ELE 510, Image Processing with Robot Vision.

Solution to Exam 7th December 2018.

Exercise 1. (25%)

a) Image sampling takes a sample of values at a discrete grid of spatial positions, usually a rectangular pattern of samples called pixels (picture elements. The quantization represent each pixel value by a finite number of bits. (See Compendium I, App. A)
b) The Bayer filter extracts the RGB color components from a pixel site in an image sensor. The pixel area is divided into 4 regions, one for Red, two for Green and one for Blue. (See Compendium I, Section 3.2)
c) This is a separable filter, the 2D impulse response is written as a product of two 1D filters, one for each spatial coordinate. (See textbook, Chapter 1 and "ConvolutionNotesLinFilters.pdf")
d) K1 is a smoothing filter (Gaussian filter). K2 is a Laplacian operator. K3 is a differentiation in the horizontal direction combined with smoothing (It is a combination of Gaussian smoothing and horizontal differentiation). (see "ELE510EdgeDetection.pdf").
e) The image gradient is a 2D vector where the first element is computed using a horizontal differentiation filter and the second element is computed using a vertical differentiation filter. (see "ELE510EdgeDetection.pdf").
f) The Harris corner detector is based on the Hessian matrix and its *eigenvalues*. A corner is detected if the following expression is above a chosen threshold: $\det(\mathcal{H}) - k\left(\frac{\mathrm{trace}(\mathcal{H})}{2}\right)^2 = \lambda_1 \cdot \lambda_2 - k\left(\frac{(\lambda_1 + \lambda_2)}{2}\right)^2$

k is a constant balancing between edge-like or corner-like structures. For an edge-like structure, one of the eigenvalues is much smaller. (see "ELE510FeaturePoints.pdf").

g) The image histogram is found by counting the number of pixels with the same intensity value and display the result as a histogram.

$h(g_k) = n_k \quad k \in \{0,1,\cdots,G-1\}$ Here n_k is the number of pixels with gray value g_k . (see “ELE510ImageEnhancementHistogram.pdf”).

h) The lines in the spectrum represent edges in the image (the line is perpendicular to the edge). The spectrum is periodic and some of the lines continue into the neighboring partition. This is due to very sharp edges, e.g. between the black coat and the background, and causes aliasing. (See LAB 2)

i) $K_x = \frac{\partial}{\partial x} G_\sigma(x,y) = -\frac{x}{2\pi\sigma^4} e^{\left(-\frac{(x^2+y^2)}{2\sigma^2}\right)}$ and similar for the y-direction.

j) This equation expresses the brightness constraint used for estimation of optical flow. (See Compendium III, section 2.2)

Exercise 2. (25%)

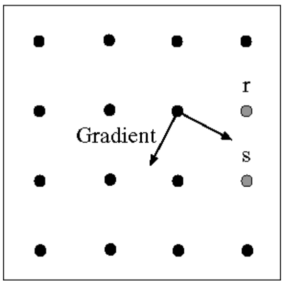
The answer to a) and b) is found in the presentation "ELE510ImageEnhancementHistogram.pdf".

a) An image is enhanced when contrast is improved, noise removed or blurring decreased. Methods are histogram manipulation, point processing, linear filters, nonlinear filters etc.
b) The method is here "histogram equalization". It is a point process where the transformation function is the cumulative distribution. This can be described by the following expression:

$$s_k = T\left(r_k\right) = \left\lfloor \frac{\left(G-1\right)}{MN} \sum_{j=0}^{k} n_j \right\rfloor \quad \text{for} \quad k \in \{0,1,\cdots,G-1\}$$

where, s_k is the new gray value and r_k the old gray value (see also Exercise 1 g) above).

