Examination #1

CS 766: Computer Vision

October 21, 2004

Last (Family) Name: ______________________________

First Name: ____________________________________

Problem	**Score**	**Out of**
1	______	15
2	______	15
3	______	20
4	______	10
5	______	15
6	______	10
Total	______	85

1. [15] **Camera Calibration**
The relationship between a 3D point at world coordinates (X,Y,Z) and its corresponding 2D pixel at image coordinates (u,v) can be defined as a projective transformation using a 3×4 camera projection matrix **P**.

(a) [3] Can the matrix **P** incorporate any lens distortions that might be in the camera? Briefly explain.

(b) [4] Give two lists, one specifying the intrinsic camera parameters and the other giving the extrinsic camera parameters.

(c) [4] Show how **P** can be decomposed into a product of matrices that contain elements expressed in terms of the intrinsic and extrinsic camera parameters.

(d) [4] Give the main steps of an algorithm for computing the matrix **P** from a single image of a known 3D "calibration object."

2. [15] **Planar Transformations**
The planar façade of a building is captured in an image taken by a camera. Assume this plane corresponds to the world coordinate frame's Z=0 plane, and scene point (X,Y) on the building projects to image pixel coordinates (u,v).

(a) [3] What is the *planar projective transformation* that describes the relationship between (X,Y) and (u,v)? Give your answer using homogeneous coordinates.

(b) [1] How many degrees of freedom does this transformation have?

(c) [1] How many point correspondences are required to determine this transformation?

(d) [2] Would having more correspondences than your answer to (c) be helpful in any way? If no, briefly explain why not. If yes, explain how they could be used.

(e) [2] Give one invariant of a planar projective transformation.

(f) [2] Give one invariant of a planar affine transformation that is not an invariant for a planar projective transformation.

(g) [4] If the building has sets of lines on it running parallel to both the X and Y axes, how could we use the corresponding lines in the image to determine if the building plane is parallel to the image plane?

3. [20] **Edge Detection**

(a) [5] Show how an approximation to the first derivative of an image can be obtained by convolving the image with the kernel [1 −1] where the image is defined as

[56 64 79 98 115 126 132 133]

Ignore computing a value for the first and last image pixels (in other words, your result will be 6 values). In addition to showing the result of the convolution, indicate where edges would be detected and why.

(b) [6] What property of the coefficients of a kernel ensures that an appropriate output is obtained for regions of constant intensity in an image when

(i) The kernel is approximating a first derivative.

(ii) The kernel is approximating a second derivative.

(iii) The kernel is approximating a Gaussian.

(c) [4] Describe a major advantage or disadvantage of using an isotropic operator instead of a non-isotropic operator with respect to the following issues:

(i) Computational efficiency.

(ii) Noise in the image.

(d) [5] What is the purpose of (i) non-maxima suppression and of (ii) hysteresis that are done in the Canny edge detector?

4. [10] **Corner Detection**

(a) [2] When would detecting corners be more appropriate than detecting edges as an initial step in an application using computer vision?

(b) [4] The Harris corner detection algorithm computes a 2×2 matrix at each pixel based on the first derivatives at that point and then computes the two eigenvalues of the matrix, λ_1 and λ_2, where $\lambda_1 \leq \lambda_2$. How can these two values be used to label each pixel as either a locally smooth region (S), an edge point (E), or a corner point (C)? Give your answer by specifying "Label pixel S if …", "Label pixel E if …" and "Label pixel C if …"

(c) [4] Given the two eigenvalues specified in (ii), explain in English the rationale and difference(s) between detecting corner points using the criterion $\lambda_1 > T_1$ (which is used in the Tomasi and Kanade algorithm) versus the criterion $\lambda_1\lambda_2 > T_2$ (which is used in the Harris algorithm), where T_1 and T_2 are appropriate thresholds.

5. [15] **Active Contours**

The energy functional that is used with active contours (snakes) usually contains the three terms: $E_{continuity}$, $E_{smoothness}$, and E_{image} (the first two terms are often combined to define a term called E_{int}).

(a) [9] For each of these three terms explain briefly what it measures, how it is defined, and what happens if the term is omitted.

(b) [3] What is the effect of giving a negative weight to $E_{continuity}$?

(c) [3] How could the energy functional definition be specified so as to cause the contour to expand?

6. [10] **Segmentation using Normalized Graph Cut**

(a) [3] Define cut(A,B) and explain intuitively what it measures.

(b) [5] Say we want to find “regions” corresponding to object boundaries by grouping connected chains of “strong” edge points that are adjacent and their gradients imply a smooth contour. Define an affinity measure for this purpose that uses the magnitude, mag($\nabla I(p)$), and direction, direc($\nabla I(p)$), of the gradient, ∇I, at a pixel p in image I to compute $\text{aff}_{\text{edge}}(x,y)$ for a pair of adjacent pixels x and y. Include any additional computational steps before computing the affinity that make sense for obtaining good results. Briefly explain your definition.

(c) [2] Say we want to segment images into regions of uniform texture using a measure of the texture at each pixel by computing some function over a $(2n+1) \times (2n+1)$ neighborhood centered on the pixel. Briefly explain why using this function to define the texture affinity between two pixels, $\text{aff}_{\text{texture}}(x,y)$, will be biased depending on how close the pixels are to a region boundary.