

A Method to Synthesize 3-Dimensional Face Corresponding to Diverse Words Expressing Facial Features through Mapping and Inspecting Validity of Mapping Function

Futoshi Sugimoto

dept. of Information Sciences and Arts

Yoyo University

Kawagoe, Japan

e-mail: f_sugi@toyo.jp

Abstract— In this study, the process to synthesize a human face based on the information of words is defined as a mapping from a word space, which is composed of the words expressing dimensions and shape of facial elements, into a physical model space where physical shape of the facial elements are formed. By introducing a concept of mapping, the use of whole words existing in the word space makes it possible to synthesize a human face based on free and uninhibited description. Furthermore, we have only to make 3-dimensional physical models corresponding to the words that are selected as training data to identify a mapping function. The others are made through the mapping. Finally, we inspect the validity of the mapping function that is obtained in this study.

Keywords- three dimensional facial synthesis; information of words; mapping function; computer graphics; GMDH.

I. INTRODUCTION

In recent years, since the importance of the facial information has been recognized, various studies and the practical use to computerize the information have been performed, and many papers and articles on them were published [1][2][3]. These researches can be divided into recognition and synthesis from the viewpoint of information processing. Our study is set in the area of the face synthesis among these researches. We aim at constructing a system to synthesize a 3-dimensional face by using computer graphics based on the information of words, which can describe facial features very freely (we call them “feature words” in this study). In this paper, we are to propose a method to form physical models of facial elements corresponding to the feature words using a mapping function, which plays a main role in our face synthesis system, and then we are also to inspect the effectiveness of the mapping function.

When we try to describe facial features of a person whom we picture to ourselves, the description is by using several distinct levels of words. Some words may describe directly and concretely the physical dimension and the shape of facial elements; while others may do so abstractly or metaphorically. In former studies on synthesis of a human face by utilizing the information of words, only few words describing directly the physical features were used, and the words explaining some degree of physical dimension, i.e.,

“slight,” “a little,” “very,” and so on, were simply added to [4].

In this study, our main aim is to synthesize a face based on free and uninhibited description. Which means we can use abstract and metaphorical words as well as concrete and physical ones as the feature words. In order to realize it, we define the process to synthesize a human face based on the information of words as a mapping from a word space to a physical model space. The word space and the physical model space are to be explained in another section in detail.

Introducing the concept of mapping enables us to synthesize the 3-dimensional physical models corresponding to diverse words [5][6][7][8]. We adopted GMDH (Group Method of Data Handling) [9] to identify a mapping function in this method. It is because the effectiveness of GMDH to identify a mapping function under the conditions whose relations are complicated and non-linear, and there are little training data, is already verified [10][11]. In this paper, we focus on inspecting the usefulness of the mapping and the validity of the mapping function that is obtained by GMDH.

The contents of the paper is as follows; In Section 2, the system we have been developing is outlined. In Section 3, the process to construct the word space and its characteristics are described. In Section 4, the physical model space is described. In Section 5, the process to make the training data and to identify the mapping function is described. In Section 6, the validity of the mapping function that is obtained in this study is inspected. Finally, in Section 7, the conclusion and the future works are presented.

II. OUTLINE OF THE SYSTEM

The outline of the facial synthesis system that we have been developing is shown in Figure 1. This system has a word space and a physical model space, which are to be explained in another sections in detail, and the process to synthesize a physical model of a human face is defined as a mapping from the word space to the physical model space. The facial elements in this research are nose, eyes, mouth, eyebrows, cheeks, jaw, and profile. The word space and the physical model space are made for each facial element, and the mapping is executed for each facial element respectively.

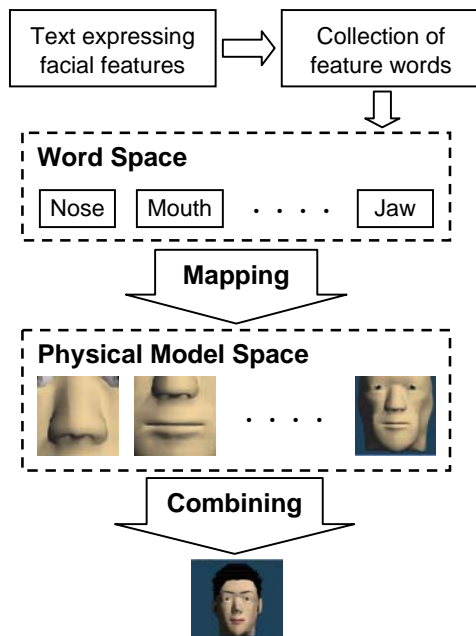


Figure 1. Outline of the system to synthesize 3-dimensional face

Before synthesizing facial elements, the feature words are collected from a sentence describing facial features or testimony of a witness, which is not, however, included in our current research. A physical model corresponding to an extracted feature word is made through mapping every individual facial element, and then a human face is synthesized through combining all physical models of facial elements together. This paper focuses on the part of making the physical models of facial elements by mapping.

