https://doi.org/10.2352/ISSN.2470-1173.2019.6.MAAP-204

Constructing glossiness perception model of computer graphics with sounds

Takumi Nakamura, Keita Hirai, Takahiko Horiuchi;
Department of Imaging Sciences, Chiba University; 1-33, Yayoi-cho, Inage-ku, Chiba-shi, Chiba, 263-8522 Japan.

Abstract

In this paper, we construct a model for cross-modal perception of glossiness by investigating the interaction between sounds and graphics. First, we conduct evaluation experiments on cross-modal glossiness perception using sounds and graphics stimuli. There are three types of stimuli in the experiments. The stimuli are visual stimuli (22 stimuli), audio stimuli (15 stimuli) and audiovisual stimuli (330 stimuli). Also, there are three sections in the experiments. The first one is a visual experiment, the second one is an audiovisual experiment, and the third one is an auditory experiment. For the evaluation of glossiness, the magnitude evaluation method is applied. Second, we analyze the influence of sounds on glossiness perception from the experimental results. The results suggest that the cross-modal perception of glossiness can be represented as a combination of visual-only perception and auditory-only perception. Then, based on the results, we construct a model by a linear sum of computer graphics and sound parameters. Finally, we confirm the feasibility of the cross-modal glossiness perception model through a validation experiment.

Introduction

In recent years, real computer graphics (CG) reproduction techniques have been developed in the research field of CG. Previous CG researches have realized realistic reproduction such as color appearance and shading based on the physical (optical) laws in the real world. Researches that reproduce not only color but also “shitsukan (i.e., material appearance / material perception)” of objects has been actively investigated. Researches to investigate the relationship between the physical laws and human shitsukan perception are achieved for realistically reproducing the shitsukan such as transparency and unevenness on the object surface on the CG. For examples, Kohko et al. [1] compared the actual material perception with the CG shading model such as the Phong reflection model and investigated how much material perception can be expressed using the reflection model. Pellacini et al. [2] examined the relationship between the physical parameters and the perceptual glossiness of the object through experiments. They also proposed a new light reflection model in CG. Ana et al. [3] reproduced material perception using a relationship between visual sensations and physical phenomena.

Concerning the shitsukan perception, humans often recognize an object and texture based on cross-modal sensations from not only visual information but also combinational information such as haptic senses when touching an object (tactile information), and auditory senses when hitting an object (sound information), etc. Fujisaki et al. conducted an experiment of the shitsukan cognition for investigating the cross-modal material recognition using visual and auditory senses [4]. They also investigated shitsukan cognition based on three senses of visual, auditory and tactile sense [5]. Rodrigo et al. [6] discovered that cross-modal shitsukan recognition was more richly recognizable compared with the recognition with one specific sense. Bonneel et al. [7] clarified perceptual details by investigating the interaction between visual and auditory regarding material similarity. They also applied their findings to CG rendering for reducing the rendering cost and improving the level of detail (LOD) of sound rather than one of a video. However, these studies of cross-modal shitsukan perception have been mainly focused on the interpretation of the relationship between material perception and brain mechanism. For reproducing shitsukan of CG, it is necessary that not only visual reproduction but also cross-modal reproduction by considering auditory sense and haptic sense are important. However, researches on them have not been investigated enough.

In this study, we generate a model of the interaction between sound effects and CG related to glossiness. In the shitsukan perception, glossiness is one of the essential components in CG rendering. Thus, we decided to treat glossiness perception in this research. We first carry out cross-modal glossiness evaluation experiments using visual, auditory and audiovisual stimuli for investigating CG and sound effects. Next, we analyze the influence of sound effects on glossiness. Based on the experimental results, we model the cross-modal glossiness and validate the feasibility of the model.

Subjective Evaluation Experiments

We prepared three types of experimental stimuli: (1) visual stimuli, (2) auditory stimuli, and (3) audiovisual stimuli. Two experiments were conducted by changing two parameters of visual stimuli. One of the parameters is a roughness (i.e., a roughness of specular reflection of object surface), and the other is specular to diffuse ratio (i.e., a ratio of diffuse reflection and specular reflection of reflected light of an object). The experiments consisted of three sections: (1) visual experiment (evaluation of 11 visual stimuli × 2 parameters), (2) audiovisual experiment (evaluation of 165 visual-auditory combined stimuli × 2 parameters), and (3) auditory experiment (evaluation of 15 auditory stimuli × 2 parameters). The parameter-changed experiments were separately conducted.

