IRSTI 28.23.15

[1]M.N. Kalimoldayev, [1]O.Zh. Mamyrbayev,
[2]A.S. Kydyrbekova, [2]N.O. Mekebayev

[1]Institute of Information and Computational Technology, Almaty, Kazakhstan
[2]al-Farabi Kazakh National University, Almaty, Kazakhstan
e-mail: kas.aizat@mail.ru

Voice verification and identification using i-vector representation

Abstract. In the area of voice recognition, many methods have been proposed over time. Automatic speaker recognition technology has reached a good level of performance, but still needs to be improved. Signature verification (SV) is one of the most common methods of identity verification in the banking sector, where for security reasons, it is very important to have an accurate method for automatic signature verification (ASV). ASV is usually solved by comparing a test signature with a registration signature(-s) signed by the person whose identity is declared in two ways: online and offline. In this study, a new i-vector based method is proposed for SV online. In the proposed method, a fixed-length vector, called an i-vector, is extracted from each signature, and then this vector is used to create a template. Several methods, such as the nuisance attribute projection and the within-class covariance normalization, are also being investigated to reduce the intra-class variation in the i-vector space. At the stage of evaluation and decision-making, they also propose to apply the support vector machine with two classes.
In this article, a new low-dimensional space, depending on the dynamics and the channel, is determined using a simple factor analysis, also known as i-vector. I-vectors have proven to be the most efficient functions for text independent speaker verification in recent studies.
Key words: i-vectors, dimensionality reduction, UBM size, speaker identification

Introduction

As time goes, voice processing technology is becoming more and more mature. Using advances in signal processing and machine learning, ASV is implemented in two ways: online and offline. With offline verification, also called static verification, we have access only to the signature image [1-3]. In such methods, we usually normalize the image size after some preprocessing, and then extract the elements from the image using a sliding window. These functions are then used to compare two signatures. On the other hand, there are online methods, also called dynamic methods, where information related to signature dynamics is provided, as well as signature image [4-6]. Dynamic information includes pressure, velocity, azimuth, etc. In these methods, changes in vertical and horizontal directions are commonly used as shape-related elements. These methods have better performance than autonomous methods and are more reliable because they use more information extracted from the signature. In addition to these advantages, signature forgery is more difficult in these methods because they use dynamic characteristics, such as speed and azimuth, which are very difficult to simulate. Our focus is on online methods. There have been a lot of SV online studies that can be grouped into two main categories:

Methods based on global signature features. These methods attempt to extract a fixed-length vector from the entire signature so that the signatures can be easily compared in vector form. These methods can be further divided into two subcategories: in the first, we try to extract global functions from the entire set of signatures. For example, in [7] Jane uses the number of strokes as a global feature. The authors use other functions such as average speed, average pressure, and the number of times the pen is lifted during signature in [8]. As a good example of Fierrez-Aguilar, [4] introduced 100 global attributes sorted by their individual discriminatory power. A subset of these functions is also used in other studies [5, 9-12]. In the second subcategory, a transformation is applied to the signature to obtain a fixed-length vector. For

example, the wavelet transform is used in [13] to extract the feature vector from the entire signature. In another study, the discrete cosine transformation (DCT) is used to obtain a fixed length feature vector [6]. The proposed method in this paper belongs to this group.

Functional methods. Methods in this category are more focused on comparing signatures and calculating the distance between two signatures. In these methods, each signature is represented using a sequence of local features extracted from it. This category can also be divided into two subcategories: the methods in the first do not perform any modeling. In fact, in these methods, a set of references is stored for each individual, and during the test, the input signature is compared to the set of references for decision-making. The most common method in this subcategory is Dynamic Time-Warping (DTW), which is used in many studies [14–17].

The second category includes methods that train a probabilistic model for each person, using signatures in his/her control set. These methods typically use probabilities for evaluation and decision making. The most common methods in this subcategory are the hidden Markov model (HMM) [18–22] and the Gaussian mixture model (GMM) [23–25].

