# 3D reconstruction using Structure for Motion
# Progress Report

Kshitij Karnawat
kshitij
kshitij@umd.edu

Hritvik Choudhari
hac
hac@umd.edu

Abhimanyu Saxena
asaxena4
asaxena4@umd.edu

Mudit Singal
msingal
msingal@umd.edu

Raajith Gadam
raajithg
raajithg@umd.edu

## Abstract

*We are working towards 3D reconstruction of indoor spaces using a pair of HDR cameras in a stereo vision configuration mounted on an indoor mobile floor robot that captures various textures and spatial features as 2D images and this data is simultaneously utilized as a feed to our algorithm which will allow us to visualize the depth map.*

## 1. Introduction

Structure from Motion (SfM) is a technique used in computer vision and photogrammetry to create 3D models from a set of 2D images captured from different viewpoints. The technique involves reconstructing the 3D structure of an object or scene by analyzing the images' geometric relationships and camera parameters.

The primary goal of our project is to create a robust and efficient SfM algorithm that can handle large data sets and provide accurate results in a reasonable amount of time while using a stereo camera configuration on a mobile floor robot integrated with Raspberry Pi. We are aiming at making the system autonomous.

To achieve this, all of us have conducted a comprehensive study of the existing SfM techniques, their limitations, and their strengths. SfM has several advantages over other techniques used for 3D reconstruction. It is a relatively low-cost technique that only requires a standard camera, making it accessible to a wider range of users. This technique can also handle a larger variety of object geometries and textures, making it more versatile. Additionally, SfM can be used in conjunction with other techniques, such as LiDAR and stereo vision, to create more accurate 3D models. However, SfM also has some limitations, such as the need for a large number of images to construct a reliable 3D model, and the sensitivity to camera calibration and lighting conditions. We are working towards improving the reliability of the model by also integrating it with the stereo vision technique and improving the algorithm.

The project also involves the implementation of several algorithms and techniques such as feature extraction, matching, bundle adjustment, and triangulation, among others.

## 2. Techniques employed

Our project aims to use a pair of HDR cameras in a stereo-vision configuration to reconstruct 3D models of indoor spaces. In this literature review, we will examine the existing research on the use of HDR cameras and stereo vision for 3D reconstruction, as well as related techniques such as feature detection and matching, bundle adjustment, and parallel processing.

HDR cameras have been used in several applications related to 3D reconstruction, such as photogrammetry and computer vision. HDR imaging can improve the quality and accuracy of the reconstructed 3D models by providing more detailed and realistic images. The use of HDR cameras has been shown to be particularly useful in scenes with high-contrast lighting, where conventional cameras may struggle to capture all the detail in both bright and dark areas.

## 3. Computing Depth Map from Stereo Images

We, humans, have evolved to be with two eyes that we can perceive depth. And when we organize cameras analogously, it's called Stereo-Vision. A stereo-vision system is generally made of two side-by-side cameras looking at the same scene, the following figure shows the setup of a stereo rig with an ideal configuration, aligned perfectly.

Stereo vision is another important technique for 3D reconstruction, which involves using multiple cameras to cap-

ture images from different viewpoints. Stereo vision can provide more accurate depth information compared to other techniques such as structure from motion (SfM). Several algorithms have been developed to perform stereo matching and generate a 3D reconstruction from the stereo images.

Feature detection and matching is a crucial steps in many 3D reconstruction algorithms, including the proposed project. Popular feature detection algorithms include SIFT, SURF, and ORB, while feature matching algorithms include brute force matching and RANSAC. The goal of feature detection and matching is to identify corresponding points in multiple images, which can then be used to generate a 3D point cloud.

Bundle adjustment is another important technique used in 3D reconstruction to refine the camera parameters and improve the accuracy of the 3D model. Bundle adjustment algorithms typically minimize the reprojection error between the 2D image points and the corresponding 3D points. Techniques such as the Levenberg-Marquardt algorithm, Gauss-Newton optimization, and the conjugate gradient method have been used for bundle adjustment. Parallel processing can be used to speed up the 3D reconstruction process by distributing the computation across multiple processors or nodes. Parallel processing can be particularly useful for large-scale 3D reconstruction tasks that involve processing a large number of images.

In summary, our proposed project aims to use HDR cameras in a stereo vision configuration to reconstruct 3D models of indoor spaces. This approach builds on existing research on the use of HDR imaging, stereo vision, and related techniques such as feature detection and matching, bundle adjustment, and parallel processing. The proposed project has several potential applications in areas such as robotics, interior design, and architecture, and has the potential to make significant contributions to the field of 3D reconstruction.

