http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 6 (8): 1146-1151, 2007
ISSN 1812-5638

Lower Face Verification Centered on Lips using Correlation Filters

Salina Abdul Samad, Dzati Athiar Ramli and Aini Hussain
Department of Electrical, Electronic and Systems Engineering, Faculty of Engineering,
University Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

Abstract: In this study, we investigate the implementation of correlation filter for lower face verification with different expression of images for each speaker. The motivation to implementing lower face images instead of face images is because the mouth is subject to fundamental changes during speech. Furthermore, the smaller size of a lower face image compared that of a face can reduce storage capacity and increase the speed of computation. The performance of lower face verification using Minimum Average Correlation Energy (MACE) filter is evaluated. The results are promising and offer good potential for lower face verification compare with face verification performance.

Key words: Correlation filter, MACE filter, lower face verification biometric

INTRODUCTION

Defining the uniqueness of a person's identity is a necessary task for the successful development of biometric systems. In speaker verification systems, the acoustic speech signal has been used as the feature for the verification process. By using speech characteristics alone, the systems perform very well in ideal or clean conditions. However, the recognition rates decrease significantly in adverse conditions such as with background noise, channel distortion or reverberation (Wark *et al*., 1999).

In order to improve the verification rate due to such effects, integration with other sources of speech information such as visible gestures from the speaker's face and body have been implemented (Yuhas *et al*., 1990). Most face recognition research is often based on static face image by assuming a neutral facial expression. However, the appearance of a face can change considerably during speech due to facial expressions. It is also well-known that visual modality of speaker's mouth region provides additional speech information which can lead to improve speaker recognition and verification system performance (Wark *et al*., 1997, 1999; Broun *et al*., 2002).

In general, visual features for automatic lipreading can be grouped into three categories which are lip contour (shape) based features, pixel (appearance) based features and a combination of both (Hennecke *et al*., 1996). For the lip contour based features, inner and outer lip contour are extracted or geometric features such as mouth height and width are used. In pixel based category, the entire image containing the speaker's mouth (Region Of Interest-ROI) is considered as informative for lipreading (Hennecke *et al*., 1996). Present investigation falls in this category.

In most automatic lipreading system, the ROI is a square containing the image pixels of the speaker's mouth region (Potamianos *et al*., 2003). The ROI can also include larger parts of the lower face, such as the jaw and check (Potamianos *et al*., 2003) or even the entire face (Neti, 2000). Potamianos *et al*. (2003) have demonstrated that using the lower face as the ROI improves lipreading significantly, over using the mouth-only ROI because it contains more information for lipreading. In this research, ROI is the lower face of the speaker's face centered around the lip center.

Several papers on speaker verification using visual information of speaker's mouth region have been reported so far. Wark and Sridharan (1998) found that speaker recognition can be obtained by using shape and intensity information from speaker's lip. Then, Wark *et al*. (1999) investigated the use of these features for robust speaker verification in the presence of background noise. The results show that the performance of the fusion of audio and visual speech information outperforms the performance of either sub-system. Broun *et al*. (2002) demonstrated that the geometric dimensions of the lips are highly effective in reducing both false acceptance and false rejection rates in speaker verification tasks. Liveness verification in audio-video speaker authentication system using dynamic lip features

Corresponding Author: Salina Abdul Samad, Department of Electrical, Electronic and Systems Engineering, Faculty of Engineering, University Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia
Tel: +603-89296590 Fax: +603-89296146

has been done by other researchers (Chetty *et al*., 2004a, b). The findings show that liveness verification can guard against replay attack.

The Minimum Average Correlation Energy (MACE) filters has been successfully applied in the field of automatic target recognition as well as in biometric verification (Savvides *et al*., 2003; Vijaya Kumar *et al*., 2002). Face verification and fingerprint verification using correlation filters have been investigated in Savvides *et al*. (2002) and Venkataramani *et al*. (2003), respectively. Savvides *et al*. (2003) performed face authentication and identification using correlation filter based on illumination variation.

In this study, we investigate the performance of lower face verification with different expression of images for each speaker by using UMACE filters. The motivation to implementing lower face images instead of face images is because the mouth is subject to fundamental changes during speech. Furthermore, the smaller size of a lower face image compared that of a face can reduce storage capacity and increase the speed of computation.

