ARTICLE IN PRESS

Neurocomputing xx (xxxx) xxxx–xxxx

Contents lists available at ScienceDirect

Neurocomputing

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

Exploiting distinctive topological constraint of local feature matching for logo image recognition

Panpan Tang, Yuxin Peng*

Institute of Computer Science and Technology, Peking University, Beijing 100871, China

ARTICLE INFO

Communicated by Zidong Wang

Keywords:
Local feature matching
Distinctive topological constraint
Feature selection
Logo image recognition

ABSTRACT

Robust local feature matching plays an important role in the challenging task of logo image recognition. Most traditional methods consider the individual local feature but ignore the affine-invariant geometric relationship among the adjacent local features, which is essential to reduce the number of mismatching. In addition, they do matching for all of the local features and ignore that many ones are insignificant, which increase the probability of mismatching and the computation complexity. To address the two limitations, we propose a robust matching method to get the better matching results by exploiting the distinctive topological constraint together with the feature selection. In the proposed method, first we employ the distinctive topological constraint to enhance the describing ability of local features, which makes full use of the affine-invariant geometric relationship among adjacent local features for more accurate local feature matching. Second, we utilize the feature selection algorithm based on the mutual information (MI), to filter out most insignificant local features before matching, which is efficient and effective to guarantee the performance of local feature matching. We evaluated the proposed method on two challenging datasets, i.e, FlickrLogos-32 and FlickrLogos-27, and achieve superior performance against the state-of-the-art methods in the literature.

1. Introduction

With the rapid expansion of digital camera and the Internet, visual data including images and videos grows explosively over the past decades. In these visual data, logos are a special class of visual objects which are closely associate with products or services. Recognition of logos in these visual data has huge commercial benefits, such as measurement of the exposure of brands in advertisement videos, protection of intellectual property in e-commerce platforms, product brand management on social media, etc. Therefore, logo image recognition has attracted increasing interests in recent years.

Logo image recognition is a specific case of image recognition, aiming at determining whether an image contains any logos and where the logos are located. However, the affine transformations such as rotation, shearing and scaling, which are caused by variety of perspectives, make logo image recognition a challenging task. Besides, occlusion, illumination change as well as diverse appearance make logo image recognition more difficult. In order to recognize logo image effectively, the most direct way is to adopt the popular methods used for image recognition, such as the boosted cascade method [1], the discriminatively trained part-based models (DPM) [2], and the deep learning based methods [3,4]. These methods have demonstrated superior performance on human face detection, pedestrian detection and general object detection, some of which can achieve relatively high precision on logo image recognition task. However, these methods are not specifically designed for logo image recognition, which restricts their performance. Firstly, the great variety of real-world logo types make the efficiency requirement of these methods critical, such as the boosted cascade method, which train a specific model for each logo type. Secondly, most methods for general object recognition are much complicated that they are not that suitable for logo recognition without specific optimization, such as DPM. Thirdly, most logo types are planar objects, which brings influence in two aspects: 1) the shapes and poses of logo are fewer than general spatial objects, and can not provide enough information to train an effective recognition model. 2) some methods in computational geometry can be adopted to help improve the recognition precision, such as the Delaunay Triangulation [5].

Recently, a number of works have tackled logo recognition using image matching based on local features [6–12] because of their robustness to affine transformations, occlusion, and illumination change. Generally, this kind of methods includes three main steps: firstly, local features such as SIFT [13], SURF [14] or FAIR-SURF [15] are extracted from images. Secondly, the correspondences between images are determined by comparing the local features in one image

* Corresponding author.
E-mail address: pengyuxin@pku.edu.cn (Y. Peng).

http://dx.doi.org/10.1016/j.neucom.2016.08.110
Received 28 March 2016; Received in revised form 7 July 2016; Accepted 8 August 2016
Available online xxxx

against the local features in the other. Finally, the similarity between two images is calculated based on the matched local features. To speed up the matching progress, local features are usually clustered and quantized into individual integer numbers called visual words. The commonly used algorithms for clustering and quantizing include K-Means, Approximate K-Means (AKM) [16], Multitask Spectral Clustering (MTSC) [17], and so on. Compared with the original features, high efficiency of the quantized features make them more suitable for large scale image recognition systems. Further, the quantized features can be represented as sparse coding to encode visual objects [18]. However, most of these works consider the individual local feature but ignore the affine-invariant geometric relationship among the adjacent local features, which is essential to enhance the describing ability of local feature and reduce the number of mismatching. In addition, they do matching for all of the local features and ignore that many ones are insignificant, which increase the probability of mismatching as well as the computation complexity. An optional solution is only considering local features in the mask region. However, not all local features in the mask region are relevant to the recognition result. Besides, the number of local features in different mask regions varies a lot, which significantly affects the recognition result. To address the two limitations, we propose a robust matching method to get the better matching results by exploiting the distinctive topological constraint together with the feature selection. Firstly, we employ the distinctive topological constraint to enhance the describing ability of local features, which makes full use of the affine-invariant geometric relationship among the adjacent local features for the effective local feature matching. Secondly, we utilize the feature selection algorithm based on the mutual information (MI), to filter out most insignificant local features before matching by measuring their importance to the final results, which is efficient and effective to guarantee the performance of local feature matching.

