IMAGE PROCESSING BASED APPROACH TO FOOD BALANCE ANALYSIS FOR PERSONAL FOOD LOGGING

Keigo Kitamura†, Chaminda de Silva‡, Toshihiko Yamasaki†, Kiyoharu Aizawa‡

Department of Information and Communication Engineering†
Interfaculty Initiative in Information Studies‡
The University of Tokyo
Email: (kitamura, chamds, yamasaki, aizawa)@hal.t.u-tokyo.ac.jp

ABSTRACT

Food images have been receiving increased attention in recent dietary control methods. We present the current status of our web-based system that can be used as a dietary management support system by ordinary Internet users. The system analyzes image archives of the user to identify images of meals. Further image analysis determines the nutritional composition of these meals and stores the data to form a Foodlog. The user can view the data in different formats, and also edit the data to correct any mistakes that occurred during image analysis. This paper presents detailed analysis of the performance of the current system and proposes an improvement of analysis by pre-classification and personalization. As a result, the accuracy of food balance estimation is significantly improved .

Keywords— Food, Image Processing, life log

1. INTRODUCTION

With the increase of personal image and video collections on the Internet, many different aspects of life can now be seen on the Web. Among them, we can find eating activities. Many people take photos of every meal for several purposes: they can share the food images, or they can improve dining habits. Regarding the latter, a study shows that taking a photograph of meal before eating can encourage weight loss [1]. If there is a way to log and extract detailed dietary information from food images, it will be beneficial to many who want to improve their dining habits.

Recently, several researches on automated food image analysis have been reported. Wu et al. attempted to analyze fast food images using *eating* videos [2]. Joutou and Yanai proposed a system that classifies food images into predetermined food menus such as hamburger and pizza [3]. We have previously presented a food image analysis system [4][5][6] that has the functions described below.

Our proposed system has two main functions: they are detection of food images and estimation of food balance. Regarding food image detection, In the estimation of food balance, SIFT features were used and the accuracy of the food image detection was 93.3% [4]. The Food Pyramid [7] is used as the basis of food balance analysis.

We propose to implement the two image analysis functions into a web based system [5]. In the case of web application with the ordinary users, a much wider variety of food images can be expected. Thus, we should analyze a wider variety of food images accurately. In the web based system, the users can correct the result of the analysis if necessary.

At the end of the first nine months, 288 users had submitted more than 5500 food images to this web based system. These images are quite diverse in both content and quality. It was observed that our pre-defined food balance analysis model [5] did not work as accurately as expected, with these images. A thorough investigation of the food images showed two main tendencies. One is that there are many images including only small piece of food. The other is that some users regularly upload nearly identical images. In order to utilize these tendencies for more accurate analysis, we can have two strategies. One is a pre-classification before estimating food balance, and the other is to personalize food image estimator. We evaluate the performance improvement obtained by employing these strategies.

The rest of this paper is organized as follows. Section 2 summarizes related research. Section 3 outlines image analysis algorithms and our web based system. Section 4 describes the nature of data image collection. Section 5 presents the proposed algorithms for pre-classification and personalized image analysis, and the results of their evaluations. Section 6 concludes the paper.

This research was supported in part by CREST of JST of Japan.

2. RELATED WORK

Image processing of food images is a emerging topic. There are a few papers that propose food image recognition. Because generic object recognition is a rather difficult problem so far, they try to make it possible within some boundary conditions.

Zhu *et al.* limited the data set to 50 images including 32 food items, such as apple and baked beans [8]. The items were on the same white dish on a checkerboard, the food item extraction was made easier than general image segmentation. Wu targeted only fast foods such as hamburgers and attempted to predict which one appeared in a video [2]. Their database contains 101 foods from nine food restaurants, such as *McDonald's, KFC, and Subway*. They used videos for testing, and the task is formulated as a problem of image matching using training images. Their system, based on SIFT feature matching, performs image recognition with an accuracy of 55%.

