Computer Vision Coursework

Q1: Propose a technique to detect salient features of your choice on the video frames above. Explain the type of features on which you will focus and justify your choice.

In this project, the Harris Corner Detector is employed to identify salient features, specifically corners, in the video frames. The Harris Corner Detector is trustworthy for its efficiency in detecting corners, which are points in an image where intensity varies significantly in multiple directions. To extract these corners, the detectHarrisFeatures() function will be employed.

The rationale behind focusing on corners as salient features is their invariance to various transformations. They exhibit invariance to rotation and scale changes, as well as robustness under different illumination. Furthermore, Corners are unique in their surroundings, making them reliable for tracking across different frames in a video sequence. These attributes make corners particularly suitable for feature matching tasks, ensuring consistent performance even under diverse image conditions.

Q2: Propose a technique to match the detected salient features between the video frames. Explain how you would approach this task and the steps you would follow.

In this approach, the Second Nearest Neighbor (SNN) method is employed to match each feature in Frame 1 with its closest and second closest features in Frame 2. This method is based on the concept of Nearest Neighbor Distance Ratio (NNDR), which helps in eliminating ambiguous matches. According to this principle, a match is deemed ambiguous and therefore discarded if the closest match is not significantly better than the second-best match, specifically if it does not exceed a set threshold. In our case, this threshold is established at 0.75.

The process begins by extracting feature descriptors from Frames 1 and 2 using the extractFeatures() function. To implement the Second Nearest Neighbor method, the knnsearch() function is used. This function identifies the two nearest neighbors in Frame 2 for each feature in Frame 1, based on the Euclidean distance. If the distance to the best match is not at least 75% smaller than that to the second-best match, the feature is considered ambiguously matched and is thus excluded. This method significantly enhances the reliability of the feature matching process, by ensuring that only the most distinct matches are considered.

Q3: Programming environment is MATLAB (a-d) and OpenCV(e):

a. Implement your proposed salient feature detector and plot the detected features on the provided pair of frames.

