Guide to answers for written examination in TSBB09 Image Sensors, 2022-01-10

Maria Magnusson, maria.magnusson@liu.se

PART I: STANDARD CAMERAS & IR SENSORS

Exercise 1
The distorsion is called chromatic aberration. A ray of white light (from the object) is decomposed into rays of different colors that intersect the image plane at different points. This can give color artefacts in the image, especially close to edges.

Exercise 2
The aim of micro-lenses is to enhance the fill factor. The fill factor is the percentage of the total area which is sensitive to light.

Exercise 3
The rolling shutter may cause the Jello effect.
When the rolling shutter is used, the image is sampled line-by-line. If the object is moving, it will be at different positions when different lines are sampled. This causes a distorsion in the final image which is called the Jello effect.

Exercise 4
Glass absorbs a large range of the IR-spectrum, while germanium does not.

Exercise 5

- An RGB color image contains three bands, the color components red, green and blue.
- A multispectral image contains several bands, often corresponding to both visual and IR light.
- A hyperspectral image contains many contiguous bands.

Exercise 6
Ideally, gain (g) and offset (o) are constant over the sensor array. In reality, both g and o vary over the sensor array. In the extreme case, this effect can imply that an isolated pixel has a value that is either very low or very high and almost independent of the incident light.

Exercise 7
The Bayer image is multiplied with 3 different masks, red (*Rmask*), green (*Gmask*) and blue (*Bmask*). *Rmaskimage*, *Gmaskimage* and *Bmaskimage* are recieved. Normalized averaging applied to *Gmaskimage* gives

$$Gimage = \frac{Gmaskimage * w}{Gmask * w}.$$

The corresponding images are shown below. (Blank pixels can be thought of as equal to 0.) A similar procedure is then performed to get *Rimage* and *Bimage*. Then *Rimage*, *Gimage* and *Bimage* can then be put together so that they constitute the three RGB-planes in a color image.

0		0		0
	0		0	
0		2		0
	2		2	
0		0		0

Gmaskimage

0	0	0	0	0
0	1	2	1	0
1	4	6	4	1
2	5	6	5	2
1	2	2	2	1

/ **8**

*Gmaskimage*w*

1		1		1
	1		1	
1		1		1
	1		1	
1		1		1

Gmask

0.5	0.5	0.5	0.5	0.5
0.5	0.5	0.5	0.5	0.5
0.5	0.5	0.5	0.5	0.5
0.5	0.5	0.5	0.5	0.5
0.5	0.5	0.5	0.5	0.5

*Gmask*w*

0	0	0	0	0
0	1	2	1	0
1	4	6	4	1
2	5	6	5	2
1	2	2	2	1

/ **4**

Gimage

Exercise 8

- SNR $= S/N = 5^2/0.05 = 500$.
- For shot noise, the variance is proportional to the signal strength.
  $\mathrm{var}(i1) = k_1 \cdot \mathrm{mean}(i1) \Rightarrow k_1 = 0.05/5 = 0.01$.
  $\mathrm{var}(i2) = k_2 \cdot \mathrm{mean}(i2) \Rightarrow k_2 = 0.05/15 = 0.0033$.
  $\mathrm{var}(i3) = k_3 \cdot \mathrm{mean}(i3) \Rightarrow k_3 = 0.25/25 = 0.01$.
  Consequently, the measurement for pixel 2 is wrong because it gives a discrepant proportionality constant.

PART II: GEOMETRY AND MULTIPLE VIEWS

Exercise 9
Let k be a constant, $k \neq 0$. Then

$$\mathbf{H} = \begin{pmatrix} k \cdot H_{11} & k \cdot H_{12} & k \cdot H_{13} \\ k \cdot H_{21} & k \cdot H_{22} & k \cdot H_{23} \\ k \cdot H_{31} & k \cdot H_{32} & k \end{pmatrix}$$

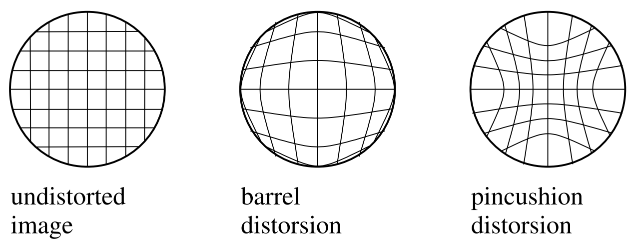
describes exactly the same homography.

