Problem set 1

1. Line Detection

1.1 Explain the problem of using the slope and y-intercept as line parameters when using the Hough transform.

1.2 When using the polar representation of lines, what does the vote of each point in the image look like in the parameter plane?

1.3 Given the point $(1, 1)$, find the vote in parameter space it will cast for angle $\theta = 0$.

1.4 Explain how lines are detected by checking the parameter plane.

1.5 Explain the trade-off regarding bin size in the parameter plane.

1.6 Describe how voting in the parameter plane can be improved if the normal at each voting point is known.

1.7 When using the Hough transform for circles, explain what should be the number of dimensions of the parameter space.

1.8 Detecting "Nearly Horizontal" Lines

- Assume an $m \times n$ image.
- Parameters of a detected line are (θ, d) where $\theta \in [45^\circ, 135^\circ]$.
- Derive the equation to compute pixel coordinates (x, y) by scanning $x \in [0, n]$ and computing:

$$y = -\frac{\cos(\theta)}{\sin(\theta)}x + \frac{d}{\sin(\theta)}$$

- Explain why x is scanned, and y is computed.

1.9 Detecting "Nearly Vertical" Lines

- Assume an $m \times n$ image.
- Parameters of a detected line are (θ, d) where $\theta \in [-45^\circ, 45^\circ]$.
- Derive the equation to compute pixel coordinates (x, y) by scanning $y \in [0, m]$ and computing:

$$x = -\frac{\sin(\theta)}{\cos(\theta)}y + \frac{d}{\cos(\theta)}$$

- Explain why y is scanned, and x is computed.

2. Model Fitting

2.1 Explain the disadvantage of using the equation $y = a \cdot x + b$ for line fitting. What kind of lines cannot be fitted accurately using this equation?

2.2 Find Coefficients for a Line

$$d = \frac{|c|}{\sqrt{a^2+b^2}}$$

- A line with a slope of 45° passes at a distance of 10 from the origin.
- Write the coefficients a, b, c in the explicit line equation $a \cdot x + b \cdot y + c = 0$.
- Verify the answer by drawing the line and checking points on it.

2.3 Line Defined by Two Points

- Given points $(10, 10)$ and $(20, 20)$, write the implicit line equation they define.
- Write the normalized normal to this line.

Slope $\Rightarrow m = \frac{x_1 - x_2}{y_2 - y_1} \Rightarrow \frac{20-10}{20-10} = 1$

$y - y_1 = m(x - x_1)$

$y - 10 = 1(x - 10)$

$y - x = 0$

Normalize → Divided by their magnitude

Normal vector = $(1, -1)$

Magnitude = $\sqrt{1^2 + (-1)^2} = \sqrt{2}$

Normalize normal = $\left(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right)$

2.4 Vector Representation of a Line

- Given a line with normal $(1, 2)$ and a distance of 2 from the origin, write the vector l representing the line in the implicit equation $l^T x = 0$.

$d = \frac{c}{|n|} \Rightarrow 2 = \frac{c}{\sqrt{1+2^2}} = 2\sqrt{5} = c$

$x + 2y + 2\sqrt{5} = 0$

$l = (1, 2, 2\sqrt{5})$

2.5 Find y-Coordinate on a Line $x + 2y + 3 = 0 \rightarrow 2 + 2y + 3 = 0 \Rightarrow y = \frac{-5}{2}$

- Given line coefficients $(1, 2, 3)$, find the y-coordinate where $x = 2$.

2.6 Line Fitting with Implicit Equation

- Explain how to fit a line using the implicit equation.
- Write the equation to solve for unknown line parameters.

2.7 Homogeneous Coordinates for Line Fitting

- Given points $\{(0, 1), (1, 3), (2, 6)\}$, write the 3×3 matrix to find line parameters in homogeneous coordinates.

2.8 Algebraic vs. Geometric Distance

- Explain the difference between algebraic and geometric distance.

2.9 Geometric Distance Approximation

- Explain how the geometric distance of a point p from an implicit curve $f(p) = 0$ is measured exactly and approximated.
- Discuss why the approximation is used.

