Vision for Multiple and Moving Cameras - Final Project

Kush Gupta

Introduction The aim of the proposed solution is to obtain a 3D reconstruction of a 3D object. The images for the object were required to capture from a selected camera and later it was calibrated. An adequate object images were taken, to have an adequate number of views for the object. This helped to extract and match feature points between different views,and to compute the fundamental matrix between views. Final step was to obtain a 3D points cloud reconstruction, and to represent the object's geometric elements over this points cloud.

1 Section 1: Obtention of the intrinsic parameters of a camera

The camera's internal parameter matrix K can be defined as below:

$$\begin{bmatrix} \mathrm{F}_x & S & \mathrm{C}_x \\ 0 & \alpha \mathrm{F}_y & \mathrm{C}_y \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \alpha & -\alpha \cos\theta & \mathrm{U}_0 \\ 0 & \beta / \sin\theta & \mathrm{V}_0 \\ 0 & 0 & 1 \end{bmatrix}$$

where F_x and F_y are the focal lengths, C_x,C_y are the offsets, S is the skew and α is the aspect ratio.

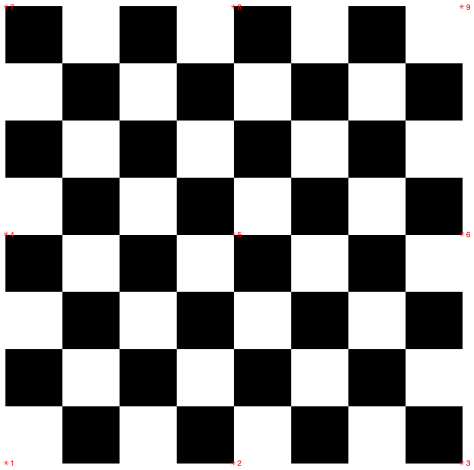

Figure 1: real checker board with marked points.

1.1 Camera Calibration: Part-A

1.Size, in millimeters, of the checkerboard(1080P) on the screen- **530x530mm**.
2.The resolution of the captured images (in pixels)= **2688x1512**.

Internal parameter Matrix-A =

$$\begin{bmatrix} 2017.27881854366 & -2.70498106891837 & 1345.67448737466 \\ 0 & 2023.25568505140 & 705.518031025625 \\ 0 & 0 & 1 \end{bmatrix}$$

- Are the pixels of your camera square?

Ans) From the obtained matrix of internal parameters A, for the selected camera we can observe that

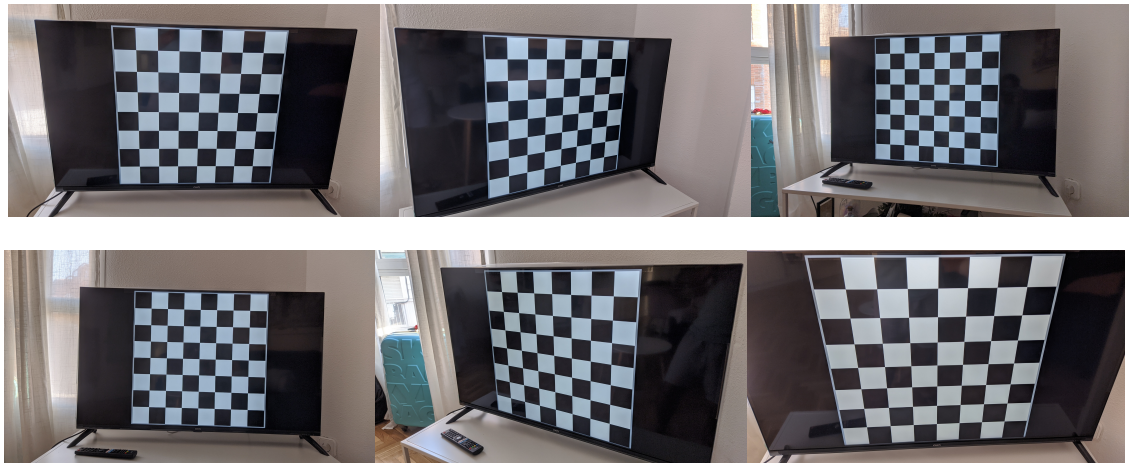

Figure 2: shows the captured images of the checker board(1080P) of size 530x530mm, for the 6 different pattern orientations.

