Part A:

Question 1:

(a) The coefficients of a digital filter designed for differentiation should sum to zero.

Solution: True. Think of a function that is constant. It's derivative is everywhere zero. This will be true only if the coefficients of the digital filter sum to zero.

(b) Ignoring image noise, all intensity profiles across edges in images of polyhedral scenes, where the objects are composed only of planar faces, are simple step functions.

Solution: False. There are many situations, even in the case of polyhedral scenes, where intensity profiles across edges are not simple step functions. For example, as in Assignment 5, inter-reflection at concave intersections of planar faces can create "roof" type intensity profiles.

(c) In edge detection, an operator that is optimal for detecting the presence of an edge will also be optimal for locating the position of that edge.

Solution: False. "Detection" and "localization" are conflicting design criteria. Doing better at detection eventually comes at the expense of localization.

(d) The number of zero-crossings of the second derivative of a function $f(x)$ filtered by a Gaussian with parameter σ is a monotonically decreasing function of σ.

Solution: True. As σ increases, no new zero crossings of the second derivative are created.

(e) If a problem is well-posed, in the sense of Hadamard, then there exists a numerically stable algorithm for computing the unique solution.

Solution: False. Well-posed, in the sense of Hadamard, means a unique, stable solution exists. But, this does not imply the existence of an algorithm to compute the solution.

(f) Physical optics suggests that there are two kinds of reflectance, diffuse (i.e., Lambertian) and specular (i.e., mirror-like). Thus, in physics-based computer vision, it is sufficient to model reflectance as a linear combination of diffuse and specular components.

Solution: False. In physics-based computer vision, one must consider what happens at the surface microstructure scale (i.e., at resolutions higher than the particular sensor can resolve as distinct). At the microstructure scale, surface roughness can, by way of inter-reflection and other phenomema, create apparent reflectance that can not be modeled at the macro scale as a linear combination of diffuse and specular components.

(g) A Lambertian material has reflectance proportional to $cos(i)$.

Solution: False. Reflectance of a Lambertian material depends also on the illumination. As we have seen, a Lambertian material under uniform hemispherical illumination has reflectance proportional to $\frac{1+cos(e)}{2}$.

(h) The reflectance map, $R(p, q)$, generalizes the definition of the bidirectional reflectance distribution function (BRDF).

Solution: False. The reflectance map, $R(p, q)$, is compiled information. It specializes the BRDF of a given material to a particular illumination and viewing situation.

(i) In photometric stereo, one often uses paint (or other surface coating) to match reflectance properties between a calibration sphere and objects whose shape is to be analyzed. One could use a white paint or a darker (but otherwise identical) grey paint. For a given configuration of light sources, the identical measurements would be obtained in either case, provided only that the light sources were made appropriately brighter when using the darker grey paint. [Hint: Recall Assignment 5].

Solution: False. As we have seen in Assignment 5, the reflectance factor, ρ, acts in a non-linear way to, among other things, alter the intensity profile across concave edges. Thus, the intensity patterns in concave regions will differ, between white and grey painted objects, in ways that can not be equalized by simple adjustment of the relative brightness of the light sources.

(j) When viewed from a fixed location, corresponding points in a sequence of images of a moving object will have exactly the same brightness values.

Solution: False. The brightness constancy constraint in optical flow does make this assumption. But, the assumption is not generally true. It is false, for example, if the moving object undergoes rotation, not just translation.

Question 2:

(a) Let $H = \{h_{ij}\}$ be the coefficients of an $n \times m$ digital filter, $i = 1, \ldots, n$, $j = 1, \ldots, m$. One can think of H as defining a 2-D array (or matrix). What does it mean for the 2-D filter, H, to be separable? Given that H is separable, what is the rank of H? [Hint: Recall that the rank of an $n \times m$ matrix is the number of linearly independent rows (or columns).]

Solution: A 2-D filter, H, is separable if and only if the number of linearly independent rows (equivalently columns) is one. Said another way, a 2-D filter, H, is separable if and only if there exists 1-D vectors A and B, respectively of dimension n and m, such that H can be written as the (outer) product

$$H = A^{\mathrm{T}} B$$

where T denotes vector transpose. Given an H that is separable, its rank is one.

(b) Some critics of Artificial Intelligence argue that humans are unlike machines because machines necessarily are "precise" in their perception. One example cited is the Mueller-Lyer illusion illustrated in Figure 1. By construction, both horizontal lines have equal length. To humans, as you hopefully agree, the lengths appear different. The critics argue that a machine is not subject to the illusion because it necessarily measures the correct length in each case.

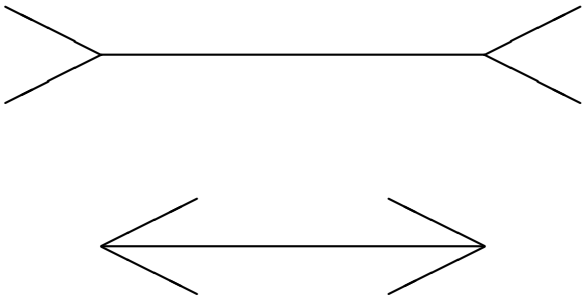

Figure 1: Mueller-Lyer Illusion

It is instructive to examine the Mueller-Lyer illusion in the context of edge detection. Based on your experience with how edge detection works at a corner, argue (one way or the other) whether a machine vision system is subject to the Mueller-Lyer illusion. [Note: It is useful to remember that edge detection is an ill-posed problem so that a smoothing (or regularizing) operator typically is required. How might this requirement influence the estimated lengths of the two horizontal lines?]

Solution: On the assumption that edge detection requires some manner of smoothing to regularize the ill-posed nature of numerical differentiation, we would expect edge detection also to be subject to the Mueller-Lyer illusion.

