Using Deep Learning to Estimate User Impressions of Designs for 3D Fabrication

Koichi Taguchi, Manabu Hashimoto, Kensuke Tobitani and Noriko Nagata

Abstract This paper proposes a method for applying three typical human impressions directly into product designs to fabricate products using 3D printers. The method automatically estimates human impressions of the three-dimensional shape of an item in terms of three representative sensibilities: “hard–soft,” “flashy–sober,” and “stable–volatile.” This technique can be used for new 3D fabrication processes that reflect the designer’s intentions directly into the shapes of products. To estimate the impressions of the shape of an object, we need to draw strong correlations between impressions, which are psychological factors, and the aspects of the shape of the object, which are physical factors. The method uses deep learning effectively to address this issue. The object being evaluated is first converted to a set of images by photographing it from 20 surrounding directions. This image set is used as input data for deep learning with parameters of human impressions of the object as supervisory signals. In experiments, we used original dataset of three-dimensional objects of a car with assigned impressions that had been quantified using the semantic differential (SD) method. The correlation coefficients between impressions estimated using this method and the supervisory signals for all the datasets were about 0.70 for “hard–soft,” about 0.61 for “flashy–sober,” and about 0.67 for “stable–unstable.”

K. Taguchi (✉) · M. Hashimoto
Chukyo University, 101-2 Yagoto Honmachi, Showa-ku, Nagoya-shi 466-8666, Aichi, Japan
e-mail: taguchi@isl.sist.chukyo-u.ac.jp

M. Hashimoto
e-mail: mana@isl.sist.chukyo-u.ac.jp

K. Tobitani · N. Nagata
Kwansei Gakuin University, 2-1 Gakuen, Sanda-shi 669-1337, Hyogo, Japan
e-mail: tobitani@kwansei.ac.jp

N. Nagata
e-mail: nagata@kwansei.ac.jp

A. Martínez-García et al. (eds.), *Progress in Optomechatronic Technologies*,
Springer Proceedings in Physics 233, https://doi.org/10.1007/978-981-32-9632-9_8

1 Introduction

Various kinds of three-dimensional printers have been developed, and their cost has decreased. Thus, we can now use them not only in factories but also at home. Furthermore, a useful database of three-dimensional models has been prepared, and everyone can utilize it via the Internet. "Personal fabrication (production using three-dimensional printers by individuals)" will likely change the traditional way of manufacturing by mass production at factories. Enabling "personal fabrication" will require skillful modeling techniques using CAD systems to make the desired three-dimensional objects. However, not everyone has such skills. Therefore, we propose a method for using impression factors to provide support in modeling complex three-dimensional objects. To design a shape of models based on impressions, we need to associate the impressions and shape of three-dimensional models. As such, we propose a method for estimating human impressions of an object to be used in supporting such modeling technologies like an expert of CAD.

Normally, we would consider that impressions of an object would be determined holistically based on shape, color, and material, but for this research, we have assumed that the shape as a dominant factor in determining the impressions of objects based on a method from Tobitani et al. [1]. We also defined three types of impressions of objects as particularly important: "hard–soft," "flashy–sober," and "stable–unstable."

Taguchi et al. [2] calculated the relationship between features and impressions using multiple regression analysis. However, it can be used only for the "hard–soft" impression factor because the features have been optimized for a specific impression. Designing which features and classifiers to optimize for various impressions is difficult. In this research, we propose a method to estimate impressions of "hard–soft," "flashy–sober," and "stable–unstable" automatically using a deep neural network (DNN) for only the shape of a three-dimensional model. Two types of convolutional network (CNN) are used currently in the field of object classification. One type is multi-view architectures, while the other is volumetric architectures. Qi et al. [3] analyzed the challenges of object classification on three-dimensional data using these two types of CNNs. As a result, the multi-view CNN architectures were proven to be more accurate than volumetric in object recognition. In addition, a person can only recognize an object from a certain viewpoint, one at a time. Therefore, we use a multi-view CNN architecture to estimate impression factors in this method.

The remainder of this paper is organized as follows. The quantification of impression factors to the objects is described in Sect. 2. The proposed method to estimate the impression factors is presented in Sect. 3. Experimental results and analysis are provided in Sect. 4. Finally, the conclusion and discussion are given in Sect. 5.

Table 1 Various bipolar adjectives (impression factors)

Bipolar adjectives (impression factors)	
ordered–chaotic	connected–disconnected
stable–unstable	dynamic–static
active–passive	healthy–unhealthy
excitable–calm	relaxed–tense
soft–hard	smooth–rough
distinct–vague	weak–strong
blunt–sharp	intense–mild
delicate–rugged	cheerful–cheerless
flashy–sober	heavy–light

2 Quantification of Impression Factors

The relationship between impression factors and three-dimensional shape needs to be clarified in order to enable intuitive manipulation of three-dimensional shapes using an impression. Tobitani et al. [1] defined the rating scale for three-dimensional shapes using the semantic differential method to quantify sensitive bipolar adjectives (impression factors). The average values of 10 people were calculated in the SD method for every 18 bipolar adjectives. The 18 bipolar adjectives for this experiment are shown in Table 1.

