Assignment 2 - Computer Vision (EEN020)

Luca Modica

November 23, 2023

1 Calibrated vs. Uncalibrated Reconstruction.

Theoretical exercise 1

The camera matrix for estimating a 3D point X is the following:

$$x = PX.$$

- x is the vector representing the point in image coordinates.
- P is the 3x4 camera matrix, is defined as follows:

$$P = K(R|t).$$

- X is the vector representing the corresponding point in the 3D world.

In our case, since the camera is *uncalibrated* (that is, the value of the intrinsic parameter K is unknown), there is a projective ambiguity of the considered 3D point. In other words, we can't uniquely determine the correspondence between the point in real-world coordinates and the related image projection. This is because the value K contains important camera internal characteristics:

- focal length,
- aspect ratio,
- skew,
- principal point.

As a consequence, estimating 3D points and cameras simultaneously under the assumption of uncalibrated cameras, X and P can always be transformed by any projective transformation T. In other words, a new solution with identical image projections can always be obtained from TX for any projective transformation T of 3D space. This is shown by the following assumption:

$$X' = TX, P' = P(T^{-1}),$$

and the following equation:

$$x' = P'X' = PT^{-1}TX = PX = x.$$

As we can see, X and X' will project to the same position in the image plane, representing the mentioned projective ambiguity.

Computer exercise 1

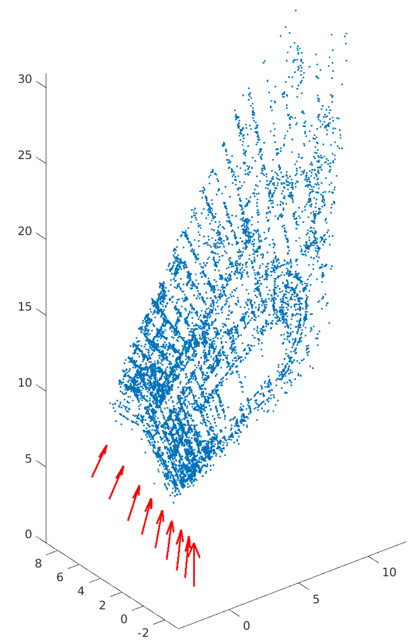


Figure 1: Plot of the 3D points of the reconstruction and the 9 available cameras. As we can see from the resulting 3D plot, the physical properties look realistic in the reconstruction.

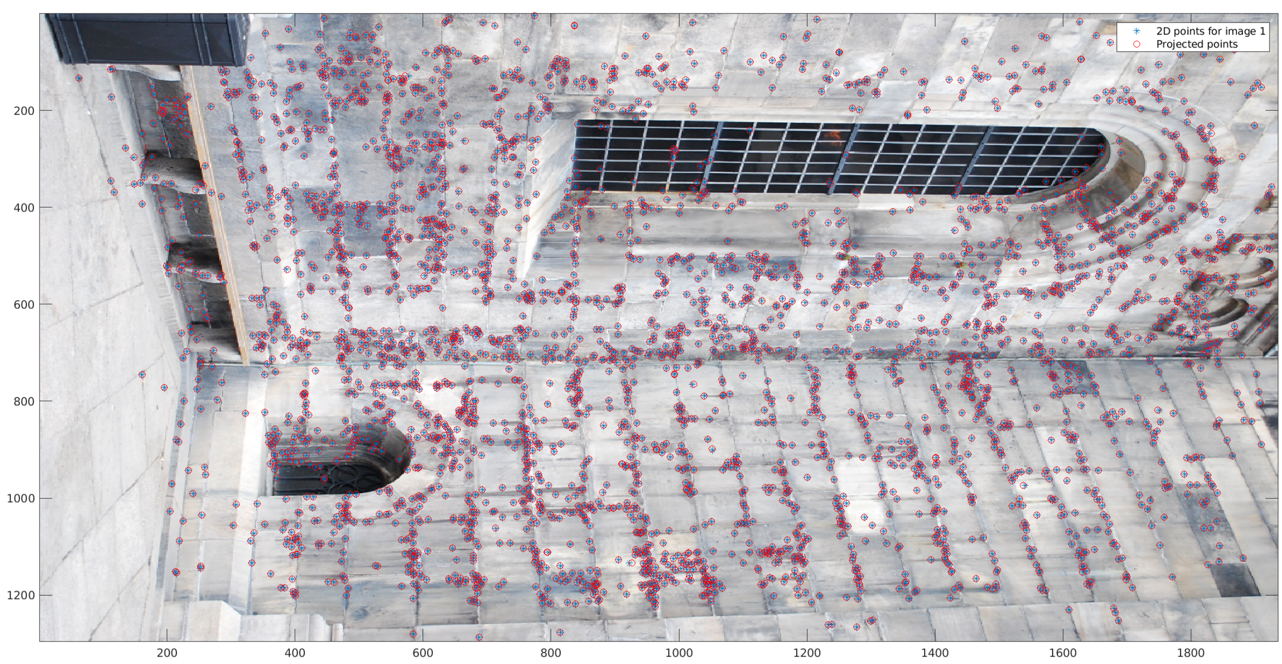
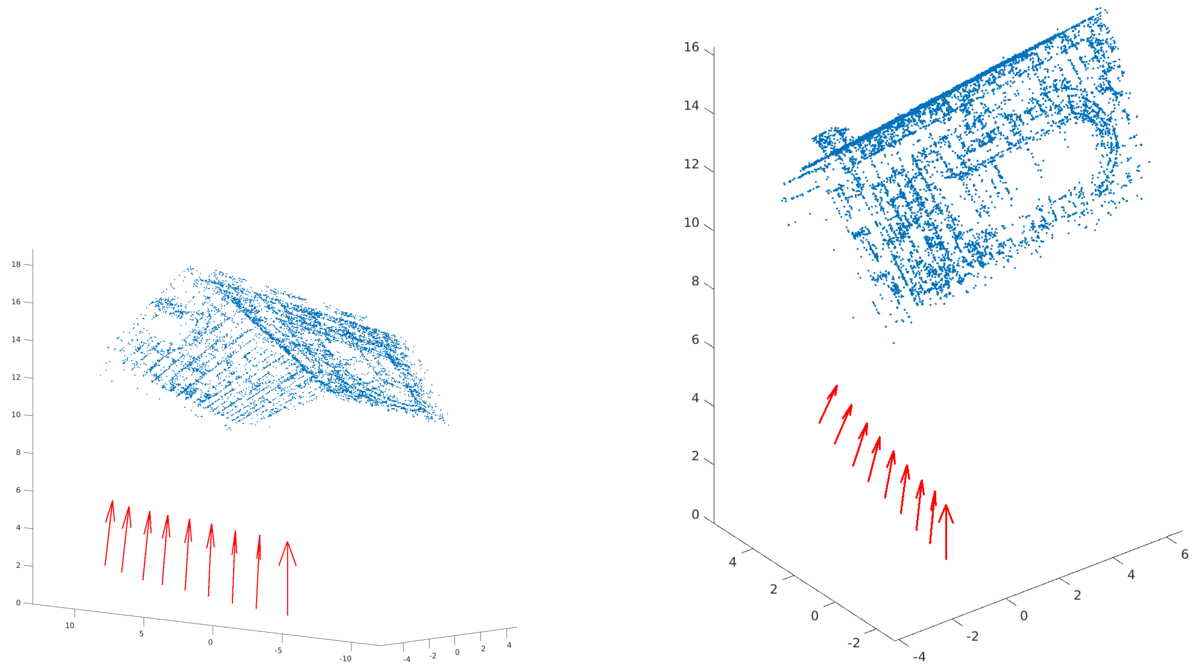