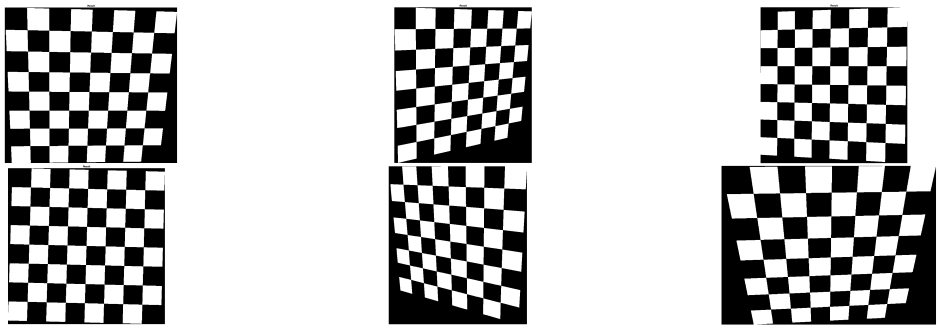

Figure 3: shows the respective homography result images for the checker board(1080P), for the 6 different pattern orientations.

F_x is almost equal to F_y with some noise. Also, we know, that pixel ratio = F_x / F_y $\sin\theta$ = 2017.2788/ 2023.2556 = 1. **Hence, we can say that the camera has square pixels**.

• Which is the degree of coincidence between the principal point and the center of the image plane?
Ans) The dimensions of the image are 2688X1512 pixels, and from the matrix of internal parameters A, we can observe that the value for C_x is approximately half of image width (2688) i.e (1345) and the value for C_y is also close to half of image height (1512) i.e 705 with some noise. **Hence, the principal point (1344,756) is very close to the center of the image plane (1345,705), with some offset in y co-ordinate (due to some noise)**.

• Are the axes of the image plane orthogonal?
Ans) To measure the skewness(θ), we know α and $-\alpha\cos\theta$ = -2.705, **solving for θ gives us a value of 89.9231, very close to 90 degrees. Hence, we can say that the image planes are orthogonal.**

1.2 Camera Calibration: Part-B

1.Size, in millimeters, of the checkerboard(720P) on the screen- **530x530mm**.
2.The resolution of the captured images (in pixels)= **2688x1512**.

Internal parameter Matrix-$A^{'}$ =

$$\begin{bmatrix} 2004.00869804025 & -5.72364419614326 & 1349.71308061241 \\ 0 & 2001.21009203381 & 732.911003272136 \\ 0 & 0 & 1 \end{bmatrix}$$

Figure 4: shows the captured images of the checker board(720P) of size 530x530mm, for the 6 different pattern orientations.

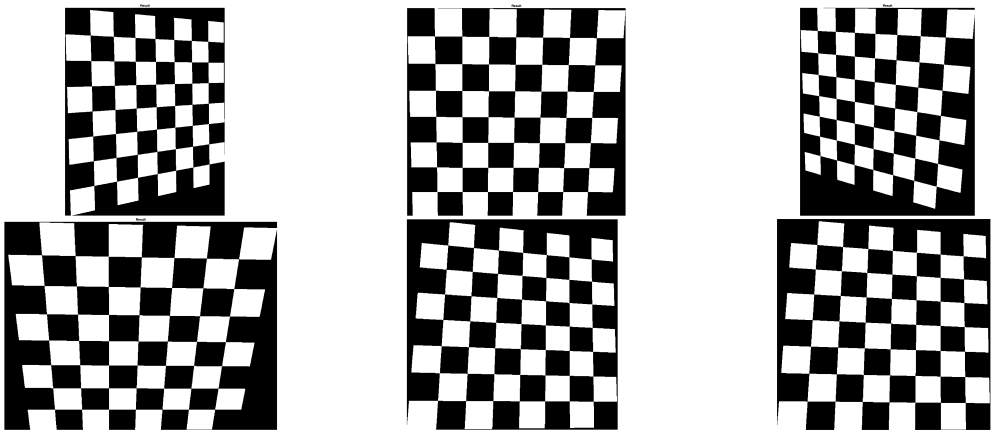

Figure 5: shows the respective homography result images for the checker board(720P), for the 6 different pattern orientations.

- Are the pixels of your camera square?

