# A Method to Integrate GMM, SVM and DTW for Speaker Recognition

Ing-Jr Ding, Chih-Ta Yen* and Da-Cheng Ou

Department of Electrical Engineering, National Formosa University, Yunlin, Taiwan, ROC.

Received 05 August 2013; received in revised form 22 August 2013; accepted 24 September 2013

## Abstract

This paper develops an effective and efficient scheme to integrate Gaussian mixture model (GMM), support vector machine (SVM), and dynamic time wrapping (DTW) for automatic speaker recognition. GMM and SVM are two popular classifiers for speaker recognition applications. DTW is a fast and simple template matching method, and it is frequently seen in applications of speech recognition. In this work, DTW does not play a role to perform speech recognition, and it will be employed to be a verifier for verification of valid speakers. The proposed combination scheme of GMM, SVM and DTW, called SVMGMM-DTW, for speaker recognition in this study is a two-phase verification process task including GMM-SVM verification of the first phase and DTW verification of the second phase. By providing a double check to verify the identity of a speaker, it will be difficult for imposters to try to pass the security protection; therefore, the safety degree of speaker recognition systems will be largely increased. A series of experiments designed on door access control applications demonstrated that the superiority of the developed SVMGMM-DTW on speaker recognition accuracy.

**Keywords:** speaker recognition, Gaussian mixture model, support vector machine, dynamic time wrapping, SVMGMM-DTW

## 1. Introduction

Biometrics technology that uses physical characteristics to perform pattern recognition has been widely developed in the recent years. Biometrics based pattern recognition includes fingerprint recognition, face recognition, iris recognition, and voiceprint recognition. Voiceprint recognition, also known as speaker recognition, is quite suitable for identity recognition due to high uniqueness of each speaker. Nowadays, speaker recognition is commonly seen in the applications of access control systems and transaction confirmation systems where confirmation of the user’s identity is required. Speaker recognition will play an important and necessary role in people’s daily life [1].

Speaker recognition can be divided into two categories: speaker identification and speaker verification. This paper focuses on the problem of speaker verification. In a speaker identification task, the purpose of the system is to determine the identity of the person. Speaker identification is generally viewed as a problem to solve “which one;” on the other hand, speaker verification is used to verify the identity of people from their uttered voices and belongs to the issue “yes or no” to be overcome. Model-based schemes have been the mainstream techniques in speaker recognition. Two popular modelling techniques are currently used, which are support vector machine (SVM) [2] and Gaussian mixture model (GMM) [3]. In general, GMM is the optimal modelling technique for speaker identification applications [4, 5] and the primary mission of the SVM approach is to

* Corresponding author. E-mail address: chihtayen@gmail.com

Tel.: 886-5-6315630; Fax: 886-5-6315609

overcome the issue of speaker verification [6, 7]. In this paper, both SVM and GMM classification models are employed to fulfil the task of speaker verification.

Dynamic time wrapping (DTW) that belongs to the category of dynamic programming is a type of optimal algorithms [8]. DTW has been widely used to solve lots of optimal problems including the typical speech recognition problem. DTW that is categorized into feature-based pattern recognition techniques does not need to establish (or train) a classification model in advance; therefore, it is generally viewed as a conceptually simple and direct recognition technique. Conventional DTW is generally used for performing speech recognition, and few studies are seen to employ the DTW technique for implement speaker recognition. In this paper, DTW is used for the purpose of speaker verification.

This paper develops an SVMGMM-DTW approach for speaker verification applications. Proposed SVMGMM-DTW employs DTW to perform speaker recognition with the support of both SVM and GMM classification models will achieve a competitive performance on recognition accuracy. In fact, the work of combining SVM and GMM for speaker recognition have been seen in the recent years such as the SVM kernel development with the support of GMM calculations [9-11]. However, all of those developed combination schemes of GMM and SVM are essentially complicated and computationally expensive. The proposed SVMGMM-DTW will provide a simple and direct scheme to merge SVM and GMM, and template matching of DTW is additionally added into SVM-GMM processing for further enhancing the overall framework of a speaker verification system. The proposed SVMGMM-DTW for speaker verification provides several advantages, as follows:

(1) To avoid unreliable decisions in conventional SVM alone or GMM alone speaker verification methods;

(2) To provide a new direction to use DTW for speaker verification applications;

(3) To offer an efficient and effective scheme to integrate SVM and GMM model-based and DTW feature-based recognition techniques.