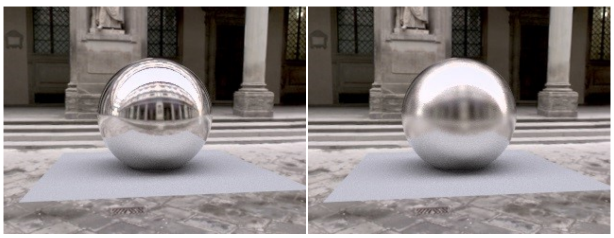
Stimuli

For the visual stimuli, we prepared CG images which were rendered with the PBRT [8]. PBRT is a rendering software that can create high-quality CG by the ray tracing method. In this experiment, the shape of the object, the color of the object, the position of the illumination and so on were all fixed, and a CG image based on the reflection characteristic of silver was rendered. The reflection characteristics of the object surface of CG were expressed using the Torrance-Sparrow model [9]. The Torrance-Sparrow model is a dichroic reflection model that describes the reflection of the surface with a linear combination of diffuse reflection and specular reflection.

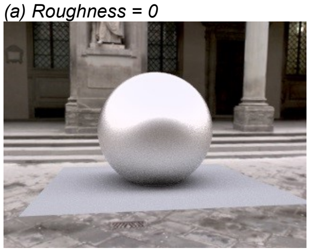
As a parameter of CG, we selected the roughness and the specular to diffuse ratio. For the stimuli with the roughness changes (the specular to diffuse ratio is fixed to 1.0), we prepared 11 types of CG images whose surface roughnesses were 0.0000, 0.0001, 0.0002, 0.0004 - 0.0256, and 0.0512. Some samples of them are shown in Figure 1. The roughness was given by PBRT. For the stimuli with the specular to diffuse ratio changes (the roughness is fixed to 0.0032), we prepared 11 types of CG images whose specular to diffuse ratio were from 0.0 to 1.0 at the interval of 0.1. Some samples of them are shown in Figure 2. The value of the specular to diffuse ratio shows the specular percentage in reflected lights on an object surface. This stimulus was generated by linearly interpolating CG with only diffuse reflection (Figure 2(a)) and CG with only specular reflection (Figure 2(c)).

As auditory stimuli, we prepared 15 sound effects by collecting free audio materials on the WEB. These auditory stimuli were mainly generated by beating a metal material and included various kinds of sound effects (for example, high sound, low sounds, etc.).

We also prepared 330 audiovisual stimuli by combining the 22 visual stimuli (11 kinds of the roughness change stimuli and 11 kinds of the specular to diffuse ratio change stimuli) and the 15 auditory stimuli.

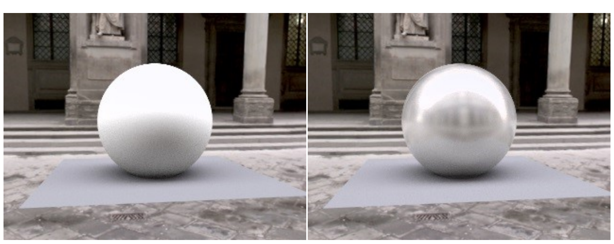


(a) Roughness = 0 *(b) Roughness = 0.0016*

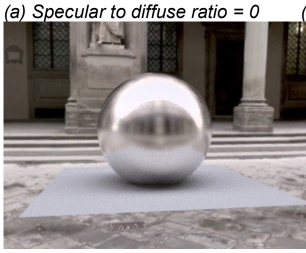


(c) Roughness = 0.0512

Figure 1. A sample of the roughness change stimuli (the specular to diffuse ratio is fixed to 1.0)



(a) Specular to diffuse ratio = 0 *(b) Specular to diffuse ratio = 0.5*



(c) Specular to diffuse ratio = 1.0

Figure 2. A sample of the specular to diffuse ratio stimuli (the roughness is fixed to 0.0032)

Procedure and evaluation method

First, we trained subjects using the dummy evaluation program before the experiment so that they become accustomed to the evaluation process. Next, three sections of the visual experiment (evaluation of 11 visual stimuli), audiovisual experiment (evaluation of 165 audiovisual stimuli), auditory experiment (evaluation of 15 auditory stimuli) were executed in order. The experiments using the stimuli with the roughness changes or with the specular to diffuse ratio changes were conducted separately.

