Vision Based Outdoor Localization of IIITD Campus

Srinidhi Hegde
2013164
IIIT-Delhi
srinidhi13164@iiitd.ac.in

Tarun Malhotra
2013112
IIIT-Delhi
tarun13112@iiitd.ac.in

Neeraj Vaishnao
2013112
IIIT-Delhi
neeraj13142@iiitd.ac.in

1. Motivation

Localization problem is one of the most prominent problems of computer vision. Interests in solving these problems developed over the years as it had wide range of applications such as controlling the motion of a robot based on the visual inputs and navigating driverless vehicles. It is heavily used in oil pipeline inspection, ocean surveying and underwater navigation, mine exploration, coral reef inspection, military applications, crime scene investigation. The idea was popularised by the competition Where am I ? by ICCV, Computer Vision Contest.

The challenges that make this problem more interesting are identification of similar entities in the image, identifying their geometric relations and querying or searching the images in a large dataset.

2. Previous Related Work

Many significant amount of work is accomplished in the field of localization. Most of the works use Where am I? ICCV Computer Vision Contests dataset to test out their application. These works are compared using the metrics used by the aforementioned contest. In [1], the idea of localization is applied on a global scale using a probabilistic models. They propose an algorithm to estimate a distribution over geographic locations from a single image using a purely data-driven scene matching approach.

The localization has been used quite frequently in urban landscape. Vision based techniques are applied in [3] to enhance the accuracy of the GPS coordinate and orientation that is estimated using magnetic sensors. Further this technique helped in improving the accuracy of projecting virtual models in the Augmented Reality application.

The problem of localization has been approached in a similar manner as ours in [4] whose detailed summary is mentioned below.

2.1. Image Based Localization in Urban Environments

The problem of image based localization comprises of three phases, location recognition, camera motion estimation (between the query and the closest reference views), and position triangulation. For localization one major requirement is getting accurate correspondences between the query and the reference views, and this is a big challenge. To obtain reliable correspondences a modified wide-baseline matching scheme is proposed. Initial matching is followed by a robust motion estimation technique capable with dealing with large number of outliers.

2.1.1 Location Recognition

This stage finds the closest views from the model database given a query view. The work uses the SIFT features proposed by D. Lowe [2]. Image D(x, y,) is obtained by taking a difference of two neighboring images in the scale space and the keypoints are detected by searching for peaks in this image. Each detected keypoint has an associated descriptor, which characterizes the gradient distribution of the local image area around it. After extracting keypoints from a query image, its descriptors are matched to those of the database views. To check if a pair of keypoints is a match, two criterions are listed. According to the first, a match is considered if the distance ratio between the closest match and second closest one is below some threshold t1. This criterion is based on the assumption that correct discriminative keypoints often have the closest neighbor significantly closer than the closest incorrect match. But in the case of buildings, due to the presence of many repetitive structures (e.g. windows), the above criterion will reject many possible matches. Hence the second criterion considers two keypoints as matched, when the cosine of the angle between their descriptors f and g is above some threshold t2. The reference views with the largest number of matches with the query view will be selected. Top 5 reference views are retained.

2.1.2 Motion Estimation

This stage computes camera motion between the query view and the reference views. The camera motion between the query view and the matched reference view is represented as g = (R, T), where R SO(3) is the rotation and T = [tx, ty, tz], T is the translation. The corresponding points obey the epipolar constraint and are related by so called essential matrix. Given the correspondences obtained in the feature matching stage, the essential matrix can be estimated using a standard eight point algorithm, which can be decomposed to 4 motion. The authors propose a novel algorithm which can successfully classify outliers and inliers with only a fraction of the computational cost of the standard RANSAC.

2.1.3 Motion Estimation

Buildings are dominant in urban scenes, and hence it is likely that corresponding points are located in planar building facades. Therefore the general motion model captured by fundamental matrix is often not appropriate and the homography model is favored. Another reason to favour homography model is that in a scene with repetitive structures, the process of inlier identification for fundamental matrix becomes more difficult.

2.1.4 Final Localization

Once the motion estimation results are produced, the top two images, that are neighbours, are considered for correct identification of the query image. For improving the accuracy from correspondence from other images, second and third nearest neighbour of the first nearest neighbour to query image are used as the two reference images. Using triangulation on all three images, that are two reference image and one query image, location of query view is obtained. If the motion between the two reference images cannot be estimated accurately then the location of these two images are interpolated.

The method developed in [4] was tested on the dataset of Where am I? ICCV Computer Vision Contest. The metric that was employed to determine the effectiveness of the technique were the localization error, that is, the difference in the original GPS coordinates and the calculated GPS coordinates. Maximum localization error that was found on employing this technique was of 16m. Furthermore they have tested the total execution time for all the views of dataset in 24 minutes.

3. Problem Statement

To estimate the GPS location (outdoors) of the user from image using matching techniques, such as homography, from image database and to interface the data through a mobile application.

4. Approach

We have divided the entire task into set of four sub-tasks which includes- data collection, database maintenance, algorithm development and analysis for image matching and mobile application development.

4.1. Data Collection

We will be using the inbuilt cameras of the mobile phones and this will be interfaced by our mobile application. There are many tools to obtain geo-tagged images on Android, we will be using one of the tools like GeoTag, GeoTag Photos Pro2, GeoCamera etc.

4.2. Database Maintenance

For now we are planning to use MySQL database to store the image dataset. This database will be linked to the mobile application to retrieve images on the go for the user reducing the response time for processing separately.

4.3. Image Matching Algorithm

As of now we will be developing the algorithm through the course of the project with some changes that are required as per our application. We will majorly be focusing on three components of algorithm design and analysis. Firstly, we will develop an efficient technique of image retrieval from our image dataset using classifiers to prune off irrelevant images. Secondly, we will detect key features such as feature corner points and edges of the building structure. Finally, we will be using homography based techniques along with triangulation to determine the GPS coordinates of the user as well as the orientation of the camera.

4.4. Mobile Application Development

Mobile application will be developed for Android platform (with minimum support version as JellyBeans).

5. Experimental Setup

We will be requiring a sufficient number of images (about 50-100) of every building in the campus. These images of dataset should be clicked from viewing angles that cover almost all the degrees of freedom. These images should have GPS data tagged into them for the sake of reverse lookup during the experiment. The GPS coordinate will represent where the user was when he/she was taking that particular image.

Some of the metrics that will be used to test the accuracy of our method of estimating GPS coordinates will be, firstly, localization error, that is, the difference in the original GPS coordinates and the calculated GPS coordinates.

For estimating the efficiency of our method we will be calculating the total execution time required for localization from set of image dataset containing a threshold number of images. If time permits we will be testing our application on the dataset and metrics provided by ICCV Computer Vision Contest.

For testing the orientation provided by our application we will project the orientation data on Google Maps with a pointer that points the direction in which the camera is oriented.

sectionDivision of Labour Srinidhi: Feature point extraction, Homography, Mobile Application Development Tarun: Classification Algorithms Neeraj: Data Collection, Database Maintenance

References

[1] J. Hays and A. A. Efros. Im2gps: estimating geographic information from a single image. In *2008 ieee conference on computer vision and pattern recognition*, pages 1–8. IEEE, 2008.

[2] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[3] G. Reitmayr and T. Drummond. Going out: robust model-based tracking for outdoor augmented reality. In *ISMAR*, volume 6, pages 109–118, 2006.

[4] W. Zhang and J. Kosecka. Image based localization in urban environments. In *3DPVT*, volume 6, pages 33–40. Citeseer, 2006.