In the area of voice recognition, the use of Gaussian mixture models (GMM) to create universal background models (UMM) and collaborative factor analysis (JFA), by far the most popular i-vector, has increased its accuracy in creating a specific dynamics model. However, sometimes we do not need to know what language the speaker speaks, because in some situations only one of them is the most important, while others are relatively less critical. Speaker verification is becoming increasingly important as a solution to secure biometric keys for industrial, forensic and government purposes, such as data encryption on mobile devices or user verification at contact centers. It seems that users are annoyed with the persistence of multiple PINs and passwords, that is, biometric data that cannot be lost or forgotten, provide significant advantages in terms of usability.

i-vector

The i-vector was first proposed for speaker recognition application, and then was applied in other applications, such as language identification, accent identification, gender recognition, age assessment, emotion recognition, sound scene classification, etc. In the main application of this method (i.e. speaker recognition), a vector of fixed length, called an i-vector, is extracted from a speech signal of arbitrary duration. In this article, we give a description of the i-vector problem and a brief overview of the initial results. We begin with a very brief description of the key components of the i-vector based on the speaker recognition system. In the following steps, this vector is used for scoring and recognition. Although i-vector is used mainly in many speech applications, it is less well known in other areas. In this article, we use the i-vector, which is usually used to recognize the speaker in the SV. Despite their different areas, speech biometrics and signature biometrics are similar in nature, as both must extract subject-specific patterns from a captured signal contaminated by changes from various irrelevant sources. Since the analysis of total variability factors is an embedded i-vector learning step that helps eliminate distractions in biometric analysis and extracts a unique identity representation vector, we expect the i-vector to be able to provide a promising solution for the signature extraction problem.

There are two reasons for using this method for SV. First, online signatures have a variable length, similar to speech signals. Using this method, we can get a fixed-length vector that facilitates the following steps in making a decision. Therefore, after extracting the temporal features from each signature, we extract the i-vector. Since we get a fixed-length vector for each signature, we can put this method in the first category above. The second reason is that a person's signatures usually differ slightly each time. These differences lead to changes within the class, which in turn increase the false rejection rate (FRR). In various applications of the i-vector in speech processing, various ways have been proposed to reduce intraclass variations, which can also be accepted in this application. Similar to the case of speaker verification, we also suggest using two different methods to reduce the undesirable effects of intra-class changes. Since there are several signature samples for each person as a reference set at the registration stage, we suggested adding them to the data used to train variation compensation methods within the class. In addition, we proposed to apply the 2nd class support vector machine (SVM) method to distinguish between i-vectors extracted from genuine and fake signatures. The experimental results showed the effectiveness of these ideas on two different databases.

Extract statistics

At this stage, for each sequence of attributes, the Baum-Welch statistic of zero and first order is calculated using UBM [26,27].

Given that Xi is a complete set of feature vectors for learning the i-th signature, the zero and first order statistics for the c-th UBM component is calculated as follows:

$$N_c\,(X_i) = \sum_t r_{i,t}^c \qquad (1)$$

The variability of the speaker or session is the variability manifested by a given speaker from one recording session to another. This type of variability is usually associated with channel effects, although this is not strictly accurate, since there are also changes within the dynamics and phonetic change. In this approach, the speech segment is represented by a low-dimensional "identity vector" (ivector – for short) extracted by factor analysis. The i-vector approach has become state-of-the-art in speaker verification, and in this paper we show that it can be successfully applied to speaker identification as well. The approach provides an elegant way to reduce multidimensional sequential input data to a low-dimensional vector of features of a fixed length, while retaining most of the relevant information [28]. The basic idea is that the session-and channel-dependent supervectors of the Gaussian mixture model cascade model (GMM) can be modeled as

$$s = m + Tw \qquad (2)$$

where m is the session-and channel-independent component of the average supervector obtained from UBM, T is the basis matrix covering the subspace encompassing the important (both for the dynamics and the session) in the supervector space, and w is the standard, normally distributed hidden variable. For each observation sequence representing the statement, our i-vector is the point estimate of the maximum a posteriori (MAP) for the hidden variable w. Our i-vector extractor learning procedure is based on the efficient implementation proposed in [29].

