EE6563 Project Progress
Footprint Recognition based on the
Spatio-Temporal Features

Saeed Kazemi[1]

University of New Brunswick

Abstract

Given present-day security concerns, many buildings have implemented robust authentication techniques. Aside from authentication to enter a building, applications such as border and airport security also administer identification. Therefore, many cities and companies provide technologies like CCTV or fingerprinting for authentication and verification. But each system has its own drawbacks. For example, due to the Covid-19 pandemic, most people wear a mask and avoid touching unnecessary surfaces. Thus, gait recognition could be a solution. In this project, we worked on temporal information of time-series data. These features used to construct a classifier. Moreover, verification mode was used for this research.

Keywords: Footprint recognition, Time series, pressure sensor

1. Introduction

Contemporary security identification has led to a plethora of biometric-based authentication systems. From palm readers at testing centers to facial recognition on smartphones, many systems are now used to regularly verify identities. These inherence-based systems are attractive in comparison to knowledge or possession-based methods of authentication because biometrics are unique, unforgettable, and far more difficult to steal than a password or swipe card.

Although biometric identification appears sophisticated when compared to something like physical keys, it does not come without its own share of caveats. For example, owing to the ongoing Covid19 pandemic, many people wear masks when outside of their house, challenging most facial recognition systems. Additionally, biometrics that rely on touch, such as fingerprinting, raise safety concerns, as the scanner may become a vector for virus transmission. Despite these setbacks, given their merits and widespread deployment, biometric identification systems are unlikely to disappear.

One behavioral biometric that has gained recent success and is worth further consideration given current constraints is gait recognition. Usage of gait recognition has grown in the security industry in recent decades due to advances in deep learning. Singh et al. [1] categorized gait recognition into two main categories, vision-

---

[1]Saeed.Kazemi@unb.ca.

based and sensor-based. In vision-based approaches, cameras capture data of a person walking for the purpose of gait recognition. Sensor-based gait recognition is performed using either wearable sensors which produce kinematic data, or floor sensors which produce kinetic data [2].

This project uses two different datasets, UoM-Gait-Dataset and Stepscan. The UoM-Gait-Dataset was obtained from iMAGiMAT sensor [3] (an optical floor sensor). This sensor contains about 160 distributed plastic optical fibers (POFs) that indicate the foot pressure signals over time. Some studies like [4] used this dataset to construct images for their research.

The second dataset that was used for this project is called the Stepscan dataset [5]. This dataset was obtained from high-resolution floor tiles that have recently been introduced by Stepscan Technologies Inc.

Figure 1 indicates three frames from this dataset. This dataset consists of a spatial-temporal tensor, X, with dimensions $S \times T \times H \times W$ where S represents the number of samples. T is the number of temporal observations or video frames; H and W are the dimension of the image in pixels. Moreover, both datasets include information on different walking speeds.

This project aims to find some temporal features from the datasets to construct a classifier for verification purposes. The nature of data in the first dataset is time series, whereas Stepscan is a video-base dataset. There are several methods for converting a tensor (like video) to 2D time-series data.

Chen et al. in [6] used contour width for defining a one-dimensional signal. They utilized some morphological operations on the background-subtracted silhouette image to extract the outer contour. Afterwards, according to the contour width of each image row, a one-dimensional signal was generated.

Another method could be that the pixel values in each frame are plotted over time. By this means, the $H * W$ time-series will be produced for each sample.

In the final method, some spatial features are extracted from each frame (e.g. centroid and maximum pressure in each image). Afterwards, we track these values over time (next frames). As a result, 3D videos with size $T \times H \times W$ are converted to the four 2D time-series data. Costilla-Reyes et al. utilized this method to combine the output of 160 distributed POFs [7].

In this research, the last mentioned method was applied to produce time-series data. Figure 2 indicates the time series extracted from the Stepscan dataset. The spatial features extracted in each frame were maximum pressure (figure 2a), the center of pressure (COP) (figures 2d and 2c), and the average pressure (figure 2b).

In general, there are two modes for footprint recognition or generally in the biometric system: verification or identification mode [8]. In verification mode, the biometric system is used for accessing buildings or data. In other words, the system compares the claimed person with its dataset to determine whether or not the claim is valid. These systems consume less processing power and time [8].

This project will focus on verification. We hope to engineer some temporal features which help us to classify participants. Furthermore, this research will be implemented in Python, and the source codes are available on the GitHub repository [9].

The rest of this research is organized as follows: Section 2 provides the relevant works and researches. Then, in Section 3, the time domain features as well as machine learning models are defined. In Section 4, we discuss the results, and we propose future work in Section 5.

