Computer Vision and Imaging [06-30213]
Formative Assignment 1

March 1, 2021

Contribution to overall module mark:	0%
Submission Method:	This assignment must be submitted through Canvas.

Name:	Zhuang Ma
Username:	ZXM097
Study program:	MSc Advance Computer Science

Part 1

Question 1.1 *Describe homogeneous coordinates and explain why they are useful for transformations between the world and camera coordinate systems, and for projections onto the camera image plane.*

1. Homogeneous coordinates is to be an original n-dimensional vector a with n + 1-dimensional vector expressed in projective geometry refers to a coordinate system used (The origin coordinate is (0,0,1)),like Euclidean geometry in a Cartesian general. *The following is the conversion of homogeneous coordinates to Euclidean coordinates:*

$$Euclidean \rightarrow Homogeneous$$

$$2D: \ (x, y) \Rightarrow \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \ 3D: \ (x, y, z) \Rightarrow \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{1}$$

$$Homogeneous \rightarrow Euclidean$$

$$2D: \ \begin{bmatrix} x \\ y \\ w \end{bmatrix} \Rightarrow (\frac{x}{w}, \frac{y}{w}), \ 3D: \ \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \Rightarrow (\frac{x}{w}, \frac{y}{w}, \frac{z}{w}) \tag{2}$$

2. Firstly, as we all know, two parallel lines cannot intersect. However, in perspective space, two straight lines can intersect. For example, the train track becomes narrower and narrower as people's sight, and the last two parallel lines intersect at infinity. If a point is at infinity, the coordinates of this point will be (∞, ∞), which becomes meaningless in Euclidean space. After introducing homogeneous coordinates, they can be

expressed mathematically. Proof: Two straight lines can intersect:
First define two parallel lines

$$\begin{cases} Ax + By + C = 0 \\ Ax + By + D = 0 \end{cases} \tag{3}$$

We know that in the Cartesian coordinate system, this system of equations has no solution, because $C \neq D$, if $C = D$, the two straight lines are the same.
In perspective space, use homogeneous coordinates $\frac{x}{w}$, $\frac{y}{w}$ instead of x, y.

$$\begin{cases} A\dfrac{x}{w} + B\dfrac{y}{w} + C = 0 \\ A\dfrac{x}{w} + B\dfrac{y}{w} + D = 0 \end{cases} \Rightarrow \begin{cases} Ax + By + Cw = 0 \\ Ax + By + Dw = 0 \end{cases} \tag{4}$$

Use the above formula to find a solution (x, y, 0), two straight lines intersect at (x, y, 0), this point is at infinity.

3. In addition, after introducing homogeneous coordinates, it is easier to perform linear geometric transformations.

   For the two most common radial transformations, translation T and rotation scaling S, translation transformation is only meaningful for points, because ordinary vectors have no concept of position, only size and direction. Rotation and scaling make sense for both vectors and points. It can be seen that homogeneous coordinates are very convenient for radial transformation.

   First, a vector represents a point in space:

$$r = [r_x, r_y, r_z] \tag{5}$$

   If we want to translate it, the translation vector is:

$$t = [t_x, t_y, t_z] \tag{6}$$

   Then the normal way is:

$$r + t = [r_x + t_x, r_y + t_y, r_z + t_z] \tag{7}$$

   If the homogeneous coordinate system is not introduced, simply use 3x3 matrix multiplication to realize the translation and find a matrix m such that:

$$m \cdot r = r + t = [r_x + t_x, r_y + t_y, r_z + t_z] \tag{8}$$

   Through the above expression (8), we find that we cannot find an m matrix that meets this expression.

   But after introducing the homogeneous coordinate system, the point r becomes:

$$r = [r_x, r_y, r_z, 1] \tag{9}$$

   Then we can find the matrix m:

$$m = \begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{10}$$

   To:

$$m \cdot r = \begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} [r_x, r_y, r_z, 1] = [r_x + t_x, r_y + t_y, r_z + t_z, 1] \tag{11}$$

In the camera model, the coordinate transformation is usually not a single one. A geometry may be designed with multiple translation, rotation, zoom and other changes in each frame. For these changes, we usually use the method of concatenating each sub-change matrix to get a final change matrix, Thereby reducing the amount of calculation. So we need to express the translation as a change matrix. Therefore, only a homogeneous coordinate system can be introduced.

Question 1.2 *Calibration of camera intrinsic and extrinsic properties often involves the use of calibration patterns which contain landmarks at known locations. List desirable properties of such calibration patterns and explain why these properties are important.*

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K[r_1, r_2, r_3, t] \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = K[r_1, r_2, t] \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \tag{12}$$

This method assumes that the calibration template plane is on the plane of the world coordinate system $Z = 0$.