In both the upper and lower cases illustrated in the illusion, the horizontal line splits into two lines at its left and right endpoints. In the upper case, smoothing will shift the point at which the horizontal line is seen to split outwards, thus making the horizontal line appear longer. In the lower case, smoothing will shift the point at which the horizontal line is seen to split inwards, thus making the horizontal line appear shorter. Said another way, even though edge detection may succeed, we would expect errors in localization at the endpoints of the horizontal lines which would cause edge detection methods (like Marr/Hildreth and Canny) to indeed be subject to the Mueller-Lyer illusion.

(c) Each of the initial NASA Landsat series of satellites carried a multispectral scanner (MSS) imaging system that recorded digital images of the earth's surface with a nominal ground resolution of 78m by 78m per pixel. The ground track of the satellite was not aligned with North-South. Thus, there was significant rotation between the actual coordinate axes of image acquisition and the natural orientation of coordinate axes typically used in mapping systems. In Canada, these Landsat MSS digital images routinely were geometrically corrected to Universal Transverse Mercator (UTM) map coordinates and resampled to a regular 50m by 50m square grid whose axes were aligned with North–South and East–West. From a image processing point of view, what would justify this oversampling? [Hint: Recall Assignment 2 and think of the ratio 78:50 as approximately $\sqrt{2} : 1$).]

Solution: On a square grid, sample spacing depends on direction. If a sample is represented every x meters in the horizontal and vertical directions, then a sample is represented only every $\sqrt{2}x$ meters in the diagonal directions.

Owing to this dependence on direction, something as (conceptually) simple as rotating the tessellation use to represent the digital data becomes complicated if one wants both to minimize loss of information and to minimize the introduction of resampling artifacts. By (careful, one-time) oversampling to a 50m grid (by the data supplier), end users can employ simple techniques (for image rotation) without that end user needing to be overly concerned about loss of information or about the introduction of unwanted artifacts.

(d) Woodham's 1990 formulation of multiple light source optical flow made use of three equations

$$\begin{array}{rcl} E_{1x}\,u \;+\; E_{1y}\,v \;+\; E_{1t} & = & 0 \\ E_{2x}\,u \;+\; E_{2y}\,v \;+\; E_{2t} & = & 0 \\ E_{3x}\,u \;+\; E_{3y}\,v \;+\; E_{3t} & = & 0 \end{array} \tag{1}$$

How did Woodham get additional equations? What are E_{ix}, E_{iy} and E_{it}, $i = 1, 2, 3$? Under what circumstances are u and v the identical variables in each of the three equations? (i.e., Why are there no subscripts on u and v in Equations (1)?)

Solution: Woodham got additional equations by simultaneously imaging a moving object under different conditions of illumination. E_{ix}, E_{iy} and E_{it}, $i = 1, 2, 3$, are respectively the partial derivatives of image irradiance, E, with respect to spatial coordinates, x and y, and time, t, under illumination condition i. The variables u and v are identical at image locations, (x, y), where there is a single underlying geometric motion corresponding to the measured optical flow at (x, y).

(e) Points in shadow cause difficulty in shape–from-shading and in the estimation of optical flow. Distinguish points in "self-shadow" from those in "cast-shadow."

Solution: A point in "self-shadow" with respect to a given light source direction, (θ_s, ϕ_s), is a point for which the angle of incidence, θ_i, with respect to that light source direction is $\geq \frac{\pi}{2}$. This is a local property corresponding to self occlusion from given the light source direction. A point in "cast-shadow" with respect to a given light source direction, (θ_s, ϕ_s), is a point for which the angle of incidence, θ_i, with respect to that light source direction is $< \frac{\pi}{2}$ but a point which receives no illumination from direction (θ_s, ϕ_s) owing to the presence of some intervening object or surface that blocks the light that would otherwise arrive from that direction. This is not a local property since it depends on something happening elsewhere in the scene.

(f) Stauffer & Grimson 2000 model the value of each pixel as a mixture of K Gaussian distributions. The value of K is limited by available memory and computational power. Stauffer & Grimson report using values of K in the range of 3–5. Why do they choose $K > 2$?

Solution: The goal of Stauffer & Grimson 2000, compared to work that preceeded them, was to model more complicated backgrounds. In particular, they wished to model backgrounds that are multimodal (including, for example, glitter off water surfaces, monitor flicker, leaf flicker from vegetation, etc). At least two Gaussian distributions are required to model bimodal backgrounds, with at least one additional Gaussian distribution required to model pixels in the foreground (i.e., pixels where there is actual motion). Thus, K is chosen to be at least 3 (i.e., $K > 2$).

Part B:

Question 3:

We have considered three approaches to edge detection based on the work of Marr & Hildreth, Torre & Poggio and Canny. In each case, the approach was based on clearly stated design criteria. For the purposes of edge detection, what is meant by:

(a) detection

(b) localization in space

(c) localization in frequency

Canny included as one of his design criteria the elimination of multiple responses to a single edge.

(d) Why is this important in Canny edge detection?

(e) How do Marr & Hildreth handle multiple responses to a single edge? (Hint: Either explain how they handle multiple responses or give an argument for why the problem does not arise with their method).

Gaussian, or Gaussian-like, convolution occurs in much of the work on edge detection that we have studied. Concerning their work, Torre & Poggio state, "This result is the simplest and most rigorous proof that a Gaussian-like filter represents the correct operation to be performed before differentiation for edge detection."

(f) Compare and contrast the way in which a Gaussian-like filter emerged as the correct operation in the Torre & Poggio approach with how the Gaussian filter emerged in the Marr & Hildreth approach and in the Canny approach.

