Person recognition based on footstep sounds using a deep convolutional neural network

Lépéshang alapú személyfelismerés mély konvolúciós neurális hálóval

Kristóf Rácz
PhD student
Géza Pattantyús-Ábrahám Doctoral School of Mechanical Engineering Sciences
Budapest University of technology and Economics, Faculty of Mechanical Engineering,
Department of Mechatronics, Optics and Engineering Informatics
racz.kristof@mogi.bme.hu

Abstract

I'm not writing an abstract just yet. Who does that? It's gonna change by the time you finish anyways.

Absztrakt

Ugyanez magyarul.

1 Introduction

Biometrics can be used to identify people: from our fingerprints and iris to our DNA, numerous biological parameters can be used to identify a person. Security systems use this to control who has access to important places or information (e.g. unlocking our phone or pc via a fingerprint sensor), and Surveillance systems have been known to track people by recognizing their face. However, these systems are not infallible: contact lenses for iris scanners, printed silicon fingerprints for fingerprint scanners, masks for face recognition or voice recording for voice detection. However, it's also been shown numerous times, that a person's gait (the way he/she walks) is unique for all individuals. Using gait as a biometric identifier has some key advantages:

- It can be detected visually (using cameras), or by the vibrations it generates (microphones for the generated sound, or sensors recording mechanical vibrations), without interrupting the activities of the person
- It is difficult to replicate an other person's patterns
- It is also difficult to mask one's gait patterns. This could provide more reliable tracking in surveillance systems, as hiding or changing one's face is much easier than hiding one's walking patterns.

Identifying someone based on the vibrations their steps generate could prove useful in today's world of smart devices. For example a corridor instrumented with vibration sensors could be used to unlock doors for the right person without interrupting their flow at all.

Although the research presented here deals with footstep sound recorded with a microphone, the principles and processes would be easily adaptable to seismic data: the only difference is the media through which the vibrations reach the sensor, and the sensors themselves.

2 Literature

A number of studies have explored the topic of identifying people based on footstep sounds.

3 Methods

3.1 Data

The recording of footstep sound were recorded from 6 individual persons, with a total of 32 individual shoe - person combination (including socks). For each combination a total of 20 takes were recorded (except for one, where accidentally only 19 takes were recorded). One take consisted of the person walking from the edge of the room towards the microphones in a straight line. The recording was stopped when the person reached the microphones, which took between 3 and 3 steps depending on the person.

3.1.1 Data collection

Data was collected in the Motion Laboratory of the Department of Mechatronics, Optic and Engineering Informatics.

For each person-shoe combination a total of 20 takes were recorded (except for one, where accidentally only 19 takes were recorded). One take consisted of the person walking from the edge of the room towards the microphones in a straight line. The recording was stopped when the person reached the microphones, which took between 3 and 3 steps depending on the person.

The recording equipment was lent to my the Kármán Studio, including: a dynamic (Shure SM58) and a condenser (AudioTechnica AT2020) microphone, a small mixer and a usb sound card, as well as the required cables and microphone stands. The signal from the two microphones were recorded simultaneously on the stereo tracks of the recordings (one microphone recorded only to the right, the other only to the left channel). The data was recorded using Audacity[1] on a 64 bit Windows 10 machine at a rate of 44.1 kHz, and 32 bit resolution. The recording setup can be seen on Figure 1.

3.1.2 Data preprocessing

Each take was exported to an individual `*.wav` file with the naming scheme `ID_shoeType_##.wav`.

3.2 Model

I choose `InceptionV3` as the basis for my model. The top part of the model was replaced with a dense feed forward network to classify the data into 6 classes. The model consisted of a flatten layer, two hidden layers with size 126 and 32, with ReLU activation functions and dropout with 0.5 probability. The output layer consisted of 6 neurons with a softmax function.