Ans) From the obtained matrix of internal parameters A, for the selected camera we can observe that F_x is almost equal to F_y with some noise. Also, we know, that pixel ratio = F_x / F_y $\sin\theta$ = 2004.0086/ 2001.2100 = 1. **Hence, we can say that the camera has square pixels**.

• Which is the degree of coincidence between the principal point and the center of the image plane?

Ans) The dimensions of the image are 2688X1512 pixels, and from the matrix of internal parameters A, we can observe that the value for C_x is approximately half of image width (2688) i.e (1349) and the value for C_y is also close to half of image height (1512) i.e 732 with some noise. **Hence, the principal point (1344,756) is very close to the center of the image plane (1349,732), with some offset in y co-ordinate (due to noise)**.

• Are the axes of the image plane orthogonal?

Ans) To measure the skewness(θ), we know α and $-\alpha \cos\theta$ = -5.72, **solving for θ gives us a value = 89.8363 , very close to 90 degrees. Hence, we can say that the image planes are orthogonal.**

Conclusion:- The internal parameter matrix-$A^{'}$ obtained with 720p checker board images contains slightly different values than internal parameter matrix-A obtained with 1080p checker board image. Since we're using the same camera, the corresponding intrinsic matrices must be the same for both the cases. Theoretically, it is expected to obtain the same intrinsic matrix for calibration object of a different size. However, for the smaller calibration object(720p checker board), the probability of the error by manually selecting the points is higher. Hence, for the subsequent tasks we'll be using A matrix obtained by the 1080p checkerboard.

2 Section 2: Finding local matches between several views of an object.

2.1 A. About the scene

The figure 6 contains montage of images captured at a resolution of 2688X1512 from different perspectives (left, center and right) of the scene. In order, to achieve faster computation and to avoid the memory issues, the images have been scaled by a factor of 0.5 (not much information is lost). The challenging tasks during capturing the images, were to take into account proper lighting conditions and to capture them at an angle such that enough point correspondences can be detected and matched between corresponding views.Also, the detected points should be robust. The captured scene contains challenges like, the reflection of light(from the middle object in pink color) which can be observed in images I8 to I11. These captured images (I8 to I11) will be a challenge for the detector to detect and matching the feature points, as the reflecting light, might cause problems while detecting and matching feature points in those images.

Figure 6: shows a mosaic representation of all the captured views.

The different view points and orientations of the scene. **Left view: I1,I2,I3. Center view: I4,I5,I6,I7. Right view: I8,I9,I10,I11.**

2.2 B. Detector and Descriptor combination

The idea to find out two distant scene images from the available different possibilities, is to have a view pair with good amount of corresponding points. So that, it is possible to achieve a good initial reconstruction from the two distant views. However, finding point matches in views having a larger angle is more difficult. Hence, Image pair **I4** and **I9** were selected as this image pair doesn't seem to have a much large angle and the images have different perspective in order to challenge the available detectors and descriptors. The selected images are appropriate to find a suitable set of the Detector, Descriptor combination.

As shown in table 1, six different efficient combinations of detector+descriptor were tested:DoH+SIFT, SURF+SURF, KAZE+KAZE, SIFT+DSP-SIFT, SURF+SIFT, SURF+KAZE. These methods were tested with below parameters:
threshold = 0.001 for detection; max ratio= 0.5 for point matching; nscales = 10; noctaves=3; Metric = 'SSD'; npoints = 350
Detectors like SIFT, DSP_SIFT and SURF, were not able to perform well for the selected images I4 and I9, due to strong changes in the perspective. Form the table 1, we can observe that KAZE+KAZE out performed all other detector descriptor combinations and handled the strong changes in the perspective well. Hence, for further analysis, we will proceed with KAZE+KAZE detector descriptor combination.

capabilities of the selected methods KAZE is a multi-scale 2D feature detection and description algorithm in nonlinear scale spaces. It out performed other approaches because other approaches detect and describe features at different scale levels by building the Gaussian scale space of an image. However, Gaussian blurring does not respect the natural boundaries of objects and smooths to the same degree both details and noise, reducing localization accuracy and distinctiveness. In contrast, KAZE detect and describe the 2D features in a nonlinear scale space by means of nonlinear diffusion filtering. In this way, it makes blurring locally adaptive to the image, reducing noise but retaining object boundaries, obtaining superior localization accuracy and distinctiveness.

