Land Mark Recognition Through Image Classification

B. Deepthi Reddy, Assistant professor, Department of information technology, JB Institute of engineering and technology, Hyderabad, Telangana

G. Yogitha, C. Anirudh Jayant, D. Goutham, JB Institute of engineering and technology, Hyderabad, Telangana.

ABSTRACT - Land mark classification and detection is useful in many social sharing websites and search engines to extract relevant data on given input image. Recent research in object recognition has used such sites as a source of image data, but the test images have been selected and labeled by hand, yielding relatively small validation sets. In this paper we study image classification on a much larger dataset of 30 million images, including nearly 2 million of which have been labeled into one of 500 categories. The dataset and categories are formed automatically from geotagged photos from Flickr, by looking for peaks in the spatial geotag distribution corresponding to frequently-photographed landmarks. We learn models for these landmarks with cloud vision API.

KEYWORDS: Object detection, Vision API, Geotagged photos.

1. INTRODUCTION

The billions of photographs in Internet-scale photo collections offer both exciting opportunities and significant challenges for computer vision, and for the area of object recognition in particular. Achieving Internet-scale object recognition and image classification is currently limited by the relatively small-scale datasets for which ground truth information is available. For instance, the widely-used PASCAL VOC 2008 dataset has about 10,000 images and 20 categories, while the LabelMe dataset is of similar size, with a larger hierarchically-organized label set. Bigger datasets such as Tiny Images have millions of images but do not include category labels, whereas other datasets make use of visual features during image selection which may bias towards certain methods. Recent work on scaling classification algorithms to Internet-sized datasets with millions of images has thus been limited to evaluating classification performance on relatively small datasets such as LabelMe.In this paper we consider image classification on much larger datasets featuring millions of images and hundreds of categories. First we develop a collection of over 30 million photos with ground-truth category labels for nearly 2 million of those images. The ground-truth labeling is done automatically based on geolocation information that is separate from the image content and the text tags that we use for classification. The key observations underlying our approach is that photos taken very near one another are likely to be of similar things. Moreover, if many people have taken photos at a given location, there is a high likelihood that they are photographing some common area of interest, or what we call a landmark.Thus we use a mean shift procedure to find peaks in the spatial distribution of geotagged photos, and then use large peaks to define the category labels. The photographs taken at these landmarks are typically quite diverse so that the labeled test datasets are challenging, with significant amounts of visual variation and a large fraction of outliers. In most cases, a landmark does not consist of any one prominent object; for example, many of the landmarks are museums, in which the photos are distributed among hundreds of exhibits. Our landmark classification problem can thus be thought of as more similar to an object category recognition problem than to a specific object recognition problem. In this we discuss the details of our dataset collection approach and compare it to some alternative techniques.

We use multiclass support vector machines to learn models for various classification tasks on this labeled dataset of nearly two million images. We use visual features based on clustering local interest point descriptors into a visual vocabulary that is used to characterize the descriptors found in each image. We also explore using the textual tags that Flickr users assign to photos as additional features.

2.LITERATURE SURVEY

Image classification using bag-of-features models has been studied extensively, however such previous work has been carried out only at much smaller scales. The work we report here uses two orders of magnitude more labeled photos – nearly two million photos as opposed to a few thousand in previous work – and one to two orders of magnitude more categories – up to 500 compared to tens in most previous work. This larger scale allows us to study how performance is affected by the number of categories and the number of training images available.

Our investigation also evaluates text tags versus image features, and considers the use of temporal context which has not received much attention in the literature. Some recent work has used large datasets, but the number of labeled photos available for evaluating performance has usually been quite small. For instance uses one million photos but only 5,000 of them have ground truth labels.

The recent work of considers a dataset with tens of millions of images, but only at thumbnail resolutions and again without labels for assessing classification accuracy.

Another line of research uses small training sets to automatically label larger image sets however such approaches generally make use of image features and machine learning techniques, and thus the resulting datasets are not independent of the kinds of features and methods that one wants to test. This raises the possibility that methods related to the ones used to create the dataset might be at an unfair advantage.

We also investigate how the visual vocabulary size affects classification performance. Although presents a technique for finding the optimal visual vocabulary size for their task, it is not clear that their method can scale to large datasets because the running time is linear in the number of images and quadratic in the number of categories.

The paper of is related to our work in that it studies geolocating photographs, but their goal is quite different from ours, as we do not try to predict location but rather just use location to derive category labels. (For instance, in our problem formulation a misclassification with a geographically proximate category is just as bad as with one that is far away. Our experiments use a standard classification paradigm and thus are comparable with many other studies. Moreover, the test set in contains only 237 images that were partially selected by hand, making it difficult to generalize the results beyond that set. In contrast we use automatically-generated test sets that contain tens or hundreds of thousands of photos, providing highly reliable estimates of performance accuracy.

2.1 OBJECT DETECTION:

Object detection is a computer technology related to computer vision and image processing that deals with detecting instances of semantic objects of a certain class (such as humans, buildings, or cars) in digital images and videos. Well-researched domains of object detection include tracking objects, for example tracking a ball during a football match, tracking movement of a cricket bat, or tracking a person in a video.