<u>Frame 1:</u>

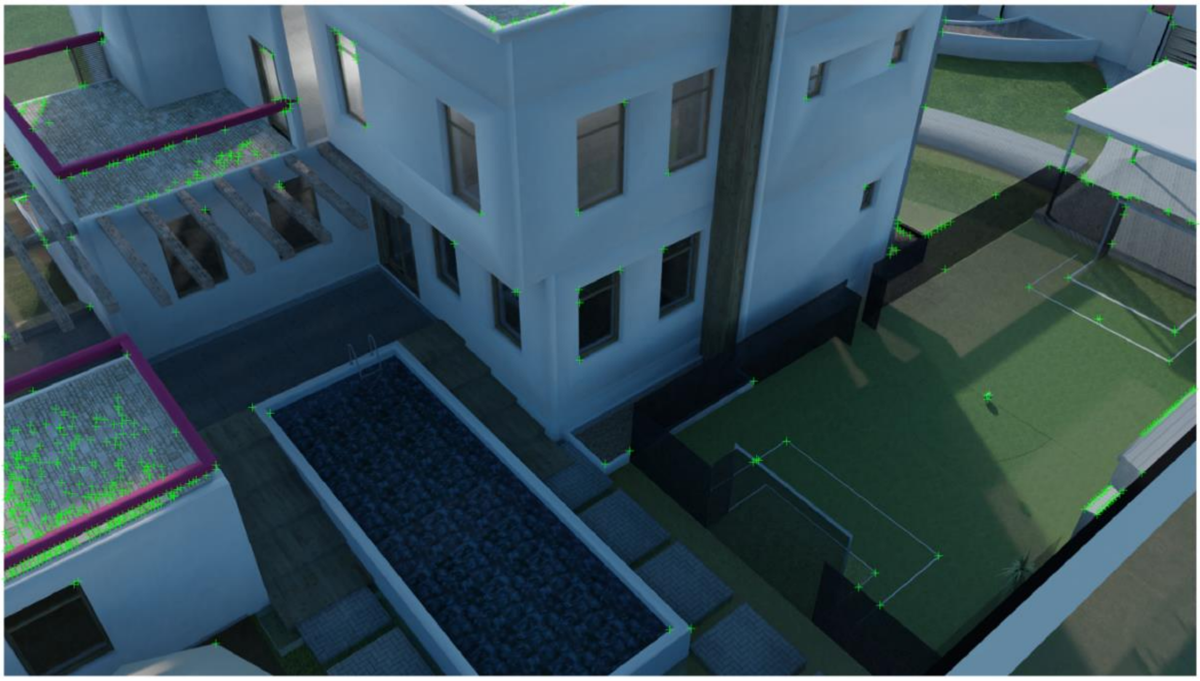

Figure 1 All Corners Detected by Harris Corner Detector in Frame 1

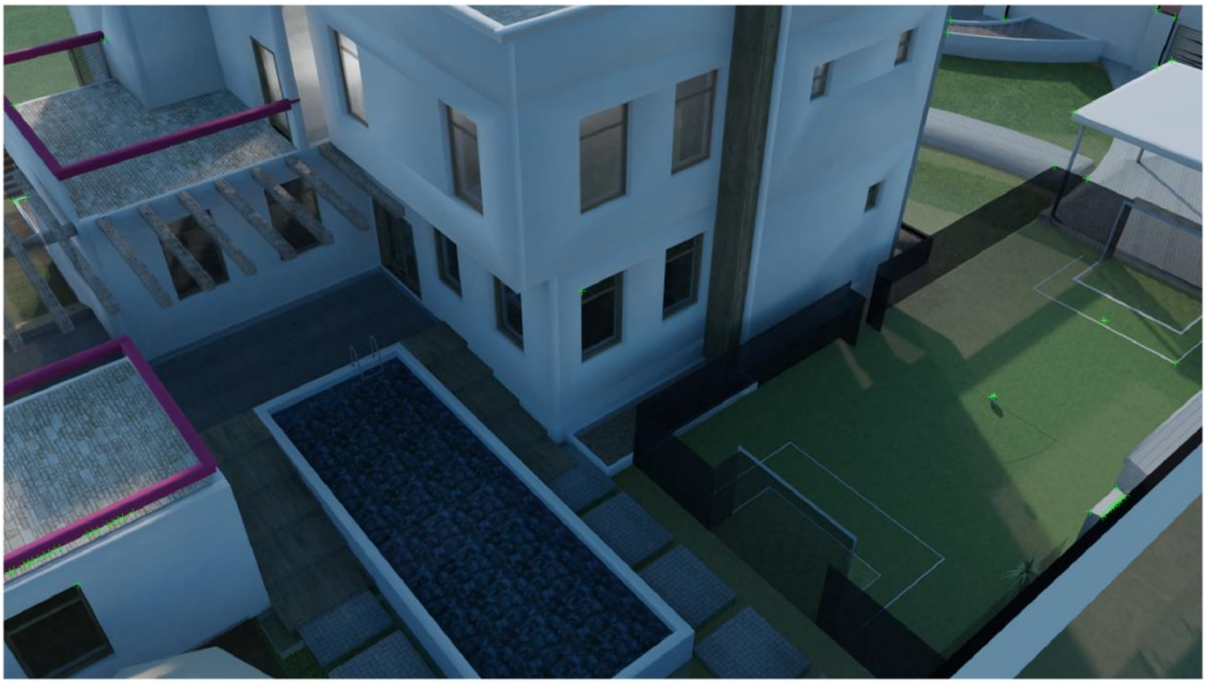

Figure 2 The Strongest 50 Corners Detected by Harris Corner Detector in Frame 1

Frame 2:

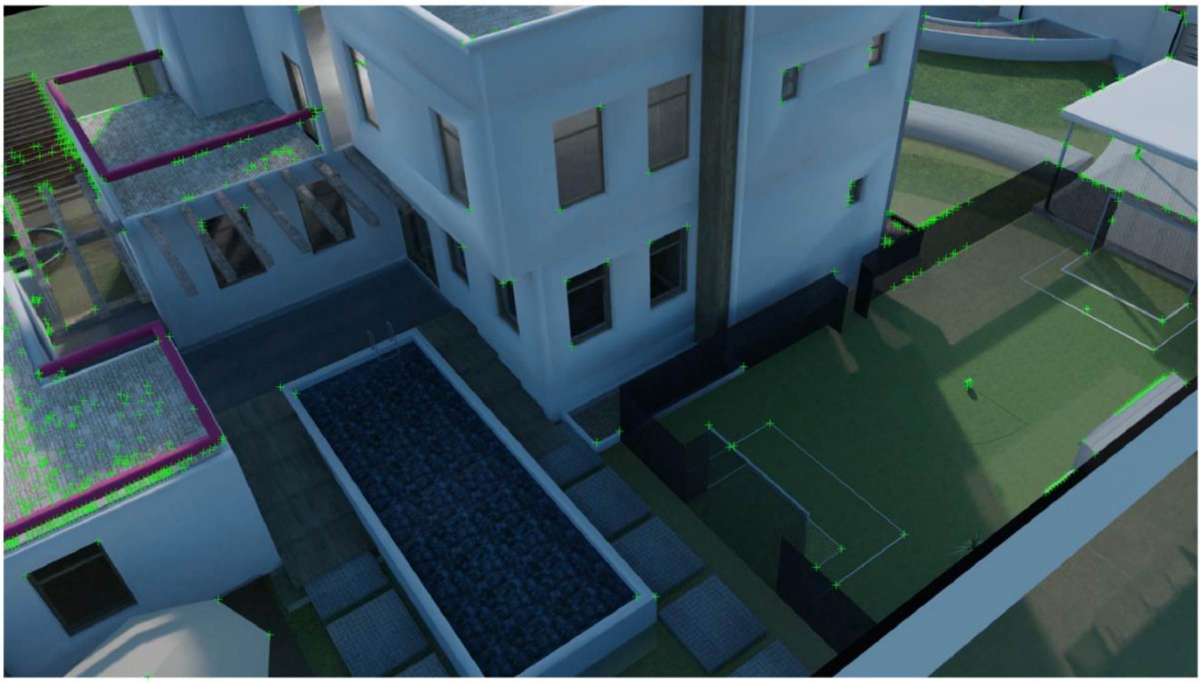

Figure 3 All Corners Detected by Harris Corner Detector in Frame 2

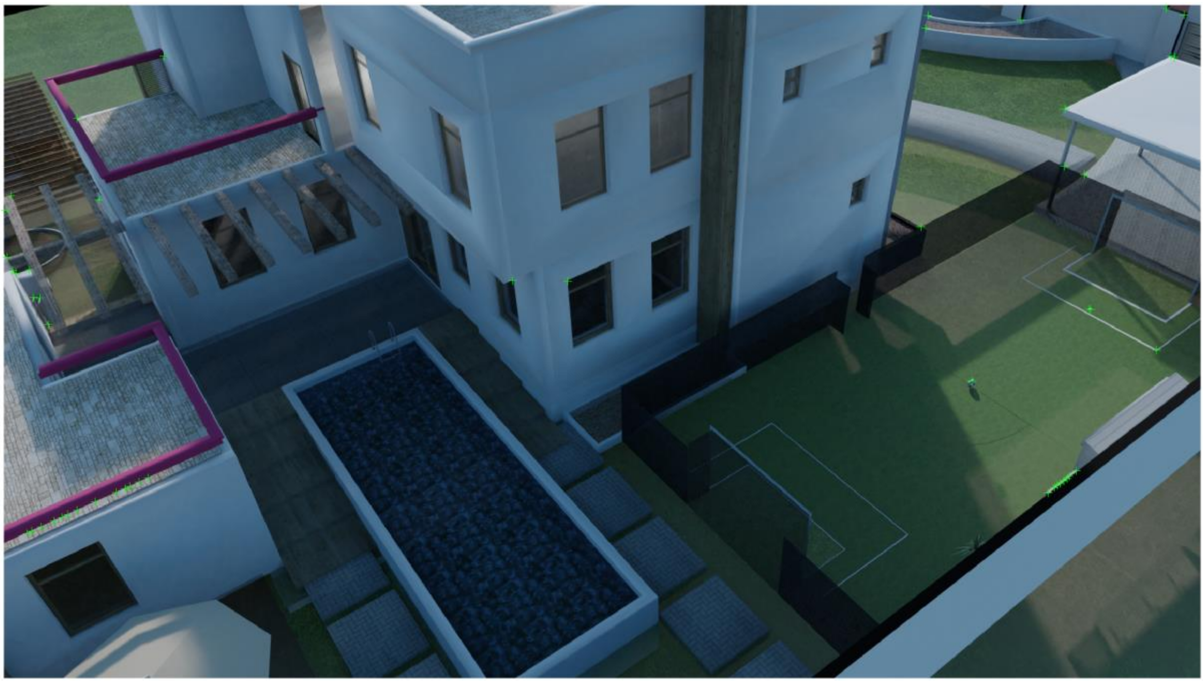

Figure 4 The Strongest 50 Corners Detected by Harris Corner Detector in Frame 2

b. Find corresponding features between the two frames and illustrate those matches. To illustrate the matches you can, for example, create a composite image (e.g centered overlay image) from the two frames.

In this composite image, there are some correct matches corners, but a group of outliers still present in the feature matching.