MATERIALS AND METHODS

Minimum Average Correlation Energy (MACE) filters: Correlation filter theory and the descriptions of the design of the correlation filter can be found in a tutorial survey paper by Vijaya Kumar (1992). According to Savvides *et al*. (2003) and Venkataramani *et al*. (2003), correlation filter evolves 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 decrease significantly to even the small changes of scale, rotation and pose of the reference image. In an effort to solve this problem, the Synthetic Discriminant Function (SDF) filter and the Equal Correlation Peak SDF (ECP, SDF) filter ware introduced which allowed several training images to be represented by a single correlation filter. SDF filter produces pre-specified values called peak constraints. This peak values corresponds to the authentic class or imposter class when an image is tested. However, the pre-specified peak values lead to misclassifications when the sidelobes are larger than the controlled values at the origin.

Savvides *et al*. (2002) developed the Minimum Average Correlation Energy (MACE) filters. This filter reduces the large sidelobes and produces a sharp peak when the test image is from the same class as the images that have been used to design the filter. There are two kinds of variants that can be used in order to obtain a sharp peak when the test image belongs to the authentic class. The first MACE filter variant minimizes the average correlation energy of the training images while constraining the correlation output at the origin to a specific value for each of the training images. The second MACE filter variant is the Unconstrained Minimum Average Correlation Energy (UMACE) filter which also minimizes the average correlation output while maximizing the correlation output at the origin (Savvides *et al*., 2003).

The solution of MACE and UMACE filter can be summarized as follows:

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

$$U_{mace} = D^{-1}m \quad (2)$$

respectively, where D is a diagonal matrix with the average power spectrum of the 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. Finally, m is a column vector containing the mean of the Fourier transforms of the training images (Savvides *et al*., 2003).

Lower face verification using correlation filter can be described as shown in Fig. 1. Each correlation filter is designed by using several training images. The test image is then cross-correlated with the designed filter (template). By analyzing the correlation output, the test image can be determined as an authentic or imposter.

Peak-to-Sidelobe Ratio (PSR) metric is used to measure the sharpness of the peak. The PSR is given by:

$$PSR = \frac{Peak - mean}{\sigma} \quad (3)$$

Here, the peak is the largest value of the test image yield from the correlation output. Mean and standard deviation are calculated from the 20×20 sidelobe region by excluding a 5×5 central mask (Savvides *et al*., 2002). Figure 2 and 3 show the sample of the correlation plane for the test image from the authentic and imposter classes, respectively.

Lower face verification: The lower face images are obtained from the facial expression database of Advanced Multimedia Processing (AMP) Lab at CMU (Sim *et al*., 2001). The lower face images are obtained by cropping the face images which are of 64×64 pixels in this database. In order to standardize all the lower face images, the coordinate (x_t, y_t) of the center of gravity (centroid) of

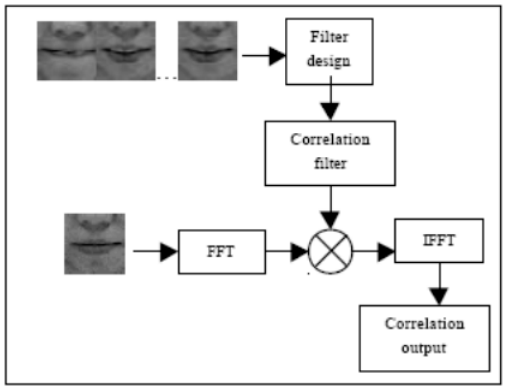

Fig. 1: The correlation filter process

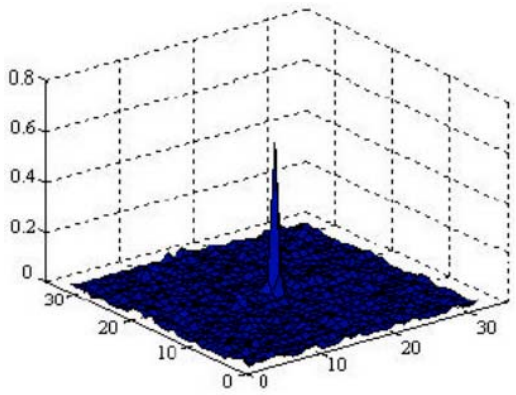

Fig. 2: Example of the correlation plane for the test image from the authentic class