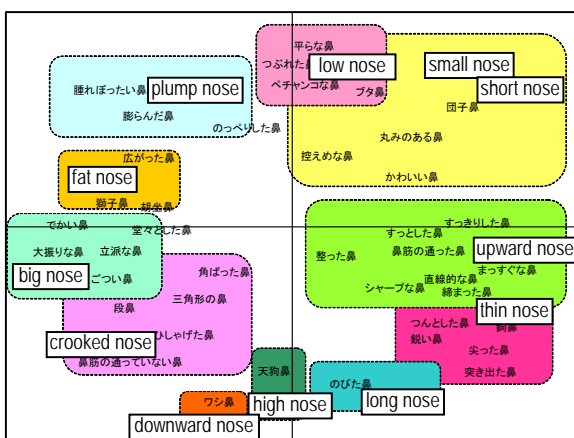


Figure 2. Word space, clustering result and training data of feature words in case of nose

III. WORD SPACE

The word space in this study is composed of the feature words, which express dimensions and shape of facial elements, and it is constructed for each facial element.

A. Construction of Word Space

In order to construct the word space, firstly many feature words were collected for each individual facial element, and secondly those words were located in a space by Multi-Dimensional Scaling method (MDS) [14] based on the similarity of those feature words. We call this space the word space. Those feature words were from a Japanese dictionary [12] and [13]. We provided four following criteria on collecting the feature words.

- (1) The nouns that express figuratively the feature of the facial elements.
- (2) The adjectives that express the feature of the facial elements when they are put before a facial element.
- (3) The mimetic words that express the feature of the facial elements.
- (4) The words that are used for expressing the feature of the facial elements in our daily conversation.

A similarity matrix among the feature words to be the input data of MDS was obtained using a method [15] of applying information theory in this study. In this method, the subjects (50 male students in our department, around 22 years old) classified the feature words based on the similarity of the impression with which the feature words are associated, and probability that the feature words are classified in the same group was found, and finally similarity among the feature words was calculated based on the probability. In the word space obtained from MDS, the more similar the feature words are, the closer they are located, while the farther, the less similar.

Every word space has six dimensions in this study. This is determined based on an indicator called “stress,” which shows how the distance relationship in the word space satisfies the similarity relationship among the feature words. Since a feature word is a point in the 6-dimensional word space, a feature word in the word space of a facial element is defined as W_i , and it is described as follows;

$$W_i = (w_1, w_2, \dots, w_6), \quad i = 1, \dots, m \quad (1)$$

Here, m is the number of the feature words in a word space, w_j expresses coordinate value of j th axis in the 6-dimensional space.

B. Characteristics of Word Space

The word space of nose is shown in Figure 2 as an example. Since we cannot find any proper English words corresponding to all Japanese feature words, we display the word space in Japanese except the training data. Although the word space is 6-dimensional in reality, it is projected on a two dimensional plane for visually understanding. The

characteristics of the word space summarized to be seen in Figure 2 are as follows;

- (1) The feature words that have completely opposite meanings stand face to face each other across the origin of coordinates.
- (2) Almost the feature words tend to be located at the edge of the word space.
- (3) The feature words that have the meaning of almost a standard feature are located near the center.

By the analogy from the characteristics mentioned above, it is appropriate to think that the “standard feature” is located at the origin of the word space, and the feature words that have an adjective that expresses the degree of feature such as “slight” and “very” are located on the straight line connecting the origin and a certain feature word.

IV. PHYSICAL MODEL SPACE

The physical model space in this study also is constructed for each facial element. It is composed of the physical shape of the facial element corresponding to each feature word.

A. Construction of Physical Model Space

The 3-dimensional geometric model of facial element corresponding to each feature word is made as a wire frame model by computer graphics (CG). In this study, the wire frame model is called the physical model of the feature words, and the space composed of the physical models is defined as physical model space. A physical model \mathbf{M}_i corresponding to a feature word \mathbf{W}_i of a facial element is a set of apexes of the wire frame model, which is described as follows;

$$\mathbf{M}_i = (\mathbf{P}_{i1}, \mathbf{P}_{i2}, \dots, \mathbf{P}_{in}) \quad (2)$$

Here, n is the number of apexes of the wire frame model for

each facial element. \mathbf{P}_{ij} is j th apex of the wire frame model, and it is composed of xyz coordinates as shown in (3).

$$\mathbf{P}_{ij} = (x_{ij}, y_{ij}, z_{ij}), \quad j = 1, \dots, n \quad (3)$$

Since the number of apexes is different from each facial element, the physical model space for each facial element has a different dimension ($3 \times$ the number of apexes) from each other.

