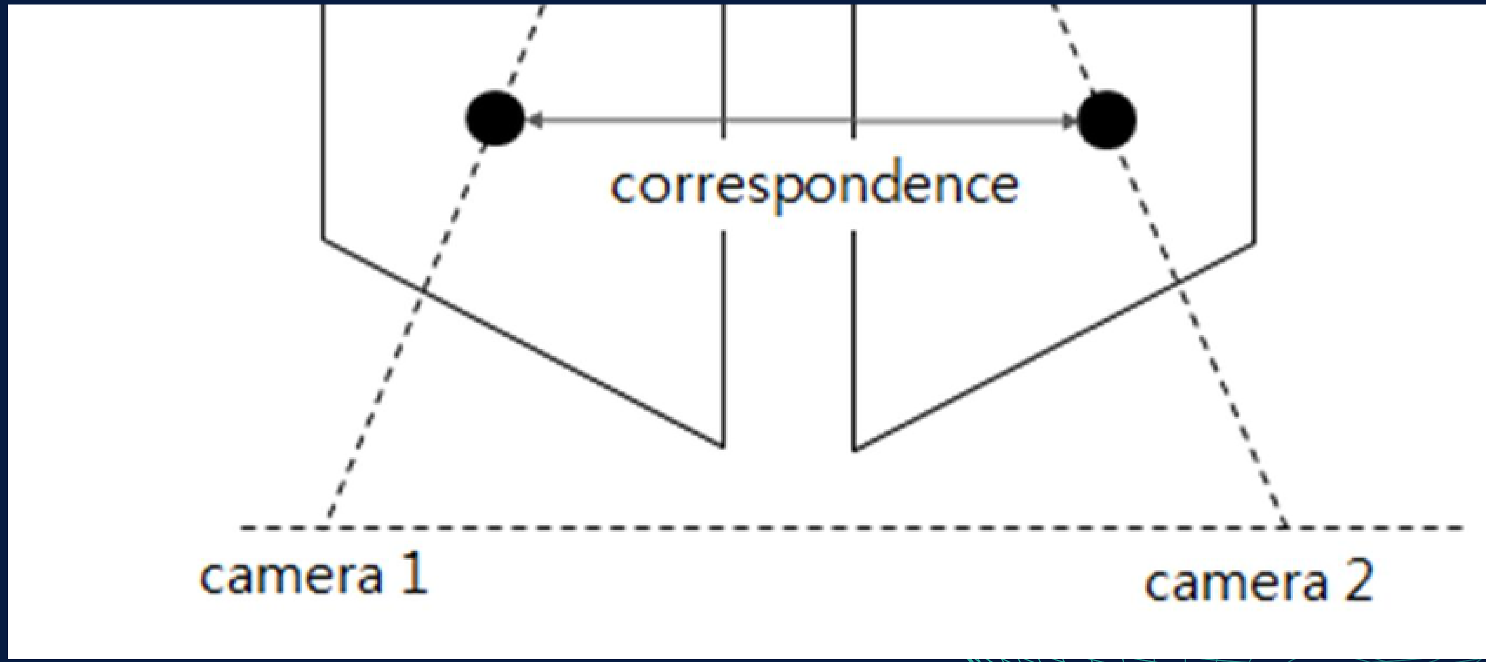
COMPUTER VISION

CSE 344

PROJECT PRESENTATION

Group No 39

PROBLEM STATEMENT



Our objective is to implement, analyze and to try to improve depth estimation through the integration of frame rectification and triangulation techniques, leveraging geometry and visual processing. By utilizing pre-trained models and existing resources, we aim to enhance the accuracy and reliability of depth estimation in various scenarios.
The challenge lies in effectively integrating these techniques and leveraging existing work done to optimize the overall depth estimation process.The solution should showcase the assembly and fine-tuning of these components to achieve superior performance.We have also analysed our results, identified where our model was failing and have tried to understand reasons behind that.

DATASET DESCRIPTION

For depth estimation, we require rectified stereo-image pairs. We obtain stereo image pairs from these datasets and obtain the rectified versions of them through our code in python.

DATASET 1

We have taken some images from the middlebury.edu 2021 dataset. This contains images taken from different orientations from the same scene. It also contains images with variations of illumination and exposure. It has 24 unique pairs of images.

DATASET 2

This dataset contains images taken from the real world by our phone camera. These also involve images taken from different viewpoints.

DESCRIPTION

For dataset 2 we have 22 images and for Middlebury dataset, there are 48 images. We have variations in illumination in these pairs of images to test how it affects our results.

DATASET DESCRIPTION AND DEMONSTRATION

Sample Images from Middlebury Dataset

Sample Images from Real World Dataset

Same Images after increasing Illuminaton

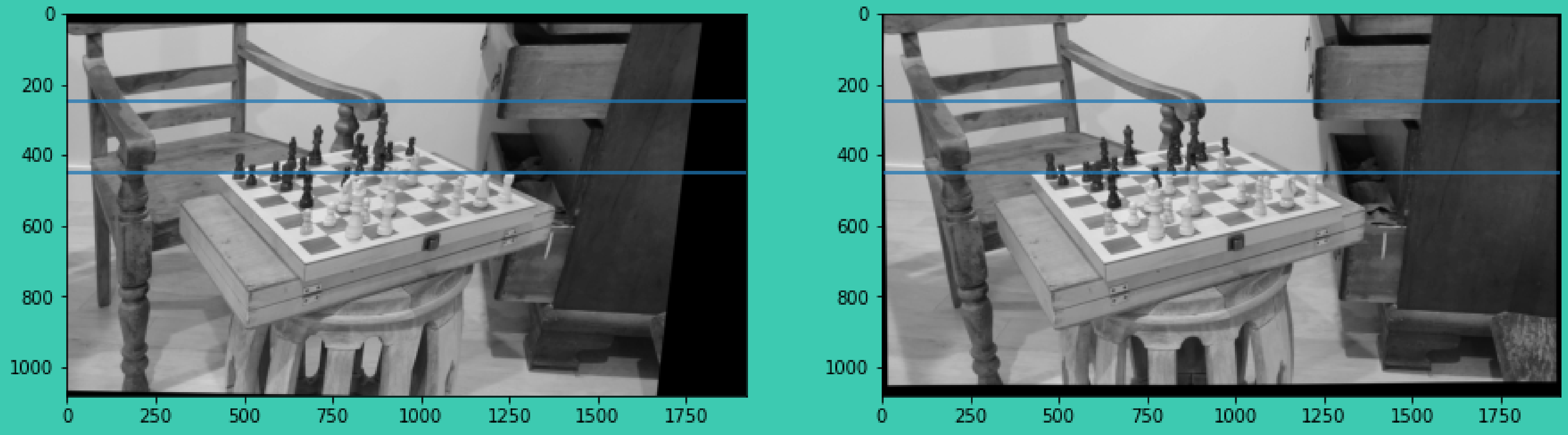
Same Images after increasing Illuminaton

Algorithm and Methodology

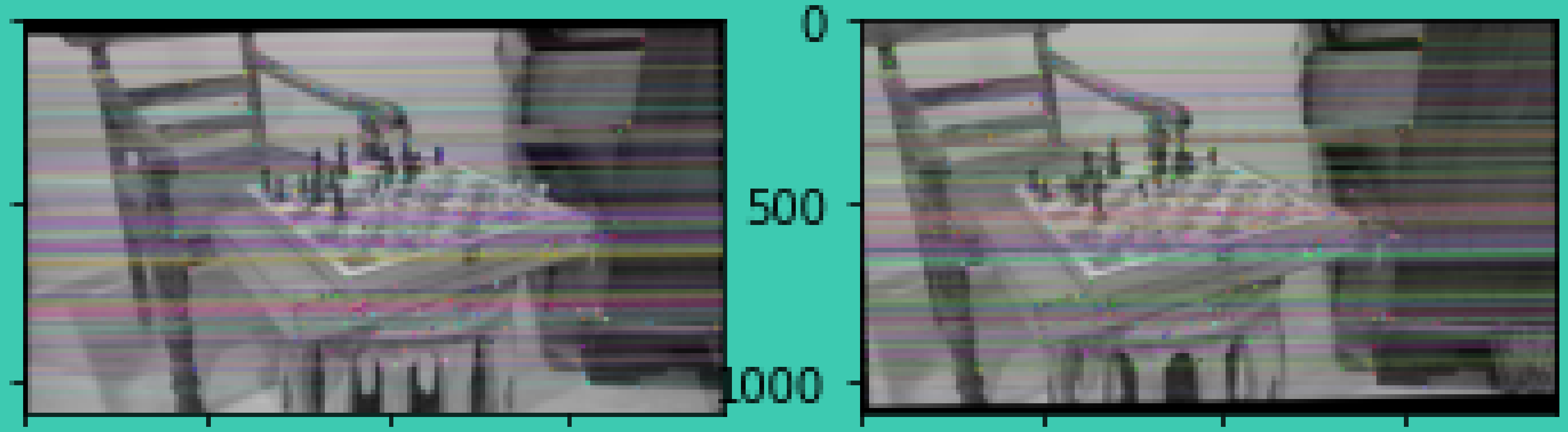
Our code performs stereo depth estimation using the SIFT feature descriptor and the StereoSGBM algorithm. Here's an overview of the steps:

