VMMC LAB EVALUATION

Santiago Herrero Melo

1. <u>Section 1: Obtention of the Intrinsic Parameters of a Camera</u>

a. *Introduction*

The first step of the assignment consists on selecting the camera with which all the work of the assignment will be developed. The camera must be used in the same conditions, not changing its parameters during the rest of the sessions. In my case, I will use the camera of my cell phone.

Now that I have selected the camera, I will proceed to its calibration in the same way as we studied during the previous laboratory sessions.

b. *Exercise 1: Calibration with the Large Checkboard*

For calibrating the camera, a large checkboard with square size of 70 mm (total length of the checkboard is, therefore, 560 mm) will be used. I have taken six photographs of the checkboard from different angles in order to accomplish a more precise calibration of the internal parameters. The images, whose resolution is 3120X4160 pixeles (the images are vertical) are the following:

1. Images for calibration with the large checkboard

We will calculate the homography matrix from each image by relating certain characteristic points of the image with those same points of a automatically generated checkboard by MatLab. All those homographies will be stored in a cell, and from these homographies, we will compute the matrix of internal parameters (A). The obtained matrix of internal parameters is the following:

$$A = \begin{pmatrix} \alpha & \beta & u0 \\ 0 & \gamma & v0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 3439.2 & 11.6061 & 1541.0 \\ 0 & 3397.2 & 2084.7 \\ 0 & 0 & 1 \end{pmatrix}$$

Taking into account how similar the focal lengths (α and β) are between them, and that the γ value is quite small, we can determine that the pixels are almost square.

The principal point is defined by u_0 and v_0. To consider that the camera is properly calibrated, these values should be as close as possible to the central point of the image, which is determined by W/2 and H/2, considering W as the width of the image and H as its height. As said previously, the width and height of the image are 3120 and 4160 pixels, respectively. Therefore, the central point would be (1560, 2080), values that are very similar of those obtained by the internal parameters matrix.

The orthogonality of the axes can be easily obtained with the following formula:

$$\gamma = \alpha * \mathrm{tg}(\varphi)$$

Where φ is the angle that form the vertical and horizontal axes that conform the pixel. By isolating φ, we obtain that its value is $\varphi = 0.19º$, which means that those axes are almost perpendicular. With these obtained values, we can consider that the obtained matrix of internal parameters is good enough for calibrating the camera.

c. *Exercise 2: Calibration with the Small Checkboard*

A second calibration has been performed with a different checkerboard of smaller size, whose squares are 36 mm wide (and therefore the size of the checkboard is 288 mm). The images, taken with the same camera, are the following:

2. Images for calibration with the small checkboard

Proceeding in the same way as in the previous exercise, we obtain the following matrix of internal parameters:

$$A' = \begin{pmatrix} \alpha & \beta & u0 \\ 0 & \gamma & v0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 3335.4 & -24.0087 & 1507.9 \\ 0 & 3346.6 & 2070.8 \\ 0 & 0 & 1 \end{pmatrix}$$

We can see that the skew value (γ) is a bit higher this time, although α and β are more similar than in the previous case. Also, in this occasion, u_0 and v_0 values are not as similar as in the previous case to the centre of the image.

The value of the angle between the vertical and the horizontal axes of the pixel is $\varphi = 0.41º$.

d. *Conclusions*

As the values of φ and γ are higher in the case of A' than the values retrieved by A, from now on the first matrix of internal parameters (A) will be the one used, as we can consider that this calibration is better, as the pixels are considered more squared and less distorted, and the location of pixels is more precise.

This difference can be explained by the type of checkboard used for calibration. Being the second checkboard smaller, the dissimilarity between points is also smaller and the error increases. These errors in the second calibration are, in part, fault of the user, as the matching points have to be selected manually, so the error is proportionally higher with a smaller checkboard.

In the following exercises, the size of the captured images is too big to be properly computed, so a resizing of the images used from section 2 has been performed. Each of the dimensions of the image have been reduced by a factor of 4 (the height and width are ¼ of the original size). Although not perfectly, this new matrix maintains the squareness of the pixel, central point and orthogonality relations previously calculated. The final matrix of internal parameters looks like this:

$$A = \begin{pmatrix} \alpha & \beta & u0 \\ 0 & \gamma & v0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 859.81 & 2.9015 & 385.26 \\ 0 & 849.3 & 521.18 \\ 0 & 0 & 1 \end{pmatrix}$$

