International Journal of Research

International Journal of Research (IJR)

e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 09, September 2015

Available at http://internationaljournalofresearch.org

A Food Recognition System for Diabetic Patients Based on an Optimized Bag-of-Features

K.Naga siva prasad

M Tech Scholor, Dept of CSE, JJ Institute of Information Technology

M Om Prakash

Asst. Professor & Head , Dept of CSE, JJ Institute of Information Technology,

Kamal narayan Kamlesh

Asst. Professor, Dept of CSE, JJ Institute of Information Technology,

Abstract—

Computer vision-based food recognition could be used to estimate a meal's carbohydrate content for diabetic patients. This study proposes a methodology for automatic food recognition, based on the bag-of-features (BoF) model. An extensive technical investigation was conducted for the identification and optimization of the best performing components involved in the BoF architecture, as well as the estimation of the corresponding parameters. For the design and evaluation of the prototype system, a visual dataset with nearly 5000 food images was created and organized into 11 classes. The optimized system computes dense local features, using the scale-invariant feature transform on the HSV color space, builds a visual dictionary of 10000 visual words by using the hierarchical k-means clustering and finally classifies the food images with a linear support vector machine classifier. The system achieved classification accuracy of the order of 78%,thus proving the feasibility of the proposed approach in a very challenging image dataset.

Index Terms—Bag of features (BoF); diabetes; feature extraction; food recognition; image classification

INTRODUCTION

THE treatment of Type 1 diabetic (T1D) patients involves exogenous insulin administration on a daily basis. A prandial insulin dose is delivered in order to compensate for the effect of a meal The estimation of the prandial dose is a complex and time-consuming task, dependent on many factors, with carbohydrate (CHO) counting being a key element.Clinical studies have shown that, in children and adolescents on intensive insulin therapy, an inaccuracy of ±10 g in CHO counting does not impair postprandial control while a±20 g variation significantly impacts postprandial glycaemia There is also evidence that even well-trained T1D patients find it difficult to estimate CHO precisely 184 adult patients on intensive insulin were surveyed with respect to the CHO content of their meals. On average, respondents overestimated the CHO contained in their breakfast by 8.5% and underestimated CHO for lunch by 28%, for dinner by 23%, and for snacks by 5%. In [5], only 23% of adolescent T1D patients estimated daily CHO within 10 g of the true amount, despite the selection of common meals. For children with T1D and their caregivers, a recent study has shown that 27% of meal estimations are inaccurate in ranges greater than ±10 g

The increased number of diabetic patients worldwide, together with their proven inability to assess their diet accurately raised the need to develop systems that will support T1D patients during CHO counting. So far, a broad spectrum of mobile phone applications have been proposed in the literature, ranging from interactive diaries to dietary monitoring based on on-body sensors The increasing processing power of the mobile devices, as well as the recent advances made in computer vision, permitted the introduction of image/video analysis-based applications for diet management In a typical scenario, the user acquires an image of the upcoming meal using the camera of his phone. The image is processed—either locally or on the server side—in order to extract a series of features describing its visual properties. The extracted features are fed to a classifier to recognize the various food types of the acquired image, which will then be used for the CHO estimation.

Recently, the bag-of-features (BoF) model was introduced into the area of computer vision as a global image descriptor for difficult classification problems BoF was derived from the bag-of-words (BoW) model. BoW is a popular way of representing documents in natural language processing [13], which ignores the order of the words belonging to a previously defined word dictionary, and considers only how frequently they appear. Similarly, in the image analysis context, an image is represented by the histogram of visual words, which are defined as representative image patches of commonly occurring visual patterns. The concept of the BoF model adequately fits the food recognition problem, since a certain food type is usually perceived as an ensemble of different visual elements mixed with specific proportions, but without any typical spatial arrangement, a fact that encourages the use of a BoF approach, instead of any direct image-matching technique. Puri *et al*. proposed a pairwise classification framework that takes advantage of the user's speech input to enhance the food recognition process. Recognition is based on the combined use of color neighborhood and maximum response features in a texton histogram model, feature selection using Adaboost, and SVM classifiers. Texton histograms resemble BoF models, using though simpler descriptors, such that histograms of all possible feature vectors can be used. In this way, the feature vector clustering procedure can be omitted; however, less information is considered by the model which might not be able to deal with high visual variation. Moreover, the proposed system requires a colored checker-board captured within the image in order to deal with varying lighting conditions. In an independently collected dataset, the system achieved accuracies from 95% to 80%, as the number of food categories increases from 2 to 20.