- The code reads two grayscale images (**img1** and **img2**) representing the sterio pair.
- SIFT keypoints and descriptors are computed for both images using the **cv.SIFT_create()** function which capture distinctive features in images.
- Keypoints are matched between the two images using the Brute-Force Matcher (**cv.BFMatcher()**). The matches are filtered based on the distance ratio to keep only the good matches.
- The fundamental matrix is computed using the RANSAC algorithm with the **cv.find FundamentalMat()** function. It describes the geometric relationship between the two camera views.
- Epipolar lines are computed for each matched pair of keypoints using the fundamental matrix and the **cv.computeCorrespondEpilines()** function. They represent the possible locations of corresponding points in the other image.
- The cv.stereoRectifyUncalibrated() function is used to compute rectification transforms for the images based on the fundamental matrix. The rectified images (img1_rectified and img2_rectified) are obtained by warping the original images.
- The StereoSGBM algorithm is used to compute the disparity map from the rectified images. The algorithm takes into account the block size, minimum disparity, maximum disparity, uniqueness ratio, and other parameters. The resulting depth map is then normalized and displayed.

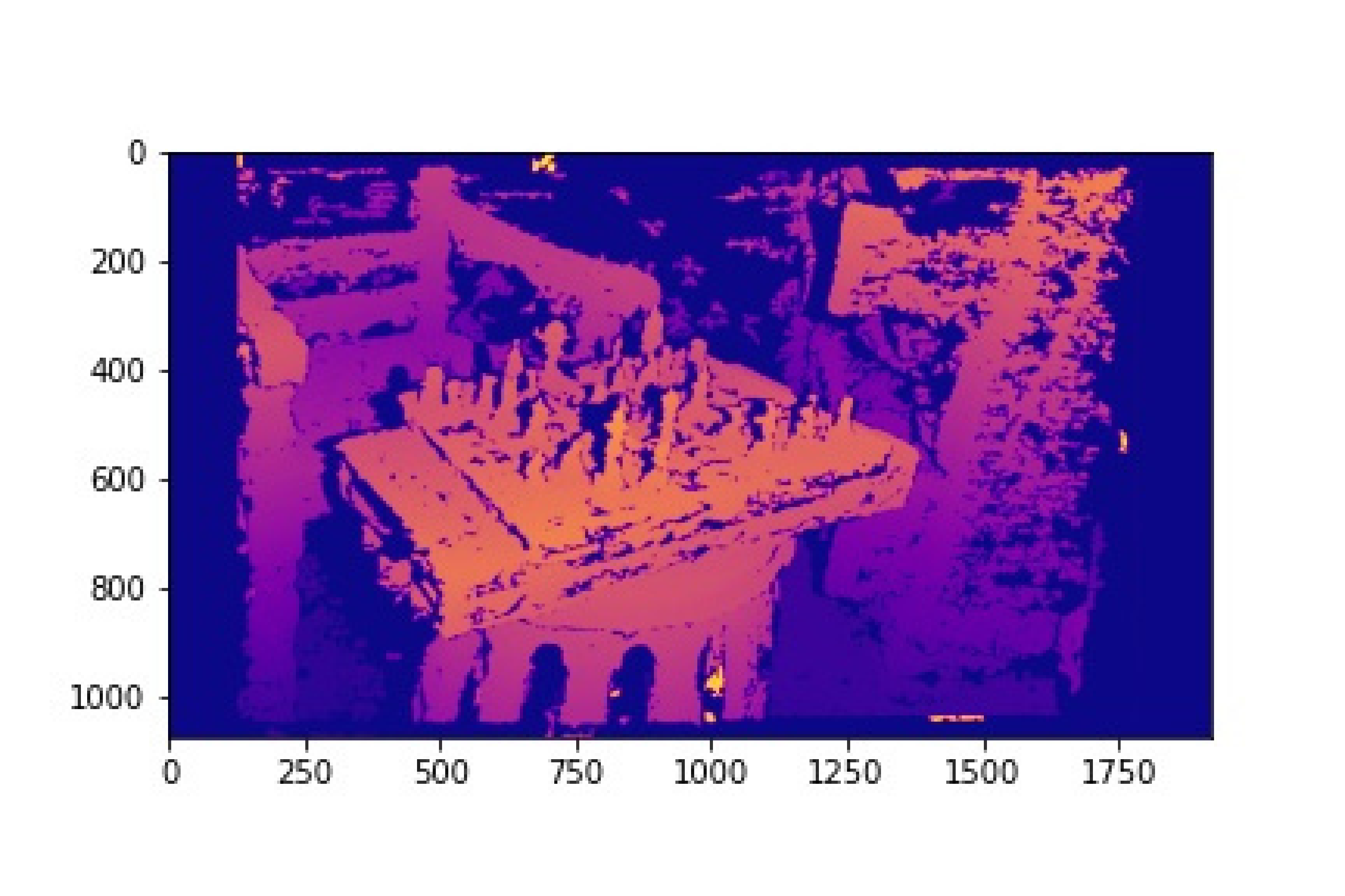
SIFT Keypoints



Rectified Images



Epilines in both images



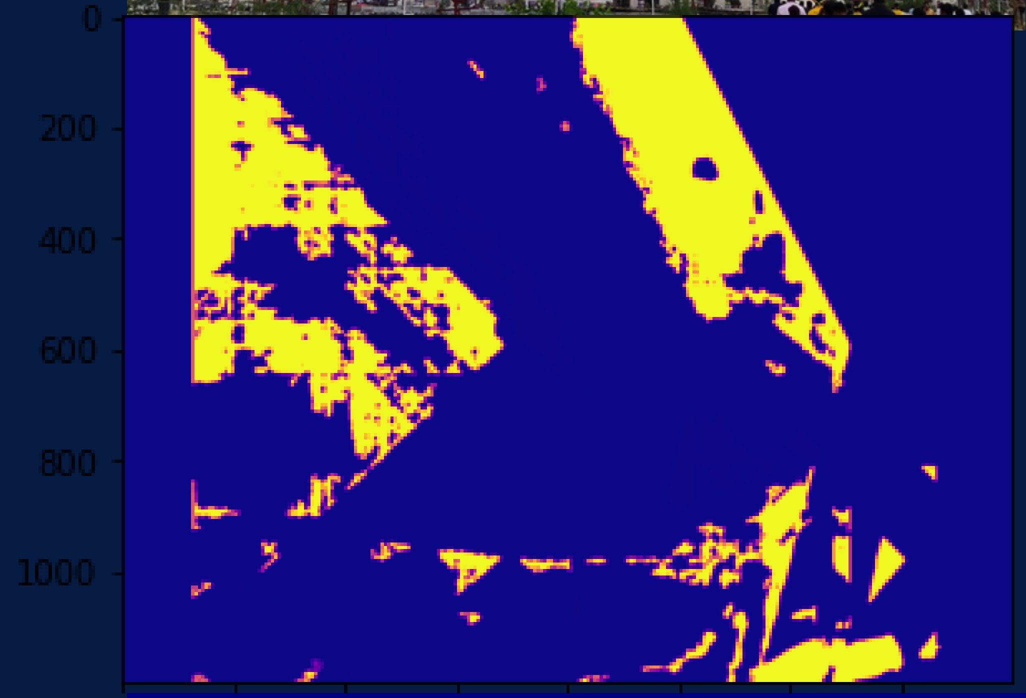
Depth Matches

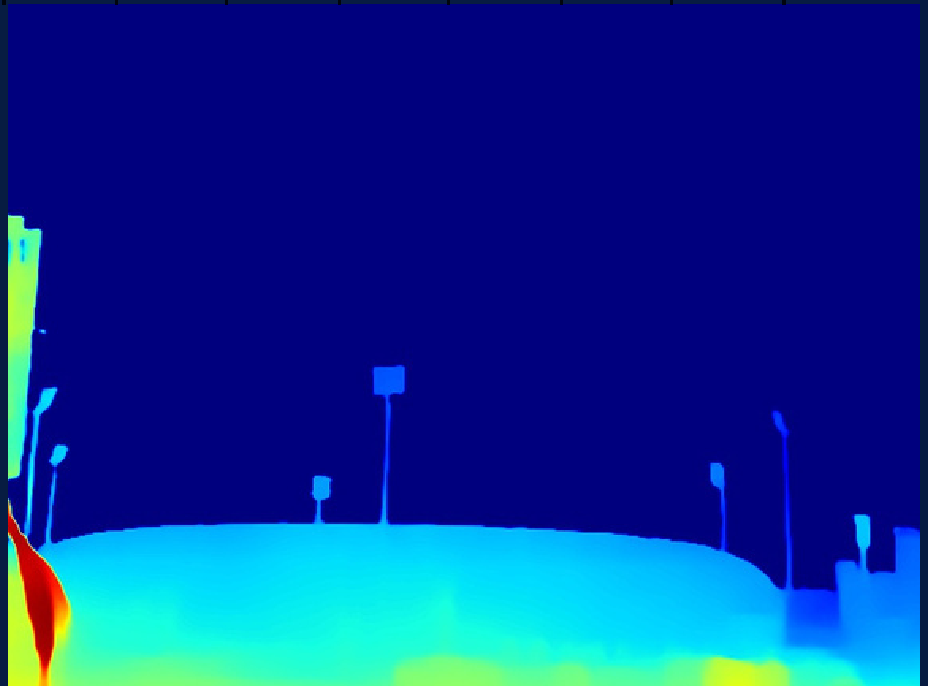
SIFT Matches

Approach 2 – Deep Learning

This code uses the train MiDaS model to estimate depth from a set of input images. It processes each image, applies the depth estimation model, and saves the resulting depth maps.

The network is trained on labeled and unlabeled data to understand the relationship between image features and depth information. By leveraging this learned knowledge, MiDaS can predict depth maps from single images, capturing the underlying 3D structure of the scene accurately.
Deep learning gives much more accurate results than the geometrical approach for all the variety of images.





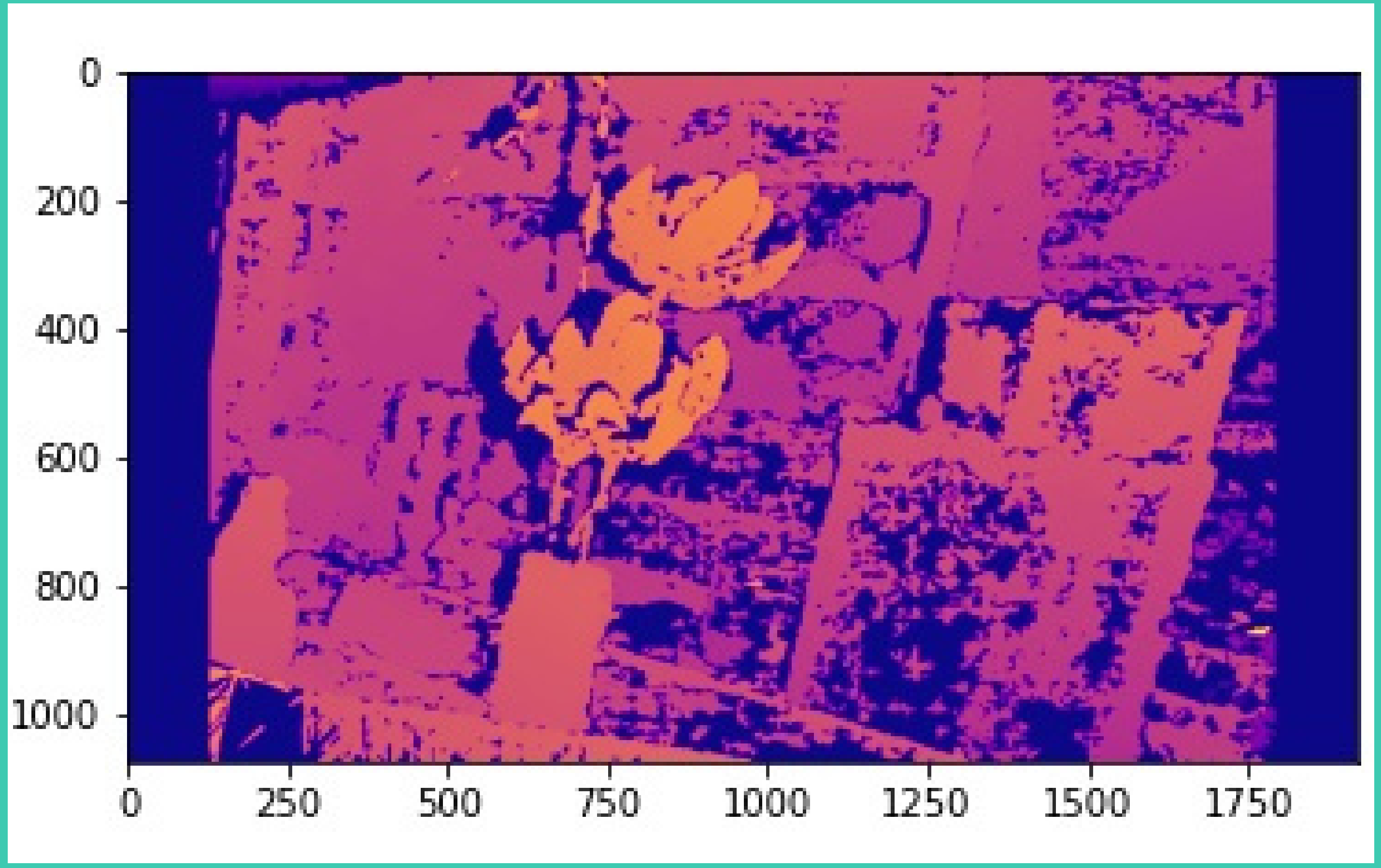
Variations in approach tried

Various techniques were explored in the code and finally only those were kept that gave the best results.