The magnitude estimation method was used to evaluate glossiness. As the reference stimuli, a CG image with the roughness = 0.0032 and only specular reflection (the specular to diffuse ratio = 1.0) was used. The reference image is shown in Figure 3. The sphere (without sound effects) located in Figure 3 was recognized as the criterion of glossy feeling as 50. The evaluation lower limit (when not feeling glossiness at all) is set to 0, and the upper limit is not set.

In the visual experiments, the subjects answered scores on glossiness for 11 visual stimuli. In the audiovisual experiment, the subjects answered glossiness scores by comparing the reference image (visual information only) with the combinational stimuli of CG images and sound effects for the 165 types of audiovisual stimuli. In the auditory experiment, subjects compared the glossiness that can be recalled from sounds with that of the reference and answered scores for the 15 auditory stimuli.

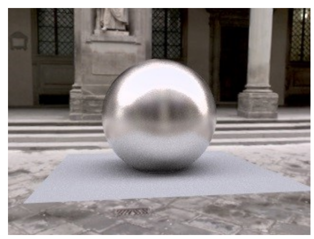


Figure 3. Reference image (roughness = 0.0032)

Environment

Stimuli were presented with monitor (EIZO ColorEdge CG-221 BK) and headphones (audio-technica ATH-MSR 7). The viewing distance is 50 cm, and the viewing angle of the stimuli is approximately 8 degrees. Through all experiments, the number of subjects was twelve (including both male and female with the twenties). Figure 4 shows the experimental environment. The experiment was conducted in a dark room.

Figure 4. Experimental situation

Experimental Results

Figure 5 shows the result of the visual experiment using the roughness. As the roughness increases, the perceptual glossiness becomes smaller. Figure 6 shows the result of the visual experiment using the specular to diffuse ratio. As the specular reflection ratio increases, the perceptual glossiness increases. The relationship between sound and glossiness is shown in Figure 7. Error bars in these figures are the standard deviations between subjects. Also, the horizontal axis of Figure 7 shows the sound feature. In this study, the frequency at peak power in the Fourier domain was used as the sound feature. When the peak frequency of the sound is around 7800 Hz, the glossiness perception to the sound effects becomes the largest. Figure 8 shows the result of the audiovisual experiment with the roughness-change stimuli. As shown in Figure 8, the result of audiovisual experiment has a similar to a combination (or sum) of single-modal results; i.e., visual and auditory experimental results as shown in Figures 5 and 7. Also, Figure 9 shows the results of the audiovisual experiment with the specular to diffuse ratio change stimuli. The result is also similar to a combination (or sum) of single-modal results (i.e., visual and auditory experimental results show in Figures 6 and 7). Then, by normalizing both results, we compare the result by the audiovisual experiment to the sum of the single-modal results by the visual experiment and the auditory experiment. The results are shown in Figure 10. The approximated shapes of both graphs are almost the same (the correlation was 0.89).

From these results, it was suggested that the perception of the glossiness was changed by combining the CG image and sound effects. In other words, by changing the combination of the physical parameters of the CG and the sound effects (the roughness, the specular to diffuse ratio and peak frequency of sound), it is possible to manipulate the cross-modal glossiness perception. Also, as shown in Figure 10, the cross-modal glossiness perception can be modeled by the linear sum of those parameters.

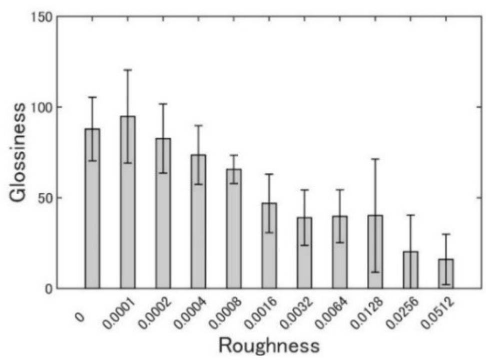


Figure 5. A result of the visual experiment (the roughness change stimuli)

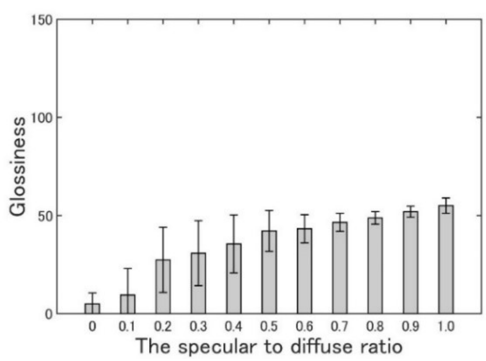


Figure 6. A result of the visual experiment (the specular to diffuse ratio change stimuli)

