Multi-space Random Mapping for Speaker Identification

Chong Tze Yuang, Andrew Teoh, David Ngo, & Michael Goh
Centre of Biometrics & Bioinformatics
Faculty of Information Science & Technology
Multimedia University
Jalan Ayer Keroh Lama 75450 Melaka
MALAYSIA

Abstract: - This paper presents a work of utilizing multi-space random mapping (MRM) to formulate a dual-factor identification system, which combines speaker biometric and personal token. Personal token will be assigned to the client to constitute a unique random subspace during enrollment and the speaker template will be generated within the random subspace. Test features will be mapped to the random subspace that is described by the personal token presented during identification. Our work has shown that MRM-system exhibits stronger discriminative ability when comparing test features to its counterfeit templates, which lied in other different random subspaces. This advantage thus contributes to better F-ratio and greater accuracy recognition. Experiments on YOHO corpus demonstrate a remarkable result where the system achieves the perfect identification rate.

Keywords:- Speaker identification, Multi-space random mapping, dual-factor authentication.

1 Introduction

Traditional token-based authentication system, e.g., smart card etc. has the major drawback of being easily fooled by the stolen token. Intruder who is holding a stolen token will trespass the security easily. Speaker recognition emerged as the more reliable authentication system based on the assumption that human speaking behavior is unique from others thus can be utilized as the biometric feature for authentication. However, the variation of the speaking manner of human is the natural disadvantage of speaker-biometric making it the least accurate biometrics compare to other static-signal biometrics such as fingerprint etc.

By combining the personal token and the speaker biometric, attacker can no more breaks through the security simply by presenting an stolen token. Since the personal token is a very unique factor, it can be hashed into the speaker biometrics, in some manner, to make the speaker feature more distinctive. Similar works can be seen in [6, 7] where the personal token is introduced to a trademark biometrics system to cure some major defect in the current security system. This paper records the work of combining the token-based authentication with the speaker recognition to gain benefit from both sides and to alleviate the drawback of them. Multi-space random mapping (MRM) is used to hash the token information to the speaker feature.

2 Multi-space Random Mapping

Multi-space random mapping (MRM) composes two stages: (a) feature extraction and (b) random projection. The feature extracted from speech raw signal is mapped to a client specified random subspace. The mapping is determined by the tokenized pseudo random numbers (PRN), which are unique from a speaker to another speaker. The random projection can be expressed as follow [1]:

$$v = \kappa R\omega . \quad (1)$$

Vector ω is the original p-order feature and R is the row-wise orthonormal standard normal distributed random matrix formed from the PRN, e.g.,

$$R_{ij} \sim N(0,1) ,\ 1 \le i \le p \text{ and } 1 \le j \le p . \quad (2)$$

The value of κ is unity since the feature dimension is remained in this projection.

During enrollment, PRN are generated and are assigned to the new-registered speaker. The PRN are arranged into a $p \times p$ matrix to be orthonormalized to construct the speaker-specified random subspace axis. Features extracted from the training speech will be mapped to the random subspace and speaker template will be produced from those random-mapped features. Thus, template of different speaker is spanned in different random subspace. Identification is performed by mapping the test feature to each of the registered subspace to match to the respective template. The arriver will be identified as the speaker whose template yields the closest distance to the test features.

Since the transformation is orthonormal, the distance between any two points that are projected to the same subspace is identical to the original distance before mapping.

$$\left\| T_1 a - T_1 b \right\|^2 = \left\| a - b \right\|^2 \quad (3)$$

Therefore the similarity between the test features and the genuine template is not altered after MRM is applied to the speaker authentication system thus preserves the intra-speaker distance.

However, by mapping two points into two different subspaces, it is almost certain that the distance between the two points will be stretched, e.g.,

$$P\left(\left\| T_1 a - T_2 b \right\|^2 > \left\| a - b \right\|^2 \right) \simeq 1 . \quad (4)$$

This has been proven empirically and the detail of the experiment will be discussed in the next section. The test features become more discriminative to the counterfeit templates for MRM-identification, thus raises the inter-speaker distance overall. Therefore, the MRM-speaker authentication shows better F-ratio, e.g., ratio of intra-speaker to inter-speaker distance, as compare to the non-MRM version.