2.10 Algebraic Distance

- Given $f(p) = 1$ and the gradient of f at the closest point is 2, compute the algebraic distance of p from the curve f.

2.11 Approximated Geometric Distance

- Given $f(p) = 1$ and the gradient of f at p is 2, find the approximated geometric distance of p from f.

2.12 Active Contours

- Write the objective function for active contours and explain its components.

2.13 Continuity and Curvature Energy

- Given three points on a contour: $p_1 = (1, 2)$, $p_2 = (2, 3)$, $p_3 = (3, 4)$, find the continuity and curvature energy at p_2.

2.14 Active Contour Tight Fitting

- For a point with high curvature (e.g., a corner), write the coefficient β to ensure tight fitting of the contour.

3. Robust Estimation

3.1 Outliers in Model Fitting

- Explain what outliers are and their fundamental problem in model fitting.

3.2 Robust Estimation Objective

- Write the objective function for robust estimation.
- Explain how it differs from the standard least squares objective.

3.3 Geman-McClure Estimator

- Write the Geman-McClure function for robust estimation and its advantages.
- Explain how the bandwidth parameter σ is iteratively adjusted.

3.4 Example: Geman-McClure Estimator

- Compute the Geman-McClure estimator for $x = 1$ and $\sigma = 1$.

3.5 RANSAC Algorithm

- Explain the principle of the RANSAC algorithm.
- Discuss whether a large or small number of points should be drawn at each attempt.

3.6 RANSAC Parameters and Trials

- Explain the parameters of the RANSAC algorithm.
- Write the formula to estimate the number of trials.

3.7 RANSAC Example

- Compute the number of experiments needed in RANSAC given:
    - Desired probability $p = 0.99$ of at least one experiment without outliers.
    - Probability $w = 0.9$ that a point is an inlier.

Problem set 2

1. Convolution Layers

1.1 Let I be a 4×4 RGB image where:

- The R channel is all 1's.
- The G channel is all 2's.
- The B channel has values:
  - 1 in the first row,
  - 2 in the second row,
  - 3 in the third row,
  - 4 in the fourth row. Compute the convolution of this image with a 3×3 filter having all ones without zero padding.

1.2 Repeat the previous question with zero padding.

1.3 Repeat the previous question using dilated (atrous) convolution with a dilation rate of 2.

1.4 Explain the template matching interpretation of convolution.

1.5 Explain how multiple scale analysis can be achieved with a fixed window size (using a pyramid).

1.6 Explain how to compensate for spatial resolution decrease using depth (number of channels) and the purpose of doing so.

1.7 Given a $128 \times 128 \times 32$ tensor and 16 convolution filters of size $3 \times 3 \times 32$, what will be the size of the resulting tensor when convolving without zero padding?

1.8 Repeat the previous question when using a stride of 2.

1.9 Explain how the number of channels can be reduced using a 1×1 convolution.

1.10 Explain the interpretation of convolution layers and the difference between early and deeper convolution layers.

1.11 Let I be the image from question 1.1. Write the result obtained using max pooling with a 2×2 filter and a stride of 2.

1.12 Explain the purpose of pooling.

1.13 Explain the purpose of data augmentation and when it is most useful.

2. CNNs

2.1 Explain the purpose of transfer learning and when it is most useful.

2.2 Explain the need for freezing the coefficients of the pre-trained network.

2.3 Explain how the coefficients of a pre-trained network can be fine-tuned.

2.4 Explain the purpose of inception blocks. Describe the solution employed in GoogleNet to address vanishing gradients in a deep network.

2.5 Explain the advantage of residual blocks. Include an explanation of how residual blocks assist with vanishing gradients.

2.6 Explain how DenseNet is constructed. Describe how DenseNet controls complexity.

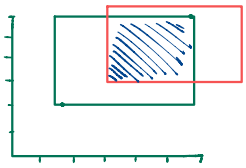
2.7 Given an image with three channels:

- The first channel has all 1's,
- The second channel has all 2's,
- The third channel has all 3's. Compute the result of a convolution with a 3×3 filter where:
- The first layer has all 1's,
- The second layer has all 2's,
- The third layer has all 3's. Repeat the computation using depth-wise separable convolution.

