

086761 - Homework 2

Yuri Feldman, 309467801 yurif@cs.technion.ac.il
Alexander Shender, 328626114 aka.sova@gmail.com

November 18, 2017

1

1.1

Substitute $\Lambda^{-1} = \Sigma$, $\eta = \Sigma^{-1}\mu$:

$$p(x) = \frac{\exp\left(-\frac{1}{2}\mu^T\Sigma^{-T}\Sigma\Sigma^{-1}\mu\right)}{\sqrt{\det 2\pi\Sigma}} \exp\left(-\frac{1}{2}x^T\Sigma^{-1}x + \mu^T\Sigma^{-1}\right) = \quad (1.1)$$

$\Sigma^{-1} = \Sigma^{-T}$ since it is symmetric

$$= \frac{\exp\left(-\frac{1}{2}(x^T\Sigma^{-1}x - 2\mu^T\Sigma^{-1} + \mu^T\Sigma^{-1}\mu)\right)}{\sqrt{\det 2\pi\Sigma}} = \quad (1.2)$$

$$= \frac{1}{\sqrt{\det 2\pi\Sigma}} \exp\left\{-\frac{1}{2}(x - \mu)^T\Sigma^{-1}(x - \mu)\right\} \quad (1.3)$$

the latter is the original multivariate Gaussian.

2

2.1

$$p(x) = N(\hat{x}_0, \Sigma_0) = \frac{1}{\sqrt{\det(2\pi\Sigma_0)}} e^{-\frac{1}{2}\|x - \hat{x}_0\|_{\Sigma_0}^2} \quad (2.1)$$

$$p(z | x) = N(h(x), \Sigma_v) = \frac{1}{\sqrt{\det(2\pi\Sigma_v)}} e^{-\frac{1}{2}\|z - h(x)\|_{\Sigma_v}^2} \quad (2.2)$$

2.2

$$p(x | z_1) = \frac{p(z_1 | x)p(x)}{\int_x p(z_1 | x)p(x)dx} \quad (2.3)$$

2.3

We focus on the exponent (omitting the $-\frac{1}{2}$ factor) in the above product, since all the rest is normalization factor (does not depend on x).

$$\|h(x) - z_1\|_{\Sigma_v}^2 + \|x - \hat{x}_0\|_{\Sigma_0}^2 \approx \quad (2.4)$$

linearization around \hat{x}_0 , H the Jacobian of h at \hat{x}_0 :

$$\approx \|h(\hat{x}_0) + H(x - \hat{x}_0) - z_1\|_{\Sigma_v}^2 + \|x - \hat{x}_0\|_{\Sigma_0}^2 = \quad (2.5)$$

$$= x^T (H^T \Sigma_v^{-1} H + \Sigma_0^{-1}) x + 2 ((h(\hat{x}_0) - H\hat{x}_0 - z_1)^T \Sigma_v^{-1} H - \hat{x}_0^T \Sigma_0^{-1}) x + \dots \quad (2.6)$$

equating terms with the exponent of a MVN, we get:

$$\Sigma_1 = (H^T \Sigma_v^{-1} H + \Sigma_0^{-1})^{-1} \quad (2.7)$$

$$\hat{x}_1 = \Sigma_1 (\Sigma_0^{-1} \hat{x}_0 - H^T \Sigma_v^{-1} (h(\hat{x}_0) - H\hat{x}_0 - z_1)) \quad (2.8)$$

Can iteratively use obtained \hat{x}_1 as the new linearization point (re-calculate jacobian H) to improve estimate until convergence (Gauss Newton).

2.4

$$p(x | z_1, z_2) \propto p(z_1, z_2 | x)p(x) = \quad (2.9)$$

assume noise v_1, v_2 independent

$$= p(z_1 | x)p(z_2 | x)p(x) \quad (2.10)$$

We again get a Gaussian, can obtain mean and variance by minimizing the exponent (again omitting the $-\frac{1}{2}$ factor)

$$\|h(x) - z_2\|_{\Sigma_v}^2 + \|h(x) - z_1\|_{\Sigma_v}^2 + \|x - \hat{x}_0\|_{\Sigma_0}^2 \quad (2.11)$$

Again need to pick a linearization point (e.g. \hat{x}_1 is a sensible one), denote \bar{x} as in the lecture, then

$$\approx \|Hx + h(\bar{x}) - H\bar{x} - z_2\|_{\Sigma_v}^2 + \|Hx + h(\bar{x}) - H\bar{x} - z_1\|_{\Sigma_v}^2 + \|x - \hat{x}_0\|_{\Sigma_0}^2 \quad (2.12)$$

Following the same derivation as before, it's easy to see that

$$\Sigma_2 = (2H^T \Sigma_v^{-1} H + \Sigma_0^{-1})^{-1} \quad (2.13)$$

$$\hat{x}_2 = \Sigma_2 \left(\Sigma_0^{-1} \hat{x}_0 - 2H^T \Sigma_v^{-1} (h(\hat{x}_0) - H\hat{x}_0 - \frac{z_1 + z_2}{2}) \right) \quad (2.14)$$

And again as in the previous clause, we can re-linearize to get better estimates with Gauss-Newton iterations.

mind the given data in the question...

you were asked to consider $p(x|z_1)$ as a known dist. (from last section)

hence it leaves the development to be

$p(z_2|x)p(x|z_1)$ rather than $p(z_2|x)p(z_1|x)p(x)$

For intuition, its equivalent for adding new factors to existing belief

[-2]

3

3.1

$$p(x_k \mid x_{k-1}, u_{k-1}) = N(f(x_{k-1}, u_{k-1}), \Sigma_w) = \frac{1}{\sqrt{\det(2\pi\Sigma_w)}} e^{-\frac{1}{2}\|x_k - f(x_{k-1}, u_{k-1})\|_{\Sigma_w}^2} \quad (3.1)$$

3.2

$$p(x_1 \mid z_1, u_0) = \frac{p(z_1 \mid x_1) \int_{x_0} p(x_1 \mid x_0, u_0) p(x_0) dx_0}{\int_{x_1, x_0} p(z_1 \mid x_1) p(x_1 \mid x_0, u_0) p(x_0) dx_1 dx_0} \quad (3.2)$$

3.3

From previous clause:

$$x_1^{MAP} \doteq \arg \max_{x_1} p(x_1 \mid z_1, u_0) = \arg \min_{x_1} -\log(p(z_1 \mid x_1)p(x_1)) = \quad (3.3)$$

$$\arg \min_{x_1} \|z_1 - h(x_1)\|_{\Sigma_v}^2 - \log \int_{x_0} p(x_1 \mid x_0, u_0) p(x_0) dx_0 \quad (3.4)$$

The integrand has the form of a gaussian, however with the left term exponent with generally nonlinear dependence on x_0 :

$$p(x_1 \mid x_0, u_0) p(x_0) \propto \exp\left\{-\frac{1}{2} \left(\|\Sigma_w^{-\frac{1}{2}}(x_1 - f(x_0, u_0))\|^2 + \|\Sigma_0^{-\frac{1}{2}}(x_0 - \hat{x}_0)\|^2 \right) \right\} \quad (3.5)$$

Linearizing f around \hat{x}_0 , h around $\hat{x}_1 = f(\hat{x}_0, u_0)$:

$$\approx \exp\left\{-\frac{1}{2} \left(\|\Sigma_w^{-\frac{1}{2}}(\hat{x}_1 + H_f(x_0 - \hat{x}_0) - x_1)\|^2 + \|\Sigma_0^{-\frac{1}{2}}(x_0 - \hat{x}_0)\|^2 \right) \right\} = \quad (3.6)$$

$$= \exp\left\{-\frac{1}{2} \left(\|\Sigma_w^{-\frac{1}{2}}(H_f - I) \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} + \Sigma_w^{-\frac{1}{2}}(\hat{x}_1 - H_f \hat{x}_0)\|^2 \right) \right\} \quad (3.7)$$

$$+ \left\| \begin{pmatrix} \Sigma_0^{-\frac{1}{2}} & 0 \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} - \Sigma_0^{-\frac{1}{2}} \hat{x}_0 \right\|^2 \} \quad (3.8)$$

this term is a Gaussian over $x_{0:1}$, and its marginalization w.r.t. x_0 yields a Gaussian in x_1 only, and everything collapses to a (linear, if we also marginalize h around \hat{x}_1)

least-squares problem, of which the solution may be refined by repeated re-linearization as in the previous problem.

