HomeWork 3

Computer Vision and Image Processing

Homography and fundamental matrix estimation

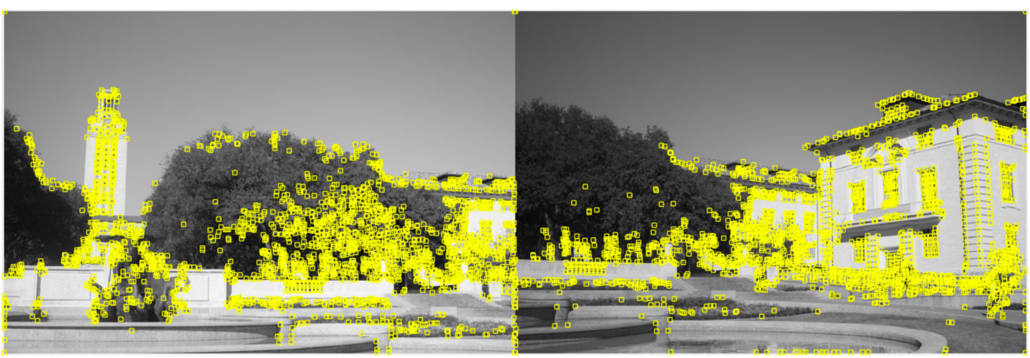
CSE-573 - Fall 2017

Deepika Chaudhary (UB IT Name: d25 , UB Person Number: 50248725)

Analysis and Results

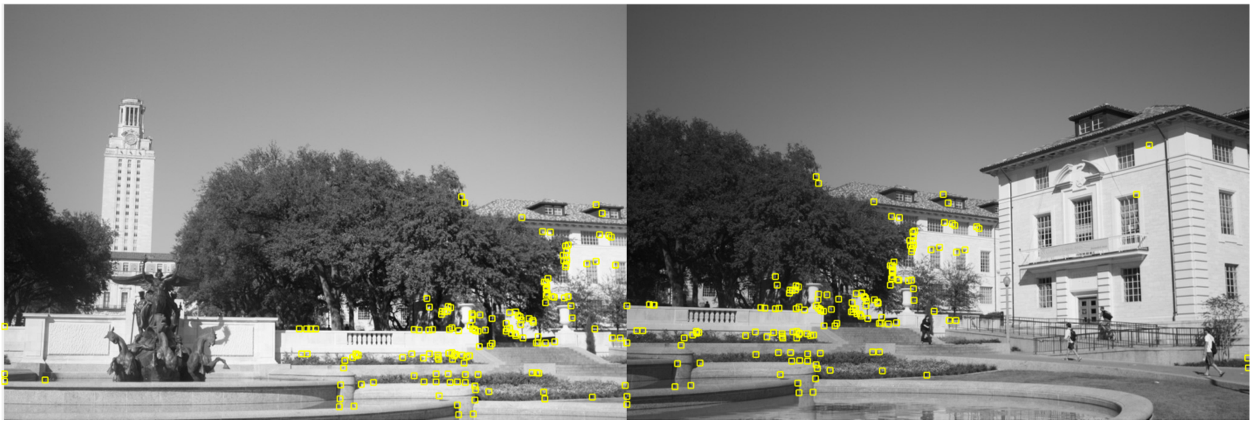
1. Homography Matrix

   a) **Estimating corners**: Harris detector was used in order to estimate the corners in each image. The coordinates of each image corners were noted down for further processing. The threshold for match used was 0.05 with radius of detection as 1 considered in non maximal suppression. The standard deviation used was 1. The result of corners detect in both the images is as below.



Features neighbor Detection: Now after the corners are generated next I calculated the matching features in both the images. Each corner generated depicts a pixel and in order to compare the same features we need information more than a pixel value as same pixel can be detected else where in the image. To solve this, I used neighbor matching techniques. Ideally the chances of having same neighborhood are minimal.
Therefore, for each corner detected pick 21 neighboring pixels in x direction and 21 in the y direction. This gives us a blob of 21 X 21 matrix. This operation is performed on every corner detected in both the images.
Now every blob of one image is compared with every other blob in image 2 using sum of squared differences.
Next the distances are sorted in ascending order and best 200 matches are picked up having the least difference between them.

