Logo Retrieval with Representation Error of Self-taught Encoding

Wei Liu[1,2, a], Yunxing Ruan[2, b], Xia Cai[2, c]

[1]*ComputerSchool of HuazhongUniversity of Science and Technology, Wuhan, 430079, China*

[2]*ComputerSchool of HuazhongNormalUniversity, Wuhan, 430074, China*

[a]*liuwei@mail.ccnu.edu.cn* ,[b]*ruanyx@mail.ccnu.edu.cn*, [c]*caixia@mail.ccnu.edu.cn;*

Abstract.

Logo retrieval in real-world scenarios has numerous potential applications in computer vision. Due to occlusion, illumination , non-rigid distortion and other reasons, the accuracy of feature matching in natural images is far lower than that in the print objects. For such a challenging task, a lot of papers have conducted very fruitful work. The algorithm finds approximate matching points in the images by locality sensitive hashing algorithm. Given matched points' position information, matched points are divided into several groups. With RANSAC algorithm, each group of matched points are divided into inlier points set and outlier points set, the candidate windows of logos can be mapped. Finally by calculating the representation error score of overlapping candidate windows, the lower score regions are eliminated, and the higher score regions are remained. The result of experiment shows that our approach can effectively locate more than one logo areas in an image, improving the recall of retrieval. And it also improves the mean Average Precision scores greatly by sorted files with representation error score.

Keywords: Logo Retrieval, Representation Error, Feature Point clustering

Introduction

Logos, as design patterns, have significant visual effect different from natural images. They are designed to promote brands, and represent the different commoditiesand services. Identification and location of logos in images is the basis of variety applications, including image understanding, intelligent navigation, target recognition, product positioning etc. At present, it is still a very challenging task, especially in natural images, such as retrieving the NIKE logo on our clothes. And due to occlusion, illumination, partial occlusion and non-rigid distortion, it is difficult to locate and identify a logo accurately.

There are several methods that focus on the logo retrieval, which can be divided into two categories: supervised learning method and unsupervised learning method. By using supervised learning algorithm,[1][2][3] extract image features from annotated database, construct model function, and then get the optimal parameters of the model, finally recognize logos and locate them. But the cost of manual annotated database is too expensive. What's more, it has no extendibility. New logos can't be found which are not in the training set. Therefore, many papers use unsupervised learning method. [4] extracts SIFT features of images, which are more accurate in searching nearest neighbor for SIFT feature is invariant in scale and rotation. Given feature points' coordinate, RANSAC algorithm [5]is used to locate the logos area. [6] divides all query feature points into positive set and negative set, and then all feature points on images are abeled positive and negative by a distance function using query feature set, finally logo areas are located by branch-and-bound searching method. [7] aims to search vehicle logo regions in the images, which are approximately above the license plates as a priori knowledge. Given logos coarse regions, saliency maps are generated, which are help to searching the exact regions with the maximum information theory.

However those algorithms of logo retrieval ignore the important fact that logo is an artificial design, which is full of edges corners, and is significant different to its background. So the representation error of the logo area must be much greater than that of the background area, if the encoder is learned from the background image mainly. The average representation error can be noted as criterion on which to sort candidate windows. The higher the score, the greater the difference between candidate window and background, and the more accurate the candidate window is.

Our algorithm has following advantages: (1) All candidate windows can be found. All the nearest neighbor feature points in an image are divided into several groups by using cluster analysis. Every group is mapped to a candidate window, which is help to improve the recall of query, especially there are more than one logo in an image. (2) The candidate window positioning inaccurate can be eliminated. All the candidate windows are sorted by average representation error score, which is accord with human cognition and for those regions overlap almost, our algorithm remains the highest score one.(3) This algorithm is more suitable for small logo retrieval. Because training sample must be extract from image during studying self-taught encoder, the probability of the small logo in image as sample must be small, so the average represent error score of small logo must be higher.

Locating candidate windows

Since SIFT feature is scale and rotate invariant, there are a lot of image feature matching algorithms using SIFT or SIFT like features to represent images. While in experiment, there are great differences for the same logo in different backgrounds. For example, the number of feature points in the areas of adidas logo in the flickrlogo dataset[8] varied from 13 to 10297, and there are only 10% images have more than 3 feature points matched each other. So the location of logo area in the rest images is a problem for having no enough matched points. On the other hand, there are too many SIFT feature points need to match each other in most images, which is too time-consuming.

