Midterm Exam CS223B

Stanford CS223B Introduction to Computer Vison, Winter 2005

Full Name: _____

Email: _____

ANSWERS

Welcome to the CS223B Midterm Exam!

- The exam is 8 pages long. Make sure your exam is not missing any sheets. The exam has a maximum score of 100 points. You have 75 minutes.
- The exam is open book, open notes (but no electronic devices that can communicate with the outside world, such as laptops).
- Write your answers in the space provided. If you need extra space, use the back of the preceding sheet.
- Write clearly and be concise.
- All points will be manually counted before certification.
- SCPD students: If you are taking this exam off campus, you have to fax it to (650) 725-1449 exactly 75 minutes after receipt. Alternatively, you can Email your answers to thrun@stanford.edu.

Question	Points
1 (20 max)	
2 (30 max)	
3 (15 max)	
4 (20 max)	
5 (15 max)	
total	

1 Calibration 20pts

We seek to calibrate an undistorted camera with a planar calibration object. The object possesses M distinct features. The location of those features in the object coordinate frame are unknown (and they are not arranged in a checkerboard).

In answering the questions below, you might want to use the following notation/equations:

$$\begin{pmatrix} \tilde{X}^C \\ \tilde{Y}^C \\ \tilde{Z}^C \end{pmatrix} = R \begin{pmatrix} X^W \\ Y^W \\ Z^W \end{pmatrix} + \begin{pmatrix} T_X \\ T_Y \\ T_Z \end{pmatrix}$$

$$R = \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \cos\varphi & 0 & \sin\varphi \\ 0 & 1 & 0 \\ -\sin\varphi & 0 & \cos\varphi \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & \sin\psi \\ 0 & -\sin\psi & \cos\psi \end{pmatrix}$$

$$\begin{pmatrix} x_{\text{im}} \\ y_{\text{im}} \end{pmatrix} = \left\lfloor \begin{pmatrix} \frac{f}{s_x}\frac{\tilde{X}^C}{\tilde{Z}^C} + o_x + 0.5 \\ \frac{f}{s_y}\frac{\tilde{Y}^C}{\tilde{Z}^C} + o_y + 0.5 \end{pmatrix} \right\rfloor$$

1.1 (5pts) What are the free parameters that can be recovered through the calibration process? Organize them into categories for the different types of parameters in calibration.

Answer: There are five intrinsic parameters but only four can be recovered. The recoverable intrinsic parameters are $\{\frac{f}{s_x}, \frac{s_x}{s_y}, o_x, o_y\}$.
There are $2M$ *parameters for the unknown feature locations on the planar calibration object. This is because we know* $\tilde{Z}_j^W = 0$ *for all features* j*. So we add the object parameters* $\{X_j^W, Y_j^W\}$ *to the set of free parameters. However, the definition of the object reference frame is somewhat arbitrary. We can therefore define* $X_1^W = Y_1^W = 0$*, and* $X_2^W = 0$*. This reduces the number of recoverable parameters by 3. This definition constrains the location of all other features on the board.*
The extrinsic parameters are $\{\phi_i, \varphi_i, \psi_i, T_{X,i}, T_{Y,i}, T_{Y,i}\}$*, for each image* i*.*
In summary, we have $4 + 6K + 2M - 3 = 6K + 2M + 1$ *unknown parameters, with* K *being the nymber of images and* M *the number of features on the calibration pad.*

1.2 (10pts) Let M be the number of features in the calibration object, and K be the number of images of the calibration pattern.

- How many parameters have to be estimated?
- How many constraints are being provided by each image of the calibration pattern?
- What are the lower bounds for M and K? Provide exact formulae, one of the form $K \geq \ldots$ and one of the form $M \geq \ldots$.
- For $K = 3$ images, what is the minimum M?

- For $M = 3$ features, what is the minimum K?