Figure 2: 3D points projected into the camera of the image 1. As we can notice, the projected 3D points appear to be close to the corresponding points of the mentioned image.



(a) Cameras with the 3D points tranformed by T_1.

(b) Cameras with the 3D points tranformed by T_2.

Figure 3: 3D points tranformed by projective tranformations T_1 and T_2, which added in general translations and a different scale compared to the original 3D points with the original cameras. While the projection with T_1 does not seem to reasonable in relation to real world coordinate, the projection with T_2 appear to be a more realistic one.

To conclude the first computer exercise, the 3D points tranformed by T_1 and T_2 were projected into the camera of the first image. As we can notice below, for both the new reconstructions, the projections and the image points align as well as the the original 3D points projected with the original cameras. This is due to projective ambiguity: for the camera of the choosen image we don't know the inner parameter K, thus the camera is unclalibrated.

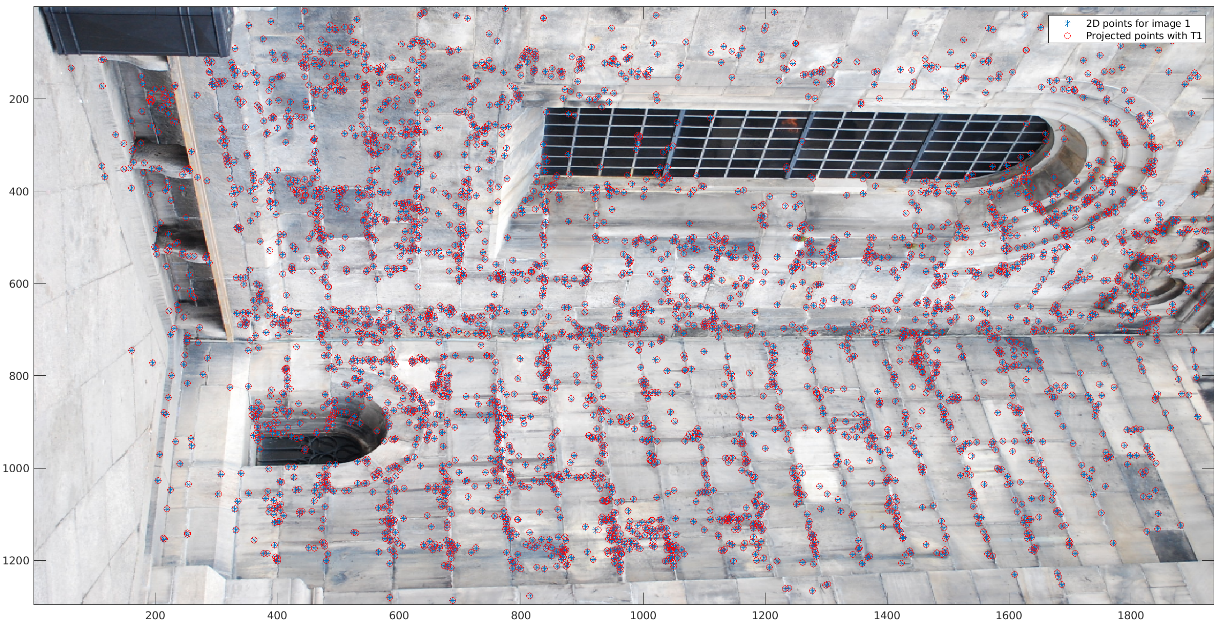


Figure 4: 3D points transformed by T_1 projected into the camera of the image 1.

