

Affective Information Processing and Recognizing Human Emotion

Fuji Ren ¹

*Department of Information Science and Intelligent Systems
The University of Tokushima
Tokushima, Japan*

*School of Information Engineering
Beijing University of Posts and Telecommunications
Beijing, China*

Abstract

Information recognition and extraction of human emotions are necessary for machines to communicate smoothly with humans and to realize emotion communications. We focus on human psychological characteristics to develop general-purpose agents that can recognize human emotion and create machine emotion. We comprehensively analyze brain waves, voice sounds and picture images that represent information included in emotion elements of phonation, facial expressions, and speech usage. We analyze and estimate many statistical data based on the latest achievements of brain science and psychology in order to derive transition networks for human psychological states. We establish a speaker word model for researching computer simulation of psychological change and emotional presentation, developing emotion interface, and establishing theoretic structure and realization method of emotion communication. A new approach for recognizing human emotion based on Mental State Transition Network will be described and one emotion estimation method based on sentence pattern of emotion occurrence events will be discussed, and some new results of the project will be given.

Keywords: Affective Information Processing, Recognizing Human Emotion, Mental State Transition Network

1 Introduction

Modern information communication mainly focuses on verbal information and to a lesser degree deals with the human emotions accompanying the message (non-verbal information). Both research and business are developing in the field of human interface technology which covers voice recognition, voice synthesis and virtual reality. However, there are many problems in affective information processing and recognizing human emotion accurately. This makes many people still feel strong resistance toward interacting with machines in business fields of terminal devices (mobile phones and car navigation systems) and medical care systems. Semantic

¹ Email: ren@is.tokushima-u.ac.jp

recognition of natural language, which is essential for a machine to communicate with a human, is not efficient enough. It is also indispensable for a machine to recognize and extract the information in human emotions (acts, facial expressions and sensibility) to actualize emotion communication without any sense of unease for the human [1],[3],[4].

Our research focuses on human mental features and aims to develop an emotion measurement model for a speaker and emotion simulation model for a computer which work as a multi-purpose agent recognizing human emotions and creating artificial emotions. To be more precise, we analyze information contained in the brain wave [36],[37], sound voice, visual image and speech pattern from the perspective of mental features [34]. We also analyze large statistical data based on the latest result of neurology and psychology in order to derive a mental state transition network. By constructing and using a word model for a speaker, we study how to simulate a change of mental state or emotional demonstrations by a computer. Our study aims to develop an emotion interface and establish a theoretical system and a method for future emotion communication.

In this paper, a new approach for recognizing human emotion based on Mental State Transition Network (MSTN) [33], [35] will be described and an emotion estimation method based on sentence pattern of emotion occurrence events will be discussed. Lastly, some new results for the project will be given.

2 Model for human emotion recognition

Traditional behavioral psychology has dealt with the recognition of human emotions, however, owing to the lack of engineering methods a significant achievement has not yet been made. On the other hand, traditional artificial intelligence has already been working on information retrieval or an inference method on the basis of human psyche and culture, however, it has not yet realized to simulate human emotions by computers.

However, these traditional studies have not been focused on the depth of affectability and have not yet succeeded in recognizing human emotion.

We present a model to recognize human emotion. The basic idea of our approach is that the MSTN should be employed to recognize human emotion using speech patterns, phonetic information and face features. Figure 1 shows the structure of affective interface [2].

The model has two parts, one is the Human Emotion Recognition Engine (HMRE) and the other is the Machine Emotion Creation Engine (MECE). The HMRE consists of linguistic information based emotion recognition module, phonetic information based emotion recognition module, and expressive information based emotion recognition module. A corpus, an ontology and an individualization DB are also employed to recognize human emotions. The MECE has three processes: sensibility language process, sensibility sound voice process, and emotion face process. A Mental State Transition Network used in both Engines and a psychological questionnaire experiment will be described in next section.

