# STRUCTURE FROM MOTION

Description of the program in StructureFromMotion.py

1. Deal with the command line arguments as stated in the README.txt for images from different views.
2. Import the specified packages from python 3.5 and OpenCV 3.1.
3. Read the input images.
4. Convert the images to gray scale as SIFT features need gray scale images for calculation of descriptors and the specified key points.
5. Create the SIFT features and store the local key points and the descriptors for both the images which will be utilized for matching.
6. BF matcher is used to match the key points and the descriptors of the specific images to each other.
7. Store all the good matches as per Lowe's ratio test with 0.7 being the threshold separation.
8. Sort these good matches in ascending order of distances for best results.
9. Draw matching lines between feature key points using the *drawkeypoints* function in OpenCV.
10. Compute the intrinsic matrix and store it as 3x3 array for both camera views.
11. It is to be noted that the intrinsic matrix is same for both the cameras used in this case and the camera intrinsic matrix is given by:

$$A = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

    where fx and fy are the focal lengths and cx and cy are the principal points.
12. Remove barrel and similar non-linear distortions for higher accuracy along with normalization to remove the intrinsic camera dependency by undistorting the points obtained after the good matching is done.
13. Use RANSAC while computing the essential matrix from the undistorted points. It is suggested having higher threshold and high probability to obtaining maximum inliers in the system of matching.

14. SVD is needed to be carried out to obtain the rotation and translation matrix by the *recoverpose* function for the undistorted points.
15. Since camera 1 is considered to be world axis aligned, the rotation matrix amounts to identity matrix with no translation.
16. Extrinsic matrix for the second camera is obtained by horizontally stacking the rotation and translation matrix obtained after SVD.
17. Final projection matrices of both camera views is of dimension 3x4 and is the dot product of the intrinsic matrix and the extrinsic matrices of the respective cameras.
18. Triangulation is the last step carried out in this process which does the 3d reconstruction after the camera parameters and the image coordinates are known.
19. For effective triangulation, normalized input points are needed to be fed with strong correlation and low outlier percentage for robust 3d reconstruction.
20. It is important to note that the output is in homogenous coordinates which needs to be normalized and stored as non-homogenous vertices for scatter plotting the output.
21. Scatter plot the output with different views for best results and effective reconstruction.

<u>Obtained Results from the above steps</u>

Output of the computed SIFT matches and values consistent with the computed essential matrix along with SVD based rotation and translation matrices calculated with 0.7 Lowe's ratio.

Rank 2 verifies that accuracy of the computed essential matrix

```
The total number of good matches found are:  1206
The total number of consistent value inliers with the essential matrix
are:  1206
The Essential matrix found is as follows:
[[ 0.00546802 -0.0531293   0.01808229]
 [-0.20464163 -0.03151648 -0.67591017]
 [-0.00866352  0.70440328 -0.02890205]]
matrix rank 2
The Rotation matrix found is as follows:
[[ 0.93227329 -0.02683345 -0.36075821]
 [ 0.00960081  0.99872918 -0.04947573]
 [ 0.36162735  0.04266133  0.93134616]]
The translation vector found is as follows:
[[ 0.99681534]
 [ 0.02340763]
 [ 0.07623165]]
```

Matches obtained after Lowe's ratio test on the SIFT points calculated in both the images.

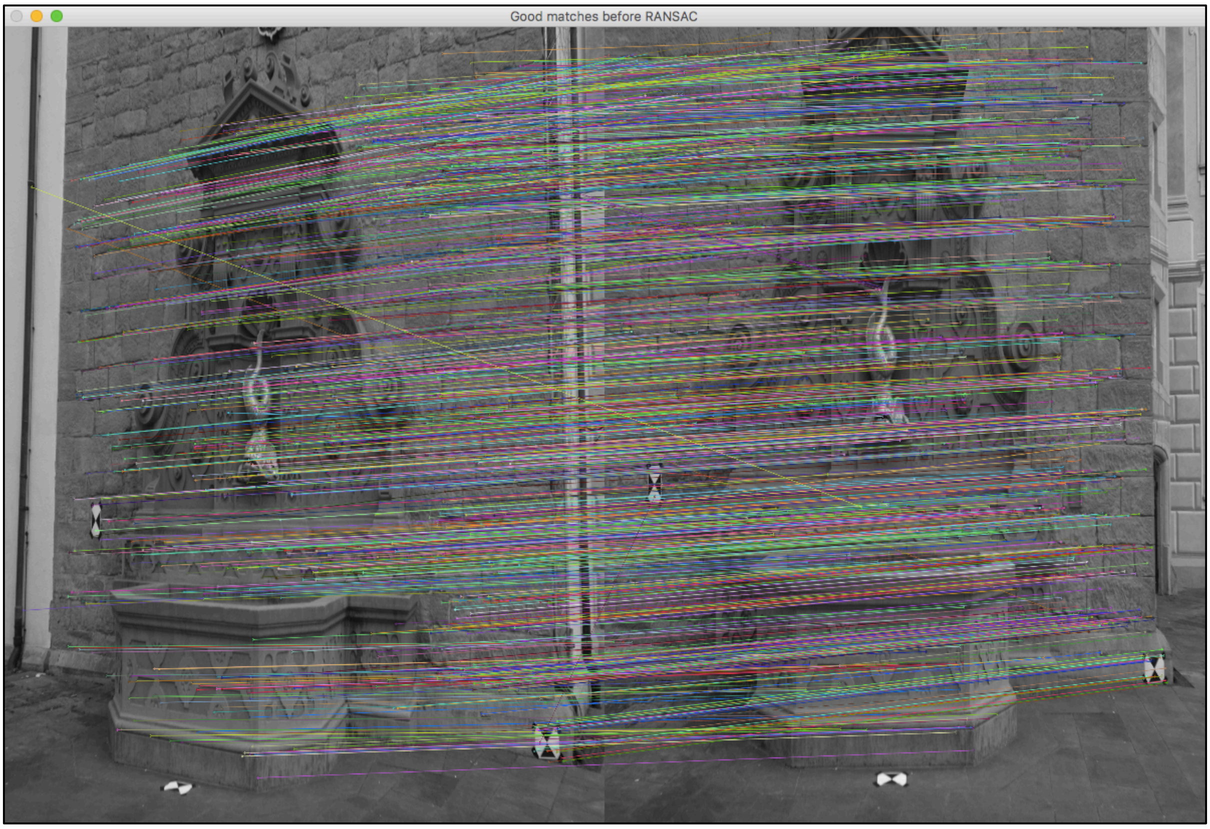


Figure 1: Bf matcher result to be fed for estimation of essential matrix.