Locality-sensitive hashing algorithm is a fast and efficient method .It can find k nearest neighbors for query features in high dimensions and in large datasets. But the k nearest neighbors are not the most exact nearest points, they include inliers and outliers. As fig.1, outliers can be eliminated by RANSAC algorithm, and the affine transformation matrix calculated is help to map the candidate window.

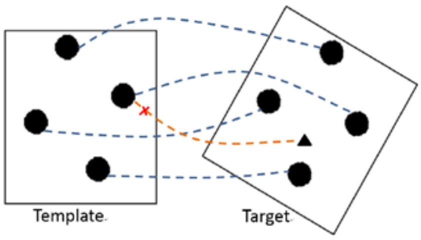

Figure1. Given matched feature point pairs, the template is mapped to the target using RANSAC algorithm. The point labeled as a black triangle is the outlier point, which can be eliminated by RANSAC algorithm.

Most of the time, there are more than one logo exist in an image. The feature points matched belonged to different logo areas must have different affine transformation matrixes accordingly. As fig. 2, one affine transformation matrix vs. several affine transformation matrixes. One affine transformation mapping all feature points to a logo area,which is labeled by dotted rectangle in fig. 2, is bound to lead missed or wrong detection. Dividing feature points to several groups and mapping them to different candidate windows will increase the recall of the retrieval, which is labeled by dashed rectangle in fig. 2.

Figure 2.The groundtruth of Cofidis logo is the green solid rectangle. The blue dashed rectangle is our result, and the red dotted rectangle is the result of one affine transformation matrixin [4]

Sorting the candidate windows

A self-taught encoder neural network [9] is an unsupervised learning algorithm that applies backpropagation, setting the target values to be equal to the inputs. The self-taught encode tries to learn a function , so as to output $\tilde{x}$ that is similar to the input x. In other words, it is trying to learn an approximation to the identity function, with constraints on the neural network which limiting the number of hidden units and limiting the number of hidden units actived. Subject to these constraints, the self-taught encoder worked as human mind, is sensitized to the edges of images, and keep high compression ratio.

Self-taught encoder learning algorithm samples patches from images themselves randomly. With no doubt, the representation error is greater if the representing object has great difference from the sample patches. Normally, the logo area in the images is smaller than the background, so the sample patches come from background mostly. And because the appearance of the logo is different significantly from background, the representation error of the logo area is greater than that of the background. As fig. 3, the patches which representation error greater than the threshold(max representation error’s 0.6 times) is plotted. All the plotted patches are centered on the logos area and edges area.

Figure 3. The points plotted in the image are the maximum representation error patches whose representation errors are greater than the threshold. Almost all the plotted points are at the edge of logo areas.

So, for almost overlapping candidate windows, the average representation error is the discriminatecriterion. The lower score regions are eliminated, the higher score regions are remained. For all candidate windows of one logo query, the result can be sorted by average representation error score inversely in order to achieve higher mean Average Precision score.

Framework

The framework of our logo retrieval algorithm includes following steps:

Step1: Find k matched point pairs for the query logo using E^2LSH [10] algorithm;

Step2: Cluster analysis for feature points matched. In an image, if there are more than 3 matched feature points, divide the feature points into several groups using cluster analysis by their positions. The number of the groups is less then s, s .

Step3: Geometric positioning. Given s groups of feature points pairs, s affine transformation matrixes are calculated by using RANSAC algorithm. Finally, the s candidate windows are mapped.

Step4: Select and sort candidate windows. For overlapping mostly candidate windows in an image, we remain the highest average representation error score windows and elimate the lower rest windows. And finally we sort all the candidate windows of a query logo.

Experiment and result

We test our algorithm on the challenging BelgaLogos dataset, which is composed of 10,000 images covering all aspects of life and current affairs. All images are resized with a maximum value of height and width equal to 800 pixels, preserving aspect ratio. A given image can contain one or several logos or no logo at all. The location of the logo have been manually annotated in the groundtruth file. There are 26 differents logos and 55 internal queries which compose the query pool Qset1. After feature extraction, the entire dataset results in a total of 24,172,440 SIFT points. And the query pool Qset1 is composed of 3308 SIFT points.

In experiment, number of nearest neighbor k is set to50, which is different to

[4] k=500. Thus, the recall is lower and precision is greater than those in [4]. The maximum of the number of groups s is set to 6. The expansion threshold is set to 20 for the fixed threshold method as [4]. There are 5 logos' result which are not included in the total score. These 5 logos are Eleclerc，Peugeot，Roche，StellaArtois，VRT, whose position in Qset1 are not according to the groundtruth file.

