A study on estimating the attractiveness of food photography

Kazuma Takahashi*, Keisuke Doman†, Yasutomo Kawanishi*, Takatsugu Hirayama*,
Ichiro Ide*, Daisuke Deguchi* and Hiroshi Murase§
*Graduate School of Information Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601 Japan
Email: takahashik@murase.m.is.nagoya-u.ac.jp, {kawanishi, hirayama, ide, murase}@is.nagoya-u.ac.jp
†School of Engineering, Chukyo University, 101 Tokodachi, Kaizu-cho, Toyota-shi, Aichi, 470-0393 Japan
Email: kdoman@sist.chukyo-u.ac.jp
‡Information Strategy Office, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601 Japan
Email: ddeguchi@nagoya-u.ac.jp

***Abstract*—This paper proposes a method for estimating the attractiveness of food photos in order to assist a user to shoot them attractively. The proposed method extracts both color and shape features from input food images, and then integrates them according to a regression scheme. By this way, the proposed method estimates the attractiveness of an unknown food photo. We also created a food image dataset taken from various 3D-angles for each food category, and set target values of their attractiveness through subjective experiments. Then, we evaluated the performance of the proposed method in two different ways of constructing the attractiveness estimator: One that constructs it for each food category, and the other that constructs a common attractiveness estimator for all food categories. Experimental results showed the effectiveness of the proposed method in addition to the necessity for adaptively selecting the estimator depending on the appearance of foods for further performance improvement.**

***Keywords*-food photo; attractiveness; framing;**

I. Introduction

The number of food photos posted on the Web has been increasing with the widespread of social networking services and cooking recipe portal sites. For users of such services, it is preferable to upload cooking recipes together with delicious-looking food photos. Most of the food photos, however, are shot by an amateur photographer. Therefore, they could have various appearances even from the identical food, which leads to various degrees of attractiveness. Figure 1 shows food photographs shot by different framing. Note that these two foods are actually identical. Figure 1(b) would look more delicious than Fig. 1(a) in term of the camera angle and its photographic framing.

In general, it is not necessarily easy to shoot delicious-looking food photos. Thus, it is useful to realize a system that recommends the best camera framing for shooting a food photo, and/or a system for selecting the most attractive one from multiple food photos. To realize such a system, it is necessary to develop a technique for quantifying the attractiveness of a food photo. In this research, we define the attractiveness as the degree of how much a food photo looks delicious.

(a) Non-attractive framing

(b) Attractive framing

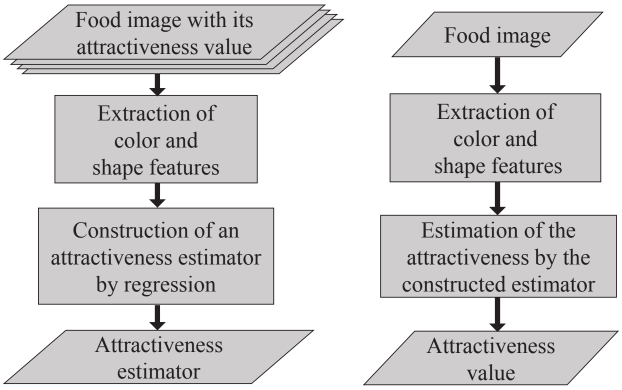
Figure 1. Photography framing of foods

Several research groups have proposed methods to classify the aesthetic quality of photos into two levels: high and low. Nishiyama et al. proposed the use of bags of color patterns in order to evaluate color harmony and color variations in local regions [1]. Tian et al. proposed a method for mining abstract aesthetic features from a massive number of training images using deep convolutional neural networks (DCNNs) [2]. These methods, however, are not for quantifying but for classifying aesthetic quality.

In the field of analysis of food attractiveness, several research groups have analyzed various factors of food attractiveness. Zellner et al. focused on the framing of various elements on the dish, and concluded that the degree of neatness was more important than the balance of the food on the dish [3], [4]. Piqueras-Fiszman et al. studied whether the shape of the plate itself influenced people's taste and flavor perception [5]. In contrast, several research groups reported the significant affect on the perceived sharpness of foods [6]–[8]. These research, however, do not consider the factors on the attractiveness of food photos.

