Journal of Information Processing Vol.20 No.3 531–536 (July 2012)

[DOI: 10.2197/ipsjjip.20.531]

Regular Paper

A Method to Synthesize Three Dimensional Face Models by Mapping from a Word Space to a Physical Model Space and the Inspection of the Mapping Function

Futoshi Sugimoto[1,a)] Makoto Murakami[1,b)] Chieko Kato[1,c)]

Received: September 4, 2011, Accepted: February 3, 2012

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-dimensinal 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: synthesis of face, feature word, word space, mapping, GMDH

1. 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 researche 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 researche. 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 (we call them "feature word" in this study). In this paper, we 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 Ref. [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 that GMDH is the very effective method to identify a mapping function under the conditions whose relations are complicated and non-linear, and there are little training data. The effectiveness of GMDH is already verified by some papers [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.

2. Outline of the System

The outline of the facial synthesis system that we have been developing is shown in **Fig. 1**. This system has a word space and a physical model space, which are to be explained in another sec-

[1] Toyo University, Kawagoe, Saitama 350–8585, Japan
[a)] f_sugi@toyo.jp
[b)] murakami_m@toyo.jp
[c)] kato-c@toyo.jp

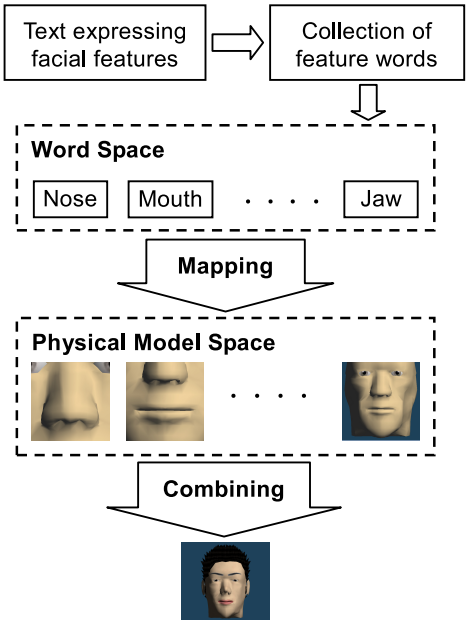


Fig. 1 Outline of the system to synthesize 3-dimensional face.

tions 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.

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.

3. 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.

3.1 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 picked up from a Japanese dictionary [12] and Ref. [13]. In picking up the feature words, we provided four following criteria and selected someone from the picked up words on the reason that they are used in everyday life, and anyone can clearly imagine facial feature from them.

(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.

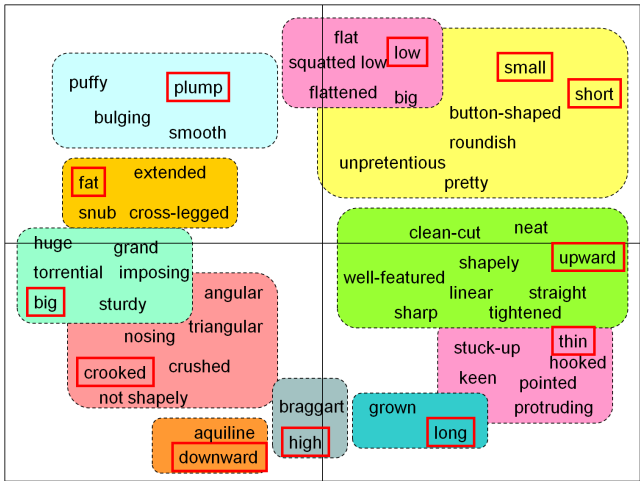


Fig. 2 Word space, clustering result and training data of feature words in case of nose.

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 $\mathbf{W}_i$, and it is described as follows;

$$\mathbf{W}_i = (w_1, w_2, \ldots, w_6), \quad i = 1, \ldots, m \tag{1}$$

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

3.2 Characteristics of Word Space

The word space of nose is shown in **Fig. 2** as an example. Although the word space is 6-dimensional in actuality, it is projected on a two dimensional plane for visually understanding. The characteristics of the word space summarized to be seen in Fig. 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 stan-

dard 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.