Could probably have a somewhat easier derivation if had marginalized the entire posterior w.r.t. x_0 .

3.4

Joint covariance $\Sigma_{0:1}$ is of the same size as $x_{0:1}x_{0:1}^T$ (that is, square, with side the sum of vector dimensions). If $x_0 \in \mathbb{R}^n$ and $x_1 \in \mathbb{R}^m$ then $\Sigma_{0:1} \in \mathbb{R}^{n+m \times n+m}$, $\Sigma_{00} \in \mathbb{R}^{n \times n}$, $\Sigma_{11} \in \mathbb{R}^{m \times m}$, $\Sigma_{01} \in \mathbb{R}^{n \times m}$.

Marginalization in covariance form: $\Sigma'_1 = \Sigma_{11}$. Marginalization in information form: $I'_1 = I_{11} - I_{01}^T I_{00}^{-1} I_{01}$ **you could have stopped here...**

We hence continue from Eq. (3.8). Denoting the integrand's information matrix as $I_{0:1}^-$, we get

$$I_{0:1}^- = \begin{pmatrix} H_f^T \Sigma_w^{-1} H_f + \Sigma_0^{-1} & -H_f^T \Sigma_w^{-1} \\ -\Sigma_w^{-1} H_f & \Sigma_w^{-1} \end{pmatrix} \quad (3.9)$$

Linearizing the "innovation" term around \hat{x}_1 we get (denote H_h the Jacobian of h at \hat{x}_1)

$$\|z_1 - h(x_1)\|_{\Sigma_v}^2 \approx \|h(\hat{x}_1) + H_h(x_1 - \hat{x}_1) - z_1\|_{\Sigma_v}^2 = \quad (3.10)$$

$$= \|\Sigma_v^{-\frac{1}{2}} (0 \quad H_h) \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} + \Sigma_v^{-\frac{1}{2}} (h(\hat{x}_1) - H_h \hat{x}_1 - z_1)\|^2 \quad (3.11)$$

This means that the information matrix of the joint distribution is:

$$I_{0:1} = \begin{pmatrix} H_f^T \Sigma_w^{-1} H_f + \Sigma_0^{-1} & -H_f^T \Sigma_w^{-1} \\ -\Sigma_w^{-1} H_f & H_h^T \Sigma_v^{-1} H_h + \Sigma_w^{-1} \end{pmatrix} \quad (3.12)$$

From here, using marginalization formula:

$$I'_1 = H_h^T \Sigma_v^{-1} H_h + \Sigma_w^{-1} - \Sigma_w^{-1} H_f (H_f^T \Sigma_w^{-1} H_f + \Sigma_0^{-1})^{-1} H_f^T \Sigma_w^{-1} \quad (3.13)$$

And Σ'_1 is the inverse of the above.

Hands-on exercises

Question 1.

(a) Based on the given information, the projection matrix is given by:

Assumption: skew = 0

$$P = K_{3 \times 3} [R | t]_{(3 \times 4)} = \begin{bmatrix} f_x & s & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} [R_G^C_{3 \times 3} | R_G^C \cdot t_{C \rightarrow G_{3 \times 1}}]$$

$$P = \begin{bmatrix} 480 & 0 & 320 \\ 0 & 480 & 270 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0.5363 & -0.8440 & 0 & -458.9384 \\ 0.8440 & 0.5363 & 0 & -243.0052 \\ 0 & 0 & 1 & 400 \end{bmatrix}$$

$$P = \begin{bmatrix} 257.42 & -405.12 & 320 & -92290.4 \\ 405.12 & 257.42 & 270 & -8642.48 \\ 0 & 0 & 1 & 400 \end{bmatrix}$$

(b) In total, the projection of a 3D point onto the screen is performed in 3 stages:

1. The points is translated and rotated into the camera coordinates using Rotation + translation matrix
2. The point is translated into the homogeneous coordinates
3. The homogenous coordinates are being normalized into the regular pixel position coordinates

Solving each step separately:

1. Convert the 3D point into Camera coordinates

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = [R | t] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \begin{bmatrix} 0.5363 & -0.8440 & 0 & -458.9384 \\ 0.8440 & 0.5363 & 0 & -243.0052 \\ 0 & 0 & 1 & 400 \end{bmatrix} \begin{bmatrix} 350 \\ -250 \\ -35 \\ 1 \end{bmatrix} = \begin{bmatrix} -60.2334 \\ -81.6802 \\ 365 \end{bmatrix}$$

2. Get the homogenous coordinates:

$$\begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix} = K_{3 \times 3} \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \begin{bmatrix} 87887.99 \\ 59343.52 \\ 365 \end{bmatrix}$$

3. Obtaining the normalized (real) coordinates:

$$\begin{bmatrix} u \\ v \\ \bar{w} \end{bmatrix} = \begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix} = \begin{bmatrix} u \\ v \\ \bar{w} \end{bmatrix} = \begin{bmatrix} 240.78 \\ 162.58 \end{bmatrix}$$

(c) Re-Projection error is defined as:

$$v = z - \pi(x, l)$$

In our case:

$$\begin{aligned} z &= \begin{bmatrix} 241.5 \\ 169 \end{bmatrix} && \text{remember its pixels => integer} \\ \pi(x, l) &= \begin{bmatrix} 240.78 \\ 162.58 \end{bmatrix} && \text{hence} \\ &&& [241; 163] \end{aligned}$$

Meaning:

$$v = \begin{bmatrix} 241.5 \\ 169 \end{bmatrix} - \begin{bmatrix} 240.78 \\ 162.58 \end{bmatrix} = \sim \begin{bmatrix} 0.71 \\ 6.41 \end{bmatrix}$$

Question 2 – (c) - camera calibration

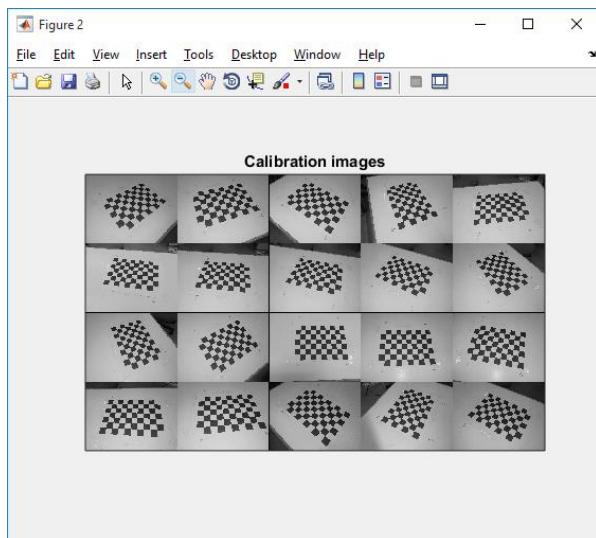
Here are presented the snapshots of the calibration process.

