International Journal of Fuzzy Logic and Intelligent Systems, vol. 8, no. 2, June 2008 pp. 105-110

Emotion Recognition Method Based on Multimodal Sensor Fusion Algorithm

Byung-Hyun Moon and Kwee-Bo Sim*

School of Electrical and Electronics Engineering, Chung-Ang University
221, Heukseok-Dong, Dongjak-Gu, Seoul, 156-756, Korea

Abstract

Human being recognizes emotion fusing information of the other speech signal, expression, gesture and bio-signal. Computer needs technologies that being recognized as human do using combined information. In this paper, we recognized five emotions (normal, happiness, anger, surprise, sadness) through speech signal and facial image, and we propose to method that fusing into emotion for emotion recognition result is applying to multimodal method. Speech signal and facial image does emotion recognition using Principal Component Analysis (PCA) method. And multimodal is fusing into emotion result applying fuzzy membership function. With our experiments, our average emotion recognition rate was 63% by using speech signals, and was 53.4% by using facial images. That is, we know that speech signal offers a better emotion recognition rate than the facial image. We proposed decision fusion method using S-type membership function to heighten the emotion recognition rate. Result of emotion recognition through proposed method, average recognized rate is 70.4%. We could know that decision fusion method offers a better emotion recognition rate than the facial image or speech signal.

Key Words : Recognition, Principal Component Analysis (PCA), Multimodal, Decision Fusion Method

1. Introduction

Today, various types of human-like machines that coexist with humans in everyday life are gradually increasing. Specially robot be used to industry achieving one's role in home. Humans, however, have the ability to have many emotional interchanges between each other as human relationships tend to be very emotional. So, if human-like machines are to be effective, these human-like machines need the ability to recognize emotions. Existing emotion recognition research has involved speech, facial images, gestures, even brain waves, and other related aspects.

Here is a summary of some of this research about emotion recognition that use speech signals. Chang-Hyun Park and Kwee-Bo Sim studied applying acoustic features (statistic of pitch, size of speech, section number, increasing Rate (IR), crossing Rate (CR)) by using Artificial Neural Network to recognize emotion [1], Weon-Goo Kim studied ways to classify emotion patterns using MFCC parameter features that be extracted in a speech signal and extracted features by applying KNN [2].

We now describe some research efforts dealing with emotion recognition that used facial image. Yong-Suk Shin experimented with emotion recognition, Shin applied a feature vector extract for the mouth, eye and eyebrow in the facial region and extracted these various features by applying PCA and Neural Network [3]. Su-Jeong Han, Myung-Geun Chun, and others experimented by applying an ICA-factorial representation method to the whole facial image. They classified a face by 6 basic expressions (happiness, sadness, anger, surprise, fear and dislike) [4]. However, there is limited research using one of humans' Bio-signals where humans recognize emotion within life. A large majority of people can recognize emotions by integrating various Bio-signals of people they deal with. Therefore, we, also, would like to have human-like machines have functions that can recognize emotions integrating various Bio-signals.

Recently, there has been some research in this area, Mingli Song has conducted some experiments involving emotion recognition from speech and facial images using Hidden Markove Model(HMM) [5]. Hatice Gunces conducted experiments on emotion recognition from a facial image and gestures using decision fusion method and feature fusion method [6]. Also De Silva performed an experiment of emotion recognition using a Fuzzy-rule based speech and facial image [7]. Hyeun-Joo Go, Dae-Jong Lee et al. attempted emotion recognition from a facial image and speech signal by using LDA and a wavelet transform, and a recognized value. He then applied a Fuzzy member function and did an emotion fusion [8].

For our experiment, we showed that rate of recognition is high for emotion recognition using of integrated various Bio-signals than using of only Bio-signal. In this paper, we explain a method of extracting features from speech and facial images in Section. We explain about an emotion recognition experiment for speech signal and facial image where we apply Principal

Manuscript received May. 17, 2008; revised Jun. 5, 2008.
* Corresponding author
This research was supported by the industry-university-research consortium program from the Seoul city and Small & Medium Business Administration(SMBA) of Korea.

Component Analysis for extracted characters in Section 3-4. As part of emotion fusion methods, we discuss, in Section 5, the decision fusion method. We present concluding remarks in Section 6.

2. Feature Extraction

2.1 Speech Signal

Features of an extracted speech signal include the statistics of pitch, the dimensions of sound, the selection number, etc. In this research, we use a pitch extract method as the most general autocorrelation approach. We extracted a pitch value every 0.1 seconds and we calculated the average of the values defined by pitch mean and variance value was acquired the equal data. We obtained the size of the sound by a magnitude estimation method. We obtained the selection number by looking for a concave extreme point of energy after finding a search starting point and ending point of a sentence from pitch contour. Divide section via smaller point than maximum of waveform/4 in the concave extreme point. We used a method in the Kwee-Bo Sim et al.'s paper [9] for the increasing rate(IR) and crossing rate(CR).

