

Online Sparse Scene Coordinates Learning for Real-Time Camera Relocalization

Nam-Duong Duong, Amine Kacete, Catherine Soladie, Pierre-Yves Richard, Jérôme Royan

▶ **To cite this version:**

Nam-Duong Duong, Amine Kacete, Catherine Soladie, Pierre-Yves Richard, Jérôme Royan. Online Sparse Scene Coordinates Learning for Real-Time Camera Relocalization. 6th International Conference on 3D Vision (3DV), Sep 2018, Verona, Italy. hal-02048750

HAL Id: hal-02048750
https://hal.science/hal-02048750

Submitted on 25 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Online Sparse Scene Coordinates Learning for Real-Time Camera Relocalization

Nam-Duong Duong, Amine Kacete, Catherine Soladie, Pierre-Yves Richard, Jérôme Royan
Institute of Research and Technology b-com
Nam-Duong.Duong@b-com.com

Abstract

Camera relocalization refers to the problematics of defining the camera pose in known scenes. It is needed in several applications such as augmented reality or robot navigation. However, current camera relocalization systems require an off-line step, that uses the labeled images to construct a scene model. This step takes time, so that is difficult to apply in the augmented reality applications. Thus, we develop an online learning for real-time camera relocalization system. We present a hybrid method combining machine learning approach and geometric approach for accurate and real-time camera relocalization from each single RGB image independently. For machine learning part, we propose a sparse scene coordinates regression forest, which is able to be fast learned during images capturing. Our system presents how easily user can experience on scenes in desktop-scale.

1. Introduction

In computer vision, the main solution of camera pose estimation for augmented reality systems is known as Simultaneously Localization And Mapping (SLAM). SLAM is a method of camera localization. It faces tracking failures in the case of fast camera motion or sudden change of viewpoint, such as in a hand-held camera. Camera relocalization is then essential.

That is why nowadays, some research works focus on camera relocalization solution from RGB images. We can split the solutions into 3 groups: geometric approaches, machine learning approaches and hybrid approaches. Geometric approaches first construct a scene model of a set of key-points attached to feature vectors by performing Structure from Motion (SfM) on a set of images of the scene. Then camera relocalization is performed from each frame independently by matching feature vector of each frame to key-frames (frame-to-frame approach) or the scene model (frame-to-model approach). To accelerate feature matching, [3] uses a visual vocabulary for efficient 2D-to-3D matching. Even so, the matching time depends on the size of the model. Machine learning approaches for camera relocalization have appeared to challenge these constraints. Deep learning based methods [2] can predict camera pose from each whole RGB image in real-time. However, limitation of these methods lies in their accuracy. Last year, [1] presented hybrid method combining deep learning approach and geometric approach for camera relocalization using only RGB images with high accuracy. However, they took more time to optimize camera pose from thousands of correspondences and consequently cannot address augmented reality applications.

In general, current works have two main limits. Firstly, learning phase requires more time to create a scene model by SfM or machine learning. Secondly, testing phase is still challenging to have a both real-time and accurate method. Our demonstration overcomes these two limits. We propose an online learning solution, that uses scene coordinates regression forest for accurate and real-time camera relocalization in augmented reality system.

2. Methodology

We present a hybrid method combining machine learning approach and geometric approach. For the former, we propose a sparse feature learning using regression forest to predict 3D world coordinates corresponding to 2D keypoint detection. That aims to define 2D-3D point correspondences. The geometric part then estimates camera pose from these correspondences.

In the training phase, our sparse feature regression forest can be learned immediately during images capturing. Our method belongs to supervised learning approaches. So that, we use a RGB-D camera and a marker (be attached in the scene) to fast estimate camera pose and to label 3D scene coordinates for each pixel. From each captured RGB-D frame with estimated camera pose from the marker, we extract a set of SURF (Speed Up Robust Features) features. Each feature is labeled by projecting 2D image coordinates in the world coordinates system. Once our system reaches sufficient amount of feature, we create a new thread to train a tree of our regression forest in the same time of capturing image. Our training step finishes when all trees are