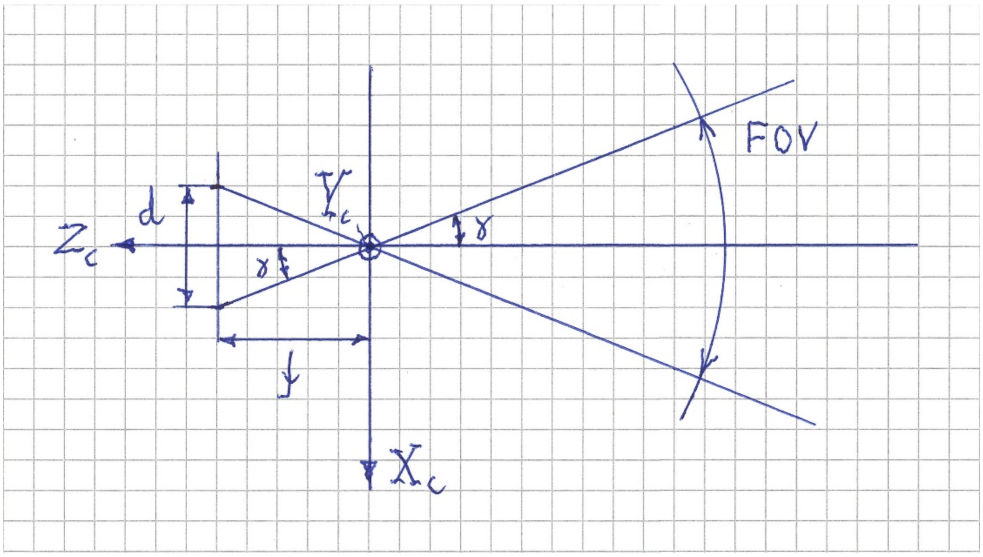
c) The answer to this and d) is found in the presentation "ELE510EdgeDetection.pdf". The Canny edge detector is based on Gaussian smoothing, computing the image gradient followed by non-maximum suppression and hysteresis thresholding.
d) The purpose of hysteresis thresholding is to connect neighbor edgels (edge points), where one may be weaker than the other, edge linking.

See Figure 

with description: Assume the marked point is an edge point. Then we construct the tangent to the edge curve (which is normal to the gradient at that point) and use this to predict the next points (here either r or s). Check the **maximum value**. At the start of a new curve a ***high threshold*** is used. To continue a curve a ***lower threshold*** is chosen.

Exercise 3. (25%)

a)



$$\mathrm{FOV} = 2\gamma \quad \tan(\gamma) = \frac{d/2}{f} \qquad \Rightarrow \qquad f = \frac{d/2}{\tan(\gamma)} = \frac{d/2}{\tan\left(\mathrm{FOV}/2\right)}$$

$$f_x = \frac{30/2}{\tan\left(53.13/2\right)} = 30.0\ \mathrm{mm} \qquad f_y = \frac{20/2}{\tan\left(36.87/2\right)} = 30.0\ \mathrm{mm}$$

The focal length is the same for the two directions, 30 mm. This is usually the case, but calibration can often show a slight difference in practise.

b) First we find the *scaling parameters*:

$$\Delta x = \frac{d_x}{N} = \frac{30 \cdot 10^{-3}}{3000} = 10 \cdot 10^{-6}\,\mathrm{m} = 10\,\mu\mathrm{m} \qquad \Delta y = \frac{d_y}{M} = \frac{20 \cdot 10^{-3}}{2000} = 10 \cdot 10^{-6}\,\mathrm{m} = 10\,\mu\mathrm{m}$$

$$\alpha = \frac{f}{\Delta x} = \frac{30 \cdot 10^{-3}}{10 \cdot 10^{-6}} = 3000 \qquad \beta = \frac{f}{\Delta y} = \frac{30 \cdot 10^{-3}}{10 \cdot 10^{-6}} = 3000$$

Then the *principle point*:

$$x_0 = \frac{N}{2} = \frac{3000}{2} = 1500 \qquad y_0 = \frac{M}{2} = \frac{2000}{2} = 1000$$

For no skew, we get **the internal calibration matrix**:

$$\mathcal{K} = \begin{pmatrix} \alpha & 0 & x_0 \\ 0 & \beta & y_0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 3000 & 0 & 1500 \\ 0 & 3000 & 1000 \\ 0 & 0 & 1 \end{pmatrix}.$$

c) In order to align the world coordinates with the camera coordinates we *rotate the world coordinates around the X-axis 90 degrees*. The origin is translated along the camera coordinates as given by the *translation vector* **t**. The result is

$$\mathcal{R} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -(-1) \\ 0 & (-1) & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}$$

$$\mathbf{t} = \begin{bmatrix} -2 \\ -1 \\ 3 \end{bmatrix}$$

The camera matrix is then:

$$\mathcal{M} = \mathcal{K}[\mathcal{R}\ \mathbf{t}] = \begin{pmatrix} 3000 & 0 & 1500 \\ 0 & 3000 & 1000 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & -2 \\ 0 & 0 & 1 & -1 \\ 0 & -1 & 0 & 3 \end{pmatrix}$$
$$= \begin{pmatrix} 3000 & -1500 & 0 & -1500 \\ 0 & -1000 & 3000 & 0 \\ 0 & -1 & 0 & 3 \end{pmatrix}.$$