B. Design of Standard Face Model

The standard face model used in this study is a Japanese man who is about 22 years old. The process to make the standard face model is as follows; First of all, the photographs of the face of 40 male university students were taken from the front and the side, and 34 items (some examples shown in Figure 3) were measured, and then the mean values of the items were calculated. Secondly, a wire frame model of the standard face was designed based on the mean values. Finally, the standard face model was completed by mapping the textures of eyes, eyebrows, lips and skin that are shown in Figure 4 on the wire frame model.

The standard face model is divided into 5 facial elements, e.g. nose, eyes, mouth, cheeks, and jaw, as shown in Figure 5, and their shape can be deformed for each element. The

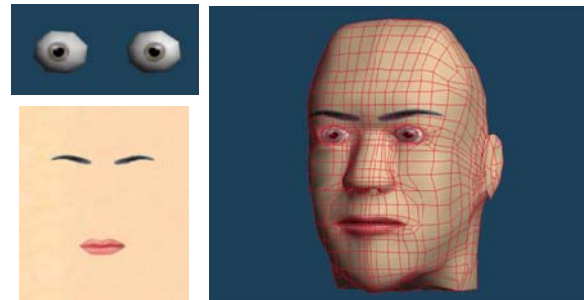


Figure 4. Texture and standard face model

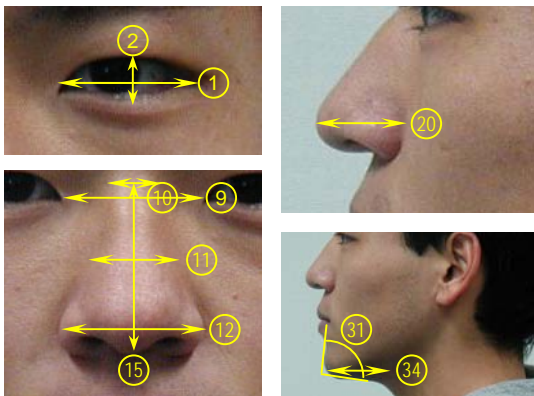


Figure 3. Examples of measurement items



Figure 5. Division into five facial elements

eyebrows are deformed by changing their shape, position and leaning of the texture, not by 3-dimensional model. It is not necessary to make the physical models for all feature words by manual labor for each facial element. Only the physical models for the feature words chosen for training data that is explained in Section 5 are needed to be made. Concerning the model of another feature words except training data, the coordinates of apexes of wire frame model are calculated by a mapping function, and the wire frame model becomes the physical model of other feature words.

V. MAPPING FUNCTION

There are many feature words in the word space of a facial element. In order to identify the mapping function, we need to select several training data from the feature words

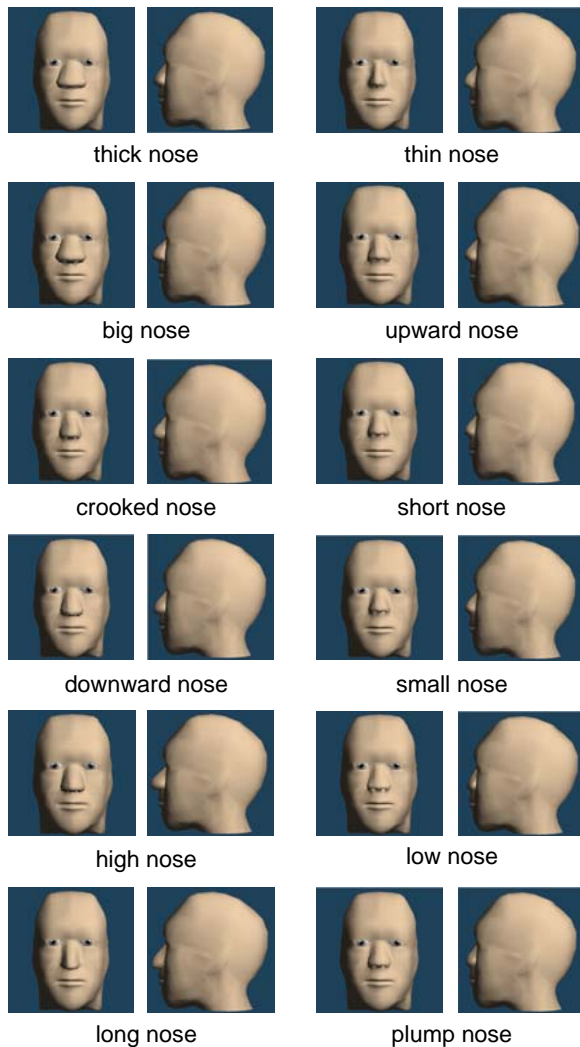


Figure 6. Physical models of nose corresponding to each training data of feature word

and to make physical models corresponding to the selected feature words.

