RANGE ESTIMATION BY OPTICAL DIFFERENTIATION

HANY FARID

A DISSERTATION

in

COMPUTER AND INFORMATION SCIENCE

Presented to the Faculties of the University of Pennsylvania in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy.

1997

Eero P. Simoncelli

Supervisor of Dissertation

Mark Steedman

Graduate Group Chairperson

This work would not have been possible were it not for:

all those who have come before us and whose shoulders we stand upon,
my current and former advisors, teachers, and mentors,
the GRASP Lab, and my many colleagues and friends (especially Mary Bravo),
and the support of my family.

With many thanks to my first and most influential teachers:
Satenik Farid, Armenoushie Karsian Sudjian, Ramy Farid and Adel Fareed

Abstract

Range Estimation by Optical Differentiation

Hany Farid

Eero P. Simoncelli

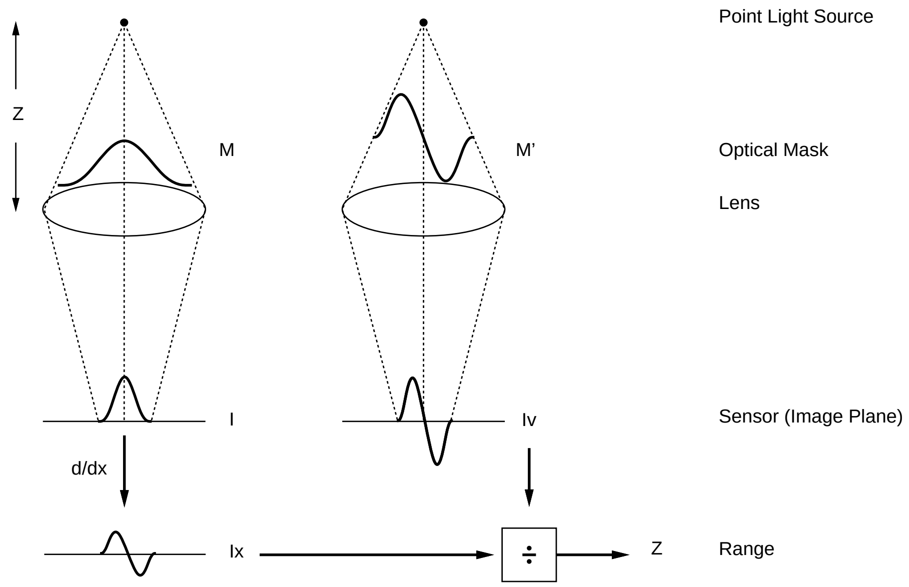
We describe a novel formulation of the range recovery problem based on computation of the differential variation in image intensities with respect to changes in camera position (or aperture size). This method uses a single stationary camera and a pair of calibrated optical masks to directly measure this differential quantity. The subsequent computation of the range image involves simple arithmetic combinations, and is suitable for real-time implementation. Both the theoretical and practical implications of this formulation are addressed.

Contents

List of Figures

Overview

This thesis focuses primarily on the problem of estimating the three-dimensional properties of visual scenes from digital imagery (i.e., range). We present a novel formulation of the range recovery problem based on computation of the differential variation in image intensities with respect to changes in camera position (or aperture size). This method uses a single stationary camera and a pair of calibrated optical masks to directly measure this differential quantity. More specifically, we show that the spatial derivative of the image formed under an optical attenuation mask, M, is related by a scale factor, α, to a second image formed under the derivative mask M'. Where the scale factor, α, is monotonically proportional to range, Z. This simple relationship is illustrated in the figure below. Note that the derivative with respect to camera position, I_v, is measured optically by simply imaging through the mask M'!



By way of introduction, the first chapter reviews the basic geometry, physics, and mathematics of image formation. Wherever possible, these concepts have been placed within a common linear algebraic framework. The final section of this chapter introduces some important new ideas regarding the design of discrete differential operators. The second chapter presents several standard range estimation techniques (stereo, motion, focus, and defocus), and concludes with the observation that all of these techniques amount to measuring changes (i.e., derivatives) with respect to different parameters. In the final chapter, we explore fully the concept of optical differentiation and its application to range estimation. All of the assumptions and constraints for this technique are made explicit and their effect on the overall system carefully studied. The theory of range estimation by optical differentiation is validated through simulation and experimentation. Portions of Chapter 1 have appeared in [Farid 97] and [Farid 96a, Farid 96b, Farid 96c], and portions of Chapter 3 have appeared in [Simoncelli 95, Simoncelli 96b, Farid 96d].

Contributions

Below are what we consider the major contributions of this work:

1. We introduce a novel formulation of the range recovery problem based on optically measuring the differential variation in image intensities with respect to viewing position (as in range from stereo or motion) or aperture size (as in range from defocus). Some advantages of this approach are:
   (a) The technique requires only a single stationary camera.
   (b) Only two images are required for the determination of range.
   (c) The computations are simple and analytic (a few 1D convolutions and arithmetic combinations of the pair of images).
   (d) Because of its simplicity, this technique is amenable to a real-time implementation.

2. As compared to classical stereo approaches, this new formulation completely avoids the difficult and computationally demanding correspondence problem. In addition, with only a single camera, the extrinsic calibration of a stereo camera pair is unnecessary in our configuration.

3. We carefully outline the constraints inherent to this system and study how violations of these constraints effect the overall system. We also develop techniques for either eliminating or relaxing many of these constraints.

4. The basic formulation and sensitivity analysis are verified extensively in simulation. We also verify this technique experimentally with a prototype range camera of our construction.

5. Building on the work of [Simoncelli 94], we have derived a set of higher-order, multi-dimensional derivative/prefilter pairs. These filters are optimally designed (in a least-squares sense) to preserve the necessary derivative relationship between the derivative and prefilter. Although somewhat peripheral to the basic range estimation work, the filters were employed in this work (and should be of general interest to others).

6. By way of introduction, we review the image formation process and formulate this process within a linear algebraic framework. Several other fundamental tools in signal and image processing are also reviewed and cast in a linear algebraic framework. Also by way of introduction, a variety of range estimation techniques are reviewed, and it is argued that each of the techniques can be thought of in a common differential framework: that of measuring change with respect to various imaging parameters.

Notation

For reference, the figure below provides the mathematical notation adopted throughout this document. Any deviations from this notation are noted.

$\mathcal{C}$	the complex numbers
$j \in \mathcal{C}$	$\sqrt{-1}$
$\mathcal{R}$	the real numbers
$\mathcal{Z}$	the integers
$\star$	convolution $(h(x) = (f \star g)(x))$
$\cdot$	inner product
$\times$	cross product
$\equiv$	defined
$\approx$	approximately
$\propto$	proportional
$\vec{v}$	n-dimensional column vector
$\vec{v}^t$	n-dimensional row vector
$M_{n \times m}$	n row by m column matrix
M^t	matrix transpose
M^{-1}	matrix inverse
$\vec{p} = (\, x \quad y \,)^t$	point in 2-D sensor coordinates
$\vec{P} = (\, X \quad Y \quad Z \,)^t$	point in 3-D world coordinates
$g(\alpha_1, ..., \alpha_n)$	n-dimensional function
$\mathcal{G}(\omega_{\alpha_1}, ..., \omega_{\alpha_2})$	Fourier transform of $g(\alpha_1, ..., \alpha_n)$
$Re(\cdot)$	real portion of a complex number
$Im(\cdot)$	imaginary portion of a complex number
$\lvert \mathcal{G}(\cdot) \rvert$	magnitude of Fourier transform: $\sqrt{Re(\mathcal{G}(\cdot))^2 + Im(\mathcal{G}(\cdot))^2}$
$\prec \mathcal{G}(\cdot)$	phase of Fourier transform: $\tan^{-1}\left(\frac{Im(\mathcal{G}(\cdot))}{Re(\mathcal{G}(\cdot))}\right)$
$f(\cdot)$	function of unspecified arguments
$f'(\cdot)$	first derivative of f with respect to its argument
$f''(\cdot)$	second derivative of f with respect to its argument
$f^n(\cdot)$	n-th derivative of f with respect to its argument
$D_{\alpha_i}(f(\alpha_1, ..., \alpha_n))$	partial derivative of f with respect to α_i
$f_{\alpha_i}(\alpha_1, ..., \alpha_n)$	partial derivative of f with respect to α_i
C^n	a function is C^n if its first n derivatives exist
$\overline{f}(\cdot)$	a discretely sampled function

Figure 0.1: Mathematical Notation.

Chapter 1

Introduction

We are primarily interested in computational methods for passive depth estimation. This is a problem with which we are all familiar; taking advantage of several cues, our visual system is able to determine, with remarkable accuracy, absolute and relative distances between objects. [1] For example, an outfielder attempting to throw out a base runner must judge the absolute distance between himself and the base to which the runner is headed. Alternatively, when placing the cap on a pen, we must make relative judgments as to the distance between two objects. In both cases, our visual system performs such tasks with remarkable ease and with seemingly little effort. Without the benefit of a few million years of evolution and the seemingly infinite wisdom of Nature, computational approaches to depth estimation have not yet achieved such remarkable performance.

Although not specifically motivated by the human visual system, we have strived to develop a computationally simple and elegant solution that would be amenable to a real-time implementation, and still allow for quantitative analysis of the assumptions, limitations and errors. The third and final chapter of this document provides the full details of our proposed solution: the anxious reader is welcome to skip to this point and avoid the first two chapters which contain a variety of fundamental ideas and tools in optics and signal/image processing (Chapter 1), as well as a review of several general classes of

[1] To illustrate the remarkable accuracy with which we are able to make relative depth judgments, consider a pair of pencils placed side by side one meter away from you. Our visual system is able to determine if the pencils are displaced by as little as 1 millimeter in depth from one another. This corresponds to a discrimination of one-tenth of one percent, and on our retina the difference in disparity is less than four ten-thousandths of a millimeter. This distance is many times smaller than the diameter of a single visual receptor!

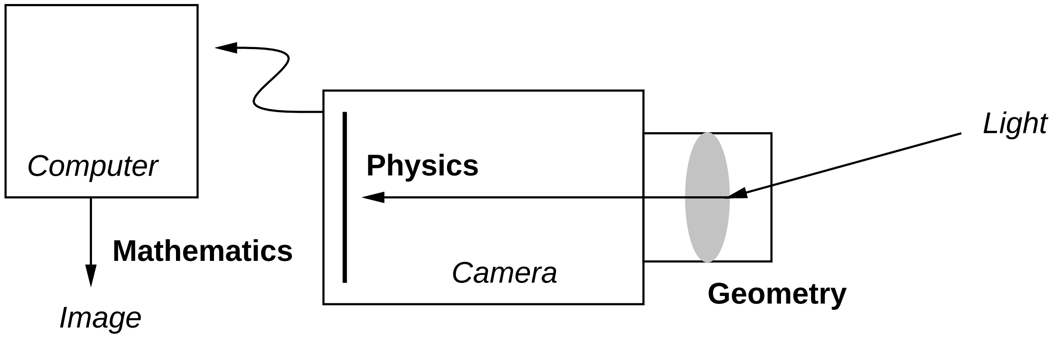


Figure 1.1: Image Formation. Illustrated are the three components of image formation which will be addressed in this chapter: Geometry, Physics, and Mathematics.

depth estimation techniques (Chapter 2). Combined, these two chapters will provide the necessary background for Chapter 3.

In order to appreciate the difficulties involved in range estimation, we begin by reviewing the process of image formation. The relevant characteristics of each stage of image formation can be conveniently described in three broad categories: geometry, physics, and mathematics. As illustrated in Figure 1.1, the geometry describes how light is collected by the camera, the physics specifies how the incoming light is stored and transformed into a digital signal, and the mathematics (i.e., sampling theory) provides us with a set of tools for understanding the transformation from a continuous signal to a discrete signal (i.e., from the intensity of the light to a digital image). Each of these stages is presented according to the standard formulations, and because many aspects of image formation are linear, these stages are recast in a linear algebraic framework.

1.1 Image Formation: Geometry

This section reviews the geometry and projection equations for four standard models of image formation: (1) pinhole under perspective projection, (2) orthographic projection, (3) para-perspective projection and (4) thin lens [2].

The following standard conventions are adopted. $\vec{P} = (\, X \quad Y \quad Z\,)^t$ denotes a point in the three-dimensional world, and $\vec{p} = (\, x \quad y\,)^t$ indicates the position of its projection onto a two-dimensional imaging sensor. The world and sensor coordinates are relative to

[2] More sophisticated models of image formation that consider thick-lens distortion (e.g., [Stevenson 95]) are not considered here.

their own coordinate systems. As illustrated in Figure 1.2, the sensor coordinate system is selected to be the plane $Z = d_s$, where d_s is the distance from the point $(\,0 \quad 0 \quad 0\,)^t$ to the sensor plane along the $Z-$axis (the optical axis), and its origin, $(\,0 \quad 0\,)^t$, lies along the Z-axis. [3]

1.1.1 Pinhole Model Under Perspective Projection

According to a pinhole camera model under perspective projection, light rays travel from a point in the three-dimensional world through an ideal, infinitesimally small, pinhole until they intersect the sensor plane (Figure 1.2). The perspective projection equations are given by:

$$x = \frac{d_s X}{Z} \qquad \text{and} \qquad y = \frac{d_s Y}{Z}, \tag{1.1}$$

where d_s is the distance from the pinhole to the sensor plane along the optical axis [4]. The perspective projection equations may be derived simply from a similar triangles argument.

Although the perspective projection equations are non-linear, they may be expressed in matrix form using the homogeneous equations:

$$\begin{pmatrix} x_s \\ y_s \\ s \end{pmatrix} = \begin{pmatrix} d_s & 0 & 0 & 0 \\ 0 & d_s & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}, \tag{1.2}$$

where the final image coordinates are given by $(\,x \quad y\,)^t = (\,\frac{x_s}{s} \quad \frac{y_s}{s}\,)^t$.

1.1.2 Orthographic Projection

Under orthographic projection, light rays from a point in the world travel parallel to the optical axis until they intersect the sensor plane (Figure 1.2). The orthographic projection

[3] The choice of a sensor coordinate system is of course arbitrary, the $Z = d_s$ plane is chosen to simplify the transformation from world to sensor coordinates.

[4] The perspective projection equations are frequently written with the parameter f, referred to as the focal length, in place of d_s. We do not adopt this convention for two reasons: (1) it is a misnomer, under the pinhole model all points are imaged in perfect focus and (2) it is inconsistent with the thin-lens model which distinguishes between focal length, f, and lens to sensor distance, d_s.

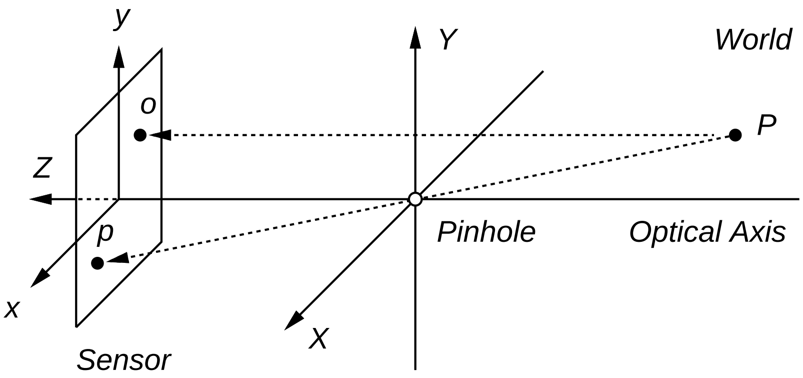


Figure 1.2: Perspective and Orthographic Projection. Illustrated are the projections of a point, $\vec{P}$, in the three-dimensional world under perspective ($\vec{p}$) and orthographic ($\vec{o}$) projection. Points in the world and sensor are specified relative to their own coordinates systems, $(X \quad Y \quad Z)$ and $(x \quad y)$, respectively. Under perspective projection, light rays pass through an infinitesimally small pinhole. Under orthographic projection, light rays travel parallel to the optical axis. See Equations (1.1) and (1.3) for the projection equations.

equations are given by:

$$x = X \qquad \text{and} \qquad y = Y. \tag{1.3}$$

Note that this simple form of the projection equations comes directly from our convenient choice of coordinate systems. These projection equations are linear and may also be expressed in matrix form as:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \tag{1.4}$$

Orthographic projection is a reasonable approximation to perspective projection when the difference in depth between points in the world is small relative to their distance to the sensor. In the special case when all the points lie on a frontal-parallel surface relative to the sensor plane (i.e., $\frac{d_s}{Z}$ is constant in Equation (1.1)), the difference between perspective and orthographic is only a scale factor, $m = \frac{d_s}{Z}$.

Between orthographic and perspective projection is para-perspective. Under this model, points in the world are projected under orthographic projection to a single plane in the world and then projected under perspective projection to the sensor plane. The para-perspective projection equations amount to orthographic projection "plus" a scale factor,

$m = \frac{d_s}{Z_0}$, where Z_0 is the position of the initial projection world plane. Since this projection is also linear, it can be represented in matrix notation:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} m & 0 & 0 \\ 0 & m & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \tag{1.5}$$

Both the orthographic and para-perspective projection equations are linear, however, the addition of a scale factor makes para-perspective a better approximation to the non-linear perspective projection equations.

Under perspective, para-perspective and orthographic projection, each point in the sensor plane corresponds to a *single* light ray striking the sensor. However, in standard imaging systems each point in the sensor corresponds to the average intensity from a *collection* of light rays. Although still an idealization, the thin lens model described next is a more accurate model of image formation.

1.1.3 Thin Lens

Light emanates from a point in all directions, and the pinhole camera model described above captures this light from a *single* direction. In contrast, a lens collects light from *many* directions and focuses the light to a small region on the sensor. An ideal thin lens brings into perfect focus (i.e., a single point) the light emanating from a point at a depth of d_o satisfying the following lens equation (Figure 1.3):

$$\frac{1}{d_o} + \frac{1}{d_s} = \frac{1}{f}, \tag{1.6}$$

where d_s is the distance to the sensor plane from the center of the lens along the optical axis, and f is the focal length of the lens [5]. Points at a depth of $Z \neq d_o$ are imaged as blurred circles [6] with a radius r:

$$r = \frac{R}{\frac{1}{f} - \frac{1}{Z}} \left| \left(\tfrac{1}{f} - \tfrac{1}{Z} \right) - d_s \right|, \tag{1.7}$$

[5] The focal length is the distance from the sensor to the lens along the optical axis where the image of an object that is infinitely far away is imaged in perfect focus. This definition is merely a restatement of Equation (1.6), where if $\frac{1}{d_o} = 0$, then $d_s = f$.

[6] The modeling of the blur as a blurred circle is only an approximation to the point spread function of the camera (i.e., the image of a point source).

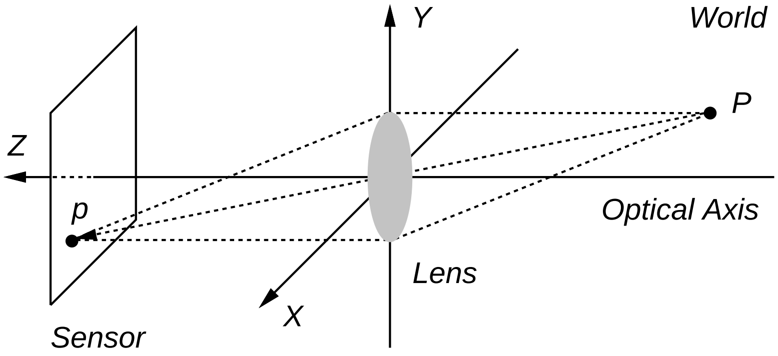


Figure 1.3: Thin Lens. Illustrated is the projection of a point, $\vec{P}$, in the three-dimensional world under the thin lens model. A lens collects light emanating from each point in the world from a continuum of directions and focuses them to a small region on the sensor plane. See Equations (1.6) and (1.7) for the projection equations.

where R is the radius of the lens. This relationship is easily derived from the imaging geometry. Equivalently, the image formation through a thin-lens may be considered in reverse. In particular, the intensity at each point in the sensor is determined from a weighted integral of the light emanating from a small surface patch in the world.

Under this model the projection of a point $\vec{P} = (X \quad Y \quad Z)^t$ in the world is centered about the point $(x \quad y)^t = (\frac{d_s X}{Z} \quad \frac{d_s Y}{Z})^t$, i.e., the perspective projection of the principle ray passing through the center of the lens (Equation (1.1)). Note that the pinhole model under perspective projection is simply a special case of the thin-lens model. In particular, a pinhole is a lens with an aperture stopped down to allow only a single light ray, the principle ray, to pass.

As in the case of perspective, orthographic, and para-perspective projection, the thin-lens equations are linear and can thus be written in matrix form:

$$\begin{pmatrix} l_2 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\frac{1}{R}\left(\frac{n_2 - n_1}{n_2}\right) & \frac{n_1}{n_2} \end{pmatrix} \begin{pmatrix} l_1 \\ \alpha_1 \end{pmatrix}, \tag{1.8}$$

where R is the radius of the lens, n_1 and n_2 are the index of refraction for air and the lens material, respectively. l_1 and l_2 are the height at which a light ray enters and exits the lens (the thin lens idealization ensures that $l_1 = l_2$). α_1 is the angle between the entering light ray and the optical axis, and α_2 is the angle between the exiting light ray and the optical axis.

1.1.4 Non-Invertibility of Image Formation

Image formation, independent of the particular model, is a three-dimensional to two-dimensional transformation (i.e., the sensor plane is two-dimensional). Inherent to such a transformation is a loss of information: the information lost is the distance to the objects in the world. Specifically, under perspective and thin-lens projection all points of the form $\vec{P}_c = (\, cX \quad cY \quad cZ \,)^t$, for any $c \in \mathcal{R}$, are projected to the same point $(\, x \quad y \,)^t$.[7] Similarly, under orthographic and para-perspective all points of the form $\vec{P}_c = (\, X \quad Y \quad cZ \,)^t$, for any $c \in \mathcal{R}$, are projected to the same point $(\, x \quad y \,)^t$. In either case, *the projection is not one-to-one and thus not invertible.*

In addition to this geometric argument for the non-invertibility of image formation, a similarly straight-forward linear algebraic argument holds. In particular, we have seen that the image formation equations may be written in matrix form as, $\vec{p} = M_{n \times m}\vec{P}$, where $n < m$ (Equations (1.2), (1.4), and (1.5)). Since the projection is from a higher dimensional space to a lower dimensional space, the matrix M is not invertible, and thus the projection is not invertible.

Up to this point we have concentrated on the geometric rules governing the formation of images through various camera models. We have neglected to discuss the process by which digital or discretely sampled images are actually formed, that is, the process by which the light rays striking the sensor plane are converted into a digital image. The following section reviews these principles, followed by the mathematics of image formation (i.e., sampling theory).

1.2 Image Formation: Physics

Along with the moon landing and Woodstock, 1969 saw the appearance of the first paper on the charge-coupled device (CCD) [Sangster 69] . Although the latter likely received less publicity, CCD technology has had a widespread technological impact over the past 25 years. Since many of the details of this technology are not critical to the understanding of our work, this section will focus only on the basic principles.

