A system to help amateurs take pictures of delicious looking food

Takao Kakimori kakimori@mm.inf.uec.ac.jp
Makoto Okabe* m.o@acm.org
Keiji Yanai yanai@cs.uec.ac.jp
Rikio Onai rikioonai@gmail.com
The University of Electro-Communications / JST, CREST*

***Abstract*—Recently, many people have begun to take pictures of meals and food either at home or in restaurants. These pictures are then uploaded to social networking services (SNS) where they are shared with friends. People want to take pictures of food that looks delicious, but they often find this difficult. This is because most people lack the knowledge required to take attractive pictures. There are many photography techniques in use, e.g., composition [1], lighting, color, focus, etc. The techniques used to take good pictures vary depending on the subject. Amateur photographers find it difficult to choose techniques and apply them appropriately. In this paper, we consider the composition of food photographs and develop a system to support amateurs taking pictures of meals and food to make the food look delicious. Our target users are food photography amateurs. Our target photographic subjects are food items on plates or dishes. Using our system, there are four steps to food photography: 1) the user provides information about the foods to be photographed, or our system automatically recognizes these food items, with the aid of a camera on a mobile phone; 2) our system suggests a composition and camera tilt that will result in a picture that makes the food look delicious; 3) the user arranges the food and dishes on the table, and sets the camera position and tilt; 4) finally, the user takes the picture. If the user is not satisfied with the suggestion, we allow the user to design a new composition quickly and easily using their mobile phone. We performed a usability study for our system followed by a subjective evaluation of the quality of the pictures taken using our system.**

I. Introduction

Most people have a mobile phone these days, and people enjoy taking pictures of their food and immediately uploading and sharing them with friends via SNS, e.g., Twitter and Facebook. Additionally, an owner of a restaurant can take pictures of their dishes to use on advertising materials or menus. These people are often amateur photographers; however, since their pictures will be seen and evaluated by the others, they want the food to look as delicious as possible.

A professional photographer knows how to adjust composition [1], lighting, color, and focus, etc., so that they take attractive pictures of food. In this paper, we consider composition, and develop a system to suggest a desirable composition to the user. The composition of a picture of food is determined by three factors [2]: 1) the number of dishes, 2) the shape of each dish, and 3) the ratio of the area of food to the area of the dish. Either the user provides this information manually, or the system automatically recognizes this information, having analyzed an image that the user has taken on their mobile phone. The system then retrieves an appropriate composition from our database and shows it to the user (Figure 1).

If the user is not satisfied with the suggested composition, the user can design a new composition and register it to the database. This can be done by performing simple gestures on the mobile phone such as pinch-in and pinch-out. The registered composition will then be shared by all users of our system.

We performed a usability study for our system. The results indicated that our system can be quickly and easily used, even by novice users. We also performed a subjective evaluation of the quality of pictures taken using our system. Our evaluation suggests that pictures taken using our system are of a higher quality than pictures taken without using our system.