4. 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.

4.1 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 word, 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;

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

Here, n is the number of apexes of the wire frame model for each facial element. $\mathbf{P}_{ij}$ is jth apex of the wire frame model, and it is composed of xyz coordinates as shown in Eq. (3).

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

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.

Although the physical model is simply constructed with wire frame, it is transformed into polygon model when it is displayed.

4.2 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 (they are shown in **Fig. 3**) were measured using an application, which can measure distance and angle of a certain part of the picture that is put on it. These distances and angles of the parts were selected according to the thought that they are necessary in order to decide the position, shape, size, and etc. of the facial elements. The distance 23 was set to 100 as a standard value, and the other distances were interpreted relative to it. On the other hand, the angles were used just they were. Then, the mean values of the items of 40 students were calculated. Secondly, a wire frame model of the standard face was designed based on the mean values. Finally, the wire frame model was transformed into a polygon model and the textures of eyes, eyebrows, lips and skin were mapped on the polygons as shown in **Fig. 4**.

The standard face model is divided into 5 facial elements, nose, eyes, mouth, cheeks, and jaw, as shown in **Fig. 5**, and their shape

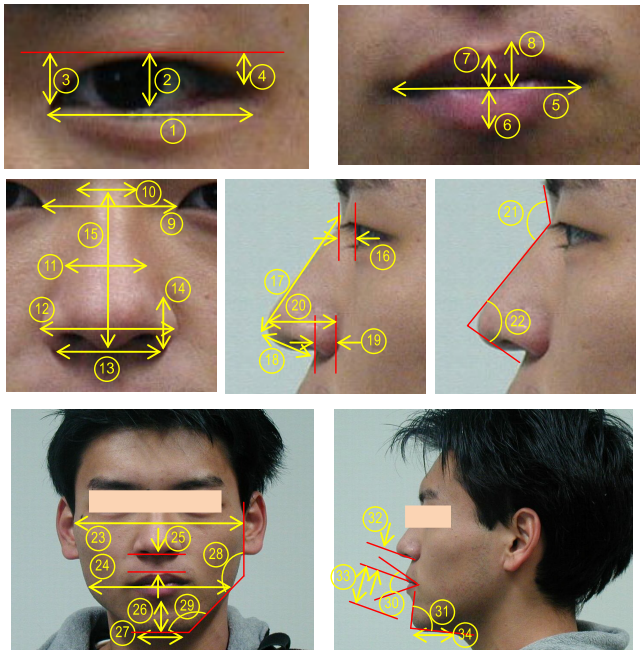


Fig. 3 All measurement items.

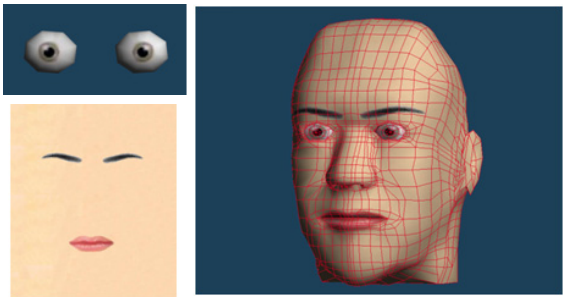


Fig. 4 Texture and standard face model.

Fig. 5 Division into five facial elements.

can be deformed for each element. The 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.

5. 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 and to make physical

models corresponding to the selected feature words.

5.1 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 word in the case of nose are shown classifying with colored area in Fig. 2, and the words that are selected for the training data are enclosed with a red square.

The training data in the physical model space corresponding to the one in the word space has 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 Section 4.2. 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.

5.2 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 the statistical method, GMDH. It is a family of inductive algorithms for computer-based mathematical modeling of multi-parametric datasets that features fully automatic structural and parametric optimization of models. The details of the method is provided in Refs. [9] and [16], so then we show briefly the algorithm in Appendix.