3 Mental state transition network

Affective communication is an important theme for developing future-generation communication systems and the methodologies for constructing an emotion interface should be established first. External information such as language and facial expressions are not enough to model human emotion, so it is necessary to combine them with more physical reactions.

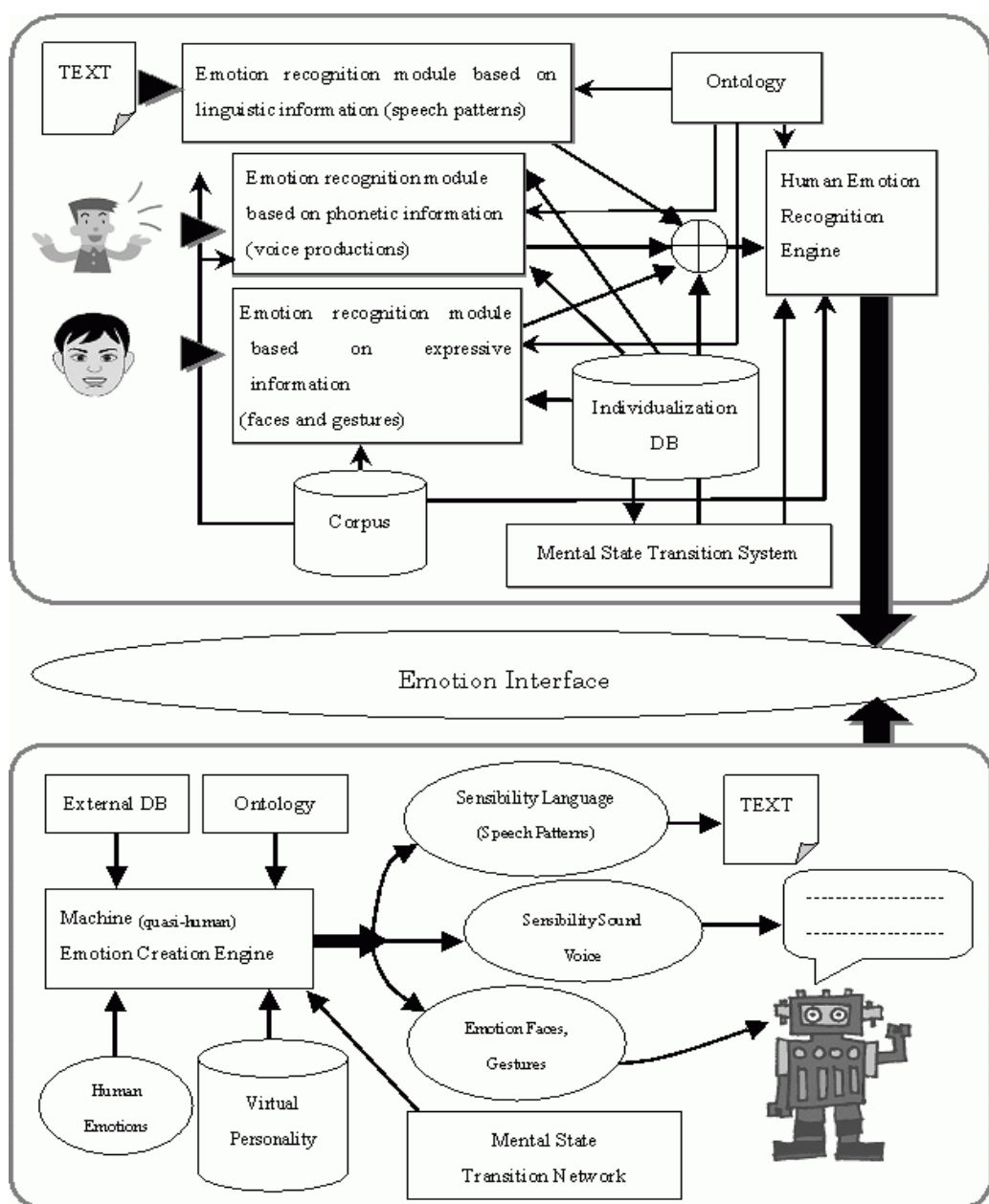


Fig. 1. The model of recognizing human emotion and creating machine emotion

We started a project to measure human emotions combining language, voice, facial expression and brain wave information. Our latest results of various experiments do not show striking achievements, however, we had an idea to approximate human physical reactions based on a particular transition network which deduces human emotions by a black box before the human brain mechanism is found out.