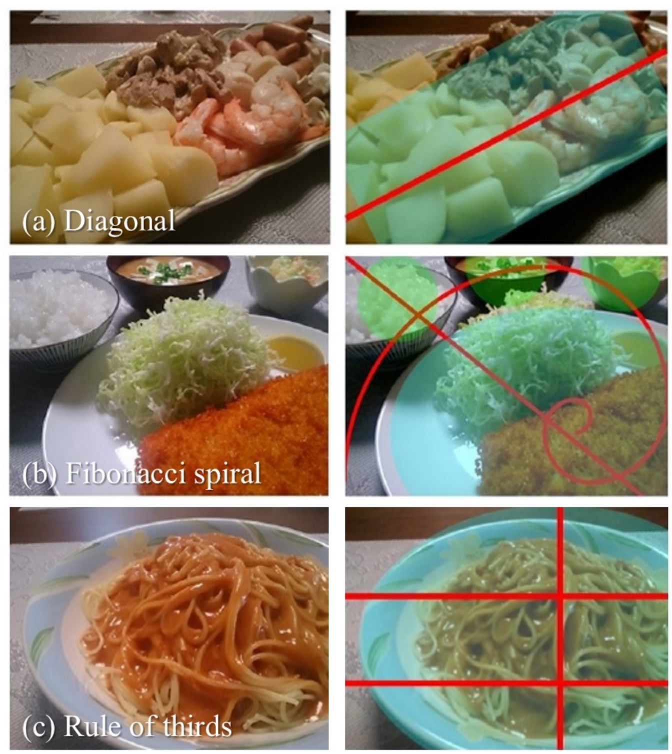

Fig. 1. Food photographs taken by users of our system. Each pair shows a picture taken by our system (left) and a suggested composition (right).

II. Previous work

A. Food recognition

Object recognition is a subject of active research in computer vision and pattern recognition. Before 2012, researchers manually designed feature vectors for the recognition of objects. For example, histograms of colors and image gradients remain popular and powerful features, e.g., SIFT [3], SURF [4], and ORB features [5]. These features are robust with respect to variations in scale, rotation, and illumination, but often fail to recognize similar objects. *Deep convolutional neural networks (DCNN)* were proposed in 2012 [6]. These automatically extract appropriate features from learning datasets. By using these, we gain a 10% improvement in the accuracy of recognition. We use a *network in network,* a particular type of DCNN, to implement our food recognizer [7]. This can recognize a dish in less than 0.5 seconds. The accuracy is 74% in the case of the first-ranked dish, and 93.5% in the case of the top-five-ranked dishes.

B. Evaluation of aesthetics

There are techniques to automatically evaluate the aesthetics of a picture [8][9]. These techniques evaluate aesthetics using DCNNs. These techniques have not been applied to the aesthetics of food pictures. In this paper we have not attempted an automatic aesthetic evaluation; instead, we performed a subjective evaluation of the quality of taken pictures.

C. Image editing

Given an input image, an image retargeting technique rearranges objects within the image and modifies its composition. Liu et al. retargeted the input image by evaluating its aesthetics based on 1) the rule of thirds, 2) the diagonal composition, 3) the balance of a picture, and 4) the size of each object [1]. This technique extracts parts from an input image but cannot rearrange objects. Our system proposes rearrangements of dishes to the user. The relighting technique allows the user to modify the lighting conditions of an image. Ren et al. used a neural network for relighting [10], which requires several hundreds of input images and is difficult for a novice user to use. Depth of focus is also important for aesthetics. For example, Liu et al. estimated the depth of a single image [11], which enabled them to later modify the depth of focus . Relighting and modification of depth of focus are useful for improving the aesthetics of a picture but are beyond the scope of this paper and our future work.

III. Taking a picture of foods

There are a number of guidelines for improving photographic composition to enhance the impact of a scene, e.g., ‘place your subject in the center’, ‘use the diagonals’ (Figure 1-a), ‘use the Fibonacci spirals’ (Figure 1-b), ‘balancing with the rule of thirds’ (Figure 1-c), etc. We assume that these guidelines are also useful for food photography. We also assume that the composition suitable for a scene is determined by the following three parameters: 1) the number of side dishes, 2) the shape of the dish (such as a circle, rectangle, oval, or square), and 3) the ratio of the area of the food to the area of the dish.

We first observed and investigated pictures of delicious looking food taken by skilled professional photographers. We learned how they take such pictures [2]. We found that the compositions of their pictures adhered to the following three conditions:

1. The existence of side dishes.
2. The shape of the main dish, which may be that of a circle, rectangle, oval, or square.
3. The area of the main dish, which can be either greater or less than twice the size of the area of the food on it. We call the former a ‘large’ dish and the latter a ‘small’ dish.

Given the information about the dishes entered by the user (Section III-A) or obtained by the automatic recognition (Section III-B), our system first determines the *basic composition* based on the conditions described above, and then shows the *final composition* to the user. First the situation is classified according to the first condition: if there are no side dishes, we use the rules defined in Table I; if there is at least one side dish, we use the rules defined in Table II. In both tables, a small rectangular or oval dish has no basic composition, because we have not found such a case through our observations.