The mapping function can be described as follows;

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

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

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

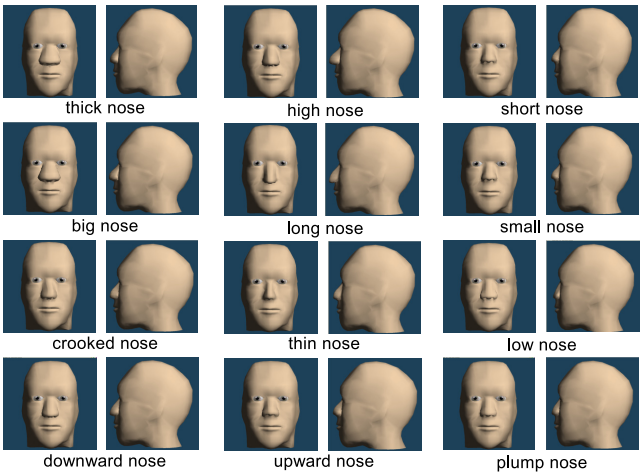


Fig. 6 Physical models of nose corresponding to each training data of feature word.

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

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

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

6. 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 **Fig. 7** was presented to 20 subjects, and they 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, 12 sets.
(4) The model belonging to the same group of the training data and the feature word having opposite meaning, 6 sets.

Case (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 1**.

The result of the questionnaire is shown in **Table 2**. There are statistically significant differences between the average degrees of case (1) and case (2) ($t = 40.04$, 19 d.f., $p < 0.001$), and between case (3) and case (4) ($t = 29.69$, 19 d.f., $p < 0.001$). The subjects agree when the feature word corresponds to the facial model. However, they disagree when the feature word doesn't correspond to the facial model. This means that the subjects are reliable. The evaluation of case (1) is slightly lower than case (3) ($t = -4.00$, 19 d.f., $p < 0.001$). 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. While, 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 empha-

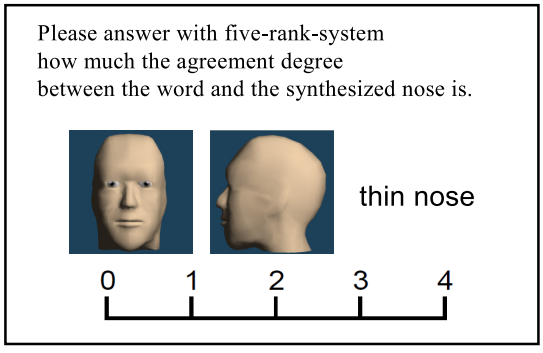


Fig. 7 An example of questionnaire used in experiment to inspect validity of physical models.

Table 1 Examples used each case in the experiment.

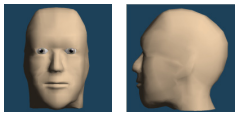
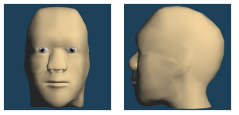
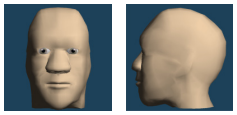
Case	Feature word	Physical model	Case	Feature word	Physical model
(1)	thin nose		(3)	flattened nose	
	thick nose			tightened nose	
(2)	thick nose		(4)	pointed nose	
	small nose			button-shaped nose	

Table 2 Result of experiment to inspect validity of physical models.

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

sized 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.

7. Conclusion and Future Work

In this paper, we propose a method to synthesize a 3-dimensional face from the information of words. 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. This paper shows the effectiveness of this method using a mapping function and establishes that the mapping function that is identified by GMDH has functioned effectively. Using this method, it becomes possible to synthesize a human face corresponding to all words in the word space.

Finally, we describe the future work 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. In our future work, we will propose a method to synthesize the physical model when the degree of feature is expressed and when the feature of facial element is described using plural feature words, and then we will shows that the method makes it possible to synthesize the physical model based on the expression even if it is very complicated. 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 race, and then we will make it possible to synthesize more various kinds of face models.

