Midterm Exam

CS 231a Spring 2015-2016

Monday, May 9th, 2016

This written exam is 80 minutes long and is open book and open notes. You may access these notes on your electronic device. However, the use of the internet during the exam is strictly forbidden. Answer the questions in the spaces provided in the exam booklet.

Honor Code Statement

1. The Honor Code is an undertaking of the students, individually and collectively:

  1. that they will not give or receive aid in examinations; that they will not give or receive unpermitted aid in class work, in the preparation of reports, or in any other work that is to be used by the instructor as the basis of grading;

  2. that they will do their share and take an active part in seeing to it that others as well as themselves uphold the spirit and letter of the Honor Code.

2. The faculty on its part manifests its confidence in the honor of its students by refraining from proctoring examinations and from taking unusual and unreasonable precautions to prevent the forms of dishonesty mentioned above. The faculty will also avoid, as far as practicable, academic procedures that create temptations to violate the Honor Code.

3. While the faculty alone has the right and obligation to set academic requirements, the students and faculty will work together to establish optimal conditions for honorable academic work.

Honor Code Acknowledgement

I acknowledge and accept the Honor Code.

Printed Name: ______________________________

SUNetID: ______________________________

Signature: ______________________________

Question	Points	Score
1	10	
2	1	
3	1	
4	1	
5	1	
6	1	
7	1	
8	1	
9	3	
10	15	
11	5	
12	20	
13	15	
Total:	75	

True/False Questions

1. [10 points] The following statements are either true or false. Circle the correct answer to the left of the question. If you circle an incorrect answer, you will be penalized one point. If you do not circle an answer, you will not be penalized.

(a) True False Consider a naive descriptor consisting of a vector of pixels around an image region of interest. Normalizing such a descriptor (i.e. making its norm equal to 1 and mean equal to 0) will make the image's pixel intensity values invariant to scaling and biasing.

Solution: True

(b) True False It is impossible to estimate the intrinsic parameters of the camera given a single image even with prior information about the scene.

Solution: False. Single view metrology provides us with methods for esimating the intrinsic parameters of the camera from a single image.

(c) True False A line in Cartesian space corresponds to a point in Hough space.

Solution: True. A point (x_1, y_1) in Cartesian space corresponds to the line $y_1 = mx_1 + n$ in Hough space.

(d) True False A good blob detector can be obtained by convolving an image with a Laplacian of Gaussians.

Solution: True.

(e) True False Pincushion distortion modifies the image only along the vertical direction and barrel distortion modifies the image only along the horizontal direction.

Solution: False. Both pincushion and barrel distortion modify the image along both the vertical and horizontal directions.

(f) True False Assume that we identified two perpendicular planes in an image of the 3D world. We estimate two vanishing points (one on each plane) in the image from a pair of parallel lines on each of these planes. Using these two vanishing points and no other information from the scene, we can in general estimate the camera parameters.

Solution: False. We need three vanishing points to setup three scalar equations to recover $\omega = (\mathbf{K}\mathbf{K}^T)^{-1}$.

(g) True False The vanishing point associated with a line in 3D space (when viewed through a pinhole camera) can never be a point at infinity.

Solution: False.

(h) True False Given a set of points, the Hough Transform can be used to fit two lines but not three lines.

Solution: False. We can keep the three most populated bins instead of the two most populated bins.

(i) True False The result of applying a scale, rotation, and translation (in that order) to a vector $\vec{v}$ is equal to the result of applying the same rotation, translation, and scale (in that order) to the same vector $\vec{v}$.

Solution: False. In general, scale, rotation, and translation do not commute.

(j) True False Space-carving can recover the true shape of any object if the object is imaged by "enough" cameras.

Solution: False. Consider a hollow sphere. Due to occlusion, space carving will never be able to infer that the inside is hollow.