METHODS DESCRIPTION

The proposed food recognition system consists of two stages: food image description and image classification .During food image description, a set of characteristics representing the visual content of the image is extracted and quantified. This set provides input to the second stage, where a classifier assigns to the image one class out of a predefined set of food classes. The design and development of both stages involves two phases: training and testing. During the training phase, the system learns from the acquired knowledge, while during the testing phase the system recognizes food types from new, unknown images.

A. Food Image Description

In order to describe the appearance of the different food classes, the BoF model was adopted, due to its proven ability to deal with high visual diversity and the absence of typical spatial arrangement within each class. BoF consists of four basic steps: 1) key point extraction, 2) local feature description, 3) learning the visual dictionary, and 4) descriptor quantization. All the steps, as presented in Fig. 1, are involved in both training and testing, except for the learning of the dictionary, which is performed only once, during the training phase.

1) Key Point Extraction: Key points are selected points on an image that define the centers of local patches where descriptors will be applied. In the current study, three different key point extraction methods were tested: interest point detectors, random sampling, and dense sampling.

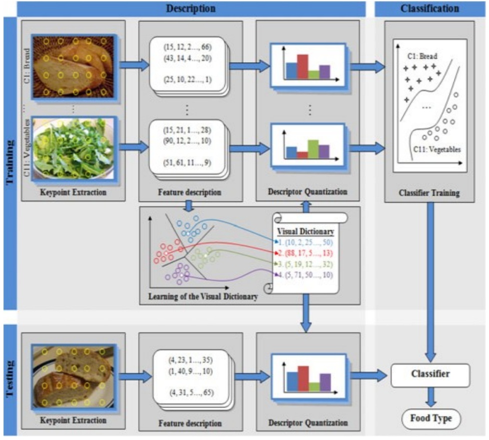


Fig. 1. Architecture of the proposed BoF-based food recognition system. The two major stages of food description and classification are illustrated within the training and testing phases. Food description includes key point extraction, feature description, descriptor quantization, and dictionary learning.

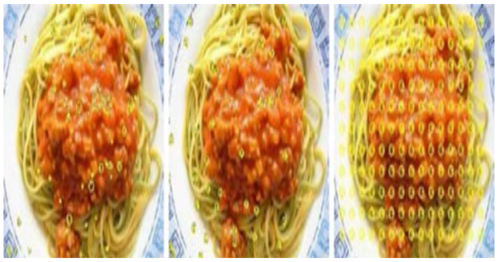


Fig. 2. Key point extraction techniques: (a) SIFT detector, (b) random sampling, and (c) dense sampling.

Interest point detectors, such as SIFT , are considered as the best choice for image matching problems where a small number of samples is required, as it provides stability under local and global image perturbations. SIFT estimates the key points by computing the maxima and minima of the difference of Gaussians (DoG), applied at different scales of the image. However, these techniques often fail to produce a sufficient number of image patches for classification problems, where the number of sampled patches constitutes the most important factor. Random and dense sampling methods have been widely used in image classification with great success since they are able to provide a BoF-based system with the required number of image patches .Random sampling is based on the random selection of point coordinates which usually follows a uniform distribution. Dense sampling is performed by extracting key points from a dense grid on the image. Any required number of points can be achieved by adjusting the spacing among them.

2) Local Feature Description: After the key point extraction, a local image descriptor is applied to a rectangular area around each key point to produce a feature vector. Identifying the appropriate descriptor size and type for a recognition problem is a challenging task that involves a number of experiments.

For the determination of the optimal descriptor size, the size\ of the object to be recognized should be considered. Although the SIFT interest point detector provides the position of the key points together with their scale, it is rarely used for image classification, as already explained. Hence, the size of the descriptor must be specified somehow after the dense or random key point sampling. A minimum size of 16 × 16 is often used as proposed in since a smaller patch would not provide sufficient information for the description. However, the use of larger sizes or combination of sizes can often give better results by resembling the multiscale image description of the SIFT detector. It should be noted that food images are scaled to a standard size, so differences in food items scale should not be extreme.