Solution:

(a) "Detection" means determining, Yes/No, whether or not there exists an edge in the region of interest. For the formulations of edge detection we have considered, detection is done densely. That is, in an $n \times n$ digital image, edge detection says Yes/No for each pixel in the image, whether or not it is an edge point.

(b) Given that we have determined that there exists an edge in a region of interest, "localization in space" addresses the issue of how well the edge points have been located (in terms of their correct spatial coordinates, (x, y)).

(c) Conceptually, we think of an image as being decomposed into multiple bandpass spatial frequency channels. Given that we have determined that there exists an edge in a region of interest, "localization in frequency" addresses the issue of the presence/absence of the edge in each of the spatial frequency channels. Some edges might exist in only one (or a small number of) frequency channels (and these are deemed localized in frequency). Other edges might exist in many (or all) frequency channels (and these are deemed not localized in frequency).

(d) Canny's method is based on extrema of a first derivative operator and the application of a threshold. For a given choice of threshold, there can be multiple local extrema above the

threshold within a small window. In such cases, it is difficult to tell whether there are indeed multiple edges in the window or merely multiple responses to a single edge. Canny explicitly included the elimination of multiple responses to a single edge (arising owing to noise) in order to accommodate multiple actual nearby edges.

(e) Marr & Hildreth is based on zero crossings of a second derivative operator. At the level of zero-crossing detection, there is no threshold involved and no explicit notion of "multiple responses to a single edge." Of course, for every zero-crossing point, the question remains, "Does it correspond to a single edge, to multiple nearby edges, or to no edge at all?" In Marr & Hildreth, this is dealt with by seeing how zero-crossings change as a function of σ. As σ increases, zero-crossings can change location, can merge with other zero-crossings and/or can disappear. For each given σ, zero-crossings in Marr & Hildreth assert the presence of an edge point at the location of those zero-crossings. By tracking what happens to those zero-crossings for different values of σ, Marr & Hildreth are able to determine whether the zero-crossings correspond to multiple edges or to no edges at all in each of the distinct spatial frequency channels that correspond to the different values of σ.

(f) The three Marr & Hildreth design criteria were localization in space, localization in frequency and rotational invariance. Localization in space and localization in frequency conflict as design criteria. But, the Gaussian is the "minimum uncertainty" (real-valued) filter that best trades-off localization in space and localization in frequency. This is how the Gaussian became the correct operation to use in Marr & Hildreth.

Canny followed Marr & Hildreth and he certainly was aware of the "minimum uncertainty" properties of the Gaussian. Canny, however, was not explicitly concerned about localization in frequency. Even for the special case of step edges, the Gaussian is not the correct operation to use according to Canny's own chosen design criteria. The derivative of the Gaussian performed well, but was sub-optimal. The Gaussian, however, is efficient to implement owing to separability and to the fact that Gaussian smoothing for larger values of σ can be achieved by repeated Gaussian smoothing with smaller values of σ. The Gaussian became the correct operation to use in Canny owing principally to its efficiency.

Torre & Poggio addressed edge detection as a problem in numerical differentiation. Specifically, they sought to interpolate a smooth, continuous function from numerical data based on their recent discovery of Tikhonov regularization (in the Soviet mathematics literature). Given that the interpolated function is regularized with a suitable stabilizing functional, differentiation of the interpolated function then becomes straightforward. For their choice of stabilizing functional and for the special case of data sampled on a regular grid, Tikhonov regularization was achieved using a linear filter whose shape was not exactly that of a Gaussian but one that was very close to that of a Gaussian. This is how the Gaussian became the correct operation to use in Torre & Poggio.

Question 4:

This problem explores the technique of matched filtering. The goal is to establish that the correlation of an idealized pattern with the input image is the "optimal" matched filter in certain well-defined situations. Let $A = \{a_{ij}\}$ and $B = \{b_{ij}\}$ be two $n \times n$ digital images representing two (idealized) patterns. Suppose further that these patterns have been grey-level normalized so that

$$\sum_{ij} {a^2}_{ij} \;=\; \sum_{ij} {b^2}_{ij}$$

An $n \times n$ digital image $C = \{c_{ij}\}$ is obtained. C is assumed to be an image either of A or B, the problem is to decide which. To make the problem interesting, suppose each pixel of C is perturbed by additive, zero mean, Gaussian noise. That is, the pixels in C are not identically those of A or B but have added to them noise where the noise comes from a Gaussian distribution with zero mean. Further, suppose that the noise acts independently on each pixel in C. Recall that if x is a random variable from a Gaussian distribution with mean μ and standard deviation σ, then the probability density function of x is given by

$$p(x) \;=\; \frac{1}{\sqrt{2\pi}\sigma}\; \exp^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Thus, for each pixel in C, the probability of observing that pixel when the pattern is really A is given by

$$p\left(c_{ij}/a_{ij}\right) \;=\; k_1\; \exp^{k_2(c_{ij}\,-\,a_{ij})^2} \quad \text{where} \quad k_1 \;=\; \frac{1}{\sqrt{2\pi}\sigma} \quad \text{and} \quad k_2 \;=\; -\frac{1}{2\sigma^2}$$

Since the noise is assumed to be independent for each pixel, the probability of observing the whole image C when the pattern is really A is just the product of the probabilities of each c_{ij} given a_{ij}. That is

$$p(C/A) \;=\; \prod_{ij} k_1\; \exp^{k_2(c_{ij}\,-\,a_{ij})^2}$$

An identical argument gives

$$p(C/B) \;=\; \prod_{ij} k_1\; \exp^{k_2(c_{ij}\,-\,b_{ij})^2}$$

It seems reasonable to decide that C is an image of A if and only if $p(C/A) \;>\; p(C/B)$.

