Visual Navigation

Winter Term 2015/2016

Dr. G. Giorgi

Prof. Dr. sc. nat. C. Günther

Technische Universität München
Lehrstuhl für Kommunikation und Navigation

Exam

September 26, 2016

- The use of the following material is permitted: lecture slides, referenced books and articles, exercises and solutions, a pocket (also programmable) calculator. Cell phones, laptops etc. . . are prohibited.
- Please indicate your name and registration number on all pages. The problems are described on 4 pages (including this one). Please also return these pages.
- The examination is divided into 3 areas. Each area typically includes several smaller problems, some of which are dependent. The subproblems marked with (*) can be solved without requiring the solution of previous subproblems.
- The maximum score is reached when a problem is solved. In the case that you cannot solve a problem completely, you have the option to provide a clear and complete description of the solution. A certain credit will also be given for such a description.

I hereby confirm that I have been informed prior to begin of the examination that I have to notify the examination supervisors immediately if sudden illness occurs during the examination. This will be noted in the examination protocol. An application for exam withdrawal has to be filed immediately at the board of examiners being in charge. A medical certificate from one of the physicians acknowledged by the Technische Universität München issued on the same day as the examination must be forwarded without delay. In case the examination is regularly completed despite of illness, a subsequent withdrawal due to illness cannot be accepted. In case the examination is ended due to illness it will not be graded.

Name: ______________________________

Matriculation number: ______________________________

Course of studies: ______________________________

München, ____________ ______________________________
(Date) (Signature)

Problem 1 *Total: 8 points*

Consider a camera characterized by the following calibration matrix:

$$\mathbf{K} = \begin{bmatrix} -f & 0 & P_x \\ 0 & -f & P_y \\ 0 & 0 & 1 \end{bmatrix} \tag{1-1}$$

with $f = 100\, px$ the focal distance and (P_x, P_y) the camera principal point offset. A simple calibration process needs to be set up in order to extract the principal point coordinates.

a) (*)[2 points] Derive a method to estimate the coordinates (P_x, P_y) from a set of spatial (three-dimensional) points *of known coordinates* and their corresponding projections onto the image focal plane. What is the minimum number of correspondences needed to unambiguously retrieve the principal point coordinates?

b) [1 point] Given the following correspondences between spatial points and their projections (both in homogeneous coordinates), compute the principal point coordinates (P_x, P_y):

$$\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \leftrightarrow \begin{pmatrix} 1 \\ 2 \\ 1 \\ 1 \end{pmatrix} \qquad \begin{pmatrix} 300 \\ 500 \\ 3 \end{pmatrix} \leftrightarrow \begin{pmatrix} 0 \\ 1 \\ 3 \\ 1 \end{pmatrix} \qquad \begin{pmatrix} -200 \\ 0 \\ 1 \end{pmatrix} \leftrightarrow \begin{pmatrix} 3 \\ 2 \\ 1 \\ 1 \end{pmatrix} \tag{1-2}$$

c) (*)[3 points)] A camera characterized by a calibration matrix $\mathbf{K}$ as in (1-1) performs a rotation of 45 degrees about the Z-axis and a translation that brings its center from the origin to the point identified by the Euclidean coordinates $(0, 1, 0)^{\mathrm{T}}$. Compute the projected pixel coordinates of the set of points lying on the line perpendicular to the reference XY plane and passing through point $(1, 1, 1)^{\mathrm{T}}$, as function of the camera principal point $(P_x, P_y)^{\mathrm{T}}$.

Hint. The rotation of magnitude θ around the Z-axis is described by the following rotation matrix:

$$\mathbf{R} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1-3}$$

d) [2 points] Compute the epipoles e and e' relative to the two-view geometry identified by camera matrices $\mathbf{P}$ and $\mathbf{P}'$, where the latter identifies the camera matrix after the movement described in the previous point has taken place.

