A METHOD TO SYNTHESIZE THREE-DIMENSIONAL FACIAL MODEL BASED ON THE INFORMATION OF WORDS EXPRESSING FACIAL FEATURES

Futoshi Sugimoto and Masahide Yoneyama
Dept. of Information and Compuiter Sciences, Toyo University, 2100 Kujirai Kawagoe 3508585, Japan

Keywords: Facial synthesis, facial features, words information, and mapping.

Abstract: Our aim is to synthesize faces based on freely-elicited expressions by expanding the range of words describing the shape of facial elements to include abstract or metaphorical expressions. We realize this by defining the synthesizing process of a human face as a mapping from a word space to a physical model space. The use of whole words existing in the word space has made it possible to synthesize a human face based on freely-elicited expressions.

1 INTRODUCTION

When we intend to describe facial elements of a person, the expressions are made by using words of several distinct levels. Some words might directly express the physical dimension and shape of facial elements, while other words might abstractly or metaphorically express the shape of facial elements. In former studies about synthesis of a human face by utilizing the information of words, few words directly expressing the physical dimension and shape were used, or the words expressing the degree of physical dimension, i.e. slight, a little, very, and so on, were used (Iwashita, S. et al, 2000) (Shan, Y. et al, 2001).

In this study, our aim is to synthesize a face based on freely-elicited expressions. This is realized by expanding the range of the words expressing the dimension and shape of the facial elements to include abstract or metaphorical expressions. We define the process where a human face is synthesized based on the word information as a mapping from a word space that is organized with the words expressing the dimension and shape of facial elements to a physical model space where physical shape of the facial elements are concretely formed. Therefore, the use of whole words existing in the word space has made it possible to synthesize a human face based on freely-elicited expressions.

2 OUTLINE OF THE SYSTEM

The outline of the facial synthesis system which we have been developing is shown in Figure 1. This system has a word space and a physical model space, which are explained in later sections in detail, and the process to synthesize a physical model of a human face is defined as a mapping from the word

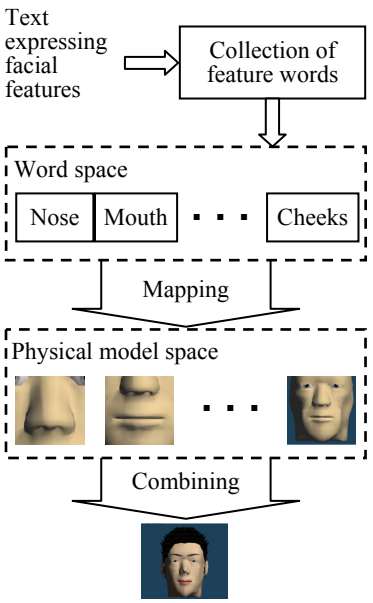

Figure 1: Outline of the facial synthesis system.

Sugimoto F. and Yoneyama M. (2007).
A METHOD TO SYNTHESIZE THREE-DIMENSIONAL FACIAL MODEL BASED ON THE INFORMATION OF WORDS EXPRESSING FACIAL FEATURES.
In *Proceedings of the Second International Conference on Computer Graphics Theory and Applications - AS/IE*, pages 221-224
DOI: 10.5220/0002075402210224

space to the physical model space. The word space and the physical model space are made for each facial element, and the mapping is executed for each facial element respectively.

Before synthesizing facial elements, the words expressing their shape (they are called the feature words) are collected from a sentence describing facial elements or testimony of a witness. This part is not included in this paper. A model corresponding to an extracted feature word is provided through mapping with respect to individual facial elements, and then a human face is synthesized through combining all physical models of facial elements together.

3 PROPOSED METHOD

We define the process where a human face is synthesized based on the word information as a mapping from a word space that is organized with the words expressing the dimension and shape of facial elements to a physical model space where physical shape of the facial elements are concretely formed. In this section, we explain the both spaces and the mapping function from the word space to the physical model space.