Multiple Choice Questions

2. [1 point] A good corner detector identifies unique key points to characterize the image. Which of the following small window of pixels would yield the best key point? **Circle the most correct answer.**

    A. A window containing a homogeneous set of pixels

    B. A window containing a single line

    C. A window containing intersecting lines

    D. A window containing a pair of parallel lines

3. [1 point] How many degrees of freedom are in the camera matrix **K** if skewness may occur, it has square pixels, and that the origin of coordinates in the image plane may not be at the principal point. **Circle the most correct answer.**

    A. 2

    B. 4

    C. 5

    D. 8

    E. 11

4. [1 point] Which of the following always hold(s) under affine transformations? **Circle all that apply.**

    A. Parallel lines remain parallel

    B. Ratio of lengths of parallel line segments remain the same

    C. Ratio of areas remain the same

    D. Perpendicular lines remain perpendicular

    E. Angles between two line segments remain the same

5. [1 point] Which of the following statements regarding projections is/are true? **Circle all that apply.**

    A. Weak perspective projection is affine

    B. Orthogonal projection is affine

    C. There exists a fixed, pre-defined magnification for perspective projections

    D. There exists a fixed, pre-defined magnification for weak perspective projections

    E. Every projected point under orthogonal projection is independent of depth

6. [1 point] Which of the following property applies if two cameras are rectified? **Circle all that apply.**

    A. Epipolar lines are vertical

    B. The cameras can can differ by an arbitrary rotation with no translation

    C. Epipolar lines will intersect at infinity

    D. Fundamental matrix reduces to an identity matrix

    E. None of the above

7. [1 point] What could be a source of error for solving the structure from motion problem by the factorization method originally described by Tomasi-Kanade ? **Circle all that apply.**

    A. There exists two cameras whose optical axes are parallel

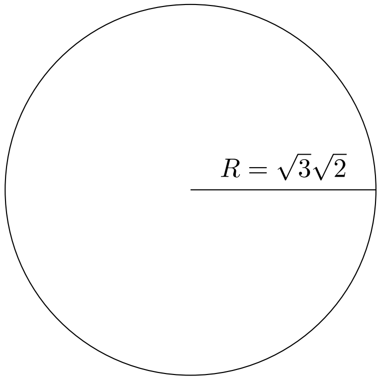
    B. Every correspondence is not visible by all the cameras

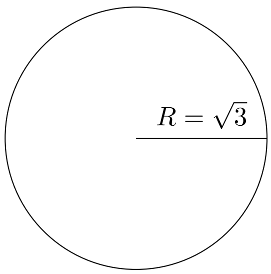
    C. There exists two cameras such that one is in the same position as another, but rotated slightly

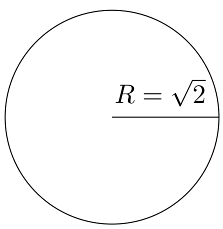
    D. Projective, not affine, cameras

    E. None of the above

8. [1 point] Which of the following circles would be mostly easily detected as a blob by a Laplace filter with a Gaussian parameterized by $\sigma = 1$? **Circle the most correct answer.**

A. 

B. 

C. 

9. [3 points] Projective, Affine, Similarity, and Isometric Transformations

(a) [2 points] Classify each of the following transformations. **Fill-in the circle corresponding to the *most specific* classification.**

i. $H_1 = \begin{bmatrix} 3/4 & -1 & 0 \\ 1 & 3/4 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ ○ Projective ○ Affine √ **Similarity** ○ Isometric

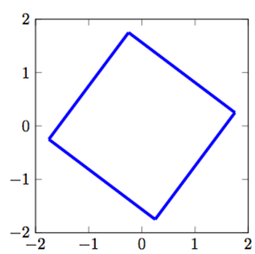
ii. $H_2 = \begin{bmatrix} 3/5 & -4/5 & 1 \\ 4/5 & 3/5 & -1 \\ 0 & 0 & 1 \end{bmatrix}$ ○ Projective ○ Affine ○ Similarity √ **Isometric**

iii. $H_3 = \begin{bmatrix} 3/16 & -1 & -1/4 \\ 1/4 & 3/4 & 1/2 \\ 1/4 & 1/4 & 1 \end{bmatrix}$ √ **Projective** ○ Affine ○ Similarity ○ Isometric

iv. $H_4 = \begin{bmatrix} 3/8 & -5/8 & 0 \\ 1/2 & 5/4 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ ○ Projective √ **Affine** ○ Similarity ○ Isometric

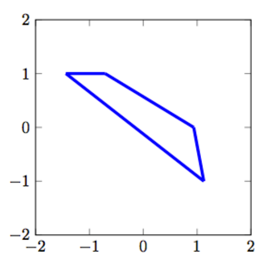
(b) [1 point] The figures below shows the outputs of applying one of transformations (projective, affine, similarity, and isometric) to a square with vertices at (1,1), (1,-1), (-1,-1), (-1,1). **Fill in the circle corresponding to the *most specific* transformation used to generate each output.**

i.