2. Section 2: Finding Local Matches between several Views of an Object

a. *Object/Scene Capture*

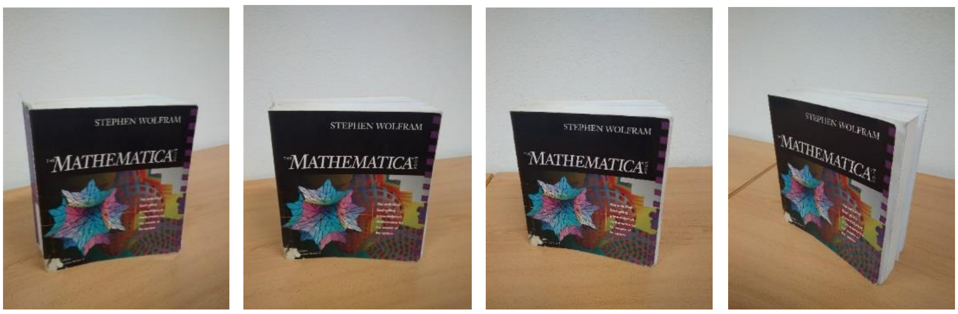
With the already calibrated camera, several images have been taken from a same scene. Some pairs of images will be used for matching, and the eight images shown in the mosaic will be used for creating a panoramic image.

3. Mosaic of the images used for creating the panorama

b. *Detection, Description and Matching of Feature Points*

We will first create a panoramic image with all the images in the mosaic of image 3. Then, we will select three of those images to pair them with one of the other images and study the results with different descriptor – detector combinations. The strange form of this panoramic image can be explained both because the images taken were vertical (but as all the procedures of calibration were made with the camera taking vertical photos, no futher computation has been performed) and because the book is the fixed point while the camera was rotating around it for taking the pictures.

4. Panoramic image



5. Images 3, 4, 5 and 7 from the set, used for evaluating matching

i. *DoH + SIFT*

6. Matched points and homographies of pairs 3-4, 3-5 and 3-7

We can see that in this case, the pair that better matches is the pair 3-4. In the case of the pair 3-7, the are so few matches that the program cannot even calculate an homography.

ii. *SURF + SURF*

7. Matched points and homographies of pairs 3-4, 3-5 and 3-7

This combination is able to match all pair of images with pretty decent solutions in all cases. Several matching points were found, especially for the first pair.

iii. *KAZE + KAZE*

8. Matched points and homographies of pairs 3-4, 3-5 and 3-7

This combination matches several points in close images, especially in the first pair, creating a good homography. In the last pair, the homography is quite deficient due to the small amount of points matched.

iv. *SIFT + DSP-SIFT*

9. Matched points and homographies of pairs 3-5, 3-5 and 3-7

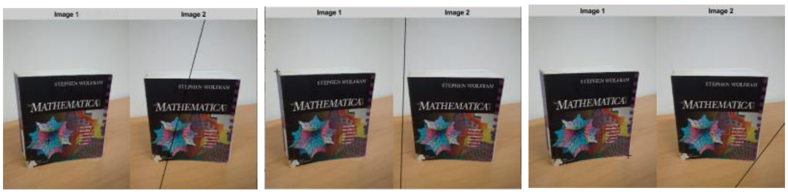
Once more, the last combination, as the images are too different, is not able to match enough points properly and returns a bad homography. In the particular case of this detector – descriptor pair, many points are matched outside the cover of the book, which is curious because those are not so textured surfaces. And, differently to the previous cases, in the second pair, many few points are matched over the cover of the book, although the homography of the two first cases are not that bad or different from other tests.

c. *Qualitative and Quantitative Evaluation*

In this section an evaluation will be performed for each of the detector – descriptor pairs with each of the image pairs. For each pair, the number of inliers and the fundamental matrix will be calculated. As the calculation of the fundamental matrix is based on RANSAC, the initialization is random, so, although similar, the results will be different each time the code is run for the same pair. The number of inliers is invariant and keeps constant for each run.