[7] Although it is true that under the thin-lens model the principle rays of all points, $\vec{P}_c$, are projected to the same sensor point, the blur radius is a function of c (i.e., the distance to the sensor plane). However, this information is lost when the light rays are averaged at the sensor plane.

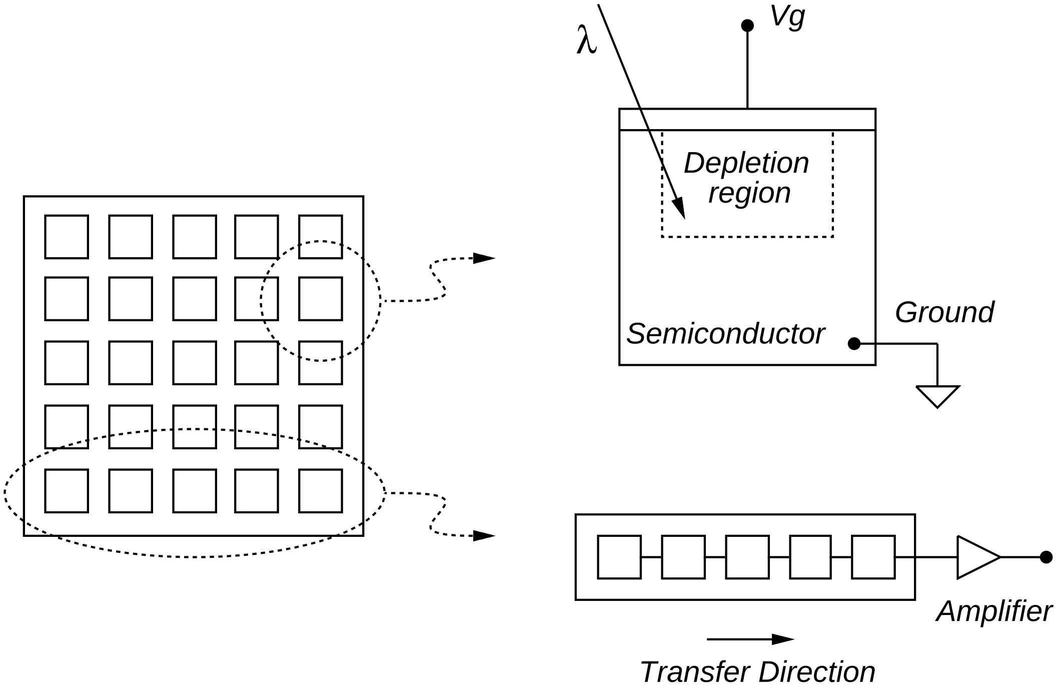


Figure 1.4: Metal-Oxide Semiconductor. Illustrated is a 5 x 5 CCD (left). Each element of the CCD array is a metal-oxide-semiconductor (MOS) capacitor that stores charge proportional to the intensity of the incoming light (top right). The stored charged is then transferred along a row of the CCD, and converted to a voltage by an amplifier (bottom right); see also Figure 1.5. An analog-to-digital converter translates the voltage into a digital number (i.e., the intensity value of a pixel); see Section 1.3 for more details on analog-to-digital conversion.

A basic CCD consists of a series of closely spaced metal-oxide-semiconductor (MOS) capacitors (Figure 1.4). A CCD is a simple charge storage and transport device: charge is stored on the MOS capacitors and then transported across these capacitors for readout and subsequent transformation to a digital image. More specifically, when a positive voltage, V_g, is applied to the surface of a P-type MOS capacitor, positive charge migrate toward ground. The region depleted of positive charge is called the depletion region (Figure 1.4). When photons (i.e., light) enter the depletion region, the electrons released are stored in this region. The value of the stored charge is proportional to the intensity of the light striking the capacitor.

A digital image is formed by transferring the stored charge from one depletion region to the next (first introduced by [Boyle 70]). The stored charge is transferred across a series of MOS capacitors (e.g., a row or column of the CCD array) by sequentially applying

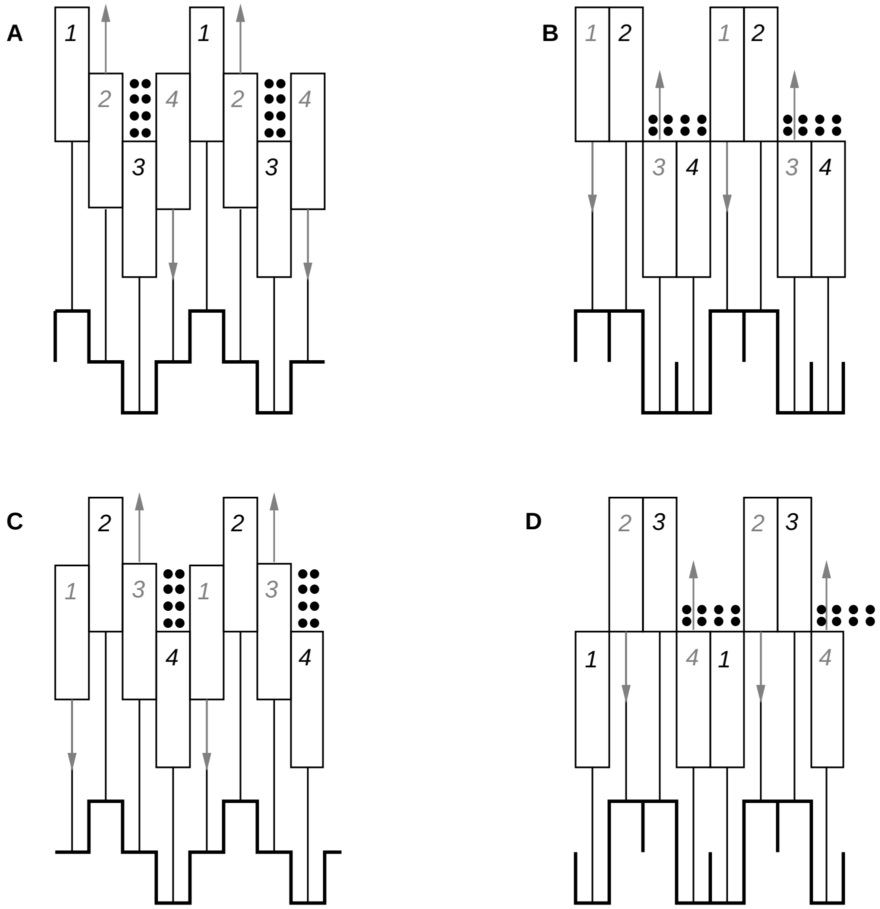


Figure 1.5: Mechanical Analog to CCD Charge Transfer. Illustrated is an analogy to the transfer of charge across a series of MOS capacitors (Figure 1.4). The number 3 pistons are analogous to the MOS capacitors, the number 1, 2, 4 pistons prevent neighboring charge from mixing as they are being transferred, and the small black dots represent the stored charge.

voltage to each MOS capacitor. A simple mechanical analog is illustrated in Figure 1.5, where the number 3 pistons are analogous to the MOS capacitors, the number 1, 2, 4 pistons prevent neighboring charges from mixing, and the small black circles represent the stored charge. As a charge passes through the last capacitor in the series, an amplifier converts the charge into a voltage. An analog-to-digital converter translates this voltage into a number (i.e., the intensity of an image pixel); see the next section for more details on analog-to-digital conversion.

1.3 Image Formation: Mathematics

Having discussed the geometry and physics of image formation, only the mathematical theory remains. In particular, this section presents the mathematical theory underlying the sampling of a continuous signal (in our case, the sampling of light by the CCD camera). Since sampling may result in a loss of information, it is important to understand precisely what information, if any, is lost during the image formation process. For a more extensive coverage of this material see [Oppenheim 89]. Discussion of this material also provides a convenient platform for reviewing two standard tools in signal and image processing, namely, convolution and Fourier transforms – these tools are used extensively in subsequent chapters. As in previous sections, each of the concepts introduced is considered within a linear algebraic framework.

1.3.1 Sampling

In the image formation process described above an analog (continuous) signal is sampled and converted into a digital (discrete) signal. Underlying this process is a beautiful and elegant mathematical theory which is presented next. For notational clarity, the presentation is in 1-D, the principles extend naturally to higher dimensions.

Throughout this section a continuous signal will be denoted as $f(x)$, and its discretely sampled counterpart will be denoted as $\overline{f}$. Mathematically, the sampling of $f(x)$ may be denoted as:

$$\overline{f}(x) \quad = \quad \sum_{k=-\infty}^{\infty} f(kT)\,\overline{\delta}(x-kT), \tag{1.9}$$

where T is the sampling period ($1/T$ is the sampling frequency), and $\overline{\delta}(t)$ is a unit-impulse function defined as:

$$\overline{\delta}(x) \quad = \quad \begin{cases} 1, & x = 0 \\ 0, & \forall x \in \mathcal{Z},\ x \neq 0 \end{cases} \tag{1.10}$$

The unit-impulse should not to be confused with the more common infinite Dirac delta function. Alternatively, the sampling process can be described by a simple product of the

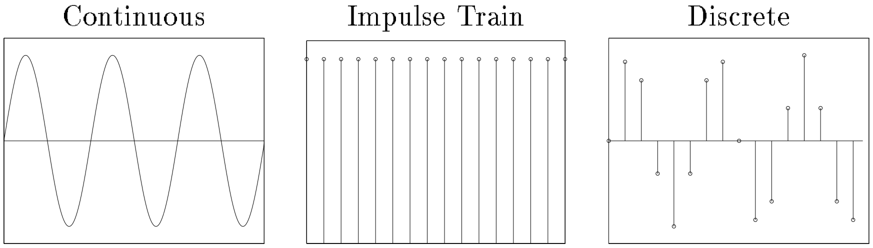


Figure 1.6: Sampling. Illustrated, from left to right: a continuous signal, $f(x) = \sin(3x)$, a unit-impulse train, $\overline{s}(x)$, and the discretely sampled function, $\overline{f}(x)$, computed by multiplying the sinusoid and the impulse train (see Equation (1.11)).

original continuous signal and a unit-impulse train:

$$\overline{f}(x) \quad = \quad f(x) \cdot \overline{s}(x), \tag{1.11}$$

where the unit-impulse train is defined as:

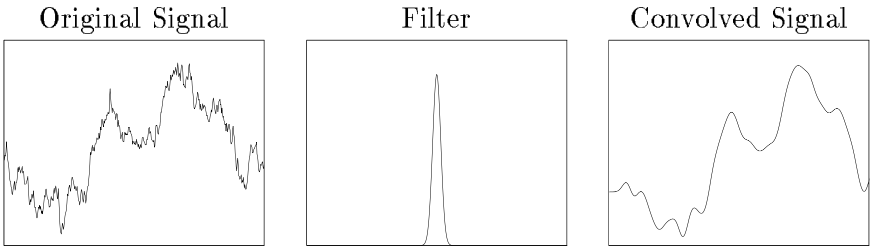
$$\overline{s}(x) \quad = \quad \begin{cases} 1, & x = kT,\ k \in \mathcal{Z} \\ 0, & \text{otherwise} \end{cases} \tag{1.12}$$

A system which implements this operation is referred to as a continuous-to-discrete-time or analog-to-digital (A/D) converter. Figure 1.6 illustrates an example this process.

In general, the sampling process is not invertible, that is, given a discrete signal, $\overline{f}(x)$, it may not be possible to *uniquely* reconstruct the original continuous signal, $f(x)$. It may be somewhat surprising then to learn that under certain conditions it is possible to reconstruct $f(x)$ *exactly* from $\overline{f}(x)$. Under such conditions, the continuous signal is *completely* characterized by its discretely sampled counterpart. To better understand how this reconstruction is accomplished, it is helpful for readers to be familiar with the basic principles of convolution and the Fourier transform. Those familiar with these basic principles should certainly skip the next two sections.

Convolution

Convolution is arguably the most fundamental operation in signal and image processing. This linear operation takes as input a signal, $\overline{f}(x)$, (for our purposes, a discrete signal),



Original Signal

Filter

Convolved Signal

Figure 1.7: Convolution. Illustrated from left to right: an arbitrary signal, a convolution filter (or kernel), and the result of applying the filter to the signal via a convolution (Equation (1.13)). Note that the effect of this particular filter is to average or blur the signal (i.e., a low-pass filter that removes the high frequency content in the signal).

and a discrete filter, $\overline{h}(x)$, and returns the result of "applying" the filter to the signal as follows:

$$\overline{g}(x) \;=\; \sum_{k=-\infty}^{\infty} \overline{f}(k)\overline{h}(x-k). \qquad (1.13)$$

This operation is frequently denoted as $\overline{g}(x) = (\overline{f} \star \overline{h})(x)$. The convolution operation amounts to "sliding" the filter across the signal, and at each position computing an inner product between the filter and the underlying signal. The result of the inner product, a scalar, is the value of the new signal for the position at which the filter is centered. Illustrated in Figure 1.7 is an example of a signal, a filter, and the result of convolving the two. Since convolution is a linear transformation it may be expressed as a matrix operation:

$$\begin{pmatrix} \overline{g}(1) \\ \overline{g}(2) \\ \dots \\ \overline{g}(n) \end{pmatrix} = \begin{pmatrix} \overline{h}(m) & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ \overline{h}(m-1) & \overline{h}(m) & 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & & & & & & & \vdots \\ \overline{h}(1) & \dots & \overline{h}(m) & 0 & 0 & 0 & \dots & 0 \\ 0 & \overline{h}(1) & \dots & \overline{h}(m) & 0 & 0 & \dots & 0 \\ \vdots & & & & & & & \vdots \\ 0 & 0 & \dots & 0 & 0 & \overline{h}(1) & \dots & \overline{h}(m) \end{pmatrix} \begin{pmatrix} \overline{f}(1) \\ \overline{f}(2) \\ \dots \\ \overline{f}(n) \end{pmatrix}$$

$$\vec{g} \;=\; M\vec{f}, \qquad (1.14)$$

where the rows of the matrix contain translated copies of the convolution filter (this expression assumes finite length signals and filters, and makes an arbitrary choice of boundary handling to ensure that the convolution matrix is square). This formulation makes it clearer that convolution is an invertible process, that is, given the filter (i.e., the matrix M) and the results of the convolution, $\vec{g}$, the original signal, $\vec{f}$ can be reconstructed *exactly*. That is, we need only invert the matrix M and premultiply with $\vec{g}$ to recover $\vec{f}$. [8]

Fourier Transform

This section begins by exploring the Fourier series and then deriving the Fourier transform. More specifically we describe the Discrete-Time Fourier Transform (just one of four possible formulations). Our discussion begins with a simple (and perhaps somewhat surprising) fact; any periodic and continuous signal (with period T) can be written as a sum of scaled and phase shifted cosines of varying frequencies:

$$f(x) \;=\; \sum_{k=0}^{\infty} c_k \cos(k\omega x + \phi_k), \tag{1.15}$$

where, $\omega = \frac{2\pi}{T}$, the coefficients c_k are in $\mathcal{R}$, and, since $f(x)$ is periodic, only $x \in [0, T]$ need be considered. It is sometimes desirable to express Equation (1.15) in terms of sine *and* cosine functions. [9] Recalling that $\cos(A+B) = \cos(A)\cos(B) - \sin(A)\sin(B)$, Equation (1.15) may be rewritten as:

$$\begin{aligned} f(x) &= \sum_{k=0}^{\infty} c_k \cos(\phi_k)\cos(k\omega x) + c_k \sin(\phi_k)\sin(k\omega x) \\ &= \sum_{k=0}^{\infty} a_k \cos(k\omega x) + b_k \sin(k\omega x). \end{aligned} \tag{1.16}$$

This is the Fourier series and the values a_k and b_k are the Fourier coefficients. Illustrated in Figure 1.8 is a simple example of a Fourier series, where the signal on the left is expressed as a weighted sum of a constant function (i.e., $k = 0$) [10] and the sine and cosine of the first and second harmonics (i.e., $k = 1, 2$). Note that the b_0 term is not present since $\sin(0) = 0$.

[8] Given that convolution preserves dimensionality (i.e., $\vec{g}$ and $\vec{f}$ have the same dimension) the convolution matrix is square, and also full rank (assuming that $\vec{h}$ is not identically zero), and therefore guaranteed to be invertible. In practice, determination of the original signal, $\vec{f}$, from the convolved signal, $\vec{g}$, also requires that the Fourier transform of the filter $\vec{h}$ be non-zero for each frequency (see next section on Fourier Transforms).

[9] One benefit of expressing the Fourier series in terms of sine and cosine is the elimination of the phase term, ϕ_k, resulting in a fixed basis set.

[10] The first term in the series (i.e., $k = 0$) is commonly referred to as the dc-component.

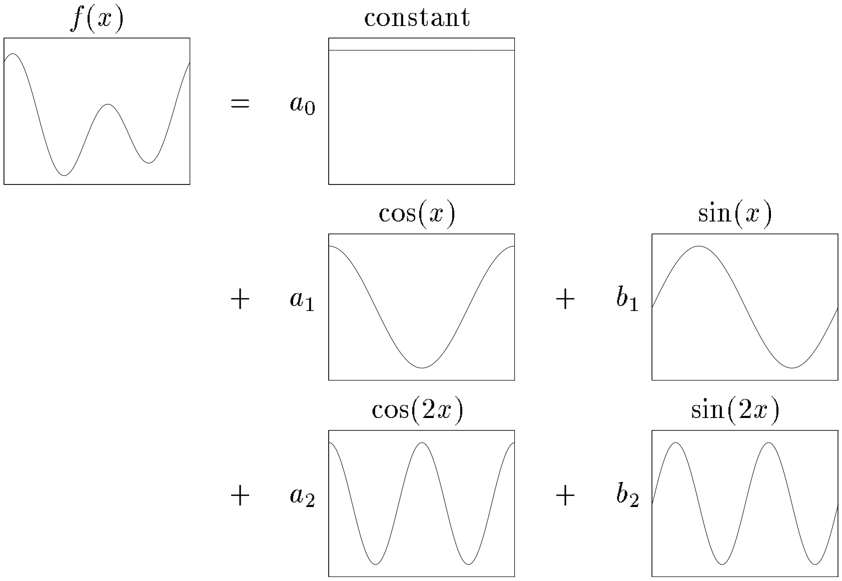


Figure 1.8: Fourier Series. Illustrated is the Fourier series expansion of the signal $f(x)$, with period $T = 2\pi$ (Equation (1.16)). True to the figure, $f(x)$ is constructed by literally adding together weighted (by a_k and b_k) combinations of the constant, sine, and cosine functions.

Equation (1.16) provides a simple expression for computing $f(x)$ *if* the coefficients a_k and b_k are known. The question now is, given $f(x)$, how can the coefficients be determined? Of course, Fourier discovered a simple method for computing the coefficients: the Fourier transform. The first coefficient in the series, a_0 (the b_0 term can be ignored since $\sin(0) = 0$) is the mean of the signal over one period (i.e. $t = 0$ to $t = T$):

$$a_0 \quad = \quad \frac{1}{T} \int_0^T dx \; f(x). \tag{1.17}$$

We can see quite easily why this must be so. Consider the mean of a signal and the mean of its Fourier series (Equation (1.16)). Since the mean value of a sine and cosine over any integral number of periods is zero, the only non-zero term remaining in the Fourier series is a_0. Therefore, in order for the equality to hold (Equation (1.16)), a_0 must be equal to the mean of $f(x)$.

The second coefficient in the series, a_1 and b_1, can be determined in a similar manner. First consider the a_1 term. Multiplying both sides of the Fourier series (Equation (1.16)) by $\cos(\omega x)$, and expanding the right-hand side yields:

$$\begin{aligned} f(x)\cos(\omega x) &= a_0\cos(\omega x) \\ &+ a_1\cos(\omega x)\cos(\omega x) + b_1\sin(\omega x)\cos(\omega x) \\ &+ a_2\cos(2\omega x)\cos(\omega x) + b_2\sin(2\omega x)\cos(\omega x) + \ldots \qquad (1.18) \end{aligned}$$

Using the identities [11] $\cos A\cos B = \frac{1}{2}(\cos(A+B)+\cos(A-B))$ and $\sin A\cos B = \frac{1}{2}(\sin(A+B)+\sin(A-B))$, the Fourier series can be rewritten as:

$$\begin{aligned} f(x)\cos(\omega x) &= a_0\cos(\omega x) \\ &+ \frac{1}{2}a_1(\cos(2\omega x)+\cos(0)) + \frac{1}{2}b_1(\sin(2\omega x)+\sin(0)) \\ &+ \frac{1}{2}a_2(\cos(3\omega x)+\cos(\omega x)) + \frac{1}{2}b_2(\sin(3\omega x)+\sin(\omega x)) + \ldots \qquad (1.19) \end{aligned}$$

Once again, we compute the mean of both sides. On the right-hand side, the mean of most terms is zero: the mean of the a_0 term is a_0 times the mean of $\cos(\omega x)$, but that is just zero. The mean of *all* of the b_k terms is also zero since $\sin(0) = 0$ and the mean of $\sin(\cdot)$ over an integral number of periods is zero. Since the mean of $\cos(\cdot)$ is zero over an integral number of periods, the mean of all of the a_k terms, *except* for a_1, is zero. In the case of the a_1 term, since $\cos(0) = 1$ the mean of the a_1 term is $\frac{1}{2}a_1$. Thus, computing the mean of $\cos(\omega x)$ times the Fourier series yields a simple expression for determining the value of the a_1 Fourier coefficient:

$$\begin{aligned} \frac{1}{T}\int_0^T dx\ f(x)\cos(\omega x) &= \frac{1}{2}a_1 \\ a_1 &= \frac{2}{T}\int_0^T dx\ f(x)\cos(\omega x). \qquad (1.20) \end{aligned}$$

Note that in a linear algebraic sense, the coefficient is computed by projecting the signal onto the basis function $\cos(\omega x)$. By multiplying both sides of Equation (1.16) by $\sin(\omega x)$, a similar expression may be determined for b_1:

$$\begin{aligned} \frac{1}{T}\int_0^T dx\ f(x)\sin(\omega x) &= \frac{1}{2}b_1 \\ b_1 &= \frac{2}{T}\int_0^T dx\ f(x)\sin(\omega x). \qquad (1.21) \end{aligned}$$

[11] The Fourier series is rewritten from a product of sines and cosines into a sum in order to simplify computing the mean.