The contribution of this study is to evaluate the result of factors affecting the i-vector, based on the speaker's sound identification. We study this in terms of parameters, where we evaluate and analyze how the various parameters of the i-vector extractor, such as the size of the Universal Background Model (UBM) and the dimension of the i-vector, affect the accuracy of speaker detection. The UBM size refers to the Gaussian component, which is the corresponding adapted component in the dynamics model. The I-vector dimension is equal to the "rank" of its own matrix. Based on Huang, a greater i-vector dimension would not give a large performance improvement of the classification, but significantly increased the computational costs. In [30] the literature discussed by reducing computations will allow efficient use of the i-vector in more applications. In this study, the recorded computation time is to investigate whether both factors affect the computation or not, and and the next direction for the next study is determined.

To record the voice in the present work, a complex of technical devices was used, the block diagram of which is shown in Fig.1. The block diagram of the system includes: microphone 1, low-frequency amplifier 2, analog-to-digital converter 3, software 4, and a computer unit 5, which records the amplitude-time and calculation of the frequency dependences of the signal.

To identify an unknown voice recording, a database of speakers was compiled and registered. Using the above equipment, voice signals were recorded, which formed the database.

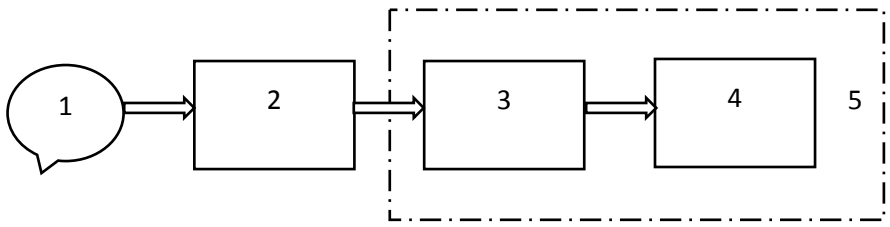

Figure1 – A block diagram of the hardware of the voice identification system

When scaling, or weighing, all experimental data are reduced to the same scale. This procedure is necessary to reduce the impact on the analysis of strongly pronounced variables. There are various ways of scaling [28], in this work standardization has been applied, since it is the most studied and tested. Standardization uses standard deviation – S_{dev}, which is one of the most commonly used weighting factors. In addition, each element of the matrix X is multiplied by the value $1/S_{dev}$:

$$X_{ik}^{scaled} = X_{ik} \frac{1}{S_{dev}}, \quad (3)$$

where X_{ik} is the i-th variable of the k -th sample, X_{ik}^{scaled} – is the i-th scaled variable of the k -th sample, S_{dev} – is the standard deviation of the sample.

Figure 2 shows the block diagram of the method used in this paper. The i-vector system consists of two main parts: the front part and the rear part. The first consists of the extraction of cepstral features and UBM learning, while the latter includes sufficient statistical calculation, T-matrix training, i-vector extraction, dimension reduction and evaluation.

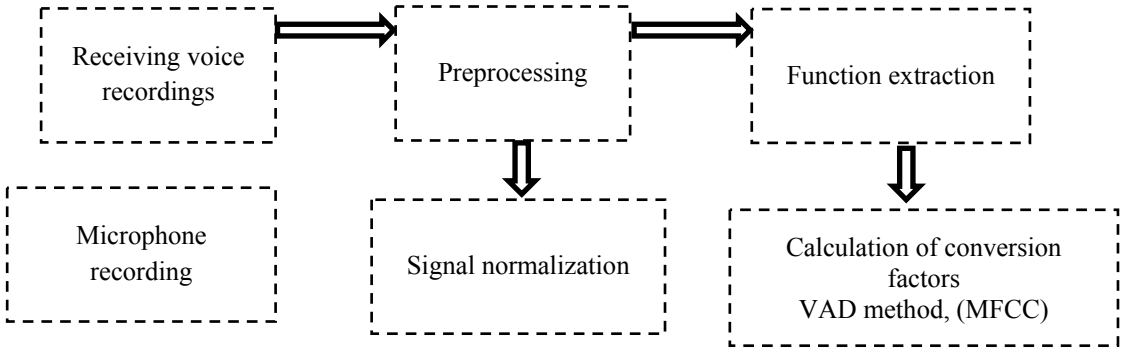

Figure 2 – Block diagram of the experiment with the speaker's identification system

Function extraction

First, simple energy speech activity detection (VAD) is performed to discard unnecessary part of the voices. Energy-based VAD is used when energy values are first calculated at the frame level, followed by data normalization and, finally, voice activity detection. A class with a higher average is considered to be the speaker's sound, and therefore the corresponding segments of the sound are preserved until smooth. Second, the characteristics of MFCC and log energy, together with their first and second order derivatives, are computed in 20 ms of Hamming window frames every 10 ms. Then the determination of the activity of the speaker's sound is applied, and the speaker's sound is normalized in accordance with the standard normal distribution.

