PLEASE HAND IN

PLEASE HAND IN

UNIVERSITY OF TORONTO
Faculty of Arts and Science

DECEMBER EXAMINATIONS 2009

CSC 487/2503H1F
Instructors: A. Jepson and D. Fleet
St. George Campus

Duration — 3 hours

Aids allowed: none

Student Number: |_|_|_|_|_|_|_|_|_|

Family Name: ______________________________

Given Name: ______________________________

*Do **not** turn this page until you have received the signal to start.*
(In the meantime, please fill out the identification section above, and read the instructions below.)

# 1:	______/15
# 2:	______/16
# 3:	______/15
# 4:	______/12
# 5:	______/14
# 6:	______/14
TOTAL:	______/86

This examination consists of 6 questions on 18 pages (including this one). *When you receive the signal to start, please make sure that your copy of the examination is complete.* If you need more space for one of your solutions, use the reverse side of the page and *indicate clearly the part of your work that should be marked.* Also, the last page is provided for scratch work.

Write your student number at the bottom of pages 2-18 of this test.

Good Luck!

Question 1. [15 MARKS]

Part (a) [3 MARKS] Principal Component Analysis (PCA) is often applied to reduce the dimensionality of input data to help mitigate problems with overfitting that are common with small datasets. In some cases, however, this use of PCA can make recognition more difficult than it would have been in the original feature space. Explain how this can happen with a simple example.

Part (b) [3 MARKS] State three key factors that make category recognition difficult in computer vision, and briefly explain the reasons why.

Question 1. (CONTINUED)

Part (c) [3 MARKS] Explain two key benefits of AdaBoost over Logistic Regression for binary classification?

Part (d) [3 MARKS] In what sense is AdaBoost a greedy algorithm? Why is this necessary?

Part (e) [3 MARKS] Say you want a feature point detector for detecting long range motion and recoverying camera and scene geometry. Give three key properties that it should have (i.e., what makes a good feature detector)?

Question 2. [16 MARKS]

Suppose you are asked to stabilize a short image sequence using an optical flow algorithm to compute the motion. But the motion in the image sequence looks fast and choppy, and a friend tells you it appears to be temporally aliased. Despite this, you first implement a gradient-based flow estimator.

Part (a) [2 MARKS] What is the gradient constraint equation? Define all variables introduced.

Part (b) [3 MARKS] Given two frames of the image sequence I_0 and I_1, and the gradient constraint equation, define an objective function for a least-squares, gradient-based estimator for translational motion.

Part (c) [2 MARKS] Because of the temporal aliasing you find that your LS approach often returns very poor estimates of translation. In a couple of sentences and/or with a figure, explain what is likely causing the problem?

Question 2. (CONTINUED)

Part (d) [4 MARKS] Describe two ways that you might be able to overcome the problem. Use just two or three sentences for each approach.

Part (e) [1 MARK] Assuming you had control over the aperture size and shutter speed of the camera, but not frame rate, how could you reduce the amount of temporal aliasing?

Part (f) [2 MARKS] The human eye has relatively little temporal aliasing, but most video cameras do. Why do we design cameras which produce so much temporal aliasing?

Part (g) [2 MARKS] For a computer vision system, would you want to use a camera that produced temporal aliasing? Why or why not?

Question 3. [15 MARKS]

It is the season for small pieces of fruit embedded in jelly! Suppose we have an orthographic image sequence in which the jelly and the fruit pieces fixed within it are deforming in an affine manner. Suppose we can track K fruit pieces. That is, for each frame t, and each piece of fruit located at $\vec{X}_k(t) \in \Re^3$, we have

$$\vec{X}_k(t) = B(t)\vec{X}_k(0), \tag{1}$$

for some 3×3 invertable matrix $B(t)$. Here, for convenience, we assume $\sum_{k=1}^{K} \vec{X}_k(0) = \vec{0}$. Suppose the internal calibration matrix for each frame t is known to be

$$M_t = s(t) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \tag{2}$$

where $s(t)$ is a positive scale factor. (In particular, we can absorb the rotation matrix $R(t)$, which is normally part of the intrinsic calibration matrix, into the transform matrix $B(t)$.) The projected image locations are then given by $\vec{x}_k(t) = M_t\vec{X}_k(t)$. Finally, suppose we track the image locations $\vec{x}_k(t)$ of the pieces of fruit, for $1 \le k \le K$ and for integer frame numbers t, $1 \le t \le T$.

Part (a) [3 MARKS] Consider the data matrix D formed from this image data in the same way as for orthographic factorization. Clearly describe how this matrix is constructed from the tracking data. In the absence of tracking noise, what is the maximal rank of D? Explain in detail.