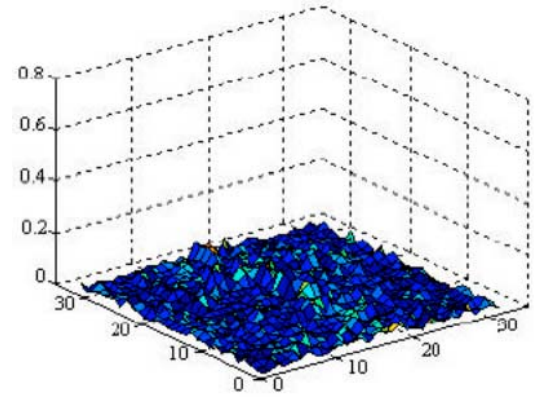

Fig. 3: Example of the correlation plane for the test image from the imposter class

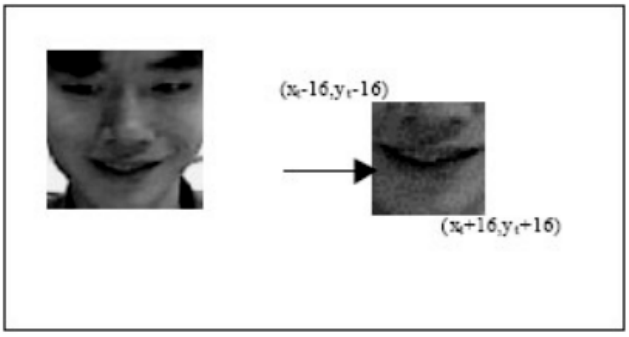

Fig. 4: The extraction of the lower face image from the face image

Fig. 5: Sample images from the lower face database of person 1

each lip image is calculated from the image. The resulting image of 32×32 pixels is cropped from this center of gravity coordinate. Figure 4 shows the extraction of the lower face image from the face image.

The lower face database consists of 10 persons where each person has 40 lower face images with varying expressions. Figure 5 shows the sample of the lower face images from the lower face database of person 1.

In this experiments, we used 3 training images for the synthesis of each person's UMACE filter. These three training images were chosen based on the largest variations among the 40 images of each person. In the test stage, for each filter (in our case we have 10 filters), we performed cross correlations of each person with 40 authentic images and another 40×9 = 360 imposter images from the other 9 persons.

The performance of the verification was measured by setting a PSR threshold value, T_0. The False Acceptance Rate (FAR) and False Rejection Rate (FRR) were then calculated as defined as below.

$$FAR = \frac{\text{No. of imposter images (PSR} > T_0)}{\text{Total imposter images}} \quad (4)$$

$$FRR = \frac{\text{No. of authentic images (PSR} < T_0)}{\text{Total authentic images}} \quad (5)$$

RESULTS AND DISCUSSION

In the experiments, the performance of each person's UMACE filter was evaluated by cross-correlating all the images in the database and their corresponding PSR were measured and recorded. Figure 6 and 7 show the best PSR performance (person 10) and the worst PSR performance (person 6), respectively. Solid line refers to the PSR of the authentic images while dotted line to the imposters. For person 10, PSR for all imposters are below 10 whereas for the authentic the values are above 30. Here, a wide margin of separation between the maximum PSR value for imposters and the minimum PSR value for the authentic offers a better potential for the verification process.

On the other hand, a small gap between the maximum PSR value for the imposters and the minimum PSR value for the authentic from person six is found. It was observed that, the image dataset from person 6 has more variations than others. Due to this, the variations in the test images were not sufficiently synthesized while designing the filter. The verification performance can be improved by using more training images for its filter. Figure 8 shows the PSR performance for person 6 by using 5 training images. A larger separation between the PSR values for the imposter and authentic is observed in this experiment compared with the experiments using 3 training images.

By using 5 training images, the maximum value for imposters is 14.53 and the minimum value for authentic is 22.85 compared with 13.3 and 15.4, respectively, while with 3 training images. However, the system is able to achieve 100% verification performance for both with 3 training images and 5 training images.

Table 1 summarizes the probability of False Acceptance Rates (FAR) and the probability of False Rejection Rates (FRR) using 3 training images by setting the PSR threshold at different levels. All persons accept person 6 and person 8, have a wide margin of separation between the maximum threshold when FRR = 0 and minimum threshold when FAR = 0.

