# Project Proposal

## 'Stereo Reconstruction'

## 1 Abstract


The aim of our project is to reconstruct a 3D mesh from a pair of stereo images with different feature detection and matching methods and provide an overview in terms of accuracy and run-time performance. First of all, we will check if the images we had found from the datasets listed below are rectified or not. If they are not, we will rectify them using the provided camera parameters and extracted features. Then, we will extract keypoint features (like SIFT [6], ORB[9], SURF[1], etc.) of the stereo images. The next step consists of matching the features. The most simple way of doing that is brute-force matching. However, we also want to leverage the fact that we have a stereo setup, so matching keypoints can actually be searched along the epipolar line. Another variant of feature extraction and matching is to compute the optical flow and especially if we want to have a 3D mesh, then we plan on using dense optical flow.

If time permits, we are also planning to use a pre-trained neural network to extract features from these images and compare to traditional keypoint descriptors. After extracting these features, we will compute the disparity map between these images for every feature extraction method that we had used. [5] [13] Afterwards, using the camera parameters, we will convert the disparity map to the depth map. Finally, using the depth map, camera intrinsics and the RGB images, we will write a 3D mesh by translating the values from pixel space to world space. Below you can find our planned processing pipeline.


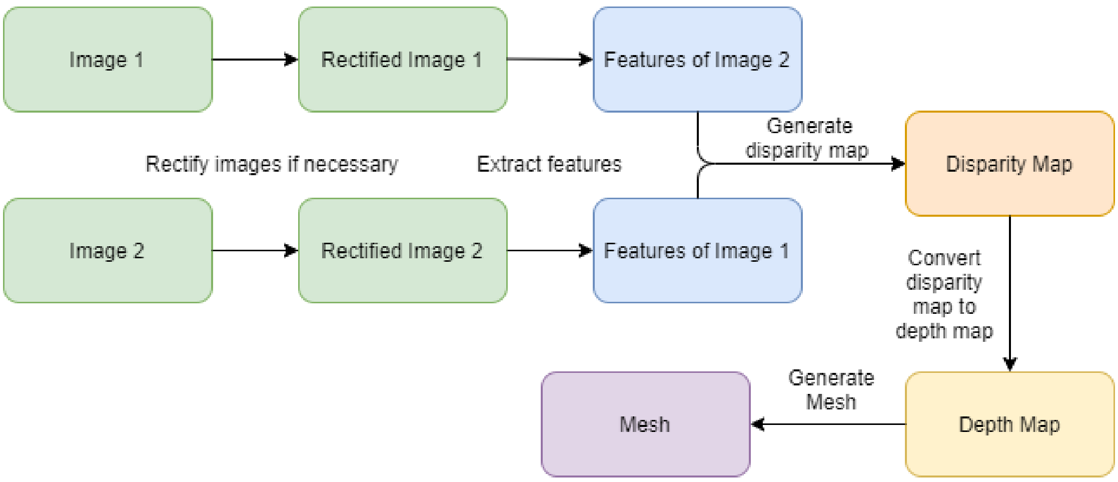


Figure 1: Method overview.

## 2 Requirements

To address the problem of stereo reconstruction, we have selected a range of datasets with different particularities. The first dataset is Middleburry's 2014 stereo dataset [11], comprising of 33 image pairs taken under several different illuminations and exposures. Additionally, the dataset consists of calibration information and the ground-truth disparity maps. The dataset was acquired with a high precision structured lightning system. This dataset is perfect for the use case of reconstructing the scene from only one pair of stereo images. Another dataset would be EuRoC [3], which contains stereo images taken from a drone, as well as calibration data and ground-truth measurements. Another popular dataset used for benchmarking computer vision algorithms is KITTI 2015 dataset [7]. The dataset consists of a total 400 scenes taken from a system mounted on a vehicle, depicting urban traffic scenes. They also provide calibration file and ground-truth in the form of disparity maps. In creating the dataset, they have used the techniques of scene flow to efficiently estimate moving object, but that is something beyond the scope of our project, although it would be interesting to compare how the dynamic objects are interfering with our proposed pipeline.

In terms of libraries and frameworks, we plan to develop our system in C++ and leverage the powerful computer vision library OpenCV [2] and PCL [10] for point cloud processing if needed. For mesh visualization, we are going to use MeshLab [4]. Interesting libraries that we may also use for inspiration are OpenMVG [8] and Theia [12], although these are mostly multi-view geometry libraries, but some ideas can be shared from this domain.

## 3 Team

Andrei Baraian 03739779
Faruk Oruc 03737463
Abhishek Saroha 03722262

## 4 Milestones

To carry out the proposed project, the milestones outline is as follows:

- 12 Jan, 2021 : Baseline Implementation and environment setup: Sparse Stereo Reconstruction and Brute-Force Matching
- 19 Jan, 2021 : Sparse Stereo Reconstruction using the Epipolar Constraint
- 26 Jan, 2021 : Dense Stereo Reconstruction
- 02 Feb, 2021 : Study on Evaluation metrics

## References


[1] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. *Computer Vision-ECCV 2006*, 3951:404–417, 07 2006.

[2] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[3] Michael Burri, Janosch Nikolic, Pascal Gohl, Thomas Schneider, Joern Rehder, Sammy Omari, Markus W Achtelik, and Roland Siegwart. The euroc micro aerial vehicle datasets. *The International Journal of Robotics Research*, 2016.

[4] Paolo Cignoni, Marco Callieri, Massimiliano Corsini, Matteo Dellepiane, Fabio Ganovelli, and Guido Ranzuglia. MeshLab: an Open-Source Mesh Processing Tool. In Vittorio Scarano, Rosario De Chiara, and Ugo Erra, editors, *Eurographics Italian Chapter Conference*. The Eurographics Association, 2008.

[5] David A. Forsyth and Jean Ponce. *Computer Vision - A Modern Approach, Second Edition.* Pitman, 2012.

[6] David Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–, 11 2004.

[7] Moritz Menze, Christian Heipke, and Andreas Geiger. Joint 3d estimation of vehicles and scene flow. In *ISPRS Workshop on Image Sequence Analysis (ISA)*, 2015.

[8] Pierre Moulon, Pascal Monasse, Romuald Perrot, and Renaud Marlet. Openmvg: Open multiple view geometry. In *International Workshop on Reproducible Research in Pattern Recognition*, pages 60–74. Springer, 2016.

[9] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. Orb: An efficient alternative to sift or surf. pages 2564–2571, 2011.

[10] R. B. Rusu and S. Cousins. 3d is here: Point cloud library (pcl). In *2011 IEEE International Conference on Robotics and Automation*, pages 1–4, 2011.

[11] Daniel Scharstein, Heiko Hirschmller, York Kitajima, Greg Krathwohl, Nera Nei, Xi Wang, and Porter Westling. High-resolution stereo datasets with subpixel-accurate ground truth. volume 8753, pages 31–42, 09 2014.

[12] Chris Sweeney. Theia multiview geometry library: Tutorial & reference. http://theia-sfm.org.

[13] Yi Zhou, Guillermo Gallego, Henri Rebecq, Laurent Kneip, Hongdong Li, and Davide Scaramuzza. Semi-dense 3d reconstruction with a stereo event camera. *CoRR*, abs/1807.07429, 2018.