USING WAVELET SCATTERING TRANSFORM TO CREATE VOICEPRINT OF A PASSWORD

B. Assanovich, E. Baniukevich

Educational Establishment “Grodno State University named after Yanka Kupala”, Grodno, Belarus

Today, new biometric technologies are increasingly being used in various protocols and interfaces that implement user identification and verification. Voice identification, which implements text-dependent and text-independent speech recognition, is widely exploited in the human-machine interface. An example is the ID R&D developer [1], owned by the Mitek group of companies, which offers an AI-based speaker recognition product IDVoice that combines three-modal biometric capture with liveness detection, digital ID issuance, and mobile authentication. The developed SDK of ID R&D produces so-called a “voiceprint” that is a template analogous to someone’s fingerprint and capable to perform user verification. Usually Shallow and Deep Neural Networks (DNN) are used in these technologies.

However, it is known [2] that for such tasks of user verification with voice signal, it must be digitized, and then a series of transformations should be performed to identify the main speech characteristics, which can then be applied to train neural networks. In recent years, several approaches to speech processing using Mel Frequency Cepstral Coefficients (MFCC) [3] and Gaussian Mixture Model (GMM) and Time-Delayed Neural Network (TDNN) that learn features from audio samples and converts them to fixed dimension vectors have been widely used.

Besides, applying these methods researches sometimes do not considered the properties of sound waves that have rich physical characteristics. The promising approach has been become the use of SincNet filters that are actually band pass filters which are derived from parameterized sinc functions [3]. The developed by authors model resulted in a significant improvement in EER score of 8.2 % with the use of vanilla SincNet with DNN fusion technique. However, recently the research interest turned again to a known promising technique that was proposed by Mallat [2] named as Wavelet Scattering Transform (WST). The process involves capturing multi-scale and invariant representations of the voice data, making it suitable for biometric applications like authentication. To evaluate the similarity for a user, the extracted features from WST can be compared using different similarity metrics.

In this work, we propose to use WST as a transformation that takes the main frequency properties of voice signal into its biometric characteristics and can possibly be used to convert voice data into a passphrase. This approach can provide to organize both the text-dependent and text-independent two-channel user identification using fusion techniques. We carried out a series of experiments where voice messages corresponding to spoken numbers in English from available dataset [4], were used as a passphrase. The range of correlation values between different voice samples versions of one user was 0.67–0.97. This proves the possibility of using WST to build single-factor or multi-factor biometric verification.

References

1. Internet Resource. Human Verification [Electronic resource]. – Access mode: https://www.idrnd.ai. – Date of access: 07.05.2024.

2. Joakim, A. Deep Scattering Spectrum / A. Joakim, S. Mallat// IEEE Trans. Signal Proc. – 2013. – Vol. 62. – P. 4114–4128.

3. M. Tripathi, D.Singh, S. Susan. Speaker Recognition Using SincNet and X-Vector Fusion. In Artificial Intelligence and Soft Computing. ICAISC 2020. Lecture Notes in Computer Science, vol 12415. Springer, Cham. 2020.

4. Internet Resource. Free Spoken Digit Dataset (FSDD). [Electronic resource]. – Access mode: https://https://github.com/Jakobovski/free-spoken-digit-dataset. – Date of access: 07.05.2024.