CS 231A Computer Vision Midterm

Monday October 31, 2011

Solution Set

1 Multiple Choice (22 points)

Each question has **ONLY ONE CORRECT OPTION** and is worth **2 points**. To discourage random guessing, **0.5 points will be deducted** for a wrong answer on multiple choice questions! Please draw a circle around the option to indicate your answer. No credit will be awarded for unclear/ambiguous answers.

1. In Canny edge detection, we will get more discontinuous edges if we make the following change to the hysteresis thresholding:

   (a) increase the high threshold

   (b) decrease the high threshold

   (c) increase the low threshold

   (d) decrease the low threshold

   (e) decrease both thresholds

   Solution c

2. Mean-shift is a nonparametric clustering method. However, this is misleading because we still have to choose

   (a) the number of clusters

   (b) the size of each cluster

   (c) the shape of each cluster

   (d) the window size

   (e) the number of outliers to allow

   Solution d

3. Which of the following actions is often applied to eliminate aliasing?

   (a) low-pass filtering

(b) high-pass filtering
(c) subtracting out the image mean
(d) better focusing
(e) PCA

Solution a

4. If you are unsure of how many clusters you have in your data, the best method to use to cluser your data would be

(a) mean-shift
(b) k-means
(c) expectation-maximization
(d) markov random field
(e) none of the above are good methods

Solution a

5. Normalized cuts is an NP-hard problem. To get around this problem, we do the following:

(a) apply k-means as an initialization
(b) allow continuous eigenvector solutions and discretize them
(c) converting from a generalized eigenvalue problem to a standard one
(d) constraining the number of cuts we make
(e) forcing the affinities to be positive

Solution b or c

6. To decrease the size of an input image with minimal content loss, we should

(a) High-pass filter and down-sample the image
(b) Crop the image
(c) Apply a hough transform
(d) Down-sample the image
(e) Low-pass filter and down-sample the image

Solution e

7. When applying a Hough transform, noise can be countered by

(a) a finer discretization of the accumulator
(b) increasing the threshold on the number of votes a valid model has to obtain
(c) decreasing the threshold on the number of votes a valid model has to obtain

(d) considering only a random subset of the points since these might be inliers

Solution b

8. In which of the following scenarios can you use a weak perspective camera model for the target object?

    (a) A squirrel passing quickly in front of you.
    (b) An airplane flying at a very high attitude.
    (c) The Hoover tower when you are taking a photo of it right in front of it.
    (d) A car beside you when you are driving.

    Solution b

9. What is the biggest benefit of image rectification for stereo matching?

    (a) Image contents are uniformly scaled to a desirable size.
    (b) All epipolar lines intersect at the vanishing point.
    (c) All epipolar lines are perfectly vertical.
    (d) All epipolar lines are perfectly horizontal.
    (e) Epipoles are moved to the center of the image.

    Solution d

10. Which of the following factor does not affect the intrinsic parameters of a camera model?

    (a) Focal length
    (b) Offset of optical center
    (c) Exposure
    (d) Image resolution

    Solution c

11. What are the degrees of freedom of the essential matrix and why?

    (a) 5; 3 dof for rotation, 3 dof for translation. Up to a scale, so 1 dof is removed.
    (b) 6; 3 dof for rotation, 3 dof for translation.
    (c) 7; a 3×3 homogeneous matrix has eight independent ratios and $det(E) = 0$ removes 1 dof.
    (d) 7; 3 dof for rotation, 3 dof for translation. Up to a scale, so 1 dof is added.

    Solution a

2 Long Answer (40 points)

12. **(10 points) Fitting Circles with RANSAC**

RANSAC is a powerful method for a wide range of model fitting problems as it is easy to implement. In class, we've seen how RANSAC can be applied to fitting lines. However, RANSAC can handle more complicated fitting problems as well, such as fitting circles. In this problem we will solve for the steps needed to fit circles with RANSAC.

(a) (1 point) What is the minimum number of points we must sample in a seed group to compute an estimate for a uniquely defined circle?

Solution

3 points, as there are 3 degrees of freedom in a circle, the (x,y) coordinates of the center and the radius.