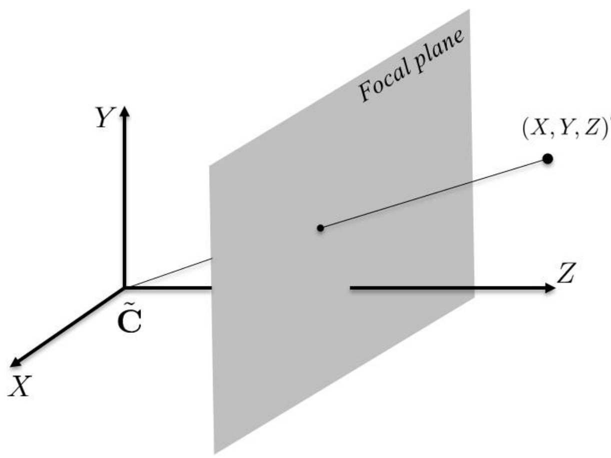

Figure 1-1: Camera local coordinate frame and focal plane.

Problem 2 *Total: 10 points*

A monocular camera is characterized by the following calibration matrix:

$$\mathbf{K} = \begin{bmatrix} -100 & 0 & 100 \\ 0 & -100 & 100 \\ 0 & 0 & 1 \end{bmatrix} \tag{2-1}$$

Its initial position defines the reference frame, so that the camera matrix can be written as

$$\mathbf{P} = \mathbf{K} \left[\mathbf{I}_3 \mid \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right] \tag{2-2}$$

The camera is then moved such that its center is now localized at point (Euclidean coordinates) $\tilde{\mathbf{C}}' = (2, 0, -1)^{\mathrm{T}}$, while performing a 90-degrees rotation about the reference Y-axis.

a) (*)[2 points] Derive the epipoles of the two-view geometry related to the above camera movement *in pixel coordinates*.

b) (*)[2 points] Compute the fundamental matrix.

c) (*)[3 points] Reconstruct the spatial coordinates of a point that is projected as $\mathbf{x} = (100, -200, 2)^{\mathrm{T}}$ (homogeneous coordinates) before the camera movement, and as $\mathbf{x}' = (-200, -300, 1)^{\mathrm{T}}$ (homogeneous coordinates) after the camera has moved. You may assume ideal conditions, so the pixel coordinates are precisely known.
Is the point reconstructed with a scale ambiguity, or are its absolute coordinates fully recovered? Justify your answer.

d) [3 points] Assume the feature detection process causes a certain amount of noise when extracting the feature pixel coordinates. This error may be described as $\mathbf{Q}_{\tilde{\mathbf{x}}} = \sigma^2 \mathbf{I}_2$ (in Euclidean coordinates). Characterize the reconstruction error when taking into account the pixel measurement error in the procedure you used in the previous point.

Notes.

I - The rotation of magnitude ϕ around the Y-axis is described by the following rotation matrix:

$$\mathbf{R} = \begin{bmatrix} \cos\phi & 0 & \sin\phi \\ 0 & 1 & 0 \\ -\sin\phi & 0 & \cos\phi \end{bmatrix} \tag{2-3}$$

II - The inverse of a 3x3 matrix $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$ can be computed as follows:

$$\mathbf{A}^{-1} = \frac{1}{\det\mathbf{A}} \begin{bmatrix} a_{33}a_{22} - a_{32}a_{23} & -(a_{33}a_{12} - a_{32}a_{13}) & a_{23}a_{12} - a_{22}a_{13} \\ -(a_{33}a_{21} - a_{31}a_{23}) & a_{33}a_{11} - a_{31}a_{13} & -(a_{23}a_{11} - a_{21}a_{13}) \\ a_{32}a_{21} - a_{31}a_{22} & -(a_{32}a_{11} - a_{31}a_{12}) & a_{22}a_{11} - a_{21}a_{12} \end{bmatrix} \tag{2-4}$$

with $\det\mathbf{A} = a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{12}a_{23}a_{31} - a_{31}a_{22}a_{13} - a_{12}a_{21}a_{33} - a_{23}a_{32}a_{11}$.