## 2. Modeling Techniques of Speaker Recognition

As mentioned before, support vector machine and Gaussian mixture model are two popular speaker recognition techniques. The first is mainly used in the area of speaker verification; the primary work of the last is to solve the problem of speaker identification. In this study, both SVM and GMM are employed to perform speaker verification tasks, which will be introduced in the following sections.

### 2.1. *Support vector machine-based speaker verification*

In this work, SVM is used for speaker verification, is often used as a data classifier generally; is based on the theory of the structural risk minimization of statistics [2]. SVM classifies new input data from a test speaker by using a separating hyperplane. If the SVM model attempts to determine whether an input datum is belonging to the group of valid speakers, it would first try to find the SVM model of the group of valid speakers in the SVM database. Next, the separating hyperplane of the SVM model of the group of valid speakers would classify the input datum as valid or invalid.

Suppose a set of labeled training points is as follows:

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n). \tag{1}$$

Each training point $x_i$ belongs to either of two classes and is given a label, $y_i \in \{-1, 1\}$ for $i = 1,2,\ldots,n$. From these training data, the hyperplane is defined by the pair ( $w$ , $b$ ), such that the point $x_i$ can be separated according to the function.

$$w \cdot x + b = 0 \tag{2}$$

$$f(x_i) = sign(w \cdot x_i + b) = \begin{cases} 1, & if \ y_i = 1 \\ -1, & if \ y_i = -1 \end{cases} \tag{3}$$

The set *S* is linearly separable if there exists a pair ( $w$ , $b$ ) such that the inequalities

$$\begin{cases} (w \cdot x_i + b) \geq 1, & if \ y_i = 1 \\ (w \cdot x_i + b) \leq -1, & if \ y_i = -1, \end{cases} \qquad i = 1,2,\ldots,n \tag{4}$$

are valid for all elements of set *S*. If the set *S* is linearly separable, a unique optimal hyperplane exists, and for this hyperplane, the margin between the projections of the training points of two different classes is maximized. If set *S* is not linearly separable, classification violations must be allowed in the SVM formulation. To deal with data which is not linearly separable, the previous analysis can be generalized by introducing some nonnegative variables $\zeta_i \geq 0$ such Eq. (4) is modified by the following formula,

$$y_i(w \cdot x_i + b) \geq 1 - \zeta_i, \quad i = 1,2,\ldots,n \tag{5}$$

Then, the solution to the optimal hyperplane problem is

$$Minimize \ \frac{1}{2} w \cdot w + C \sum_{i=1}^{n} \zeta_i \tag{6}$$

where *C* is a constant. The Lagrangian method [2] can be used in searching for the optimal hyperplane in Eq. (6). Fig. 1 depicts that a SVM separation hyperplane is used to classify the input data that is acquired from a test speaker as the valid speaker's or the imposter's. In Fig. 1, the black point indicates the category of the valid speaker, and the white point denotes the category of the imposter.

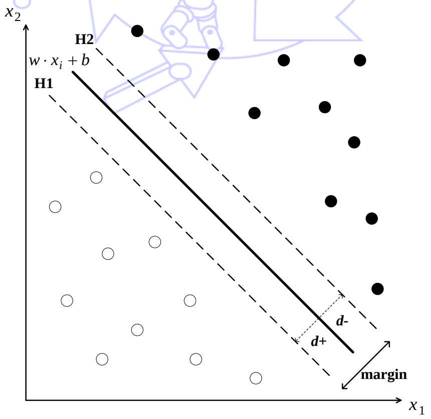

Fig. 1 Classification of speakers by a SVM hyperplane (black points: valid speakers, white points: imposters)

### 2.2. Gaussian mixture model-based speaker verification

GMM models are adopted in the development of speaker verification systems in this study. Mathematically, a GMM is a weighted sum of *M* Gaussians, denoted as [3]

$$\lambda = \{w_i, \mu_i, \Sigma_i\}, \ i = 1, 2, \ldots, M \ , \ \sum_{i=1}^{M} w_i = 1 \tag{7}$$

where $w_i$ is the weight, $\mu_i$ is the mean and $\Sigma_i$ is the covariance.