TABLE I
Rules we use if there is no side dishes.

Shape of the main dish	Size	Basic composition
Circle	Large	(a) Rule of thirds
	Small	(b) Centering
Rectangle or oval	Large	(c) Diagonal
	Small	N/A
Square	Large	(d) Diagonal + centering
	Small	(e) Diagonal + centering

TABLE II
Rules we use if there is at least one side dish.

Shape of the main dish	Size	Basic composition
Circle	Large	(a) Diagonal + Fibonacci spiral
	Small	(b) Diagonal + rule of thirds
Rectangle or oval	Large	(c) Diagonal
	Small	N/A
Square	Large	(d) Diagonal + centering
	Small	(e) Diagonal

Figure 2 shows the final compositions in a case where there are no side dishes. Figure 3 shows the final compositions in a case where there is at least one side dish. The system displays the final composition on the screen of the user’s mobile phone. We created a composition database to aid the specification of the final composition. In this database the positions, rotations, and scales of the main dish and side dishes are registered for all possibilities admitted by the combination of items and the basic compositions. In the case of a large circular main dish without any side dishes, we calculate the position and scale of

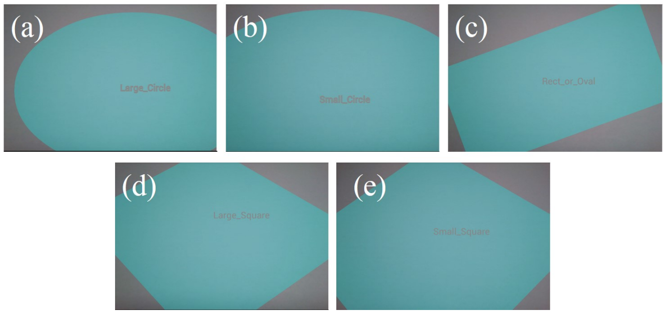

Fig. 2. Each figure shows the final composition corresponding to the basic composition in Table I. For example, when the main dish is large and circular, the system determines the basic composition according to the rule of thirds, and then makes a final composition like that shown in (a).

the main dish using the rule of thirds, as shown in Figure 2 (a), which is registered in the database. Given a large circular main dish with two side dishes, we calculate the positions and scales of the dishes based on the diagonal and the Fibonacci spiral: we place side dishes along the diagonal and the spiral, and place the main dish at the end point of the spiral. In the

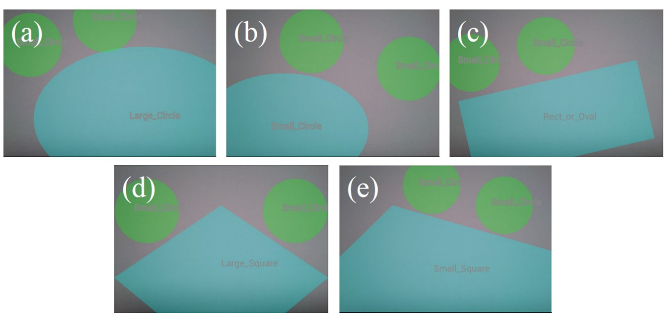

Fig. 3. Each figure shows the final composition corresponding to the basic composition in Table II. For example, when the shape of the main dish is large and circular, the system determines the basic composition to be a combination of the diagonal and the Fibonacci spiral, and then suggests a final composition similar to that shown in (a).

case of a rectangular side dish, we also take rotations into account: the side dish rotates along the spiral (Figure 4). We

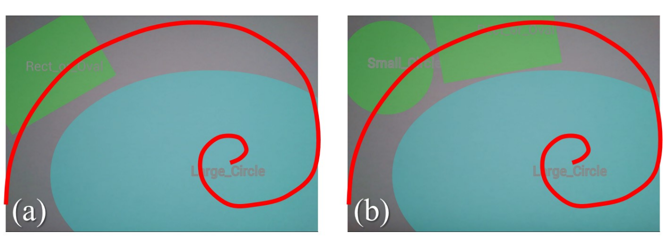

Fig. 4. The red line represents the Fibonacci spiral. The rectangular side dish rotates along with the spiral.

also ascertain which camera angle will be best for taking the picture. (Section III-C).

