# Automatic Dietary Assessment from Fast Food Categorization

Lei Yang[1, 2], Nanning Zheng[1], Hong Cheng[1,2], John D. Fernstrom[3], Mingui Sun[4], Jie Yang[2]
[1] Institute of Artificial Intelligence and Robotics, Xian Jiaotong University, P.R.China
[2] Human Computer Interaction Institute, Carnegie Mellon University
Departments of [3]Psychiatry and [4]Neurosurgery, University of Pittsburgh

***Abstract*—This paper presents a novel approach for automatic dietary assessment from images captured from eating activities. We propose to directly estimate calories of fast food from its category and formulate dietary assessment as an object categorization problem. We use a modified bag of feature model for fast food categorization. We evaluate the proposed method in a database of McDonald fast food. The experimental results are encouraging and promising.**

## I. INTRODUCTION

Accurately acquiring diet data from free-living individuals is an essential requirement to understand the etiology of obesity and develop effective treatment methods for patients. Currently, dietary assessment is largely dependent on self-reported data from patients. However, this type of data is often inaccurate and biased. Thus the data might not be able to accurately reflect the habitual behavior of obesity patients in real life. A goal of this research is to develop a system which could assess food intake and energy expenditure automatically and objectively. We are developing a unified, miniature sensor device which combines with a microscopic video camera and other sensors such as an accelerometer, an oximeter, a semiconductor thermistor, and a microphone, etc [6]. The video camera is configured to record the same scene as the wearer perceives. The device will be used for automatically capturing eating/drinking activities as well as physical activities. Besides the hardware, we also need to develop algorithms and tools for processing and analyzing recorded multimedia data. In this paper, we present an approach for automatic dietary assessment from fast food categorization.

Fast food is the food that can be prepared and served quickly. Every day about one quarter of the U.S. population eat fast food [4]. Many fast food restaurants have standardized their food ingredients. Thus, for a given category, we could know its major nutrition facts directly (e.g., calories, fat, etc.). This provides us an easier way for automatic dietary assessment. Instead of analyzing the ingredients, we could obtain the nutrition facts directly from categorizing the food, i.e., from food to calories. Suppose that a subject has worn a video device which has captured the food eaten. If the system can automatically recognize the food with the known ingredients, it knows how many calories the subject has taken. In this way, we can formulate the dietary assessment problem as an object categorization problem. We further employ computer vision and pattern recognition techniques to solve this problem. Fig. 1 illustrates the basic concept of the proposed approach. In the rest of the paper, we describe the main algorithm (section II), present system implementation and the experiment results (section III), and make conclusions and discuss the future work (Section IV).

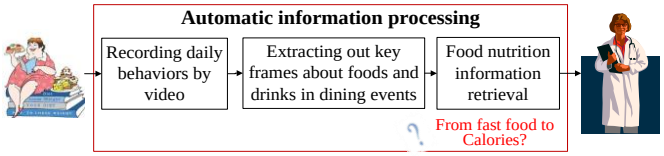



Figure 1. The automatic diet data acquiring process from video

## II. FOOD CATEGORIZATION

To determine food categories is actually a classical object categorization problem, which is an active research topic in both computer vision and pattern recognition fields. Recently, methods using "bag of features" (BoF) model achieve much success for object categorization. Thus we utilize a modified BoF approach in this project to recognize food categories.

### *A. Bag of Features*

Object categorization using BoF consists of four steps: 1) images are sampled into a collection of "features"; 2) features are mapped to a finite vocabulary based on their appearance; 3) a statistic, or signature, of such visual words is computed; 4) the signatures are fed into a classifier for labeling. All four steps can be implemented in different ways. In this paper, we adopt random sampling, SIFT descriptor, hierarchical K-means clustering algorithm, and the simple Nearest Neighbor voting classifier respectively. The process is similar to the procedure described in [5].

### *B. Multiple Segmentation*

In classical BoF methods, all visual words in an image are placed into a single histogram, without considering their spatial relationships. In order to preserve some spatial relationship, we use multiple image segmentation to produce groups of related visual features as suggested by [3].

Sufficient segmentations of each input image are produced with a high chance of obtaining a few "good" segments that will contain potential foods. We choose the mean-shift segmentation framework [1] due to its better performance and efficiency [2]. We totally get 33 segmentations, by varing spatial_band=5,7,9 and range_band = 1,3,5,7,9,11,13,15,17,19,21.However, even with multiple segmentations, it is not always possible to get a complicated food object as a single segment. We consider a larger collection of segments by merging up to 3 segments that are adjacent and from the same segmentation.

### *C. Integrating Bag of Features and Segmentation*

We integrate segmentation with BoF as follows. Each segment is regarded as a stand-alone image by masking and zero padding the original image. Then the signature of the segment is computed as in regular BoF, but any features that fall entirely outside its boundary are discarded.

Let $I$ be a test food image and its $q$-$th$ segment $S_q$, $I_{ic}$ be the $i$-$th$ training image of the $c$-$th$ category. Let $\phi(I)$ (or $\phi(s)$) be the signature of image $I$ (or segment $S$) and $\Omega(I)$ (or $\Omega(S)$) the number of features extracted form image $I$ (or segment $S$).