To determine the GMM model parameters for a certain sound class, the E-M algorithm is readily applicable. It is noted that before running the E-M algorithm, a crucial job is to initialize the model first, i.e., to assign starting values to the parameters, which can be realized by a binary splitting vector quantization algorithm. With the initial model parameter settings, the E-M process starts iteratively maximizing the likelihood estimate of the training data by adjusting the initial model parameters; specifically, the expectation and the maximization steps in the E-M process are repeated so that the parameter set $\lambda = \{w_i, \mu_i, \Sigma_i\}, \ i = 1, 2, \ldots, M$ of the GMM converges to an equilibrium state. In general, the number of iterations of the E-M algorithm typically goes as high as several thousands. In this work, two GMM models for representing "the valid speakers" and "the imposters" are established. At the end of the training phase of GMM models, two sets of $\lambda$ parameters (2 GMM models, that is), $\lambda_1$ and $\lambda_2$, are finally determined. Fig. 2 shows two GMM models, the valid speaker's and the imposter's, established for speaker verification tasks.

After the training, the speaker recognition procedure can be executed based on two trained GMM models of $\lambda_1$ and $\lambda_2$. Note that the classifier deployed here is basically a GMM classifier consisting of two separate GMM models, one for the group of valid speakers, and the other for the group of imposters. During the recognition phase of speaker verification, the class of $X$, covering $n$ acoustic feature vectors of $D$ dimensions, $X = \{x_i \mid i = 1, 2, \ldots, n\}$, is determined by maximizing a *posteriori* probability $P(\lambda_s \mid X)$,

$$\hat{s} = \max_{s=\{1,2\}} P(\lambda_s \mid X) = \max_{s=\{1,2\}} \frac{f(X \mid \lambda_s)}{f(X)} \cdot P(\lambda_s) \tag{8}$$

Note that

$$f(x_i \mid \lambda_s) = \sum_{j=1}^{M} w_j \cdot b_j(x_i) \tag{9}$$

and

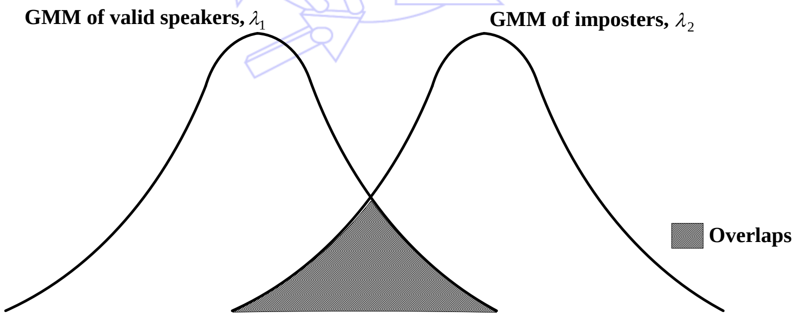
GMM of valid speakers, $\lambda_1$
GMM of imposters, $\lambda_2$
▨ Overlaps

Fig. 2 GMM-based speaker verification by establishing two GMM models of the valid speaker and the imposter

$$b_j(x_i) = \frac{1}{(2\pi)^{D/2} \cdot |\Sigma_s|^{1/2}} \cdot \exp\left\{ -\frac{1}{2}(x_i - \mu_s)^T (\Sigma_s)^{-1} (x_i - \mu_s) \right\} \tag{10}$$

And at the end of the recognition procedure of speaker verification, the signal $X$ is then classified as one of the two sound classes indicated by $\hat{s}$.

## 3. Proposed SVMGMM-DTW for Speaker Verification

This section will introduce speaker verification using the proposed SVMGMM-DTW method. Proposed SVMGMM -DTW firstly combines the above-mentioned GMM and SVM classifiers by a voting scheme to form a Voting-SVMGMM verification scheme, and then the Voting-SVMGMM is further merged by a DTW decision support scheme. DTW technique is generally seen to perform template matching in the application of speech recognition, and designed to fulfill speaker recognition tasks in this work.