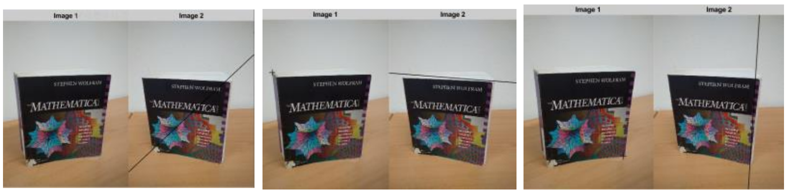
i. *DoH + SIFT*

$$\begin{pmatrix} -8.411 * 10^{-5} & -2.595 * 10^{-4} & 0.4726 \\ 2.647 * 10^{-4} & 3.216 * 10^{-6} & -0.0086 \\ -0.4278 & 0.0020 & 0.7965 \end{pmatrix}$$



10. Fundamental matrix and qualitative evaluation for pair 3-4

$$\begin{pmatrix} -1.446 * 10^{-6} & -1.884 * 10^{-5} & 0.0072 \\ 2.188 * 10^{-5} & 7.008 * 10^{-7} & -0.0143 \\ -0.0073 & 0.0112 & 0.9998 \end{pmatrix}$$

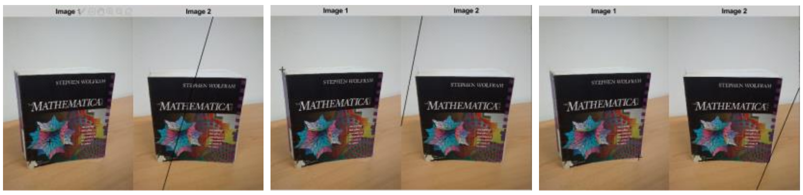


11. Fundamental matrix and qualitative evaluation for pair 3-5

In the case of the last pair, not enough points were matched. For calculating the fundamental matrix, at least 8 matches are needed, and this pair didn't achieve that minimum. Therefore, no fundamental matrix could be calculated.

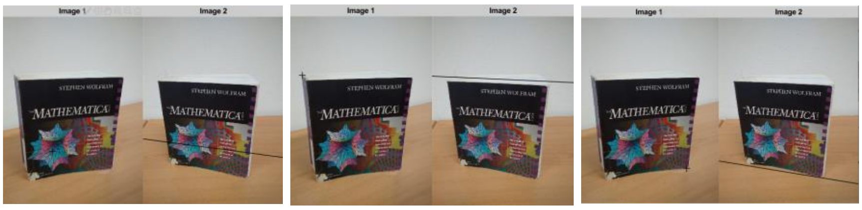
ii. *SURF + SURF*

$$\begin{pmatrix} 1.494 * 10^{-4} & 1.896 * 10^{-4} & -0.8569 \\ -1.859 * 10^{-4} & -7.281 * 10^{-6} & -0.1367 \\ 0.7747 & 0.1386 & 0.3086 \end{pmatrix}$$



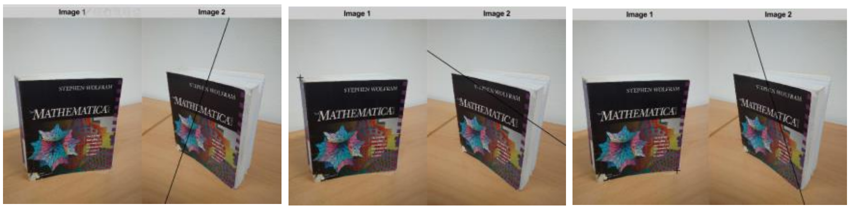
12. Fundamental matrix and qualitative evaluation for pair 3-4

$$\begin{pmatrix} -1.041 * 10^{-7} & 2.409 * 10^{-6} & -2.226 * 10^{-4} \\ 3.257 * 10^{-6} & -1.878 * 10^{-7} & -0.0125 \\ -9.056 * 10^{-4} & 0.0096 & 0.9999 \end{pmatrix}$$



13. Fundamental matrix and qualitative evaluation for pair 3-5

$$\begin{pmatrix} 6.405 * 10^{-6} & 2.154 * 10^{-5} & -0.0116 \\ -1.691 * 10^{-5} & -4.294 * 10^{-7} & 0.0073 \\ 0.0047 & -0.0073 & 0.9999 \end{pmatrix}$$

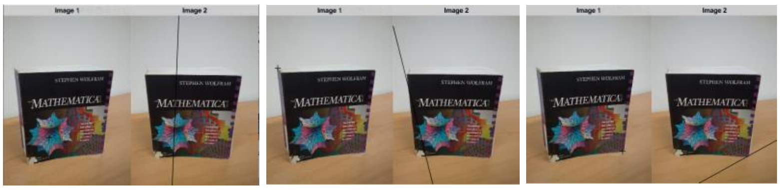


14. Fundamental matrix and qualitative evaluation for pair 3-7

With this combination, we can see pretty good and precise results for every pair, especially for the second one, where the epipolar lines do very good matchings between the images. Also, we can see that the epipolar point is located inside the image, and not outside as in other cases of the transformation matrix.