Moreover, the **normalized camera matrix**:

$$\mathcal{M}_{norm} = \begin{pmatrix} 1000 & -500 & 0 & -500 \\ 0 & -333.3 & 1000 & 0 \\ 0 & -0.3333 & 0 & 1 \end{pmatrix}.$$

d) The four corners of robot B in world coordinates are:

$$P_1 = \begin{bmatrix} 1.5 \\ 15 \\ 0 \\ 1 \end{bmatrix} \quad P_2 = \begin{bmatrix} 2.5 \\ 15 \\ 0 \\ 1 \end{bmatrix} \quad P_3 = \begin{bmatrix} 1.5 \\ 15 \\ 0.5 \\ 1 \end{bmatrix} \quad P_4 = \begin{bmatrix} 2.5 \\ 15 \\ 0.5 \\ 1 \end{bmatrix}$$

The corresponding points in the image plane of camera A is:

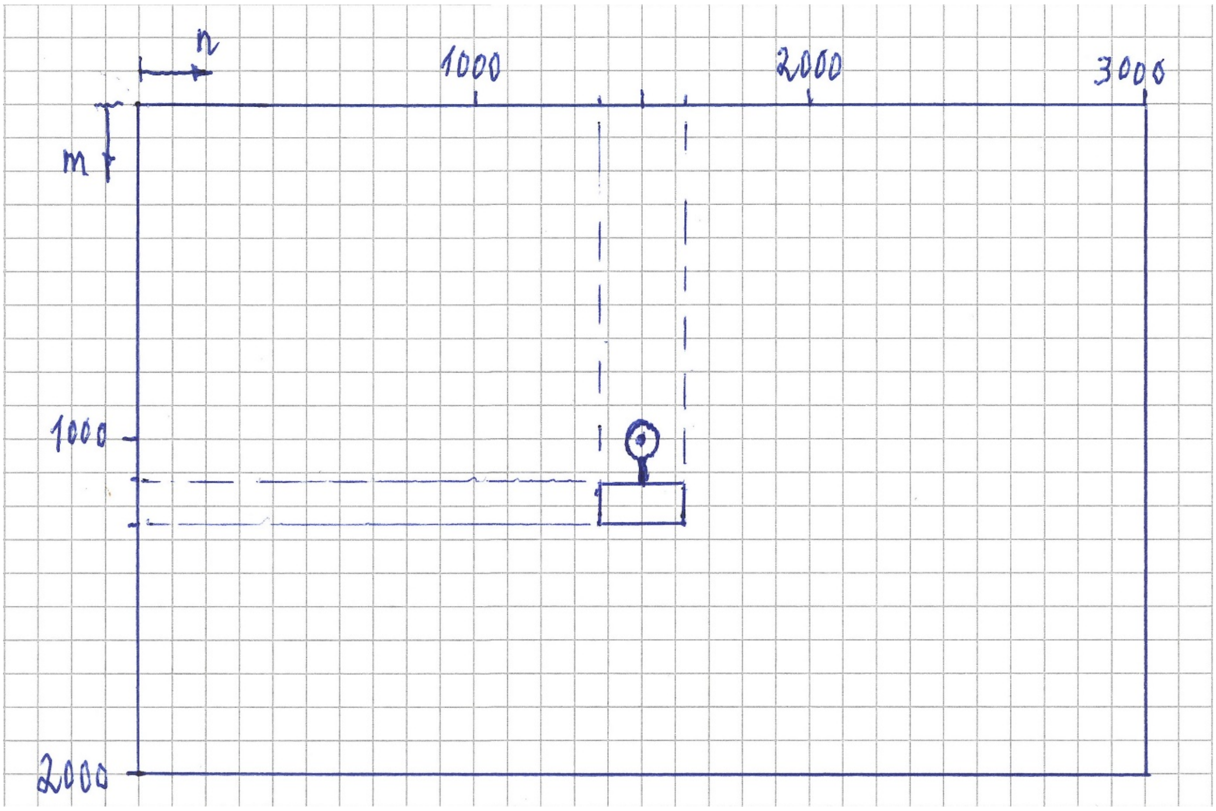
$$\lambda p_k = \mathcal{M}_{norm} P_k = \begin{pmatrix} 1000 & -500 & 0 & -500 \\ 0 & -333.3 & 1000 & 0 \\ 0 & -0.3333 & 0 & 1 \end{pmatrix} P_k \quad \text{for} \quad k \in \{1,2,3,4\}$$

