Filtering 3D Keypoints Using GIST For Accurate Image-Based Localization

Charbel Azzi[1]
cazzi@uwaterloo.ca
Daniel Asmar[2]
da20@aub.edu.lb
Adel Fakih[1]
afakih@uwaterloo.ca
John Zelek[1]
jzelek@uwaterloo

[1] Systems Design Engineering
University of Waterloo
Waterloo, Canada

[2] Vision and Robotics Lab
Mechanical Engineering Department
American University of Beirut
Beirut, Lebanon

Image-Based localization (IBL) addresses the problem of estimating the 6 DoF camera pose in an environment, given a query image and a representation of the scene. The tree-based approach [2] is the standard solution for IBL. When dealing with large-scale environments, the need to reduce the search space of the tree-based becomes the main focus. Sattler et al. [3] reduced the search space by clustering the 3D points into bag-of-words. This approach is well known to trade accuracy for speed due to the quantization effect. Recently, Kendall et.al [1] 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 propose the Gist-based Search Space Reduction (GSSR) system to reduce the search space by finding candidates keyframes in the database, then match against the 3D points that are only seen from these candidates.

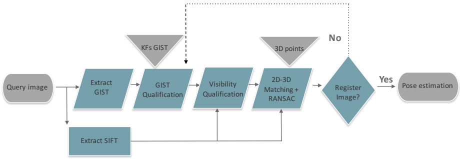

Figure 1: GSSR system overview.

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 using Eq. 1:

$$F_k = \frac{\sum_{i=1}^{N} P_i(KF_i, KF_k)}{N}, \tag{1}$$

where N is the total number of candidate keyframes and P_i is the number of 3D points in common between the tested candidate KF_k and the keyframe KF_i at i. If the ratio F_k is high enough the candidate keyframe 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.

Table 1: GSSR benchmarked against tree-based [2], PoseNet [1] and ACG Localizer [3].

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 major standard datasets validates the advantages of GSSR. Table 1 shows that GSSR scores the best localization accuracy among all the approaches on the Cambridge 5 Scenes dataset. Note that GSSR has 10 times better accuracy than PoseNet and significantly better than ACG Localizer which is one of the best feature-based IBL systems. GSSR was able to accurately localize a query image in less than 0.1 sec which makes it the fastest feature-based IBL system for large-scale scenes.

[1] 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.

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

[3] 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.