Problem 3 *Total: 12 points*

Consider a camera-based positioning system used on an autonomous underground train that travels perfectly straight paths. A series of visual markers consisting of signs displaying a recognizable pattern (similar to a QR code, in which a single anchor reference point can be assigned) are placed on the right wall of the tunnel in which the train drives (assume the tunnel having a square section). A side-looking camera installed on the train is employed to recognize and store the positions and characteristics of the landmarks. To reduce the installation cost, the position of the markers is not planned in advance, thus the system has to autonomously build a map of the signs.
You are required to design the VSLAM core algorithm for the map-building process using a Kalman filter. The parameters of the problem are:

- Only an image-based navigation system is used, without aiding of any other sensor.
- Consider a realistic constant-acceleration motion model.
- The side-looking camera always points perpendicularly to the longitudinal axis of the train.
- The signs are spread so that at most only one sign falls within the camera field of view at any given time.
- An internal algorithm is used to recognize the markers and compute their positions *in the robot's local reference frame, built as follows: axis X' directed along-track, axis Z' directed towards the wall, perpendicularly to X', and axis Y' directed upwards, so to complete the right-hand frame*. Furthermore, since the same sign is tracked for the whole time it remains within the camera field of view, an estimation of the relative speed between the camera and the sign can also be extracted. The time interval at which the system operates is $\Delta t = 1$s.
- The signs are placed on the tunnel's wall, whose distance from the railway is constant. Assume that the distance between the camera and the wall is constant and equals to d along the camera Z' axis. The along-track position and height of the signs are initially unknown.

a) (*)[4 points] Design a suitable *prediction* step for the Kalman filter that will be used in this application. Credit will be given if the number of parameters used is minimal.

b) [4 points] Design a suitable *observation model* for the Kalman filter that will be used in this application. Credit will be given if all the available information that can be extracted from continuously tracked signs is used.

c) [2 points] Using the Kalman Filter built in the previous two points, compute the updated state estimate at $t = 1$ s starting from the following initial conditions:

  - The train initial position ($t = 0$) is at the origin of the chosen reference frame; the initial velocity is null; the initial acceleration measures 3 m/s^2 and it is directed towards the along-track direction.
  - The initial uncertainty of the state can be considered null (that is, the initial state is deterministically known, with no associated uncertainty).
  - No marker has been pre-loaded in the map.

d) [2 points] At $t = 1$ s, a sign enters the field of view of the camera, with anchor point estimated to be at following coordinates: $(X', Y', Z') = (2, 0, d)$ m. Compute the augmented state vector and the associated covariance matrix, considering that the aforementioned coordinates have been estimated with an error characterized by

$$\mathbf{Q} = \sigma_{m_1}^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \tag{3-1}$$

Visual Navigation

Winter Term 2015/2016

Dr. G. Giorgi

Prof. Dr. sc. nat. C. Günther

Technische Universität München
Lehrstuhl für Kommunikation und Navigation

Exam

September 26, 2016

- The use of the following material is permitted: lecture slides, referenced books and articles, exercises and solutions, a pocket (also programmable) calculator. Cell phones, laptops etc... are prohibited.
- Please indicate your name and registration number on all pages. The problems are described on 4 pages. Please also return those pages.
- The examination is subdivided into 3 areas. Each area typically includes several smaller problems, some of which are dependent.
- The maximum score is reached when a problem is solved. In the case that you cannot solve a problem completely, you have the option to provide a clear and complete description of the solution. A certain credit will also be given for such a description.

I hereby confirm that I have been informed prior to begin of the examination that I have to notify the examination supervisors immediately if sudden illness occurs during the examination. This will be noted in the examination protocol. An application for exam withdrawal has to be filed immediately at the board of examiners being in charge. A medical certificate from one of the physicians acknowledged by the Technische Universität München issued on the same day as the examination must be forwarded without delay. In case the examination is regularly completed despite of illness, a subsequent withdrawal due to illness cannot be accepted. In case the examination is ended due to illness it will not be graded.