## CASE 1: 3D RECONSTRUCTION USING ALL GOOD MATCHES AT 0.7 LOWE'S RATIO WITH 1206 MATCHES.

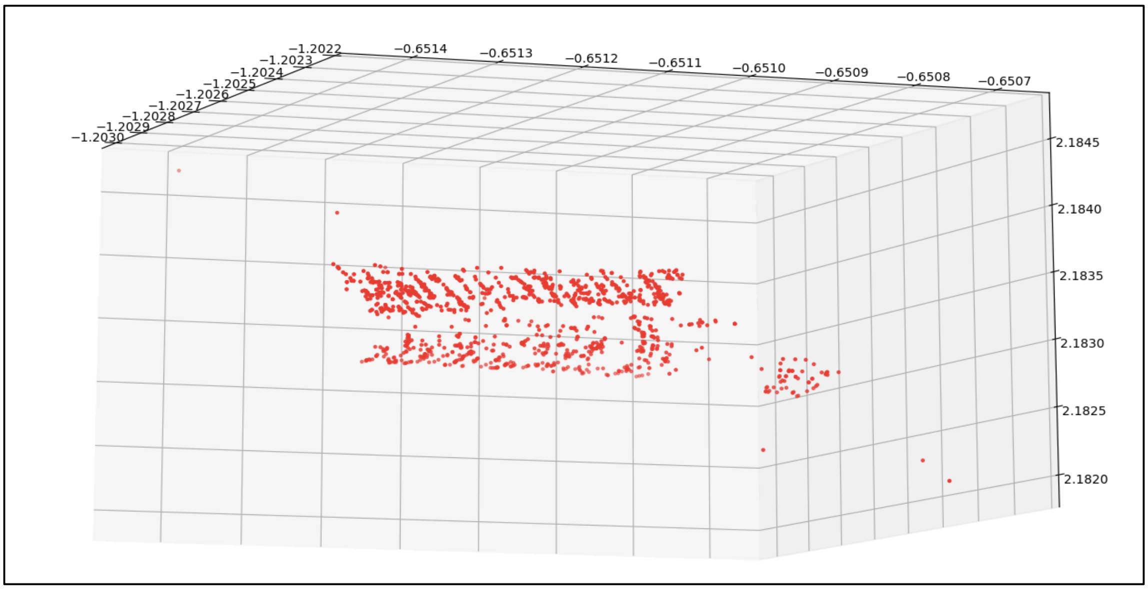


Figure 2: Side view of reconstructed 3d points

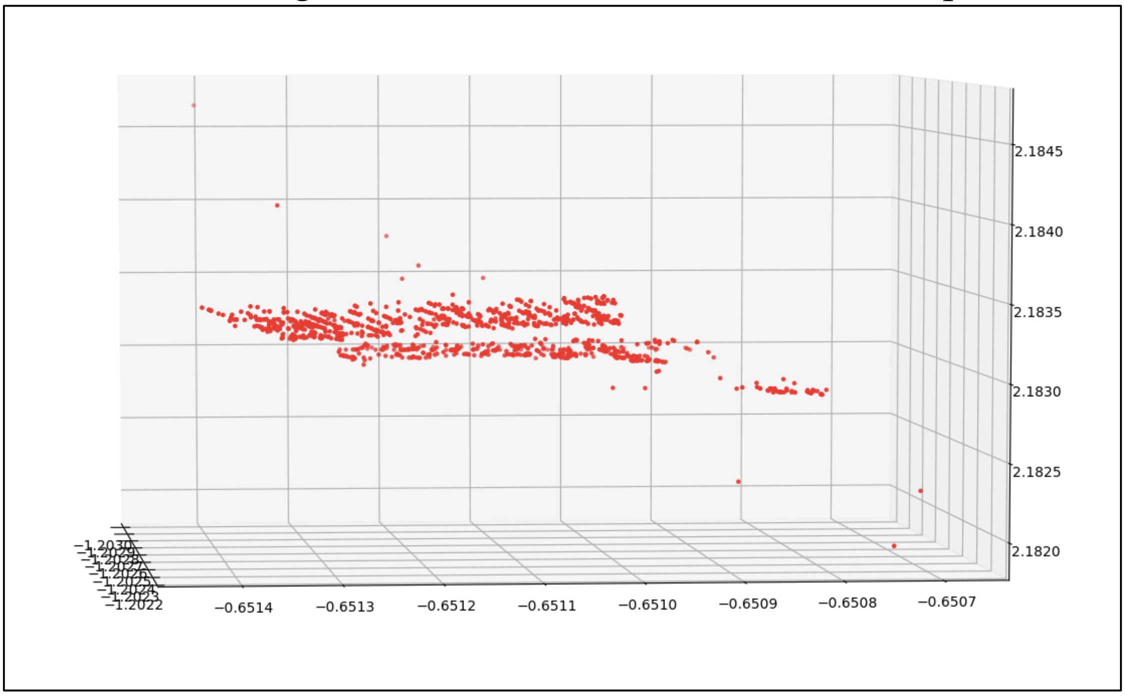


Figure 3: Side view of reconstructed 3d points clearing showing 3 separate depths of the wall, fountain and the floor.