2.8 Describe how MobileNets make computations faster.

3. Object Detection

3.1 Explain the two tasks that need to be achieved in object detection.



3.2 Given:

- A detected object bounding box defined by the corners $(2, 2)$ and $(6, 6)$,
- A ground-truth bounding box defined by the corners $(3, 3)$ and $(7, 7)$, Compute the IoU similarity metric and Jaccard distance.

3.3 Given a dataset of images where some contain faces, you train an object detection algorithm that outputs a detection box and the probability of a face in the box. Using a 0.5 IoU threshold, compute $AP_{0.5}$ (average precision) for the following precision-recall pairs:
$(1, 0), (1, 0.2), (0.6, 0.4), (0.6, 0.6), (0, 0.8), (0, 1)$.

3.4 Explain why detection box coordinates are normalized to be between 0 and 1.

3.5 Explain the different terms in the loss function needed for an object detection network.

3.6 Given a grid cell object detection with a 3×3 grid and 10 detection boxes at each cell, write the size of the output tensor for the algorithm.

3.7 Explain the difference between single-shot and two-shot approaches.

3.8 Describe the different terms in the loss function of the YOLO object detector.

3.9 Explain how ROI-pooling is done and its purpose.

3.10 Explain non-maximum suppression in the context of object detection and the need for it.

3.11 Explain the 3 loss terms in Mask R-CNN.

4. Semantic Segmentation

4.1 Explain the difference between semantic segmentation and instance segmentation.

4.2 Given a 5×5 image and a 3×3 filter, compute the size of the matrix that can multiply the vectorized image (1D) to produce the convolution results.

4.3 Compute the size of the transpose convolution matrix from the previous question.

4.4 Explain the need for skip connections in U-Net and how information is propagated along skip connections.

4.5 Explain the DeepLab network architecture.

4.6 Explain the metric used for evaluating semantic segmentation results.

Problem Set 3

1. Camera Calibration 1

1.1 Given the projection equation $p = MP$, explain the problems of forward projection, calibration, and reconstruction. Which problem is the easiest? Which is the most difficult?

1.2 Explain the necessary input for camera calibration.

1.3 Explain the steps in the non-coplanar calibration algorithm.

1.4 Given a known projection matrix with rows $(1, 2, 3, 4)$, $(1, 0, 3, 4)$, $(1, 1, 1, 1)$ and a world point $P_i = (1, 2, 3)$, compute the 2D image coordinates of P_i after projecting it. Convert P_i to homogeneous coordinates before projecting.

1.5 Given the corresponding world-image points: $(1, 2, 3) \leftrightarrow (100, 200)$, write the first two lines of the matrix that needs to be formed for solving the unknown projection matrix M.

1.6 What is the minimal number of points necessary to find a unique solution for M? How is the solution obtained?

1.7 Explain the principle used to extract the unknown camera parameters from the projection matrix M.

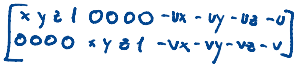
1.8 Explain how to compute the quality of the projection matrix M estimate.

1.9 Explain the principle of planar camera calibration. How does planar calibration differ from non-coplanar calibration?

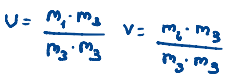
1.10 Explain the difference between the homography (2D projective map) H and the projection matrix M. What assumption ensures we are dealing with homography matrices?

2. Camera Calibration 2

2.1 Given a pair of corresponding image and world points $(1, 2)$ and $(3, 4, 5)$ respectively, write the first two rows of the matrix that must be constructed to solve for the unknown coefficients of the projection matrix.

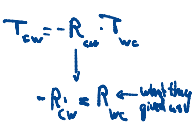


2.2 Given the estimated projection matrix M with rows $(1, 2, 3, 4)$, $(2, 3, 4, 5)$, $(3, 4, 5, 6)$, find the camera parameters u_0 and v_0 (coordinates of the principal point) of the camera.