	DoH+SIFT	SURF+SURF	KAZE+KAZE	SIFT+DSP-SIFT	SURF+SIFT	SURF+KAZE
Number of inliners in calculating the **homography** transform	80	190	**1182**	166	148	201
Number of inliners in calculating the **fundamental** matrix	45	168	**779**	100	91	138
Number of corresponding points	90	335	**1558**	200	181	276

Table 1: showing the number of in-liners for different detector descriptor combination.

After selecting the KAZE+KAZE as the detector-descriptor combination, i used it to test on different image pairs to measure it's robustness and performance for different image pairs and observed that it works very well for different image pairs and it matches enough number of points even for images with strong perspective changes. Even for the challenging scenario where in images (I8 to I11) we had observed reflection. It was able to detect and match a large number of points. The number of in-liners obtained for different image pairs can be seen in table 2.

The textbfestimated homography matrices tform21 and tform12 appears to be correct as there were enough point correspondences. The transformation of image 2 with respect to image 1 can be observed in the left image refer 8 and the transformation of image 1 with respect to image 2 can be observed in right image 8.

Estimated Fundamental matrix:- The rank of the obtained fundamental matrix F is 2 (checked using matlab rank function). Since the rank of F is 2 and the epipolar lines are spinning around one point refer 9, the estimation of the fundamental matrix F appears to be correct.

	I1 I11	I2 I10	I3 I9	I4 I8	I5 I7	I6 I11	I4 I11	I1 I8	I2 I7	I1 I6	I5 I9	I3 I8
Number of in-liners in calculating the **homography** transform	195	558	184	1486	1813	277	207	1383	203	1585	1459	163
Number of in-liners in calculating the **fundamental** matrix	144	442	108	1134	1817	255	166	962	139	1334	1160	100
Number of corresponding points	288	884	216	2267	3634	509	331	1924	278	2667	2319	199

Table 2: comparison between the number of inliners for homograpghy, number of inliners for the fundamental matrix and number of corresponding points for different selected image pairs.

The estimated homography matrices:
tform21=

$$\begin{bmatrix} 0.61513531 & -0.10905320 & -0.00028692785 \\ -0.093186609 & 0.75640196 & -1.7789820e-05 \\ 99.369102 & 107.82619 & 1 \end{bmatrix}$$

tform12=

$$\begin{bmatrix} 1.6606549 & 0.17094769 & 0.00047987694 \\ 0.19142853 & 1.4048941 & 7.0529932e-05 \\ -183.94606 & -167.62819 & 1 \end{bmatrix}$$

The estimated fundamental matrix F =

$$\begin{bmatrix} -1.08320e-07 & -2.12542e-05 & 0.00784957 \\ 2.26957e-05 & -2.23697e-06 & -0.01668251 \\ -0.00874678 & 0.01572209 & 0.99966819 \end{bmatrix}$$

Figure 7: shows the point correspondences between the image pair I4 and I9.

Conclusion: In conclusion, image pair **I4** and **I9** were selected as they, doesn't have a much larger angle for the initial reconstruction and have strong perspectives changes in order to challenge the different combinations of detectors and descriptors. After, an regressive evaluation it was determined that KAZE+KAZE out performed as a detector descriptor combination. This combination was able to cope with strong perspective changes and matched a good enough number of points. Using I4 and I9 image pair and KAZE for detecting and matching points we obtained a good fundamental matrix with rank of 2, and the estimated homography matrices appeared to be correct.

Figure 8: (Left) transformed image 2 w.r.t 1, (Right) transformed image 1 w.r.t 2

Figure 9: screen captures of the vgg_gui_F.m GUI, showing the epipole in the image pair.

3 Section 3: 3D reconstruction and calibration

In section 2, we found out that the images I4 and I9 had a strong perspective change and had enough number of matched points for initial reconstruction. Hence, in order to select the other set of images, i performed a comparison among different image pairs and noted their respective point correspondences which can be seen in table 3. For the N-view point matching, images subset (I4,I5,I6, I7,I8,I9) were used, to have a good 3D point cloud reconstruction. The detected interest points in these images can be seen in fig 10. For matching, the max ratio value was 0.8. Higher max ratio value is selected in order to increase the number of the point correspondences.

Figure 10: Row-1(Left to Right) I4,I5,I6, Row-2(Left to Right)I7,I8,I9. Images used for the N-view point matching, and the detected interest points in each of them.

