

# Caloric and Nutritional Information Using Image Classification of Restaurant Food

Arne Bech

12/10/2010

## Abstract

Self-reported calorie estimation tends to be inaccurate and unreliable, while accurate automated estimation is expensive. Close to half of all Americans are trying to loose weight, and a significant number therefore try to track their calorie intake, self-monitored or with professionals. The Smart Remote Food Photography Method (SRFPM) archives significant accuracy of classification of fast food using SVM's. This will allow the cost of accurate automated estimation to go down by reducing the amount of work the operators at the estimation facility has to do.

## 1 Introduction

Over the last decade in the US the percentage of obese americans have increased by 70% [4]. It is an alarming trend that can cause significant decrease of life quality for the affected as well as increased health cost. This is one of the reasons weight loss is a common topic with close to half of all americans trying to loose weight [1].

Regardless if a person trying to loose weight is on a self-monitored program or working with professionals (e.g. dietician), they will often track their calorie intake for use as a motivation factor to eat less as well as for analyzing their current eating habits. However, there are several problems with self reported calorie intake. Research has shown that people tend to underreport the number of calories they are consuming [6] and the hassle of calculating and reporting the caloric content of every meal often results in unreported meals.

Previous research has been done that could be useful in this area such as image based food classification and estimation of caloric content directly from an image [7] [5]. Problems with these approaches have been both complexity and low accuracy. Another approach is the Remote Food Photography Method (RFPM). It involves users uploading pictures of their meals, and then having professionals estimate the calorie content. The professionally estimated calorie values were shown to be significantly more accurate than self-reported values [8]. One problem with this approach might be the expense of having to hire professionals to do the estimations, particular compared to self-reporting.

With this project we seek to look into the intersection between these two different areas of research, machine learning and RFPM, to come up with a solution that improves accuracy and lowers cost of calorie estimation compared to machine learning and RFPM respectively. Since this is an assignment for a machine learning class, we will look at how we can use machine learning in this new integrated approach. In particular we will look at using machine learning to classify fast food based on images and geolocations.

## 2 The Problem

The approach we will take is to use machine learning to assist the human operator that is responsible for estimating calories. An use case would be an operator that receives a picture of a meal, in regular RFPM he or she would have to figure out the food type and nutritional values using reference material and

then determine the amount of food. With our approach the same use case would be an operator that receives the food image with a description of what it is e.g. the description could be "Chicken nuggets from KFC - 300 Calories a piece", then all the operator needs to do is to count the chicken nuggets. However, in more difficult cases, such as home-cooked meals the operator would only receive the image. One way of understanding this approach is to consider that machine learning can deal with the "low-hanging fruit" (e.g. more easily classifiable).

I call this new combined approach Smart Remote Food Photography Method (SRFPM).

If we limit ourselves to restaurant foods (which would be significant share of all meals), we can easily see that we have a new and very important feature we can use - location. The location can be used to reverse lookup the restaurant name (e.g. through Google Maps API ). We can justify this by realizing that user submitting pictures of their meals will very likely use a smartphone, which in most cases have A-GPS support built in.

Through the rest of this paper we'll set up and evaluate the accuracy of machine learning classification given location and a image.

### 3 Method

#### 3.1 Data

Training and testing data is taken from the Pittsburgh Fast-food Image Dataset [3] which consist of over 4500 images of fast food, with corresponding labels and names of the restaurant they belong to. The data we selected for this project is described in table 1.

To limit the scope of this project, all backgrounds have been removed from the pictures (e.g. figure 1).

#### 3.2 Extracted Features

##### 3.2.1 Location

This is used to lookup the restaurant in which the picture was taken. For this project the



(a) Pizza

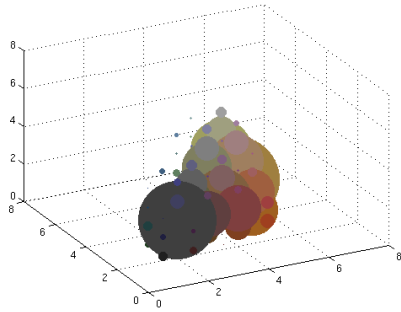


(b) Mask

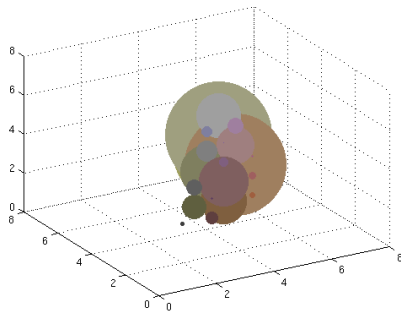
Figure 1: Background is set to black (a), mask used in (b)

