Estimation of the Attractiveness of Food Photography Focusing on Main Ingredients

Kazuma Takahashi*
Graduate School of Infomation Science,
Nagoya University, Japan

Keisuke Doman
School of Engineering,
Chukyo University, Japan
kdoman@sist.chukyo-u.ac.jp

Yasutomo Kawanishi
Graduate School of Informatics,
Nagoya University, Japan
kawanishi@i.nagoya-u.ac.jp

Takatsugu Hirayama
Graduate School of Informatics,
Nagoya University, Japan
hirayama@i.nagoya-u.ac.jp

Ichiro Ide
Graduate School of Informatics,
Nagoya University, Japan
ide@i.nagoya-u.ac.jp

Daisuke Deguchi
Information Strategy Office,
Nagoya University, Japan
ddeguchi@nagoya-u.jp

Hiroshi Murase
Graduate School of Informatics,
Nagoya University, Japan
murase@i.nagoya-u.ac.jp

ABSTRACT

This research aims to develop a method to estimate the attractiveness of a food photo. The proposed method extracts two kinds of image features: 1) those focused on the appearance of the main ingredient, and 2) those focused on the impression of the entire food photo. The former is newly introduced in this paper, whereas the latter is based on previous research. The proposed method integrates these image features with a regression scheme to estimate the attractiveness of an arbitrary food photo. We have also built and released a food image dataset composed of images of ten food categories taken from 36 angles named NU FOOD 360x10. The images were assigned target values of their attractiveness through subjective experiments. Experimental results showed the effectiveness of integrating both kinds of image features.

CCS CONCEPTS

• **Human-centered computing** → *Visual analytics*;

KEYWORDS

Food photography, attractiveness, framing

ACM Reference format:
Kazuma Takahashi, Keisuke Doman, Yasutomo Kawanishi, Takatsugu Hirayama, Ichiro Ide, Daisuke Deguchi, and Hiroshi Murase. 2017. Estimation of the Attractiveness of Food Photography Focusing on Main Ingredients. In *Proceedings of CEA2017, Melbourne, Australia, August 20, 2017,* 6 pages.
DOI: 10.1145/3106668.3106670

*Currently at Fuji Xerox Co., Ltd.

CEA2017, Melbourne, Australia

DOI: 10.1145/3106668.3106670

(a) Non-attractive framing (b) Attractive framing

Figure 1: Photographic framing of a food.

1 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. Users of such services prefer to upload delicious-looking food photos together with other contents such as cooking recipes, comments, or reviews. Most of the food photos, however, are shot by an amateur photographer, which leads to various degrees of attractiveness. Figure 1b would look more attractive from the point of deliciousness than Fig. 1a in terms of camera angle and its photographic framing, although these two photos are actually obtained by shooting the same food. Note that we define the attractiveness as the degree of how much a food photo looks delicious. It is not necessarily easy for an amateur photographer to shoot attractive food photos, mainly because decision of camera framing is not always easy. Thus, it would be useful to realize a system that can recommend the best camera framing for shooting an attractive food photo and/or a system for selecting the most attractive one from a list of food photos. This research aims to develop a technique to quantify the attractiveness of a food photo.

Although there has been much research on food image understanding, most of them dealed with the task of retrieval and classification [3]. Some researchers have proposed methods to classify the aesthetic quality of general 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 [9]. Tian et al. proposed a method to construct a classification model for each query image using deep convolutional neural networks (DCNNs) [15]. However, these methods do not consider the food-specific attractiveness discussed in [12].

Sakiyama et al. have proposed a method for making food photos attractive by post-processing [11]. They tried to do so by post-super-imposing adding a bubble and steam animation into food photos. This method, however, is not for supporting photography itself, but for conversion of already shot photos, so it cannot change the shooting angle.