From the third row we get $\lambda = -4$ and the resulting image points in pixels are

$$\tilde{p}_1 = \begin{bmatrix} 1625 \\ 1250 \end{bmatrix} \quad \tilde{p}_2 = \begin{bmatrix} 1375 \\ 1250 \end{bmatrix} \quad \tilde{p}_3 = \begin{bmatrix} 1625 \\ 1125 \end{bmatrix} \quad \tilde{p}_4 = \begin{bmatrix} 1375 \\ 1125 \end{bmatrix} \quad \text{where}$$

$$\tilde{p}_k = \frac{1}{\lambda} \begin{bmatrix} \lambda p_k(1) \\ \lambda p_k(2) \end{bmatrix} \quad \text{for} \quad k \in \{1,2,3,4\}$$

The camera center of robot B is on the optical axis of camera A and therefore imaged to the center of the image plane. The result is:



Note how the image axis's in pixels have been flipped to give an "upright" image of the object.

e) Rotation of the camera A around the Y_C axis:

$$\mathbf{P}_{c2} = \mathcal{R}_2 \mathbf{P}_{c1} \quad \text{where} \quad \mathcal{R}_2 = \begin{pmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{pmatrix}$$

$$\mathbf{P}_{c1} = \mathcal{R}_1 \mathbf{P}_w + \mathbf{t}_1 \quad \Rightarrow \quad \mathbf{P}_{c2} = \mathcal{R}_2 \left(\mathcal{R}_1 \mathbf{P}_w + \mathbf{t}_1 \right) = \mathcal{R}_2 \mathcal{R}_1 \mathbf{P}_w + \mathcal{R}_2 \mathbf{t}_1 \quad \Rightarrow$$

$$\mathcal{R}_{new} = \mathcal{R}_2 \mathcal{R}_1 = \begin{pmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \sin(\theta) & 0 \\ 0 & 0 & 1 \\ \sin(\theta) & -\cos(\theta) & 0 \end{pmatrix}$$

$$\mathbf{t}_{new} = \mathcal{R}_2 \mathbf{t}_1 = \begin{pmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{pmatrix} \begin{bmatrix} -2 \\ -1 \\ 3 \end{bmatrix} = \begin{bmatrix} -2\cos(\theta) - 3\sin(\theta) \\ -1 \\ -2\sin(\theta) + 3\cos(\theta) \end{bmatrix}.$$

Exercise 4. (25%)

a) The robot B moves towards A in a straight line. The distance (Y-axis) for the lights on B in world coordinates as a function of time (frame number) is

$$Y(t) = 15 - V \cdot t \quad t = \frac{1}{40} k \quad k \in \{0, 1, 2, \cdots\} \text{ for } t \geq 0$$

$$Y(k) = 15 - 2 \cdot \frac{1}{40} k = 15 - 0.05 \cdot k$$

The two light points are

$$P_{l1} = \begin{bmatrix} 1.5 \\ 15 - 0.05k \\ 0.5 \\ 1 \end{bmatrix} \quad P_{l2} = \begin{bmatrix} 2.5 \\ 15 - 0.05k \\ 0.5 \\ 1 \end{bmatrix}$$

We use the camera matrix and find the image points as a function of frame number (time) for the right light, P_{l1} :

$$\lambda p(k) = \mathcal{M}_{norm} P(k) = \begin{pmatrix} 1000 & -500 & 0 & -500 \\ 0 & -333.3 & 1000 & 0 \\ 0 & -0.3333 & 0 & 1 \end{pmatrix} \begin{bmatrix} 1.5 \\ 15 - 0.05 \cdot k \\ 0.5 \\ 1 \end{bmatrix} \quad \text{for} \quad k \in \{0, 1, 2, 3 \cdots\}$$

The third row give the value of λ and the two first rows divided by λ give *x(k)* and *y(k)*.