Clearly, this argument holds for any coefficient. In particular, the a_k and b_k coefficients are determined by computing the mean of the Fourier series multiplied by $\cos(k\omega x)$ and $\sin(k\omega x)$, respectively. The general form for computing the Fourier coefficients (i.e., the Fourier transform) is:

$$a_k = \frac{2}{T}\int_0^T dx\ f(x)\cos(k\omega x) \qquad \text{and} \qquad b_k = \frac{2}{T}\int_0^T dx\ f(x)\sin(k\omega x). \tag{1.22}$$

We may adopt a more compact representation for the Fourier series by recalling what Feynman calls "the most remarkable formula in Mathematics" [Feynman 77]:

$$e^{j\omega} \ = \ \cos(\omega) + j\sin(\omega), \tag{1.23}$$

where j is the complex value $\sqrt{-1}$ (frequently denoted by mathematicians as i). This relationship allows us to express the Fourier series (Equation (1.16)) in terms of the more compact complex exponential:

$$\begin{aligned} f(x) &= Re\left(\sum_{k=0}^{\infty}(a_k - jb_k)(\cos(k\omega x) + j\sin(k\omega x))\right) \\ &= Re\left(\sum_{k=0}^{\infty} c_k e^{jk\omega x}\right), \end{aligned} \tag{1.24}$$

where the Fourier transform is also expressed in terms of complex exponentials:

$$c_k \ = \ \frac{2}{T}\int_0^T dx\ f(x)e^{-jk\omega x}. \tag{1.25}$$

Due to the symmetry/anti-symmetry of the cosine and sine function (i.e., $\cos(\omega) = \cos(-\omega)$ and $\sin(\omega) = -\sin(-\omega)$), the Fourier series can be written over all ω, positive and negative, with a scale adjustment in the Fourier transform:

$$f(x) = \sum_{k=-\infty}^{\infty} c_k e^{jk\omega x} \qquad \text{and} \qquad c_k = \frac{1}{T}\int_0^T dx\ f(x)e^{-jk\omega x}. \tag{1.26}$$

As notation is rarely standard, it is useful to comment on various notations that one may expect to see elsewhere. First, note that Equation (1.26) is continuous in the space domain and discrete in the frequency domain. Such a transform is referred to as the Discrete-Time Fourier Transform (DTFT), just one of four possible formulations. Instead of the c_k notation, we will adopt the more traditional $F(\omega) = \frac{1}{T}\int_0^T f(x)e^{-j\omega x}$ notation for

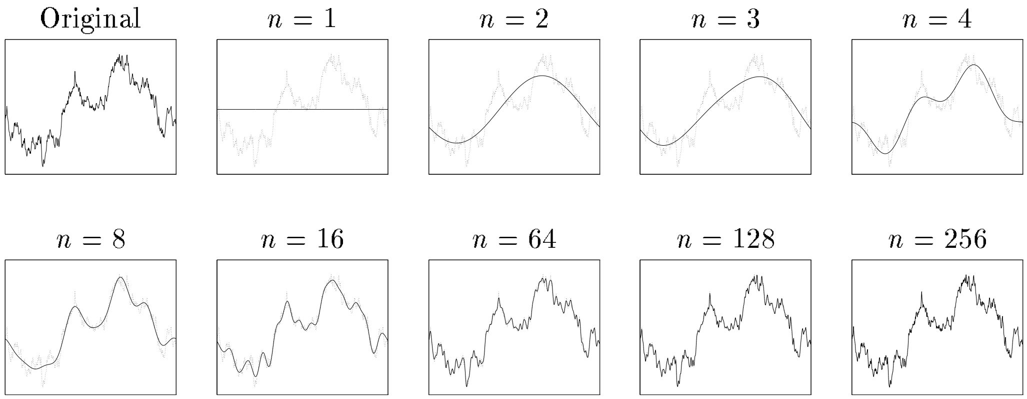


Figure 1.9: A Fourier Series. Illustrated is a signal (upper left) and its partial reconstruction from its Fourier series. Each panel illustrates a partial reconstruction of the signal from the first $n-1$ frequencies and the dc-term. Note that the reconstruction based on only the dc-term ($n = 1$) is a constant function, equal to the mean of the signal. The reconstruction based on the dc-term and one frequency ($n = 2$) is a single phase and amplitude modulated raised sinusoid. The reconstruction based on the dc-term and first 255 frequencies ($n = 256$) is an exact reconstruction of the original signal which was composed of 512 samples. A simple transmission scheme may transmit a "coarse" version of the signal by transmitting only the coefficients of the first few frequencies but this allows only a partial reconstruction of the signal (i.e., the high frequency information is lost).

the Fourier transform of a signal, $f(x)$. In general the Fourier transform, $F(\omega)$, is complex and may be expressed in terms of its real and imaginary parts:

$$F(\omega) \quad = \quad F_R(\omega) + jF_I(\omega), \tag{1.27}$$

or in polar coordinates in terms of magnitude, $|F(\omega)|$, and phase, $\prec F(\omega)$:

$$\begin{aligned} F(\omega) &= |F(\omega)| \cos(\prec F(\omega)) + j|F(\omega)| \sin(\prec F(\omega)) \\ &= |F(\omega)| e^{j \prec F(\omega)}, \end{aligned} \tag{1.28}$$

where the magnitude and phase are given by:

$$|F(\omega)| \quad = \quad \sqrt{F_R(\omega)^2 + F_I(\omega)^2} \tag{1.29}$$

$$\prec F\omega) \quad = \quad \tan^{-1}\left(\frac{F_I(\omega)}{F_R(\omega)}\right). \tag{1.30}$$

It is usually with respect to the magnitude and phase that the Fourier transform of a signal is studied, and is commonly referred to as Fourier or frequency analysis.

As has been the case so often before, the Fourier transform is also a linear operation and can be expressed in matrix notation:

$$\begin{aligned}
\begin{pmatrix} \overline{F}(0) \\ \overline{F}(\omega) \\ \overline{F}(2\omega) \\ \dots \\ \overline{F}(T\omega) \end{pmatrix} &= \frac{1}{T} \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ e^{j0} & e^{j\omega} & e^{j2\omega} & \dots & e^{jT\omega} \\ e^{j0} & e^{j2\omega} & e^{j4\omega} & \dots & e^{j2T\omega} \\ \vdots & & & & \vdots \\ e^{j0} & e^{jT\omega} & e^{j2T\omega} & \dots & e^{jT^2\omega} \end{pmatrix} \begin{pmatrix} \overline{f}(0) \\ \overline{f}(1) \\ \overline{f}(2) \\ \dots \\ \overline{f}(T) \end{pmatrix} \\
\vec{F} &= M\vec{f}, \qquad (1.31)
\end{aligned}$$

where $\omega = \frac{2\pi}{T}$. Once again, it is straight-forward to see that this linear transform is invertible: the matrix M is square, and the basis functions (i.e., the rows of M) are orthonormal, and thus the matrix is invertible. Note that this matrix formulation is for the Discrete Fourier transform (DFT), that is, discrete in both the space and frequency domain, whereas our discussion has been based on the DTFT (continuous in space and discrete in frequency), which of course cannot be represented in matrix notation. Nonetheless, both the DTFT and DFT are linear transforms.

It may appear that the Fourier transform is an overcomplete representation since an n-dimensional signal in the space domain is represented by a $2n$-dimensional signal in the frequency domain. That is, the Fourier transform is a n-dimensional *complex* valued signal, consisting of a real and imaginary component at each of the n samples. Similarly, the Fourier transform may appear overcomplete since it is equivalently represented by n magnitude and n phase components. However this is not the case, the Fourier transform is a rank preserving transform. The seemingly overcomplete representation is due to the symmetry properties of the Fourier transform, that is, for a real-valued signal, the Fourier transform is symmetric with respect to its origin (i.e., for each frequency component, the magnitude and phase for ω and $-\omega$ are equivalent). [12]

The Fourier transform has proven to be a useful and powerful tool for studying signals and images. For example, the classic noise removal method (the Wiener filter) is typically

[12] Another other useful symmetry property to note about the Fourier transform is that for a symmetric function in the space domain, the Fourier transform is purely real (i.e., consists of only cosine terms), and for an anti-symmetric function, the Fourier transform is purely imaginary (i.e., consists of only sinusoid terms). This of course makes sense, since the cosine and sinusoids exhibit the same symmetry/anti-symmetry.

derived by observing that "natural" images tend to have a $\frac{1}{\omega}$ frequency response (i.e., images tend to have more power in the low frequencies). Standard image compression schemes (e.g., JPEG) compute a block (or local) Fourier transform and send only the coefficients of the low frequencies [13]. A simple example of this is illustrated in Figure 1.9. Some motion estimation techniques operate in the frequency domain by exploiting the fact that the temporal frequency response of a translating pattern lies along a line passing through the origin, and that the orientation of the line is proportional to speed. And, as we will see later, several range estimation techniques (e.g., range from focus and defocus) rely on analyzing the frequency response of several images taken under different optical settings.

1.3.2 Sampling and Reconstruction

With an understanding of the convolution operator and basic Fourier theory, we are now prepared to complete the sampling story: the reconstruction of a continuous signal from its discretely sampled counterpart. Special attention will be given to what information, if any, is lost during the transformation from a continuous to a discrete representation.

Recall that in the space (or time) domain, a continuous signal, $f(x)$, is sampled at a rate T by multiplying by an impulse train, $\overline{s}(x)$:

$$\overline{f}(x) \quad = \quad f(x) \cdot \overline{s}(x). \tag{1.32}$$

In the frequency domain, the above sampling operation is given by:

$$\overline{F}(\omega) \quad = \quad (F \star \overline{S})(\omega), \tag{1.33}$$

where $\star$ is the convolution operator, $F(\omega)$ is the Discrete-Time Fourier Transform (DTFT) of $f(x)$ (as presented in the previous section), and $\overline{F}(\omega)$ and $\overline{S}(\omega)$, are the Discrete Fourier Transforms (DFT) of $\overline{f}(x)$ and $\overline{s}(x)$, respectively. [14] The Fourier transform of an impulse

[13] Actually, JPEG employs a block Discrete Cosine Transform (DCT), which has a slightly different basis than the Fourier Transform but is similar in spirit.

[14] As noted, the Fourier transform of the product of two signals is the convolution of their individual Fourier transforms. The dual property also holds, the Fourier transform of the convolution of two signals is the product of their individual Fourier transforms. A proof for the latter is given here. Consider the Fourier series (Equation (1.26)) of a signal, $g(x)$: $g(x) = \sum_{\omega} G(\omega) e^{j\omega x}$. Our goal is then to prove that if $G(\omega)$ is equal to the convolution of, say, $F(\omega)$ and $H(\omega)$, then $g(x) = f(x) \cdot h(x)$. Substituting $(F \star H)(\omega)$

train, $\overline{s}(x)$, is again an impulse train [Oppenheim 83]:

$$\overline{S}(\omega) \;=\; \sum_{k=-\infty}^{\infty} \overline{\delta}(\omega - k\omega_s), \tag{1.34}$$

where $\omega_s = \frac{2\pi}{T}$. Note that the spacing between impulses in the frequency domain is *inversely* proportional to the spacing, T, in the space domain. Now, convolving an impulse train, $\overline{S}(\omega)$, with $F(\omega)$ in Equation (1.33) gives:

$$\overline{F}(\omega) \;=\; \sum_{k=-\infty}^{\infty} F(\omega - k\omega_s), \tag{1.35}$$

an infinite superposition of translated copies of $F(\omega)$ (i.e., the Fourier transform of the original continuous signal):

Illustrated in Figure 1.10 is an example of the effects of sampling in the space and frequency domains. At the top of this figure is a continuous signal, and below it are sampled version of the signal sampled at a rate T and $2T$. Note that when sampling at a rate T the copies of the Fourier transform overlap, while at twice this sampling rate, the copies no longer overlap. This is due to the fact that by increasing the sampling rate, the distance between the impulses *decreases* in the spatial domain, and *increases* in the frequency domain. Since sampling in the frequency domain is defined as a convolution with the impulse train, this increase in the distance between impulses makes it less likely that the copies of the Fourier transform will overlap.

As will be seen shortly, if the copies of the Fourier transform do not overlap (i.e., the sampling rate is sufficiently high), then the continuous signal can theoretically be recovered *exactly* from its discretely sampled counterpart. That is, no information is lost during

into the Fourier series for $G(\omega)$ followed by a few algebraic manipulations gives:

$$\begin{aligned}
g(x) &= \sum_{\omega} (F \star H)(\omega) e^{j\omega x} \\
&= \sum_{\omega} \left(\sum_{k} F(k) H(\omega - k) \right) e^{j\omega x} \;=\; \sum_{k} F(k) \sum_{\omega} H(\omega - k) e^{j\omega x} \\
&= \sum_{k} F(k) \sum_{l} H(l) e^{j\omega(l+k)} \;=\; \sum_{k} F(k) e^{j\omega k} \sum_{l} H(l) e^{j\omega l} \\
&= f(x) \cdot h(x).
\end{aligned}$$

Recall that in the space domain, the sampled signal $\overline{f}$ could be determined from the convolved signal, $\overline{g}$, by multiplying by the inverse of the convolution matrix. In the frequency domain, this operation reduces to a division by the Fourier transform of the convolution kernel, $H(\omega)$. As such, convolution is invertible only if the Fourier transform of the kernel is non-zero.

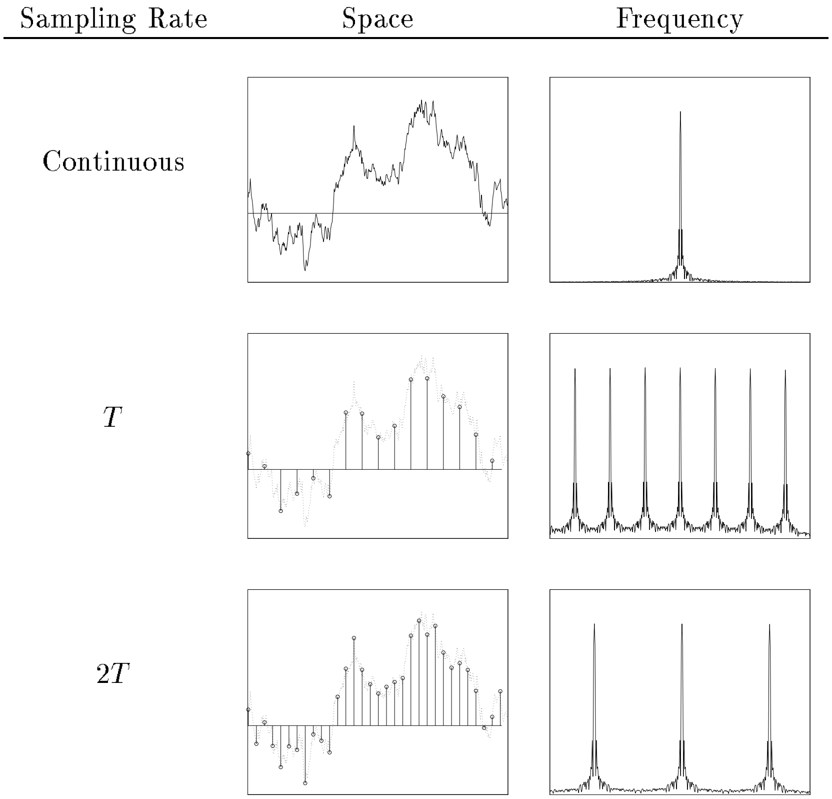


Figure 1.10: Sampling in Space and Frequency. Illustrated is a continuous and sampled (at a rate T and $2T$) signal (left) and its Fourier transform (right). First note that sampling in the space domain leads to periodicity in the frequency domain (i.e., the Fourier transform of the sampled signal is an infinite superposition of copies of the original Fourier transform). When sampled at a rate T, the copies of the Fourier transform overlap and information is lost. When sampled at a rate $2T$, the copies of the Fourier transform do not overlap, and a perfect copy of the Fourier transform of the original continuous signal remains in tact (see also Figure 1.11).

sampling. On the other hand, if the copies of the Fourier transform do overlap, the continuous signal cannot be recovered exactly, and in this case, the reconstructed signal will be said to be aliased.

Since the Fourier transform of the sampled signal contains copies of the Fourier transform of the original *continuous* signal, we need only extract one of these copies and inverse Fourier transform this copy to return to the original signal. The extraction of a single copy can be accomplished by multiplying the Fourier transform of the sampled signal with an ideal reconstruction boxcar filter (Figure 1.11). This filter has a value of 1 over the span

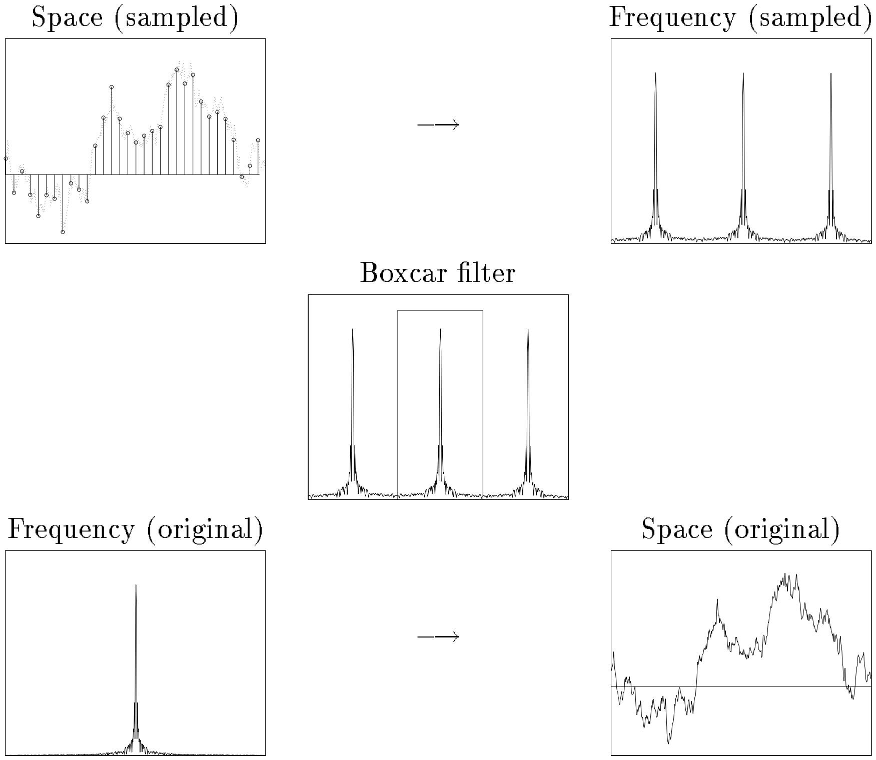


Figure 1.11: Sampling and Reconstruction in the Frequency Domain. Illustrated is a sampled signal and its Fourier transform (top), the application of a boxcar filter to extract a single copy of the Fourier transform (middle) and the reconstruction of the original signal (bottom). See also Figure 1.10.

of non-zero frequencies and is 0 elsewhere. This filter returns a single copy of the Fourier transform. Inverse Fourier transforming returns the original signal. This sequence of steps is illustrated in Figure 1.11.

The precise conditions under which a continuous signal can be recovered *exactly* from its samples can now be stated formally: let ω_N be the highest non-zero frequency component of the Fourier transform, $F(\omega)$, of the original continuous signal, $f(x)$, and let the sampling rate T equal $\frac{2\pi}{\omega_s}$. The replicas of $F(\omega)$ will not overlap when $\omega_s > 2\omega_N$ [15] in which case $f(x)$ can be recovered exactly from its samples. If this inequality does not hold, then the copies will overlap and $f(x)$ cannot be recovered exactly from its samples, and is said to be aliased. In other words, if the original signal is appropriately *bandlimited* (i.e., the Fourier series in Equation (1.15) consists of a finite and small number of frequencies), then it can

[15] The frequency ω_N is commonly referred to as the Nyquist frequency and $2\omega_N$ as the Nyquist rate.

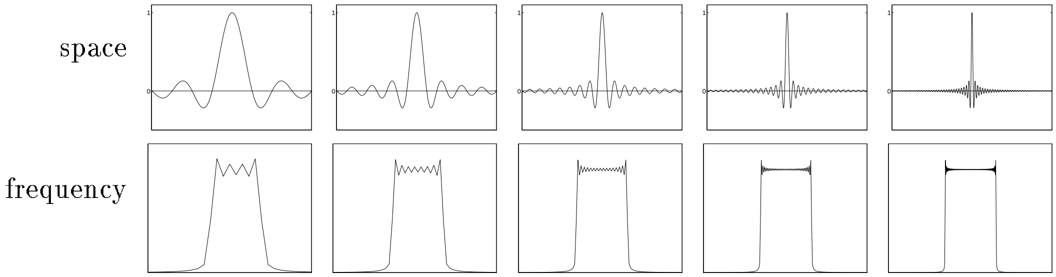


Figure 1.12: Ideal Reconstruction Function. Illustrated are truncated ideal reconstruction functions ($\frac{\sin(\pi x/T)}{\pi x/T}$) in both the space and frequency domain. If a signal is sampled above the Nyquist rate, then convolution with an *infinite* version of this function will return an *exact* copy of the original continuous signal. However, convolution with a truncated copy will introduce artifacts in the high frequencies (i.e., the Fourier transform is not a perfect boxcar function, it is not identically 1 over the range of non-zero frequencies, see also Figure 1.11). Unrelated to these artifacts, note that when centered about any integral number of the sampling period, the function has a zero value at each integer value, except at the point at which it is centered, where the value is one. This of course makes sense, since it is precisely at the integer values that the value of the continuous signal is known. At non-integer values, the continuous signal is computed from a weighted average of neighboring samples.

be sampled at or above the Nyquist rate without any loss of information.

The frequency-domain reconstruction of a continuous signal has a parallel in the space domain. Recalling that multiplication in the frequency domain is equivalent to convolution in the space domain, the original signal can be reconstructed by convolving the sampled signal with the inverse Fourier transform of the ideal boxcar function. This function (Figure 1.12) can be expressed analytically as, $\frac{\sin(\pi x/T)}{\pi x/T}$, where T is the sampling rate. Note that this filter has infinite spatial extent, and is thus often impractical to implement. A finite length filter can be constructed by sampling and truncating this function, however, as illustrated in Figure 1.12, this introduces errors in the high frequencies (i.e., the Fourier transform of the truncated filter is not a perfect boxcar filter). A further problem with the ideal reconstruction filter is that since it falls to zero quite slowly, the effects of truncation are significant for even fairly large filters. To overcome these problems, a “gentler” filter is often employed, frequently a Gaussian.

Finally, in keeping with the linear algebraic formulations of the previous stages of image formation, we note that sampling is also a linear operation and can thus be expressed in terms of simple matrix manipulations. Since such a formulation will require finite length signals, we will concern ourselves with the subsampling and reconstruction of a discrete signal (e.g., starting with a n-dimensional signal, subsample to a $n/2$-dimensional signal, and then reconstruct the original n-dimensional signal). Although this formulation is slightly different than the continuous to discrete sampling described above, the principles are similar.