In this experiment, conditions such as background, materials, lighting, and so on were kept constant, and the range of impression coefficients was −3.0 to 3.0. We analyzed using major factor analysis (PFA) and Varimax rotation. According to the literature, the impression coefficients of the three-dimensional shape can be expressed by three bipolar shapes: "hard-soft," "flashy-saw," and "stable-unstable." The experimental dataset was abstract objects manufactured by a professional designer. When generating these three-dimensional shapes, the bipolar adjectives mentioned above were presented as guidelines for shape design. Therefore, we focused on these most relevant impression factors for the three-dimensional shapes.

3 Impression Estimation Method

The flow of our method is shown in Fig. 1.

Our method uses a deep learning convolutional network with multi-viewpoint images to estimate human impressions of three-dimensional object. This method includes both the learning and estimation modules. The method is assumed to have supervised learning, so the three-dimensional models and ground-truth (impression factors) need to be made to correspond. In the learning module, we used the semantic differential (SD) method [4] to estimate values for the impressions of each object, and we used these values as supervisory signals. Incidentally, they are based on the

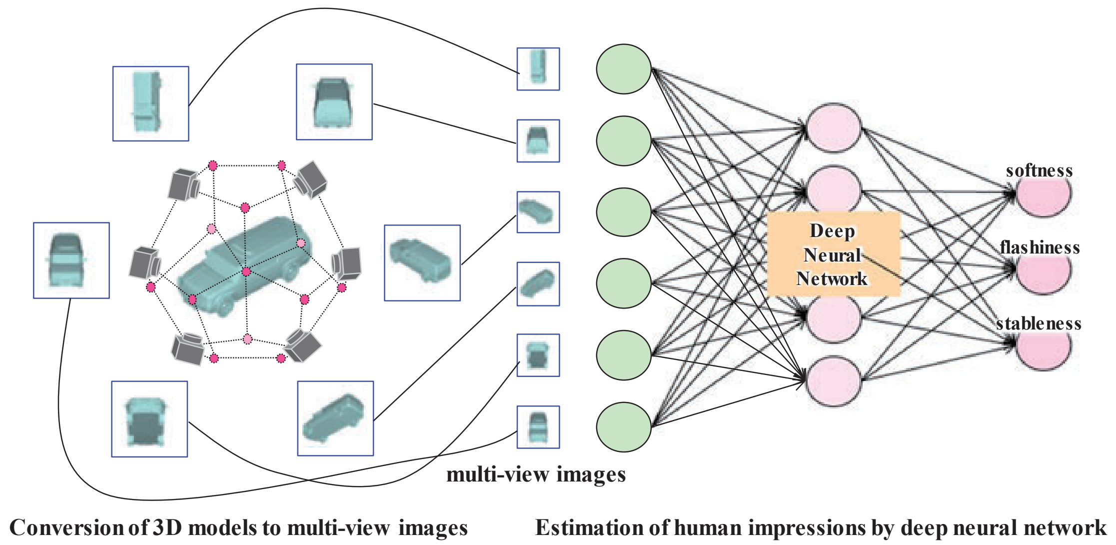

Fig. 1 Flow of our method for estimating impression factors

contents of Sect. 2. The input data for the network were the multiple images taken across viewpoints. Figure 2 shows an example of virtual viewpoints set around a three-dimensional object.

The multi-view images represent three-dimensional shapes using multiple views of three-dimensional objects that were generated by rendering. The virtual viewpoints of plurality were installed around the objects. In this study, the input data consisted of a set of images of the objects taken from 20 directions spaced equally around them. Therefore, the input data were grayscale images taken from 20 viewpoints. Here, an area in which a three-dimensional model did not exist was defined as the background and supplemented with a numerical value. The network consisting of impression estimations was achieved using this image set as the input for deep learning and the training to minimize the difference between the estimated results and the supervisory signals. However, in our method, there is a scale invariance due to transformation of

