Tastes and Textures Estimation of Foods Based on the Analysis of Its Ingredients List and Image

Hiroki Matsunaga[1], Keisuke Doman[1,2], Takatsugu Hirayama[1], Ichiro Ide[1(✉)], Daisuke Deguchi[1,3], and Hiroshi Murase[1]

[1] Graduate School of Information Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan
matsunagah@murase.m.is.nagoya-u.ac.jp, kdoman@sist.chukyo-u.ac.jp,{hirayama,ide,murase}@is.nagoya-u.ac.jp, ddeguchi@nagoya-u.jp

[2] School of Engineering, Chukyo University, 101 Tokodachi, Kaizu-cho, Toyota 470-0393, Japan

[3] Information Strategy Office, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan

Abstract. Recently, the number of cooking recipes on the Web is increasing. However, it is difficult to search them by tastes or textures although they are actually important considering the nature of the contents. Therefore, we propose a method for estimating the tastes and the textures of a cooking recipe by analyzing them. Concretely, the proposed method refers to an ingredients feature from the "ingredients list" and image features from the "food image" in a cooking recipe. We confirmed the effectiveness of the proposed method through an experiment.

1 Introduction

Recently, the number of cooking recipes on the Web is increasing. An example of a cooking recipe posted on the Web is shown in Fig. 1. Currently, users would usually search from a large number of cooking recipes for those that suit their requirements by means of keywords matching with the recipe title or the list of ingredients. However, it is difficult to search cooking recipes by tastes or textures although they are actually important considering the nature of the contents. Labeling each recipe with its tastes and textures could be a solution, but we cannot expect all users who post cooking recipes on the Web to do so.

As related work, a tastes sensor has been developed by Tahara and Toko [8]. It imitates the biological effects on the surface of the human tongues, and measures the tastes of food in the aspect of the five basic tastes; *sweet*, *sour*, *salty*, *bitter*, and *umami*. However, normal users that post cooking recipes on the Web cannot make use of this sensor casually, since it is very expensive. In addition, it cannot measure textures.

H. Matsunaga—Currently at IVIS, Inc.

V. Murino et al. (Eds.): ICIAP 2015 Workshops, LNCS 9281, pp. 326–333, 2015.
DOI: 10.1007/978-3-319-23222-5_40

Juicy Japanese-style hamburger

Ingredients list

Minced beef	160–200 g
Leek	1/2
Shiitake mushrooms	4
Egg	1
Butter	1 piece

Preparation steps

1. Mince the leek and the Shiitake mushrooms.	**2.** Warm the pan over medium heat, and melt the butter.	**3.** Add **1.** in **2.** and stir. Add salt to let the moist evaporate.	**4.** Put into a plate to cool it down, once it gets starchy.
5. After 10 min., put the minced meet, egg, and **4.** in a bowl and mix them well with a fork. Season with salt, pepper, and soy sauce.	**6.** After leaving 10 min., divide it in two with a fork and knead two putties, and then fry them on the pan.	**7.** Reverse them and fry, once the meat juice starts simmering around the rim of the putties.	**8.** Finally, dish up the hamburgers into a plate.

Fig. 1. Example of a cooking recipe posted on the Web[1] by one of the authors.

Thus, we are aiming at estimating the tastes and the textures of a food by analyzing cooking recipes. Concretely, the proposed method refers to an ingredients feature from the "ingredients list" and image features from the "food image" in a cooking recipe.

In the following sections, we first introduce the proposed method on tastes estimation in Sect. 2, and then report its results in Sect. 3.1. In addition, we also report the result of applying the same scheme to textures estimation in Sect. 3.2. Finally, we conclude the paper in Sect. 4.