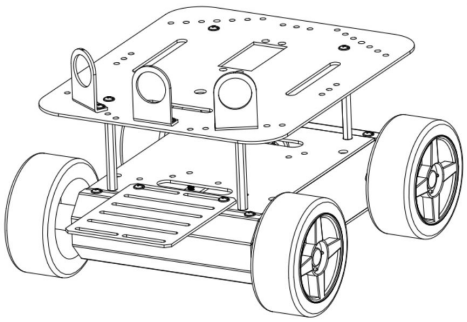

Figure 2. Robot model

## 4. Hardware

Our hardware component comprises the following stack:

| SR.NO. | HARDWARE COMPONENT | SPECIFICATIONS |
|---|---|---|
| 1 | Jetson Nano | Quad-core ARM Cortex-A57 with 4GB RAM |
| 2 | Raspberry Pi Camera V2 | 3.68 x 2.76 mm sensor; f = 3.04 mm |
| 3 | Robot Chassis | DF Robot Pirate (8.27 x 7.48 x 3.15 inches) |
| 4 | Arduino Nano | ATmega328 microcontroller |
| 5 | Adafruit BNO055 IMU | 9-DOF measurement with ARM Cortex-M0 based processor |

Figure 1. Hardware stack

## 5. Approach

Below is the flow chart that briefly describes the process to obtain a 3D point cloud from 2D images:

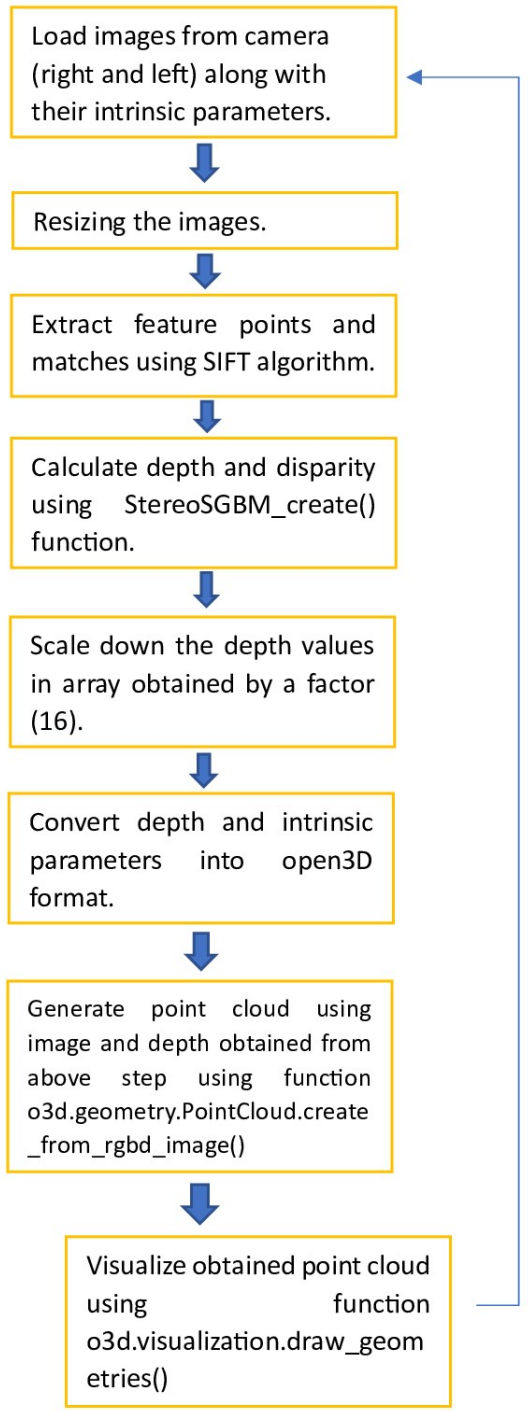


Figure 3. Flow chart

## 6. Results

Our dataset contains hundreds of images captured by us exclusively for the purpose of this project, which is a collection of diverse varieties in textures and lighting. We are aiming to generalize our model by reconstructing various different indoor spaces. One such common point of interest or observation in an indoor environment is a chess board which generally has very strong features and can be used for the computation of various camera characteristics.

The images used to generate the point cloud are taken using a OnePlus 7 camera.

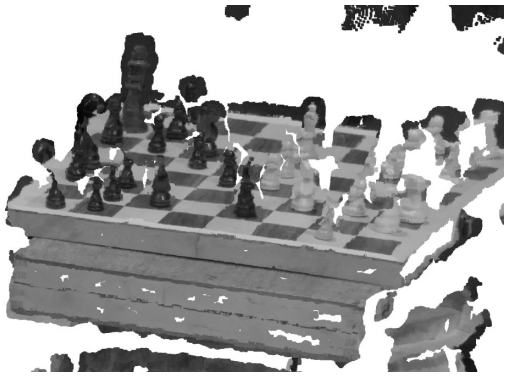

Figure 4. Point Cloud for chess3 from the Middlebury Dataset zoomed in for detail [5]