Name: ______________________________

Matriculation number: ______________________________

Course of studies: ______________________________

München, ____________ ______________________________
(Date) (Signature)

Problem 1 – Solution

a) A spatial point $\mathbf{X}$ is projected onto the focal plane as

$$\mathbf{x} = \mathbf{K}\,[\mathbf{R} \mid \mathbf{t}]\,\mathbf{X} \tag{1-1}$$

where both $\mathbf{x}$ and $\mathbf{X} = (X, Y, Z, 1)^{\mathrm{T}}$ are given in homogeneous coordinates, $\mathbf{R}$ is the camera rotation with respect to the reference frame and $\mathbf{t}$ the translation vector. In order to calibrate the camera by retrieving the principal point coordinates, we take as reference frame the camera frame, such that $\mathbf{R} = \mathbf{I}_3$ and $\mathbf{t} = \mathbf{0}$, giving

$$\mathbf{x} = \mathbf{K}\,[\mathbf{I}_3 \mid \mathbf{0}]\,\mathbf{X} = \begin{pmatrix} P_x Z - fX \\ P_y Z - fY \\ Z \end{pmatrix} = \begin{pmatrix} P_x Z - 100X \\ P_y Z - 100Y \\ Z \end{pmatrix} \tag{1-2}$$

Each correspondence $\mathbf{x} \leftrightarrow \mathbf{X}$ gives two independent identities, so that a single spatial point of known coordinates can be used to obtain the value of P_x and P_y.

b) Substitution of the first correspondence

$$\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \leftrightarrow \begin{pmatrix} 1 \\ 2 \\ 1 \\ 1 \end{pmatrix} \tag{1-3}$$

in (1-2) gives

$$\begin{cases} 0 = P_x - 100 \\ 0 = P_y - 200 \end{cases} \tag{1-4}$$

whose solution is $P_x = 100$ px and $P_y = 200$ px. Note that any of the other two correspondences could have been used.

c) The rotation matrix about the Z-axis reads

$$\mathbf{R} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} \sqrt{2} & -\sqrt{2} & 0 \\ \sqrt{2} & \sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1-5}$$

After the given translation and rotation, the camera matrix $\mathbf{P}'$ reads

$$\mathbf{P}' = \mathbf{K}\mathbf{R}\left[\mathbf{I} \mid -\tilde{\mathbf{C}}\right] = \begin{bmatrix} -50\sqrt{2} & 50\sqrt{2} & P_x \\ -50\sqrt{2} & -50\sqrt{2} & P_y \\ 0 & 0 & 1 \end{bmatrix} \left[\mathbf{I} \mid \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}\right] \tag{1-6}$$

The set of points lying on the line perpendicular to the reference XY plane and passing through point $(1, 1, 1)^{\mathrm{T}}$ are simply identified by Euclidean coordinates $\mathbf{X} = (1, 1, Z)^{\mathrm{T}}$. Therefore, the projection of these points is found as

$$\mathbf{x} = \mathbf{P}'\mathbf{X} = \begin{pmatrix} P_x Z - 50\sqrt{2} \\ P_y Z - 50\sqrt{2} \\ Z \end{pmatrix} \tag{1-7}$$

The actual pixel coordinates of such projections are then

$$\tilde{\mathbf{x}} = \begin{pmatrix} P_x - \frac{50\sqrt{2}}{Z} \\ P_y - \frac{50\sqrt{2}}{Z} \end{pmatrix} \tag{1-8}$$

d) One obtains

$$\mathbf{e} = \mathbf{P}\mathbf{C}' = \mathbf{K}\left[\mathbf{I} \mid \mathbf{0}\right] \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} = \begin{bmatrix} -100 & 0 & P_x \\ 0 & -100 & P_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ -100 \\ 0 \end{pmatrix} \tag{1-9}$$

and

$$\mathbf{e}' = \mathbf{P}'\mathbf{C} = \mathbf{KR}\left[\mathbf{I} \mid \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}\right]\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -50\sqrt{2} \\ 50\sqrt{2} \\ 0 \end{pmatrix} \tag{1-10}$$

Note that the epipole $\mathbf{e}$ is the vanishing point of the Y-axis (obviously since the pure translation was parallel to the Y-axis), whereas the epipole $\mathbf{e}'$ is the vanishing point on the line that bisects the XY-plane (again obvious, since a rotation of 45 degrees was performed about the Z-axis).