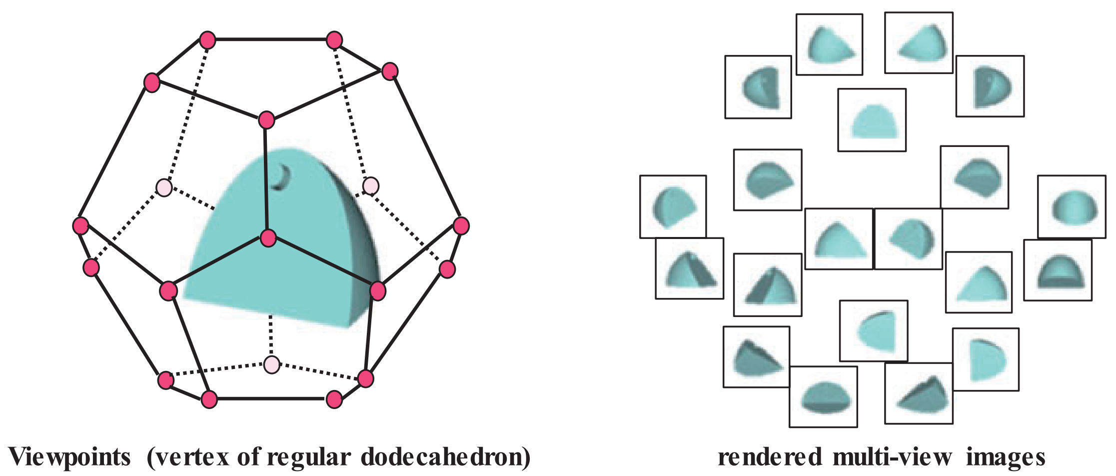

Fig. 2 Virtual viewpoints set around a three-dimensional object and the rendered images

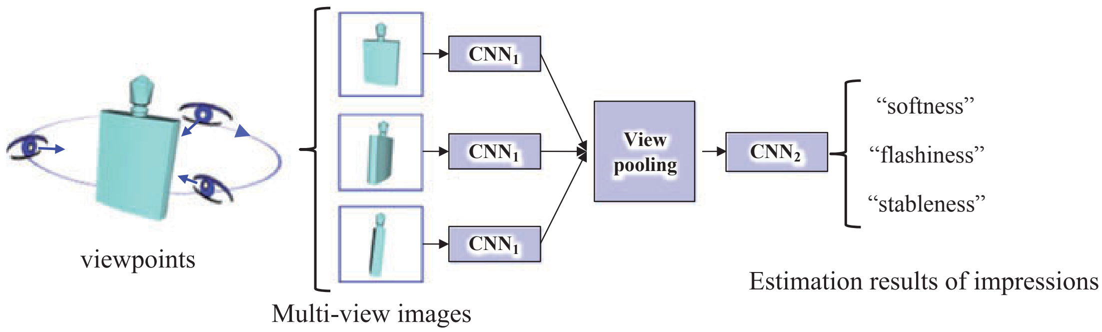

Fig. 3 Constitution of the network

three-dimensional objects to multi-view images. Therefore, we unified the size of the three-dimensional object presented to the subject when quantifying the impression of a person with respect to a three-dimensional object by SD method [4]. The network is shown in Fig. 3.

Our method consists of two stages: the first stage is with five convolutional layers and three pooling layers and the second stage is with a view-pooling layer, which integrated the multi-viewpoint image, and three fully connected layers. The error function used was softmax, and the optimization method used was Adam [5].

4 Experimental Results of Impression Estimations

In our experiments, we used the SD method [3] to associate supervisory signals with three-dimensional objects. For the objects, we used ModelNet40 [6] in specific classes: a car, vase, and chair. Each three-dimensional object was evaluated by 20–40 people, and the −3 to 3 was used for the supervisory signals. In the experiments, the proposed method was evaluated using correlation coefficients. The experimental datasets were evaluated using 20 three-dimensional objects randomly. Figure 4 shows two distributions of relationships between the estimated results and the supervisory signals. The blue dots represent the results of the previous method, and the red dots

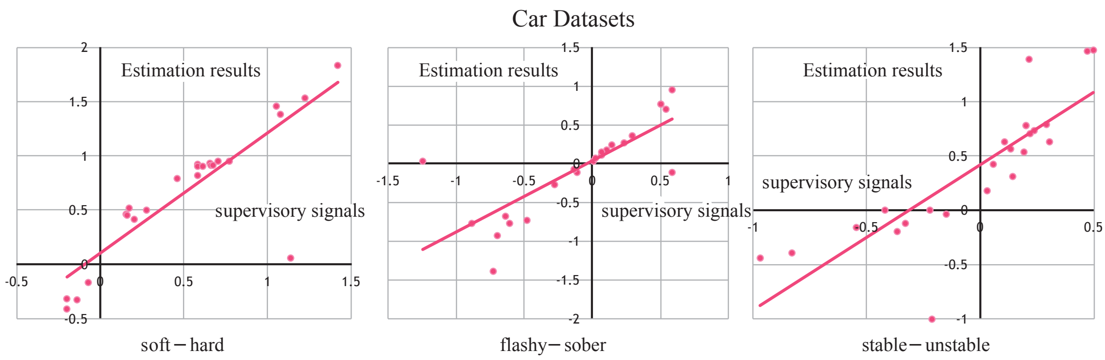

Fig. 4 Relationships between the estimated results and supervisory signals

Table 2 Correlation coefficients between estimation results and supervisory signals

Datasets	Correlation coefficients for impressions		
	Hard-soft	Sober-flashy	Unstable-stable
Car	0.70	0.61	0.67

represent the results of our new method. Also, Table 2 shows the estimation results and correlation coefficients of the supervisory signals.

