Practice Exam Questions

For your exercise only. Do not turn in for grading
Last modified on Wednesday 13th February, 2019

Overview

Question Why is computer vision considered a hard problem as "seeing" seems to come naturally to us without effort? Name at least three factors that make computer vision harder than it appears to be.

Question Name at least three academic disciplines that overlap somewhat with computer vision. State why researchers in these related fields and computer vision might be be working on some common problems and what these problems may be.

Question Name at least three topics that are often covered in an undergraduate course in computer vision (or topics that you are interested in learning).

Linear Filter

Question For a 1D image which is all zero except for a single-pixel non-zero spike in the middle, what happen if you pass a 3-wide box filter over it once? twice? or as many times as possible? What happens if you pass a 3-wide median filter over it once? twice? or as many times as you want? How do you compare the performance of an average filter with a median filter based on this result?

Question For a 1D image which is all zero except for a 3-pixel-wide non-zero spike in the middle, what happen if you pass a 3-wide box filter over it once? twice? or as many times as possible? What happens if you pass a 3-wide median filter over it once? twice? or as many times as you want? How do you compare the performance of an average filter with a median filter based on this result?

Question If you have a 2D image which is all zero except for a nonzero 1x1 spike in the middle, what happen if you pass a 3x3 median filter over it once? twice? or as many times as possible ? Do the 1D and 2D median filters do different things to the 1D and 2D 1x1 spikes?

Question If you have a 2D image which is all zero except for a nonzero 3x3 spike in the middle, what happen if you pass a 3x3 median filter over it once? twice? or as many times as possible ? Do the 1D and 2D median filters do different things to the 1D and 2D 3x3 spikes? What happen if you have an $n \times n$ spike where $n > 3$?

Question If you have a completely random image, in the sense that each pixel's intensity is an independent, normally distributed random variable with mean zero and standard deviation of 1 (so the pixel values can be negative). Estimate the percentage of pixels that will give rise to a response that is larger than 3 by passing a discrete 1D derivative operator of the form $|I_{i+1,j} - I_{i,j}|$.

Edge Detection

Question You are given some "idealized" mathematical functions called Ramp, Step, and Impulse as shown in Fig. 1(a). Can you "massage" these basic functions to build a general ramp function, as shown in Fig. 1(b)?

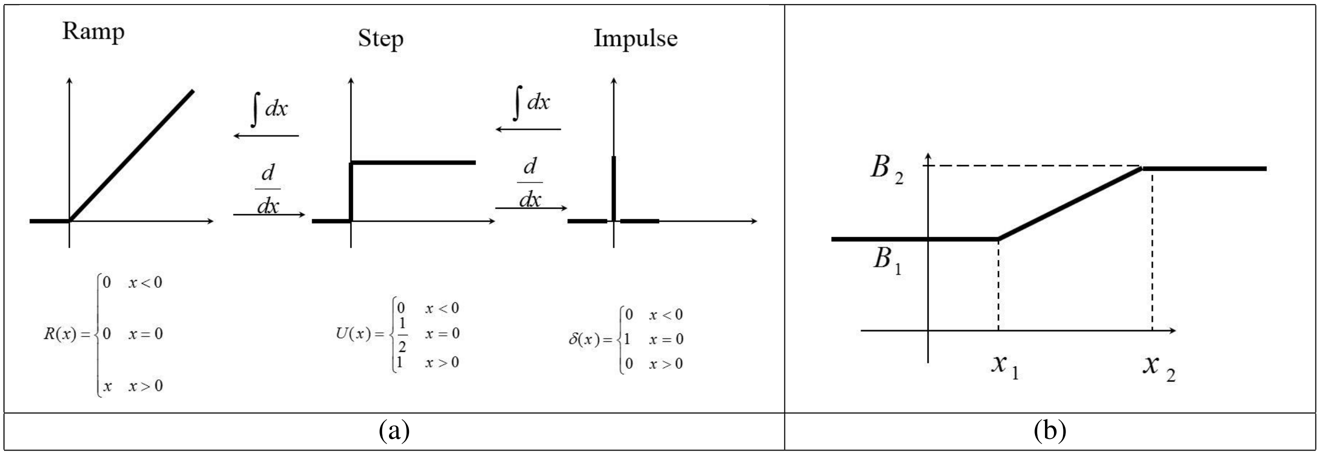

Ramp
Step
Impulse
∫dx
d/dx
∫dx
d/dx
R(x) = {0 x < 0; 0 x = 0; x x > 0}
U(x) = {0 x < 0; 1/2 x = 0; 1 x > 0}
δ(x) = {0 x < 0; 1 x = 0; 0 x > 0}
B2
B1
x1
x2
(a)
(b)

Figure 1: (a) Idealized image profiles ramp, step, and impulse, and (b) a general ramp function to approximate a 1D edge.