Question 3. (CONTINUED)

Part (b) [2 MARKS] Suppose we attempt to use the metric upgrade procedure in the lecture notes, despite the fact that it was derived for rigid 3D motion, not 3D affine motion as we have here. What constraints on the camera matrix were used to compute this (rigid) metric upgrade?

Part (c) [2 MARKS] In class we obtained an equation for a 3×3 symmetric matrix $Q = AA^T$ which imposed the constraints on the camera matrix identified in part (b) above. How exactly was this Q computed in the rigid case? Briefly describe how, and under what conditions, A can be computed from Q (in the rigid case).

Question 3. (CONTINUED)

Part (d) [6 MARKS] Suppose we can factor the matrix Q from part (c) in the form $Q = AA^T$ for some real-valued 3×3 matrix A. Answer the following four questions:

1. Briefly describe how the matrix A is applied to the factors of D in the rigid case to obtain a metric upgrade. What is the resulting shape matrix, and what are the reconstructed camera matrices for each frame t?
2. Now suppose we have computed Q and A in exactly the same way as for the rigid case, but here the data is taken from the affinely deforming jello described earlier. Given the resulting reconstructed shape and camera matrices, would you expect large reconstruction errors of the image points $\vec{x}_k(t)$ for noiseless image data? Explain.
3. In what sense is the shape matrix a reconstruction for the initial points $\{\vec{X}_k(0)\}$? Explain.
4. How well do the camera matrices fit the model the model $M(t)B(t)$ above? Explain.

More space on next page.

Question 3. (CONTINUED)

Part (e) [2 MARKS] Given noiseless tracking, briefly summarize how we can distinguish 3D affine motion (as in the jello case) from 3D rigid motion (e.g., frozen jello).

Question 4. [12 MARKS]

Suppose we are running a particle filter and, at the current frame t, we approximate the filtering distribution $p(\vec{x}_t|\vec{z}_{1:t})$ by a weighted set of samples $\{(\vec{x}_k, w_k)\}_{k=1}^K$. Here the weights $w_k > 0$ sum to one.

Part (a) [1 MARK] Estimate the mean of the filtering distribution using the set of samples $\{(\vec{x}_k, w_k)\}_{k=1}^K$.

Part (b) [2 MARKS] Show that the estimate of the mean in part (a), say $\vec{m}$, minimizes $E_S[||\vec{x} - \vec{m}||^2]$ where $||\cdot||$ denotes Euclidean length (2-norm), and $E_S[\cdot]$ denotes expected value over the sampled distribution represented by $\{(\vec{x}_k, w_k)\}_{k=1}^K$. That is, $E_S[||\vec{x} - \vec{m}||^2] = \sum_{k=1}^K w_k ||\vec{x}_k - \vec{m}||^2$.

Question 4. (continued)

Part (c) [7 marks] For reasons that were made clear in assignment 3, we may wish to identify different peaks in the filtering distribution $p(\vec{x}_t|\vec{z}_{1:t})$. Following the approach of part (b) above, consider minimizing $\mathcal{O}(\vec{m})$ defined as:

$$\mathcal{O}(\vec{m}) = E_S[\rho(||\vec{x} - \vec{m}||)] = \sum_{k=1}^{K} w_k \rho(||\vec{x}_k - \vec{m}||). \qquad (3)$$

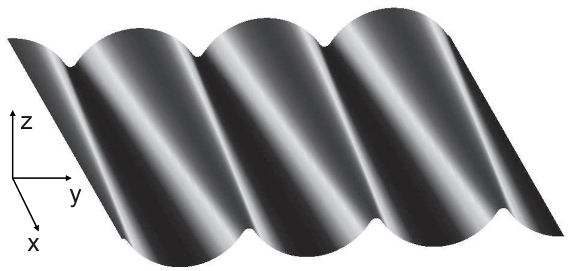
Using the Geman-McLure estimator $\rho(e) = e^2/(e^2 + \sigma^2)$ derive an iteratively reweighted least squares algorithm for locally minimizing $\mathcal{O}(\vec{m})$

Part (d) [2 marks] Where else in this course have we discussed a similar update equation using samples from a discretely represented distribution?

Question 5. [14 MARKS]

Here we consider the Lambertian shading model. Throughout this question suppose that the camera has been calibrated so that when a uniform radiance distribution is imaged, the resulting image is a positive constant (i.e., we remove any roll-off of the intensities towards the sides and corners of the image).

Suppose we have a surface $z(x, y)$ with continuous first derivatives, such as the one depicted below (which is **not** shaded with a Lambertian model to make the shape more apparent), and the camera is located far above the surface looking down (i.e., the optical axis is in the $-z$ direction, which is a different view than the one below).