This is done in file ‘pick_feature_neighbours.m’. The best matches generated are displayed below.



Estimating Homography Matrix: Next task is to calculate the homography matrix in order to determine the features of one image to another using the formula:

$$\begin{bmatrix} wx' \\ wy' \\ w \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

This is done in order to estimate the similarity between 2 images on the same planar surface. The goal is to minimize Ah = 0.
To estimate homography RANSAC method was used using the algorithm

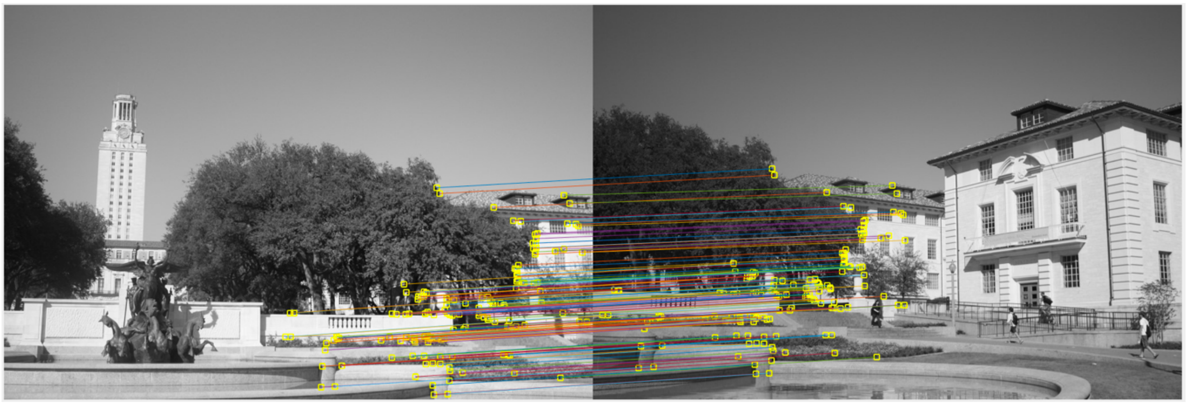
Algorithm:
1. Select random sample of minimum required size to fit model
2. Compute a putative model from sample set
3. Compute the set of inliers to this model from whole data set.
Repeat 1-3 until model with the most inliers over all samples is found.

The best matches were estimated from feature neighbor method above. Next in order to find the best fit model for all the matches randomly pick 4 matched pairs from both the images. Basically h11,h12…h33 are the 9 unknowns that we need to calculate. Here h33 =1 so we are left with 8 unknowns that is why 4 random points are picked.
After picking 4 random points the homography matrix is estimated and now this model is used to test the no of inliers and outliers points. Inliers points are the ones that successfully fits the generated model while outliers re the ones that doesn’t fit the model. The homography matrix with maximum inlier to total matches ratio is considered to be the best model and the transformation of source to destination is done using this model.