In the field of photography, Kakimori et al. developed a system that shows a user the guideline for arranging dishes in photographic framing [5]. Although the system may be useful for an amateur photographer to arrange dishes, the system neither recommends the best camera angle for each dish nor evaluates the attractiveness of food photos. Michel et al. reported that there is a camera angle from which a food looks the most attractive [8], and the rotation angle in particular is one of the key factors when deciding the framing. On the other hand, we proposed a method for estimating the attractiveness of food photos in order to propose the best camera framing based on image features [13]. This method extracted several color and shape features to evaluate the impression of the entire food photo. We also confirmed the effectiveness of the method through experiments on three food categories. This research, however, did not distinguish the image features of a main ingredient. As we can see in Fig. 1, it is actually better to shoot focusing on the main ingredient in the front. Thus, the appearance of the main ingredient should affect the attractiveness of a food photo even for the same food.

Therefore in this paper, we introduce additional image features to those proposed in [13] focusing on the main ingredient of a food for more accurate attractiveness estimation. The proposed method extracts two kinds of image features: 1) those that evaluate the appearance of the main ingredient, and 2) those that evaluate the impression of the entire food photo. The former is newly introduced in this paper, whereas the latter is a minor improvement of the image features introduced in [13]. The proposed method integrates these image features with a regression scheme. In addition, we built a food image dataset (named "NU FOOD 360x10") composed of images of ten food categories which has been released to the public[1]. Note that in this paper, we focus on a situation in which only one dish is to be captured.

The contribution of this paper is two-fold. One is the introduction of image features on the appearance of main ingredients, which leads to the improvement of the accuracy of food attractiveness estimation. The other is the construction of a food image dataset "NU FOOD 360x10" which has been released to the public to facilitate research in the field.

This paper is organized as follows. Section 2 describes the details of the proposed method. Then, dataset construction by human subjects is described in Section 3. Next, the results of evaluating the proposed method is reported in Section 4. Finally, Section 5 concludes this paper.

[1]NU FOOD 360x10: http://www.murase.is.i.nagoya-u.ac.jp/nufood/

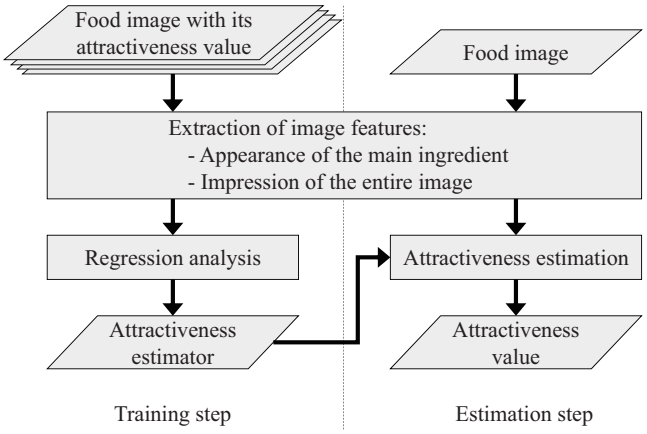


Figure 2: Process-flow of the proposed method.

(a) Input image **(b) Dish region R_d** **(c) Main ingredient region R_m**

Figure 3: Example of the result of region segmentation for an "Eel rice-bowl" image.

2 PROPOSED METHOD

The process-flow of the proposed method is shown in Fig. 2. The proposed method is composed of two steps: training step and estimation step. The training step constructs an attractiveness estimator using food images with their attractiveness values based on a regression framework. The proposed method uses Random Regression Forests [6] for the regression. Here, the objective variable is the attractiveness value of the food photo, and the explanatory variables are the image feature values described below. The estimation step estimates the attractiveness of an input food image using the attractiveness estimator.

The following sections describe the procedure of image feature extraction in the proposed method.

