GrillCam: A Real-time Eating Action Recognition System

Koichi Okamoto and Keiji Yanai

The University of Electro-Communications, Tokyo
1-5-1 Chofu, Tokyo 182-8585, JAPAN
{okamoto-k@mm.inf.uec.ac.jp,yanai@cs.uec.ac.jp}

Abstract. In this demo, we demonstrate a mobile real-time eating action recognition system, GrillCam. It continuously recognizes user's eating action and estimates categories of eaten food items during mealtime. With this system, we can get to know total amount of eaten food items, and can calculate total calorie intake of eaten foods even for the meals where the amount of foods to be eaten is not decided before starting eating. The system implemented on a smartphone continuously monitors eating actions during mealtime. It detects the moment when a user eats foods, extract food regions near the user's mouth and classify them. As a prototype system, we implemented a mobile system the target of which are Japanese-style meals, "Yakiniku" and "Oden". It can recognize five different kinds of ingredients for each of "Yakiniku" and "Oden" in the real-time way with classification rates, 87.7% and 80.8%, respectively. It was evaluated as being superior to the baseline system which employed no eating action recognition by user study.

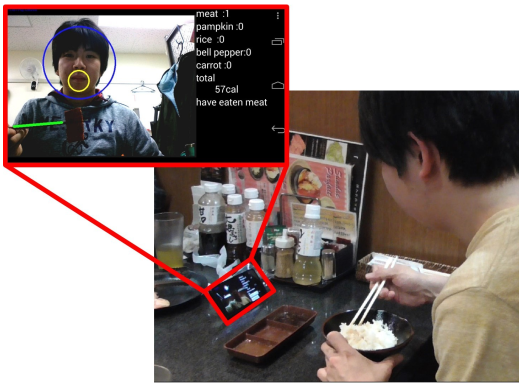
Keywords: Mobile Food Recognition, Eating Action Recognition, Food Recording System

1 Introduction

In recent years, due to a rise in healthy thinking on eating, many people take care of eating and foods, and some people record daily diet regularly. To assist them, many mobile applications for recording everyday meals have been released so far. Some of them employ food image recognition, which enable users to record daily foods only by taking photos.

We has proposed a mobile food recording system which has a 100-kind food recognition engine so far [1]. Since the recognition engine employed the state-of-the-art image recognition method, Fisher Vector and liner SVM classifiers, the classification rate was relatively high. The top-5 classification rate for 100 classes was 79.2%. However, it required taking a meal photo before eating, and all the foods taken as a photo had to be served before eating.

Then, we proposed a mobile food recognition application, GrillCam, which was applicable for the case that the amount of food eaten by one person was not decided before eating such as sharing large dishes or barbecue-style meal (grilling meats and vegetables while eating) before [2]. After that, we improved the proposed system in terms of the recognition accuracy and target meals.



meat :1
pampkin :0
rice :0
bell pepper:0
carrot :0
total
57cal
have eaten meat

Fig. 1. A typical usage of the proposed system. A smartphone on which the proposed system is running is put toward a user's face. It continuously detects eating action and estimates food categories of eaten foods.

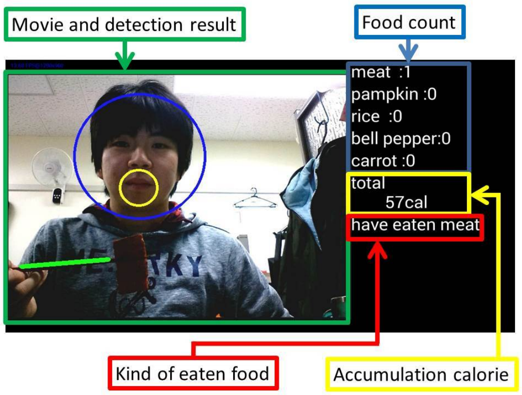
2 System Overview

Figure 1 shows a typical scene when we use the proposed system implemented on a Android smartphone. In this scene, the user is having a "Yakiniku" meal which is a Japanese-style barbecue of grilling thin-sliced beef and vegetables on the hot plate. When we use it, we stand it with slight tilt in front of a user who is eating so that the built-it inner camera of the smartphone faces to the user's face. Although it is common for image recognition application for a smartphone to use a backside camera, we use an inner camera to record user's action by the camera and show information of eaten food items to the user at the same time.

The screen-shot image of the UI of the proposed application is shown in Figure 2. A user can always check what the system is recognizing at the time regarding face, mouth and chopsticks. A blue circle, a yellow circle and a green line shown in the detected result area represent detected face region, mouth region and detected chopsticks region, respectively. On the right side of the screen, total calorie intake and the number of eaten items are displayed. A user can check how many calories he/she has taken while eating.

The proposed system performs eating action recognition according the following processing flow:

1. Detect a user's face and mouth.
2. Detect and track chopsticks.
3. Segment out a region candidate region (bounding box) around the tip of the chopsticks at the moment when the tip of the chopsticks is approaching the mouth.
4. Recognize a food item category for the segmented region.
5. Calculate and accumulate food calorie intake, and display it on the screen.
6. Repeat this procedure until the meal is finished.



meat :1
pampkin :0
rice :0
bell pepper:0
carrot :0
total
57cal
have eaten meat

Fig. 2. The screen of the proposed mobile application, GrillCam.

In the current implementation, to detect face and mouth, we used a standard face and mouth detector in the OpenCV Library. To detect chopsticks, we used background subtraction and the Hough transform.

