Towards Global Localization Using Global Descriptors

Charbel Azzi   University of Waterloo, ON, Canada
Daniel Asmar   American University of Beirut, Beirut, Lebanon
Adel Fakih   University of Waterloo, ON, Canada
John Zelek   University of Waterloo, ON, Canada

Abstract

Image-Based Localization (IBL) is the problem of estimating the 3D pose of a camera with respect to a 3D representation of the scene. IBL, despite being a trivial problem for small scenes, becomes quite challenging as the size of the scene grows. Aside from the computational burden, matching against a very large number of 3D keypoints spanning a wide variety of viewpoints, illumination, and areas is a very unreliable process that results in a large number of outliers and ambiguous situations. In recent years, a number of approaches have attempted to address the problem using paradigms such as bag-of-words, features co-occurrence and others, with varying degrees of success. This paper explores the use of global descriptors, in particular GIST, to tackle this problem. We present a system that relies on a similarity measure derived from GIST to qualify a limited number of 3D points for the matching process, hence reducing the problem to its small size counterpart. Our results on a standard dataset show that our system can achieve better localization accuracy than the state of the art at a fraction of the computational cost, which can used towards global localization.

1 Introduction

We are interested in the problem of given an initial 3D map of an environment, where are we at a particular time? This may be a kidnapped robot problem where we are dropped into the environment or it may be the problem of keeping track of the position given that we know where we were in the previous instant. Accurate localization is essential if we are to determine if there are dynamic obstacles or changes to our map. We are assuming that the original map was created with a Structure-from-Motion procedure (i.e., Visual SFM) and that we have the original images, poses from which images were captured as well as the keypoints features and their correspondences in neighboring images.

The tree-based approach [6], also known as the Fast Approximate Nearest Neighbors (FLANN), is the standard solution for the search space reduction in IBL. This approach aims to speed up the search in 2D-3D by finding the approximate nearest neighbors with respect to all the 3D points in the map. When dealing with environments that have thousands or hundreds of thousands of 3D points, matching against all those points becomes computationally expensive. Thus, the need to reduce the search space of the tree-based approach becomes the main focus. One of the notable solutions for reducing the search space is the use of Bag-of-Words (BoW) [4]. This approach is well known to trade accuracy for speed due to the clustering effect, where the best correspondences might be missed especially in IBL [7]. On the other hand, Kendall et.al [5] used deep convolutional neural networks to solve the problem. The accuracy of this approach is enough for location recognition applications, but is not enough to compete with the accuracy of the main IBL systems. In this paper, we present a novel solution that achieves better accuracy then the state-of-the-art, while maintaining considerable speedups. Recently, Azzi et.al [1, 2, 3] started using global descriptors to reduce the search space in IBL which returns the best accuracy in IBL. In this paper, we present further progress using global descriptors that can be used towards global localization.

2 Approach

Figure 1 shows an overview of the proposed system, where the GIST distance between the query and all the keyframes is computed. If the distance is below a certain threshold than the keyframe is considered a candidate match. In order to remove outlier keyframes each candidate keyframe is checked and the candidate one qualifies for localization. The 3D points of those candidates will be matched to the SIFT features of the image before removing the outliers via RANSAC. Only images with enough inliers will qualify to the pose estimation step.
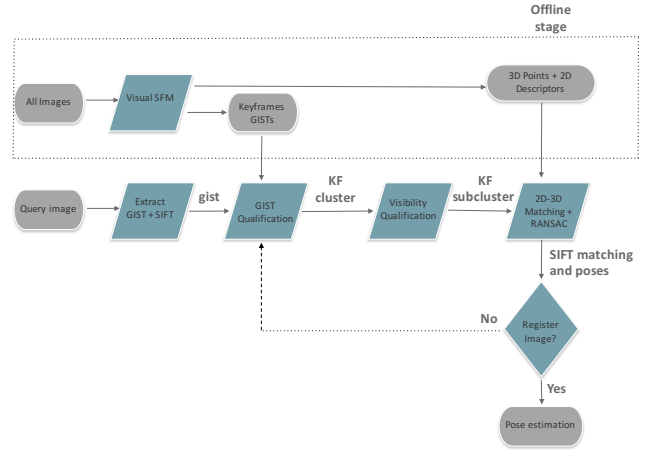

Fig. 1: GSSR system overview.