To solve the problem of identification of the individual, the analysis of the individual frequency spectrum of voice signals is the main one. In this setting, the first two factors (the amplitude and duration of the signal) are random and need to get rid of them. To do this, all signals were reduced to one amplitude, that is, the amplitude normalization was performed:

$$a_i^{norm} = \frac{a_i}{a_{max}} \quad (4)$$

where a_i is the measured amplitude, a_i^{norm} is the maximum amplitude, normalized amplitude, i = 0,1, ..., k.

In order to remove the second factor (speech rate), time normalization was performed. The second factor was taken into account programmatically by using the same number of samples. The amplitude-frequency characteristics in the form of the recorded audio signal spectrum were analyzed directly. The frequency spectra had the form of the dependence of the amplitude A from the frequency f.

When recording voice signals in real conditions, it is possible to impose random factors, including both external mechanical noise and hardware noise. Median filtering [31] was used for their suppression, which consisted in exclusion from the initial emission signal.

The current lack of a clear systematization of voice features, as well as the existence of a large number of voice characteristics of various levels, such as the basic tone [32], formant frequencies

[33,34] and others [35,36], is a certain difficulty in choosing most informative features and characteristics for a specific identification method and requires a separate study. This section provides qualitative and quantitative estimates for the selection of informative voice characteristics. The difference in voice timbres is described by different frequency spectra of voice signals. Fourier decomposition is a natural mathematical apparatus for frequency spectrum analysis. Processing of data representing numerical amplitude-time dependences can be carried out using discrete Fourier transform:

$$A(k) = \sum_{n=0}^{N-1} a(n) e^{-j2\pi \frac{k}{N} n},$$

$$k = 0,1,2,\dots,N-1 \quad (5)$$

$$a(n) = \frac{1}{N} \sum_{k=0}^{N-1} A(k) e^{-j2\pi \frac{k}{N} n},$$

$$k = 0,1,2,\dots,N-1 \quad (6)$$

where A(k), a(n) are direct and inverse discrete Fourier transforms, respectively, k and n are sample numbers, and N is the number of samples. Coefficients A (k) can be used precisely as the elements of the matrix X, forming rows in this matrix.

I-vector extraction

I-Vector based systems. As explained earlier, at present, the i-vector in the space of complete variability has become a modern approach to voice recognition.

This method, which was introduced after its predecessor, the joint factor analysis, can be considered as a method of extracting a compact representation with a fixed length in the presence of an arbitrary length signal. The extracted compact unit vector can then be used either to measure similarity based on vector distance or as input for any further feature transformation or modeling. There are certain steps to extract the i-vector from the signal. First the features should be extracted from the input signal, then the Baum – Welch statistics should be extracted from the features, and finally the i-vector is calculated using these statistics. We will explain these steps in detail below.

For each statement, the corresponding feature sequence is eventually transformed into an i-vector using a GMM-based i-vector extractor with three different UBM-sized components trained from the combined features from all the samples included in our training data (Fig.3). Three UBM sizes that make up 32, 64 and 128 components.