To segment out a food candidate region, in the previous implementation, we assumed a food item is located on the tip of the detected chopsticks, which was relatively a straight way. If lines of chopsticks were detected as being longer than their actual length, a food region was also estimated incorrectly.

Therefore, we improved a step to estimate a food candidate bounding box by taking into account the center of the mass of binarized images as follows (Figure 3):

1. Segment out a large rectangular region around the tip of the chopsticks.
2. Binarize the region.
3. Calculate the center of mass in the region.
4. Extract a bounding box around the estimated center.

To recognize a food category of the food item being picked by chopsticks. In the previous implementation [2], we used a Bag-of-Features representation of the ORB [3] features and HSV color histogram. However, the accuracy was not enough.

In the current implementation, we adopt ORB and HSV color histogram, and each of the local descriptors are represented by Fisher Vector, and one-vs-rest linear SVM as a classifier for rapid recognition.

To calculate the total calorie intake, we use the pre-defined standard calorie values on each food item. To estimate food calorie intake precisely, we will estimate the volume of the detected food items in the future work.

3 Experiments

We have implemented the proposed system as an Android application. In the experiments, we used Google Nexus 5 (2.3GHz QuadCore, Android 4.4) as a

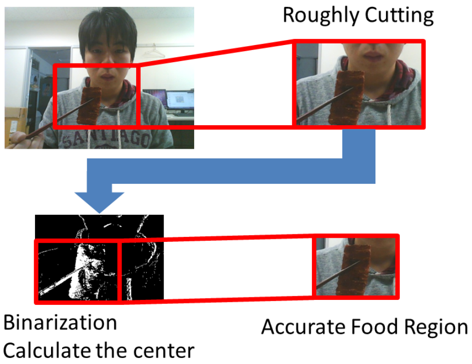


Fig. 3. Food region extraction

target smartphone. We chose "Yakiniku" and "Oden" meal as target domains for the demo system implementation. "Yakiniku" is a Japanese-style barbecue meal to eat baked sliced meats and vegetables, while "Oden" is a Japanese winter dish consisting of several ingredients in a light, soy-flavored dashi broth. In the experiments, we evaluate classification accuracy and the system usability.

For the implementation, we selected the following five typical food items in Yakiniku: "meat", "rice", "pumpkin", "bell pepper" and "carrot". Further, we selected the following five typical food items in Oden: "radish", "egg", "Hanpen(Boiled Fish Cake)", "Konjac" and "chiikuwa (Grilled Fish cake)". Figure 4 shows kind of Yakiniku, and Figure 5 shows Oden. We stored typical calorie values on the above food items on the system to calculate total calories of eaten food items.

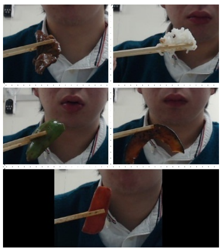


Fig. 4. The target food items for "Yakiniku".

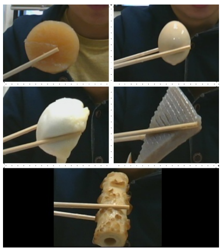


Fig. 5. The target food items for "Oden".

For training, we prepared more than one hundred images for each of the food items. As shown in Table 1, the result of the classification rate at Yakiniku and Oden. In the previous implementation, the system was able to recognize only

Yakiniku. The classification rate by the previous system with a conventional bag-of-features was 74.8%, while the rate by the proposed system employing Fisher Vector is 87.7%. The rate was improved by 12.9 points.

Table 1. The classification rate at Yakiniku and Oden.

Yakiniku [2]	Yakiniku	Oden
74.8%	87.7%	80.8%

Table 2. The mean and standard deviation of the five-step evaluation score.

baseline	GrillCam
2.36 ± 1.12	4.36 ± 1.41

We made a simple user study as well. For comparison, we prepared a baseline system which had no image recognition function and instead required users to touch food item buttons on the screen every time they eat a food item. We asked eleven users to eat five kinds of food items in front of both the proposed system and the baseline system, and to evaluate both systems in five-step score regarding their usability. As shown in Table 2, the score of the proposed system outperformed the score of the baseline greatly. This indicates the effectiveness of the proposed system.

4 Conclusions

In this paper, we proposed a new-style food recording system, GrillCam, which employs real-time eating action recognition and food categorization for meal scene. The system continuously monitors user's eating action and estimates the calorie intake in a real-time way.

This enables us to estimate total calories of the food the amount of which is undecided before eating. This special feature is totally different from existing image-recognition-based food calorie estimation systems which requires taking meal photos before starting to eat.

Although in the current implementation the kinds of mean scenes and the number of food items are still limited, we believe the proposed system, GrillCam, will be more practical by extending it so as to recognize various meal scenes such as sharing large platters, conveyor-belt-style sushi and hot-pot-style meals like sukiyaki. To do that, we will extend it so as to detect fork, knife and hands as well as chopsticks.

References

1. Kawano, Y., Yanai, K.: Real-time mobile food recognition system. In: Proc. of CVPR International Workshop on Mobile Vision (IWMV) (2013).
2. Okamoto, K., Yanai, K.: Real-time eating action recognition system on a smartphone. In: Proc. of ICME Workshop on Mobile Multimedia Computing (2014).
3. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: an efficient alternative to SIFT or SURF. In: Proc. of IEEE International Conference on Computer Vision (2011).