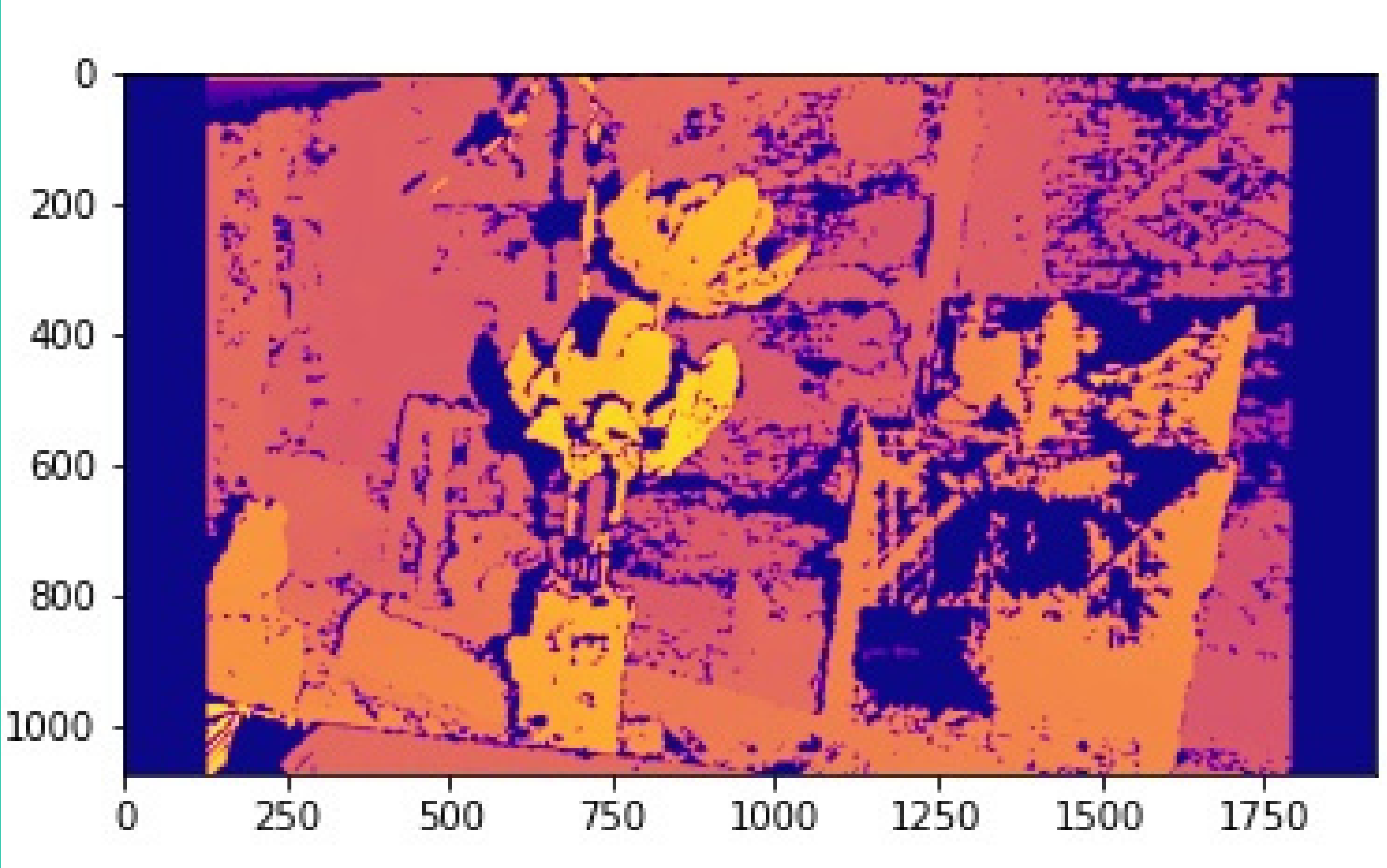
Color vs Grayscale images

We tried both RGB and grayscale images, Grayscale images work better for the StereoSGBM algorithm for several reasons:

1. Simplified Intensity Comparison as grayscale images have single chanel
2. Elimination of Color Inconsistencies.
3. Reduced Computational Complexity



Using FLANN Index



Using Bruteforce

Keypoint correspondence matching

FLANN Index was tested as an alternative to bruteforce Matcher. It performs a k-nearest neighbor search to find the best matches between the descriptors of the two images.

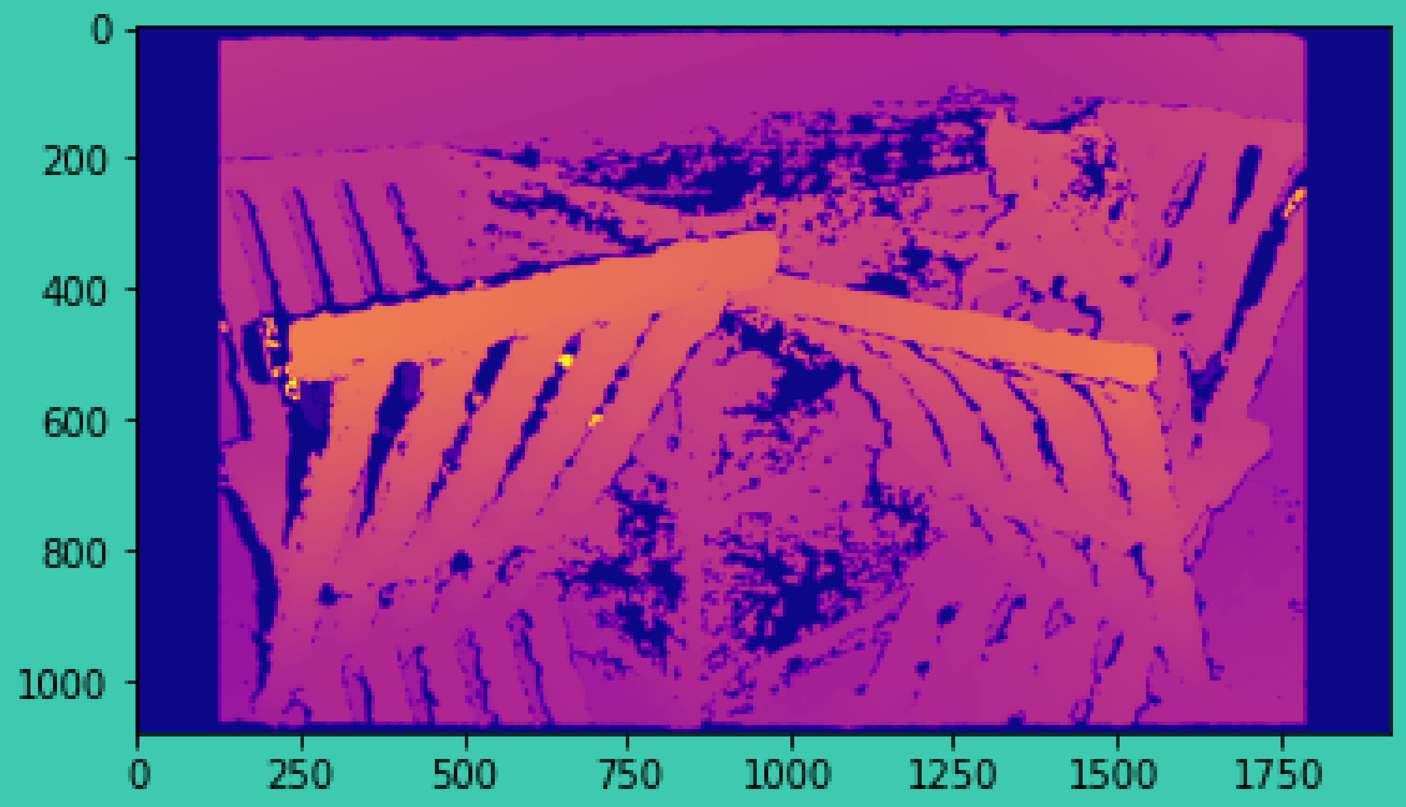
Brute-Force Matcher compares each descriptor from the first image with all descriptors from the second image, which can be computationally expensive but provides more accurate results and hence has been finally used.

Variations in approach tried

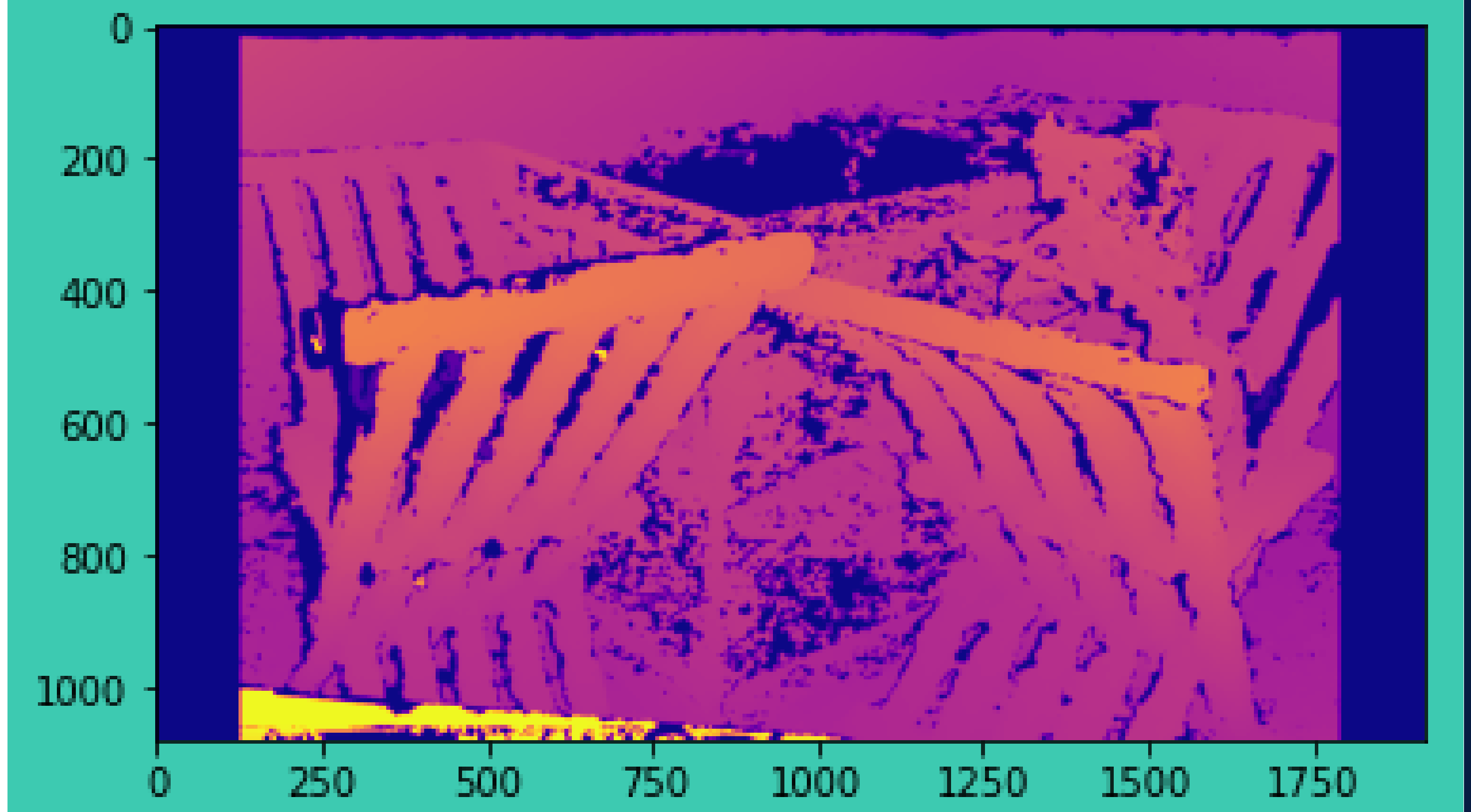
Various techniques were explored in the code and finally only those were kept that gave the best results.