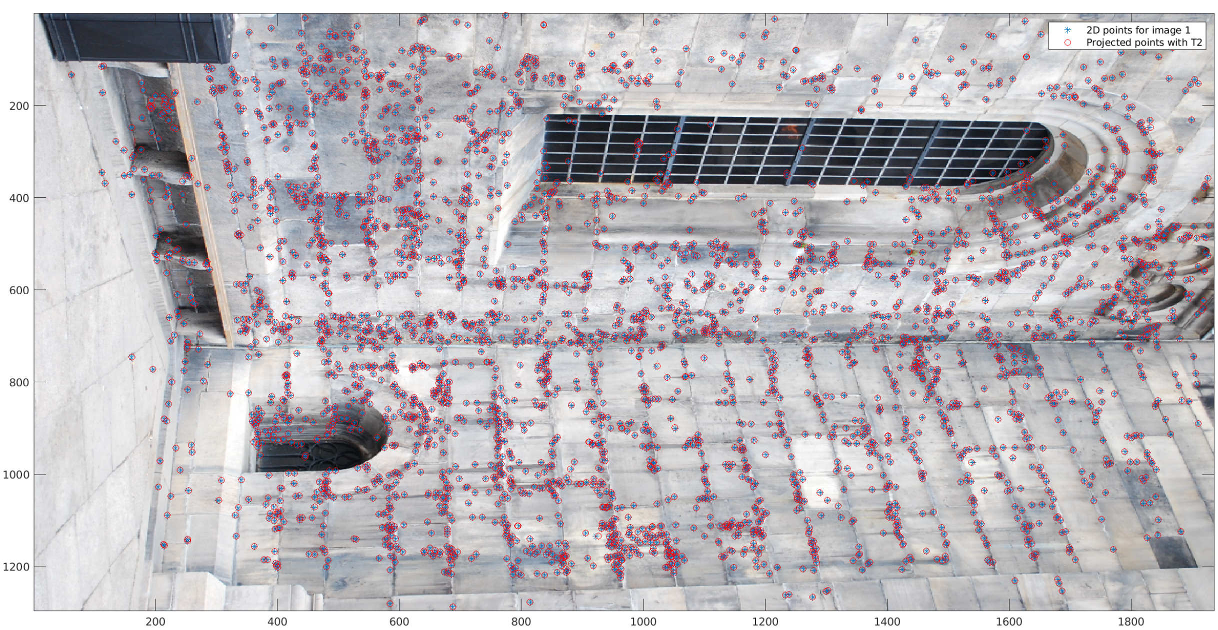


Figure 5: 3D points transformed by T_2 projected into the camera of the image 1.

Theoretical exercise 2

Using calibrated cameras (that is, having known the value of the inner matrix K) means that we know the following information related to the camera itself:

- f - focal length,
- γ - aspect ratio,
- s - skew,
- (x_0, y_0) - principal point.

All of this known information reduce the ambiguity of the projection of the 3D point, thus we would not get the same projective distortion in as in an uncalibrated camera scenario.

Despite how the previous projective ambiguity is affected when the cameras are calibrated, there is still ambiguity to the reconstruction. This is due to the unknown scale of a 3D point: the reconstruction can be determined up to a projective transformation. If $\lambda x = PX$, then for any projective transformation we have:

$$X' = H^{-1}X.$$

2 Camera Calibration.

Theoretical exercise 3

We verify with the following code in MATLAB:

```
K_inv = adjoint(K) / det(K);
```

that:

$$K^{-1} = \begin{pmatrix} 1/f & 0 & -x_0/f \\ 0 & 1/f & -y_0/f \\ 0 & 0 & 1 \end{pmatrix}.$$

Moreover, by multiplying A and B, we also verify that:

$$K^{-1} = AB.$$

The matrices A and B represent 2 different transformations:

- A contains the information related to the focal length,
- B contains the value of the camera principal point.

From the geometric point of view, given the point $x :=$ image coordinates, those 2 tranformations change coordinates in the image using:

$$x' = K^{-1}x$$

obtaining a so called **normalized (calibrated) camera**. In other words:

$$x' = K^{-1}K(R|T)X = (R|T)X,$$

where X is the the related 3D point.

With this transformation, the principal point (x_0, y_0) will end up to the principal axes $((0, 0, 1))$. In general, a point with ditance f to the principal point end up to to the coordinates $(-x_0/f, -y_0/f)$, which correspond to the original 2D location of the camera sensor.

Now, suppose a camera with resolution 800×600 and

$$K = \begin{pmatrix} 400 & 0 & 400 \\ 0 & 400 & 300 \\ 0 & 0 & 1 \end{pmatrix}.$$

The related normalized points of $x_1 = (0, 300)$ and $x_2 = (800, 300)$ are the following:

$$x'_1 = K^{-1}x_1 = (-1, 0),$$

$$x'_2 = K^{-1}x_2 = (1, 0).$$

Considering the related obtained point in $\mathbb{P}^2$:

$$x'_1 = (-1, 0, 1),$$

$$x'_2 = (1, 0, 1),$$

the angle between the viewing rays of these points can be computed as follows:

$$\text{angle} = \arctan(x'_1 \cdot x'_2) = 0.$$

Now, let's consider the following calibrated camera $P = K[R|t]$ and its normalized version $P' = [R|t]$.

- The camera center C is the 3D point from which all the camera's viewing rays emanate. for this reason, we can compute C by taking the null-space of P, since C is a point such that:

$$PC = 0.$$

- the principal axis can be retrieved by taking the first 3 components of the last row of P (in other words, $P_{1:3}$).

