for AR/VR at IEEE Computer Vision and Pattern Recognition (CVPR) 2019*, 2019. [Online]. Available: https://mixedreality.cs.cornell.edu/s/NewTitle_May1_MediaPipe_CVPR_CV4ARVR_Workshop_2019.pdf.

APPENDIX

A. Results Log

TABLE I
CLASSIFICATION WITH SILHOUETTE FEATURES: *Width Profiles (WP), Area (A), Perimeter (P), Compactness (C), Dispersion (D), Elliptical Fourier Descriptors (EFD).*

Feature Vector	Correct Classifications (CCR * 22)				
	k-NN	*RF*	*LR*	*SVM*	*DT*
WP:50	16	11	11	9	11
WP:100	14	12	11	8	9
WP:200	14	12	11	8	7
WP:300	13	9	14	8	9
WP:400	13	12	13	8	8
WP:500	13	12	13	8	8
WP:1000	13	13	14	8	12
WP:50, APCD	3	12	11	9	9
WP:100, APCD	3	14	11	9	8
WP:200, APCD	3	13	11	9	8
WP:300, APCD	4	10	13	9	8
WP:400, APCD	4	14	13	9	7
WP:50, EFD:10	16	10	10	9	8
WP:100, EFD:10	14	11	12	8	6
WP:200, EFD:10	14	12	13	8	12
WP:300, EFD:10	13	9	14	8	9
WP:400, EFD:10	13	16	14	8	12
WP:500, EFD:10	13	11	14	8	6
WP:1000, EFD:10	13	13	17	8	8
WP:1000, EFD:10, APCD	6	14	17	9	7
WP:1000, EFD:5	13	11	14	8	9
WP:1000, EFD:15	13	10	13	8	8

TABLE II
CLASSIFIACTION WITH POSE FEATURES: *Mode (xy,xyz,xy+xyz), Visibility/Confidence (V), Body measurement ratios.*

Feature Vector	Correct Classifications (CCR * 22)				
	k-NN	*RF*	*LR*	*SVM*	*DT*
xy	6	9	2	2	11
xy, V	3	7	3	1	10
xyz	5	9	4	1	9
xyz, V	5	9	4	0	8
xy+xyz	5	7	5	1	9
xy+xyz, V	3	8	4	0	9
Ratios	3	0	0	0	0
Ratios, V	1	1	0	1	2
Ratios	3	0	0	0	0
Ratios, xy	4	6	3	0	12
Ratios, xyz	5	7	5	0	8
Ratios, xy +xyz	5	6	5	1	6
Ratios, xy, V	2	4	3	1	9

TABLE III
CLASSIFICATION WITH COMBINED FEATURES: PREVIOUS SILHOUETTE FEATURES + PREVIOUS POSE FEATURES. DUE TO SO MANY FEATURE COMBINATIONS POSSIBLE, THE TABLE ONLY INCLUDES NOTABLE COMBINATIONS. AUTO-SCRIPTING WAS USED TO OPTIMISE RESULTS.

Feature Vector	Correct Classifications (CCR * 22)				
	k-NN	*RF*	*LR*	*SVM*	*DT*
WP:1000, EFD:10, xy, Ratios	11	10	14	8	10
WP:1000, EFD:10, APCD, xy, Ratios	7	14	14	9	10
WP:1000, EFD:10, APCD, xyz, Ratios	7	13	14	9	8
WP:1000, EFD:10, APCD, xy, Ratios, V	7	14	16	9	7
WP:1000, EFD:10, APCD, xyz, Ratios, V	7	12	14	9	6
WP:1000, EFD:10, APCD, xy+xyz, Ratios	7	13	14	9	9
WP:1000, EFD:10, APCD, xy+xyz, Ratios, V	7	13	16	9	8
WP:1000, EFD:10, APCD, xy+xyz, V	7	13	17	9	12
WP:1000, EFD:10, Ratios	11	11	17	8	5
WP:1000, EFD:10, APCD, Ratios	7	9	17	9	4
WP:1000, EFD:10, Ratios, V	11	12	17	8	8
WP:2000, EFD:10, Ratios, V	11	12	17	8	7
WP:400, EFD:10, xy, Ratios	10	16	14	8	11
WP:2000, xy, Ratios	11	18	14	8	6

TABLE IV
VERIFICATION PERFORMANCE. *Equal error rate (EER)* AND THE *Correct Classification Rate* AT SPECIFIC *EER* THRESHOLD ACROSS FEATURE VECTORS. RUNS WITH COMBINED FEATURES USE THE BEST PARAMETERS FROM MULTIPLE RUNS OF THAT SPECIFIC COMBINATION.

Feature Vector	Euclidean Distance		Cosine Similarity	
	EER	*CCR*	*EER*	*CCR*
WP:50	0.64	0.35	0.47	0.53
WP:200	0.65	0.35	0.47	0.53
WP:500	0.65	0.35	0.47	0.53
WP:1000	0.66	0.35	0.47	0.53
WP:2000	0.66	0.35	0.47	0.53
EFD (all variations)	0.57	0.44	0.59	0.42
APCD	0.71	0.29	0.51	0.49
Hu	0.70	0.30	0.67	0.28
AR, SL	0.66	0.33	0.68	0.32
Height:100	0.50	0.50	0.50	0.50
Height:10	0.48	0.52	0.50	0.50
Height:200	0.49	0.50	0.51	0.49
Ratios	0.66	0.34	0.66	0.34
xy	0.62	0.38	0.60	0.40
xyz	0.63	0.37	0.63	0.37
xy+xyz	0.63	0.37	0.61	0.39
WP:50, Height:100	0.52	0.49	0.50	0.50
WP:100, EFD	0.65	0.35	0.46	0.54
Height:10, EFD	0.48	0.52	0.50	0.50
WP:100, EFD, PCA	0.60	0.40	0.61	0.40

TABLE V
CLASSIFICATION AFTER PCA. *PCA* APPLIED TO THE BEST-PERFORMING FEATURES FOR EACH CLASSIFIER. EACH ROW CONTAINS THE BEST RESULTS FROM MULTIPLE RUNS ACROSS A RANGE OF *PCA* COMPONENTS (N).

Feature Vector	PCA n	Correct Classifications (CCR * 22)				
		k-NN	*RF*	*LR*	*SVM*	*DT*
WP:50, EFD:10	30	10	5	6	9	2
WP:1000, EFD:10, APCD	30	14	8	13	9	7
WP:1000, EFD:10, APCD, xy+xyz, V	50	14	10	13	9	3
WP:2000, xy, Ratios	50	14	12	13	9	5