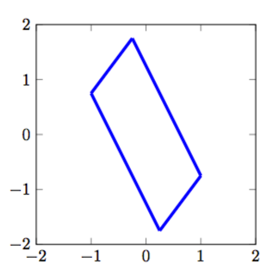
○ Projective ○ Affine √ **Similarity** ○ Isometric

ii.



√ **Projective** ○ Affine ○ Similarity ○ Isometric

iii.



○ Projective √ **Affine** ○ Similarity ○ Isometric

Camera Models and Calibration

10. [15 points] Suppose we want to solve for the camera matrix $\mathbf{K}$ and that we know that the world coordinate system is the same as the camera coordinate system. Assume the matrix $\mathbf{K}$ has the structure outlined below. Note that K_{33} is an unknown. Assume that we are given n correspondences. Each correspondence consists of a world point (x_i, y_i, z_i) and its projection (u_i, v_i) for $i = 1, \ldots, n$.

$$\mathbf{K} = \begin{pmatrix} K_{11} & K_{12} & K_{13} \\ 0 & K_{22} & K_{23} \\ 0 & 0 & K_{33} \end{pmatrix}$$

(a) [2 points] What is the minimum number of correspondences needed to solve for the unknowns in the matrix $\mathbf{K}$?

Solution: Given the structure of the camera matrix ($K_{21} = K_{31} = K_{32} = 0$) and the information that K_{33} should be treated as an unknown, we have six unknowns. Each correspondence provides two equations and thus we would need at least three correspondences.

(b) [8 points] Set up an equation of the form $\mathbf{Ax} = \mathbf{0}$ to solve for the unknowns in $\mathbf{K}$ (where $\mathbf{A}$ is a matrix, and $\mathbf{x}$ and $\mathbf{0}$ are vectors). Be specific about what the matrix $\mathbf{A}$ and vector $\mathbf{x}$ are.

Solution: Each correspondence gives us an equation of the form

$$K_{33} z_i \begin{bmatrix} u_i \\ v_i \end{bmatrix} - \begin{bmatrix} K_{11}x_i + K_{12}y_i + K_{13}z_i \\ K_{22}y_i + K_{23}z_i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

and we can rearrange to obtain

$$\begin{bmatrix} x_i & y_i & z_i & 0 & 0 & -z_i u_i \\ 0 & 0 & 0 & y_i & z_i & -z_i v_i \end{bmatrix} \begin{bmatrix} K_{11} \\ K_{12} \\ K_{13} \\ K_{22} \\ K_{23} \\ K_{33} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Thus, our linear system has the form $\mathbf{Ax} = \mathbf{0}$ with

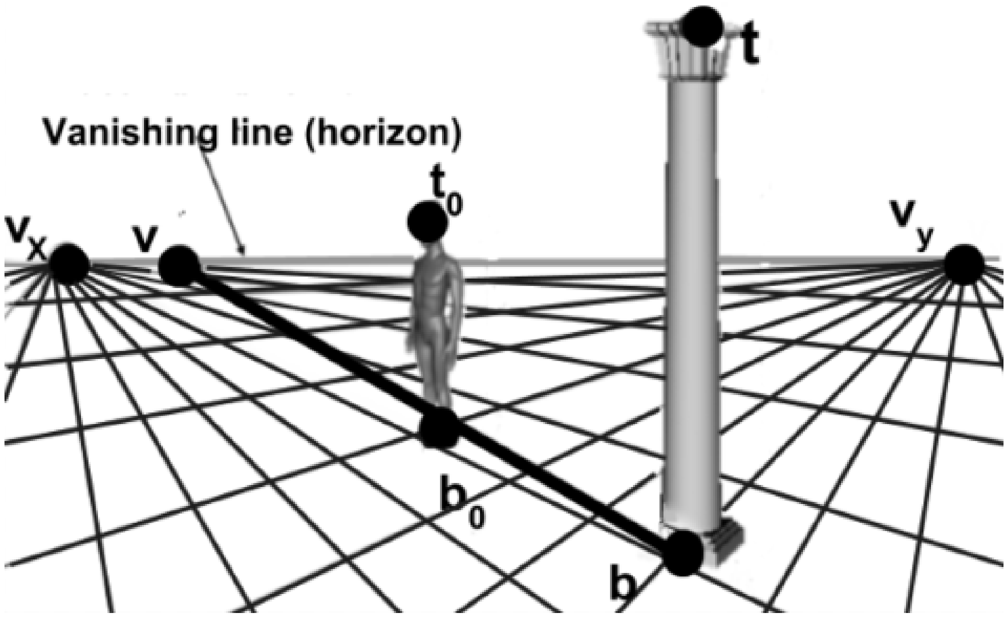
$$\mathbf{A} = \begin{bmatrix} x_1 & y_1 & z_1 & 0 & 0 & -z_1 u_1 \\ 0 & 0 & 0 & y_1 & z_1 & -z_1 v_1 \\ x_2 & y_2 & z_2 & 0 & 0 & -z_2 u_2 \\ 0 & 0 & 0 & y_2 & z_2 & -z_2 v_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n & y_n & z_n & 0 & 0 & -z_n u_n \\ 0 & 0 & 0 & y_n & z_n & -z_n v_n \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} K_{11} \\ K_{12} \\ K_{13} \\ K_{22} \\ K_{23} \\ K_{33} \end{bmatrix}.$$

