LETTER

Estimation of the Attractiveness of Food Photography Based on Image Features

Kazuma TAKAHASHI[†*], **Tatsumi HATTORI**[††a)], ***Nonmembers***, **Keisuke DOMAN**[††b)],
Yasutomo KAWANISHI[†††c)], **Takatsugu HIRAYAMA**[††††d)], ***Members***, **Ichiro IDE**[†††e)], ***Senior Member***,
Daisuke DEGUCHI[†††††f)], ***Member***, ***and*** **Hiroshi MURASE**[†††g)], ***Fellow***

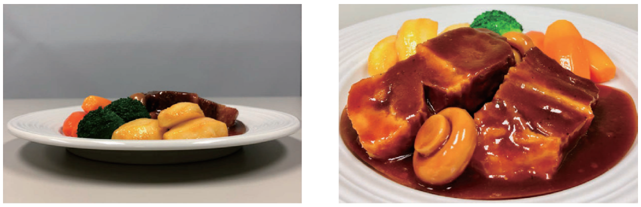
SUMMARY We introduce a method to estimate the attractiveness of a food photo. It extracts image features focusing on the appearances of 1) the entire food, and 2) the main ingredients. To estimate the attractiveness of an arbitrary food photo, these features are integrated in a regression scheme. We also constructed and released a food image dataset composed of images of ten food categories taken from 36 angles and accompanied with attractiveness values. Evaluation results showed the effectiveness of integrating the two kinds of image features.

key words: *food photography, attractiveness, photographic framing*

1. Introduction

Many food photos are posted on the Web such as social media and cooking recipe sites. Their users would prefer to upload delicious-looking (hereafter, attractive) food photos to attract social attention. For example, Fig. 1 (b) would most likely attract more viewers than Fig. 1 (a) because of camera angle and photographic framing, although these are photos of the same dish. Since these kinds of decision are not always easy for an amateur photographer to make, we are trying 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 food photo from a list. For such purposes, this paper proposes a method that quantifies the attractiveness of a given food photo.

Most previous research on food image understanding studies the task of retrieval and classification [1]. Meanwhile, there is research on the classification of the aesthetic quality of general photos into two levels: high or low. Tian et al. proposed a method that constructs a classification model for each query image using Deep Convolutional Neural Networks (DCNNs) [2], and targets general photos. The aesthetic quality is related but different from the food attractiveness, and their method does not consider food-specific attractiveness discussed in [3]. For photography support considering food-specific attractiveness, Kakimori et al. developed a system that presents a user a guideline for arranging dishes in photographic framing [4]. Although this 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 [5], and the rotation angle, in particular, is one of the key factors when deciding the photographic framing. Focusing on this point, in this paper, we propose a method for estimating the attractiveness of food photos by integrating two kinds of image features.



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

Fig. 1 Photographic framing of a dish.

This paper summarizes our work presented in [7] as an extended work of [6], and is organized as follows. Section 2 describes the details of the proposed method. Then, dataset construction through subjective experiments is introduced in Sect. 3. Next, the evaluation of the proposed method is reported in Sect. 4. Finally, Sect. 5 concludes this paper.

2. Attractiveness Estimation Method

Figure 2 shows the process-flow of the proposed method: The training step constructs an attractiveness estimator using food images accompanied with attractiveness values in a regression framework, while the estimation step estimates the attractiveness of an input image using the estimator. Both steps use several image features extracted from an input image, which reflect the appearance difference caused by the difference of the camera angle, and should be suitable

Manuscript received October 18, 2018.
Manuscript revised February 24, 2019.
Manuscript publicized May 7, 2019.

[†]The author is with Graduate School of Information Science, Nagoya University, Nagoya-shi, 464–8601 Japan.

[††]The authors are with Graduate School of Engineering, Chukyo University, Toyota-shi, 470–0393 Japan.

[†††]The authors are with Graduate School of Informatics, Nagoya University, Nagoya-shi, 464–8601 Japan.

[††††]The author is with Institute of Innovation for Future Society, Nagoya University, Nagoya-shi, 464–8601 Japan.

[†††††]The author is with Information Strategy Office, Nagoya University, Nagoya-shi, 464–8601 Japan.