Sakiyama et al. have proposed a method for making food photos attractive by post-processing. They tried to do so by post-super-imposing adding bubble and steam animation into food photos [9]. This method, however, is not for supporting photography but for conversion with post-processing.

In the field of photography, Michel et al. reported that



(a) Training step (b) Estimation step

Figure 2. Process flow of the proposed method.

there is a camera angle from which a food looks the most delicious [10]. In other words, it is important to consider the rotational angle when deciding the framing. However, their work is only an analysis, and not a method for evaluating the attractiveness of food photos taken from various camera angles. Kakimori et al. developed a system to recommend an amateur photographer to shoot a delicious-looking food photo [11]. This system supports the arrangement of dishes in photographic framing. This system, however, dose not recommend the best rotational angle for each dish.

Therefore, in this research, we propose a method for estimating the attractiveness of food photos for recommending the best camera framing based on image features by focusing on the rotation angle. The most important contribution of this paper is to quantify the attractiveness of a food photo. Note that in this paper, we focus on a situation in which only one dish is to be captured.

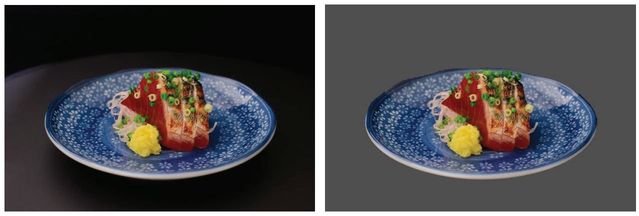
This paper is organized as follows. Section II describes the details of the proposed method. Then, dataset construction by human subjects is described in Section III. Next, the results of evaluating the proposed method is reported in Section IV. Finally, Section V concludes this paper.

II. Estimating the Attractiveness of Food Photos from Image Features

The proposed medthod extracts both color and shape features from a food image, and then estimates its attractiveness. The process flow of the proposed method is shown in Fig. 2. The training step constructs an attractiveness estimator using food images with their attractiveness values. The estimation step estimates the attractiveness of an input food image using the attractiveness estimator. The following sections describe the details of each step.

A. Training step: Construction of an attractiveness estimator

The training step constructs an attractiveness estimator as shown in Fig. 2(a). First, food images with their attractiveness values are input. Next, from each input image, the region containing the dish is detected as the dish region using GrabCut [12], as shown in Fig. 3. Then, color and shape features are extracted from the dish region. Finally, the attractiveness of a food photo is estimated by using Support Vector Regression (SVR) [13]. Here, the objective variable is the attractiveness value of the food photo and the explanatory variables are the image feature values. Details of the features used in the proposed method are described below.



(a) Input (b) Output

Figure 3. Example of foreground extraction using GrabCut.

1) Color feature: The color components of a food affects its attractiveness. It is said that there is a relationship between the color distribution of a food and our appetite. That is, warm colors such as red, orange and yellow stimulate our appetite, whereas cool colors such as blue, purple, gray and black lose our appetite [14]. We consider that foods stimulating our appetite is attractive. In addition, Nishiyama et al. reported that the colors of photos have a significant influence on their perceived quality from the viewpoint of color harmony. Considering the above, we consider color feature in order to evaluate the color distribution of food.

The proposed method uses the color difference in the CIELAB color space which is designed to approximate human visual perception. First, the proposed method calculates the L*a*b color histogram from an input image, and then calculates the most frequent color $G = (L, a, b)$. Note that each of L*a*b components is quantized into eight levels ($0 \leq L, a, b \leq 7$) to reduce the number of dimensions of the color feature. Next, the proposed method divides the input image into 5×5 blocks, and calculates the most frequent color $R_i = (\widetilde{L}_i, \widetilde{a}_i, \widetilde{b}_i)$ and its frequency F_{R_i} in each block. Here, i is the index of each block. Note that $0 \leq i \leq 24$, $0 \leq \widetilde{L}_i, \widetilde{a}_i, \widetilde{b}_i \leq 7$. Then, the color difference D_i between R_i and G is calculated as

$$D_i = F_{R_i} \sqrt{(L - \widetilde{L}_i)^2 + (a - \widetilde{a}_i)^2 + (b - \widetilde{b}_i)^2}. \quad (1)$$

F_{R_i} is a weight coefficient for the color difference between R_i and G. Finally, the following 25-dimensional vector $\boldsymbol{D}$ is composed and is used as the color feature.

$$\boldsymbol{D} = (D_0, D_1, \ldots, D_{24}). \quad (2)$$

2) Shape feature: The shape and the arrangement of ingredients affect the visual appearance of food photos, which makes a difference in the best camera angle. Therefore, we consider shape feature in order to evaluate the 3D volume of food and the orientation of its ingredients.