Additional manual processing was also taken into account to achieve more accurate food image recognition. Martin *et al.* have developed a semi-automated application to estimate the food intake for participants [9]. This is one of the projects in Remote Food Photography Method (RFPM) [10], which has proven to be reliable and valid for measuring free-living people's accurate food intake. They asked the subjects to place a reference card next to their food before taking pictures. In addition, the user manually selects the food region for each food in the training process. Because they aimed to estimating accurate food intake, additional processing is necessary. Joutou and Yanai proposed a system that estimates 50 food menus, such as hamburger and pizza [3]. Each menu consists of 100 example images in a database. Training images are selected so that they can classify effectively. They mainly include only one target food. With multiple kernel learning, they have achieved the 61% classification rate for the 50 menu food.

The researches above have set boundary conditions for their work. On the other hand, we aim to build a system that ordinary web users can use in their daily life. Once public users take and upload food images, a wide variety of food images is expected. The works above are not yet ready for such data. For instance, for the system in [3], the accuracy of classification fell down to 37% when the system was opened. It is necessary to collect a large number of food images of real users and conduct experiments with them. Therefore, we have been developing FoodLog system as a web application and keep collecting diverse food images.

Table 1. Research on food images.

Ref.	Target	Manual processing	Image data collection	# of images
[9]	Food Items	Put on the fixed plate	Captured by them	50
[2]	Fast Food	none	Captured by them	N.A.
[10]	Patients food	Place reference card	Captured by them	N.A.
[3]	50 menus	Clipping for training	Web	100 / menu

3. FOODLOG SYSTEM

In order to build a practical food logging system, we implemented two main image recognition functions: detecting food images in a collection of daily images and analyzing the images to estimate the attributes of food.

3.1. Detecting Food Images

Our system detects food images from user daily images uploaded. The system does not require users to separate food images manually or upload them to a different location from their normal photo-sharing web site. Therefore, this step is essential.

Detecting food images is based on the global features and the local features which are combined and learned by putting into Support Vector Machine (SVM). In our previous works [4], it has been demonstrated that two types of global features are suitable. One is the information regarding circles detected by Hough transforms, and the other is an average value in color space. We also utilize a *Bag-of-Features* (BoF) model which is the representation of image content using local features. Recently, representing models using local features has been proven to be successful [11] [12]. The BoF [13] represents images as collections of independent local patches, and vector-quantizes them as histogram vectors. As the local features, most researches use SIFT features [14] which detect and describe the local gradients. We apply the BoF model and the preceding global features, and have detected food images by 93.3 % accuracy. Five hundred images were used in this experiment. They consist of both images collected from the web and images taken by the authors themselves.

Table 2. Food balance guide (SV=servings).

	Grains	Vegetable	Meat& Beans	Milk	Fruits
Intake/ day	5−7SV	5-6SV	3−5SV	2SV	2SV
1SV	rice bread	vegetable mushroom	egg nut	yogurt cheese	peach orange
2SV	noodle pasta		fish	milk	apple pear
3SV			meat		

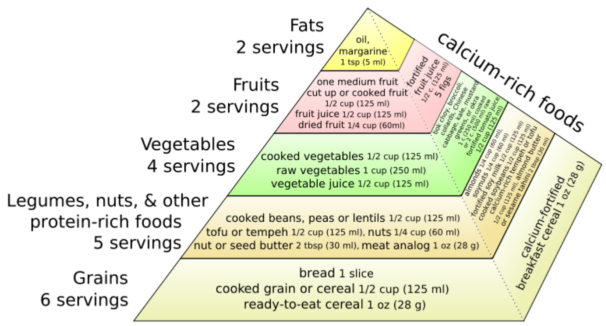

Fig 1. Food balance guide [15] and Food pyramid [16]

3.2. Estimating Food Balance

After the food-only images are detected, the system can make the multimedia diary. The main goal of food image analysis is to estimate and visualize the nutrition information corresponding to the food shown in the images. We employ "Food Balance Guide [15]" as the model for dietary composition. It is determined by Ministry of AFF Japan, and it is similar to "My pyramid [16]" made by the United States Department of Agriculture and the similar models are used around the world. The food balance guide and the food pyramid are shown in Fig 1. It categorizes food into five groups; grains, vegetable, meat & beans, milk, and fruit. In these groups, the quantity of each food is defined by the original unit, *serving* (SV). Examples are shown in Table 1.