[*]Presently, with Fuji Xerox Co., Ltd.

a) E-mail: hattori.t@md.sist.chukyo-u.ac.jp
b) E-mail: kdoman@sist.chukyo-u.ac.jp
c) E-mail: kawanishi@i.nagoya-u.ac.jp
d) E-mail: takatsugu.hirayama@nagoya-u.jp
e) E-mail: ide@i.nagoya-u.ac.jp
f) E-mail: ddeguchi@nagoya-u.jp
g) E-mail: murase@i.nagoya-u.ac.jp

DOI: 10.1587/transinf.2018EDL8219

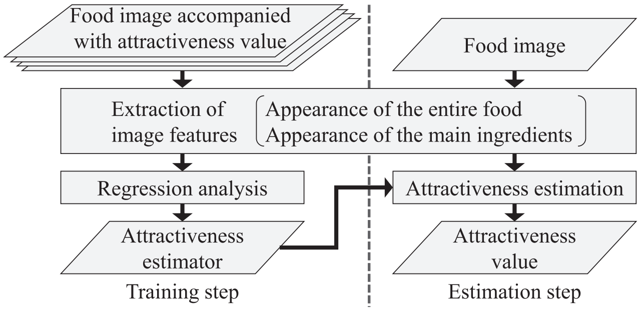

Food image accompanied with attractiveness value
Food image
Extraction of image features
Appearance of the entire food
Appearance of the main ingredients
Regression analysis
Attractiveness estimation
Attractiveness estimator
Attractiveness value
Training step
Estimation step

Fig. 2 Process-flow of the attractiveness estimation method.

for the attractiveness estimation. The details are described below.

2.1 Region for Image Feature Extraction

By using an algorithm such as GrabCut [8], an input image is segmented into two regions: 1) the dish region R_d for extracting the image features to evaluate the appearance of the entire food, and 2) the main ingredients region R_m for extracting the image features to evaluate the appearance of the main ingredient. We expect that the latter will be selected manually through user interaction.

2.2 Image Features: Appearance of the Entire Food

The following image features are extracted from the dish region R_d in an input image.

- **Color Feature $\boldsymbol{C}$**: The color distribution in R_d is represented by $\boldsymbol{C} = (c_1, c_2, \ldots, c_{100})$ where c_i $(i = 1 \ldots 100)$ is the frequency-weighted Euclidean distance from the most frequent CIELAB color in R_d to that in a local region i obtained by radially dividing the input image.
- **Shape Feature $\boldsymbol{E}$**: The edge strength in R_d is represented by $\boldsymbol{E} = (e_1, e_2, \ldots, e_{100})$ where e_j $(j = 1 \ldots 100)$ is the maximum edge strength in a block j obtained by equally dividing the input image.
- **Color and Shape Feature $\boldsymbol{A}$**: The appearance of the food is represented by $\boldsymbol{A}$, a 4,096-dimensional Deep Convolutional Activation Feature (DeCAF) [9].

2.3 Image Features: Appearance of the Main Ingredients

The following image features are extracted from the main ingredients region R_m in an input image.

- **Size Feature S**: The apparent size of the main ingredients is represented by the area ratio S of R_m to R_d.
- **Position Features P_x and P_y**: The relative position of the main ingredients is represented by the x- and y-directional differences, P_x and P_y, respectively, between the gravity centers of R_d and R_m.
- **Shape Feature $\boldsymbol{O}$**: The edge orientation of the main ingredients is represented by a 36-bin orientation histogram $\boldsymbol{O} = (O_1, O_2, \ldots, O_{36})$.
- **Moment Feature $\boldsymbol{M}$**: The orientation statistics of the main ingredients are represented by the first to the fourth central moments $\boldsymbol{M} = (M_1, M_2, M_3, M_4)$ of $\boldsymbol{O}$, where M_1, M_2, M_3, and M_4 are the average, the variance, the skewness, and the kurtosis of $\boldsymbol{O}$, respectively.

2.4 Training and Estimation

Random Regression Forests [10] is used as an estimator based on regression in which the objective variable is the attractiveness value of a food photo and the explanatory variables are the image features $(\boldsymbol{C}, \boldsymbol{E}, \boldsymbol{A}, S, P_x, P_y, \boldsymbol{O}, \boldsymbol{M})$. Within the regression framework, once the relation between the attractiveness values and the image features is trained, the regressor can estimate the attractiveness value of an arbitrary input image only from its image features.

3. Dataset Construction through Subjective Experiments

We conducted subjective experiments to construct an image dataset accompanied with attractiveness values for constructing the attractiveness estimator, and released it to the public[†]. The dataset is composed of ten food categories: 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, considering the variation of the appearance in both color and shape. Details of the experimental method and results are described below.