We hypothecate that human emotions are placed on several states and they transit between several discrete states. Here we call these states as “mental state”. Human mental states can transit from one state to another state on a certain condition. Frequencies of transition among these states are not the same, however, there exists a certain expectation value without considering external causes. Analysis of large data and human personality information allow building the following network module representing mental state transitions as Fig.2 shows. In Fig.2, 0 means Serene, 1 means Happy, 2 means Sad, 3 means Angry, 4 means Disgust, 5 means Fear and 6 means Surprise.

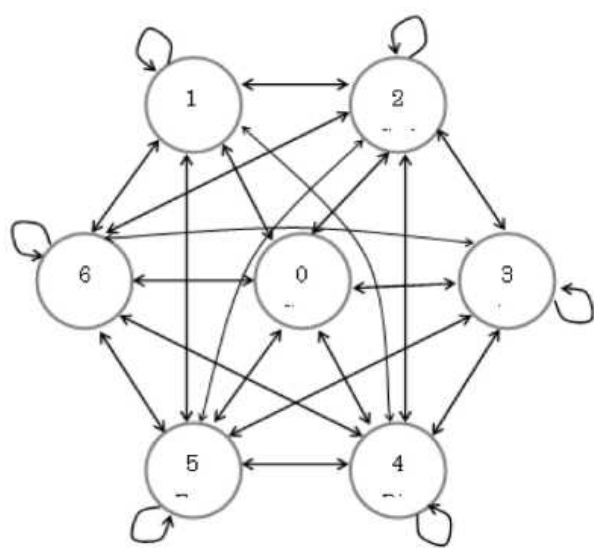


Fig. 2. Concept of Mental State Transition Network

We obtained the probability distribution in our model through an experiment using a psychological questionnaire. In the experiment, we had about 200 participants recruited primarily from different high schools and universities in China and Japan respectively. The psychological experiment required participants to fill out a table which was designed for creating transitions among seven emotional (mental) states.

Table 1 shows the MSTN model introduced from the psychological experiment.

Table 1
Mental State Transition Network

	1	2	3	4	5	6	7
1	0.421	0.213	0.084	0.190	0.056	0.050	0.047
2	0.362	0.509	0.296	0.264	0.262	0.244	0.252
3	0.061	0.090	0.320	0.091	0.123	0.137	0.092
4	0.060	0.055	0.058	0.243	0.075	0.101	0.056
5	0.027	0.039	0.108	0.086	0.293	0.096	0.164
6	0.034	0.051	0.064	0.076	0.069	0.279	0.075
7	0.032	0.042	0.068	0.048	0.121	0.092	0.313

4 Emotion energy acquired from phonetic information

We call an appearance information an emotion energy. At present, the emotion energies are acquired from three parts, the linguistic information, the phonetic information and the expressive information. This section describes a method for acquiring emotion energy from phonetic information.

Human emotions such as anger, sorrow, joy, laughter or excitement are considered to reside in intonations or rhythms in voice. It is important for human emotion recognition to use voice information. In traditional studies which attempted to recognize emotions from voice information, voice recognition Hidden Markov Model (HMM) are commonly used for emotion recognition. As the virtual personality on computer should respond to the users naturally, duration, speed, pausing, amplitude and basic frequency of speech have been analyzed for recent emotion recognition studies. Neural net or tones are also utilized for emotion analysis. These studies, however, are facing difficulties in recognizing emotional ups and downs or emotional transitions, because general emotion analysis based on voice recognition cannot recognize the expressions from the intonation. Although there is a new approach based on phoneme, it is still difficult to recognize emotion using a neural net, which cannot trace the judgment and lacks reproduction ability and accuracy.