(b) (5 points) If we obtain a good solution that has few outliers, we want to refit the circle using all of the inliers, and not just the seed group. For speed purposes, we would like to keep the same center of the estimated circle from the seed group, and simply refit the radius of the circle. The equation for our circle is given as:

$$(x - x_c)^2 + (y - y_c)^2 = r^2$$

where (x_c, y_c) are the coordinates of the center of the circle, and r is the radius. The error we would like to minimize is given by:

$$E(x_c, y_c, r) = \sum_{i=1}^{m} (\sqrt{(x_i - x_c)^2 + (y_i - y_c)^2} - r)^2$$

where m is the number of inliers. Derive the new radius r that minimizes this least squares error function.

Solution

$$\frac{\partial E}{\partial r} = -2 \sum_{i=1}^{m} (\sqrt{(x_i - x_c)^2 + (y_i - y_c)^2} - r)$$

$$r = \frac{1}{m} \sum_{i=1}^{m} (\sqrt{(x_i - x_c)^2 + (y_i - y_c)^2})$$

(c) (4 points) One of the benefits to RANSAC is that we are able to calculate the failure rate for a given number of samples. Suppose we know that 30% of our data is outliers. How many times do we need to sample to assure with probability 20% that we have at least one sample being all inliers? You can leave your answer in terms of log functions.

Solution

$$1 - 0.2 = (1 - (0.7)^3)^k$$

$$k = \frac{\log(1-0.2)}{\log(1-(0.7)^3)}$$

13. **(10 points) Fundamental Matrix**

When you show your friend around Stanford campus, he takes a picture I of the Hoover tower. Later he wants to take a bigger picture of the Hoover tower, but his camera has no zoom-in function, so he walks forward for a short distance and takes a new picture I'. Assume there is only a forward translation of the camera perpendicular to the image plane and the movement distance is d.

(a) (5 points) Find the essential matrix F between I and I' in terms of d.

Solution
In general, $E = \hat{t}R$, where $\hat{t}$ is a skew-symmetric matrix related to translation vector. For a forward translating camera, we have

$$R = I, \hat{t} = \begin{bmatrix} 0 & -d & 0 \\ d & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

therefore,

$$E = \begin{bmatrix} 0 & -d & 0 \\ d & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

(b) (5 points) The tip of the Hoover tower on I is $p = (x\ y)$, what is its corresponding epipolar line l in I'? Express l in terms of x, y and d.

Solution
From $l = Ep$, the epipolar line for point $p = [x\ y\ 1]$ is

$$l = \begin{bmatrix} 0 & -d & 0 \\ d & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} -yd \\ xd \\ 0 \end{bmatrix}$$

14. **(10 points) Linear Filter**

In this problem, you will explore how to separate a 2D filter kernel into two 1D filter kernels. Matrix K is a discrete, separable 2D filter kernel of size $k \times k$. Assume k is an odd number. After applying filter K on an image I, we get a resulting image I_K.

(a) (1 point) Given an image point (x, y), find its value in the resulting image, $I_K(x, y)$. Express your answer in terms of I, k, K, x and y. You don't need to consider the case when (x, y) is near the image boundary.

Solution
$I_K(x, y) = \sum_{i=1}^{k} \sum_{j=1}^{k} K_{ij} I(x - i + \frac{k}{2}, y - j + \frac{k}{2})$

(b) (5 points) One property of this separable kernel matrix K is that it can be expressed as the product of two vectors $g \in \mathbb{R}^{k\times 1}$ and $h \in \mathbb{R}^{1\times k}$, which can also be regarded as two 1D filter kernels. In other words, $K = gh$. The resulting image we get by first applying g and then applying h to the image I is I_{gh}. Show that $I_K = I_{gh}$.

Solution

$$\begin{aligned}
I_K(x,y) &= \sum_{i=1}^{k}\sum_{j=1}^{k} K_{ij} I(x-i+\frac{k}{2}, y-j+\frac{k}{2}) \\
&= \sum_{i=1}^{k}\sum_{j=1}^{k} g_i h_j I(x-i+\frac{k}{2}, y-j+\frac{k}{2}) \\
&= \sum_{j=1}^{k} h_j \sum_{i=1}^{k} g_i I(x-i+\frac{k}{2}, y-j+\frac{k}{2}) \\
&= \sum_{j=1}^{k} h_j I(x, y-j+\frac{k}{2}) \\
&= I_{gh}(x,y).
\end{aligned}$$

(c) (4 points) Suppose the size of the image is $N \times N$, estimate the number of operations (an operation is an addition or multiplication of two numbers) saved if we apply the 1D filters g and h sequentially instead of applying the 2D filter K. Express your answer in terms of N and k. *Ignore the image boundary cases so you don't need to do special calculations for the pixels near the image boundary.*

Solution
For the 2D filter, there are k^2 multiplication operations and k^2-1 addition operations for each pixel. In total, $N^2(2k^2-1)$.