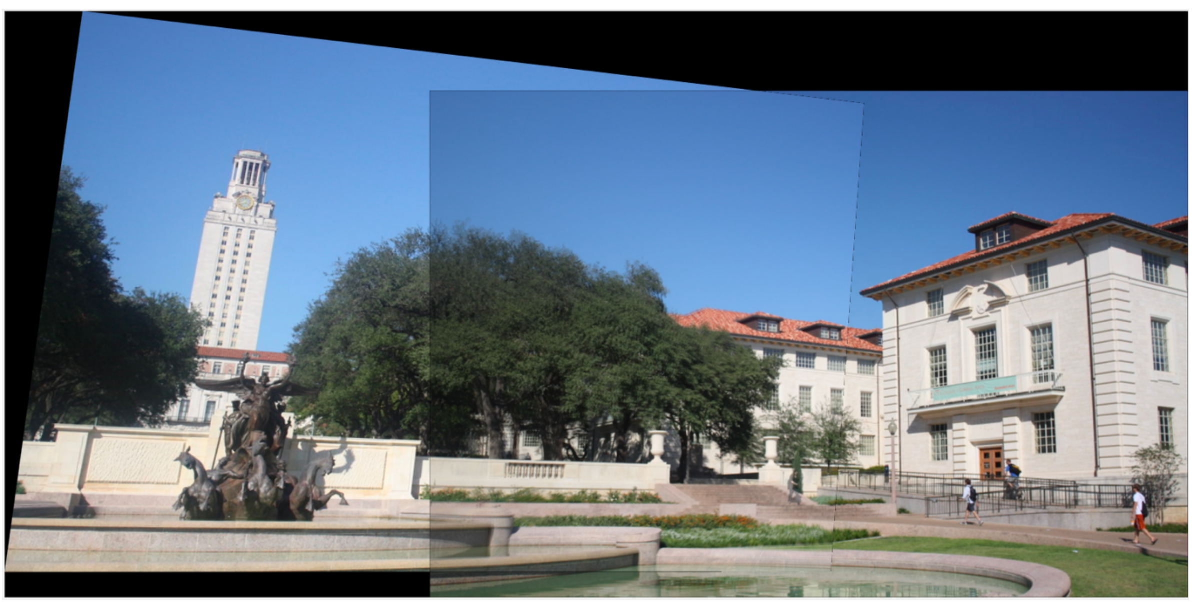
b) For the uttower pair provided

No of inliers are estimated by code was: 144/200
The average residuals for the inliers is : 6.5.

The location of inlier matches is depicted in the image below:



c) Stitching results:



2. **Fundamental Matrix Estimation:** In order to calculate the translation and rotation when image is calculated from different physical points. The fundamental matrix in uncalibrated camera case id determined by epipolar lines estimation. This is done by formula x1'F (x2)

= 0. The fundamental matrix in this case is calculated by 8 point algorithm. Here 8 random points are taken and fundamental matrix is generated from them.

Here we have eight unknowns which can be easily solved by Singular Value Decomposition (SVD), as shown in Exercise 5.5. Applying SVD to A yields the decomposition A = UDV'. The homogeneous least-squares solution corresponds to the least singular vector, which is given by the last column of V.

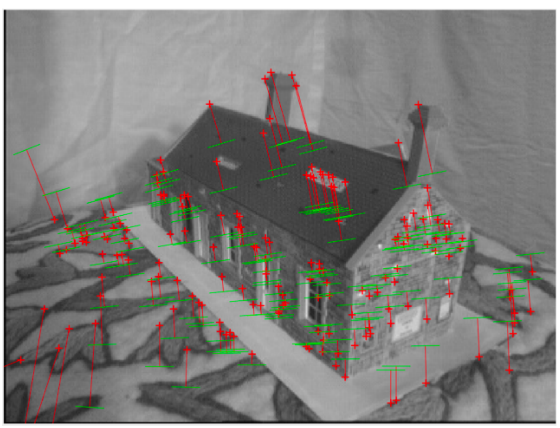
In the presence of noise, the matrix F estimated this way will however not satisfy the rank-2 constraint. This means that there will be no real epipoles through which all epipolar lines pass, but the intersection will be spread out over a small region. In order to enforce the rank-2 constraint, we therefore again apply SVD to F and set the smallest singular value D33 to zero.

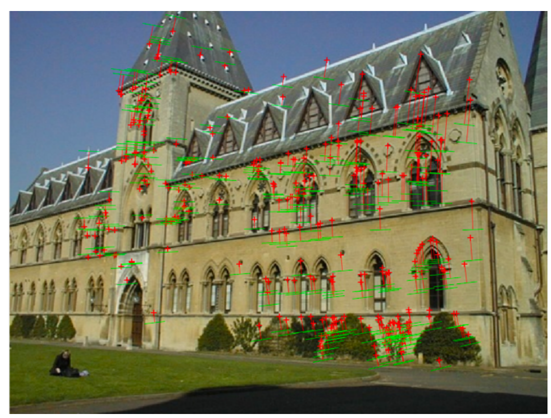
$$\mathbf{Af} = 0$$

$$\begin{bmatrix} u_1u_1' & u_1v_1' & u_1 & v_1u_1' & v_1v_1' & v_1 & u_1' & v_1' & 1 \\ u_2u_2' & u_2v_2' & u_2 & v_2u_2' & v_2v_2' & v_2 & u_2' & v_2' & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_Nu_N' & u_Nv_N' & u_N & v_Nu_N' & v_Nv_N' & v_N & u_N' & v_N' & 1 \end{bmatrix} \begin{bmatrix} F_{11} \\ F_{21} \\ \vdots \\ F_{33} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

