cs512 Assignment 4: Review Questions

Amit Nikam
Department of Computer Science
Illinois Institute of Technology

November 28,2020.

---

1 Camera calibration 1:

1.a Forward projection is the process of imaging, i.e. capturing an image in 2D from a 3D system, this is the easiest. Calibration is the process of identifying the external and internal parameters of the camera, and this is difficult. The reconstruction is formation of an object in 3D from multiple 2D images, this is the most difficult problem.

1.b The necessary input for camera calibration is mapped corresponding points in 3D world coordinates and 2D image coordinates(pixels).

1.c The steps in the non-coplanar calibration are one to find the projection matrix M and second then finding the parameters from M.

1.d Image coordinates are (18/7 , 14/7) ~ (3,2)

1.e The first two lines of the matrix are [1,2,3,1,0,0,0,0,-100,-200,-300,-100; 0,0,0,0,1,2,3,1,-200,-400,-600,-200]

1.f Minimum number of points that is necessary to be able to find a unique solution for M is 6 as the unknown values are 12. This solution is obtained by taking SVD of the point matrix that is created, and then taking the column of V that belongs to the zero singular value of this matrix.

1.g This principal is of singular value decomposition. A matrix can be represented as a decomposition of three matrices U,D,V as $A=UDV^T$. U and V are orthogonal matrices of mxm and nxn dimension respectively, while D is Diagonal matrix with singular values on diagonal and mxn dimension. Columns of V are eigenvectors fo A^TA and thus we take the column of V that belongs to zero singluar value of A since this will give us zero eigenvector of A^TA.

1.h We can measure the quality by projecting world points onto the image to get the image equivalent point with our solution $\acute{M}$ and then look at the distance between the actual point and the projected point. If the sum of the L2 of this difference is low, then the quality is good.

1.i In planar camera calibration we estimate the parameters using a 2D calibrator. In this, since the calibration points are all in the same plane in real world, we use a 2D homography between the calibration plane and image for each view. Since one view will not be sufficient, we need multiple views during input. This is the major difference between the planar and non-coplanar methods.

1.j The homography H captures the projective map between points on 2D calibration plane in 3D world space and image, assuming that the Z coordinates for the calibrator are 0. Meanwhile $\breve{M}$ finds the projection map between 3D world points and image.

2 Camera calibration 2:

2.a The first two rows are [3,4,5,1,0,0,0,0,-3,-4,-5,-1; 0,0,0,0,3,4,5,1,-6,-8,-10,-2].

2.b U_0 = 26/50 = 0.52; V_0 = 38/50 = 0.76

2.c The projection error is of 1.7331. Taking square root we get pixel error of 1.32 ~ 1-2 pixel error.

2.d The rotation of camera w.r.t world R = [6,0,0,0 ; 0,1,0,0 ; 0,0,1,0 ; 0,0,0,1]
The translation of camera w.r.t world is T = $[-6,-2,-3,-1]^T$

2.e The first two rows are = [3,4,1,0,0,0,-3,-4,-1 ; 0,0,0,3,4,1,-6,-8,-2]

3 Multiple view geometry 1:

3.a Sparse stereo matching is a feature-based correspondence approach. In this w find feature points, localize and then match them to find corresponding points. Advantage is that it is capable of handling large disperities. In dense stereo matching we capture all patches and instead of distance between features vectors we measire correlation or SSD. Disadvantage is that there are many ore points, but plus side is it is good for small disperities.

3.b In cross correlation the two windows are multiplied with each other and the sum of such multiplication is taken. If this sum is a high value then the match / correlation is good. In SSD we pretty much do the same, just instead of sum of multiplications we take the sum of L2 differences. A smaller value indicates good correlation. The risk in taking the entire image as the search space is that there will be many correlating points and computations. To reduce the search to a line, we use the epipolar equation from the concept of epipolar plane.

3.c The depth of the 3D point is 333.3334.

3.d Ambiguity problem occurs when some points in one camera image wrongly correlates to some points in the second camera image. This leads to generation of projections that might seem right but are wrong as their depth will be wrong. These depths are very different than actual depths.

3.e Expression for the rotation and translation of the right camera with respect to the left camera is derived from TR = $R_l \, T_l^{-1} \, T_r \, R_r$ equation. After solving this we get the expressions as:

Rotation (R) = $R_l^T R_r$
Translation (T) = $R_l^T (T_r - T_l)$

4 Multiple view geometry 2:

4.a The depth of point is 6.6667mm. (Less than focal length, inside camera)

4.b The matrix $[A]_X$ is = $\begin{bmatrix} 0 & -3 & 2 \\ 3 & 0 & -1 \\ -2 & 1 & 0 \end{bmatrix}$

4.c Outcome value of the eight point equation = 68

4.d Row in fundamental matrix = [2, 4, 2, 3, 6, 3, 1, 2, 1]