For each of the 1D filters, there are k multiplication operations and $k-1$ addition operations for each pixel. In total, $N^2(4k-2)$.

So the number of operations saved is $N^2(2k^2-4k+1)$

15. **(10 points) Perspective Projection**

In figure 1, there are two parallel lines l_1 and l_2 lying on the same plane Π. l_1' and l_2' are their projections through the optical center O on the image plane Π'. Let's define plane Π by $y=c$, line l_1 by equation $ax+bz=d_1$, and line l_2 by equation $ax+bz=d_2$.

(a) (3 points) For any point $P=(x,y)$ on l_1 or l_2, use the perspective projection equation below to find the projected point $P'=(x',y')$ on the image plane. f' is the focal length of the camera. Express your answer in terms of a, b, c, d, z and f'.

$$\begin{cases} x' = f'\frac{x}{z} \\ y' = f'\frac{y}{z} \end{cases} \tag{1}$$

Solution

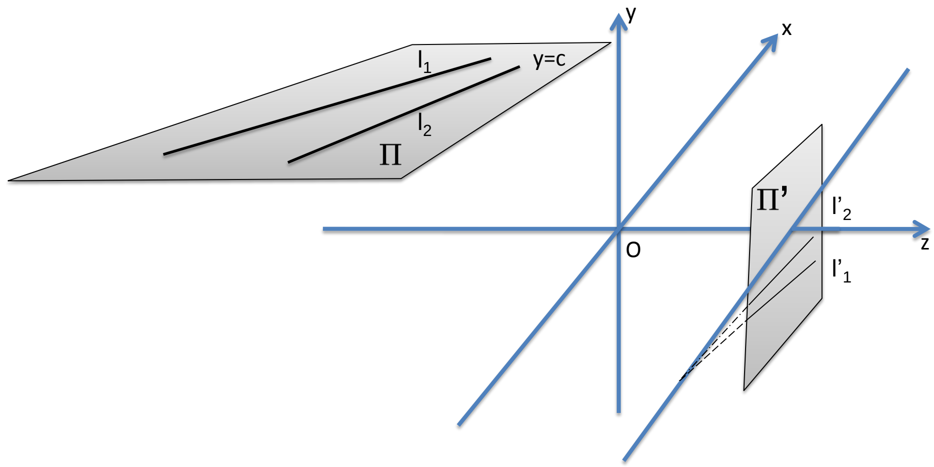

Figure 1: Perspective Projection

According to the perspective projection equation, a point on l projects onto the image point defined by

$$\begin{cases} x' = f'\frac{x}{z} = f'\frac{d-bz}{az}, \\ y' = f'\frac{y}{z} = f'\frac{c}{z}. \end{cases}$$

(b) (7 points) It turns out l_1' and l_2' appear to converge on the intersection of the image plane Π' given by $z = f'$ and the plane $y = 0$. Explain why.

Solution
This is a parametric representation of the image δ of the line Δ with z as the parameter. This image is in fact only a half-line since when $z \to -\infty$, it stops at the point $(x', y') = (-f'\frac{b}{a}, 0)$ on the x' axis of the image plane. This is the vanishing point associated with all parallel lines with slope $-\frac{b}{a}$ in the plane Π. All vanishing points lie on the x' axis, which is the horizon line in this case.

3 Short Answer (38 points)

16. **(5 points)** Given a dataset that consists of images of the Hoover tower, your task is to learn a classifier to detect the Hoover tower in new images. You implement PCA to reduce the dimensionality of your data, but find that your performance in detecting the Hoover tower significantly drops in comparison to your method on the original input

data. A sample of your input training images are given in Fig. 2. Why is the performance suffering?

