Finding Main Object in Images based on Google and SIFT Keypoints

Bodi Yuan

Abstract

Object recognition is a popular topic and it has been studied for a long time. And it is also useful in many fields. There are lots of methods for object recognition relying on image matching, machine learning, and pattern recognition. In this paper, the main task is to recognize and localize the main object in an arbitrary image with an obvious main object. And the key idea is to use google to search the original image for similar images and keep some most similar images. Comparing the original image with each of those images based on SIFT algorithm (including SIFT keypoints and SIFT descriptor), some keypoints on the main object will match more times than those not on the main object, since those similar images main object usually are similar to the original images. The final step is to keep those more times matching keypoints and to use them representing where the main object is in the original images. Experimental results show that this method can recognize the main object in most arbitrary images without image preprocessing or any learning algorithms.

I. INTRODUCTION

The recognition and localization of objects is an important task for computer vision. For image matching, there are several feature detectors and descriptor algorithm, such as SIFT (Scale Invariant Feature Transform) features [1], [2], SURF features (Speeded Up Robust Features) [3], Harris corner [4]. On the other hand, google has done a really awesome work on image searching. For instance, if you search an exact image, it can give you a bunch of similar images. There are lots of research on object recognition. [6] introduces a method combine Harris corner and SIFT descriptor. [5] gives an feature descriptor algorithm which has an simple descriptor with scale invariant feature. [7] introduces a method based on artificial neural networks. However, all these presented method can work under particular conditions or need to learn the specific object before recognizing. In this paper, in order to recognize an arbitrary object in an arbitrary image with an obvious main object, I combine image searching and image matching together. For an arbitrary input image, using google to search this image and can get lots of similar images. Comparing the original image with those similar images based on SIFT features, the main object will be recognized and localized.

II. PROPOSED METHOD

A. Search similar images

The method in this paper is to use Google to recognize the main object in images. At first, the image going to be recognized should be a good image, which means this image has an obvious main object in the image. Only this can guarantee Google find some closed results. Then use google to search this image and get some similar result. And the most similar results will be chose, such as top 100 results, as a database shown in Fig. 1.

Bodi Yuan: PhD candidate in Mechanical Engineering Department, University of California, Berkeley. Email addresses: bodiyuan@berkeley.edu
B. Match keypoints

The next step is to find lots of keypoints in both the original image and database image. In this paper, I use SIFT keypoints and SIFT descriptor, which is rotation invariant and scale invariant. Since there is a bunch of images similar to the original image, each image in database will have some keypoints matching with the original image. Then compare each the image in the database with the original image. The result of keypoints in original image is shown in Fig. 2.

C. Keep the most matching points

Since most images in the database have the similar main object with the original image, those keypoints on the main object have more probability to get matching and will spontaneously have more times matching with other points after the whole comparing process. If any point has more than a threshold times matching, then keep this point. Otherwise, delete this point. After this step, as shown in Fig. 3, there will be small amount keypoints left and most lie on the main object.
D. Filter

Sometimes some keypoints in the background will also occasionally have many matching points. Therefore, a filter algorithm should be implemented to improve the performance of this method. Since we just need to find the only one object in the image, we can assume that most points should be close to each other and those points which are far away from most points are probably noises. For each point, calculate the distance to the other points and add them together as the whole distance for this point. Then the point with the highest whole distance will be deleted. Repeat this process until the exact percent points left. (In this paper, I choose 80 percent as the threshold) Finally using those points, the main object can be recognized and it is easy to get to know where it is (Fig. 4).

III. Experiments

The proposed method has been tested through a series of experiments. All of them performed on an Intel CPU Core i7-3667U processor with 2 GHz and 4 GB RAM. The operating system adopted is Windows 8 and the programming language is Matlab. The results are shown below.
As we can see, in all of these experiments, filter is an important step and it deletes lots of noise points and ensures that most points lie on the main object. In fact, we can also enhance the threshold of matching times to improve the accuracy of this algorithm. The experiments results show that most the keypoints left after filter lie on the main object or are very close to the main object. Using these keypoints is easy to recognize the main object and also localize where it is.
Besides, suppose there is an arbitrary image with more than one main object. If we can divide this image to several parts and each part contains one main object, then we can use this paper’s method to find out the main object in each part. Thus we can recognize and localize each part in an arbitrary image. Here, I just use an simple example (Fig. 6) as test and simply segment the image by hand. And then I test the proposed algorithm for each part separately. Fig. 7, 8, 9 show the result. Using the results shown in Fig. 9, we can know that there are a dog and a cat in the original image and also know where the dog and the cat are.
IV. CONCLUSION

This paper has presented a method for automatically recognizing the main object in images, which does not need any preprocessing or particular environment. The method is easy to realize and can work well under general conditions. As for the future work, if we can combine this method with an effective image segmentation method, then we can automatically label each part in images and also know where each object is.

REFERENCES

Fig. 9: original images; keypoints in the original images; the most matching keypoints; results after filter