2 Tastes Estimation Method

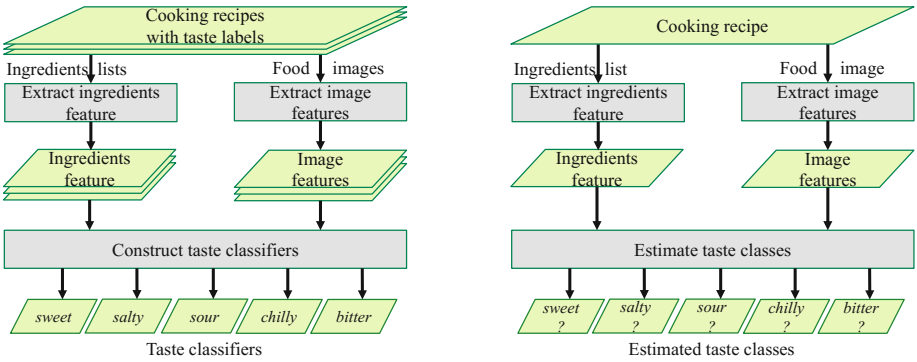
As shown in Fig. 1, a typical cooking recipe posted on the Web is composed of a "title", a "food image", an "ingredients list", and "preparation steps". The proposed method estimates the tastes of a food in two steps referring to the "ingredients list" and the "food image"; the training step and the estimation step, as shown in Fig. 2 and described below. Note that here, we assume that the structure of a cooking recipe could be automatically analyzed, and the "food image" and the "ingredients list" are available for immediate processing.

2.1 Training Step

Classifiers are constructed for each taste class. Each classifier is a one-versus-rest classifier that judges whether the food has the specific taste or not.

The process flow of the training step is shown in Fig. 2(a). First, an ingredients feature is extracted from the "ingredients list". Next, image features are extracted from the "food image". Classifiers for each taste class are constructed using these two features extracted from a large-number of cooking recipes with taste labels.

[1] Translated from http://cookpad.com/recipe/1452708/.



(a) Training step (b) Estimation step

Fig. 2. Process flow of the proposed method.

Ingredients Feature. First, an ingredients dictionary is built by accumulating all ingredients that appear in a cooking recipe dataset. Next, an ingredients feature vector is formed for each cooking recipe, as a binary vector with the value 1 for ingredients that appear in the recipe, and 0 for all the others.

Image Features. First, in order to extract image features precisely, regions that include plates and tables are cropped. Here, GrabCut [7] was employed following the practice in the food recognition system proposed by Kawano et al. [5]. In our case, GrabCut is given the entire input image as the initial region. Next, as image features, Hue-Saturation histogram, Hue-Saturation correlogram [4], BoF representation [1] of SIFT features [6], and HOG [2] are extracted.

Taste Classifiers. An SVM [9] classifier for each taste class is constructed, that learns the features extracted from the cooking recipes with manually labeled taste labels.

2.2 Estimation Step

As in the training step, an ingredients feature and image features are extracted from an input cooking recipe, and then each of the trained SVM classifiers estimates if it has the corresponding taste or not.

3 Experiments

We evaluated the effectiveness of the proposed method through tastes estimation experiments in Sect. 3.1. In addition, for reference, we also report preliminary results on the application of the same scheme to textures in Sect. 3.2.

Table 1. Number of cooking recipes labeled with taste classes by subjects.

Taste class	*sweet*	*sour*	*chilly*	*salty*	*bitter*
Number of cooking recipes	1,254	366	241	537	213

Table 2. Number of cooking recipes labeled with taste classes by referring to user comments.

Taste class	*sweet*	*sour*	*chilly*	*salty*	*bitter*
Number of cooking recipes	4,849	1,093	907	495	362

3.1 Tastes Estimation Experiments

Construction of Datasets. We first constructed a dataset of cooking recipes labeled by human subjects through a subjective experiment. First, 2,700 cooking recipes were randomly selected from 440,000 cooking recipes in the "Rakuten recipe dataset"[2]. Then, 45 Japanese male and female subjects were asked to label them with taste classes. Here, each subject was presented the "title", the "food image", and the "ingredients list" of 60 cooking recipes, and was asked to choose up to two out of five taste classes; *sweet*, *sour*, *chilly*, *salty*, and *bitter*. The subjects were also allowed to choose "unknown" if they could not decide, in which case, the corresponding cooking recipe was excluded from the dataset. As a result, we obtained 1,827 cooking recipes labeled with taste classes. The number of the cooking recipes labeled with each taste class in this dataset is shown in Table 1.

However, since manual labeling requires sufficient amount of man-power, it is difficult to create a larger dataset. Thus, we also constructed a second dataset by referring to expressions related to each taste class from user comments posted to each cooking recipe. This allows us to create a larger dataset automatically, although it may degrade the reliability of the labels.

First, morphological analysis was applied to the comments. Next, each word was matched with a dictionary of taste-related expressions for each taste class, which we prepared manually beforehand. If a match was found, the corresponding taste class was labeled, and if not, the cooking recipe was excluded from the dataset. Note that multiple labeling was allowed. As a result, we obtained 7,706 cooking recipes with taste labels as the second dataset. The number of cooking recipes labeled with each taste class in this dataset is shown in Table 2.