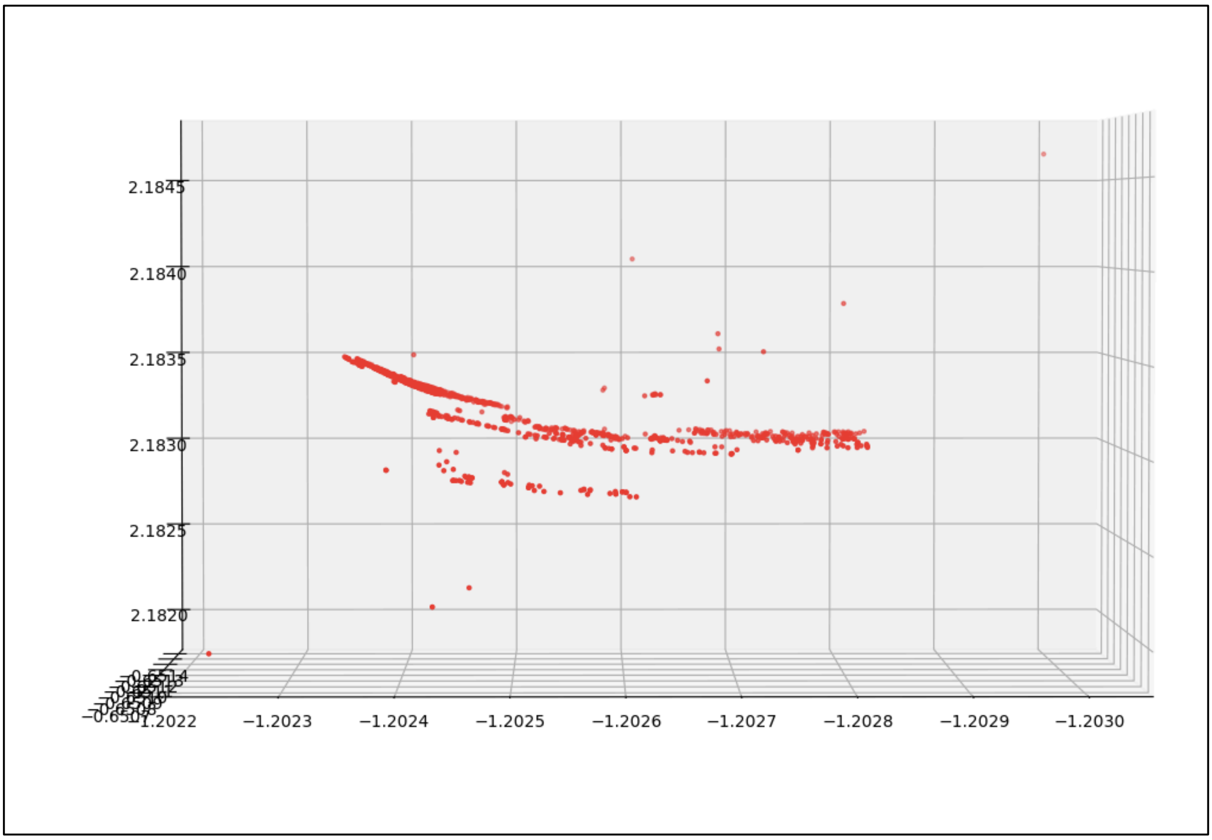


Figure 4: Bottom view of reconstructed 3d points

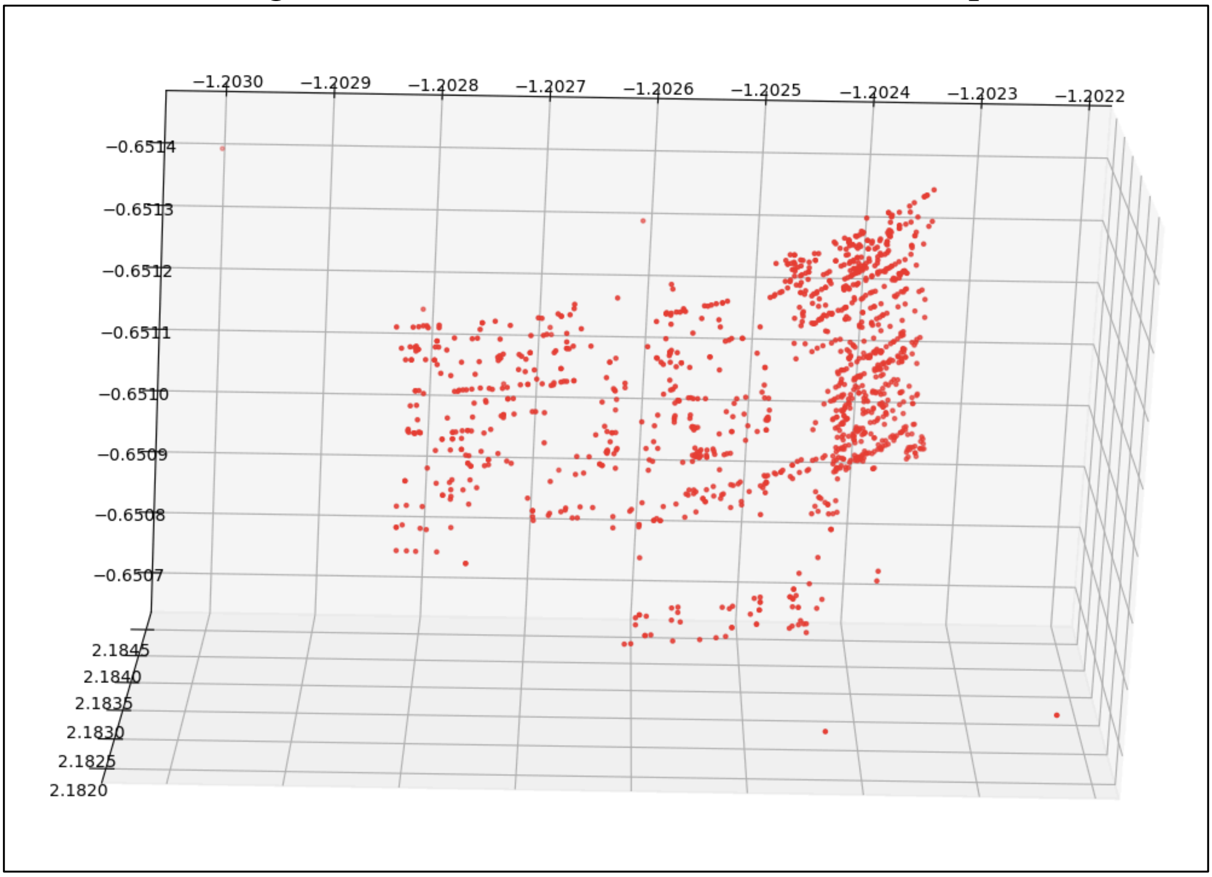


Figure 5: Left view of reconstructed 3d points

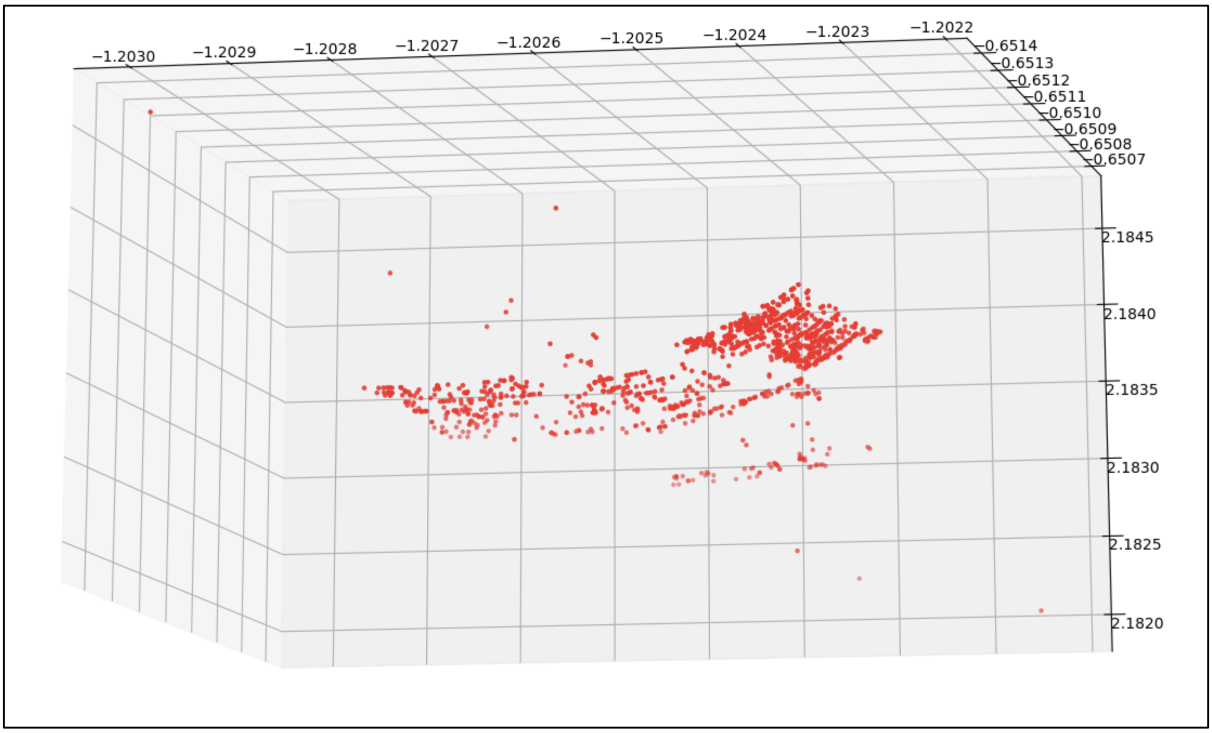


Figure 6: Top view of reconstructed 3d points

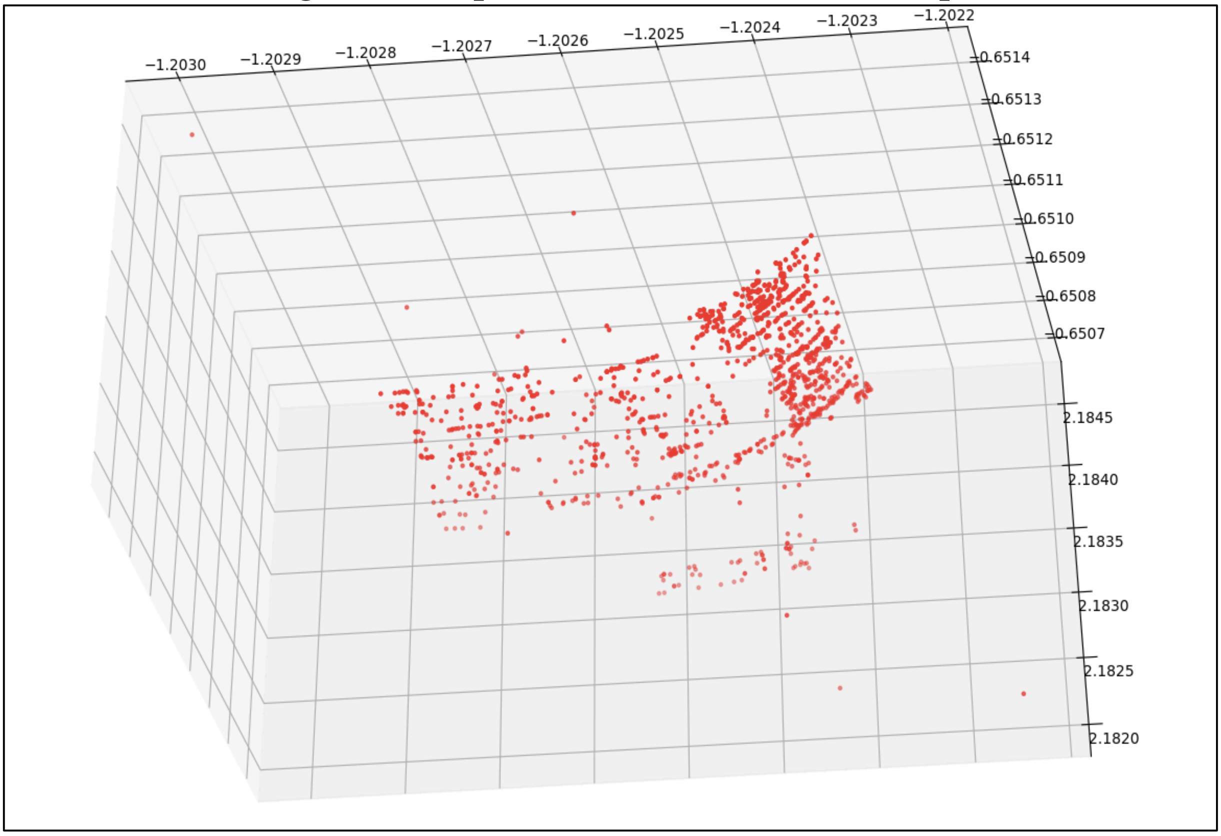


Figure 7: Top view of reconstructed 3d points

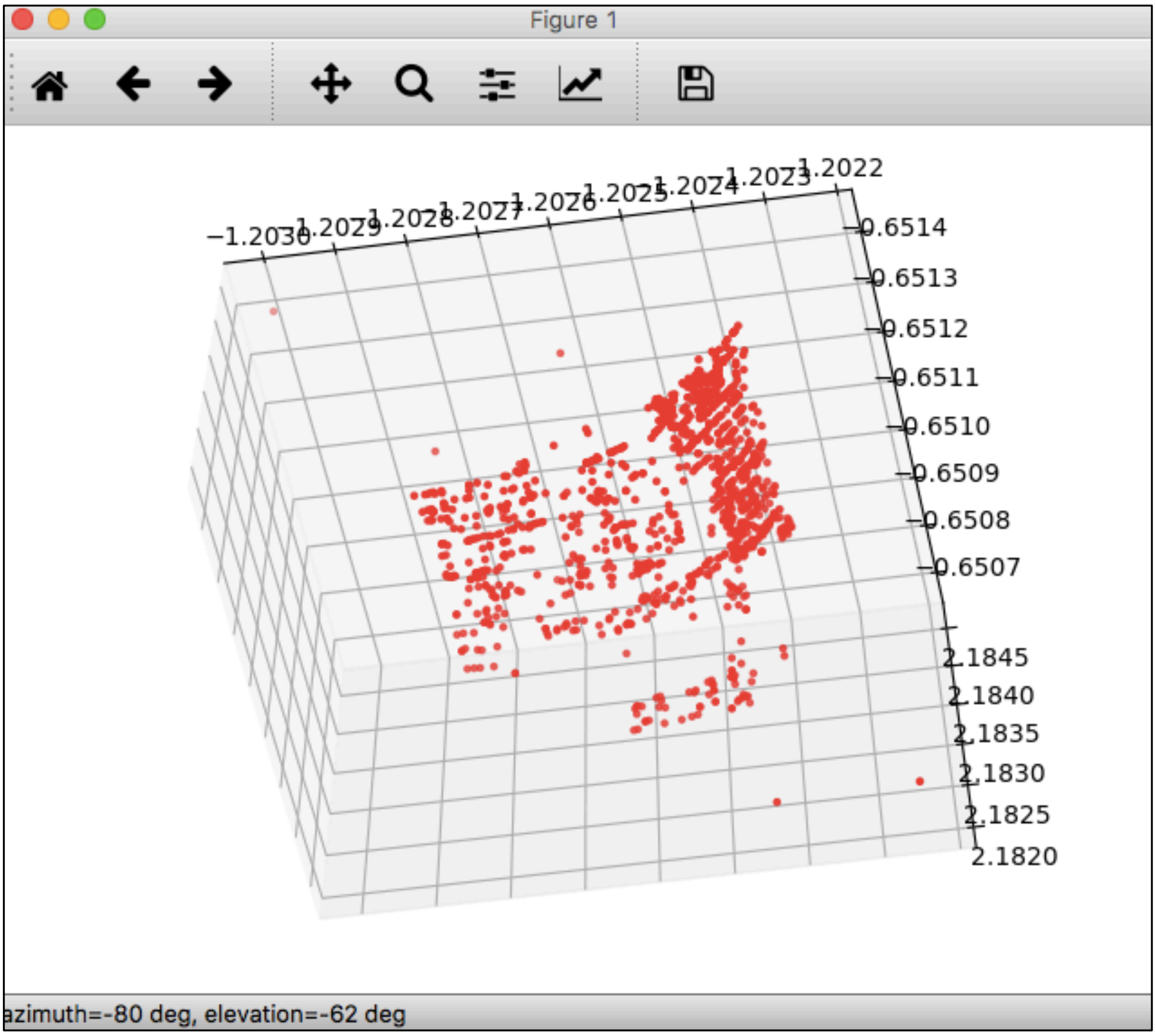


Figure 8: Front view of reconstructed 3d points

## CASE 2: 3D RECONSTRUCTION USING ALL GOOD MATCHES AT 0.8 LOWE'S RATIO WITH 1446 MATCHES.

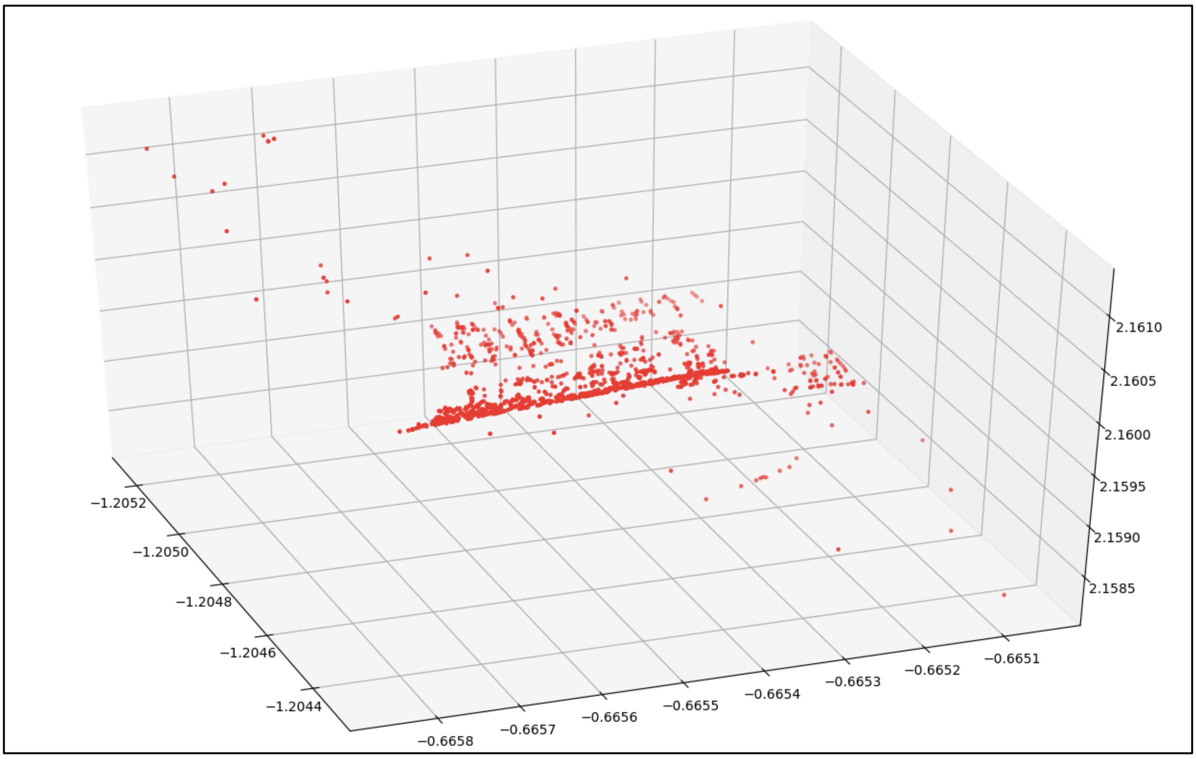


Figure 9: Left view of reconstructed 3d points

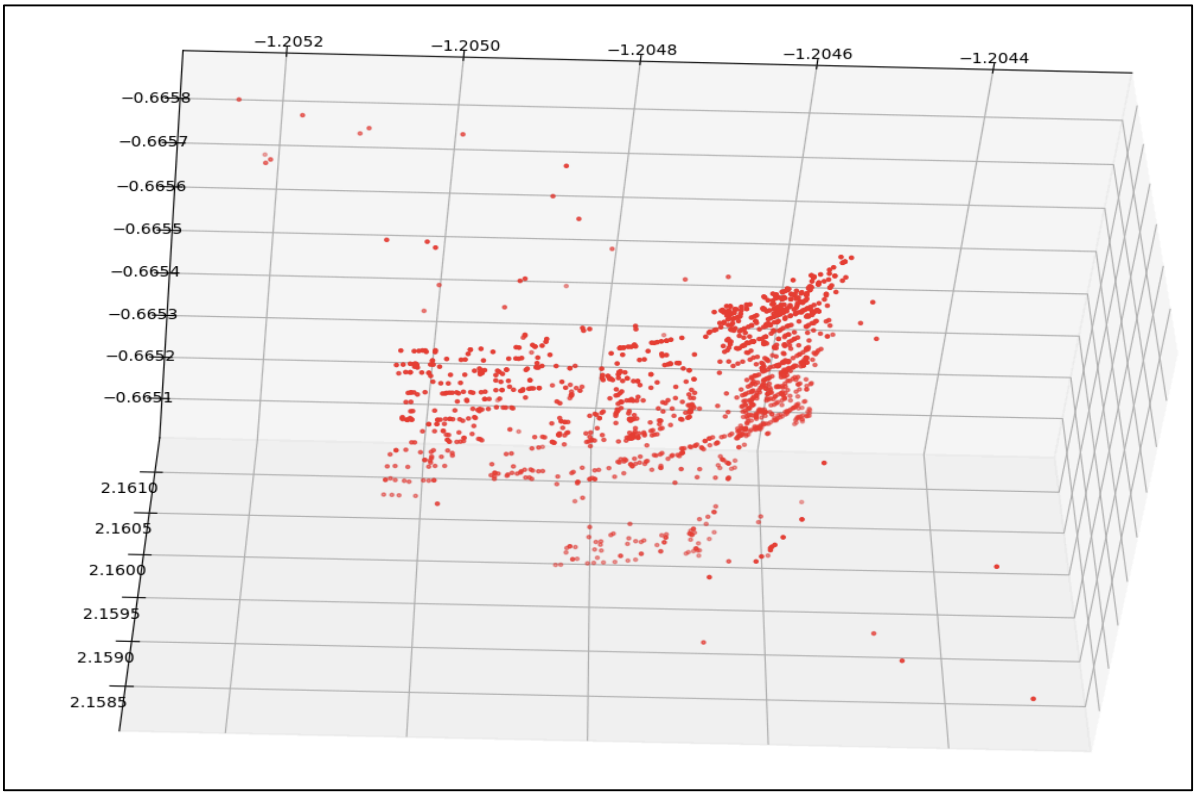