### *3.1. The utilization of DTW for speaker recognition*

Conventional DTW is generally used for speech recognition procedures. In this paper, DTW is used for speaker verification. Dynamic time warping categorized into dynamic programming techniques is a nonlinear warping algorithms that mixes time warping and appropriate template matching calculations [8]. Figure 3 illustrates how the DTW algorithm is used to search for an optimal path between the testing data and the referenced template. As shown in Fig. 3, when computing the similarity degree between the testing data and the referenced template, the lower distortion between both of them means the higher similarity degree. The following explains how conventional DTW is performed on speaker verification in this work. The test utterance from a new speaker for identity verification is composed of *T* frames; an arbitrary frame (a feature vector) is denoted as *t*; the referenced template consists in *R* frames; the arbitrary frame in the referenced template is indicated as *r*; the distortion between *T* and *R* frames can be represented by $d[T(t), R(r)]$; the starting point of the overall comparison path is $(T(1), R(1)) = (1,1)$, and the ending point of the path is $(T(M), R(N)) = (T, R)$.

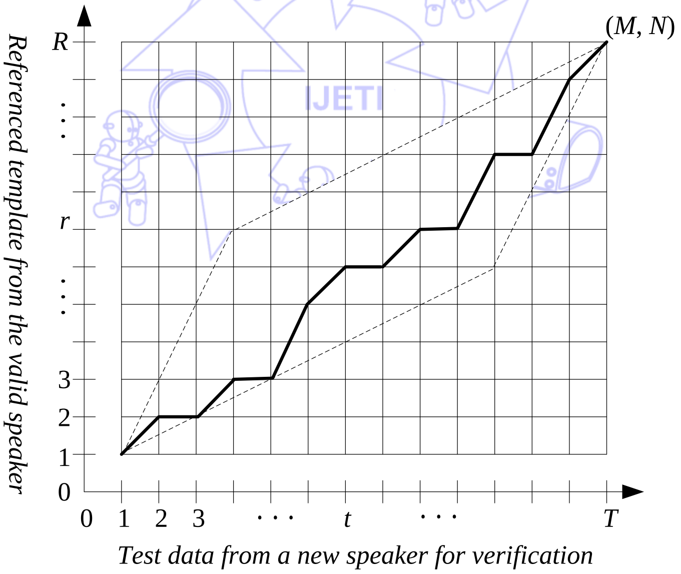
(M, N)
R
r
3
2
1
0
0
1
2
3
· · ·
t
· · ·
T
Referenced template from the valid speaker
Test data from a new speaker for verification

Fig. 3 Template matching operations of DTW used for speaker verification tasks

Based on the above settings of DTW operations, the DTW distance of the optimal comparison path is derived using Eq. (11). Note that the frame *t* in the testing data is generally not equal to the frame *r* in the referenced template in indexes.

$$(optimal)\, D = \min \sum_{m=1}^{M} d(T(m), R(m)) \tag{11}$$

Assume that the point $(T(0), R(0)) = (0,0)$ and $d(0,0) = 0$, the accumulated distance that selects the optimal source path can be represented as follows,

$$min\, D(t,r) = \min_{(t-1,r-1)} \{min\, D(t-1,r-1) + d(t,r)\} \tag{12}$$

The $min\, D(t,r)$ in Eq. (12) indicates the shortest distance from the starting position to the position ($t$, $r$). In Fig. 3, the black solid line is represented as the DTW optimal matching path with the distance derived using Eq. (12), and the dotted line is the global path search constraint to be used to effectively reduce the searching time to acquire an overall optimal path on DTW operations.

Note that after finishing DTW template matching operations, a decision that the speaker is valid or invalid is made according to the similarity degree between the test data and the referenced template. As shown in Fig. 3, the distortion between $T$ and $R$ frames, i.e. $d[T(M), R(N)]$, indicates such the similarity degree. The distortion between test data and referenced template is sometimes also called DTW distance. If the DTW distance is small, the test data from a new speaker is like to referenced template from a valid speaker; therefore, the speaker will be verified as a valid speaker, and then accepted by the system. Conversely, the speaker will be viewed as an imposter and rejected immediately by the system due to a large value of the DTW distance. In this paper, the verification decision of DTW speaker recognition is mainly dependent on the value of the calculated DTW distance.

### *3.2. Proposed SVMGMM-DTW*

Fig. 4 shows the proposed SVMGMM-DTW method for speaker verification applications. To ensure the validity of a new speaker, proposed SVMGMM-DTW is composed of two verification phases, the first phase of combined SVM and GMM verification schemes, and the second phase of the DTW verification scheme. As shown in Fig. 4, the test utterance from a new speaker is request for verification of validity, and the decision of speaker verification with proposed SVMGMM-DTW is made as either “the imposter” or “the valid speaker.” The test speaker is categorized into the group of valid speakers finally only when the test data from the speaker is to pass both the first verification phase and the second verification phase. If the test data is invalid due to a strange speaker to the system or too much background noise contained in the data, the data will be verified as the imposter’s and be neglected immediately in the first phase and no any verification procedure will be continued. Otherwise, the second verification phase will run and the DTW template matching technique will go on to perform verification of speaker’s identity. By providing a double-checking procedure of speaker verification, proposed SVMGMM-DTW will make an extremely reliable recognition decision, and any imposter that tries to pass the system will be effectively prohibited.