3.3 Training

I tried to follow the general steps of transfer learning on a CNN, but training the classification model on the bottleneck features of InceptionV3 with the ImageNet wights proved to be of little use: it was clear that the feature extraction needed for the spectrograms is different than what is used for normal photographs.

Still, ImageNet weights proved to be useful as a pretraining of the network??

[1] https://www.audacityteam.org

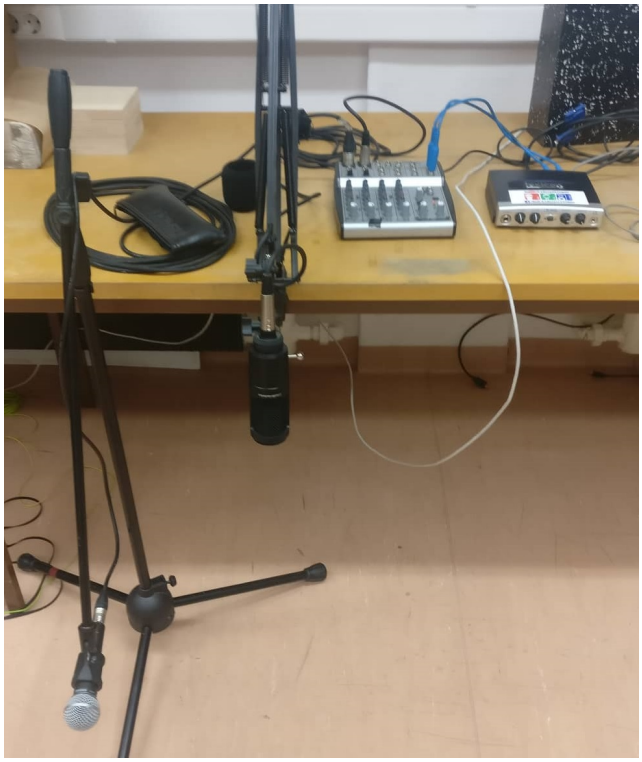

Figure 1: Audio recording setup with two microphones, a mixer and an USB sound card

3.4 Hyper-parameter optimization

3.5 Results

4 Outlook, future plans

4.1 Figures

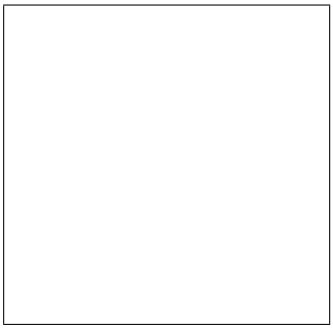

Figure 2: Sample figure caption.

4.2 Tables

All tables must be centered, neat, clean and legible. The table number and title always appear before the table. See Table 1.

This package was used to typeset Table 1.

Table 1: Sample table title

	Part	
Name	Description	Size (μm)
Dendrite	Input terminal	$\sim$100
Axon	Output terminal	$\sim$10
Soma	Cell body	up to 10^6

Acknowledgments

Use unnumbered third level headings for the acknowledgments. All acknowledgments go at the end of the paper. Do not include acknowledgments in the anonymized submission, only in the final paper.

References

References follow the acknowledgments. Use unnumbered first-level heading for the references. Any choice of citation style is acceptable as long as you are consistent. It is permissible to reduce the font size to `small` (9 point) when listing the references. **Remember that you can use more than eight pages as long as the additional pages contain *only* cited references.**

[1] Alexander, J.A. & Mozer, M.C. (1995) Template-based algorithms for connectionist rule extraction. In G. Tesauro, D.S. Touretzky and T.K. Leen (eds.), *Advances in Neural Information Processing Systems 7*, pp. 609–616. Cambridge, MA: MIT Press.

[2] Bower, J.M. & Beeman, D. (1995) *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System.* New York: TELOS/Springer–Verlag.

[3] Hasselmo, M.E., Schnell, E. & Barkai, E. (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *Journal of Neuroscience* **15**(7):5249-5262.