1 Numerical Exercise

1. Explain the Dolly-Zoom Effect.

   Solution
   The Dolly-Zoom effect (see `https://youtu.be/u5JBlwlnJX0?t=44` for a video) is the result of a change in focal length combined with a camera translation. In plain English, the camera is zoomed out while being moved forward, which results in the background becoming more compressed (seeming more distance) compared to the foreground.
   To understand the effect better, consider the two sketches shown in Figure 1. In the first (zoomed) position, all points p_1 to p_4 are visible in the image plane, with the background points being roughly three times further apart than the foreground points p_1 and p_2. When the camera is moved closer to Position 2 and zoomed out, the foreground points are still visible at the same points in the image plane, the background point p_4 has now moved much closer to p_2, while p_3 disappeared completely. Hence, the foreground remained the same while the background got compressed.

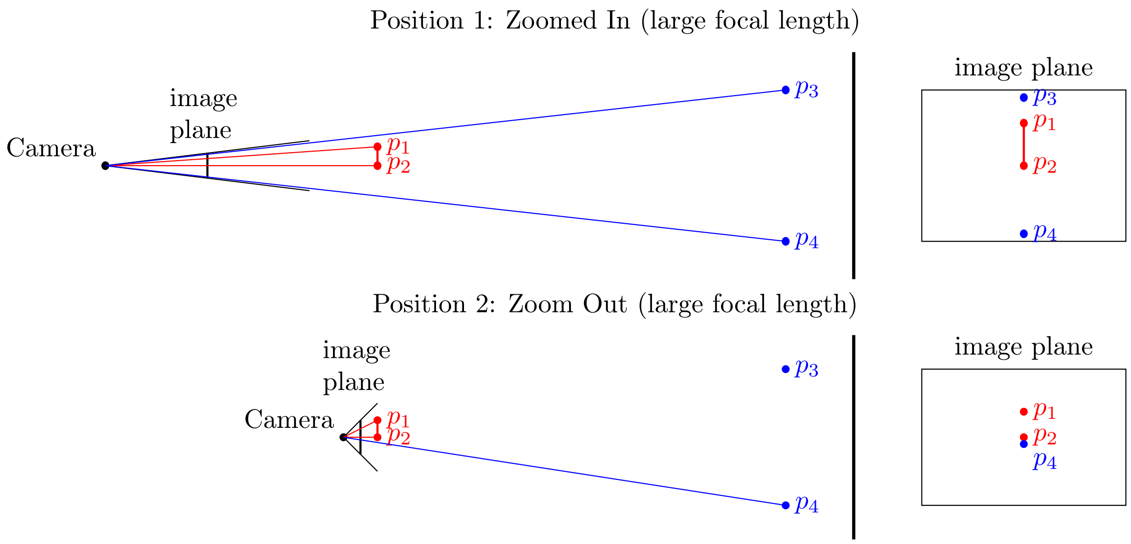

Figure 1: Illustration of the Dolly Zoom

2. What is the condition on the elements of a projective transformation H such that parallel lines remain parallel?

   Solution
   Parallel lines remaining parallel implies that, for all the directions, points at infinity (which are the intersections of the parallel lines) are again mapped to points at infinity. In homogeneous coordinetes, the points at infinity are defined as $[a \;\; b \;\; 0]$[1]. Hence, $[a \;\; b \;\; 0]$ is mapped to $[a' \;\; b' \;\; 0]$ under the projective transformation H. Consequently $H \; [a \;\; b \;\; 0]^\top$ should have the third entry zero.

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}, \qquad h_{31}a + h_{32}b = 0 \;\forall a, b \Longrightarrow h_{31} = h_{32} = 0$$

3. Write the projective transformation H that corresponds to doubling the focal length.

   Solution
   The components h_{11} and h_{22} correspond to the focal length. Therefore,

$$H = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

[1] `https://en.wikipedia.org/wiki/Homogeneous_coordinates`

4. Consider a single camera C with the following intrinsic parameters:
   - f the focal length $[m]$
   - k_x and k_y the resolution in $[\frac{pixels}{m}]$ of the camera in x and y direction
   - c_x and c_y in $[pixels]$ as the x and y coordinates of the camera central point

   and the following extrinsic parameters:
   - R the rotation matrix of the camera to the world coordinate system, and
   - t the translation vector between the optical center of the camera and the center of the world coordinate system.

   In homogeneous coordinates the projection $\hat{p}$ of a point p onto the image plane is then computed as

   $$\hat{p} = MP = K[R|t]p$$

   with the camera calibration matrix K

   $$K = \begin{pmatrix} fk_x & 0 & c_x \\ 0 & fk_y & c_y \\ 0 & 0 & 1 \end{pmatrix}$$

   - Which parameters can be calibrated given a planar calibration pattern?
   - How can they be estimated?
   - Which parameters cannot be calibrated? How can this problem be tackled?

   Solution
   The DLT-method discussed in the lectures can be used to estimate the complete calibration matrix K from a planar pattern. However, the products fk_x and fk_y can only be estimated jointly as the focal length in pixels. To actually obtain f, k_x and k_y seperately one could, for example, get the pixel distance (pixel spacing) from the manufacturer and use this information to convert the focal length in pixels into a metric focal length.

5. Discuss the use of regular squares and regular circles as calibration patterns.
   - Which points would you use as calibration points?
   - How would you detect them?
   - Which approach is more accurate?

   Solution
   As we have not discussed feature detection in class yet, we will limit ourselves to basic ways on how to detect possible calibration points. If the size of the circles is known, a simple yet effective strategy could be to detect a circle by maximizing the correlation of the image with a circular kernel of the correct size. The correlation would be maximal, if the kernel is perfectly aligned with a circle in the image. By "shifting" a circlular pattern over the image and measuring the correlation at each position, we can detect the circles. For the square pattern, the "Harris Corner" detector (explained in Lecture 04) would be suitable. For an accurate calibration, accurate point detection is the key ingredient. For the circles, their center is the only thing that can be reliably detected. However, in constrast to the checkerboard pattern, the center is not as clearly identifyable as the corners of the black and white squares. Thus the positional accuracy of the circle-center detection will be much lower than the accuracy of the squares-corner detection. Consequently, one would prefer to use squares.