ABSTRACT

In the demo, we demonstrate a large-scale food recognition system employing high-dimensional Fisher Vector and linear one-vs-rest classifiers. Since all the processes on image recognition perform on a smartphone, the system does not require an external image recognition server, and runs on an ordinary smartphone in a real-time way.

The proposed system can recognize 256 kinds of food by using the UEC-Food256 food image dataset we built by ourselves recently as a training dataset. To implement an image recognition system employing high-dimensional features on mobile devices, we propose linear weight compression method to save memory. In the experiments, we proved that the proposed compression methods make a little performance loss, while we can reduce the amount of weight vectors to 1/8. The proposed system has not only food recognition function but also the functions of estimation of food calorie and nutritious and recording a user’s eating habits.

In the experiments with 100 kinds of food categories, we have achieved the 74.4% classification rate for the top 5 category candidates. The prototype system is open to the public as an Android-based smartphone application.

Categories and Subject Descriptors
H.4.0 [Information Systems Applications]: General

Keywords
mobile visual recognition; food image classification; mobile application

1. INTRODUCTION

In recent years, food habit recording services for smartphones have become popular. They can awake users’ food habit problems such as bad food balance and unhealthy food trend, which is useful for disease prevention and diet. However, most of such services require selecting eaten food items from hierarchical menus by hand, which is too time-consuming and troublesome for most of the people to continue using such services for a long period.

Most of the existing mobile image recognition systems such as Google Goggles need to send images to high-performance servers, which must makes communication delay, requires communication costs, and the availability of which depends on network conditions. On the other hand, image recognition on the client side, that is, on a smartphone is much more promising in terms of availability, communication cost, delay, and server costs. Due to recent rapid progress of smartphones, they have obtained enough computational power for real-time image recognition. Then, by taking advantage of rich computational power of recent smartphones as well as recent advanced object recognition techniques, in this demo, we propose a real-time food recognition system which runs on a common smartphone. To boost accuracy and speed of food image recognition, we adopt high-dimensional Fisher Vector and linear classifiers with compressed weights, and implement a system as a multi-threaded system for using quad CPU cores effectively.

Since the recognition process on the proposed system is performed repeatedly about twice a second, a user can search for good position of a smartphone camera to recognize foods accurately by moving it continuously without pushing a camera shutter button. This is a big advantage of a real-time image recognition system on a mobile device.

We have implemented this system as an Android smartphone application so as to use quad CPU cores effectively for real-time recognition. As a result, the system takes only 0.19 second for one-time food recognition. In the experiments, we have achieved the 74.4% classification rate for the top 5 category candidates which outperformed the classification rate of our previous server-side food image recognition system.

To summarize novelties of the proposed system, it consists of three folds: The novelty of the proposed system is (1) recognizing a large number of food categories, 256 foods (2) adopting the state-of-the-art Fisher Vector with modified HoG (RootHoG) and color patches, and (3) using classifier weight compression which we newly proposed to realize a large-scale recognition with high dimensional features.
2. SYSTEM OVERVIEW

The final objective of the proposed system is to support users to record daily foods and check their food eating habits. To do that easily, we embedded a large-scale food image recognition engine on the proposed system.

The processing flow of the proposed system is as follows:

1. A user points a smartphone camera toward food items before eating them. The system is continuously acquiring frame images from the camera device in the background.

2. Food recognition is carried out continuously about twice per second for the captured frame images.

3. We extract RootHoG patches and color patches, and code them into Fisher Vector (FV) representation with level-1 Spatial Pyramid. The total dimension of the FV is 35,840. We apply one-vs-rest linear classifiers with compressed weights to the FV, and obtain top five food candidates. Note that linear classifier weights are trained in advance with the UEC-Food256 data set.

4. After food image recognition, the top five food item candidates are shown on the screen as shown in Figure 1.

5. A user selects food items from the food candidate list by touching on the screen, if found. Before selecting food items, a user can indicate relative rough volume of selected food item by the slider on the right bottom on the screen for calorie estimation. If not, user moves a smartphone slightly and go back to 1.

6. The name, calorie and nutrition of the recognized food items are shown on the screen and are recorded in the system.

3. IMPLEMENTATION AND EVALUATION

In this section, we describe the implementation of the prototype system briefly.

3.1 New Food Dataset: UEC-Food256

UEC-Food256 is a newly-constructed food image dataset, which consists of 256 kinds of foods. Each category has more than 100 images. It has been built by extending UEC-Food100 [1] with transfer learning and crowd-sourcing automatically. The detail is submitted to ECCV2014 as a technical paper.

3.2 RootHoG and Color Patches with Fisher Vectors

We use RootHOG patches and Color patches as local features, and code them into Fisher Vectors (FV) with level-1 Spatial Pyramid after applying PCA. RootHOG is an element-wise square root of the L1 normalized HOG, which is inspired by “RootSIFT” [2]. In fact, in the preliminary experiments, RootHOG led to the better performance than original HOG. Finally, we obtain a 35,840-dim feature vector for each image.

3.3 Linear Classifier with Weight Compression

We adopt 1-vs-rest linear classifiers for estimation of food categories. To train linear classifier weights, we use AROW [3].

Since we have 256 food categories, the total amount of required memory to store all the classifier weight vectors in the float representation is 35MB. This is too much for one Android application. Then, we propose a scalar-based compression method for weight vectors of linear classifiers. This method has a good characteristic that we do not need to uncompress the compressed vector when evaluating the corresponding classifier.

4. REFERENCES