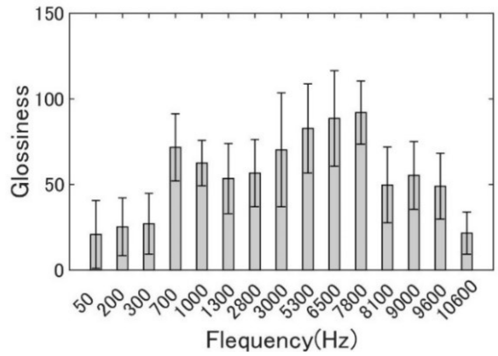


Figure 7. A result of the auditory experiment (peak frequency of sound change stimuli and gloss perception)

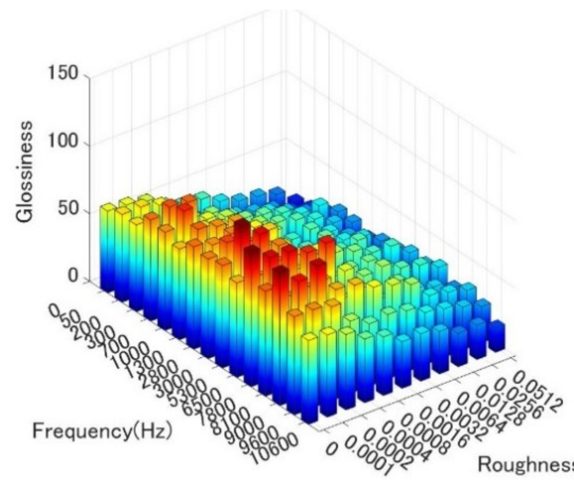


Figure 8. Glossiness regarding the roughness and peak frequency of sound

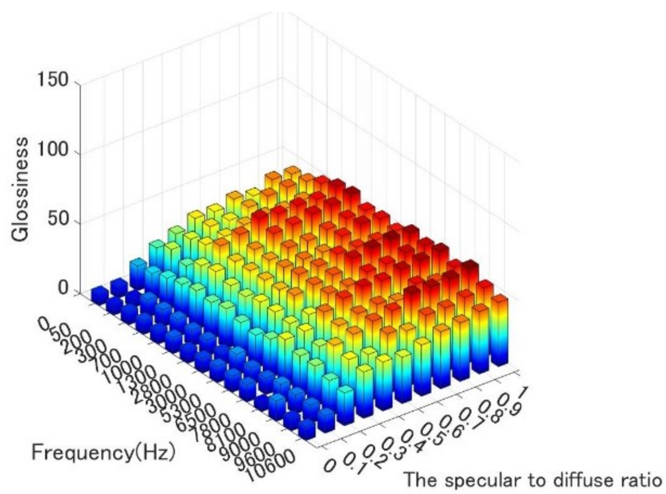
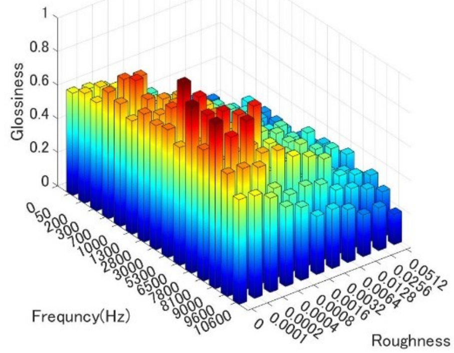
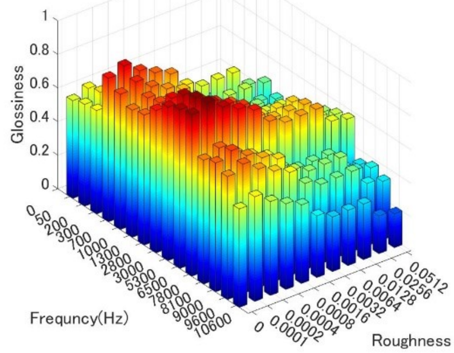


Figure 9. Glossiness regarding the specular to diffuse ratio and peak frequency of sound



(a) A result of audiovisual experiment (normalization)



(b) A result of visual experiment + Result of the auditory experiment (normalization)
Figure 10. Comparison of cross-modal with a sum of single-modal