3 Speaker Identification System

The following block diagram (Fig. 1) depicts the speaker identification system in this work. MRM scheme is built on the vector quantization (VQ) speaker recognition framework [3]. Collected speech waveform is blocked into 240-sample frame with 160 overlapped with adjacent frames. Each frame is categorized into speech and non-speech frame based on the energy profile. Speech frames are selected to extract the linear predictive (LP) cepstrum [2]

During enrollment, a new PRN sequence is generated to form the random transformation to map the features extracted from the training speech signal. The feature vectors in the new subspace are clustered into 16 clusters using modified K-means (MKM) algorithm [3]. Each cluster contributes one centroid to be stored as the speaker's template.

In the identification session, test speech waveform and the PRN sequence are inputted to the identifier. Feature extracted from the test speech is mapped to the random subspace that is described by the PRN. Nearest-neighborhood matching is performed to compare the projected test feature to each registered templates [4] in each random subspace.

4 Database

The experiment was conducted on YOHO speaker verification corpus [5]. The speech tokens from the first enrollment session are used to generate the template from all 138 speakers. All speech tokens from testing session are used to evaluate the system. Thus each speaker template is generated from 24 tokens and there are total 5,520 (40 tokens × 138 speakers) trails of identification testing.

5 Experiments and Discussion

The fact that there is such a great possibility that the distance between two points will be stretched resulted from MRM, as stated in (3), has been proven through experiments. The experiment has been carried out by running 50,000 trials to collect the amount of distance that is extended between two points resulted from MRM. Two points are picked randomly in every trial and to be mapped into two different random subspaces. The random mapping is described by the orthonormal normal distributed random matrix, which is refreshed in every trial. The distance between the two points is read before and after mapping and their difference are recorded to form the distribution histogram.

Fig. 2 shows the histograms of the distributions of the distance between the two random-picked points, a_i and b_i, before and after MRM, mapped by R and T, and the amount of distance stretched. The dimension of the point is 20 orders.

Fig. 2(a) and 2(b) show the distribution of the original distance, $d_i = \|a_i - b_i\|^2$, and the projected distance $m_i = \|R_i a_i - T_i b_i\|^2$, respectively. Histogram in Fig. 2(c) shows the distribution of the amount of extension of the distance resulted from MRM, e.g.,

$$\begin{aligned} e_i &= m_i - d_i \\ &= \|R_i a_i - T_i b_i\|^2 - \|a_i - b_i\|^2 \end{aligned}.$$

As observed, almost all trial stretches the distance, e.g. $e_i > 0$, thus the following conclusion is drawn.

$$\begin{aligned} &P(e > 0) \simeq 1 \\ &P\left(\|Ra - Tb\|^2 > \|a - b\|^2\right) \simeq 1 \end{aligned} \quad (5)$$

The experiment is repeated in the dimension of 10, 15, and 50 orders and the respective histograms are shown in Fig. 3, Fig. 4, and Fig. 5, respectively.

When the random mapping is executed at higher dimension order, more trials were experienced distance extension. At the order of 50, distance is extended in all 50,000 trials. Thus, we deduce a conclusion that it is more certain that MRM will extend the distance at higher order of dimension.

In order to monitor the effect of MRM to the speaker identification system, experiments are carried out to compare the performance of the MRM-system with the non-MRM system. Without engaging MRM, the identification system follows the traditional VQ-speaker recognition framework [3]. The accuracy and the F-ratio of the system at different feature order are shown in Fig. 6.

The MRM-system scores the perfect identification rate at feature order from 8 to 20. For MRM-identification system, higher feature order yields better F-ratio. This is contrast with the manner of the F-ratio shown by the original system, where higher order produces weaker F-ratio. The phenomenon is due to the behavior of MRM where higher order MRM stretches greater distance between two points. At higher feature order, the inter-speaker matching yields better discrimination thus contributing to lower F-ratio.

The improvement of F-ratio contributed by MRM can also be observed from the histogram of the imposter and genuine distance. These histograms for MRM system and non-MRM system are shown in Fig. 7 and Fig. 8 for feature order 10 and 15.

Same conclusions are drawn from the observation on the histogram of the genuine and imposter distance. MRM isolates the genuine distribution from the imposter distribution indicating a perfect recognition rate will be achieved. The separation between the genuine and imposter distribution can be further distanced by increasing the feature order as shown by Fig. 8.

6 Conclusion

In this work, MRM system has been shown to possess remarkable discriminative ability in counterfeit comparison thus contribute to better accuracy recognition. MRM has been incorporated with the classical VQ speaker identification system to form a dual-factors identification system that combines the speaker biometric and the personal token. Experiments on YOHO corpus has scored perfect identification rate.