1. First, the images of the checkerboard plane have been taken from various camera poses. The camera used is the Huawei P9 dual-lenses camera. Thus, in order to keep the process authentic, one of the lenses was covered. Each of the corners was marked by a number for a more convenient calibration process.

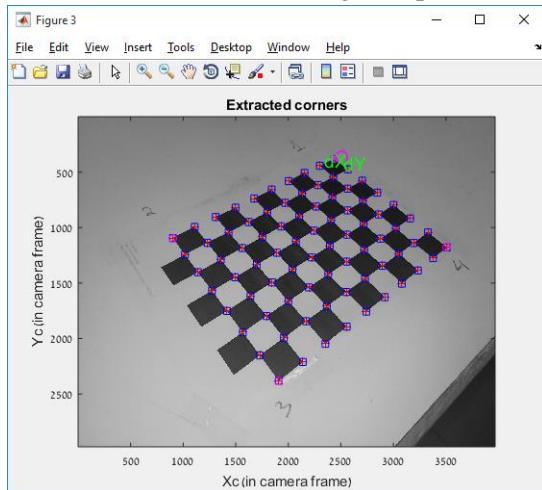


Figure 1 : the camera used in the calibration process

2. Following the instructions, the images are being uploaded into the calibration program:



3. Corner extraction of the images, repeated on all the images



4. Calibration parameters are obtained:

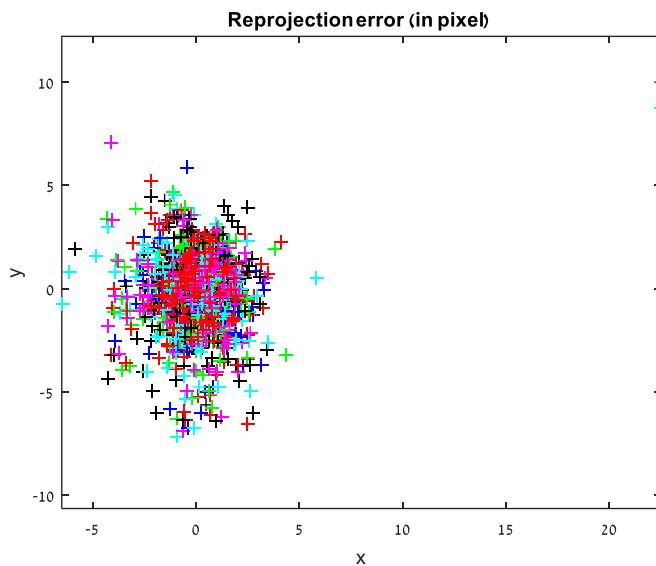
```
Focal Length:      fc = [ 3189.24204  3189.24204 ]
Principal point: cc = [ 1983.50000  1487.50000 ]
Skew:            alpha_c = [ 0.00000 ] => angle of pixel = 90.00000 degrees
Distortion:       kc = [ 0.00000  0.00000  0.00000  0.00000  0.00000 ]
```

Main calibration optimization procedure - Number of images: 19
 Gradient descent iterations: 1...2...3...4...5...6...7...8...9...10...11...12...13...14...15...16...17...18...19...20...21...done
 Estimation of uncertainties...done

Calibration results after optimization (with uncertainties):

```
Focal Length:      fc = [ 3020.08805  3056.13647 ] +/- [ 39.75145  35.31200 ]
Principal point: cc = [ 1999.65339  1592.93877 ] +/- [ 9.42775  35.21215 ]
Skew:            alpha_c = [ 0.00000 ] +/- [ 0.00000 ] => angle of pixel axes = 90.00000 +/- 0.00000 degrees
Distortion:       kc = [ 0.10233  -0.31015  -0.00415  0.00271  0.00000 ] +/- [ 0.00874  0.03295  0.00096  0.00121
0.00000 ]
Pixel error:     err = [ 1.41622  1.74597 ]
```

5. The reprojection error looks quiet large:



Our aim is to decrease it to the biggest scale.

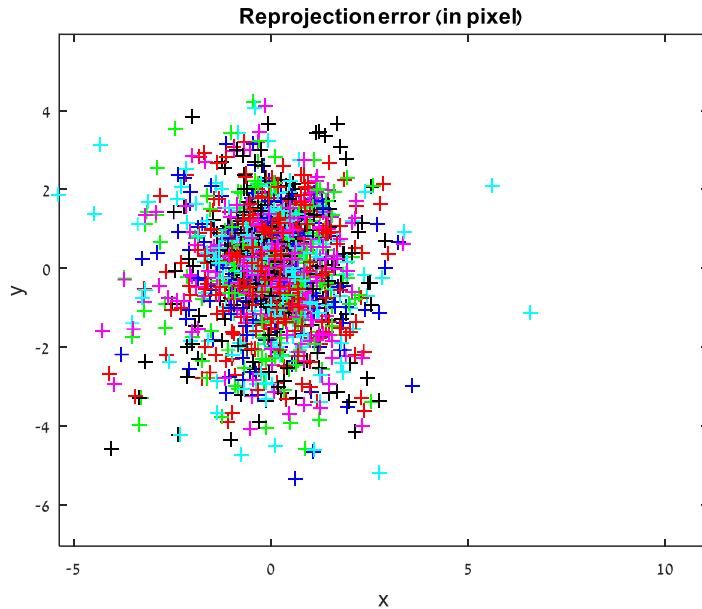
6. Using the Recomp.corners function provided in the GUI, a new calibration is performed, resulting in the following calibration results:

Calibration results after optimization (with uncertainties):

```
Focal Length:      fc = [ 3043.25574  3075.15744 ] +/- [ 28.86633  26.23904 ]
Principal point: cc = [ 2000.49865  1569.94521 ] +/- [ 7.74148  24.72921 ]
Skew:            alpha_c = [ 0.00000 ] +/- [ 0.00000 ] => angle of pixel axes = 90.00000 +/- 0.00000 degrees
Distortion:       kc = [ 0.12149 -0.40301 -0.00409  0.00292  0.00000 ] +/- [ 0.00753  0.02930  0.00080  0.00097
0.00000 ]
Pixel error:     err = [ 1.20546  1.47363 ]
```

It may be observed that the Pixel error has decreased in comparison to the first calibration results.

