International Journal of Interdisciplinary Innovative Research &Development (IJIIRD)
ISSN: 2456-236X
Vol. **06** Special Issue **01** | **2021**

A Review Paper on Image Localization Using Google Maps

Ankita Punde[1] , Sampada Rohankar[2], Komal Ghonge[3], Akash Sonone [4] , Vaibhav Wagh[5], Prasad Kulkarni[6], Apeksha Ghonge[7]

[1,2,3,4,5,6,7] *BE Student, Electrical Engineering, Padmashri Dr. VBKCOE, Malkapur, Maharashtra, India*

ABSTRACT

In this paper we propose an approach for image localization that combines visual odometer with map information from Open Street Maps to provide robust and accurate estimates for the image position. The main contribution of this work comes from the incorporation of the map data as an additional cue into the observation model of a Monte Carlo Localization framework. The resulting approach is able to compensate for the drift that visual odometer accumulates over time, significantly improving localization quality. As our results indicate, the proposed approach outperforms current state-of the- art visual odometer approaches, indicating in parallel the potential that map data can bring to the global localization task. The implemented scoring algorithm can efficiently give the matching scores between a query image and all possible database images. Upon searching a new approximate orthogonal image, a set of scaling and rotations are first selected, and the visual words are transformed and matched against the database. The best locations along with scales and rotations are determined from the query results of the different set of transformed visual words. Experiments show a high success rate and high speed in searching map databases for aerial images from different datasets.

Keywords*: Orthogonal satellite imagery, Image localization, Image feature, SIFT, Map database indexing.*

……

1. INTRODUCTION

With the rising popularity of smart phones and the increasing use of social media like Face book and Twitter, more and more pictures are available on the internet. In some cases when an incident has occurred photos are placed online before the emergency services are noticed.[3] These photos can provide valuable information about the incident, such as the location. The location in this case can be an exact GPS coordinate, but also an estimation of the location within a given range. In addition to law enforcement, automatic location detection can also apply to the following services: automatic adding geographical information to a photo, tracking people or extending the photo plug-in on social networks with the automatic tagging of photos.

The accuracy of the location may very per application. To add geo information to a photo, an accurate location is needed. Unfortunately, in most cases geographical information about the picture is not available. Even if an image contains geodetic, it is still not completely reliable, as there is no way to verify the location data of the used device was up to date at the time the picture was taken. Occasionally, it is possible to manually determine the location by the recognition of known buildings and landmarks.[5] However, in most cases this is not possible and also not completely reliable. A solution based purely on the actual image data would not super from problems like this. Image recognition however, is notoriously midcult and computationally expensive.

One of the main contributions of this paper is to propose a new feature indexing for geo-located features in a map and uses the extended visual words on map to index 2D location grids. Unlike the general image retrieval problem, geographical size and geographical rotation of features in map database can be recovered. Visual words with sizes and rotations differentiate features at different scales and different orientations, which leads to a more efficient indexing and retrieval system.

2. RELATED WORK

A great deal of research exists in the area of computer vision and image recognition. As computing power becomes more readily available, it becomes possible to process larger datasets. One _led where very large datasets are used is image recognition. The technology used for general image matching can also be used for deferent image-based solutions to real-life problems. [2] The possibilities for image based localization are explored. Since our research is supervised by TNO, the technology described in this paper has been a major nuance on our research. Technologies suggested in the report include using a feature database to look up septic element from the query image, and using a geometric matching algorithm on a subset of the database to pick the most accurate.

2.1 Invariant Image Feature

In recent years, there has been an extensive investigation in local invariant image features to achieve robustness to viewpoint changes. Lowe's Scale Invariant Feature Transform (SIFT) extracts distinctive scale and rotation invariant features from the DOG (difference of Gaussian) scale space of images, and describes the corresponding normalized image patches with SIFT descriptors, which are 128D vectors constructed from the local gradient histograms (Lowe, 2004). Figure 2 shows an example image with SIFT features. While SIFT detector handles only 2D similarity transformation,[6] some other feature detectors (e.g. MSER) go beyond to achieve invariance to affine changes. A good overview and evaluation of such affine invariant features can be found in (Mikolajczyk et al., 2005). SIFT descriptor

Fig -1: Illustration of the image localization problem.

2.2 Image Search with Visual Words

We will give a short outline of the visual word based image to introduce the terminology, which is used in this paper. The image search proceeds by a local approach, by detecting local features, computing a description (feature) vector and matching with a database of feature vectors (see illustration in Figure 4). Each local detection is described by a SIFT feature vector. Each SIFT feature vector is then quantized with the so-called vocabulary tree. It assigns a single integer value, denoted visual word (VW) to a SIFT feature vector. This eases matching a lot. Instead of computing distances between SIFT feature vectors only integer values have to be compared. Each image is then represented as a set of visual words. The set is denoted as a document vector which is a *v*-dimensional vector where *v* is the number of possible visual words.