Experimental Method. We conducted experiments to evaluate the effectiveness of the proposed method using each of the two datasets. When constructing a classifier for a taste class, cooking recipes labeled with the corresponding taste

[2] Rakuten Inc., "Rakuten datasets," http://www.nii.ac.jp/cscenter/idr/rakuten/rakuten.html

label were used as positive samples, while all the others were used as negative samples. Since there was a large difference between the numbers of positive and negative samples in the datasets, each class was weighted according to the inverse of the sample size in the SVM training step. We compared the proposed method with two comparative methods; one that used only the ingredients feature, and one that used only image features. Each method was evaluated through eight-fold cross validation. Precision, recall, and F-measure were used as the evaluation criteria.

Results. The experimental results from the first dataset labeled by the subjects are shown in Table 3. From the results, we confirmed the effectiveness of the proposed method for all the taste classes.

The experimental results from the second dataset labeled by referring to user comments are likewise shown in Table 4. Since the results were similar and sometimes better than those obtained from the first dataset, we considered that the larger size of the dataset contributed more than the degradation of the labels.

Discussion. For the *salty* class labeled by subjects, the highest F-measure was obtained when only the ingredients feature was used. A representative ingredient that causes a food to become *salty* will be "salt". Actually, many cooking recipes labeled as *salty* contained "salt". This would be a reason that the ingredients feature was effective to estimate the *salty* class, while it lead to lower accuracy when using only the image feature, since "salt" is usually not visually perceivable. Thus, selecting different features for each class, could improve the accuracy.

3.2 Application to Textures Estimation

In order to evaluate if the proposed scheme could also be applied to textures estimation, we performed a similar experiment as that in Sect. 3.1 for textures. According to a research by Hayakawa et al. [3], there are 445 texture expressions in the Japanese language. Here, we targeted the following five texture expressions that were most frequently used in the comments; *shaki-shaki*, *fuwa-fuwa*, *toro-toro*, *saku-saku*, and *hoku-hoku*.

Construction of a Dataset. To label the cooking recipes with the five texture classes, we applied the same procedure as that for the second dataset created in Sect. 3.1 that was labeled by referring to user comments. As a result, we obtained 5,219 cooking recipes with texture labels. The number of cooking recipes labeled with each texture class is shown in Table 5.

Experimental Method. We conducted an experiment to evaluate the applicability of the proposed scheme to textures estimation using the dataset, in the same manner as in Sect. 3.1.

Table 3. Estimation results from taste classes labeled by subjects.

(a) *sweet* class

Method	Precision	Recall	F-measure
Proposed method	0.813	0.838	**0.825**
Ingredients feature	**0.818**	0.828	0.822
Image features	0.701	**0.928**	0.798

(b) *sour* class

Precision	Recall	F-measure
0.405	0.390	**0.397**
0.393	0.336	0.362
0.209	**0.672**	0.319

(c) *chilly* class

Method	Precision	Recall	F-measure
Proposed method	**0.393**	0.220	**0.282**
Ingredients feature	0.325	**0.227**	0.256
Image features	0.227	0.104	0.142

(d) *salty* class

Precision	Recall	F-measure
0.538	**0.545**	0.542
0.561	0.533	**0.547**
0.337	0.384	0.359

(e) *bitter* class

Method	Precision	Recall	F-measure
Proposed method	**0.409**	0.399	**0.404**
Ingredients feature	0.342	**0.418**	0.376
Image features	0.246	0.192	0.216

Table 4. Estimation results from taste classes labeled by referring to user comments.

(a) *sweet* class

Method	Precision	Recall	F-measure
Proposed method	**0.755**	**0.844**	**0.797**
Ingredients feature	0.743	0.837	0.787
Image features	0.706	0.703	0.705

(b) *sour* class

Precision	Recall	F-measure
0.408	0.410	**0.409**
0.552	0.282	0.373
0.167	**0.485**	0.243

(c) *chilly* class

Method	Precision	Recall	F-measure
Proposed method	0.511	0.260	0.345
Ingredients feature	**0.576**	0.294	**0.388**
Image features	0.196	**0.615**	0.298

(d) *salty* class

Precision	Recall	F-measure
0.398	0.225	**0.287**
0.348	0.091	0.144
0.089	**0.503**	0.152

(e) *bitter* class

Method	Precision	Recall	F-measure
Proposed method	0.680	0.350	**0.462**
Ingredients feature	**0.777**	0.329	**0.462**
Image features	0.086	**0.439**	0.144

Table 5. Number of the cooking recipes labeled with texture classes by referring to user comments.