2.1 Region for Image Feature Extraction

Each input image is segmented into the following two regions, for example, by using GrabCut [10] as shown in Fig. 3. One is the dish region R_d which contains the entire dish including all the ingredients, which is used for extracting the image features to evaluate the impression of the entire food photo. The other is the main ingredient region R_m which contains only the ingredient characterizing the food, which is used for extracting the image features to evaluate the appearance of the main ingredient. We suppose that the main ingredient region can be selected manually by a user via an interface such as a smartphone in the estimation step.

2.2 Image Features: Appearance of the Main Ingredients

In order to decide an attractive framing, a photographer should consider the appearance of main ingredients such as the apparent size in the photo, the arrangement, and the orientation. The following image features S, P_x, P_y, $\boldsymbol{O}$ and $\boldsymbol{M}$ are extracted from the main ingredient region R_m in an input image.

2.2.1 Size Feature: Apparent Size of the Main Ingredients. The proposed method calculates the area ratio S of the main ingredient region R_m to the dish region R_d as,

$$S = \frac{|R_m|}{|R_d|}. \tag{1}$$

2.2.2 Position Feature: Relative Position of the Main Ingredients. The proposed method calculates the x- and y-directional difference, P_x and P_y, between the gravity centers of the dish region (x_d, y_d) and the main ingredient region (x_m, y_m) as,

$$P_x = x_d - x_m, \tag{2}$$

$$P_y = y_d - y_m. \tag{3}$$

2.2.3 Shape Feature: Orientation of the Main Ingredients. The proposed method calculates the strength and the orientation of the gradient, and then assemble an orientation histogram. The orientation here is quantized into 36 levels by dividing the range of the gradient angles [0,360) into 36 orientations in order to reduce the number of dimensions of the image feature. Finally, the following 36-dimensional vector $\boldsymbol{O} = (O_1, O_2, \ldots, O_{36})$ is obtained.

2.2.4 Moment Feature: Orientation Statistics of the Main Ingredients. The proposed method also calculates the first to the fourth central moments $\boldsymbol{M} = (M_1, M_2, M_3, M_4)$ of the orientation histogram $\boldsymbol{O}$, where M_1, M_2, M_3, and M_4 are the average, the variance, the kurtosis, and the skewness of $\boldsymbol{O}$, respectively.

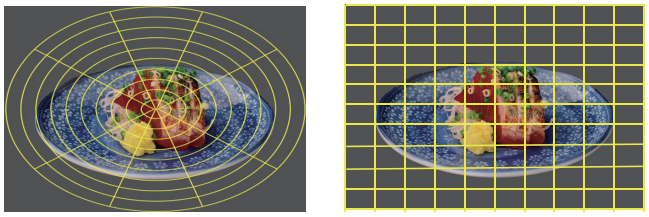
2.3 Image Features: Impression of the Entire Food Photo

The following image features $\boldsymbol{C}$, $\boldsymbol{E}$ and $\boldsymbol{A}$ are extracted from the dish region in an input image.

2.3.1 Color Feature: Color Difference in the CIELAB Color Space. It is known that there is a relationship between the color distribution of a food and our appetite [7]. Thus, the proposed method considers the color difference in the CIELAB color space, which is designed to approximate human visual perception. Note that this feature has been introduced in the previous research [13].

The proposed method first calculates the most frequent color (L, a, b) in the CIELAB color space from the dish region. Each of the color channels here is quantized into eight levels ($1 \leq L, a, b \leq 8$) to reduce the number of dimensions of the feature vector. Next, the proposed method divides the input image into 100 radial local regions as shown in Fig. 4a, and calculates the most frequent color (L_i, a_i, b_i) and its frequency F_i in each local region. Here, i is the index of each block, $1 \leq i \leq 100$, and $1 \leq L_i, a_i, b_i \leq 8$. Then, the color difference C_i is calculated as

$$C_i = F_i \sqrt{(L - L_i)^2 + (a - a_i)^2 + (b - b_i)^2}. \tag{4}$$



(a) Division for the color feature extraction

(b) Division for the edge feature extraction

Figure 4: Region division for extracting image features on the impression of the entire food photo.