2.2 Facial Image

To form a feature vector, first, the face area must be detected in a color image eliminating the background image from the total image. Second, we detect lips, eyes and eyebrows by applying a skin ton accumulation algorithm to the face dimension with this method. Third, we then use a converter after to GRAY image, form feature vector. A detailed explanation of such a method is in Kwee-Bo Sim et al.'s paper [10].

3. An Emotion Recognition using Speech Signal

In this paper, we make a experiment of emotion recognition through Principal Component Analysis, Fig.1 shows training process. When size of data is $N \times N$ and the number of recognition candidate data is M, step1 composes recognition candidate data vector (Γ_m) of set(S) of recognition candidate by row vector ($N^2 \times 1$). Step 2 do normalize data via preset average and dispersion to reduce error. Step 3 calculates average data vector(Ψ) from data vector set of recognition candidates. Step4 calculated difference(Φ_i) of recognition candidates' data vector and average data vector. Step 5 calculate covariance procession (C) from Φ_i of $N^2 \times 1$ vector. Step 6 calculates eigenvalue (λ_i) and eigenvector (μ_i) by covariance procession. This paper we composed of extracted feature vector by average of pitch, size of speech, number of section, increasing rate (IR) and crossing rate(CR) about speech sample. We composed of recognition candidate data vector set after we line up feature vectors by 5×1 row vector. We project user's speech signal in eigenvector that is acquired through PCA training process.

Through such experiment, we measured distance between user's speech signal data vector and training data vectors. Because of most similar to user's speech signal, the speech signal that distance being the smallest is decided for emotion recognition result and we know the result of test by Fig. 2. Fig. 3 shows an emotion recognition System for speech signal.

