Practice Final

CMSC 426

The final will be cumulative for the whole semester. You should consider the 1st and 2nd Practice midterms as providing notes and practice for the final. In addition, the final will cover the following material, which we have discussed since the second midterm.

- Stereo Matching
  - Understand what a disparity map is. Understand constraints on matching, including the epipolar constraint, the ordering constraint, the non-negative disparity constraint, the uniqueness constraint, the photometric constraint (matching pixels have similar intensity) and the smoothness constraint (neighboring pixels have similar disparity).
  - Understand how dynamic programming can be used for stereo matching.
  - Understand how window-based correlation can be used for stereo matching.
- Motion Flow Fields
  - Understand the flow fields caused by simple camera motions. Eg., what is the flow field when the camera just rotates about the x, y, or z axis, or just translates.
  - Focus of expansion and its relationship to the epipole.
- Matching with Optical Flow
  - Understand the optical flow equation and how to use it to derive a constraint (linear equation) on the motion of a point.
  - Understand the Lucas-Kanade algorithm, and how it uses an assumption of constant flow in a small window to solve for the flow.
  - The aperture problem. When is this present? What is its relation to the Lucas-Kanade algorithm, and to corner detection.
- The Essential Matrix, and epipolar geometry.
  - Understand what the essential matrix means. How can it be used to determine whether two image points come from the same scene point. What is its relationship to the epipolar constraint?
  - Understand how to build the essential matrix when the camera translates but does not rotate.
  - Understand how to build an essential matrix when there is also rotation.
- Image coordinates and world coordinates. You should understand what we mean when we represent an image point with world coordinates, and when we represent an image with image coordinates, relative to the camera.
- 3D Rotation Matrices
  - Know that a 3D rotation matrix has rows (and columns) that are orthonormal.
  - Be able to construct a 3D rotation matrix for simple rotations.
- The final will NOT include other material on linear lighting and motion, object detection or object recognition.

This is also a good time to take stock of some themes that run throughout the course. There are a few main concepts, mostly mathematical, that underlie much of the material we have discussed:

- **Correlation and Convolution**
  These show up in many ways throughout the semester. First, we discussed how to use convolution to smooth an image, or to find image derivatives. Then, in the texture problem set, we saw that correlation has a close relationship with matching based on the sum-of-square distances. SSD matching shows up again when we are matching between images, as in stereo matching.
- **Image Gradients**
  Image gradients are a fundamental way in which we measure how an image is changing. Understanding how to compute an image gradient, and getting an intuition for its properties is very important. This includes, for example, understanding that the direction of the image gradient encodes the direction in which the image is changing most rapidly, while the magnitude of the gradient tells us how rapidly the image is changing. Image gradients are basic to edge detection, but they also show up in SIFT descriptors, and in optical flow.
- **Histograms**
  Histograms are a way of capturing the aggregate properties of an image, or a region of an image. They are also a way of building a probability distribution to model what is going on in an image. In addition to constructing a histogram from image information, we have also learned about smoothing histograms using Kernel Density Estimation (KDE) to build a probability distribution, and we have talked about using vector quantization to build a discrete histogram in a high-dimensional space. Histograms are used in image processing, such as in histogram equalization. Vector quantization is used for color quantization, and also to build histograms that describe textures, or for classification with bag-of-words approaches. Histograms are also used in building SIFT descriptors, while histograms combined with KDE are used for background subtraction.
- **Representing Motion with Matrices**
  We have talked about how to represent rotation and translation in a matrix. This includes understanding what a rotation matrix is and how to build one. We've also talked about affine transformations, similarity transformations, and scaled orthographic projection with matrices. 2D matrix operations are used to align images, as in mosaicing. 3D matrices are needed for 3D operations, including 3D structure-from-motion and the Essential matrix.
- **3D Geometry: Perspective projection and epipolar geometry**
  The core of this is understanding the pinhole camera model, and how it can be used to determine the relationship between a camera, a 3D point, and its 2D image. This is fundamental in any vision task that attempts to recover the 3D structure of the world. Next, we have talked about the relationship between a 3D scene and two cameras. This gives rise to epipolar geometry, which we have used to constrain stereo matching, understand flow fields and the focus of expansion, and build the Essential matrix.

Practice Problems

1. Stereo Matching
   a. Which of the following disparity maps violate the ordering constraint or the uniqueness constraint.

[0 0 1 1 1 2 2 2 2 3]

Uniqueness is violated. When one pixel has disparity of 0, and the next has disparity of 1, they both match the same point.

[0 0 X 1 1 X 2 2 X 3 3]

No constraints are violated.

[0 1 1 X X 3 3]

Uniqueness is violated for the same reason as the first example.

[X X 1 X X 5]

Ordering is violated. The point with disparity 1 matches the second pixel in the right image. The point with disparity of 5 matches the first point in the right image.

   b. Consider the following 1D images, assuming an occlusion cost and photometric cost as used in the problem set. What are the best matches, if we enforce the ordering constaint, the uniqueness constraint, and the non-negative disparity constraint? What if we enforce any two of these constraints? What if we enforce any one of these constraints? What if we enforce none?

L: [1 1]
R: [1 1]

Let's just do a couple of examples.

With no constraints violated, the disparity must be [0 0]

Allowing ordering to be violated doesn't add any possible matches.