The results of epipoles generated are shown below:

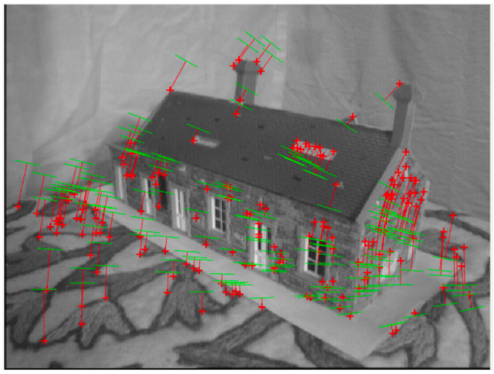
a) **UnNormalized Image using ground truth matches**:



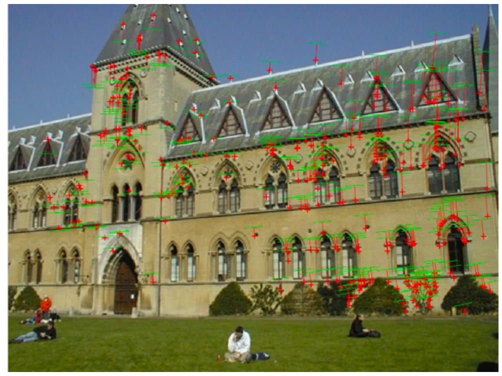


Residual value = 1.1003e+03

Residual value = 255.5795

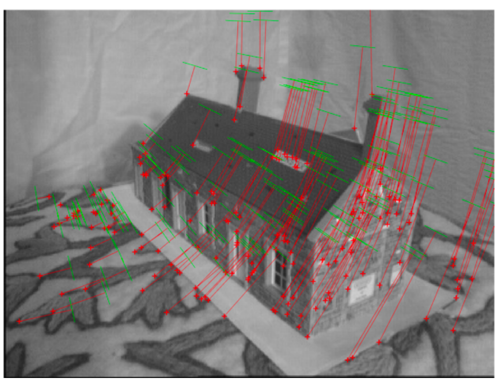


Residual value = 696.8593

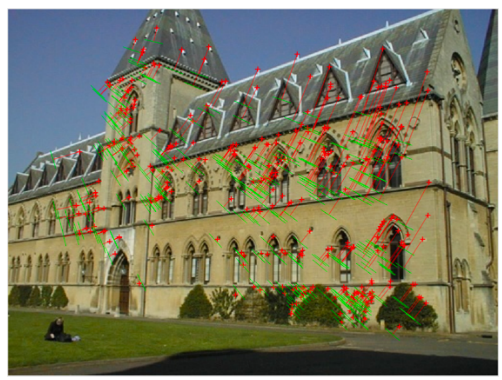


Residual Value = 241.8695

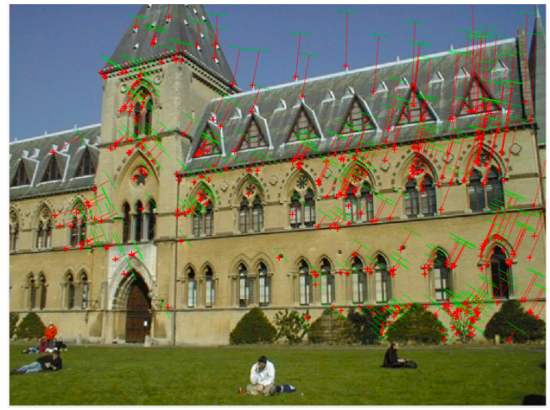
Normalized Image using ground truth matches:



residual value = 3.4863e+03



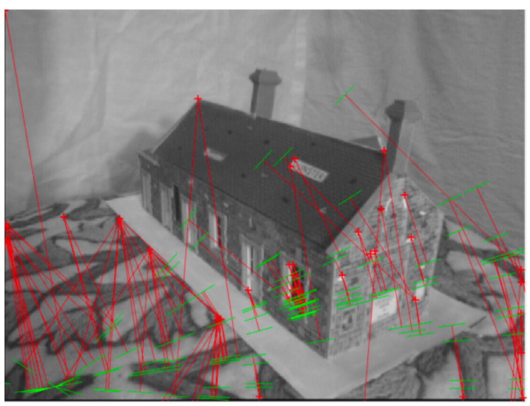
residual value = 755.9031



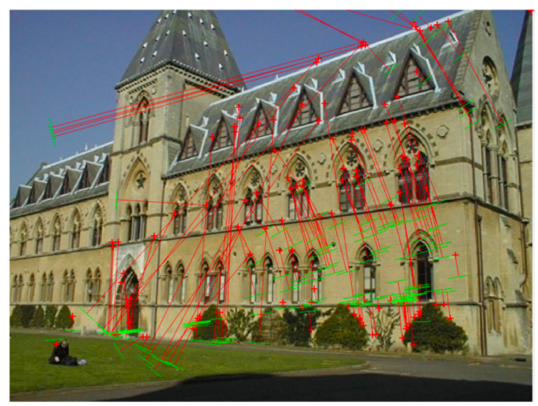
residual value = 3.0116e+03

residual value = 1.0236e+03

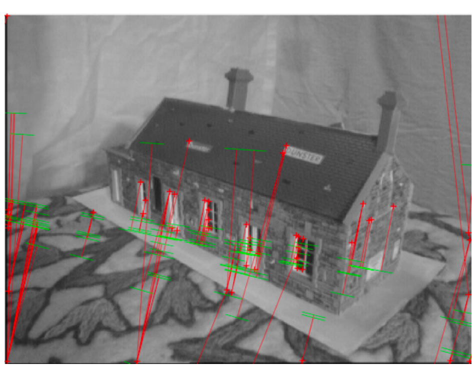
b) Normalized Image without using ground truth matches(RANSAC):



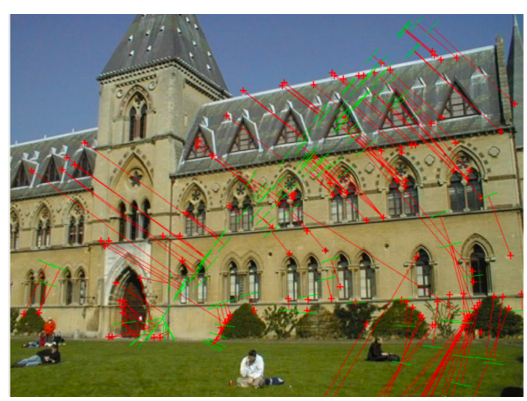
residual value = 4.8038e+03



residual value = 1.0763e+04



residual value = 5.6712e+03



residual value = 6.7212e+03

No of inlier matches generated by RANSAC is 164/200. Which is less than ground truth matches but givea good ratio to generate a model.

The results of ground truth matches are way better than Ransac method because the ground truth matches are all inliers while the matches generated by our Ransac method are not that good because the ratio of match is low. The data that we have is not that accurate because it depends on different threshold values. It has many parameters that should be tuned before desired results are generated. And moreover it can't be used in case the inlier/outlier ratio is small.

c) Visualization of 3D points centers using triangulation.

Camera centers are calculated simply by taking the SVD of the camera matrix and taking the last column of V. Triangulation is done by calculating the least squares of each matching pair of points. Given a 2D point correspondence x1, x2 in homogeneous coordinates, the 3D point location X is given as follows

$$\begin{aligned} \lambda_1 \mathbf{x_1} &= \mathbf{P_1 X} \\ \lambda_2 \mathbf{x_2} &= \mathbf{P_2 X}. \end{aligned}$$

We can now build the cross-product of each point with both sides of the equation and obtain

$$\begin{aligned} \mathbf{x_1} \times \mathbf{P_1 X} &= [\mathbf{x_1}_\times]\mathbf{P_1 X} &= 0 \\ \mathbf{x_2} \times \mathbf{P_2 X} &= [\mathbf{x_2}_\times]\mathbf{P_2 X} &= 0, \end{aligned}$$