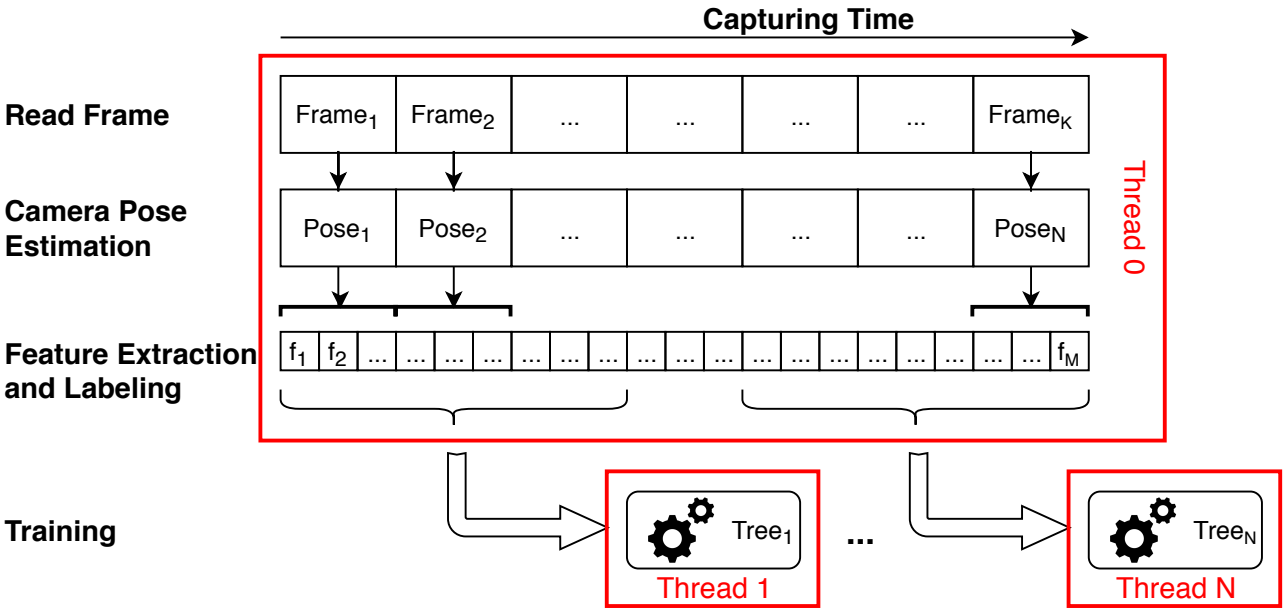


Figure 1. Our on-line training system. Our sparse feature regression forest is learned by multi-threads in the same time of capturing data.

learned. Capturing and training phase takes about 2 minutes for a desktop-scale scene with the configuration of our sparse feature regression forest including: extraction up to 500 SURF features for each image; 4 trees in the forest; 16 depth of each tree. Figure 1 shows the online learning of our system.

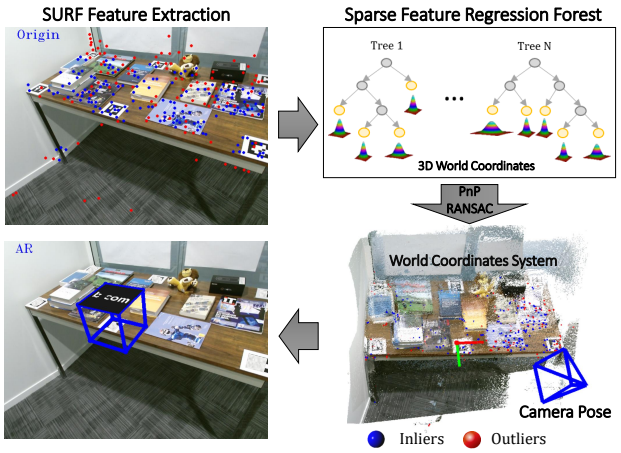


Figure 2. Our real-time camera relocalization in the augmented reality.

In the testing phase, we use only RGB camera. All features extracted from RGB image pass through our regression forest to obtain a set of 2D-3D correspondences. We then use Perspective-n-Point (PnP) and RANdom SAmple Consensus (RANSAC) algorithms to calculate camera pose. Our method can perform in real-time at $50ms$ per frame. The testing time of our demonstration is shown in Figure 2.

3. Conclusion

In this demonstration, we presented our hybrid method combining machine learning and geometric approach, where machine learning part is online learned during capturing data. Our method shows both accurate and real-time result for camera relocalization from only RGB images.

4. Videos

We made a demonstrative video [1] to show the training and testing performance of our method for camera relocalization. The video [2] presents the ability to help user experience easily thanks to our solution. Finally, the video [3] shows the robustness of our method to challenges of generalization, occlusion and illumination changes.

References

[1] E. Brachmann, A. Krull, S. Nowozin, J. Shotton, F. Michel, S. Gumhold, and C. Rother. Dsac - differentiable ransac for camera localization. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 1

[2] A. Kendall and R. Cipolla. Geometric loss functions for camera pose regression with deep learning. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 1

[3] T. Sattler, B. Leibe, and L. Kobbelt. Efficient & effective prioritized matching for large-scale image-based localization. *IEEE transactions on pattern analysis and machine intelligence*, 39(9):1744–1756, 2017. 1

[1] https://youtu.be/mBcpzHEkQLc
[2] https://youtu.be/fsyGV1Ca5Uo
[3] https://youtu.be/2oOL-BSemmQ