2004－ＤＰＳ－117　（30）
2004－CSEC－　24　（30）
2004／3／5

Implementation and Evaluation of Feeling Communication System Based on Individual User Model

Kouhei Hayashi, Ying Dai, Yoshitaka Shibata
Faculty of Software and Information Science, Iwate Prefectural University
g231a028@edu.soft.iwate-pu.ac.jp, dai@soft.iwate-pu.ac.jp, shibata@iwate-pu.ac.jp

Recently, as advent of Internet and broadband network technologies, remote communication system such as video conference system, has been popular to communicate each other at the remote places. In such system, although understanding of the real feeling of the partner can be usually understood from his facial expression, the difference of the individual facial expression is always generated and prevents them from the mutual understanding. This difference becomes more serious specially when both are not familiar each other. On order to resolve this problem, we suggest a feeling communication system to attain real and smooth communication by feeling in consideration of individual user model. In our system, the real feeling of the partner is extracted by recognition part, transmitted through the network and synthesize into the facial expression using individual user model. This paper describes the system architecture and design issued and methodology to realize a total system. In order to verify the usefulness of suggested method, a prototype system is constructed to evaluate its functionality and performance.

個人モデルに基づく感情通信システムにおける実装と評価

林 耕平　戴 瑩　柴田 義孝
岩手県立大学ソフトウェア情報学部
g231a028@edu.soft.iwate-pu.ac.jp, dai@soft.iwate-pu.ac.jp, shibata@iwate-pu.ac.jp

近年インターネット回線の広帯域化に伴い、ビデオ会議システム等の遠隔コミュニケーションシステムの利用の機会が増大している。これにより、実際に一度も会ったことのない人とのコミュニケーションの機会も増えるものと思われる。このようなシステムの多くはお互いの感情の伝達を表情に頼る部分が多いが、人間の表情には個人差がある。このような個人差は時に相手の感情の理解を妨げる要因となりうる。特によく知らない相手とのコミュニケーションでは表情の個人差は重大な問題になると考えられる。本研究においてはこれまでこの表情の個人差を考慮し、遠隔コミュニケーションにおける感情の通信を円滑に行うことを目的とした「感情通信システム」の研究を行ってきた。このシステムでは表情の個人差を「個人モデル」として数値化し、それに基づいた表情の認識・合成を行う。本稿ではこの感情通信システムの実装と性能評価について述べる。

1. Introduction

Recently, many communication systems between remote places are developed by the spread of computer networks. Therefore, the opportunity of communication with the people who have not actually met so far increases. However, smooth communication can not be realized without correct understanding partner's feeling. The individual difference in facial expression causes mutual misunderstanding. And even if they feel the same feelings, not all of them necessarily show the same expression. Moreover, even if it looks the same expression, the recognized feeling may vary in person to person. Although many remote communication systems are communicating based on the facial expression, there are few systems in consideration of the individual difference of facial expression. Therefore, the communication prevention by the individual difference of facial expression poses a problem. The purpose of our research is to solve the problem in this communication gap between the user's prevention. So far we have investigated developed the "Feeling communication System"[1], [2] which eliminates the individual difference of expression and makes communication smoothly. Feeling Communication System can reduce individual difference based on the "Individual User Model" which expressed the individual difference of expression as a model. This paper describes the architecture and implementation of a system.

2. Outline of Feeling Communication System

Our "Feeling communication System" consists of two parts. One is "Analysis Part" and another is "Synthesis Part." In an "Analysis Part", a user's face expression is captured with a camera and the feature of expression operation is extracted. Then, feeling is analyzed by the extracted feature. Facial Extraction process is carried out using Individual User Model. It is possible for this to extract a user's feeling from the expression in which individuality is included. A user's extracted feeling information is transmitted to a communication partner's system. Moreover, in "Synthesis Part", the feeling information received from a communication partner's system is changed into expression information. Then, expression information is drawn as 3-D computer graphics. Also in "Synthesis Part", those feeling information is changed into expression information, an "Individual User Model" is used. Thus, by looking the generated expression, a user can understand a partner's feeling correctly.

