https://doi.org/10.2352/ISSN.2470-1173.2018.13.IPAS-194

An Estimation Method of Human Impression Factors for Objects from their 3D Shapes Using a Deep Neural Network

Koichi Taguchi, Manabu Hashimoto; Chukyo University; Aichi, Japan
Kensuke Tobitani, Noriko Nagata; Kwansei Gakuin University; Hyogo, Japan

Abstract

In this paper, we propose a method for automatically estimating three typical human-impression factors, "hard-soft", "flashy-sober", and "stable-unstable" which are obtained from objects by analyzing their three-dimensional shapes. By realizing this method, a designer's will in directly shaping an object can be reflected during the design process. Here, the focus is highly correlating human impressions to the three-dimensional shape representation of objects. Previous work includes a method for estimating human impressions by using specially designed features and linear classifiers. However, it can be used for only the "hard-soft" impression factor because the feature has been optimized for this impression. The performance of this method is relatively low, and its processing time is low. In addition to, there is a serious problem in which this method can be used for only a particular impression factor. The purpose of this research is to propose a new method that can apply to all three typical impression factors mentioned above. First, we use a single RGB image that was acquired from a specific view direction instead of general three-dimensional mesh data from the range finder. This enables a very simple system consisting of a single camera. Second, we use a deep neural network as a nonlinear classifier. For our experiment, a lot of learning sample images with numerical human-impression factors were used. As for annotating correct impression factors as ground-truths, we utilized the SD (semantic differential) method, which is very popular in the field of psychological statistics. We have shown that the success rate of the proposed method is 83% for "hard-soft", 78% for "flashy-sober", and 80% for "stable-unstable" when using test images that are not included in the learning data.

Introduction

Recently, various kind of three-dimensional printers have been developed and they became cheaper than before[1]. Accordingly, we can use them very easily not only in a factory but also in our home[2]. Furthermore, useful database of three-dimensional models has been prepared and everyone can utilize it via Internet[3]. It is expected that "personal fabrication(production using three-dimensional printers, etc., by individuals)" will change traditional way of manufacturing such as mass production (at factories). In order to realize "personal fabrication", skillful modeling techniques using CAD systems are required to make three-dimensional objects that you like[4, 5]. However, not everyone has skillful modeling techniques. Therefore, we propose a method for using impression factors to support complex three-dimensional modeling[6]. An example of manufacturing based on impressions for realizing personal fabrication using 3D printers is shown in figure 1.

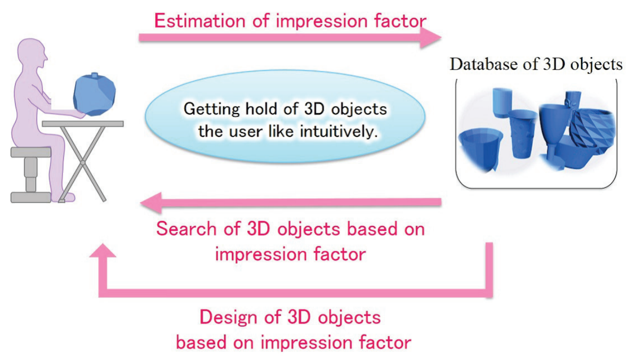

Figure 1. *An example of manufacturing based on impressions for realizing personal fabrication using 3D printers*

In order to design an shape of models based on impressions, we need to associate impressions and shape of three-dimensional models. Originally, it seems that the impression factors of a person with respect to an object are determined comprehensively according to shape, color, material, etc. In this research, we assume that the most dominant factor in determining impression factors is shape. Based on the previous method[7], we focus on the most relevant impression factors to the objects: "hard-soft", "flashy-sober", and "stable-unstable". Taguchi et al[8]. calculates relationship between feature and impressions by multiple regression analysis. However, it can be used for only the "hard-soft" impression factor because the feature has been optimized for a specific impression. It is difficult that designing which feature and classifiers optimized for various impression. In this research, we propose a method to estimate automatically impressions of "hard-soft", "flashy-sober", "stable-unstable" using Deep Neural Network(DNN) to only shape of a three-dimensional model.