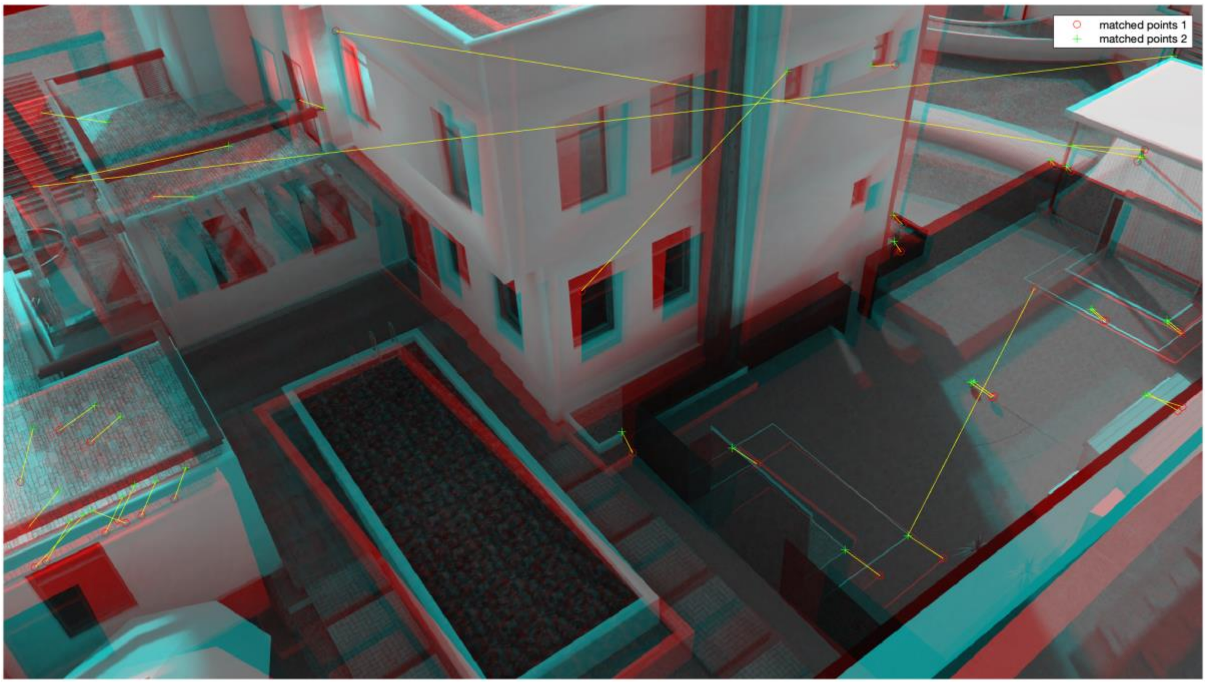

Figure 5 Feature Matching

c. Compare the estimated fundamental matrices and explain any possible disagreement between the two methods. Which method is more accurate? Justify your answer and suggest how you could improve the least accurate method.

The fundamental matrix derived using matched features comes directly from image data, satisfying the epipolar constraint between frames 2 and 1. RANSAC is used to mitigate outliers, but this estimation is still sensitive to them, heavily depending on the quality of feature matching.