Finally, a 100-dimensional vector $\boldsymbol{C} = (C_1, C_2, \ldots, C_{100})$ is obtained.

2.3.2 Shape Feature: Orientation and Strength of Edge. The shape and the arrangement of ingredients affect the visual appearance of food photos, which makes a difference of the best camera angle. Thus, the proposed method considers the orientation and the strength of the orientation histogram.

The proposed method first divides an input image into 10×10 blocks as shown in Fig. 4b. Next, the proposed method calculates the maximum edge strength e_j and gradient orientation n_j from each block. Here, j is the index of each block, and $1 \leq j \leq 100$. The orientation from each block is quantized into 36 levels to reduce the number of dimensions of the feature vector. Finally, a 100-dimensional vector $\boldsymbol{E} = (e_1 n_1, e_2 n_2, \ldots, e_{100} n_{100})$ is obtained.

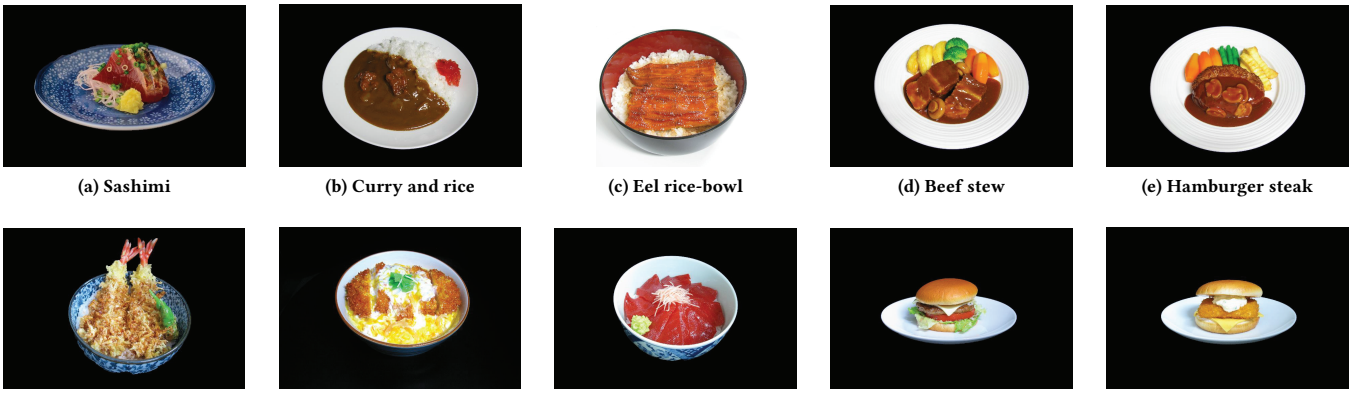
2.3.3 Color and Shape Feature: Deep Convolutional Activation Feature (DeCAF). DeCAF [2] is the weight data on the neural network trained with ImageNet [1], which includes 1,000 categories of objects. The network is composed of eight layers. The first five are convolutional layers and the remaining are fully-connected layers. The proposed method normalizes the 4,096-dimensional output values of the seventh layer into [0,1], and uses them as an image feature $\boldsymbol{A}$.

3 DATASET CONSTRUCTION BY SUBJECTIVE EXPERIMENTS

We conducted subjective experiments in order to make an image dataset with attractiveness values (NU FOOD 360x10[1]) for constructing the attractiveness estimator. We took the paired comparison approach in order to determine the attractiveness value for each image. The details of the experimental method and results are described below.

3.1 Food Categories

We added seven food categories into the previous dataset [13], and consequently built a larger dataset of ten food categories shown in Fig. 5. These food categories were selected considering the variation of the appearance in both color and shape. Note that we used plastic food samples instead of real ones considering both convenience and reproducibility.