Recently, there are roughly Convolutional Network Network (CNN) of two types in the field of object classification. One of CNN is the multi view architectures while another is the volumetric architectures. Charles et al[9]. analyzed challenges of object classification on three-dimensional data by using CNNs of two types. As a result, it is proved that multi view CNN architecture is more accurate than volumetric CNN architecture in object recognition. In addition, a person can only recognize an object from a certain viewpoint one by one. Therefore, we use a multi view CNN architecture to estimate impression factors in this method.

The remainder of this paper is organized as follows. Related works are described in Section 2. Quantification of impression factor to the objects is described in Section 3. The proposed method to estimate impression factor is presented in Section 4. Experimental results and analysis are provided in Section 5. Finally, the conclusion and discussion are given in Section 6.

Related Work

Shape Descriptors to Estimate Impression Factor

In order to design an shape of models on the basis of impression factors, we need to associate impression factors and shape of three-dimensional models. Taguchi el al[8]. focused on the distribution of local curvature as a feature. Comparing the overall roundness of " hard " impression of objects and " soft " one , we found that they are clearly different. Figure 2 shows the shape descriptors to estimate impression factor.

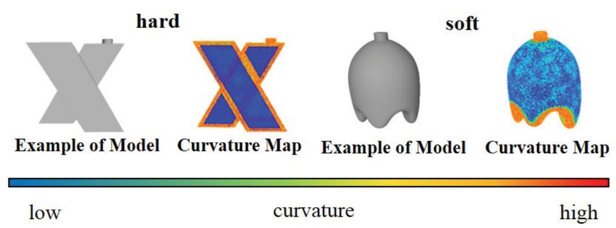

Figure 2. *Example of feature focusing on the overall curvature of the three-dimensional point cloud*

However, it can be used for only the " hard-soft " impression factor because the feature has been optimized for a specific impression. In addition, human labor is increased because the time which required for visual analysis and the trial number is increased in proportion to the number of objects. Therefore, it is difficult that designing which feature and classifiers optimized for various impression.

Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are one of the machine learning architecture. This architecture demonstrates performance by learning a lot of RGB images. It is mainly used in the field of computer vision. In the recent, it is applied not only to object detection, scene recognition, but also to text and image generation, material recognition and various fields[9-17].In addition, they become obvious that they are able to learn better performance than handcraft feature in the tasks of computer vision. Therefore, we speculate that applications to estimation of impression factors are expected.

CNN on 3D Data

In the recent, new CNNs architecture to process three-dimensional data is proposed. They are roughly divided into two categories. Figure 3 shows the approach of voxel CNN and MVCNN. One is a method using three-dimensional voxel grid, and the other is a method using multi-view images.

Wu et al[10]. propose 3DShapeNets to represent a geometric three-dimensional shape as a probability distribution of binary variables on a three-dimensional voxel grid (voxel CNN).

On the other hand, Su et al[11]. proposed CNN architecture to estimate three-dimensional representation by using 2-dimensional images of multiple rendered views (MVCNN). This approach is a CNN that can collaborate on all rendered views.

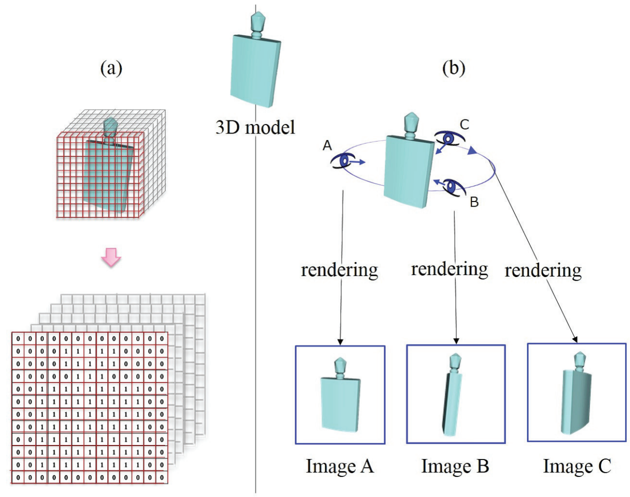

Figure 3. *Example of Convolutional Network Networks（CNNs）of two types on three-dimensional data : (a) using 3D voxel grid like 3DShapeNets, (b) using 2-dimensional images of multiple rendered views*