Figure 10: Front view of reconstructed 3d points

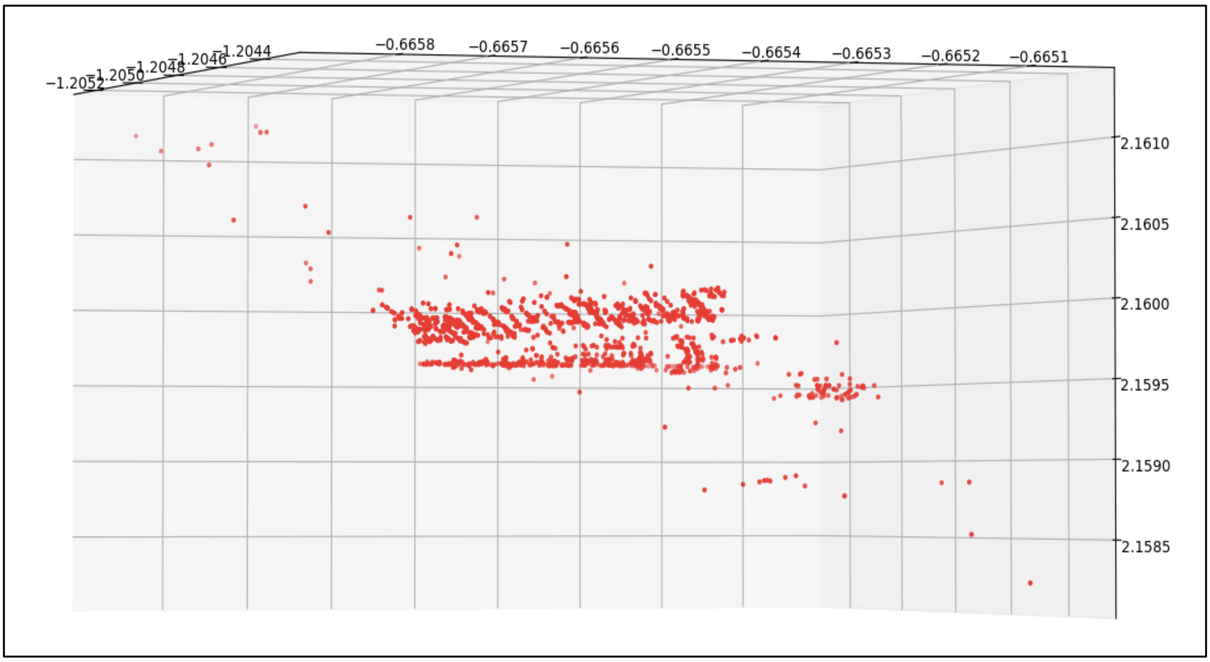


Figure 11: Side view of reconstructed 3d points

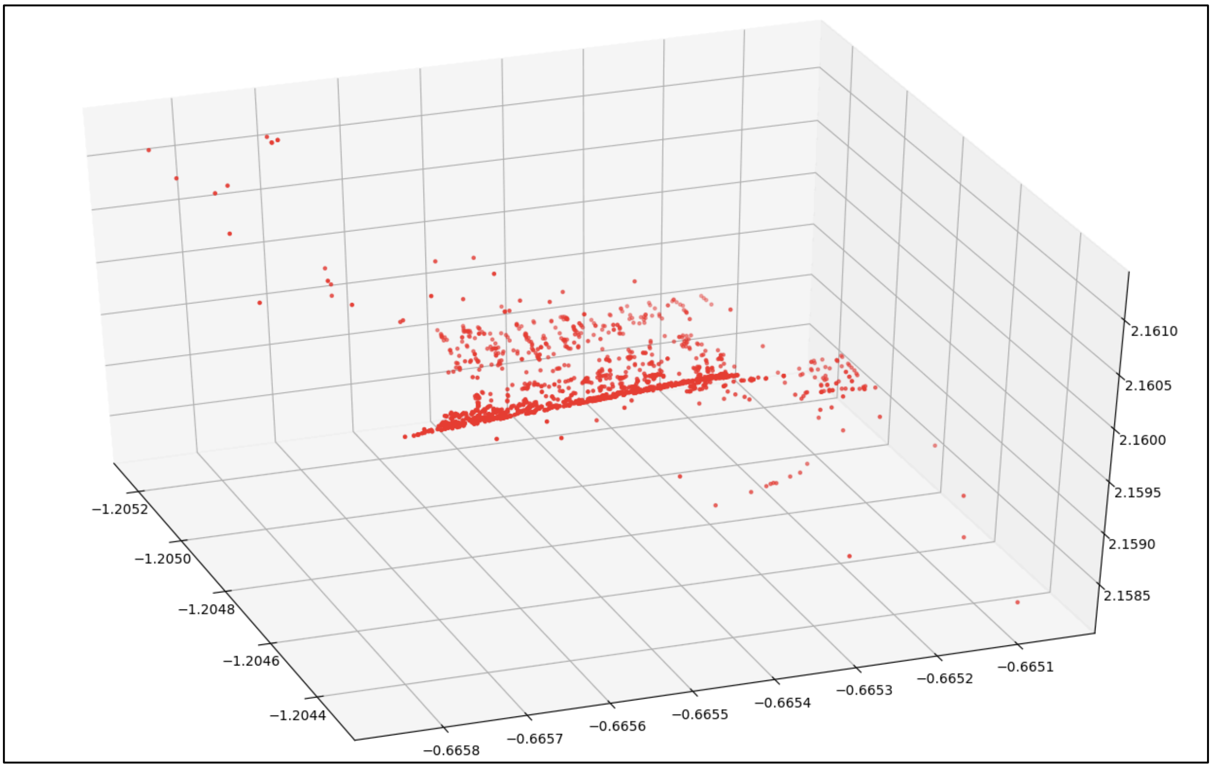


Figure 12: Side view of reconstructed 3d points

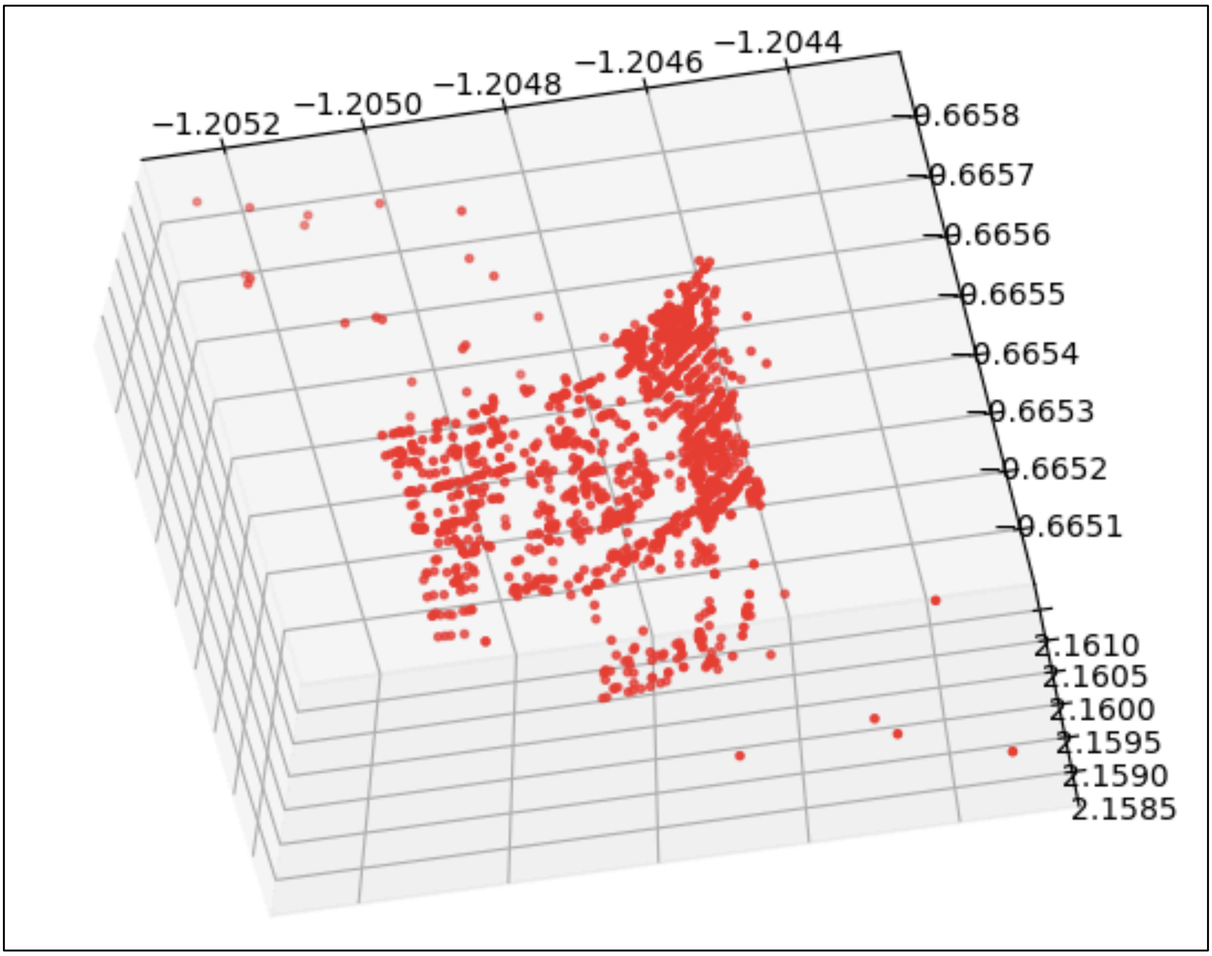


Figure 13: Front view of reconstructed 3d points

## CASE 3: 3D RECONSTRUCTION USING TOP 1000 GOOD MATCHES AT 0.7 LOWE'S RATIO.

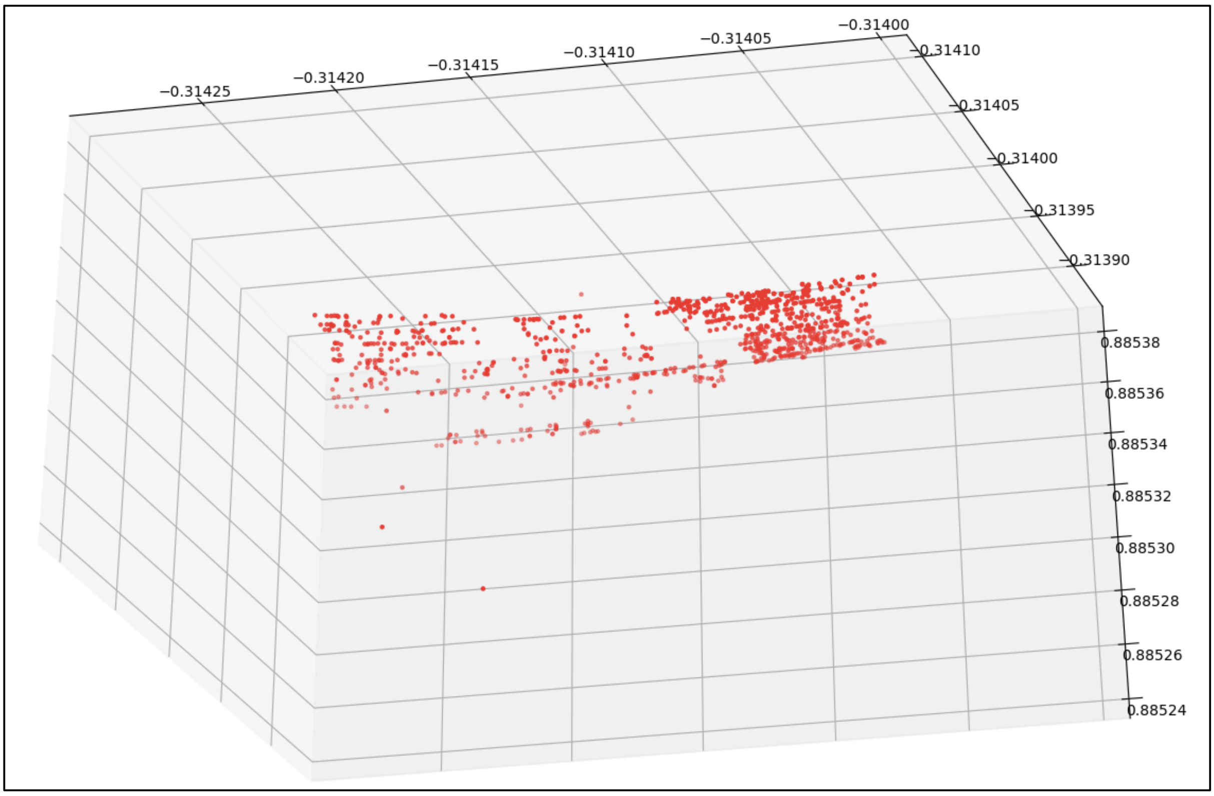


Figure 14: Top view of reconstructed 3d points

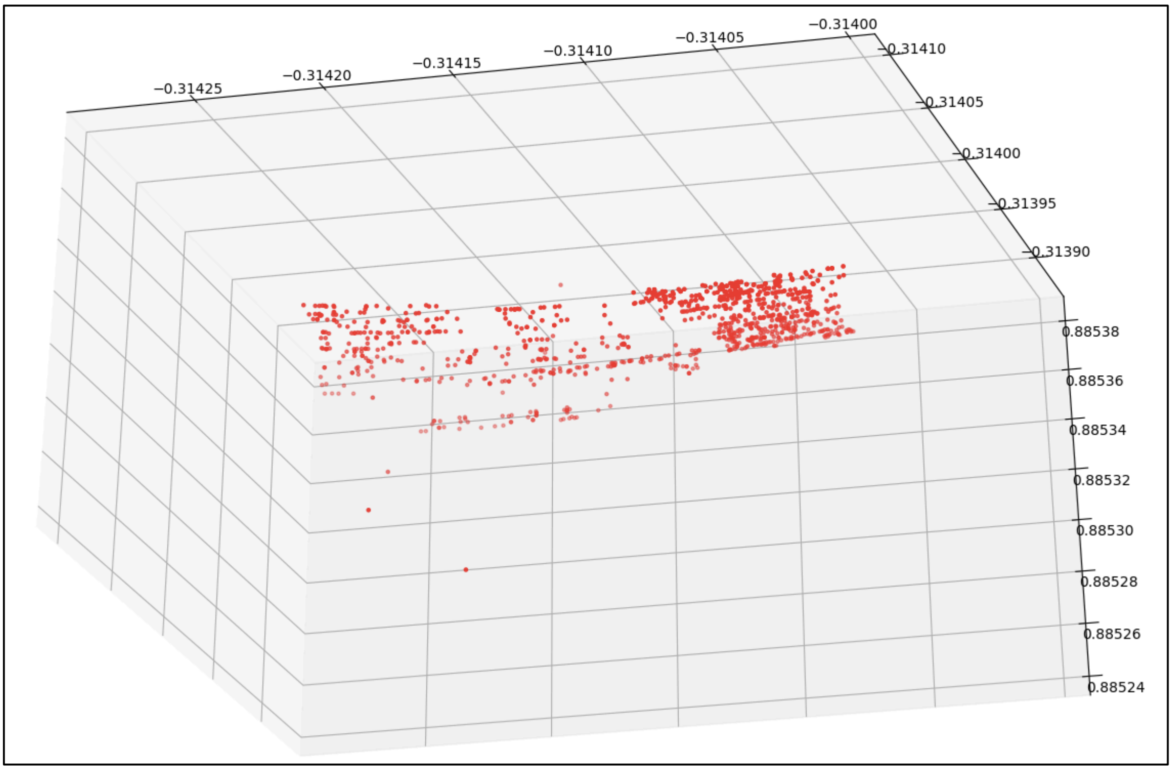


Figure 15: top view of reconstructed 3d points with 1000 good points

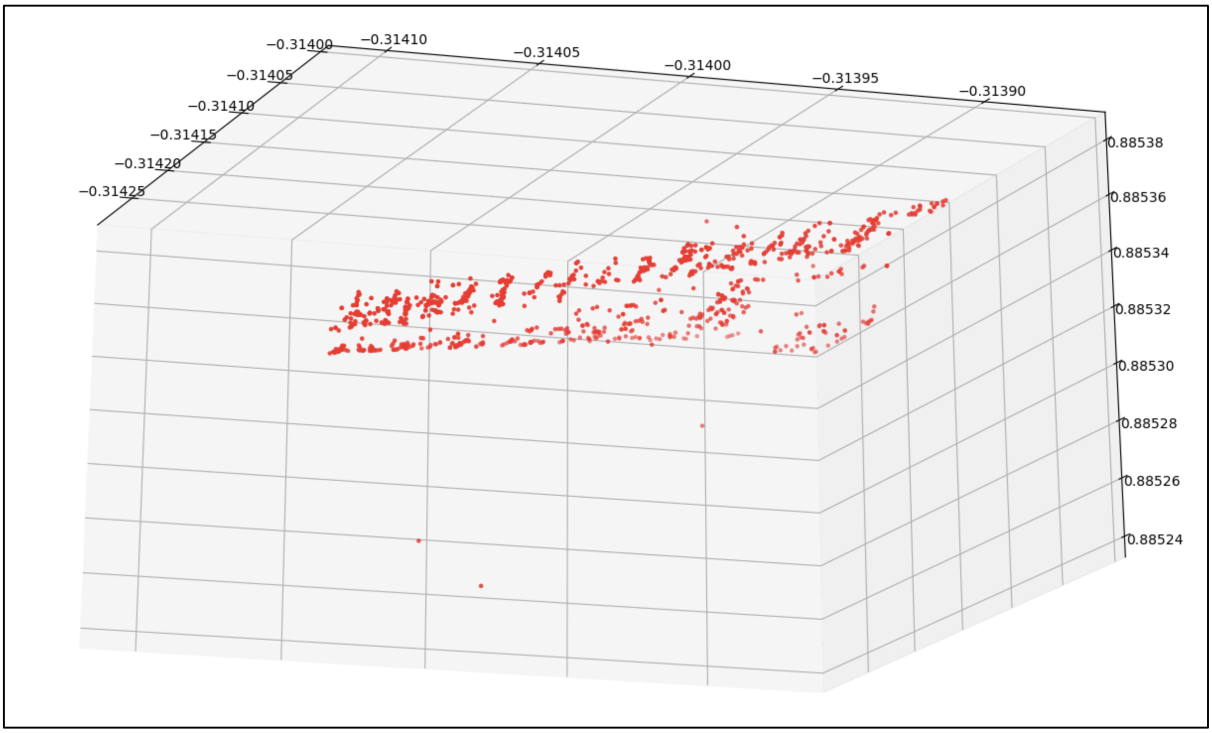


