Technische Universität München
Fakultät für Informatik
PD Dr. Slobodan Ilic

WS 19/20
Exercise Sheet 2
November 14, 2019

Exercises in Tracking & Detection

Task 1 Repetition Camera Models

a) **Homogeneous Coordinates** Why do we need them?

Answer To express translation (an affine transformation) as a matrix vector product and to express infinity in the perspective geometry.

b) **Internal calibration matrix** What is its purpose and properties? How to express the pinhole camera model's perspective projection and transformation to pixel coordinates in terms of f, k_u, k_v, u_0, v_0 in the internal calibration matrix. How do changes in focal length f affect an image?

Answer To express the projection from 3D points in the camera coordinate frame to 2D points on the image plane.

$$m = KM_C = \begin{bmatrix} k_u f & 0 & u_0 \\ 0 & k_v f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} k_u f X + Z u_0 \\ k_v f Y + Z v_0 \\ Z \end{bmatrix} = \begin{bmatrix} k_u f \frac{X}{Z} + u_0 \\ k_v f \frac{Y}{Z} + v_0 \end{bmatrix}$$

Changes in focal length affect how zoomed in the image will be. A large focal lenght means that the same object in 3D appears bigger (and maybe clipped) in the image. See `http://ksimek.github.io/2013/08/13/intrinsic/` for a simulator.

c) **External calibration matrix** What is its purpose and properties? What is an Euclidean transformation $[R, t]$, how many parameters does it have? **Answer** To model the change from world to camera coordinate system. One way to represent the full rigid body motion, consisting of translation and rotation and having 3 degrees of freedom, are homogeneous coordinates and 4×4 matrices, a representation for **SE(3)**, see chapter 2 of *Y. Ma, S. Soatto, J. Kosecka, and S.S. Sastry. An Invitation to 3-D Vision: From Images to Geometric Models.*

$$\boxed{\mathrm{SE}(3) = \left\{ \mathrm{T} = \begin{bmatrix} \mathrm{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{4\times 4} \;\middle|\; \mathrm{R} \in \mathrm{SO}(3), \mathbf{t} \in \mathbb{R}^3 \right\}}$$

$$\boxed{\mathrm{SO}(3) = \{ \mathrm{R} \in \mathbb{R}^{3\times 3} \mid \mathrm{R}^T \mathrm{R} = \mathrm{I}, \det(\mathrm{R}) = +1 \}}$$

d) **Camera Distortion** What types are there and how importand are they? Why are they not part of the intrinsic matrix? How to compensate for these effects?

Answer Radial important (=) shape, tangential (can be neglected). They are non-linear effects, thus cannot be modelled using matrix multiplication which allows linear (or affine in case of homogeneous coordinates) transformations only. To compensate, either undistort images using look-up or distort the model projection.

Task 2 Camera Calibration and Pose Estimation algorithms

a) What is the error we try to minimize + Equation?

Answer Reprojection Error

$$\min_{R,T} \sum_i \|\mathbf{A}(\mathbf{R}\mathbf{M}_i + \mathbf{T}) - \mathbf{m}_i\|^2 = \min_{R,T} \sum_k \|\mathbf{A}(\mathbf{R}\mathbf{P}^k + \mathbf{T}) - \mathbf{p}^k\|^2$$
$$= \min_{R,T} \sum_k \|\Delta p_{i-1,i}\|^2 = \min_{R,T} \sum_k \|\hat{p}_i^k - p_i^k\|^2 = \min_{R,T} \sum_k \|p_{i-1}^k - p_i^k\|$$

b) When can we use DLT?

Answer Planar structure, intrinsics must not be known. Gives $\mathrm{P} = [\mathrm{P}_3|\mathbf{c}_4]$, compute absolute conic AA^T, do a Cholesky decomposition to get $\mathrm{R} = \mathrm{A}^{-1}\mathrm{P}_3$, not necessarily $\in \mathrm{SO}(3)$, do ortho-normalization using SVD of $\mathrm{R} = \mathrm{UV}^T$.

c) DLT: Why is the null vector not a valid solution?

Answer Because all points would be projected to 0, which is not a valid projection matrix.

d) When can we use PnP?

Answer arbitrary 3D structure, intrinsics must be known

e) What is the minimum number of correspondences required to estimate the camera pose? Briefly explain your reasoning.

- for DLT **Answer** 12 elements of the projection matrix, but due linear dependence of the rows of this matrix, we this matrix defined up to a scale, therefore there are 11 unknowns. Each correspondence gives 2 equations -> 6 correspondences points must be known. In practice much more needed. Note that we optimize here the elements of projection matrix, not on its parameterization (3 parameters for intrinsics and 6 for the pose()rotation and translation) . When projection matrix is parameterized and represented as $\mathrm{P} = \mathrm{A}[\mathrm{R}|\mathbf{t}]$ then we use non-linear optimization to find solution.
- for P3P/PnP **Answer** 6 correspondences, where each correspondence gives 2 equations -> 3 corresponding points must be known. BUT yield up to 4 solutions -> 4th correspondence needed because of ambiguity.

f) Summary. Fill out the missing entries in the following table to summarize your findings about the algorithms

algorithm	#corresp. points	intrinsics
Umeyama*	4	unknown

(*) only briefly covered in lecture slide 43 as Euclidean displacement from World 3D to Camera 3D points, no minimization of reprojection error but of 3D euclidean distance, e.g. for point cloud alignment and without scale. In that case it is called Kabsch algorithm. Note that either Kabsch nor Umeyama methods for point cloud registration has been presented at the lectures.

Answer

algorithm	correspondences	intrinsics
DLT	6 2D-3D	uncalibrated
PnP, P3P	4 2D-3D	calibrated
Kabsch method, Umeyama	3 3D-3D	none