Towards Robust Vision Based SLAM System in Endoscopy with Learning Based Descriptor

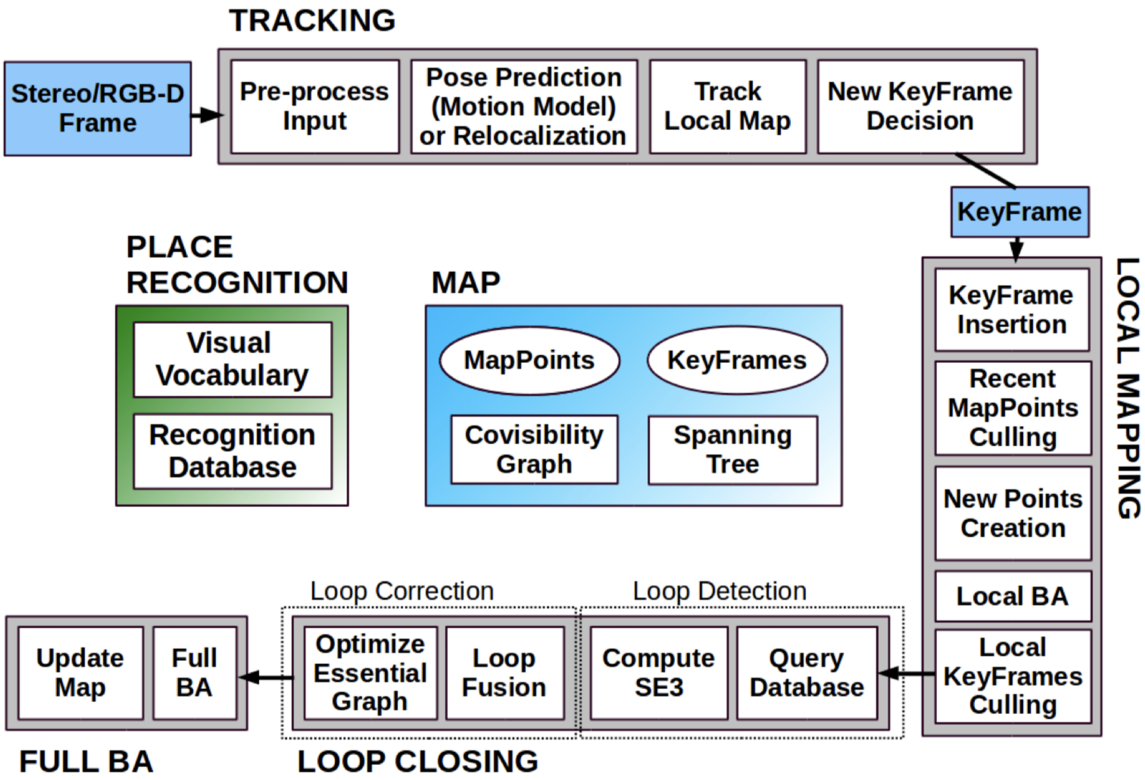
Basic Info (->Summary)

- Student Name: Yiping Zheng
- Mentor: Xingtong Liu
- Size group: 1
- Skills: C++/Pytorch/Python/ORB-SLAM Architecture

