World Applied Sciences Journal 9 (3): 259-267, 2010
ISSN 1818-4952

A Correlation Filter Based Biometric Speaker Authentication Systems

[1]*Dzati Athiar Ramli,* [2]*Salina Abdul Samad and* [2]*Aini Hussain*

[1]School of Electrical & Electronics Engineering, USM Engineering Campus,
Universiti Sains Malaysia, 14300, Nibong Tebal, Pulau Pinang, Malaysia
[2]Department of Electrical, Electronic and Systems Engineering, Faculty of Engineering,
Universiti Kebangsaan Malaysia, 43600, UKM Bangi, Malaysia

Abstract: In this study, we propose a novel approach using Unconstrained Minimum Average Correlation Energy (UMACE) filters as classifiers in an audio-visual biometric speaker authentication system. A modified version of the speech signal in the form of its spectrographic image is used as the audio feature since UMACE filters require 2D-image representation. Lip-reading data are utilized as a second modality, or source, in order to assist the system performance under acoustically degraded condition. In order to extract maximum information from the spectrogram image and distinguish inter-speaker variability hence minimizes the number of training images for filter synthesis, two pre-processing steps i.e. exclusion of low energies and morphological image processing are implemented on the spectrogram. We also recommend a multi-sample fusion for each modality in order to enhance the reliability of the single system performance. Experimental results show that the UMACE filter-based speaker authentication system can be a viable multi-sample multi-source biometric authentication system.

Key words: Correlation filters · Multi modal · Multi sample · Spectrographic · lip-reading

INTRODUCTION

Biometrics is a technology that intends to make use of physiological and behavioral characteristics to identify or authenticate an individual. Traditional ways to verify people require keys, smart cards (something that one has) and passwords (something that one knows). Due to the problem of keys and smart cards that may be stolen and passwords that may be forgotten, biometric features become an alternative technique for person verification [1]. According to Reynolds [2], the benefits of using speech signal trait for biometric systems are that it is natural and easy to produce, requiring little custom hardware, has low computation requirement and is highly accurate (in clean noise-free conditions). But, in uncontrolled conditions, the reliability of the system falls severely as the signal to noise ratio (SNR) of the speech signal drops. This becomes the main problem for utilizing speech signals for biometric systems. Furthermore, since voice is categorized as a behavioral signal, the signal is likely to vary in time due to the change of speaking rates, health and emotional conditions of speakers. Different microphones and channels also affect the accuracy of the system performance. Consequently, the implementation of biometric systems has to appropriately discriminate the biometric features from one individual to another and at the same time, the systems also need to deal with the distortions of the features.

One of the approaches to solve these problems is by incorporating fusion technique to the biometric system architecture. Enhancement of the system performance by using fusion approach has been reported in many studies recently. Keyhanipour *et al.* [3] used the combination of content and context features by employing neural framework to improve the performances of web ranking. Hassan *et al.* [4] reported that the integration of global positioning system and inertial navigation system leads to accurate navigation technology. Adaptive neuro fuzzy inference system has been used for the fusion scheme. In this paper, we focus on multi-sample and multi-modal systems using audio and visual features. The purpose of developing a multi-sample subsystem is to reduce the noise in biometric features caused by intra-class variations and distortions such as the changing of speaking rates. Whereas the execution of a multi-modal

Corresponding Author: Dzati Athiar Ramli, School of Electrical & Electronics Engineering, USM Engineering Campus, Universiti Sains Malaysia, 14300, Nibong Tebal, Pulau Pinang, Malaysia, E-mail: dzati@eng.usm.my

subsystem is to overcome the problem when either one of the modalities is inoperative or fully degraded for example, adversely distorted speech signals as demonstrated in [5-7]. Multi-modal biometric systems are also implemented in order to improve overall system performance as reported in [8-10]. Speech and face modalities have been used in these studies. For each modality, we propose Unconstrained Minimum Average Correlation Energy (UMACE) filters as classifiers. Common classifiers for audio-visual applications include Euclidean distance, DTW, ANN, HMM and SVM as reported in [1, 11]. The motivation of implementing correlation filters as classifiers is because it is robust in presence of distortions such as illumination changes and facial expressions [12, 13]. In addition, the correlation-filter approach uses the entire image as features and hence all the pixel intensity of the image are used for verification and this gives an advantage because no tedious features extraction process is needed [14]. Other characteristics of this advanced correlation filters are shift invariance, closed form expression and capability to suppress imposter using a universal threshold [12, 13]. Correlation filters have been successfully applied in biometric systems for visual application such as face verification and fingerprint verification as reported in [15, 16]. Lower face verification and lip movement for person identification using UMACE filters have been implemented in [17, 18], respectively. A study of using UMACE filters in speaker verification for speech signal as features can be found in [19].