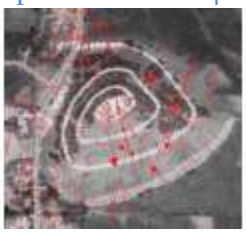

Fig- 2: SIFT features shown as arrows

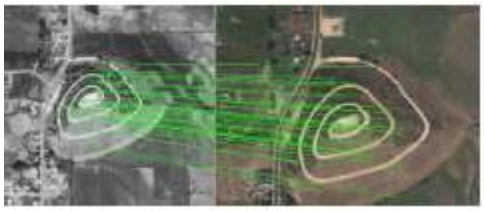

Fig -3: SIFT matches

3. METHODS

This section describes the complete chain used in our concept for localization based on image matching. The rest step is to collect images for building the ground-truth database from Google Street view. Subsequently, descriptors are extracted from those images. For this the Open CV library [4] is used. Next, the extracted descriptors are used to build a data structure that is searchable quick and ancient. From this data structure the actual matching and localization can be performed. How is an acceptable match denned? The requirements for a successful match can be deferent from application to application.[7] Depending on the application, an acceptable result can be identifying the correct city. In other applications, a match can be a street, a location within 20 meters or even the exact camera position and viewing angle. For some applications, the algorithm needs to return one location that has to be correct, whereas for other purposes having the correct result in the rest 10 suggestions is accurate enough. When implementing a system, these requirements need to be denned, in order for the system to be able to produce the desired results.[8] In the experimental concept, three deferent match methods were implemented, that have deferent uses, performance and accuracy.

3.1 Ground-truth database with geo-tagged images

Fig- 4: Created cutouts from a Google Street view panorama

As can be seen in Figure 4, much of the detail in the panorama is lost, because only two cutouts are made. In most cases that is not a problem, because 8.699 panoramas are made every 100 meter5. Detail that is lost in a panorama is mostly still covered in another panorama. But there are cases where unique objects are complete discarded and not stored in the image database, which possibly leads to a decrease in matching accuracy.

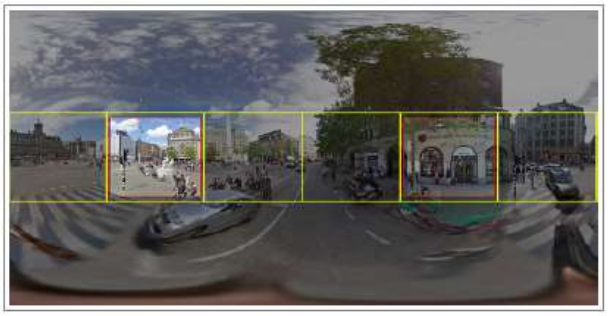

Fig -5: Created cutouts from a Google Street view panorama

As shown in the above panorama Figure 5, the loss of unique objects can be overcome by creating cutouts from more and larger view angles. The objects drawn in red are the cutouts created in the current setup. The objects drawn in yellow are the cutouts that need to be made to cover all the unique elements in the panorama.

Fig -6: Visual crawler result for "the Dam" in Amsterdam

Although not all objects are covered by the two cutouts that are currently made, like the "Royal Palace on the Dam" (most left in figure 5) they can still be in the image database, due the fact that multiple panoramas are created in that area, as shown in figure 6. Each point on the map represents the location where a panorama is made[5].The circle drawn in red, is the location of the panorama in figure 5, which does not cover the "Royal Palace on the Dam" and the "National Monument the Dam", but they are still covered by the cutouts made from the panoramas drawn in the blue rectangle.

3.2 Matching

With the descriptor tree built as described earlier, it is possible to perform lookups from the tree. When performing a lookup, descriptors are extracted from the query image. Then each descriptor is looked up in the tree. From the root, the distance from the descriptor to the centers of the children of the root is calculated, and the child with the center closest to the query descriptor is selected.

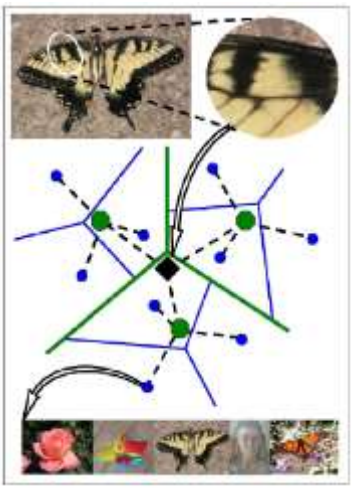

Fig -7: Schematic visualization of the descriptor tree.

