Original Article
International Journal of Fuzzy Logic and Intelligent Systems
Vol. 13, No. 3, September 2013, pp. 164-170
http://dx.doi.org/10.5391/IJFIS.2013.13.3.164

ISSN(Print) 1598-2645
ISSN(Online) 2093-744X

Improved Bimodal Speech Recognition Study Based on Product Hidden Markov Model

Su Mei Xi[1,2] and Young Im Cho[1]
[1]College of Information Technology, The University of Suwon, Hwaseong, Korea
[2]School of Science, Qilu University of Technology, Jinan, China

Abstract

Recent years have been higher demands for automatic speech recognition (ASR) systems that are able to operate robustly in an acoustically noisy environment. This paper proposes an improved product hidden markov model (HMM) used for bimodal speech recognition. A two-dimensional training model is built based on dependently trained audio-HMM and visual-HMM, reflecting the asynchronous characteristics of the audio and video streams. A weight coefficient is introduced to adjust the weight of the video and audio streams automatically according to differences in the noise environment. Experimental results show that compared with other bimodal speech recognition approaches, this approach obtains better speech recognition performance.

Keywords: Feature extraction, Bimodal speech recognition, Product hidden Markov model, Weight coefficient

1. Introduction

Speech recognition technology has made great progress in recent decades, and automatic speech recognition (ASR) systems have become increasingly widespread. Since an ASR system is vulnerable to speech noise, and since almost all voice signals contain noise, ASR identification performance using only audio information cannot meet the need. Therefore, developing a robust speech recognition system in a noisy environment is an urgent problem. Developing an integration strategy for audio and visual information is one of the many challenges facing an audio-visual (bimodal) ASR system. From the point of view of perception, video information corresponding to audio information can improve a person's understanding of a speaker's voice. In a noisy environment or for hearing-impaired listeners, video information is a useful complement.

Generally, the audio-visual ASR (AVSR) systems work by the following procedures. First, the acoustic and the visual signals of speech are recorded by a microphone and a camera, respectively. Then, each signal is converted into an appropriate form of compact features. Finally, the two modalities are integrated for recognition of the given speech. Integration of acoustic and visual information aims at obtaining as good recognition results as possible in noisy circumstances. It can take place either before the two information sources are processed by a recognizer early integration (EI) or after they are classified independently late integration (LI). LI has been shown to be preferable because of its better performance and robustness than EI [1], and psychological supports [2].

Received: May. 30, 2013
Revised : Sep. 9, 2013
Accepted: Sep. 14, 2013

Correspondence to: Young Im Cho
(ycho@suwon.ac.kr)

Considering the asynchronous nature of speech signal and video signal, we put forward an improved product hidden Markov model (HMM) in this paper, used for implementing the bimodal voice recognition of Chinese words, which formulate an improved HMM as a multi-stream HMM. Moreover, we control the stream weights of the audio-visual HMM by the generalized Pareto distribution (GPD) algorithm [1, 3], in order to adaptively optimize the audio-visual ASR. According to the corresponding relation between the weight coefficient and instantaneous signal-to-noise ratio (SNR), it can adjust the weight ratio of audio stream and video stream adaptively.

Experimental results demonstrate that the importance of audio features is far higher than that of video features in a quiet environment, while in the presence of noise, video features make an important contribution to speech recognition [4-6].

In Section 2, we introduce some related research works about speech recognition technology. In Section 3, we introduce the classical speech feature parameter extraction approach. In Section 4, we put forward our improved HMM, and, in Section 5, we present our weight optimization approach based on the GDP algorithm. In Section 6, we implement our improved bimodal product HMM system, and, show some experiment results between other ASR systems and the improved HMM. Section 7 is conclusion and our future work.

2. Related Works

In speaker recognition, features are extracted from speech signals to form feature vectors, and statistical patter n recognition methods are applied to model the distribution of the feature vectors in the feature space. Speakers are recognized by pattern matching of the statistical distribution of their feature vectors with target models. Speaker verification (SVR) is the task of deciding, upon receiving tested feature vectors, whether to accept or reject a speaker hypothesis, according to the speaker's model. Mel-frequency cepstral coefficients (MFCC) [7] are a popular feature-extraction method for speech signal processing, and Gaussian mixture models (GMM) have become a dominant approach for statistical modeling of speech feature vectors for text-independent SVR [8].