The task is to estimate the number of servings from each of the above categories, in the food shown in a given image. We formulate it as a two-step classification. In the first step, a food image is divided into 300 blocks, and a feature vector composed of color histogram and DCT coefficients are extracted from each block. The blocks are categorized into the five groups defined in the food balance guide (grains, vegetables, meat & beans, fruits, and milk) or "others (non-food)". Then, histograms of the five ingredients are formed, and the food balance is determined in the second step. SVM is used in both stages. In our preliminary experiment, the accuracy of classification was 73%. The system performed

Table 3.FoodLog users' activities (first 9 months)

# of all users	# of all images	# of all food images	# of modified food images	# of new food images /day
288	9281	5695	1009	10-20

well on fruits and milk because the number of servings is very biased and either 0 or 1 in the given data. On the other hand, the number of servings of grains, meats & beans and vegetables were more difficult to estimate. Their accuracy turned out to be less than 50%.

3.3. Interfaces

Figure 5(a) shows the web application that we developed by incorporating the above mentioned algorithms. It has been developed as a web mash-up with *Flickr* [17]. With image recognition, the users can get their multimedia meal diaries by taking and uploading food images. In addition to image recognition, we have implemented two more functions: modifying the recognition result and visualizing the log.
(note: At the begging of May 2010, a new system will be in operation, which does not use Flickr anymore.)

3.3.1. Correction

The automatic image analysis results are not always correct. Then, the system provides an easy-to-use interface to verify and correct such errors. As shown in Fig 5 (d) the images which are classified as food images are on the upper part and the non-food images are on the lower part. The users can correct errors in detection easily by drag and drops, as the blue line shows. Fig 5 (e) shows the interface for modifying the result of estimating food balance. Users can modify the number of servings for each category, using sidebars.

3.3.2. Visualization

Once users take and upload their food images and modify them regularly, they can check their dietary habits in various ways. The system gives them some visualization tools. The users can check daily, weekly, and monthly data. Each page shows food images and food balance taken during the corresponding period. The upper part of Fig 5 (b) demonstrates that the user ate too much meat and beans, and too little vegetables. Fig 5 (c) shows a weekly summary for checking what time the user took meals on each day.

4. THE ACCURACY OF IMAGE RECOGNITION

The web application described in Section 3.3, called FoodLog, has been open to the public since March 2009. In

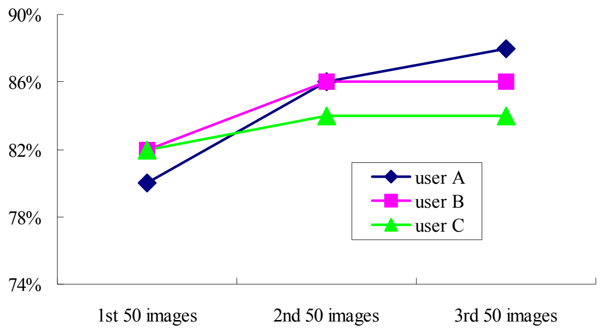

Fig. 2. The change of the accuracy of food image detector on each user's images

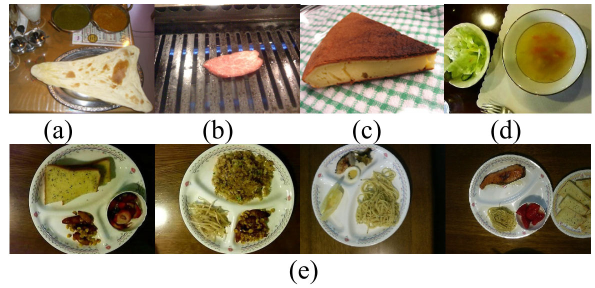

Fig.3. Examples of uploaded food images: (a) - (d) are the unexpected kind of food images, (b) are the almost same food images uploaded by the same users.

Table 4.Image recognition result (The number of images applied on making the model)

Model	The accuracy of food images detector	The accuracy of food balance estimator
Preliminary	93% (500images)	50% (100 images)
After release	85% (500images)	37% (100images)
After re-training	92% (8000images)	38% (800images)

the following sub-sections, we discuss the usage and performance of this system in this nine months.