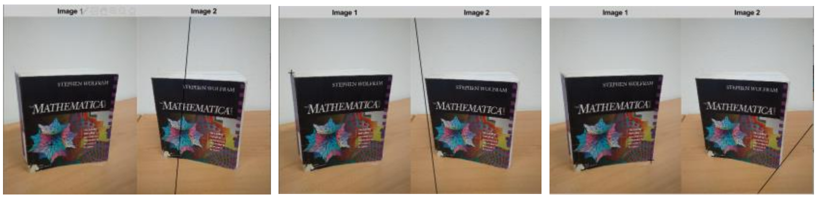
iii. *KAZE + KAZE*

$$\begin{pmatrix} -2.23 * 10^{-5} & -1.117 * 10^{-4} & 0.129 \\ 1.165 * 10^{-4} & 4.806 * 10^{-7} & -0.0315 \\ -0.1182 & 0.0277 & 0.9838 \end{pmatrix}$$



15. Fundamental matrix and qualitative evaluation for pair 3-4

$$\begin{pmatrix} -2.905 * 10^{-5} & -5.539 * 10^{-5} & 0.0907 \\ 5.426 * 10^{-5} & 2.77 * 10^{-6} & -0.0138 \\ -0.0682 & 0.0071 & 0.9934 \end{pmatrix}$$

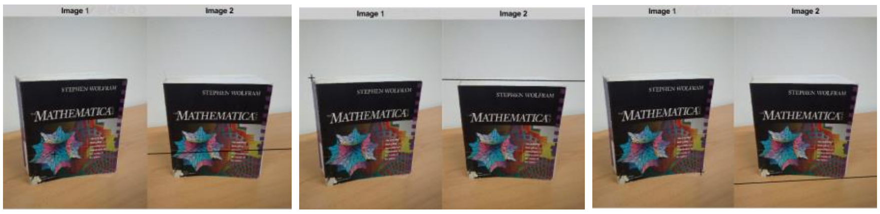


16. Fundamental matrix and qualitative evaluation for pair 3-5

For the last pair, as very few points were matched, no fundamental matrix could be computed. In the other pairs of images, a great amount of inliers were computed with a good precision of epipolar point computation, although not perfect.

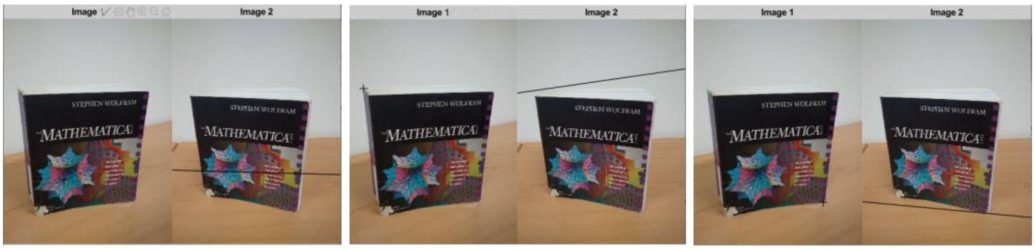
iv. *SIFT + DSP-SIFT*

$$\begin{pmatrix} -1.36 * 10^{-7} & -2.331 * 10^{-6} & 0.001 \\ 7.38 * 10^{-6} & -3.996 * 10^{-7} & -0.0229 \\ -0.002 & 0.0205 & 0.9995 \end{pmatrix}$$



17.Fundamental matrix and qualitative evaluation for pair 3-4

$$\begin{pmatrix} 9.712 * 10^{-7} & 4.283 * 10^{-6} & -0.0031 \\ 1.456 * 10^{-6} & -3.662 * 10^{-7} & -0.0018 \\ 0.0011 & 0.0092 & 0.9999 \end{pmatrix}$$



18. Fundamental matrix and qualitative evaluation for pair 3-5

The few amount of points matched by the 3-7 pair makes impossible the computation of the fundamental matrix. It can be seen that the calculation of this matrix for the previous pairs of images is quite precise.

d. *Selection*

Descriptor – Detector	Images Used	Number of points	Number of inliers
DoH - SIFT	3 and 4	373	187
	3 and 5	73	37
	3 and 7	1	0
SURF – SURF	3 and 4	380	190
	3 and 5	227	114
	3 and 7	41	21
KAZE – KAZE	3 and 4	1119	560
	3 and 5	437	219
	3 and 7	13	0
SIFT – DSP-SIFT	3 and 4	542	271
	3 and 5	84	42
	3 and 7	12	0

For every detector – descriptor pair studied, the pair of images that has shown better results is the one formed by **the third and the fourth images**, probably because consecutive images provide better matches. When having to select between detector - descriptor pairs, the homographies obtained are very similar. Although in some cases an improvement can be felt when applying the fundamental matrix, the parameter that has made me decide finally for the combination **KAZE+KAZE** are the inliers, which are much higher in this case respect to the other combinations, although for every combination the ratio of inliers over the total detected points is around 50%.