Constructing a Model for Cross-modal Glossiness Perception

Experimental results showed that the linear sum of visual and auditory parameters can express the perceptual glossiness in cross-modal, and can manipulate the perceptual glossiness by changing their combination. Based on the findings, by analyzing the experimental results, we create a model formula for predicting the cross-modal glossiness perception. Specifically, by optimizing the variables of the roughness, the specular to diffuse ratio and peak frequency of sound, a formula for representing human glossiness perception is modeled using these three physical characteristic quantities.

$$G_{av} = a_{av}O(r) + b_{av}P(m) + c_{av}Q(f) + d_{av} \quad (1)$$

$$O(r) = a_r log r + b_r \quad (2)$$

$$P(m) = a_m m + b_m \quad (3)$$

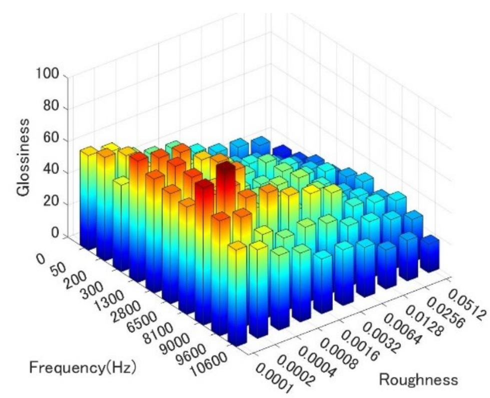
$$Q(f) = a_f f^2 + b_f f + c_f \quad (4)$$

G_{av} in the Equation (1) represents a cross-modal glossiness score. O(r), P(m) and Q(f) are single-modal glossiness models regarding the roughness r, the specular to diffuse ratio m and the peak frequency of sound f. These single-modal models are expressed by the Equation (2) to (4). The glossiness model concerning the roughness r, the specular to diffuse ratio m, and the peak frequency of sound f was found by approximating the relationship between them and the subjective glossiness scores in the previous section. A cross-modal glossiness model G_{av} was constructed from the linear sum of these glossiness models. a, b, c, and d in each equation represent coefficients, and their values are shown in Table 1.

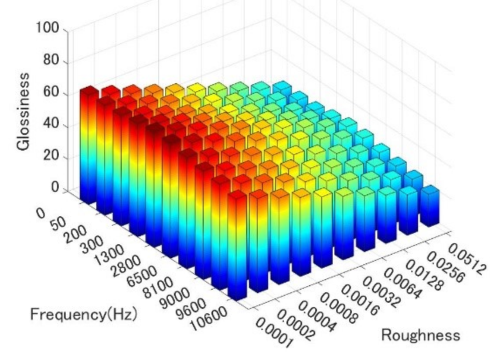
Then, we compared the evaluated values obtained in the subjective experiments with the predicted values obtained from the model. Each graph is shown in Figure 11. Figure 11 shows data excluding outliers in the construction of the model. The predicted value was calculated by fixing the specular to diffuse ratio to 1.0. Although the frequency axis is somewhat different, the approximate shapes of both graphs are almost the same. The correlation was also high (0.89). Then, it is considered that our model can predict a cross-modal glossiness perception using the CG and sound features.

Table 1: The Value of coefficients for Equation (1)-(4)

Coefficient name	Value
a_{av}	0.57
b_{av}	0.56
c_{av}	0.12
d_{av}	-18.59
a_r	-8.474
b_r	-22.59
a_m	47.59
b_m	12.22
a_f	-2.0×10^{-6}
b_f	0.018
c_f	21.67



(a) Evaluated value of audiovisual experiment (excluding outliers)



(b) A predicted value obtained from the model
Figure 11. Comparison of evaluated value and a predicted value

Validation Experiment

A validation experiment was conducted to confirm the feasibility of the cross-modal glossiness perception model. The evaluation experiment in the previous section was carried out with either the roughness or the specular to diffuse ratio as a fixed value. On the other hand, the verification experiment uses stimuli by randomly combining all parameters for verifying whether the cross-modal glossiness perception changes according to the change of the combination of the parameters. There are ten experimental stimuli in total. The combination of parameters was decided so that the predicted value was distributed from a low value to a high value. The experimental setting was the same as in the evaluation experiment described above.