3.1 Word Space

Many feature words were collected with respect to individual facial elements, e.g. mouth, nose, eyes, eye-brows, cheeks, jaw, and profile, from a Japanese dictionary (Kindaichi, 2001). Then, every word space with respect to individual facial elements is formed. The procedure of forming the word space is as follows. First of all, a similarity matrix among the feature words is obtained from an experiment using subjects, based on the similarity of the shape recalled by the words. Second, a spatial layout of the feature words is obtained by inputting the similarity matrix into Multi-Dimensional Scaling method (MDS). This spatial layout of the feature words based on the similarity matrix of the feature words is the word space. Figure 2 is the word space of nose. The figure is projected on a two dimensional plane for recognizing it visually. The word space is characterized by the following facts;

(1) The origin is neutral, the farther away a feature word is from the origin, the greater the characteristic of the feature word.
(2) Similar words are arranged close together, while dissimilar words are arranged further away from each other.
(3) Every word space has six dimensions. This is determined based on an indicator called "stress," which shows how the distance relationship in the word space satisfies the similarity relationship between the feature words.
(4) A feature word $\mathbf{W}_i$ in the word space is described as follows;

$$\mathbf{W}_i = (w_1, w_2, ..., w_6), \quad (1)$$

i=1, . . . , *the number of feature words*

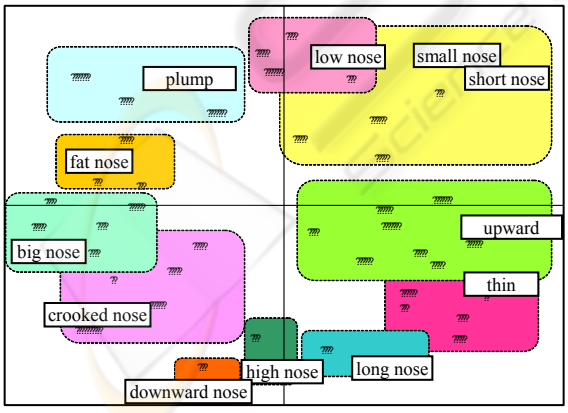

Figure 2: Word space and feature words as training data in the case of nose. Since we can not find the correct English words corresponding to all Japanese feature words, we display the word space in Japanese except the training data.

3.2 Physical Model Space

Concrete shapes of the facial elements on 3-dimensional computer graphics (CG) are determined by xyz coordinates of apexes of a wire frame model. Although the number of apexes of each facial element is different from each other, a set of xyz coordinates of all the apexes in the wire frame model becomes the parameters of the physical model space. A physical model $\mathbf{M}_i$ of a facial element is described as following;

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

here, n is the number of apexes of each facial element. $\mathbf{P}_{ij}$ is jth apex of a wire frame model.

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

3.3 Mapping Function

Several feature words are extracted as training data from the word space equally in space for each individual facial element respectively, and the wire frame models corresponding to the extracted feature words are built up by means of a CG tool. Table 1

Table 1: Feature words as training data in the case of nose and physical models corresponding to the words.

thick nose		thin nose	
big nose		upward nose	
crooked nose		short nose	
downward nose		small nose	
high nose		low nose	
long nose		plump nose	

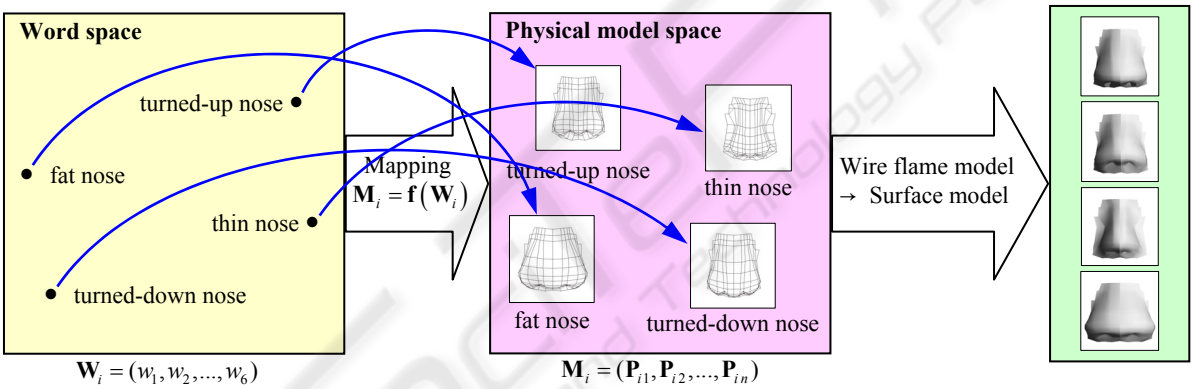

Figure 3: Example of the mapping process to synthesize physical models in the case of nose.

shows the twelve extracted feature words in the word space of nose and the physical models corresponding to the words. A set of xyz coordinates of all the apexes in this wire frame model becomes the parameters of the physical model. Then, we identify the mapping function **f** based on the training data using a statistical method, Group Method of Data Handling (GMDH) (Ivakhnenko, 1971). The mapping function is described as follows;

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

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

The actual mapping functions are the following, and the number of functions is $3 \times n$.