A. Training Data

It is necessary that several feature words are extracted for training data from the word space equally in space for each individual facial element respectively. At first, the feature words were classified using cluster analysis based on Euclid distance among the feature words in the word space. Next, the representative was selected from each cluster, and they became the training data. The clusters of the feature words in the case of nose are shown classifying with colored area in Figure 2, and the words that are selected for the training data are enclosed with a square. Only the training data words are displayed in English for the reason mentioned in Section 3.

The training data in the physical model space corresponding to the one in the word space is necessary to be made. Several photographs, which have the facial element having the impression with which the training data is associated were picked up from 40 photographs mentioned in Sections 4. The physical models of the training data were made by manual labor based on the average value of the measured items of the selected photographs. Figure 6 shows the twelve training data words of nose and the physical models corresponding to the words.

B. Identification of Mapping Function

A set of xyz coordinates of all the apexes in the wire frame model becomes the parameters of the physical model space. We identify the mapping function from the training data using a statistical method, GMDH. The mapping function can be described as follows;

$$\mathbf{M}_i = \mathbf{f}(\mathbf{W}_i) \quad (4)$$

Since a physical model \mathbf{M}_i is a set of apexes of the wire frame model as shown in equation (2), the mapping function for each apex becomes as follows;

$$\mathbf{P}_{ij} = \mathbf{f}_j(\mathbf{W}_i) \quad (5)$$

Furthermore, since an apex \mathbf{P}_{ij} is composed of xyz coordinates as shown in equation (3), the actual mapping function becomes as follows;

$$x_{ij} = f_{ij}(\mathbf{W}_i), y_{ij} = f_{yj}(\mathbf{W}_i), z_{ij} = f_{zj}(\mathbf{W}_i) \quad (6)$$

A set of functions are obtained for each individual facial element respectively. The number of mapping function for each facial element is 3 x the number of apexes of the wire frame model.

VI. INSPECTION OF MAPPING FUNCTION

We inspect the validity of the physical models, which are made by this system using the mapping function in this section. Therefore, the questionnaire including 36 sets such as shown in Figure 7 was presented to 20 subjects, and they

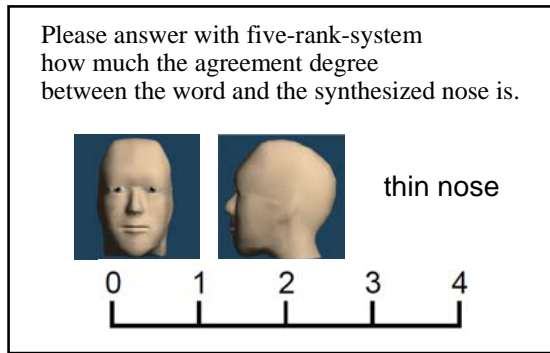


Figure 7. An example of questionnaire used in experiment to inspect validity of physical models

	Case	Agreement degree				
		0	1	2	3	4
Training data	(1)	3.09 ★				
	(2)	★ 0.38				
Other data	(3)	3.30 ★				
	(4)	★ 0.75				

















Figure 8. Result of experiment to inspect validity of physical models

were required to evaluate the agreement degree between the feature word and the physical model with five-rank-system. The subjects are male students in our department, and are around 22 years old. The 36 sets that combine the feature word and the physical model are as follows;

- (1) The training data model and the feature word corresponding to it, 12 sets.
 - (2) The training data model and the feature word having opposite meaning, 6 sets.
 - (3) The model belonging to the same group of the training data and the feature word corresponding to it, 12sets.
 - (4) The model belonging to the same group of the training data and the feature word having opposite meaning, 6 sets.
- (2) and (4) are inserted into the questionnaire in order to check the reliability of the subjects. The examples that used in each case in the experiment are shown in Table I.