The proposed method calculates the first to the fourth central moments of an orientation histogram from an input image. These four kinds of central moments represent the average, the variance, the kurtosis, and the skewness of the orientation histogram, which are used to roughly represent the shape of a food.

The detailed process is as follows: First, the edge image is calculated from an input image, and then the magnitude $m(x, y)$ and the orientation $\theta(x, y)$ for each pixel (x, y) in the edge image is calculated as

$$m(x, y) = \sqrt{f_h(x, y)^2 + f_v(x, y)^2}, \tag{3}$$

$$\theta(x, y) = \tan^{-1} \frac{f_h(x, y)}{f_v(x, y)}, \tag{4}$$

where $f_h(x, y)$ and $f_v(x, y)$ are the horizontal and the vertical edge images calculated from the input image. The orientation $\theta(x, y)$ is quantized into 36 levels to reduce the number of dimensions of the shape feature. Note that the above procedure is applied only to the pixels inside the dish region but excluding those within 5 pixels distance around its contour. Next, the proposed method calculates the average M_1, the variance M_2, the kurtosis M_3 and the skewness M_4 of the orientation histogram as

$$M_1 = \frac{\sum_{i=1}^{I} H(i)}{n}, \tag{5}$$

$$M_2 = \frac{\sum_{i=1}^{I} (H(i) - M_1)^2}{n}, \tag{6}$$

$$M_3 = \frac{\sum_{i=1}^{I} (H(i) - M_1)^3}{n\sigma^3}, \tag{7}$$

$$M_4 = \frac{\sum_{i=1}^{I} (H(i) - M_1)^4}{n\sigma^4}, \tag{8}$$

where I is the number of bins of the histogram, $H(i)$ is the value of the i-th bin, and n is the number of pixels in the dish region. Also, $\sigma = \sqrt{M_2}$ represents the standard deviation of the orientation histogram. Finally, the following vector $\boldsymbol{M}$ is composed, and is used as the shape feature.

$$\boldsymbol{M} = (M_1, M_2, M_3, M_4). \tag{9}$$

B. Estimation step: Estimation of the attractiveness values of food photos

The estimation step estimates the attractiveness value of a food photo using the attractiveness estimator. In the same manner as the training step, the proposed method extracts the color and the shape features from the dish region in an input image. Then, the attractiveness value of the food image is calculated by the attractiveness estimator.