For audio front-end module, we propose a modified version of the speech signal in the form of its spectrographic as audio feature to the system since UMACE filters utilize 2D image representation. Spectrographic image is a modified spectrogram image that represents the time-varying spectrum of speech signal which contains personal information of speakers. Originally, human experts manually analyzed the spectrogram images for semiautomatic speaker recognition [20]. Compared to this traditional way which is more tedious and time consuming, our UMACE filter-based method gives advantages in terms of simplicity and is fully automatic.

For the visual front-end module, lip-reading features are employed in the system. Lip-reading features are the sequence of lip images while the speaker utters the words for example, zero to nine. The advantages of utilizing lip-reading features together with speech signals are due to the simplicity process of data collection and the cost effective factor since they can be simultaneously captured using the same hardware, i.e., digital video camera. We choose lip-reading features instead of static lip images because the movement of the lip contains extra information i.e. behavioral and physical characteristics, while static lips contain just physical characteristics. Furthermore, the implementation of UMACE filter is appropriate for lip-reading verification so as to deal with the changing of lip appearance in the lip sequence. In addition, the use of lip features, compared with face, can also minimize the storage capacity and increases the speed of computation as well. Several researches using lip information as features to recognition systems have been reported. Reference [21] uses shape and intensity information from a person's lip in a speaker recognition system. The utilization of geometric dimension such as height, width and angle of speaker's mouth as features was investigated by Broun *et al.* [22]. Apart from lip contour-based features, pixel-based features i.e. DCT have also been experimented as features for person recognition in [5].

Finally, for the fusion and verification module, two step processes are executed. In the first case, a multi-sample fusion scheme is performed by fusing the scores from several samples from each modality by using an average operator. While in the second step, the final score is given by the combination of the hard score outputs of the audio and visual verification system using AND and OR operators. We then evaluate the proposed system in clean noise-free condition and under noisy conditions. Different levels of SNR of speech signals, ranging from clean to 10dB are experimented so as to simulate the real life conditions.

The database used in this study is the Audio-Visual Digit Database (2001) [23]. The database consists of video and the corresponding audio of people reciting digits zero to nine. The video of each person is stored as a sequence of JPEG images with a resolution of 512 x 384 pixels while the corresponding audio provided is a monophonic, 16 bit, 32 kHz, WAV format.

CORRELATION FILTERS CLASSIFIERS

The theory of correlation filters and further description can be found in a tutorial survey paper by Vijaya Kumar [24]. As described in [13], correlation filters evolve from matched filters which are optimal for detecting a known reference image in the presence of additive white Gaussian noise. However, the detection rate of matched filters drops significantly to even the small changes of scale, rotation and pose of the reference

image. With the intention to deal with these limitations, the Synthetic Discriminant Function (SDF) filter and the Equal Correlation Peak SDF (ECP SDF) filter are introduced. Several training images are utilized as a linear combination to represent a single correlation filter. In this case, a pre-specified value called peak constraint is obtained by ECP SDF filter that corresponds to the authentic class or imposter class when an image is tested. However, the pre-specified peak values lead to misclassifications when the side-lobes are larger than the controlled values at the origin. To address this problem, advanced correlation filters are introduced as reported in [13, 14].

Minimum Average Correlation Energy (MACE) filter is developed to address the problem stated above by reducing the large side-lobes and to produce a sharp peak when the test image is from the same class as the images that have been used to design the filter. To facilitate this condition, the MACE filter variant minimizes the average correlation energy of the training images while constraining the correlation output at the origin to have a pre-specified value. Assume that there are N training images and each image is of size $d_1 x d_2$. The solution of MACE filter is given as in equation (1).

$$\mathbf{H}_{mace} = \mathbf{D}^{-1}\mathbf{X}\left(\mathbf{X}^{+}\mathbf{D}^{-1}\mathbf{X}\right)^{-1}\mathbf{c} \quad (1)$$

by minimizing the average correlation energy (ACE), $E_1 = H^+DH$ while satisfying the constrains, $X^+H = c$. D is a diagonal matrix with the average power spectrum of N training images placed along the diagonal elements. X consists of the Fourier transform of the training images lexicographically re-ordered and placed along each column. c is a column vector of length N containing the desired correlation output at the origin for each training images.

Instead of constraining the correlation output at the origin to have a pre-specified value, the Unconstrained Minimum Average Correlation Energy (UMACE) filter focus on minimizing the average correlation output and relaxing the constrains, $X^+H = c$ by requiring only the average correlation height (ACH). The average correlation height (ACH) is defined as equation (2).

$$ACH = \left|\frac{1}{N}\sum_{i=1}^{N} H^{+}x_i\right| = \left|H^{+}m\right| \quad (2)$$

The unconstrained MACE (UMACE) filter is then found by maximizing the correlation output at the origin and this optimization leads to the following filter equation in equation (3).

$$U_{mace} = D^{-1}m \quad (3)$$

Where m is a column vector containing the mean of the Fourier transforms of N training images.

