Identifying Individuals Through Egocentric Motion: A Study Using Body Worn Cameras

Sameer Hans
Digital Security
EURECOM
Biot, France
hans@eurecom.fr

Jean-Luc Dugelay
Digital Security
EURECOM
Biot, France
dugelay@eurecom.fr

Mohd Rizal Mohd Isa
Faculty of Defense Science and Technology
Universiti Pertahanan Nasional Malaysia (UPNM)
Kuala Lumpur, Malaysia
rizal@upnm.edu.my

2025 13th International Workshop on Biometrics and Forensics (IWBF) | 979-8-3315-3715-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/IWBF63717.2025.11113448

***Abstract*—Body worn cameras (BWCs) have become increasingly integrated into various professional contexts during the last decade, especially in law enforcement. BWCs are useful instruments for improving security, accountability, and transparency by offering real-time, first-person perspective recordings of conversations and events. They record enormous volumes of video, which can provide important insights into how people behave and act in various situations.**

User identification through egocentric motion analysis offers a novel perspective in biometrics, leveraging the unique motion patterns captured by BWCs. In this study, we provide FALEBego, a novel dataset using body worn camera consisting of egocentric motion of different users. We also provide some preliminary tests on the proposed dataset. This work includes two distinct insights: (1) introduction of a dataset comprising egocentric motion of 23 subjects recorded with a chest-mounted BWC, capturing their distinct walking patterns, and (2) the dataset is utilized to present a comparative analysis of different deep learning architectures for user identification based on egocentric motion data. This study highlights the potential of egocentric motion as a biometric modality and provides insights into the effectiveness of different architectures in this emerging domain. The complete dataset is available for research purposes and can be accessed by contacting the authors directly.

***Index Terms*—Body Worn Camera, Egocentric motion, Law enforcement, Surveillance**

I. Introduction

The growth of Body Worn Cameras (BWCs) has revolutionized various domains, including law enforcement, healthcare, and sports, by providing first-person perspectives on activities and events. Beyond their use for video documentation and evidence collection [15], they are an essential tool for law enforcement to improve accountability [14] and transparency [6]. BWCs have unlocked new possibilities for understanding egocentric motion. Unlike third-person observational data, egocentric motion captures an individual's movements directly from their point of view, encapsulating gait, posture, and dynamic movement patterns. These features are inherently distinct across individuals, making them a rich data source for identity recognition tasks.

Identifying users through their egocentric motion is an emerging research area that leverages the combination of computer vision and human motion analysis. Traditional methods for user recognition have predominantly focused on biometric identifiers such as facial features, fingerprints, or iris patterns. However, egocentric motion offers an alternative modality, particularly useful in scenarios where conventional biometric data may be unavailable or occluded, such as during active movement, adverse lighting conditions, or while wearing personal protective equipment.

This work introduces FALEBego[1], a novel dataset for user identification by their egocentric motion using BWCs. This study aims to explore the feasibility of identifying users based on their egocentric motion using BWCs. By employing state-of-the-art computer vision techniques, such as optical flow and deep learning models, we analyze the motion patterns to achieve robust user recognition. We compare various deep learning architectures based on C3D [16], I3D [3], SlowFast network [9], and TimeSformer [1], to evaluate the performance of different models for this specific identification task.

The findings from this research have significant implications for security, authentication, and personalized technology applications. For instance, user recognition through egocentric motion can be utilized for access control in secure facilities, hands-free authentication in augmented reality systems, or personalized assistance in wearable technologies. By advancing the understanding of egocentric motion as a biometric modality, this study opens new avenues for identity recognition that extend beyond traditional biometrics and harness the full potential of wearable camera systems.

The paper is organized as follows. In section II we survey related work on egocentric activities and BWCs. In section III, we introduce the steps followed in the data collection. We report our experimental setup and implementation in section IV. The experiments and the results are presented in section V. Finally, the conclusions and future work follow in section VI.

II. Related Works

User identification based on motion patterns has emerged as a promising research area in biometric authentication. Traditional approaches in gait recognition, which analyze walking patterns, have been adapted for egocentric scenarios. Methods leveraging wearable sensors or cameras have shown that motion patterns can serve as unique biometric identifiers.

[1]To obtain the dataset, please visit https://faleb.eurecom.fr/

The ability to extract spatial and temporal information from video data has been significantly enhanced by models like Convolutional Neural Networks (CNNs) [16] and Recurrent Neural Networks (RNNs). However, the application of multi-modal approaches, combining RGB and optical flow, to user identification from BWC data is still an emerging field.