Figure 16: Side view of reconstructed 3d points

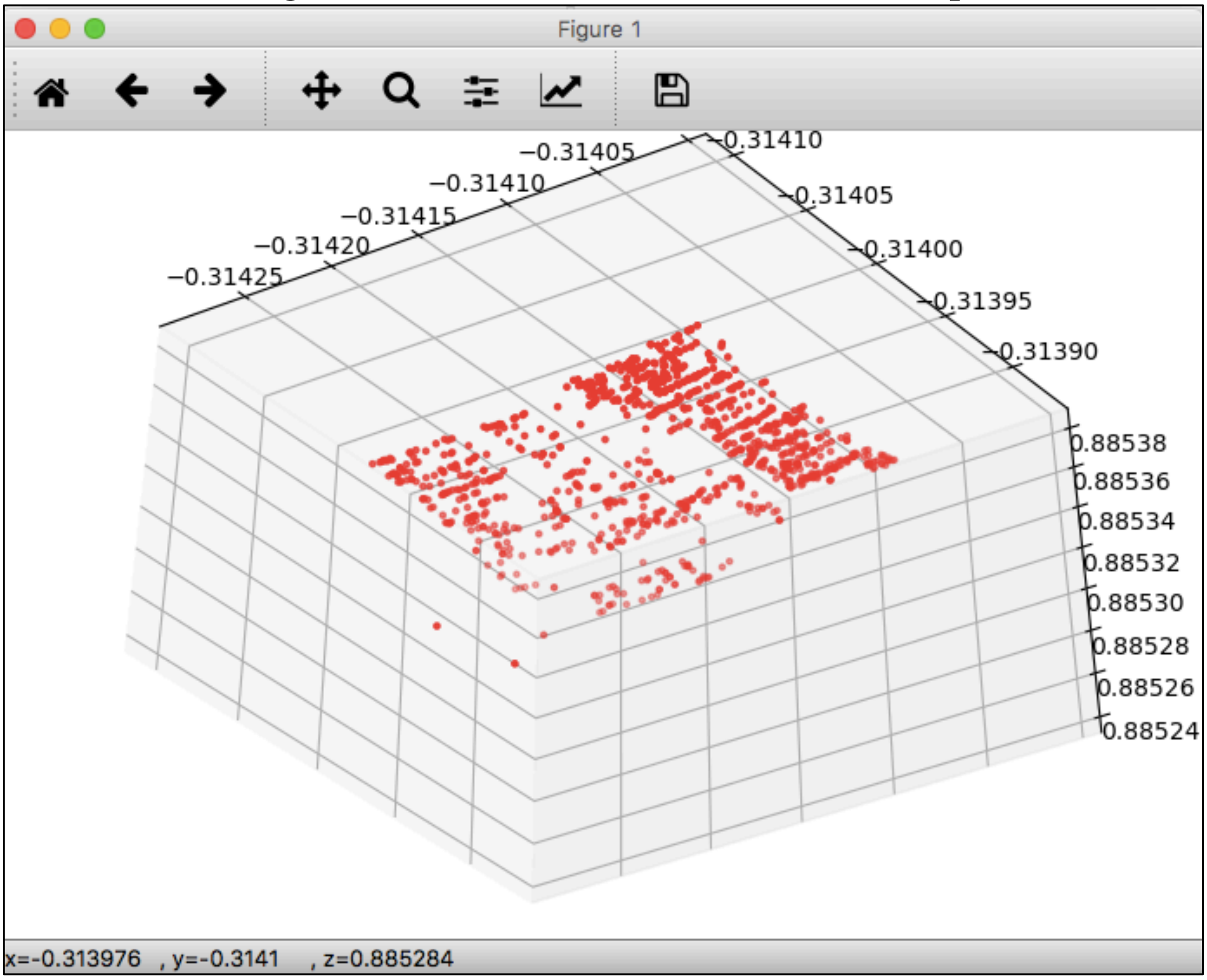


Figure 17: Front view of reconstructed 3d points

## Analysis of the obtained Results from the above steps

The above results clearly indicate that the obtained essential matrix is of rank 2 and can be used to compute the extrinsic matrix of the second camera view using SVD. The 3D reconstruction is possible by then using these camera matrices and image points efficiently.

Three cases are described in the above results with Figure 1displaying the corresponding matches found out between the different camera views followed by the 3 cases.