Image Set	Matched points
I4,I5,I6,I7,I8,I9,I10,I11	763
I4,I5,I6,I7,I8,I9,I10	1298
I4,I5,I6,I7,I8,I9	**2076**
I4,I5,I6,I7,I8	4448
I1,I2,I3,I4,I8,I9,I10,I11	15
I1,I2,I3,I8,I9,I10	22
I2,I3,I4,I8,I9,I10	28
I2,I3,I4,I5,I6,I7,I8	28
I2,I3,I4,I5,I6,I7	34
I5,I6,I7,I8,I9	2552

Table 3: comparison between the correspondence points between different images pairs to find a subset of images from the scene for 3D point cloud reconstruction.

3.1 Initial projective reconstruction

Figure 11: (Left)I4, (Right)I9; images used for the estimation of the fundamental matrix, with the detected interest points and point matches.

Residual re-projection error. 8 point algorithm = 190.3581
Pixel error: mean = [-1.09500 -5.64202]
Pixel error: std = [8.84364 16.41824]

The corresponding reprojection error Histogram of initial projective reconstruction can seen in figure 12.

3.2 Improving the initial reconstruction by Projective Bundle Adjustment.

Residual reprojection error, after re-sectioning = 494.9883
Pixel error: mean = [-0.59600 2.58236]
Pixel error: std = [23.90479 20.28774]
The corresponding reprojection error Histogram after re-sectioning step can seen in figure 13.

Reprojection error, after Bundle Adjustment = 63.4991
Pixel error: mean = [-0.00000 -0.00000]
Pixel error: std = [9.13793 6.59595]

The corresponding reprojection error Histogram, after the projective bundle adjustment step can seen in figure 14.

Justification for the different re-projection error values:- The error for the initial reconstruction was calculated using the 8-point algorithm on the data points of the two end cameras. The **error value after the initial reconstruction was 190.3581**. After the initial reconstruction (using two end cameras), we did re-sectioning using all the available cameras. Since, we have increased

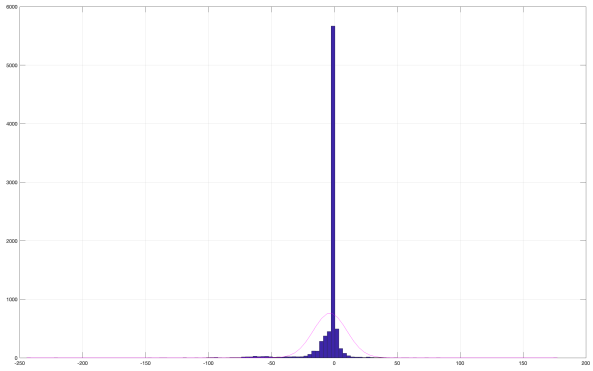

Figure 12: Reprojection error Histogram of initial projective reconstruction.

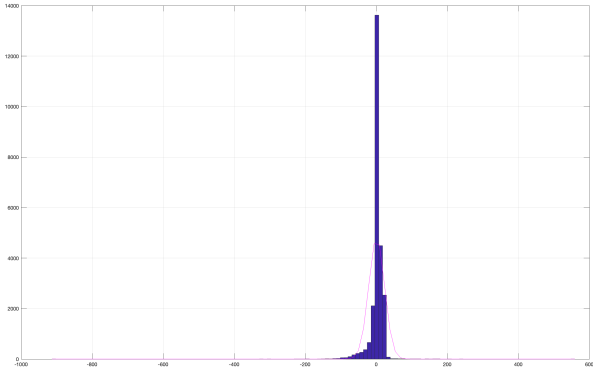

Figure 13: Reprojection error Histogram after re-sectioning step.

the number of cameras (using all the cameras), the re-projection error will be summed up for all cameras and hence, the error after re-sectioning should increase. We observed the same thing that **error increased from 190.3581 to 494.9883 after re-sectioning**. After performing the bundle adjustment step, the re-projection error must reduce, because we are using the 3D points to calibrate all the cameras, which must minimize the error in all views. As expected, **the error decreased from 494.9883 to 63.4991 after projective bundle adjustment**.

3.3 Euclidean reconstruction of the scene

Reprojection error, after euclidean Re-construction = 1531.7998
Pixel error: mean = [-8.31782 4.87388]
Pixel error: std = [15.81900 52.16449]
The corresponding reprojection error histogram, for the euclidean reconstruction of the scene can seen in figure 15.