Very limited work exists on image processing by BWCs. Despite advancements, egocentric user identification presents several challenges. Motion patterns captured from BWCs are influenced by factors [7] such as camera placement, walking speed, and environmental conditions. Additionally, the lack of large-scale, publicly available datasets specific to egocentric user identification has hindered progress in this domain. Early works in this domain focused on action recognition, where models leveraged the spatial and temporal information in egocentric video data to identify activities. For instance, datasets like EPIC-Kitchens [8] have been instrumental in driving research on egocentric video analysis. These studies emphasized the challenges associated with egocentric data, which distinguish egocentric vision from traditional third-person video analysis.

Reference [10] presented an approach to identifying photographers using egocentric video, leveraging camera motion as a unique identifier. Their method capitalized on sparse optical flow vectors to model body motion as a distinctive feature of each photographer. They experimented with Linear Predictive Coding with a kernel-based SVM, achieving 81% recognition accuracy, and a CNN that improved accuracy to 90% on the recognition task (6 people).

The study [12] provides a significant contribution to the field of egocentric video analysis by exploring ego-motion classification for body-worn videos. The authors categorize ego-motions based on similarity transformations between successive video frames. Their method extracts motion features such as horizontal displacement, rotation, and zoom, along with frequency analysis, to capture periodicity in motion. These features are then classified using graph-based semi-supervised and unsupervised learning algorithms, which achieve high accuracy on choreographed videos and real-world data.

The study [5] focuses on first-party action recognition in body-worn videos. They investigate the identification of ego-activities in first-person video and suggest a system that uses hand-crafted features and a graph-based semi-supervised learning technique to classify ego-activities in body-worn video footage. They achieve comparable performance to supervised methods on public datasets, however the challenges include insufficient training data and law-specific actions.

The research [17] introduced a study on egocentric hand identification. The study utilized several modalities, including RGB, depth, and hand segmentation masks, to investigate how features such as hand shape, skin texture, and motion contribute to person identification. They demonstrated that even with constrained data like binary hand silhouettes, reasonable identification accuracy could be achieved. However, their study focused specifically on hand gestures and required the hands to be visible in the egocentric view, limiting its generalizability to scenarios involving broader motion patterns.

Fig. 1: Cammpro I826 Body Camera.

Reference [4] presents a multimodal dataset for human action detection that makes use of wearable sensors and a depth camera to identify actions more precisely. It has 880 sequences of 22 human acts carried out by 5 participants. Based on depth images of various actions, the recognition rate ranges from 58% to 97%. Although the number of participants in this dataset is extremely small, it is nonetheless helpful as a starting point.

In this paper, we present a comprehensive dataset comprising egocentric motion recordings from 23 subjects, specifically designed to evaluate user identification performance using BWCs. To the best of our knowledge, this is the first publicly available dataset leveraging egocentric motion captured by BWCs, particularly suitable for user identification task just by egocentric motion. No prior studies in the literature have explored or published findings utilizing police BWCs for user identification in such practical settings.

III. DATASET COLLECTION

For the data collection, students from UPNM volunteered. The users recorded using Cammpro[2] I826 Body camera (as shown in Figure 1), which was securely mounted at the center of the chest of the user [2]. The recordings were carried out over multiple sessions spread across a week. All the recordings were done with a video resolution of 2304×1296 pixels at 30 fps.

The activity was recorded in an outdoor setting. It was divided into 2 scenarios. Two endpoints (A and B) were designated at opposite ends of the campus, approximately an 8-minute walk apart. In the first scenario, the user walked from point A to point B at a normal pace. For the second scenario, they followed the same path back (B to A) in a slow jogging pace. Before starting, all participants received clear instructions on how to perform the tasks. A total of 23 subjects participated in the activity, with each subject contributing two egocentric videos: one approximately 8 minutes long, capturing their walk from A to B, and another around 5 minutes long, documenting their slow jog from B to A.

[2]https://www.cammpro.com/

IV. SETUP

A. Preprocessing - RGB

To ensure a sufficient amount of training data, the videos are divided into sequences of 4 seconds, which are adequate to capture a few steps of the user's motion. This accounts to around 200 videos per user. In total, we have 4724 videos in total for the 23 subjects. The videos are then converted into frames and organized based on the two scenarios (walking and slow jogging). At 30 fps, each 4-second sequence results in 120 frames.