A recently developed method for overcoming model mismatch is to use a reverberant speech database for training target models [9]. This method was tested on an adaptive-GMM (AGMM)-based SVR system [10] with reverberant speech, with various values of reverberation time (RT). Matching of RT between training and testing data was reported to reduce the equal-error rate (EER) from 16.44% to 9.9%, on average, when using both Z-norm and T-norm score normalizations. However, the study in [9] did not investigate the effect of GMM order on SVR performance under reverberation conditions. In fact, it may be difficult to find such research studies on this effect in the literature.

The audio and visual fusion techniques investigated in previous work include feature fusion, model fusion, or decision fusion. In feature fusion, the combined audio-visual feature vectors are obtained by the concatenation of the audio and visual features, followed by a dimensionality reduction transform [11]. The resulting observation sequences are then modeled using one HMM [12]. A model fusion system based on multi-stream HMM was proposed in [13]. The multi-stream HMM assumes that audio and video sequences are state synchronous but allows the audio and video components to have different contribution to the overall observation likelihood. However, it is well known that the acoustic features of speech are delayed from the visual features of speech, and assuming state synchronous models can be inaccurate. We proposed an audio visual bimodal that uses a product HMM. The audio visual product HMM can be seen as an extension of the multi-stream HMM that allows for audio-video state asynchrony. Decision fusion systems model independently the audio and video sequences using two HMMs, and combine the likelihood of each observation sequence based on the reliability of each modality [11].

3. Speech Feature Parameter Extraction

3.1 Audio Feature Extraction

The normalized energy, MFCC and linear predictive cepstrum coefficients (LPCC) of speech describe the prosodic features, timbre features and perceived features, respectively, so they are selected as audio feature parameters in this paper.

The MFCC computation formula is as follows:

$$\mathrm{MFCC}\,(\mathrm{t},\mathrm{i}) = \sqrt{\frac{2}{N}} \sum_{j=1}^{N} \lg\left[E_{mel}\,(t,j)\right] \cos[i\,(j-0.5)\,\frac{\pi}{N}] \tag{1}$$

where N is the number of triangular filters; $\mathrm{E_{mel}}$(t,j) is the output energy for the j^{-th} filter at t time, $\{\mathrm{MFCC(t,i)}\}_{i=1,2,\ldots,p}$ is the corresponding MFCC parameters at t time, and P of $\{\mathrm{MFCC(t,i)}\}_{i=1,2,\ldots,p}$ is order.

The LPCC computation formula is as follows:

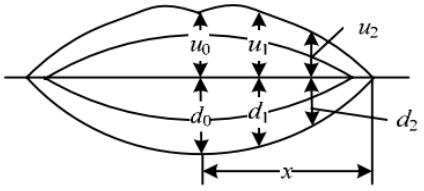

Figure 1. Mouth parameters.

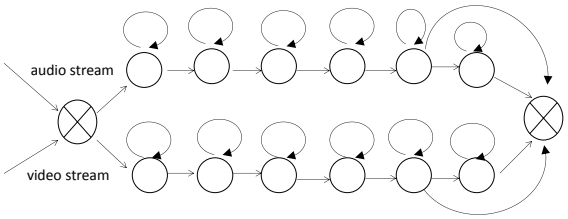

Figure 2. Double stream hidden Markov model model for word speech recognition (⊗ is synchronization point).

$$\begin{aligned}\mathrm{LPCC}\,(\mathrm{t},\mathrm{i}) =& \mathrm{LPC}\,(\mathrm{t},\mathrm{i}) \\ &+\sum_{k=1}^{i-1}\frac{k-i}{i}\mathrm{LPCC}\,(t,i-k)\,\mathrm{LPC}(t,k)\end{aligned} \tag{2}$$

where LPC (t, k) is the $\mathrm{k}^{-\mathrm{th}}$linear prediction coefficient at time t, $\{\mathrm{LPCC(t,i)}\}_{i=1,2,\ldots,p}$ is the corresponding LPCC parameters at time t, and P of $\{\mathrm{LPCC(t,i)}\}_{i=1,2,\ldots,p}$ is order.