Comments on the Euclidean reconstruction: We have obtained four possible solutions. Solution 2 was the real illustration of the captured scene. The reconstruction seemed accurate visually. The 3D point cloud reconstruction can be observed in figure 16. The different elements in the scene can be distinguished in the 3D world and the visualization of the scene in the 3D world can be observed in the figure 17. However, the illustration had some pixel errors, (some points behind the camera center).

Conclusion: In conclusion, we can see that the euclidean reconstruction was very realistic. It illustrates well the relative distances and the depth of the scene. There were 4 distinguishable parts in the scene (object-1, object-2, object-3 and some points in the background). With the figures 16, 17 it was illustrated that the relative distances between the objects were well-preserved. The detected points on the objects remained planar. The 3D point cloud reconstruction was done for a relatively high number of matched points (2076), due to which we have obtained a nice and realistic 3D point cloud reconstruction with small errors.

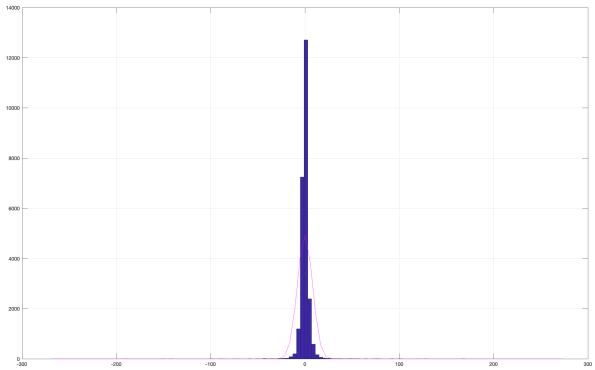

Figure 14: Reprojection error histogram, after the projective bundle adjustment step.

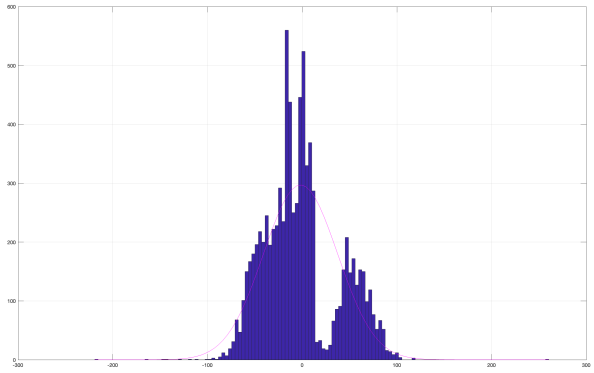

Figure 15: Reprojection error histogram for euclidean reconstruction of the scene.