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

A set of functions are obtained for each individual facial element respectively. Figure 3 shows an example of the mapping process to synthesize physical models in the case of nose.

3.4 Calculation to Compound the Feature Words

The shape of facial elements is occasionally expressed by plural feature words. In addition, the shape is often expressed by adding some words expressing the degree of the feature to the feature word, i.e. very chubby cheeks, slightly-slanted eyes. In such a case, the positions occupied by these graded feature words in the word space are obtained by a grading calculation. Furthermore, each physical model corresponding to the individual graded feature word is obtained by means of the mapping function mentioned above, and integrating those physical models by means of morphing of CG in the

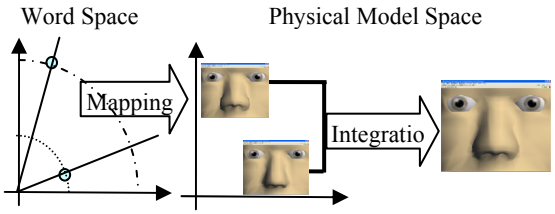

Figure 4: The process of grading calculation in the word space, the mapping from the word space to the physical model space, and the integrating physical models corresponding to the individual graded feature words in the case of nose.

physical model space making it possible to obtain a final physical model with respect to the individual facial elements. Figure 4 shows the process of the grading calculation, the mapping, and the integration.

4 CONCLUSIONS

We asked a subject to watch a photograph and describe the looks freely, and synthesized a personal face by the proposed system based on the description. Three samples of photographs and synthesized faces are shown in Figure 5. The looks description of each photograph is as follows;

(a) "He has slant eyes, a thin nose, and an oblong mouth. His jaw is sharp and cheeks sink."

(b) "He has very thin eyes, a straight lined nose, and a little edge upped mouth. His jaw is sturdy and cheeks a little chubby."

(c) "He has very downward-slanting eyes and eyebrows, a little big nose, and chubby cheeks. His jaw is round."

Twenty subjects were shown one synthesized face and ten of photos including a photo corresponding to the synthesized face at the same time, and asked to select a correct one from among ten photos. All synthesized faces were considered similar to the corresponding original faces.

5 FUTURE WORKS

We briefly explain about the future prospects for this system. When we describe a person's face, we also use words to describe the skin or hair, such as "pale cheeks" or "a 7:3 hair parting (like Japanese typical businessmen)." We must collect such words and prepare models using them. We have already prepared phy\sical models appropriate for feature words from some 50 portrait photos, and intend to

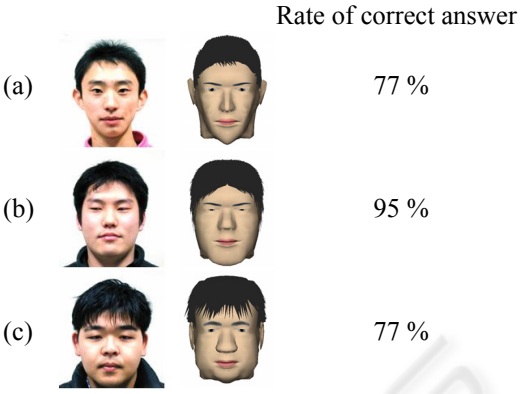

Figure 5: Three samples of photographs and synthesized faces based on a testimony.

use the face database of 300 people to prepare more physical models appropriate for feature words. At the same time, we also intend to draw more accurate CG models.

At the moment, the parameters of the physical model space are absolute coordinates of the apexes of wire frame models, but we plan to replace the absolute coordinates with relative variations from the standard model and we believe that we will be able to cope with differences in gender or age by changing the standard models.

REFERENCES

Ivakhnenko, A.G., 1971. Polynomial theory of complex systems. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol.SMC-1, No.4. pp.364-378.

Iwashita, S. et al, 2000. Facial caricature drawing based on the subjective impressions with linguistic expressions, *IEICE Transactions on Information and Systems*, Vol.J83-D-I, No.8, pp.891-900. (in Japanese)

Shan, Y. et al, 2001. Model-Based Bundle Adjustment with Application to Face Modeling. *Proceedings of the 8th International Conference on Computer Vision*, Vol. II, Vancouver, Canada, pp.644-651.

Kindaichi, K., 2001, Japanese Dictionary, the Fifth Edition, Sanseido.