Table1mAP of All logos in Qset1

Logo Name	Baseline	Ours
Adidas	0.6301	**0.6303**
Adidas-text	0.6969	**0.8182**
Base	0.2411	**0.4725**
Bouigues	**0.6667**	0.5
Citroen	**0.4792**	0.3714
Citroen-text	0.2223	**0.3909**
CocaCola	0.231	**0.2794**
Cofidis	0.0329	**0.3184**
Dexia	0.2256	**0.3444**
Ecusson	0.52	0.3333
Ferrari	0.6443	0.483
Gucci	1	1
Kia	0.246	**0.5595**
Mercedes	0.4844	**0.5324**
Nike	1	1
President	0.4544	**1**
Puma	**1**	0.6235
Puma-text	**0.0773**	0.0499
Quick	0.4298	**0.5628**
SNCF	0.4537	**0.55**
TNT	0.3435	**0.4516**
All	0.48	**0.53**

Table 2 Recall of All logos in Qset1

Logo name	Baseline	Ours
Adidas	0.2109	**0.2449**
Adidas-text	**0.1587**	0.1429
Base	0.1543	**0.1605**
Bouigues	0.1429	0.1429
Citroen	0.0385	0.0385
Citroen-text	0.1269	**0.1421**
CocaCola	0.1636	**0.3636**
Cofidis	0.2000	**0.8333**
Dexia	0.2383	**0.3277**
Ecusson	0.0306	**0.0408**
Ferrari	**0.3506**	0.3247
Gucci	0.5000	0.5000
Kia	0.3404	**0.6454**
Mercedes	**0.1163**	0.1047
Nike	0.0085	0.0085
President	0.5714	**0.7143**
Puma	0.0446	**0.1146**
Puma-text	**0.2593**	0.2222
Quick	0.5088	**0.7544**
SNCF	0.4286	0.4286
TNT	0.4706	**0.6275**
All	0.2411	**0.3277**

Table 1 gives the mean Average Precision scores of the baseline and our algorithm. The result show that the mAP is consistently higher for most queries with up to 10% improvements.

Table 2 gives the recall of the two methods. Our method can deal with the scenarios that contain more than one logos in an image. Using cluster analysis, more candidate windows can be found which improve the recall of retrieval.

Some results of different logo retrieval have been shown in fig. 4.

Figure 4 some results of logo retrieval in BelgaLogos dataset

Conclusion

Logo retrieval in natural images is a challenging problem in computer vision. We test our algorithm on the challenging BelgaLogos dataset,the result show our method can effectively locate more than one logo in an image, and by using average representation error of self-taught encoder, our method is effective to improvethe recall and precision of retrieval superior to existing method.

Reference

[1] Kuo-Wei Li , Shu-Yuan Chen , Songzhi Su,Der-Jyh Duh , Hongbo Zhang , Shaozi Li. Logo detection with extendibility and discrimination.Multimedia Tools and Applications [J].2013.4.Volume 72, Issue 2.1285-1310

[2] Jerome Revaud,MatthijsDouze,CordeliaSchmid.Correlation-Based Burstiness

for Logo Retrieval. ACM Multimedia. 2012.965-968

[3]KlebanJ, XieX, MaWY. Spatial pyramid for logo detection in natural scenes. In:Proc.Int'l IEEE Multimedia & Expo(2008)

[4] Alexis Joly, Olivier Buisson. Logo Retrieval with aContrario Visual Query Expansion. ACM Multimedia.2009. 581-584

[5] Fischler, M.A. and Bolles, R.C.Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Communications of the ACM, 24(6)1981. 381–395,

[6] JingjingMeng , Junsong Yuan , Yuning Jiang , NityaNarasimhan , VenuVasudevan ,YingWu. Interactive Visual Object Search through Mutual Information Maximization. ACM Multimedia, 2010.10.25-29

[7] Songan Mao, Mao Ye, Xue Li, Feng Pang, JingleiZhou.Rapid vehicle logo region detection based on information theory. Computers & Electrical Engineering, Volume 39, Issue 3. April 2013.863-872

[8] Stefan Romberg, Lluis Garcia Pueyo, Rainer Lienhart, and Roelof van Zwol. Scalable logo recognition in real-world images.ICMR. 2011.

[9] RajatRaina, Alexis Battle, Honglak Lee, Benjamin Packer and Andrew Y. Ng. Self-taught learning: Transfer learning from unlabeled data. ICML.2007.

[10]M. Datar, N. Immorlica, P. Indyk, and V. Mirrokni. Locality-sensitive hashing scheme based on p-stable distributions. DIMACS Workshop on Streaming Data Analysis and Mining, 2003.