2. Literature Review

DARPA, the Defense Advanced Research Projects Agency of the USA, started to research gait recognition by vision data in the early 2000s [2]. Besides vision data [6], some studies have instead used accelerometry from smartphones [10], audio [11], and underfoot pressures data [12].
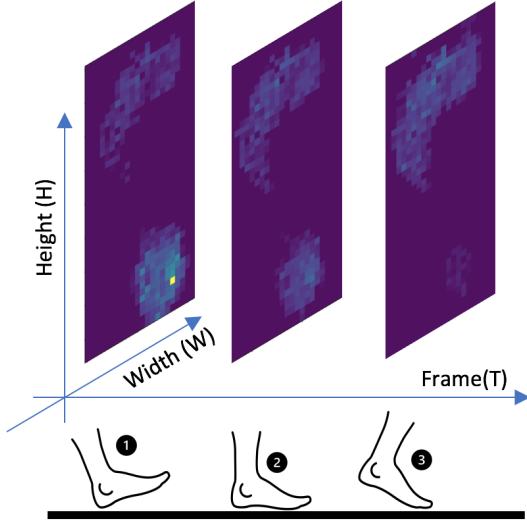
Figure 1: Different frames of footprint video in Stepscan dataset.

Addlesee in [13] used a new sensor (Active floor) for the first investigations into footprint recognition. This sensor was a square carpet tile maintained at the corners by some load cells and supplied the Ground Reaction Forces (GRFs). Orr and Abowd [14] extracted ten temporal features from the GRFs curve.

Moustakidis et al. [15] extracted temporal features from the wavelet decomposition of GRFs and then applied a kernel-based support vector machine. These studies have limitations in terms of the small sample sizes used for classification (e.g., 15 [14], 10 [15], and 15 [16]), and moderate classification rates ($CR < 90\%$).

Pataky in [17] could achieve a 99.6% classification rate in a 104-participant dataset. This result was based on spatial alignment and automated dimensionality reduction. He used a template image that was made in [18]. Pataky named this template the Munster-104 template.
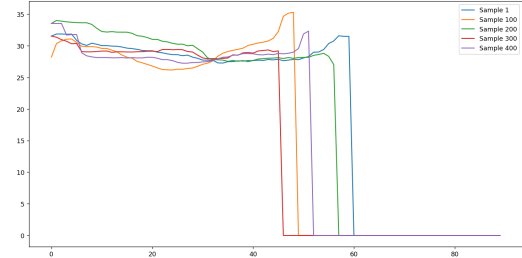
In 2015, Cantoral-Ceballos [3] introduced an intelligent carpet system. This carpet system (iMAGiMAT) worked based on the deformation of 116 distributed POFs. So that applying pres-
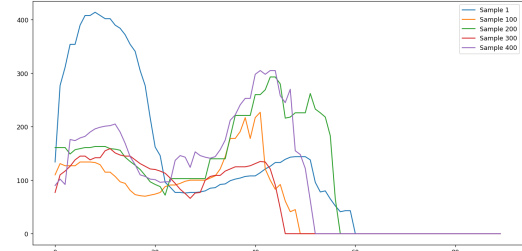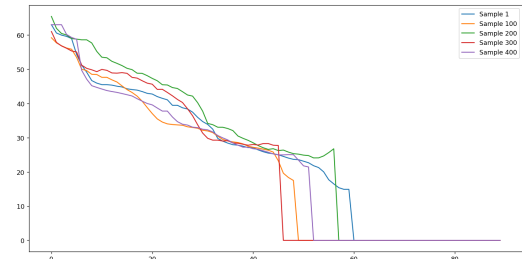

(a) The maximum pressure in each frame
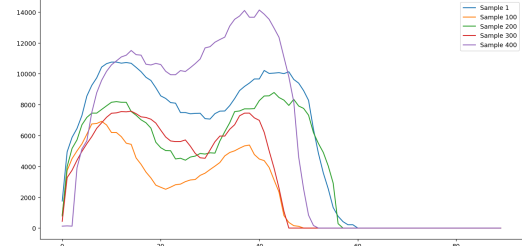
(b) The average pressure in each frame

(c) The x position in the center of pressure (COP) in each frame

(d) The y position in the center of pressure (COP) in each frame

Figure 2: The time series extracted from the Stepscan dataset based on four spatial features. The horizontal axis indicates the frame number.

3

sure to this system would change the intensity of the transmitted light. Thus the nature of the output of this sensor is time-series data.