The normalized version P' has the same camera center C and the same principal axis $P_{1:3}$. This for the following reasons.

- the camera center C is still in the null-space of P'; this since $P'C = 0$ is the same of $PC = 0$ up to a scale factor contained in K.
- Since the camera has not changed, the line connecting the camera center and the principal point would be the same as in the original camera. Therefore, also the principal axis will be the same as in the original case.

3 Direct Linear Transformation DLT

Theoretical Exercise 5

- To show that the following linear least square system

  $$\min_v ||Mv||^2$$

  always has the minimum value 0, we will take the derivative of the objective function and set it to 0, to search for stationary points.

  The derivative of the objective function, in respect of v, is the following:

  $$\frac{d}{dv}[||Mv||^2] = \frac{d}{dv}[(Mv)^T Mv] = \frac{d}{dv}[v^T M^T Mv] = 2M^T Mv.$$

  After setting the derivative to 0, we obtain:

  $$2M^T Mv = 0 \rightarrow v = \frac{1}{2}(M^T M)^{-1}0 = 0.$$

  That is, the solution to $\min_v ||Mv||^2$ is $v = 0$, that is the 0-vector.

- To show that if M has a singular value decomposition $M = U\Sigma V^T$ then

  $$||Mv||^2 = ||\Sigma V^T v||^2$$

  We can evaluate the following:

  $$||Mv||^2 = (Mv)^T Mv = (U\Sigma V^T v)^T U\Sigma V^T v = v^T V\Sigma^T U^T U\Sigma V^T v.$$

  Since U is orthoganal, $U^T U = I$. Thus, we can write:

  $$v^T V\Sigma^T U^T U\Sigma V^T v = v^T V\Sigma^T \Sigma V^T v = (\Sigma V^T v)\Sigma V^T v = ||\Sigma V^T v||^2,$$

  satisfying the equation above.

  To show instead that $||V^T v||^2 = 1$ if $||v||^2 = 1$ in the same context, we can write the following:

  $$||V^T v||^2 = (V^T v)^T V^T v = v^T V V^T v.$$

  Since also V is an orthogonal matrix, $VV^T = I$. this implies:

  $$v^T V V^T v = v^T v = ||v||^2.$$

  Where the last result is equal to 1 if $||v||^2 = 1$.

- Let's consider this new optimization problem:

  $$\min_{||\tilde{v}||^2=1} ||\Sigma\tilde{v}||^2.$$

  Recalling the previous point, M has a singular value decomposition $M = U\Sigma V^T$, where U and V are orthogonal and thus they don't affect the norm of a vector. So, referring the previous minimization problem, we can write:

  $$\min_{||v||^2=1} ||Mv||^2 = \min_{||v||^2=1} ||U\Sigma V^T v||^2 = \min_{||v||^2=1} ||\Sigma v||^2,$$

  where the last result will give us the same minimal value of $\min_{||\tilde{v}||^2=1} ||\Sigma\tilde{v}||^2$ and allows us to obtain a solution to $\min_{||v||^2=1} ||Mv||^2$ from $\min_{||\tilde{v}||^2=1} ||\Sigma\tilde{v}||^2$.

  Moreover, there are always at least two solutions to these (equivalent) problems. This is because, since U and V are orthogonal, in addition to their scaling they can be different matrices that brings to different solutions, that still satisfies the minimization problem.

Theoretical Exercise 6

Having the following relations:

$$\tilde{x} \sim Nx,$$
$$\tilde{x} \sim \tilde{P}X,$$

We can replace $\tilde{x}$ obtaining:

$$Nx \sim \tilde{P}X.$$

For the exercise, we need the value of x from the camera $\tilde{P}$. Thus, we will divide both parts by N obtaining:

$$x \sim N^{-1}\tilde{P}X.$$

Finally, to compute the camera P that solves the original problem

$$x \sim PX$$

from P, we can just replace equations and divide by X:

$$PX \sim N^{-1}\tilde{P}X \rightarrow P \sim N^{-1}\tilde{P}.$$

So, $P = N^{-1}\tilde{P}$.