EXAMPLES OF CLUSTERED IMAGE PATCHES AND VISUAL WORDS

	Image patches				Visual word (cluster center)
Cluster 1			...		(35, 2,..., 12)
Cluster 2			...		(3, 82,..., 72)
Cluster 3			...		(16, 9,..., 42)

SIFT: SIFT considers a region of 16 × 16 pixels around a given key point. Then, the region is divided into 4 × 4 subregions and for each of them an eight-bin histogram of the intensity gradient orientation is computed, leading to a 128-dimensional feature vector. SIFT features provide remarkably informative texture description, and are invariant to light intensity changes. Despite the superior texture description capabilities of the original SIFT, its inability to capture any color information constitutes a problem for the description of many objects, including foods.

Color SIFT variants: Color SIFT is computed similarly to SIFT, but with one major difference. Instead of using just the intensity image, the histograms of gradient orientations are computed in various color channels, so the resulting descriptor constitutes the concatenation of the individual descriptors. Thus, the size of the created feature vector is 128 × *NC*, where *NC* is the number of color channels used. The variants used for experimentation in the framework of this study operate in the RGB, HSV, Hue, Opponent, and C-Invariant color spaces [24], producing the rgbSIFT, hsvSIFT, hueSIFT, opponentSIFT, and cSIFT, respectively. In addition, rgSIFT has also been used; this applies SIFT only in the Rnorm and Gnorm components of the normalized RGB color model.

3) Learning the Visual Dictionary: Once the descriptors of each training image patch have been computed, the most representative patches need to be identified which will constitute the system's visual words. To this end, the feature vectors that correspond to the various patches are clustered in a predefined number of clusters. The centers of the created clusters constitute the visualwords of the dictionary, while the entire clustering procedure is known as the dictionary learning. Table I provides some examples of clustered image patches and visual words.

The most common clustering technique used for the creation of visual dictionaries is the *k*-means clustering algorithm.

4) Descriptors Quantization: Descriptor quantization is the procedure of assigning a feature vector to the closest visual word of a predefined visual vocabulary. Once the visual dictionary is learnt, each descriptor of an image is quantized and the histogram of visual word occurrences serves as a global description of the image. Then, the histogram values are usually scaled to [0 1] and fed to the classifier either for training or testing. The efficiency of this part of the system is crucial, since it affects processing times for both training and testing. The complexity of the descriptor quantization mainly depends on the dimensions of the descriptor and the number of visual words.

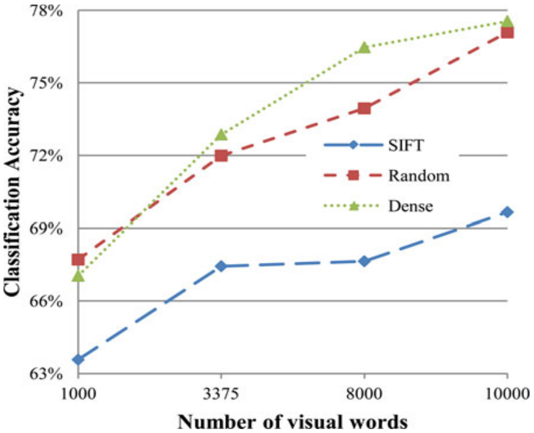
B. Food Image Classification

The image classification stage is involved in both training and testing phases. In order to identify the appropriate classifier for the specific problem, several experiments with three supervised classification methods were conducted: SVM, ANN, and Random Forests (RF).

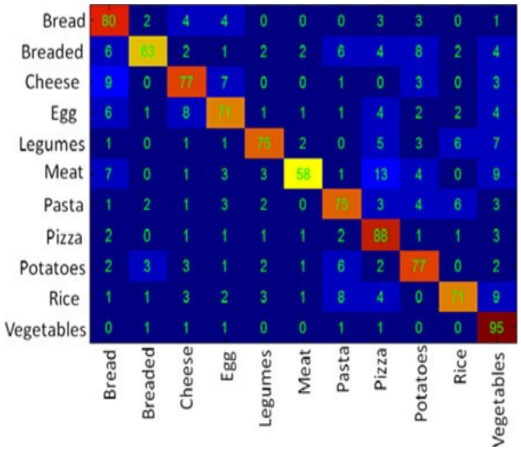
Food Image Dataset

For the experimental needs of the system developed a dataset of 4868 color images was created by collecting images from the web. The food types and their categorization were identified in collaboration with the Department for Endocrinology, Diabetology and Clinical Nutrition of Bern University Hospital, Inselspital.