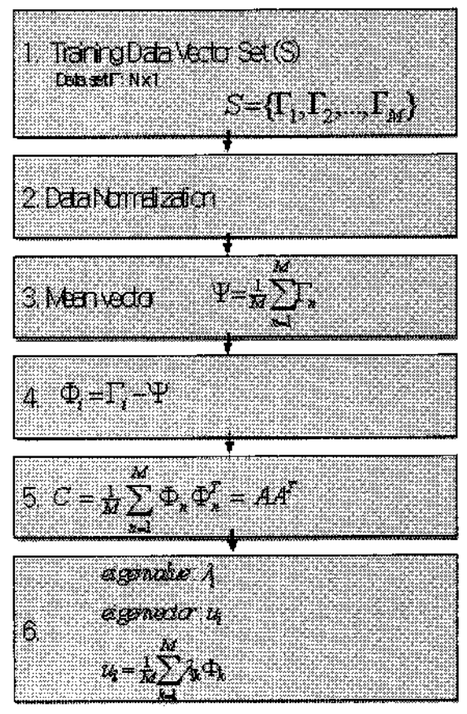


Fig. 1. Training process using the PCA

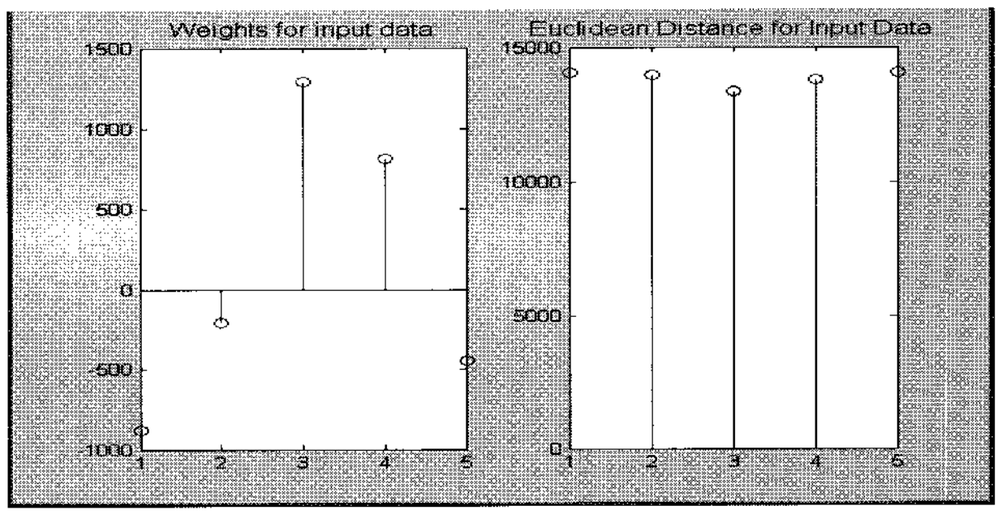


Fig. 2. PCA result of speech signal

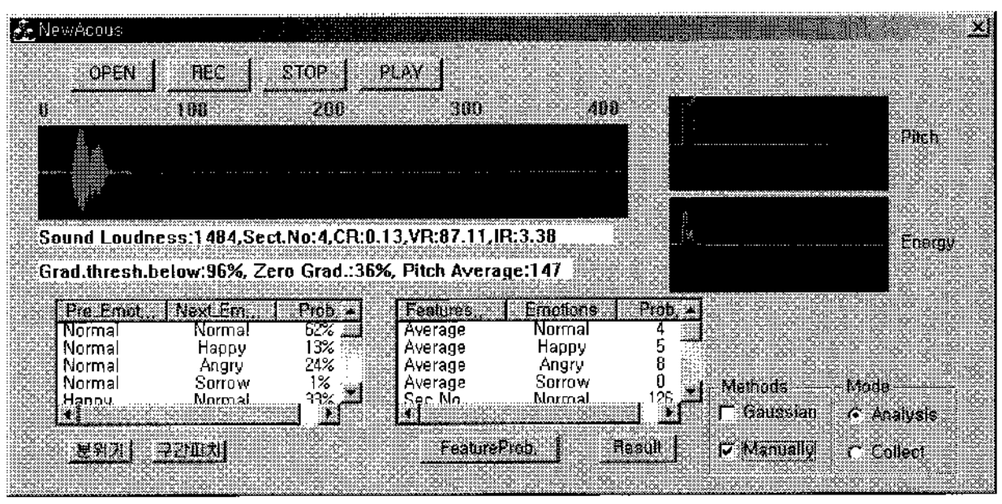


Fig. 3. An emotion recognition system for speech

We make an experiment of emotion recognition with each particularly emotion 50 times using this constructed system. Fig. 4 shows the experiment's result.

Through the experiment result, we know that the recognition rate of other emotions is better than normal, because an

emotional factor is insufficient in feature vectors. The average emotion recognition rate of this experiment was 63%. This result is not behind the other studied recognize rate.

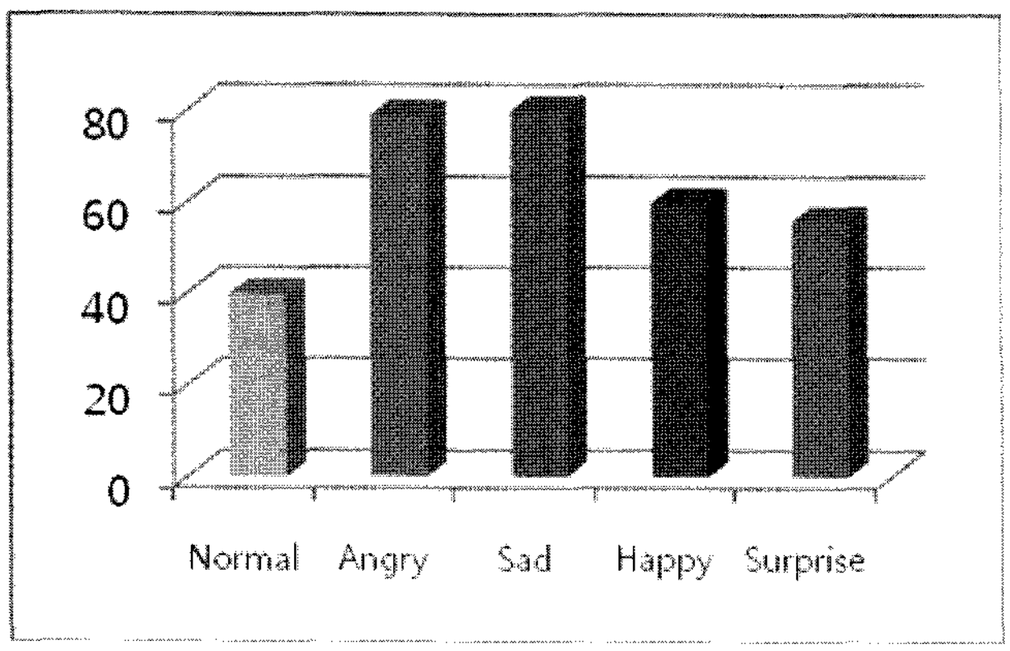


Fig. 4. The emotion recognition rate of speech signal

4. An Emotion Recognition using Facial Image

Facial image is data that consist of multidimensional feature vectors. PCA is most representative method that provides the needed multivariate data process method. PCA can be used to reduce the data to a low dimension while keeping information of a high dimension to recognize the multidimensional feature vector as well as other properties. PCA gets a principal axis by statistical method and it is method to reduce dimension that do to project feature vector for principal axis direction.

We took picture showing five expression (normal, happy, anger, surprise and sad) of five men (graduate students from 25~31 years of age). Fig. 5 displays some of these collected pictures.

Fig. 5. Representative face image of the various emotions used for the experiments

If a new facial picture is the input, we, then, take a private projection for the eigenface and from this, we find a component value. From these values for new facial pictures, we compare the Euclidean distance with the weight in the eigenface of candidate face images.