(c) [5 points] Explain how to solve for the unknowns in the camera matrix $\mathbf{K}$. Make sure $K_{33} = 1$.

Solution: We want to minimize $\|Ax\|$ subject to $\|x\| = 1$. Thus, we use SVD on A and obtain the eigenvector corresponding to the smallest eigenvalue to give us a scaled version of $(K_{11}, K_{12}, K_{13}, K_{22}, K_{23}, K_{33})$. To ensure that $K_{33} = 1$, we divide each element in the eigenvector by the element corresponding to K_{33} (in our eigenvector). The result is a vector of the form $(K_{11}, K_{12}, K_{13}, K_{22}, K_{23}, K_{33} = 1)$.

Single View Metrology

11. [5 points] Let t_0 and b_0 be the end points of an object (a man) in the image plane. Let t and b be the end points of another object (a column) in the image plane. Let v_x and v_y be the horizontal vanishing points. Also, v is the point where a line passing through b and b_0 meets the horizon line.



(a) [2 points] Find an expression of the normal vector of the ground plane (i.e. the floor to which the pillar and the man are perpendicular) in terms of v_x and v_y assuming the camera is canonical?

> **Solution:** Since for a canonical camera, $\mathbf{K} = \mathbf{I}$. Thus, the normal of the ground plane is simply $\frac{v_y \times v_x}{||v_y \times v_x||}$

(b) [2 points] Let one object (the man) have height H in the physical world and let the other object (the column) have height R in the physical world. Let p be the point in the image plane where the line passing through t and t_0 meets the horizon. What is the expression for the point p in terms of t, t_0, v_x, v_y and v if $H \neq R$?

> **Solution:**
>
> $$p = (t \times t_0) \times (v_x \times v_y)$$

(c) [1 point] What happens if $H = R$?

> **Solution:** If $H = R$, then $\mathbf{p} = \mathbf{v}$ since the line passing through $\mathbf{r}$ and $\mathbf{t}_0$ would be parallel to the line passing through $\mathbf{b}$ and $\mathbf{b}_0$.

Epipolar Geometry

12. [20 points] Consider the point $\mathbf{X}$ expressed with respect to the world coordinate system. Suppose that the transformation between two cameras is described by a rotation matrix $\mathbf{R}$ and a translation vector $\mathbf{t}$ so that $\mathbf{X}' = \mathbf{R}\mathbf{X} + \mathbf{t}$. Assume that the first camera coordinate system is equivalent to the world coordinate system and that its camera matrix is $\mathbf{K}$. The camera calibration matrix of the second camera is $\mathbf{K}'$.

    (a) [4 points] Write the camera projection matrices $\mathbf{M}$ and $\mathbf{M}'$ in terms of $\mathbf{K}$, $\mathbf{K}'$, $\mathbf{R}$, and $\mathbf{t}$.

    Solution:

    $$\mathbf{M} = \mathbf{K}\left[\mathbf{I}|\mathbf{0}\right]$$
    $$\mathbf{M}' = \mathbf{K}'\left[\mathbf{R}|\mathbf{t}\right]$$

    (b) [4 points] What are the expressions for the epipoles, $\mathbf{e}$ and $\mathbf{e}'$ in terms of one or more of the following: $\mathbf{M}$, $\mathbf{M}'$, $\mathbf{R}$, and $\mathbf{t}$.