The first phase of SVMGMM-DTW is to use a combined scheme of SVM and GMM classifiers for verifying the speaker’s test data. The so-called Voting-SVMGMM is to use a voting algorithm to consider all classification results of these two classifiers. In the Voting-SVMGMM, all of the classification results of speech data derived using the GMM are evaluated within the framework of the voting scheme as those calculated using the SVM. Note that using the voting scheme to GMM, the likelihood evaluation result of each frame categorized as either the valid speaker class or the imposter class, will be appropriate for providing a simple and direct fusion way to combine SVM and GMM classification results. When completing the work of performing Voting-SVMGMM, votes of acceptance and votes of rejection will be determined. Votes of acceptance derived from Voting-SVMGMM come from the accumulation of SVM valid speaker classification votes and GMM valid speaker classification votes; votes of rejection obtained from Voting-SVMGMM are the aggregation of SVM imposter classification votes and GMM imposter classification votes. When votes of acceptance is more than votes of rejection, the test speaker is a candidate of valid speakers and the test data will be fed into the second verification phase again for advanced evaluation. Otherwise, the test speaker is verified as an imposter, and then rejected immediately in this phase.

The second phase of SVMGMM-DTW is to employ DTW for dealing with speaker verification. As mentioned in the previous section, the calculated DTW distance derived from DTW template matching operations is used to evaluate the validity of the test speaker's data. As can be seen in Fig. 4, there are two DTW template matching operations carried out in the second phase of SVMGMM-DTW. One DTW operation is to fulfill the matching of imposters, and the other one is to conduct the matching of valid speakers. In the operation of imposter matching, if the calculated DTW distance between the test speaker's data and the pre-established imposter's data is smaller than the setting value of the threshold, the test speaker is like to the pre-established imposter and then rejected immediately by the system. Otherwise, the test data will continue the following DTW operation of valid speaker matching. In the valid speaker matching operation, if the calculated DTW distance between the test speaker's data and the valid speaker's data is larger than the setting value of the threshold, the test speaker is not like to the valid speaker, is strange to the verification system, is refused by the system. Otherwise, the test speaker will be verified as a valid speaker and accepted finally by the system.