Charles et al[9]. analyzed challenges of object classification on three-dimensional data by using CNNs of two types. As a result, it is proved that multi view CNN architecture is more accurate than volumetric CNN architecture in object classification. In addition, a person can only recognize an object from a certain viewpoint one by one. In order to perceive the three-dimensional shapes, people confirm the object from plurality of viewpoints. Therefore, we assume that MVCNN is more appropriate than voxel CNN to estimate the impression factors.

Quantification of Impression Factor

In order to realize intuitive manipulation of three-dimensional shapes by an impression, the relation between impression factors and three-dimensional shape need to be clarified. In order to quantify sensitive bipolar adjectives (impression factor), Tobitani et al [7]. defined the rating scale for three-dimensional shapes by using semantic differential method. The average value of 10 people was calculated in the SD method per one of the 18 bipolar adjectives. The 18 bipolar adjectives for this experiment is shown in 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

That time, the conditions such as background, material and lighting are kept constant and range of impression factor is -3.0 to

3.0. They were analyzed by a principal factor analysis (PFA) and varimax rotation. It has been suggested that impression factors to three-dimensional shapes are represented by three bipolar adjectives about "hard-soft", "flashy-sober", "stable-unstable". Therefore, we focus on the most relevant impression factors to the three-dimensional shapes: "hard-soft", "flashy-sober", and "stable-unstable".

In addition, experimental datasets are three-dimensional objects of cosmetic bottle that professional designers produced. When they produced the three-dimensional shapes, three bipolar adjectives ("stable-unstable", "flashy-sober", "soft-hard") were presented as guidelines for designing shapes. Example of three-dimensional shapes of cosmetic representing three bipolar adjectives is shown figure 4.

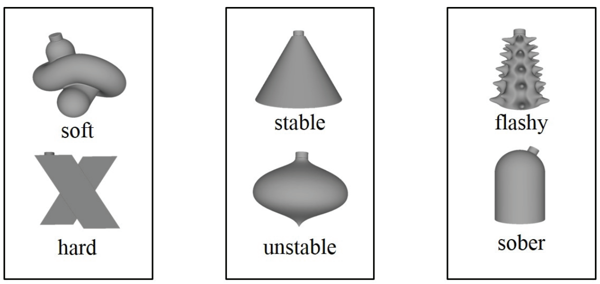

Figure 4. *Example of three-dimensional shapes of cosmetic bottle representing three bipolar adjectives(impression factors)*

A total of 90 three-dimensional objects granted impression factor were produced by them.

Method to Estimate Impression Factor

This paper focuses on a method to estimate the impression factors on three-dimensional models. We propose a network that estimates impression factors using MVCNN. The flow of method is shown in Figure 5.

This method is the Impression factor module, the learning module, and the estimation module. This method is assumed to have supervised learning, it is necessary to correspond the three-dimensional models and ground-truth (Impression factors).

In the Impression factor module, at first, we set the impression factors for each object as a ground-truth by using the SD method. Incidentally, these are based on the contents which described in Section 3.

In the learning module, the multiple images taken from multiple viewpoints are used. They represent three-dimensional shapes by using multiple views of three-dimensional objects which was generated by rendering. The virtual viewpoints of plurality are installed around the three-dimensional objects. In this study, we create 20 rendered views (multiple images) by placing 20 virtual cameras around the three-dimensional objects. Therefore, the input data are RGB images taken from 20 viewpoints. Here, an area in which the three-dimensional model does not exist is defined as a background and supplemented with a numerical value. In addition, online databases of three-dimensional models normally are stored as polygon meshes which are collections of points. Since the proposed method does not need a three-dimensional point cloud, acquisition of data is easy. Figure 6 shows an example of a virtual viewpoints set around a three-dimensional object.