The face image where the distance amounts to being the smallest decides the emotion recognition result and we know the result of test by Fig. 6.

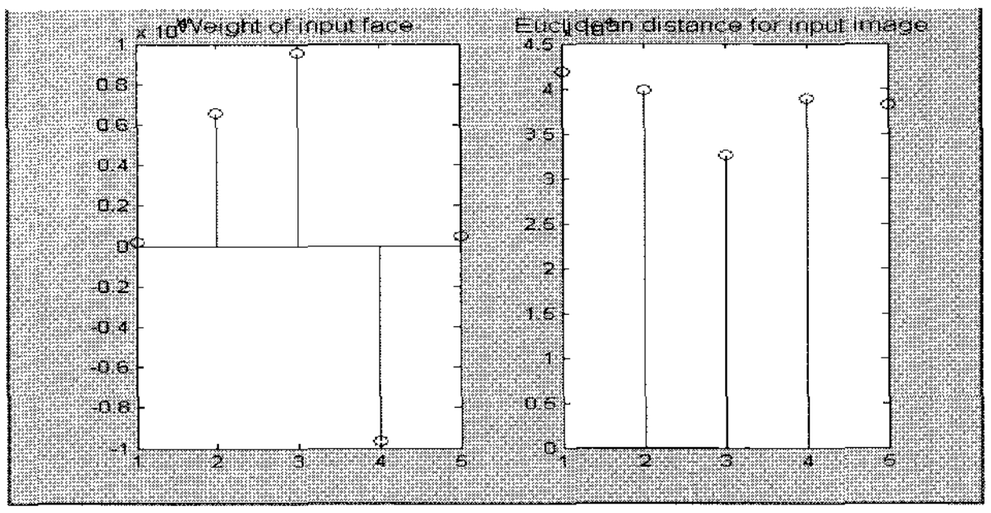


Fig. 6. PCA Result of facial image

In our research described in this paper, we constructed such series processes by using the system as illustrated in Fig. 7.

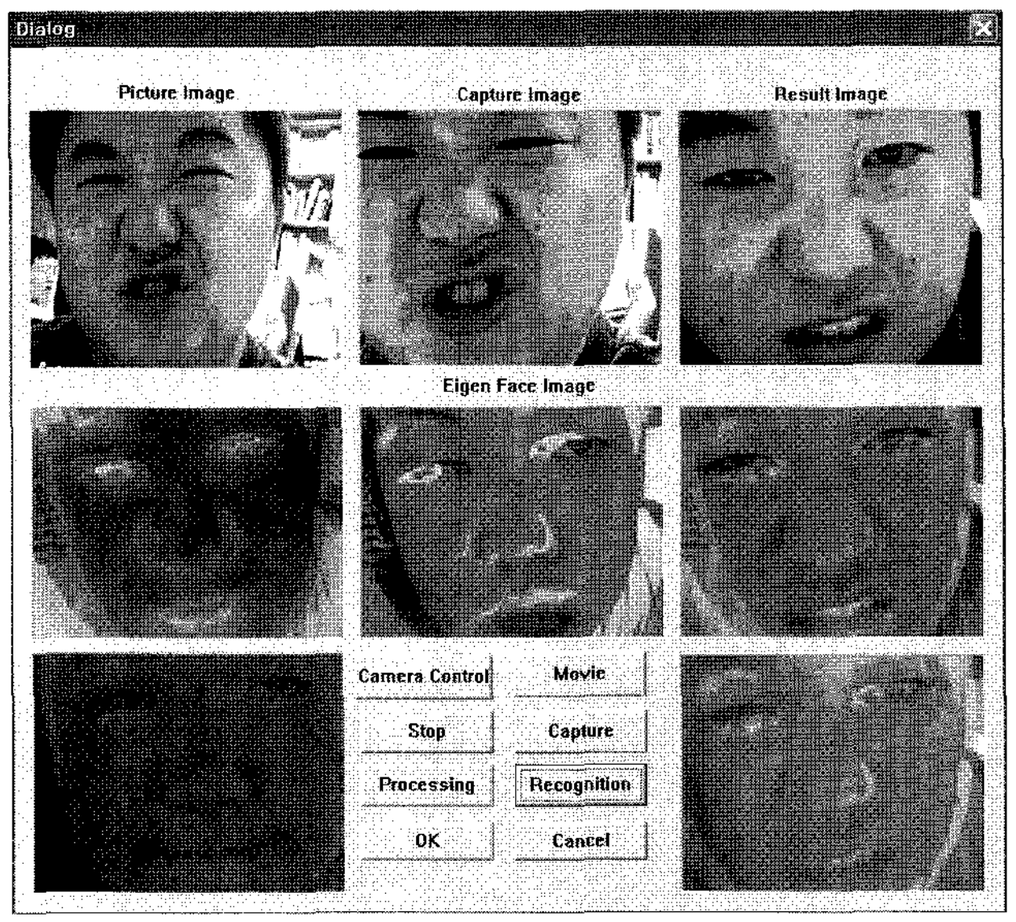


Fig. 7. Face image emotion recognition system

We experimented on emotion recognition with each particularly emotion 50 times using this constructed system. Fig. 8 shows the experiment's result.