3.1 Photographing Method

We shot photos of plastic food samples instead of real ones considering both convenience and reproducibility, from various 3D-angles while keeping a fixed distance between the camera and the sample. Here, we shot from three elevation angles: 30, 60, and 90 deg., where 0 and 90 deg. correspond to shooting from the side and the top of the dish, respectively. We set an arbitrary rotation angle as 0 deg., and then shot from 0 to 330 deg. with the step of 30 deg. in a clockwise direction around the center of the sample. As a result, we obtained 36 food photos in total for each food category.

3.2 Determination of Attractiveness Values by Paired Comparison

We used Thurstone's paired comparison method [11] to determine the attractiveness values of food photos. This method was developed for sensory tests, and is used to determine an interval scale for perceived quality. An image pair out of ${}_{36}\mathrm{C}_2 = 630$ pairs for each food category was presented at a time to participants who were asked to respond with which image looked more delicious. Participants were 28 Computer Science-major students in their 20s, out

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

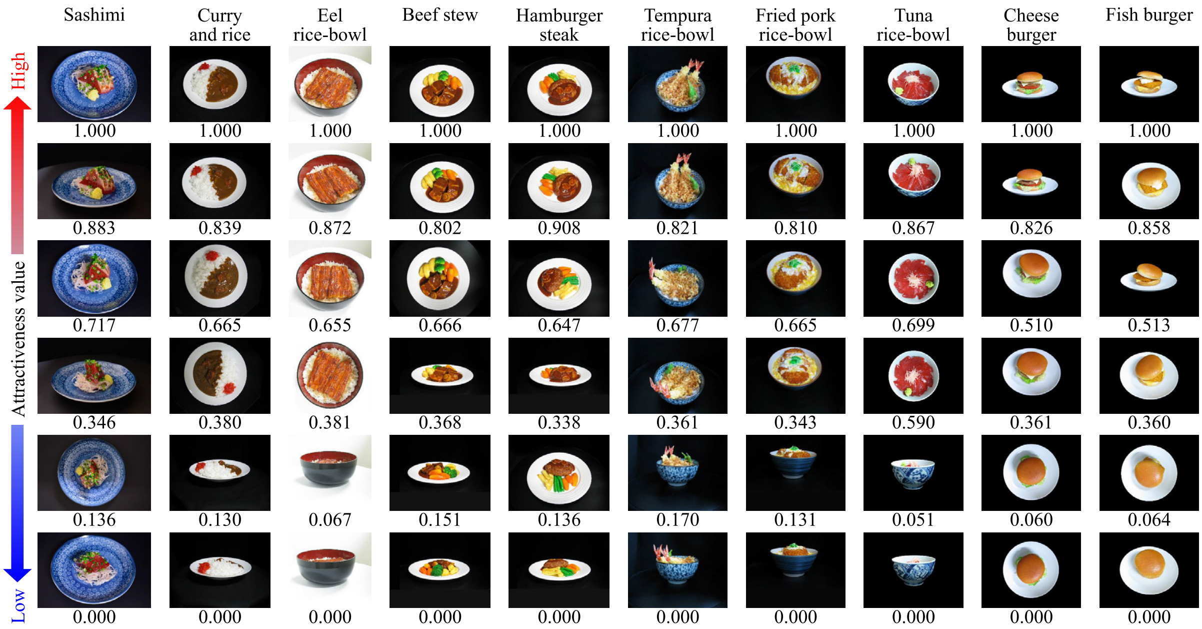


Fig. 3 Attractiveness values for each image in each food category.

of which nine subjects were assigned for each food category. We finally obtained three to four responses for each image pair resulting in 2,150 responses in total for each food category.

The obtained attractiveness values normalized into the range of [0, 1] are shown in Fig. 3, which are used as target values for the regression in the proposed method.

4. Evaluation Experiment

We evaluated the effectiveness of the proposed method through experiments.

4.1 Method

We applied a leave-one-out scheme with the dataset described in Sect. 3. The dish and the main ingredients regions for feature extraction were precisely labeled by using GrabCut [8] and manual selection for each food category in order to prevent its effect on the estimation accuracy, considering the purpose of this research.

We compared the estimation accuracy of the proposed method with that of a comparative method based on [2], which was originally designed to classify the aesthetic quality of general photos using DCNNs, in order to confirm the necessity of a food-specific method for attractiveness estimation. For each method, we evaluated the Mean Absolute Error (MAE) between the estimated and the target values for the attractiveness of food photos.

4.2 Results

The results are summarized in Table 1. The average MAE of the proposed method was 0.087, whereas that of the comparative one was 0.344. The proposed method outperformed the comparative one for all food categories. From the results, we can conclude the following important points: 1) the necessity for considering the attractiveness specific to food photos, and 2) the effectiveness of integrating the appearances of both the entire food and the main ingredients.