(a) Sashimi (b) Curry and rice (c) Eel rice-bowl (d) Beef stew (e) Hamburger steak

(f) Tempura rice-bowl (g) Fried pork rice-bowl (h) Tuna rice-bowl (i) Cheese burger (j) Fish burger

Figure 5: Food categories used for evaluation experiments.

3.2 Photographing Method

We shot food photos from various 3D-angles with the setting shown in Fig. 6, namely, Sashimi, Curry and rice, Eel rice-bowl, Beef stew, Hamburger steak, Tempura rice-bowl, Fried pork rice-bowl, Tuna rice-bowl, Cheese burger, and Fish burger. The apparatus was equipped with a turn-table so that it could change the elevation angle and rotation angle while keeping a fixed distance between the camera and the subject. Note that shooting from 0 and 90 elevation angles corresponds to shooting from the side and the top of the dish, respectively. We shot food photos from three elevation angles: 30, 60, and 90 degrees. Also, we set an arbitrary rotation angle as 0 degrees, and then shot from 0 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.

3.3 Determination of Attractiveness Values by Paired Comparison

We used Thurstone's paired comparison method [14] in order to determine the attractiveness values of food photos. This method was developed for sensory test, and 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. 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." Human subjects were 28 Computer Science-major students in their 20s, out of which nine subjects were assigned for each food category. We finally obtained three or four responses for each image pair and 2,150 responses in total for each food category.

The attractiveness values obtained in the experiment are shown in Fig. 7. Note that these values were normalized into the range of [0,1], and were used as target values for the regression in the proposed method.

4 EXPERIMENTS

We evaluated the effectiveness of the proposed method through experiments.

Table 1: Experimental results: Mean Absolute Error (MAE) in the range of [0,1].

Category	Tian et al. [15]	Proposed
Sashimi	0.330	0.128
Curry and rice	0.214	0.087
Eel rice-bowl	0.383	0.068
Beef stew	0.349	0.086
Hamburger steak	0.258	0.095
Tempura rice-bowl	0.405	0.124
Fried pork rice-bowl	0.326	0.097
Tuna rice-bowl	0.297	0.054
Cheese burger	0.438	0.065
Fish burger	0.441	0.071
Average	0.344	0.087

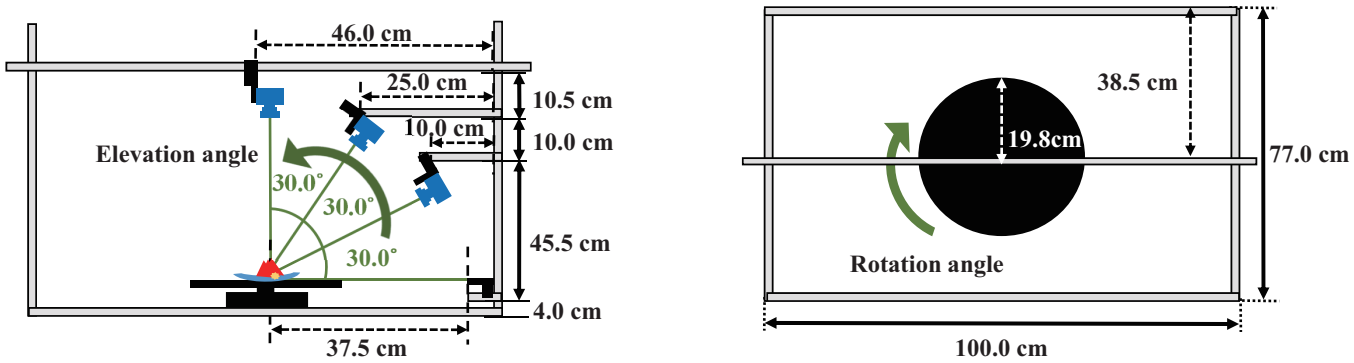
4.1 Method

We applied a leave-one-out scheme with the dataset described in Section 3 for training and evaluating the attractiveness estimator proposed in this paper. The main ingredient region for feature extraction was manually labelled for each food category.