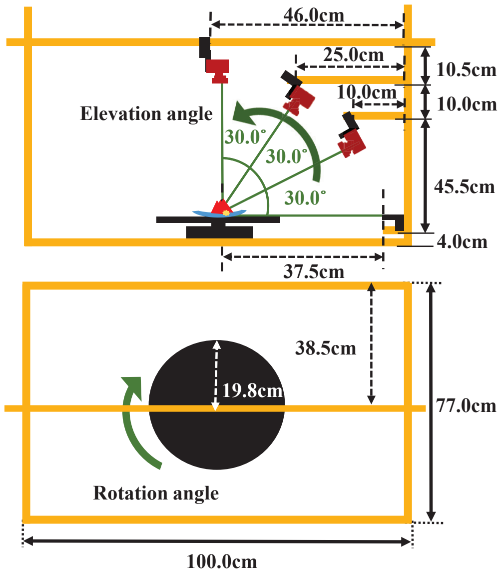


Figure 4. Camera setting for the experiments: tabular drawing

III. Dataset construction by subjective experiments

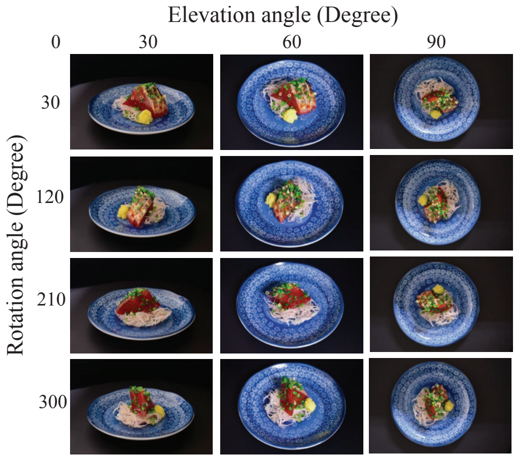
We conducted subjective experiments to make an image dataset with attractiveness value for constructing the attractiveness estimator. Here, we took the paired comparison approach and measured the attractiveness values for each image. Experimental conditions and results are described below.

A. Target food categories

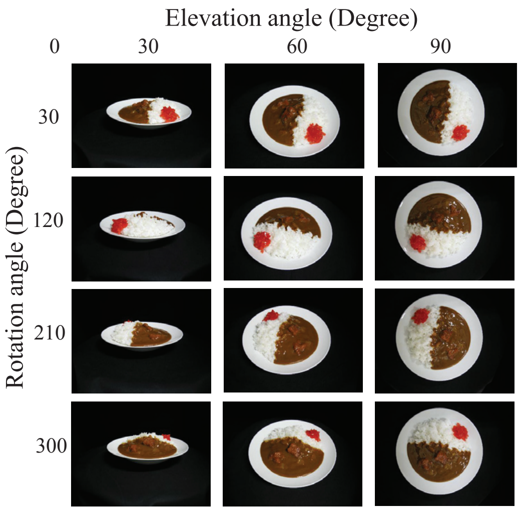
We selected three food categories for the evaluation of the proposed method considering the variation of the appearance in both color and shape: "Seared bonito sashimi", "Curry rice", and "Eel rice bowl". Note that we used plastic food samples instead of real ones as subjects of food photos, considering both convenience and reproducibility.

B. Photographing method

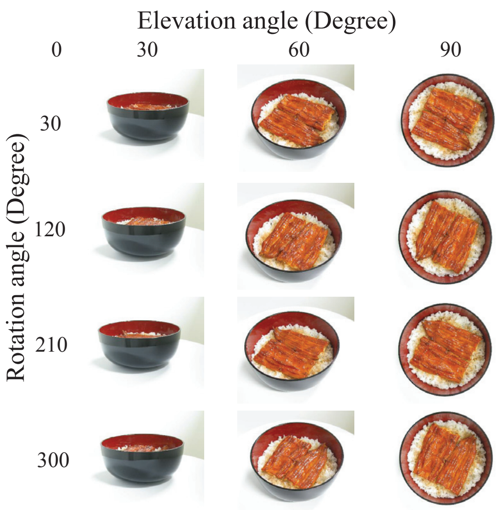
We shot food photos from various 3D-angles with the photographing apparatus designed as Fig. 4. The apparatus was equipped with a turn-table. Accordingly, it could change the elevation angle and rotation angle for photography while keeping a fixed distance between the camera and the subject. Note that photography from 0 and 90 elevation angles corresponds to photographing from the side and the top of the subject, respectively. We shot food photos from three elevation angles: 30, 60 and 90 degrees. Also, we set an arbitrary rotation angle as 0 degrees. We then photographed from 0 degrees to 330 degrees with the step of 30 degrees in clockwise direction around the center of the subject. As a result, we obtained 36 food photos in total for each food category. Figure 5 shows a part of the image datasets.