3.2 Video Feature Extraction

We select lip parameters as video features, a segment video image using a two-dimensional fast thresholding segmentation algorithm, and lip feature extraction parameters [14], as shown in Figure 1, where x is the distance from the labial center line to the edge, u_0 and d_0 are the heights of the upper and lower halves, respectively, of the mouth centerline, and u_1,d_1,u_2,d_2 are the respective heights of the corresponding x-third point. The original lip parameter is v_t= [x, u_0, d_0, u_1, d_1, u_2, d_2] for the image frame at t time.

4. Improved Product HMM

Auditory and visual features have some synchronicity, with some asynchrony within a certain range. When people talk, mouth movement has already begun before the voice, and it takes time to close the mouth and return to the natural state after the voice, so visual information is usually ahead of auditory information by about 120 ms [15], which is close to the average duration of a phoneme.

Therefore, asynchrony can be permitted in the auditory and visual training model. This paper proposes an improved product HMM model based on the product HMM method. For Chinese words recognition, which usually corresponds to five or six states, only one state migration is allowed between the audio and video streams as a result of the presence of asynchrony. Figure 2 shows the HMM model for the audio and video streams, and Figure 3 shows the topology of the corresponding product

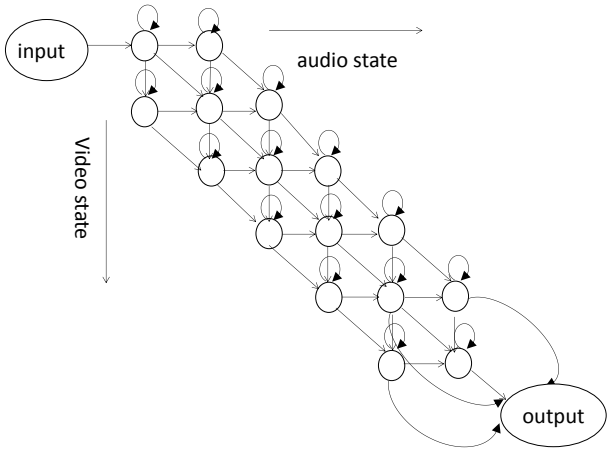

Figure 3. Topological structure of the proposed product hidden Markov model.

HMM model.

We extract the video features and audio features during the training stage. Generally speaking, the frame rate of the phonetic features is higher than the frame rate of the video features, so we use interpolation for them after video-feature extraction in order to ensure training synchronicity in the data flow.

A bimodal speech feature vector consists of the observation vectors of audio features and video features. According to Bayes' theorem, the classification result for the maximum posterior probability is

$$W^{*}=arg\max_{W} P(W|O^{a},O^{v}) = \frac{P\,(O^{a},O^{v}\mid W)\,P(W)}{P(O^{a},O^{v})} \tag{3}$$

where W denotes some word and O^{a} and O^{v}denote the vector sequence of audio and video features, respectively; if they are independent of each other, the joint output probability is $\mathrm{P}\,(O^{a},O^{v}) = \mathrm{P}\,(O^{a}\mid \mathrm{W})\,\mathrm{P}\,(O^{v}\mid \mathrm{W})$.

A multiple-model method needs to combine the audio and video streams in terms of formulas. For the improved product HMM method, we assume that the audio and video streams are conditionally independent, so observation vector and transition

probabilities at t time are

$$O^t{=}O_t^a \otimes O_t^v \tag{4}$$

$$a_{im,jn}{=}a_{ij}^a \times a_{mn}^v \tag{5}$$

where a_{ij}^a is the transition probability from state i to state j among the audio HMM and a_{mn}^v is the transition probability from state m to state n in the video HMM. The output probability of state ij is

$$P_{ij}\left(O^a, O^v \mid W\right){=}P_i(O^a|W^a)^\lambda P_j(O^v|W^v)^{1-\lambda} \tag{6}$$