Part (a) [4 MARKS] Given $z(x, y)$, the direction $\vec{L}$ to a distant light source, and the viewing direction $\vec{V}$, what is the reflected radiance distribution (per unit cross-sectional area perpendicular to $\vec{V}$) that is scattered from this Lambertian surface from the neighbourhood of the point $(x, y, z(x, y))$? Define all the terms you use, along with their physical units.

Part (b) [2 MARKS] Suppose the surface is known to be a plane and the only unknowns in the equation in part (a) are the position and orientation of this plane. Discuss exactly which aspects of this plane are determined by the equation in part (a).

Question 5. (CONTINUED)

Part (c) [2 MARKS] Now suppose we fold a uniformly painted Lambertian plane to form a non-planar, continuous, piecewise planar surface. This surface consists of just the two half-planes joined along a single line. Suppose both sides of this fold are clearly visible in the image. Given the brightness calibration above, is it possible for such a surface to generate a constant (non-zero) image when lit with a single distant light source? Describe the constraints on the normals for the two planes that make this possible or impossible.

Part (d) [3 MARKS] Give a simple concrete example of a smoothly curved (non-planar), uniformly painted, Lambertian surface which, when lit from a particular direction, generates a constant (non-zero) image. Explain.

Question 5. (CONTINUED)

Part (e) [3 MARKS] Describe how the surface depicted in the figure displayed at the beginning of this question could generate a **constant**, strictly positive, image when it is lit with a single distant light source in a particular direction L, and when the surface reflectance is Lambertian with a constant, positive albedo. (As always, we assume the image brightness has been calibrated as described above.) You can ignore any inter-reflection. If it is not possible to generate a constant image in this manner, explain. If it is possible, explain what you think the surface might be and what light source direction you would have to use.

Question 6. [14 MARKS]

In the context of Bayesian filtering let $\mathbf{x}_t$ and $\mathbf{z}_t$ denote the state and observation at time t, and let the state and observation histories be $\mathbf{x}_{1:t}$ and $\mathbf{z}_{1:t}$. The recursive filtering equations are then given by

$$p(\mathbf{x}_t \,|\, \mathbf{z}_{1:t}) = c\, p(\mathbf{z}_t \,|\, \mathbf{x}_t)\; p(\mathbf{x}_t \,|\, \mathbf{z}_{1:t-1}) \tag{4}$$

$$p(\mathbf{x}_t \,|\, \mathbf{z}_{1:t-1}) = \int_{\mathbf{x}_{t-1}} p(\mathbf{x}_t \,|\, \mathbf{x}_{t-1})\; p(\mathbf{x}_{t-1} \,|\, \mathbf{z}_{1:t-1})\, d\mathbf{x}_{t-1} \tag{5}$$

Recall that in a particle filter we maintain a weighted set of samples in order to represent a particular probability distribution, say $p(\mathbf{x})$, and that these weighted samples can be used to provide an approximation for $\int g(\mathbf{x})p(\mathbf{x})d\mathbf{x}$, i.e., the expected value of some test function $g(\mathbf{x})$ under the distribution $p(\mathbf{x})$.

Part (a) [2 MARKS] State mathematically and name the two key assumptions that permit one to express the filtering equation in the recursive fashion given in Equations (4) and (5).

Part (b) [2 MARKS] Derive a simple expression for c in terms of $p(\mathbf{z}_t \,|\, \mathbf{x}_t)\; p(\mathbf{x}_t \,|\, \mathbf{z}_{1:t-1})$.

Part (c) [2 MARKS] In a particle filter where one draws samples from $p(\mathbf{x}_t \,|\, \mathbf{z}_{1:t-1})$, in theory what are the exact *importance weights* required to ensure a properly weighted sample set for $p(\mathbf{x}_t \,|\, \mathbf{z}_{1:t})$? In one sentence briefly justify why this is the case. (Hint: You need c.)

Question 6. (CONTINUED)

Part (d) [4 MARKS] In a practical particle filter how are the importance weights typically approximated? Justify this approximation (e.g., based on your answers to parts (b) and (c)).

Part (e) [2 MARKS] Under what circumstances might the approximation for c be poor? What are the potential consequences of this circumstance for a particle filter?

Question 6. (CONTINUED)

Part (f) [2 MARKS] In your assignment 3, when trying to maintain a multimodal distribution, you found that, invariably, all particles tend to migrate to a single mode. If we knew c, in absolute terms, explain why you might be able to determine when significant modes of the filtering distribution are not well represented.

For scratch work.

Total Marks = 86