(a) Prove that $p(C/A) \;>\; p(C/B)$ if and only if $\sum_{ij} c_{ij}a_{ij} \;>\; \sum_{ij} c_{ij}b_{ij}$. [Hint: Instead of comparing $p(C/A)$ and $p(C/B)$ directly as the products of exponentials, it is useful to take logarithms and compare the resulting sums. For $x \;>\; 0$ and $y \;>\; 0$, $x \;>\; y$ if and only if $\ln(x) \;>\; \ln(y)$].

Thus, in order to decide whether C is A or B it is sufficient to compare the correlation of A with C to the correlation of B with C. Said another way, A and B are their own best matched filters.

(b) Suppose C is an $n \times n$ digital image and A and B are $m \times m$ digital patterns with $n \;\gg\; m$. Given the result in part (a), describe, as concisely as you can, how you would go about finding instances of patterns A and B in the image C?

(c) What would be the effect on matched filtering, as implemented in part (b) above, if the patterns A and B were not grey-level normalized (i.e., $\sum_{ij} {a^2}_{ij} \neq \sum_{ij} {b^2}_{ij}$)?

(d) Is matched filtering, as implemented in part (b) above, likely to be robust with respect to differences in spatial scale between image and pattern? What about differences in rotation between image and pattern?

(e) Given the task of optical character recognition (OCR), how might you design a font to take best advantage of the results derived in this problem?

Solution:

(a)

$$\begin{aligned} p(C/A) &= \prod_{ij} k_1 \exp^{k_2 (c_{ij} - a_{ij})^2} \\ &= {k_1}^{n^2} \prod_{ij} \exp^{k_2 (c_{ij} - a_{ij})^2} \end{aligned}$$

Therefore,

$$\begin{aligned} \ln p(C/A) &= n^2 \ln k_1 + \sum_{ij} k_2 (c_{ij} - a_{ij})^2 \\ &= n^2 \ln k_1 + k_2 \sum_{ij} {c_{ij}}^2 + k_2 \sum_{ij} {a_{ij}}^2 - 2k_2 \sum_{ij} c_{ij} a_{ij} \end{aligned}$$

with the analogous result holding for $\ln p(C/B)$. Given that $\sum_{ij} {a^2}_{ij} = \sum_{ij} {b^2}_{ij}$, we obtain

$$\begin{aligned} & p(C/A) > p(C/B) \\ \Leftrightarrow \quad & -2k_2 \textstyle\sum_{ij} c_{ij} a_{ij} > -2k_2 \sum_{ij} c_{ij} b_{ij} \\ \Leftrightarrow \quad & k_2 \textstyle\sum_{ij} c_{ij} a_{ij} < k_2 \sum_{ij} c_{ij} b_{ij} \\ \Leftrightarrow \quad & \textstyle\sum_{ij} c_{ij} a_{ij} > \sum_{ij} c_{ij} b_{ij} \qquad \text{since} \quad k_2 < 0 \end{aligned}$$

(b) One can use each of the $m \times m$ digital patterns, A and B, as correlation filters to apply to the $n \times n$ digital image, C. One would then look for local maxima in the result and apply a threshold, T. That is, local maxima in the correlation of A with C above the threshold, T, would be detected as instances of A in C and local maxima in the correlation of B with C above the threshold, T, would be detected as instances of B in C.

Note: One might choose the threshold, T, to achieve a desired probability for correct detection. This, however, requires taking both the noise variance, σ^2, and each local window's $\sum_{ij} {c_{ij}}^2$ into account. On the other hand, determining if $p(C/A) > p(C/B)$ is independent of σ the local window's $\sum_{ij} {c_{ij}}^2$.

(c) If the patterns A and B were not grey-level normalized then one can not determine whether $p(C/A) > p(C/B)$ by comparison of $\sum_{ij} c_{ij} a_{ij}$ to $\sum_{ij} c_{ij} b_{ij}$ alone. One idea would be to use separate thresholds, $\mathrm{T_A}$ and $\mathrm{T_B}$, adjusted to accommodate for the differences in $\sum_{ij} {a^2}_{ij}$ and $\sum_{ij} {b^2}_{ij}$.

Aside: Reworking the analysis in part (a), one can show

$$\begin{aligned} & p(C/A) \;>\; p(C/B) \\ \Leftrightarrow \;\; & \textstyle\sum_{ij} c_{ij}a_{ij} \;-\; \sum_{ij} c_{ij}b_{ij} \;>\; \frac{1}{2}(\sum_{ij} {a^2}_{ij} \;-\; \sum_{ij} {b^2}_{ij}) \end{aligned}$$

(d) No. Matched filtering, as implemented in part (b), is not robust either to differences in spatial scale between image and pattern or to differences in rotation between image and pattern.

(e) We would like a fixed size font so that it is easy to scale and align characters, during imaging, to a standard coordinate system. Further, and more importantly in the context of this problem, we would like to design each character in the font so that grey-level normalization comes for free. That is, for any two font characters, A and B, we would like their $m \times m$ digital patterns to be such that $\sum_{ij} {a^2}_{ij} \;=\; \sum_{ij} {b^2}_{ij}$. That way, all the OCR character detections can be done effeciently with linear filtering, without (non-linear) normalization.

Question 5:

This problem demonstrates that, in certain circumstances, photometric stereo allows one to determine information about surface material as well as about surface orientation. As we have seen, the bidirectional reflectance distribution function (BRDF) of an ideal lossless Lambertian material is $f_r = 1/\pi$. In this context, lossless means that 100% of the surface irradiance is reflected. Suppose instead that only a fraction, ρ, $0 \leq \rho \leq 1$, of the irradiance is reflected. Then the BRDF of a Lambertian material is $f_r = \rho/\pi$. The fraction ρ is a reflectance factor that often is referred to as the "albedo."