We also compare the result of lower face verification versus face verification performed by Savvides *et al.* (2002) in terms of its error percentages based on PSR values, storage capacity and speed. In terms of PSR values, comparison between the performance of lower face verification and face verification are given by Table 2 and 3, respectively. For lower face verification, it shows that all 10 persons are 100% correctly classified whereas for face verification, the percentage of FAR and FRR are 1 and 2.6%, respectively.

In term of storage capacity, for each lower face image which is of size 32×32 pixels, the memory space has been reduce to 75% instead of using 64×64 pixels face image. In

Table 1: Probability of False Accepted Rate (FAR) and False Rejection Rate (FRR) at different PSR thresholds (10 to 16) for all 10 persons

FAR, FRR = 0				FRR, FAR = 0		
10	11	12	13	14	15	16
0	0	0	0	0	0	0.025
0.006	0.003	0	0	0	0	0
0	0	0	0	0.025	0.075	0.175
0.008	0	0	0	0	0	0
0.006	0.003	0	0	0	0	0
0.1	0.041	0.019	0.008	0	0	0.025
0.022	0.003	0	0	0	0	0
0.013	0.003	0	0	0	0.025	0.025
0	0	0	0	0	0	0
0.006	0	0	0	0	0	0

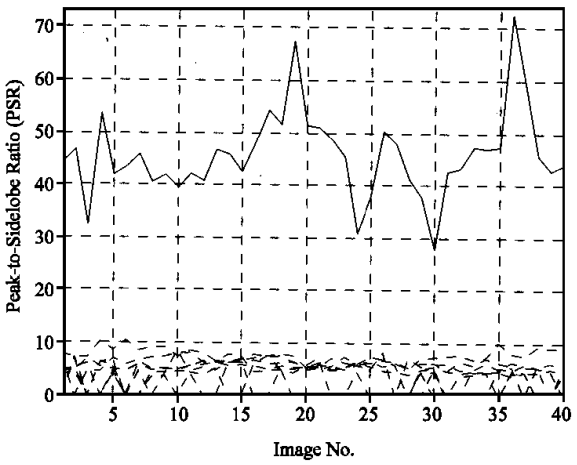

Fig. 6: The best PSR performance using 3 training images (person 10)

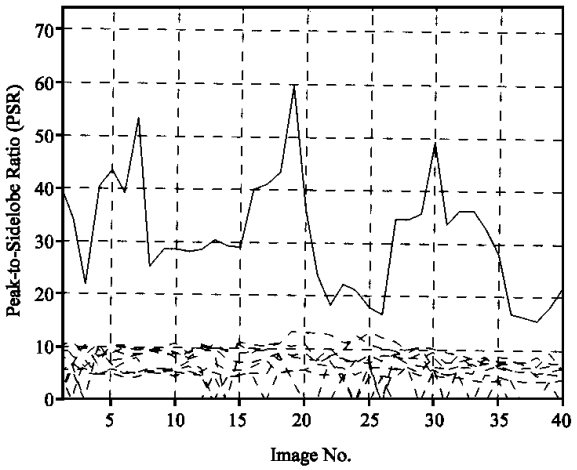

Fig. 7: The worst PSR performance using 3 training images (person 6)

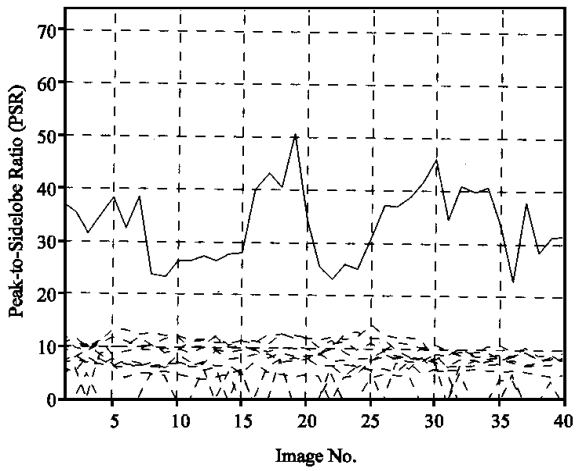

Fig. 8: PSR performance of person 6 using 5 training images

Table 2: Error percentage for lower face verification for all 10 persons using 3 training images

Person	1	2	3	4	5	6	7	8	9	10
FAR, FRR = 0	0	0	0	0	0	0	0	0	0	0
FRR, FAR = 0	0	0	0	0	0	0	0	0	0	0

Table 3: Error percentage for face verification performance for all 10 person using 3 training images

Person	1	2	3	4	5	6	7	8	9	10
FAR, FRR = 0	0	0	0	1	0	0	0	0	0	0
FRR, FAR = 0	0	0	0	2.6	0	0	0	0	0	0

addition, the speed of computation while running each test image increase up to 51.6% while performing the experiment with lower face verification compared with face verification.