The weight coefficient λ ($0 < \lambda < 1$) reflects the different weights of the two modes, which depend on the recognition performance of each mode under the different noise conditions. Zhao et al.'s experiments [16] showed the following linear relationship.

$$\lambda = 0.017 \cdot \mathrm{SNR} + 0.4 \tag{7}$$

The weight coefficient is greater ($\lambda > 0.825$) when the speech signal noise is smaller (SNR>25 dB), illustrating that audio information plays a larger role in the decision-making. The weight coefficient decreases with an increase of noise, illustrating that the proportion of video information increases gradually in the decision-making. When SNR = 5 dB, $\lambda \approx 0.5$, illustrating that they have the same importance at that moment.

5. Weight Optimization Based on the GPD Algorithm

Using the existing training data, based on formula (7), and calculating the weight coefficient according to the GDP algorithm [17], this training algorithm defines a misclassification distance that provides the correct distance between the class information and other information. The misclassification distance is computed using a smooth loss function and is minimized.

For a well-trained product HMM model, according to the N-best recognized hypotheses, supposing x as the unknown word vector, $L_c^{(x)}(\lambda)$ as the logarithmic likelihood values of correctly identifying the x in the model, $L_n^{(x)}(\lambda)$ as the logarithmic likelihood values of the N-best candidate vector of misrecognized words, the misclassification distance is

$$d^{(x)}\left(\lambda\right) = -L_c^{(x)}\left(\lambda\right) + \lg\{\frac{1}{N}\sum_{n=1}^{N}\exp[\eta L_n^{(x)}(\lambda)]\}^{\frac{1}{\eta}} \tag{8}$$

Where η is a smoothing parameter and N is the total candidate number. The total loss function of x after smoothing is

$$\mathrm{Lost}\left(\lambda\right) = \sum_{x=1}^{X}\frac{1}{1+\exp[-ad^{(x)}(\lambda)]}, a{>}0 \tag{9}$$

The purpose of training is to minimize Lost(λ) so as to minimize the error. The recursive formula of weight is

$$\lambda_{k+1}{=}\lambda_k{-}\varepsilon_k U_k \nabla \mathrm{Lost}\left(\lambda\right), k{=}\,1, 2, \ldots \tag{10}$$

The condition is $\varepsilon_k{>}0$, $\sum_{k=1}^{\infty}\varepsilon_k{=}\infty$ and $\sum_{k=1}^{\infty}\varepsilon_k^2{<}\infty$. $\{\mathrm{U}_k\}$ is a finite positive matrix sequence. The algorithm converges as k$\rightarrow\infty$. The recursion stops, and the final weight is obtained when the difference of the recursive value is smaller than a given threshold.

6. Experiment Results and Analysis

6.1 Experiment Dataset

We constructed a bimodal corpus and selected from seven people (five for male and two for female). The corpus contains 50 Chinese words, totaling 750 words for the seven people, including 550 words for training and the others for recognition. As needed, we added some noises of different intensity for recognition speech words. The sampling rate of the speech signal is 22.05 kHz. The quantitative value is 16 bits. The frame length of the speech frame is 28 ms. The frame shift is 14 ms, using a Hamming window as the window function. In order to ensure the synchronization of the video and audio streams after the extraction of video and audio features, we interpolated the video features and input these feature parameters into the improved product HMM, shown as in Figure 4. The video features were the lip parameter v_t and the dynamic parameter v_t, totaling 14 dimensions. To determine the final scheme of the audio features, we preselected three sets of features as follows:

(1) MFCC feature: MFCC (14 dimensions) + ΔMFCC (14 dimensions) + normalized audio energy, totaling 29 dimensions;

(2) LPCC feature: LPCC (14 dimensions) + ΔLPCC (14 dimensions) + normalized audio energy, totaling 29 dimensions;

(3) MFCC-LPCC joint feature: MFCC (14 dimensions) + ΔMFCC (14 dimensions) + LPCC (14 dimensions) + ΔLPCC (14 dimensions) + normalized audio energy, totaling 57 dimensions.