From Case 1, figures 2 through 8 define the different views of the reconstructed 3D points with Lowe's ratio of 0.7 which results in 1206 samples with all of them being inliers, hence the construction is robust enough. The different views elaborate the walls, floor and fountain separately as 3 lines. With the said focal length and principal point, unless bundle adjustment is carried out multiple times, the result is not highly accurate. The result could be better if bundle adjustment along with better inlier count is obtained. Since lesser SIFT points with only 2 views and lesser bundle adjustment and correction of error is carried out, exact reconstruction is not possible.

From Case2, figures 9 through 13 define the different views of the reconstructed 3D points with Lowe's ratio of 0.8 which results in 1446 good matches but larger outliers. After reconstruction as seen in the figures, denser point cloud is obtained but the number of outliers also increase which adds on to the inaccuracy. Hence, we can conclude that according empirical finding 0.7 is the ideal ratio and SIFT matches with this ratio help in obtaining the optimal results.

From Case 3, figures 14 through 17 define the different views of the reconstructed 3D points with Lowe's ratio of 0.7 with top1000 good matches. It is a sparse map of the matches which shows that predominantly the background wall matches are of high stability than the fountain or the floor while reconstruction. It is to be noted that these top1000 matches are based on the minimal distance between the key points. Hence larger density is concentrated towards the walls in the background. The later 206 points account largely to the foreground objects.

Hence using all the cases we can conclude that 0.7 is the ideal ratio for reconstruction along with sorted good matches showing the robustness of the matches between different matched objects. Better reconstruction is possible with more camera views and larger bundle adjustment for eliminating the error created while matching only 2 views.