A. Manual input

Before taking food pictures, the user can provide information about the dishes to be photographed manually by pushing buttons (Figure 5). The user specifies the shape of the main dish by pushing the 'circle', 'rectangle or oval', or 'square' button. The system also has buttons to specify the ratio of the area of the dish to that of the food to indicate, for instance, whether they have a 'large' or a 'small' dish. To provide further information about each side dish, additional buttons can be used; for example there are buttons that specify glasses and cups as side dishes. (Figure 5).

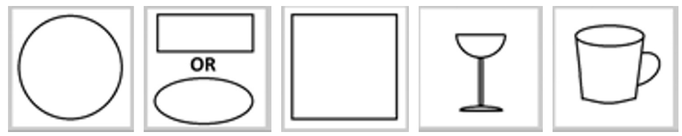

Fig. 5. The user pushes the button to provide information about the dishes. From left to right, the buttons to specify a circular shape, a rectangular or oval shape, a square shape, a glass as a side dish, and a cup as a side dish.

B. Automatic food recognition

Our system can recognize dishes automatically [7]. When using this function, the user does not have to push buttons. As this function recognizes the names of the items of food on each dish, it enables our system to make more flexible suggestions to the user.

When using this function, the user takes a shot of each dish, one by one. The system then shows the names of the top five-ranked dish names suggested by the automatic recognition. The accuracy is 74% for the most highly ranked dish and 93.5% for the top-five ranked dishes. The user then selects the correct name of the dish manually. The number of types of food that the system can recognize is 100. We also constructed a database containing the typical shape and size of each food item and corresponding dish. For example, Chinese noodles tend to be in a large circular dish and soup tends to be in a small circular bowl. By querying the database with the name of the food, we can determine the shape and size ratio of the dish. The drawback of this strategy is that we cannot take unusual cases into account, for example, soup in a square dish. In such a case, the user has to go to the manual input mode. However, since such cases rarely happen, our simple strategy based on this kind of mapping is still powerful in many of the more usual cases.

C. Taking a picture of foods

After providing information about the dishes, the system shows the final composition on the screen of the user's mobile phone (Figure 6). The user arranges the dishes on the table according to the composition so that each dish fits to the corresponding object in the screen. As shown in Figure 1, the final composition overlays the image from the camera. After arrangement, the user adjusts the tilt of the camera and takes the picture by pushing the shutter button. During this process, the 'Tilt' icon at the bottom right corner is useful: the color of

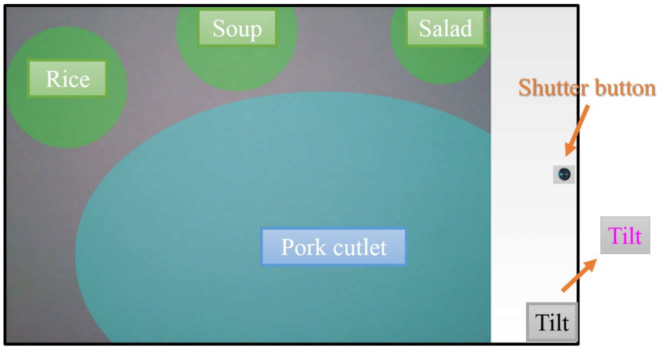

Fig. 6. The user interface to take the picture of foods.

the 'Tilt' icon changes to purple when the tilt of the camera is close to the best tilt angle, as determined by the system.

IV. Interactive Design of a New Composition

In the case that the user wishes to photograph food items or dishes that are not known to the system, our system allows the user to interactively design a new composition. For example, assume the user wants to design a composition of the dishes shown in Figure 7 (a). The user first takes a picture of the dishes. The user then pushes the 'rectangle or oval' button of Figure 5, and the system enters the mode where the user can design the rectangle by pinch-in and pinch-out gestures (Figure 7-b). After finishing the design of the rectangle for the main dish, the user then designs the composition for the other dishes in a similar manner (Figure 7-c).