(a) Seared bonito sashimi



(b) Curry rice



(c) Eel rice bowl

Figure 5. Example of food images for each food category

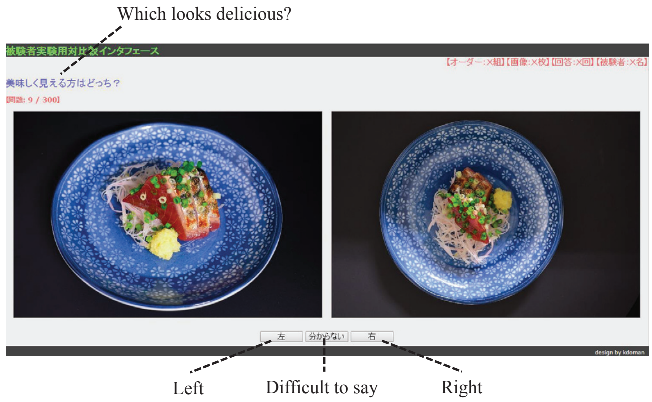


Figure 6. Interface used in the subjective experiments

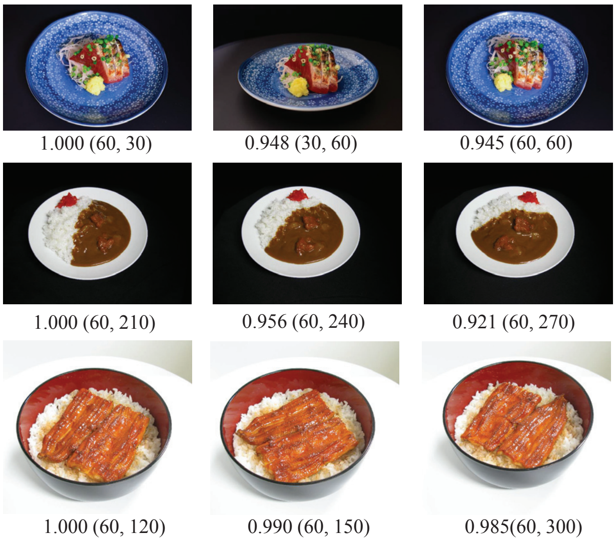
C. Determination of attractiveness values by paired comparison

We used Thurstone's method [15] in order to determine the attractiveness values of food photos. Thurstone's method is a kind of paired comparison methods for sensory test. It can be used to determine an interval scale for perceived quality. In the experiments, the number of image pairs were ${}_{36}C_2 = 630$ for each food category. Figure 6 shows the interface for the experiments. An image pair was shown at a time to human subjects, and they were asked to respond which image looked more delicious by selecting one of the buttons: "Left", "Right", or "Difficult to say." Note that human subjects in the experiments were sixteen males and four females in their 20s. As a result, we obtained three to four responses for each image pair and 2,150 responses in total. Then, we calculated the attractiveness values for each image according to Thurstone's method, and normalized them into the range [0, 1]. These values were used as target values for the regression in the proposed method.

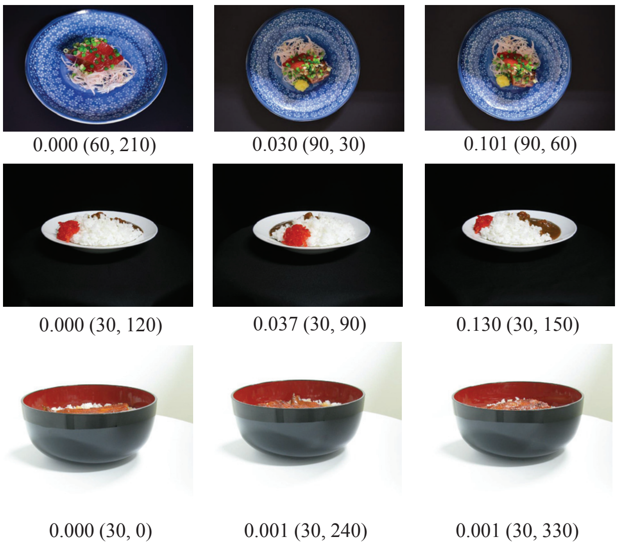
Figure 7 shows the top three and the least three images in the attractiveness values in each image dataset. For "Seared bonito sashimi", the human subjects tended to prefer framings with the grilled surface and the ginger to the front side of the dish. For "Curry rice", they tended to prefer framings with Fukujinzuke (pickled vegetables) to the back side. For "Eel rice bowl", they tended to prefer framings taken from high elevation degrees with the cut line of the eel facing the front. In addition, they did not generally prefer framings in which all of the ingredients could not be visually recognized due to occlusion.

IV. Experiments

We analyzed the effectiveness and the generalization performance of the proposed method which extracted both the color and the shape features from food images. The experimental method and the results are described below.