We know that the experiment result is the recognition rate of surprise and anger rather than sadness, because do not look for well difference with normal of expression and method to better the emotion recognition rate is emotion that expression change use reliable facial image data to learning data particularly.

The average emotion recognition rate of this experiment was 53.4%. This result was not better when comparing it with the recognition rate of another paper [11].

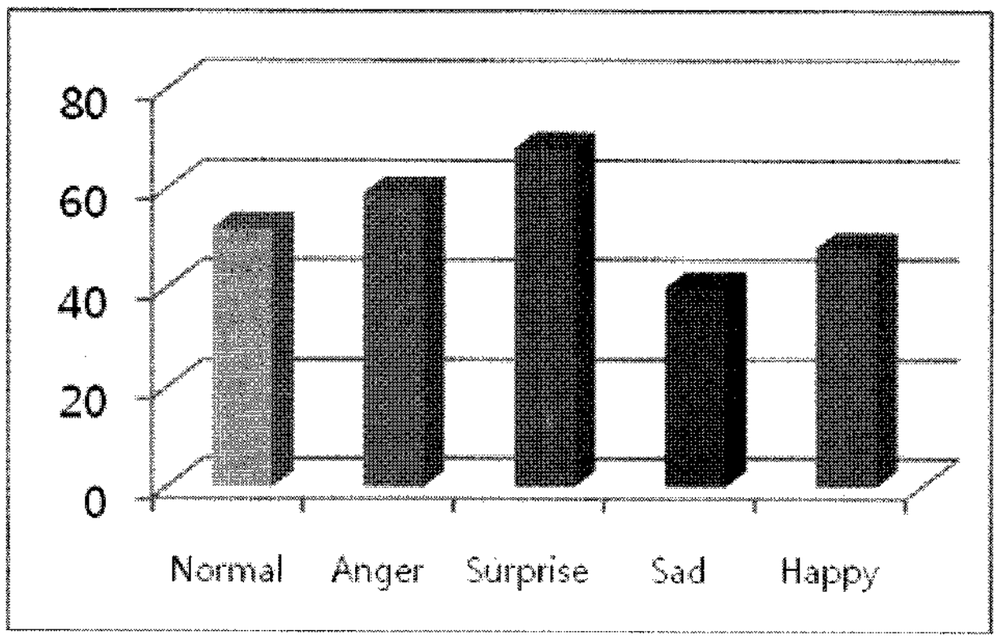


Fig. 8. PCA based on emotion recognition rate

5. An Emotion Recognition using Multimodal Fusion Method

In general, the multimodal sensor fusion method can be divided by decision fusion method and feature fusion method [12]. The decision fusion method fuses the recognized result value for each emotion recognition system. This method has the advantage that is easy to fuse different Bio-Action signals. Fig. 9 is a high level flow chart of this decision method.

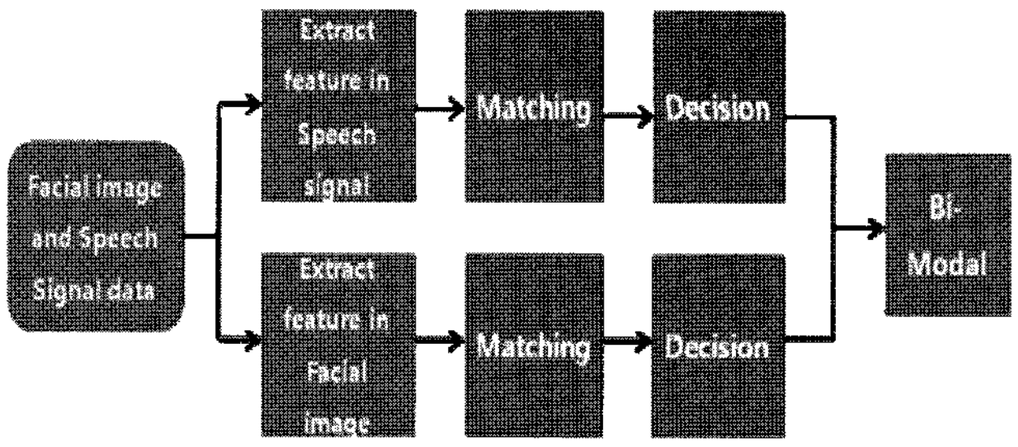


Fig. 10. The feature fusion method

The feature fusion method fuses extracted feature vectors for each of the Bio-Action signals and then this method constructs a new feature vector. It needs a measurement method for good performance similarity for emotion recognition. Fig. 10 shows a high level flow chart of the feature fusion method.