Computer Exercise 2

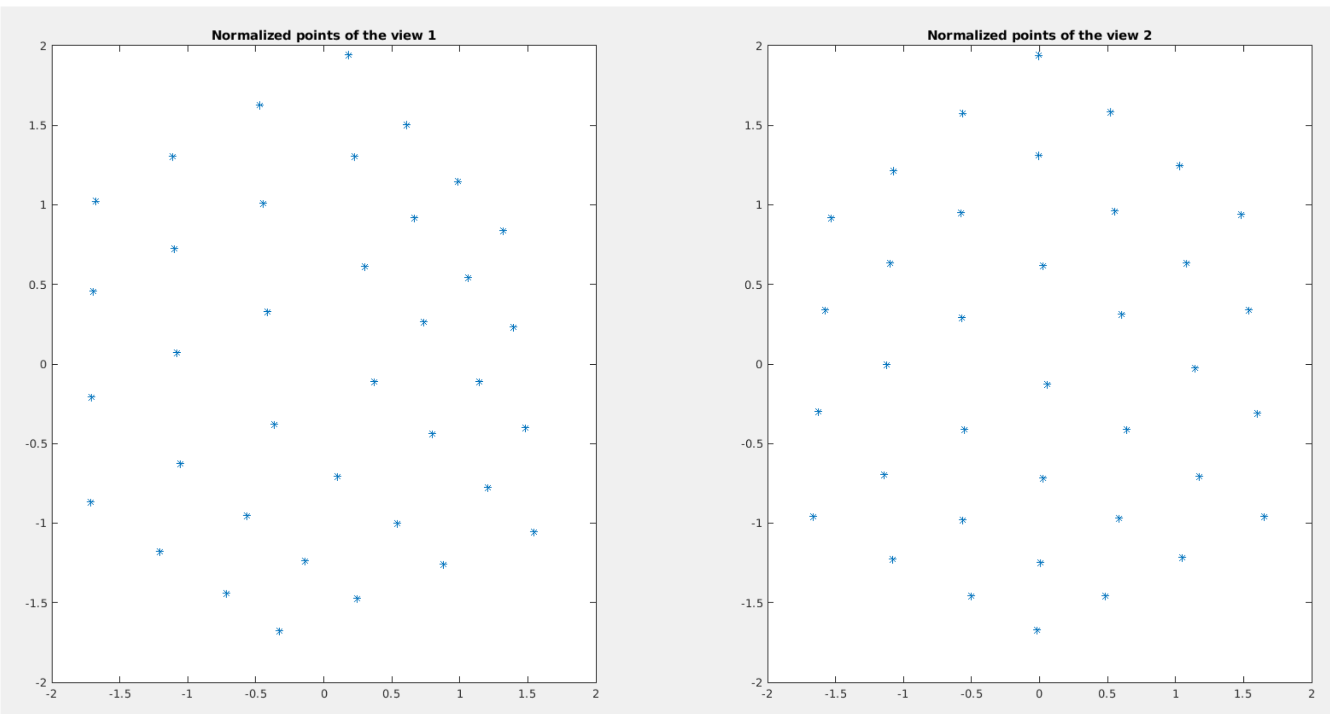


Figure 6: Normalized 3D points, subtracting the meaning and re-scaling by the standard deviation of the original ones.

Mean and standard deviation used for normalization in view 1:

- mean of coordinate x: 1014.8534
- mean of coordinate y: 839.0383
- standard deviation of coordinate x: 196.6099
- standard deviation of coordinate y: 198.5218.

Mean and standard deviation used for normalization in view 2:

- mean of coordinate x: 930.9657
- mean of coordinate y: 795.186
- standard deviation of coordinate x: 198.4141
- standard deviation of coordinate y: 199.4361.

Moreover, for each view we verify that the points are around (0, 0) by checking that the mean (for each coordinate) is equal or close to the 0-value; with a similar methodology, we compute the standard deviation of the normalized points to check that those are equal to 1, obtaining the desired value.

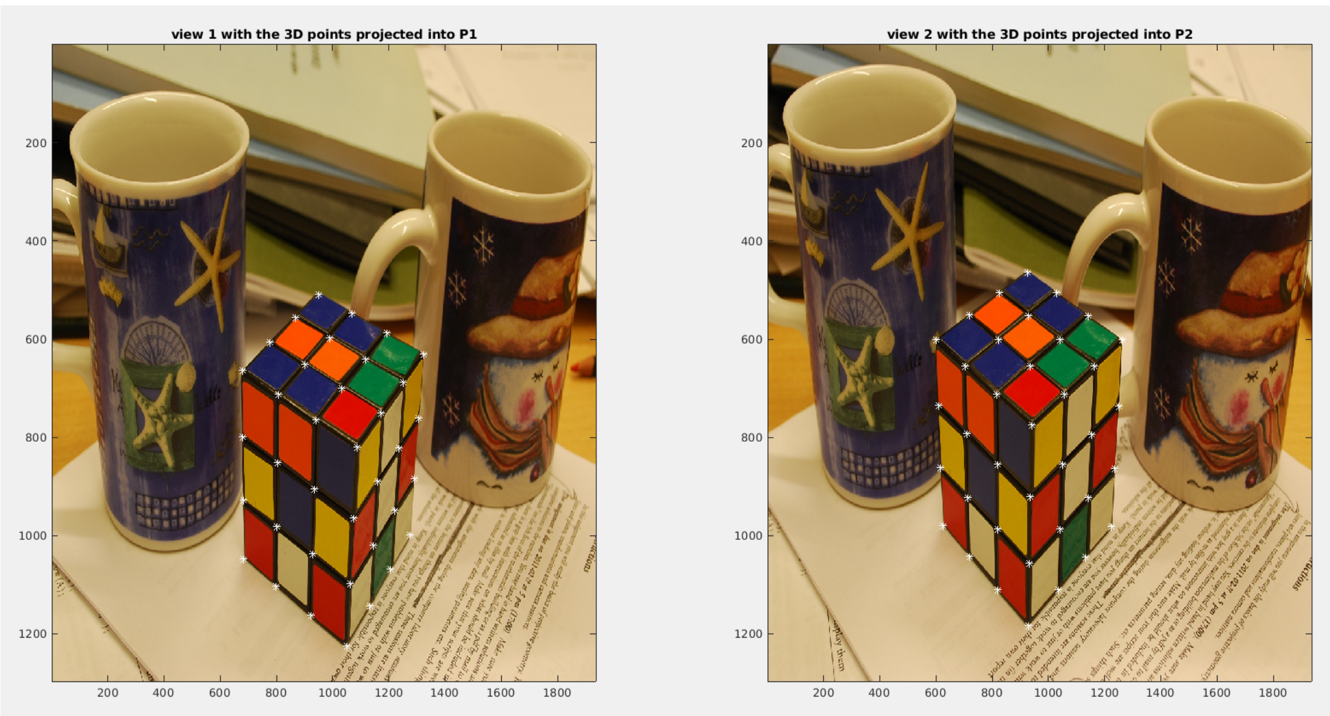


Figure 7: Image views, with the projected 3D points (hgighlighted in white) projected into the respective de-normalized estimated cameras from the DLT method. The points appear to be close to each other, following the shape of the Rubrik cube in both cases.