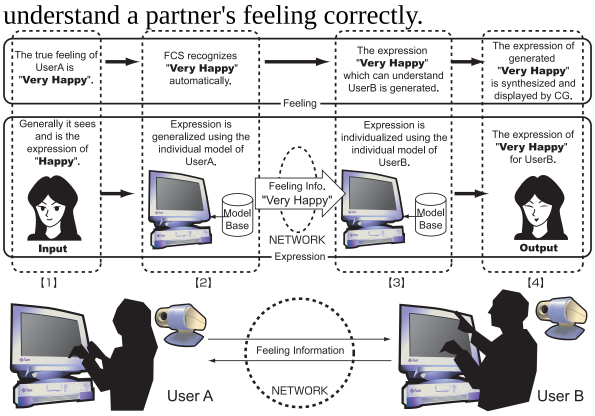

Fig 1 Example of remote communication by Feeling Communication System based on Individual User Model.

Fig.1 shows an example of the on-line communication using the suggested this system. User A and User B are communicating at remote places each other. Here, it is assumed that User A expresses her feeling moderately and User B recognizes partner's expression moderately. Moreover, both users do not know each other individuality. The both persons could not communicate smoothly, if communication using an ordinal TV is performed in such a situation. However, because the individual difference in facial expression is eliminate by the communication using a Feeling Communication System based on Individual User Model, it is possible to communicate smoothly.

3. Analysis Part

Analysis Part extracts a user's feeling from the facial image in which a user's individuality is included correctly. This is achieved by the expression analysis based on an Individual User Model. Fig. 2 shows system architecture of Analysis Part.

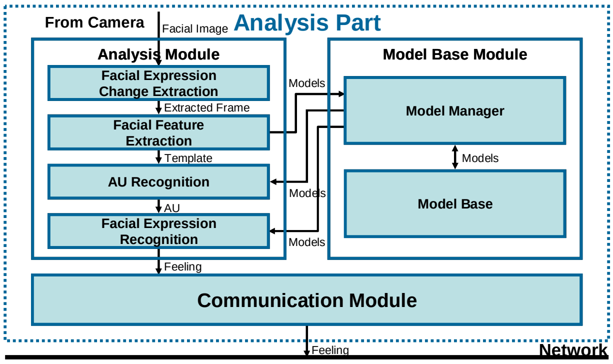

Fig 2 System architecture of Analysis Part.

Analysis Part consists of Analysis Module, Model Base Module, and Communication Module. In Facial Expression Change Extraction in Analysis Module, the frame with facial expression change is detected and captured. In Facial Feature Extraction, the features of facial expression operation are extracted and a template is output. In AU Recognition, Action Unit is recognized from the template. The AU which Ekman defined by Facial Action Coding System (FACS), is the minimum unit of distinguishable expression behavior which is independent mutually. In Facial Expression Recognition, the feeling is predicted based on the combination of AU. Model Manager of Model Base Module loads or stores an Individual User Model, and the Individual User Model is stored in Model Base. Communication Module manages communication with a communication partner's system. The operation and realization of the method of each module are shown below.

3.1. Facial Expression Change Extraction

In order to extract change of expression, a RGB color picture is changed into YIQ expression. The portion of people's skin is extracted from orange after that using I components including the color of image to cyanogen. Moreover, the difference of the continuous two frames of pictures is calculated to extract the portion of face with expression change.

3.2. Facial Feature Extraction[1][3]

The features of a motion are calculated by the optical flow of the continuation picture which had changed in face expression. Based on the histogram of optical flow, the templates of facial action features are gained. Because facial action features mainly appears in the surroundings of a mouth the surroundings of an eye, the histograms are calculated on the upper region and the lower region of face respectively. The intensity of AU and the expression which show the degree of feeling are defined.

Fig. 4 is an example of a y-axis template of the picture showing the feature of the motion of "pulling the both ends of a lip horizontally". The circumference of the upper lip goes up upwards and the feature to which the circumference of the lower lip falls downward a little is shown.

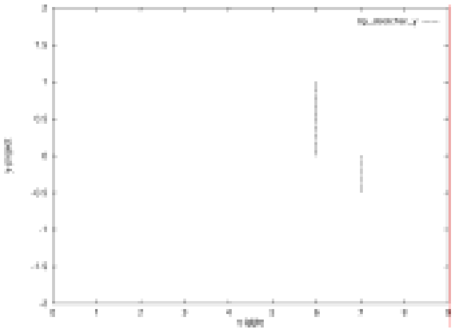

Fig 3 Example of lip stretcher y template.

3.3. AU Recognition[1][3]

AU is recognized using the Back Propagation algorithm (BP) of three layer neural network.

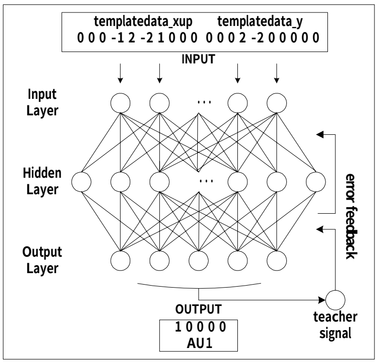

Fig 4 The neural network which recognizes AU in the upper half of a face.