Reference:

[1] Johnson, W. B. & Lindenstrauss, J., Extension of Lipschitz mapping into a Hilbert Space, *Conference in modern analysis and probability. Amer. Math. Soc., Providence, R.I.*, 1984, pp. 189-206.

[2] Makhoul, J., Linear Prediction: A tutorial review, *Proc. IEEE*, vol. 63, 1975, pp. 561-580.

[3] Soong, F. K., Rosenberg, A. E., Rabiner, L. R., and Juang, B.-H., A vector quantization approach to speaker recognition, *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, 1985, pp. 387-390.

[4] Higgins, A. Bhaler, L., and Porter, J., Voice identification using nearest neighbor distance measure, *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 1993, pp. 375-378.

[5] Higgins, A., YOHO speaker verification, *presented at the Speech Research Symp.*,1990.

[6] Teoh, A. B. J., Ngo, D. C. L. & Goh, A, Quantized multispace random for two-factor identity authentication, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, under review.

[7] Savvides, M., Kumar, B.V.K.V., & Khosla, P.K., Cancelable biometric filters for face recognition, *Proc. 2004 ICPR*, 2004, 3: 922-925.

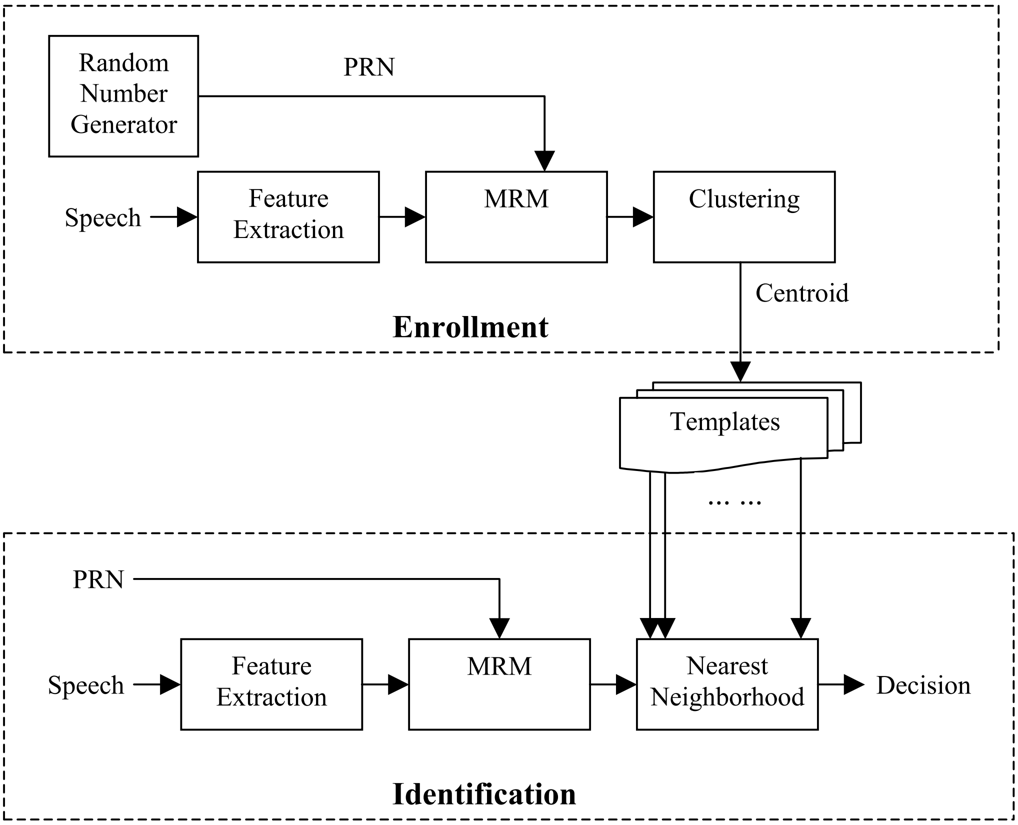
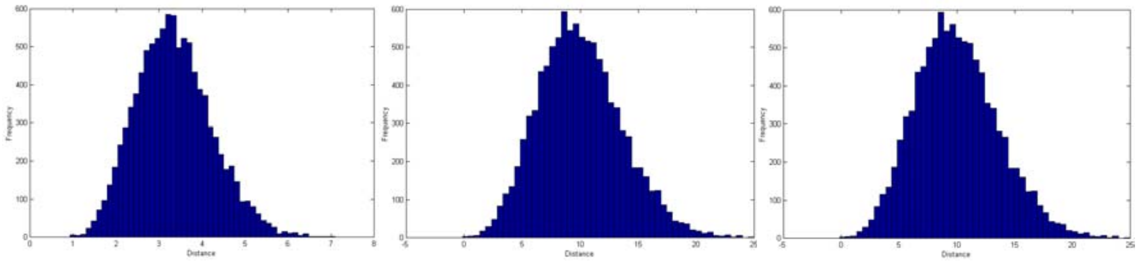
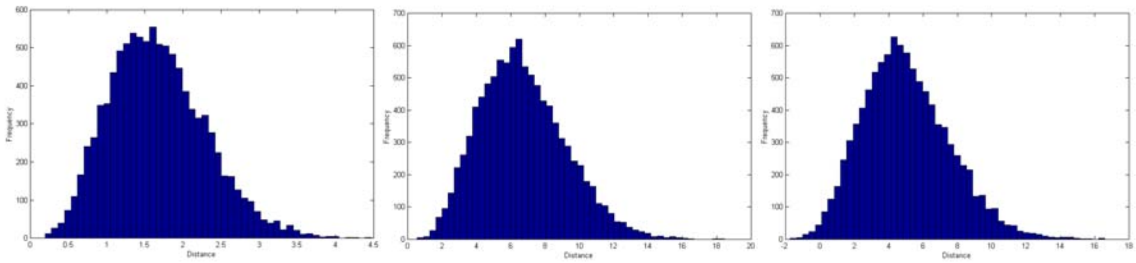