Background, Specific Aims, and Significance

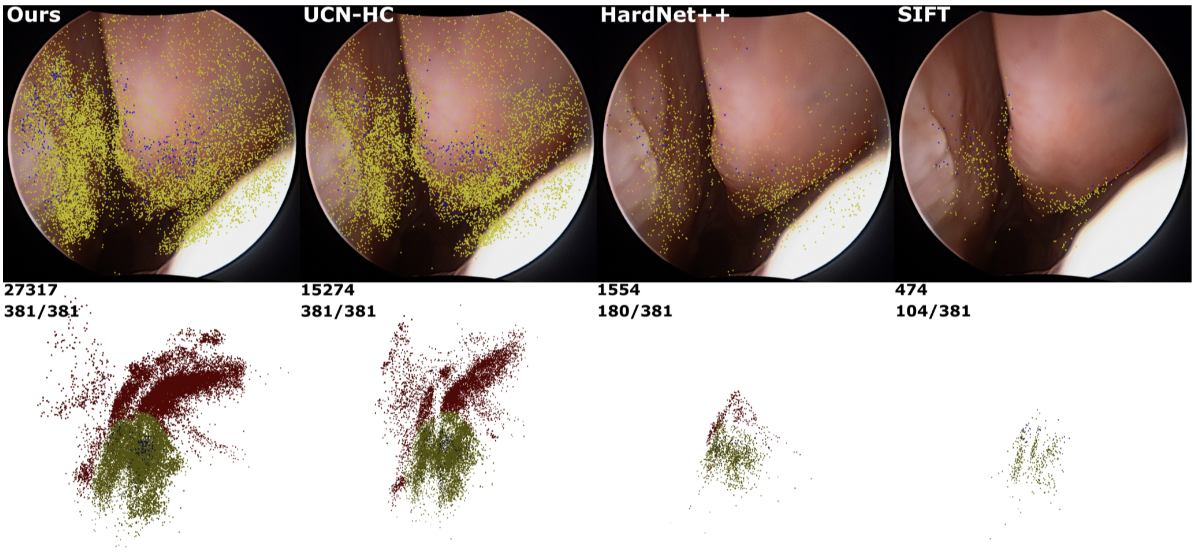
Surgical navigation is widely used in clinical application to provide accuracy and dexterity for surgeons to handle tools. There are two kinds of methods of surgical navigation, which are marker based and vision based method. The marker based method is the traditional one. For example, in sinus surgery, we put marker on the endoscope and on patient's head. CT registration, external tracking information. The defect of marker based navigation method is that it's an indirect registration, which result in a bigger error, over 2mm. However, in sinus surgery, there are many small tissue structures which are less than 1mm. As you can see, the traditional marker-based navigation system may not be accurate enough for surgeons to avoid critical structure on patient's body like artery, eye balls, brain, neural etc.

Vision-based method is developed in order to achieve higher accuracy in surgical navigation, by using the direct video information to recover the relative pose of endoscope with regard to patient's body. A common intermediate step in this procedure is called SLAM (Simultaneously Localization and Mapping), which is to compute the relative camera pose and a sparse point cloud representing the structure of the surrounding environment from a video sequence. It first detects feature points from every video frame and tries to find a transformation to match feature points between two continuous frames together, from which the camera pose and 3D point cloud can be computed.



In conclusion, SLAM system is the key element of achieving high accuracy in (vision-based) surgical navigation. And since feature extraction is at the beginning of the SLAM pipeline, how well this procedure can be done has a significant influence on all other procedures and final results. Therefore, finding robust feature descriptor, from which feature points are extracted, is a key to the SLAM system. Many feature descriptors have been proposed such as SIFT, SURF, BRIEF, ORB, etc. All the traditional feature desciptors uses merely local texture information, which can be good enough in most indoor or outdoor navigation scenes. However, in surgical navigation scene, we are often dealing with images of tissue surfaces which are smooth and don't contain very much local texture information, causing the traditional feature descriptors to perform not well enough. Feature points extracted can be very sparse and repetitive, which makes the following feature matching procedure not very robust and causing the whole system to be error prone.

With the rise of deep learning, there's a chance to design a new kind of feature descriptor, we call it dense learning descriptor, by training a nerual network. It combines the global anatomical information with the local texture information and it computes a feature description for every pixel of the image. we can get a more robust feature mathing performance in the endoscope scene and this can result in a better overall performance of surgical navigation.



However,
intermediate step: video cam pose, sparse 3D anatomical structure.
Online camera traj. 3D structure —> SLAM motivation
feature descriptor, local texture information, endoscopic video, sparse, repetitive.
SLAM system
feature matching not very robust,
learning based to global information(anatomical) local texture information
robust better feature matching module performance.