Costilla-Reyes et al. [19] extracted five features directly from raw data of the iMAGi-MAT sensor. These features were spatial Average (SA), standard deviation (SD), adjacent mean (AM), cumulative sum (CS), and cumulative product (CP). They implemented 14 various machine learning methods for classification. The best result belonged to the Random Forest model with a validation score of $90.84 \pm 2.46\%$.

Later, in [7] they used an end-to-end convolutional neural network to extract Spatio-temporal features automatically. This technique increased their score by about 7 percent.

3. Features Extraction and Classifiers

In this section, we describe algorithms along with their results. These are including Features Extraction and Selection, hyper-parameters of the classifier, and evaluation methodology. Afterward, the results are reported.

3.1. Features Extraction and Selection

As mentioned before, our goal is to develop a classification model. For this classification task, we use about 34 features in four categories. These feature sets are explained briefly here, and more details about them can be found in Appendix A.

After extracting features from each time series, some low variance and high-correlated features were eliminated. Feature selection causes the complexity of the model to reduce. Ten percent of handcrafted features were set aside for testing our classifier, and others were divided into 10-fold cross-validation for evaluation and training.

To have equally balanced classes in the test set and cross-validation set, the methods Stratified and StratifiedKFold were used. As a result, all subjects were included in the test result.

3.1.1. Temporal Features

The first set of features extracted was temporal features. In this set, we focused on features that related to the time axis. Features like Entropy, Absolute energy, Centroid, Area under the curve fall into this group. The number of features extracted was 11.

3.1.2. Statistical Features

Statistical information was another feature set that was extracted from the dataset. Min, Max, variance, and standard deviation were some of the statistical features. The total number of features in this group is about 10 for each time series.

3.1.3. Spectral Features

In the third category, both FFT and wavelet transform were used to extract spectral information from the dataset. Not only the time complexity but also the number of features were more than two other feature sets. Max power spectrum, Maximum frequency, Spectral centroid, Wavelet energy, and FFT mean coefficient were spectral features extracted from the dataset.

3.1.4. Autoregressive model (AR) coefficients

The final set of features in this research was the coefficients of AR. In this set, the first-order differencing used to make our data stationary. Then based on significant lag on Partial Auto-Correlation function (PACF), the order of the model was selected. This approach extracted two features for each time-series signal.

3.2. Machine learning algorithms

These time-series features were fed to four different types of machine learning models with tuned hyper-parameters. The hyper-parameters tuning was implemented by grid search and performance evaluation with a 10 StratifiedKFold cross-validation. Table 1 shows these models along with their best-tuned hyper-parameters.

4

3.3. Results Progress

Results of four different machine learning algorithms in several types of features are shown in table 1. For comparing algorithms, the accuracy metric was used.

The LDA classifier's best performance was a 69.2% accuracy, as shown in table 1. This result comes from classifying the combination of all features. The kNN algorithm had similar performance on all kinds of features except AR features. All algorithms had the worst results on this group of features.

The SVM performed best with the mixture of all features, at 56% accuracy (see table 1). All other feature subsets achieved lower than this amount. The worst performance for all these classifiers belonged to the random forest classifier with about 30%.

4. Discussion Progress

The results indicate that spectral features are more powerful features for all machine learning algorithms. The reason behind this might be the number of features in this set. Based on table A.2, there are about 300 features for this set, much more than two others.

Another feature that has a significant effect on the accuracy is inter stride distance. This feature has been added only to all features and increased the accuracy of LDA from 60% to 69%.

The results might suggest that using AR features are not significant impact on accuracy. The accuracy of LDA without these features reduced to 68.57%.

Since we used one window for extracting temporal and statistical features, these results are not reliable for the quality of these two sets. Furthermore, it should be taken into account that the times-series signals were included many zeros due to the fact that videos were not normalized in the time axis. Thus, these zeros could bring a negative impact on temporal and statistical features.

5. Remaining Work

In recent years, improvements in the computational process of computers and other benefits of Deep Neural Networks (DNN) have caused many researchers (like [20] and [7]) to move towards DNN for the classification of time series data. Therefore, these algorithms will be reviewed in a later stage of this project.

Because the entire time series were considered for extracting features, we have had one value for each attribute. For instance, while local maximums could be good features, the algorithm returned only the global maximum as a Max feature. We could slide a window on the signal to extract handcrafted features. It causes only a part of the data to consider in each moment. In other words, because each data includes several stages of the walking cycle, features should be extracted according to these periods.