The reprojection error has decreased as well:



7. Analyzing some of the critical errors (in this example – the points in teal color in the reprojection error graph from step 6):

Pixel error: err = [1.20546 1.47363] (all active images)

Selected image: 17

Selected point index: 11

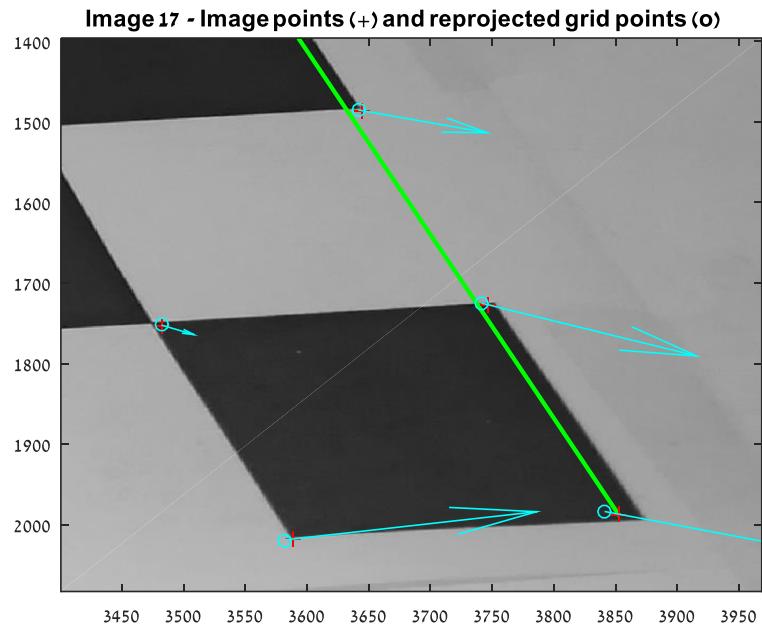
Pattern coordinates (in units of (dX,dY)): (X,Y)=(0,6)

Image coordinates (in pixel): (3746.05,1726.11)

Pixel error = (5.61498,2.11331)

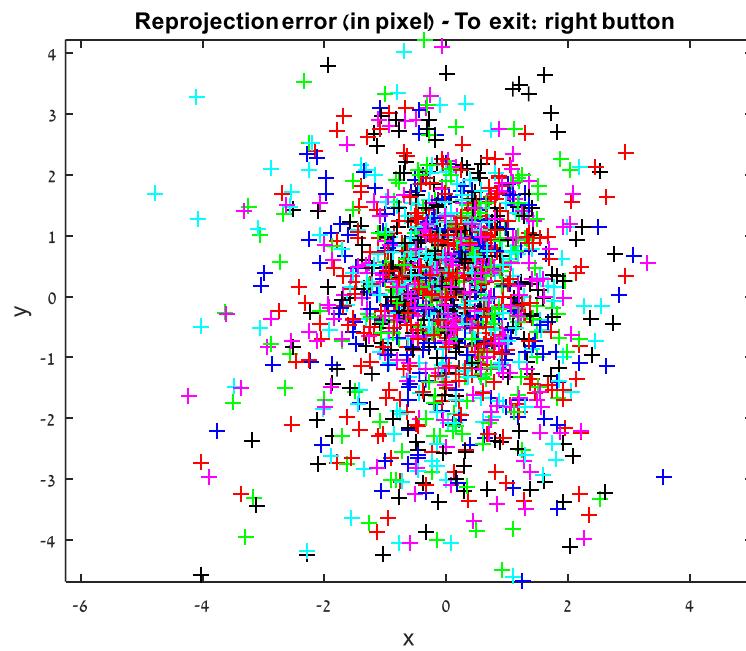
Window size: (wintx,winty) = (5,5)

As it may be seen, image 17 was not processed in a correct way, perhaps the corners were poorly extracted. By checking the image, it is validated:



Using a larger values for $wintx$ and $winty$ parameters should help to increase the ability to find the problem.

By using the values $\begin{cases} wintx = 8 \\ winty = 8 \end{cases}$, the following performance improvement is noticed:



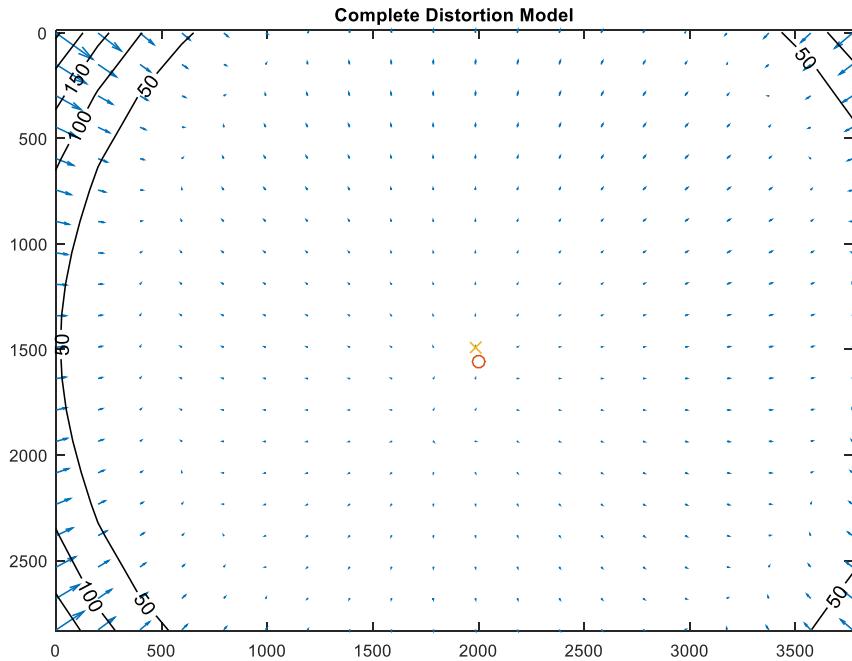
By looking at numerical representation of the calibration, it is obtained:

Calibration results after optimization (with uncertainties):

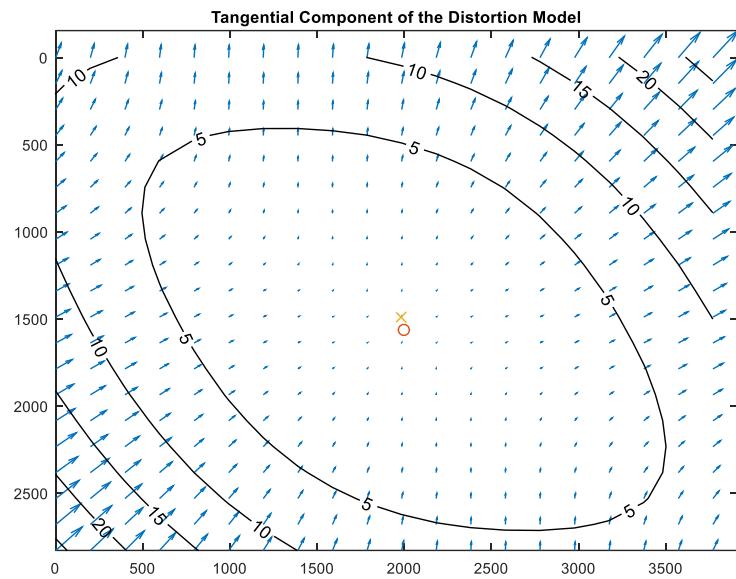
Focal Length: $fc = [3064.80707 \ 3093.11220] +/- [24.15153 \ 22.46772]$
 Principal point: $cc = [1999.02148 \ 1548.76380] +/- [7.03138 \ 19.94824]$
 Skew: $\alpha_c = [0.00000] +/- [0.00000] \Rightarrow \text{angle of pixel axes} = 90.00000 +/- 0.00000 \text{ degrees}$
 Distortion: $kc = [0.13995 \ -0.49685 \ -0.00421 \ 0.00270 \ 0.00000] +/- [0.00721 \ 0.02887 \ 0.00074 \ 0.00086 \ 0.00000]$
 Pixel error: $err = [1.09649 \ 1.41861]$

