論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Large-Scale Visual Localization Based on Image Retrieval and Local 3D Reconstruction
著者(和文)	田平創
Author(English)	Hajime Taira
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第11421号, 授与年月日:2020年3月26日, 学位の種別:課程博士, 審査員:奥富 正敏,蜂屋 弘之,倉林 大輔,大山 真司,塚越 秀行
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第11421号, Conferred date:2020/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

系・コース： Department of, Graduate major in	システム制御 システム制御	系 コース	申請学位（専攻分野）： Academic Degree Requested	博士 Doctor of	（ 工学 ）
学生氏名： Student's Name	田平創		指導教員（主）： Academic Supervisor(main)	奥富正敏	
			指導教員（副）： Academic Supervisor(sub)		

要旨（英文 800 語程度）
Thesis Summary（approx.800 English Words ）

Finding 6-Degree-of-Freedom (6DoF) pose of a camera (visual localization) is a key ability of robotic intelligent system and autonomous navigation. Because of the wide spreading of image data and internet sharing, the target scene for the localization problem becomes larger in both indoor and outdoor scenarios. Previous approaches that require a 3D database capturing the whole scene to estimate the camera pose are impractical or sometimes infeasible for such large-scale visual localization problems.

In this thesis, we introduce a visual localization pipeline designed for large-scale environments. Compared to other approaches using the large 3D database, we first seek the relevant location of an input image using an image database, which allows scaling to the larger scene. We next compute an accurate 6DoF camera pose using local geometric information around the relevant location. Contributions of this thesis can be divided into three parts; 1) We propose two feature matching algorithms for the general purpose of wide-baseline image matching but can also be beneficial for visual localization. 2) We construct a large-scale image dataset in an urban city-scale scene while preparing precise 6DoF camera poses of input images to evaluate the performance of visual localization. Using this new dataset, we evaluate previous approaches for visual localization and show the large-scale 3D model is not necessary for accurate camera pose estimation. 3) We also evaluate our pipeline on a large-scale indoor scenario. To deal with several difficulties raised by indoor scenes, we propose a new algorithm verifying the estimated camera pose using a pre-captured local 3D point cloud. Our method achieves state-of-the-art performance on our indoor dataset. Altogether, we tackled for scaling visual localization problem to much larger environments. We confirmed our approach effectively addresses the issue based on efficient image retrieval and local 3D reconstruction. Also, two new datasets we constructed through the thesis depict several challenging situations for visual localization induced in large-scale scenarios and will be beneficial for further researches in this area.

This thesis is organized as follows.

In Chapter 2, we propose two feature matching algorithms that can potentially contribute to visual localization. Sec. 2.2 introduce a feature matching for spherical panoramic images which capture the scene efficiently despite of large image distortion. Our method synthesizes multiple undistorted images for each panoramic image and perform feature matching on them to mitigate the side effect of image distortion. We show that our method provides accurate and rich correspondences between panoramic images and can contribute to build a rich 3D model via SfM pipeline. We also propose a feature matching designed for image pairs taken from very different viewpoints (Section 2.3). We again generate multiple synthetic images that approximate potential appearance changes on the image induced by view changes, to collect the variations of keypoints and feature vectors.

Chap. 3 gives a detailed description of our contributions for outdoor city-scale visual localization. We first create a new dataset for visual localization in an urban situation that includes a set of accurate camera poses for evaluation (Sec. 3.2). We propose a localization pipeline that computes 6DoF camera pose of a query using only 2D image database (Sec. 3.4), which constructs our main contribution in this thesis. Sec. 3.5 provides various evaluation of our method and others on our new dataset, and shows the benefits of our approach on large-scale outdoor scenarios.

Chap. 4 describes our contributions for large-scale indoor visual localization, on which the rich 3D point clouds and additional information are available. Because of the lack of existing dataset for indoor scenario, we again create a dataset composed of RGBD database images and user photos taken by smartphone, along with its accurate camera poses (Sec. 4.2). In Sec. 4.3, We propose a localization pipeline exploiting rich 3D database to estimate and verify the camera pose for the input image. Sec. 4.4 additionally provides details of our investigations for pose verification step. We validate our method effectively scales to a large-scale indoor environment and achieves state-of-the-art performance (Sec. 4.5).

Finally, we add concluding remarks and potential future works of this area in Chap. 5.

備考 ： 論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。
Note : Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).
注意：論文要旨は、東工大リサーチリポジトリ(T2R2)にてインターネット公表されますので、公表可能な範囲の内容で作成してください。
Attention: Thesis Summary will be published on Tokyo Tech Research Repository Website (T2R2).