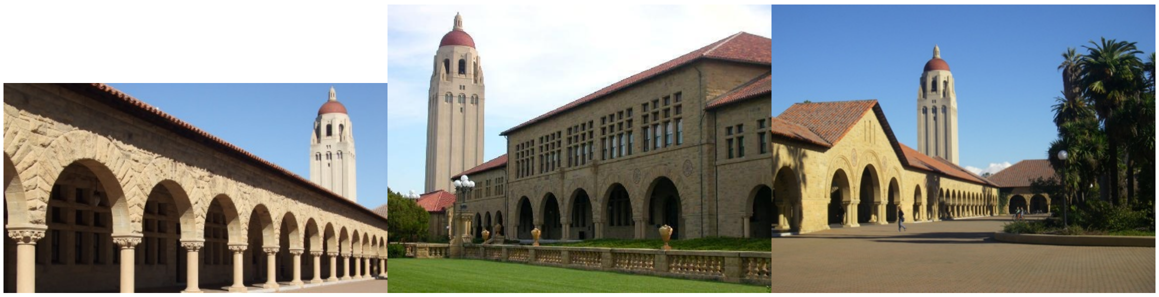

Figure 2: Example of input images

Solution

The Hoover tower in the images are not aligned, thus applying PCA to reduce the dimensionality here will not preserve the performance of the algorithms since we are trying to extract some signal out of the tower.

17. **(5 points)** You are using k-means clustering in color space to segment an image. However, you notice that although pixels of similar color are indeed clustered together into the same clusters, there are many discontiguous regions because these pixels are often not directly next to each other. Describe a method to overcome this problem in the k-means framework.

Solution

Concatenate the coordinates (x, y) with the color features as input to the k-means algorithm.

18. **(5 points)** To do face detection on your webcam, you implement boosting to learn a face detector using a variety of rectangle filters similar to the Viola-Jones detector. Some of the weak classifiers perform very well, resulting in near perfect performance, while some do even worse than random. As you are selecting your classifiers, you suddenly find that at a certain iteration k, the new classifier being selected and added in takes on a negative weight α_k in the final additive model. Explain why the negative weight appears, and justify your answer.

Solution

The negative weights appear because of the classifiers that perform worse than random. Their β_k value is greater than 1, causing the α_k value to be negative. An intuitive explanation for this is that we can invert the decision by a classifier that performs worse than chance to get a classifier better than chance.

19. **(5 points)** As shown in figure 3, a point Q is observed in a known (i.e. intrinsic parameters are calibrated) affine camera with image plane Π_1. Then you translate the camera parallel to the image plane with a known translation to a new image plane Π_2 and observe it again.

    (a) (2 points) Draw the image points Q'_1 and Q'_2 on Π_1 and Π_2 on the left figure of Fig. 3. Is it possible to find the depth of the 3D point Q in this scenario? Briefly explain why.

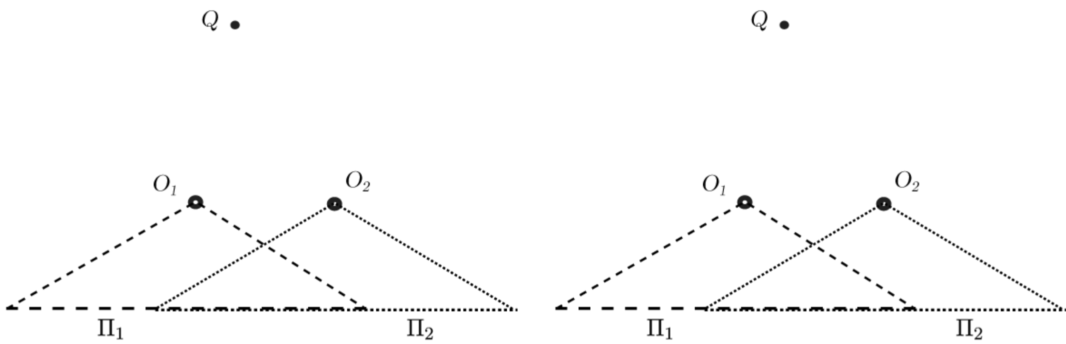

Figure 3: 3D point reconstruction

Solution

The solution of both parts is in figure 4.

No. We cannot determine point Q because it can be any 3D point on the line.