Fig. 7. The user interface for designing a new composition.

V. Results and Discussion

We implemented our system on a mobile phone with the Android operating system (OS). Since the computational costs of our algorithms are low, our system works interactively and the user can obtain immediate feedback from the system. We performed subjective evaluations of the usability of our system and the quality of the pictures taken using our system.

A. Usability study

The subjects were seven men in their twenties. They were already familiar with operations on mobile phones with Android OS. No subjects had used our system previously. First of all, we gave each subject a short tutorial on our system. We prepared two types of dishes for our experiments: one consisted of a Japanese pork cutlet as the main dish, and rice, miso soup, and salad as side dishes (Figure 8-left). The other consisted of either Chinese fried rice or dumplings; the user could select the main dish from the two (Figure 8-right). After the tutorial, we asked each subject to use our system to take pictures of these dishes. Figure 9 shows the pictures taken by the subjects using our system; even a novice user of our system could take pictures with a composition similar to the final composition (Figure 3) suggested by the system. This implies that our system allows a novice user to quickly and easily take a well-composed picture.

Fig. 8. Dishes that we prepared for the subjective evaluation.

Fig. 9. Pictures taken by the subjects using our system.

B. Quality assessment

In this experiment, we considered three groups of subjects: groups A, B, and C. There were three subjects in group A, three subjects in group B, and eighteen subjects in the group C. We asked the subjects in group A to take pictures using our system while subjects in group B took pictures without using our system. We told all the subjects to make the food in the pictures look as delicious as possible whilst including all the dishes. Groups A and B used the same device to take pictures.

We then asked subjects in group C to evaluate the pictures taken by groups A and B. Figure 10 shows the pictures. We performed a subjective evaluation as follows: we showed each row of Figure 10 to the subject, and asked them to vote for the picture with the most delicious looking food. Figure 10 (a), 10 (c), and 10 (f) were taken by the group A, i.e., using our system. Figure 10 (b), 10 (d), and 10 (e) were taken by group *B*, i.e., without using our system.

The results from the vote are shown in Figure 11. For each row of Figure 10, all the subjects voted for the picture taken by group *A*. We also asked each subject to vote for the best and worst pictures. The result is shown in Figure 12. Figure 10 (b) received 16 ‘worst’ votes. On the other hand, the ‘best’ votes were distributed relatively uniformly among the pictures taken by our system, which proves that our system enables users to take reasonably attractive pictures.

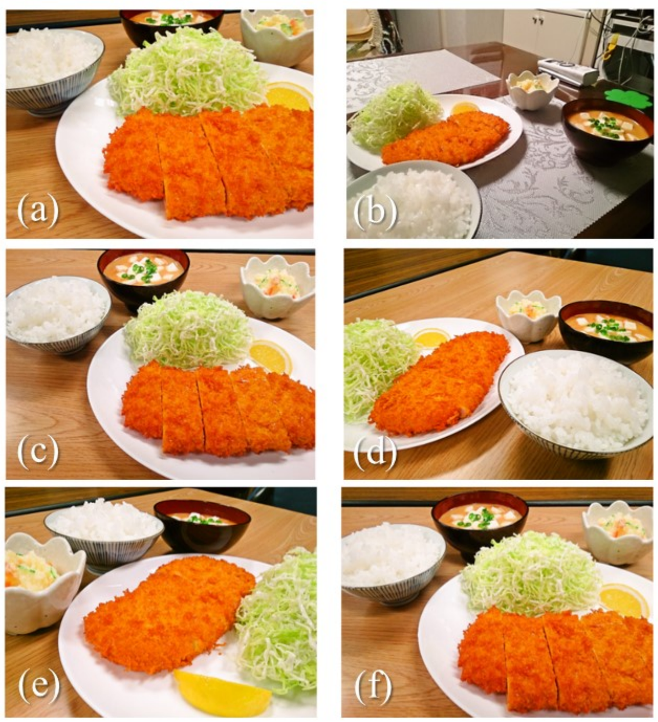

Fig. 10. Pictures taken by subjects using our system.