Restaurant	# Different Foods	Total Images
Aarby's	11	205
KFC	11	192
McDonalds	10	209
Pizza Hut	11	204
Quiznos	8	132
Subway	10	172

Table 1: Dataset used for this project



(a) Pizza Type 1



(b) Pizza Type 2

Figure 2: Visualization of RGB Bins for two different pizza types for  $n = 8$

restaurant is known, so the lookup process itself is not directly implemented.

### 3.2.2 RGB Bins

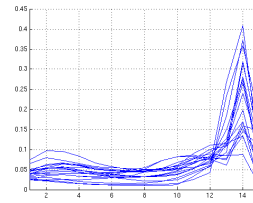
Sampling the 3D RGB space into a  $n * n * n$  matrix where each element is the normalized count of colors of that type (figure 2). This is then converted to a vector by appending each element of the 3D matrix.

### 3.2.3 Average Color

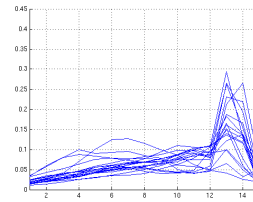
The average color of the image.

### 3.2.4 Intensity Histogram

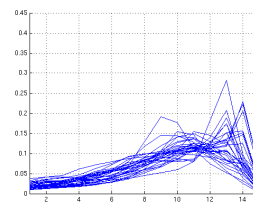
Using  $n$  coefficients, this describes the normalized grayscale histogram for the image (see figure 3).



(a) Histogram 1



(b) Histogram 2



(c) Histogram 3

Figure 3: Averaged Intensity Histograms from three different pizza types with  $n = 15$

### 3.2.5 Bag of Features (SIFT)

A set of SIFT descriptors are extracted from each image using VLFeat library [9]. For each restaurant all descriptors are combined into a large matrix and put through a k-means algorithm to extract  $k$  centroids.

The final feature vector for each image is a binary vector of size  $k$  where each element corresponds to a centroid. If one or more descriptors are mapped to a particular centroid then the respective element in the feature vector is set to one.

### 3.3 Algorithm

To classify the different foods, this project relies on SVM through the libsvm library [2]. In order to evaluate the different features we construct a set of feature vectors for each image. The feature vectors are described in table 2.

To clarify further notation we define a dataset to be the all the feature vectors of one

V	Length	Kernel	Description
1	512	Linear	RGB Bins
2	3	RBF	Avg. Color
3	15	RBF	Intensities
4	18	RBF	Avg. Color + Intensity
5	530	Linear	RGB Bins + Avg. Color + Intensity
6	9000	Linear	SIFT
7	9530	Linear	RGB Bins + Avg. Color + Intensity + SIFT

Table 2: The different feature vectors used

V	Arby's	KFC	McDonalds	Pizza Hut	Quiznos	Subway
1	69.8%	80.2%	81.3%	86.8%	65.9%	70.9%
2	59.0%	72.4%	67.5%	66.2%	62.9%	53.5%
3	57.6%	68.8%	62.2%	56.4%	55.3%	53.5%
4	65.4%	78.6%	71.8%	77.9%	58.3%	62.2%
5	67.3%	85.4%	83.7%	87.7%	66.7%	71.5%
6	87.8%	89.6%	80.4%	87.7%	90.2%	73.3%
7	87.8%	91.7%	81.3%	91.2%	89.4%	76.7%

Table 3: Accuracies for different fast-foods for different feature vectors using 5-fold cross validation

type for one restaurant.

To preprocess the data, we scale it so that it is between 0 and 1. We then run a mutual information algorithm, sorting the dataset such that the first elements of feature vectors is where we expect there to be the most useful information.