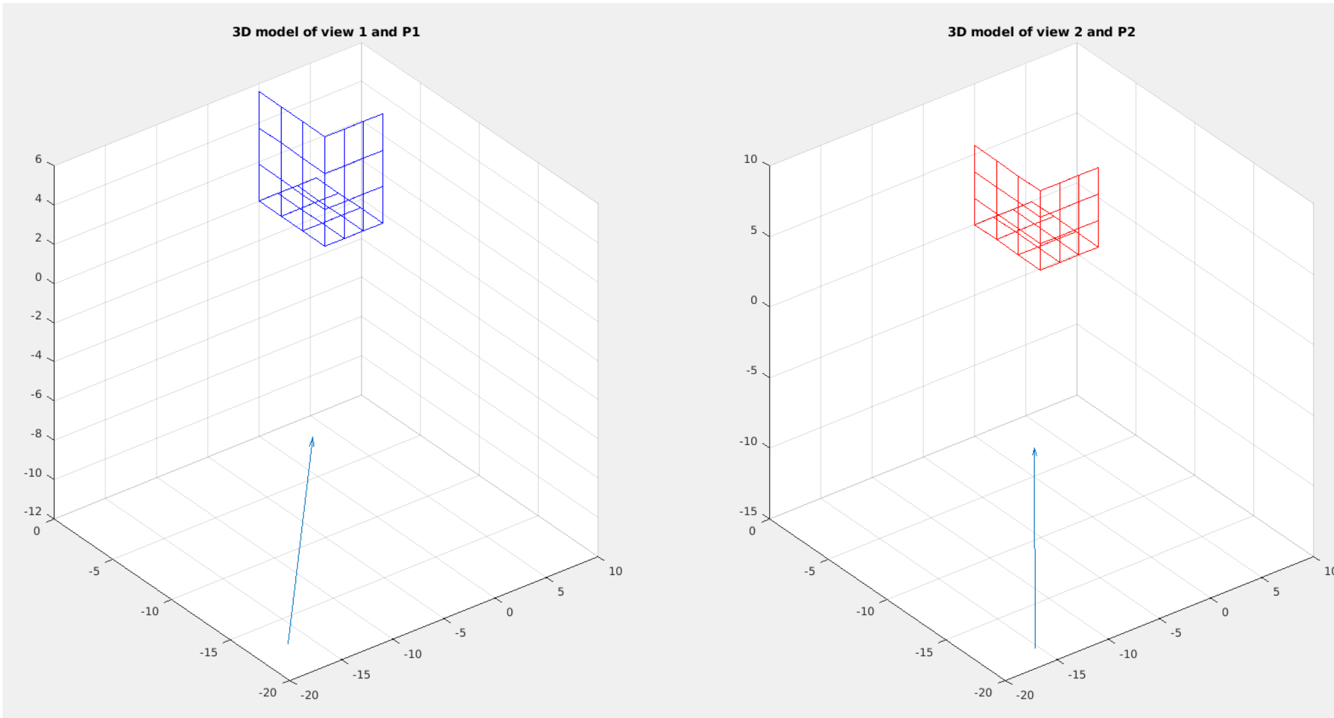


Figure 8: 3D version of the Rubrik cubes and the respective camera centers and principal axes, one for each view. The 3D plot appear to be reasonable, since the camera matrices appear to point in the right side of the of the 3D cube (in relation to the 2D image view). Moreover, the point the 2 cameras are pointing to are the related principal point.

Normalized calibration matrix for the camera P_1:

$$K_1 \approx \begin{pmatrix} 37.4451 & -0.1048 & 980.2049 \\ 0 & 37.4173 & 694.0193 \\ 0 & 0 & 1 \end{pmatrix}$$

In the general case, we obtained the calibration matrix from P_1 and P_2 by using the RQ decomposition method. We can be confident that the results obtained are the "true" parameter, for the following reasons.

- Plotting the normalized principal points to the respective image views the results appear to be reasonable, with the point on the line of the principal axis of the camera (for each view).

