FOOTSTEP SOUNDS AS BIOMETRIC AUTHENTICATION USING A DEEP CONVOLUTIONAL NEURAL NETWORK

Kristóf RÁCZ[1], **Rita M. KISS**[2]

[1] Budapest University of Technology and Economics, Department of Mechatronics, Optics and Mechanical Engineering Informatics, 1111 Műegyetem rkp. 3, Budapest, Hungary, E-mail: racz.kristof@mogi.bme.hu;

[2] Budapest University of Technology and Economics, Department of Mechatronics, Optics and Mechanical Engineering Informatics, 1111 Műegyetem rkp. 3, Budapest, Hungary, E-mail: rita.kiss@mogi.bme.hu;

1. Introduction

Various biometric authentication methods exist with different levels of security, such as fingerprint and retina scanners or face recognition [1]. Gait can also be used as a biometric authentication method [2]. Despite that the identification by gait is possible to perform unobtrusively, remotely and covertly, there are no prominent examples of real-world applications using it.

Typical gait measurements use video images or stereophotogrammetry. However, much of the information in gait is suspected to be also present in the sound or vibrations generated by walking. Although extracting this information is a lot less intuitive than from images, neural networks are capable of high levels of abstraction from data. The present study aims to examine how well a modern deep neural network can identify people based on the sound of their footsteps. This could have serious implication in todays world where smart devices such as phones can easily collect and analyze sound and vibration data, even without the knowledge of the user.

2. Methods

The footstep sounds of six individuals (3 male, 3 female, age 21-30 years, height 162-183 cm, body mass 58-85 kg) have been recorded in a total of 32 different configurations of footwear, with 20 takes for each configuration for a total of 640 sound samples. The recording equipment consisted of an AudioTechnica AT2020 condenser microphone connected to a laptop via a USB audio interface. Sound was recorded at 44.1 kHz with 24-bit resolution. The samples were filtered for background noise and converted into Mel spectrograms (Fig. 1) using python.

The spectrograms served as the input for a deep convolutional neural network. Using the Keras deep learning python library, transfer learning was performed on the InceptionV3 [3] image classification architecture, which was adapted for classification into six classes, one for each participant. The spectrograms were split into train-validate-test sets with 16-2-2 samples for every person–footwear combination. For training, the SGD optimizer was used with a learning rate of 0.001 and batch size of 16 based on hyperparameter optimization. Training time was approximately one hour in an online GPU accelerated runtime in the Google Colaboratory service (https://colab.research.google.com).

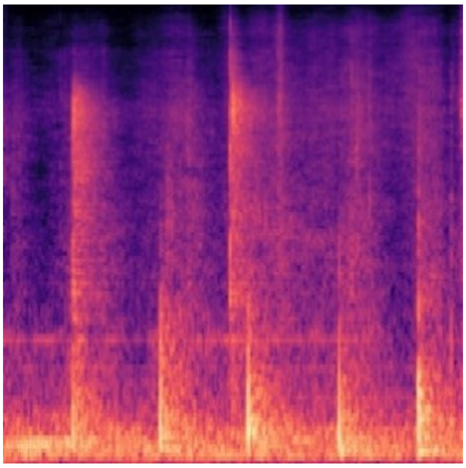

Fig. 1. Example of a spectrogram generated from a sound recording of footsteps.

3. Results

The network achieved 100% accuracy on both the training and validation datasets at the end of training, and 98% accuracy on the test dataset, which translates to one misclassified sample out of 64. The confusion matrix can be seen on Fig. 2.

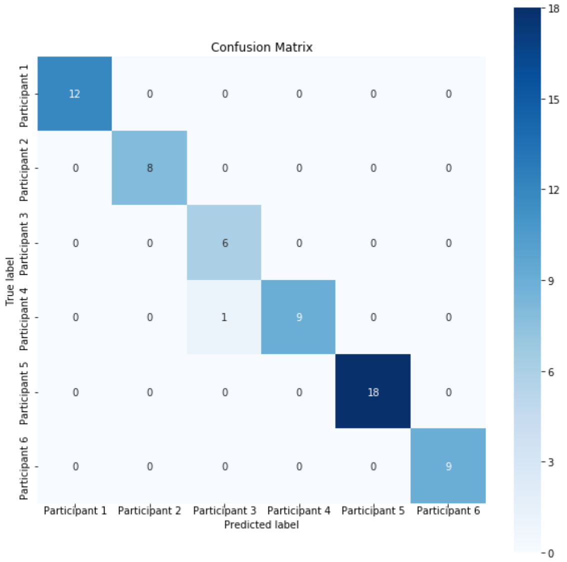

Fig. 2. Confusion matrix of the trained network for the test dataset.

4. Discussion

The unusually high level of accuracy indicates one or more of the following problems:

- the dataset is too small and homogenous,
- the network is more complex than it needs to be for this task (approximately 26.5 million parameters),
- there is something that is more characteristic of the person than their footstep sound in the recordings (for example the noise generated by the trousers worn).

Possible limiting factors, that will be tested in future include tolerance for noisy environments, shorter sound samples, more persons to identify or recognizing someone in new footwear. All these factors influence the robustness of the system, which determines usability in real world authentication scenarios. Possible applications in data collection also include determining the gender, height or body mass of a person based on their footstep sounds.

This should raise some serious ethical concerns regarding personal data and privacy. As mobile devices collect data almost non-stop, this could provide large data companies even more insight into our lives. Gait is a complex system that is not only representative of a person, but also changes based on our mood and health, data which could be very valuable to certain companies.

5. Conclusions

Deep neural networks can be used to extract information about the person based on the sound of their footsteps. This can be used in authentication/recognition but can also be potentially dangerous in today's world filled with smart devices, where privacy is an ever-growing concern.

Acknowledgements

This work was supported by the Hungarian Scientific Research Fund OTKA [grant number K135042].

References

[1] Chinegram, K. Various Biometric Authentication Techniques: A Review. *J Biom Biostat*, 2017, 8(5)

[2] Goffredo, M., Bouchrika, I., Carter, J.N., Nixon, M.S. Self-Calibrating View-Invariant Gait Biometrics. *IEEE transactions on systems, man and cybernetics part B – cybernetics,* 2010, 40(4), 997-1008

[3] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z. Rethinking the Inception Architecture for Computer Vision. *arXiv:1512.00567.*