Problem 2 – Solution

a) The initial camera matrix is

$$\mathbf{P} = \begin{bmatrix} -100 & 0 & 100 & 0 \\ 0 & -100 & 100 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{2-1}$$

After the movement, we obtain

$$\mathbf{P}' = \mathbf{KR}\left[\mathbf{I}_3 \mid -\tilde{\mathbf{C}}\right] = \begin{bmatrix} -100 & 0 & -100 & 100 \\ -100 & -100 & 0 & 200 \\ -1 & 0 & 0 & 2 \end{bmatrix} \tag{2-2}$$

The epipoles are then

$$\mathbf{e} = \mathbf{P}\tilde{\mathbf{C}}' = \begin{pmatrix} -300 \\ -100 \\ -1 \end{pmatrix} \quad \rightarrow \quad \tilde{\mathbf{e}} = \begin{pmatrix} 300 \\ 100 \end{pmatrix} \tag{2-3}$$

and

$$\mathbf{e}' = \mathbf{P}'\tilde{\mathbf{C}} = \begin{pmatrix} 100 \\ 200 \\ 2 \end{pmatrix} \quad \rightarrow \quad \tilde{\mathbf{e}}' = \begin{pmatrix} 50 \\ 100 \end{pmatrix} \tag{2-4}$$

b) The fundamental matrix is found as

$$\mathbf{F} = \boldsymbol{\Omega}_{\mathbf{e}'}\mathbf{P}'\mathbf{P}^{+} \tag{2-5}$$

with $\boldsymbol{\Omega}_{\mathbf{e}'}$ the skew-symmetric matrix associated to the epipole $\mathbf{e}'$.
The pseudoinverse $\mathbf{P}^{+}$ is

$$\mathbf{P}^{+} = \mathbf{P}^{\mathrm{T}}(\mathbf{P}\mathbf{P}^{\mathrm{T}})^{-1} = \begin{bmatrix} -0.01 & 0 & 1 \\ 0 & -0.01 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \tag{2-6}$$

Thus

$$\mathbf{F} = \begin{bmatrix} 0 & -2 & 200 \\ 1 & 0 & -300 \\ -100 & 100 & 0.0002 \end{bmatrix} \tag{2-7}$$

c) From the camera matrices before and after the movement

$$\mathbf{P} = \begin{bmatrix} -100 & 0 & 100 & 0 \\ 0 & -100 & 100 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad ; \quad \mathbf{P}' = \begin{bmatrix} -100 & 0 & -100 & 100 \\ -100 & -100 & 0 & 200 \\ -1 & 0 & 0 & 2 \end{bmatrix} \tag{2-8}$$

we can take the first two identities of each projection $\mathbf{x} = \mathbf{P}\mathbf{X}$ and $\mathbf{x}' = \mathbf{P}'\mathbf{X}$ to obtain

$$\begin{cases} 100 = -100X + 100Z \\ -200 = -100Y + 100Z \\ -200 = -100X - 100Z + 100 \\ -300 = -100X - 100Y + 200 \end{cases} \tag{2-9}$$

Since we only have three unknowns, we can use just the first three identities, and by re-ordering we obtain

$$\begin{cases} 100 = -100X + 100Z \\ -200 = -100Y + 100Z \\ -300 = -100X - 100Z \end{cases} \tag{2-10}$$

Solving the system gives $\tilde{\mathbf{X}} = (1, 4, 2)^{\mathrm{T}}$. There is no scale ambiguity in the reconstruction of the 3D coordinates, although we operate with a monocular camera. The reason is that the exact extent of the camera movement is known in this case, enabling us to reconstruct the 3D coordinates of the observed point as we were observing it from a stereo camera of known geometry.