    Solution: The epipoles are the projections of the other camera center.

    $$\mathbf{e} = \mathbf{M}\begin{pmatrix} -\mathbf{R}^T\mathbf{t} \\ \mathbf{1} \end{pmatrix} = -\mathbf{K}\mathbf{R}^T\mathbf{t}$$
    $$\mathbf{e}' = \mathbf{M}'\begin{pmatrix} \mathbf{0} \\ \mathbf{1} \end{pmatrix} = \mathbf{K}'\mathbf{t}$$

    (c) [8 points] An alternate formulation of the fundamental matrix, $\mathbf{F} = [\mathbf{e}']_\times \mathbf{K}'\mathbf{R}\mathbf{K}^{-1}$. Using the results from the earlier parts, show that $\mathbf{e}' \times \mathbf{x}' = \mathbf{F}\mathbf{x}$ where $\mathbf{x}'$ is the image of $\mathbf{X}$ in the second camera coordinate system.

    Solution:

    $$\begin{aligned} \mathbf{F}\mathbf{x} &= [\mathbf{e}']_\times \mathbf{K}'\mathbf{R}\mathbf{K}^{-1}\mathbf{x} \\ &= [\mathbf{e}']_\times \mathbf{K}'\mathbf{R}\mathbf{K}^{-1}\mathbf{K}\begin{bmatrix}\mathbf{I} & \mathbf{0}\end{bmatrix}\mathbf{X} \\ &= [\mathbf{e}']_\times \mathbf{K}'\mathbf{R}\begin{bmatrix}\mathbf{I} & \mathbf{0}\end{bmatrix}\mathbf{X} \\ &= [\mathbf{e}']_\times \mathbf{K}'\begin{bmatrix}\mathbf{R} & \mathbf{0}\end{bmatrix}\mathbf{X} \end{aligned}$$

    We recall that in Part b we showed that $e' = K't$ and that the cross product of a vector with itself is zero. Therefore:

    $$\begin{aligned} \mathbf{F}\mathbf{x} &= [\mathbf{e}']_\times \mathbf{K}'\begin{bmatrix}R & 0\end{bmatrix}\mathbf{X} \\ &= [\mathbf{e}']_\times \mathbf{K}'\begin{bmatrix}R & t\end{bmatrix}\mathbf{X} \\ &= [\mathbf{e}']_\times \mathbf{x}' \\ &= \mathbf{e}' \times \mathbf{x}' \end{aligned}$$

    (d) [4 points] What is the geometric interpretation of $\mathbf{e}' \times \mathbf{x}'$?

Solution: $\mathbf{l}' = \mathbf{e}' \times \mathbf{x}'$ is the epipolar line that passes through epipole $\mathbf{e}'$ and point $\mathbf{x}'$ in the image of the second camera.

Plane Fitting: Hough Voting, RANSAC, and Least-Squares

13. [15 points] Assume that you have captured a point cloud with a single dominant plane (e.g. the front wall of a building) at unknown orientation, plus smaller numbers of other scene points (e.g. trees, poles, a street, etc.) that are not part of this plane. Each point in the point cloud is represented by (x_i, y_i, z_i) coordinates. The equation of a plane is given by $ax + by + cz + d = 0$.

    (a) [5 points] Now, we would like to fit a plane to these points using least squares. Write the linear system whose solution minimizes the objective function in a least squares sense.

    Solution:

    $$\begin{bmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ \vdots & \vdots & \vdots & 1 \\ x_n & y_n & z_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \mathbf{0} \text{ subject to } \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} \neq \mathbf{0}$$

    (b) [5 points] Describe how the RANSAC algorithm could be used to fit a plane to these points in the scene.

    Solution: We can choose a randomized subset of consisting of three points from our point cloud and use them to form a triangle. Then, we can take the cross product of the triangle edge vectors to obtain a normal and the barycenter of the triangle to obtain an origin that together define a plane. Using this definition, we find all points that are within some threshold distance along the normal of the plane and denote these as the set of inliers. We repeat the process until the model with the most inliers has been found.

    (c) [5 points] Define a Hough transform based algorithm to fit a plane to these points in the scene. That is, define the dimensions of your Hough space, a procedure for mapping the scene points (i.e. the (x, y, z) coordinates for each pixel) into this space, and how the plane is determined.

    Solution: The Hough space is defined by the parameters $\alpha = a/d$, $\beta = b/d$, and $\gamma = c/d$. Then, each point corresponds to a plane $\alpha x + \beta y + \gamma z + 1 = 0$ in Hough space. We can partition our Hough space using a uniform 3D grid, find all Hough cells through which the plane corresponding to each point passes through, and then find the Hough cell with the most intersections. The center of this cell will correspond to a set of parameters that determine the plane's orientation.