A small part from an image ends up in a certain leaf in the tree. This leaf contains pointers to di_erent source images of the similar descriptors. [9]

3.3 Geometrical matching

The previously discussed ways of matching are based on finding similar descriptors that occur in the images. Geometrical matching not only looks at the occurrence of similar descriptors, but also looks at their relative positions. This is done by creating a homograph between the matching descriptors, using Open CV.

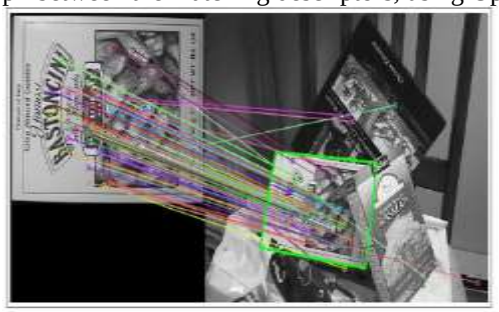

Fig -8: Image showing a projective transformation from one image to another.

4. DISCUSSION

The paper proposed a new visual word for indexing orthogonal satellite imagery and the associated method for image-based localization. The proposed visual word incorporates the geographic information of image features, and gives stronger for indexing images.

A scoring implementation is designed to match a query image to a part of a large image (represented as multiple tiles), which is significantly different to standard image retrieval. Scale and rotations are recovered together with location by matching the proposed visual words. In this research an attempt was made to design and build a system to perform image localization based on looking up an image from a large database of descriptors.

As a real world example of a useful implementation of such a system could be a situation where the police and an image of an accident on Twitter, and no information about the location is present. On average 1436 descriptors are detected in each cutout. SIFT has no limit on the number of discovered descriptors in an image. Also descriptors detected in non unique elements, like the sky, trees, water and streets are currently saved.[6] An algorithm is needed which can determine which SIFT descriptors in an image are important and the ones less important.

5. CONCLUSION

In this paper we addressed the problem of finding the exact GPS location of images. We leveraged a large-scale structured image dataset covering the whole 360◦ view captured automatically from Google Maps Street View. We proposed a method for relocating single images, specifically examining how the accuracy of current localization methods degenerates when applied to large-scale problems. First, we indexed the SIFT descriptors of the reference images in a tree; said tree is later queried by the SIFT descriptors of a query image in order to find each individual query descriptor's nearest neighbour. We proposed a dynamic pruning 14 method which employed GPS locations to remove unreliable query descriptors if many similar reference descriptors exist in disparate areas. Finally, a novel approach - using the proximity information of images - was proposed in order to localize groups of images. First, each image in the image group was localized individually, followed by the localization of the rest of the images in the group within the neighbourhood of the found location. Later, the location of each image within the rough area (Limited Subset) with the highest value was selected as the exact location of each image.

REFERENCES

[1] Fraundorfer, F., Stew´enius, H. and Nist´er, D., 2007. A binning scheme for fast hard drive based image search. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition.
[2] James Hays and Alexei A. Efros. Im2gps: estimating geographic information from a single image. In Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2008
[3] Lowe, D., 2008. image features from scale-invariant keypoints. International Journal of Computer Vision , pp.
[4] T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T. and Van Gool, L., 2009. A comparison of affine region detectors. International Journal of Computer Vision 45(1-2), pp. 33–62.
[5] Stew´enius, H., 2010. Scalable recognition with a vocabulary tree. In: Proc. IEEE, New York City, New York, Vol. 2.
[6] Genc, Y., 2010. Gpubased video feature tracking and matching..
[7] Carl Doersch, Saurabh Singh, Abhinav Gupta, Josef Sivic, and Alexei A. Efros. ACM Transactions on Graphics (SIGGRAPH), 31(4), 2012.
[8] Jan Salmen, Sebastian Houben, and Marc Schlipsing. Google street view images support the development of vision-based driver assistance systems. pages 891{895, 2012.
[9] M. A. Wong. A K-means clustering algorithm. Applied Statistics, In Proceedings of the International Conference on Computer Vision Volume 2, USA, 2013. IEEE Computer Society
[10] Lowe, D.G.: Distinctive image features from scale-invariant keypoints. IJCV 60 (2013)
[11] Vedaldi, A., Fulkerson, B.: VLFeat: An open and portable library of computer vision algorithms. http://www.vlfeat.org/ (2014)
[12] Muja, M., Lowe, D.G.: Fast approximate nearest neighbors with automatic algorithm configuration. In: VISAPP. (2014)
[13] Balanda, K.P., MacGillivray, H.L.: Kurtosis: A critical review. The American Statistician 42 (2014) 111–119