We now show that three light source photometric stereo is sufficient to determine both the surface orientation and the reflectance factor, ρ, for Lambertian materials, even when ρ varies from point to point on the object surface. Suppose we obtain three images under three different light source conditions. Let

$$\begin{aligned} \mathbf{s}_1 &= [s_{11}, s_{12}, s_{13}]^{\mathrm{T}} \\ \mathbf{s}_2 &= [s_{21}, s_{22}, s_{23}]^{\mathrm{T}} \\ \mathbf{s}_3 &= [s_{31}, s_{32}, s_{33}]^{\mathrm{T}} \end{aligned} \tag{2}$$

be unit column vectors ($^{\mathrm{T}}$ denotes vector transpose) defining the directions of three distant, point sources of illumination. Define the 3×3 matrix, $\mathbf{S}$, by

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{31} & s_{32} & s_{33} \end{bmatrix} \tag{3}$$

Let $\mathbf{E} = [E_1, E_2, E_3]^{\mathrm{T}}$ be the column vector of irradiance values at an image point, (x, y), in each of the three images. Let $\mathbf{n} = [n_1, n_2, n_3]^{\mathrm{T}}$ be the column vector corresponding to a unit surface normal at (x, y) and let ρ be the associated reflectance factor.

(a) Express image irradiance, $E_i(x, y)$, as a function of ρ, $\mathbf{s}_i$ and $\mathbf{n}$ for $i = 1, 2, 3$. [Hint: Recall that the cosine of the angle between two vectors is the dot product divided by the product of the magnitudes. Also, since the absolute strength of the light sources is not specified, keep things simple by assuming the three light sources each have irradiance $E_0/\pi = 1$ where E_0 is the irradiance of the light source measured perpendicular to the beam of incident light].

(b) Combine the results of part (a) into a single matrix equation that expresses $\mathbf{E}$ as a function of ρ, $\mathbf{S}$ and $\mathbf{n}$.

(c) Indicate how the matrix equation of part (b) can be used to solve for ρ.

(d) Indicate how the matrix equation of part (b) and the result of part (c) can then be used to solve for $\mathbf{n}$.

(e) Under what conditions will the solution determined in parts (c) and (d) fail to exist?

Solution:

(a) For $i = 1, 2, 3$

$$E_i(x, y) = \rho \, \mathbf{s}_i \, \mathbf{n}$$

(b)

$$\mathbf{E} = \rho \, \mathbf{S} \, \mathbf{n}$$

(c) Assuming the inverse, $\mathbf{S}^{-1}$, exists, we obtain

$$\rho \, \mathbf{n} = \mathbf{S}^{-1} \, \mathbf{E}$$

Since $|\mathbf{n}| = 1$, we can take the magnitude of each side of the above equation to obtain

$$\rho = \rho \, |\mathbf{n}| = \left| \mathbf{S}^{-1} \, \mathbf{E} \right|$$

(d)

$$\mathbf{n} = \frac{\mathbf{S}^{-1} \, \mathbf{E}}{\rho} = \frac{\mathbf{S}^{-1} \, \mathbf{E}}{|\mathbf{S}^{-1} \, \mathbf{E}|}$$

(e) The solution determined in parts c) and d) will fail to exist if $\mathbf{S}^{-1}$ fails to exist. For $\mathbf{S}^{-1}$ to exist, it is necessary that the three vectors, $\mathbf{s}_i = [s_{i1}, s_{i2}, s_{i3}]$, $i = 1, 2, 3$, be linearly independent. That is, the three directions to the distant point sources of illumination must not be co-planar.

Aside: The solution also only works locally for surface points that are visible to all three light sources.

Question 6:

In this problem, we explore edge detection at a corner. We use Gaussian smoothing and detect edges based on zero crossings of the second directional derivative along the gradient. To keep the problem tractable, we consider a single, isolated 2-D corner. Let $F(x, y)$ be the 2-D "image" defined by

$$F(x,y) \;=\; \begin{cases} 1 & x > 0 \;\; y > 0 \\ 0 & \text{otherwise} \end{cases}$$

$F(x, y)$ defines a "corner" at $x = 0$, $y = 0$, corresponding to the location where two (unit) step edges meet, one along the positive x axis and the other along the positive y axis. Thus, $F(x, y)$ can be written as

$$F(x,y) \;=\; f(x)\, f(y)$$

where

$$f(t) \;=\; \begin{cases} 1 & t > 0 \\ 0 & \text{otherwise} \end{cases}$$

First, we smooth the image, $F(x, y)$, with a 2-D Gaussian. The scale parameter, σ, of the Gaussian controls the degree of smoothing. But, notice that $F(x, y) = F(sx, sy)$, for all $s > 0$. This means that $F(x, y)$ is self-similar at all scales. Thus, it is sufficient to consider the case $\sigma = 1$. Let $G(x, y)$ be the 2-D Gaussian (with $\sigma = 1$) defined by

$$G(x,y) \;=\; \frac{1}{2\pi}\, \exp^{-\dfrac{x^2+y^2}{2}}$$

$G(x, y)$ can be written as

$$G(x,y) \;=\; g(x)\, g(y)$$

where

$$g(t) \;=\; \frac{1}{\sqrt{2\pi}}\, \exp^{-\dfrac{t^2}{2}}$$

Note that $g(t)$ is the 1-D Gaussian (with $\sigma = 1$). Recall that convolution commutes so that

$$G(x,y) * F(x,y) \;=\; F(x,y) * G(x,y)$$

($*$ denotes convolution). Also, recall that 2-D and 1-D convolutions are defined, respectively, as