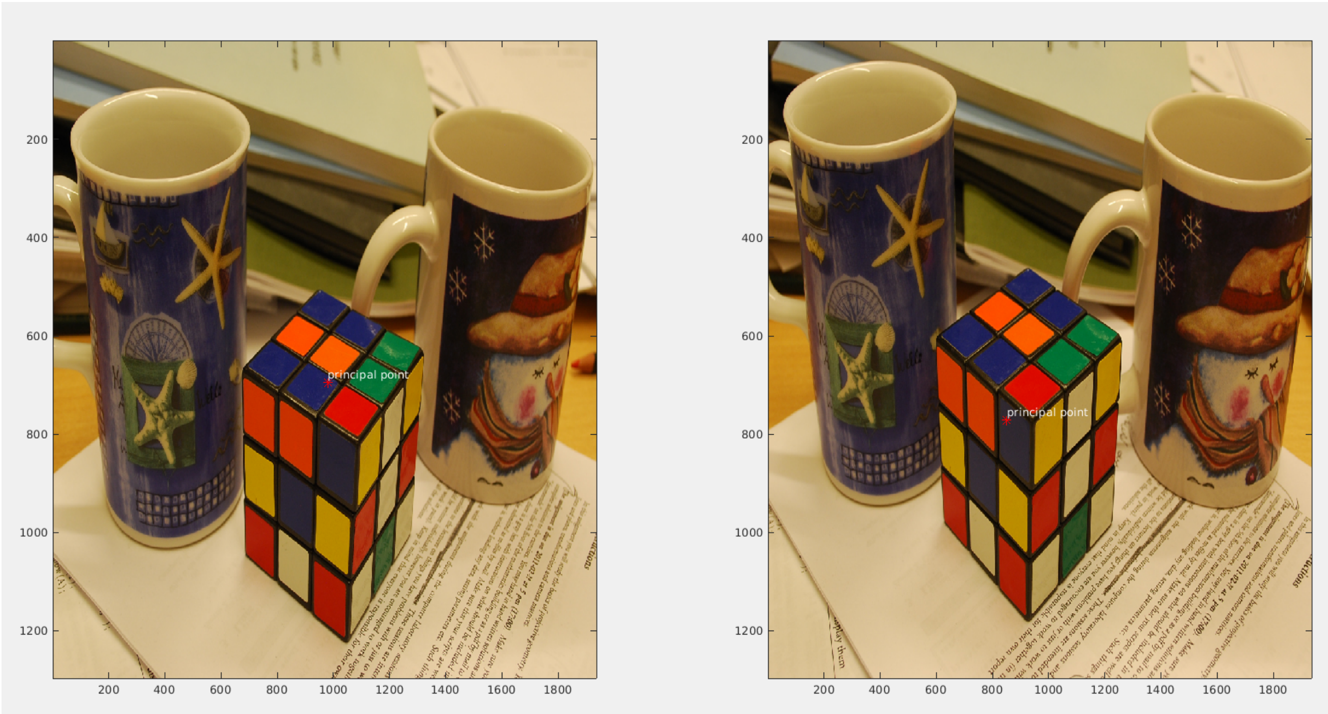


Figure 9: Image views, with their plotted principal point (highlighted in red).

- The matrices follow the upper triangular assumption.
- By normalizing the camera, we will obtain back the estimated camera from the DLT method. In other words, taking for example the first camera:

$$K_1^{-1}P_1 = \tilde{P}_1.$$

Moreover there is no ambiguity as in uncalibrated cases, for the information that the calibration matrix gave us about the camera (from the focal length to eventual skew). Another reason is that the calibration matrices are derived from a camera estimated using DLT method, which use images of a known object by default to eliminate the projective ambiguity.

Optional part

- RMS for points without standardization: 3.571624074460156.
- RMS for normalized points with standardization: 3.571187330646256.
- RMS for filtered points without standardization: 4.191299473674268.
- RMS for normalized filtered points with standardization: 4.187240926018546.

Based on the experiments, standardized points allows us to yield better numerical results for camera resectioning, especially if we have fewer 3D points of the known object to perform DLT. In other words, if we have fewer 3D points the difference between the normalized and non-normalized version is bigger than using all the points.

4 Triangulation using DLT

Computer Exercise 4

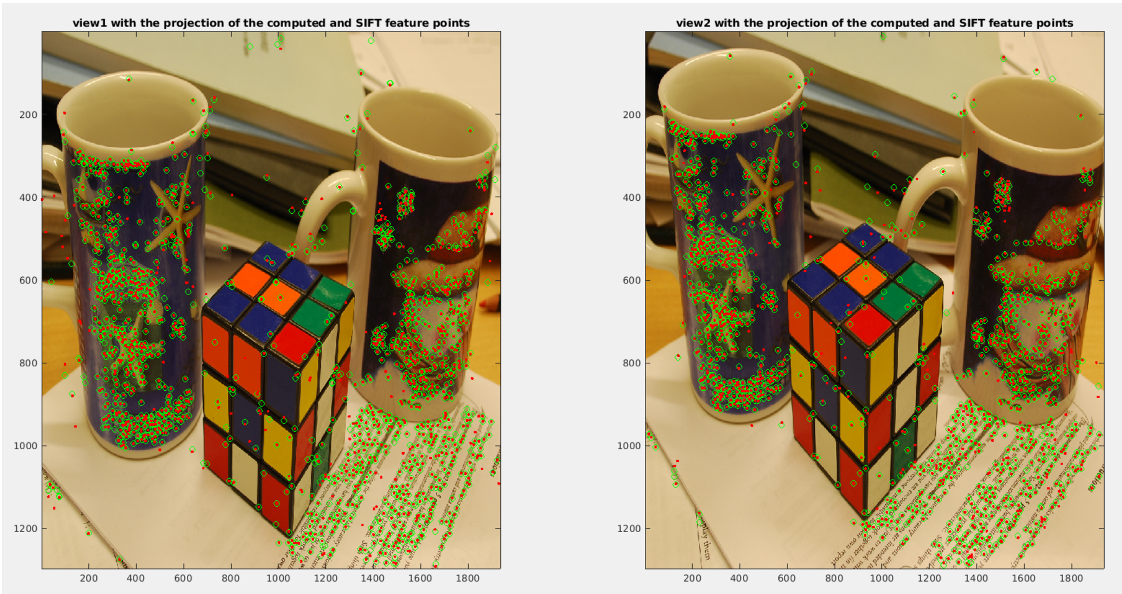


Figure 10: SIFT points (in green) and computed points (in red) into the related views. As we can notice, the results of the triangulation is really close to the extracted points from the SIFT algorithm.