$$\lambda(k) = -0.3333(15-0.05\cdot k)+1 = -4+\frac{0.05}{3}k$$

$$\lambda x(k) = 1500-500(15-0.05\cdot k)-500 = 1500-500\left[(15-0.05\cdot k)+1\right]$$

$$= -6000+500\cdot 0.05k-500 = 1500\left(-4+\frac{0.05}{3}k\right)-500 \quad \Rightarrow$$

$$x(k) = 1500+\frac{500}{\left(4-\frac{0.05}{3}k\right)},$$

$$\lambda y(k) = -333.3(15-0.05\cdot k)+1000\cdot 0.5$$

$$= 1000\left[-0.3333(15-0.05\cdot k)+1\right]-500 \quad \Rightarrow$$

$$y(k) = 1000+\frac{500}{\left(4-\frac{0.05}{3}k\right)}.$$

We see that the change in *x* and *y* is the same as a function of frame number *k*.

b) The frame number corresponding to *t* = 4.0 seconds is

$$t = \frac{1}{40}k \quad \Rightarrow \quad k = 40\cdot t$$

$$k_{4s} = 40\cdot 4 = 160$$

We then find the optical flow by subtracting the position at frame number (160+1) from the position at frame 160. Except for the stationary value the change in x and y is the same. Therefor the velocity vector has two equal components such that the direction is at an angle of 45 degrees. Figure 5 in the exam text can confirm this. The track to the right corresponds to the right light. The resulting optical flow at time *t* = 4.0 seconds is

$$\mathbf{v}(t=4.0\mathrm{s})=\begin{bmatrix} x(161)-x(160) \\ y(161)-y(160) \end{bmatrix}=500\left(\frac{1}{4-\frac{0.05}{3}161}-\frac{1}{4-\frac{0.05}{3}160}\right)\begin{bmatrix}1\\1\end{bmatrix}$$

$$=500\frac{0.15}{15.8}\begin{bmatrix}1\\1\end{bmatrix}=4.63\begin{bmatrix}1\\1\end{bmatrix}\text{pixels.}$$

c) The optical flow is a **diverging field** because the motion is along the optical axis. The point in the center of Figure 5 is the "**Focus of Expansion**", **FOE**, because we have an expanding velocity field when the motion is towards the scene (robot A). For other points in the scene, the magnitude of the velocity is inversely to the depth and direction radially out from the FOE. The FOE is here in the center (identical to the principle point) because there is no motion in the X_C or Y_C direction (perpendicular to the optical axis).

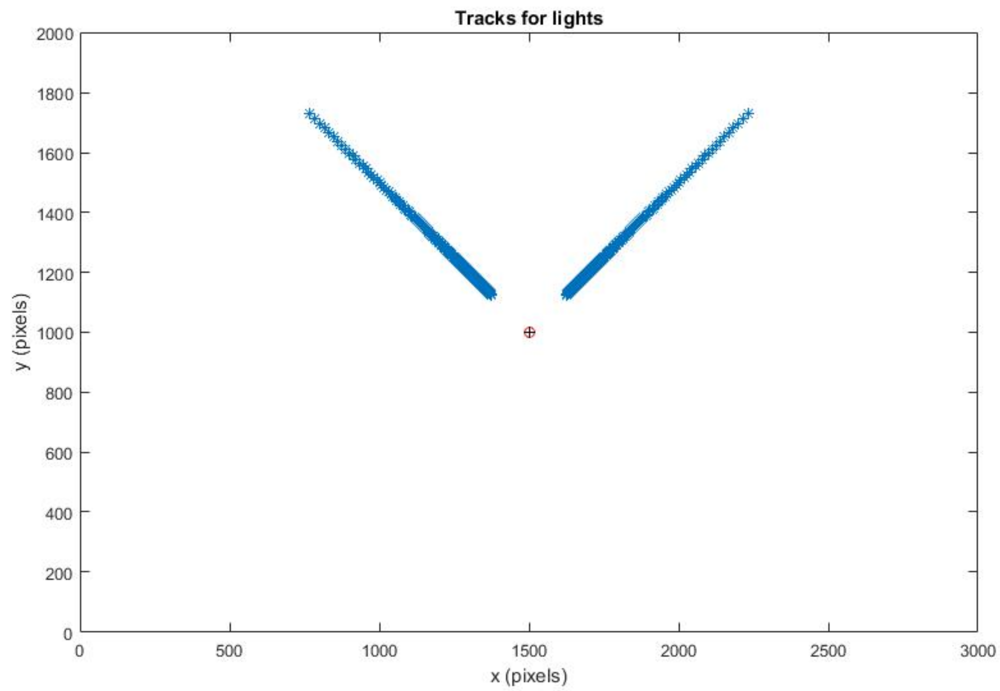


Figure 5.