$$F(x,y) * G(x,y) \;=\; \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(x-u, y-v)\, G(u,v)\, du\, dv$$

and

$$f(x) * g(x) \;=\; \int_{-\infty}^{\infty} f(x-u)\, g(u)\, du$$

(a) Show that

$$G(x,y) * F(x,y) \;=\; \phi(x)\, \phi(y) \tag{4}$$

where

$$\phi(t) \;=\; \int_{-\infty}^{t} g(u)\, du$$

Let $\phi'(t)$ denote the derivative of $\phi(t)$. By the fundamental theorem of calculus,

$$\frac{d}{dx}\int_a^x g(u)\,du \;=\; g(x)$$

for any $x > a$ so that

$$\phi'(t) \;=\; g(t) \tag{5}$$

Recall that the second directional derivative along the gradient of a 2-D function, f, is given by

$$\frac{\partial^2}{\partial n^2} \;=\; \frac{{f_x}^2 f_{xx} \;+\; 2 f_x f_y f_{xy} \;+\; {f_y}^2 f_{yy}}{{f_x}^2 \;+\; {f_y}^2} \tag{6}$$

(subscripts denote partial differentiation).

(b) Is $\dfrac{\partial^2}{\partial n^2}$ a linear operator? [Note: No proof is required].

(c) Does $\dfrac{\partial^2}{\partial n^2}(G(x,y) * F(x,y)) \;=\; (\dfrac{\partial^2}{\partial n^2}G(x,y)) * F(x,y)$? [Note: No proof is required].

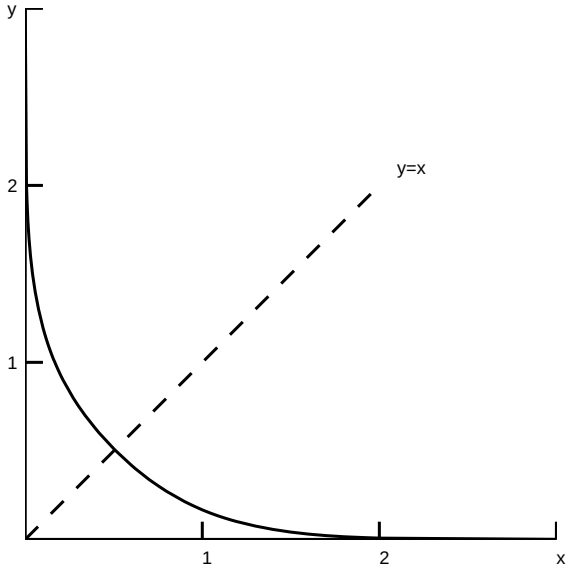

Figure 2: The 2-D corner and its bisector, the line $y = x$. Curved line shows detected edge points displaced near the corner at $x = 0$ and $y = 0$.

Edges are detected at zero crossings of $\frac{\partial^2}{\partial n^2}(G(x,y) * F(x,y))$. One could apply equation (6) to equation (4) to determine zeros directly. This was done to generate Figure 2. Note, however, that the problem is symmetric about the corner bisector (i.e., about the line $y = x$). It should be easy to convince yourself that, owing to symmetry, the zero crossing point corresponding to the corner at $x = 0$, $y = 0$, must lie along the line $y = x$. Also, owing to symmetry, the direction of the gradient is along the line $y = x$. Along the bisector, $G(x,y) * F(x,y) \;=\; \phi(x)\,\phi(x) \;=\; \phi^2(x)$ so that

$$\frac{\partial^2}{\partial n^2}(G(x,y) * F(x,y)) \;=\; \frac{d^2\phi^2(x)}{dx^2}$$

(d) Prove that $\frac{d^2\phi^2(x)}{dx^2} \;=\; 0$ if and only if

$$g(x) \;-\; x\,\phi(x) \;=\; 0 \tag{7}$$

[Hint: Use equation (5)].

Solving equation (7) numerically (to 10 decimal places) gives $x = 0.5060544690$, which we will say is approximately $x = 1/2$. Thus, the corner at point $(0, 0)$ is displaced to the point $(1/2, 1/2)$. If we now let σ be arbitrary, rather than $\sigma = 1$, then the self-similarity argument given above allows us to conclude that the corner at point $(0, 0)$ is displaced to the point $(\sigma/2, \sigma/2)$. Indeed, Figure 2 is self-similar at all scales and we can interpret scale on the x and y axes to be in units of σ.

Fredrik Bergholm of Sweden analyzed this and other special cases of 2-D image features. In particular, he also considered corners where the interior angle at the corner ranged from 0 to π radians, in addition to the $\pi/2$ case considered here.

(e) Use your intuition, rather than formal analysis, to argue whether the displacement at a corner with interior angle $\pi/4$ radians is going to be more (or less) than $(\sigma/2, \sigma/2)$.

Solution:

(a)
$$\begin{aligned} G(x,y)*F(x,y) &= F(x,y)*G(x,y) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(x-u, y-v)\, G(u,v)\, du\, dv \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x-u)\, f(y-v)\, g(u)\, g(v)\, du\, dv \\ &= \int_{-\infty}^{\infty} f(x-u)\, g(u)\, du \int_{-\infty}^{\infty} f(y-v)\, g(v)\, dv \\ &= \int_{-\infty}^{x} g(u)\, du \int_{-\infty}^{y} g(v)\, dv \\ &= \phi(x)\, \phi(y) \end{aligned}$$

(b) $\dfrac{\partial^2}{\partial n^2}$ is not a linear operator.