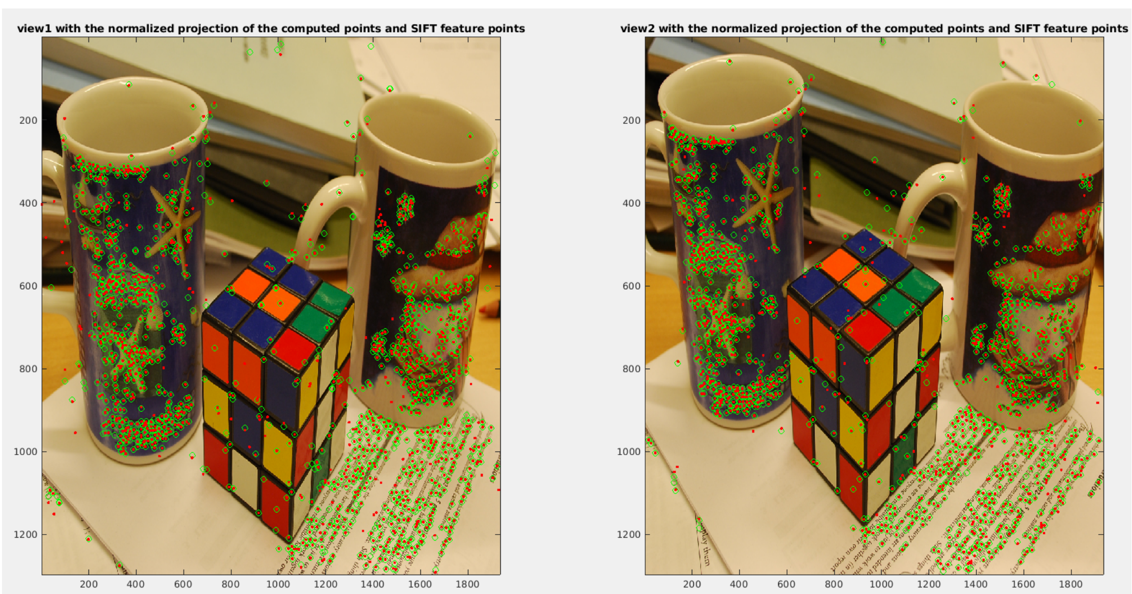


Figure 11: SIFT points (in green) and computed points (in green) into the related views, after normalize both the points and the camera. Compared to the previous plot, we can notice a slight improvement after the normalization, with the computed 3D points closer to the SIFT ones.

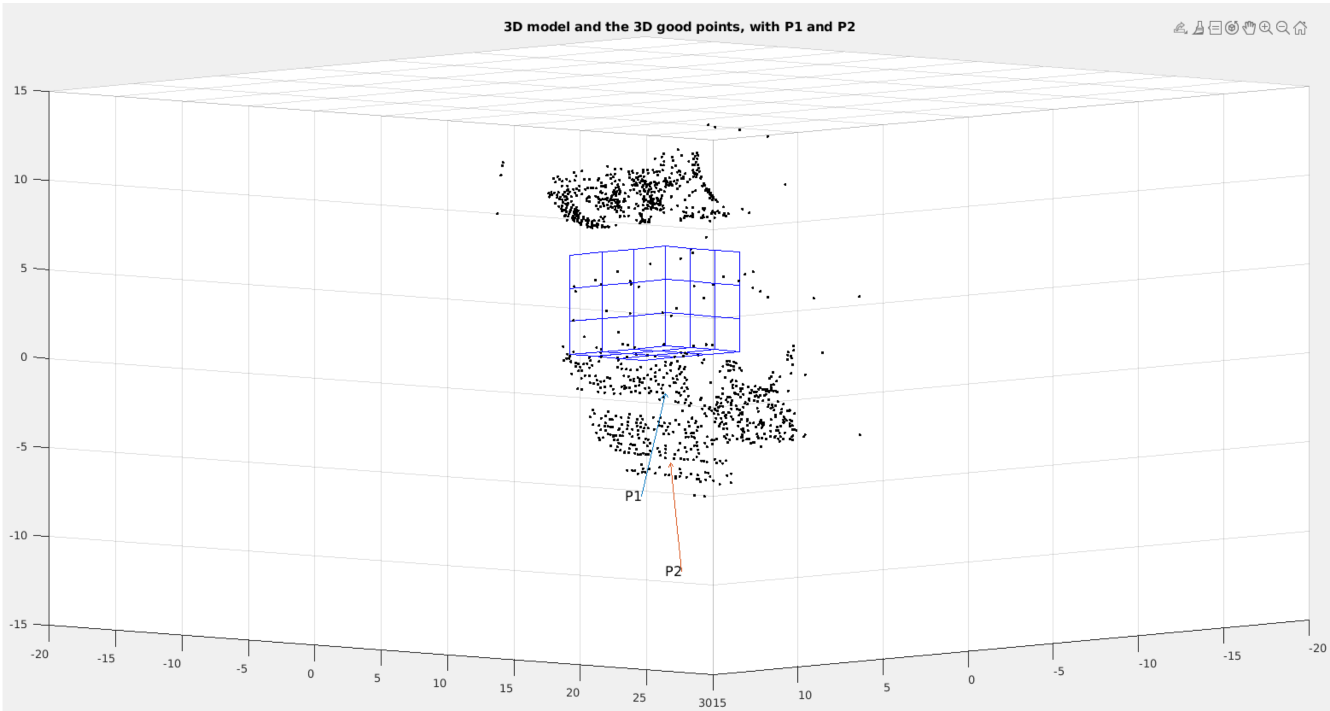


Figure 12: remaining 3D points after the filtering (the "good points"), the cameras and the cube model in one and the same 3D plot. Related to the black 3D points, we can distinguish dominant objects, like the cups and the paper of the image views.