Assuming that we have calculated the zero and first order statistics using (7) and (8), we can calculate the a posteriori covariance matrix [i.e. (w_i, w_j)], average (i.e. $E[w_i]$ and second moment (i.e. $[w_i w_i^t]$ for wi, using the following relations:

$$Cov(w_i, w_j) = (I + \sum_c N_c(X_i) T_c^t \sum_c^{-1} T_c)^{-1} \quad (7)$$

$$E[w_i] = \left(Cov(w_i, w_j) \sum_c T_c^t \sum_c^{-1} F_c(X_i)\right)^{-1} \quad (8)$$

$$E(w_i, w_i^t) = Cov(w_i, w_j) + E[w_i] E[wi]^t \quad (9)$$

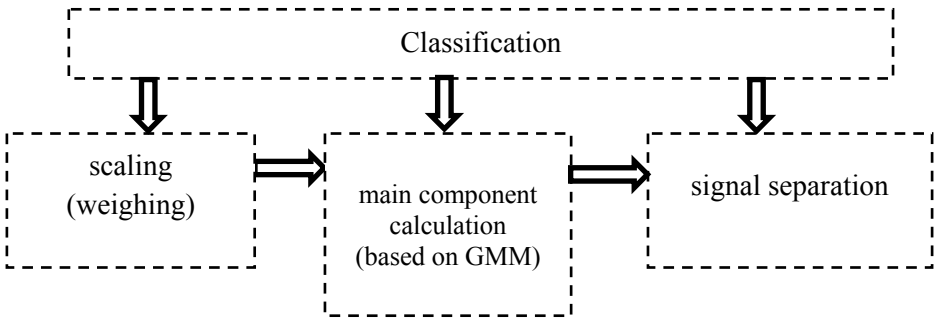

Figure 3 – Using projection methods

Post processing

Since the simulation of i-vectors contains information about the dynamics and channel variability in the same space at the same time, the channel compensation technique in the common factor space is required to eliminate undesirable effects. Channel-compensated approaches play a major role in I-vector speaker recognition systems. Therefore, channel compensation is necessary to ensure that test data obtained from different channels can be properly evaluated by loudspeaker models. For channel compensation to be possible, channel variability should be modeled explicitly.

Before calculating the verification estimates, channel bleaching, linear discriminant analysis (LDA) and within-class covariance normalization (WCCN) were performed to compensate for the channel.

We used the same dataset to train the total variability matrix to evaluate the LDA and WCCN matrices. Since the extracted i-vectors contain variations both within and between accents, the goal of dimension reduction is to project the i-vectors into a space where the variability between accents is maximal and the variability within the accent is minimized. In this study, three different measurements were experimented with: 100, 200, and 400. Thus, we optimized the parameters of the i-vector to experiment and evaluate the result.

Scoring

Finally, the identification result from the system is given by calculating the similarity score. The simplest and fastest counting function, that is, the cosine distance, is calculated between the i-vectors from the dynamics model and the i-vector from the test segment. The decision-making and evaluation process is then computed, and the system performance is then represented using accuracy 91,11%, CMC curves, and detection error tradeoff (DET).

t-SNE for individual signatures using raw i-vectors (i.e., without applying any transformations) (Figure 4)

