Neurocomputing 140 (2014) 242–251

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Food image classification using local appearance and global structural information

Duc Thanh Nguyen [a], Zhimin Zong [a], Philip O. Ogunbona [a], Yasmine Probst [b], Wanqing Li [a,*]

[a] School of Computer Science and Software Engineering, University of Wollongong, NSW 2522, Australia
[b] School of Health Sciences, University of Wollongong, NSW 2522, Australia

ARTICLE INFO

Article history:
Received 27 April 2013
Received in revised form
11 December 2013
Accepted 3 March 2014
Communicated by Jinhui Tang
Available online 13 April 2014

Keywords:
Food image classification
Local binary pattern
Non-redundant local binary pattern
Shape context

ABSTRACT

This paper proposes food image classification methods exploiting both local appearance and global structural information of food objects. The contribution of the paper is threefold. First, non-redundant local binary pattern (NRLBP) is used to describe the local appearance information of food objects. Second, the structural information of food objects is represented by the spatial relationship between interest points and encoded using a shape context descriptor formed from those interest points. Third, we propose two methods of integrating appearance and structural information for the description and classification of food images. We evaluated the proposed methods on two datasets. Experimental results verified that the combination of local appearance and structural features can improve classification performance.

1. Introduction

The high incidence of obesity has been linked to an imbalanced food intake [1]. It is believed that a better understanding of the aetiology and effective health management programs should be developed through a better food-intake reporting system. Conventionally, this has been achieved manually through self-reporting or recording from observation. However, numerous studies have revealed that data obtained by these means seriously underestimates food intake, and thus does not accurately reflect the habitual eating behaviour of humans in real life [2–4].

Recently, image processing and pattern recognition techniques have been applied to improve the accuracy and efficiency of food intake reporting through automatic image-based food recognition systems [5]. In these systems, a comprehensive nutrition database is used to generate a daily food intake report for individuals based on computerised recognition of food images. Motivated by the importance of the health related issues associated with food intake and the progress made to date in the application of pattern recognition-based methods, this paper focusses on developing a food image recognition and classification method. Such a recognition and classification tool forms the core of a computerised food intake reporting system. We note that the problem of food recognition is not a simple test case of object recognition and this has been observed in a number of food recognition and classification research publications [6–8,4]. Largely, this is because of the possible variations in appearance (colour, texture, and shape) and viewpoints of food images. The problem is also exacerbated by the complexity of the recording environment, e.g. uncontrolled photographing conditions and illumination conditions.

Generally speaking, the state-of-the-art methods for food image recognition and classification have used descriptors that mainly exploit appearance-based features including colour [6], texture [7] and shape [8,4] in describing food objects. While several of these appearance-based descriptors have been successful in existing food image recognition and classification methods, structural information of food objects has been ignored. It would seem that structural information is as important as appearance information. Moreover, a combination of appearance-based features and structural features would enhance the recognition performance. In this paper, we propose to combine both local appearance and global structure in the description and classification of food images. The contributions of this paper are summarised as follows:

- To take advantage of texture as a discriminative feature in describing the appearance information of food objects, we propose the use of non-redundant local binary pattern (NRLBP) to encode the local textures of food images.
- In order to describe the structural information of food objects, we use the scale-invariant interest points [9] and employ

* Corresponding author.
E-mail addresses: dtn156@uowmail.edu.au (D.T. Nguyen), philipo@uow.edu.au (P.O. Ogunbona), yasmine@uow.edu.au (Y. Probst), wanqing@uow.edu.au (W. Li).

http://dx.doi.org/10.1016/j.neucom.2014.03.017
0925-2312/

a shape context descriptor [10] to encode the spatial relationship between interest points.

- We propose two different methods to integrate both the local appearance and global structural information in describing and classifying food images.

The proposed methods were evaluated on two different datasets: the Pittsburgh Fast-Food Image (PFI) dataset [6] and a new dataset we collected with other food categories. Experimental results showed that combining both the local appearance and global structural information could enhance the classification accuracy and outperform the baselines [6] provided with the PFI dataset.

The rest of this paper is organised as follows. In Section 2 we provide a brief review of existing work on food image classification. Section 3 presents basic elements such as appearance-based features and structural features as well as how to combine those features in describing and classifying food images. Experimental results along with comparative analysis are presented in Section 4. Section 5 concludes the paper and discusses future work.