Exercise 10

$$\mathbf{A} = \begin{pmatrix} 2000 & 0 & 499.5 \\ 0 & 1980 & 299.5 \\ 0 & 0 & 1 \end{pmatrix}.$$

Exercise 11
The height of the truck is 4 m. Let the distance to the truck be W. The normalized image coordinates are then $(\text{length}/W, 4/W)$, where length is unknown.
$4/W \cdot 1000 = 200 \Rightarrow W = 20$.
$\text{length}/W \cdot 900 = 500$.
$\text{length}/20 \cdot 900 = 500 \Rightarrow \text{length} = 10/9 \text{ m} \approx 11.11 \text{ m}$.

Exercise 12
The figure illustrates an undistorted image of a grid as well two distorted images of the grid, one with barrel distorsion and one with pincushion distortion.



Exercise 13
It is possible to find a unique decomposition of $\mathbf{C}$ into its internal $\mathbf{A}$ and external $[\mathbf{R}\ \mathbf{t}]$ parameters. This decomposition is referred to as *camera resectioning* and it works because $\mathbf{A}$ is an upper triangular 3×3 matrix and $\mathbf{R}$ is a rotational matrix and consequently also an orthogonal matrix.

Exercise 14
The images must share a common camera centre (origin), i.e. the images must be taken from the same view-point but in different directions. Given that the objects in the images are far away, the camera centres do not have to be exactly at the same point. Each image can be transformed into another by a (rotational) homography.
Regard a case with two images which do NOT share a common camera centre. In the overlapping area some objects might be seen in one image but not in the other image, because objects may occlude each other.

Exercise 15
We have measured the point (X_1, Y_1, Z_1) in the world. It corresponds to the point (u_1, v_1) in the image. Therefore

$$s \begin{pmatrix} u_1 \\ v_1 \\ 1 \end{pmatrix} = \begin{pmatrix} C_{11} & C_{12} & C_{13} & C_{14} \\ C_{21} & C_{22} & C_{23} & C_{24} \\ C_{31} & C_{32} & C_{33} & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \\ 1 \end{pmatrix}.$$

This gives three equations

$$su_1 = C_{11}X_1 + C_{12}Y_1 + C_{13}Z_1 + C_{14}, \tag{1}$$
$$sv_1 = C_{21}X_1 + C_{22}Y_1 + C_{23}Z_1 + C_{24}, \tag{2}$$
$$s = C_{31}X_1 + C_{32}Y_1 + C_{33}Z_1 + 1. \tag{3}$$

Equation (3) inserted into (1) and (3) inserted into (2) gives

$$u_1 = C_{11}X_1 + C_{12}Y_1 + C_{13}Z_1 + C_{14} - C_{31}X_1u_1 - C_{32}Y_1u_1 + C_{33}Z_1u_1$$
$$v_1 = C_{21}X_1 + C_{22}Y_1 + C_{23}Z_1 + C_{24} - C_{31}X_1v_1 - C_{32}Y_1v_1 + C_{33}Z_1u_1.$$

This is the two upper rows in equation (3) in the examination. The other rows are received in a similar way by using other corresponding points (X_i, Y_i, Z_i) and (u_i, v_i), $i = 2, ..., N$ in the world and image respectively.

Exercise 16
One end-point of the sword is at $(u_a, v_a) = (220, 460) \Rightarrow$

$$\begin{pmatrix} 7.6160 \\ 53.4751 \\ 1.8620 \end{pmatrix} = \begin{pmatrix} 0.0769 & -0.0002 & -9.21 \\ 0.0159 & 0.138 & -13.5029 \\ 0.0003 & 0.0022 & 0.784 \end{pmatrix} \cdot \begin{pmatrix} 220 \\ 460 \\ 1 \end{pmatrix} \Rightarrow$$
$$\begin{pmatrix} X_a \\ Y_a \\ 1 \end{pmatrix} = \begin{pmatrix} 7.6160 \\ 53.4751 \\ 1.8620 \end{pmatrix} /1.8620 = \begin{pmatrix} 4.0902 \\ 28.7192 \\ 1 \end{pmatrix}.$$