Correlation filters are synthesized in the frequency domain using Fast Fourier Transform (FFT). For each person in the database, the corresponding correlation filter is designed using several training images. The number of training images used depends on the variation among the training images. In the verification task, the test image in its FFT form is then cross-correlated with the corresponding designed filter of the claimed person. The correlation output is obtained by calculating the inverse FFT of the cross-correlation value. By analyzing the correlation output, the test image can be determined as an authentic or imposter. Peak-to-Side-lobes ratio (PSR) metric is used to measure the sharpness of the peak. Here, the peak is the largest value of the test image of the correlation output. Mean and standard deviation are calculated from the 20x20 side-lobes region by excluding a 5x5 central mask [13].

AUDIO AND VISUAL FRONT-END MODULES

A spectrogram is an image representing the time-varying spectrum of a signal. The vertical axis (y) shows frequency, the horizontal axis (x) represents time and the pixel intensity or color represents the amount of energy (acoustic peaks) in the frequency band y, at time x. In the 1960s, voiceprint analysis, which is semiautomatic speaker recognition, used spectrograms [20]. The computation of spectrogram was done using MATLAB programming (version 7, Release 14). The process of producing spectrogram image is described as follows.

Sampling Process: Sampling process converts the continuous speech signal which denoted as $x_a(t)$ periodically so as to produce sampled sequence $x(n)=x_a(nT)$ $-\infty < n < \infty$. n are integer values, T(sec) is called as sampling period and its reciprocal, $F = \frac{1}{T}(Hz)$ is termed as sampling frequency.

Pre-Emphasis Task: The sampled speech signal is then filtered using a high-pass filter by using equation $H(z) = 1 - a.z^{-1}$, $0 \le a \le 1$, with a = 0.95. A pre-emphasis of high frequencies is required to compress the signal dynamic range by flattening the spectral tilt in order to raise the SNR.

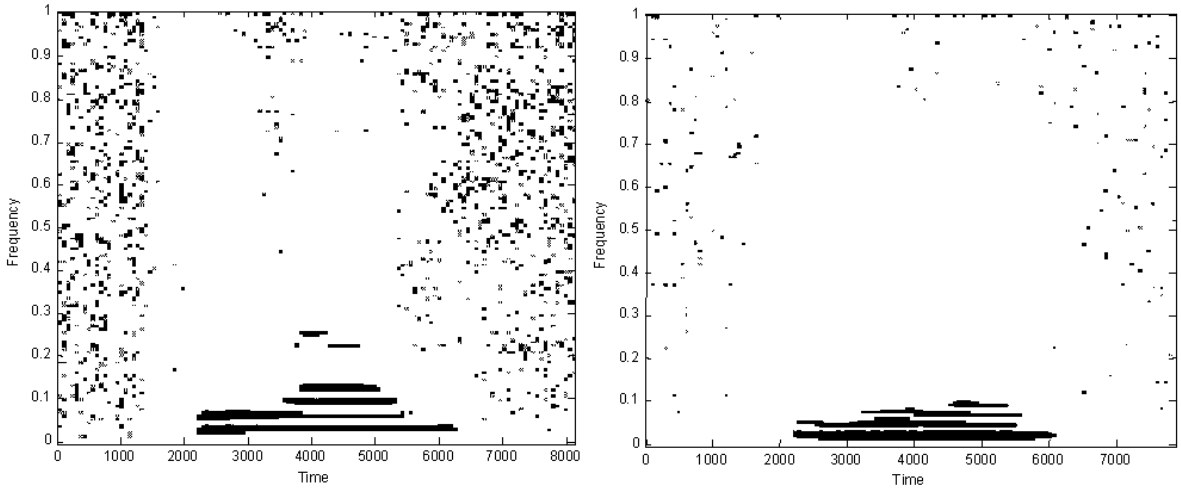

Fig. 1: Samples of spectrogram images

Framing and Windowing Task: The ideal window function should have a very narrow main lobe which increases the resolution and no side lobes or frequency leakage. This task using Hamming window with frame length is equal to 20ms and 50% overlapping is executed on the signal.

FFT Computation Task: D epending on the number of samples per frame, this task uses 1024-point FFT computation and this value determines the frequencies at which the discrete-time Fourier transform is computed.

Spectrogram Image: Finally, the logarithm of energy (acoustic peak) of each frequency bin is calculated. Samples of spectrogram images for word *zero* from different subjects in the database are illustrated as in Figure 1.