As part of the preprocessing pipeline, the extracted frames are resized to a standard resolution of 224×224 pixels to ensure uniformity and compatibility with various model architectures. During training, input clips are randomly cropped into $16 \times 112 \times 112$ patches, enabling both spatial and temporal jittering to improve generalization. These augmentation strategies help to mitigate overfitting and enhance the model's ability to identify users under varied conditions.

To further prepare the data for the task, normalization is applied to the pixel values of the frames using the mean and standard deviation of the ImageNet [13] dataset, for standardizing pixel intensity values across all channels. The data is then shuffled and split into training, validation, and test sets in a 65:15:20 ratio. During this split, a subject can appear in all the sets.

B. Preprocessing - Optical Flow

The same 4-second videos are used for computing optical flow. We calculate optical flow using the Farneback method, which estimates motion between two consecutive frames of a video. The first frame is converted to grayscale, which serves as the reference for calculating motion in the subsequent frame. The Farneback optical flow algorithm is used to compute the dense motion field between the current and next grayscale frames. The output is a flow field, a 2D vector for each pixel showing motion in horizontal and vertical direction.

Figure 2 provides a visualization of sample frames obtained after preprocessing. For visualization of optical flow, we convert first into HSV image and then into RGB image. This allows the optical flow to be visualized.

C. Implementation

For our experiments, we used C3D, I3D, SlowFast network, and TimeSformer models. These models were chosen for their respective strengths: C3D as a baseline model, I3D for its popularity and proven performance, SlowFast Network for its advancements in the field, and TimeSformer for its novelty.

- **C3D:** We implement a pretrained C3D model, trained on the Sports-1M dataset [11], which comprises 1.1 million sports videos belonging to one of the 487 sports categories. To evaluate the performance of the model and gain some insights on the videos by BWCs, the model is fine-tuned on our dataset. The SGD optimizer is used for training. The learning rate is fixed as 0.001 after various experiments. The initial layers of the model are frozen, and we add fc8 layer to match the number of classes in our dataset. This layer is trained from scratch using random weights.
- **I3D:** We experiment with I3D architecture pretrained on Kinetics-400 [3] dataset. This model is used to initialize our network, where we replace the final projection layer to match the number of classes in our dataset. The model was trained using the CrossEntropy loss function, and optimized using the Adam optimizer with a learning rate of 0.001. During training, both training and validation metrics, including loss, accuracy, precision, recall, and F1-score, were monitored to evaluate the performance. We also ensured that each input video was processed as a stack of frames, allowing the I3D model to leverage its 3D convolutional layers to capture temporal dynamics.
- **SlowFast Network:** The SlowFast network operates by processing video inputs through two pathways: the slow pathway, which samples frames at a lower frame rate to capture long-range temporal patterns, and the fast pathway, which processes higher frame-rate sequences to capture finer motion details. The SlowFast network is pretrained on Kinetics-400 in our implementation. The final fully connected layer of the network is replaced with a new layer corresponding to the number of classes in the dataset. To accommodate both fast and slow temporal dynamics, the video frames are split into two pathways, with the slow pathway subsampling every fourth frame, while the fast pathway uses all the frames. The training uses cross-entropy loss to minimize classification error, with an Adam optimizer tuned on a learning rate of 0.001.
- **TimeSformer:** The TimeSformer is a deep learning architecture designed specifically for video understanding tasks like action recognition. It applies transformers directly to the spatial and temporal dimensions of the video. It processes video frames as a sequence of patches, integrating attention mechanisms across both time and space, allowing for more efficient and scalable learning of video features. In our implementation, the model is pretrained on Kinetics-400. We fine-tune the model by modifying the classifier head and employing techniques like mixed precision training to handle GPU memory constraints. During training, the model's performance is evaluated on validation data after each epoch, and its performance is seen on test data every 5 epochs to monitor accuracy and loss, aiming to improve the classification performance.

V. EXPERIMENTS

A. RGB user recognition

To evaluate the effectiveness of user identification based on egocentric motion, we conducted a series of experiments using each model on our custom egocentric dataset. The models were trained for 20 epochs, and metrics of accuracy, precision, recall, F1-score, and loss were tracked for training, validation, and testing phases.