Let $\vec{f}_n$ be a n-dimensional signal and $\vec{g}_m$ be a m-dimensional signal ($m < n$) generated by subsampling by a factor of n/m. This linear operation can be expressed as a matrix operation:

$$\begin{aligned} \vec{g}_m &= \begin{pmatrix} 1 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 & 0 & 0 & 0 \\ \vdots & & & & \ddots & & & & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & 1 \end{pmatrix}_{m \times n} \vec{f}_n \\ &= S_{m \times n} \vec{f}_n, \end{aligned} \tag{1.36}$$

where in this example $m = n/2$, and S is referred to as the sampling matrix. Note that the spacing between the non-zero entries between rows determines the subsampling rate, in this case we are subsampling by a factor of two and every other sample is discarded. Now, recall the constraint that allowed us to exactly reconstruct a sampled signal: for a fixed sampling rate, the original signal must be bandlimited, that is, it must be expressible as a *finite* sum of basis functions (in our case, sines and cosines), where the number of basis functions is related to the sampling rate. In matrix notation, this constraint may be expressed as:

$$\vec{f}_n = B_{n \times m} \vec{w}_m, \tag{1.37}$$

where the columns of the matrix B contain the sampled basis functions, and the vector $\vec{w}$ is the appropriate weighting of each basis function (e.g., the Fourier transform). In this example, the original n-dimensional signal $\vec{f}$ is written as a linear combination of m basis functions, that is, it actually lies in a m-dimensional subspace of the larger n-dimensional space. Combining this expression with the sampling matrix (Equation (1.36)) yields:

$$\begin{aligned}\vec{g}_m &= S_{m\times n}\left(B_{n\times m}\vec{w}_m\right) \\ &= M_{m\times m}\vec{w}_m, \qquad (1.38)\end{aligned}$$

where the columns of the matrix M contain the sampled basis functions. Since the original basis set spans the full m-dimensional space, the matrix M is guaranteed to be invertible. In addition, if the basis is chosen to be orthonormal, then the inverse is simply M^t, and:

$$\vec{w}_m = M^{-1}_{m\times m}\vec{g}^t_m. \qquad (1.39)$$

Finally, given the weights $\vec{w}$, the original n-dimensional signal, $\vec{f}$ can be reconstructed (Equation (1.37)). That is, given the subsampled signal, $\vec{g}$, the original signal can be reconstructed *exactly*: $\vec{f} = BM^{-1}\vec{g}$.

This result should not be entirely surprising in that the original n-dimensional signal is contained in a m-dimensional subspace, where $m < n$ (i.e., the original signal can be expressed as a linear combination of m basis functions, Equation (1.37)). As such, the original n-dimensional representation of the signal is overcomplete, and only m-samples are required to fully represent the signal. [16]

This brings to an end this section on digital image formation. Before beginning our study of how three-dimensional properties of the world can be recovered from such images, we take another slight detour. Since we are interested in differential approaches to the range

[16] Consider the following simple example. Let $\vec{f} = (2 \;\; 0 \;\; 4 \;\; 0)^t$, and $\vec{g} = (2 \;\; 4)^t$ obtained by subsampling $\vec{f}$ by a factor of two $\left(\vec{g} = S\vec{f}, \text{ where } S = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}\right)$. Noticing that $\vec{f}$ actually lies in a two-dimensional subspace (i.e., a plane) of the original four-dimensional space, $\vec{f}$ is expressed as a linear combination of the two canonical basis vectors (the columns of the matrix B): $\vec{f} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 2 \\ 4 \end{pmatrix} = B\vec{w}$. According to our derivation above, the original vector $\vec{f}$ can be reconstructed from the subsampled version, $\vec{g}$ as:

$$\vec{f} = B(SB)^{-1}\vec{g} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}\left(\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}\right)^{-1}\begin{pmatrix} 2 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{-1}\begin{pmatrix} 2 \\ 4 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 4 \\ 0 \end{pmatrix}.$$

Of course, in this example the canonical basis was used instead of the Fourier basis (i.e., sines and cosines). Nonetheless, it should be clear that the subsampling and reconstruction are independent of the particular choice of basis (i.e., the matrix B).

recovery problem, it is worthwhile to first discuss some issues involved in the differential analysis of digital imagery.

1.4 Image Derivatives

Although the details of this next section are not essential to the general understanding of subsequent chapters, it is important to appreciate the difficulties that arise when measuring differential quantities. Furthermore we believe that the careful design of derivative filters is essential for many vision and image processing tasks.

We begin by considering a continuous and differentiable function, $f : \mathcal{R}^2 \rightarrow \mathcal{R}$, and the discretely sampled function $\overline{f} : \mathcal{Z}^2 \rightarrow \mathcal{R}$. The question which we address is that of computing partial derivatives of the discretely sampled function, $\overline{f}(x,y)$. Strictly speaking, such a sampled function cannot be differentiated. But, as was shown in the previous section, if the original underlying continuous function, $f(x,y)$, is sampled above the Nyquist rate, then the *sampled* derivative of the original continuous function can be determined. More specifically, if the function $\overline{f}(x,y)$ was sampled at a rate T, above the Nyquist rate, then the original continuous function can be reconstructed *exactly* as:

$$f(x,y) \quad = \quad (S \star \overline{f})(x,y), \tag{1.40}$$

where $S(x,y) = \frac{\sin(xT)\sin(yT)}{xyT^2}$ is the ideal-sinc function, and $\star$ is the convolution operator defined as, $(S \star \overline{f})(x,y) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} S(x-i, y-j)\overline{f}(i,j)$. Since both sides of the above equation are continuous, the differential operator, $D(\cdot)$, can be applied to both sides of this equation. For example, the partial derivative with respect to x is:

$$\begin{aligned} D_x(f) &= D_x(S \star \overline{f}) \\ &= D_x(S) \star \overline{f} \\ &= S_x \star \overline{f} \\ &\equiv f_x, \end{aligned} \tag{1.41}$$

where for notational clarity, the spatial parameters are dropped. The final derivative

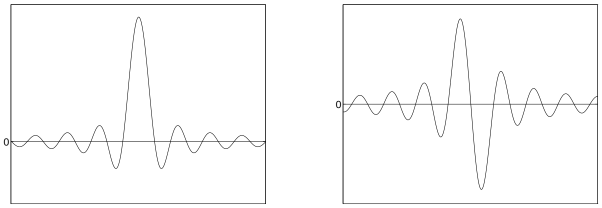


Figure 1.13: Ideal Reconstruction Function and its Derivative. Illustrated on the left is the truncated ideal reconstruction function ($p(x) = \frac{\sin(\pi x/T)}{\pi x/T}$) and on the right, its derivative ($D(p(x)) = \frac{\pi^2 x/T^2 \cos(\pi x/T) - \pi/T \sin(\pi x/T)}{(\pi x/T)^2}$). Note that in addition to being infinite in extent, the functions fall off gradually from the origin. As a result, truncation introduces artifacts in the high-frequencies (see also Figure 1.12).

measurement is determined by sampling the right-hand side of the above equation:

$$\begin{aligned} \overline{D_x(f)} &= \overline{(S_x \star \overline{f})} \\ &\equiv \overline{f}_x. \end{aligned} \tag{1.42}$$

Alternatively, and of more practical interest, the continuous function S_x can be sampled first and then applied to the sampled signal:

$$\overline{f}_x = \overline{S}_x \star \overline{f}. \tag{1.43}$$

In a similar way, the partial derivative of f with respect to y is given by $\overline{f}_y = \overline{S}_y \star \overline{f}$. Of course, the ideal-sinc function, S, is spatially infinite in extent making it computationally intractable to implement (Figure 1.13). In addition, this function has considerable energy in its tails, and as a result truncation introduces substantial artifacts. A more localized function, P, may be chosen and its sampled partial derivatives, P_x or P_y, applied as in Equation (1.43). Typically one would like to choose such a function based on the bandlimitedness of the function f and the sampling rate.

A second constraint that one may like to impose on the function P is that of xy-separability. A function is said to be xy-separable if it can be expressed as a product of two "one-dimensional" functions:

$$P(x, y) = p_1(x, y) \cdot p_2(x, y), \tag{1.44}$$

where, p_1 is independent of y and p_2 is independent of x. This constraint clearly has important implications in terms of computational efficiency. For if a function has this property then the 2-D convolution can be replaced with a pair of 1-D convolutions. [17] A further benefit of this constraint is that the 2-D design problem has been reduced to a simpler 1-D design problem, and as a result, the filters will extend naturally to multi-dimensional functions. Given this separability constraint, let's see how it affects the derivative calculation in Equation (1.41), (as before, the spatial parameters are dropped for notational clarity):

$$\begin{aligned} D_x(f) &= D_x(P \star \overline{f}) \\ &= D_x((p_1 \cdot p_2) \star \overline{f}) \\ &= D_x(p_1 \star p_2 \star \overline{f}) \\ &= D_x(p_1) \star (p_2 \star \overline{f}) \\ &= d_1 \star (p_2 \star \overline{f}), \end{aligned} \tag{1.45}$$

where the convolution with d_1 is along the x dimension, and the convolution with p_2 is along the y dimension (i.e., the convolutions are one-dimensional, see footnote 17). When computing directional derivatives we can ensure that the directional differentiation and prefiltering along various axis are the same by strengthening the separability constraint such that $p_1(u, \cdot) = p_2(\cdot, u)$ where the one-dimensional function is denoted simply as $p(u)$. Under this notation, the above equation can be expressed as:

$$D_x(f) \;=\; d \star (p \star \overline{f}), \tag{1.46}$$

where $D(p(u)) = d(u)$. As will be seen shortly, this seemingly simple constraint will turn out to play a key role in our filter design. Note again that the convolutions with d and p are one-dimensional along the x and y dimensions, respectively. As in Equation (1.43), the final sampled derivative is obtained by first sampling the continuous functions d and

[17]

$$\begin{aligned} (P \star \overline{f})(x,y) &= \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} P(x-i,y-j)\overline{f}(i,j) = \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} p_1(x-i,y-j)p_2(x-i,y-j)\overline{f}(i,j) \\ &= \sum_{i=-\infty}^{\infty} p_1(x-i,y) \sum_{j=-\infty}^{\infty} p_2(x,y-j)\overline{f}(i,j) = \sum_{i=-\infty}^{\infty} p_1(x-i,y)(p_2 \star \overline{f})(i,y) \\ &= (p_1 \star (p_2 \star \overline{f}))(x,y), \end{aligned}$$

where, since p_1 and p_2 are dependent on only one of their variables (i.e., P is xy-separable), the convolution with p_1 is along the x dimension, and the convolution with p_2 is along the y dimension, i.e., 1-D convolutions.

p and then convolving them with $\overline{f}$ along the appropriate dimensions. For example, the partial derivatives in x and y are given by:

$$\overline{f}_x = \overline{d}^t \star (\overline{p} \star \overline{f}) \qquad \text{and} \qquad \overline{f}_y = \overline{d} \star (\overline{p}^t \star \overline{f}), \tag{1.47}$$

where $\overline{d}$ and $\overline{p}$ are column vectors applied along the y dimension, and $\overline{d}^t$ and $\overline{p}^t$ are row vectors applied along the x dimension.

It may appear that, as before, the filter design problem reduces to simply choosing an interpolation function $P(x, y) = p_1(x, y) \cdot p_2(x, y)$, with the additional constraint that $p_1(u, \cdot) = p_2(\cdot, u) \equiv p(u)$. For any such function, the required derivative relationship between the resulting 1-D functions, $d(u)$ and $p(u)$, is automatically satisfied (i.e., $D(p(u)) = d(u)$). For example, a unit-variant Gaussian, $P(x, y) = \frac{1}{2\pi^2} e^{-(x^2+y^2)/2}$ (which is separable, with $p(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$) leads to the pair of functions:

$$d(u) = \frac{-u}{\sqrt{2\pi}} e^{-u^2/2} \qquad \text{and} \qquad p(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}, \tag{1.48}$$

which clearly satisfies the required derivative relationship. But, recall that the actual filters, $\overline{d}$ and $\overline{p}$, are gotten by sampling these continuous functions. Due to artifacts introduced by sampling, the derivative relationship between these functions will typically be destroyed. To address this problem, we propose to simultaneously design a pair of filters that optimally (in a least-squares sense) preserves the required derivative relationship. A least-squares solution is formulated for the joint design of a 1-D prefilter and derivative filter pair, $\overline{p}$ and $\overline{d}$ (as in [Simoncelli 94]). This filter pair is designed so as to optimally preserve the derivative relationship: $\overline{D(p(u))} = \overline{d(u)}$. We begin with the design of a first-order derivative and prefilter, and then show how the basic constraint and design can be extended to higher-order derivative filters.

In the frequency domain, the derivative relationship between the derivative, $d(u)$, and prefilter, $p(u)$, becomes:

$$j\omega \mathcal{P}(\omega) \;\; = \;\; \mathcal{D}(\omega), \tag{1.49}$$

where, $\mathcal{P}(\omega)$ and $\mathcal{D}(\omega)$ are the Fourier transforms of the derivative and prefilter, respectively. [18] We write a weighted least-squares error function to be minimized:

$$E(\mathcal{P}, \mathcal{D}) \;\; = \;\; \int d\omega \;\; [W(\omega)(j\omega\mathcal{P}(\omega) - \mathcal{D}(\omega))]^2 \, , \tag{1.50}$$

[18] It is straight-forward to show that differentiation in the space domain is equivalent to multiplication by

where $W(\omega)$ is the frequency weighting function. We write a discrete approximation of Equation (1.50) over the m-vectors $\vec{p}$ and $\vec{d}$ containing the sampled derivative and prefilter, respectively:

$$E(\vec{p}, \vec{d}) \quad = \quad |W(F'\vec{p} - F\vec{d})|^2, \tag{1.51}$$

where the columns of the matrix $F_{n\times m}$ contain the first m Fourier basis functions, the matrix $F'_{n\times m}$ is $j\omega F_{n\times m}$ (i.e., an approximation to the discrete-time Fourier transform, DTFT), and $W_{n\times n}$ is a diagonal frequency weighting matrix. Note that the dimension m is determined by the filter size and the dimension n is the sampling rate of the continuous Fourier basis functions, which should be chosen to be sufficiently large to avoid sampling artifacts. Equation (1.51) can be expressed more concisely as:

$$E(\vec{u}) \quad = \quad |M\vec{u}|^2, \tag{1.52}$$

where the matrix M and the vector $\vec{u}$ are constructed by "packing together" matrices and vectors:

$$M = (\, WF' \quad | \quad -WF \,) \qquad \text{and} \qquad \vec{u} = \begin{pmatrix} \vec{p} \\ \vec{d} \end{pmatrix}. \tag{1.53}$$

The minimal unit vector $\vec{u}$ is then simply the minimal-eigenvalue eigenvector of the matrix M^tM. [19] The derivative and prefilter are then normalized so that the prefilter has

an imaginary ramp in the frequency domain. Consider first the Fourier series representation of the function $f(x)$: $f(x) = \int d\omega \;\, F(\omega)e^{j\omega x}$, where $F(\omega)$ is the Fourier transform of $f(x)$. Computing the derivative with respect to x gives: $\frac{df(x)}{dx} = \frac{d}{dx}\int d\omega \;\, F(\omega)e^{j\omega x} = \int d\omega \;\, F(\omega)\frac{d}{dx}e^{j\omega x} = \int d\omega \;\, j\omega F(\omega)e^{j\omega x}$. That is, differentiation in the space domain is equivalent to multiplication of the Fourier transform by an imaginary ramp, $j\omega$.

[19] An error function of the form $E(\vec{u}) = |M\vec{u}|^2$ can be minimized analytically using eigenvector techniques. In particular, the minimal unit vector, $\vec{u}$, is the minimal-eigenvalue eigenvector of the matrix M^tM; (see [Strang 88] for more details). Since we are interested in the minimal unit vector, the error function can be expressed as $E(\vec{u}) = \frac{|M\vec{u}|^2}{\vec{u}^t\vec{u}}$. Expanding the numerator gives: $E(\vec{u}) = \frac{\vec{u}^tM^tM\vec{u}}{\vec{u}^t\vec{u}}$. The matrix M can be decomposed, using a singular value decomposition (SVD), into a product of three matrices: $M = O_1DO_2$, where O_1 and O_2 are orthonormal, and D is a diagonal matrix. Substituting into the numerator of $E(\vec{u})$ gives $\vec{u}^t(O_2^tDO_1^t)(O_1DO_2)\vec{u} = \vec{u}^tO_2^tD^2O_2\vec{u}$ (note that since O_1 is orthonormal, $O_1^tO_1$ is the identity matrix, and since D is diagonal, the matrix $D^2 = DD$ is computed by squaring each of the diagonal elements of the matrix D).

The numerator of our error function, $\vec{u}^tO_2^tD^2O_2\vec{u}$, may be interpreted geometrically as a rotation (i.e., O_2 is orthonormal), dilation (i.e., D is diagonal), and rotation of our space of possible solutions (i.e., vectors, $\vec{u}$ lying on the unit circle). Illustrated below is an example for a two-dimensional space. The resulting space of solutions is an ellipse (in higher-dimensions, an ellipsoid). Finally, in order to minimize $E(\vec{u})$, we would like to maximize, the denominator of our error function, $\vec{u}^t\vec{u}$. In the geometric sense, we are looking for the vector lying along the unit sum. See [Farid 97] for a constrained least-squares minimization (using Lagrange multipliers) that incorporates this normalization into the error function.

We have created a matched derivative/prefilter set using the above design criteria (throughout, a frequency weighting function of $\frac{1}{|\omega|}$ was used, approximating the power spectrum of most "natural" images). Illustrated in Figure 1.14 are the magnitudes of the Fourier transforms of the derivative filter, $\vec{d}$, and $|j\omega|$ times the prefilter, $\vec{p}$ (i.e., the frequency domain derivative of the prefilter). If the filters were perfectly matched (i.e., one is the derivative of the other), then the two should coincide exactly: notice that this is nearly the case for the 5-tap filters (see Figure 1.17 for filter tap values). For comparison, filter pairs based on a truncated sinc function are also shown in Figure 1.14.

Higher-order derivative filters may be designed using a similar strategy. The design of an N^{th}-order derivative is similar to that of the first-order derivative, with the constraint of Equation (1.49) replaced with:

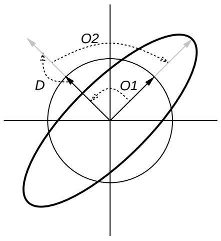
$$(j\omega)^N \mathcal{P}(\omega) \quad = \quad \mathcal{D}_N(\omega), \tag{1.54}$$

that is, the derivative filter is constrained to be the N^{th} derivative of the prefilter. If, on the other hand, we are interested in the simultaneous design of the first N derivative filters, the following set of constraints may be considered:

$$(j\omega)^{n-m} \mathcal{D}_m(\omega) \quad = \quad \mathcal{D}_n(\omega), \tag{1.55}$$

for $0 \leq n, m \leq N$ and $m < n$ and where $\mathcal{D}_n(\omega)$ denotes the Fourier transform of the n^{th} derivative filter and $\mathcal{D}_0(\omega)$ denotes the Fourier transform of the prefilter. Using the full

major axis of the ellipse, this vector is the minimal-eigenvalue eigenvector of the matrix M^tM.



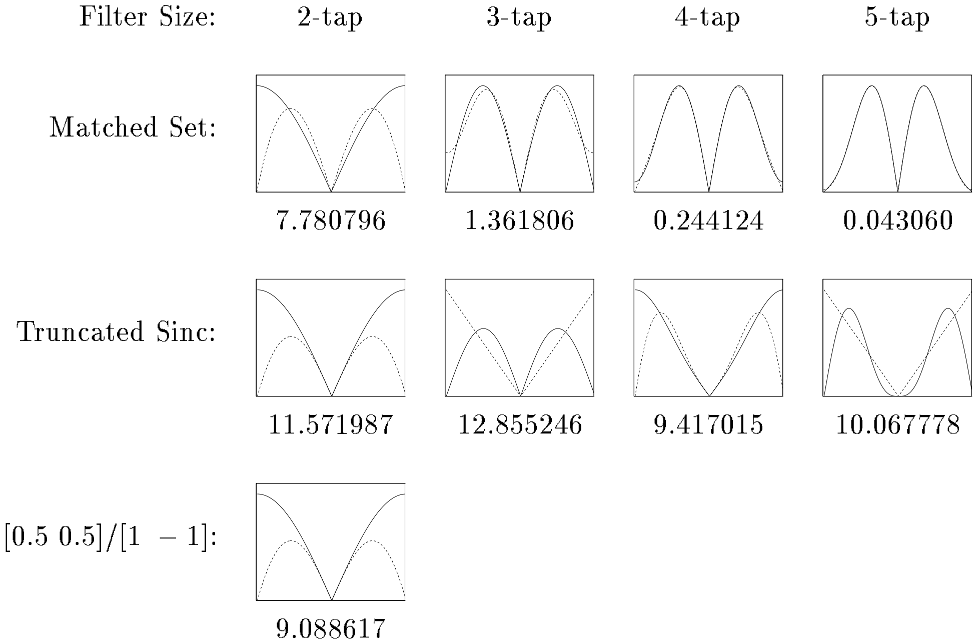


Figure 1.14: Prefilter and First-Order Derivative Filter Pairs (frequency domain). Illustrated in each panel is the magnitude of the Fourier transform of the derivative filter (solid line) and the frequency-domain derivative of the prefilter (dashed line), that is, the prefilter multiplied by $|j\omega|$. Shown are matched filters based on a total least-squares optimal design criteria (Equation (1.50)), a truncated ideal reconstruction prefilter and its derivative, and a commonly used 2-tap filter pair. If the filters were perfectly matched then the curves would coincide: this is nearly the case for the optimally designed 5-tap filters. Beneath each plot is the rms error.

set of $\frac{N(N+1)}{2}$ constraints, the equivalent matrix, M, in Equation (1.53) takes the form:

$$M = \left(\begin{array}{c|c|c|c|c|c|c|c}
j\omega WF & -WF & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & j\omega WF & -WF & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & j\omega WF & -WF & \cdots & 0 & 0 & 0 \\
\vdots & & & & \ddots & & & \vdots \\
0 & 0 & 0 & 0 & \cdots & j\omega WF & -WF & 0 \\
0 & 0 & 0 & 0 & \cdots & 0 & j\omega WF & -WF \\
(j\omega)^2 WF & 0 & -WF & 0 & \cdots & 0 & 0 & 0 \\
0 & (j\omega)^2 WF & 0 & -WF & \cdots & 0 & 0 & 0 \\
\vdots & & & & \ddots & & & \vdots \\
0 & 0 & 0 & & \cdots & (j\omega)^2 WF & 0 & -WF \\
(j\omega)^3 WF & 0 & 0 & -WF & \cdots & 0 & 0 & 0 \\
\vdots & & & & \ddots & & & \vdots \\
(j\omega)^n WF & 0 & 0 & & \cdots & 0 & 0 & -WF
\end{array} \right) \qquad (1.56)$$

where, as before, the columns of the matrix $F_{n \times m}$ contain the first m Fourier basis

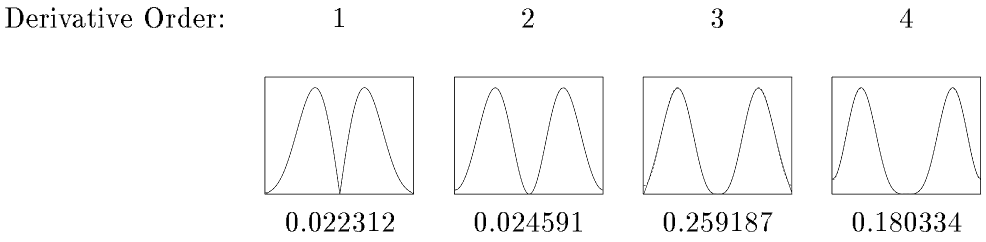


Figure 1.15: 7-tap Prefilter and First Through Fourth Derivative Filter Pairs (frequency domain). Illustrated is the magnitude of the Fourier transform of each derivative filter (solid line), and the frequency-domain derivative of the prefilter, i.e., the prefilter multiplied by $|j\omega|^n$ (dashed line). If the filters were perfectly matched then the curves would coincide exactly: notice that this is nearly the case for these filters; since the matches are nearly perfect, it is difficult to see both curves. Beneath each plot is the rms error.

functions, $W_{n \times n}$ is a diagonal frequency weighting matrix, and the vector $\vec{u}$ is given by:

$$\vec{u} \;=\; (\vec{p} \;\;|\;\; \vec{d}_1 \;\;|\;\; \vec{d}_2 \;\;|\;\; \ldots \;\;|\;\; \vec{d}_N\,)^t\,, \tag{1.57}$$

where, $\vec{p}$ is the prefilter and $\vec{d}_n$ is the n^{th} derivative filter. Again, the unit vector, $\vec{u}$, that minimizes the error function $E(\vec{u}) = |M\vec{u}|^2$ is the minimal-eigenvalue eigenvector of the matrix M^tM.