(a) Top three images in the attractiveness values for each image dataset. The numbers indicate the attractiveness value (elevation angle, rotation angle).



(b) Least three images in the attractiveness for each image dataset. The numbers indicate the attractiveness value (elevation angle, rotation angle).

Figure 7. Examples of the results from the subjective experiments.

A. Method

The attractiveness estimator is constructed in two different ways by changing the training data: One way was to create it for each food category, and the other way was to create a common attractiveness estimator for all food categories. We also compared the performance by different usage of features for the proposed method: One used only the color feature described in Section II-A1, and the other used only the shape feature described in Section II-A2. Here, the estimation accuracy was evaluated through leave-one-out cross-validation; One image was used for testing while the remaining ones were used for training, and this was repeated for each image. In the experiments, LIBSVM [16] was used as an implementation of SVR, and a linear kernel was used. The feature values calculated by each method were normalized into the range [0, 1]. As for the evaluation criteria, we used Mean Absolute Error (MAE) between the estimated values by each method and the target values determined through the subjective experiments.

Table I
EXPERIMENTAL RESULTS: ESTIMATION ERROR WHEN CREATING AN ESTIMATOR FOR EACH FOOD CATEGORY

Method (feature)	**Mean Absolute Error**		
	Seared bonito sashimi	**Curry rice**	**Eel rice bowl**
Comparative 1 (color)	0.187	0.201	0.117
Comparative 2 (shape)	0.213	0.174	**0.095**
Proposed (color + shape)	**0.180**	**0.146**	0.117

Table II
EXPERIMENTAL RESULTS: ESTIMATION ERROR WHEN CREATING AN ESTIMATOR FOR ALL FOOD CATEGORIES

Method (feature)	**Mean Absolute Error**		
	Seared bonito sashimi	**Curry rice**	**Eel rice bowl**
Comparative 1 (color)	**0.230**	0.188	0.163
Comparative 2 (shape)	0.271	0.182	0.324
Proposed (color + shape)	0.232	**0.181**	**0.153**

B. Results

Estimation errors of two attractiveness estimators are described in the following sections.

1) Estimation error when creating an estimator for each food category: Table I shows the estimation error when an attractiveness estimator was constructed for each food category. The proposed method outperformed the comparative methods for the food categories "Seared bonito sashimi" and "Curry rice" where MAEs were 0.180 and 0.146, respectively. However, comparative method 2 which used only the shape feature outperformed the proposed method for the food category "Eel rice bowl" where MAE was 0.095.

2) Estimation error when creating a common estimator for all food categories: Table II shows the estimation error when a common attractiveness estimator was constructed for all food categories. The proposed method outperformed the comparative ones for the food categories "Curry rice" and "Eel rice bowl" where MAEs were 0.181 and 0.153, respectively.

Comparative method 1 outperformed the proposed method for the food category "Seared bonito sashimi" where MAEs was 0.230. The minimum estimation error by the estimator created for each food category was lower than that by the common estimator created for all food categories.

C. Discussions

As mentioned before, we constructed the attractiveness estimator in two different ways, and obtained a higher estimation accuracy in terms of the overall MAE by creating it for each food category. This would be because the attractiveness of food is affected by various factors including colors and shapes, and their significance to the attractiveness may also vary for each food category. For example, the comparative method 2 which used only the shape feature outperformed the proposed method for the food category "Eel rice bowl." This would be because its hue is relatively uniform. Accordingly, it was more important to consider the shape feature than color features for evaluating its attractiveness. Thus, depending on the food category, it is better to adaptively select image features for evaluating the attractiveness the framing of food photos.

There may be, however, various appearances even in the same food category due to the differences of dishes, arrangement, and so on. Accordingly, it would be better to categorize foods focusing on their appearance rather than their category names, and then construct an attractiveness estimator for each appearance category.

V. CONCLUSION

This paper proposed a method for estimating the attractiveness of a food photo toward a system for shooting attractive ones from the viewpoint of color and shape. We created an image dataset of food photos with their attractiveness values for each food category through subjective experiments. We also compared the performance of the proposed method in two different ways of constructing an attractiveness estimator: One was to construct an attractiveness estimator for each food category, and the other was to construct a common attractiveness estimator for all of the food categories. Experimental results showed the effectiveness of the proposed method in addition to the necessity for adaptively selecting the estimator for each food category for further performance improvement.