Fig. 5 shows the network structure in the upper half of a face. Template data is input into an input layer with 20 units. As for the output layer with five units, each corresponds to one AU. It learns using the study data which made the group AU to the template data and it which were prepared beforehand. When the template data of an x-axis and a y-axis is input into the input layer of each network by this, the calculation will be performed by the neural network and AU will be outputted from an output layer.

3.4. Facial Expression Recognition[1][3]

By Associative Memory Model, expression is recognized from AU. In an Associative Memory Model, when x^k and y^k are memorized by the pair, the associative vector y^k can recollect from the key vector x^k.

$$\left[y^1, y^2, \ldots, y^K\right] = M\left[x^1, x^2, \ldots, x^K\right] \tag{1}$$

About six basic expression of "Happy", "Sadness", "Anger", "Fear", "Disgust", and "Surprise", an associative vector expresses one memory pattern by one vector, respectively, is what coded the pattern to 1 and 0, and expresses it with $y=(y_1, y_2, \ldots, y_n)^T$. Table 1 is a correspondence table of six basic expression and an associative vector. Moreover, what expressed the existence of AU with the vector $x=(x^1, x^2, \ldots, x^{44})$ of 1 and 0 serves as a key vector. And the memory matrix M in which $\|Y-MX\|^2$ becomes the minimum from the associative vector Y and the rate X of AU generating is estimated. Here, since an output is expressed with the vector of 1 and 0, the quantum function is defined as follows.

$$\Psi(u) = \begin{cases} 1 & \text{if } u > 0; \\ 0 & \text{otherwise}, \end{cases} \tag{2}$$

Then, the remembrance by the memory procession and the key vector can be expressed as follows.

$$z = \Psi(\hat{M}x - \theta) \tag{3}$$

Z makes the thing near an associative vector a recognition result by the threshold θ, and the feeling is outputted.

attribute value	association vector
Happiness (H)	$(100000)^T$
Sadness (SA)	$(010000)^T$
Surprise (SR)	$(001000)^T$
Fear (F)	$(000100)^T$
Anger (A)	$(000010)^T$
Disgust (D)	$(000001)^T$

Table 1 Associative vector

3.5. Model Base Module

The weights of BP neural network and the associative memory model are managed by the model base module. By adjusting the values of afoot weights, the individual BP model and associative memory model can be gotten by the model base module.

4. Synthesis Part

Based on the feeling information received from a communication partner's system through the network, "Synthesis Part" generates and displays expression for a user to understand a partner's feeling correctly.

Fig. 6 is system architecture of the Synthesis Part.

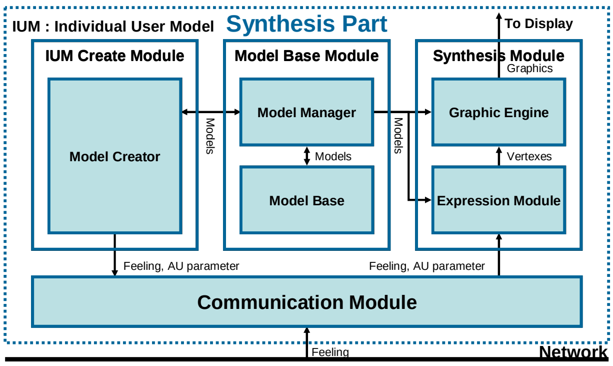

Fig 5 System architecture of Synthesis Part.

The Synthesis Part consists of Communication Module, Synthesis Module, Model Base Module, and IUM Create Module. Data transmission with the feeling information between communication partners is managed in the Communication Module. In the Expression Module of Synthesis Module, the vertex coordinates of the wire-frame model after expression change were calculated based on the received feeling information. Facial expression is realized in Graphic Engine. Model Manager of Model Base Module performs loading and storing of an Individual User Model, wire-frame model, and texture, and stores each model in Model Base. An Individual User Model is created in Model Creator of IUM Create Module. The function and the realization techniques of each module are described below.

4.1. Communication Module

This module performs management of communication with a partner's system, and management of each of other module. The feeling information received from the partner's system includes "feeling type" and "level". Communication Module receives these two parameters and passes them to Expression Module.

