Final Exam Study Guide

Description

As indicated in class the goal of the exam is to encourage you to review the course material. This study guide is not guaranteed to be comprehensive, **just because some subject is not on the guide doesn't mean that material is not on the exam**. The sample questions are representative of the questions to be asked. (These are actually, perhaps, a little more ambiguous and take longer to answer. The real exam questions are quicker to answer). The slides and the assigned readings in Forsyth and Ponce are considered the material that can be in the final exam.

Questions Digest

The questions/notes below are representative of content that will be covered by the exam.

1. ## LINEAR SYSTEMS

    a. Make sure you understand what makes certain image operations linear and what are some operators we use in, say edge detection, that are not linear.

    b. Describe how you might do edge detection using at least two operations – first a linear one followed by some number of non-linear ones – that would find edges in a slightly noisy image.

    c. What's the difference between Gaussian noise and salt and pepper noise? Why does a linear filter work well to reduce the noise for the Gaussian case but not the other?

    d. How is sharpening done using filtering? And would it matter whether you used convolution or correlation?

    e. What are two ways to compute gradients in an image that has some noise in it?

    f. What can you do during edge detection to account for the fact that some edges vary in contrast along the edge – that is sometimes they are strong and sometimes weak.

2. ## DATA STRUCTURES

    a. A standard Hough transform performs voting for a parametric shape. Why are we doing voting and why does it work?

    b. How can the Hough transform help in identifying lines, circles, and other shapes. You should be able to interpret a Hough Accumulator and determine what shapes are

present along with details of each one (location, orientation, size, etc.)

c.   A friend needs to find the pool balls in an image of a pool table. Would a Hough transform be a good idea? Why/why not? Would RANSAC be better?

3. FREQUENCY

a.   Fourier analysis decomposes images according to a basis set. What is that basis set?

b.   How does the Fourier transform encode the magnitude and phase of sinusoidal component of a signal?

c.   Is the Fourier transform a linear operation?   Why or why not?

d.   Why does convolving an image with a Gaussian **attenuate** the high frequencies?

e.   What is aliasing and when does it happen? Draw a picture that explains it in terms of a comb filter doing the sampling and the effect of that operation in the frequency domain.

f.   What is the relation between a Gaussian pyramid and aliasing? In particular, why can you reduce the size at each step and not lose (hardly) any information?

4. CAMERA MODELS and CALIBRATION

a.   What is the role of an "aperture" in a typical camera? Why would you want a large aperture? Why would you want a small one?

b.   Related: How is depth of field related to aperture size?

c.   Zooming the lens (changing the focal length) is not the same as moving closer with the camera. Why?
OR: Why does a person's nose look so big compared to their face if you take an image closer to them than further away?

d.   Perspective projection: A point in 3D at location <X,Y,Z> in the cameras coordinate system appears where in the image? And, what assumptions about the intrinsics did you just make?

e.   Why do all lines parallel to each other converge to the same point in an image?

f.   How many degrees of freedom are in the extrinsics and intrinsics? What are they?

g. How many 3D points need to be observed to do absolute calibration? Why?

h. Write the perspective projection equation as a 3x1 = [3x4] * [4x1] How many unknowns are in the above equation?

i. One way to solve for the unknowns is to view some points whose 3D position is known and whose 2D position is recorded. How many equations do I get per viewed world point? If I have, say, 10 points, how would I solve for those unknowns.

5. N-VIEWS

a. What is an affine transform? And how many pairs of matching points between to images do I need to solve for it?

b. What is a homography? And how many pairs of matching points between to images do I need to solve for it?

c. Draw a picture that describes rectifying a plane – i.e. why you can convert the image a slanted plane such as the face of a building into an image of building as if you were viewing it head-on.

6. STEREO

a. Given two cameras and a point P in the world, draw out the epipolar plane geometry.

b. What is an epipole?

c. What is the difference between the *essential* matrix and the *fundamental* matrix?

d. We view some world point P with two parallel cameras separated by baseline B meters, and with a focal length of *f*. If the world point P is located horizontally at x_L in the left image (in the same units as *f*) and x_R in the right image the *disparity d* is $(x_L - x_R)$. Write the formula for the depth *Z* of P in terms of *d,* B, and *f*.

e. What are some constraints about the viewed surface or that matching that reduce the search in looking for stereo matches?

f. What's the difference between normalized correlation and regular (cross) correlation?

g. What does random dot stereograms tells us about human stereopsis?