Sample images of the developed visual dataset. The dataset contains nearly 5000 food images organized into 11 classes.

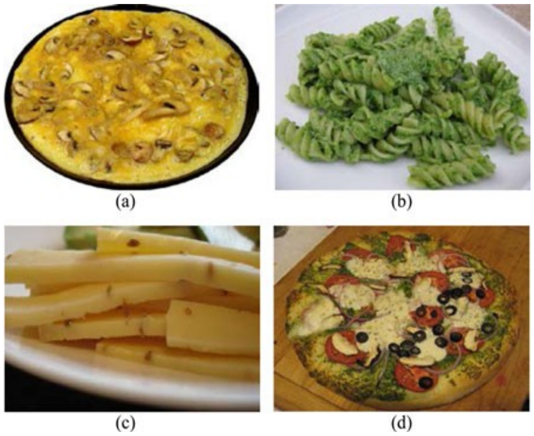


Comparison of sampling strategies in terms of overall classification accuracy. SIFT key point detector, random, and dense sampling have been used.



	Bread	Breaded	Cheese	Egg	Legumes	Meat	Pasta	Pizza	Potatoes	Rice	Vegetables
Bread	80	2	4	4	0	0	0	3	3	0	1
Breaded	6	63	2	1	2	2	6	4	8	2	4
Cheese	9	0	77	7	0	0	1	0	3	0	3
Egg	6	1	8	71	1	1	1	4	2	2	4
Legumes	1	0	1	1	75	2	0	5	3	6	7
Meat	7	0	1	3	3	58	1	13	4	0	9
Pasta	1	2	1	3	2	0	75	3	4	6	3
Pizza	2	0	1	1	1	1	2	88	1	1	3
Potatoes	2	3	3	1	2	1	6	2	77	0	2
Rice	1	1	3	2	3	1	8	4	0	71	9
Vegetables	0	1	1	1	0	0	1	1	0	0	95

Confusion matrix of the proposed optimized system. The entry in the *i*th row and *j*th column corresponds to the percentage of images from class *I* that was classified as class *j*.



Examples of incorrectly classified images: (a) eggs classified as pizza, (b) pasta classified as vegetable, (c) cheese classified as potatoes, and (d) pizza\ classified as vegetables.

CONCLUSION

In this paper, we propose a BoF-based system for food image classification, as a first step toward the development of a portable application, providing dietary advice to diabetic patients through automatic CHO counting. A series of five major experiments was carried out for choosing and optimizing the involved components and parameters of the system. The experiments were conducted on a newly constructed food image dataset with 4868 images of central European food belonging to 11 different food classes.

REFERENCES

[1] American Diabetes Association, “Standards of medical care in diabetes- 2010,” *Diabetes Care*, vol. 33, no. 1, pp. S11–S61, 2010.

[2] C. E. Smart, K. Ross, J. A. Edge, C. E. Collins, K. Colyvas, and B. R. King, “Children and adolescents on intensive insulin therapy maintain postprandial glycaemic control without precise carbohydrate counting,” *Diabetic Med.*, vol. 26, no. 3, pp. 279–285, 2009.

[3] C. E. Smart, B. R. King, P. McElduff, and C. E. Collins, "In children using intensive insulin therapy, a 20-g variation in carbohydrate amount significantly impacts on postprandial glycaemia," *Diabetic Med.*, vol. 29, no. 7, pp. e21–e24, Jul. 2012.

[4] M. Graff, T. Gross, S. Juth, and J. Charlson, "How well are individuals on intensive insulin therapy counting carbohydrates?" *Diabetes Res. Clinical Practice*, vol. 50, suppl. 1, pp. 238–239, 2000.

[5] F. K. Bishop, D. M. Maahs, G. Spiegel, D. Owen, G. J. Klingensmith, A. Bortsov, J. Thomas, and E. J. Mayer-Davis, "The carbohydrate counting in adolescents with type 1 diabetes (CCAT) study," *Diabetes Spectr.*, vol. 22, no. 1, pp. 56–62, 2009.

K.Naga siva prasad.
M Tech Scholor, Dept of CSE,
JJ Institute of Information Technology,

M Om Prakash
Asst. Professor & Head , Dept of CSE,
JJ Institute of Information Technology,

Kamal narayan Kamlesh
Asst. Professor, Dept of CSE,
JJ Institute of Information Technology,