The supervisory signal in Fig. 4 is an average value evaluated by 20–40 people. The correlation coefficients between impressions estimated using this method and the supervisory signals for all the datasets were about 0.70 for "hard–soft," about 0.61 for "flashy–sober," and about 0.67 for "stable–unstable." As a result, the proposed method confirmed a strong correlation with each impression. When we analyzed three-dimensional objects with large difference between estimate results and supervisory signals, the variation in supervisory signals for the three-dimensional object was large.

5 Conclusion

We proposed a method for automatically estimating three typical human impression factors, "hard–soft," "flashy–sober," and "stable–unstable," which were obtained from objects by analyzing the shapes of three-dimensional models. The results of the experiment show that the correlation coefficients between the impressions estimated using this method and the supervisory signals for the original datasets of car used were about 0.70 for "hard–soft," about 0.61 for "flashy–sober," and 0.67 for "stable–unstable."

In future work, we will clarify the impression structure of people and estimate using the RGB image, because our method cannot consider the texture and material.

Acknowledgments This research was partially supported by the Center of Innovation Program from the Japan Science and Technology Agency, JST.

References

1. K. Tobitani, S. Akizuki, K. Katahira, M. Hashimoto, N. Nagata, A comparison study on 3D features in terms of effective representation for impression of shape, in *The 2nd International Conference on Digital Fabrication*, No. 22 (2016)
2. K. Taguchi, K. Sasaki, M. Hashimoto, K. Tobitani, N. Nagata, A proposal of 3D local feature for estimating human's impression factor to shape of object, in *International Workshop on Advanced Image Technology* (2017)
3. C.R. Qi, H. Su, M. Niesner, A. Dai, M. Yan, L.J. Guibas, Volumetric and multi-view CNNs for object classification on 3D data, in *IEEE Conference on Computer Vision and Pattern Recognition* (2016)

4. C.E. Osgood, G.J. Suci, P.H. Tannenbaum, *The Measurement of Meaning* (University of Illinois Press, Oxford, England, 1957)
5. D.P. Kingma, Adam: a method for stochastic optimization, in *International Conference on Learning Representations* (2015)
6. Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, J. Xiao, 3D ShapeNets: a deep representation for volumetric shapes, in IEEE *Conference on Computer Vision and Pattern Recognition* (2016), pp. 1912–1920

Mr. Koichi Taguchi received his B.E. degree from Chukyo University, Aichi, Japan, in 2017. He is now a graduate student of the university and is very interested in technologies about object recognition and its application to sensitivity information processing. He is a member of the Japanese societies of IPSJ Japan and JSPE.

Manabu Hashimoto received his B.E. and M.E. degrees from Osaka University, Osaka, Japan, in 1985 and 1987. He joined the Mitsubishi Electric Corporation in 1987, and he has been engaged in research on robot vision, image recognition, pattern recognition, and human sensing in the Manufacturing Development Laboratory, Industrial Electronics and Systems Development Laboratory, and the Advanced Technology R&D Center of the company. He received his Ph.D. degree from Osaka University in research on 3-D object recognition in 2000. Since 2008, he has been a professor of Chukyo University, and he is the Dean of the School of Engineering. He received the Technical Innovation Award in 1998 from the Robotics Society of Japan, the Excellent Academic Award in 2012 from the Symposium on Sensing via Image Information, and the Best Paper Award of IWAIT2017 and IWAIT2018. He is a member of the IEEE and Japanese societies of IEICE, IPSJ, IEEJ, and RSJ, among others.

Kensuke Tobitani received his doctoral degree in engineering from Gifu University, Japan, in 2009. He is currently a lecturer at Kwansei Gakuin University. His research interests include computer vision, computer graphics, and Kansei information processing.

Noriko Nagata received her BS degree in mathematics from Kyoto University in 1983 and her Ph.D. degree in systems engineering from Osaka University in 1996. She was a researcher at the Industrial Electronics and Systems Laboratory of Mitsubishi Electric Corporation from 1983 to 2003. She joined Kwansei Gakuin University in 2003 as an associate professor. She is currently a professor in the Department of Human System Interaction and a director of the Research Center for Kansei Value Creation. In 2009, she was a visiting scholar at Purdue University. Her research interests include Kansei (affective) information processing, computer graphics, and multimedia systems. She is a member of the IEEE and ACM.