3 Results

Table 1: The candidate KFs (inside the left and right boxes) for the formed clusters for the Kings College scene from Figure 2(b) around the query image 300 (top image).
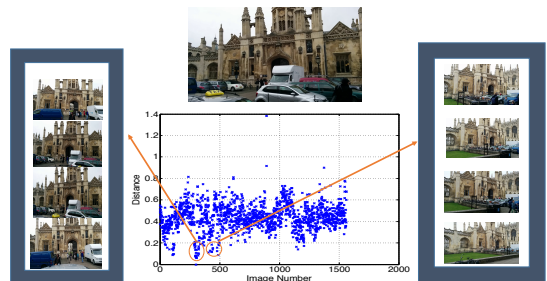

Table 2: GSSR benchmarked against tree-based [6], PoseNet [5] and ACG Localizer [7].

Dataset	# Train Images	# Query Images	Median Error				Average Time (s)			
			Tree-Based	PoseNet	ACG Localizer	GSSR	Tree-Based	PoseNet	ACG Localizer	GSSR
Kings College	1220	343	0.229m, 0.194deg	1.992m, 3.261deg	0.910m, 0.761deg	0.213m, 0.188deg	1.8	0.01	0.74	0.102
StMarys Church	1487	530	0.180m, 0.424deg	2.645m, 5.102deg	0.662m, 0.634deg	0.175m, 0.309deg	4.852	0.01	0.51	0.105
Old Hospital	895	182	0.341m, 0.272deg	2.441m, 2.923deg	1.044m, 0.846deg	0.299m, 0.228deg	1.179	0.01	0.33	0.065
Shop Façade	231	103	0.138m, 0.220deg	1.490m, 4.299deg	0.548m, 0.575deg	0.140m, 0.217deg	0.302	0.01	0.39	0.044
Street	3015	2923	0.410m, 0.672deg	3.910m, 3.75deg	0.972m, 1.004deg	0.364m, 0.539deg	11.43	0.01	0.77	0.109
Average			0.260m, 0.358deg	2.496m, 3.867deg	0.872m, 0.772deg	0.238m, 0.296deg	3.91	0.01	0.548	0.085

Experimental results on a standard dataset validates the advantages of GSSR. Table 2 shows that GSSR scores the best localization accuracy among all the approaches on the Cambridge 5 Scenes dataset. GSSR was able to accurately localize a query image in less than 0.1 sec which makes it the fastest feature-based IBL system. These results show the powerful aspect of using global descriptors to reduce the search space in large city scale environments. This proves that IBL can be one of the solutions for global localization such as relocalization, loop closure and place recognition.

References

[1] Charbel Azzi, Daniel Asmar, Adel Fakih, and John Zelek. Filtering 3d keypoints using gist for accurate image-based localiza-

tion. In *Proceedings of the British Machine Vision Conference (BMVC)*. BMVA Press, September 2016.

[2] Charbel Azzi, John Zelek, Daniel Asmar, and Adel Fakih. Image-based localization using context. *Vision Letters*, 1(1), 2015.

[3] Charbel Azzi, John Zelek, Daniel Asmar, and Adel Fakih. Global descriptors reduce the image-based localization search space. *RSS Workshop on Visual Place Recognition*, 2016.

[4] Mark Cummins and Paul Newman. Fab-map: Probabilistic localization and mapping in the space of appearance. *The International Journal of Robotics Research*, 27(6):647–665, 2008.

[5] Alex Kendall, Matthew Grimes, and Roberto Cipolla. Posenet: A convolutional network for real-time 6-dof camera relocalization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2938–2946, 2015.

[6] Marius Muja and David G Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. *VISAPP (1)*, 2:331–340, 2009.

[7] Torsten Sattler, Bastian Leibe, and Leif Kobbelt. Improving image-based localization by active correspondence search. In *Computer Vision–ECCV 2012*, pages 752–765. Springer, 2012.