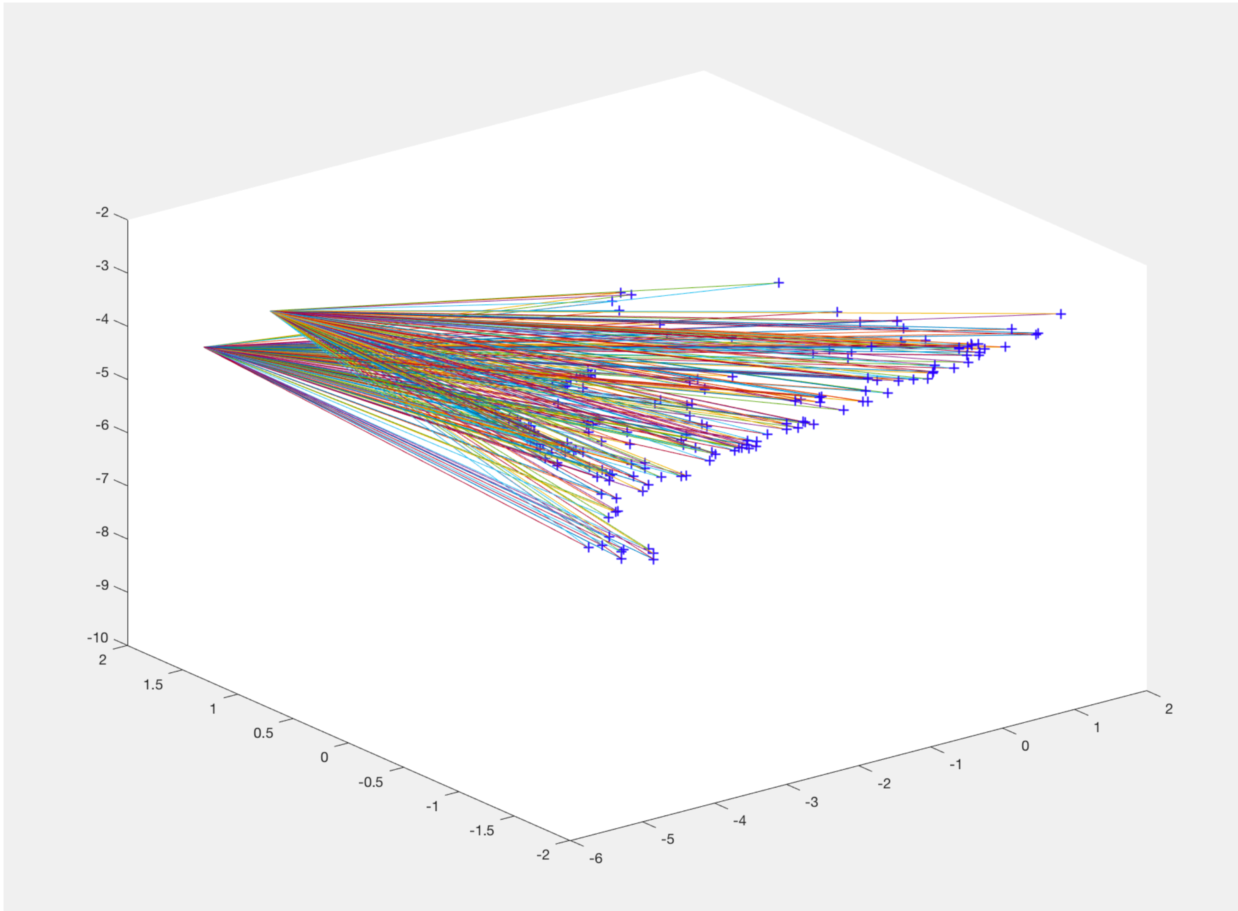
where we used the skew-symmetrix matrices [xi×] to replace the cross products

$$\mathbf{a} \times \mathbf{b} = [\mathbf{a}_\times]\mathbf{b} = \begin{bmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{bmatrix} \mathbf{b}.$$

The centers calculated from the given camera matrices were:

P1_c = [-4.8224 1.4807 -3.6490 1]
P2_c = [-5.0495 1.9316 -4.6627 1]

The results of each point in 3D is generated as:



where blue + marks denote the 3D coordinates.
The 2 points from where the points are generating are the camera centers.
The average residual is calculated by sum of squared distances: The sum of square distance of each point in image is calculated from its 3D point estimated by using camera matrix.

The average residual distance of 2D point of image 1 to 3d point X is
8.3508e+04

The average residual distance of 2D point of image 2 to 3d point X is
7.6965e+04