We compared the estimation accuracy of the proposed method with that of a comparative method based on [15], which was designed to classify the aesthetic quality of general photos using deep convolutional neural networks (DCNNs). For each method, we evaluated the Mean Absolute Error (MAE) between the estimated values and the target values for the attractiveness of food photos.

4.2 Results

Experimental results are summarized in Table 1. The average MAE of the proposed method was 0.087, whereas that of the comparative method [15] was 0.344. The proposed method outperformed the comparative one for all food categories. From the results, we can see at least the following two important things: One is the necessity for considering the attractiveness specific to food photos.



(a) Side view (b) Top view

Figure 6: Camera setting for the experiments.

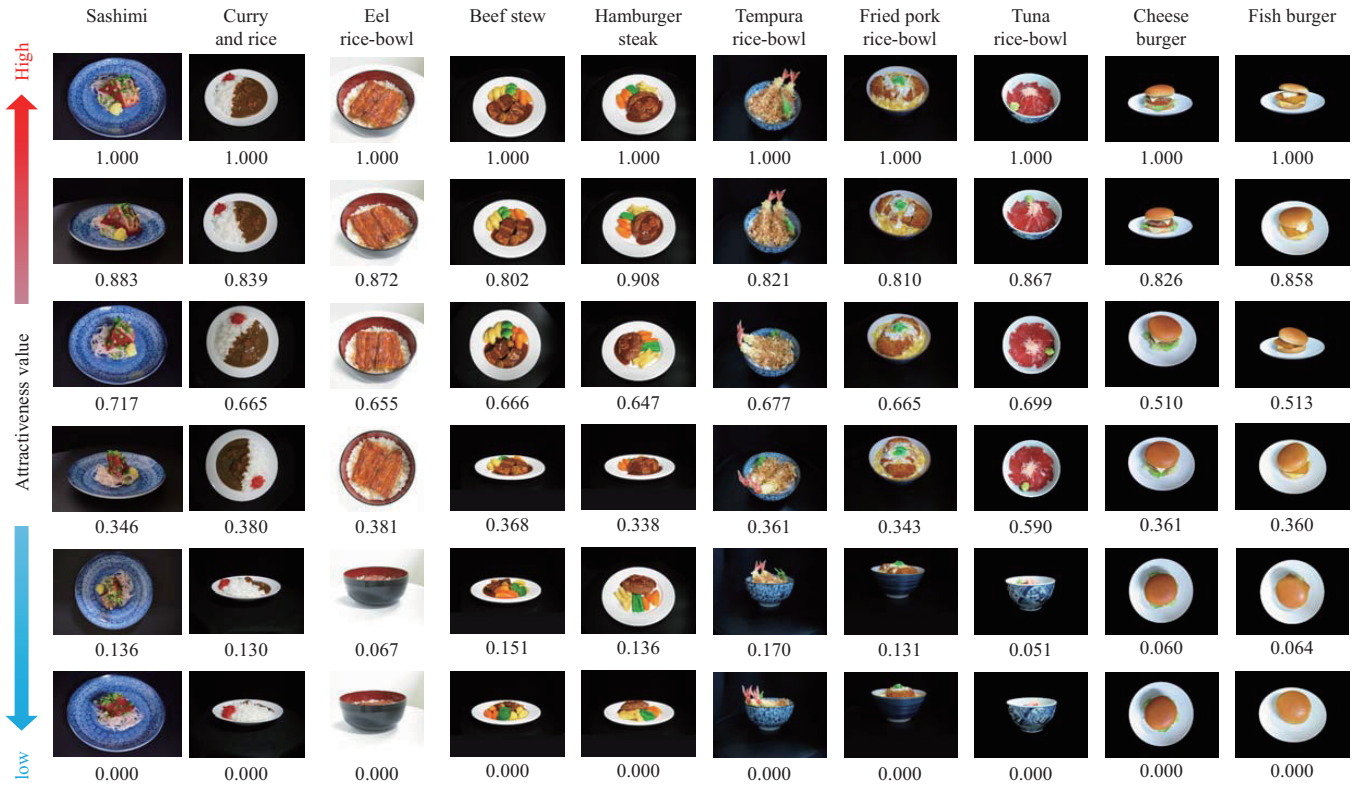


Figure 7: Attractiveness values for each image in each food category.