4.1. Users' Behavior and Activities

As shown in Table 3, in the nine months, there are 288 users and more than 5,500 food images have been uploaded. Fig. 2 shows the changes of the accuracy of food image detector. The points of 1st 50 images shows the accuracy on 50 images that each user uploads to FoodLog immediately after they start to use the system. The points of 2nd 50 images shows the accuracy on 50 images which user uploads following the 1st 50 images. We can find that the accuracy of the 2nd and 3rd 50 images is higher than that of 1st 50 images. By this figure, it is considered that users get more accustomed, and they know better how to take their food images detected more correctly. In fact, because the food images that are taken just above are most probable to be detected as food images, some users started to take food images just above meanwhile after they began to use the system.

However, the average accuracy of food image detection is 85.3%. It is lower than during the preliminary experiments. The middle column in Table 4 shows the results for the pre-defined food image analysis model. Here, we define the accuracy of food balance estimator as the accuracies of three users' three main categories; grains, meat & beans, and vegetables. The three users upload their food images and correct the results every day. We can find that both accuracies are lower than those for the preliminary experiments that are shown in the upper column in Table 4. This is because there appeared a much wider variety of food image compared to that of the training database. In such cases, the analysis tends to produce wrong results. Therefore, it is necessary to modify this non-general model.

4.2. Expanded database

In order to produce a more general detection/classification model, almost all kind of uploaded images should be included in the training. Regarding the food image detection, we have achieved 93.6% accuracy by 10-fold cross validation and the new 1000 test images are recognized with 91.8% accuracy. Regarding the food balance estimator for 800 food images, we have achieved 40.6% accuracy by 10-fold cross validation and the new 100 test images are estimated with 38.2% accuracy.

5. IMPROVEMENT OF IMAGE RECOGNITION

To improve the accuracy of balance estimation, we introduce a pre-classification of food images and a personalized estimator.

5.1. Food Image Classification

There are so many images including only small piece of food as shown in Figs. 3 (b), (c) and (d). In fact, 29% of all food images include less than 1 SV in all categories: such images are defined as "small food" images. For example, Fig 3(b) includes only 1 SV meat & beans, (c) includes only 1 SV grains and (d) includes only 1 SV vegetables. Therefore, before estimating each category's food balance, we classify food images into two: small food images and others. The image classification algorithm is almost the same as that of food image detector: the global features and local features are combined and learned by putting them into SVM. The information of circles detected by Hough transforms is useful especially because the number of circles represents the number of the plates and bowls. It is expected that the fewer circles the image includes, the less food

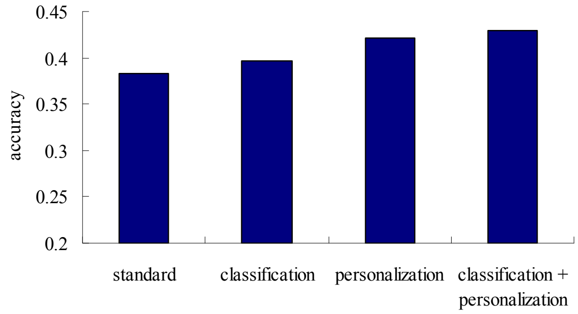

Fig. 4. The improvement of the accuracy of food balance estimator

balance the meal includes. However, local features are somewhat different from that of food image detector. Our previous work shows that the local features for food images need not only the information of gradient but also the information of color [6]. Finally, we apply the BoF model with modified SIFT features and the preceding global features to detect small food images with 83.3 % accuracy.

After that, each category's balance is estimated by the same way. In other words, we choose 1 or 0 SV in estimating small food images and choose from 1.5 to 3 SV in estimating the other images. When we apply image classification and estimate each category, the accuracy of the food balance estimator is improved upto 40%.

5.2. Personalization

The user's dietary habits are not the same and they are reflected to their food images (Fig 3e). Therefore, we want to choose training datasets that fit the user's dietary habits and make a personalized estimator. Because our system has an error correction interface, it is expected that the system can obtain the user's correction data. The more training datasets include the corresponding user's images, the more accurate the food balance estimation will become.

When the same user's food images are utilized as the training images and the test images, the results of the food balance estimation improves from 37% to 42% on average. As a result, we find that the personalized image analysis using the users' own data is meaningful.