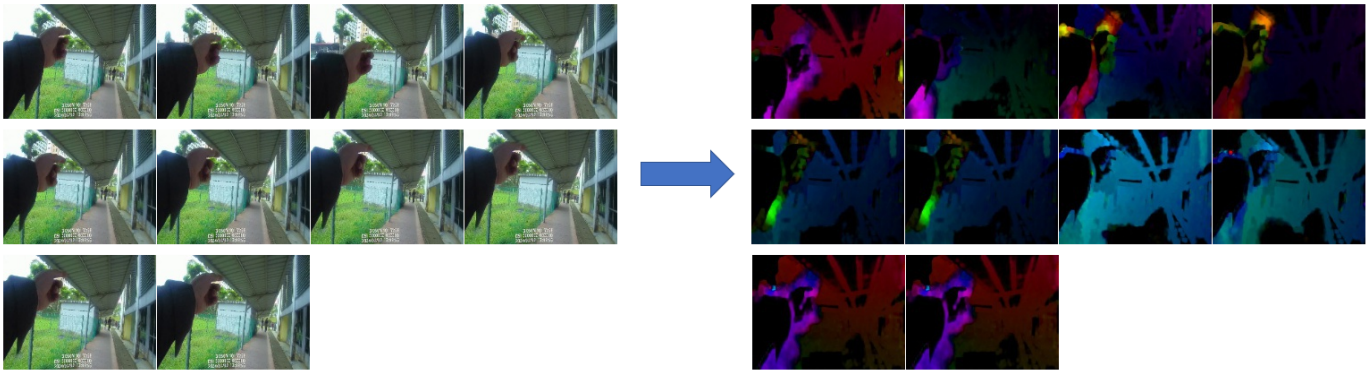

Fig. 2: **Frames**: For a particular subject, the rgb and optical flow frames. We show 10 consecutive frames. the rgb frames show the walk and pattern for a particular time sequence. For optical flow, each frame represents motion information in terms of direction (hue) and speed (intensity).

Among the models, I3D demonstrated the most consistent performance, achieving a test accuracy of 89.9% and a balanced F1-score of 0.90, showcasing its robust ability to capture temporal dynamics in egocentric motion. The TimeSformer model, with its transformer-based architecture, also achieved competitive results, with a test accuracy of 89.23% and F1-score of 0.89, demonstrating its capability to model long-range temporal dependencies. While SlowFast exhibited slightly lower performance with a test accuracy of 88%, it maintained a precision and recall of 88%. C3D demonstrated the lowest test accuracy at 85.36%, highlighting its limitations in capturing the complexity of egocentric motion. Validation accuracy remained consistent across the models, further emphasizing the generalization ability of I3D, SlowFast, and TimeSformer.

These experiments highlight the strength of I3D and TimeSformer in leveraging temporal motion patterns, while traditional 3D convolutional models like C3D lag in performance.

Table I shows the accuracy of different models across all the phases as discussed above.

TABLE I: Models comparison.

Models	Train	Validation	Test
C3D	92.7	86.05	85.36
I3D	98	90.33	**89.9**
SlowFast	96.5	90.38	88
TimeSformer	97.36	90.43	89.23

B. OPTICAL FLOW user recognition

The optical flow frames capture the motion between consecutive frames, revealing how the subject moves through space. This dynamic information is crucial for identifying users based on their movement patterns, such as walking, running, or other locomotion behaviors. However, optical flow alone lacks the rich spatial details of the subject's appearance. Once the optical flow frames were extracted, we experiment with I3D model in order to verify the user, where we receive an accuracy of 61.97% on the test set. Instead of using raw RGB frames (which contain detailed visual information), the model only had access to motion patterns. In situations where users exhibit similar motion patterns, optical flow frames do not provide enough discriminative information to reliably identify individuals, which explains the lower accuracy observed when using optical flow in isolation.

C. RGB + OPTICAL FLOW user recognition

To potentially improve the performance of user identification in egocentric video data, we experiment with the combination of RGB and optical flow frames. We employed a two-stream I3D architecture. The model processes both RGB and optical flow frames in parallel, leveraging spatial and temporal information to recognize actions more effectively. RGB frames capture the appearance information, such as textures, colors, and spatial structures, while optical flow frames focus on motion patterns and temporal dynamics by encoding pixel-level displacements between consecutive frames. During testing, both models process their respective inputs independently, and their outputs are averaged to produce a final prediction.