A 7-tap prefilter and a set of matched derivative filters (first through fourth-order) were created using the above design criteria (a frequency weighting function of $\frac{1}{|\omega|}$ was again used). Illustrated in Figure 1.15 are the magnitudes of the Fourier transforms of each derivative filter, $\vec{d}_n$, and $|j\omega|^n$ times the prefilter, $\vec{p}$. If the filters were perfectly matched (i.e., if the derivative filter were the n^{th} derivative of the prefilter), then the two should coincide exactly: notice that this is nearly the case for each of the derivatives (see Figure 1.17 for filter tap values).

It is important to mention that these filters are *not* guaranteed to return the exact derivative of the original continuous function. In fact, this was not even one of our design criteria. Rather, we are more interested in ensuring that the directional derivatives are properly matched. That is, that when computing directional derivatives they are with respect to the same underlying function. This relationship is especially important in a variety of tasks that compare directional derivatives, for example, computation of the gradient, orientation analysis, and motion estimation.

1.5 Summary

We began this chapter by arguing that a thorough understanding of the image formation process would aid us in the subsequent processing of digital images. Various aspects of image formation were discussed: the geometry of several camera models, the physics of charge-coupled devices (CCD), and the mathematics of sampling theory. In addition to the standard formulations, we showed that most of the various stages of image formation are linear and can therefore be considered within a linear algebraic framework. The discussion on image formation also provided a convenient platform for introducing two fundamental tools to signal and image processing, convolution and Fourier transforms, each of which will be used extensively in subsequent chapters. We concluded with some important issues in the design of digital derivative filters. Perhaps the most important point of this chapter is that image formation is a 3-D to 2-D transformation, and that inherent to such a transformation is a loss of information about the depth of objects in the world. This information can, however, be recovered from multiple images, and it is this recovery process that will be the central theme of the subsequent chapters.

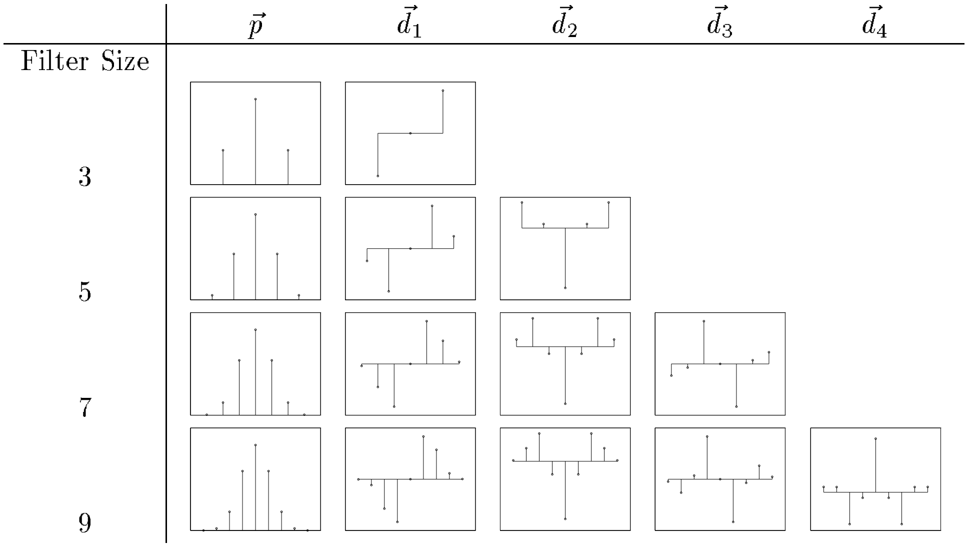


Figure 1.16: Prefilter and Derivative Filter Pairs (space domain). Illustrated are 3- to 9-tap matched prefilter ($\vec{p}$) and first through fourth derivative filter pairs ($\vec{d}_1$, $\vec{d}_2$, $\vec{d}_3$, and $\vec{d}_4$). See Figure 1.17 for actual tap values.

	-4	-3	-2	-1	0	1	2	3	4
$\vec{p}$				0.22274	0.55451	0.22274			
$\vec{d}_1$				-0.45805	0.00000	0.45805			
$\vec{p}$			0.02475	0.24629	0.45789	0.24629	0.02475		
$\vec{d}_1$			-0.09205	-0.31838	0.00000	0.31838	0.09205		
$\vec{d}_2$			0.23426	0.03754	-0.54599	0.03754	0.23426		
$\vec{p}$		0.00194	0.05706	0.24788	0.38622	0.24788	0.05706	0.00194	
$\vec{d}_1$		-0.01087	-0.12159	-0.22471	0.00000	0.22471	0.12159	0.01087	
$\vec{d}_2$		0.04247	0.16700	-0.04150	-0.33525	-0.04150	0.16700	0.04247	
$\vec{d}_3$		-0.11353	-0.03588	0.41492	0.00000	-0.41492	0.03588	0.11353	
$\vec{p}$	0.00011	0.00840	0.07614	0.24158	0.34755	0.24158	0.07614	0.00840	0.00011
$\vec{d}_1$	-0.00077	-0.02454	-0.12261	-0.17808	0.00000	0.17808	0.12261	0.02454	0.00077
$\vec{d}_2$	0.00401	0.05676	0.12045	-0.05646	-0.24952	-0.05646	0.12045	0.05676	0.00401
$\vec{d}_3$	-0.01631	-0.09080	0.02470	0.28794	0.00000	-0.28794	-0.02470	0.09080	0.01631
$\vec{d}_4$	0.04967	0.04872	-0.29623	-0.05163	0.49880	-0.05163	-0.29623	0.04872	0.04967

Figure 1.17: Prefilter and Derivative Filter Taps. Filter taps correspond to the matched prefilter ($\vec{p}$) and first through fourth derivative filter pairs ($\vec{d}_1$, $\vec{d}_2$, $\vec{d}_3$, and $\vec{d}_4$) illustrated in Figure 1.16.

Chapter 2

Range Estimation: Overview

2.1 Introduction

In the previous chapter we saw that image formation is a three-dimensional to two-dimensional transformation. The transformation is not one-to-one and thus not invertible: lost in this transformation is the three-dimensional structure of the world (as illustrated in Figure 2.1). Assuming no prior information and a standard imaging system, the full three-dimensional structure of the world *cannot* be recovered from a single two-dimensional image.

In addition to revealing where information is lost, the image formation models presented in the previous chapter also suggest methods for recovering this information. For example, under perspective projection (Equation (1.1)), the projection of a point in the world is inversely proportional to its depth. As such, the range can be estimated by measuring the change in the projection of a point from different viewing positions. Based on the measurement techniques, such approaches are referred to broadly as range from stereo (see [Dhond 89, Koschan 93, Ozanian 95] for reviews) or motion (e.g., [Koenderink 91, Heeger 92, Tomasi 92, Costeira 95]).

The thin-lens model suggests a different method for recovering range. Recall that under the thin-lens model (Equation (1.7)), a point in the world is imaged as a blurred circle where the radius of the blur circle is a function of its depth. Range can be estimated by measuring the change in the amount of blur as a function of different optical settings.

Based on the particular changes in optical settings, such approaches are referred to as range from focus (e.g., [Krotkov 87, Grossman 87]) or defocus (e.g., [Pentland 87, Subbarao 88, Xiong 93, Nayar 95]).

A common property shared by these techniques is that of *measuring change*. As such, we will argue that it is natural to consider each of these techniques within a *differential* framework. In doing so, we will observe that the different range estimation techniques amount to computing *discrete* approximations to a derivative (with respect to different parameters). At this point, the significance of the previous section on image derivatives should be more apparent, and later on we will borrow heavily from many of the ideas presented there. But first, this chapter reviews the principles of range from stereo, motion, focus, and defocus as described above.

2.2 Range from Stereo

When a point $\vec{P}$ in the 3-D world is projected onto a pair of spatially offset imaging sensors, its image will fall on different relative locations on the two sensors (Figure 2.1). This difference is a function of depth: points closer to the sensor will be more disparate than more distant points. By measuring the difference or disparity between the projection of the same point onto a pair of imaging sensors (binocular stereo) range can be estimated. Although there are many variations, this section outlines a basic system of range from stereo.

Consider the binocular stereo configuration in Figure 2.1. In this figure the sensor nodal points are separated by a distance b, the baseline, and are a distance d_s from the sensor plane. The world point, $\vec{P} = (\, X \quad Y \quad Z\,)^t$, is at a distance Z from the sensor plane and the disparity between the image of $\vec{P}$ in both sensors, $\vec{p}_1 = (\, x_1 \quad y_1\,)^t$ and $\vec{p}_2 = (\, x_2 \quad y_2\,)^t$, is denoted by Δ. Similar triangles yields the relationship $\frac{Z}{b} = \frac{d_s}{\Delta}$. Combined with the perspective projection equations (Equation (1.1)), this relationship gives a simple expression for determining the position of the point in the three-dimensional world:

$$X = \frac{d_s x_1}{Z}, \qquad Y = \frac{d_s y_1}{Z} \qquad \text{and} \qquad Z = \frac{d_s b}{\Delta}. \tag{2.1}$$

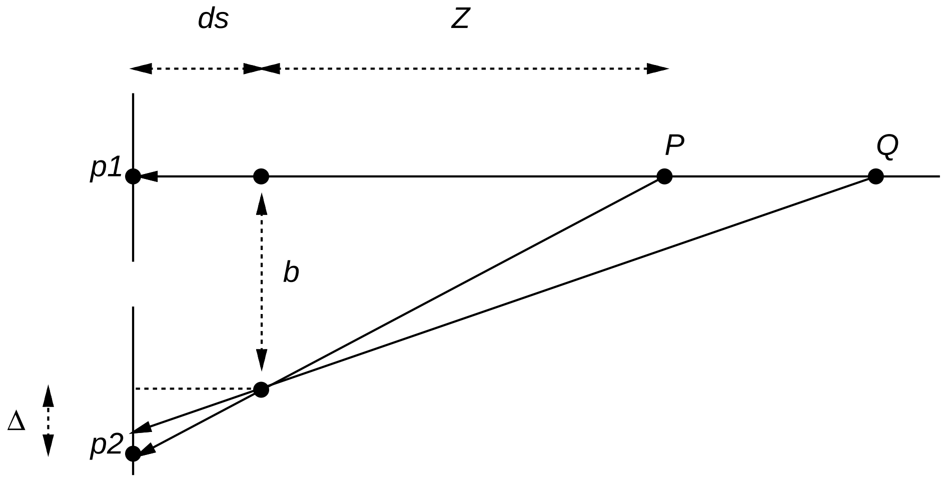


Figure 2.1: Range from Stereo. Illustrated are a pair of spatially offset pinhole cameras (by an amount b). The image of a point in the world, $\vec{P}$, on the camera sensors are denoted as $\vec{p}_1$ and $\vec{p}_2$. First, note that although the points $\vec{P}$ and $\vec{Q}$ are at different depths, they are imaged onto the same point in the sensor, $\vec{p}_1$. As such, the distance to these points cannot be determined from a single image. However, when imaged through a pair of spatially offset cameras, points near the sensor will produce a greater disparity, Δ, than more distant points. The distance, Z, to the point $\vec{P}$ can then be estimated from the relative positions of its projection onto the sensors (Equation (2.1)).

Assuming a known baseline, b, and sensor to nodal point distance, d_s, range, Z, can be estimated if the correspondence between the projection of the point in both images is known.

Given the projection of the point in one image $I_1(\cdot)$, (centered at (x_1, y_1)), the corresponding projection in the other image, $I_2(\cdot)$, can be determined by finding the point (x_2, y_2) that minimizes the following sum of squared differences (SSD) metric:

$$c(x_2, y_2) \quad = \quad \frac{\sum_{x=-n/2}^{n/2} \sum_{y=-n/2}^{n/2} (I_1(x_1 + x, y_1 + y) - I_2(x_2 + x, y_2 + y))^2}{n^2}, \qquad (2.2)$$

where the summation is performed over a small, $n \times n$, image patch centered at (x_1, y_1).

By limiting the geometry of the sensors so that their nodal points are only translated in the horizontal direction (i.e., a parallel optical axis configuration), the correspondence search can be simplified. In particular, the corresponding match for a point (x_1, y_1) is of the form (x_2, y_1), that is, the matching point lies along the same horizontal scan line in the digital image. With such a configuration, the correspondence search will yield a horizontal disparity. By computing the horizontal or vertical disparity at each small image patch in an image, the complete structure of the imaged three-dimensional world can be determined.

Stereo images are frequently displaced horizontally or vertically in order to simplify the correspondence search. In particular, with such a configuration, searching is performed along a single horizontal or vertical digital scan line (i.e., the epi-polar line). Of course, any pair (or more) of spatially offset cameras will suffice. As illustrated in Figure 2.2, a nonparallel stereo geometry leads to an oblique epi-polar line, complicating the correspondence search. [1] Also illustrated in this figure is a trinocular and multiocular stereo configuration where matching is performed across multiple images.

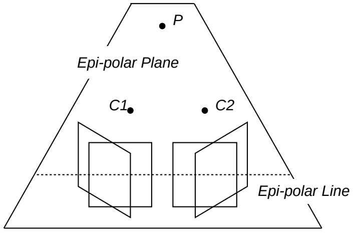
Due to the inherently discrete nature of the search for corresponding points (Equation (2.2)), this technique is frequently referred to as *discrete* matching. In contrast, we may consider a *differential* formulation of the correspondence problem. In particular, we write a quadratic error function (in 1-D, the extension to 2-D is straight-forward) over the disparity, 2Δ:

$$E(\Delta) \;=\; \sum_x \left(I_1(x+\Delta) - I_2(x-\Delta)\right)^2, \qquad (2.3)$$

where the summation is over the spatial image parameter, x. Performing a Taylor series expansion [2] and throwing away the higher-order terms gives:

$$\begin{aligned} E(\Delta) &\approx \sum_x \left((I_1(x) + \Delta I_1'(x)) - (I_2(x) - \Delta I_2'(x))\right)^2 \\ &= \sum_x \left(I_1(x) - I_2(x) + \Delta(I_1'(x) + I_2'(x))\right)^2 \end{aligned}$$

[1] Consider the parallel-axis stereo pair of cameras illustrated below (with camera centers, C_1 and C_2). The epi-polar plane is the plane passing through the point P and the camera centers, C_1 and C_2. The epi-polar line is then defined to be the intersection of the epi-polar and image planes. Note that in the case of a parallel axis geometry, the epi-polar line coincides with a single horizontal scan line of the digital image. In the case of a non-parallel axis geometry, the epi-polar line does not coincide with a horizontal scan line.



[2] The Taylor series expansion of $f(x+a)$ is given by the infinite sum: $f(x+a) = f(x) + \frac{af'(x)}{1!} + \frac{a^2 f''(x)}{2!} + \ldots + \frac{a^n f^n(x)}{n!} + \ldots$.

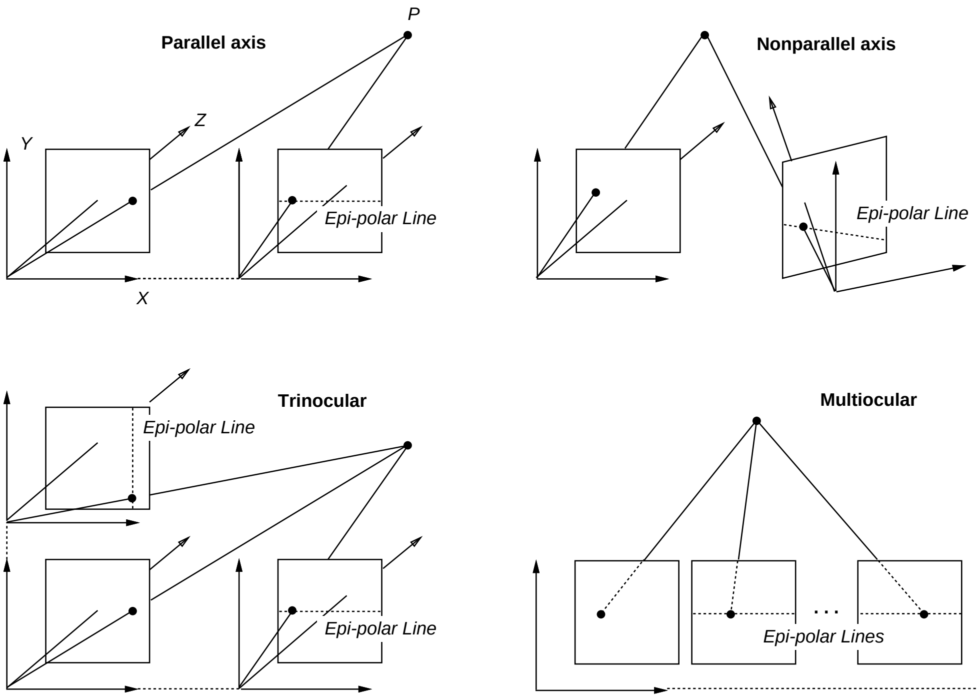


Figure 2.2: Stereo configurations. Illustrated are several possible stereo configurations. (1) binocular parallel optical axis, (2) binocular non-parallel optical axis, (3) trinocular, and (4) multiocular. The epi-polar line defines the constraint line for correspondence matching.

$$= \sum_x \left(I_d(x) + \Delta I_s'(x)\right)^2, \tag{2.4}$$

where, the subscripts s and d denote sum and difference, respectively. The Taylor series is truncated so that the resulting error function is linear and can be minimized analytically. Taking the derivative with respect to Δ, setting equal to zero and solving for Δ yields the minimal solution:

$$\begin{aligned} \frac{\partial E}{\partial \Delta} &= \sum_x 2I_s'(x)\left(I_d(x) + \Delta I_s'(x)\right)^2 \\ \Delta &= -\frac{\sum_x I_d(x) I_s'(x)}{\sum_x I_s'(x)^2} \end{aligned} \tag{2.5}$$

In 2-D, this expression is simply $\Delta = -\frac{\sum_{(x,y)} I_d(x,y) I_s'(x,y)}{\sum_{(x,y)} I_s'(x,y)^2}$, where $I_d(x,y) = I_1(x,y) - I_2(x,y)$, and $I_s'(x,y)$ is the partial spatial derivative, $\frac{\partial I_s(x,y)}{\partial x}$.

This differential formulation provides an analytic solution to the correspondence problem: disparity, Δ, is computed via a simple arithmetic combination of the sum and difference of a stereo pair of images. Note however that due to the truncated Taylor series expansion (Equation (2.4)), this formulation assumes small disparities between the stereo pair. That is, the *truncated* Taylor series expansion, $I(x+\Delta)$, is only valid for small values of Δ (for example, trivially, the truncated expansion is exact for $\Delta = 0$).

2.3 Range from Motion

The stereo formulation described above is based on estimating range from a pair of images. This section reviews the basic principles of estimating range from several images, termed range from motion. The presentation is provided in two parts: (1) estimating motion in the image plane due to arbitrary camera motions, and (2) determining range from the motion estimate, with known and unknown camera motions. Although the formulations will appear different, we will show that range from stereo may be considered as a constrained version of range from motion.

2.3.1 Motion Estimation

The standard differential formulation of motion estimation is based on the brightness constancy assumption: the brightness of the image of a point in the world is constant when viewed from different positions [Horn 81, Horn 86]. The collection of images is usually parameterized with respect to a temporal parameter, and denoted as $I(x,y,t)$. According to our assumption, the derivative of the image intensity function, $I(x,y,t)$, [3] with respect to time should be zero for each position in the world (assuming a static scene and moving camera). This leads to the following constraint for estimating local changes in the image, $\vec{v}(x,y,t)$ (the constraint is written for a fixed point in space and time: the spatial and temporal parameters are omitted for notational convenience):

[3] The intensity function, $I(x,y,t)$, is parameterized by its spatial parameters, x and y, and a temporal parameter, t. Although we adopt this standard notation, we note that the notation frequently leads to confusion. In particular the spatial parameters should be denoted as functions of t, $I(x(t),y(t),t)$, otherwise, taking partial derivatives of these parameters with respect to t is meaningless (as in Equation (2.6)).

$$\begin{aligned} 0 &= \frac{dI}{dt} \\ &= \frac{\partial I}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial I}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial I}{\partial t} \\ &= \frac{\partial I}{\partial x} v_x + \frac{\partial I}{\partial y} v_y + \frac{\partial I}{\partial t} \\ &= \vec{I}_s^t \cdot \vec{v} + I_t \\ \vec{I}_s^t \vec{v} &= -I_t, \end{aligned} \tag{2.6}$$

where $\vec{I_s} = (\, I_x \quad I_y \,)$ and I_t are the spatial and temporal derivatives, respectively.

Equation (2.6) provides a *single* constraint in *two* unknowns, the horizontal (v_x) and vertical (v_y) velocity components. One method for solving this underconstrained equation is to impose a local smoothness constraint. A quadratic error function based on the brightness constancy constraint for several points (indexed by i) over a small region in the image can be written as:

$$E(\vec{v}) \;=\; \sum_i [\vec{I}_s^t(x_i, y_i, t) \cdot \vec{v}(x_i, y_i, t) + I_t(x_i, y_i, t)]^2. \tag{2.7}$$

Note that this error function is similar to the differential stereo formulation, where I_s' in Equation (2.4) is replaced by the vector of partial spatial derivatives, $\vec{I_s}$, and the difference image I_d (an approximation to a temporal derivative) is replaced with a temporal derivative computed from several images over time, I_t. With vectors in place of scalars, the minimizing solution will include vectors and matrices instead of only scalars. To compute a linear least-squares estimate of $\vec{v}(x_i, y_i, t)$ at each point in space and time, the gradient of this error function is computed:

$$\begin{aligned} \vec{\nabla} E(\vec{v}) &= 2 \sum \vec{I}_s [\vec{I}_s^t \cdot \vec{v} + I_t] \\ &= 2 \sum \left[\begin{pmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{pmatrix} \vec{v} + \begin{pmatrix} I_x I_t \\ I_y I_t \end{pmatrix} \right] \\ &= 2 \left[\begin{pmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{pmatrix} \vec{v} + \begin{pmatrix} \sum I_x I_t \\ \sum I_y I_t \end{pmatrix} \right] \\ &= 2[M\vec{v} + \vec{b}], \end{aligned} \tag{2.8}$$

where for notational convenience, the spatial and temporal parameters are dropped. Setting this equation equal to zero gives the least-squares estimate of velocity:

$$\vec{v} \;=\; -M^{-1}\vec{b}. \tag{2.9}$$

Of course, the matrix M is not guaranteed to be invertible. If the intensity variation in a local image patch varies only one-dimensionally (e.g., $I_x = 0$ or $I_y = 0$) or zero-dimensionally ($I_x = 0$ and $I_y = 0$), then M is *not* invertible (i.e., M is rank deficient or singular). These singularities are sometimes referred to as the aperture and blank wall problem, respectively. The matrix M *is* invertible if and only if the intensity variation is two-dimensional.