Before performing the verification task using UMACE filter, the spectrogram needs to go through a pre-processing phase. In this study, we attempt to reduce the number of training images during filter synthesis. This aim can be achieved by minimizing the variability in the image. In order to accomplish this goal, we optimize the appearance of the spectrogram so that the filter will be able to extract maximum information from the image and distinguish the inter-speaker variability. Two pre-processing steps have been implemented on the spectrogram, i.e. exclusion of the low energies and morphological image processing. For the first task, in order to reduce the variation in image visibility only the high energies are retained in the spectrogram. This process is done by matching the FFT magnitudes during the computation of spectrograms. The objective of the second task is to remove the noisy spot in the image. This task is done by using a morphological opening process. In order to recover back the original shape of the image, morphological closing is then utilized on the image. Besides that, morphology technology is applied in order to acquire a smooth contour line as described in reference [25].

Our spectrographic image database consists of 10 groups of spectrographic images (zero to nine) of 25 persons with 46 images per group of size 32x32 pixels, thus 11500 images in total. For each person, we used 6 training images for the synthesis of the UMACE filter and the other 40 images were used for the testing process. These six training images were chosen based on the largest variations among the 46 images. We use 250 filters which represent each word for the 25 persons. In the testing stage, we performed cross correlations of each corresponding word with 40 authentic images and another 40x24=960 imposter images from the other 24 persons. Summary of the audio front end module is described as in Figure 2.

In order to locate the lips on a face, techniques for face detection and lip localization have been used in this study [26, 27]. In the first task, we implement a color-based technique and template matching algorithm to segment human skin regions from non-skin color. The skin likelihood is computed and then transformed to the skin-segmented image (binary image) and finally, the face region is then portrayed in a rectangle.