2.2 GEOTAGGED PHOTOGRAPH:

A geotagged photograph is a photograph which is associated with a geographical location by geotagging. Usually this is done by assigning at least a latitude and longitude to the image, and optionally altitude, compass bearing and other fields may also be included. every part of a picture can be tied to a geographic location, every part of a picture can be tied to a geographic location, When geotagged photos are uploaded to online sharing communities such as Flickr, the photo can be placed onto a map to view the location the photo was taken. In this way, users can browse photos from a map, search for photos from a given area, and find related photos of the same place from other users.

3. ARCHITECTURE

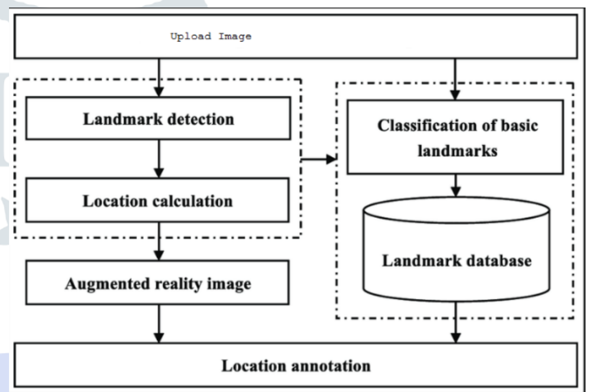

Fig 3.1 System Architecture

4.EXISTING SYSTEM:

Achieving Internet-scale object recognition and image classification is currently limited by the relatively small-scale datasets for which ground truth information is available. For instance, the widely-used PASCAL VOC 2008 dataset has about 10,000 images and 20 categories, while the LabelMe dataset is of similar size, with a larger hierarchically-organized label set. Bigger datasets such as Tiny Images have millions of images but do not include category labels, whereas other datasets make use of visual features during image selection which may bias towards certain method. Recent work on scaling classification algorithms to Internet-sized datasets with millions of images has thus been limited to evaluating classification performance on relatively small datasets such as LabelMe.

4.1 DISADVANTAGES:

- Most of the existing systems used to work on type of data inside image but no based on landmark classification.

5. PROPOSED SYSTEM:

In proposed system when input image is given we are extracting information from given image and training data in the way as information location is extracted from image and features are compared using API and accurate location information and land mark information is displayed.

5.1 ADVANTAGES:

- Automatic extracting of location and land mark form image which helps in searching related images and related data.
- Results are accurate for most of the images.

6. CODE & OUTPUT RESULTS

App.py:

```
import argparse
import io
import os
from google.cloud import vision
from google.cloud.vision import types
from google.oauth2 import service_account

credential_path = "C:/Users/dell/Desktop/vision/Visionsentiment-
47b8abe50af1.json"
os.environ['GOOGLE_APPLICATION_CREDENTIALS'] = credential_path

def main(image_file):
    # Instantiates a client
    client = vision.ImageAnnotatorClient()
    # Loads the image into memory
    with io.open(image_file, 'rb') as image_file:
        content = image_file.read()
    image = types.Image(content=content)
    # Performs label detection on the image file
    response = client.logo_detection(image=image)
    annotations = response.logo_annotations
    for annotation in annotations:
        print(annotation.description)
if __name__ == '__main__':
    parser = argparse.ArgumentParser()
    parser.add_argument('image_file', help='The image you\'d like
to label.')
    args = parser.parse_args()
    main(args.image_file)
```

Fig 6.1: Coding

Landmark	Random tags	Image
Effiel tower	Effiel City Travel night	
London eye	stone cross london	
Big Ben	westminster London ben	

7. CONCLUSION

We have presented a means of creating large labeled image datasets from geotagged image collections, and experimented with a set of over 30 million images of which nearly 2 million are labeled. Our experiments demonstrate that multiclass SVM classifiers using SIFT-based bag-of-word features achieve quite good classification rates for largescale problems, with accuracy that in some cases is comparable to that of humans on the same task. We also show that using a structured SVM to classify the stream of photos taken by a photographer, rather than classifying individual photos, yields dramatic improvement in the classification rate. Such temporal context is just one kind of potential contextual information provided by photo sharing sites. When these image-based classification results are combined with text features from tagging, the accuracy can be hundreds of times the random guessing baseline. Together these results demonstrate the power of large labeled datasets and the potential for classification of Internet-scale image collections.

8. REFERENCES

[1] B. Collins, J. Deng, K. Li, and L. Fei-Fei. Towards scalable dataset construction: An active learning approach. In ECCV,2008.

[2] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. PAMI, 2002.

[3] K. Crammer and Y. Singer. On the algorithmic implementation of multiclass kernel-based vector machines. JMLR,2001.

[4] D. Crandall, L. Backstrom, D. Huttenlocher, and J. Kleinberg.Mapping the world's photos. In WWW, 2009.

[5] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In ECCVWorkshop on Statistical Learning in Computer Vision, 2004.