Final Exam

ECE/CS 181b

June 12, 2008

Name: ______________________________

This is a closed book/notes examination. Calculators and other devices with memory are not allowed.

Instructions: All questions on this exam are weighted equally. To receive full credit, answer the following questions clearly and completely, providing analytical derivations when required. If the space provided is insufficient, you may use extra paper, but make sure your name and the question number is on it.

For this exam, we will adopt the following conventions (READ CAREFULLY):

- Boldface letters indicate points ($\mathbf{x}$), lines ($\mathbf{l}$), vectors ($\mathbf{p}$) or matrices ($\mathbf{H}$).
- All the quantities related to the second camera are identified by the superscript *prime* ($'$).
- The notation $\mathbf{H} : A \mapsto A'$ means a matrix $\mathbf{H}$ which maps A to A'.
- The notation $\mathbf{x} \leftrightarrow \mathbf{x}'$ means that $\mathbf{x}$ and $\mathbf{x}'$ are corresponding elements (points, lines, etc.) in the two image planes.

QUESTION 1: Edge Detection.
This question is about the Laplacian of the Gaussian operator, used in the context of edge detection:

1. Describe how you would use the LoG operator to detect edges in images.
   Answer: *Mainly two-step process. First the image is filtered with a second derivative of a gaussian (the filter is isotropic, therefore the computation can be performed efficiently in the frequency domain). Second the zero crossings of the output are identified as edge points.*

2. Describe the LoG operator in terms of the order of the derivatives that it computes. Comment about the main advantages and disadvantages of this.
   Answer: *LoG is a second derivative type of operator. Advantages: good localization. Disadvantages: sensitive to noise.*

3. What parameter has to be tuned by the user?
   Answer: *LoG requires to set the σ, that is the scale of the Gaussian blurring.*

4. What is the Laplacian Pyramid? Why is it used and how it is constructed in practice?
   Answer: *The Laplacian pyramid is a pyramid of the outputs of LoG filters at increasing scales. It is used to detect edges at multiple scales. Usually it is approximated by the DoG pyramid.9*

QUESTION 2: Camera Projection.

1. Under what conditions will a line viewed with a pinhole camera have its vanishing point at infinity?
   Answer: *The line has to be parallel to the image plane.*

2. A scene point at coordinates $(400, 600, 1200)$ is perspectively projected into an image at coordinates $(24, 36)$, where both coordinates are given in millimeters in the camera coordinate frame and the cameras principal point is at coordinates $(0, 0, f)$ (i.e. , $u_0 = 0$ and $v_0 = 0$). Assuming the aspect ratio of the pixels in the camera is 1, what is the focal length of the camera? (Note: the aspect ratio is dened as the ratio between the width and the height of a pixel; i.e. , k_u/k_v.)
   Answer: $u = \frac{fx}{z}$ *so* $f = \frac{uz}{x} = 24 * 1000/400 = 72$ *mm.*

QUESTION 3: SIFT.
Describe in details the following steps of the SIFT feature matching procedure

1. Detection of the keypoints
2. Orientation assignment to each keypoint.
3. Creation of the 128 dimensional descriptor.
4. Descriptor matching.

~~**QUESTION 4**: Optical Flow~~

Let $I(x, y, t)$ be a video sequence taken by a rigidly moving camera observing a rigid, static and Lambertian scene. In class we discussed the brightness constant equation:

$$I(x + u, y + v, t + 1) - I(x, y, t) = 0$$

which yields the equation used to estimate the optical flow:

$$I_x u + I_y v + I_t = 0$$

Let us consider a slightly different case. Assume that between two consecutive views there is an affine change in the image intensities, i.e. the brightness constancy constraint reads:

$$I(x + u, y + v, t + 1) = aI(x, y, t) + b \tag{1}$$

where $u(x, y)$ and $v(x, y)$ represent the optical flow (motion parameters) and $a(x, y)$ and $b(x, y)$ represent photometric parameters. Propose an algorithm for estimating (u, v, a, b) from the image brightness I and its spatio-temporal derivatives I_x, I_y, I_t. Clearly explain the steps. What is the minimum size of a window around each pixel that allows one to solve the problem?