Figure 12 is a scatter diagram showing the relationship between the evaluated values and the predicted values in the validation experiment, and the orange line is the approximate curve obtained by the least squares method. Since the relationship between the evaluated value and the predicted value can be approximated by a straight line, it can be said that the correlation between them is high. ctually, tThe correlation between the evaluation value and the predicted value was 0.93. However, since the slope of the approximate curve is 0.63, it can be seen that this model is not suitable for predicting glossiness. From these discussions, this model is effective for estimating the glossiness relatively.

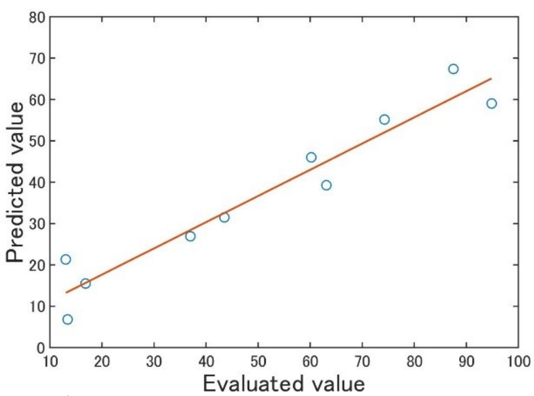


Figure 12. Scatter diagram and the approximate curve in the validation experiment

Conclusions

In this research, we analyzed the cross-modal perception of glossiness. For the development of a cross-modal model concerning the glossiness for sound effects and CG, evaluation experiments were carried out using CG images with changing the parameters of the roughness and the specular to diffuse ratio and sounds effects with various frequency features. As a result, it was found that by combining visual sense and auditory sense, evaluation of cross-modal glossiness perception changes as same as when evaluating them through single-modal perception. That is, by changing the combination of the feature amount of the image and that of the sound, it was found that there is a possibility to manipulate the human sense of glossiness.

Based on the experimental results, a model for predicting cross-modal glossiness perception was constructed from the roughness, the specular to diffuse ratio (which are the image feature amount) and the peak frequency (which is the feature quantity of the sound). A validation experiment was carried out using stimuli with various combinations of the roughness, the specular to diffuse ratio and peak frequency of sound. From the above results, it was confirmed that the model has good accuracy, and it was confirmed that by using this model, it is possible to manipulate the human sense of glossiness.

The cross-modal glossiness model constructed in this study is considered to make it possible to improve the shitsukan researches in future CG creation.

References

[1] S. Kohko, K. kawakami, "Study of Texture Appearance Using Computer Graphics," The illuminating Institute of Japan, vol. 79. no.5, pp.204-211, 1995

[2] P. Fabio, J. A. Ferwerda, and D. P. Greenberg, "Toward a psychophysically-based light reflection model for image synthesis," Proceedings of the 27th annual conference on Computer graphics and interactive techniques, pp.55-64, 2000.

[3] A. Serrano, D. Gutierrez, K. Myszkowski, H. Seidel, and B. Masia, "An intuitive control space for material appearance," SIGGRAPH Asia, vol. 35, no. 6, pp.186, 2016

[4] W. Fujisaki, N. Goda, I. Motoyoshi, H. Komatsu, and S. Nishida, "Audiovisual integration in the human perception of materials," Journal of Vision, vol. 14, no. 4, 2014

[5] W. Fujisaki, M. Tokita, and K. Kariya, "Perception of the material properties of wood based on vision, audition, and touch," Vision Research vol. 109, Part B, pp.185-200, 2015

[6] M. Rodrigo, I. Julian, W. Michael, H. Matthias, "Multimodal Perception of Material Properties," SIGGRAPH, pp.33-40, 2015

[7] N. Bonneel, C. Suied, I. Viaud-Delmon, G. Drettakis, "Audio-visual perception of materials for virtual environments," AGM TAP, vol. 7, no. 1, 2017

[8] PHYSICALLY BASED RENDERING. http://www.pbrt.org/ (access date 2018/02/08)

[9] K. E. Torrance, and E. M. Sparrow, "Theory for Off-Specular Reflection From Roughened Surfaces," J. Opt. Soc. Am. vol. 57, no. 9, pp.1105-1114, 1967.

Author Biography

Takumi Nakamura received the B.E. degree from Chiba University in March 2018. He is currently a student of a master's course program in Chiba University. His research interests are material perception and imaging technologies. Especially, he is now constructing a model for the shitsukan cross-modal perception by investigating the interaction between sounds and graphics.