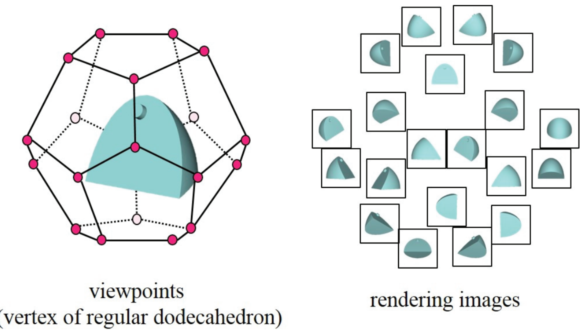

Figure 6. *Virtual viewpoints(vertex of regular dodecahedron) set around a three-dimensional object and its rendering image*

In this method, a multi-view image is generated by placing 20 virtual cameras at 20 vertices of a dodecahedron surrounding a shape. The color image to be input to the DNN was photographed from the specific viewpoint in the direction of the center of gravity of the object with respect to each three-dimensional object in which the impression was set. As a result, multiple two-dimensional images descriptors are obtained for one three-dimensional shape per view, so some form of integration is required for recognition tasks. We treat the photographed image as the same impression as the impression-set three-dimensional object, and estimate the impression on the object shape by learning with DNN. These integration and network configuration uses the method of Su et al[11].

In the estimation module, a feeling impression is estimated by using a multi-viewpoint image generated from a three-dimensional model in the network generated by the learning module. The process of generating an image is the same as that of the learning module.

Exprements

Dataset of the Experiment

The object used for this experiment is shown in Figure 7.

This dataset consists of the objects which is shown in Section 2. Therefore, each object has impression factor ("hard-soft", "flashy-sober", "stable-unstable") in advance. This dataset consists of a total of 90 the objects. Incidentally, these objects are motifs of cosmetic bottles. We consider that their objects using in this study is appropriate because bottles are that user carry always and each user have different preferences to shape.

The Result of Experiments

At first, multi-viewpoint image was taken from the vertex (virtual viewpoint) of the regular dodecahedron toward the center of gravity of each object. In other words, 20 images taken from different viewpoints for each object are generated. The taken multi-viewpoint image was classified into learning data and test data for each object. The learning data consists of 50 objects and test data consists of 40 objects.

As an experiment, each impression factor was estimated by three methods using these data. We compared multi-view