Another is the effectiveness of integrating both the appearance of the main ingredients and the impression of the entire food photo.

4.3 Discussion

We investigated the effectiveness of each image feature in more detail. Table 2 shows the MAEs when using only one of the image features. For reference, this table also includes the MAEs when using all the image features on the appearance of the main ingredients (Section 2.2) and the impression of the entire food photo (Section 2.3), denoted as "All". The average MAE when using only DeCAF was 0.090, which showed the best performance among these nine methods including "All"s. The second best in the methods except for "All"s was the orientation of the main ingredient named "Shape" (the fourth column from the left). The effective image feature depended on the food category. This suggests that more accurate estimation can be achieved by switching the attractiveness estimators depending on an input food image. One approach for

Table 2: Experimental results: Mean Absolute Error (MAE) in the range of [0,1] (Bold indicates the lowest error for each category, and "All" indicates the combination of all the image features of the same kind).

Category	Appearance of the main ingredients					Impression of the entire food photo			
	Size	Position	Shape	Moment	All	Color	Shape	DeCAF	All
Sashimi	0.192	0.236	0.132	0.169	0.130	0.264	0.208	**0.123**	0.125
Curry and rice	0.153	0.164	0.118	0.109	0.120	0.165	0.096	0.092	**0.087**
Eel rice-bowl	0.088	0.110	0.077	0.115	0.077	0.173	0.069	**0.061**	0.068
Beef stew	0.195	0.155	0.140	0.149	0.133	0.158	0.154	**0.084**	0.086
Hamburger steak	0.186	0.152	0.126	0.171	0.118	0.264	0.158	0.097	**0.095**
Tempura rice-bowl	0.183	0.158	**0.101**	0.138	0.112	0.279	0.235	0.127	0.123
Fried pork rice-bowl	0.160	0.102	0.100	0.119	0.095	0.244	0.114	**0.094**	0.098
Tuna rice-bowl	0.038	**0.032**	0.039	0.039	0.039	0.196	0.059	0.055	0.055
Cheese burger	0.095	0.084	0.118	0.148	0.117	0.219	0.068	**0.065**	0.068
Fish burger	0.099	0.159	0.059	0.130	**0.057**	0.285	0.201	0.104	0.107
Average	0.139	0.135	0.101	0.129	0.100	0.225	0.136	**0.090**	0.091

realizing such an idea is to recognize the food categories of input images. This approach will construct attractiveness estimators specific for various food categories in advance, and selects one corresponding to the food category of an input image. For example, Hassannejad et al. reported that the recognition accuracy of 88.28%, 81.45%, and 76.17% were achieved as top-1 accuracies on ETH Food-101, UEC FOOD 100, and UEC FOOD 256, respectively [4]. The approach would be, however, not realistic if we needed to switch the estimators because the number of food categories is almost uncountable, and various appearances of ingredients and toppings exist even within a food category. Thus, it would be better to switch depending on the food appearance on the color and shape features instead of the food category.

5 CONCLUSION

This paper proposed a method for estimating the attractiveness of food photos. The proposed method integrated two kinds of image features: the appearance of the main ingredient and the impression of the entire food photo. Also, an image dataset (NU FOOD 360x10[1]) for food sample photos with their attractiveness values was built through subjective experiments. We confirmed the effectiveness of the proposed method, and suggested the necessity for adaptively switching attractiveness estimators.