CONCLUSIONS

In this study, we have evaluated the performance of lower face verification using correlation filter. We have found that the results are promising and it can be an alternative solution for performing lower face verification. The finding also revealed that UMACE filter are tolerant to different expressions and have the ability to suppress imposter. Compared with face verification, it also maintains a good performance in term of the error percentages, storage capacity and speed of computation.

REFERENCES

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, May 2002, Orlando.

Chetty, G. and M. Wagner, 2004a. Liveness verification in audio-video speaker authentication. Proceeding of International Conference on Spoken Language Processing ICSLP 04, 4-8 Oct 2004, Jeju Island, Korea.

Chetty, G. and M. Wagner, 2004b. Automated lip feature extraction for liveness verification in audio-video authentication. Proceeding of Image and Vision Computing, 2004, New Zealand.

Hennecke, M.E., D.G. Stork and K.V. Prasad, 1996. Visionary Speech: Looking Ahead to Practical Speechreading Systems. Speechreading by Humans and Machines. Lecture Notes in Computer Science, Springer Verlag, pp: 331-349.

Neti, C., 2000. Audio visual speech recognition. Center for Language and Speech Processing. The Johns Hopkins University, Final Workshop 2000 Report. www.clsp.jhu.edu/ws2000/group/av_speech/.

Potamianos, G. and C. Neti, 2001. Improved ROI and within frame discriminant features for lipreading. Proceeding of the International Conference on Image Processing, 7-10 Oct 2001, Thessaloni, Greece.

Potamianos, G., C. Neti, G. Gravier, A. Garg and A.W. Senior, 2003. Recent advances in the automatic recognition of audio-visual speech. Proceeding of the IEEE, 91: 1306-1326.

Savvides, M., B.V.K. Vijaya Kumar and P. Khosla, 2002. Face verification using correlation filters. Proceeding of 3rd IEEE Automatic Identification Advanced Technologies, 2002.

Savvides, M. and B.V.K. Vijaya Kumar, 2003. Efficient design of advanced correlation filters for robust distortion-tolerant face recognition. Proceeding of the IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS'03), 2003.

Savvides, M., K. Venkataramani and B.V.K. Kumar, 2003. Incremental updating of advanced correlation filters for biometric authentication system. Proceeding of ICME, 3: 229-232.

Sim, T., S. Baker and M. Bsat, 2001. The CMU Pose, Illumination and Expression (PIE) database of human faces. Tech Report CMU-RI-TR-01-02, Robotics Institute, Carnergie Mellon University.

Venkataramani, K. and B.V.K. Vijaya Kumar, 2003. Fingerprint Verification Using Correlation Filters. System. AVBPA 2003, 9-11 June 2003, Guildfart, UK.

Vijaya Kumar, B.V.K., 1992. Tutorial survey of composite filter designs for optical correlators. Applied Optics, 31: 4773-4801.

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 (ICIP), 22-25 Sept 2002, Rochester, New York.

Wark, T., D. Thambiratnam and S. Sridharan, 1997. Person authentication using lip information. Proceeding of IEEE on Speech and Image Technologies for Computing and Telecomunications (TENCON), 2-4 Dec 1997, Brisbane, Australia.

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, 12-15 May 1998, Seattle, USA.

Wark, T., S. Sridharan and V. Chandran, 1999. The use of speech and lip modalities for robust speaker verification under adverse conditions. IEEE International Conference on Acoustics Speech and Signal Processing, 7-11 June 1999, Florence, Italy.

Yuhas, B.P., M.H. Jr. Goldstein, T.J. Sejnowski and R.E. Jenkins, 1990. Neural network models of sensory integration for improved vowel recognition. Proc. IEEE, 78: 1658-1668.