Technische Universität München
Fakultät für Informatik
Prof. Dr. N. Navab

WiSe 2011/2012
Exercise Sheet 2
November 10, 2011

Exercises in 3D Computer Vision I

Exercise 1 Recap: Camera Models

The projection matrix $\mathbf{P} \in \mathbb{R}^{3\times 4}$ of a camera relates points $\mathbf{X}$ in the 3D world (in homogeneous coordinates) to their images $\mathbf{x}$ on the camera plane as $\mathbf{x} = \mathbf{PX}$. Intrinsic parameters (such as the focal length) and extrinsic parameters (such as the location of the camera in the world) are encapsulated in the projection matrix. The projection matrix an be decomposed as $\mathbf{P} = \mathbf{KR}[\mathbf{I}|-\tilde{\mathbf{C}}]$, where $\tilde{\mathbf{C}}$ is the camera center in the world coordinate system, $\mathbf{R}$ is the rotation matrix representing the orientation of the camera and $\mathbf{K}$ is the matrix of intrinsic camera parameters.

a) Give the entries of the intrinsic parameters matrix $\mathbf{K}$ for each of the following camera models: (a) basic *pinhole camera*, (b) *CCD camera*, (c) *finite projective camera*. How many degrees of freedom does $\mathbf{P}$ have in each of the cases?

- Pinhole Camera:

$$\mathbf{K}_{\text{pinhole}} = \begin{bmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{bmatrix}, \tag{1}$$

  where f is the focal length of the camera and p_x, p_y is the principal point in image coordinates. $\mathbf{P}$ has 9 degrees of freedom.

- CCD Camera:

$$\mathbf{K}_{\text{ccd}} = \begin{bmatrix} \alpha_x & & x_0 \\ & \alpha_y & y_0 \\ & & 1 \end{bmatrix}, \tag{2}$$

  with $\alpha_x = fm_x$, $\alpha_y = fm_y$, $x_0 = m_x p_x$, $y_0 = m_y p_y$. m_x and m_y are the number of pixels per unit in image coordinates. $\mathbf{P}$ has 10 degrees of freedom.

- Finite Projective Camera:

$$\mathbf{K}_{\text{finite}} = \begin{bmatrix} \alpha_x & s & x_0 \\ & \alpha_y & y_0 \\ & & 1 \end{bmatrix}, \tag{3}$$

  where s is the skew parameter. $\mathbf{P}$ has 11 degrees of freedom.

b) Explain the geometric meaning of the columns and rows of $\mathbf{P}$. *Hint:* Use the following notation: columns are written $\mathbf{p}_i$, such that $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4]$. Rows are written $\mathbf{r}_i^\top$, such that $\mathbf{P} = [\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3]^\top$. Then, for the columns, consider the products of the form $\mathbf{p}_1 = [\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4](1,0,0,0)^\top$. For the rows, consider the products $\mathbf{r}_i^\top \mathbf{X}$.

- Columns are particular image points: $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ are the vanishing points for the 3 coordinate axes of the world coordinate system. These columns are the images of the directions of the world corrdinate axes, e.g. $(1,0,0,0)$ for the X axis. $\mathbf{p}_4$ is the image of the world coordinate origin $(0,0,0,1)$.

- Rows are particular world planes: For instance, all points on the plane $\mathbf{r}_1^\top$ satisfy $\mathbf{r}_1^\top \mathbf{X} = 0$ and are therefore imaged at $\mathbf{PX} = (0, y, w)^\top$ (i.e. on the image y axis). Similarly, the points on the plane $\mathbf{r}_2^\top$ are projected on the image x axis. In other words, the world plane $\mathbf{r}_1^\top$ is defined by the camera center and the line $x = 0$ in the image, $\mathbf{r}_2^\top$ is defined by the camera center and the line $y = 0$. The plane $\mathbf{r}_3^\top$ is the principal plane, i.e. the plane through the camera center and parallel to the image plane. All points on this plane are imaged at $\mathbf{PX} = (x, y, 0)^\top$, i.e. on the line at infinity.

c) Outline an algorithm to determine the projection matrix $\mathbf{P}$ from point correspondences $\{\mathbf{x}_i \leftrightarrow \mathbf{X}_i\}$ between points $\mathbf{X}_i$ in 3D space and points $\mathbf{p}_i$ on the 2D image. How many points are required?