d) Expression (2-10) is a linear system $\mathbf{y} = \mathbf{M}\tilde{\mathbf{X}}$, with $\mathbf{y} = (100, -200, -300)^{\mathrm{T}}$ and

$$\mathbf{M} = \begin{bmatrix} -100 & 0 & 100 \\ 0 & -100 & 100 \\ -100 & 0 & -100 \end{bmatrix} \tag{2-11}$$

The error propagation reads then

$$\mathbf{Q}_{\tilde{\mathbf{X}}} = \left(\mathbf{M}^{\mathrm{T}}\mathbf{Q}_{\mathbf{y}}^{-1}\mathbf{M}\right)^{-1} \tag{2-12}$$

Assuming no correlation across images, the covariance matrix $\mathbf{Q}_{\mathbf{y}}^{-1}$ is diagonal:

$$\mathbf{Q}_{\mathbf{y}} = \sigma^2\mathbf{I}_3 \tag{2-13}$$

which then gives

$$\mathbf{Q}_{\tilde{\mathbf{X}}} = 5 \cdot 10^{-5}\sigma^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 1 & 1 \end{bmatrix} \tag{2-14}$$

Problem 3 – Solution

a) The problem is strictly one-dimensional: the along-track coordinate, hereafter referred to as s, is sufficient to locate the train. Using a constant-acceleration model, we will need to include speed ($\dot{s}$) and acceleration ($\ddot{s}$) in the state vector. There is no need to model the train attitude. A reasonable choice of for the state vector parameters is then

$$\mathbf{x} = \begin{pmatrix} s \\ \dot{s} \\ \ddot{s} \\ \mathbf{m} \end{pmatrix} \tag{3-1}$$

where $\mathbf{m}$ is the $2m$-vector containing the 2D coordinates of the stored m features:

$$\mathbf{m} = \begin{pmatrix} \mathbf{m}_1 \\ \vdots \\ \mathbf{m}_m \end{pmatrix} \quad \text{with} \quad \mathbf{m}_i = \begin{pmatrix} s_i \\ h_i \end{pmatrix} \tag{3-2}$$

with s_i the along-track coordinate of landmark i and h_i its height (no need of course to store a third coordinate of the landmarks: their distance to the camera, and thus the track, is constant, d). In other terms, just the along-track

coordinate and a height parameter are sufficient to locate each landmark.
A suitable motion model can be formed as

$$\mathbf{x}_t = \mathbf{A}_t \mathbf{x}_{t-1} + \boldsymbol{\epsilon} \quad , \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{R}_t) \tag{3-3}$$

with

$$\mathbf{A}_t = \begin{bmatrix} 1 & \Delta t & \frac{1}{2}(\Delta t)^2 & \mathbf{0} \\ 0 & 1 & \Delta t & \mathbf{0} \\ 0 & 0 & 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{2m} \end{bmatrix} \tag{3-4}$$

in which we assumed that the features are static.
We can inject now some process noise, to account for the uncertainty in the train predicted location, speed and acceleration. A reasonable noise matrix is

$$\mathbf{R}_t = \begin{bmatrix} (\Delta t)^4 \sigma_p^2 & (\Delta t)^3 \sigma_p^2 & (\Delta t)^2 \sigma_p^2 & \mathbf{0} \\ (\Delta t)^3 \sigma_p^2 & (\Delta t)^2 \sigma_p^2 & (\Delta t) \sigma_p^2 & \mathbf{0} \\ (\Delta t)^2 \sigma_p^2 & (\Delta t) \sigma_p^2 & \sigma_p^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{3-5}$$

in which we model the uncertainty on the acceleration with σ_p^2 and we propagate it into velocity and position uncertainty by integration. We do not inject any error in the map part. We could also trade some accuracy in the error propagation for numerical efficiency by removing the off-diagonal elements:

$$\mathbf{R}_t = \begin{bmatrix} (\Delta t)^4 \sigma_p^2 & 0 & 0 & \mathbf{0} \\ 0 & (\Delta t)^2 \sigma_p^2 & 0 & \mathbf{0} \\ 0 & 0 & \sigma_p^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{3-6}$$

b) The camera, pointed rightward with respect to the along-track direction, captures the signs anchor points, enabling a reconstruction of their 2D coordinates *in a local frame* as

$$\mathbf{m}_i' = \begin{pmatrix} X_i' \\ Y_i' \\ Z_i' \end{pmatrix} \tag{3-7}$$

Given the problem constraints, the coordinate X_i' equals the difference between the global along-track coordinate of the landmark s_i and the train along-track position (here identified with the camera center) s: $X_i' = s_i - s$. The coordinate Y' can be equalized to the height of the landmark anchor point: $Y' = h_i$. The third coordinate Z' is, as per problem definition, constant: $Z' = d$, and we will disregard it in the design of the observation model. This can be given as

$$\begin{pmatrix} X_i' \\ Y_i' \end{pmatrix} = \begin{pmatrix} s_i' \\ h_i' \end{pmatrix} = \begin{pmatrix} s_i \\ h_i \end{pmatrix} - \begin{pmatrix} s \\ 0 \end{pmatrix} \tag{3-8}$$

Furthermore, during the time the same sign stays within the camera field of view, an estimation of the train speed $\dot{s}'$ in the camera frame becomes available. This can be directly modeled as an observation of the train actual speed, so that the observation model modifies as

$$\begin{pmatrix} s_i' \\ h_i' \\ \dot{s}' \end{pmatrix} = \begin{pmatrix} s_i \\ h_i \\ \dot{s} \end{pmatrix} - \begin{pmatrix} s \\ 0 \\ 0 \end{pmatrix} \tag{3-9}$$

Re-writing the above model in terms of state vector gives

$$\mathbf{z}_{i,t} = \begin{pmatrix} s_i' \\ h_i' \\ \dot{s}' \end{pmatrix} = \underbrace{\begin{bmatrix} -1 & 0 & 0 & \mathbf{0} & 1 & 0 & \mathbf{0} \\ 0 & 0 & 0 & \mathbf{0} & 0 & 1 & \mathbf{0} \\ 0 & 1 & 0 & \mathbf{0} & 0 & 0 & \mathbf{0} \end{bmatrix}}_{\mathbf{C}_t} \mathbf{x}_t \tag{3-10}$$

If the landmark just entered the camera field-of-view, no estimation of the speed is available, and the observation model reduces to

$$\mathbf{z}_{i,t} = \begin{pmatrix} s_i' \\ h_i \end{pmatrix} = \underbrace{\begin{bmatrix} -1 & 0 & 0 & \mathbf{0} & 1 & 0 & \mathbf{0} \\ 0 & 0 & 0 & \mathbf{0} & 0 & 1 & \mathbf{0} \end{bmatrix}}_{\mathbf{C}_t} \mathbf{x}_t \tag{3-11}$$

c) The initial (mean) robot's state vector can be written as (no landmarks are stored at $t = 0$)

$$\boldsymbol{\mu}_0 = \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix} \tag{3-12}$$

Its covariance, as per problem data, is

$$\boldsymbol{\Sigma}_0 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{3-13}$$

The predicted (mean) value of the state vector is (with $\Delta t = 1$)

$$\bar{\boldsymbol{\mu}}_1 = \mathbf{A}_1 \boldsymbol{\mu}_0 = \begin{bmatrix} 1 & 1 & \frac{1}{2} \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix} = \begin{pmatrix} 1.5 \\ 3 \\ 3 \end{pmatrix} \tag{3-14}$$