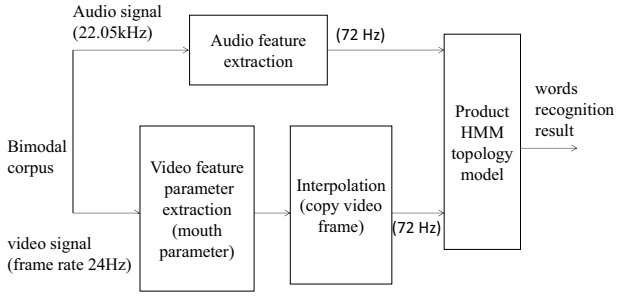

Figure 4. Proposed bimodal speech recognition process.

Table 1. Bimodal speech recognition rate of audio feature parameters for different SNR

SNR/dB	MFCC feature	LPCC feature	MFCC-LPCC joint feature
5	58.8	53.5	62.4
20	73.6	70.1	78.9
Clean	90.5	85.7	93.8

SNR, signal-to-noise ratio; MFCC, Mel-frequency cepstral coefficient; LPCC, linear predictive cepstrum coefficient.

6.2 Experiment Results

Under different SNR conditions, according to the recognition result (Table 1), we selected the MFCC-LPCC joint feature to train modal and recognize speech.

The convergence performance of the GPD algorithm depends on the choice of the feature parameters. After many experiments, we set $\mathrm{N} = 2$ in formula (8), $\alpha = 0.1$ in formula (9), $\epsilon_k = 50/\mathrm{k}$ in formula (10), and convergence threshold T_c= 0.01. The recursion stopped when the recursive interpolation $\nabla \mathrm{W}=\mathrm{W}_{k+1}-\mathrm{W}_k < \mathrm{T}_c$.

Four speech recognition schemes were proposed for comparing the bimodal speech recognition performance of the different methods.

(1) For the single-modal speech recognition of the audio, the single audio model adopts the classical left-to-right no-cross HMM modal [18]. Considering that the training objects are Chinese words, five or six states are selected and the output probability density function is a four-dimension mixed Gaussian density distribution. The audio parameters are the aforementioned 57 dimension MFCC-LPCC joint feature parameters.

(2) Based on the EI model [19], the joint feature vector is composed of the audio feature vector and video feature vectors, which are inputted to the single HMM model for training, with the same modal architecture as in solution (1).

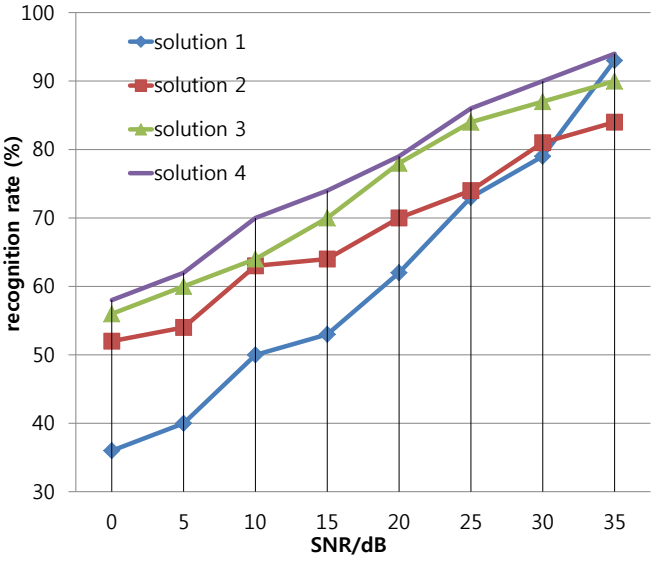

Figure 5. Speech recognition rate comparison of four solutions for different signal-to-noise ratios (SNRs).

(3) The multiple HMM model, mixing audio and video features on the state layer, assigning different weights to the audio and video streams according to the principle of formula (6) [5,6], requires special restrictions so as to maintain synchronization between the video and voice stream.

(4) This solution achieves word recognition through the improved product HMM model (Figure 3) proposed here. Based on the assumption that the audio and video streams are independent of each other, it allows for a step-state deviation between them.

To facilitate the performance comparison, the audio and video feature parameters of solutions (2) and (3) are the same as that of solution (4), and the recognition results are as shown in Figure 5.