Fig. 1 Block Diagram of Speaker Identification System



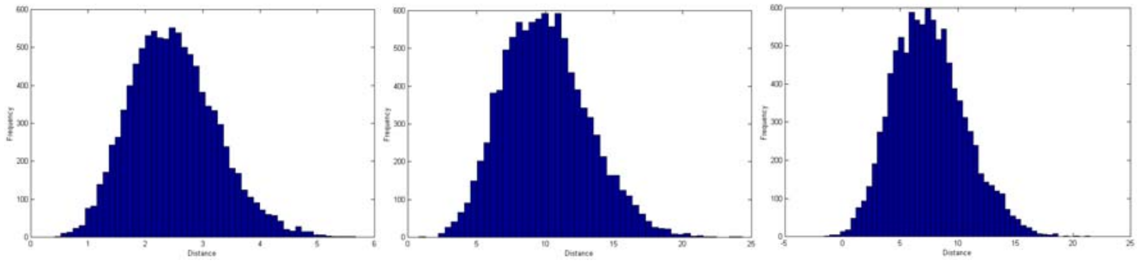
(a) Original distance (b) Projected distance (c) Extension of distance

Fig. 2 MRM at Order of 20



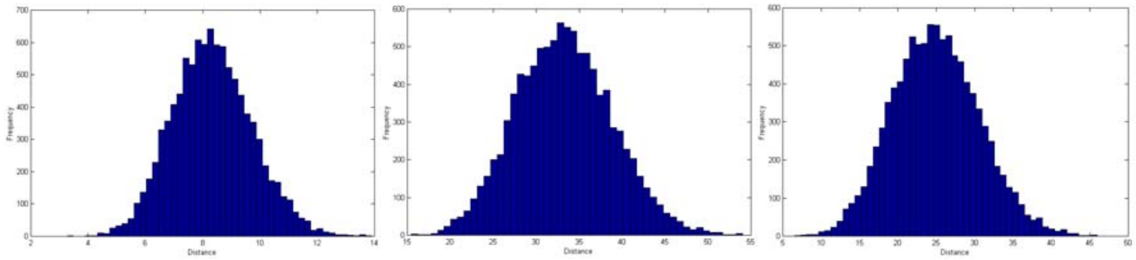
(a) Original distance (b) Projected distance (c) Extension of distance