2.3 Given a pair of corresponding image and world points $(1, 2)$ and $(3, 4, 5)$ respectively, and an estimated projection matrix M with rows $(1, 2, 3, 4)$, $(2, 3, 4, 5)$, $(3, 4, 5, 6)$, find the projection error of the matrix M.

2.4 After performing calibration, let the obtained rotation of the world with respect to the camera be given by $I + Q$, where I is a 4×4 identity matrix and Q is a 4×4 matrix having 5 in its $(0, 0)$ element and zeros elsewhere. Let the obtained translation of the world with respect to the camera be given by the vector $(1, 2, 3)$. Compute the rotation and translation of the camera with respect to the



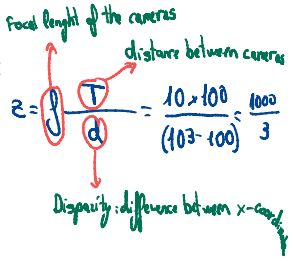
world.

2.5 Given a pair of corresponding image and world points $(1, 2)$ and $(3, 4, 0)$ respectively, used for planar calibration, write the first two rows of the matrix that must be constructed to solve for the unknown coefficients of the homography matrix.

3. Multiple View Geometry 1

3.1 Explain the difference between sparse and dense stereo matching. What are the advantages/disadvantages of each approach?

3.2 Explain how normalized cross-correlation (NCC) and sum of square distances (SSD) can be used for point matching. What is the risk of allowing the search space to be the entire image? How can the search space be reduced to a line?



3.3 Given an axis-aligned stereo pair with corresponding points $(100, 200)$ and $(103, 200)$ in the left and right images, compute the depth (z-coordinate) of the 3D point that produced this projection. Assume the focal length of both cameras is 10, the baseline is 100, and the system uses camera coordinates.

3.4 Explain the ambiguity problem in stereo matching.

3.5 Given R_l, T_l (rotation and translation of the left camera with respect to the world) and R_r, T_r (rotation and translation of the right camera with respect to the world), write the expression for the rotation and translation of the right camera with respect to the left camera.

4. Multiple View Geometry 2

4.1 Given an axis-aligned stereo system with a focal length of $10\,\mathrm{mm}$ and a baseline of $20\,\mathrm{mm}$, compute the depth of a point with a disparity of $30\,\mathrm{mm}$.

4.2 Let $A = (1, 2, 3)$ and $B = (2, 3, 4)$. Write the matrix that, when multiplied by B, results in the cross product $A \times B$.

4.3 Let F be a fundamental matrix with rows $(1, 2, 3), (2, 3, 4), (3, 4, 5)$. Let $(1, 2)$ and $(2, 3)$ be corresponding left and right points. Compute the value of $p_r^T F p_l$.

4.4 Given corresponding left and right points $(1, 2)$ and $(2, 3)$ respectively, write the respective row in the matrix that must be formed to solve for the fundamental matrix.

Problem Set 4

1. Stereo

1.1 Explain the difference between sparse and dense stereo matching. What are the advantages and disadvantages of each approach?

1.2 Explain how normalized cross-correlation (NCC) and sum of square distances (SSD) can be used for point matching. What is the risk of allowing the search space to be the entire image? How can the search space be reduced to a line?

1.3 Given an axis-aligned stereo pair with corresponding points $(100, 200)$ and $(103, 200)$ in the left and right images respectively, compute the depth (z-coordinate) of the 3D point that produced this projection. Assume that the focal length of both cameras is 10, the baseline is 100, and the system uses camera coordinates.

1.4 Explain the ambiguity problem in stereo matching.

1.5 Given R_l, T_l and R_r, T_r (rotation and translation of the left and right cameras with respect to the world), write the expression for the rotation and translation of the right camera with respect to the left camera.

2. Epipolar Geometry

2.1 Explain what the epipole is. How are epipolar lines "formed" in the image?

2.2 Write the expression of the essential matrix E. Given a set of corresponding points p_l, p_r (in camera coordinates), write the epipolar constraint equation using the essential matrix E.