When using known camera parameters, the estimation is independent of image and feature matching quality. Its accuracy largely dependent on the precision of camera calibration.

In this case, the difference between the fundamental matrices Frobenius norm is only 0.001861, indicating minimal discrepancy between the two methods. Even though there isn’t a definitive answer as to which method yields a better result, in this instance, the estimation using camera calibration might be worse. These given parameters may not be derived from a large number of calibration images covering sufficient viewpoints. The fundamental matrix yields better result because SNN with NNDR, and RANSAC are utilized here to facilitate accuracy.

d. Find the correctly matched points that meet the epipolar constraint and illustrate these matches. Briefly explain how these matches have been identified.

By using the fundamental matrix computed from the matched feature points, the estimateFundamentalMatrix() function simultaneously provides inliers that indicate the matches satisfying the epipolar constraint. Points fulfilling this constraint will lie along their corresponding epipolar lines in the other image.

These inliers are identified using the RANSAC method. RANSAC operates by iteratively selecting random subsets of matches, computing the fundamental matrix for each subset, and then evaluating all matches against this matrix. Matches are considered inliers if the distance from the points to their respective epipolar lines is minimal. Eventually, the model with the largest set of inliers is selected.

This approach allows for the visualization of only those matched points that are correctly aligned according to the epipolar constraint, as presented in Figure 6.

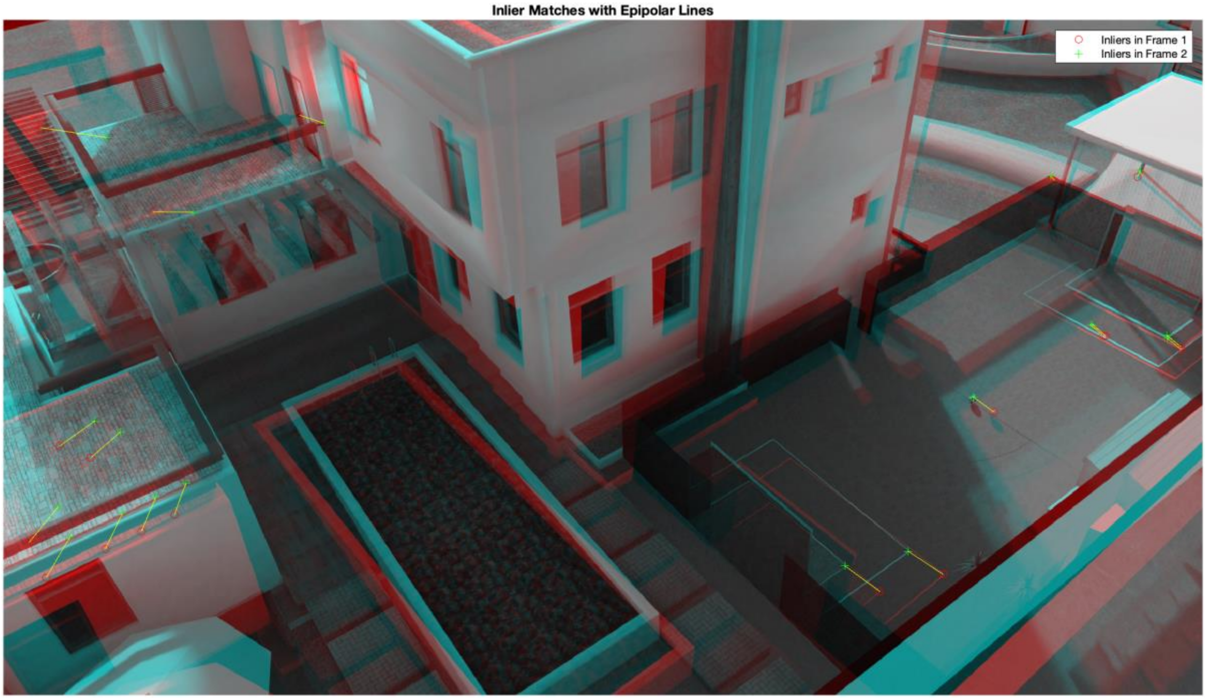

Figure 6 Feature Matching after Epipolar Constraint

e. Estimate the area of the swimming pool and the length (touchline) of the football field. (hint: you can establish the disparity map between these frames or you can apply 3D surface reconstruction)