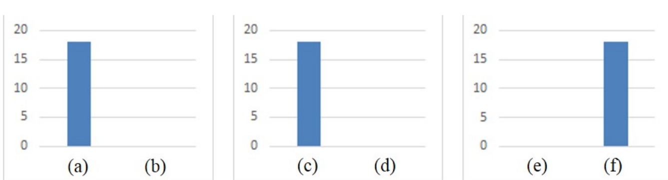

Fig. 11. The results of the vote for a picture from each row of Figure 10.

Finally, we show some other results: pictures of various foods taken with the help of our system (Figure 13).

VI. Conclusion

We proposed an interactive system to help people to take pictures of delicious looking food. Given information about the dishes, such as the number, the shape, and the ratio of dish to food, the system retrieves the best composition from the database and shows the suggestion to the user. Our system also supports users in designing new compositions.

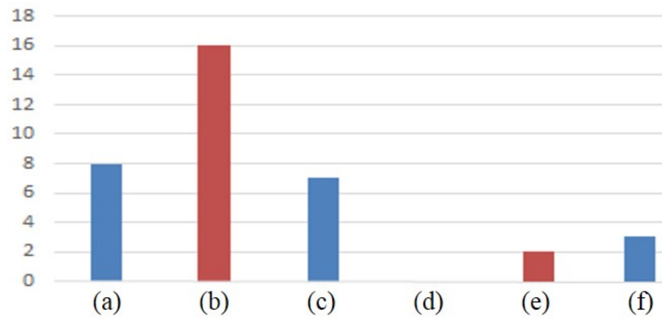

Fig. 12. The results of the vote for the best and worst pictures in Figure 10.

Fig. 13. Pictures of various foods taken by our system.

An interesting suggestion from one of the subjects regarding the usability study was that they wanted to see multiple composition candidates and then select their favorite. The extension of our system to show multiple candidate compositions will be straightforward, so we will undertake this task in the future.

We considered the composition of food items and dishes with regard to creating attractive pictures in this paper. Other factors such as lighting and color correction will be considered in future work. Also, we would like to implement a user interface that enables depth of focus to be edited during post-production.

References

[1] L. Liu, R. Chen, L. Wolf, and D. Cohen-Or, “Optimizing photo composition,” *Computer Graphics Forum*, vol. 29, no. 2, pp. 469–478, 2010.

[2] N. S. Young, “Food photography: From snapshots to great shots,” *Peachpit Press*, 2011.

[3] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[4] H. Bay, T. Tuytelaars, and L. V. Gool, “Surf: Speeded up robust features,” in *Proc. of ECCV*, 2006, pp. 404–417.

[5] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “Orb: an efficient alternative to sift or surf,” in *Proc. of ICCV*, 2011.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems 25*, 2012, pp. 1097–1105.

[7] K. Yanai and Y. Kawano, “Food image recognition using deep convolutional network with pre-training and fine-tuning,” in *Proc. of IEEE International Conference on Multimedia Expo Workshops (ICMEW) 2015*, 2015, pp. 1–6.

[8] X. Lu, Z. Lin, H. Jin, J. Yang, and J. Z. Wang, “Rapid: Rating pictorial aesthetics using deep learning,” in *Proc. of MM ’14*, 2014, pp. 457–466.

[9] X. Lu, Z. Lin, X. Shen, R. Mech, and J. Z. Wang, "Deep multi-patch aggregation network for image style, aesthetics, and quality estimation," in *Proc. of ICCV '15*, 2015.

[10] P. Ren, Y. Dong, S. Lin, X. Tong, and B. Guo, "Image based relighting using neural networks," *ACM Trans. Graph.*, vol. 34, no. 4, pp. 111:1–111:12, 2015.

[11] F. Liu, C. Shen, and G. Lin, "Deep convolutional neural fields for depth estimation from a single image," *CoRR*, vol. abs/1411.6387, 2014.