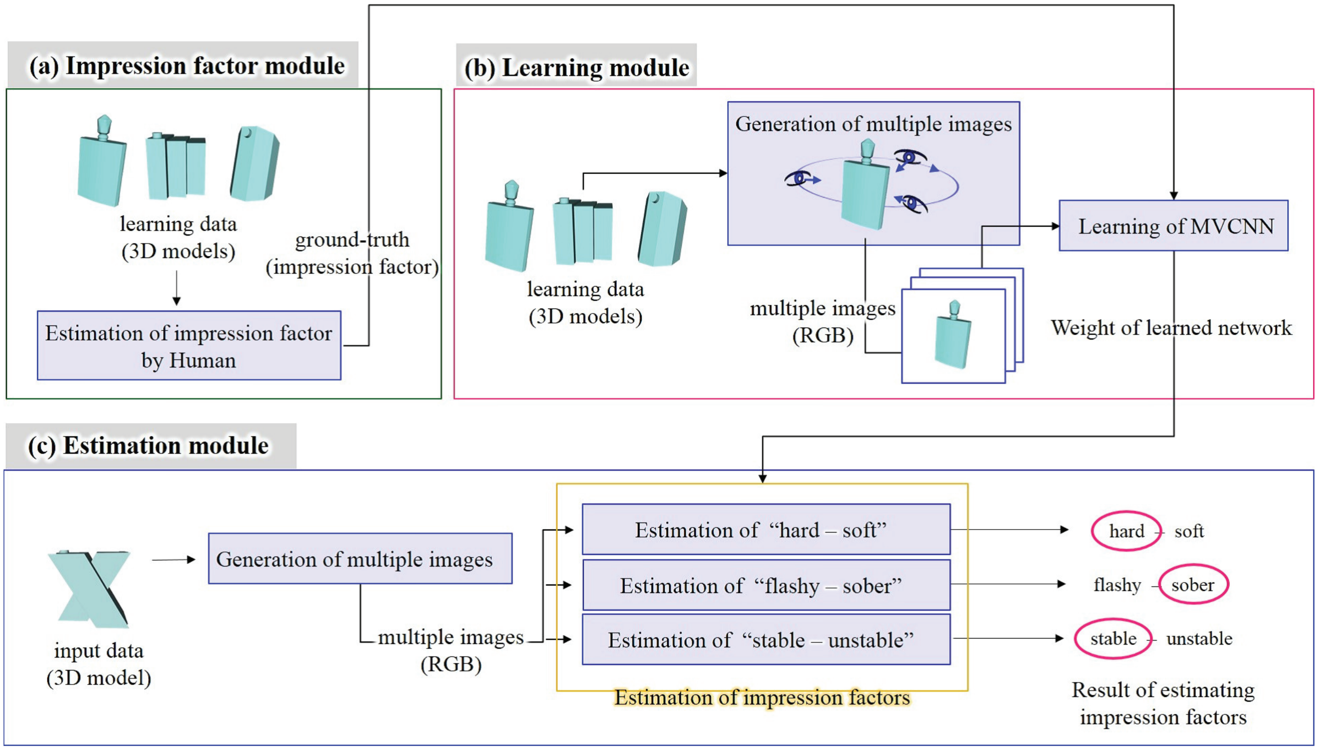

Figure 5. *Proposed method's flow to estimate impression factors: (a) In impression factor module, correspondence the three-dimensional models and ground-truth (Impression factors) by human (b) In learning module, Learning of impression factors and multiple images (c) In estimation module, estimation of impression factors on unlearned data*

CNN and Taguchi[8], VGGNet[12] for estimating impression factor.The method of Taguchi[8] estimates by designing heuristic feature to each impression factor and using multiple regression analysis.

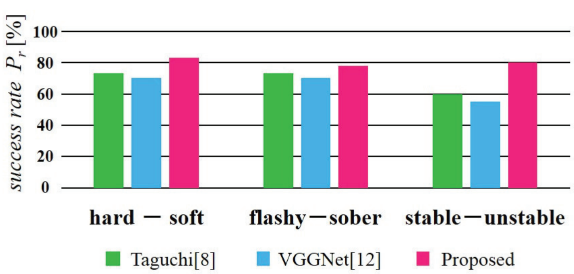

Figure 8. *Success rate of each impression factor by using each method*

Success rate of each impression factor by using each method

Method	Impression factors		
	hard-soft[%]	flashy-sober[%]	stable-unstable[%]
Feature[8]	73	73	60
VGGnet[12]	70	70	55
Proposed Method	83	78	80

VGGNet[12] is one of the CNNs. In this experiment, VGGNet learn and classify one by one of the multi-view images for estimating each impression factor. In this experiment, the success rate shows the coincidence rate between the results of learning the deep neural network using the training data with the ground truth and outputting the impression factors to an unknown object in two stages and the ground truth. Table 2 and figure 8 shows the success rate of each impression factor by using each method.

As shown in Table 2 and Figure 8, the success rate of each impression factor by using our method is confirmed that " hard-soft " is 83%, " flashy-sober " is 78% and " stable-unstable " is 80% in test data. The success rate of multi-view CNN method was about 80%. It improved about 12% and better than proposed method, single-view CNN method. Thereby, we confirmed usefulness of using multi-view images to estimate impression factors. In addition, the estimating success rate of each impression factor by using our method is confirmed that "hard-soft" is 98%, "flashy -sober " is 98% and " stable-unstable " is 98% in learning data.