Illumination changes

We tested our code on images with high brightess and low brightness and in most cases higher illumination leads to better results.
Higher image brightness improves contrast, making keypoints more distinct and aiding keypoint detection. Stronger keypoints lead to better matching.
Brighter images also have more detailed texture patterns, enhancing feature descriptors for improved matching.



Before increasing illumination

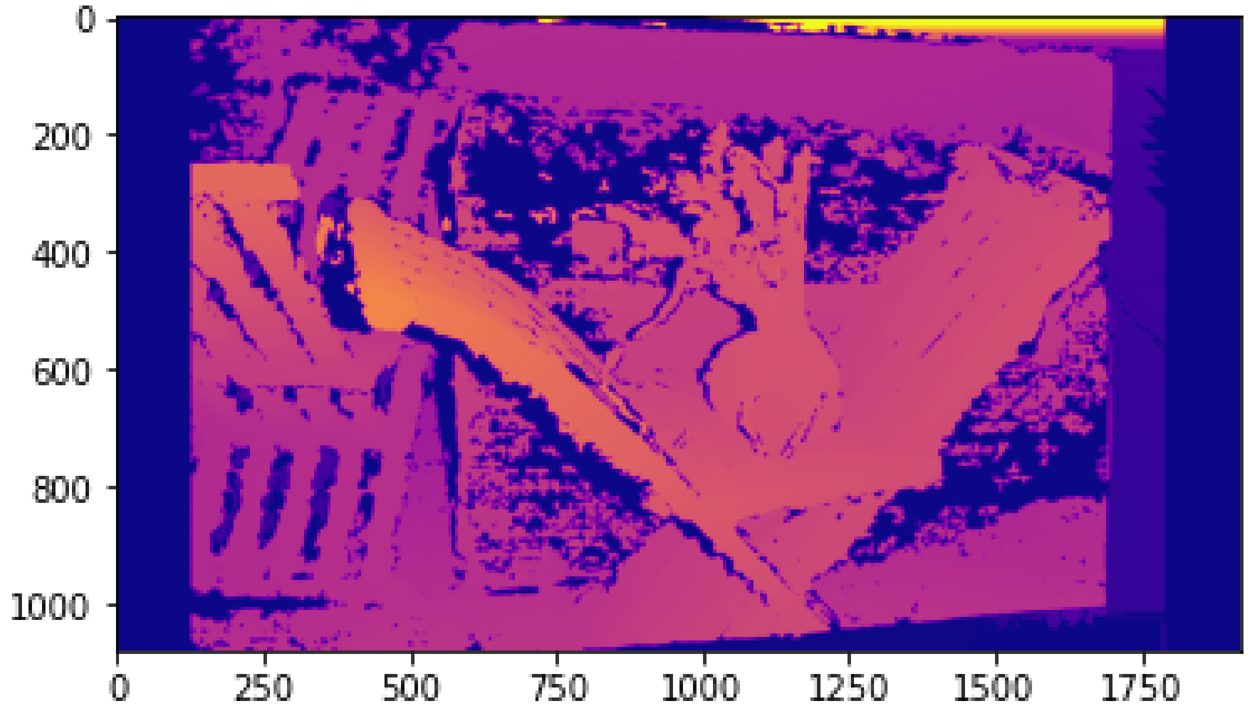