Several other motion estimation techniques may be found in the literature (e.g., matching techniques similar to the SSD stereo formulation, spatio-temporal energy models, frequency domain regression, frequency domain phase estimation), see [Simoncelli 93] for a review and unification of these approaches.

Up to this point, we have only described a method for measuring the motion in the image plane due to an arbitrary camera motion. What remains is to show how to recover the structure (i.e., range) given the velocity measurements and camera motion, where the case of a known and unknown camera motion will be considered separately.

2.3.2 Range Estimation (with known camera motion)

Assuming a pinhole model under perspective projection (Section 1.1), the image of a point, $\vec{P} = (\, X \quad Y \quad Z\,)^t$, is given by:

$$\begin{pmatrix} x \\ y \end{pmatrix} \;=\; \begin{pmatrix} \frac{d_s X}{Z} \\ \frac{d_s Y}{Z} \end{pmatrix}. \tag{2.10}$$

As the camera undergoes an arbitrary small motion relative to a static scene, the new projection of $\vec{P}$ can be expressed differentially by taking the temporal derivative of Equation (2.10): [4]

$$\begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix} \;=\; \begin{pmatrix} \frac{d_s Z(dX/dt) - d_s X(dZ/dt)}{Z^2} \\ \frac{d_s Z(dY/dt) - d_s Y(dZ/dt)}{Z^2} \end{pmatrix}. \tag{2.11}$$

[4] Note again that the spatial parameters, x and y, and X, Y, and Z, are actually functions of time, t.

As before, we denote the motion of a fixed point in space and time in the image plane as a velocity vector (again, spatial and temporal parameters are omitted):

$$\vec{v} = \begin{pmatrix} v_x \\ v_y \end{pmatrix} = \begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix}, \tag{2.12}$$

where $\vec{v}$ is the quantity measured above in Equation (2.9).

The differential motion of $\vec{P}$ in the world can be expressed as a rigid-body transformation, parameterized by an axis of rotation, $\vec{R} = (\, R_x \quad R_y \quad R_z \,)^t$, and a translation vector, $\vec{T} = (\, T_x \quad T_y \quad T_z \,)^t$:

$$\begin{aligned} \vec{V} &= (\, \tfrac{dX}{dt} \quad \tfrac{dY}{dt} \quad \tfrac{dZ}{dt} \,)^t \\ &= -(\vec{R}^t \times \vec{P}^t + \vec{T}^t). \end{aligned} \tag{2.13}$$

As shown in [Heeger 92], substituting the motion components in world coordinates (Equation (2.13)) into the expression for the motion in image coordinates (Equation (2.11)) yields:

$$\begin{aligned} \vec{v} &= \begin{pmatrix} \frac{xy}{d_s} & -d_s - \frac{x^2}{d_s} & y \\ d_s + \frac{y^2}{d_s} & -\frac{xy}{d_s} & -x \end{pmatrix} \begin{pmatrix} R_x \\ R_y \\ R_z \end{pmatrix} + \frac{1}{Z} \begin{pmatrix} -d_s & 0 & x \\ 0 & -d_s & y \end{pmatrix} \begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix} \\ &= B\vec{R} + \frac{1}{Z} A\vec{T}. \end{aligned} \tag{2.14}$$

With known camera parameter, d_s, and camera motion, $\vec{R}$ and $\vec{T}$, and an estimate of velocity, $\vec{v}$, at each point in space, (x, y), the complete structure of the imaged three-dimensional world can be estimated by solving for Z in Equation (2.14).

We have seen that the differential stereo and motion error functions are similar, and should therefore not be surprised to discover that although the formulations differ, range from stereo and motion are closely related. In fact, range from stereo with a small baseline is equivalent to range from motion with a constrained and known camera motion. In particular, a constrained Equation (2.14) is equivalent to Equation (2.1). Consider the equation relating motion in the image plane to the camera motion and to range (Equation (2.14)), where the camera is constrained to have zero rotation and a translation confined to the x

dimension (i.e., a stereo configuration):

$$\begin{aligned}\begin{pmatrix} v_x \\ v_y \end{pmatrix} &= B\vec{0} + \frac{1}{Z}\begin{pmatrix} -d_s & 0 & x \\ 0 & -d_s & y \end{pmatrix}\begin{pmatrix} T_x \\ 0 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} -\frac{d_s T_x}{Z} \\ 0 \end{pmatrix}. \end{aligned} \tag{2.15}$$

Extracting the horizontal velocity component and letting $b = T_x$, yields the familiar range from stereo equation: $v_x = -\frac{d_s b}{Z}$, where v_x is synonymous with the disparity, Δ. The negative sign cancels with the negative sign in the definition of v_x (Equation (2.6)).

Note also that range from stereo suffers from the same type of singularities as range from motion. In particular, if the intensity variation in a local image patch does not vary or varies only one-dimensionally, then the correspondence search will yield multiple matches. As a result, accurate disparity estimates at such regions are impossible. [5]

2.3.3 Range Estimation (with unknown camera motion)

In the previous section we described a method for recovering range when the camera motion is known. If the camera motion is unknown, then the recovery of range is not as straight-forward. There are several techniques for simultaneously estimating range and camera motion. One such technique, chosen for its simplicity, is reproduced here ([Tomasi 92]).

Unlike the differential motion algorithm discussed above, this algorithm assumes that specific "features" are tracked as the camera moves, and that these features belong to a single rigid object (see [Costeira 95] for a relaxation of the latter constraint). It is also assumed that the images are formed under an orthographic projection model (Section 1.1). The camera coordinate system is denoted by $(\, i_f \quad j_f \quad k_f \,)^t$, where k_f is the optical axis, and where the subscript, $f = [1, F]$, refers to the frame number. The position of feature points in the image plane are denoted as $(\, x_{f_p} \quad y_{f_p} \,)^t$, where $p = [1, N]$ refers to a feature being tracked over frame f. The corresponding feature point in the world is denoted as

[5] Coarse-to-fine algorithms are helpful in overcoming the singularities due to the blank wall and aperture problem.

$\vec{P}_p = (\, X_p \quad Y_p \quad Z_p \,)^t$, and of course is not indexed by frame because its position is constant over time.

Under orthographic projection, a point, $\vec{P}_p$, at frame, f, is projected to:

$$\begin{pmatrix} x_{f_p} \\ y_{f_p} \end{pmatrix} = \begin{pmatrix} \vec{i}_f \cdot (\vec{P}_p - \vec{T}_f) \\ \vec{j}_f \cdot (\vec{P}_p - \vec{T}_f) \end{pmatrix}, \tag{2.16}$$

where $\vec{T}_f$ is the vector between the origin of the camera (at frame f) and the world coordinate system. Note that this projection is defined for N feature points, across F frames. For reasons that will become clear in a moment the x and y centroids are subtracted, for each frame, from each image coordinate:

$$\begin{pmatrix} u_{f_p} \\ v_{f_p} \end{pmatrix} = \begin{pmatrix} x_{f_p} - \frac{1}{N}\sum_{p=1}^{N} x_{f_p} \\ y_{f_p} - \frac{1}{N}\sum_{p=1}^{N} y_{f_p} \end{pmatrix} \tag{2.17}$$

Combining the transformed and original image coordinates, and performing some simple algebraic manipulations yields:

$$\begin{aligned} \begin{pmatrix} u_{f_p} \\ v_{f_p} \end{pmatrix} &= \begin{pmatrix} \vec{i}_f \cdot (\vec{P}_p - \bar{P}) \\ \vec{j}_f \cdot (\vec{P}_p - \bar{P}) \end{pmatrix} \\ &= \begin{pmatrix} \vec{i}_f \cdot \vec{Q}_p \\ \vec{j}_f \cdot \vec{Q}_p \end{pmatrix}, \end{aligned} \tag{2.18}$$

where $\bar{P}$ is the centroid of the points in the world. Note that something very nice has occurred: the translation vector, $\vec{T}_f$, no longer appears in our system of equations. This is obviously attractive, since it is assumed that the motion of the camera is unknown. The above equation provides a set of linear equations in the image coordinate system over F frames, and the image and world coordinates of N features points over F frames. This set of equations can be expressed in matrix form:

$$\begin{aligned} W_{2F\times N} &= (\, \vec{i}_1 \ldots \vec{i}_F \vec{j}_1 \ldots \vec{j}_f \,)^t_{2F\times 3} (\, \vec{Q}_1 \ldots \vec{Q}_N \,)_{3\times N} \\ &= MS, \end{aligned} \tag{2.19}$$

where the matrix M embodies the "motion" and the matrix S embodies the "structure". The matrix W contains our measurements: the image coordinates of N points tracked over F frames. In particular, row f of the matrix contains the x-coordinates of the N^{th} feature point at frame f, and row $f + F$ contains its y-coordinate. Although it is known that the measurement matrix, W, is a product of a "motion" and "structure" matrix, we have no method for separating these component matrices. In the previous algorithm, the motion was known, making it easy to solve for the structure. Now, the motion is not known and so both the motion and structure must be solved for simultaneously.

We observe that under ideal conditions (i.e., no noise in the measurements) the matrix M is rank 3 (i.e., M is a product of a $2F \times 3$ and $3 \times N$ matrix). Consider now a singular value decomposition (SVD) of the measurement matrix: $M = O_1 D O_2$, where O_1 and O_2 are orthonormal and D is a diagonal matrix. These matrices are partitioned as follows:

$$O_1 = (\, O'_{1(2F\times 3)} \quad O''_{1(2F\times N-3)} \,)_{2F\times N} \quad D = \begin{pmatrix} D'_{3\times 3} & 0_{3\times N-3} \\ 0_{N-3\times 3} & D''_{N-3\times N-3} \end{pmatrix}_{N\times N} \quad O_2 = \begin{pmatrix} O'_{2(3\times N)} \\ O''_{2(N-3\times N)} \end{pmatrix}_{N\times N} \tag{2.20}$$

where $M = O_1 D O_2 = O'_1 D' O'_2 + O''_1 D'' O''_2$. Since W is rank 3, the second term in this summation is 0, and the motion and structure matrices are simply $M = O'_1 (D')^{1/2}$ and $S = (D')^{1/2} O'_2$. This only provides a solution up to a linear transformation, that is, the decomposition $W = MS$ is not unique since for any invertible matrix A, $W = (MA)(A^{-1}S) = MS$. A unique solution can be determined by restricting the rows of the motion matrix, M, to be orthonormal (i.e., they define an orthogonal coordinate system).

Both range from stereo and range from motion rely on observing changes in the appearance of the world across multiple viewpoints. In the case of range from focus and range from defocus, the camera remains stationary and it is the change in the appearance of the world with respect to different optical settings that is used to estimate range.

2.4 Range from Focus

Our impression may be that our entire retinal image is in clear focus, this is commonly referred to as having an infinite depth of field. However, imaging devices (including the human eye) do not have an infinite depth of field: objects projected onto an imaging sensor plane are only in focus if they lie within a relatively narrow band of distances from the imaging sensor's focal plane (the depth of field). If they lie outside this range they are imaged out of focus. Furthermore, the amount of defocus is a simple function of range. This section reviews the basic principles for recovering range from focus.

Recall that under the thin lens model, a point at a depth Z will be imaged as a blurred circle with a radius:

$$r = \frac{R}{\frac{1}{f} - \frac{1}{Z}} \left| \left(\frac{1}{f} - \frac{1}{Z} \right) - d_s \right|, \tag{2.21}$$

where R is the radius of the lens, d_s is the lens to sensor distance, and f is the focal length. Note that the amount of blurring, r, is a simple function of the range, Z. In order to recover the three-dimensional structure of the world we must first model the effects of blurring on an otherwise perfectly focused image (i.e., an image of the world taken through an ideal pinhole camera). Blurring is usually modeled by a simple convolution: a defocused image, $I_d(\cdot)$, can be expressed as the convolution of a blurring function, $h(\cdot)$, and a perfectly focused image, $I_f(\cdot)$:

$$I_d(x, y) = h(x, y, \sigma) \star I_f(x, y). \tag{2.22}$$

The blurring function, meant to approximate the point-spread function [6] of a camera, is frequently taken to be a two-dimensional Gaussian:

$$h(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma}}, \tag{2.23}$$

where σ is monotonically related to the blur radius, r, which in turn is related to range.

Note that implicit in the convolution (Equation (2.22)) is the assumption that the world consists of a single frontal-parallel surface. In particular, the blurring function, $h(x, y, \sigma)$, is assumed to be constant over the entire image. Of course, the world does not generally consist of a single frontal-parallel surface, and the blurring parameter, σ, depends on the spatial parameters, (x, y). This assumption will be addressed later; for now, we will put it aside and continue.

Recall that under the thin lens model of image formation, only surfaces lying within the depth of field will be imaged in perfect focus (i.e., $\sigma \to 0$). A straight-forward method for recovering range presents itself: since the depth of field can be altered in a systematic fashion by varying the sensor's internal parameters, the range of a point in the world can be estimated by determining the lens to sensor distance d at which it is imaged in perfect focus (Figure 2.3). What is needed then is a method for measuring the amount of blur in an image.

[6] The point spread function of a camera describes the relative attenuation of incoming light as a function of lens position. This function can be described by the blurred image of a single point light source.

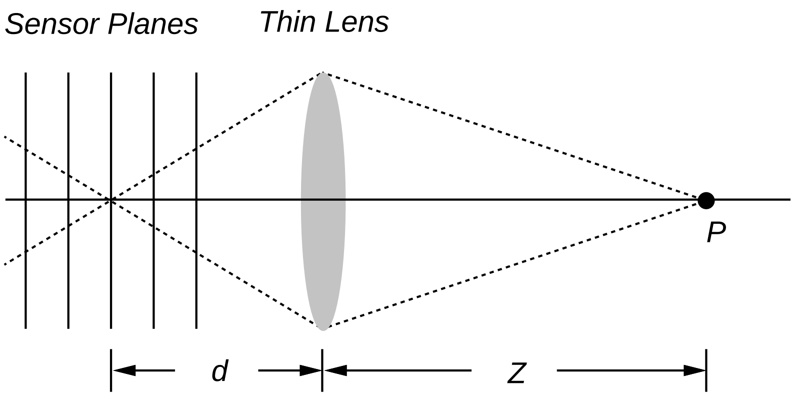


Figure 2.3: Range from Focus. Illustrated is a sequence of images of a point, $\vec{P}$, taken with varying lens to sensor distances, d_s. The point will be imaged in perfect focus, when the sensor plane is a distance of d from the lens. For smaller or larger values of d, the same point will be imaged out of focus.

Our model of blurring (Equations (2.22) and (2.23)) is equivalent to low-pass filtering. Therefore, an operator, $o(x, y)$, for measuring the amount of blur in an image should be sensitive to *high* frequencies because a patch in the image will be maximally focused when its high frequency content is maximal. Convolution with any of a variety of standard high-pass filters may be used for measuring the amount of blur (e.g., the Laplace operator, $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$). More specifically, the amount of blur at a point (x_0, y_0) in the image may be computed as the sum of the responses of convolving with $o(x, y)$ in a $n \times n$ neighborhood around (x_0, y_0):

$$\sum_{x=-n/2}^{n/2} \sum_{y=-n/2}^{n/2} (o(x, y) \star I_d(x_0 + x, y_0 + y))^2. \tag{2.24}$$

Without prior knowledge of the frequency content of the scene, a sequence of images (usually between 50 and 100) with varying amounts of blur must be acquired. As illustrated in Figure 2.3 this sequence of images may be obtained by varying the lens to sensor distance. The focus operator is then applied to a $n \times n$ patch in *each* image in the sequence (Equation (2.24)). From an off-line calibration stage the lens to sensor distance for the image at which the focus operator is maximal is then used to estimate range. By repeating this process for each $n \times n$ image patch the complete structure of the imaged three-dimensional world can be estimated.

We will show next how explicit knowledge of either the blurring function or frequency

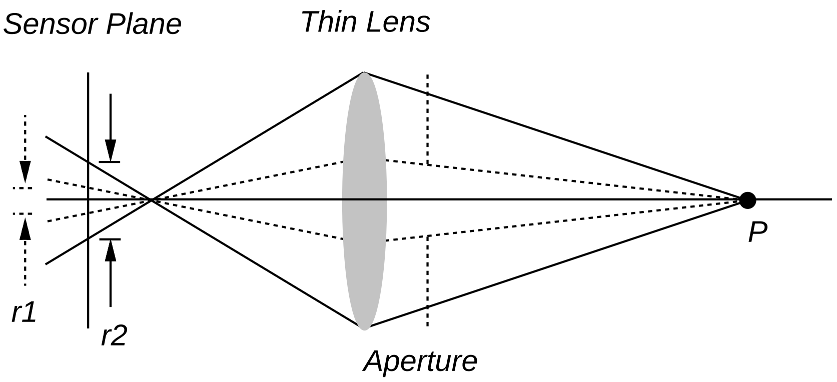


Figure 2.4: Range from Defocus. Illustrated is the image of a point, $\vec{P}$, taken with a full aperture ($r2$, solid lines) and a quarter aperture ($r1$, dashed lines). Note that as the aperture is stopped down, the radius of the blur circle is reduced accordingly.

content of the scene can reduce the number of measurements required for recovering range.

2.5 Range from Defocus

Assuming an image with N pixels, our model of image blurring (Equations (2.22)) provides N constraints (the intensity at each pixel in the measured image, $I_d(x,y)$) in $2N$ unknowns (for each pixel in the image, the desired blur parameter, σ, and the intensity of the perfectly focused image, $I_f(x,y)$). Additional constraints can be added by making more measurements of the scene. In particular, the same scene can be imaged with varying sensor parameters (Figure 2.4). Each new measurement provides an additional N constraints from the intensity at each pixel. Thus, a minimum of two measurements are required to solve for the N blur parameters. A difficulty arises in that $I_f(x,y)$ can only be measured by imaging through a physically unrealizable pinhole camera. In order to estimate the blur parameters, the immeasurable and unknown $I_f(x,y)$ must be eliminated from the system of constraints. This section outlines how this can be accomplished.

Recall, that the blurring of an image, $I_d(x,y)$, was modeled as a convolution with a blurring function, $h(x,y,\sigma)$ (Equations (2.22) and (2.23)). Consider now a pair of images acquired with different aperture sizes (Figure 2.4), and consequently with different amounts of blur:

$$
\begin{aligned}
I_1(x,y) &= h(x,y,\sigma_1) \star I_f(x,y), && (2.25)\\
I_2(x,y) &= h(x,y,\sigma_2) \star I_f(x,y). && (2.26)
\end{aligned}
$$

From this pair of equations (in three unknowns, σ_1, σ_2 and $I_f(x,y)$), we wish to solve for σ_1 or σ_2, which is monotonically related to range. In order to solve for the desired parameters, the third unknown, $I_f(x,y)$, must be eliminated. To this end, consider the ratio of the power of the Fourier transforms of $I_1(x,y)$ and $I_2(x,y)$:

$$
\begin{aligned}
\frac{\mathcal{I}_1(\omega_x,\omega_y)}{\mathcal{I}_2(\omega_x,\omega_y)} &= \frac{\mathcal{H}(\omega_x,\omega_y,\hat{\sigma}_1)\cdot\mathcal{I}_f(\omega_x,\omega_y)}{\mathcal{H}(\omega_x,\omega_y,\hat{\sigma}_2)\cdot\mathcal{I}_f(\omega_x,\omega_y)}\\
&= \frac{\mathcal{H}(\omega_x,\omega_y,\hat{\sigma}_1)}{\mathcal{H}(\omega_x,\omega_y,\hat{\sigma}_2)}, && (2.27)
\end{aligned}
$$

where $\hat{\sigma}_i = \frac{1}{2\pi\sigma_i^2}$. Note that in the frequency domain, the unknown $I_f(x,y)$ cancels! Taking the natural log [7] of both sides:

$$
\begin{aligned}
\ln\left(\frac{\mathcal{I}_1(\omega_x,\omega_y)}{\mathcal{I}_2(\omega_x,\omega_y)}\right) &= \ln\left(\frac{\mathcal{H}(\omega_x,\omega_y,\hat{\sigma}_1)}{\mathcal{H}(\omega_x,\omega_y,\hat{\sigma}_2)}\right)\\
&= -\frac{1}{2}(\omega_x^2+\omega_y^2)(\sigma_1^2-\sigma_2^2). && (2.28)
\end{aligned}
$$

If σ_1 can be expressed as $\sigma_1 = \beta\sigma_2$, for some constant $\beta \in \mathcal{R}$ (i.e., the ratio of the aperture sizes is known), then σ_2 can be solved for directly in Equation (2.28). From the imaging geometry the range, Z, to the point is given by:

$$
Z = \frac{F d_s}{d_s - f - \sigma_1 F}, \qquad (2.29)
$$

where d_s is the distance between the lens and sensor plane, f is the focal length of the lens, and F is the f-number of the lens(the ratio of focal length to lens diameter).

The blur parameter, σ_1, can be estimated by first computing the Fourier transform of an image pair and then computing the log of the ratio of these transforms. Note that implicit in this calculation is the assumption that the world consists of a single frontal-parallel surface (as was the case in range from focus), that is, the Fourier transform is a *global* transform operating on the entire image. In order to get a more local estimate of range,

[7] By computing the natural log of the ratio of the blurring functions, the blurring function, $\mathcal{H}(\cdot)$, is constrained to be an exponential in the frequency domain, and therefore, in the spatial domain as well.

a windowed short-time Fourier transform (STFT) is typically used. [8] By computing a STFT for each $n \times n$ image patch, the complete structure of the imaged three-dimensional world can be estimated.

2.6 Summary

In this chapter we have seen that measurements of four image properties (viewpoint, time, lens to sensor distance, and aperture size) can be used to estimate range. We have avoided discussing the multitude of (often minor) variations within each of these classes. To the contrary, we argue that these techniques share a common and fundamental property: that of *measuring change*. Specifically, each range estimation technique measures changes in the appearance of the world with respect to different parameters: viewing position (stereo), time or viewing position (motion), lens to sensor distance (focus), or aperture size (defocus). Either implicitly or explicitly, each of these techniques amount to computing a derivative (i.e., measuring change) with respect to their relevant parameters. More precisely, a *discrete* approximation to a derivative is computed (e.g., stereo approximates a derivative with respect to viewpoint from just two samples). One may be tempted to argue that techniques such as range from motion are differential, however, in that all of these techniques sample along the dimension to be differentiated, none of the aforementioned techniques are *strictly* differential.