The two-stream I3D model was initialized with pretrained weights from the Kinetics-400 dataset. Dropout regularization with a probability of 0.5 was added to the fully connected layers to mitigate overfitting. The model was fine-tuned on the dataset using the Adam optimizer with a learning rate of 0.001 and weight decay of 1e-4 to incorporate L2 regularization. Training was performed for 20 epochs, and the performance was monitored using the cross-entropy loss function. The final prediction was made by averaging the softmax outputs of the RGB and optical flow streams.

The model achieved an accuracy of **91.19%** on the test set with a corresponding cross-entropy loss of 0.31, outperforming all single-stream architectures tested, including RGB-based I3D, C3D, SlowFast, and TimeSformer. The observed performance boost can be attributed to the complementary nature of the features learned from both modalities. The RGB frames help the model distinguish between users based on their appearance-related features, while the motion-related features from the optical flow provide additional context to differentiate individuals based on their movement patterns. The

fusion of both types of information allows the model to learn more robust, discriminative features, improving its ability to identify users with greater accuracy. This result highlights the advantage of integrating spatial and motion-based temporal features in the two-stream setup.

D. DOMAIN SHIFT user recognition testing

In this experiment, we aim to evaluate the ability to identify users based on egocentric motion patterns under domain shift conditions. Specifically, we investigate whether training on walking motion and testing on jogging can provide meaningful identification results.

While the test accuracy (62.89%) and other metrics indicate that the results are not entirely random, the observed performance highlights the challenge of transferring motion-specific features between distinct activities (walking to jogging). The performance remains above random chance. This suggests that user-specific motion patterns learned from walking retain some discriminative power when applied to jogging. However, the performance gap highlights the difficulty of transferring motion-specific features between distinct activities. This demonstrates the potential for using egocentric motion as a robust identifier even under changing activity conditions.

This experiment underscores the feasibility of user identification through egocentric motion, with implications for applications where domain adaptation between activities is required. Further exploration of domain generalization and activity-invariant feature learning techniques could improve performance in such scenarios.

VI. CONCLUSION

In this study, we introduce a novel dataset aimed at user identification through egocentric motion analysis, captured using BWCs. The dataset comprises recordings from 23 distinct subjects, with each video representing a unique individual's egocentric motion patterns. By providing synchronized RGB and optical flow frames, this dataset enables researchers to explore user identification tasks based on motion dynamics and visual features. Our dataset is specifically tailored for user identification by egocentric view, and represents a significant step forward in the field of egocentric video analysis.

We evaluate the dataset using state-of-the-art models, including single-stream and two-stream architectures, to establish robust benchmarks. The two-stream I3D model, leveraging RGB and optical flow modalities, achieved the best performance. This result highlights the advantage of combining spatial and temporal information for user identification tasks.

Our dataset sets a new benchmark for research in egocentric motion analysis and user identification. The strong performance of the two-stream I3D model establishes a solid baseline. This dataset provides a unique opportunity to develop and refine algorithms for egocentric motion-based identification in real-world scenarios. We believe this dataset will serve as a valuable resource for advancing the development of biometric systems and motion-centric video understanding in wearable camera applications.

Future work will explore advanced fusion techniques, such as attention mechanisms, to further enhance multi-stream architectures. Additionally, as a part of our ongoing research on BWCs, we aim to extend the dataset for other tasks, based on face recognition and multi-action recognition in complex environments, with actions that are uniquely relevant to law enforcement scenarios.

REFERENCES

[1] Bertasius, G., Wang, H., Torresani, L.: Is space-time attention all you need for video understanding? In: Proceedings of the International Conference on Machine Learning (ICML) (July 2021)

[2] Bryan, J.: Effects of Movement on Biometric Facial Recognition in Body-Worn Cameras. PhD thesis, Purdue University Graduate School (5 2020). https://doi.org/10.25394/PGS.12227372.v1

[3] Carreira, J., Zisserman, A.: Quo vadis, action recognition? a new model and the kinetics dataset. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (July 2018)

[4] Chao, X., Hou, Z., Mo, Y.: Czu-mhad: A multimodal dataset for human action recognition utilizing a depth camera and 10 wearable inertial sensors. IEEE Sensors Journal **22**, 1–1 (04 2022). https://doi.org/10.1109/JSEN.2022.3150225

