M2 Master student internship

Discovering new objects for visual localization

Marie-Odile BERGER, Gilles SIMON
TANGRAM team, Centre Inria de l'université de Lorraine (Nancy)

1 General information

Duration	5/6 months
Localization	Nancy (france)
advisors	Marie-Odile BERGER, Gilles SIMON
Address	LORIA/INRIA, Campus Scientifique - BP 239, 54506 Vandœuvre-lès-Nancy
Email	`marie-odile.berger@inria.fr` `gilles.simon@loria.fr`

Website of the TANGRAM team : `https://team.inria.fr/tangram`

2 Motivations

Recent work by the TANGRAM team has demonstrated the value of using object detection to aid visual relocation in a SLAM (Simultaneous Localisation And Mapping) context. The method, called OA-SLAM for object-assisted SLAM, is described in [1] and a video showing the system in use is available at `https://www.youtube.com/watch?v=L1HEL4kLJ3g`. However, this work is limited by the fact that the detectable objects must be known to the detector, i.e. the vocabulary used (defined by the object categories) is closed. The aim of this project is to extend the method to discover new objects in a variety of environments. For example, a factory may have many objects (valves, sensors, etc.) that are not recognised by the usual networks such as YOLO, which prevents the current localisation method from working.

The aim of this internship is therefore to investigate both recent advances in image segmentation (Segment Anything – SA [2]) and methods for extending a vocabulary based on proposals for boxes likely to contain objects [3, 4]. Compared to existing work, the proposed method will be able to take advantage of (i) the possibility of reconstructing objects or specific points on objects via SLAM to reduce the number of tracking hypotheses, (ii) the fact that the OA-SLAM method is based on an approximate modelling of objects by ellipsoids and therefore does not require perfectly delimited segmentations.

3 General objectives

Conventional segmentation approaches frequently suffer from over-segmentation, which can greatly complicate the tracking of regions and the discovery of promising objects. We will therefore first study the results of SA [2] (completeness of segmented parts, repeatability between close images, dots) on video sequences in a variety of environments. Based on the observations, we will then develop a method for tracking the segmented parts using different criteria such as those presented in [3, 4]. Integration with SLAM will enable us to benefit from additional criteria, since the approximate representation of objects should make it possible to define criteria that are more effective than the usual criteria based on a fine delimitation of objects.

Finally, we will study to what extent the labelling obtained automatically using the proposed method allows new learning guided by the model [5].

4 Working context

The trainee will join the TANGRAM team common to LORIA and Inria Nancy Grand Est. He/she will benefit from the team's research environment and expertise in image processing and analysis. A body of work and code on object SLAM is also available in the team and will serve as a basis for the development of the internship work.

5 Profile/ How to apply

Profile: Master's degree in computer science or applied mathematics. Even preliminary skills in computer vision and machine learning will be appreciated.

How to apply: Please send an email to marie-odile.berger.fr and gilles.simon@loria.fr. Please attach the following documents to your application: Curriculum vitae, transcripts of MSc and BSc university records.

References

[1] M. Zins, G. Simon, M.-O. Berger. OA-SLAM: Leveraging Objects for Camera Relocalization in Visual SLAM. *ISMAR 2022 - 21st IEEE International Symposium on Mixed and Augmented Reality*, Singapore, 2022.

[2] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollar, R. Girshick. Segment Anything. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 4015-4026.

[3] A. Osep, W. Mehner, P. Voigtlaender, B. Leibe. Track, Then Decide: Category-Agnostic Vision-Based Multi-Object Tracking *IEEE International Conference on Robotics and Automation (ICRA)*, 2018.

[4] Y. Du, Y. Xiao, V. Lepetit. Learning to Better Segment Objects from Unseen Classes with Unlabeled Videos. *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 3355-3364.

[5] M. Zins, G. Simon, M.-O. Berger. Object-Based Visual Camera Pose Estimation From Ellipsoidal Model and 3D-Aware Ellipse Prediction. *International Journal of Computer Vision*, 2022, 130, pp. 1107-1126.