Our research aims to construct an emotion interface, which requires construction of the models for human emotion recognition and artificial emotion creation. As mentioned before, we comprehensively analyze information contained in the brain wave, sound voice, visual image and speech patterns from the perspective of mental features. We also analyze and presume large statistic data based on the latest result of neuroscience and psychology in order to derive mental state transition network.

As for extracting emotion elements from speaker's voice, we did not use spectrum learning by using a neural network or a HMM, instead, we used a mental state transition network recognizing the human emotion based on its intonation or transition amounts and by obtaining the information about its mental state transition. We conducted several experiments about optional voice, emotion analysis

for music, emotion spectrum and emotional recognition from a series of emotional dialogues, and we also detected information about emotion transition from a series of dialogues. By those experiments we examined and modified the adjustability of the measurement method comparing the actual human emotion transition.

5 Emotion energy acquired from expressive information

The expressive information in this paper indicates facial expressions. A prototype system for identifying facial expressions by using facial features is presented. The system recognizes 7 facial expressions. The 7 facial expressions are made up of 6 basic emotional expressions (happiness, sadness, surprise, fear, anger, and dislike) and one non-expression. The Facial Action Coding System (FACS) is used to make the resulting system robust. For identification, the shortest distance between the input features and features stored in a dictionary is used. From these facial expressions a user’s intention can be extrapolated and used to improve the human computer interaction experience [7],[8],[9],[10].

5.1 Facial Action Coding System (FACS)

Facial expressions play an important role in communication by relaying non-verbal information about the physical and mental state of the speaker [13],[14],[15],[16],[17],[18],[19],[20]. Detecting the expressions and quantitatively describing them are important for research in multiple fields, such as medicine and psychology [21],[26],[27],[28],[29],[30]. The FACS was designed in the 70’s by Ekman and Friesen [5],[6],[11],[22],[23],[24][25]. It is based on years of psychological investigation and experimentation and still the most widely used as the robust method for describing facial behaviors. Its use has spread outside of the psychological and clinical fields.

FACS consists of a set of 44 visually discriminable independent Action Units (AUs). Expressions can be described by combining multiple AUs. Each expression is given a score that is made up of the list of AUs. Table 2, shows some examples of AUs.

No.	Action Unit	No.	Action Unit
1	Inner Brow Raiser	7	Lip Tightener
2	Outer Brow Raiser	10	Upper Lip Raiser
4	Brow Lower	17	Lower Jaw Raiser
5	Upper Lid Raiser	45	Blink

Table 2
Example Action Units

5.2 Modeling of Face Pictures

A dictionary is built for each user and is comprised of 7 images, see table 3. The feature points are extracted for each image and the values are normalized. For cononicalization the distance between the tails of the eyes is used as it does not get affected by facial expressions [31]. After calculating the feature values, a difference vector is calculated between the expressionless image and each of the basic images.

Image	Description
F0	Expressionless
F1	Expression of Happiness
F2	Expression of Surprise
F3	Expression of Dislike
F4	Expression of Fear
F5	Expression of Anger

Table 3
Expression Image Dictionary

The matrix X , see equation 1, is computed and stored for each individual. In the matrix i corresponds to the index of the 7 expression images and j points the feature value with that index. In the case of $i = 0$, the values for the expressionless image are used. The average vector is also computed for each user in the database. In order to save space and calculation time, the image is not stored in the database and only the vector is stored.

$$(1) \quad X = \begin{pmatrix} X_{0,0} & X_{0,1} & \cdots & X_{0,j} \\ X_{1,1} & X_{1,1} & \cdots & X_{1,j} \\ \vdots & \vdots & \ddots & \vdots \\ X_{i,0} & X_{i,1} & \cdots & X_{i,j} \end{pmatrix}$$

5.3 Expression Judgment

For judging the expression in an image the minimum distance identification method discussed in [32] is coupled with FACS. The expressions are ranked according to the number of the matched feature values. When the match is not above the given threshold or when the differences between the 1st ranked and other expressions are small, minimum distance identification is used. The minimum distance identification method is used on the input vector and those in the dictionary. If the FACS expression and the minimum distance identification expression are the same then that expression is the answer. If they differ then the answer will be the one with the larger distance from the 2nd best candidate.

