Abstract

Recent years have seen an exponential increase in the use of mobile devices. Since many of the mobile devices are equipped with a camera and are connected to the internet, localization in an urban environment using landmark images is gaining popularity. The idea is simple. A tourist takes images of a landmark where he or she is standing with a mobile camera which are then transmitted to a server where the image(s) are matched against a database of landmark images for that locality. If a match is found, relevant information such as background information on the landmark, nearby transit facilities or information on other important landmarks nearby is sent back. This type of application has tremendous potential as a mobile city guide or navigation aid. In this project, we investigate the use of local invariant shape features and global features such as colour and texture for the recognition task as evident from literature and present various retrieval techniques. A variety of descriptors for landmark recognition and scene classification are discussed. Insights into vocabulary building and weighting schemes for representing landmark images are provided that can help in boosting recognition rates.

Contents

List of Figures

List of Tables

1 Introduction

Automatic landmark recognition from images can enable better localization in an urban environment. Since most people nowadays carry mobile devices that are connected to the internet, mobile landmark identication is gaining popularity. The basic idea is a person captures some photos of the place where he or she is standing with a mobile device. These photos are then transmitted to a server over the internet where they are matched against a database of landmarks. If a match is found, background information about the landmark and other relevant information is sent back.

Figure 1: application of landmark recognition system

This type of application is immensely useful as a mobile city guide. Some of this information can even be overlaid on live camera frames thereby enabling augmented reality. For example, user focuses his camera on some important buildings near by and the names of the buildings pop up as annotations on the camera frame. This can be very useful for navigation in an urban setting as most people use landmarks as an important means of nding their way in a city.

The use of augmented reality based annotation on camera images in order to provide easy landmark-based navigation instructions from start to end point. A similar effort can be observed in, where the authors enable landmark-based navigation between two buildings in a university campus setting.

2 Literature review

we have used various resources for collecting the information about developing the machine learning model and gone through various research papers for identifying the best model that suits our requirements most of the research papers are from microsoft that are published at the image-net conference in 2015 that helped us a lot at coming to the coclusion of our model we have chosen. And also went through various blogs and web articles for finding a way through creating, cleaning and feature detection of the data.

2.1 Data collection

we have created our dataset both training dataset and testing dataset from kaggle which is a an online community of data scientists and machine learners, owned by Google. Kaggle allows users to find and publish data sets, explore and build models in a web-based data-science environment, work with other data scientists and machine learning engineers, and enter competitions to solve data science challenges. Kaggle got its start by offering machine learning competitions and now also offers a public data platform, a cloud-based workbench for data science, and short form AI education.

2.2 Algorithm Design

we have reffered various reasearch papers to compare the performance of the algorithms so as to choose the one with best performance there by came to conclusion to use ResNet(residual network).This ResNet architecture was more successful that traditional, hand-crafted feature learning on the ImageNet. Their DCNN, named AlexNet, contained 8 neural network layers, 5 convolutional and 3 fully-connected. This laid the foundational for the traditional CNN, a convolutional layer followed by an activation function followed by a max pooling operation, (sometimes the pooling operation is omitted to preserve the spatial resolution of the image).

Much of the success of Deep Neural Networks has been accredited to these additional layers. The intuition behind their function is that these layers progressively learn more complex features. The first layer learns edges, the second layer learns shapes, the third layer learns objects, the fourth layer learns eyes, and so on. Despite the popular meme shared in AI communities from the Inception movie stating that “We need to go Deeper”, He et al. [2] empirically show that there is a maximum threshold for depth with the traditional CNN model.

3 Conclusion

In this project we have observed that feature selection is an important step in the recognition pipeline that is often ignored or undervalued. Insights into vocabulary building and weighting schemes for representing landmark images are provided that can help in boosting recognition rates. Some effective approaches to landmark classication or nding images are explored which resulted in identifying a best algorithm that provides optimal results in terms of training and performance.

References

1. Relja Arandjelovic and Andrew Zisserman. Three things ev- eryone should know to improve object retrieval. In 2012 IEEE Conference on Computer Vision and Pattern Recognition, pages 2911–2918. IEEE, 2012.

2. Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In European conference on computer vision, pages 404–417. Springer, 2006.

3. Ondrej Chum, James Philbin, Josef Sivic, Michael Isard, and Andrew Zisserman. Total recall: Automatic query expansion with a generative feature model for object retrieval. In 2007 IEEE 11th International Conference on Computer Vision, pages 1–8. IEEE, 2007.

4. Gabriella Csurka, Christopher Dance, Lixin Fan, Jutta Willamowski, and Cedric Bray. Visual categorization with bags of keypoints. In Workshop on statistical learning in computer vision, ECCV, volume 1, pages 1–2. Prague, 2004.

5. Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. arXiv preprint arXiv:1801.07698, 2018.

6. Herve J egou and Ond rej Chum. Negative evidences and cooccurences in image retrieval: The benefit of pca and whitening. In European conference on computer vision, pages 774–787. Springer, 2012