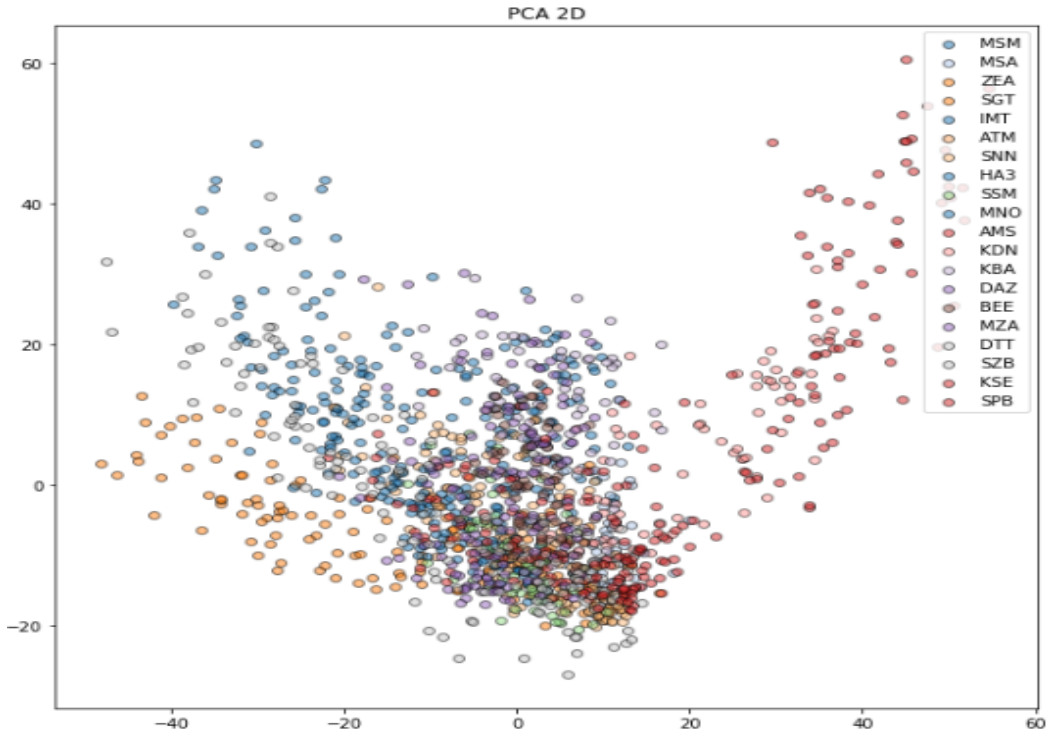

Figure 4 – Two-dimensional data representation

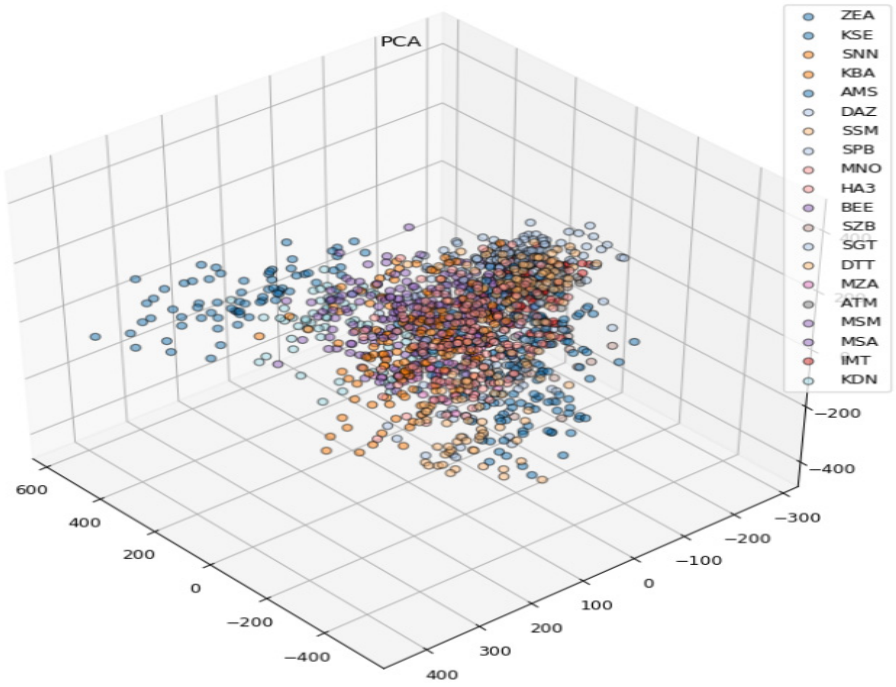

Figure 5 – Three-dimensional data representation

As shown in Fig. 5, for both projection methods, all points are combined into compact areas not intersecting with each other, each area corresponds to the records of one speaker. This indicates that both methods have provided a reliable separation of speakers by voice data.

Note that the direct comparison of the above graphs of accounts for the methods of main components (Fig. 4) and projections on latent structures (Fig. 5) does not allow to quantify these calculation options. The advantage of one method over another can be estimated from the residual dispersion graph.

Result and discussion

A series of experiments was conducted to study the effect of the number of UBM components, vector dimension, and post-processing techniques. These experiments were conducted using a set of voices – a set of open source toolkit and extensible tools for recognition of the modern level. 50 votes taken from the database were used for the evaluation.

The adaptation of projection methods of the main components and projections on latent structures in relation to the analysis of acoustic signals in technologies of personal identification by voice has been carried out.

The speech database of voice data intended for tasks of voice identification and differing in considerable number of repetitions of phrases by the same speakers is created. The use of this database by increasing the number of repetitions provides a more accurate assessment of the identification result.