4.2. Expression Module

First, the feeling information received from Communication Module is changed into expression information in this module. Expression information is the combination of AU in "basic expression", and those intensity. In this system, combining basic expression like a montage generates other complicated expression. In order to combine basic expression, this system divides people's face into three domains, "upper" "middle" and the "bottom". Table 2 is example of generates expression. This expression is “compound”.

Domains of a face	Basic Expression	AU
Top	Neutral	-
Middle	Neutral	-
Bottom	Disgust	9, 10

Table 2 The example of the expression composition using three domains of a face. (contempt)

Second, based on the combination of AU, the vertex coordinates after modification of a 3-D polygon model are computed. This uses the muscle model method[5]. The muscle model method is the technique to simulate the muscles of the face called mimic muscles. Real expression modification is possible by using this technique. The computed vertex coordinates are sent to Graphic Engine.

4.3. Graphic Engine

In this module, a face model is transformed based on the vertex coordinates computed by Expression Module. The face model used by this system is a polygon model obtained using the 3-D scanner, and consists of 1928 polygons. Texture mapping is also performed by this module.

4.6. Model Base Module

An "Individual User Model", "face model", and "texture" are stored in Model Base. Model Manager accesses Model Base and loads these models on the memory. Thereby, the "Individual User Model" becomes usable in Expression Module, and the "face model" and the "texture" become usable in Graphic Engine.

4.5. IUM Creator

Individual User Model Creator (IUM Creator) is a module which creates the Individual User Model for Synthesis Part. The Individual User Model for synthesis describes the level of AU to a user's basic expression of operation. From the general model (AU which Ekman defined by FACS should put together.) showing an average expression, an Individual User Model adjusts and creates the level of AU of operation, and the combination of AU. The followings are formulas to create an Individual User Model from a general model. First, each basic motion level of AU is given to AUL_n from AUL_1. And the intensity adjustment parameter is given to K_n from K_1. Initially 1s are given to the parameters corresponding to AUs in which basic expression can be attained, and 0 is given as an initial value of an intensity adjustment parameter in addition to it. Here, when the user cannot be satisfied with this general model, an interface as shown in Fig. 6 is used, and an intensity adjustment parameter is changed. The scrollbar of the right window of Fig. 6 supports to adjust each intensity adjustment parameter of AU. The value of a parameter can be changed by moving this from/to the right and to/from the left. Moreover, the changed value is dynamically reflected in the expression of a left window. The user creates and saves the expression which carried out in this way and he could be satisfied. An Individual User Model will be computed by the left side if the saved intensity adjustment parameter is applied to a formula (4).

$$\begin{pmatrix} AUL'_1 \\ AUL'_2 \\ \vdots \\ AUL'_{n-1} \\ AUL'_n \end{pmatrix} = \begin{pmatrix} K_1 & & & & \\ & K_2 & & 0 & \\ & & \ddots & & \\ & 0 & & K_{n-1} & \\ & & & & K_n \end{pmatrix} \begin{pmatrix} AUL_1 \\ AUL_2 \\ \vdots \\ AUL_{n-1} \\ AUL_n \end{pmatrix} \qquad (4)$$

Fig 6 IUM Creator

5. Prototype System of Synthesis Part

In order to evaluate usefulness of our suggested method, we constructed a prototype system, which was implemented on a graphics engine, Octane in SGI. The prototype system was implemented by both C language and OpenGL. As the initial evaluation model of our suggested system, only the synthesis part was developed whole the analysis part was emulated to extract the correct feeling of the sender side through the analytical part which was carried out in previous research [3]. Therefore, it is assumed that the receiver side just can receive the information with the extracted feeling from sender side. The 3D facial model is made by capturing the actual human the 3D scanner. In our prototype system, 10 facial expressions including “doubt”, “interest”, “paralysis”, “apprehensions”, “frightened”, “contempt”, “distrust”, “upset”, “haughty”, “decision”, in addition to the basic one can be generated. All of the extended facial expression can be generated by combining those facial expressions. As a tool to create the individual model, “IUM Creator”, was introduced as shown in Fig. 6. Through the IUM Creator, the parameters of the AUs in the user model can be interactively adjusted until he could be satisfied with the facial expressions with the partner.