The objective is to find a 3×4 camera matrix $\mathbf{P}$ such that $\mathbf{x}_i = \mathbf{PX}_i$ for all i. This task is very similar to computing a homography in 2D, except that now we are computing a 3×4 matrix instead of a 3×3 matrix – we can thus use the DLT algorithm.

- Write $\mathbf{p}_i \times \mathbf{PX}_i = \mathbf{0}$. The homogeneous vectors $\mathbf{x}_i$ and $\mathbf{PX}_i$ will not be identical, but will have the same direction. Therefore their cross product must be zero.
- Set up a system of linear equations $\mathbf{A}_i\mathbf{p} = \mathbf{0}$ for every point correspondence $\{\mathbf{x}_i \leftrightarrow \mathbf{X}_i\}$. Here, $\mathbf{p}$ is a 12×1 vector containing the sought entries of $\mathbf{P}$ and $\mathbf{A}_i$ is a 2×12 matrix.
- Create a system $\mathbf{Ap} = \mathbf{0}$ by stacking up the matrices $\mathbf{A}_i$ for all point correspondences. Minimize $\|\mathbf{Ap}\|$ under the constraint that $\|\mathbf{p}\| = 1$ using the SVD decomposition.

Since $\mathbf{P}$ is a homogeneous 3×4 matrix, it has 12 entries and 11 degrees of freedom (the representation is up to scale). Each point correspondence between the world and the image provides 2 constraints (x and y coordinates in the image). This leads to the result that we need at least 5.5 point correspondences in order to compute the camera matrix. So $\mathbf{A}$ is a $2n \times 12$ matrix, with n being the number of used point correspondences.

Exercise 2 Projection Matrices

a) **Decomposition of the Camera Matrix**

Now that we have $\mathbf{P}$ we can extract the intrinsic and extrinsic parameters from it. We do so by decomposing $\mathbf{P}$ into $\mathbf{P} = \mathbf{KR}\left[\mathbf{I}|\mathbf{R}^\top\mathbf{t}\right]$. Here $\mathbf{K}$ is a 3×3 upper triangular matrix containing the intrinsic parameters. $\mathbf{R}$ and $\mathbf{t}$ represent the extrinsic parameters: the rotation (3×3 matrix $\mathbf{R}$) and translation (3×1 vector $\mathbf{t}$). $\mathbf{R}$ and $\mathbf{t}$ take all points from the world coordinate system into the camera coordinate system of the camera.

We can get $\mathbf{K}$ and $\mathbf{R}$ by applying the QR-decomposition to the left 3×3 submatrix of $\mathbf{P}$. The QR-decomposition gives for a matrix $\mathbf{M}$ the decomposition $\mathbf{M} = \mathbf{QR}$ such that $\mathbf{Q}$ is an orthogonal matrix and $\mathbf{R}$ is an upper triangular matrix. Let us assume we have $\mathbf{P} = [\mathbf{KR}|\mathbf{Kt}] = [\mathbf{M}|\mathbf{p}_4]$. By feeding $\mathbf{M}^{-1}$ to the QR-decomposition, we get $\mathbf{M}^{-1} = \mathbf{QR}$. It follows that $\mathbf{M} = (\mathbf{QR})^{-1} = \mathbf{R}^{-1}\mathbf{Q}^{-1}$. Since $\mathbf{Q}$ is orthogonal $\mathbf{Q}^{-1} = \mathbf{Q}^\top$ and thus $\mathbf{Q}^\top$ is orthogonal and can be interpreted as a rotation matrix. Since the inverse of an upper triangular remains upper triangular we can interpret $\mathbf{R}^{-1}$ as the matrix containing the intrinsic parameters.

After doing the QR-decomposition, it is crucial to assure that the signs of the focal length from the computed matrix $\mathbf{K}$ are positive since the focal length can not be negative in the pinhole model.

Finally, the translation $\mathbf{t}$ is given by: $\mathbf{t} = \mathbf{K}^{-1}\mathbf{p}_4$.

b) **Recovery of Camera Center from P**

For the camera center, the equation $\mathbf{Pc} = 0$ holds since $\mathbf{c}$ is the only point that can't be imaged by the camera. So solving $\mathbf{Pc} = \mathbf{0}$ yields $\mathbf{c}$. [1]

If $\mathbf{P}$ is already decomposed as $\mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}]$ then we can use the relationship $\mathbf{c} = -\mathbf{R}^\top \mathbf{t}$.

[1] By the way: $\mathbf{c}$ can be described as a point in which the three planes described by the rows of $\mathbf{P}$ intersect.