The minimum distance identification method is used when FACS is not capable of determining the correct expression. Some examples of this situation is when both right and left facial expression is asymmetry or when the expression is individually characteristic. Since this method requires vector data from training images only people who have previously had their expressions put in the dictionary are able to be directly determined. However, a non-registered user's expressions can still be determined.

The first step in determining a non-registered user's expression is to calculate the feature values in their expressionless face. Then, this is compared with the values of every expressionless feature in the dictionary. The person in the dictionary with the closest vector data will then be used for further determination.

5.4 Judgment Using FACS

In [24], AUs are mapped to expressions. Each AU is labeled as operating (ON) or non-operating (OFF). The expression is determined by the agreement of ON and OFF AUs. The expression with the highest agreement is chosen as the candidate expressions.

5.5 Judgment Using Minimum Distance Identification

Judgment using minimum distance identification uses the vector data stored in the dictionary. The input image's feature values are compared to the expressionless image in the dictionary using the Euclidean distance. The expression with the minimum distance and the expressions within 5% of the minimum distance are scored. The expression with the highest score is chosen as the candidate expression.

6 Emotion energy acquired from speech patterns

An emotion of a speaker can be recognized based on speech patterns in most cases [34]. For example, the sentence: "The doctor told me that the injury is very serious" contains *fear and anxiety*. The sentence: "Ms. Hanako Tokushima, mother of Mr. Taro Tokushima passed away peacefully on Wednesday" contains *sorrow* and "The board of directors of the promotion strategy for information communication research and development has given final approval to fund your research application" contains *joy*, respectively.

The basic idea of acquiring emotion energy from speech patterns can be describe as follows.

For conversations, an "emotion dictionary", "image value database" and "favor value database" are identified, and "emotion attribute", "attribute image value" and "likability" are decided for each word in the conversations. Next, the "modifier dictionary" enlarges or reduces an emotion attribute for each noun or verb. The sentence pattern is searched for in the "emotion occurrence phenomenon dictionary." When the same pattern is found in the dictionary, the emotion attribute value is set for the sentence according to the emotion occurrence rule. The emotion parameter

is calculated and one emotion is judged.

Fig.3 shows a structure of the emotion estimation based on speech patterns.