Given that these range estimation techniques amount to computing discrete approximations to a derivative, we may borrow from the issues that arose in the design of the optimal derivative filters (Section 1.4). In so doing, we have arrived at a technique for computing derivatives with respect to viewing position and aperture size which does not rely on a discrete approximation - the derivatives are *directly measured optically* from a single stationary camera. The full exploration of these techniques is the subject of the final chapter.

[8] A windowed short-time Fourier transform (STFT) is employed for computing a *local* Fourier transform. Note that this windowing leads to blurring in the space and frequency domains. In particular, multiplicative windowing (by $W(x,y)$) in the space domain is equivalent to convolution in the frequency domain. Rewriting Equation (2.27) with the windowing yields: $\frac{\mathcal{I}_1(\omega_x,\omega_y)\star\mathcal{W}(\omega_x,\omega_y)}{\mathcal{I}_2(\omega_x,\omega_y)\star\mathcal{W}(\omega_x,\omega_y)} = \frac{(\mathcal{H}(\omega_x,\omega_y,\hat{\sigma}_1)\cdot\mathcal{I}_f(\omega_x,\omega_y))\star\mathcal{W}(\omega_x,\omega_y)}{(\mathcal{H}(\omega_x,\omega_y,\hat{\sigma}_2)\cdot\mathcal{I}_f(\omega_x,\omega_y))\star\mathcal{W}(\omega_x,\omega_y)}$. Note that $\mathcal{I}_f(\omega_x,\omega_y)$ no longer cancels, and convolution with the windowing function leads to blurring and subsequent errors in range estimation.

Chapter 3

Range Estimation by Optical Differentiation

3.1 Introduction

We are now prepared to connect Chapters 1 and 2 in what will be the central contribution of our work. In Chapter 2 we presented various, seemingly unrelated, techniques for estimating range, and concluded that each of the techniques can be thought of in terms of *measuring change* with respect to various parameters. It is natural then to consider these range estimation techniques within a differential framework. To do this it is necessary to apply the continuous differential operator to discrete functions as discussed in Chapter 1. As we will see, this calculation extends beyond just computing spatial derivatives in the image plane.

We begin by revisiting the standard binocular stereo formulation of Section 2.2. Recall that the disparity, Δ, (inversely proportional to range) can be determined from a pair of images taken from spatially offset viewing positions:

$$\Delta = -\frac{I(x,y,v_1) - I(x,y,v_2)}{\frac{\partial}{\partial x}(I(x,y,v_1) + I(x,y,v_2))}, \tag{3.1}$$

where, for simplicity, the summation over spatial position x and y is dropped, and the subscript notation for denoting a stereo image pair, $I_1(x,y)$ and $I_2(x,y)$, is replaced with

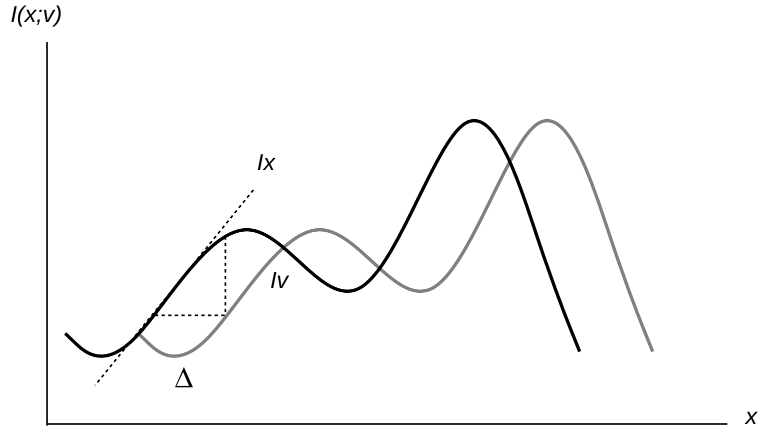


Figure 3.1: Viewpoint and Spatial Derivatives. Shown is a one-dimensional signal as it may appear from two distinct viewpoints (black and gray curve). The disparity, Δ, is a simple function of the viewpoint, I_v, and spatial, I_x, derivative: $\Delta = \frac{I_v}{I_x}$ (see Equation (3.1)).

$I(x, y, v_1)$ and $I(x, y, v_2)$, respectively. First we should convince ourselves that Equation (3.1) amounts to computing derivatives of $I(\cdot)$. Clearly, the denominator is a derivative with respect to the spatial parameter, x. Perhaps less obvious, the numerator, the *difference* between the stereo image pair, is an approximation to a derivative with respect to viewpoint. [1]

It may not be immediately obvious why range is proportional to a ratio of these derivatives, but as illustrated in Figure 3.1, this ratio makes perfect sense. Shown in this figure is a one-dimensional intensity function as it may appear from two distinct viewpoints. Note first that the signal is translated by a constant amount, Δ, which of course is inversely proportional to range. The measurable quantities are the viewpoint, I_v, and the spatial derivatives, I_x. From this figure it is clear that $I_x = \frac{I_v}{\Delta}$, or $\Delta = \frac{I_v}{I_x}$ (i.e., the "slope" is the ratio of the change "y" and the change in "x"). This example also illustrates the discrete nature of these derivative measurements.

Since the intensity function, $I(\cdot)$, is discrete (i.e., sampled in both space and viewpoint) let's now examine Equation (3.1) more closely in light of our discussion on the proper technique for computing derivatives of a discrete function. To compute the denominator, interpolation in y and v are required, followed by differentiation in x (hence the summation,

[1] The approximation of a derivative by a simple difference can be most easily seen by the standard definition of differentiation: $f'(x) = \lim_{\varepsilon \to 0} \frac{f(x) - f(x+\varepsilon)}{\varepsilon}$, where the approximation is computed by letting $\varepsilon = 1$.

the simplest form of interpolation, over viewpoint, i.e., $I(x,y,v_1)$ and $I(x,y,v_2)$). Although not expressed explicitly, it is assumed that the resulting summed image is interpolated in y before applying the directional derivative operator in x. To compute the numerator, we again require interpolation in x and y followed by differentiation in v. Given only two samples with respect to viewpoint, v, the simple difference is the best we can hope to accomplish in approximating the viewpoint derivative. Again, it is assumed that after subtracting the images, they are properly interpolated with respect to x and y.

So, we have seen that range can be computed as the ratio of the viewpoint derivative and spatial derivatives of the intensity function, $I(x,y,v)$. In the case of binocular stereo, a generally crude approximation to the viewpoint derivative (a difference) is computed from just two samples along the viewpoint dimension, v. To compute a more accurate viewpoint derivative the intensity function can be sampled at three or more viewing positions, as in the case of multiocular stereo (or range from motion). Illustrated in Figure 3.2 is a 2-D example of such a configuration. Shown is the projection of a single point in the world imaged through five spatially offset pinhole cameras. Also shown is the subsequent processing for computing the spatial and viewpoint derivatives (see figure caption for more details). We naturally expect the accuracy of the viewpoint derivative to improve with increasing number of views. Of course, the expense, physical size, and calibration of a system containing numerous cameras may become prohibitive. As an alternative, we propose a simple technique for measuring the viewpoint and spatial derivatives from a *single stationary* camera and a pair of optical masks.

It is generally assumed that computing a viewpoint derivative requires multiple views, either from two or more cameras, or by moving a single camera. This assumption typically comes from thinking in terms of a pinhole camera model (Section 1.1). In particular, under the pinhole camera model, a camera captures the projection of the world from a *single* viewpoint. However, under the more realistic thin-lens model, a *single* camera collects light from a *continuum* of viewpoints (Section 1.1). Of course, all this information is lost once the lens focuses the light and the CCD sensor integrates over the different viewpoints. Nonetheless, at the *front* of the lens, the information is there and available to

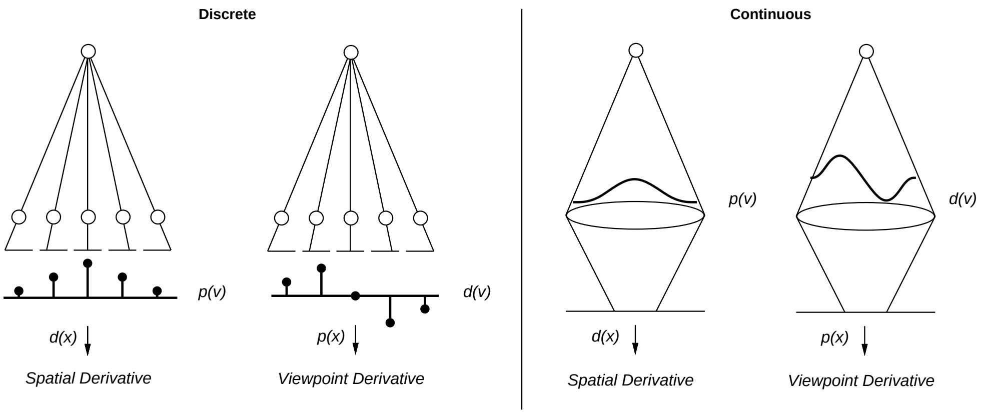


Figure 3.2: Differentiating with Respect to Viewpoint and Space. Illustrated on the left is the projection of a single point through five spatially offset pinhole cameras. The spatial derivative is computed by first interpolating over viewpoint, $p(v)$, (i.e., a linear combination of the five images, where the weighting of each image is specified by the height of the impulse directly beneath the image). A standard derivative filter, $d(x)$, is then applied to compute the spatial derivative. Similarly, the viewpoint derivative is computed by differentiating over viewpoint, $d(v)$, followed by interpolating over the spatial parameter, $p(x)$. Shown on the right are the same set of operations but now computed along a continuum of viewpoints. In particular, the interpolation and differentiation across viewpoint are accomplished by placing a variable opacity optical attenuation mask directly in front of the camera lens.

us. It is precisely this information that we propose to exploit. [2]

Consider the image of a single point in the world, but this time, imaged through a single thin-lens imaging system (Figure 3.2). First, note that unlike a pinhole camera, the lens collects light from a continuum of viewpoints before focusing the light onto the sensor. Now consider what would happen if an optical attenuation mask were placed directly in front of the lens. The functional form of the optical mask takes on values in the range 0 to 1, where a value of 0 attenuates the incoming light ray fully, and a value of 1 passes the light ray unattenuated. The spatial derivative can then be computed by first imaging through the mask, $M(u)$, which blurs over viewpoint and then by differentiating in the spatial parameter. Likewise, the viewpoint derivative can be computed by imaging through a derivative mask, $M'(u)$, [3] and then blurring across the spatial parameter (Figure 3.2). Note that this set of calculations is simply a continuous version of the five-camera stereo setup also illustrated in Figure 3.2. That is, the "weighting of the light" is identical, but

[2] These ideas were inspired by the work of Adelson and Wang in [Adelson 92], see Section 3.5 for more details on their work.

[3] In practice, the mask $M'(v)$ cannot be used directly as an attenuation mask since it contains negative values. This issue is addressed in Section 3.3.3.

now we are operating with a continuum of viewpoints (across the diameter of the lens) as opposed to a set of discrete views. This simple intuition is formalized in the next section.

3.2 Optical Differentiation

We have seen, at least intuitively, that a viewpoint and spatial derivative (and hence range) may computed from a single stationary camera and a pair of optical attenuation masks. In this section these ideas are formalized, and later, a variant of this technique is discussed.

3.2.1 Viewpoint Derivatives

We begin by considering a simplified world consisting of a single, uniform intensity point light source. This assumption is made only to simplify the explanation and accompanying mathematics and will be relaxed later. The image of such a point light source through an optical attenuation mask, $M(\cdot)$, is a scaled and dilated version of the mask function, $M(\cdot)$:

$$I(x) \quad = \quad \frac{1}{\alpha} M \left(\frac{x}{\alpha} \right), \tag{3.2}$$

where the parameter α is monotonically related to range, and is derived from the imaging geometry:

$$\alpha \quad = \quad 1 - \frac{d_s}{f} + \frac{d_s}{Z}, \tag{3.3}$$

d_s is the distance between the lens and sensor, and f is the focal length of the lens as in Figure 3.3. Intuitively, Equation (3.2) makes sense: as the point moves further from the lens, the image becomes more blurred (i.e., shorter (scaled) and broader (dilated)). Since the amount of light entering the lens does not significantly change with depth, the scaling and dilation must be such that the area of the mask image is independent of depth (i.e., α): $\int dx \ \frac{1}{\alpha} M \left(\frac{x}{\alpha} \right) = \int dv \ M(v)$ (this is easily proven by letting $u = \frac{x}{\alpha}$ and substituting this into the left-hand side of the equality).

With such a system the *effective* viewpoint may be altered by translating the mask, while leaving the lens and sensor stationary, i.e., imaging through the mask $M(v + v_0)$.

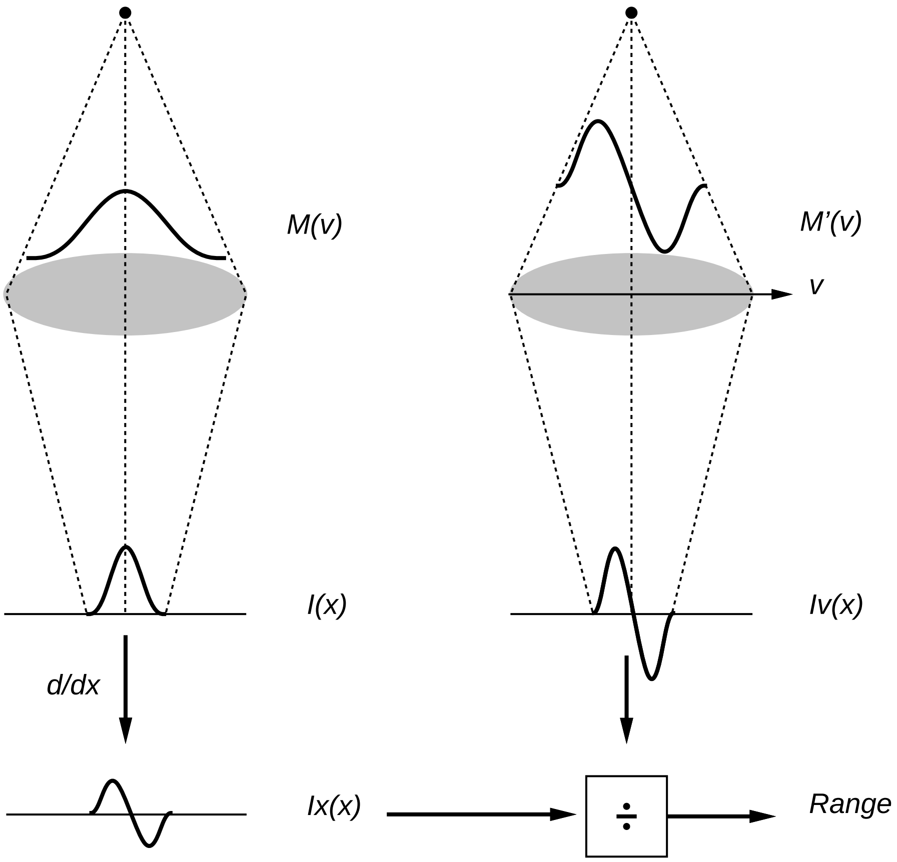


Figure 3.3: Range Estimation by Optical Differentiation. Illustrated are the images of the same uniform intensity point light source imaged through a pair of spatially varying optical attenuation masks. Note that unlike a pinhole camera, the lens collects light from a *continuum* of viewpoints. This allows the viewpoint derivative to be measured directly by imaging through the mask $M'(v)$. The corresponding spatial derivative is computed by first imaging through the mask $M(v)$ and then computing a spatial derivative in the image plane. The ratio of the viewpoint and spatial derivatives is proportional to range (Equation (3.5)).

The *differential* change in the image with respect to viewpoint may thus be measured by imaging through the derivative of this mask, $M'(v)$, (evaluated at $v_0 = 0$). Let's now convince ourselves that range can be determined by imaging through this pair of optical masks. The pair of images formed under the optical mask $M(\cdot)$ and its derivative, $M'(\cdot)$ (Figure 3.3) are simply:

$$I(x) = \frac{1}{\alpha} M\left(\frac{x}{\alpha}\right) \quad \text{and} \quad I_v(x) = \frac{1}{\alpha} M'\left(\frac{x}{\alpha}\right). \tag{3.4}$$

As expected, the spatial derivative of $I(x)$ turns out to be closely related to $I_v(x)$:

$$\begin{aligned} I_x &= \frac{\partial}{\partial x}\left[\frac{1}{\alpha}M\left(\frac{x}{\alpha}\right)\right] \\ &= \frac{1}{\alpha^2}M'\left(\frac{x}{\alpha}\right) \\ &= \frac{1}{\alpha}\ I_v(x) \\ \alpha &= \frac{I_v(x)}{I_x(x)}, \end{aligned} \tag{3.5}$$

where α is monotonically proportional to range, Z (Equation (3.3)). Note that this is precisely the definition with which we began: range is proportional to a ratio of the viewpoint and spatial derivative. The difference being that, here, the viewpoint derivative is *measured directly* by imaging through the mask $M'(\cdot)$ – it is the image $I_v(x)$!

Equation (3.5) embodies the fundamental relationship used for the differential computation of range in a world consisting of a single point light source. A complex scene consisting of a collection of many such light sources imaged through an optical mask appears as a superposition of scaled and dilated versions of the mask function. In particular, we can write an expression for the image formed under the masks $M(\cdot)$ and $M'(\cdot)$ by integrating over the images of the visible points, $\vec{p}$:

$$I(x) = \int dx_p\ \frac{1}{\alpha_p}M\left(\frac{x-x_p}{\alpha_p}\right)L\left(\frac{x_p}{\alpha_p}\right) \quad \text{and} \quad I_v(x) = \int dx_p\ \frac{1}{\alpha_p}M'\left(\frac{x-x_p}{\alpha_p}\right)L\left(\frac{x_p}{\alpha_p}\right) \tag{3.6}$$

where α_p is still monotonically related to range (Equation (3.3)), and the integral is performed over the variable x_p, the position in the sensor of a point $\vec{p}$ projected through the center of the lens. The function $L(\cdot)$ specifies the light intensity at each point, $\vec{p}$, and it is assumed that, for each point, this intensity function is uniform across the optical mask (i.e., $L'(x_p) = 0$, for all $\vec{p}$). Once again, the spatial derivative of $I(x)$ is closely related to $I_v(x)$:

$$I_x(x) \quad = \quad \int dx_p\ \frac{1}{\alpha_p^2}M'\left(\frac{x-x_p}{\alpha_p}\right)L\left(\frac{x_p}{\alpha_p}\right) \tag{3.7}$$

As before, the viewpoint and spatial derivatives ($I_v(x)$ and $I_x(x)$, respectively) differ only in a multiplicative term of α_p. Unfortunately, solving for α_p is nontrivial, since it is embedded in the integrand and depends on the integration variable. Consider, however, the special case where all points in the world lie on a frontal-parallel surface relative to the sensor (in practice, this assumption need only be made locally, see Section 3.3). Under this condition, the scaling parameter α_p is the same for all points x_p and:

$$I_v(x) = \frac{1}{\alpha}\int dx_p \; M'\left(\frac{x - x_p}{\alpha}\right) L\left(\frac{x_p}{\alpha_p}\right) \quad \text{and} \quad I_x(x) = \frac{1}{\alpha^2}\int dx_p \; M'\left(\frac{x - x_p}{\alpha}\right) L\left(\frac{x_p}{\alpha_p}\right) \tag{3.8}$$

The scaling parameter, α, may then be expressed as:

$$\alpha \;=\; \frac{I_v(x)}{I_x(x)}, \tag{3.9}$$

where α is monotonically proportional to range, Z (Equation (3.3)). As before, range is determined from a ratio of the viewpoint and spatial derivative, where now the former is measured directly by imaging through the derivative mask, $M'(\cdot)$.

Least-Squares Solution

In order to deal with singularities (i.e., $I_x(x) = 0$ in Equation (3.5)), a least-squares estimator can be used for α (as in [Lucas 81]). Specifically, the quadratic error function $E(\alpha) = \sum_p (I_v(x) - \alpha I_x(x))^2$ can be minimized, where the summation is performed over a small patch in the image, p. Taking the derivative with respect to α, setting equal to zero and solving for α yields the minimal solution:

$$\alpha \;=\; \frac{\sum_p I_v(x) I_x(x)}{\sum_p I_x(x)^2}. \tag{3.10}$$

By integrating over a small patch in the image, the least-squares solution avoids singularities when the spatial derivative, $I_x(x)$, is zero at a single point in the image. However, since the denominator still contains an $I_x(x)$ term (integrated over a small image patch), a singularity still exists when $I_x(x)$ is zero over the entire image patch (e.g., the "blank wall" or "aperture" problem). In order to avoid this singularity, a small constant may be added to the denominator:

$$\alpha \;=\; \frac{\sum_p I_v(x) I_x(x)}{\left(\sum_p I_x(x)^2\right) + \varepsilon}. \tag{3.11}$$

The choice of the constant ε and its effect on the estimate of α are discussed in Section 3.2.3.

3-D Formulation

This technique extends easily to a three-dimensional world [4]: we need only consider two-dimensional masks $M(u, w)$, and the horizontal partial derivative $M'(u, w) = \partial M(u, w)/\partial u$. For a more robust implementation, the vertical partial derivative mask $\partial M(u, w)/\partial w$ may also be included. The least-squares error function becomes:

$$E(\alpha) \;=\; \sum_p (I_{v_x} - \alpha I_x)^2 + (I_{v_y} - \alpha I_y)^2. \tag{3.12}$$

Solving for the minimizing α gives:

$$\alpha \;=\; \frac{\sum_p (I_{v_x} I_x + I_{v_y} I_y)}{\left(\sum_p (I_x^2 + I_y^2)\right) + \varepsilon}. \tag{3.13}$$

Through several simplifications and assumptions we have arrived at an elegant and efficient scheme for estimating range using a matched pair of optical attenuation masks placed in front of a single stationary camera. The spatial derivative of the image formed under the first mask is related by a scale factor to a second image created with the derivative of the first optical mask. This scale factor is monotonically related to range. As we have seen several times before, range is computed from a ratio of the viewpoint and spatial derivative. But the key difference here is that, instead of approximating the viewpoint derivative with a simple difference, the viewpoint derivative is *measured directly* by imaging through the derivative mask: it *is* the image $I_v(x)$! As a result, the viewpoint derivative is no longer approximated from samples, but is measured directly from a continuum of viewpoints at the front of the lens.

In the next few sections we will discuss the implications of the various simplifications and assumptions that were made, verify the technique through simulations and experiments and analyze the errors and sensitivity of this approach. But first, we note an interesting variant of the optical viewpoint derivative approach based this time on direct measurements of the derivative with respect to aperture size (as in range from defocus, Section 2.5).

[4] Note that in 2-D, the image of a point source is $\frac{1}{\alpha}M(\frac{x}{\alpha})$, and in 3-D, it is $\frac{1}{\alpha^2}M(\frac{x}{\alpha}, \frac{y}{\alpha})$.