As it may be observed, as a consequence from using all those tools, the pixel error has decreased from the initial $[1.41622 \ 1.74597]$ to $[1.09649 \ 1.41861]$

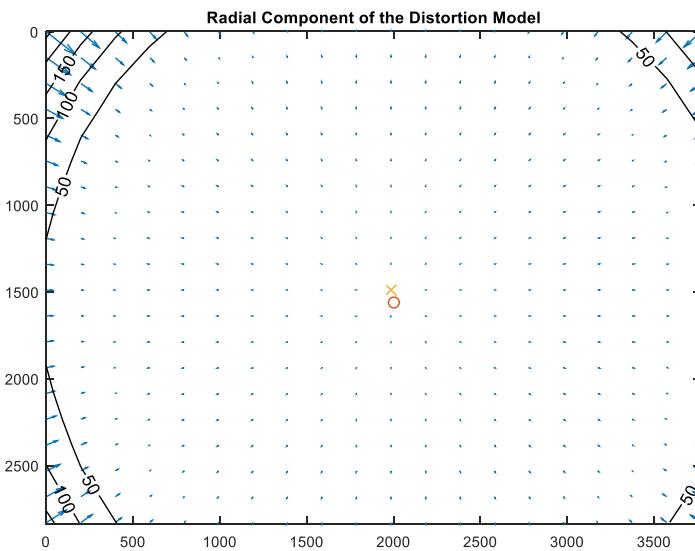
8. Visualizing the distortions creates the following plots. The guess is that indeed, since the smartphone possesses two cameras, the distortion is “shifted” away from the second camera, so that together they create a full “symmetric” distortion model.



Pixel error	= [1.207, 1.472]	
Focal Length	= (3053.92, 3084.16)	+/- [29.44, 26.62]
Principal Point	= (1998.19, 1560)	+/- [7.87, 25.49]
Skew	= 0	+/- 0
Radial coefficients	= (0.1196, -0.3962, 0)	+/- [0.00752, 0.02929, 0]
Tangential coefficients	= (-0.004411, 0.002749)	+/- [0.0008022, 0.000984]



Pixel error	= [1.207, 1.472]	
Focal Length	= (3053.92, 3084.16)	+/- [29.44, 26.62]
Principal Point	= (1998.19, 1560)	+/- [7.87, 25.49]
Skew	= 0	+/- 0
Radial coefficients	= (0.1196, -0.3962, 0)	+/- [0.00752, 0.02929, 0]
Tangential coefficients	= (-0.004411, 0.002749)	+/- [0.0008022, 0.000984]



Pixel error	= [1.207, 1.472]	
Focal Length	= (3053.92, 3084.16)	+/- [29.44, 26.62]
Principal Point	= (1998.19, 1560)	+/- [7.87, 25.49]
Skew	= 0	+/- 0
Radial coefficients	= (0.1196, -0.3962, 0)	+/- [0.00752, 0.02929, 0]
Tangential coefficients	= (-0.004411, 0.002749)	+/- [0.0008022, 0.000984]

9. Using one of the images, the extrinsic parameters were also calculated:

Extrinsic parameters:

Translation vector: $Tc_{ext} = [157.196418 \quad -33.157908 \quad 330.557975]$
 Rotation vector: $omc_{ext} = [0.919513 \quad -2.736548 \quad 0.820951]$
 Rotation matrix: $Rc_{ext} = [-0.803385 \quad -0.594159 \quad 0.039334 \quad -0.517696 \quad 0.664303 \quad -0.539160 \quad 0.294217 \quad -0.453516 \quad -0.841285]$
 Pixel error: $err = [1.70132 \quad 1.94694]$

Question 2 – (d)

By taking the data provided by the toolbox in the previous section of the question, the calibration matrix K is obtained: (values are shortened up to the closest natural number)

$$K = \begin{bmatrix} f_x & s & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 3065 & 0 & 1999 \\ 0 & 3093 & 1549 \\ 0 & 0 & 1 \end{bmatrix}$$

If, for example, we use the image that the extrinsic parameters were extracted for (Question 3 (c) – 9), we obtain the following projection matrix:

$$\begin{aligned} K [R \mid t] &= \begin{bmatrix} f_x & s & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} [R_{3 \times 3} \mid t_{3 \times 1}] = \\ &= \begin{bmatrix} 3065 & 0 & 1999 \\ 0 & 3093 & 1549 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -0.803385 & -0.594159 & 0.039334 & 157.196418 \\ -0.517696 & 0.664303 & 0.039334 & -33.157908 \\ 0.294217 & -0.453516 & -0.841285 & 330.557975 \end{bmatrix} = M_{3 \times 4} \\ M_{3 \times 4} &= \begin{bmatrix} -1874.23 & -2727.67 & -1561.17 & 1142592.4 \\ -1145.49 & 1352.19 & -1181.49 & 409476.89 \\ 0.294 & -0.4535 & -0.8412 & 330.55 \end{bmatrix} \end{aligned}$$

This matrix M is unique for each of the images, since the extrinsic parameters do change for every camera pose.

Question 3

- (a) The images were taken
- (b) Using the provided vlfeat library features and descriptors have been extracted.

- Each of the feature vectors is a 4-row vector:
$$\begin{cases} (f(1:2)) - \text{location of the feature} \\ f(3) - \text{scale} \\ f(4) - \text{orientation} \end{cases}$$
- Each of the descriptor vectors is a 128 elements vectors, describing orientation histogram (for 8 orientations) for each of the 16 cells in the vicinity of the feature ($16 * 8 = 128$).

The vlfeat library allows the usage of customized parameters for feature extraction:

- Peak threshold
- Edge threshold

They define the threshold for a potential feature to be declared a feature. In our example, 57 features were extracted for #1 and 64 features were extracted for Image #2:

