Rider Posture-based Continuous Authentication with Few-Shot learning for Mobility Scooters*

Devan Shah [1] [†], Ruoqi Huang [2], Tingting Chen [2], Murtuza Jadliwala [3]

[1] Princeton University
[2] California State Polytechnic University, Pomona
[3] University of Texas, San Antonio
ds6237@princeton.edu, {ruoqihuang, tingtingchen}@cpp.edu, murtuza.jadliwala@utsa.edu

Abstract

Current practice of mobility scooter user authentication using physical keys and traditional password-based one-time security mechanisms cannot meet the needs of many mobility scooter riders, especially senior citizens having issues in recalling memory. New seamless authentication approaches are needed to provide ongoing protection for mobility scooters against takeovers and unauthorized access. Existing continuous authentication techniques do not work well in a mobility scooter setting due to issues such as user comfort, deployment cost and enrollment time, among others. In that direction, our contributions in this research effort are two-fold: (i) we propose a novel system that incorporates advances in few-shot learning, hierarchical processing, and contextual embedding to establish continuous authentication for mobility scooter riders using only posture data. This security system, trained on data collected from real mobility scooter rides, demonstrates quick enrollment and easy deployability, while successfully serving as an unobtrusive first layer of security. (ii) we provide to the research community the largest publicly available repository of mobility scooter riders' body key-points data to enable further research in this direction.

Introduction

Although mobility scooters provide powerful ways to help people with mobility challenges, especially senior citizens with meeting their transportation needs, there is very little research focusing on improving mobility scooter security in smart and connected communities compared to other powered micromobility vehicles (Vinayaga-Sureshkanth et al. 2020). Current practice of user authentication using physical keys and traditional password-based one-time security mechanisms fall short in accommodating many mobility scooter users, such as those having issues in recalling memory due to age-related diseases and dysfunctions. There is a clear need for new user-friendly and seamless authentication approaches that provide ongoing protection for mobility scooters against takeovers and unauthorized access.

*This work is supported partly by grants NSF CNS #2318671, NSF CNS #2050826

†Postal Address: 6869 Frist Campus Center, Princeton, New Jersey, 08544 USA. Phone Number: +1 (845) 596-5179.

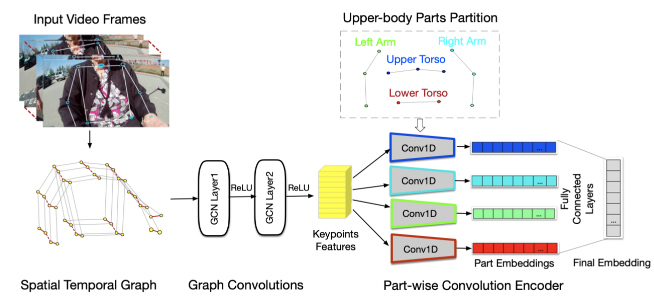

Figure 1: Framework to Generate Upper-body Movement Embeddings of Mobility Scooter Riding

Recent advances in deep learning models and techniques have enabled accurate continuous user authentication using behavioral biometrics such as gait, keystroke, pulse or touch. However, much research in tangential fields uses special sensors or has a long enrollment time, and thus cannot address the challenges in the mobility scooter setting with users' need for riding comfort and ease of deployment.

In this work, for easier and unobtrusive deployment, we leverage riders' posture data (excluding the face) captured from user-facing cameras to extract features for continuous user-authentication of mobility scooter riders. We propose a novel deep architecture for learning user-specific embedding vectors from few training samples by employing the sequence of users' (or riders') upper-body keypoint coordinates detected during enrollment. Prior works such as (Coskun et al. 2018) also utilize deep models in gait analysis for person identification. However, user postures on mobility scooters show unique characteristics in the movements of upper body parts, different from those in full-body movements. We leverage Graph Neural Networks (Li, Zhao, and Ma 2020) and a body-part-based hierarchical encoding structure, which have strengths in extracting features representing unique spatial correlations of upper body keypoints when riding mobility scooters.

Deep Representation of Mobility Scooter Riding Postures

Model Architecture

The continuous video stream of the mobility scooter rider is processed into 128-frame segments, each segment producing one authentication decision. As shown in Figure 1, we process each frame to gather the coordinates of 9 torso keypoints (left/right shoulders, left/right elbows, left-/right wrists, left/right hips and neck). With a sequence of 9 keypoints detected from 128 frames, we perform spatio-temporal graph convolutions on the graph derived from naturally connecting the joints at each frame and connecting joints with themselves in the prior and posterior frame. Such networks intuitively allow each joint to "understand" its spatial and temporal relationship with surrounding joints.