Answer: *We can rewrite (1) as:*

$$I(x + u, y + v, t + 1) - I(x, y, t) = aI(x, y, t) - I(x, y, t) + b$$

or

$$I_x u + I_y v + I_t - aI(x, y, t) + I(x, y, t) - b = 0$$

which can be written in matrix form as:

$$\begin{bmatrix} I_x & I_y & I & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 - a \\ -b \end{bmatrix} = I_t$$

Since we have four unknowns we need to stack 4 or more of these equations and then solve the correspondent overdetermined system in a least square sense.

QUESTION 5: PCA for Face Recognition

1. What is the purpose of using PCA during the face recognition based on eigen-faces?

2. PCA is a linear transformation, in the sense that we project the data on a sub-space using a set of basis vectors. What is the main characteristic of these vectors that makes PCA different w.r.t. other linear transformations and optimal in least square sense?

3. To compute the basis vector we need to compute the eigenvectors of the covariance matrix $\mathbf{C} = \mathbf{A}\mathbf{A}^{\mathbf{T}}$, which most of the times is unfeasible. We compute instead the eigenvectors of $\mathbf{A}^{\mathbf{T}}\mathbf{A}$, which is much smaller. Show what is the relationship between the eigenvectors of $\mathbf{A}^{\mathbf{T}}\mathbf{A}$ and the ones of $\mathbf{A}\mathbf{A}^{\mathbf{T}}$.

~~**QUESTION 6**: Shape from Shading~~

1. Give the definition of *Lambertian* and *Specular* surfaces?
2. Explain why the full moon looks more like a disk rather than a sphere.
3. What is photometric stereo?
4. What are the differences between photometric stereo and traditional binocular stereo?

QUESTION 7: Epipolar Geometry

Consider are three image planes A_1, A_2 and A_3, with known homography matrices $\mathbf{H_{31}}$: $B_1 \rightarrow B_3$ and $\mathbf{H_{32}}$: $B_2 \rightarrow B_3$. Assume also you know the position of the three epipoles $\mathbf{e_1}$, $\mathbf{e_2}$ and $\mathbf{e_3}$.

- Compute the expression of the two epipolar lines, $\mathbf{l_2}$ and $\mathbf{l_3}$, corresponding to the known point $\mathbf{x_1}$ (Note you can do this without known the two fundamental matrices $\mathbf{F_{21}}$ and $\mathbf{F_{31}}$).
- Derive now the expression for $\mathbf{F_{21}}$ and $\mathbf{F_{31}}$ exploiting your results from the previous point.

~~**QUESTION 8**: Optical Flow~~

Consider a camera moving along the optical axis toward a planar surface at right angles to the optical axis. Assume that the velocity of the camera is W and Z is its distance to the plane.

- Compute the optical flow $(u(x, y), v(x, y))$ in the image.
- Is the optical flow stationary (that is, independent of time)?
- Is the Laplacian of the optical flow zero?
- How could you predict the time to impact (Note that W and Z are not known, you have only (u, v) as observables?

QUESTION 9: Multiview Stereo

- Prove that the pixel *disparity* of a 3D scene point is inversely proportional to the distance of the scene point from the camera.
- In trinocular stereo, one can use the epipolar constraint to obtain accurate and robust matches. If you add one more camera (i.e. a total of four cameras), how will it help/affect the matching algorithm? Discuss the advantages and disadvantages between the two configurations.

QUESTION 10: Image Geometry and Photometry

- Suppose that an image is created by a camera in a certain world. Now imagine the same camera placed in a similar world in which everything is twice as large and all distances between objects have also doubled. Compare the new image with the one formed in the original world. Assume perspective projection.
- Suppose that an image is created by a camera in a certain world. Now imagine the same camera placed in a similar world in which everything has half the reflectance and the incident light has been doubled. Compare the new image with the one formed in the original world.

QUESTION 11: SIFT

- What is the advantage in using local features like SIFT instead of global description of the image like the one obtained using eigenfaces?
- What is the idea behind the multi-resolution approach used to build the Laplacian pyramid? What problem are we trying to solve by doing that?
- Why are we interested in the eigenvalues of the image Hessian to prune out the key point candidates? What information is contained in those eigenvalues?
- You know that once the SIFT feature vector for a keypoint is found, it is normalized to have unit norm. Which desirable property of SIFT comes about due to this feature vector normalization step?