

Midterm Exam

CS223b

Stanford CS223b Computer Vision, Winter 2004

Feb. 18, 2004

Full Name: _____

Email: _____

- This exam has 7 pages. Make sure your exam is not missing any sheets, and write your name on every page.
- The exam is closed book, closed notes.
- The exam has a maximum score of 75 points. You have 60 minutes.
- Write your answers in the space provided. If you need extra space, use the back of the preceding sheet.
- Please write clearly and be concise. Most questions can be answered in 1-2 sentences.

Good luck!

1	(5 max)	
2	(10 max)	
3	(20 max)	
4	(30 max)	
5	(10 max)	
6	(15 extra max)	
total	(75 max + 15 extra)	

1 Feature Detection

5pts

When a very good lens is coupled with a standard CCD camera, will defocussing the lens (1) improve or (2) worsen edge localization? Explain why (1-2 sentences).

Answer: (1) If the lens is good and sharply in focus, its bandwidth may exceed the sampling rate of the CCD. Defocussing the lens will narrow the band of the signal (like applying a lowpass filter) and reduce aliasing.

2 Active Contours

10pts

The snake energy function consists of three terms, E_{cont} , E_{curv} , and E_{image} .

1. For each of these terms, discuss what happens when the term is omitted?

Answer: The main function of E_{cont} is to keep the points uniformly distributed along the snake; removing it endangers that. It also makes the snake less willing to shrink.

Removing E_{curv} takes away the snake's desire to stay smooth, allowing it to become jagged and to overfit noise.

Removing E_{image} makes the snake shrink to a point.

2. What is the effect of giving a negative weight to E_{cont} ?

Answer: The points are forced into a non-uniform distribution along the snake, plus the snake will have a tendency to grow. Because of the first property, a negative coefficient should not be used for a snake started inside an object.

3 Optical Flow

20pts

Given the brightness constancy equations:

$$\frac{dE}{dx} \frac{dx}{dt} + \frac{dE}{dy} \frac{dy}{dt} + \frac{dE}{dt} = 0$$

1. Explain intuitively what each of the five derivatives in the brightness constancy equation measures.

Answer:

- $\frac{dE}{dx}$ is the x component of the image gradient.
- $\frac{dE}{dy}$ is the y component of the image gradient.
- $\frac{dx}{dt}$ is the x component of the motion field
- $\frac{dy}{dt}$ is the y component of the motion field
- $\frac{dE}{dt}$ is the change in image brightness over time

2. Give three different concise phenomena in the physical world that will invalidate the brightness constancy equation.

Answer: Television, barber pole, spinning Lambertian sphere, occlusions, specular highlights on a glossy surface.

3. What is the aperture problem, and how does it relate to this equation?

Answer: The aperture problem is the ambiguity in the direction of the motion field along an isobrightness contour when a small image patch is being observed during motion. Only the component of the motion field along the image intensity brightness can be determined. The orthogonal component is entirely unconstrained. This can be seen in equation (1) which is a single scalar equation with two unknowns.

4. Suppose that the camera only translates, that you know the direction of translation, and that the rest of the world is stationary. Explain how you would use this knowledge to address the aperture problem.

Answer: The knowledge of the direction of translation could be used as an additional constraint on the motion field to give us 2 equations with two unknown.

4 Stereo and Epipolar Geometry

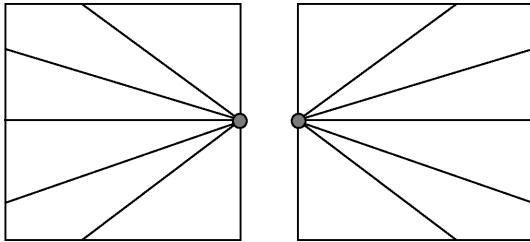
30pts

1. Given two ideal pinhole cameras where:

- the baseline of the two cameras is parallel to their scanlines,
- the optical axes of the two cameras intersect to form an angle of 90 degrees,
- the two centers of projection are at equal distances from the intersection of the optical axis, and
- the field of view of each camera is 90 degrees.

Draw the epipoles and a few epipolar lines.

Answer:



2. What is the size of the fundamental matrix, the number of degrees of freedom in it, and the minimum number of point correspondences to compute it using linear algebra, assuming sufficiently many images?

Answer: The size of the fundamental matrix is 3×3 . There are 7 degrees of freedom in the fundamental matrix. There are 9 total. Subtract 1 account for scale. Subtract another to account for that fact that its less than full rank ($\det(F) = 0$). One needs 8 point correspondences to calculate F using linear algebra.

3. Under what conditions is the essential matrix identical to the fundamental matrix?

Answer: $E = M_r^{-T} E M_l^{-1}$. Or, in other words, when the pixels expressed in camera coordinates are exactly the same as those expressed in pixel coordinates.