The covariance matrix of the predicted state is

$$\begin{aligned} \bar{\boldsymbol{\Sigma}}_1 &= \mathbf{A}_1 \boldsymbol{\Sigma}_0 \mathbf{A}_1^{\mathrm{T}} + \mathbf{R}_1 \\ &= \mathbf{R}_1 \\ &= \sigma_p^2 \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \end{aligned} \tag{3-15}$$

when using (3-5), or

$$\bar{\boldsymbol{\Sigma}}_1 = \sigma_p^2 \mathbf{I}_3 \tag{3-16}$$

when using (3-6).
Since no landmarks are observed, the updated state coincides with the predicted state:

$$\boldsymbol{\mu}_1 = \bar{\boldsymbol{\mu}}_1 = \begin{pmatrix} 1.5 \\ 3 \\ 3 \end{pmatrix} \tag{3-17}$$

and

$$\boldsymbol{\Sigma}_1 = \bar{\boldsymbol{\Sigma}}_1 \tag{3-18}$$

d) The landmark coordinates in the camera frame are $(X', Y', Z') = (2, 0, d)$ m. Following the above reasonings, the landmark position in the global frame becomes

$$\mathbf{m}_1 = \begin{pmatrix} s_{i=1} \\ h_{i=1} \end{pmatrix} = \begin{pmatrix} s'_{i=1} + s_1 \\ h_{i=1} \end{pmatrix} = \begin{pmatrix} 2 + 1.5 \\ 0 \end{pmatrix} = \begin{pmatrix} 3.5 \\ 0 \end{pmatrix} \tag{3-19}$$

The augmented state vector reads then

$$\mathbf{x}_{1,a} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{m}_1 \end{pmatrix} \quad \text{with} \quad \boldsymbol{\mu}_{1,a} = \begin{pmatrix} 1.5 \\ 3 \\ 3 \\ 3.5 \\ 0 \end{pmatrix} \tag{3-20}$$

In order to build the augmented covariance matrix $\boldsymbol{\Sigma}_{1,a}$ we need to propagate the noise on both $\mathbf{x}_1$ and the observed coordinates $(s_{i=1}, h_{i=1})^{\mathrm{T}}$. The easiest way to do this is to write the linear system used to build the augmented state vector $\mathbf{x}_{1,a}$:

$$\boldsymbol{\mu}_{1,a} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \mathbf{m}_1 \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\mu}_1 + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} s'_{i=1} \\ h_{i=1} \end{pmatrix} \tag{3-21}$$

From which, using (3-15),

$$
\begin{aligned}
\boldsymbol{\Sigma}_{1,a} &= \begin{bmatrix}1&0&0\\0&1&0\\0&0&1\\1&0&0\\0&0&0\end{bmatrix} \boldsymbol{\Sigma}_1 \begin{bmatrix}1&0&0\\0&1&0\\0&0&1\\1&0&0\\0&0&0\end{bmatrix}^{\mathrm{T}} + \begin{bmatrix}0&0\\0&0\\0&0\\1&0\\0&1\end{bmatrix} \begin{bmatrix}\sigma_{m_1}^2 & 0\\ 0 & \sigma_{m_1}^2\end{bmatrix} \begin{bmatrix}0&0\\0&0\\0&0\\1&0\\0&1\end{bmatrix}^{\mathrm{T}} \\
&= \begin{bmatrix} \boldsymbol{\Sigma}_1 & \sigma_p^2 \begin{bmatrix}1&0\\0&0\\0&0\end{bmatrix} \\ \sigma_p^2 \begin{bmatrix}1&0&0\\0&0&0\end{bmatrix} & \begin{bmatrix}\sigma_{m_1}^2+\sigma_p^2 & 0\\ 0 & \sigma_{m_1}^2\end{bmatrix}\end{bmatrix}
\end{aligned} \tag{3-22}
$$

The result is expected: the subset of the covariance matrix related to the location, speed and acceleration is unchanged; the covariance of the estimated landmark position in the map increases due to the uncertainty on the train updated position, but only on the along-track coordinate; a correlation is present between the location of the train and the first coordinate $s_{i=1}$ of the landmark due to the linear relationship (3-19).