7. SHADING

    a. What is Lambertian shading? And what does it say is the relation between the incident light angle, the normal, the viewing direction and brightness?

    b. If a surface is Lambertian, how many known light sources would you need to turn on (one at a time) to unambiguously figure out the orientation of the surface at each visible point?

    c. In photometric stereo under a Lambertian assumption there are 3 degrees of freedom at every point on the surface so we need at least 3 light sources. What are the 3 degrees of freedom? (Hint: two have to do with geometry.)

8. FEATURES

    a. We say that point descriptors should be both "invariant" and "distinctive"? What do we mean by "invariant" and why is it good?

    b. Harris features are referred to as "*Harris corners*" and are found by looking at a *2nd moment matrix.* Why and why? And what does it mean if the largest eigenvalues of that matrix is much, much, much bigger than the second one?

    c. How can we make a feature detector (like SIFT) mostly invariant to illumination?

    d. Are Harris corners invariant to rotation? Why or why not? What about SIFT features?

9. MODEL FITTING

    a. In using RANSAC to do, say, a panorama, what are putative matches? How do you get them? Why do you need them?

    b. Suppose we are using RANSAC to find circles. Our inputs might be points or oriented edge elements. What would the argument be as to why points are better? What would the argument be as to why the oriented edge elements would be better?

10. SEGMENTATION

    a. How can segmentation be thought of as a clustering problem? How do you get geometry into that approach?

b.   What does Mean Shift do and how does it relate to segmentation?

11. MOTION

a.   What is the Brightness constancy constraint equation and what are the unknowns?

b.   What is the aperture problem in considering image motion?

c.   What is the relation between the Lucas and Kanade optic flow method and finding the Harris corners.

d.   Lucas and Kanade is the optic flow method based upon gradients. What are the assumptions of the method? And what can be done to apply the algorithm when those assumptions are false.

e.   How would you work the knowledge that there is affine flow only into the LK method?

12. TRACKING

a.   Tracking is iterating between Prediction and Correction.   In terms of the observations, prediction can be written as:

$$P\left(X_t \middle| Y_0 = y_0, Y_1 = y_1, \ldots, Y_{t-1} = y_{t-1}\right)$$

Write out a similar expression for the correction step.

b.   In such tracking what is the role of the dynamics model? The likelihood (observation) model?

c.   There are two independence (or conditional independence) assumptions in the tracking we did (Kalman or Particle). What are they? Hint – one has to do with the states, the other with the observations.

d.   The Kalman filter imposes Gaussian distributions for the state estimation and two other model elements. What are those elements?

e.   Particle filters first sample from a weighted distribution of particle, each particle being representative of the state. After that sample is picked, what is done to the sample *before considering the measurements*

13. CLASSIFICATION

   a. If we reduce the number of dimensions of a signal using PCA, we first subtract off the mean. Why?

   b. What's the difference between generative models and discriminative models for classification? Which relies on Bayes rule and how?

   c. What's a *cascade (filter)* and how is it used with boosting for face detection?

   d. What are integral images and why are they so useful?

   e. What is the *Kernel trick*? And how do we make use of it with SVMs?

   f. How do we define the "bag of words" that is used for recognition?

14. ACTIVITY

   a. An HMM   is defined by a triple written in class as (A,B,Π) but in the book as (P,Q,Π) . What is each of these? (Or "What are the three elements that make up an HMM?" if you can't remember which is which.)

   b. What are the three fundamental problems to be solved when using an HMM? And what is the forward algorithm?

   c. If N is the no. of states and T is the number of observations (one per time step), the forward algorithm gives a recursive method of computing the probability of a given HMM producing the observation sequence (written as P(O|λ)). What is the computational complexity of that computation in terms of N and T?

   d. And just how are HMMs used in activity recognition?

15. MORPHOLOGY

   a. How are OPEN and CLOSE defined in terms of Dilate and Erode?

   b. What is the effect of using a bigger structuring element when doing a CLOSE operation as opposed to a smaller one?

   c. What is the maximum number of repeated CLOSE operations you can perform until you stop seeing any further effects?