Future work includes: (1) introducing additional color and/or shape features for performance improvement, (2) introducing a framework for clustering the appearance of foods and introducing additional image features, (3) considering the color harmony between foods and dishes, and the size of the photographic subject, and (4) realizing an assistance system that interactively guides the user to the best camera angle for photography.

ACKNOWLEDGMENT

Parts of this research were supported by JSPS Grant-in-Aid for Scientific Research and MSR-Core12 program. We appreciate the members of the laboratory for taking part in the experiment.

REFERENCES

[1] M. Nishiyama, T. Okabe, I. Sato, and Y. Sato. (2011, June) "Aesthetic quality classification of photographs based on color harmony," *Proc. 2011 IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 33–40.

[2] X. Tian, Z. Dong, K. Yang, and T. Mei. (2015, Oct.) "Query-dependent aesthetic model with deep learning for photo quality assessment," *IEEE Trans. on Multimedia*, vol. 17, no. 11, pp. 2035–2048.

[3] D. A. Zellner, M. Lankford, L. Ambrose, and P. Locher. (2010, Jul.) "Art on the plate: Effect of balance and color on attractiveness of, willingness to try and liking for food," *Food Quality and Preference*, vol. 2, no. 5, pp. 575–578.

[4] D. A. Zellner, E. Siemers, V. Teran, R. Conroy, M. Lankford, A. Agrafiotis, L. Ambrose, and P. Locher. (2011, Dec.) "Neatness counts. How plating affects liking for the taste of food," *Appetite*, vol. 57, no. 3, pp. 642–648.

[5] B. Piqueras-Fiszman, J. Alcaide, E. Roura, and C. Spence. (2012, Apr.) "Is it the plate or is it the food? Assessing the influence of the color (black or white) and shape of the plate on the perception of the food placed on it," *Food Quality and Preference*, vol. 24, no. 1, pp. 205–208.

[6] O. Deroy and D. Valentin. (2011, Sep.) "Tasting liquid shapes: investigating the sensory basis of cross-modal correspondences," *Chemosensory Perception*, vol. 4, no. 3, pp. 80–90.

[7] S. A. Day. (2011) "The human sensoria and synaesthetic approach to cooking," *Collapes*, vol. 7, pp. 378–409.

[8] C. Spence and O. Deroy. (2014) "Tasting shapes: A review of four hypotheses," *Theoria et Historia Scientiarum*, vol. 10, pp. 207–238.

[9] S. Sakiyama, M. Okabe, and R. Onai. (2013, Dec.) "Animating picture of cooking using video example of bubble and image deformation," *Proc. 21st Workshop on Interactive Systems and Software*, pp. 79–84.

[10] C. Michel, A. T. Woods, M. Neuhauser, A. Landgraf, and C. Spence. (2015, Sep.) "Rotating plates: Online study demonstrates the importance of orientation in the plating of food," *Food Quality and Preference*, vol. 44, pp. 194–202.

[11] T. Kakimori, M. Okabe, K. Yanai and R. Onai. (2015, Nov.) "A system to support the amateurs to take a delicious-looking picture of foods," *Proc. SIGGRAPH ASIA Symposium on Mobile Graphics and Interactive Applications*, no. 28.

[12] C. Rother, V. Kolmogorov, and A. Blake. (2004, Aug.) "GrabCut: Interactive foreground extraction using iterated graph cuts," *ACM Trans. on Graphics*, vol. 23, no. 3, pp. 309–314.

[13] A. J. Smola and B. Schölkopf. (1998, Aug.) "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222.

[14] F. H. Mahnke. (1996) "Color, environment, and human response: An interdisciplinary understanding of color and its use as a beneficial element in the design of architectural environment," *John Wiley & Sons*.

[15] L. L. Thurstone. (1927, Jul.) "Psychophysical analysis," *American J. of Psychology*, vol. 38, no. 3, pp. 368–389.

[16] C.-C. Chang and C.-J. Lin. (2011, Apr.) "LIBSVM: A library for support vector machine," *ACM Trans. on Intelligent Systems Technology*, vol. 2, no. 27, pp. 1–27.