3. Section 3: 3D Reconstruction and Calibration

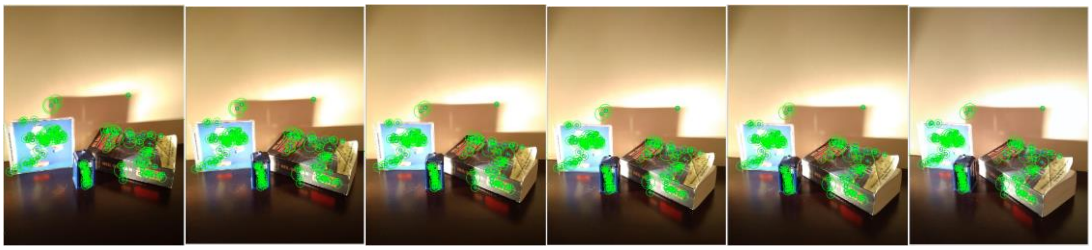
For the purpose of this section, the dataset used is different, as the images used in the previous section were not enough representative as they only contained a single object. The images used for this section are the following, containing three different objects with sharp right angles.

19. Montage of the images used for the 3D reconstruction

a. *Compute consistent point matchings among the views*

For doing so, the provided **n_view_matching** function has been used. The point locations and features used in this function are obtained from the detector and descriptor matching previously performed, which returns a matrix of points, the detected points to be reconstructed. For further operations, these points have to be homogeneized, as we only have two coordinates of them.

Different amount of points have been detected for each image, and the final obtained correspondence is for 176 points, with which we will work from now on.



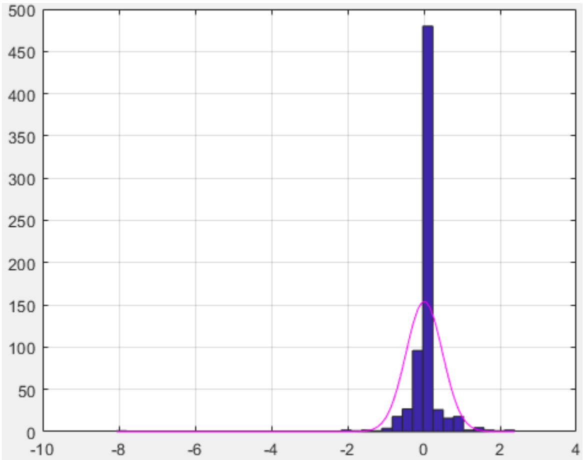
20. Distribution of the matched points for all the images

b. *Fundamental Matrix and Initial Projective Reconstruction*

The initial projective reconstruction has been calculated with the first and last (first and sixth) cameras used for the composition. From this projective reconstruction, the following fundamental matrix was obtained, with a reprojection error of 0.22766:

$$\begin{pmatrix} -1.694 * 10^{-7} & 1.823 * 10^{-6} & -1.398 * 10^{-4} \\ 2.43 * 10^{-6} & 6.233 * 10^{-7} & -0.0091 \\ -0.001 & 0.0067 & 0.9999 \end{pmatrix}$$

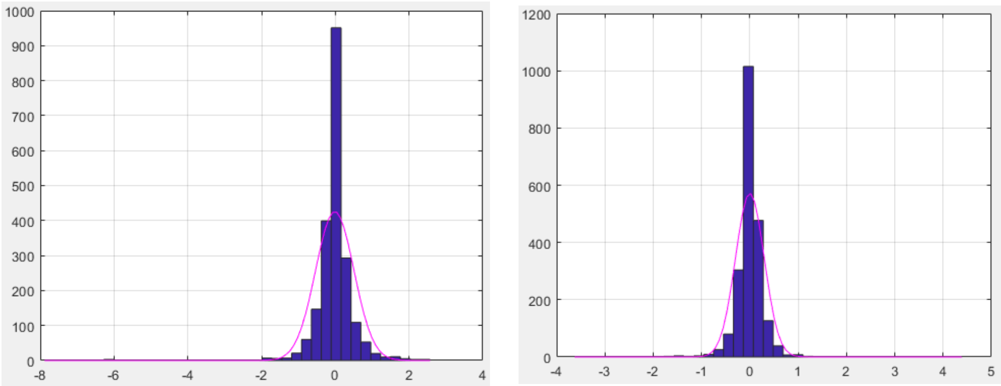
The error histogram is as follows:



21. Error histogram for the initial projective reconstruction

c. *Projective Bundle Adjustment*

This time, all the images were used, in order to improve the initial projective reconstruction. For doing so, first a resectioning step was performed. The purpose of this step is to obtain the projection matrices for every view (six in this case). The reprojection error after this step is of 0.27018. Afterwards, the bundle adjustment was performed, obtaining new reprojected mtrices and new coordinates of reprojected points. The reprojection error after this step is 0.087888.



22. Error histograms after resectioning step and after bundle adjustment

d. *Re-Compute the Fundamental Matrix*

After the projective bundle adjustment, a new fundamental matrix had to be calculated. For obtaining this matrix, the two first cameras (the two first matrices of the bundle adjustment) were used. The function employed for computing this fundamental matrix was the **vgg_F_from_P** function, and the matrix looks as follows. For further computations, the matrix is normalized.

$$\begin{pmatrix} -1.204 * 10^{-8} & 5.065 * 10^{-7} & -2.522 * 10^{-4} \\ 3.269 * 10^{-8} & 6.145 * 10^{-8} & -0.0035 \\ 5.93 * 10^{-5} & 0.0032 & 0.1695 \end{pmatrix}$$

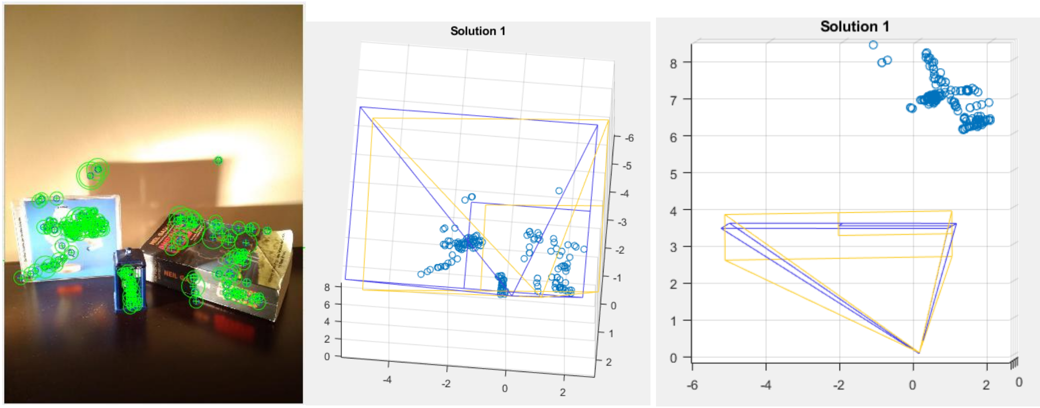
e. *Euclidean Reconstruction*

The final step is to use the properties of the essential matrix between two cameras to obtain the final euclidean reconstruction of the composition. To obtain this matrix, the fundamental matrix that was calculated in the previous step will be used with the matrix of internal parameters of the camera used, which was calculated in the first section of this report. This matrix has the following form:

$$\begin{pmatrix} 0.0525 & 2.1818 & 0.0833 \\ 0.141 & 0.2693 & -17.2395 \\ 0.4107 & 16.9439 & 0.2914 \end{pmatrix}$$

With this essential matrix, four reconstructions of two cameras will be calculated, taking into account that the first camera is fixed, and the second one can vary between two locations and two rotations. Once this is obtained, the final reconstruction will be computed and shown, as a cloud of points with the two cameras used for this last step. For computing the homogeneous coordinates of the points, the reprojected points after the bundle adjustment will be used to do the triangulation.

Once these homogeneous coordinates for the euclidean transformation are computed, also the projection matrices for the euclidean transformation have to be calculated. Once both of those elements, are obtained, the final coordinates of the 3D points are computed, showing the final reconstruction.



23. Correspondences between the points matched in the image and their 3-D reconstruction

As it can be inferred from the previous images, the 3D reconstruction is quite precise, showing the points in a correctly located distribution. Also, we can differenciate between two different planes, one vertical that corresponds to the CD cover and the TARDIS, and o diagonal (almost horizontal) plane that corresponds to the cover of the book.