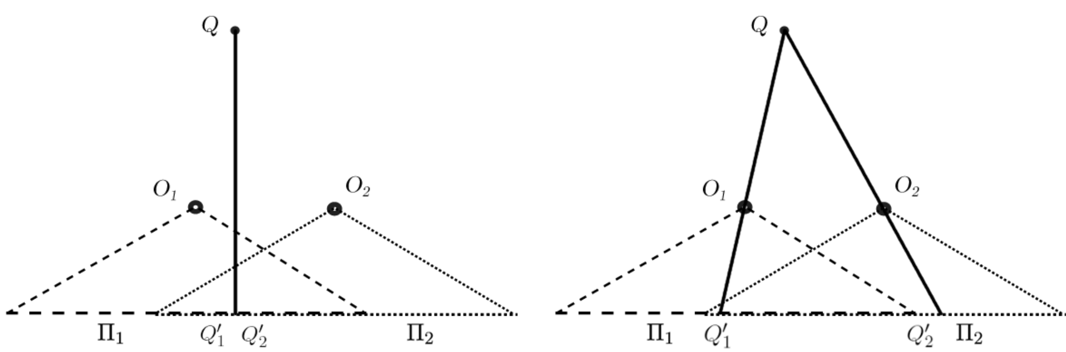

Figure 4: 3D point reconstruction

    (b) (3 points) What if this is a perspective camera? Draw Q'_1 and Q'_2 on the right figure of Fig. 3. Is it possible to find the depth of the 3D point Q in this scenario? Briefly explain why.

Solution

Yes, because we can do triangulation in this case.

20. **(6 points) Fundamental matrix estimation**

(a) (2 points) What is the rank of the fundamental matrix?

Solution
The rank is 2.

(b) (4 points) In the 8-point algorithm, what math technique is used to enforce the estimated fundamental matrix to have the proper rank? Explain how this math technique is used to enforce the proper matrix rank.

Solution
In the 8-point algorithm, SVD can be used to enforce the estimated F has rank 2. Specifically, we compute the SVD decomposition $F = U\Sigma V$, and we then zero out diagnals of Σ except for the two largest singular values to obtain $\tilde{\Sigma}$. We can reconstruct $F = U\tilde{\Sigma}V$.

Note: All of the following questions require you to specify whether the given statement is true or false and provide an explanation. No credit will be awarded without a valid explanation. Please limit your explanation to **1 − 2 lines**.

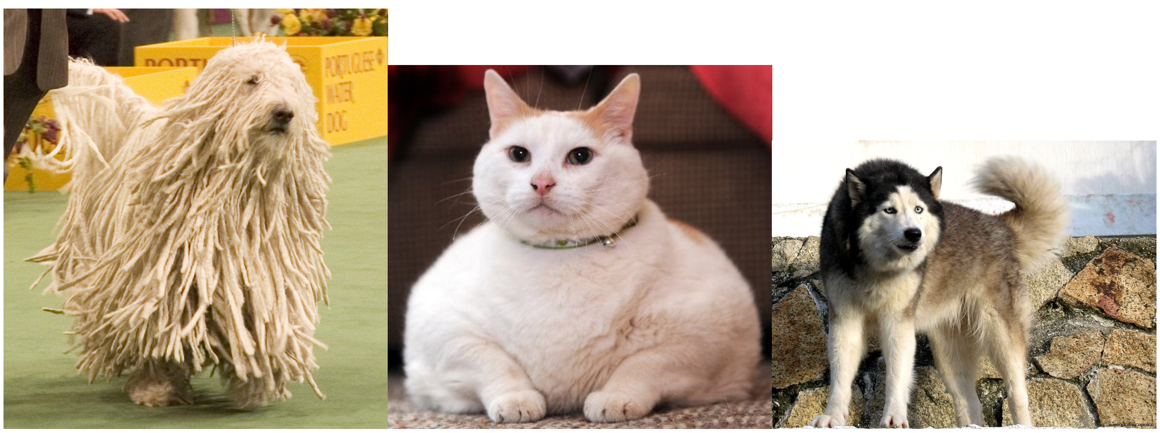

Figure 5: Example of input images

21. **(3 points)** True / False : Given a set of 3 images as shown in Fig. 5, finding and labeling the image in the center as "containing cat" is considered a *detection* task in recognition. Why or why not?

Solution
False. We are not localizing the cat in the image, so this is considered classification.

22. **(3 points)** True / False : When doing face recognition, we are given a vector of features (usually pixel values) as the representation for each of the images in our training set. In order to compute the eigenfaces, we concatenate each of these vectors together into a matrix, and then find the eigenvectors of this matrix. Why or why not?

Solution
False. We compute the eigenfaces by finding the eigenvectors of the covariance matrix.

23. **(3 points)** True / False : Both Eigenfaces and Fisherfaces are unsupervised methods. That is, they are able to operate on data without having to provide labels for the instances. Why or why not?

    Solution
    False. Eigenfaces are indeed an unsupervised method, but Fisherfaces utilize class labels in the formulation.

24. **(3 points)** True / False : If we initialize the k-means clustering algorithm with the same number of clusters but different starting positions for the centers, the algorithm will always converge to the same solution. Why or why not?

    Solution
    False. Different initializations will result in different clusters because they are local minima.