In our research, we used a decision fusion method. This method is applied various algorithms. But in this paper, we used method such as the fuzzy member function for fusion.

Weights were assigned representing the recognized emotion state with each speech signal and facial image. Each of the emotion states and weights represent as follows:

$$\text{Image weight} = w_{i(normal)}, w_{i(happy)}, w_{i(surprise)}, w_{i(sad)}, w_{i(anger)}$$

$$\text{Speech weight} = w_{s(normal)}, w_{s(happy)}, w_{s(surprise)}, w_{s(sad)}, w_{s(anger)}$$

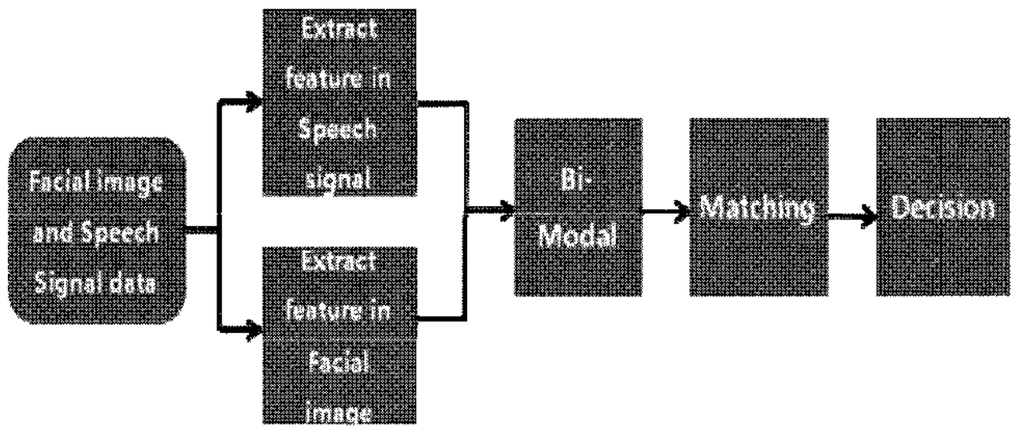


Fig. 10. The feature fusion method

To calculate these weights, in our research, we recognized emotions through the training data of facial images, speech signals and we calculated average of emotion values that being comprised it

We extract feature vector through the speech signal and facial image and we recognize emotion. We select the same emotion state to recognized emotion state for learning data.

About selected emotion values, Member degree of emotion values that is recognized calculate for the most general S-type member function and such series process represented as figural by Fig. 11.

Equation (2) represents the output of the actual system we researched. I is the emotion output of the facial image and S is the emotion output of the speech signal.

The speech signal, the facial image and weight calculate multiplication, then the output appears about the emotion state.

There are two methods that display the latest emotion state by using this result value.

The first method selects a maximum value for the result value about the same emotion. The second method displays the value by adding the result value about the same emotion. Finally, we proposed that we use this method in a prior case.

$$\begin{aligned} O_{narnal} &= w_{i(nornal)} I_{narnal} + w_{s(nornal)} S_{normal} \\ O_{happy} &= w_{i(happy)} I_{happy} + w_{s(happy)} S_{happy} \\ O_{surprise} &= w_{i(surprise)} I_{surprise} + w_{s(surprise)} S_{surprise} \\ O_{sad} &= w_{i(sad)} I_{(sad)} + w_{s(sad)} S_{sad} \\ O_{anger} &= w_{i(anger)} I_{anger} + w_{s(anger)} S_{anger} \end{aligned} \quad (1)$$

In conclusion we get the emotions state that desired by a selection of the maximum among each output value of emotion state. It is as in Equation (2).

$$\begin{aligned} &System\ Output = \\ &\quad Max\{O_{normal}, O_{happy}, O_{surprise}, O_{sad}, O_{anger}\} \end{aligned} \quad (2)$$

Fig. 12 shows an emotion recognition rate using the decision fusion method. The average emotion recognition rate of this experiment was 70.4%. We can get better one than result using of speech signal and facial image.

