DEPTH ESTIMATION USING STEREO CAMERAS

Santosh Vasa, Ashwin Unnikrishnan

INTRODUCTION

Depth map data is used in a variety of applications, including object avoidance, 3D reconstruction, localization, and more. This is possible with a stereo setup of two cameras. Lidars and radars can be used to create depth maps, but they are expensive pieces of equipment. To retrieve different views, stereovision employs two horizontally (or arbitrarily) displaced cameras. The disparity between the two image views allows us to estimate the depth of the image at each pixel location.

COMPUTATIONAL PERSPECTIVE:

Uncalibrated Stereo: Although the intrinsic parameters of each camera are known in uncalibrated stereo setups, the extrinsic parameters relative to both cameras are unknown. These extrinsic parameters are needed for the 3D – construction of the scene (3d point cloud). To calibrate the setup, we will need to compute the following.

Good features to track: To compute the fundamental matrix, we will need at least 8 good features that match both images taken from each camera. Using SIFT or SURF features can provide good results without the aperture problem. We use the RANSAC algorithm to reduce the number of bad matches by removing the outliers. We will investigate more algorithms in this space and try out state-of-the-art algorithms.

Fundamental Matrix: The fundamental matrix can map a pixel location in the left image to a line segment in the right image based on the epipolar constraint. There are eight unknown parameters in this matrix. Using 8 feature detection points, such as those provided by SIFT, we can obtain 8 equations for solving the fundamental matrix.

Template Matching: We can find the depth correspondence or registration of all the matching points in the images using template matching. We will investigate the most efficient and accurate methods of accomplishing this. The disparity map of the scene, which can be used to estimate the depth map, is formed by the corresponding points between two images.

Triangulation: We must find the 3-dimensional coordinates of the points in either the left or right camera coordinate system using the intrinsic parameters of both cameras. This is the inverse of project 4, in which we projected 3D points onto the 2D image plane. Using the intrinsic and relative extrinsic parameters, we can find four equations with three unknowns that can project two-dimensional points into three dimensions. This whole process lets us map arbitrary views to depth maps of the scene.

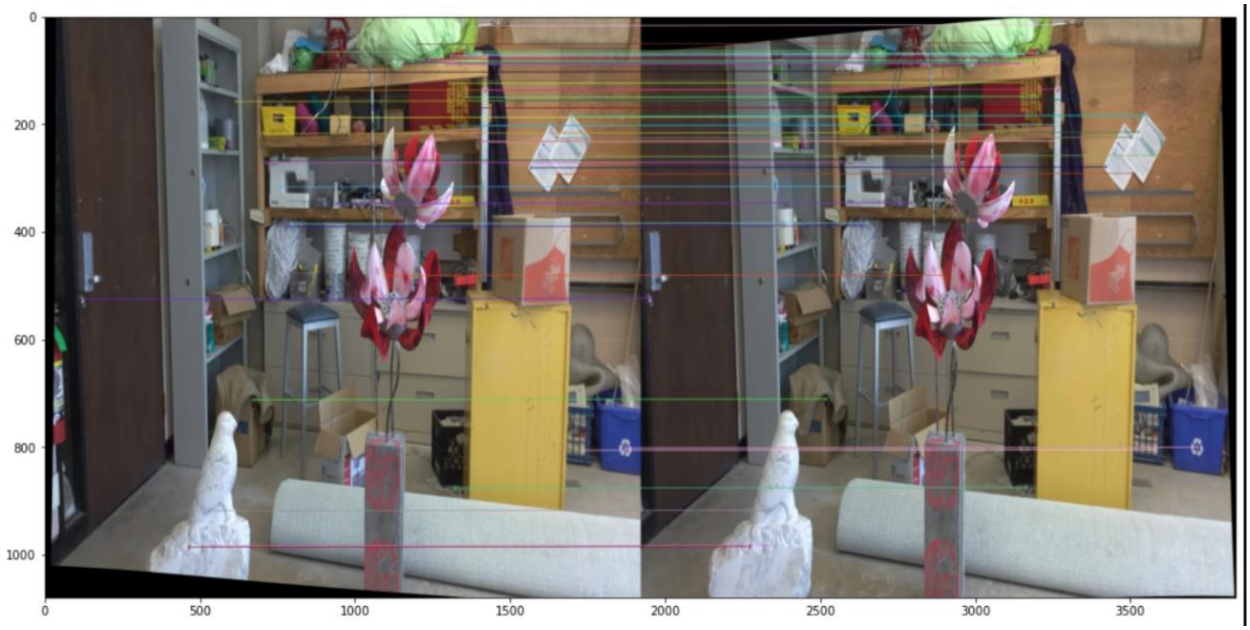
WHY IT IS INTERESTING:

Applications: Depth maps can be used for a variety of purposes. Using the additional depth information obtained from the stereo setup, we can improve the quality of face recognition. Identifying the scale and measurements of objects in an image can provide more information for better classifying in object detection. Simple setups like this can be used by self-operating home robots to estimate and localize themselves in indoor environments.

Learning: This project will introduce us to the geometry of computer vision. After successfully completing this project, we will be able to pursue more advanced projects such as 3D – Scene Reconstruction, Localization and Mapping, Augmented Reality on any surface, Obstacle Avoidance, and others. These are also some of the topics we hope to cover in our upcoming Advanced Computer Vision course this fall.

Work Updates:

- **Check-in 1** (4/18/2022)
  - We are currently using the dataset from **middlebury** and implementing different feature detectors available online. For example, the below image is an output of SIFT (Brute Force Matcher) algorithm that maps feature points between two images in the stereo. The feature points will be used to build the essential and the fundamental matrix.



  - Next, we will try using different feature point matching algorithms like RANSAC and try to remove bad feature points and improve the SIFT algorithm. This will improve our run time as well as we won't be comparing each feature point on the left with each feature point on the right.
  - As part of our live setup, we are setting up two web cameras and calculating the baseline. Once the pipeline is set with the above dataset, we will calibrate our camera and then use the images to find the disparity and calculate the depth.