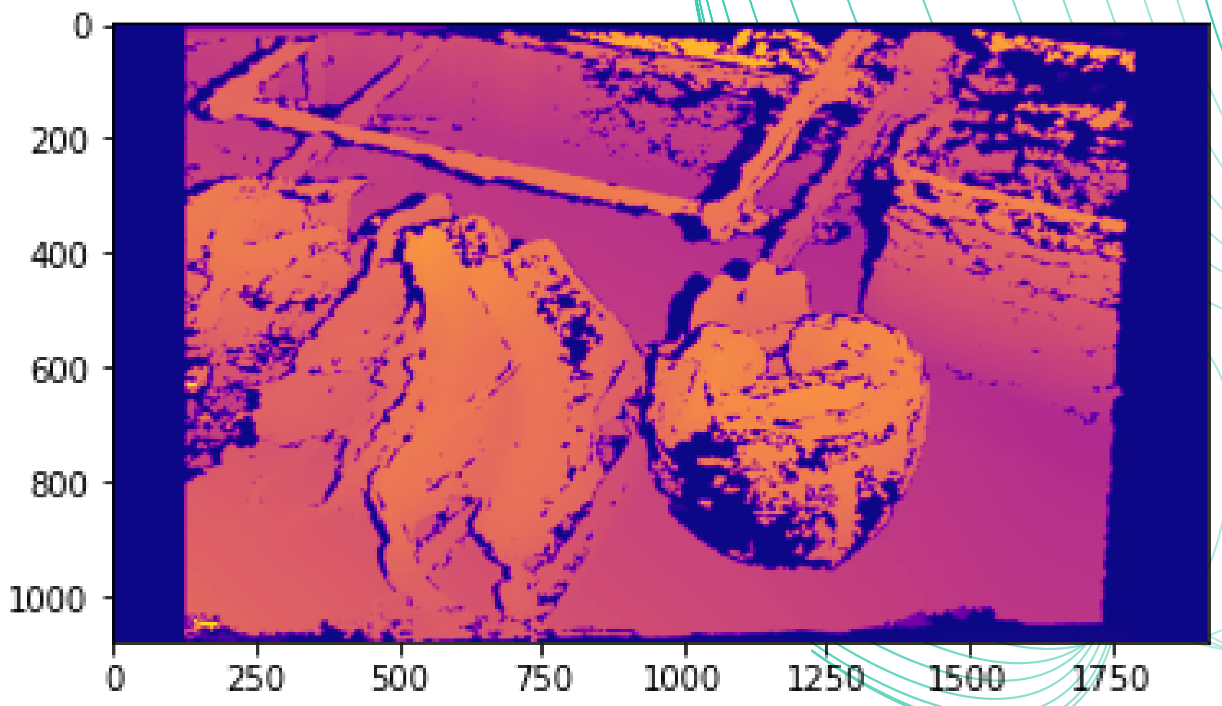
Afttter increasing illumination

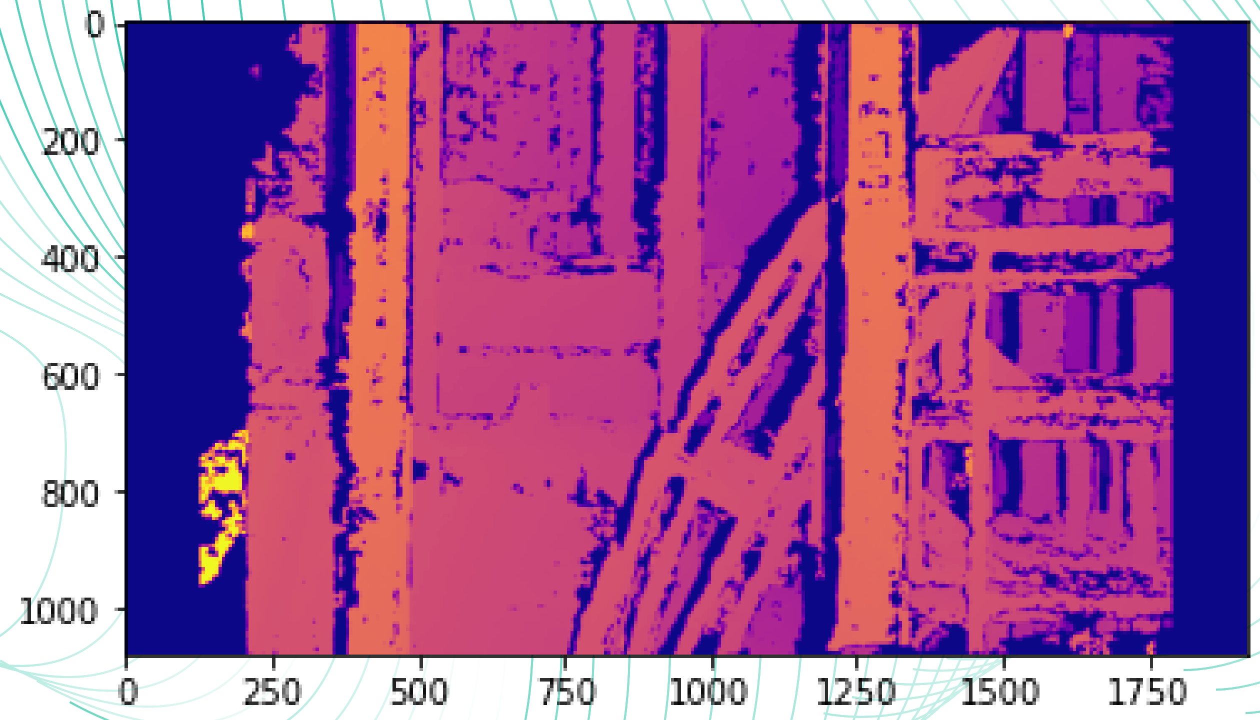
RANSAC for Fundamental matrix calculation and tuning paramters

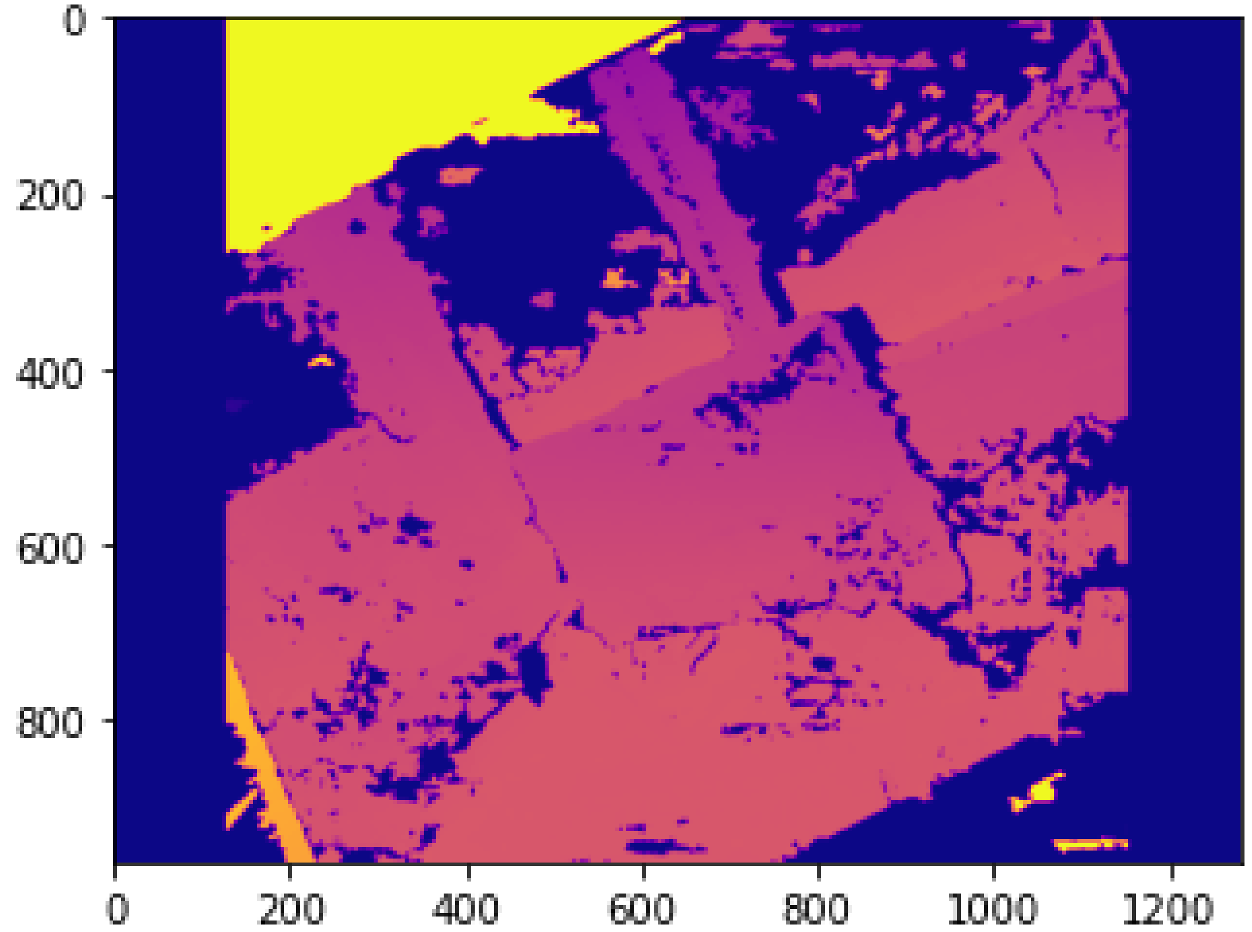
The fundamental matrix was also computed using the normalized 8-point algorithm with cv.FM_8POINT.
However, RANSAC gave better results, and it is more robust to outliers and was used in the final implementation. We also fine-tuned parameters in our opencv functions to get optimal results