Future work includes the study on a realistic and effective way of switching estimators for more accurate estimation. In addition, we will focus on other photography parameters such as zooming, lighting, and blurring.

ACKNOWLEDGMENTS

Parts of this research were supported by JSPS Grant-in-Aid for Scientific Research and MSR-Core12 program. We would like to thank Drs. Tao Mei and Jianlong Fu at Microsoft Research Asia for providing evaluation results on our dataset based on [15]. We would also like to thank the subjects for participating in the experiment, and Mr. Tatsumi Hattori at Chukyo University for helping us with the evaluation.

REFERENCES

[1] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In *Proc. 2009 IEEE Conference on Computer Vision and Pattern Recognition*. 248–255.

[2] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. 2014. DeCAF: A deep convolutional activation feature for generic visual recognition. In *Proc. 31st International Conference on Machine Learning*. 647–655.

[3] Giovanni Maria Farinella, Dario Allegra, Marco Moltisanti, Filippo Stanco, and Sebastiano Battiato. 2016. Retrieval and classification of food images. *Computers in Biology and Medicine* 77 (2016), 23–39.

[4] Hamid Hassannejad, Guido Matrella, Paolo Ciampolini, Ilaria De Munari, Monica Mordonini, and Stefano Cagnoni. 2016. Food image recognition using very deep convolutional networks. In *Proc. 2nd International Workshop on Multimedia Assisted Dietary Management*. 41–49.

[5] Takao Kakimori, Makoto Okabe, Keiji Yanai, and Rikio Onai. 2015. A system to support the amateurs to take a delicious-looking picture of foods. In *Proc. Symp. on Mobile Graphics and Interactive Applications at SIGGRAPH Asia 2015*. 28.

[6] Andy Liaw and Matthew Wiener. 2002. Classification and regression by randomForest. *R News* 2, 3 (Dec. 2002), 18–22.

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

[8] Charles Michel, Andy T. Woods, Markus Neuhäuser, Alberto Landgraf, and Charles Spence. 2015. Rotating plates: Online study demonstrates the importance of orientation in the plating of food. *Food Quality and Preference* 44 (Sept. 2015), 194–202.

[9] Masashi Nishiyama, Takahiro Okabe, Imari Sato, and Yoichi Sato. 2011. Aesthetic quality classification of photographs based on color harmony. In *Proc. 2011 IEEE Conference on Computer Vision and Pattern Recognition*. 33–40.

[10] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. 2004. GrabCut — Interactive foreground extraction using iterated graph cuts. *ACM Trans. on Graphics — Proc. ACM SIGGRAPH 2004* 23, 3 (Aug. 2004), 309–314.

[11] Syohei Sakiyama, Makoto Okabe, and Rikio Onai. 2014. Animating images of cooking using video examples and image deformation. In *Mathematical Progress in Expressive Image Synthesis I (Mathematics for Industry)*, Vol. 4. Springer Japan, 171–176.

[12] Charles Spence and Betina Piqueras-Fiszman. 2014. *The perfect meal: The multisensory science of food and dining.* WILEY Blackwell.

[13] Kazuma Takahashi, Keisuke Doman, Yasutomo Kawanishi, Takatsugu Hirayama, Ichiro Ide, Daisuke Deguchi, and Hiroshi Murase. 2016. A study on estimating the attractiveness of food photography. In *Proc. 2nd IEEE International Conference on Multimedia Big Data*. 444–449.

[14] L. L. Thurstone. 1927. Psychophysical analysis. *American Journal of Psychology* 38, 3 (July 1927), 368–389.

[15] Xinmei Tian, Zhe Dong, Kuiyuan Yang, and Tao Mei. 2015. Query-dependent aesthetic model with deep learning for photo quality assessment. *IEEE Trans. on Multimedia* 17, 11 (Nov. 2015), 2035–2048.