[5] Chen, H., Li, H., Song, A., Haberland, M., Akar, O., Dhillon, A., Zhou, T., Bertozzi, A.L., Brantingham, P.J.: Semi-supervised first-person activity recognition in body-worn video. arXiv preprint arXiv:1904.09062 (2019)

[6] Choi, S., Michalski, N.D., Snyder, J.A.: The "civilizing" effect of body-worn cameras on police-civilian interactions: Examining the current evidence, potential moderators, and methodological limitations. Criminal Justice Review **48**(1), 21–47 (2023). https://doi.org/10.1177/07340168221093549

[7] Corso, J.J., Alahi, A., Grauman, K., Hager, G.D., Morency, L.P., Sawhney, H., Sheikh, Y.: Video analysis for body-worn cameras in law enforcement. arXiv preprint arXiv:1604.03130 (2018)

[8] Damen, D., Doughty, H., Farinella, G., Fidler, S., Furnari, A., Kazakos, E., Moltisanti, D., Munro, J., Perrett, T., Price, W., Wray, M.: The epic-kitchens dataset: Collection, challenges and baselines. IEEE Transactions on Pattern Analysis and Machine Intelligence **PP**, 1–1 (05 2020). https://doi.org/10.1109/TPAMI.2020.2991965

[9] Feichtenhofer, C., Fan, H., Malik, J., He, K.: Slowfast networks for video recognition. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 6201–6210 (2019). https://doi.org/10.1109/ICCV.2019.00630

[10] Hoshen, Y., Peleg, S.: An egocentric look at video photographer identity. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 4284–4292 (2016), https://api.semanticscholar.org/CorpusID:14144375

[11] Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Fei-Fei, L.: Large-scale video classification with convolutional neural networks. In: CVPR (2014)

[12] Meng, Z., Sánchez, J., Morel, J.M., Bertozzi, A.L., Brantingham, P.J.: Ego-motion classification for body-worn videos. In: Tai, X.C., Bae, E., Lysaker, M. (eds.) Imaging, Vision and Learning Based on Optimization and PDEs. pp. 221–239. Springer International Publishing, Cham (2018)

[13] Ridnik, T., Ben-Baruch, E., Noy, A., Zelnik-Manor, L.: Imagenet-21k pretraining for the masses (2021), https://arxiv.org/abs/2104.10972

[14] Suat Cubukcu, Nusret Sahin, E.T., Topalli, V.: The effect of body-worn cameras on the adjudication of citizen complaints of police misconduct. Justice Quarterly **40**(7), 999–1023 (2023). https://doi.org/10.1080/07418825.2023.2222789

[15] Todak, N., Gaub, J.E., White, M.D.: Testing the evidentiary value of police body-worn cameras in misdemeanor court. Crime & Delinquency **70**(4), 1249–1273 (2024). https://doi.org/10.1177/00111287221120185

[16] Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M.: Learning spatiotemporal features with 3d convolutional networks. In: 2015 IEEE International Conference on Computer Vision (ICCV). pp. 4489–4497 (2015). https://doi.org/10.1109/ICCV.2015.510

[17] Tsutsui, S., Fu, Y., Crandall, D.: Whose hand is this? Person Identification from Egocentric Hand Gestures . In: 2021 IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 3398–3407. IEEE Computer Society, Los Alamitos, CA, USA (Jan 2021). https://doi.org/10.1109/WACV48630.2021.00344

APPENDIX

We show the results of the best performing model, in the case of single-stream networks (I3D) for our specific task of user identification. Figure 3 shows the ROC curve for I3D model. The results are based on Table I.

To further assess the biometric reliability of our egocentric motion-based user identification system, we compute the False Match Rate (FMR) and False Non-Match Rate (FNMR).

FMR: measures the probability that a different subject (impostor) is mistakenly classified as the target subject. It is the estimated error of a biometric authentication system in which it incorrectly matches two entirely different individuals and identifies them as the same person.

$$FMR = \frac{FP}{\text{Imposter Attempts}}$$

- **False Positives (FP)**: These are cases where the model predicts a sample to belong to the wrong subject (i.e., the predicted label is different from the true label).
- **Impostor Attempts**: This is the total number of attempts where the true subject is not the same as the predicted subject (i.e., the total number of predictions for other subjects excluding the correct one).

FNMR: FNMR measures the probability that the model fails to recognize a match when it should have.

$$FNMR = \frac{FN}{\text{Genuine Attempts}}$$

- **False Negatives (FN)**: These are cases where the model incorrectly classifies a genuine match (same subject) as a mismatch (different subject).
- **Genuine Attempts**: refer to the total number of classification attempts where a subject is being compared against their own identity.

