

## Exercises in 3D Computer Vision

### Exercise 20      Mosaicing

Merging two or more images of a scene where the images have just slight rotational differences is called mosaicing. The mosaicing creates a panorama image of several independent images by computing a 2D homography on point correspondences and applying it to the whole image(s). Figure 1 shows a mosaicing applied to two images of a building.



Figure 1: Mosaicing on two images of the TUM building in Garching in order to obtain a panorama image

In this exercise, we will perform mosaicing on two images, i.e. one image is transformed such that it fits the other image. An extension to more than two images is made easily by declaring one image as reference image and computing the homography of all other images to the reference image. An algorithm for mosaicing is given as follows:

- (i) Determine point correspondences between the two images  $\{\mathbf{x}'_i \leftrightarrow \mathbf{x}_i\}$ .
- (ii) Compute a 2D homography from the point correspondences.
- (iii) Warp one image in order to create a panorama view of the two images.
  - a) Think of ways to determine point correspondences in the two images.
  - b) One method to solve for the homography is the DLT algorithm. Sometimes it happens that the coordinate system of two images is not equal or the scaling is different. Is the DLT algorithm invariant to such changes? Is there a possibility to improve the outcome of the DLT algorithm in this respect?
  - c) What has to be considered for image warping and how is it performed?

### Exercise 21      Projection Matrices

Determining the projection matrix of a camera is crucial for many 3D computer vision applications. Informally, it relates the 3D world to the imaged 2D world, which allows a user to switch between 2D and 3D, e.g. recover 3D objects of features only visible in the

2D images. The calibration of a camera determines the camera's geometry, i.e. where it is located in 3D space and what focal length (or other internal parameters) it uses to acquire images. This information is all encapsulated in a **projection** matrix  $\mathbf{P} \in \mathbb{R}^{3 \times 4}$  that takes each 3D point  $\mathcal{X}^*$  to the corresponding 2D point  $\mathbf{p}$  on the image (plane).

- a) Explain the computer vision camera model by describing the image formation pipeline from 3D world coordinates to 2D pixel coordinates.
- b) The matrix  $\mathbf{P}$  has 4 columns and 3 rows. Explain their geometric meaning.
- c) How can  $\mathbf{P}$  be determined? How many d.o.f. has the matrix  $\mathbf{P}$ ? How many points are needed?
- d) **(H)** Here is an example for estimating  $\mathbf{P}$ : We define a cube in 3D space with one corner in the origin of the world coordinate system and an edge length of 4 [cm]. Thus, the 3D points can be defined as

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 4 \end{pmatrix}, \begin{pmatrix} 0 \\ 4 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 4 \\ 4 \end{pmatrix}, \begin{pmatrix} -4 \\ 0 \\ 4 \end{pmatrix}, \begin{pmatrix} -4 \\ 4 \\ 4 \end{pmatrix}, \begin{pmatrix} -4 \\ 4 \\ 0 \end{pmatrix}, \begin{pmatrix} -4 \\ 0 \\ 0 \end{pmatrix}.$$

The corresponding corner points on the image plane are

$$\begin{pmatrix} -8.4853 \\ 5.1962 \end{pmatrix}, \begin{pmatrix} -3.9470 \\ 2.5648 \end{pmatrix}, \begin{pmatrix} -5.6569 \\ 5.1962 \end{pmatrix}, \begin{pmatrix} -3.0127 \\ 3.1877 \end{pmatrix}, \\ \begin{pmatrix} -3.8140 \\ 1.3821 \end{pmatrix}, \begin{pmatrix} -3.1142 \\ 2.0820 \end{pmatrix}, \begin{pmatrix} -5.0212 \\ 3.1877 \end{pmatrix}, \begin{pmatrix} -6.5783 \\ 2.5648 \end{pmatrix}.$$

Determine the camera projection matrix  $\mathbf{P}$  from the point correspondences.

- e) Now that  $\mathbf{P}$  is computed, one would like to decompose it such that all parameters can be recovered. The extrinsic camera parameters determine the camera's pose in 3D space, and consist of a translation and a rotation that takes the world coordinate system to the camera coordinate system. The intrinsic camera parameters represent the camera's focal length, the aspect ratio (ratio of pixel size in x- and y-direction), and the principal point. Determine extrinsic and intrinsic parameters from  $\mathbf{P}$ .
- f) Determine the camera center from  $\mathbf{P}$ .