Comparison of various informative parameters of voice signals used as a feature vector in projection analysis methods has been carried out. The residual dispersions were calculated that showed the preference for the use of voice identification of the Mel-cepstral decomposition coefficients, which improve the separation of the source signals according to their features and reduce the contribution of random distortions.

We found that the accuracy increases as the dimension of the i-vector increases. In addition, our results showed that the UBM with smaller size outperforms larger UBM. In addition, the result of the time calculation shows that the processing takes longer when the dimension of the i-vector increases and the size of UBM is larger.

Conclusion

In this article, we studied how the i-vector extractor parameter, such as the UBM size and i-vector dimension, affects the accuracy of voice

identification. As for the parameters, the highest accuracy was achieved when using UBM with Gaussians and an i-vector dimension. They are similar to those reported in the general voice recognition literature. As for the data, we found that the selection of UBM training data is the most important part, followed by the dimension of the i-vector. This is understandable because the earlier components of the system affect the quality of the remaining steps.

For further research, we propose to study the effect for a larger i-vector dimension and a larger UBM size. For this study, we do not do this because of the long computation time, because we use the small size of the speaker's database. In the following studies, we can reduce the computation time by exploring other factors that influence this, and add additional data to further study this effect of the experiment.

This work carried out in the framework of the project "Development of technologies for multilingual automatic speech recognition using deep neural networks".

References

1. Kalera, M.K., Srihari, S., Xu, A . "Verifying and Identifying Signatures Offline Using Distance Statistics" Int. J. Pattern Recognit. Artif. Intel. 18, (07) (2004): 1339–1360.
2. Singh, J., Sharma, M . "Verifying signatures offline using neural networks." J. Inf. I-manager. Technol.1, (4) (2012): 35.
3. Daramola, S.A., Ibiyemi, TS.S . "Autonomous signature recognition using the hidden Markov model (HMM)." Int. J. Comput. Appl. 10, (2) (2010): 17–22.
4. Firres-Aguilar, J., Nanni, L., Lopez, Penalba, J., et al . "An online signature verification system based on a combination of local and global information." Audio and video biometric authentication of identity (2005): 523–532.
5. Nanni, L . "Advanced Improved multi-match method for online signature verification with global functions and tokenized random numbers." Neurocomputing, 69, (16) (2006): 2402-2406.
6. Liu Y., Yang, Z., Yang, L . "Online Signature Verification Based on DCT and Sparse Representation," IEEE Trans. Cybern. 45, (11), (2015): 2498-2511.
7. Jane AK, Griss FD, Connell S.D . "Online Signature Verification." Pattern Recognit. 35, (12) (2002): 2963-2972.
8. Lee.L.L., Berger.T., Aviczer.E. "Reliable Online Human Signature Verification Systems." IEEE Trans. Pattern Anal. Max Intell. 18, (6) (1996): 643–647.
9. Nanni, L., Lumini, A. "Parsen window classifier ensemble for online signature verification." Neurocomputing 68 (2005): 217–224.
10. Lei, H., Govindaraju, W . "Comparative study of the consistency of functions in verifying signatures online." Pattern Recognit. Lett. 26, (15) (2005): 2483-2489.
11. Richiardi, J., Ketabdar, H., Drygajlo, A . "Selecting local and global functions for online signature verification." 2005 Proc. Eighth int. Conf. Analysis and recognition of documents (2005): 625–629.
12. Nanni, L. "Experimental comparison of single-class classifiers for online signature verification," Neurocomputing 69, (7) (2006): 869–873.
13. Leytman DZ, George S.E. "Verification of a handwritten signature online using wavelets and neural back distribution networks". 2001 Proc. Sixth Int. Conf. Analysis and recognition of documents (2001): 992–996.
14. Kholmatov A., Yanikoglu B. "Identity Authentication Using the Improved Signature Verification Method on the Internet", Pattern Recognit. Lett. 26, (15) (2005): 2400–2408.
15. Vivaracho-Pascual, C., Faundez-Zanuy, M., Pascual, JM "An effective low-cost approach for recognizing signatures online based on the normalization of length and fractional distances." Pattern Recognit. 42, (1) (2009): 183–193.
16. Sato, Y., Kogure, K. "Online signature verification based on the form, movement, and pressure of the letter." Proc. Sixth Int. Conf. Pattern Recognition.(1982): 823–826.
17. R. Martens, Klezen L. "Optimization of dynamic programming for online signature verification". 1997 Proc. Fourth Int. Conf. Document Analysis and Recognition 2 (1997): 53–656.
18. Firres J., Ortega-Garcia J., Ramos D. "Online Signature Verification Based on HMM: Character Extraction and Signature Modeling." Pattern Recognit. Lett., 28, (16) (2007): 2325–2334.
19. Dolfing, J., Aarts, E., Van Oosterhout, J. "Online Signature Verification Using Hidden Markov Models." 1998 Proc. Fourteenth Int. Conf. Pattern Recognition 2 (1998): 1309–1312.
20. Van BL, Garcia-Salichetti S., Dorizzi B. B. "On the use of the Viterbi path along with