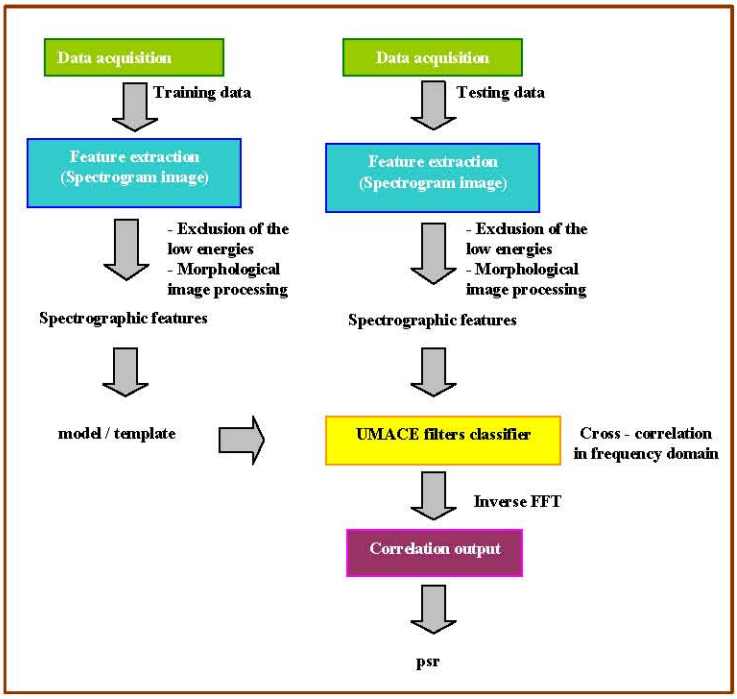

Fig. 2: Audio front end module

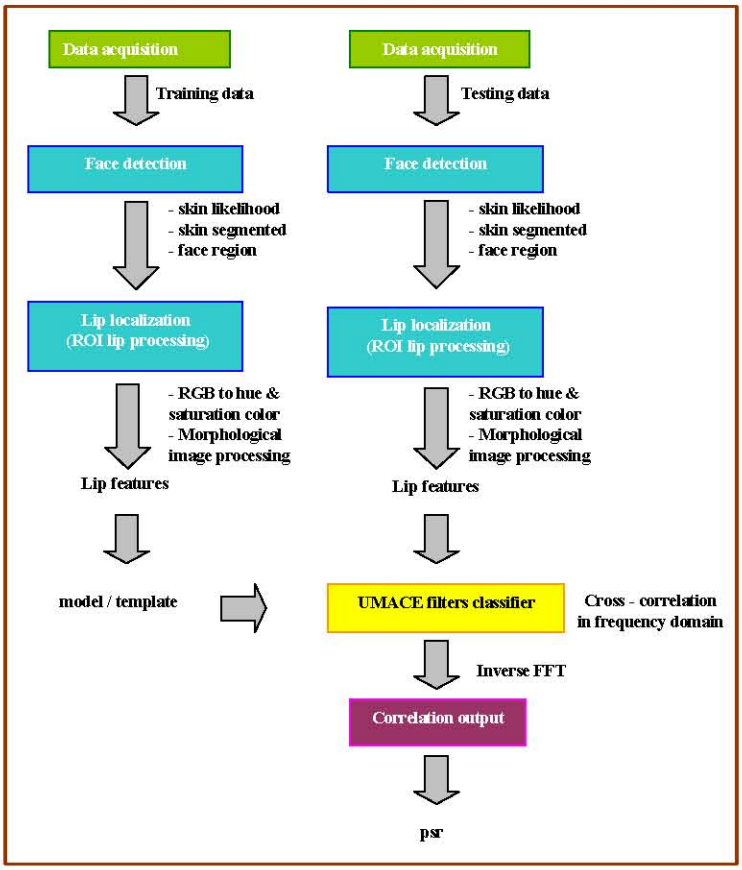

Fig. 3: Visual front end module

For the lip localization task, hue/saturation color thresholding has been employed in order to differentiate the lip area from the face [26, 27]. According to Matthews *et al.* [28], the detection of the lip in hue and saturation color is much easier owing to its robustness under wide range of lip colors and varying illumination condition. Furthermore, processing in color format images confirms high parameter sensitivity. Direct conversion to binary form causes information loss as stated in [29]. From the hue-saturation image, a binary image is then computed by setting the threshold values, $H_0 = 0.04$ and $S_0 = 0.1$. H_0 is the threshold value assigned for hue color while S_0 is the threshold value for saturation color. These values are defined by running the experiment using the validation data from the database when the system achieved the optimum performance. By applying morphological image processing, the largest blob is determined as a lip region. The lip regions of 64 x 64 pixels are then extracted for evaluation. For lip-reading features, for each person, 41 sequences of images (with 20 images per sequence) have been utilized. One of the 41 sequences from each person in the database is chosen as training images to synthesize of the UMACE filters. In our case, we have 25 filters which represent each person. 20 images per sequence are used for multi-sample fusion. During the testing stage, for each filter, we performed cross correlations with 800 authentic images (40 authentic sequences) and another 800x24=19200 imposter images (960 imposter sequences) from the other 24 persons. Summary of the visual front end module is described as in Figure 3.

FUSION VERIFICATION MODULES

Multi-sample fusion considers a combination of scores from several samples that are extracted from the same modality. Although this technique employs many data samples, but it does not give any burden on users during data collection because a single and long sample of an utterance and lip-reading frames sequence can be simply separated into a number of short samples. Implementations of the multi-sample fusion approach can be found in [30, 31]. The studies revealed that integrating the scores of multiple samples can boost the performance of biometric systems. Kuncheva [32] investigated six operators for scores combining i.e. average, median, majority vote, maximum, minimum and Oracle. Among the six operators, performance using the average operator outperforms the other operators. As demonstrated in [31], by combining scores from N number of samples, the error due to the classification can be reduced by factor N.

Using average operator, lets assume the score for every sample from an utterance is denoted as s_n :n=1,...,N. Then, by considering each utterance, $\hat{F} = f(s_1 s_2,...,s_N)$ is defined as the fused estimate score while f is the average fusion scheme. Subsequently, the fused estimate score $\hat{F}$ is calculated and these scores from both audio and visual modules are then passed to the multi-modal fusion phase.

Multi-modal fusion deals with the integration of the decision scores from the audio and visual verification systems. A simple fusion scheme namely the voting