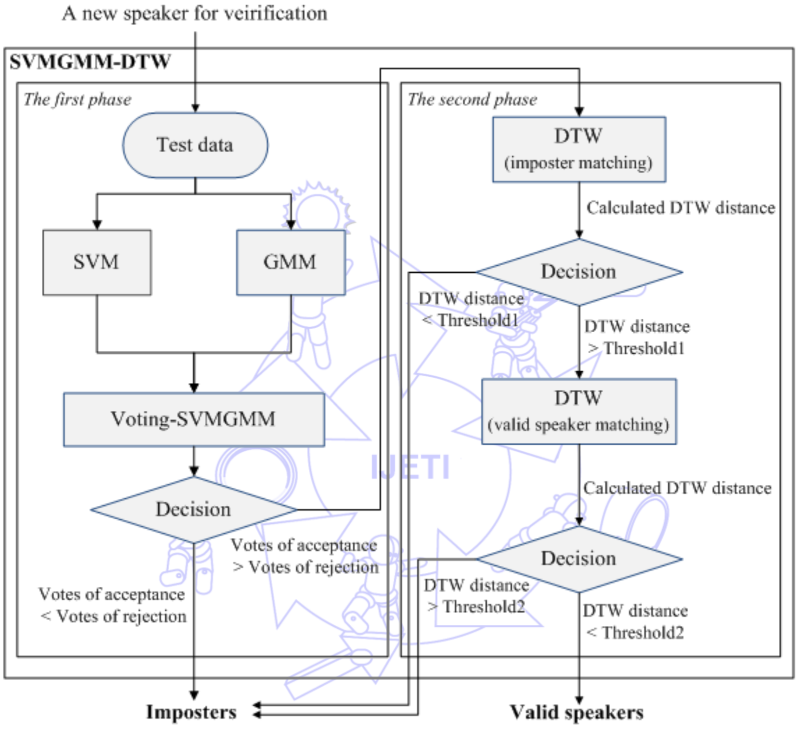

Fig. 4 Frameworks of proposed SVMGMM-DTW speaker verification

## 4. Experiments and Results

Speaker verification experiments with the proposed SVMGMM-DTW method are carried out in a database that is composed of DB-1, DB-2… and DB-10. Each of 10 databases contains both valid speakers' and imposters' speech data. Table 1 shows the ratio between numbers of the valid speaker and numbers of the imposter in each different database. As can be seen in Table 1, the experimental database for speaker verification is well-designed with different settings of valid speakers and imposters in each database. Speaker verification experiments in this work are designed using an access control system application in which the test speaker in each database is requested to utter his (or her) names in Mandarin to be the key for identity verification. The speech data from each test speaker is collected by recording from a microphone. The speech frames are 20-ms wide with a 10-ms overlap. The sampling rate is 44.1kbps. 16-bits of resolutions and mono setting of channels are adopted in this experiment. For each frame, a 10-dimensional feature vector was extracted; the feature vector for each frame is a 10-dimensional cepstral vector.

Table 1 Designed database for speaker verification containing
the different ratio of valid speakers and imposters

| Database for speaker verification | The ratio of numbers of valid speakers and numbers of imposters |
|---|---|
| DB-1 | 7 : 9 |
| DB-2 | 6 : 10 |
| DB-3 | 7 : 7 |
| DB-4 | 8 : 8 |
| DB-5 | 6 : 7 |
| DB-6 | 9 : 7 |
| DB-7 | 10 : 6 |
| DB-8 | 5 : 11 |
| DB-9 | 11 : 5 |
| DB-10 | 4 : 4 |

Speaker verification experiments contain two phases, the training phase and the recognition phase. In the training phase, the main work in this phase is to establish SVM and GMM classification models. DTW templates for pattern matching are also established in the training phase. The training data are collected from 26 speakers where some speakers are chosen as the valid speakers and the other speakers are imposters. Each of the 26 speakers is asked to offer 10 utterances of his (or her) names in Mandarin, 260 utterances in total, as the training data for the establishment of SVM model, GMM model and DTW templates. In this training phase, an SVM classification model for verifying the speaker's utterance, two GMM vocal models that represent the valid speakers and the imposters, and a set of DTW templates to match the utterance from the test speaker are developed using the collected 260 utterances.

Table 2 Comparative experimental results of GMM alone, SVM alone and
Voting-SVMGMM on recognition performances

| Test database | Recognition rates | | |
|---|---|---|---|
| | Speaker verification methods | | |
| | GMM alone | SVM alone | Voting-SVMGMM |
| DB-1 | 65.30% | 69.10% | 74.33% |
| DB-2 | 62.34% | 65.37% | 72.59% |
| DB-3 | 64.57% | 64.93% | 76.32% |
| DB-4 | 64.26% | 67.29% | 74.73% |
| DB-5 | 66.62% | 66.15% | 73.94% |
| DB-6 | 62.84% | 68.09% | 75.04% |
| DB-7 | 62.73% | 68.96% | 74.55% |
| DB-8 | 62.83% | 65.53% | 67.99% |
| DB-9 | 60.45% | 66.74% | 71.71% |
| DB-10 | 66.24% | 68.81% | 72.50% |
| Avg. | 63.82% | 67.10% | 73.37% |

In the recognition phase, each of the 26 speakers is asked again to provide additional 10 utterances of his (or her) names in Mandarin for the performance comparison of speaker recognition methods. There are 260 utterances in total to be as the test data in the recognition phase. Note that the training data in the training phase and the test data in the recognition phase are completely different. Table 2 shows the recognition performance of the GMM alone, the SVM alone and the Voting-SVMGMM. In the average instance involving 10 test databases, the Voting-SVMGMM approach demonstrated the highest recognition rate of 73.37%, followed by the SVM alone at 67.1%, and the GMM alone has the worst recognition accuracy of 63.82%. The competitive superiority of proposed SVMGMM-DTW is shown in Table 3. As shown in Table 3, speaker verification using DTW technique alone without any verification support of SVM, GMM, or combined SVM and

GMM will be still dissatisfactory on recognition performance. The proposed SVMGMM-DTW method to perform DTW speaker recognition with the help of Voting-SVMGMM verification will achieve the best recognition rate of 73.7% among all speaker verification methods.