With ordering and non-negative disparity violated, the disparity could also be [-1 1]

With uniqueness violated, we can have [0 1]

2. Motion Flow Fields
    a. How can you tell the difference between the motion flow caused by a translation in the x direction, and a rotation about the y axis?

With translation in the x direction, all flow vectors are horizontal. With rotation, flow vectors in the upper right part of the image will be mostly horizontal, but also point upwards a bit. Similarly, in the upper left, the will point downwards slightly. On the bottom also they will point away from the image center.

    b. Suppose a camera is translating in the direction (1,1,1). Draw the resulting flow field. How would this differ from the flow caused by a translation of (2,2,2)?

The flow vectors will point away from the point (1,1), which is the focus of expansion.

The flow will look similar, but since the camera is moving faster, the flow vectors will be twice as long. (However, if you don't know how far away points are, you won't know if the camera is moving faster, or the points are just closer.)

3. Matching with optical flow.
    a. Suppose a first image has intensities described by $H(x,y) = x^2+y^2$. A second image has intensities described by $H(x,y) = (x-1)^2+(y-1)^2$. Use the optical flow equation to determine a linear constraint on the optical flow at the points (7,3) and (3,7), and then assume that the flow is the same at both points and solve for the flow.

$H(7,3) = 58$. $I(7,3) = 40$. So I_t at that point is -18. The image gradient for I is (2x-1, 2y-1), so at (7,3) it is (13,5). So we get the equation $18 = 13u + 5v$. At the point (3,7) we also get I_t=-18, and a gradient of (5,13) so we get $18 = 5v + 13u$. Solving these, we can find $(u,v) = (1,1)$.

4. The Essential Matrix
    a. Suppose you see a point at (2,0) in the first image and at (4,0) in the second image. You see another point at (5,4) in the first image and at (5,2) in the second image. Assume the camera is translating, but not rotating.
        i. What are the flow vectors for each point? Where do they intersect? What is the focus of expansion? What does this tell us about the direction of the motion?

The flow vectors are (2,0) and (-2,0). They intersect at (5,0). This is the focus of expansion. This means that motion is along the vector (5,0,1).

        ii. Notice that the points are moving towards the focus of expansion. What does this tell us about the motion?

That we are moving backwards, in the direction (-5,0,-1).

iii. Use the direction of the motion to determine the Essential matrix.

$$\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & -5 \\ 0 & 5 & 0 \end{pmatrix}$$

iv. Use two equations derived from the points to determine the essential matrix.

$$(2,0,1)^T \begin{pmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{pmatrix} \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix} = 0$$

Gives us $-4T_y+8T_y=0$. *So* $T_y=0$

$$(5,4,1)^T \begin{pmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{pmatrix} \begin{pmatrix} 5 \\ 2 \\ 1 \end{pmatrix} = 0$$

Gives us $20T_z-5T_y-10T_z+2T_x+5T_y-4T_x=0$. *Combining these two equations we get:*

$10T_z-2T_x=0$. *Up to a scale factor, this gives us the solution (-5,0,-1).*

5. Coordinate systems.
   a. Suppose a camera has a focal point of (1,0,1) and an image plane of z = 2. Suppose a point appears in the image with world coordinates (3,4,2). What are its image coordinates?

The center of the image, in the image plane, will be at (1,0,2). So, relative to this, the coordinates will be (2,4).

   b. Suppose you have a camera with a focal point of (3,7,2) and an image plane of y = 6, and another camera with a focal point of (3,7, 12) and an image plane of y = 6. There is a point in the scene at the origin. Where does this appear in the two cameras, in image coordinates. What is the disparity?

A line from (0,0,0) to (3,7,2) has the equation (x,y,z) = t(3,7,2). To appear in the image, y = 6, we have t = 6/7 and the image appears at (18/7, 6, 12/7). The center of the image is at (3,6,2). The problem does not specify the directions of the x and y axis in the camera, but we will set them to be x and z. (That is, the camera has rotated around the x axis, so the x direction doesn't change, and the old y axis is now pointing in the direction *of the z axis. This causes the old z axis to point in the negative y direction, so that the image plane is in the negative y direction from the focal point. Then the image coordinates will be (-3/7, -2/7).*

When the focal point is at (3,7,12), (x,y,z) = t(3,7,12), and the point appears at (18/7, 6, 72/7). The camera center is (3,6,12), and the points image coordinates are (-3/7, -12/7). The epipolar lines are horizontal in this case, and the disparity is 10/7. Note that the baseline is 10, so this gives us the correct depth of (10/7)/10 = 7.

6. 3D rotations
   a. Which of the following matrices represent 3D rotations? Explain your answer for each.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

This is a rotation, just the identity transformation, meaning a rotation of 0 degrees.

$$\begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

Also a rotation. Note that the rows are orthonormal.

$$\begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 \end{pmatrix}$$

No. The second and third rows are not orthogonal.

$$\begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ -\sqrt{2}/\sqrt{3} & 1/\sqrt{6} & 1/\sqrt{6} \end{pmatrix}$$

Yes, the rows are orthonormal. (By the way, the determinant of this matrix is 1. If the rows are orthonormal, the determinant could be -1, which would mean that there is a rotation and a reflection.)