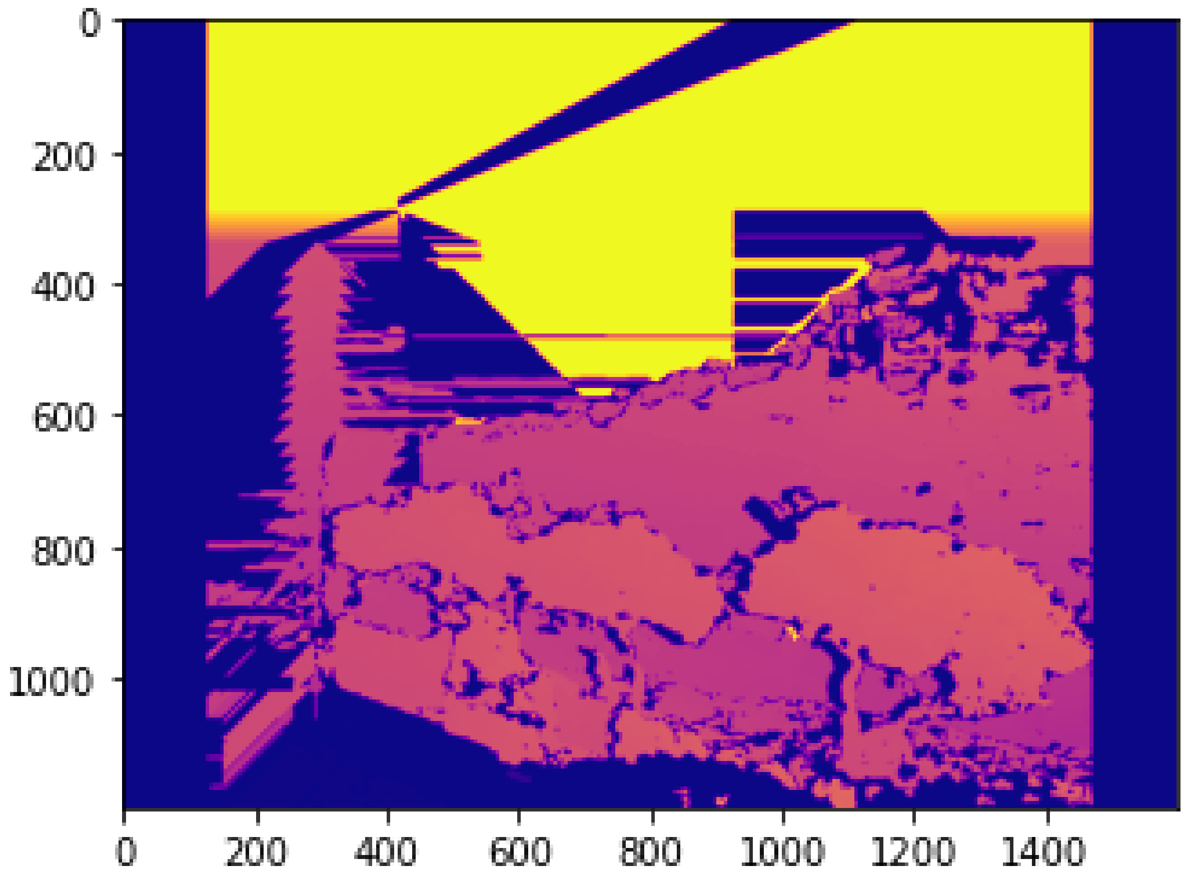
Results Obtained

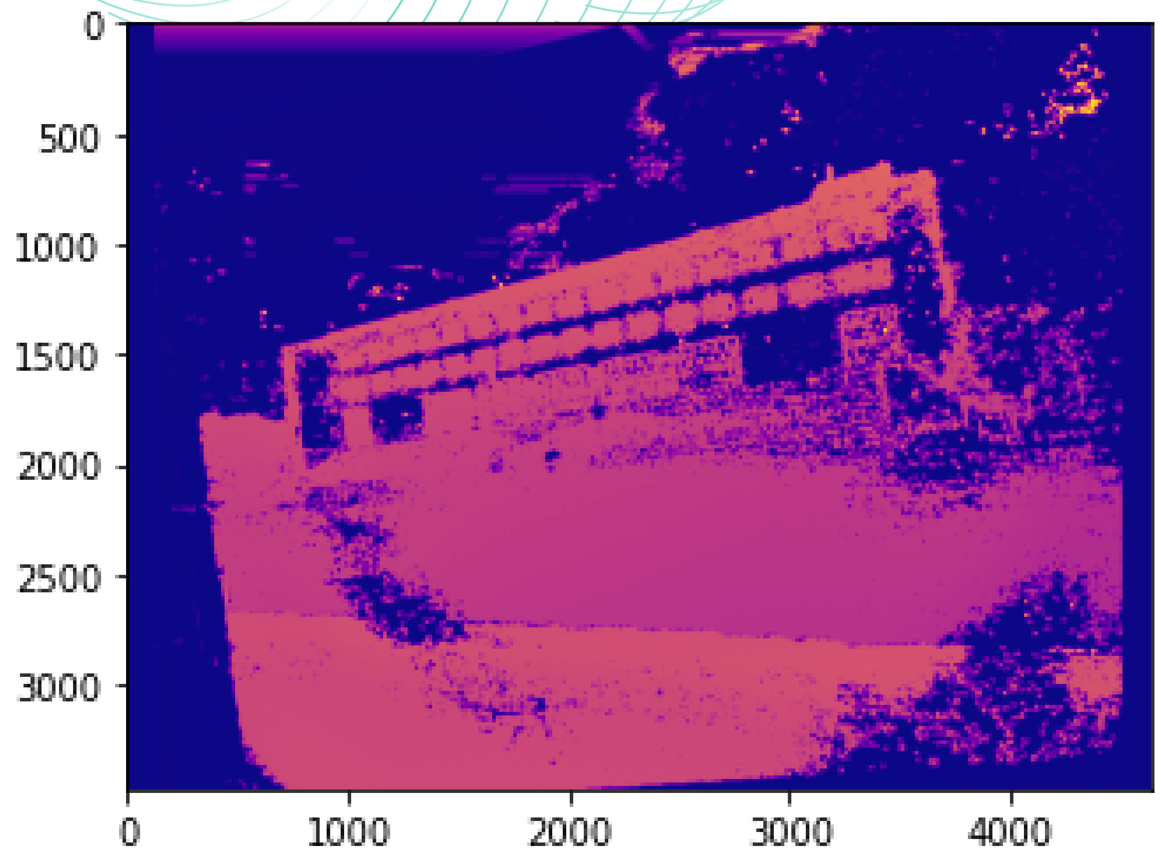


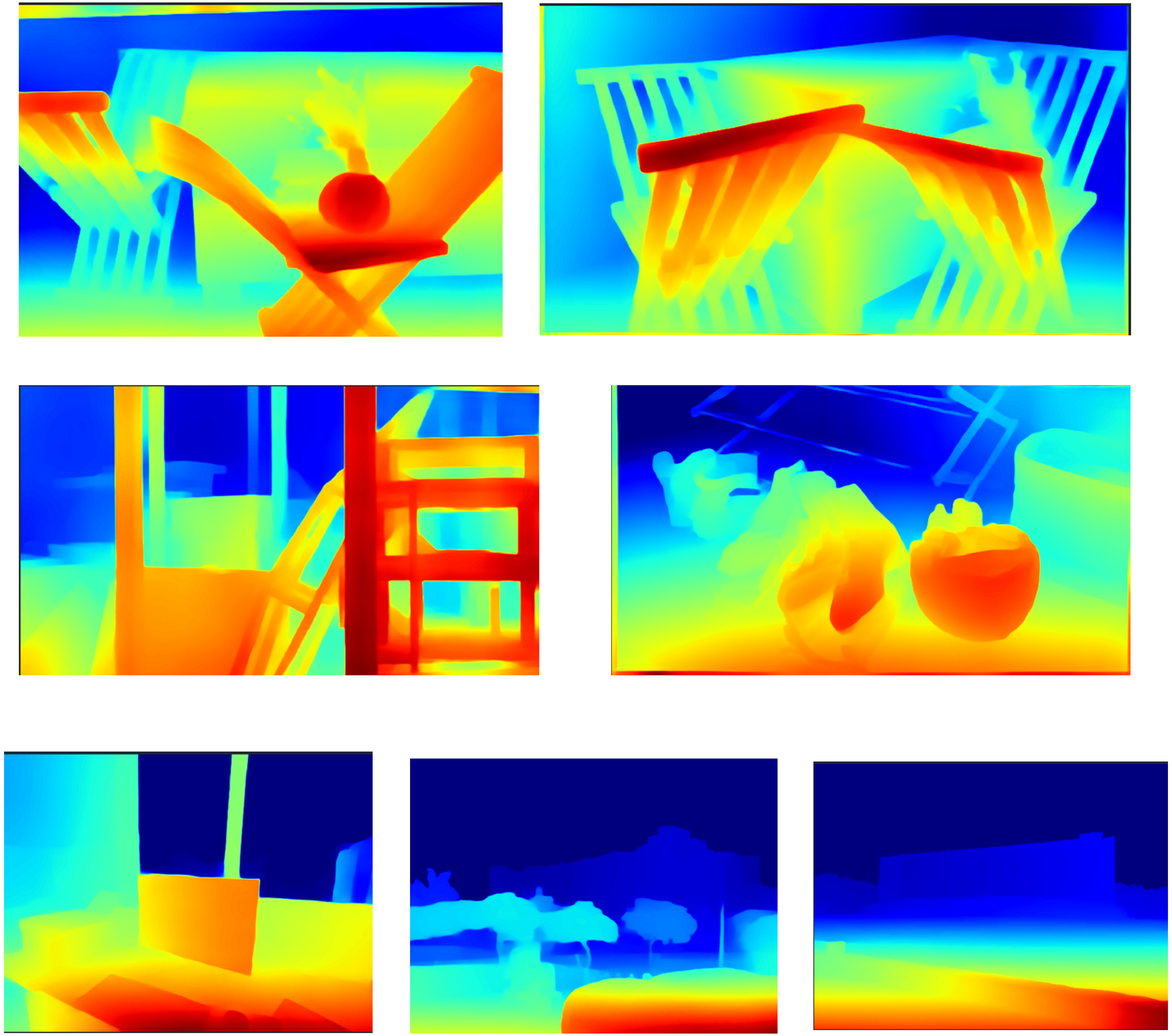












ANALYSIS OF RESULTS

Visual Analysis

- We have visualized the depth map using a colormap that represents depth values using "plasma" colormap to assign different colors to different depth values.
- This visualization helps in understanding the spatial distribution of depth in the scene.
- Blue Corresponds to a farther point in the image here
- While Red Corresponds to closer points in the image.