Question Continue from the previous question, what kind of response will you get by passing the general ramp edge in Fig. 1(b) through a 1D derivative operator once? and twice? Explain based on your answer why there are edge detectors using either the first or the second derivative and what should these operators look for in the filter response.

Question Continue from the previous question, state at least three reasons why such simple mathematical models and edge operators can fail for real-world images.

Edge Linking

Question Hough transform is often used in the vehicular safety application as an important component of a lane departure warning system. A typical LDW system is illustrated graphically in Fig. 2, where a front-mounted camera constantly scans the road in front to locate lane marking (solid or dotted white or yellow lines shown in Fig. 2). If the car crosses a detected lane without the driver activates the turn signal, the LDW system gives an audio alert.
How can Hough transform be used in such a scenario? Will the "vanilla" Hough transform described in lecture work well? How can you improve the performance to reduce false positives?

Question If you are asked to design a Hough transform method to detect regular hexagons (regular means that all edges are of the same length and all interior angles are 120^o), how would you do that? Be creative with your solution. Any scheme that uses some kind of "voting" principle like Hough transform is acceptable.

Question Yet another way to link edge points is to fit the edge points with an analytic equation. E.g., you know the points should lie on a circle; you just don't know where and how large the circle is.
Consider the problem of fitting edge points to a general conic section, given by $ax^2 + bxy + cy^2 + dx + ey + f = 0$. If you are given a number of edge points $(x_i, y_i), i = 1 \cdots, n$ from your edge detector, you would like to "best fit" these points by the conic model. How should you do it? More specifically, how do you solve for the unknown coefficients $(a, \cdots, f)$ in terms of the given data points $(x_i, y_i), i = 1 \cdots, n$? What is the minimum number of edge points you need to solve for the coefficient?

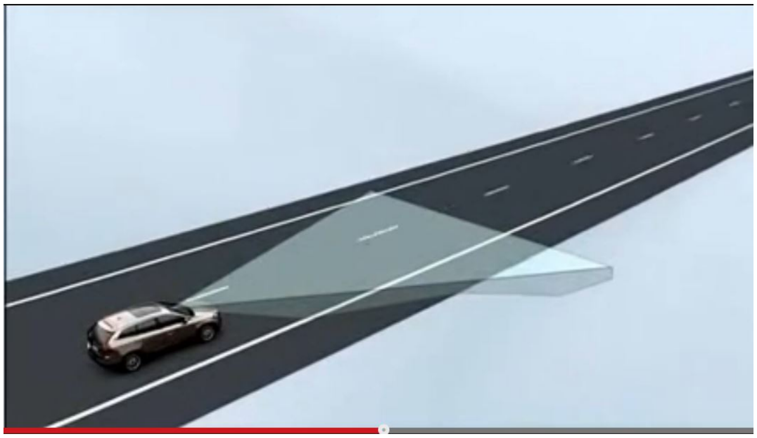


Figure 2: A graphic illustration of a lane departure warning system.

Question Continue from the previous question: mathematically, it might be easier to consider a similar problem of fitting an explicit function $y = f(x)$ instead of an implicit function $f(x, y) = 0$ as in the previous question. If you know that the the edge points suppose to lie on an $n - th$ order polynomial, or $y = f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$, how would you solve for the unknown coefficients? How many edge points do you need to solve the equation?

Question One common complication of such a curve fitting problem in the two previous questions is that you do not really know which points actually belong to the curve. In a busy image, you can detect a large number of edge points and only a subset of these edge points actually belong to the curve (e.g., a circle). How can you find what edge points should be included/excluded from the fitting process?

Question Try practicing your calculus skill (particularly in using the chain rule for derivatives) to derive the force equation of a snake deformable template that allows for rotation.