Texture class	*shaki-shaki*	*fuwa-fuwa*	*toro-toro*	*saku-saku*	*hoku-hoku*
Number of cooking recipes	1,445	1,353	843	828	750

Table 6. Estimation accuracy for texture classes labeled by referring to user comments.

(a) *shaki-shaki* class

Method	Precision	Recall	F-measure
Proposed method	0.767	0.689	0.726
Ingredients feature	**0.778**	**0.691**	**0.732**
Image features	0.487	0.544	0.514

(b) *fuwa-fuwa* class

Precision	Recall	F-measure
0.708	0.593	**0.645**
0.702	0.593	0.643
0.317	**0.678**	0.432

(c) *toro-toro* class

Method	Precision	Recall	F-measure
Proposed method	0.282	0.507	0.363
Ingredients feature	**0.289**	0.547	**0.378**
Image features	0.207	**0.603**	0.310

(d) *saku-saku* class

Precision	Recall	F-measure
0.642	0.465	**0.539**
0.639	0.448	0.526
0.245	**0.587**	0.346

(e) *hoku-hoku* class

Method	Precision	Recall	F-measure
Proposed method	0.771	0.598	0.650
Ingredients feature	**0.773**	0.601	**0.660**
Image features	0.224	**0.649**	0.333

Results. The experimental results are shown in Table 6. Compared with the experimental results of the tastes estimation in Sect. 3.1, the F-measures were in the same level, so we confirmed that the proposed method could also be applied to textures estimation.

Discussion. For some texture classes, the highest F-measure was obtained when only the ingredients feature was used. We found that in many cases, cooking recipes that were selected as positive samples in each texture class were those on a specific dish. For example, most cooking recipes on a "salad" were labeled with the *shaki-shaki* class. Some dishes often share the same ingredients and follow similar preparation steps, but they could have various visual appearances, so indeed image features may not necessarily be effective in such cases.

In this experiment, we selected only five out of the 445 texture expressions, so in order to truly confirm the effectiveness of the proposed scheme for textures

estimation, we need to extend the number of texture classes. However, it may be difficult to do so due to the insufficient numbers of positive samples available.

4 Conclusion

We proposed an estimation method of tastes and textures from cooking recipes. The proposed method analyzed the text feature extracted from the "ingredients list" and the image features extracted from the "food image" in a cooking recipe.

Experimental results showed the effectiveness of the proposed method for all taste classes. The proposed scheme also showed its extensibility to textures estimation. Future work includes introducing additional information, such as the "preparation steps" and the quantity of ingredients.

Acknowledgments. Part of this work was supported by Grant-in-Aid for Scientific Research (24240028). We thank Rakuten Inc. for providing their recipe contents.

References

1. Csurka, G., Bray, C., Dance, C., Fan, L.: Visual categorization with bags of keypoints. In: Proc. ECCV 2004 Workshop on Statistical Learning in Computer Vision, pp. 59–74 (May 2004)
2. Dalal, N., Triggs, W.: Histograms of oriented gradients for human detection. In: Proc. 2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pp. 886–893 (June 2005)
3. Hayakawa, F., Kazami, Y., Nishinari, K., Ioku, K., Akuzawa, S., Yamano, Y., Baba, Y., Kohyama, K.: Classification of Japanese texture terms. J. of Texture Studies **44**(2), 140–159 (2013)
4. Huang, J., Kumar, S.R., Mitra, M., Jing, W., Zabih, Z.: Image indexing using color correlogram. In: Proc. 1997 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pp. 762–768 (June 1997)
5. Kawano, Y., Yanai, K.: Foodcam: A real-time food recognition system on a smartphone. Multimedia Tools and Applications, 1–27 (April 2014)
6. Lowe, D.: Object recognition from local scale-invariant features. In: Proc. 1999 IEEE Int. Conf. on Computer Vision, pp. 1150–1157 (September 1999)
7. Rother, C., Kolmogorov, V., Blake, A.: Grabcut: Interactive foreground extraction using iterated graphcuts. ACM Trans. on Graphics **23**(3), 309–314 (2004)
8. Tahara, Y., Toko, K.: Electronic tongues –A review. IEEE Sensors J. **13**(8), 3001–3011 (2013)
9. Vapnik, V.: The nature of statistical learning theory. Springer, New York (1998)