information about the probability of MMOs for online signature verification", IEEE Trans. Syst. Cybern Man B, Cybern. 37, (5) (2007): 1237–1247.

21. Rua, EA, Castro, JLA. "Online Signature Verification Based on Generative Models", IEEE Trans. Syst. Cybern Man B, Cybern. 42, (4) (2012): 1231–1242.

22. Jan L., Vijazha B., Prasad R . "Using Hidden Markov Models for Verifying Signatures", Pattern Recognit. 28, (2) (1995): 161-170.

23. Richiardi, J., Drygajlo, A. "Gaussian Mixture Models for Online Signature Verification." Proc. 2003 ACM SIGMM Workshop on Biometric Methods and Applications (2003): 115–122.

24. Miguel-Hurtado, O., Mengibar-Pozo, L., Lorenz, MG et al . "Signature verification online using models with dynamic time distortion and Gaussian mixture". 2007 41st Annual IEEE Int. Carnahan Conf. Security Technology (2007): 23–29.

25. Kenny P., Owel P., Dehak N., et al "Investigation of the variability of interspecific systems in speaker verification." IEEE Trans. Audio Speech Lang. Process. 16, (5), (2008): 980-988.

26. Kenny P., Bulian G., Dumushel P. "Modeling Your Own Voices with Rare Learning Data", IEEE Trans. Speech Audio Process., 13, (3), (2005): 345–354.

27. Hamm A., Hennebert J., Ingold R. "Models of a Gaussian mixture for verifying the signature of an abyss". Machine learning for multimodal interaction (2006): 102–113

28. Zhen Huang, Yu-Chei Cheng, Kehuang Li, Ville Hautamaki, Chin-Hui Li. "Blind segmentation method for detecting acoustic events based on I-vector". Proc. Annu. Conf. Int. Speech Comm. Assoc. INTERSPESH August: (2013): 2282–2286.

29. Ondrey Glembeck, Lukas Burghet, Pavel Mateyka, Martin Carafiat and Patrick Kenny "Simplification and optimization of i-vector extraction". ICASSP, IEEE Int. Conf. Acoust. The process of the speech signal (2011): 4516–4519.

30. Ayvazyan, S.A. "Applied statistics: Classification and reduction of dimension." Finance and Statistics (1989): 607

31. Grigoryan, R.L. "Comparison of various ways to assess the similarity of the distributions of the pitch frequency in the problem of speaker identification by his speech." Speech technology no. 3 (2010): 35–38.

32. Golubinsky, A.N. "Method of estimating formant frequencies based on the polyharmonic mathematical model of the speech signal." Speech technology no. 3 (2010): 29–34.

33. Konev, A.A. "On one algorithm for estimating formant frequencies in the interval of closed vocal folds." Speech technology no. 3 (2010): 50–53.

34. Golubinsky, A.N. "The model of a speech signal in the form of a pulse AM-oscillations with several carriers to verify the personality by voice." Control Systems and Information Technologies no.4 (2007): 86–91.

35. Roldugin, S.V. "Models of speech signals for voice identification." Radiotekhnika no. 11 (2002): 79–81.

36. Azarov, I.S. "Calculation of instantaneous harmonic parameters of a speech signal." Speech technology no. 1 (2008): 67–77.