2. Related work

In food image classification, colour has been considered as one of the important features. For example, Chen et al. [6] employed a $4 \times 4 \times 4$-bin RGB colour histogram (each bin corresponds to one of the components Red, Green, and Blue) to describe food images. Each pixel in the food image was then mapped to its closest bin in the histogram to generate a 64-dimensional feature vector representing that food image. The 64-dimensional feature vectors of all training food images were used to train a support vector machine (SVM) classifier for food image classification.

For texture information, Gabor texture features extracted on local regions of 3×3 and 4×4 and at various scales and orientations were employed in the work of Joutou and Yanai [7]. Similar to the colour histogram, the texture features of all local regions were concatenated to create a richer and higher dimensional feature vector to describe a food image.

Shape information is also used in classification of food objects. For example, in [8], the size (counted as the area) and shape (represented by the ratio of the difference between the major and minor axes, and their sum) of bread objects were used as classification cues. However, this method requires a prior knowledge of the size and shape of food objects while such information may not always be available especially in a multi-food recognition environment with possible occlusion and variable shape.

Recently, the popular scale-invariant feature transform (SIFT) descriptor, introduced by Lowe [9], was employed to encode the local shape of food objects. For instance, in [4], interest points and their local features were extracted using the SIFT detector. The classification then proceeded frame-by-frame by matching individual SIFT features from a newly acquired food image to a database of pre-trained features. This process is similar to matching key points SIFT descriptors in [9]. The advantages of SIFT descriptors are well documented. First, SIFT features extracted at interest points are local features with high informative content. Second, they are stable under local and global perturbations in the image domain. In particular, SIFT features are invariant to image scale and rotation, and have been shown to provide robust matching across a substantial range of affine distortion, change in 3D viewpoint, addition of noise, and change in illumination.

Similar to [4], but instead of encoding food images directly using SIFT features, the Bag-of-Features (BoF) approach was employed in [6,7]. This approach is inspired by the Bag-of-Words (BoW) approach devised originally for text classification [11]. In the BoF approach, codewords are represented by image features (e.g. SIFT features in this case). Each food image is then represented as a histogram of codewords defined by a codebook. The histogram is considered as a feature vector and used to train a discriminative classifier, e.g. SVM. In [7], colour histogram, Gabor texture features, and SIFT features were together employed to train a multiple kernel learning (MKL) SVM in which a sub-kernel was assigned to each type of features.

In general, extracting features at interest points [4,6,7] has advantages in capturing the local information of food images and coping with the deformation of the shape of food objects. However, to date, existing work has not considered how to exploit the spatial relationship between interest points despite the potential importance of this information for object recognition. The topology of interest points embodies the structural information of objects and thus when combined with appearance information becomes a powerful feature to discriminate an object from others.

3. Proposed food image classification

In this paper, we explore combining both local appearance and global structural information for enhancing the description and classification of food images. In particular, the SIFT detector [9] is used to detect interest points. Non-redundant local binary pattern (NRLBP) [12] is employed as the local textural descriptor and extracted at interest points to describe the appearance information of food objects. The topology of interest points represents the structural information of food objects and is encoded as a shape context descriptor [10]. We propose two food image classification methods to integrate the local appearance information with the global structural information. The first method extends the Bag-of-Features approach [4,6,7] by including the structural information. The second method is based on our previous work in [13]. However, different from [13], in the second method, NRLBP is used. In the rest of this section, we briefly describe the NRLBP (Section 3.1) and the shape context (Section 3.2). The two classification methods will then be presented (Section 3.3).

3.1. Local binary pattern (LBP) and non-redundant local binary pattern (NR-LBP)

Local binary pattern (LBP) is an effective descriptor to describe local texture [14] and has been widely used in a range of applications including texture classification [15], object recognition [16] and detection [17,18,12]. The success of the LBP descriptor has been due to its robustness under illumination changes, computational simplicity and discriminative power. Fig. 1 represents an example of the LBP in which the LBP code of the centre pixel (in red colour and value 20) is obtained by comparing its intensity with neighbouring pixels' intensities. The neighbouring pixels whose intensities are equal or higher than the central pixel's are labelled as “1”; otherwise as “0”.

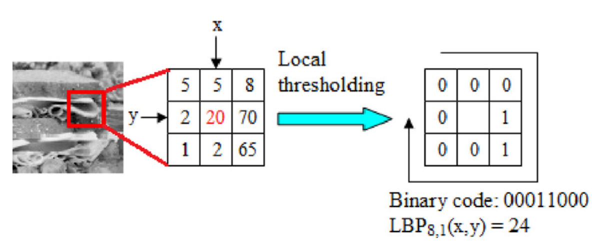

Fig. 1. An illustration of the $LBP_{8,1}$ descriptor. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this paper.)