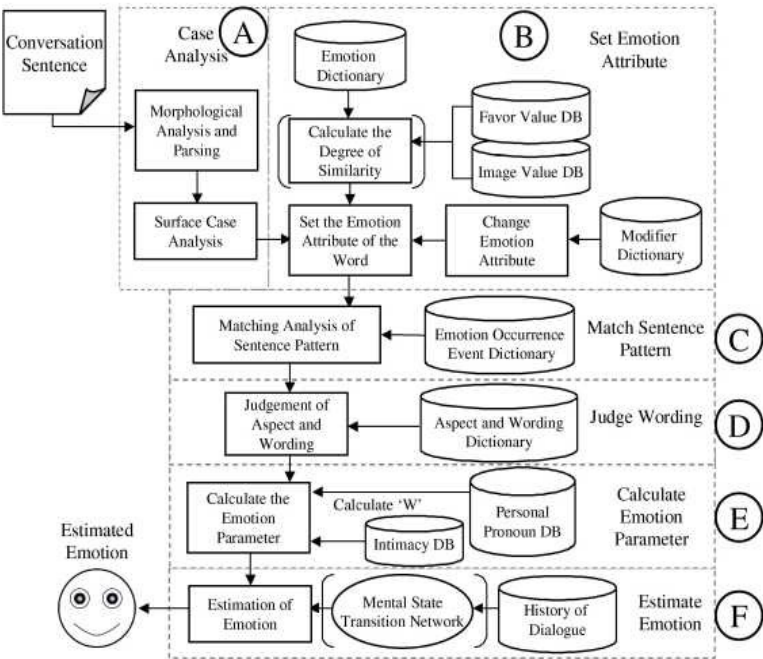


Fig. 3. Structure of the emotion estimation system

7 Conclusion

This paper has proposed a new paradigm for research on human emotion recognition and artificial emotion creation.

Traditional emotion studies were not focused on integration of internal and external aspects, because they were technically difficult. The research presented in this paper has developed a new paradigm for human mental state transition network and methods for human emotion recognition, artificial emotion creation, mental state transition of artificial emotion, and affective presentation. These are unseen creations which combine external feature information and physical reactions (mental state transition). The new methods are expected to largely contribute to the international society by developing future communication business.

Existing methods detect rhythms from voice sound (pronunciation), utilize emotion recognition (anger, joy, and sorrow) in order to divide recognition dictionaries

of voice sound, and then replace the dictionaries according to the emotion. However, the estimation of human emotions based on voice sound is limited. This paper uses speech patterns besides voice sound to develop a new method for human emotion recognition. Utilizing speech patterns for speaker's emotion measurement is a unique feature of our study. Emotion communication is an important issue for developing next-generation communication systems. However, the existing models including the external information of voice sound or language are far from modeling and transmitting human emotions. Our research integrates external information and internal emotion transition mechanism to recognize human emotion, which shows the novelty of our method.

Emotion recognition will provide computers with adequate conversations considering the user's condition, which will enable better computer-based services. For instance, medical and welfare services can be improved by recognizing patients' and the elderly people's emotions. Computers can figure out problems and also improve rehabilitations. It might be even possible to diagnose mental disease automatically. Moreover, integration with voice sound recognition technology will widen its application to communication style robot and realize healing conversations desired especially for elderly people, which will create new industries such as robotics, car navigation systems, and call center services and will eventually realize a barrier-free ubiquitous computing society.

Modern information communication has been mainly focused on verbal information, information quantified with 0's and 1's on the network. Recent studies emphasize the importance of "non-verbal information." The effectiveness of the communication is largely affected by who is in charge of the communication. For example, persuasion by boy/girl friend would be much more effective than by others. Such real-life examples prove the importance of considering emotion. However, traditional information communication techniques have not dealt with human emotion. Our research, which suggests both human emotion recognition model and artificial emotion creation model, has significant values and potentials in the sense that it would give the direction for development of next-generation information communication techniques and improvement of human communication.

This paper has introduced the abstract of our project: Human Emotion Recognition and Artificial Emotion Creation. The mental state transition network has been presented along with some results of the psychological experiments.

The following are some open problems and future work.

1. An individualization DB has been constructed, as shown in Fig.1. However, how does one copy with the individual difference in the mental state transition network?
2. How do we build an ontology and use it for emotion recognition?
3. How do we acquire the appearance emotion energy correctly?
4. Should we consider weights for each appearance emotion energy, such as emotion energy from linguistic information, from the phonetic information, and from the expressive information?

5. Should corporal information, for example, brain waves, be employed in the process? If yes, how do we use the corporal information to recognize the human emotion?

Utilizing the emotion energy acquired from external information in the proposed model, completing the mental state transition network, and developing an emotion interface will also be the future work.

Acknowledgement

Many thanks to Dr. Shingo Kuroiwa, David B. Bracewell, Junko Minato and all our colleagues participating in this project. The experimental systems described in the paper were developed by Kazuyuki Matsumoto (emotion measurement based on speech pattern), Shunji Mitsuyoshi (emotion estimation based on sound voice), Nobuo Nagao, Takahiro Kuroda and Jia Ma (facial expression recognition), Hua Xiang and Peilin jiang (the psychological experiment).