Moreover, cropping the heel or toe area in each video could separate features based on the walking stages (pressure area). By this means, each video will be divided into N sub-area, and then for each region, we reconstruct time-series signals and extract features. As a result, these features belong only to that area or stage of walking.

In this research, regardless of eliminating high-correlated and low variance features, we should use a more complex feature selection algorithm to eliminated irrelevant and redundant ones. This reduction might increase the accuracy.

References

[1] J. P. Singh, S. Jain, S. Arora, U. P. Singh, A Survey of Behavioral Biometric Gait Recognition: Current Success and Future Perspectives, Archives of Computational Methods in Engineering (2019). `doi:10.1007/s11831-019-09375-3`.

[2] P. Connor, A. Ross, Biometric recognition by gait: A survey of modalities and features, Computer Vision and Image Understanding 167 (2018) 1–27. `doi:10.1016/j.cviu.2018.01.007`.

Table 1: Machine learning models along with its best-tuned hyper-parameters.

Classifier	*Features*[*]	Hyper-parameter(s)	Accuraccy
LDA	All	'n_components': 10	69.28%
kNN	All	'n_neighbors': 9, 'metric': manhattan,'weights': distance	42.85%
SVM	All	'kernel': rbf, 'decision_function_shape': ovr, 'C': 1000, 'gamma': 0.0001	60.00%
RFC[**]	All	'criterion': entropy, 'max_depth': 11	28.57%
LDA	temporal	'n_components': 10	47.43%
kNN	temporal	'n_neighbors': 21, 'metric': manhattan,'weights': distance	34.29%
SVM	temporal	'kernel': rbf, 'decision_function_shape': ovr, 'C': 10, 'gamma': 0.01	46.29%
RFC	temporal	'criterion': entropy, 'max_depth': 9	28.57%
LDA	statistical	'n_components': 10	48.00%
kNN	statistical	'n_neighbors': 15, 'metric': manhattan,'weights': distance	36.57%
SVM	statistical	'kernel': linear, 'decision_function_shape': ovr, 'C': 10, 'gamma': 1	43.43%
RFC	statistical	'criterion': gini, 'max_depth': 23	26.86%
LDA	spectral	'n_components': 10	57.71%
kNN	spectral	'n_neighbors': 7, 'metric': manhattan,'weights': distance	39.43%
SVM	spectral	'kernel': linear, 'decision_function_shape': ovr, 'C': 0.1, 'gamma': 1	48.57%
RFC	spectral	'criterion': entropy, 'max_depth': 7	32.00%
LDA	AR	'n_components': 1	4.57%
kNN	AR	'n_neighbors': 21, 'metric': manhattan,'weights': distance	2.86%
SVM	AR	'kernel': linear, 'decision_function_shape': ovr, 'C': 1000, 'gamma': 1	8.00%
RFC	AR	'criterion': gini, 'max_depth': 19	1.14%

* All features were standardized.
** Random Forest Classifier

[3] J. A. Cantoral-Ceballos, N. Nurgiyatna, P. Wright, J. Vaughan, C. Brown-Wilson, P. J. Scully, K. B. Ozanyan, J. A. Cantoral-Ceballos, N. Nurgiyatna, P. Wright, K. B. Ozanyan, J. Vaughan, P. J. Scully, Intelligent Carpet System, Based on Photonic Guided-Path Tomography, for Gait and Balance Monitoring in Home Environments, IEEE SENSORS JOURNAL 15 (1) (2015) 279. doi:10.1109/JSEN.2014.2341455. URL http://ieeexplore.ieee.org.

[4] O. Costilla-Reyes, R. Vera-Rodriguez, A. S. Alharthi, S. U. Yunas, K. B. Ozanyan, Deep learning in gait analysis for security and healthcare, Studies in Computational Intelligence 865 (2020) 299–334. doi:10.1007/978-3-030-31760-7{_}10.

[5] P. C. Connor, Comparing and combining underfoot pressure features for shod and unshod gait biometrics, in: 2015 IEEE International Symposium on Technologies for Homeland Security, HST 2015, Institute of Electrical and Electronics Engineers Inc., 2015. doi:10.1109/THS.2015.7225338.

[6] C. Chen, J. Liang, H. Zhao, H. Hu, Gait recognition using Hidden Markov model, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 4221 LNCS (60402038) (2006) 399–407. doi:10.1007/11881070{_}56.

[7] O. Costilla-Reyes, P. Scully, K. B. Ozanyan, Deep Neural Networks for Learning Spatio-Temporal Features From Tomography Sensors, IEEE Transactions on Industrial Electronics 65 (1) (2018) 645–653. doi:10.1109/TIE.2017.2716907.