2D Analysis

Question Prove mathematically that the valley finding intuition for histogram-based foreground and background segmentation is correct. That is, if the intensity distribution of foreground and background pixels can both be modeled a Gaussian distribution where the separation of the means is "large enough" to see a valley in between the two peaks, then segmenting the scene on the lowest valley intensity value produces the smallest classification error.

Question Does the valley finding intuition always work for segmenting Gaussian distributed foreground and background pixels? If not, provide a counterexample.

Features and Stitching

Question Similarity of two feature vectors is often defined as the Euclidean distance between the two feature vectors. The smaller the Euclidean dstance, the more similar the two feature vectors are. In that sense, all elements of the feature vectors are considered equally important and reliable. What happens if elements in the feature vectors are not equally important or reliable? How would you change the distance formula to account for that?

Question Yet another problem is the spread of the elements in a feature vector. Some elements may have very large range while others may have very small ranges. Will different dynamic ranges affect how you use the feature vectors?

Question You can compute distance between feature vectors another way by comparing their directions. Two feature vectors pointing in the same direction are considered more similar than two feature vectors pointing in the opposite directions. Are these two similarity definitions (distance and direction) equivalent?

Question How would you solve the 360^o panorama stitching problem where the residual error in alignment can come back to bite you when you try to connect the head back to the head?

Recognition using NN

Question In some sense a CNN (convolutional NN) is a special case of an FNN (fully connected NN). In an FNN, every input neuron is connected to every output neuron and every connection is allowed a different weight. In a CNN, the connection pattern is usually more localized (e.g., $3 \times 3, 7 \times 7$). Furthermore the same pattern is used everywhere in the image.

If CNN is a special case of FNN, why don't we use FNN everywhere? Can it recover more general topology and connection patterns through training?

Question For the network design used in program #4, compute the number of variables (both w - neuron-neuron connection and b neuron biases) in each layer.

Question To find the right network parameters, people by-and-large use gradient descent search: starting from an initial guess of the parameters, go in the negative gradient direction of the loss function, and the step iterates until the gradient becomes small enough. Or $\omega_{new} = \omega_{old} - \eta \nabla L$, where L is the loss function, ω represents the vector of system parameters and η is the learning rate.

Does "the gradient becomes small enough" condition guarantees a global minimum? If not, does it guarantee at least a local minimum?

You probably suspect that given gradient descent is the workhorse of NN optimization and the vanilla equation above looks suspiciously fragile (also recall your experience from program #2 in using gradient descent with explicit Euler method), there must be many improvements of the basic scheme. Do a web search and see if you can find more sophisticated and robust versions of such a search technique.

Question Recall that for a classification problem, the final loss function is defined as the cross entropy of 2 probability distributions: one is the ground-truth label $\hat{y}$, which is usually a one-hot vector of size n, n being the number of classes. The other is the network prediction $f(x)$, where x is the input feature. Again, $f(x)$ is a vector of length n, which is usually the output of a softmax function at the end of the network.

Why we cannot define the loss function as the L-2 norm of the two vectors $||f(x) - \hat{y}||$? You might want to do a web search for the answer as it might not be obvious why L-2 norm won't work.

Optical Flow

Question Suppose that along an image scan line, the intensity function can be approximated by $I(x) = x + U(x-3)x^2 - U(x-6)x^3$, where U is the Step function shown in Fig. 1(a). Further assume that there is

no y movement. The optical-flow equation can then be simplified from $I_x u + I_y v + I_t = 0$ to $I_x u + I_t = 0$, i.e., a 1D case.
Assume that the u velocity of the object can change anywhere from, say, 0 pixel/sec to -10 pixel/sec (the object moves to the left). You approximate I_x by using the forward differencing formula $I_x \approx I_{i+1,j} - I_{i,j}$. Calculate the error in the optical-flow equation $|I_x u + I_t|$ as a function of the velocity of the object and the given intensity model for $x \geq 0$. What does the calculation tell you?

Question Now assume that you "squeeze" the object length-wise by a factor of 10, that is x' = x/10, how will the intensity profile given above change?

If you still use the same formula $I_{x'} \approx I_{i+1,j} - I_{i,j}$ to approximate $I_{x'}$, how will your error in the optical-flow equation change? How does "squeeze" an object length-wise correspond to the operations in building a Gaussian pyramid in a hierarchical optical-flow computation?