FNMR was computed per subject and overall (by dividing the total FN by the total genuine attempts across all subjects). Similarly, overall FMR was calculated by dividing the total FP by the total impostor attempts, considering only the cases where a misclassification occurred. In addition to reporting the overall performance, we also provided a per-subject breakdown of FMR and FNMR to offer a deeper insight into how well the system performs across different individuals. In total, we get 89 FP and 86 FN for the test set with 941 video samples. The overall FNMR of 0.0965 suggests that **9.65%** of genuine attempts were misclassified, while the overall FMR of 0.0999 implies that approximately **10%** of impostor attempts were falsely accepted. The model seems to perform relatively well overall, but there are significant discrepancies between individual subjects. Some subjects are identified correctly with minimal errors, while others have higher rates of misclassification. Figure 4 shows the FMR and FNMR per subject using the I3D model.

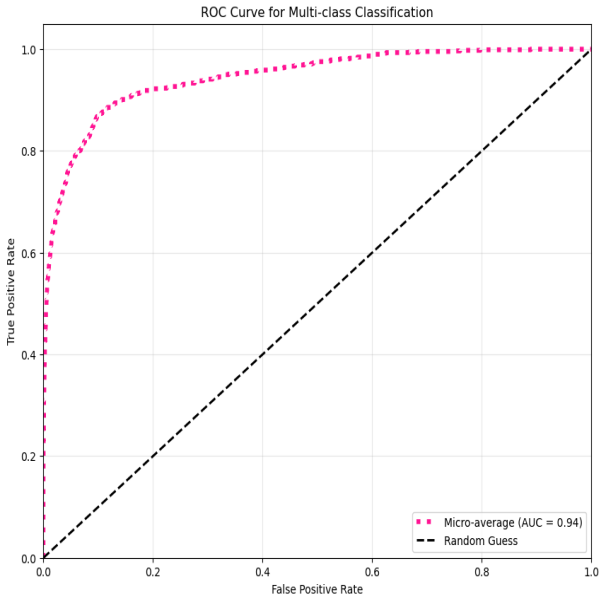

Fig. 3: The micro-average ROC curve. The AUC is measured at 0.9837, indicating an excellent ability of the model to distinguish between classes. This curve aggregates all predictions across all classes into a single evaluation, ensuring that each instance contributes equally to the overall performance.

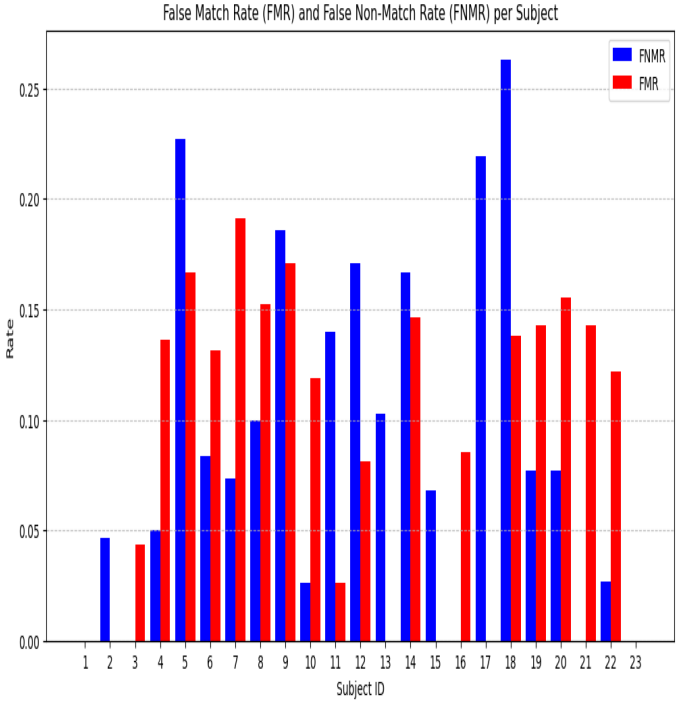

Fig. 4: The FMR variation across subjects highlights the vulnerability of certain subjects to false matches, indicating possible feature similarities among different identities. The FNMR trend shows that some subjects experience significantly higher false rejections, suggesting challenges in recognizing genuine attempts, possibly due to intra-class variations.