References

[1] Kaneko, M. (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] Valenti, R., Jaimes, A. and Sebe, N.: Facial Expression Recognition as a Creative Interface, *Proc. 2008 International Conference on Intelligent User Interfaces*, pp.433–434 (2008).

[3] Kumano, S., Otsuka, K., Yamato, J., Maeda, E. and Sato, Y.: Pose-Invariant Facial Expression Recognition using Variable-Intensity Templates, *International Journal of Computer Vision*, Vol.83, No.2, pp.178–194 (2009).

[4] Iwashita, S. and Onisawa, T.: Facial Caricature Drawing with Personal Impressions, *Proc. 5th International Conference on Soft Computing*, pp.209–212 (1998).

[5] Hoshino, Y. and Sugimoto, F.: Forming 3-Dimensional Face Model Based on the Information of Words Expressing Facial Feature, *Proc. 68th Annual Convention IPS Japan*, Vol.4, pp.69–70 (2006). (in Japanese)

[6] Mochida, K. and Sugimoto, F.: A System to synthesize 3-Dimensional Face Based on the Flexible Verbal Expression, *Proc. 2006 Symposium of Human Interface Society*, pp.1139–1142 (2006). (in Japanese)

[7] Sugimoto, F. and Yoneyama, M.: 3D Face Synthesis Based on the Information of Words Expressing Facial Features, *Proc. IEEE Workshop on Computational Intelligence in Virtual Environments*, pp.1–6 (2009).

[8] Sugimoto, F.: A Method to Visualize Information of Words Expressing Facial Features, *Proc. IEEE Computer Society 2010 5th International Multi-conference on Computing in the Global Information Technology*, pp.169–174 (2010).

[9] Ivakhnenko, A.G.: Polynomial theory of complex systems, *IEEE Trans. Syst., Man, and Cybernetics*, Vol.SMC-1, No.4, pp.364–378 (1971).

[10] Honda, N., Sugimoto, F. and Aida, S.: 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] Liao, Z.G., Liu, W.F. and Xiang, G.Y.: The prediction research on the vehicles for business transport of Guangxi in China with the GMDH algorithm method, *Proc. 2010 International Conference on Computer Application and System Modeling*, Vol.3, pp.36–38 (2010).

[12] Kindaichi, K.: *Japanese Dictionary, 5th Edition*, Sanseido (2001). (in

Japanese)
[13] Konomiya, R. (Ed.): *Dictionary to Explain Body*, Shintensya (2002). (in Japanese)
[14] Hayashi, C. and Akuto, A.: *Multi-Dimensional Scaling Method*, pp.3–160, Science Co. (1976). (in Japanese)
[15] Saito, T.: *Multi-Dimensional Scaling Method*, pp.197–200, Asakura-syoten (1980). (in Japanese)
[16] Farlow, S.J.: The GMDH Algorithm of Ivakonenko, *The American Statistician*, Vol.35, No.4, pp.210–215 (1981).

Appendix

Figure A·1 illustrates a summary of the model construction by GMDH. The basic technique of GMDH algorithm is a self-organization method. It fundamentally consists of the following steps:

(1) Split a dataset including a dependent variable y and independent variables $x_1, x_2, \ldots, x_m$ into a training dataset and a checking dataset.
(2) Make every two variable pairs taking out from m independent variables. The number of pairs is $s = {}_m\mathrm{C}_2$.
(3) Estimate the coefficients a_k $(k = 0, \ldots, 5)$ of all transfer functions $G(x_i, x_j)$ to using training dataset. Here, the transfer function is defined as follows.

$$G(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2 \qquad \text{(A.1)}$$

(4) Compute mean square error between y and prediction $(u_1, u_2, \ldots, u_s)$ of each transfer function using checking dataset.
(5) Sort out the predictions in ascending order of mean square error and select p predictions $(u'_1, u'_2, \ldots, u'_p)$.
(6) Set the selected predictions in the first layer to new input variables for the next layer, and sort out again the prediction $(v_1, v_2, \ldots, v_s)$ that are estimated in the second layer.
(7) Build up a multi-layer structure by applying steps (2)–(6).
(8) When the mean square error become larger than that of the previous layer, stop adding layers and choose the transfer function having the minimum mean square error in the highest layer as the final model output.