6.3 Result Comparison and Analysis

In a low-noise environment, the difference in the recognition rate between the single-mode audio recognition method (solution 1) and the bimodal recognition method (solution 2-solution 4) is not large. When noise increases, the gap in recognition rate between single-mode and double-mode recognition will increase. This shows that in high-noise environments, the video information contributes more to the speech recognition rate.

For the bimodal recognition method, the EI model-based method did not assign the weights of video and audio information dynamically, leading to the lowest recognition rate for this method among all the bimodal recognition methods. The

solution 3 in [5] only considered the speech recognition of isolated words, while we consider recognizing Chinese phrases, increasing the recognition length. In contrast to our method, [5] used a neural network model, which was slower in actual speech recognition. Furthermore, [5] did not consider the asynchronous nature of the video and audio signals, but simply used the weighted fusion of the speech and video streams. Different from other methods, the method proposed in this paper considered the asynchrony, as shown in Figure 5, the speech recognition rate was slightly higher than in solution 3.

7. Conclusion

In this paper, with the aim of achieving effective speech recognition in noisy environments, a product HMM-based bimodal speech model allowing a one-step state offset to adapt to the asynchronous nature of the video signal and audio signal is proposed.

According to the corresponding relation between the weight coefficient and instantaneous SNR, the model can adjust the weight ratio of the audio and video streams adaptively. We selected a 50 Chinese 2-digit word corpuses as training and identification data, in contrast to other types of programs. The result showed that our proposed model can ensure the accuracy and robustness of speech recognition in a noisy environment.

Although we have shown effectiveness of the proposed bimodal HMM method on the Chinese 2-digit word recognition tasks, this scheme can be extended for multiword or continuous speech recognition tasks. In such cases, it would be a problem that, from the two modalities, we have unmanageably many possible word or phoneme sequence hypotheses to be considered for weighted integration. Also, more complicated interactions between the modalities can be modeled by using cross-modal associations and influences, where we still can use the proposed integration method for adaptive robustness. With these considerations, further investigation of applying the proposed system to complex tasks such as multiword or continuous speech recognition is in progress.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

This work was supported by three projects of the Shandong Province Higher Educational Science and Technology Program (J12LN09), China, Ji'nan Youth Science and Technology Star Project (No.20120104), China, and by a Natural Science Fundation Project of Shandong Province, China (No. zr2011fm028).

References

[1] B. V. Dasarathy, "Sensor fusion potential exploitation: innovative architectures and illustrative applications," *Proceedings of the IEEE,* vol. 85, no. 1, pp. 24-38, Jan. 1997. http://dx.doi.org/10.1109/5.554206

[2] D. W. Massaro, "Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry," *Hillsdale, NJ: Lawrence Erlbaum*, 1987.

[3] J. S. Lee and C. H. Park, "Training hidden Markov models by hybrid simulated annealing for visual speech recognition," in *Proceedings of 2006 IEEE International Conference on Systems, Man and Cybernetics*, Taipei, 2006, pp. 198-202, Oct. 2006.

[4] K. Kumatani, S. Nakamura, and K. Shikano, "An adaptive integration based on product HMM for audio-visual speech recognition," in *Proceedings of 2001 IEEE International Conference on Multimedia and Expo*, Tokyo, 2001, pp. 813-816. http://dx.doi.org/10.1109/ICME.2001.1237846

[5] J. S. Lee and C. H. Park. "Robust audio-visual speech recognition based on late integration," *IEEE Transactions on Multimedia*, vol. 10, no. 5, pp. 767-779, Aug. 2008. http://dx.doi.org/10.1109/TMM.2008.922789

[6] S. Dupont and J. Luettin, "Audio-visual speech modeling for continuous speech recognition," *IEEE Transactions on Multimedia*, vol. 2, no. 3, pp. 141-151, Sep. 2000. http://dx.doi.org/10.1109/6046.865479

[7] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 28, no. 4, pp. 357-366, Aug. 1980. http://dx.doi.org/10.1109/TASSP.1980.1163420

[8] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Processing,* vol. 10, no. 1, pp. 19-41, Jan. 2000.