4. If E is the essential matrix returned by the 8-point algorithm, what do the solutions to the following two systems represent: $E x = 0$ and $E' x = 0$.

Answer: The right and left epipoles, respectively. Consider $E x = 0$. The epipolar constraint can be expressed as follows: $p_l E p_r = 0$. When this equations hold, we know that p_l and p_r lie on corresponding epipolar lines. But consider a p_r such that $E p_r = 0$ already. This suggests that no matter what p_l I choose, p_r must lie on a corresponding epipolar line. This is only true for one unique p_r , a point through which all epipolar lines must pass. That point is the right epipole. The argument is similar for the left epipole.

5. Imagine two pinhole cameras with focal length f , co-planar image planes and baseline of length b . Suppose both cameras have a pixel size of s , and for the sake of this question we assume pixels are either black or white (so that we can't do the sub-pixel-accuracy trick). Give a mathematical expression for the (approximate) z -error induced by the limited pixel resolution. State how this error depends on z and on f (linear? quadratic? exponential?).

Answer: There are many ways to derive this, e.g., by approximating the exact formula through Taylor expansion. Here is a simple derivation, which is also approximate: If the point is at depth z centered between both co-planar cameras, it describes a right triangle whose legs are $\frac{b}{2} + x$ and $f + z$. Usually x is negligible in comparison to $\frac{b}{2}$, so $\frac{b}{2}$ really determines the accuracy of the depth estimate. Let's now assume that the depth error for image coordinate $x + s$ is the same in magnitude as for the image coordinate $x - s$; this is reasonable for points far out. Then we have

$$\frac{f + z}{\frac{b}{2}} = \frac{f + z + \delta}{\frac{b}{2} + s} \quad (1)$$

which resolves to $\delta = 2s \frac{f+z}{b}$. That is, linear in z and f .

6. Give one advantage and one disadvantage of using a feature-based stereo algorithm compared to a correlation-based (or intensity-based) algorithm.

*Answer: Advantage: it is faster, less sensitive to illumination changes, and specularities.
Disadvantage: returns sparse disparity maps*

5 Structure From Motion

10pts

Suppose we perform SFM in an environment with N stationary point features, of which no four are coplanar. Suppose further that the absolute location of k static features is known, whereas the location of $N - k$ is unknown.

1. For $m = 10$ camera images, what is the minimum value of k that enables us to reconstruct the absolute location of all features?

Answer: The answer is simple: $k = 3$ points are usually sufficient to localize each camera (6 constraints - 6 intrinsic parameters). $k = 2$ points would be insufficient since the entire structure could be rotated round the axis defined by these two points.

As a footnote, there might exist configurations under which even $k = 3$ points provide ambiguities. For example, this may happen when two points are equidistant to a third, and the camera is positioned exactly on the plane that separates those two points. Then there is a discrete ambiguity. But these cases are rare...

2. Suppose we add a moving feature to the environment. Can we detect which feature moves? If yes, argue how. If no, provide a counterexample.

Answer: It's actually possible to make a feature look stationary by moving in proportion to the camera motion. Here's how you show this: pick a point feature in your environment, and hold your hand so this feature becomes occluded. Then move your head and your hand, maintaining the occlusion of this point feature. If you do this really carefully with a point-sized occluder, the resulting occluder will be indistinguishable from a static point in the environment.

6 Extra Points: Acoustic Orthographic Structure from Motion 15pts

This is a tricky question! In class, we discussed the SFM solution to the orthographic SFM problem, and in particular the Tomasi/Kanade algorithm for recover the structure and camera extrinsics of m cameras from n known correspondences under orthographic projections. Suppose we exchange cameras by omni-directional microphones: The problem is to find out the locations of m microphones from the joint detection times of n sound signals originating from n unknown locations. Assume the sound sources are all very(!) far away, and assume that the correspondence is known. Can you devise a SVD solution to this problem in affine geometry? Can you explain how to obtain a solution in Euclidean geometry? How many claps are needed to localize m microphones up to an affine/Euclidean transformation?

Answer: Eric Berger found a nice answer: The relative arrival time—mapped into distances using the speed of sound—is an affine coordinate system, even one that retains distances. Simply pick the first three measurements and orthogonalize this basis.

The SVD answer is based on the insight that for a sound very far away, only the angle of the incoming sound wave matters. Thus, the “structure” is 1-D for each sound source when confined to a plane, and 2-D for a sound in 3-D space. The relative arrival time of a sound between different sensors is then a projection of the coordinate differences onto the incoming sound direction. This projection involves sin and cos, but when replaced with an arbitrary matrix becomes a linear operation. SVD is then applicable just as in the Tomasi/Kanade paper.