Figure 5. Point Cloud for chess3 from the Middulebury Dataset[5] with color

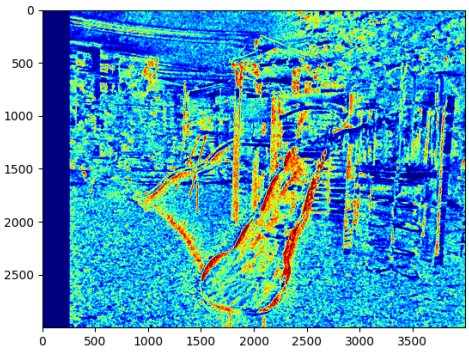


Figure 6. Depth on Custom Data

Figure 7. Point Cloud for Custom Data with color

Figure 8. Point Cloud (Left side view) for Custom Data with color

Figure 9. Point Cloud (Right side view) for Custom Data with color

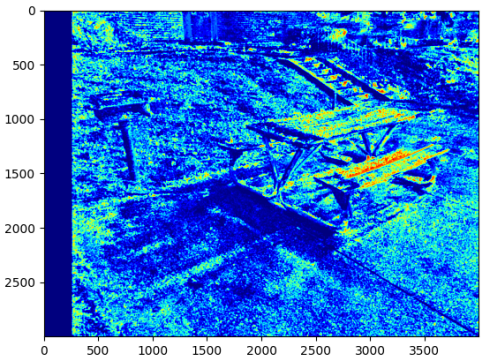


Figure 10. Depth on Custom Data

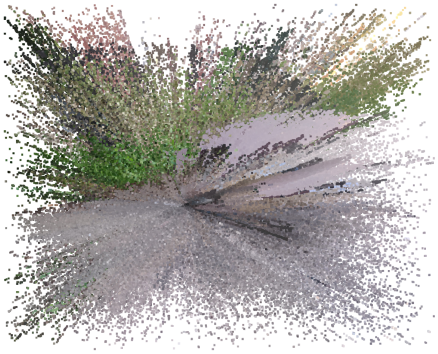

Figure 11. Point Cloud for Custom Data with color

Figure 12. Point Cloud for Custom Data with color

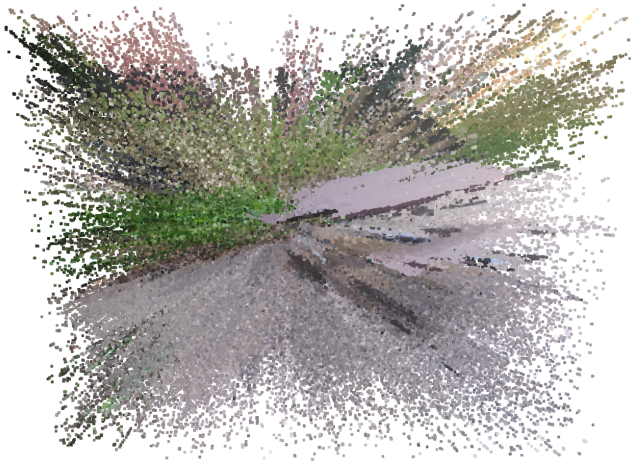

Figure 13. Point Cloud for Custom Data with color

## 7. Fallback goals

1. The image quality from the Raspberry-pi camera ($640 \times 480$) pixels is not up to the mark compared to high-end cameras like Intel Real-sense or smartphone cameras ($4000 \times 3000$) pixels. This creates a huge difference in the depth map and generated point cloud quality as the features that we are able to extract using SIFT are around 150 and more than 15000 for the Raspberry-pi camera and OnePlus 7 camera respectively. In that case, we shall go for a better camera than the originally decided Raspberry-pi camera to get better results.
2. We originally decided to build a room map using 3D reconstructed point cloud using SfM technique. Depending on whether our stereo camera setting is giving us desired results or not, we may switch our use case to 3D object reconstruction using the same technique.

## 8. References


1. https://github.com/facebookresearch/Replica-Dataset
2. https://towardsdatascience.com/3-d-reconstruction-with-vision-ef0f80cbb299
3. https://www.cs.cornell.edu/projects/bigsfm/
4. https://www.youtube.com/watchv=DoZJaqBzSso&ab_channel=NicolaiNielsen
5. http://www.open3d.org/docs/release/tutorial/geometry/pointcloud.html
6. https://docs.opencv.org/3.4/d9/d0c/group__calib3d.html#ga1bc1152bd57d63bc524204f21fde6e02
7. http://www.open3d.org/docs/release/tutorial/geometry/pointcloud.html
8. Peter Anderson, Qi Wu, Damien Teney, Jake Bruce, Mark Johnson, Niko Sunderhauf, Ian Reid, Stephen Gould, and Anton van den Hen- ¨ gel. Vision-and-language navigation: Interpreting visually-grounded navigation instructions in real environments. In CVPR, 2018.
9. Stanislaw Antol, Aishwarya Agrawal, Jiasen Lu, Margaret Mitchell, Dhruv Batra, C. Lawrence Zitnick, and Devi Parikh. VQA: Visual Question Answering. In ICCV, 2015.
10. D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nesic, X. Wang, and P. Westling. High-resolution stereo datasets with subpixel-accurate ground truth. In German Conference on Pattern Recognition (GCPR 2014), Münster, Germany, September 2014.