  Answer: We have to estimate $4+6K+2M-3$ *parameters from* $2KM$ *constraints. Hence we have*

$$\begin{aligned} 2KM \geq 1+6K+2M &\iff 2KM-6K \geq 1+2M \\ &\iff K(2M-6) \geq 1+2M \\ &\iff K \geq \frac{1+2M}{2M-6} = \frac{7+2M-6}{2M-6} = 1+\frac{7}{2M-6} \end{aligned} \quad (1)$$

  and (assuming $M > 3$*):*

$$\begin{aligned} K \geq 1+\frac{7}{2M-6} &\iff K-1 \geq \frac{7}{2M-6} \\ &\iff 2M-6 \geq \frac{7}{K-1} \\ &\iff 2M \geq 6+\frac{7}{K-1} \\ &\iff M \geq 3+\frac{7}{2K-2} \end{aligned}$$

  For $K = 3$ *images, we need* $M \geq 5$ *features. With only* $M = 3$ *features, we cannot calibrate, no matter how many images. We need at least* $M = 4$ *features.*

1.3 (5pts) Provide an algorithm for performing the calibration. It shall be legitimate to say "*minimize the following expression over the variables* $x\ldots$*"* or "*compute the eigenvalues of matrix* Y*"* without providing details on the algorithm. You may also use standard algorithms such as SVD.

  Answer: Let $\langle \hat{x}_{\rm im}(i,j), \hat{y}_{\rm im}(i,j) \rangle$ *be the image location of the* j*-th feature in the* i*-th image. Then minimize*

$$\sum_{i,j} \left| \begin{pmatrix} x_{\rm im}(i,j) - \hat{x}_{\rm im}(i,j) \\ y_{\rm im}(i,j) - \hat{y}_{\rm im}(i,j) \end{pmatrix} \right|_2^2$$

  in the free parameters above.

  To be precise, our notation has to be annotated:

$$\begin{pmatrix} \tilde{X}^C(i,j) \\ \tilde{Y}^C(i,j) \\ \tilde{Z}^C(i,j) \end{pmatrix} = R(i) \begin{pmatrix} X^W(j) \\ Y^W(j) \\ 0 \end{pmatrix} + \begin{pmatrix} T_X(i) \\ T_Y(i) \\ T_Z(i) \end{pmatrix}$$

$$\begin{pmatrix} x_{\rm im}(i,j) \\ y_{\rm im}(i,j) \end{pmatrix} = \left\lfloor \begin{pmatrix} \frac{f}{s_x} \frac{\tilde{X}^C(i,j)}{\tilde{Z}^C(i,j)} + o_x + 0.5 \\ \frac{f}{s_y} \frac{\tilde{Y}^C(i,j)}{\tilde{Z}^C(i,j)} + o_y + 0.5 \end{pmatrix} \right\rfloor$$

2 Perspective Geometry 30pts

Consider a scene that contains 5 collinear features, A, B, C, D, and E, where collinearity is defined in 3D coordinates. We know that under perspective projection, the projected features are also collinear in the 2D image plane.

2.1 (5pts) Suppose A, B, C, D, and E are all equidistant. Will the *order* of the projected points (along the projected line) always be the same in the camera image? If yes, argue why this must be the case. If no, provide a counterexample.

Answer: Yes, they will. It's the direct result of the fact that the projection is linear (and it also follows from the math below).

2.2 (5pts) Will the observed features also be equidistant in the image plane? If yes, argue why. If not, argue why not.

Answer: The answer is no, with the exception of uninteresting degenerate cases. Distant objects appear smaller under perspective projection. The ratio is given by the math below (Question 2.5): $\lambda = (Z_0 + \mu w)^{-1}$.

2.3 (5pts) What are the implications of the previous answer for stereo vision? Specifically, suppose we find five collinear features a, b, c, d, e in the left image, will they also be collinear in the right image? If yes, argue why; if not, provide a counterexample.

Answer: Collinearity in the 2D image space does not imply collinearity in 3D coordinates. When points are not collinear, they can be viewed from a direction where the projection is also not collinear.

2.4 (5pts) Suppose the projected points a, b, c, d, e are collinear in both stereo images. Can we expect that the order is preserved? Again, argue the correctness or provide a counterexample.

Answer: Unfortunately not. The answer was already given in class when we observed that a foreground object can shift relative to the background.

2.5 (10pts) *Difficult! Try not to spend too much time on this question.* Prove the claim in the introduction: under perspective projection the projected features are collinear in the 2D image plane.