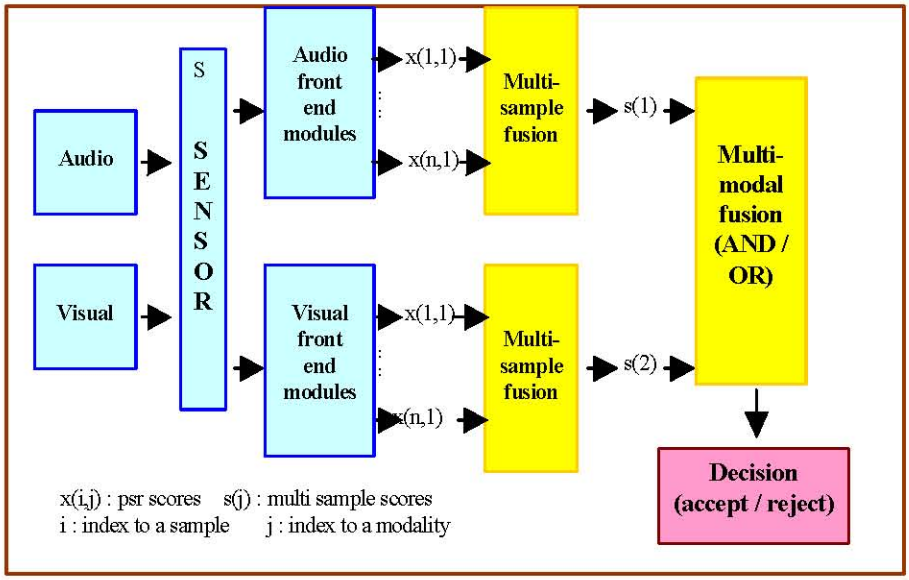

Fig. 4: The architecture of the audio-visual multi-sample fusion

technique is implemented to the fusion process. Voting is a classical empirical technique where the global decision rule is obtained simply by fusing the hard decisions made by two experts. A hard decision is a binary score that only returns either a 0 or 1. This technique accepts the identity claimed by the person under test if at least k-out-of-2 experts decide that the person is genuine [33]. When k=1, this is called the OR rule. The identity claimed is accepted if at least one of the 2 experts decides that the person under the test is authentic. Intuitively, the OR rule leads to the acceptance as fairly simple and rejection as rather difficult. When k=2, it is called the AND rule. The identity claimed is accepted only if both experts decide that the person under test is authentic. The acceptance will be rather difficult and rejection will be undemanding when using the AND rule. The overall architecture of the system is illustrated in Figure 4.

RESULTS AND DISCUSSIONS

For each modality, the score value is calculated via averaging the correlation output values of testing data. By setting a score threshold value, S_0 for each person, the score is then compared to this threshold value, S_0 for accept or reject decision. The false acceptance rate (FAR) and false rejection rate (FRR) are then calculated and overall performance is calculated by combining these two errors into total success rate (TSR).

The overall single-modal systems performance based on TSR percentages using single-sample and multi-sample fusion approaches is compared in Table 1. Excellent improvements are observed when the multi-sample fusion approach is implemented in both modalities. For lip-reading verification, after implementing the multi-sample fusion approach, the result increases by 7.51% and for spectrographic features, the improvement of the performance is observed as 6.83%. The results also show the importance of executing the pre-processing steps i.e. exclusion of the low energies and morphological process to the spectrogram images in order to achieve good performance for the single-sample approach. The performance improves by 19.63% after implementing the pre-processing steps as shown in Table 1.

A wide margin of separation between the maximum PSR values for imposters and the minimum PSR values for the authentic can reduce the percentage of false acceptance rates as well as false rejection rates and therefore gives a better verification performance. For each person lip-reading verification evaluation based on FRR,

Table 1: Performance of single-sample and multi-sample approach for single-modal system

		After pre-processing	
Features	Before pre-processing	Single- sample	Multi-sample
Lip sequence	-	92.29%	99.8%
Spectrographic	73.14%	92.77%	99.6%

Table 2: Performance of multi-sample fusion

Feature	FRR	FAR	TSR
Lip sequence	0.9%	0.1%	99.8%
Spectrographic	5.1%	0.204%	99.6%

Table 3: Performance of multi-sample audio system with decreasing SNR

	Clean	30dB	20dB	10dB
FRR	5.1%	5.7%	13.3%	100%
FAR	0.204%	0.208%	0.558%	0%

Table 4: Error percentages of audio system and fusion system with decreasing SNRs

		clean	30dB	20dB	10dB
FRR	Audio only	5.1	5.7	13.3	100
	OR	0.1	0.2	0.3	0.9
	AND	5.9	6.5	13.9	100
FAR	Audio only	0.204	0.208	0.558	0
	OR	0.295	0.3	0.633	0.29
	AND	0.008	0.008	0.025	0

we found that except for person 17, each authentic person in the database is correctly authenticated. It was observed that images for person 17 were severely corrupted by uneven illumination. The correlation outputs produced very low PSR values for all images in the sequence. The FRR percentage from person 17 is 22.5%. In terms of FAR percentages, an error of 0.4% is observed from persons 3, 10, 15 and 0.8% error is found for person 18. The rest of the people in the database are correctly verified. For the evaluation using spectrographic features, we found that error rate percentages based on FRR are below than 5% except for persons 1 and 3, which are found as 17% and 15%, respectively. Based on FAR, the error percentages are below than 0.8%. We observe that the performance of person 1 and 3 are due to severe corrupted speech signals and inconsistent speaking rate in several data from those people. The overall system performance based on FAR, FRR and TSR percentages are summarized in Table 2.