Compared to existing literature, the distinctive contributions of this paper lie on three aspects: (1) the topological constraint proposed in this paper is distinctive, which is effective to reduce mismatching between logo images and is different from those in the existing literature. (2) We utilize feature selection algorithm based mutual information to filter those irrelevant features that disturb the matching process. This practice is effective and is not seen in the relevant literature. (3) We realize logo localization based on the optimized matched local features, which has not been done by the similar methods. Fig. 1 shows the process of the proposed method. We carry out experiments on two challenging logo recognition benchmarks to evaluate the proposed method, and experimental results show its superior performance against the state-of-the-art methods in the literature.

This paper includes the preliminary work in [19] but significantly extends it in the following ways. Firstly, we implement logo localization based on the matched local features optimized by the proposed topological constraint. Secondly, we compare the proposed method with more state-of-the-art methods, including the deep learning based method, in order to extensively validate the effectiveness of the proposed method.

The rest of the paper is organized as follows. After reviewing the related work in Section 2, we present the details of the proposed distinctive topological constraint and recognition framework in Section 3. Section 4 reports extensive experimental results that validate the effectiveness of the proposed method. Finally, we conclude the paper in Section 5.

2. Relate work

By application of more powerful features, sophisticated models and effective recognition strategies, great progress has been achieved for image recognition over the past years. Nevertheless, as one of the special case, logo image recognition has not been addressed with enough emphasis, and only limited attention has been paid in the literature. Generally, there are mainly two kinds of methods for logo image recognition: one is the popular methods for general image recognition, and the other is based on local feature matching.

There are a lot of methods for general image recognition in the literature, which we can find from the popular visual challenges such as PASCAL VOC [20] and ILSVRC [21] over the past decades. Some of the methods have demonstrated superior performance for many years and are regarded as the landmark methods, such as the boosted cascade method [1], the discriminatively trained part-based models (DPM) [2], and the deep learning based methods [3,4]. The boosted cascade method is a classical method for face detection. It extracts simple features from images and utilizes the AdaBoost based learning algorithm to yield extremely efficient classifiers. Then it combines increasingly more complex classifiers in a "cascade" which allows background regions of the image to be quickly discarded while spending more computation on promising object-like regions. The restrictions on employing this method in logo image recognition lies on two aspects: first, it requires at least hundreds of positive samples in the training stage, which are not easily available in practice. Second, the limited shapes and poses of logo usually can not provide enough information to train an effective detector. The DPM model has won champions in the PASCAL VOC challenges from 2007 to 2009, which represents object as a mixture of several deformable parts. It sums the score of all the part detectors within a given window, and identifies the window as a positive object if the summed score is greater than a pre-defined threshold. However, the high complexity of DPM restricts its generalization in logo recognition, since the test stage alone will cost about one week on a regular PC for the 20 categories of PASCAL VOC. To speed up DPM, Guo et al. [22] extracted saliency map from the origin image to get the candidate detection area. However, the strategy is not suitable for logo, because it is always tiny and clings to some other object. Recently, some deep learning based methods have achieved superior performance in image recognition, such as the RCNN [3] series. Besides computer vision, deep learning has incomparable effect in some other fields, whose effectiveness and efficiency rely heavily on the amount of training data and GPU, respectively.

Recently, logo image recognition based on local feature matching has attract the attention of researchers. On one hand, local feature like SIFT is robust to scale, rotation, and illumination changes. On the other hand, compared with the general methods for image recognition, local feature matching based methods are relatively simple but effective on logo recognition. The key step of this kind of methods is to identify the correspondences between images by comparing the local features in one image against the local features in the other, and mismatching of local features is the biggest obstacle to restrict the performance, which means that two matched local features locate at two different objects or different regions of the same object. To enhance the discriminative ability of individual local feature and reduce the number of mismatching, Huang et al. [23] employed multiple Gaussian distributions to capture objects' global spatial structure, which alleviates the sensitivity to object shifting. Zhou et al. [24] proposed a novel scheme, which is called spatial coding, to encode the spatial relationships among local features. Since it is specifically for partial-duplicate image search and is based on the assumption that two matched images share the same or similar spatial layout, spatial coding is not inherently suitable for recognition of real-world images. Romberg et al. [9] proposed a Bundle min-Hashing (BmH) method by aggregating multiple adjacent local features into bundles. However, it ignores the relative position between the local features in the bundle, which restricts the improvement of describing ability to some extent. Kalantidis et al. [6,8] proposed to group local features into triples and represent the triples as signatures, which captures both visual appearance and local geometry. However, the proposed geometric constraint in [6,8] is sensitive to rotation changes which are typical cases in real-world images. Similarly, Wan et al. [10] proposed a Tree-based Shape Descriptor (TSD) to encode