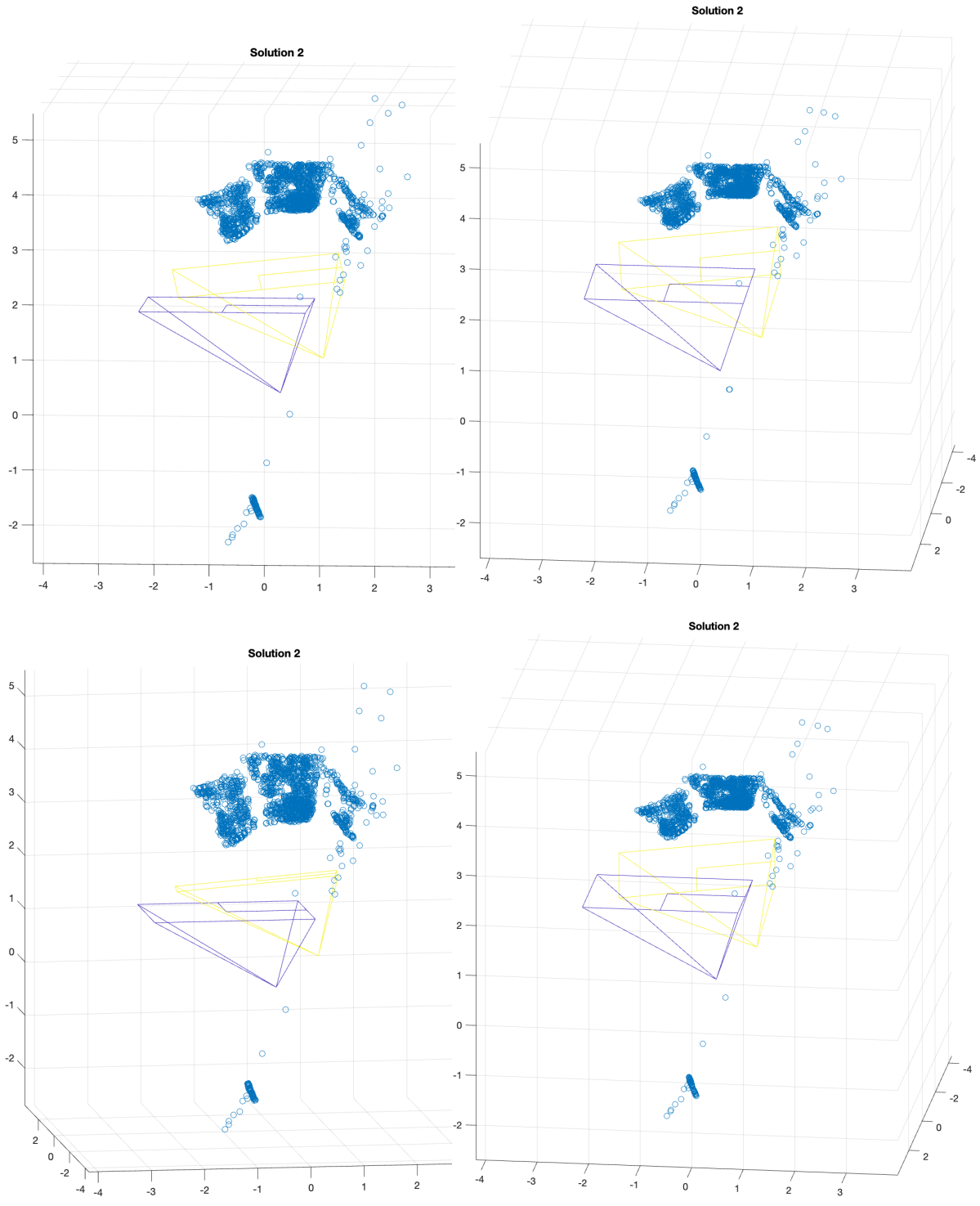

Figure 16: results showing several viewpoints of the 3D point cloud reconstruction.

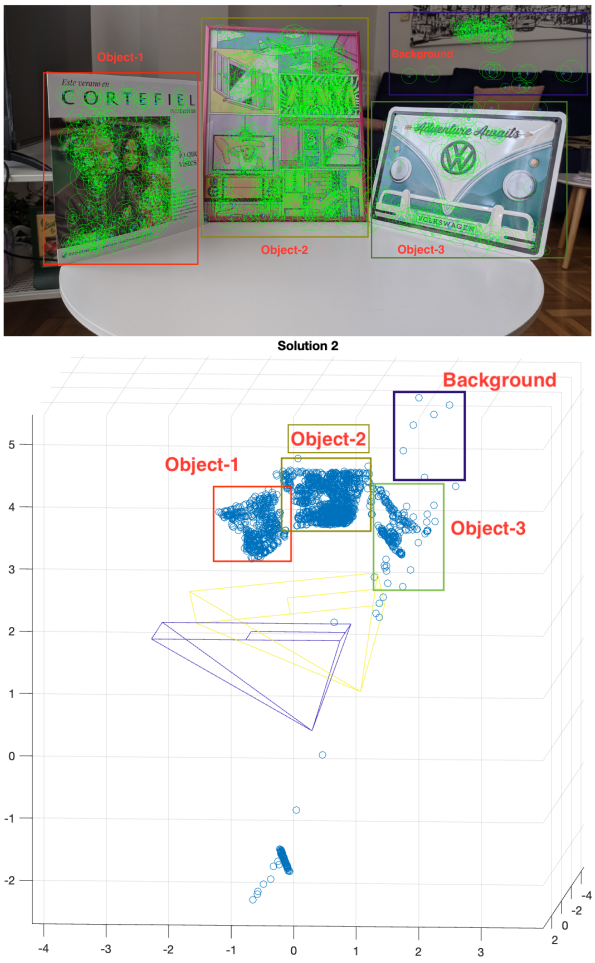

Figure 17: visualization of the scene in the 3D world. (Top) objects in the scene (Image 5) and (bottom) corresponding objects in the reconstructed 3D point cloud.