Table 3 shows the overall error percentages of the system using spectrographic features corrupted by lowering the Signal to Noise Ratio (SNR) from 30dB to 10dB. The system still performs well with the SNRs of 30dB and 20dB. One of the valuable point that we discovered by using spectrographic features through our

proposed UMACE filters-based classifiers is that the small blobs produced in the spectrographic image when the signal corrupted by white noise are diminished after applying the morphological processes. In other words, the corrupted spectrographic image is nearly restored to its clean noise-free condition. Nevertheless, at 10dB, the system totally fails to verify the all authentic persons resulting in PSR values below the threshold and this leads to all test data verified as imposters. Tables 4 compares the error percentages based on FRR and FAR for the single-modal (audio) system and the multi-modal system.

From the results, we can conclude that, the OR operator leads to acceptance being rather easy and this is excellent for the authentic access case since it means that the FRR is likely to be small. But, this is unfavorable with respect to the protection against potential imposters since the FAR tends to be higher. On the other hand, the AND operator make acceptance difficult and this is critical to the authentic access but good with respect to the protection against imposters. In our case, OR operator is preferable due to its ability to reduce the FRR error of multi-modal system in both clean and acoustically degraded conditions. A big improvement based on FRR is observed where 100% error is reduced to 0.9% by using OR operator while the error remains 100% for the AND operator. Although the OR operator introduces a bit higher FAR compared to AND operator but performing multi-modal system using the OR operator is required to maintain the reliability of the overall performance.

CONCLUSION

This study has shown that our novel spectrographic features can be used as features in a speaker authentication system using the UMACE filters as the classifiers. The implementation of multi-sample approach is required to enhance the performance of single-modal systems. This approach does not give any burden on users during data collection because a single and long sample of an utterance and lip-reading frames sequence can be simply separated into a number of short samples. It is also shown that lip-reading features can be employed to increase the performance of the system with acoustically degraded signal inputs. This concludes that the proposed UMACE filter-based authentication system can be an alternative technique to be implemented in a multi-modal biometric speaker authentication system due to its ability to perform well in image variations as well as in noisy conditions. Future study will be devoted to the improvement of the multi-modal fusion decision schemes in order to be more adaptive in various level of SNR and the increment of number of person and image.

ACKNOWLEDGMENT

This research is supported by the following research grants: Fundamental Research Grant Scheme, Malaysian Ministry of Higher Education, FRGS UKM-KK-02-FRGS0036-2006, Science Fund, Malaysian Ministry of Science, Technology and Innovation, 01-01-02-SF0374 and Incentive Grant Universiti Sains Malaysia.

REFERENCES

1. Campbell, J.P., 1997. Speaker Recognition: A Tutorial, Proceeding of the IEEE, 85: 1437-1462.
2. Reynolds, D.A., 2002. An Overview of Automatic Speaker Recognition Technology, Proceeding of IEEE on Acoustics Speech and Signal Processing, 4: 4072-4075.
3. Keyhanipour, A.H., M. Piroozmand, K. Badie, 2009. A Neural Framework for Web Ranking Using Combination of Content and context Features, World Appl. Sci. J., 6(1): 6-15.
4. Hassan, A.M. and S. Khairulmizam, 2009. Integration of Global Positioning System and Inertial Navigation System with Different Sampling Rate using Adaptive Neuro Fuzzy Inference System, World Appl. Sci. J., 7(Special Issue of Computer & IT): 98-106.
5. Fox, N.A. and R.B. Reilly, 2004. Robust Multi-Modal Person Identification with Tolerance of Facial Expression, Proceeding of IEEE International Conference on System, Man and Cybernetics, pp: 580-585.
6. Sanderson, C. and K.K. Kuldip, 1999. Multi-Modal Person Verification System Based on Face Profile and Speech, International Symposium on Signal Processing and its Applications, pp: 947-950.
7. Brunelli, R. and D. Falavigna, 1995. Personal Identification using Multiple Cue, IEEE Trans. on Pattern Analysis and Machine Intelligence, 17(3): 955-966.
8. Teoh, A., S.A. Samad and A. Hussein, 2004. Nearest Neighbourhood Classifiers in a Bimodal Biometric Verification System Fusion Decision Scheme, Journal of Research and Practice in Information Technology, Australian Computer Society, 36(1): 47-62.
9. Teoh, A., S.A. Samad and A. Hussein, 2004. Neighbourhood Classifiers in Biometric Fusion, International, Journal of the Computer, the Internet and Management, Assumption University of Thailand, 12(1): 23-36.