This research has been partially supported by the Ministry of Education, Culture, Sports, Science and Technology of Japan under Grant-in-Aid for Scientific Research(B), 19300029.

References

- [1] Zhiliang Wang, The Modeling of Artificial Psychology, *Proceedings of The International Conference on Artificial Intelligence*, pp1253, June 24-27, 2002.
- [2] Fuji Ren, Recognize Human Emotion and Creating Machine Emotion, Invited paper, *Information*, Vol.8, No.1, pp. 7-20, January 2005
- [3] Rosalind W. Picard, *Affective Computing*, Preface, pp2, pp22, pp190. The MIT Press Cambridge, Massachusetts London, England, 1997.
- [4] R. Cowie, E. Douglas-Cowie, N. TSApatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. Taylor. Emotion Recognition in Human-Computer Interaction, *IEEE Sig. Proc. Mag.*, Vol. 18 (1), pp.32-80, Jan 2001.
- [5] P. Ekman, R.W. Levenson, and W.V. Friesen, Automatic Nervous System Activity Distinguishes Among Emotion, *Science*, 221: 1208-1210, Sep. 1983.
- [6] W.M. Winton, L. Putnam, and R. Krauss, Facial and Autonomic Manifestations of the Dimensional Structure of Emotion, *Journal of Experimental Social Psychology*, 20:195-216
- [7] N.H. Frijda The Emotion, *The studies in Emotion and Social Interaction*, Cambridge University Press, Cambridge, 1986.
- [8] Plutchik, R. Emotions, *A Psychoevolutionary Synthesis*, New York: Harper & Row. 1980
- [9] Russell, J.A, A Circumplex Model of Affect, *Journal of Personality and Social Psychology*, 39, 1161-1178. 1980
- [10] P. R. Kleinginna, Jr. and A. M. Kleinginna, A Categorized list of Emotion Definitions, with Suggestions for a Consensual Definition, *Motivation and Emotion*, 5(4):345-379, 1981
- [11] P. Ekman, Universals and Cultural differences in Facial Expressions of Emotion, *In J.Cole (Ed.), Nebraska Symposium on Motivation*, Vol.19, Lincoln: University of Nebraska Press. Pp207-283 1972.
- [12] Goldstein MD, Strube MJ. Independence Revisted, The Relation between Positive and Negative Affect in A Naturalistic Setting, *Pers. Soc. Psychol. Bull.* 20:57-64.1994
- [13] Oatley, K., Jenkins, J.M, *Understanding Emotions*. Blackwell 1996
- [14] Yamata Takeshi, Murai Junnichi, Yokuwakaru Sinritookei, *Understanding psychological statistic*, pp. 32-80. ISBN 4-623-03999-4 Printed in Japan.2004

