# NA 568 Team 2 Final Project Proposal

Kun Sun, Ning Xu, Parth Chopra, Poorani Ravindhiran, Tribhi Kathuria

## I. Problem and Motivation

Visual Simultaneous Localization and Mapping [vSLAM] methods[1] use rich sensor data from cameras to perform localization and mapping. This is used for applications in robotics, augmented reality and navigation. For long-term autonomy, *loop closure* is an important and challenging part of the vSLAM pipeline, and a subject of ongoing research. Loop closure aims to associate new observations with previously visited places on a map, so as to decrease accumulating uncertainty in prior pose estimates, and obtain a better global map estimate. This project aims to present a comparative study of two visual place recognition techniques for loop closure, namely Bag-of-Words[2], and unsupervised Deep Learning [3]. We build upon the ORB-SLAM2 framework[4] for the same.

## II. Approach

To compare different techniques for place recognition, we use the ORB-SLAM2 framework, which uses visual sensor data to compute a trajectory and sparse 3D reconstruction. The ORB-SLAM2 framework is a generally robust SLAM pipeline commonly used in real-time deployments. It is easy to re-use with the open-source multi-threaded implementation, and ROS packaging available. In addition, it can be used for multiple camera configurations such as monocular, stereo and RGB-D.

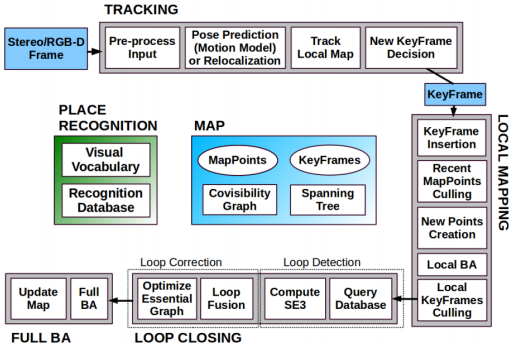


Fig. 1: ORB-SLAM2 Framework

The first technique we evaluate is Bag-of-Words (BoW), which is a feature extraction method and can be used to find matches between keyframes by building frequency histograms.

The second technique we evaluate is Lightweight Unsupervised Deep loop closure. To overcome issues in ConvNet-based supervised learning and unsupervised learning implementations, this technique combines -

- An autoencoder architecture to learn features in an unsupervised manner, and
- Traditional vision elements like Histogram of Oriented Gradients (HOG) features for dimensionality reduction, making the process lightweight; and for illumination/geometry invariance as well as randomized projective transformations for viewpoint invariance.

## III. Feasibility Evaluation

We judge the feasibility of the project according to the problem complexity, time and available resources.

- In the original complex structure of ORB-SLAM2, our team will work on modules like the visual preprocessing unit, loop detection, and loop closure proposal module as shown in Fig. 1.
- Our team will build on available resources including the online code repository for ORB-SLAM2, our place recognition approaches and off-the-shelf API's for image processing such as Sklearn/OpenCV. For the back-end optimization, GTSAM will be used.

## IV. Expected Results

We aim to evaluate Bag-of-Words[2] and unsupervised Deep Learning [3] as place recognition techniques for performing loop closure in the ORB-SLAM2 framework. With testing on the TUM RGB-D indoor dataset [5], and the NCLT Vision & LiDAR dataset [6], we will obtain benchmarks for

- Accuracy of loop closure proposals
- Root Mean-Square Error of pose estimates
- Computational cost

In the above process, our team aims to gain exposure to working with the existing codebase of ORB-SLAM2, and create API's to change loop closure methods in the framework in Figure 1.

## References

[1] Takafumi Taketomi, Hideaki Uchiyama, and Sei Ikeda. Visual slam algorithms: a survey from 2010 to 2016. *IPSJ Transactions on Computer Vision and Applications*, 9(1):16, Jun 2017.

[2] N. Kejriwal, S. Kumar, and T. Shibata. High performance loop closure detection using bag of word pairs. *Robotics and Autonomous Systems*, 77:55–65, 2015.

[3] Nathaniel Merrill and Guoquan Huang. Lightweight unsupervised deep loop closure. In *Proceedings of Robotics: Science and Systems*, Pittsburgh, Pennsylvania, June 2018.

[4] Raúl Mur-Artal and Juan D. Tardós. ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras. *IEEE Transactions on Robotics*, 33(5):1255–1262, 2017.

[5] Jürgen Sturm. Tum rgb-d dataset. Available at https://vision.in.tum.de/data/datasets/rgbd-dataset.

[6] Nicholas Carlevaris-Bianco, Arash K. Ushani, and Ryan M. Eustice. University of Michigan North Campus long-term vision and lidar dataset. *International Journal of Robotics Research*, 35(9):1023–1035, 2015.