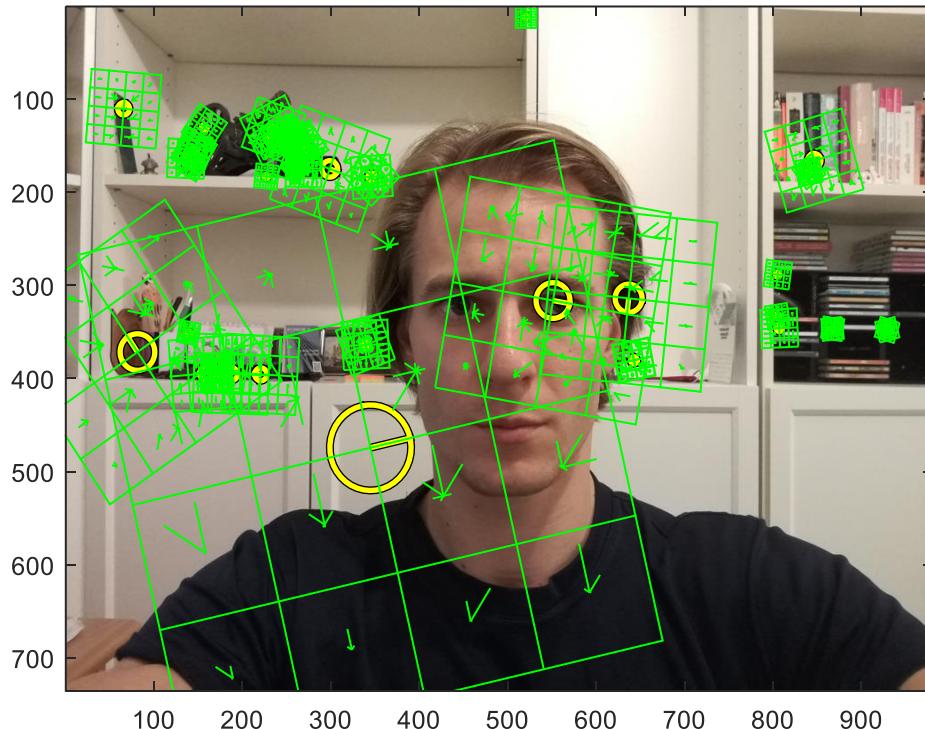


Figure 1 Image 1: 57 features detected

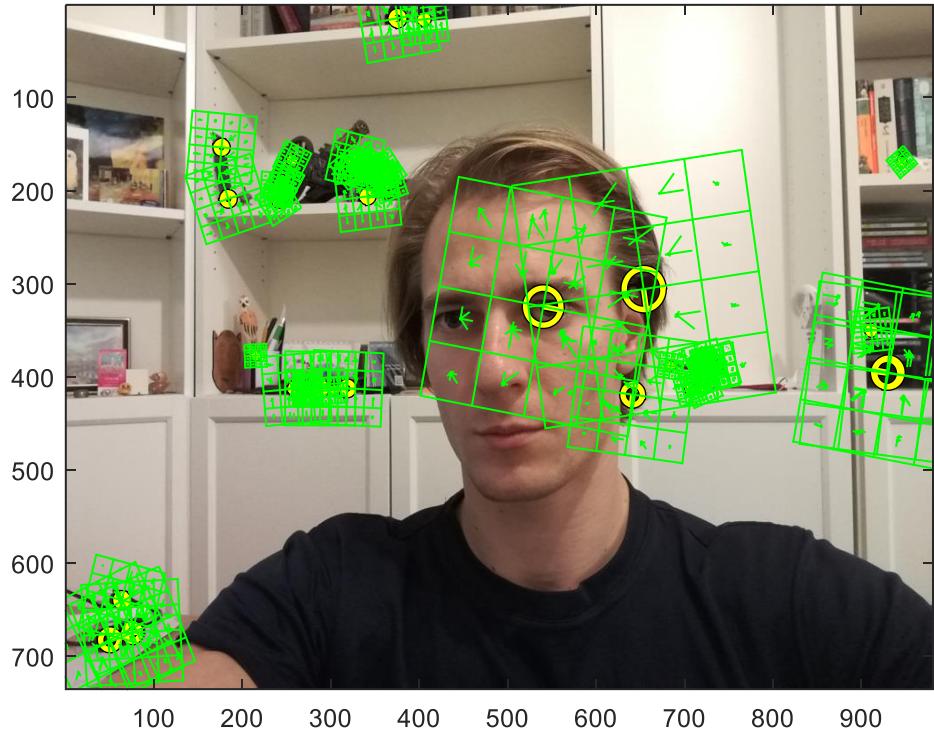
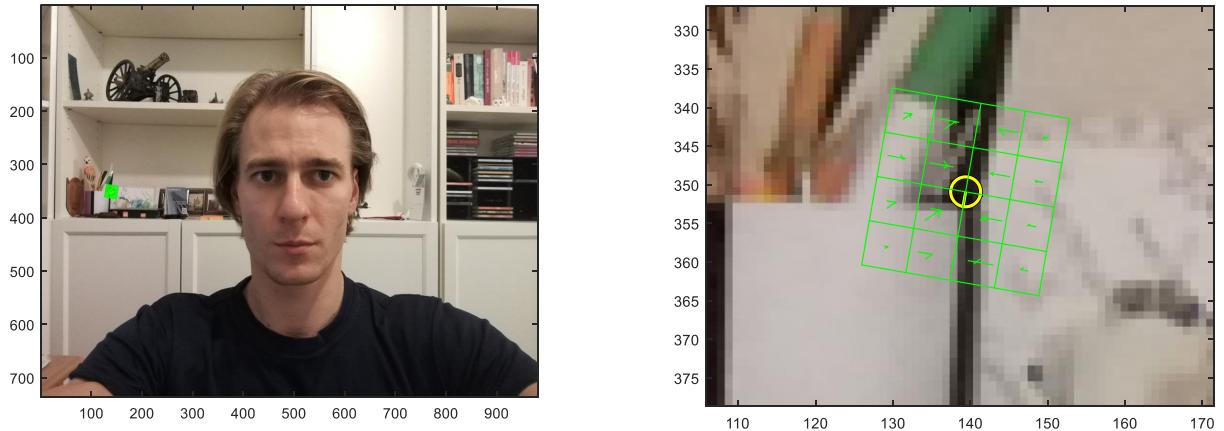


Figure 2 Image2: 64 features detected

Scale and orientation for a representative feature: select arbitrary index 1. The feature 1 on Image1 is the following:
(right image is the close-up)

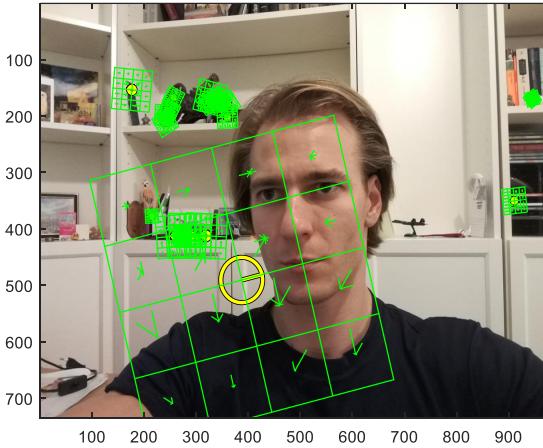
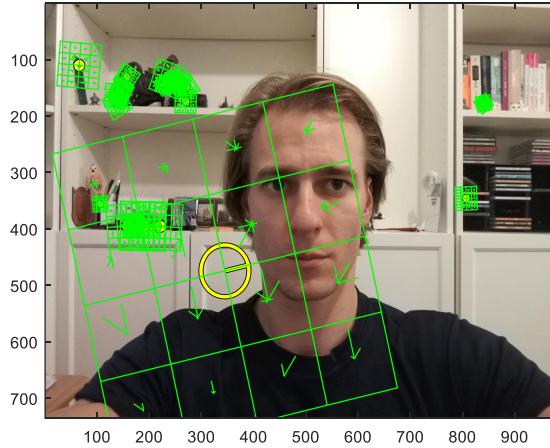


The representative feature is then :
$$\begin{cases} x \\ y \\ scale \\ orientation \end{cases} = \begin{cases} 139.4 \\ 350.91 \\ 1.937 \\ 0.172 \end{cases}$$
; with scale = 1.937 & orientation = 0.172

(c) The vlfeat library does possess the function for the feature matching, but since the question asks to perform the calculation as explained in class, the algorithm is written manually, while the library function helps us to validate the results. The algorithm for feature matching is the following:

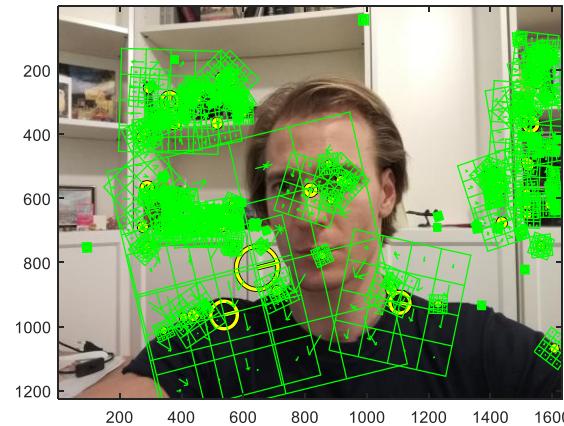
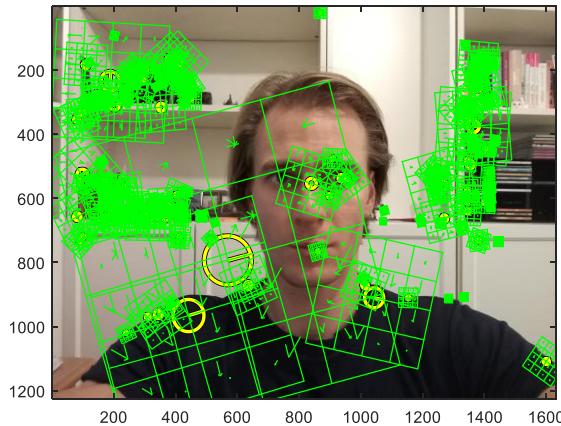
1. Calculate the Euclidean distance from every feature descriptor (128 row vector) in Image 1 to every feature descriptor in Image 2.
2. For every feature in Image 1, find the minimum Euclidean distance out of Euclidean distances calculated to every feature in Image 2 (calculated in step 1)
3. Use the threshold parameter to validate this feature is “unique” enough. Meaning, the minimal Euclidean distance multiplied by threshold is still smaller than any other Euclidean distance to other features. Default value in vlfeat library is 1.5
4. If the condition in Step 3 holds, add this couple of matched features (their indexes) to the matched array.

The features matched by this way using a certain threshold (1.5) may be observed in the following image:



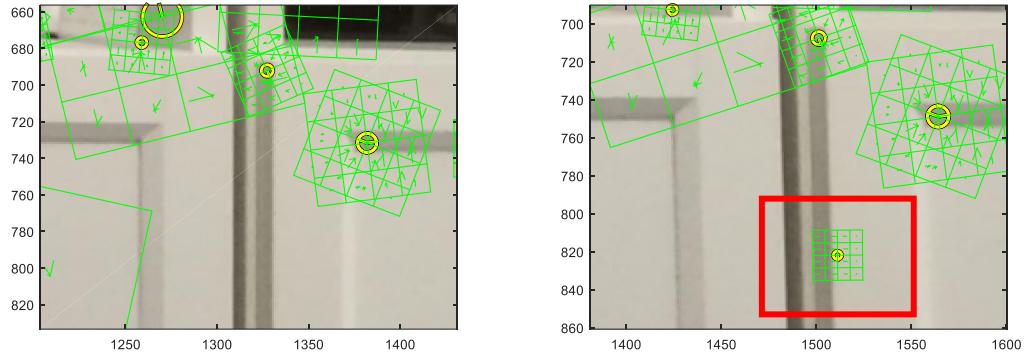
By the first look, it is indeed observable that those features are corresponding to each other. No outliers were found on this image, since the threshold parameters were set quiet to make the features amount small (for easier visual inspection).

In order to find the outliers, the higher resolution image was used, with smaller threshold parameters for feature extraction. As a result, in the following specific example, for Image 1 and 2 : 1188 and 1210 features were extracted, with 237 matches:

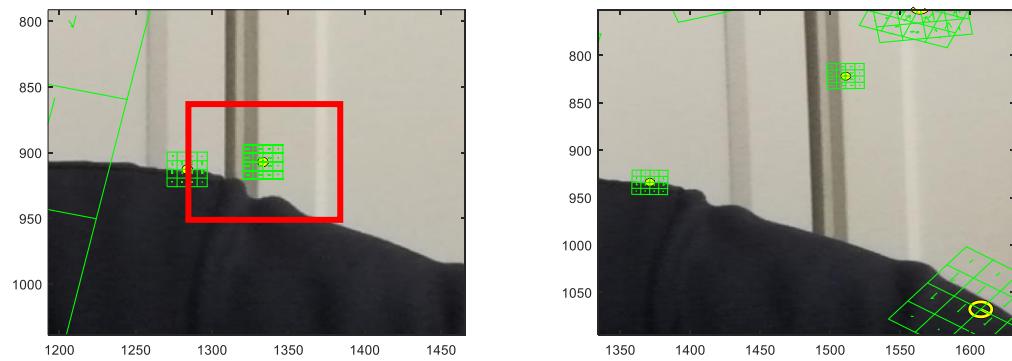


After the fast inspection, the following are the two examples of the outliers: (marked with red boxes) – meaning this feature is identified as match with some other feature from another image, but as we can see, it does not appear in the second image where it should be.

Outlier 1 :

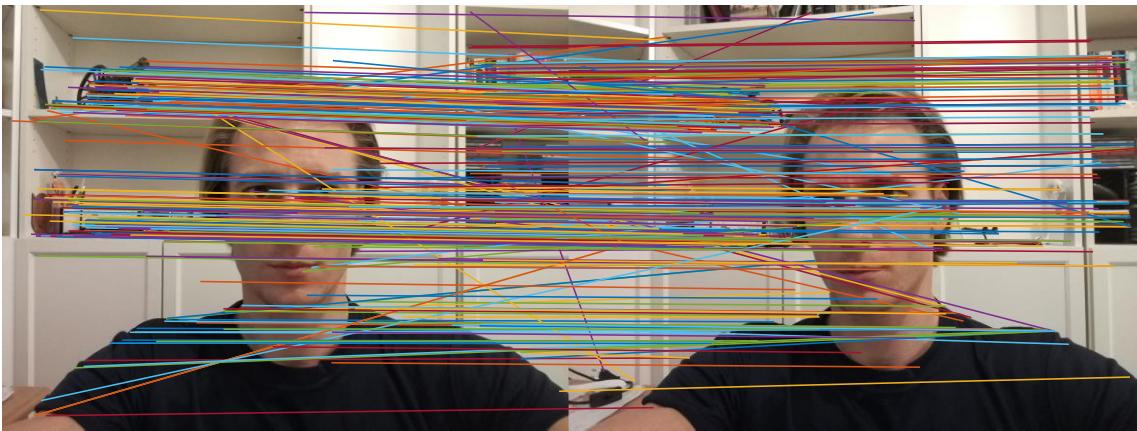


Outlier 2 :