- [15] S. Akamatsu, Computer Recognition of human face - A survey, (in Japanese), *Transactions of IEICE, D-II*, vol. J 80-DII no. 8, 2031-2046, 1997.
- [16] Y. Uchino, Morphological Analysis System Chasen, *Information processing*, vol.41 no.11, 1208-1214, 2000.
- [17] K. Mera, T. Ichimura, T. Aizawa, T. Yamashita: Invoking Emotions in a Dialog System based on Word Impressions, *Transactions of the Japanese Society for Artificial Intelligence*, vol.17 No.3 A, 186-195, 2002.
- [18] A. Mehrabian, *Silent Messages: Implicit Communication of Emotions and Attitudes*, Wadsworth, Belmont, California.
- [19] H. Yamada and S. Shibui, The Relationship between Visual Information and Affective Meaning from Facial Expressions of Emotion, *Perception*, 27, 133, 1998.
- [20] J. Zhang, Y. Yan, and M. Lades, Face Recognition: Eigenface, Elastic Matching, and Neural Nets, *Proc. of IEEE Conf.*, vol. 85, no. 9, 1422-1435, 1997.
- [21] M. Yamaguchi, T. Kato, and S. Akamatsu, Relationship between Physical Traits and Subjective Impressions of the Face - Age and Sex Information, *Transactions of IEICE*, HC94-89, 1995.
- [22] P. Ekman, *Darwin and Facial Expression: A century of research in review*, Academic Press, N.Y., 1973.
- [23] P. Ekman and W.V. Friesen, Pictures of Facial Affect, *Human Interaction Laboratory*, University of California Medical Center, San Francisco, 1976.
- [24] P. Ekman, W.V. Friesen, Riki Kudo, *et al*, Unmasking the Face, *Seishinsyobo*, Tokyo, April, 1987.
- [25] P. Ekman and W.V. Friesen, *Facial action coding system : a technique for the measurement of facial movement*, Consulting Psychologists Press, 1978.
- [26] Matsubashi, Fujimoto, Nakamura, Minami, The proposal of an effective correction HSV table color system for face domain extraction, *Transactions of The Institute of Image Information and Television Engineers*, vol. 49, no.6, 787-787, 1995.
- [27] M. Kass, A. Witkin, and D. Terzopoulos, Snakes: Active contour models, *Int. J. Computer Vision*, vol. 1, 321-331, 1998.
- [28] Y. Mitsukura, M. Fukumi, and N. Akamatsu, A Design of Face Detection System by using Lip Detection Neural Network and Skin Distinction Neural Network, *IEEE SMC*, 2789-2793, 2000.
- [29] Sekioka, Y. Yokogawa, N. Funabiki, T. Higashino, T. Yamada, and Y. Mori, A Proposal of a lip Contour Approximation Method Using the Function Synthesis, *Transactions of IEICE*, vol. J84-D-II, no.3, March, 2001.
- [30] Y. Shiga, H. Ebine, O. Nakamura, On Extraction of the Feature Amounts from Facial Parts for the Recognition of Expressions, *Human Interface 84-17, Information Media*, 35-17, Aug 1999.
- [31] H. Shimoda, T. Kunihiro, and H. Yoshikawa, A Prototype of a Real-time Expression Recognition System from Dynamic Facial Image, *Journal of Human Interface Society*, vol. 1, no.2, 25-32, 1999.
- [32] K. Ishii, S. Ueda, E. Maeda, and H. Murase, *Intelligible pattern recognition*, Ohmsha, Tokyo, 1998.
- [33] Peilin Jiang, Hua Xiang, Fuji Ren, Shingo Kuroiwa, An Advanced Mental State Transition Network and Psychological Experiments, *Lecture Notes in Computer Science*, Volume 3824, pp.1026-1035, Springer-Verlag 2005.
- [34] Kazuyuki Matsumoto, Fuji Ren, Shingo Kuroiwa, An Algorithm for Measuring Human Emotions based on Context and Sentence Pattern, *Proceedings of the third International Conference on Information*, pp.215-218, 2004.
- [35] Fuji Ren, Dianbing Jian, Hua Xiang, Shingo Kuroiwa, Tetsuya Tanioka, Zhong Zhang, Chengqing Zong, Mental State Transition Network and Psychological Experiments, *the International Conference on Artificial Intelligence and Soft Computing*, pp.439-444, 2005.
- [36] Shinichi Chiba, Fuji Ren, Shingo Kuroiwa, et al. Analysis of electroencephalographic activity in condition of emotional activation in human, *the International Conference on Artificial Intelligence and Soft Computing*, pp.445-450, 2005.
- [37] Kyoko Osaka, Shin-ichi Chiba, et al., Estimating Emotion Changes Using Electroencephalographic Activities and its Clinical Application, *Proceedings of 2005 IEEE International Conference on Natural Language Processing and Knowledge Engineering*, pp.830-834, 2005.