Question Lucas-Kanade method is considered one of the most robust optical-flow methods, as it solves a "a flow vector per a local region" instead of a "a flow vector for every pixel" problem. Prove mathematically that even with Lucas-Kanade, it is not possible to recover the flow vectors correctly if the local gradient fields are somewhat "de-generate." Define your own de-generation conditions in the proof.

Camera & Projection Geometry

Question What are the main advantage and disadvantage of using a small pin-hole camera?

Question Using a pin-hole (perspective) camera model, show mathematically that parallel lines in 3D project into convergent lines in the image plane. Find the coordinates of the vanishing point as a function of the line direction and show that the vanishing point is not a function of line position. Are there any parallel lines in space that stay parallel after perspective projection? If so, what are these lines?

Question Using a parallel projection model, show mathematically that parallel lines in 3D stay parallel after projection.

Question Using an Affine camera model, show that parallel lines stay parallel, but perpendicular lines do not necessarily stay perpendicular.

Question A scene point at coordinates (400,600,1200) is projected in a pinhole camera at image coordinates (24,26), where both are given in millimeters in the cameras reference frame and the image coordinates have their origin at the principal point of the camera. Assuming the aspect ratio of the pixels is 1, what is the focal length of the camera? (Recall, the aspect ratio is defined as the ratio between the width and the height of a pixel.)

Question Continue from the previous question, now assuming the aspect ratio is not 1 and the same scene point projects instead to (24,24), what is the aspect ratio of the pixels in this camera?

Question Show that (1) if the 3D scene is planar or (2) if the scene distance is very large, the projection equation can be simplified into a homography transform.

Question Under what conditions can a general homography transform be further simplified to be an Affine transform?

Question Show that in a viewer centered coordinate system (i.e., x pointing right, y pointing up and z pointing toward the viewer), all horizontal lines, regardless of their height z intersect with the horizon. What is the equation of the horizon?

Question Continue from the previous question, if we fix the coordinate system as before, but tilt the camera downward by an angle of θ, where will all horizontal lines intersect?

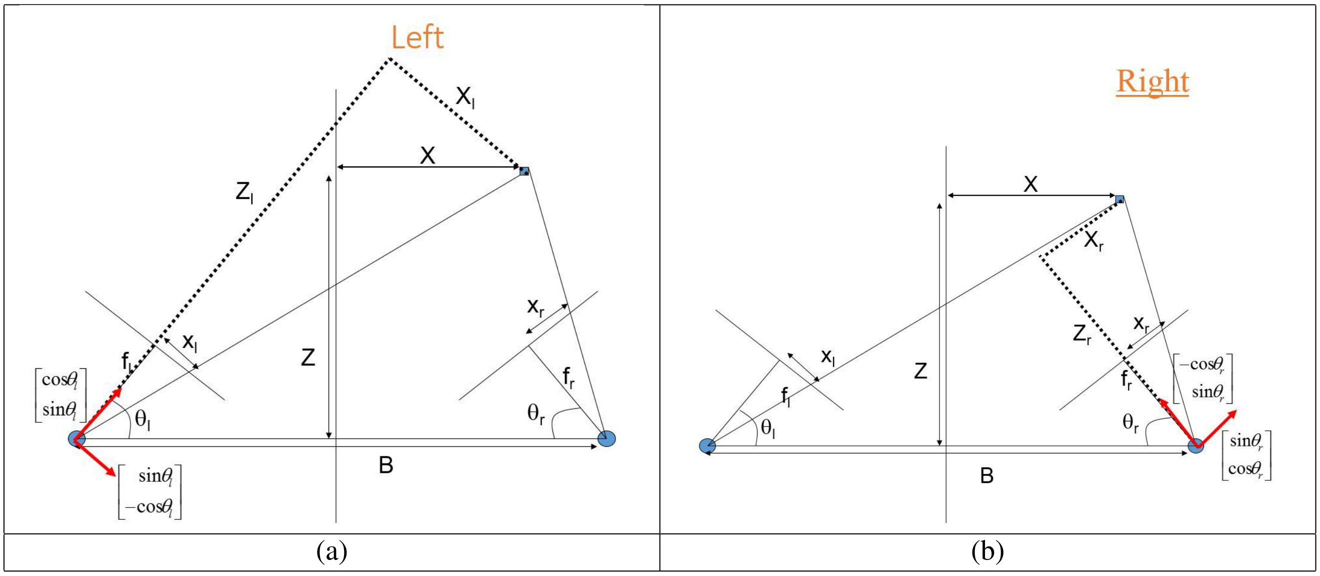
Stereo

Question What is the resolvable distance of a standard stereo configuration? Depth of an object locate at a distance that induces zero disparity cannot be resolved.