Segments are classified based on the nearest neighbor rule. Define the distance of the test segment $S_q$ to class c as:

$$d(S_q,c)=\min_i d(S_q,I_{ic})=\min_i \left\|\phi(S_q)-\phi(I_{ic})\right\|_1 . \quad (1)$$

We assign the segment $S_q$ to its closest category $c_1(S_q)$:

$$c_1(S_q)=\arg\min_c d(S_q,c) . \quad (2)$$

In order to combine segment labels into a unique image label, we define the second best labeling segment $S_q$ first:

$$c_2(S_q)=\arg\min_{c\neq c_1(S_q)} d(S_q,c) . \quad (3)$$

Then we compare the distance of Sq to c1 and c2, defining:

$$p(c_1(S_q)\,|\,S_q)=(1-r)+r/C,\ r=\frac{d(S_q,c_1(S_q))}{d(S_q,c_2(S_q))} . \quad (4)$$

$C$ is the number of categories. For other labels, $c\neq c_1(S_q)$:

$$p(c\,|\,S_q)=\frac{1-p(c_1(S_q)\,|\,S_q)}{C-1} . \quad (5)$$

Let $\{S_1,\ldots,S_k\}$ be all the segments of a test image $I$. The label of the food image $I$ is then given by (6):

$$C(I)=\arg\max_c \sum_{q=1}^{K} p(c\,|\,S_q)\omega(S_q),\ \omega(S_q)=\Omega(S_q)/\Omega(S_{\max}) , \quad (6)$$

where $S_{max}$ is the largest segment (in number of features).

## III. Experiment Results and Systems

### A. System Implantation for "From Food to Calories"

A miniature digital camera attached to the body is utilized to acquire video data. Once key frames that contain food are extracted from the video sequence, we can then recognize their categories and infer calories from the database. We choose McDonald fast food for test because of its popularity, standard production, and detailed nutrition information offered. The nutrition information of McDonald food could be easily found from its website (four categories of them are listed in table I). We have ignored the volume and content difference between foods in a general category in this paper. Although this estimation process is coarse, it is the first attempt toward the final goal of automatically dietary assessment.

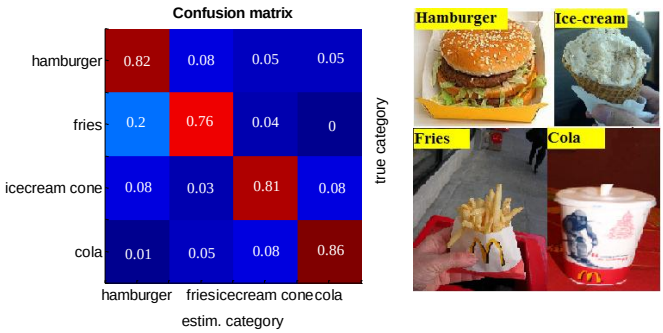
### B. Food image categorization results

To demonstrate the feasibility of proposed food categorization method, we collect a dataset with four most common fast foods in McDonald®: hamburger, fries, ice-cream cone, and Coco Cola. There are 100 images in each category. Even though the categories of fast foods are small and the database is limited, as our first attempt, the experimental results are encouraging and promising (fig.2).

20 images are used for training and the remained 80 images are for test in each category. The statistic results as well as some examples of correctly classified images are shown in fig.2. We have achieved an average categorization accuracy of 81.25% in total.

Table I. Calorie Information of Some MacDonald Fast Food

| Menu Item | Calories(C) |
|---|---|
| Hamburger | 250 |
| Medium French Fries | 380 |
| Ice Cream Cone | 150 |
| Coca-Cola Classic (Medium) | 210 |




a. Confusion matrixes of food categorization accuracy

b. Some Correct food labeling results

Figure 2. Experiment Results

## IV. Conclusions

This research is to build up a system which could assess food intake and energy expenditure automatically and objectively. We are developing a unified, miniature sensor device that is configured to record the same scene as the wearer perceives. The system will provide reliable information for dietary assessment. In this paper, we have proposed a novel method for automatic dietary assessment from images. We directly estimate calories of fast food from its category and formulate dietary assessment as an object categorization problem. We use a modified bag of feature model for fast food categorization. We have demonstrated the feasibility of the proposed method in a database of McDonald fast food. We will improve accuracy of the food categorization algorithm and work on more categories of food.

## Acknowledgements

This work is partially supported by NIH under Genes, Environment and Health Initiative (GEI) (grant No. U01 HL91736). The first author is supported by China State Scholarship fund and the work was performed in Carnegie Mellon University.

## References

[1] D. Comaniciu, and P.Meer, "Mean Shift: A Robust Approach toward Feature Space Analysis," IEEE Trans. Pattern Analysis and Machine Intelligence, 24(5), pp. 603-619, 2002.

[2] T. Malisiewicz, and A. Efros, "Improving Spatial Support for Objects via Multiple Segmentations," British Machine Vision Conference 2007, Sep.2007.

[3] A. Rabinovich, A. Vedaldi, Carolina Galleguillos, Eric Wiewiora, and Serge Belongie, "Objects in Context," ICCV 2007, 2007.

[4] E. Schlosser, Fast Food Nation: The Dark Side of the All-American Meal, Houghton Mifflin Books, 2001

[5] A.Vedaldi, http://vision.ucla.edu/vedaldi/code/bag/bag.html.

[6] N. Yao, R. Sclabassi, Q. Liu, J. Yang, J. Fernstrom, M. Fernstrom, and et al., "A Video Processing Approach to the Study of Obesity," Proceedings of IEEE ICME 2007, pp. 1727-1730, 2007.