10. Teoh, A., S.A. Samad and A. Hussein, 2005. Fusion Decision for a Bimodal Biometric Verification System using Support Vector Machine and its variations, ASEAN Journal on Science and Technology for Development, ASEAN Committee on Science and Technology 19(1): 1-16.
11. Campbell, W.M., 2003. A SVM/HMM System for Speaker Recognition, Proceeding of IEEE on Acoustics Speech and Signal Processing, 2: 209-212.
12. Vijaya Kumar, B.V.K., M. Savvides, K. Venkataramani and C. Xie, 2002. Spatial Frequency Domain Image Processing for Biometric Recognition, Proceeding of International Conference on Image Processing, 1: 53-56.
13. Savvides, M., K. Venkataramani and B.V.K. Vijaya Kumar, 2003. Incremental Updating of Advanced Correlation Filters for Biometric Authentication System, Proceeding of ICME, 3: 229-232.
14. Venkataramani, K. and B.V.K. Vijaya Kumar, 2004. Performance of Composite Correlation Filters in Fingerprint Verification, Optical Engineering, pp: 1820-1827.
15. Savvides, M., B.V.K. Vijaya Kumar and P. Khosla, 2002. Face Verification using Correlation Filters, Proceeding of Third IEEE Automatic Identification Advanced Technologies, pp: 56-61.
16. Venkataramani, K. and B.V.K. Vijaya Kumar, 2003. Fingerprint Verification using Correlation Filters, System AVBPA, pp: 886-894.
17. Samad, S.A., D.A. Ramli and A. Hussain, 2007. Lower Face Verification Centered on Lips using Correlation Filters, Information Technology J., 6(8): 1146-1151.
18. Samad, S.A., D.A. Ramli and A. Hussain, 2007. Person Identification using Lip Motion Sequence, In: B. Apolloni, R.J. Howlett and L. Jain, (eds.): Knowledge-Based Intelligent Information Engineering Systems. Lecture Notes in Computer Science, Vol. 4692. Springer-Verlag, Berlin Heidelberg New York, pp: 839-846.
19. Samad, S.A., D.A. Ramli and A. Hussain, 2007. A Multi-Sample Single-Source Model using Spectrographic Features for Biometric Authentication, IEEE International Conference on Information, Communications and Signal Processing, CD ROM.
20. Klevents, K. and R.D. Rodman, 1997. Voice Recognition - Background of voice recognition, Artech House Publisher, INC Norwood.
21. Wark, T. and S. Sridharan, 1998. A Syntactic Approach to Automatic Lip Feature Extraction for Speaker Identification, IEEE International Conference on Acoustics Speech and Signal Processing, 6: 3693-3696.
22. Broun, C.C., X. Zhang, R.M. Mersereau and M. Clements, 2002. Automatic Speechreading with Application to Speaker Verification, IEEE International Conference on Acoustics Speech and Signal Processing, 1: 685-688.
23. Sanderson, C. and K.K. Paliwal, 2001. Noise Compensation in a Multi-Modal Verification System, Proceeding of International Conference on Acoustics, Speech and Signal Processing, pp: 157-160.
24. Vijaya Kumar, B.V.K., 1992. Tutorial Survey of Composite Filter Designs for Optical Correlators, Applied Optics, 31: 4773-4801.
25. Ming, X., W. Xiaopei and H. Quanping, 2009. Algorithm Based on Point Feature for Fingerprint Image Segmentation, World Appl. Sci. J., 7(Special Issue of Computer & IT): 168-174.
26. Chetty, G. and M. Wagner, 2004. Liveness Verification in Audio-Video Speaker Authentication, Proceeding of International Conference on Spoken Language Processing ICSLP 04, pp: 2509-2512.
27. Chetty, G. and M. Wagner, 2004. Automated Lip Feature Extraction for Liveness Verification in Audio-Video Authentication, Proceeding of Image and Vision Computing, pp: 17-22.
28. Matthews, I., J. Cootes, J. Bangham, S. Cox and R. Harvey, 2002. Extraction of Visual Features for Lip-reading, IEEE Trans. on Pattern Analysis and Machine Intelligence, 24(2): 198-213.
29. Rasras, R.J., I.E.L. Emary and D.E. Skopin, 2009. Parallel Processing of ART1 Neural Network Algorithm and Application for Recognition of Color Images, World Appl. Sci. J., 7(8): 1071-1076.
30. Cheung, M.C., M.W. Mak and S.Y. Kung, 2004. Multi-Sample Data-Dependent Fusion of Sorted Score Sequences for Biometric verification, Proceeding of the IEEE Conference on Acoustics Speech and Signal Processing, pp: 229-232.
31. Poh, N., S. Bengio and J. Korczak, 0000. A Multi-Sample Multi-Source Model for Biometric Authentication, Proceeding of the IEEE 12th Workshop on Neural Networks for Signal Processing, pp: 375-384.
32. Kuncheva, L.I., 2001. A theoretical Study on Six Classifier Fusion Strategies, Proceeding of the IEEE Transaction on Pattern Analysis and Machine Intelligence, pp: 348-353.
33. Darasathy, B.V., 1994. Decision Fusion, IEEE Computer Society Press.