Conclusion

We proposed a method for automatically estimating three typical human-impression factors, " hard-soft, " " flashy-sober, " and " stable-unstable, " which are obtained from objects by analyzing their shapes of three-dimensional models. In results of the experiment, we confirmed that success rate of impression factors is " hard-soft " is 83%, " flashy-sober, " is 78%, and " stable-unstable " is 80% by using multi-view CNN method. The success rate of multi-view CNN method was about 80%. It improved

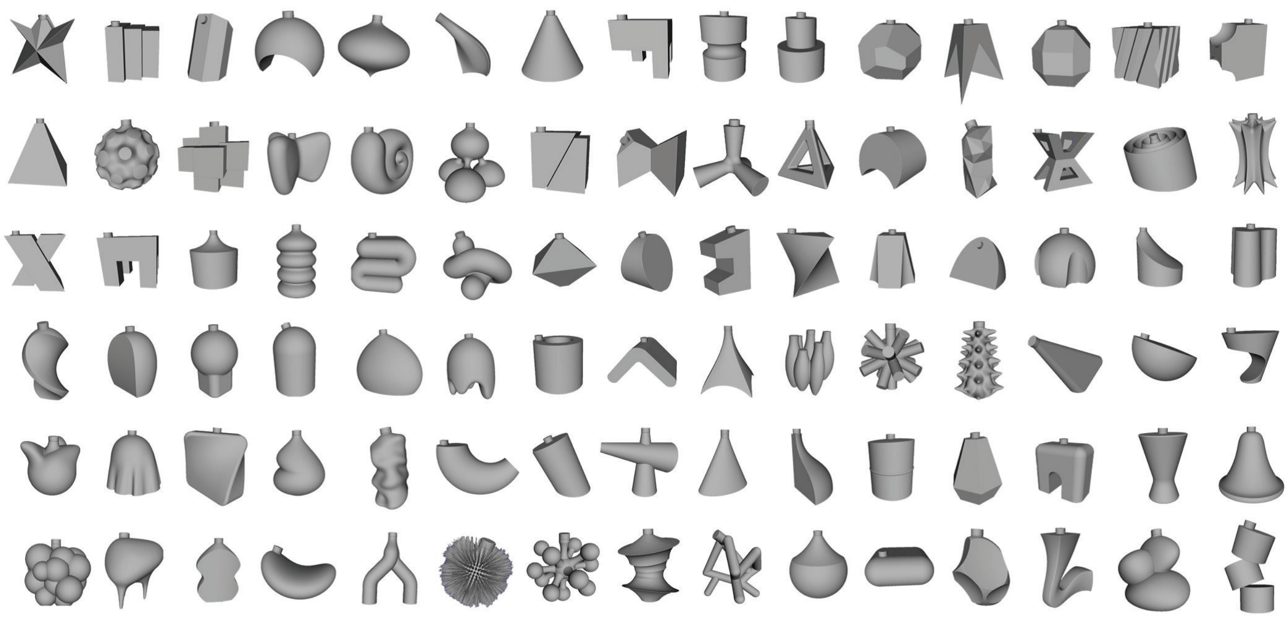

Figure 7. *Experimental datasets of three-dimensional models of cosmetic bottle that professional designers produced*

about 12% and better than proposed method, single-view CNN method.

In the future work, we are going to analysis associate the appearances of three-dimensional models from view points and impression factors. Thereby, we will believe that their clarification is the trigger to reflect impression factors during the design process. In addition, we are going to create many three-dimensional models which granted impression factors.

Novelty

The novelty of this paper was to associate representative impression factors such as "hard-soft", "flashy-sober", and "stable-unstable " for an object with the three-dimensional object shape.

Acknowledgments

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

References

[1] S. Mueller, Y. Ochiai, L. Devendorf, M. Gannon, S. Coros, P. Baudisch, " CrossFAB: Bridging the Gap between Personal Fabrication Research in HCI, " Computer Graphics, Robotics, Art, Architecture, and Material Science, The ACM Conference on Human Factors in Computing Systems (CHI), pp. 3431-3437, 2016.

[2] G. S. Morlock, " Personal Fabrication, " Communications of the ACM, Vol. 53 No. 10, pp. 14-15, 2010.

[3] Y. Shen, K. Dou, J. Gu, " RoCuModel An Iterative Tangible Modeling System, " International Conference on Tangible, Embedded and Embodied Interaction (TEI), pp. 73-76, 2014.