Answer: A convenient answer involves writing the perspective projection homogeneous coordinates:

$$\lambda \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{pmatrix} = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

Now consider the equation of a line parameterized by the variable μ*:*

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} X_0 \\ Y_0 \\ Z_0 \end{pmatrix} + \mu \begin{pmatrix} u \\ v \\ w \end{pmatrix}$$

And plug this into our projection equation:

$$\begin{pmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{pmatrix} = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix} \left[\begin{pmatrix} X_0 \\ Y_0 \\ Z_0 \end{pmatrix} + \mu \begin{pmatrix} u \\ v \\ w \end{pmatrix} \right]$$

$$= \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_0 \\ Y_0 \\ Z_0 \end{pmatrix} + \mu \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \\ w \end{pmatrix}$$

This is a linear equation in μ. The linearity of the projection

$$\lambda \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{pmatrix}$$

now follows from the fact that both $\tilde{x}$ and $\tilde{y}$ are multiplied by the same factor λ. λ is of course not *linear in μ, since perspective projection does not preserve distances.*

An alternative elegant proof came from one of the students: The points in 3D space form a line. This line and the focal point define a plane. The plane intersects the image plane; and this intersection itself is a line. Since all projections go through the focal point, the images of the 3D points must lie on this line.

3 Stereopsis 15pts

Imagine a stereo camera rig of two cameras with focal length f, whose image planes are coplanar, and whose baseline is b. Provide a mathematical expression that relates the pixel size s to the minimum distinguishable depth Z. Specifically, at what rate changes the depth resolution with the absolute depth Z of an object. Assume $Z \gg f$, $Z \gg b$, and the object is centered in front of one of the cameras. *It is okay to be approximate if your approximation is sound and well-explained.*

Answer: Place two objects on the optical axis of the left camera, so that they both are projected onto the image center. Call their depths Z, and $Z + \Delta Z$, respectively. In the right image, the first object will be projected at location x off the image center (we are ignoring y in the analysis). The second object is projected one pixel away, to location $x + s$.

In approximation, we assume that the rays entering the right camera from both objects are parallel. Then we have two similar triangles:

$$\begin{aligned}
\frac{f+Z+\Delta Z}{b+x+s} \approx \frac{f+Z}{b+x} &\iff f+Z+\Delta Z \approx \frac{(f+Z)\,(b+x+s)}{b+x} \\
&\iff \Delta Z \approx (f+Z)\,(\frac{b+x+s}{b+x}-1) \approx \frac{(f+Z)\,s}{b+x} \\
&\approx s\,\frac{f+Z}{b} \approx s\,\frac{Z}{b}
\end{aligned}$$

Thus, the answer is linear with ratio Z/b.

4 True or False? 20pts

Correct answer is 1 point per question; a false answer results in minus 1 point.

Answer:

	FALSE	*Straight lines in a cylindrical projection always remain straight.*
	FALSE	*An image may have no more than three vanishing points*
	FALSE	*Any perspective projection is also an affine transform.*
TRUE		*Any affine transform is also a perspective projection.*
TRUE		*Least squares calibration techniques tend to yield superior calibration results over the algebraic methods discussed in Trucco/Verri.*
TRUE		*The focal length of a camera can be determined by rotating it (approximately) around its focal point.*
TRUE		*The image center of a distortion-free camera can be determined by analyzing vanishing points of rectrahedral objects of unknown size (e.g., a cube).*
	FALSE	*Every Gaussian filter G can be decomposed into two 1-D filters that are aligned with the vertical and horizontal axes of the image, respectively.*
	FALSE	*The Canny edge detector is a linear filter.*
TRUE		*Parallax refers to the apparent shifting of an object when viewed from different directions.*
TRUE		*The smaller the baseline of a stereo rig, the lower the error rate for the correspondence step.*
TRUE		*The fundamental matrix is not invertible.*
	FALSE	*The epipolar plane is normal to the optical axis.*
	FALSE	*To estimate the range of a straight edge in the image, 2 cameras are always sufficient.*
	FALSE	*Optical flow is the preferred technique for establishing correspondence in a stereo rig.*
TRUE		*Corners found by the Harris corner detector tend to be unaffected from the aperture effect.*
	FALSE	*Image stitching requires image warping.*
	FALSE	*When using exactly eight points, the "Eight Point Algorithm" requires that no three points are coplanar.*
	FALSE	*For Structure from Motion (SFM), we need at least three images when points are in correspondence.*
TRUE		*The RANSAC algorithm addresses the correspondence problem.*

5 Vision Systems 15pts

You are tasked to develop a vision system for building safer cars. In particular, your system should warn a human driver when in danger of colliding with a pedestrian. The system should work day and night, and in poor weather. Outline a system for performing this task. For each component, explain the problem that is being solved, the approach brought to bear, and the type results to expect. You will be graded based on the soundness of your proposal and your reasoning.