The model then contains 4 residual convolutional encoders, one for each body part (i.e., Upper Torso, Lower Torso, Left Arm, and Right Arm). Each convolution is followed by the ReLU activation function and batch normalization, and the encoder contains 5 residual convolutions and 2 max pooling layers. The residual connections prevent over-processing and addresses the vanishing gradient problem. Having such encoders enables a pyramid of feature extraction. Each encoder produces a 64-dimensional segment embedding via global average pooling of the final 64 channels. Given these four embeddings, a two-layer fully connected network with a ReLU activation then produces a single 64-dimensional embedding for the 128-frame video sample.

Model Training

Triplet Metric Loss is used to minimize Euclidean distance between embeddings from the same user, described as: $L(x^a, x^p, x^n) = \max(||f(x^a) - f(x^p)||_2 - ||f(x^a) - f(x^n)||_2 + \alpha, 0)$, where x^a represents an anchor embedding, x^p is an embedding from the same user as x^a, and x^n is an embedding from a different user. α represents the mar-gin between positive and negative pairs, and f is the model function. To minimize the loss, the model aims to achieve $||f(x^a) - f(x^p)||_2 + \alpha < ||f(x^a) - f(x^n)||_2$, thus embeddings from the same rider will be close, whereas those from different riders will be further away, with an enforced margin between samples from different classes of α.

Rather than using all pairs of positive and negative samples, we accelerate model training by performing Triplet Mining. For each anchor in a batch, we choose from the batch the closest positive embedding and the closest negative embedding that is not closer than the positive embedding to produce challenging triplets for the model to train on.

Authentication Decision

To enroll in the authentication system, the user rides the mobility scooter for 3 minutes, and random video segment samples of user riding are processed to produce embeddings, which are stored collectively. A rider is authenticated based on the Euclidean distance of the embedding of their current riding and the user's enrollment embeddings.

Data Collection and Data Set

We gathered mobility scooter riding data from 42 individuals on campus as described in Table I. They completed the riding tasks in an average of 15 minutes, which included forward riding, backwards riding, 90° and 45° left and right turns, 360° rotations, both on-pavement and on-grass riding, and sudden acceleration and deceleration. The sequence of numbers in Figure 2 depicts the riding route.

Table 1: Participants #

Age	Female	Male
18-25	4	30
26-60	2	2
>60	3	1
Total	9	33

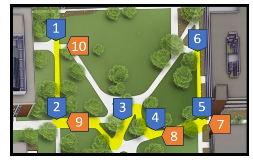

Figure 2: Riding Route

After filtering out video segments where participants are not following the tasks, we have approximately 10 hours and 9 minutes of recorded video footage, and a total of 1.1 mil-lion frames. The dataset is available at github.com/Mobility-Scooter-Project/Public-Data.

Evaluation

We test two pose estimation models MediaPipe and MoveNet (both from Google) to generate the keypoint coordinates in our system and other components and experiment settings are kept the same. Table 2 illustrates the Area Under the Curve (AUC) metric for the ROC curve of the authentication system. The two system variations both yield high accuracy using few enrollment samples. We also note the system based on MediaPipe has greater variability and is less accurate than when using MoveNet.

Table 2: ROC AUC of the Authentication System

Pose Est.	Enrollment Samples				
	1	5	10	20	40
MoveNet	0.976	0.986	0.990	0.987	0.990
MediaPipe	0.98.1	0.896	0.962	0.960	0.964

Conclusion

This work provides a continuous authentication system for mobility scooter riders. Our model leverages spatio-temporal graph convolutions before a hierarchical encoding structure to produce embeddings and is trained with Triplet Metric Loss. Experimental results based on real mobility scooter riders' data show that our system is easy to deploy and accurate.

References

Coskun, H.; Tan, D. J.; Conjeti, S.; Navab, N.; and Tombari, F. 2018. Human Motion Analysis with Deep Metric Learning. arXiv:1807.11176.

Li, N.; Zhao, X.; and Ma, C. 2020. JointsGate: Gate Recognition based on Graph Convolutional Networks and Joints Relationship Pyramid Mapping.

Vinayaga-Sureshkanth, N.; Wijewickrama, R.; Maiti, A.; and Jadliwala, M. 2020. Security and privacy challenges in upcoming intelligent urban micromobility transportation systems. In Proceedings of the Second ACM Workshop on Automotive and Aerial Vehicle Security, 31–35.