Linear kernels have a cost parameter. RBF kernels also have an additional  $\gamma$  parameter. On top of that we need to know if we should reduce the size of the feature vectors (e.g. disregarding elements with lowest mutual information). To find the good values for these parameters we run a 3D grid search (2D with linear kernels) over a range of parameters.

All accuracies are computed using 5-fold cross validation.

## 4 Results

The accuracies for the different feature vectors for different restaurants are reported in Table 3.

Parameter selections are shown for KFC and Pizza Hut for feature vector 7 (i.e. all features combined) in table 4 and table 5. Its interest-

ing that the best vectors are pruned from over 9000 elements per image to only 660 and 360 respectively after being sorted by mutual information value to get the best accuracy from the SVM's.

## 5 Discussion

This SVM based system shows that we get significant accuracy in image classification when supplemented with location information, and in this case it would do very well classifying fast food before sending it off to an operator. With KFC and Pizza Hut, we were able to archive over 90 % accuracy combining all of the features. In general the increased accuracy can probably be attributed to two main factors, the dataset is smaller and the SVM's are optimized for the particular food at a restaurant. The accuracy could also be potentially improved, if images were taken from a certain angle (e.g. from the top). Images in these datasets are taken from a variety of angles.

This algorithm could be straightforwardly implemented in a current RFFPM system, dy-

Parameter	Value
Cost	0.084
Pruned Feature Length	660
Accuracy	91.2%

Table 4: Pizza Hut: Feature Vector 7

Parameter	Value
Cost	0.084
Pruned Feature Length	360
Accuracy	91.7%

Table 5: KFC: Feature Vector 7

namically training on new classified images as they are processed by the operators.

While the results are good, it is important to realize the limitations of the accuracy reported in this project. There are several factors that favors higher accuracy that might be removed in a real world scenario. In particular the datasets contains only images taken in the laboratory which means that the lighting is somewhat consistent in all images, and bad images (blurry, wrong exposure etc.) have been removed. The dataset also contains only a subset of the foods offered at the restaurants, which will to some degree inflate the accuracy scores since we have fewer objects to match between.

Other interesting areas to explore in this area would be a similar project, but based off sit down restaurants, not fast-food, as well as learning the typical meal choices of users and further narrow the dataset by applying a prior probability distribution.

## References

- [1] CL Bish, HM Blanck, MK Serdula, M Marcus, HW Kohl, and LK Khan. Diet and physical activity behaviors among americans trying to lose weight: 2000 behavioral risk factor surveillance system. *Obes Res*, 13:596–607, 2005.
- [2] Chih-Chung Chang and Chih-Jen Lin. *LIBSVM: a library for support vector machines*, 2001.
- [3] Mei Chen, Kapil Dhingra, Wen Wu, Lei Yang, Rahul Sukthankar, and Jie Yang. Pfid: Pittsburgh fast-food image dataset. 2010.
- [4] Eric A Finkelstein, Ian C Fiebelkorn, and Guijing Wang. National medical spending attributable to overweight and obesity: How much, and who’s paying? *Health Affairs Web Exclusive*, May 2003.
- [5] King-Shy Goh, Edward Chang, and Kwang-Ting Cheng. Svm binary classifier ensembles for image classification. In *Proceedings of the tenth international conference on Information and knowledge management, CIKM ’01*, pages 395–402, New York, NY, USA, 2001. ACM.
- [6] Michael E. Holmstrup, Kay Stearns-Bruening, and Timothy J. Fairchild. Caloric estimation bias of realistic meal and beverage preparations, 2008.
- [7] C.K. Martin, S. Kaya, and B.K. Gunturk. Quantification of food intake using food image analysis. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pages 6869 –6872, 2009.
- [8] Corby K. Martin, Hongmei Han, Sandra M. Coulon, H. Raymond Allen, Catherine M. Champagne, and Stephen D. Anton. A novel method to remotely measure food intake of free-living individuals in real time: the remote food photography method. *British Journal of Nutrition*, 101(03):446–456, 2009.
- [9] A. Vedaldi and B. Fulkerson. Vlfeat: An open and portable library of computer vision algorithms. <http://www.vlfeat.org/>, 2008.