Among them, K is the parameter matrix in the camera, $\bar{M} = [X, Y, 1]^{\mathrm{T}}$ is the homogeneous coordinate of the point on the template plane, $\bar{m} = [u, v, 1]^{\mathrm{T}}$ is the homogeneous coordinate corresponding to the point on the template plane being projected onto the image plane, $[r_1, r_2, r_2]$ and t They are the rotation matrix and translation of the camera coordinate system relative to the world coordinate system.

$$H = [h_1, h_2, h_3] = \lambda K[r_1, r_2, t] \tag{13}$$

$$r_1 = \frac{1}{\lambda} K^{-1} h_1,\ r_2 = \frac{1}{\lambda} K^{-1} h_2 \tag{14}$$

According to the nature of the rotation matrix, that is, $r_1^{\mathrm{T}} r_2 = 0$ and $||r_1|| = ||r_2|| = 1$, each image can obtain the following two constraints on the parameter matrix:

$$h_1^{\mathrm{T}} K^{-\mathrm{T}} K^{-1} h_2 = 0 \tag{15}$$

$$h_1^{\mathrm{T}} K^{-\mathrm{T}} K^{-1} h_1 = h_2^{\mathrm{T}} K^{-\mathrm{T}} K^{-1} h_2 \tag{16}$$

Since the camera has 5 unknown parameters, when the number of images obtained by the camera is greater than 3, the linear unique solution K can be obtained.

Part 2

For this task, you will have to submit the following file:

- code **username_formativeassignment1_part2.m**
- output image as username_formativeassignment1_1.jpg

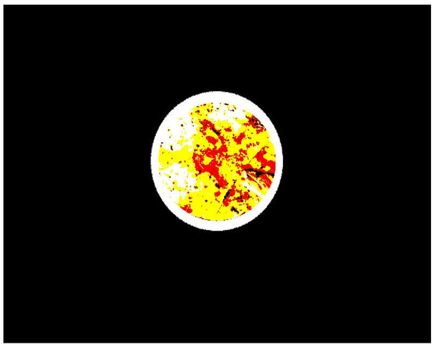

Figure 1: ZhuangMa formative assignment 1.1

- histogram of the difference image as username_formativeassignment1_2.jpg

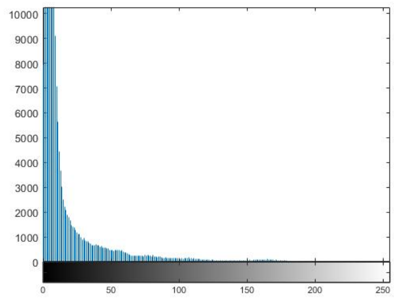

Figure 2: ZhuangMa formative assignment 1.2

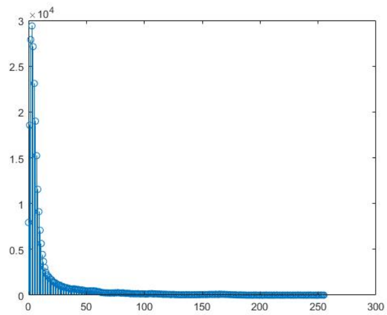

Figure 3: ZhuangMa formative assignment 1.2

- output image as username_formativeassignment1_3.jpg

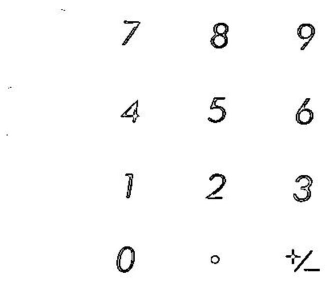

Figure 4: ZhuangMa formative assignment 1.3

Part 3

For this task, you will have to submit the following file:

- code **username_formativeassignment1_part3.m**

Question 3.1 *What is the use of the erosion and dilatation? When do we use them and what can we expect as an output?*

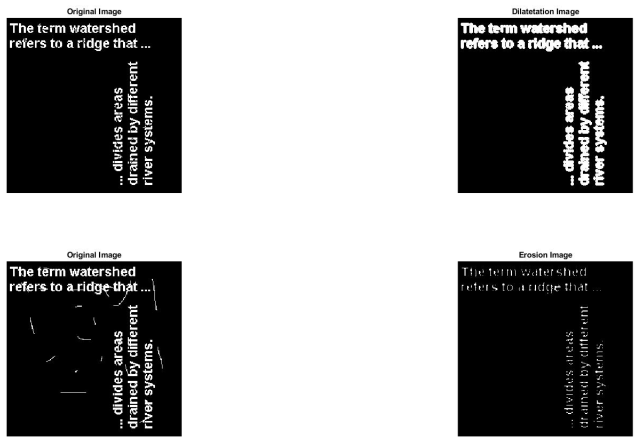

Figure 5: ZhuangMa formative assignment 3.2

1. Erosion and dilatation have the following uses:

   - Eliminate noise
   - Isolate independent image elements, join adjacent elements in the image.
   - Find obvious maximum or minimum areas in the image
   - Find the gradient of the image

2. In summary, use "Erosion" when dealing with "defective" images, and use "Erosion" when dealing with "bulging" images.

- When it is necessary to deal with noise, for example, some letters on a string of characters are not written in a standard manner. There may be strokes that do not enter the letter. In this case, the image can be denoised by erosion. In the example just now. The final result can filter out strokes that do not belong to a certain letter.
- When the image is not continuous. For example, for a string of characters, the strokes that should belong to a certain letter are not written due to writing problems, resulting in incomplete letters. At this time, the image can be processed using the dilatation operation. The final result of the above example can make the letters complete.
- When you need to extract image edge information. For example, to obtain the edge information of a string of hand strings, the image needs to be defective first, and the original image is used for the erosion operation. Finally, the image after the expansion operation is used to subtract the image after the erosion operation. The required string edge information.