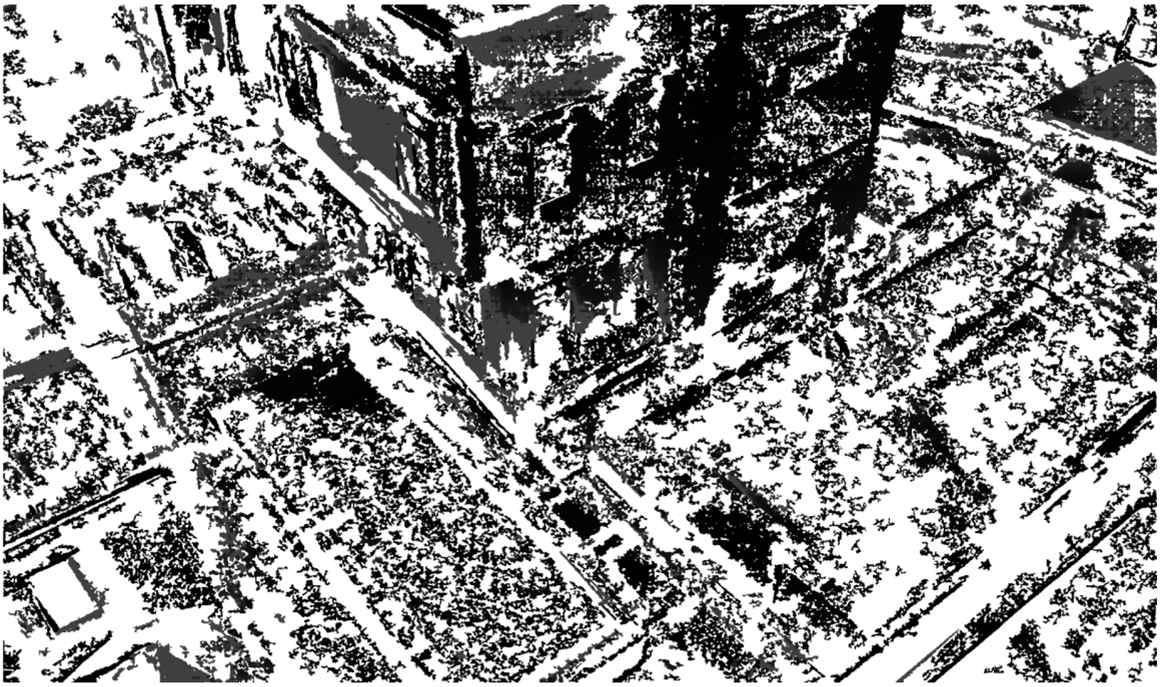

Figure 7 Depth Map Calculated from Disparity Map (Figure 10)

The touchline of the football field is estimated to be 93.569 meters.

The area of the swimming pool is estimated to be 3831.1524914480137 square meters, assuming it is rectangular and calculated by width multiplied by length.

The proposed methodology is that manually selecting two points along the touchline on the depth map. For the swimming pool, three points representing three vertices are selected to obtain the width and length. Once these points are identified, the real-world distance between them can be measured based on the depth values.

The average focal length and principal points are used here because of the stereo matching involving two cameras. To locate the start and end of the touchline, a manual selection function has been implemented. This function reads the depth graph, allowing users to manually select the points they believe are the appropriate borders on the interface, and then stores these points' coordinates automatically.