5.3. Comparison

Fig. 4 and Table 5 summarize the accuracy of food balance estimator. Figure 4 shows results for four estimators; standard, classification, personalizarion, and classificarion and personalization. "standard" is the one for the experiments explained in 3.2. "Classification" is the estimator that makes use of preclassification before food balance estimation described in 5.2, and "personaliztion" is the one described in 5.3. We use the both food image pre-

Table 5. The accuracy for each user by "pre-classification and personalized estimator" (changes from "standard")

	user A	user B	user C	Average
Grains	46% (+12%)	48% (+6%)	50% (+6%)	48% (+8%)
Meats & Beans	32% (0)	38% (-2%)	24% (+2%)	31% (0)
Vegetabl es	52% (0)	50% (+10%)	50% (+20%)	50% (+10%)
Average	43% (+4%)	45% (+5%)	41% (+9%)	43% (+6%)

classification and the personalization estimator together, "classification and personalization", in which the system classifies food images into small food images and others and it estimates the food balance by the estimator personalized for each user. We have achieved 43% accuracy with this estimator. Table 5 shows the accuracy for each user. It is found that the accuracy of each user are almost same by using personalized estimator. However, the accuracy improvement of each category is totally different. In the case of meat & beans, it seems difficult to improve. Because the accuracy of meat & beans is lower than that of other two categories, it is considered that meat & beans has little relevance to image features.

6. CONCLUSION

In this paper, we have presented our foodlog system and the evaluation of our web-based system. We also showed that a image pre-classification and a personalized estimator can contribute to improving the food balance estimation. By image pre-classification, the accuracy of each category improves by 3% and the online-training using user's own food images yielded an improvement of 4%. The overall accuracy with both techniques is 44%.

The goal of this work is to extract information in order to provide people beneficial information about their dietary habits. In order to achieve this goal, we need to correct a large number of food images with evaluation. In addition to this, we have to utilize users' manual inputs efficiently. We are going to research not only on analyzing the food images, but also on analyzing communities based on meals..

REFERENCES

[1] L. Zepeda, and D. Deal, "Think before you eat: photographic food diaries as intervention tools to change dietary decision