The result of the questionnaire is shown in Figure 8. The evaluation of case (1) is slightly lower than case (3). The reason is that since the words directly express physical shape, they are selected as the training data, so it seems that the subjects may expect a more typical shape. On the other hand,

TABLE I. EXAMPLES USED EACH CASE IN THE EXPERIMENT

Case	Feature word	Physical model	
(1)	thin nose		
	thick nose		
(2)	thick nose		
	small nose		
(3)	flattened nose		
	tight nose		
(4)	pointed nose		
	snub nose		

since the other words except the training data are abstract, the subjects tend to easily agree. The portrait resembles the person himself if the characterized part is emphasized more than necessary. In the same way as this effect, the training data model may be the extreme geometric model that the part is emphasized more in our system too. However, it may be said that the purpose to make the physical models corresponding to the feature words except training data using mapping function can be accomplished enough.

VII. CONCLUSION AND FUTURE WORK

We propose a method to synthesize a 3-dimensional face from the information of words in this paper. This method allows a user to use the words abstractly or figuratively expressing the physical shape of facial elements as well as the words directly expressing it. The characteristics of this method is that it defines the process where a human face is synthesized based on the information of words as a mapping

from the word space to the physical model space. Using this method, it becomes possible to synthesize a human face corresponding to all words in the word space.

Finally, we describe the future challenge and prospect. The process to synthesize human 3-dimensional face by combining the physical models of facial elements together is already completed, so we will perform the evaluation immediately in the future and publish it in an article. Although the current standard face model is a Japanese man who is about 22 years old, we will make it for each sex, age and for typical shape of facial profile, and then we will make it possible to synthesize more various face models.

REFERENCES

- [1] M. Kaneko Ed., "Now, How Interesting a "Face" Is!; Processing of Facial Images and Its Application," The Journal of the Institute of Image Information and Television Engineers, Vol. 62, No. 12, pp. 13-41, 2008. (in Japanese)
- [2] R. Valenti, A. Jaimés and N. Sebe, "Facial Expression Recognition as a Creative Interface," Proceedings of 2008 International Conference on Intelligent User Interfaces, pp. 433-434, 2008.
- [3] S. Kumano, K. Otsuka K, J. Yamato, E. Maeda, and Y. Sato, "Pose-Invariant Facial Expression Recognition using Variable-Intensity Templates," International Journal of Computer Vision, Vol. 83, No. 2, pp. 178-194, 2009.
- [4] S. Iwashita and T. Onisawa, "Facial Caricature Drawing with Personal Impressions," Proceedings of 5th International Conference on Soft Computing, pp. 209-212, 1998.
- [5] Y. Hoshino and F. Sugimoto F, "Forming 3-Dimensional Face Model Based on the Information of Words Expressing Facial Feature," Proceedings of the 68th Annual Convention IPS Japan, Vol. 4, pp. 69-70, 2006. (in Japanese)
- [6] K. Mochida and F. Sugimoto, "A System to synthesize 3-Dimensional Face Based on the Flexible Verbal Expression," Proceedings of the 2006 Symposium of Human Interface Society, pp. 1139-1142, 2006. (in Japanese)
- [7] F. Sugimoto and M. Yoneyama, "3D Face Synthesis Based on the Information of Words Expressing Facial Features," Proceedings of IEEE Workshop on Computational Intelligence in Virtual Environments, pp. 1-6, 2009.
- [8] F. Sugimoto, "A Method to Visualize Information of Words Expressing Facial Features," Proceedings of IEEE Computer Society 2010 Fifth International Multi-conference on Computing in the Global Information Technology, pp. 169-174, 2010.
- [9] A. G. Ivakhnenko, "Polynomial theory of complex systems," IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-1, No. 4, pp. 364-378, 1971.
- [10] N. Honda, F. Sugimoto, and S. Aida, "Analysis of Cartoon Faces and Design of Face Pattern Used in Faces Method," The Bulletin of the University of Electro-Communications, Vol. 31, No. 1, pp. 1-10, 1980. (in Japanese)
- [11] A. G. Liao, W. F. Liu, and G. Y. Xiang, "The prediction research on the vehicles for business transport of Guangxi in China with the GMDH algorithm method," Proceedings of 2010 International Conference on Computer Application and System Modeling, Vol. 3, pp. 36-38, 2010.
- [12] K. Kindaichi, Japanese Dictionary, the Fifth Edition, Sanseido, 2001. (in Japanese)
- [13] R. Konomiya Ed., Dictionary to Explain Body, Shintensya , 2002. (in Japanese)
- [14] C. Hayashi A. and Akuto, Multi-Dimensional Scaling Method, pp. 3-160, Science Co., 1976. (in Japanese)
- [15] T. Saito, Multi-Dimensional Scaling Method, pp. 197-200, Asakurasyoten, 1980. (in Japanese)