Question We know that disparity is inversely proportional to the apparent depth (the further away an object is, the smaller the disparity shift). Hence, the depth resolution is not linear either. Depth resolution at a distance Z in space is defined as the minimum change of Z, or ΔZ, that induces a disparity shift that is different from that at distance Z. Derive the equation for ΔZ for the standard stereo configuration and plot ΔZ as a function of Z. What does the plot tell you?

Question Consider a more general stereo configuration shown in Fig. 3. Instead of two cameras looking straight ahead and in parallel, the optical axes of the two cameras turn toward each other and converge at an angle (this is similar to the "fixation" point of human vision).

Question Give some reasons why a dynamic-programming based stereo matching algorithm cannot really guarantee the optimal solution, even when dynamic-programming is supposed to look at all possible solutions intelligently.


Left
X_l
X
Z_l
x_l
f_l
Z
x_r
f_r
$\begin{bmatrix}\cos\theta_l\\ \sin\theta_l\end{bmatrix}$
θ_l
θ_r
B
$\begin{bmatrix}\sin\theta_l\\ -\cos\theta_l\end{bmatrix}$
Right
X
X_r
x_r
Z_r
x_l
f_l
Z
f_r
$\begin{bmatrix}-\cos\theta_r\\ \sin\theta_r\end{bmatrix}$
θ_l
θ_r
$\begin{bmatrix}\sin\theta_r\\ \cos\theta_r\end{bmatrix}$
B

(a) (b)

Figure 3: General stereo configuration for the (a) left and (b) right cameras.

Derive mathematically an expression for x_l in terms of f_l, θ_l, X_l, Z_l and B.
Derive mathematically an expression for x_r in terms of f_r, θ_r, X_r, Z_r and B.

Question Continue from above, what is the formula for disparity $d = x_l - x_r$ for this general configuration? As a reality check, verify that your disparity equation reduces to the standard parallel-axis stero equation when $\theta_l = \theta_r = 90^o$ and $f_l = f_r = f$.

Structure from Motion

Question Derive mathematically the equations for the intersection of two lines in space. Each line results from the back projection of a 2D feature from an image. The first camera has the camera matrix as an identity matrix and the second camera has the camera matrix defined by R and T — the translation and rotation between the two camera frames.

Question If you move the camera in such a way as shown in Fig. 4, or the two snapshots were taken symmetrically with the two optical axes coplanar and making an angle of 90^o. Both snapshots have a focal length f and the principal point in both images is at (0, 0). Sketch the epipolar geometry, indicating clearly the positions of the epipoles and the epipolar lines.

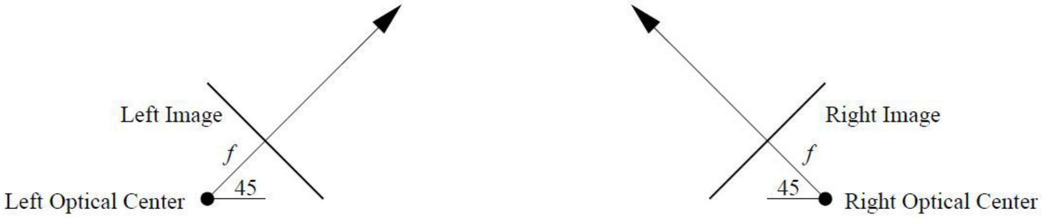


Figure 4: A simple 2-view SfM configuration.

Question Solve algebraically for the coordinates of the epipole in the left image.

Question Given the fundamental matrix, F, relating two images, how can it be used to determine if a point, p = [u v 1], in the left image corresponds to a point, p = [u v 1], in the right image?

Question How can the fundamental matrix be used to solve for the epipoles in the two images?

Question How many point correspondences are required to estimate F using linear techniques?