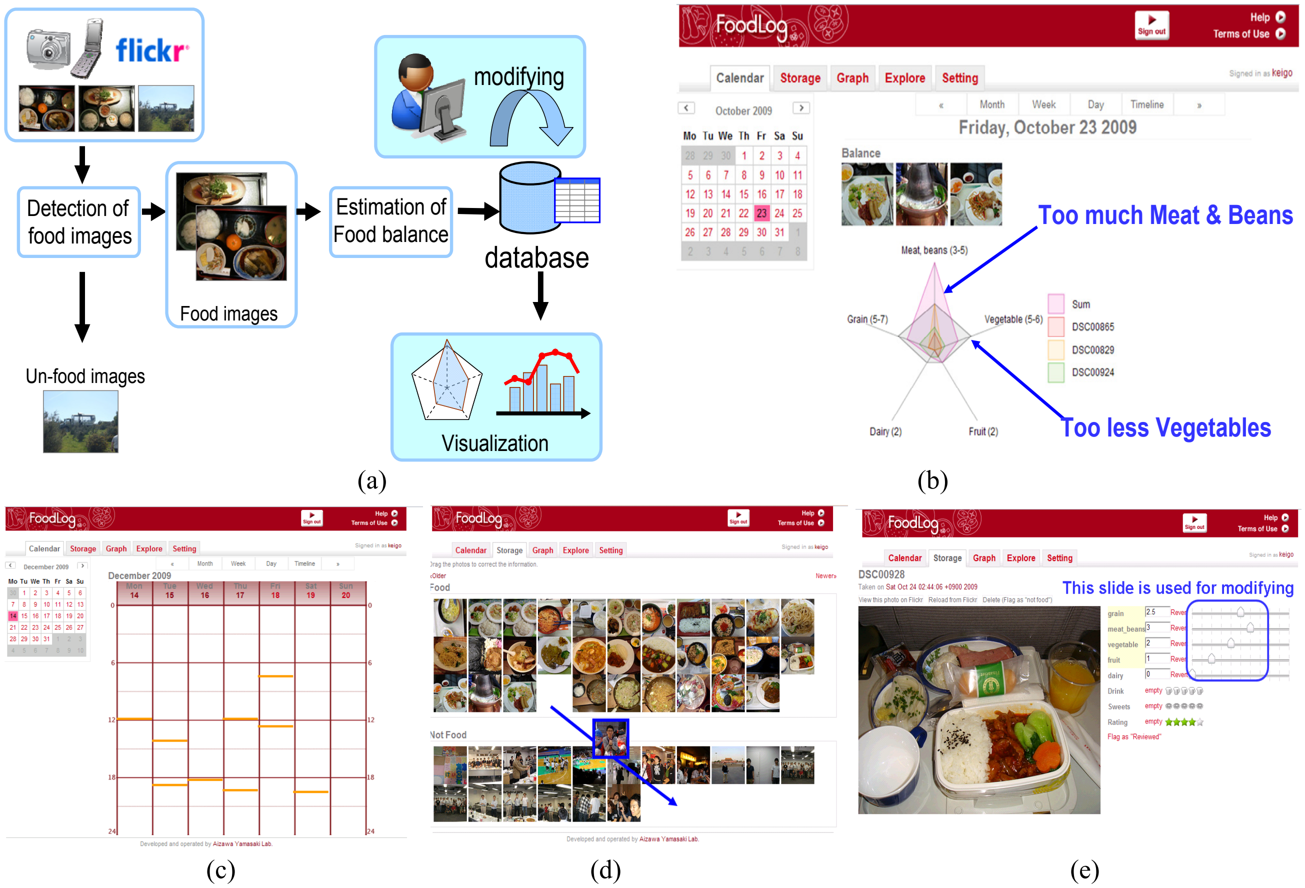

Fig 5. (a) Flow chart of the system (b) and (c) show interfaces which are daily food images and weekly change of food balance. (d) and (e) show modifying interfaces. (d) enables users to correct the result of food image detection. (e) gives users interface for correct the result of image analysis.

making and attitudes," International Journal of Consumer Studies, Vol. 32, No. 6, pp. 692-698, Nov, 2008

[2] W. Wu and J. Yang, "Fast food recognition from videos of eating for calorie estimation" In Proc. of IEEE International Conference on Multimedia & Expo, pp 1210-1213, June 2009

[3] T. Joutou and K. Yanai, "A food image recognition system with multiple kernel learning", In Proc. of IEEE International Conference on Image Processing, pp 285-288, September 2009

[4] K. Kitamura et al., "Food log by analyzing food images", Proceeding of the 16th ACM International conference on Multimedia, pp. 999–1000, 2008.

[5] FoodLog http://www.foodlog.jp/

[6] K. Kitamura et al., "FoodLog: Capture, Analysis and Retrieval of Personal Food Images via the Web", ACM Multimedia Workshop on Cooking and Eating Activities, pp. 23-30, 2009.

[7] mypyramid.gov, United States Department of Agriculture, http://www.mypyramid.gov/

[8] F. Zhua, et al., "Technology-assisted dietary assessment" In Proc. of SPIE, Vol. 6814, 681411, 2008

[9] C K. Martin et al., "Quantification of Food Intake Using Food Image Analysis". In Proc of IEEE Engineering in Medicine and Biology Society, pp. 6869-6873, September 2009

[10] C K. Martin et al., "A novel method to remotely measure food intake of free-living individuals in real time: the remote food photography method", British Journal of Nutrition, Vol. 101, pp. 446-456, 2009

[11] L. Fei-Fei and P. Parona, "A bayesian hierarchical model for learning natural scene categories," Proc. of IEEE Computer Vision and Pattern Recognition, pp. 524-531, 2005

[12] H. Berg et al., "SVM-KNN: discriminative nearest neighbor classification for visual category recognition," Proc. of IEEE Computer Vision and Pattern Recognition, pp. 2126-2136, 2006

[13] G. Csurka et al., "Visual categorization with bags of keypoints, Proc. ECCV Workshop on Statistical Learning in Computer Vision, 2004, pp. 59− 74.

[14]D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision Vol. 60, No.2, pp. 91-110, 2004

[15] Food Balance Guide http://www.j-balanceguide.com/ Ministy of Agriculture, Forestry and Fisheries, Japan

[16] V. Messina et al., "A New Food Guide: For North American Vegetarians" Canadian Journal of Dietetic Practice and Research Vol. 64, No.2, pp. 82-86, 2003 http://commons.wikimedia.org/wiki/File:Vegan_food_pyramid.svg

[17] Flickr, http://www.flickr.com