(c) $\dfrac{\partial^2}{\partial n^2}(G(x,y)*F(x,y)) \neq (\dfrac{\partial^2}{\partial n^2}G(x,y))*F(x,y)$

(d)
$$\begin{aligned} \frac{d\phi^2(x)}{dx} &= 2\,\phi(x)\,\phi'(x) \\ &= 2\,\phi(x)\,g(x) \end{aligned}$$
$$\text{so that} \quad \begin{aligned} \frac{d^2\phi^2(x)}{dx^2} &= 2\,g^2(x) + 2\,\phi(x)\,g'(x) \\ &= 2\,g^2(x) - 2\,\phi(x)\,x\,g(x) \\ &= 2\,g(x)\,[g(x) - x\,\phi(x)] \end{aligned}$$

Now, $g(x) > 0$ for all values of x. Therefore, $\frac{d^2\phi^2(x)}{dx^2} = 0$ if and only if $g(x) - x\,\phi(x) = 0$.

(e) Displacement at a corner with interior angle $\pi/4$ will be more than $(\sigma/2, \sigma/2)$.

Question 7:

In this problem, we develop a "direct" method to determine time to contact relative to a planar surface. The method is based on analysis of the 2-D motion field resulting from translational motion under perspective projection and the constant brightness assumption.

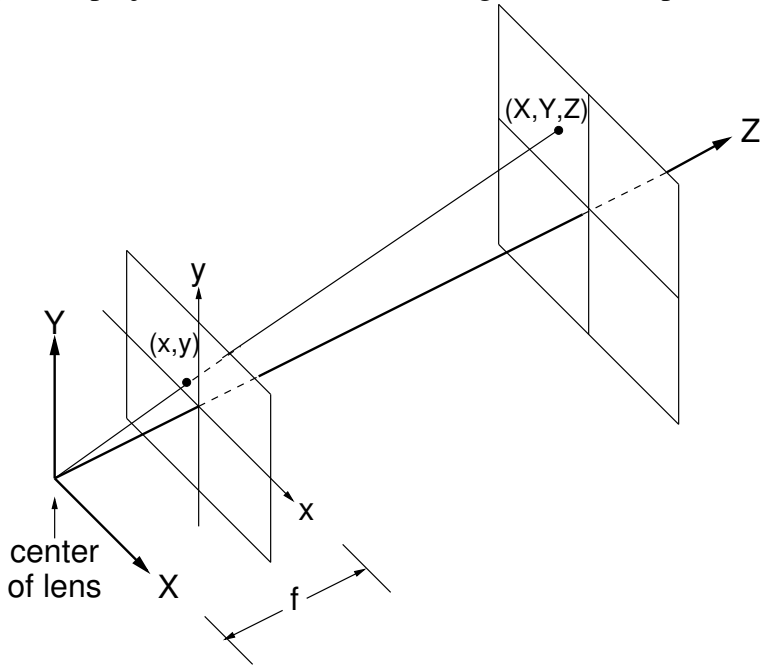

Figure 3: Perspective Projection

First, we deal with perspective projection. Figure 3 illustrates. 3-D point (X, Y, Z) projects to 2-D image point (x, y) where

$$x = f\frac{X}{Z} \text{ and } y = f\frac{Y}{Z} \tag{8}$$

Here, uppercase coordinates refer to quantities in the 3-D world and lowercase coordinates refer to quantities in the 2-D image plane. Let

$$[u, v] = \left[\frac{dx}{dt}, \frac{dy}{dt}\right] \quad \text{and} \quad [U, V, W] = \left[\frac{dX}{dt}, \frac{dY}{dt}, \frac{dZ}{dt}\right]$$

$[u, v]$ is the 2-D motion field. $[U, V, W]$ is the 3-D velocity of a point on the planar surface relative to the camera center of lens (COL) (which is opposite to the motion of the camera relative to the planar surface).

Differentiating the perspective projection Equations (8) with respect to time, we obtain

$$u = f\left(\frac{U}{Z} - \frac{X}{Z}\frac{W}{Z}\right) \text{ and } v = f\left(\frac{V}{Z} - \frac{Y}{Z}\frac{W}{Z}\right)$$

which can be re-written as

$$u = f\left(\frac{U}{Z} - \frac{x}{f}\frac{W}{Z}\right) \text{ and } v = f\left(\frac{V}{Z} - \frac{y}{f}\frac{W}{Z}\right)$$

or

$$u \;=\; \frac{1}{Z}\,(f\,U \;-\; x\,W) \;\text{ and }\; v \;=\; \frac{1}{Z}\,(f\,V \;-\; y\,W) \tag{9}$$

Consider the simple case of translational camera motion along the optical axis towards a planar surface oriented perpendicular to the optical axis.

(a) For this simple case, $U \;=\; V \;=\; 0$. Why?

Time to contact (TTC) is defined as the time remaining before the camera COL reaches the planar surface being viewed if the relative motion between the camera and the surface continues without change (i.e., if the relative motion continues at constant velocity).

Thus, time to contact, T, is the distance (to contact), Z, divided by the velocity of the camera COL, $-W$. That is,

$$T \;=\; -\frac{Z}{W} \tag{10}$$

Subsequent analysis will develop a method to solve for the inverse of TTC. Denote this inverse as C where

$$C \;=\; \frac{1}{T} \;=\; -\frac{W}{Z} \tag{11}$$

Suppose there is some object in the planar surface with (linear) size, S. Let the size of its image be s. Under perspective projection, we obtain

$$\frac{s}{f} \;=\; \frac{S}{Z}$$

which we re-write as

$$s\,Z \;=\; f\,S \tag{12}$$

Now let's differentiate Equation (12) with respect to time.

(b) The derivative of the right-hand-side (RHS) of Equation (12) is 0. Why?