Background: SLAM in endoscope scene has two major defects, tissue deformation, and smooth and repetitive textures. Learning descriptor has the potential to tackle the latter defects.

Specific Aims: Using learning descriptor to overcome the defects of repetitive and smooth textures in SLAM tasks in endoscopic scenario.

Significance: Improve the robustness and accuracy of SLAM task in endoscope scenario.

Technical Approach

Integrate the learning descriptor with the SOTA SLAM architecture.

** motivation
ORB - (local) descriptor - keypoint detector - potential point of interest - adjacent region
dense descriptor

结构上 - how to generate

repeatablity
sparse key point

CNN

**feature matching
new frame

candidate keypoint location 3D mapping location
new descriptor map matching

** local descriptor 用在哪些模块，pair-wise feature matching, loop close, relocalization, replace with global descriptor.

Milestones and Status

Milestone 1: Getting started
Get the test dataset and set up the software environment (ORB-SLAM + Linux + Pytorch)
Planned Date: Feb. 6 ~ Feb. 10 (4 days)
Status: 90%

Milestone 2: Implement the algorithm
Planned Date: Feb. 11 ~ Feb. 29 (20 days):
Status: 0%

Milestone 3: Tune parameters and Evaluate the performance
Planned Date: Mar. 1 ~ Mar. 20 (20 days)
Status: 0%

Milestone 4: Test on various datasets
Planned Date: Mar. 21 ~ Mar. 31 (10 days)
Status: 0%

Milestone 5: Summarize result and writing report
Planned Date: Apr. 1 ~ Apr. 15 (15 days)
Status: 0%

Milestone 6: Prepare for the poster session.
Planned Date: Apr. 15 ~ ? (15 days) (3~4 days are enough)
Status: 0%

Deliverables

- **MInimum**: An implementation of a modern SLAM system with learning descriptor integrated. Expected by Mar.1.
- **Expected**: Fine tuning the parameter of the system to achieve better performance than state of the art SLAM system working in endoscope scene. Expected by Apr. 1
- **Maximum**: Assessment of improvement of system performance over SOTA systems and write a paper over it. Expected by May 1.

Dependencies

Obtaining datasets - contact Xingtong
Obtaining NN of the learning descriptor - contact Xingtong
setup the ORB SLAM running environement

Project Bibliography

- Project Reading list ( )
  Sinus SfM
  1. S. Leonard, A. Sinha, A. Reiter, M. Ishii, G. L. Gallia, R. H. Taylor, et al. Evaluation and stability analysis of video-based navigation system for functional endoscopic sinus surgery on in vivo clinical data. 37(10):2185–2195, Oct. 2018
     Stomach SfM
  2. A. R. Widya, Y. Monno, K. Imahori, M. Okutomi, S. Suzuki, T. Gotoda, and K. Miki. 3D reconstruction of whole stomach from endoscope video using structure-from-motion. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pages 3900– 3904, 2019
     Abdomen SLAM
  3. O. G. Grasa, E. Bernal, S. Casado, I. Gil, and J. Montiel. Visual slam for handheld monocular endoscope. IEEE transactions on medical imaging, 33(1):135–146, 2013.
  4. N. Mahmoud, I. Cirauqui, A. Hostettler, C. Doignon, L. Soler, J. Marescaux, and J. M. M. Montiel. Orbslam-based endoscope tracking and 3d reconstruction. In CARE@MICCAI, 2016.
     Oral Cavity SLAM
  5. L. Qiu and H. Ren. Endoscope navigation and 3D reconstruction of oral cavity by visual slam with mitigated data scarcity. In Proceedings of the IEEE Conference on Computer Vi- sion and Pattern Recognition Workshops, pages 2197–2204, 2018.
- Page references ([1][2])

Reports and Presentations

Log

???

Other Resources and Project Files

???