Fig. 3 MRM at Order of 10



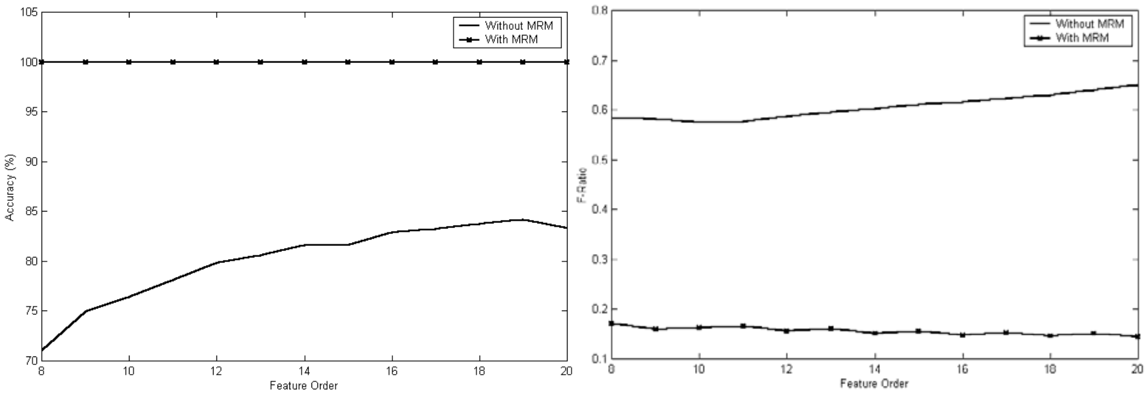
(a) Original distance (b) Projected distance (c) Extension of distance

Fig. 4 MRM at Order of 15



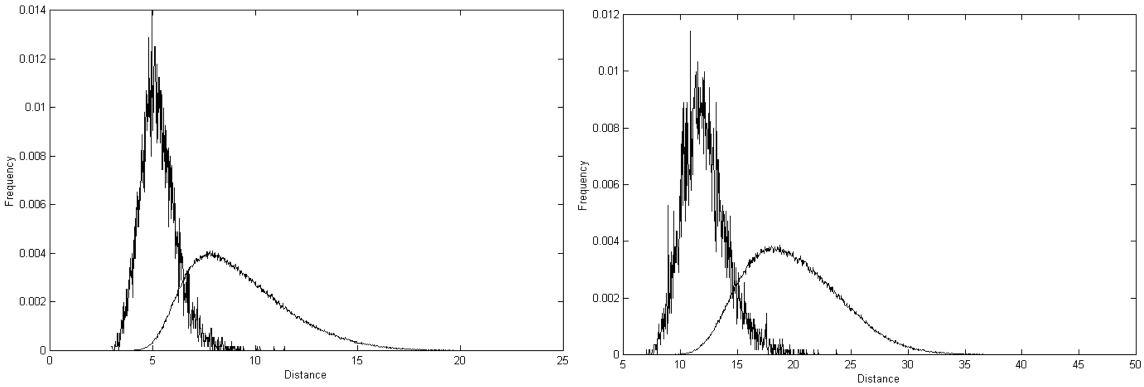
(a) Original distance (b) Projected distance (c) Extension of distance

Fig. 5 MRM at Order of 50



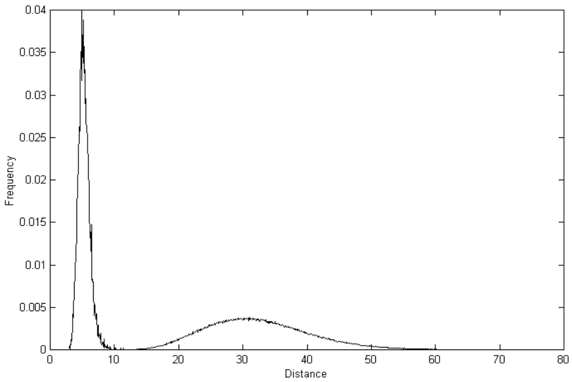
(a) Identification rate (b) F-ratio

Fig. 6 Performance of the system with MRM and without MRM

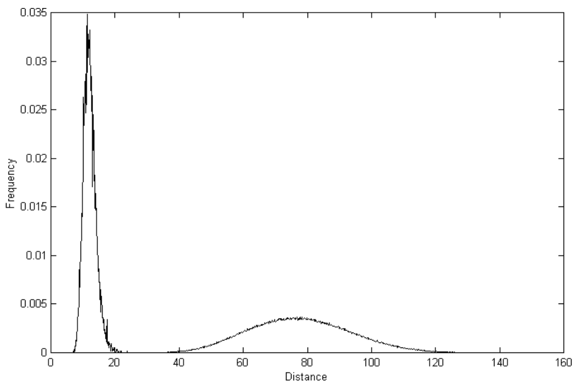


(a) Feature order 10 (b) Feature order 15

Fig. 7 Histogram for imposter and genuine distance for non-MRM identification system



(a) Feature order 10



(b) Feature order 15

Fig. 8 Histogram for imposter and genuine distance for MRM identification system