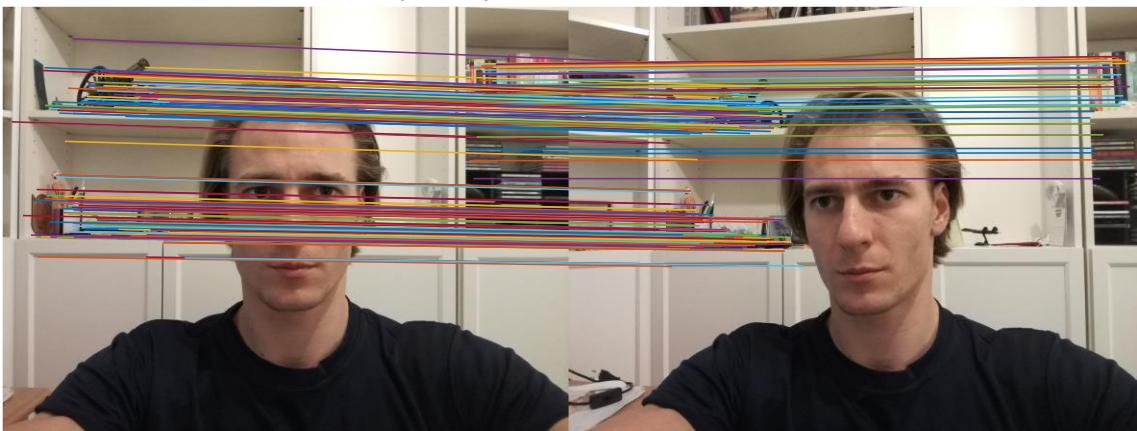


Using another function, provided by the vlfeat library (although which does not appear in the standard library, but its code is public on their website), the “SIFT_MOSAIC”, we may witness the outliers more clearly. In the following image it may be visible how some tentative matches are found, but later declared “outliers” and do not appear in the “inliers” section. The percentage ratio of the inliers also shown, allowing to adjust the parameters:

395 tentative matches

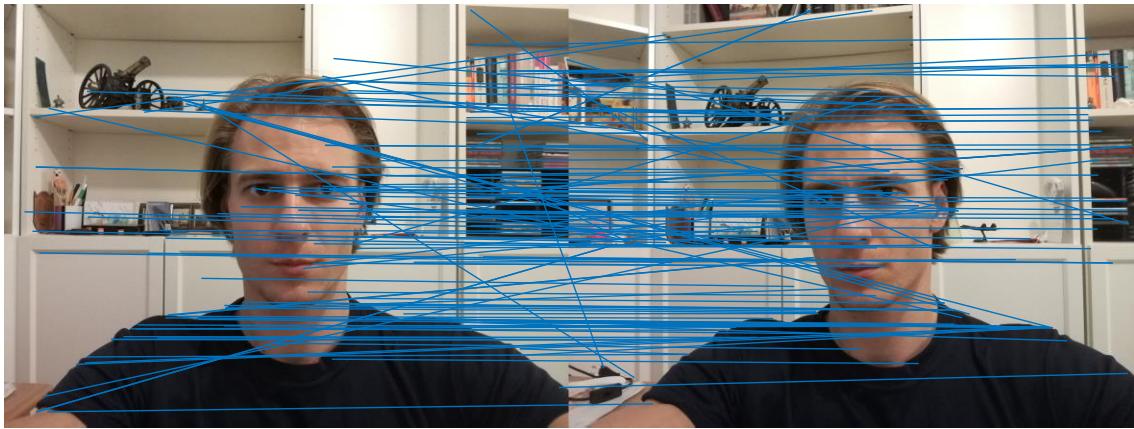


221 (55.95%) inliner matches out of 395



Plotting just the outliers gives the following figure.

145 Outlier matches

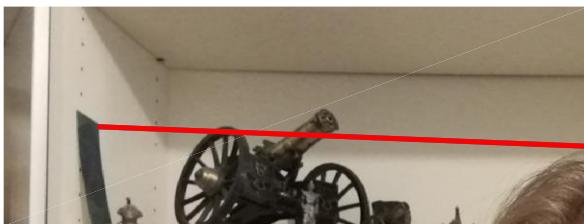


Answering the question, we represent some representative inlier and outlier matches, and discuss possible reasons for their appearance (the outliers):

Inlier representative match: (1st figure shows the general picture, 2nd picture the close-up)



Close-up (looks almost same on both pictures, surely):

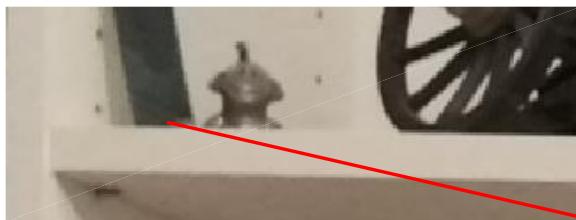


Outliers' representative matches: although apparently Euclidean Distance between 2 feature descriptions represented here was small enough for this couple to be declared a “match”, we can see that this is certainly not the case, and the point is not the same.

Example #1:



Close-up Left Image:



Close-up Right Image:



Example #2:



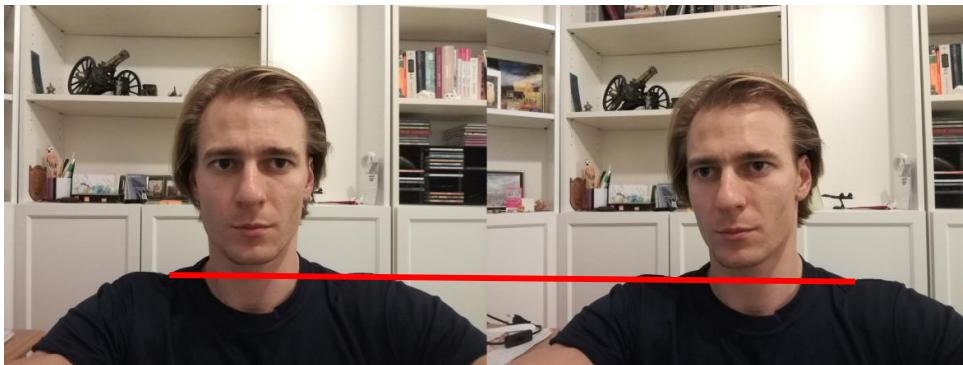
Close-up left image:



Close-up right image:

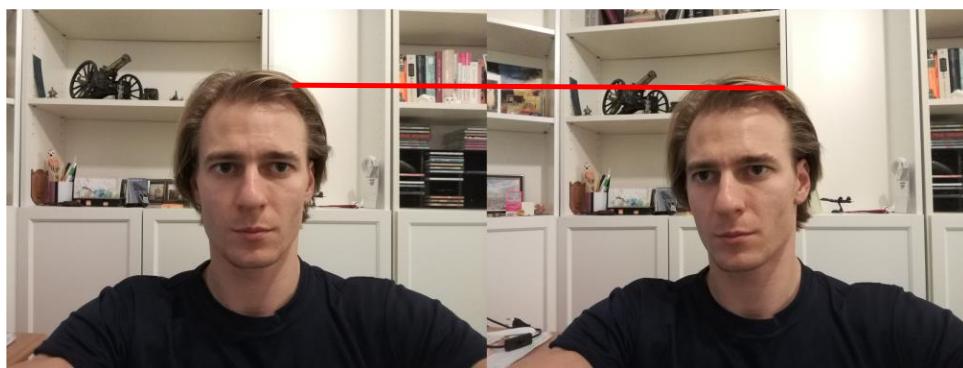


Example #3



I think, no close-up image needed here.

Example #4:



Explanation for possible outliers appearance: several reasons may lead to the creation of the outliers. Here are presented some we have come up with:

1. Since Feature description use the SIFT technique, which is based on the edge orientation, same features may have similar descriptions, although are in fact representing completely different objects. (may be seen in Example 1&2).
2. For the same reason as in 1, but the objects are indeed similar in the image, or appear several times in the same image – both having the similar edges. For example, the shelves behind me, or the eyes. Example 3 shows
3. The environment is dynamic. Each frame is taken at different camera pose, and at different time. A similar feature may appear in a certain image, while it wasn't present in the previous image, but a feature with similar description was. This may be the reason for Example 4 (but not 100%)