For the football touchline:

- Selected Point: (X: 1295, Y: 786)
- Selected Point: (X: 1741, Y: 463)

For the swimming pool:

- Selected Point: (X: 421, Y: 642)
- Selected Point: (X: 609, Y: 541)
- Selected Point: (X: 1072, Y: 987)

4. Optional: Illustrate the disparity map and the rectification result for the above video frames.

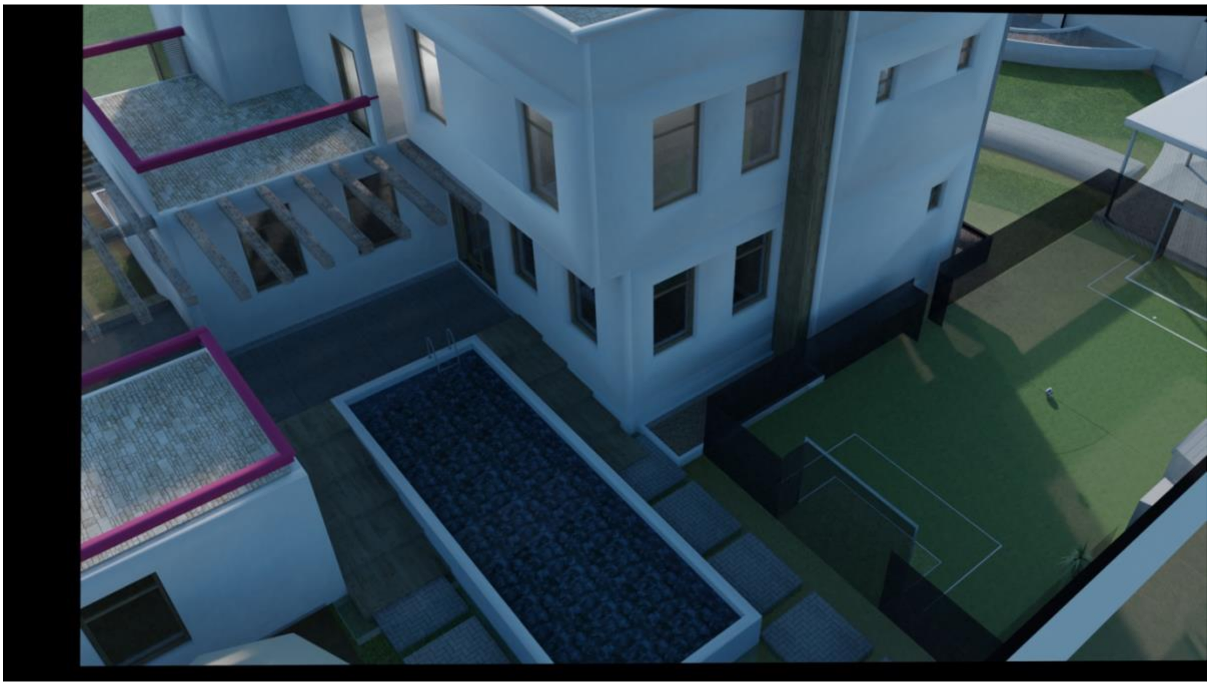

Figure 8 Rectified Frame 1

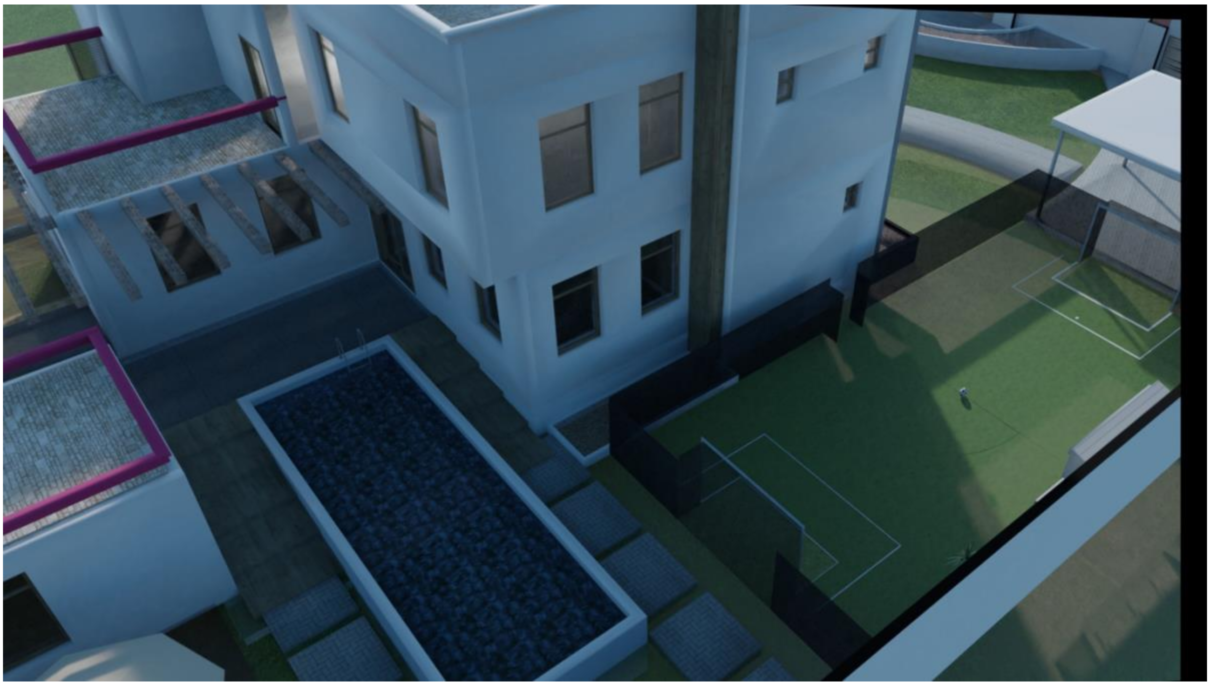

Figure 9 Rectified Frame 2

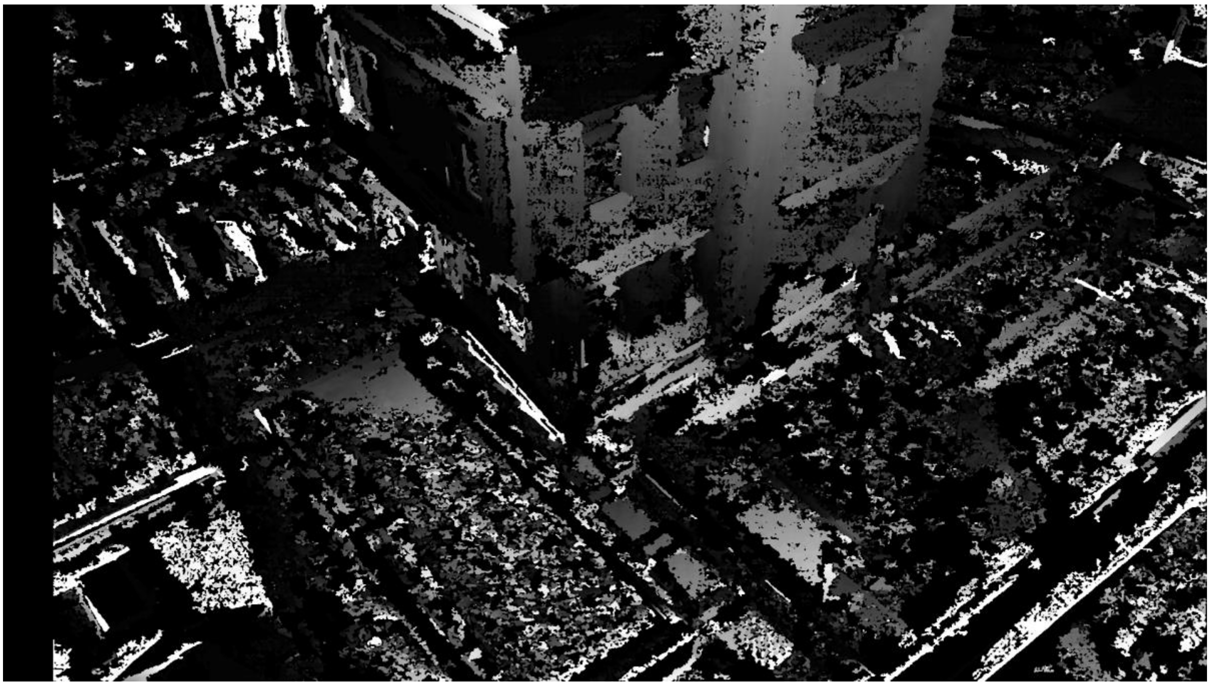

Figure 10 Disparity Map