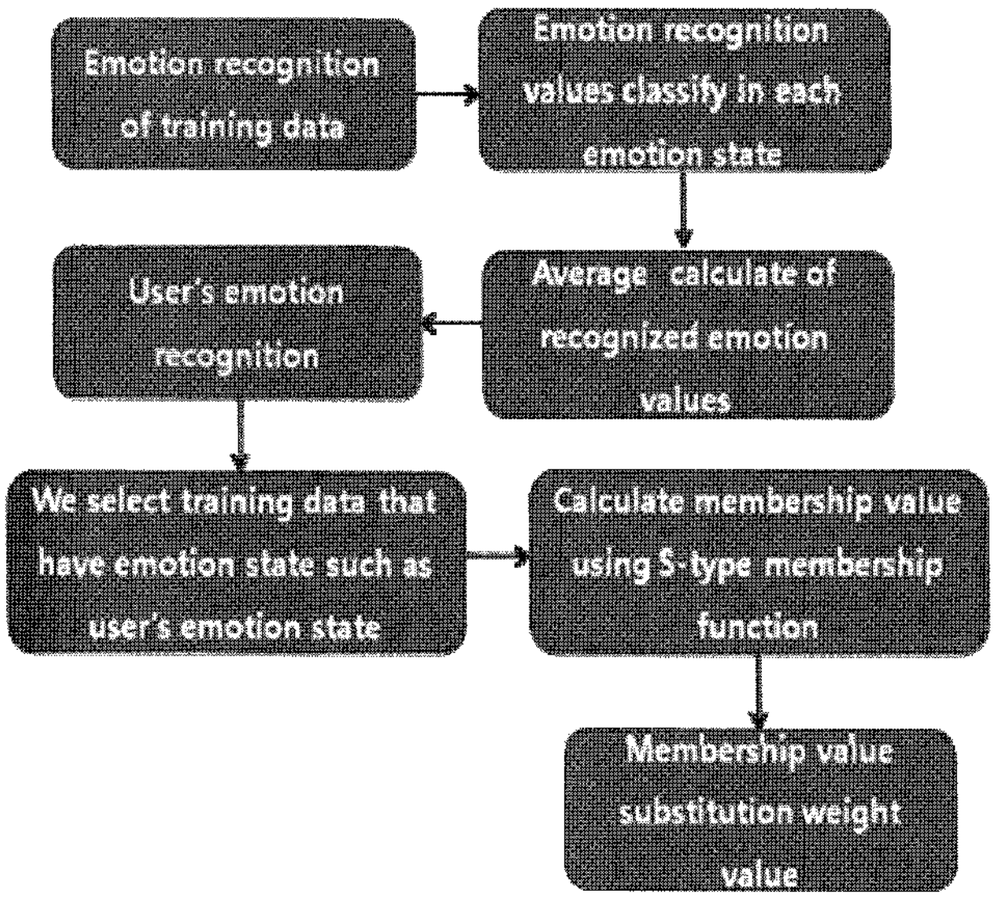


Fig. 11. The weight calculate method

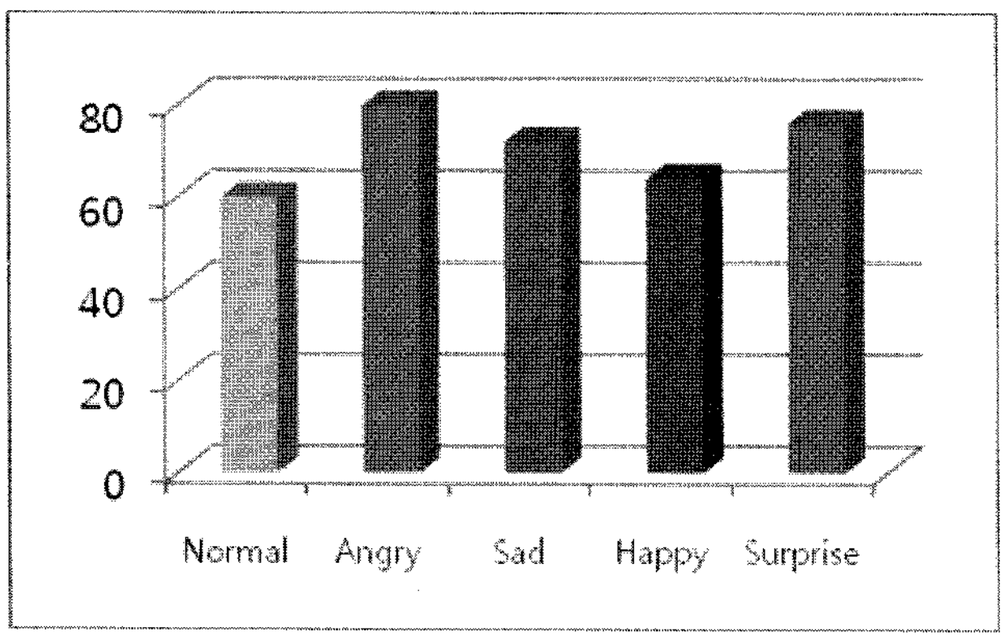


Fig. 12. Multimodal based on emotion recognition rate

6. Conclusion

In this paper we extracted to feature that used to a speech signal and facial image. Applying Principal Component Analysis for Extracted feature, we make an experiment on emotion recognition about five emotions (Normal, happy, sad, anger, surprise). We propose to method using a fusion method that involving a decision fusion method using an S-type member function. In our experiment, the average emotion recognition rate was 63% using a speech signal and the average emotion recognition rate was 53.4% using a facial image. That is, we know that a speech signal provides a better emotion recognition rate than does a facial image. We proposed decision fusion method using S-type membership function to heighten the emotion recognition rate. The recognized rate of average through proposed method was 70.4%. We can know that emotion recognizing case using of various bio-signals fusion get better than using of a bio-signal. After research, we are going to propose feature fusion method and make an experiment. And we are going to propose superior emotion fusion method after compare it with decision fusion method.