Table 3 Recognition accuracy comparisons of DTW-alone and proposed SVMGMM-DTW speaker verification

| Test database | Recognition rates | |
|---|---|---|
| | Speaker verification methods | |
| | DTW alone | Proposed SVMGMM-DTW |
| DB-1 | 74.78% | 73.52% |
| DB-2 | 66.92% | 72.96% |
| DB-3 | 69.27% | 76.31% |
| DB-4 | 69.76% | 75.88% |
| DB-5 | 71.23% | 72.80% |
| DB-6 | 72.90% | 75.35% |
| DB-7 | 71.46% | 75.34% |
| DB-8 | 72.96% | 68.22% |
| DB-9 | 72.70% | 73.85% |
| DB-10 | 77.32% | 72.75% |
| Avg. | 71.93% | 73.70% |

## 5. Conclusions

In this paper, a combination of SVM, GMM and DTW, called SVMGMM-DTW, is proposed for speaker verification. The developed SVMGMM-DTW method containing two verification phases, the combined SVM and GMM phase, and the DTW phase, can provide a double-checking scheme to the test data from a new speaker. Speaker verification systems with presented SVMGMM-DTW will make an extremely reliable recognition decision, and the imposter that tries to pass the system will be effectively rejected. Proposed SVMGMM-DTW employs DTW to perform speaker recognition with the support of both SVM and GMM classification models achieves a competitive performance on recognition accuracy. Experimental results on speaker verification demonstrate the superiority of developed SVMGMM-DTW.

## References

[1] B. K. Sy, “Secure computation for biometric data security — application to speaker verification,” IEEE Systems Journal, vol. 3, no. 4, pp. 451–460, 2009.

[2] C. J. C. Burges, “A tutorial on support vector machines for pattern recognition,” Data Mining and Knowledge Discovery, vol. 2, no. 2, pp. 121–167, 1998.

[3] D. A. Reynolds and R. C. Rose, “Robust text-independent speaker identification using Gaussian mixture models,” IEEE Transactions on Speech and Audio Processing, vol. 3, no. 1, pp. 72–83, 1995.

[4] L. Burget, P. Matejka, P. Schwarz, O. Glembek, and J. Cernocky, “Analysis of feature extraction and channel compensation in a GMM speaker recognition system,” IEEE Transactions on Audio, Speech, and Language Processing, vol. 15, no. 7, pp. 1979–1986, 2007.

[5] J. C. Wang, C. H. Yang, J. F. Wang and H. P. Lee, “Robust speaker identification and verification,” IEEE Computational Intelligence Magazine, vol. 2, no. 2, pp. 52–59, 2007.

[6] J. Louradour, K. Daoudi and F. Bach, “Feature space mahalanobis sequence kernels: Application to SVM speaker verification,” IEEE Transactions on Audio, Speech, and Language Processing, vol. 15, no. 8, pp. 2465–2475, 2007.

[7] W. M. Campbell, J. P. Campbell, T. P. Gleason, D. A. Reynolds and W. Shen, “Speaker verification using support vector machines and high-level features,” IEEE Transactions on Audio, Speech, and Language Processing, vol. 15, no. 7, pp. 2085–2094, 2007.

[8] H. Sakoe and S. Chiba, “Dynamic programming algorithm optimization for spoken word recognition,” IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 26, no. 1, pp. 43–49, 1978.

[9] C. H. You, K. A. Lee and H. Li, “GMM-SVM kernel with a Bhattacharyya-based distance for speaker recognition,” IEEE Transactions on Audio, Speech, and Language Processing, vol. 18, no. 6, pp. 1300–1312, 2010.

[10] C. H. You, K. A. Lee and H. Li, “An SVM kernel with GMM-supervector based on the Bhattacharyya distance for speaker recognition,” IEEE Signal Processing Letters, vol. 16, no. 1, pp. 49–52, 2009.

[11] C. Longworth and M. J. F. Gales, “Combining derivative and parametric kernels for speaker verification,” IEEE Transactions on Audio, Speech, and Language Processing, vol. 17, no. 4, pp. 748–757, 2009.