Fig 7 Example of synthesis expression "anger". (Top : anger, Middle : anger, Bottom : anger)

Fig 8 Example of synthesis expression "anger". (Top : anger, Middle : anger, Bottom : fear)

5.1. Evaluation

In order to verify the effectiveness of our suggested expression synthesis process based on the user model in this experiment, the recognition rates of the feeling for both cases with and without the user model were compared. As the experiment method, the by total 13 people in the 20th men and women was carried out how they could recognize the correct feeling from the synthesized facial for various feeling on the display. By assuming the first communication with the objective partner, no a priori information with him was provides. In this experiment, only the basic facial expressions were investigated by the following procedure:

1) Explain the procedure of the experiment.
2) Show various facial expressions without both any knowledge about the feelings and the user model, and obtain the feeling as answer to the presented facial expression.
3) Also show various facial expression while telling the kind of the objective feelings but not using the IUM and obtain the feelings as answer to the presented facial expression.
4) Also show various facial expression while telling the kind of the objective feelings and using the IUM and obtain the feeling as answer to the presented facial expression.

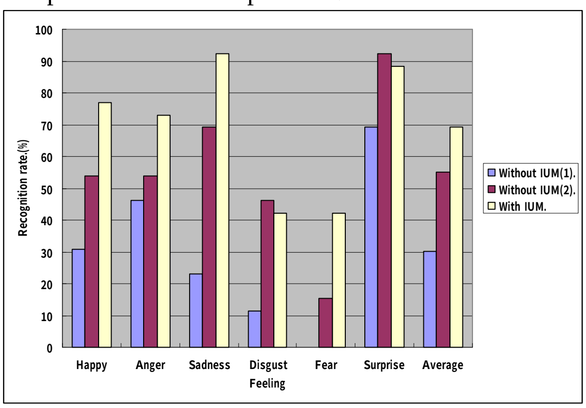

Fig 9 Comparison of the recognition rate of feeling.

The result of this experiment is shown in Fig. 9. From Fig. 9, it is founded that the recognition rate varies for each feeling. Particularly, the recognition rates are relatively higher for “anger” and “surprise” and lower for “disgust” and “fear”. The reasons why the recognition rate is lower for “disgust” and “fear” are due to the kind and the intensity of AUs for both “disgust” and “fear” are quite similar. The same answer that the difference between “disgust” and “fear” was not clear could be obtained from questionnaire result.

We can also see that the recognition rate using the Individual User Model is higher than that without the Individual User Model except for “surprise” and “disgust”. This reason with why the expression rate for ”disgust” is due to the similarity of both Aus as obtained from the questionnaire result. On the other hand, the reason with case of “surprise” is considered that the creation of the Individual User Model failed, particularly because the opening mouse as typical feature of “surprise” was too wider than usual one. As a result, the difference between “surprise” and “happy” cannot be identified as all of the tested persons regarded “surprise” expression as “happy” from the questionnaire result.

Fig 10 "Surprise" expression. (Without IUM)

Fig 11 "Surprise" expression. (With IUM)

6. Conclusion

In this paper, the feeling communication system which can reduce the miscommunication on the computer communication network due to the influence by the difference on nationality or culture was proposed. The system architecture and methodology to realized whole system was explained. In particular, the implementation of the synthesis part of the whole system was precisely described while the analysis part was emulated. From the result of this experiment, improvement has been checked to the rate of recognition of the feeling by use of an individual model. When the rate of recognition of the whole feeling was compared, in the experiment when not using an individual model, it was 55.1%, and in the experiment using an individual model, it is 69.2% and the rate of recognition improved about 14% by use of an individual model. However, the user made a mistake not creating an individual model well in feeling, and it has recognized, and also when the rate of recognition fell, a certain thing was understood. After that we will also implement the analysis part. This system can be useful not only for application of feeling communication but application of remote medical system where the facial expression of the patients in hospitals can be automatically recognized and transmitted to the nursing office to quickly treat him.

Reference

[1] Kouhei Hayashi, Shizuho Katahira, Ying Dai and Yoshitaka Shibata; Research of feeling communication system based on individual model, DPS, Vol.2002, No.54, pp115-120,(in Japanese) (2002).

[2] Hidetoshi Kuraishi, Yoshitaka Shibata; Feeling Communication System using Facial Expression Analysis/Synthesis based on Individual Model, IEEE Proceeding of ICOIN-10, pp211-220 (1996).

[3] Y. Dai, S. Katahira, Y. Shibata, D. Cai; Understanding of facial expressions by the hierarchical recognition approach of man-machine collaboration, Proc. of DMS 2002, pp160-166 (2002)

[4] Ekman, P. & Friesen, W. V.; Unmasking the face. A guide to recognizing emotions from facial clues. Englewood Cliffs, New Jersey: Prentice-Hall(1975)

[5] Waters, K. ; A muscle model for animating three-dimensional facial expression, Computer Graphics, Vol.21, No.4, pp.17-24(1987).