$E = R^T [T]_\times$

$p_r^T E p_l = 0$

2.3 Write the expression of the fundamental matrix F. Given a set of corresponding points p_l, p_r (in image coordinates), write the epipolar constraint equation using the fundamental matrix F.

$F = K_r^{-T} E K_l^{-1}$

$p_r F p_l = 0$

2.4 What is the rank of the essential and fundamental matrices? Why?

2.5 Given the point p_l in the left image and the fundamental matrix F, what is the corresponding right epipolar line?

2.6 Given the point p_r in the right image and the fundamental matrix F, what is the corresponding left epipolar line?

2.7 Explain what a weak calibration of a stereo pair is.

2.8 Given corresponding points $(100, 200)$ and $(50, 100)$ in the left and right images respectively, write the first line in the matrix that has to be formed to solve for the unknown fundamental matrix using the 8-point algorithm. Do not normalize the points.

2.9 Explain how to normalize the points in the 8-point algorithm for estimating the fundamental matrix. Why is this necessary? Explain how the fundamental matrix of the original points can be recovered from the fundamental matrix of the normalized points.

2.10 Explain how the epipoles can be recovered from the fundamental matrix.

3. Reconstruction

3.1 Explain how a stereo pair can be rectified. What happens after the pair is rectified?

3.2 Explain the different approaches for reconstruction depending on what is known about the cameras.

3.3 Let the right image be rotated by R and translated by T with respect to the left image. Let p_l and p_r be corresponding points in the left and right images. Write the matrix that must be formed to solve for the coefficients (a, b, c) of the corresponding triangulated 3D point.

3.4 Using the coefficients (a, b, c) of the triangulated point, write the formula to compute the 3D point.

3.5 Explain why there is an unknown scale in Euclidean reconstruction. How can this unknown scale be removed?

3.6 Given the estimated essential matrix E, explain how this matrix can be normalized to have a baseline of 1.

3.7 Explain how the unknown signs of rotation and translation can be determined when using Euclidean reconstruction.

Problem Set 5

1. Motion

1.1 Explain the difference between 3D motion vectors, 2D projected motion vectors, and observed 2D motion vectors (optical flow). Is it possible that motion in 3D will not produce optical flow vectors?

1.2 What will be the projected motion field in a video taken by a car driving on a straight road and looking to the side (assume that objects in the scene do not move)? Where will projected motion vectors be larger?

1.3 What will be the projected motion field in a video taken by an airplane aiming to land at a fixed point while looking forward (assume that objects in the scene do not move)? Where will projected motion vectors be larger?

1.4 Write the fundamental motion projection equation relating 3D motion vectors V to 2D projected motion vectors v, the focal length f, and the position of the object point P. Assume the z-coordinate of the object point is z, and that the z-component of the 3D motion vector is V_z. What is the z-component of the projected motion vector v, according to this equation?

1.5 Assuming 3D motion with translational velocity τ and rotational velocity ω, write the equation for the projected translational and rotational motion.

1.6 Explain what the projected motion field looks like in the case of pure translational motion. Make sure to distinguish between the cases where there is or there is no translational component in z.

1.7 Write the equations for the instantaneous epipole.

1.8 Explain when motion parallax is created, and write the relative motion field equations.

2. Optical Flow

2.1 Write the optical flow constraint equation (OFCE). What is the basic assumption that is used to derive this equation?

2.2 Explain the aperture problem. What part of the motion can we hope to recover based on a single point?

2.3 Explain how the aperture problem is addressed in block-based optical flow estimation methods.

2.4 Write the objective function of block-based optical flow estimation in a patch. Write the system of equations that must be solved in order to find the optical flow in the patch. What is the purpose of weighted block methods? How do the weights modify the solution?

2.5 Explain the advantage of an affine motion model. Write the objective function for the affine model, then write the solution. How are the motion vectors in a patch recovered once the affine motion parameters are recovered?