[4] C. Weichel, M. Lau, D. Kim, N. Villar, H. Gellersen, " MixFab: A Mixed-Reality Environment for Personal Fabrication, " The ACM Conference on Human Factors in Computing Systems (CHI), pp. 3855-3864, 2014.

[5] P. Baudisch, S. Mueller, "Personal Fabrication State of the Art Future Research, " The ACM Conference on Human Factors in Computing Systems (CHI), pp. 936-939, 2016.

[6] J. Rajapakse, H. Jayasinghe, Y. Tokuyama, K. Miyata, A. Marasinghe, " The Effect Variable Base on Perceiving Gender Kansei of 3D-shape, " International Conference on Biometrics and Kansei Engineering(ICBAKE), pp. 136-14, 2009.

[7] K. Tobitani, S. Akizuki, K. Katahira, M. Hashimoto, and N. Nagata, " A Comparison Study on 3D Features in Terms of Effective Representation for Impression of Shape," The 2nd International Conference on Digital Fabrication (ICDF), No. 22, 2016.

[8] Koichi Taguchi, Kosuke Sasaki, Manabu Hashimoto, Kensuke Tobitani, Noriko Nagata, " A proposal of 3D local feature for estimating human' s impression factor to shape of object, " International Workshop on Advanced Image Technology(IWAIT), 2017.

[9] Charles R. Qi, Hao Su, Matthias Niesner, Angela Dai, Mengyuan Yan, Leonidas J. Guibas, " Volumetric and Multi-View CNNs for Object Classification on 3D Data, " IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2016.

[10] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, Jianxiong Xiao, " 3D ShapeNets : A Deep Representation for Volumetric Shapes, " IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp. 1912-1920, 2016.

[11] Hang Su, Subhransu Maji, Evangelos Kalogerakis, Erik Learned-Miller, " Multi-view Convolutional Neural Networks for 3D Shape Recognition, " the International Conference on Computer Vision(ICCV), pp. 945-953, 2015.

[12] Karen Simonyan, Andrew Zisserman, " VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE SCALE IMAGE RECOGNITION, " the International Conference on Learning Representations(ICLR), pp.1-14, 2015.

[13] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, " ImageNet Classification with Deep Convolutional Neural Network, " Vol. 1,

pp.1097-1105, 2012.
[14] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich ,"Going Deeper with Convolutions," IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp. 1-9, 2015.
[15] Yan Zhang, Mete Ozay, Xing Liu and Takayuki Okatani, "Integrating Deep Features for Material Recognition, " International Conference on Pattern Recognition(ICPR), 2016.
[16] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, " Generative Adversarial Nets, " International Conference on Neural Information Processing Systems(NIPS),Vol. 2, pp. 2672-2680, 2014.
[17] Andrej Karpathy, Li Fei-Fei, " Deep Visual-Semantic Alignments for Generating Image Descriptions, " IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Vol. 39, pp. 664-676, 2015.

Author Biography

Mr. Koichi Taguchi received the B.E. degrees from Chukyo University, Aichi, Japan, in 2017.Now he is 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 Japanese societies of IPSJ Japan and JSPE.

Manabu Hashimoto received the B.E. and M.E. degrees from Osaka University, Osaka, Japan, in 1985 and 1987 respectively. He joined Mitsubishi Electric Corporation in 1987, and he has been engaged in research on robot vision,image recognition, pattern recognition, and human sensing in Manufacturing Development Laboratory,Industrial Electronics and Systems Development Laboratory, and Advanced Technology R&D Center of the company. He received 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 School of Engineering. He received the Technical Innovation Award in 1998 from the Robotics Society of Japan, Excellent Academic Award in 2012 from Symposium on Sensing via Image Information, and the Best Paper Award of IWAIT2017 and IWAIT2018. He is a member of IEEE and Japanese societies of IEICE, IPSJ, IEEJ, RSJ, and 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 the BS degree in mathematics from Kyoto University in 1983 and the PhD 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 IEEE and ACM.