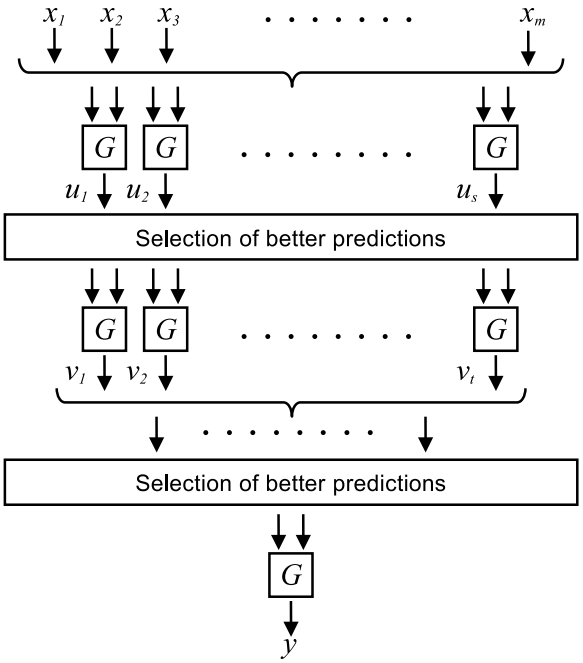
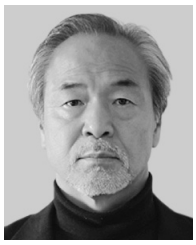
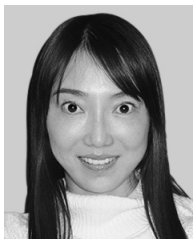


Fig. A·1 Summary of the model construction by GMDH.



Futoshi Sugimoto received his B.S. degree in communication systems engineering and M.S. degree in management engineering from the University of Electro-Communications, Tokyo, Japan, in 1975 and 1978, respectively, and Ph.D. degree in computer science from Toyo University, Tokyo, Japan, in 1998. In 1978, he joined Toyo University as a Research Associate in the Department of Information and Computer Sciences. From 1984 to 1999, he was an Assistant Professor, from 2000 to 2005, was an Associate Professor, and from 2006 to 2008, was an Professor in the same department. From 2009, he has been a Professor in the Department of Information Sciences and Arts. From April 2000 to March 2001, he was an exchange fellow in the University of Montana, USA. His current research interests are in cognitive engineering and human interface. Dr. Sugimoto is a member of the Institute of Image Information and Television Engineers, IPSJ, and Human Interface Society (Japan).

Makoto Murakami received his Bachelor, Master, and Ph.D. Degrees in Information Science from Waseda University, Tokyo, Japan in 1997, 1999, and 2003, respectively. Currently he is working as an Associate Professor in the Department of Information Sciences and Arts, Toyo University, Saitama, Japan. His research interest includes image processing, speech processing, and human interface. He is a member of IPSJ, IEICE, IEEE CS, and ACM.



Chieko Kato graduated from the Faculty of Literature, Shirayuri Women's University in 1997, and received her M.A. from the Tokyo University and Dr. Eng. degree from Hosei University in 1999 and 2007, respectively. She served 2003 to 2006 as an Assistant Professor at the Oita Prefectural Junior College of Arts and Culture. She currently teaches at Toyo University, which she joined in 2006 as an Assistant Professor, and was promoted to an Associate Professor in 2007. Her research areas include clinical psychology and psychological statistics. She is a member of IEICE Japan, Design Research Association, the Japanese Society of Psychopathology of Expression and Arts Therapy.