[8] A. K. Jain, A. Ross, S. Prabhakar, An Introduction to Biometric Recognition, IEEE

Transactions on Circuits and Systems for Video Technology 14 (1) (2004) 4–20. doi:10.1109/TCSVT.2003.818349.

[9] SKazemii/EE6563.
URL https://github.com/SKazemii/EE6563

[10] J. Mäntyjärvi, M. Lindholm, E. Vild-jiounaite, S. M. Mäkelä, H. Ailisto, Identifying users of portable devices from gait pattern with accelerometers, in: ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, Vol. II, 2005. doi:10.1109/ICASSP.2005.1415569.

[11] J. T. Geiger, M. Hofmann, B. Schuller, G. Rigoll, Gait-based person identification by spectral, cepstral and energy-related audio features, in: ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2013, pp. 458–462. doi:10.1109/ICASSP.2013.6637689.

[12] K. Nakajima, Y. Mizukami, K. Tanaka, T. Tamura, Footprint-Based Personal Recognition, Tech. Rep. 11 (2000).

[13] M. D. Addlesee, A. Jones, F. Livesey, F. Samaria, The ORL active floor, IEEE Personal Communications 4 (5) (1997) 35–41. doi:10.1109/98.626980.

[14] R. J. Orr, G. D. Abowd, The Smart Floor: A Mechanism for Natural User Identification and Tracking, Tech. rep. (2000).

[15] S. P. Moustakidis, J. B. Theocharis, G. Giakas, Subject recognition based on ground reaction force measurements of gait signals, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 38 (6) (2008) 1476–1485. doi:10.1109/TSMCB.2008.927722.

[16] L. Middleton, A. A. Buss, A. Bazin, M. S. Nixon, A floor sensor system for gait recog-

nition, Tech. rep.
URL www.tekscan.com

[17] T. C. Pataky, T. Mu, K. Bosch, D. Rosenbaum, J. Y. Goulermas, Gait recognition: Highly unique dynamic plantar pressure patterns among 104 individuals, Journal of the Royal Society Interface 9 (69) (2012) 790–800. doi:10.1098/rsif.2011.0430.

[18] T. C. Pataky, K. Bosch, T. Mu, N. L. Keijsers, V. Segers, D. Rosenbaum, J. Y. Goulermas, An anatomically unbiased foot template for inter-subject plantar pressure evaluation, Gait and Posture 33 (3) (2011) 418–422. doi:10.1016/j.gaitpost.2010.12.015.

[19] O. Costilla-Reyes, P. Scully, K. B. Ozanyan, Temporal Pattern Recognition in Gait Activities Recorded with a Footprint Imaging Sensor System, IEEE Sensors Journal 16 (24) (2016) 8815–8822. doi:10.1109/JSEN.2016.2583260.

[20] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, P. A. Muller, Deep learning for time series classification: a review, Data Mining and Knowledge Discovery 33 (4) (2019) 917–963. doi:10.1007/s10618-019-00619-1.
URL https://doi.org/10.1007/s10618-019-00619-1

Appendix A. List of features

The list of features from several categories used in this project (Table A.2):

Table A.2: list of features

#	Features Name	Categories Name	# features
1	Histogram	Statistical	10
2	Max	Statistical	1
3	Mean	Statistical	1
4	Mean absolute deviation	Statistical	1
5	Median	Statistical	1
6	Median abs deviation	Statistical	1
7	Min	Statistical	1
8	Root mean square	Statistical	1
9	Standard deviation	Statistical	1
10	Variance	Statistical	1
11	AR coefficients	AR	8
12	Absolute energy	Temporal	1
13	Area under the curve	Temporal	1
14	Centroid	Temporal	1
15	Entropy	Temporal	1
16	Mean absolute diff	Temporal	1
17	Mean diff	Temporal	1
18	Median absolute diff	Temporal	1
19	Median diff	Temporal	1
20	Negative turning points	Temporal	1
21	Positive turning points	Temporal	1
22	Slope	Temporal	1
23	FFT mean coefficient	Spectral	256
24	Max power spectrum	Spectral	1
25	Maximum frequency	Spectral	1
26	Median frequency	Spectral	1
27	Spectral centroid	Spectral	1
28	Spectral entropy	Spectral	1
29	Wavelet abs mean	Spectral	10
30	Wavelet energy	Spectral	10
31	Wavelet stand deviation	Spectral	10
32	Wavelet entropy	Spectral	1
33	Wavelet variance	Spectral	10
34	Inter stride	Temporal	1
	The total number of features		341