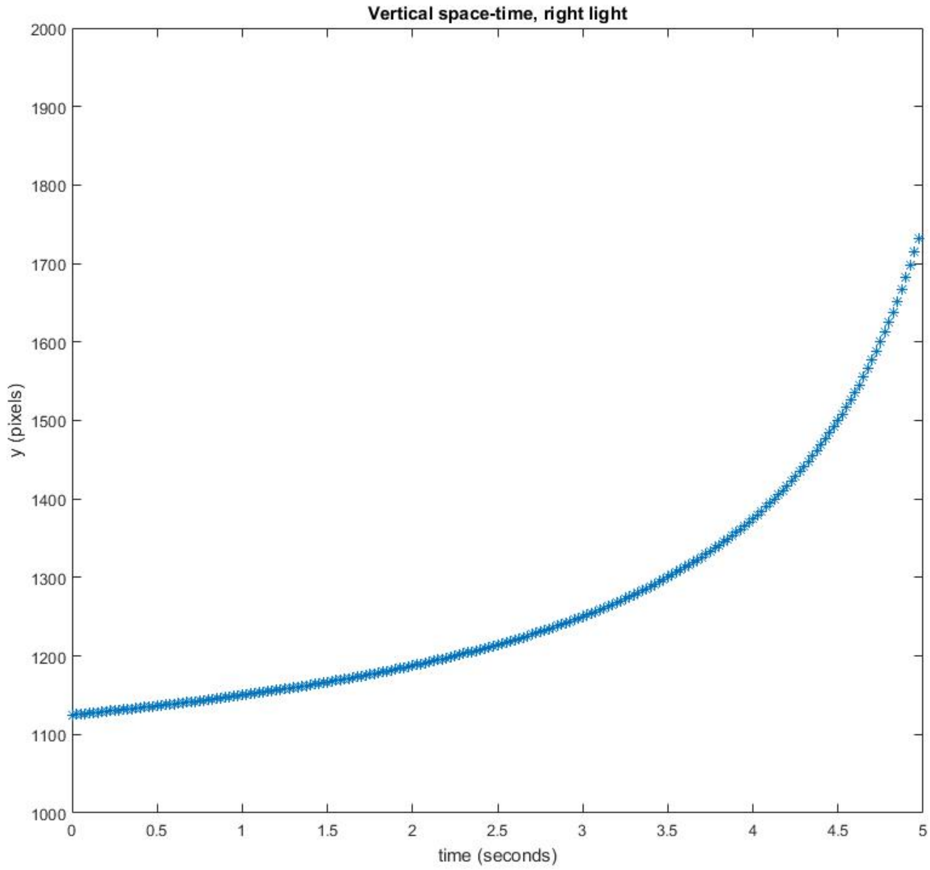


Figure 6.

d) The time to collision can be found from a length in the image and its time derivative (See Compendium III, section 1.1):

$$\tau = \frac{l(t)}{l'(t)}$$

For the case in this exercise let's use the length between the two lights on robot B. At *t* = 0 seconds we have $x(0)=1500+\frac{500}{4}=1500+125$. As the two lights are symmetric around the center in the image (see Figure 5), the distance at time zero is $2\cdot 125=250$ pixels. One time step ahead we have $x(1)=1500+\frac{500}{4-0.05/3}=1500+125.523$, i.e. the length is now $2\cdot 125.523=251.046$ pixels. This gives

$$\tau = \frac{l(t)}{l'(t)} = \frac{l(0)}{\Delta l / \Delta t} = \Delta t \cdot \frac{l(0)}{l(1) - l(0)} = 0.025 \cdot \frac{250}{251.046 - 250}\,\text{s}$$
$$= 5.975\,\text{s}$$

The distance from the camera center A to the front of robot B (the position of the lights) is 12.0 m at time zero. With a speed of 2 m/s, the time to collision should be 6.0 seconds. The difference between this and the answer above corresponds to one time increment, Δt . This is reasonable since we used frame 1 in the computation of the derivative!