If we apply the product rule for differentiation to the left-hand-side (LHS) of Equation (12) and combine this with the result of part (b) we obtain

$$s\,W \;+\; Z\,\frac{ds}{dt} \;=\; 0 \tag{13}$$

(c) Show that

$$T \;=\; s \Big/ \frac{ds}{dt} \tag{14}$$

[Hint: Combine the results of Equations (10) and (13).]

Equation (14) suggests the possibility of determining TTC based on image measurements alone. The challenge, of course, is to develop a method that is robust.

Now, consider image brightness, $E(x, y, t)$. As with Horn & Schunck 1981, we apply the chain rule for differentiation to $E(x, y, t)$ to obtain

$$\frac{dE}{dt} \;=\; E_x\, u \;+\; E_y\, v \;+\; E_t$$

where subscripts denote partial differentiation. Suppose $\dfrac{dE}{dt} = 0$. This is the "constant brightness assumption."

(d) Explain, in simple English, what the "constant brightness assumption" means.

(e) Is the "constant brightness assumption" reasonable for the simple case of translational camera motion along the optical axis towards a planar surface oriented perpendicular to the optical axis? Briefly justify your answer.

With the constant brightness assumption, we obtain the (classic) optical flow constraint equation

$$E_x\, u \;+\; E_y\, v \;+\; E_t \;=\; 0 \tag{15}$$

Recalling the results in part (a) and Equation (9), we substitute $u = -x(W/Z)$ and $v = -y(W/Z)$ into Equation (15) to obtain

$$-\frac{W}{Z}(x\, E_x \;+\; y\, E_y) \;+\; E_t \;=\; 0$$

or

$$C\, G \;+\; E_t \;=\; 0$$

where C is the inverse of TTC, as in Equation (11), and G is the "radial gradient" $(x\, E_x \;+\; y\, E_y)$.

Finally, we formulate a least squares method to minimize

$$\sum (C\, G \;+\; E_t)^2 \tag{16}$$

Minimizing Equation (16) is straightforward. We differentiate with respect to C and set the result to zero to obtain

$$\sum (C\, G \;+\; E_t)\, G \;=\; 0 \tag{17}$$

or

$$C \sum G^2 \;=\; -\sum G\, E_t \tag{18}$$

so that C, the inverse of TTC, is

$$C \;=\; -\frac{\sum G\, E_t}{\sum G^2} \tag{19}$$

(f) We have not been specific about what region of the image to sum over in Equations (16)–(19). Over what region of the image should we sum? Briefly justify your answer.

The method derived here can be generalized to allow translational motion other than along the optical axis toward a planar surface at an arbitrary orientation to the optical axis. The mathematical details become more complicated (and thus beyond the scope of an examination question) but the essence of the method remains.

(g) For the simple case of translational camera motion along the optical axis towards a planar surface oriented perpendicular to the optical axis, is the estimation of TTC via Equation (19) likely to be robust? Briefly justify your answer.

Solution:

(a) Given that the camera motion is purely translational along the optical axis then, by definition,

$$U \;=\; \frac{dX}{dt} \;=\; 0 \;\text{ and }\; V \;=\; \frac{dY}{dt} \;=\; 0$$

(i.e., There is no component of motion in the X and Y directions, respectively).

(b) The size, S, of the object in the world is not changing with time. Similarly, the focal length of the camera, f, is not changing with time. Thus, the product, fS, is not changing with time. That is,

$$\frac{d(fS)}{dt} \;=\; 0$$

(c) We have

$$s\,W \;=\; -\,Z\,\frac{ds}{dt}$$

so that

$$T \;=\; -\,\frac{Z}{W} \;=\; s \Big/ \frac{ds}{dt}$$

as required.

(d) The "constant brightness assumption" says that (measured) image irradiance does not change as a consequence of motion.

That is, if a point in an image, (x_0, y_0), has brightness α at time t_0 and it moves to point (x_1, y_1) at time t_1, then the brightness of (x_1, y_1) at time t_1 also is α.

(e) Yes. By construction, the motion is purely translational. Therefore, under an assumption of orthographic projection, the gradient, (p, q), at each point on the planar surface does not change as a consequence of motion. (In our simple case, the gradient remains $p = 0$, $q = 0$).

To the extent that an image irradiance equation, $E(x, y) \;=\; R(p, q)$, holds at each point on the planar surface, then image irradiance will not change as a consequence of motion.

Note: It need not be the case that the same image irradiance equation applies to the entire surface. Rather, what is required is that image irradiance for each point on the object surface depends only on the gradient, (p, q).

Aside: Perspective projection is a second order effect that can become significant as the camera comes close to the planar surface. Under perspective projection, the gradient corresponding to the direction to the viewer is not always $p = 0$, $q = 0$. Rather, it also depends on image coordinates, (x, y). This, in turn, means that the "constant brightness assumption" will not be satisfied exactly, even when the motion is purely translational. The effect is small as long as the change in image position, (x, y), as a consequence of motion remains small.

(f) We should be summing over all points in the image for which the basic assumptions of our problem hold.

That is, we should be summing over all visible points on the planar surface that the camera is approaching orthogonally (i.e., perpendicular to the surface).

(g) Yes. There is a single parameter of motion to estimate, namely time to contact (TTC). Measurements required at each pixel, (x, y), corresponding to points on the planar surface, are estimates of the partial derivatives, E_x, E_y and E_t, all of which can be estimated robustly. Finally, all the points on the planar surface contribute independently to the estimation of TTC.

In the end, lots of measurements combine to estimate a single parameter.

For further information, see B.K.P. Horn, Y. Fang and I. Masaki, "Time to Contact Relative to a Planar Surface," in *Proc. IEEE Intelligent Vehicles Symposium*, pp. 68–74, 2007.