Disparity range

We can check the range of disparities present in the depth map. A larger range indicates a wider depth coverage and better ability to capture both near and far objects. Its value is in the range of 0 to 255.
For this image its 255.

Comparing with ground truth or DL based results

If ground truth is available to us then we can find the difference between the ground truth and our depth map using metrics like 2 norm or we can compare our results with some highly accurate DL results.

ANALYSIS OF FAILURE CASES

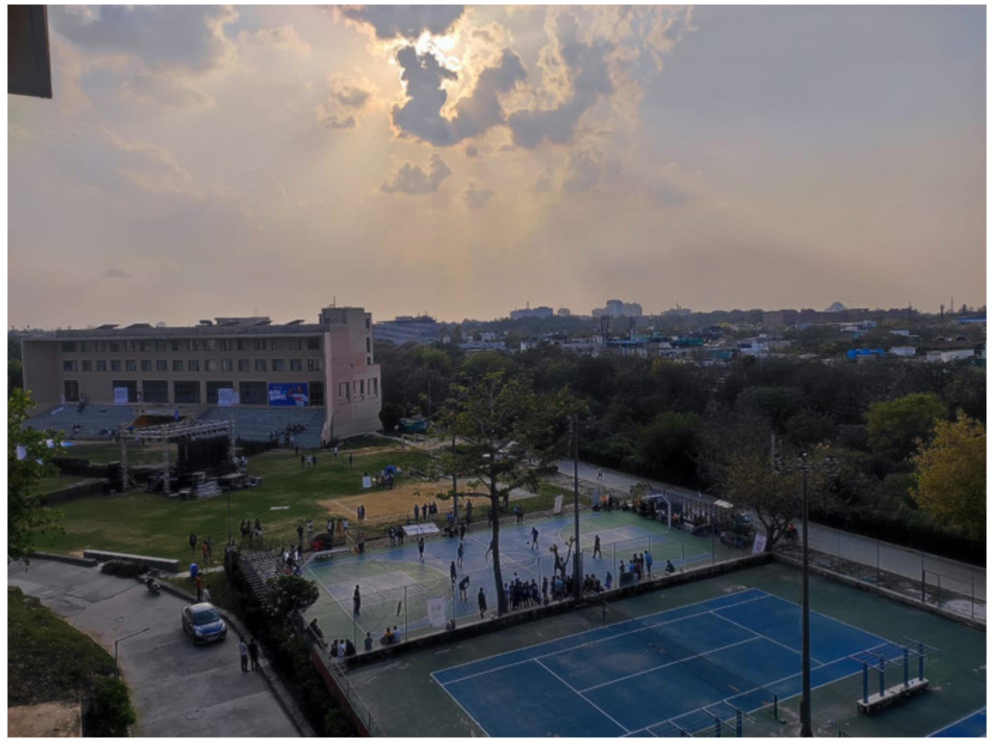
Reflective or textureless surfaces: Reflective or textureless surfaces can cause difficulties in feature extraction and matching, leading to inaccurate fundamental matrix estimation and depth estimation.
Sky, plain wall and similar spaces in images dont lead to good results

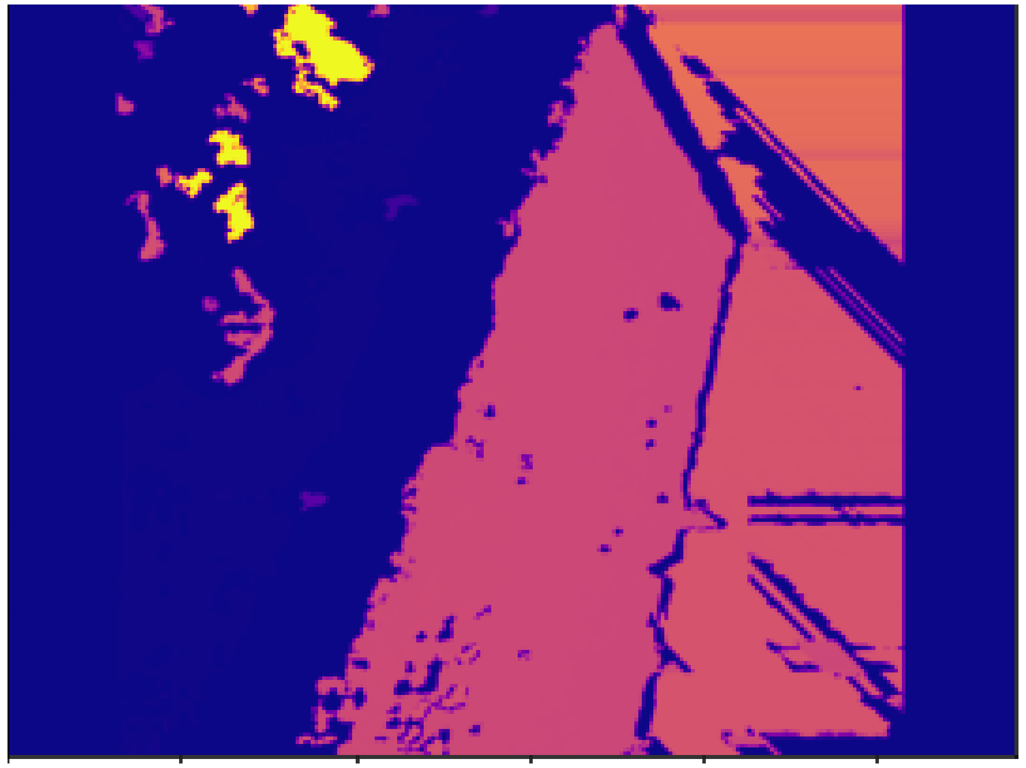
Images which are too bright. Overexposure in the images can lead to noise, which can affect the performance of stereo vision. When an image is overexposed there can be loss of detail and information.

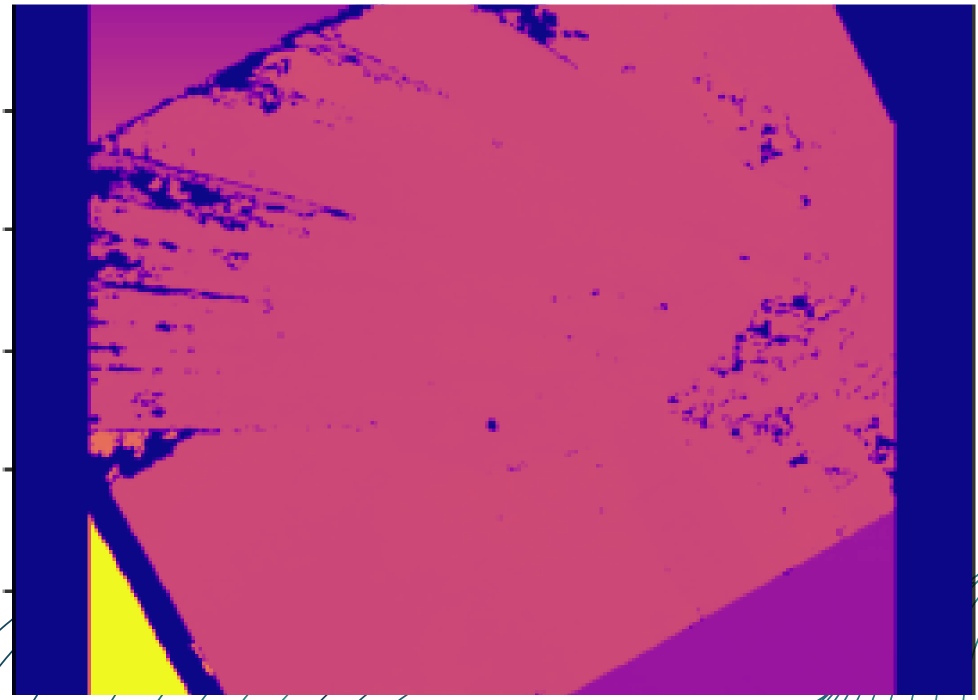
Objects like trees, vehicles and humans don't get detected properly, reasons

1. Occlusion: Trees can have complex structures with overlapping branches, leaves, and other elements.
2. Thin structures:
3. Depth discontinuities

Failure due to trees







Failure due to human beings

Individual Contributions

Pritish Poswal - Creating real world dataset and finding existing dataset, improving two view stereo method code and making presenation

Vibhor Agarwal - Finding relevant work already present, writing base code for two view stereo approach and DL based approach

Ishita Sindhwani - Trying and testing variations in two view approach, evaluating the results,making presentation