Reference

[1] Chang-Hyun Park, Kwee-Bo Sim "The Pattern Recognition Methods for Emotion Recognition with Speech Signal", *Journal of Control, Automation, and Systems Engineering*, vol. 12, no. 3, pp. 284-288, Mar. 2006.

[2] Weon-Goo Kim "Comparison of feature parameters for emotion recognition using speech signal", *Journal of Institute of Electronics Engineers of Korea(IEEK)*, vol. 40, no. 5, pp. 371 - 377, 2003.

[3] Young-Suk Shin "The Effect of Facial Expression Recognition Based on the Dimensions of Emotion using PCA Representation and Neural Networks," *Lecture Notes in Computer Science(LNCS)*, Vol. 3656, pp. 1133-1140, 2005.

[4] Su-Jeong Han, Keun-Chang Kwak, Hyoun-Joo Go, Sung-Suk Kim, Myung-Geun Chun " Facial Expression Recognition using ICA-Factorial Representation Method", *Journal of Korea Fuzzy Logic and Intelligent Systems Society(KFIS)*, vol.13, no.3, pp. 371-376, 2003.

[5] Mingli Song. Jiajun Bu, Chun Chen and Nan Li, "Audio-Visual based Emotion Recognition," *Proc. of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04)*, Vol. 2, pp. 1020-1125, 2004.

[6] Hatice Guncs and Massimo Piccardi, "Bi-Modal Emotion Recognitionfrom Expressive Face and Boby Gessture" *Journal of Network and Computer Applications*, pp. 1~12, 2006.

[7] J De Silval, Pei chi Ng., "Bimodal Emotion Recognition", *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, pp 332~335, 2000.

[8] Hyeun-Joo Go, Dae-Jong Lee, Myung-Geun Chun " An Emotion Recognition Method using Facial Expression and Speech Signal", *Journal of Korea Information Science Society(KISS)*, vol. 31, no. 6, pp. 799-807, 2004

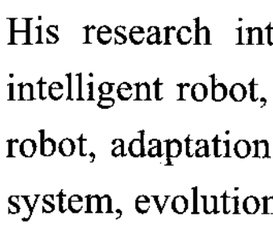
[9] Kwee-Bo Sim, Chang-Hyun Park, "Analyzing the element of emotion recognition from speech", *Journal of Korea Fuzzy Logic and Intelligent Systems Society(KFIS)*, vol. 11, no. 6, pp.510 - 515, 2001. 12.

[10] Kwee-Bo Sim and Dong-Hun Lee, "Emotion Recognition and Expression System of Mobile Robot Based on 2D Facial Image", *Journal of Control, Automation, and Systems Engineering*, vol. 13, no. 4, pp. 371-376, Apr. 2007.

[11] Ho-Duck Kim, Hyun-Chang Yang, Chang-Hyun Park, and Kwee-Bo Sim, "Emotion Recognition Method of Facial Image using PCA ", *Journal of Korea Fuzzy Logic and Intelligent Systems Society(KFIS)*, vol.16, no.6, pp. 772-776, Dec. 2006.

[12] Myung-Hun Kim, Chi-Geun Lee, In-Mi So, Sung-Tae Jung "Design and Implementation of a Bio-modal user Recognition

System using Face and Audio," *Journal of Korea Society of Computer and information(KSCI)*, vol. 10, no. 5, pp. 353-361, 2005.



Byung-Hyun Moon

He received his B.S degree in the Department of Electrical and Electronics Engineering from Chung-Ang University, Seoul, Korea, in 2007. He is currently Master course in the School of Electrical and Electronics Engineering from Chung-Ang University. His research interests include intelligent System, ubiquitous intelligent robot, emotion recognition and expression system of robot, adaptation and learning algorithm, neural network, fuzzy system, evolutionary computation, etc.

Kwee-Bo Sim

He received his B.S. and M.S. degrees in the Department of Electronic Engineering from Chung-Ang University, Korea, in 1984 and 1986 respectively, and Ph.D. degree in the Department of Electronics Engineering from The University of Tokyo, Japan, in 1990. Since 1991, he is currently a Professor. His research interests include intelligent System, ubiquitous intelligent robot, emotion recognition and expression system of robot, intelligent surveillance system, intelligent home, ubiquitous computing and sensor network, adaptation and learning algorithm, neural network, fuzzy system, evolutionary computation, multi-agent and distributed autonomous robot system, evolvable hardware and embedded system, etc. He is a member of IEEE, SICE, RSJ, KITE, KIEE, KIIS, and ICROS Fellow.