[9] I. Peer , B. Rafaely, and Y. Zigel, "Reverberation matching for speaker recognition," in *Proceedings of 2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, Las Vegas, NV, 2008, pp. 4829-4832. http://dx.doi.org/10.1109/ICASSP.2008.4518738

[10] F. Bimbot, J. F. Bonastre, C. Fredouille, G. Gravier, I. Magrin-Chagnolleau, S. Meignier, T. Merlin, J. Ortega-Garcia, D. Petrovska-Delacretaz, and D. A. Reynolds, "A tutorial on text-independent speaker verification," *EURASIP Journal on Advances in Signal Processing*, vol. 2004, no. 4, pp. 430-451, Apr. 2004. http://dx.doi.org/10.1155/S1110865704310024

[11] C. Neti, G. Potamianos, J. Luettin, I. Matthews, H. Glotin, D. Vergyri, J. Sison, A. Mashari, and J. Zhou, "Audio visual speech recognition," in *Final Workshop 2000 Report, Center for Language and Speech Processing*, Baltimore, 2000.

[12] L. R. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*, Englewood Cliffs, NJ: PTR Prentice Hall, 1993.

[13] J. Luettin, G. Potamianos, and C. Neti, "Asynchronous stream modeling for large vocabulary audio-visual speech recognition," in *Proceedings of 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, UT, 2001, pp. 169-172. http://dx.doi.org/10.1109/ICASSP.2001.940794

[14] H. Zhao, C. Tang, and T. Yu, "Fast thresholding segmentation for image with high noise," in *Proceedings of 2008 International Conference on Information and Automation*, Changsha, 2008, pp. 290-295. http://dx.doi.org/10.1109/ICINFA.2008.4608013

[15] Lei Xie and D. Jiang, "Audio-visual synthesis and synchronous asynchronous experimental research for bimodal speech recognition," *Journal of Northwestern Polytechnical University*, vol. 22, no. 2, pp.171-175, 2004.

[16] H. Zhao, Y. Gu, and C. Tang, "Research of relationship between weight coefficient of product HMM and instantaneous SNR in bimodal speech recognition", *Journal of Computer Application*, vol. 29, pp. 279-285, 2009.

[17] A. Adjoudani and C. Benot, "On the integration of auditory and visual parameters in an HMM-based ASR," in *Proceedings NATO ASI Conference on Speechreading by Man and Machine: Models, Systems and Applications*, 1995, pp. 461-471.

[18] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE,* vol. 77, no. 2, pp. 257-286, Feb. 1989. http://dx.doi.org/10.1109/5.18626

[19] C. Bregler and S. M. Omohundro, "Nonlinear manifold learning for visual speech recognition," in *Proceedings of 1995 5th International Conference on Computer Vision*, Cambridge, MA, 1995, pp. 494-499. http://dx.doi.org/10.1109/ICCV.1995.466899

Su mei Xi

Su Mei Xi is a Ph.D. candidate in the Department of Computer Science at the University of Suwon in Korea, and a lecturer in the School of Science at Qilu University of Technology in China (inservice personnel pursuing doctoral degree). Her research interests include artificial intelligence, information retrieval, and multimedia processing. She received her M.S. degree in Computer Science from Shandong University in China in 2009, and her B.S. degree in Computer Science from Shandong University of Science and Technology in China in 2001. Her research areas include intelligent system, ubiquitous system, information retrieval, multimedia processing, fuzzy system, etc.

E-mail: xsm@suwon.ac.kr

Young Im Cho

Young Im Cho is an associate professor in the Department of Computer Science at the University of Suwon in Korea. Her research interests area are artificial intelligence, pattern recognition, information retrieval, ubiquitous system etc. She received her B.S. degree in Computer Science from Korea University in 1988, her M.S. degree in Computer Science from Korea University in 1990, and her Ph.D. degree in Computer Science from Korea University in 1994. She received post-doc. degree at the university of Massachusetts in USA in 2000. She worked at Samsung electronics company at 1995. Her research areas include intelligent system, ubiquitous system, information retrieval, neural network, fuzzy system, etc.

E-mail: ycho@suwn.ac.kr