The other end-point is at $(u_b, v_b) = (400, 380) \Rightarrow$

$$\begin{pmatrix} 21.4740 \\ 45.2971 \\ 1.7400 \end{pmatrix} = \begin{pmatrix} 0.0769 & -0.0002 & -9.21 \\ 0.0159 & 0.138 & -13.5029 \\ 0.0003 & 0.0022 & 0.784 \end{pmatrix} \cdot \begin{pmatrix} 400 \\ 380 \\ 1 \end{pmatrix} \Rightarrow$$
$$\begin{pmatrix} X_b \\ Y_b \\ 1 \end{pmatrix} = \begin{pmatrix} 21.4740 \\ 45.2971 \\ 1.7400 \end{pmatrix} /1.7400 = \begin{pmatrix} 12.3414 \\ 26.0328 \\ 1 \end{pmatrix}.$$

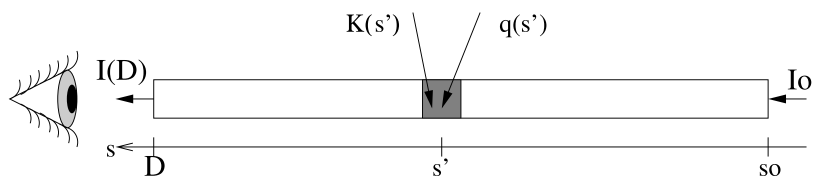
The length is therefore $\sqrt{(12.3414 - 4.0902)^2 + (26.0328 - 28.7192)^2} \approx 8.68$ dm.

PART III: NON-STANDARD IMAGE SENSORS

Exercise 17
A homography.

Exercise 18
8 patterns, since $2^8 = 256$.

Exercise 19
$\kappa(s)$ is the absorbtion function along a ray in the s-direction through the volume and $q(s)$ is the emission function along that ray.
The figure illustrates what was mentioned in the exercise.
Also, a volume element at $s = s'$ is marked, it can both absorb light $\kappa(s')$ and emit light $q(s')$.



Exercise 20
Suppose that object A is closer than object B. If you swap the images, object B will seem to be closer than object A.

Exercise 21
An event camera has significantly better temporal resolution and consequently less motion blur than a standard camera. In an event camera, the pixels operate independently and asynchrounously. Only those pixels that have changed needs to be updated at every time position. In a standard camera, the whole frame is updated at every time position.

Exercise 22
An HDR image can be produced from three images which have been captured with different exposure times.
V_0, the one with longest exposure time, will be saturated in bright areas, but dark areas will have good contrast.
V_1, the one with medium exposure time, will have good contrast for medium bright areas.
V_2, the one with shortest exposure time, will have good contrast in bright areas, whereas dark areas will be hardly visible.
V_{HDR}, the HDR image, can be created as the mean of the three images:

$$V_{\mathrm{HDR}} = \frac{V_0 + V_1 + V_2}{3}$$

Exercise 23

The program has been modified with the red text, see below.

```
for z=-127 to 128
      for x=-127 to 128
            S(x,z):=0;
            y:=-128;
            do
                  y:=y+1;
            while (f(x,y,z)<77 and y<128)
            if (f(x,y,z)<77 or fy(x,y,z)<0)
                  S(x,z):=0;
            else
                  magn := sqrt(fx(x,y,z)^2 + fy(x,y,z)^2 + fz(x,y,z)^2);
                  cosphi := max(fy(x,y,z) / magn, 0);
                  cosphi2 := cos(2*arccos(cosphi));
                  k := 0.5;
                  if (cosphi2 < 0) k := 0;
                  S(x,z) := 0.5*cosphi + k*cosphi2^2;
            end
      end
end
```

Exercise 24

a) MPI occurs when light can take several paths from emitter to sensor: partly the direct path, partly indirect paths when the light bounces more than once in the scene. When the arrival time of the light is measured, the distance comes to be overestimated, if the light that has taken an indirect path is included in the measurement.

b) • Only the *ToF* camera has MPI. It emits many light rays in many directions at once. Then MPI, as described in a), may happen.
• At a certain time point, *Lidar* sends out a pulse in only **one** direction and then recieves in only **one** direction. It can therefore not get MPI.
• A *sheet-of-light* camera illuminates a line, and can therefore get light from indirect paths. However, the camera does not calculate distance from arrival time.