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

Category	Tian et al. [2]	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

For reference, the relations between the target and the estimated values for Eel rice-bowl and Fish burger are shown in Fig. 4. We can see that the target values greatly change depending on the camera angle, and the estimated values accurately follow the target values for every camera angle.

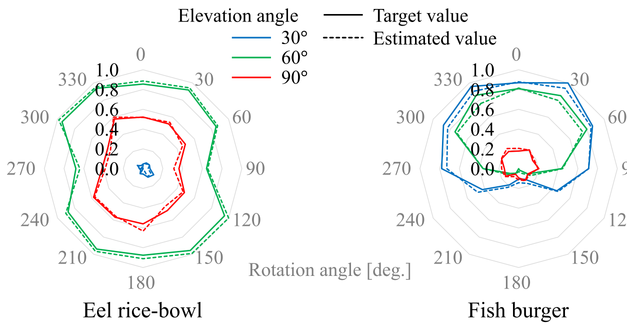
The performance of the proposed method (MAE = 0.087) shows that on average, it can be applied to a system that evaluates the attractiveness in ten levels, where the absolute error should be less than 0.1 for at worst one-level evaluation error.

4.3 Discussion

We investigate the effectiveness of each image feature. Table 2 shows the MAEs when using only one of the image

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



Eel rice-bowl

Fish burger

Fig. 4 Relation between the target and the estimated values.

features. This table also includes the MAEs when using all the image features together, denoted as "All".

The average MAE when using only DeCAF was 0.090, which was the best among all nine features including "All". The second best excluding "All" was the orientation of the main ingredients named "Shape". In some cases, the highest accuracy was obtained by using one or some of the image features, which showed that the proposed image features were suitable for attractiveness estimation. In addition, the effective image feature depended on the food category. This suggests that more accurate estimation could be achieved by switching the attractiveness estimators (i.e. the combination of the image features) based on the result of food category recognition or appearance clustering of an input image.

5. Conclusion

We proposed a method for estimating the attractiveness of food photos. The proposed method integrated two kinds of image features: the appearances of the entire food and the main ingredients. Also, an image dataset for food sample photos accompanied with target attractiveness values assigned through subjective experiments was constructed and released. Through an evaluation experiment, 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 KAKENHI 16H02846 and 17K12719, and MSR-Core12 program. We would like to thank Drs. Tao Mei and Jianlong Fu at Microsoft Research Asia for providing results of their method [2] on our dataset.

References

[1] G.M. Farinella, D. Allegra, M. Moltisanti, F. Stanco, and S. Battiato, "Retrieval and classification of food images," Comput. Biol. Med., vol.77, pp.23–39, 2016.

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

[3] C. Spence and B. Piqueras-Fiszman, The perfect meal: The multisensory science of food and dining, Wiley Blackwell, July 2014.

[4] T. Kakimori, M. Okabe, K. Yanai, and R. Onai, "A system to support the amateurs to take a delicious-looking picture of foods," Proc. Symp. on Mobile Graphics and Interactive Applications at SIGGRAPH Asia 2015, p.28, Nov. 2015.

[5] C. Michel, A.T. Woods, M. Neuhäuser, A. Landgraf, and C. Spence, "Rotating plates: Online study demonstrates the importance of orientation in the plating of food," Food Qual. Prefer., vol.44, pp.194–202, Sept. 2015.

[6] K. Takahashi, K. Doman, Y. Kawanishi, T. Hirayama, I. Ide, D. Deguchi, and H. Murase, "A study on estimating the attractiveness of food photography," Proc. 2nd IEEE Int. Conf. on Multimedia Big Data, pp.444–449, April 2016.

[7] K. Takahashi, K. Doman, Y. Kawanishi, T. Hirayama, I. Ide, D. Deguchi, and H. Murase, "Estimation of the attractiveness of food photography focusing on main ingredients," Proc. 9th Workshop on Multimedia for Cooking and Eating Activities, pp.1–6, Aug. 2017.

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

[9] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, "DeCAF: A deep convolutional activation feature for generic visual recognition," Proc. 31st Int. Conf. on Machine Learning, pp.647–655, June 2014.

[10] A. Liaw and M. Wiener, "Classification and regression by randomForest," R News, vol.2, no.3, pp.18–22, Dec. 2002.

[11] L.L. Thurstone, "Psychophysical analysis," Am. J. Psychol., vol.38, no.3, pp.368–389, July 1927.