3.2.2 Aperture Size Derivatives

In the previous section range was computed based on a viewpoint derivative. Here we show a similar technique for estimating range that is based on the derivative with respect to aperture size (as in range from defocus). Given the optical mask, $M(u)$, the *effective* aperture size may be altered by dilating the mask while leaving the lens and sensor stationary, i.e., imaging through the mask $\frac{1}{a}M\left(\frac{u}{a}\right)$. The additional $\frac{1}{a}$ term out front is an area preserving term and is used to ensure that the changes in the aperture size do not depend on the intensity of the point light source being imaged. The *differential* change in the image with respect to aperture size may be measured by imaging through the derivative of the mask with respect to the dilation parameter, a:

$$M_a(u) \quad = \quad -\tfrac{1}{a^2}M\left(\tfrac{u}{a}\right) - \tfrac{u}{a^3}M'\left(\tfrac{u}{a}\right). \tag{3.14}$$

For notational simplicity, this function is evaluated at $a = 1$, giving the masks $M(u)$ and $M_a(u) = -M(u) - uM'(u)$. Now, consider a pair of images formed through these masks:

$$I(x) = \frac{1}{\alpha}M\left(\frac{x}{\alpha}\right) \quad \text{and} \quad I_a(x) = -\frac{1}{\alpha}M\left(\frac{x}{\alpha}\right) - \frac{x}{\alpha^2}M'\left(\frac{x}{\alpha}\right). \tag{3.15}$$

As suggested from the solution to the dilation equation (see Section 3.3.3 for more details), we consider the second partial derivative of $I(x)$:

$$I_x(x) = \frac{1}{\alpha^2}M'\left(\frac{x}{\alpha}\right) \quad \text{and} \quad I_{xx}(x) = \frac{1}{\alpha^3}M''\left(\frac{x}{\alpha}\right). \tag{3.16}$$

This series of steps is illustrated in Figure 3.4. Note that unlike the viewpoint formulation (Equation (3.5)), there appears to be no clear relationship between the aperture size and spatial derivatives ($I_a(x)$ and $I_{xx}(x)$, respectively). However, consider a mask function which satisfies the constraint that $M''(u) = -(M(u) + uM'(u))$ (e.g., a Gaussian satisfies this constraint, see Section 3.3.3):

$$\begin{aligned} I_a(x) &= -\frac{1}{\alpha}M\left(\frac{x}{\alpha}\right) - \frac{x}{\alpha^2}M'\left(\frac{x}{\alpha}\right) \\ &= -\frac{1}{\alpha}\left(M\left(\frac{x}{\alpha}\right) + \frac{x}{\alpha}M'\left(\frac{x}{\alpha}\right)\right) \\ &= \frac{1}{\alpha}M''\left(\frac{x}{\alpha}\right) \qquad (3.17) \end{aligned}$$

and

$$\begin{aligned} I_{xx}(x) &= \frac{1}{\alpha^2}\left(\frac{1}{\alpha}M''\left(\frac{x}{\alpha}\right)\right) \\ &= \frac{1}{\alpha^2}I_a(x) \qquad (3.18) \end{aligned}$$

Now, the second-order spatial derivative, $I_{xx}(x)$, is related to the aperture size derivative by only a multiplicative factor of α^2. Given the constraint on the optical mask, range can now be estimated as:

$$\alpha = \sqrt{\frac{I_a(x)}{I_{xx}(x)}}, \qquad (3.19)$$

where of course, α is monotonically proportional to range (Equation (3.3)).

Least-Squares Solution

As in the case of the viewpoint derivative formulation, a least-squares estimator can be used for estimating α^2. Specifically, the quadratic error function $E(\alpha^2) = \sum_p (I_a(x) - \alpha^2 I_{xx}(x))^2$ is minimized, where the summation is performed over a small patch in the image, p. Taking the derivative with respect to α^2, setting equal to zero and solving for α^2 yields the minimal solution:

$$\alpha^2 = \frac{\sum_p I_a(x) I_{xx}(x)}{\sum_p I_{xx}(x)^2}. \qquad (3.20)$$

By integrating over a small patch in the image, the least-squares solution avoids singularities when the spatial derivative, $I_{xx}(x)$, is zero at a single point in the image. However, since the denominator still contains an $I_{xx}(x)$ term (integrated over a small image patch), a singularity still exists when $I_{xx}(x)$ is zero over the entire image patch. In order to avoid this singularity, a small constant may be added to the denominator:

$$\alpha^2 = \frac{\sum_p I_a(x) I_{xx}(x)}{\left(\sum_p I_{xx}(x)^2\right) + \varepsilon}. \qquad (3.21)$$

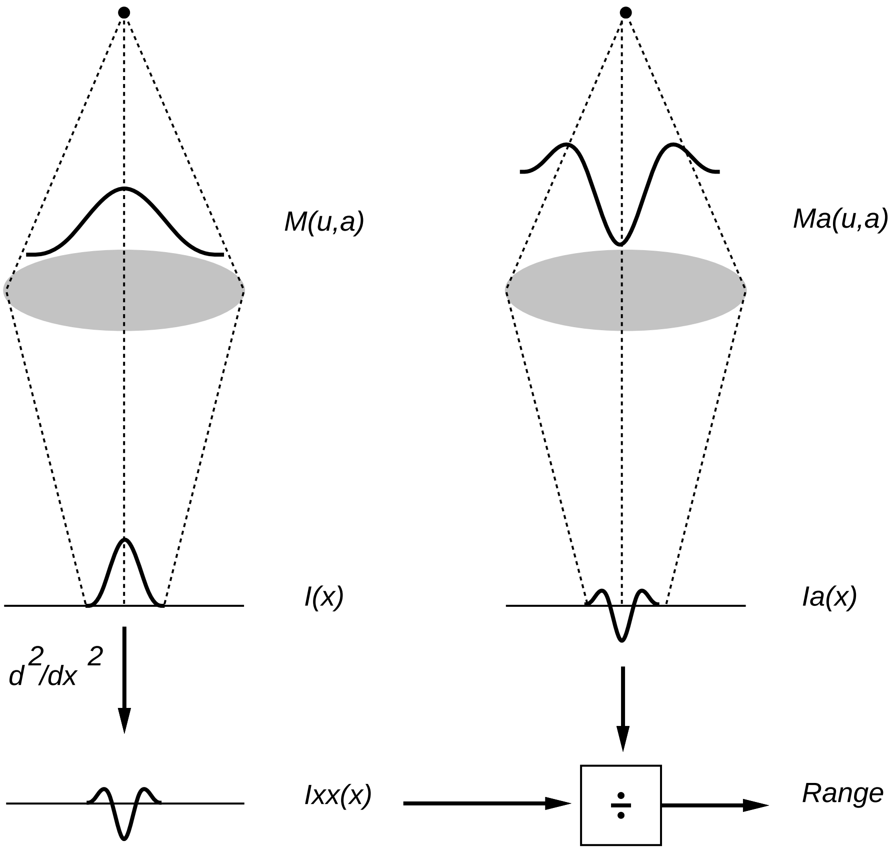


Figure 3.4: Range Estimation by Optical Differentiation. Illustrated is the same uniform intensity point light source imaged through a pair of spatially varying optical attenuation masks. The second mask is the derivative of the first mask with respect to a dilation (e.g., in the case of a Gaussian, the derivative is taken with respect to the standard deviation). Here the derivative with respect to aperture size is measured directly, it is the image $I_a(x)$! The ratio of this derivative to the second-order spatial derivative of the first image, $I(x)$, is proportional to range (Equation (3.19)).

The choice of the constant ε and its effect on the estimate of α^2 are discussed in Section 3.2.3

3-D Formulation

Extending this formulation to a three-dimensional world, and considering Gaussian-based optical masks gives:

$$G(u, w, \sigma) = \frac{1}{\sigma^2} e^{-(u^2+w^2)/2\sigma^2}, \tag{3.22}$$

and its derivative with respect to a dilation (i.e., σ):

$$\begin{aligned} G_\sigma(u,w,\sigma) &= \frac{\partial}{\partial\sigma} G(u,w,\sigma) \\ &= -\frac{2}{\sigma^3} e^{-(u^2+w^2)/2\sigma^2} + \frac{(u^2+w^2)}{\sigma^5} e^{-(u^2+w^2)/2\sigma^2}. \end{aligned} \tag{3.23}$$

Let $I(x,y)$ and $I_\sigma(x,y)$ be the images obtained through the masks $G(\cdot)$ and $G_\sigma(\cdot)$, respectively. It can be shown, and follows from above, that these two images obey the following constraint:

$$I_\sigma(x,y) \;=\; \alpha^2 \sigma (I_{xx}(x,y) + I_{yy}(x,y)), \tag{3.24}$$

where $I_{xx}(x,y)$ and $I_{yy}(x,y)$ correspond to the horizontal and vertical second spatial derivatives of $I(x,y)$. The least-squares error function becomes:

$$E(\alpha^2) \;=\; \sum_p (I_\sigma(x,y) - \alpha^2\sigma(I_{xx}(x,y) + I_{yy}(x,y)))^2 \tag{3.25}$$

Solving for the minimizing α gives:

$$\alpha^2 \;=\; \frac{\sum_p I_\sigma(x,y)(I_{xx}(x,y) + I_{yy}(x,y))}{\left(\sum_p \sigma(I_{xx}(x,y) + I_{yy}(x,y))^2\right) + \varepsilon}. \tag{3.26}$$

There are several notable differences between this formulation based on aperture size derivatives and the previous formulation based on viewpoint derivatives. First, a second-order spatial derivative is required here, not only a first-order. Second, the ratio of the aperture size derivative and spatial derivative is proportional to the square of the parameter α which is proportional to range. As such, only the absolute value of α can be determined. Physically, this translates into an ambiguity in points equally spaced on either side of the focal plane. A second look at the optical masks reveals why this must be so. Whereas the viewpoint derivative mask is anti-symmetric with respect to the center of the lens (Figure 3.3), the aperture size derivative mask is symmetric (Figure 3.4). As a result, points equally spaced on either side of the focal plane will differ by a sign in the case of the anti-symmetric mask, but will appear identical in the case of the symmetric mask. This ambiguity may be eliminated by focusing the camera at infinity, ensuring that $\alpha > 0$. However in practice this may be impractical since the scene will be imaged completely out of focus, resulting in a poor spatial derivative signal.

3.2.3 Maximum Likelihood Estimation

In the least-squares estimation of α (Equations (3.11) and (3.21)) a small constant, ε, is added to the denominator to avoid singularities when the spatial derivative, I_x, is identically zero over the patch of integration. Here, we derive a systematic solution for this small constant based on a maximum likelihood estimator (MLE). In this formulation, we borrow heavily from the maximum likelihood motion estimation of [Simoncelli 93].

We will consider the 2-D optical differentiation with respect to viewpoint, where the parameter to be estimated, α, is given by the, now familiar, constraint:

$$I_x \alpha - I_v = 0, \tag{3.27}$$

where I_x and I_v are the spatial and viewpoint derivatives, respectively. Because of noise, filter and optical mask inaccuracies, etc., these derivatives are only approximations to the true derivatives, which are denoted as $\hat{I}_x$ and $\hat{I}_v$. The relationship between the true and measured quantities is made explicit by introducing a set of additive random variables:

$$I_x = \hat{I}_x + n_1 \quad \text{and} \quad I_v = \hat{I}_v + n_2. \tag{3.28}$$

The above constraint can then be expressed in terms of the measured quantities rather than the ideal quantities as follows:

$$\begin{aligned} 0 &= \hat{I}_x \alpha - \hat{I}_v \\ &= (I_x - n_1)\alpha - (I_v - n_2) \\ I_x \alpha - I_v &= n_2 - \alpha n_1. \end{aligned} \tag{3.29}$$

Unlike the previous ideal constraint, this constraint gives us a probabilistic relationship between the true and measured parameters and accounts for errors in the derivative measurements. The probability distributions for the random variables n_1 and n_2 are chosen so that α can be solved analytically, and without any knowledge of their true distributions! Thus, it is assumed that these random variables have independent zero-mean Gaussian distributions. Under these assumptions, the above constraint takes the form:

$$I_x \alpha - I_v = n_2 - \alpha n_1. \tag{3.30}$$

The right-hand side of this constraint equation is a zero-mean Gaussian random variable with variance $\lambda_2 - \alpha\lambda_1$, where λ_1 and λ_2 are the variances corresponding to the random variables n_1 and n_2, respectively. This constraint can now be interpreted as a conditional probability:

$$P(I_v \mid \alpha, I_x) \;=\; \exp\left(-\tfrac{1}{2}\frac{(I_x\alpha - I_v)^2}{\lambda_2 - \alpha\lambda_1}\right). \qquad (3.31)$$

This conditional probability can be rewritten, according to Bayes' rule [5], to give a conditional probability on the desired parameter α:

$$P(\alpha \mid I_x, I_v) \;=\; \frac{P(I_v \mid \alpha, I_x)P(\alpha)}{P(I_v)}. \qquad (3.32)$$

The prior distribution, $P(\alpha)$, is again chosen to be a zero-mean Gaussian with variance λ_p. Because the denominator in the above conditional probability is only a normalization factor and does not affect the relative probabilities, it can be ignored. Expanding the above expression gives:

$$\begin{aligned}
P(\alpha \mid I_x, I_v) &= \exp\left(-\tfrac{1}{2}\frac{(I_x\alpha - I_v)^2}{\lambda_2 - \alpha\lambda_1}\right)\exp\left(-\tfrac{1}{2}\frac{\alpha^2}{\lambda_p}\right) \\
&= \exp\left(-\tfrac{1}{2}\left[\left(\frac{I_x^2}{\lambda_2 - \alpha\lambda_1} + \frac{1}{\lambda_p}\right)\alpha^2 - \left(\frac{2I_xI_v}{\lambda_2 - \alpha\lambda_1}\right)\alpha + \left(\frac{I_v^2}{\lambda_2 - \alpha\lambda_1}\right)\right]\right) \\
&= \exp\left(-\tfrac{1}{2}\frac{(\alpha - \mu_\alpha)^2}{\lambda_\alpha}\right), \qquad (3.33)
\end{aligned}$$

where the mean, μ_α, and variance, λ_α, of this Gaussian distribution are determined by completing the square in the above exponential:

$$\mu_\alpha = \frac{I_xI_v}{(\lambda_2 - \alpha\lambda_1)\lambda_\alpha} \quad \text{and} \quad \lambda_\alpha = \frac{I_x^2}{\lambda_2 - \alpha\lambda_2} + \frac{1}{\lambda_p}. \qquad (3.34)$$

That is, the resulting probability distribution is a Gaussian parameterized by its mean and variance, and the maximum likelihood estimate (MLE) is simply the mean, μ_α. Expanding the MLE gives:

$$\begin{aligned}
\mu_\alpha &= \frac{I_xI_v}{I_x^2 + \frac{\lambda_2 - \alpha\lambda_1}{\lambda_p}} \\
&= \frac{I_xI_v}{I_x^2 + \varepsilon}, \qquad (3.35)
\end{aligned}$$

[5] According to Bayes' rule, $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$.

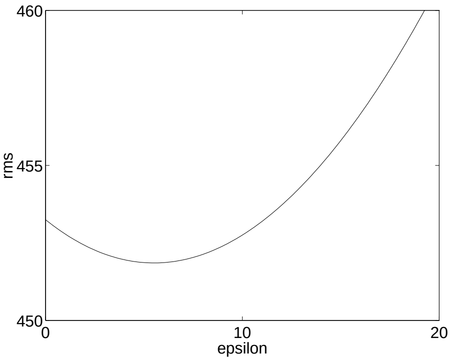


Figure 3.5: Illustrated is an example of the effect of varying ε in the maximum likelihood estimator (see Equation (3.36)). Empirically we have found that rms errors are minimized by choosing ε to be 0.02 of the mean of the denominator in the least-squares estimator (Equation (3.10)).

which now looks very much like the original least-squares formulation (Equation (3.11)). This is not surprising since a MLE with a Gaussian distribution is equivalent to a least-squares estimate. As before, the MLE can be determined by integrating over a small neighborhood in the image. If we assume that the noise at each point in the neighborhood is independent, zero-mean Gaussian, then the MLE takes the form:

$$\mu_\alpha \quad = \quad \frac{\sum_P I_x I_v}{\sum_P (I_x^2 + \varepsilon)}. \tag{3.36}$$

The assumption that the noise within a neighborhood is independent is unlikely to hold, but it provides a reasonable first approximation and allows for an analytic solution.

In this formulation, the arbitrary choice of ε in the least-squares solution has been replaced with a term based on a probabilistic analysis of the expected noise in the derivative measurements. Of course, since we have no evidence that the underlying distributions are independent, Gaussian, or zero-mean, the choice of ε is still equally arbitrary: the variances for the derivative measurements, λ_i, and the prior must still be chosen. Nonetheless, if we were so inclined, it would be possible to make some measurements and try to approximate the observed probability distributions with the required Gaussian distributions. In lieu of such measurements, we have found empirically that on average the rms error can be minimized by choosing ε to be 0.02 of the mean of denominator in the least-squares estimator (i.e., $\sum_P I_x^2$). Illustrated in Figure 3.5 is a simple example of how the rms error varies with the choice of ε.

3.2.4 Coarse To Fine

A major drawback of our direct differential range estimator and most other passive range estimation techniques is the inability to estimate range in regions with minimal or no texture. A standard solution to this problem is to implement a coarse-to-fine algorithm, also referred to as multi-resolution, scale-space or pyramid algorithms (e.g.,[Marr 79, Grimson 81, Terzopoulos 85, Lim 87, Hoff 89, Vleeschauwer 93, Gokstorp 95, Menard 96], or as in the motion estimation of [Simoncelli 93]). We expect that any of these approaches would be of benefit to our system.

3.2.5 "Steerable" Derivatives

We note here an interesting extension to the basic optical viewpoint differentiation formulation of Section 3.2.1. In the basic formulation, we focused exclusively on the horizontal, x, and vertical, y, viewpoint/spatial derivatives. However, this was a somewhat arbitrary choice, as any directional derivative would have sufficed. Is there any benefit to using one directional derivative over another? Possibly, but it may be difficult to determine, a priori, which directional derivative optical mask to use, and even then the choice may depend on the spatial location in the image. However this need not concern us since given the horizontal and vertical viewpoint/spatial derivatives, the derivative at *any* orientation can be determined *exactly* from a linear combination of these two. [6]

[6] The simplest example of this principle can be found by examining the directional derivative of a unit-variant 2-D Gaussian, $G(x,y) = e^{-(x^2+y^2)}$. Ignoring the scaling parameters, the partial derivatives in x and y are given by:

$$G_x(x,y) = -2xe^{-(x^2+y^2)} \quad \text{and} \quad G_y(x,y) = -2ye^{-(x^2+y^2)}.$$

Consider a polar representation of these derivatives (with radial portion, $r = \sqrt{x^2+y^2}$, and angular portion, $\theta = \tan^{-1}(y/x)$):

$$G_x(r,\theta) = -2re^{-r^2}\cos(\theta) \quad \text{and} \quad G_y(r,\theta) = -2re^{-r^2}\sin(\theta).$$

Note first that these derivatives are rotated copies of each other, that is, $G_x(r,\theta-\pi/2) = G_y(r,\theta)$ (this is trivial to see in that $\cos(\theta-\pi/2) = \sin(\theta)$). Now, consider, the directional derivative rotated to an *arbitrary* orientation, ϕ:

$$\begin{aligned} G_x(r,\theta-\phi) &= -2re^{-r^2}\cos(\theta-\phi) \\ &= -2re^{-r^2}(\cos(\theta)\cos(\phi)+\sin(\theta)\sin(\phi)) \\ &= \cos(\phi)(-2re^{-r^2}\cos(\theta)) + \sin(\phi)(-2re^{-r^2}\sin(\theta)) \\ &= \cos(\phi)G_x(r,\theta) + \sin(\phi)G_y(r,\theta). \end{aligned}$$

For our purposes, this property of the directional derivative may be exploited to adaptively and locally determine the direction at which the spatial (or viewpoint) derivative is strongest. These directional derivatives can then used in the same manner as in Equation (3.5). More specifically, consider the collection of images taken through the mask, $M(x, y)$, and the horizontal and vertical derivative masks, $M_x(x, y)$ and $M_y(x, y)$. These images are denoted as $I(x, y)$, $I_{v_x}(x, y)$, and $I_{v_y}(x, y)$, and the horizontal and vertical spatial derivatives of $I(x, y)$ are denoted as $I_x(x, y)$ and $I_y(x, y)$. The orientation, θ_s, and magnitude, r_s, of the spatial derivative in the direction of maximal change (i.e., the gradient) can be determined at each position in the image as:

$$\theta_s = \tan^{-1}\left(\frac{I_y}{I_x}\right) \quad \text{and} \quad r_s = \sqrt{I_x^2 + I_y^2}, \tag{3.37}$$

where the spatial parameters, x and y, are dropped for notational convenience. Similarly, the viewpoint derivative in the direction of maximal change is given by:

$$\theta_v = \tan^{-1}\left(\frac{I_{v_y}}{I_{v_x}}\right) \quad \text{and} \quad r_v = \sqrt{I_{v_x}^2 + I_{v_y}^2}. \tag{3.38}$$

For the purposes of range estimation, we require a viewpoint and spatial derivative in the *same* direction. As such, the viewpoint derivative can be "steered" to the orientation of maximal spatial derivative change, θ_s as:

$$I_{v_\theta} \quad = \quad r_v \cos(\theta_v - \theta_s), \tag{3.39}$$

The ratio of this derivative with the spatial derivative at the same orientation, $I_\theta = r_s$, gives an estimate of range (Equation (3.5)). By locally computing the directional derivative in the direction of maximal spatial variation, we benefit by adaptively (but still analytically) finding the strongest spatial derivative signal, boosting the signal-to-noise ratio and potentially avoiding singularities in the estimate of range.

That is, the directional derivative at any orientation, ϕ, can be synthesized from a linear combination of the directional derivatives, $G_x(x, y)$ and $G_y(x, y)$. In other words, these two directional derivatives form a complete (and, if properly scaled, orthonormal) basis for the complete set of rotations of the directional derivative. Note also that we only require a basis set of size two since the angular portion of this function is bandlimited: it contains only the first harmonic. Thus according to Nyquist (Section 1.3), only two samples are required to fully represent the full set of rotations. Also note that the x and y partial derivatives form the canonical basis, but that any two distinct directional derivatives will fully span the space. These principles were termed steerability by Freeman and Adelson in [Freeman 91] and applied to a variety of computer vision problems, see also [Danielsson 80, Knutsson 93, Koenderink 87] for some earlier work, [Simoncelli 92, Perona 92, Beil 94, Simoncelli 96a, Perona 95, Hel-Or 96] for extensions and generalizations of these principles and [Farid 96c] for a tutorial.

3.3 Assumptions and Constraints

As illustrated in Figure 3.6, several assumptions and constraints were imposed in deriving the differential techniques of the previous section (here, we consider the viewpoint derivative formulation, the same basic constraints hold for the aperture size formulation - any additional constraints will be noted). More specifically, computation of the matched spatial and viewpoint derivatives (I_x and I_v) assumes that the (1) intensity of each point in the world is uniform across the diameter of the lens (i.e., the brightness constancy assumption), (2) that the functional form of the optical masks takes on values in the range $[0, 1]$, (3) that the optical masks maintain a derivative relationship (i.e., any non-linearities in the mask formation have been corrected), (4) that the imaging process is linear, and (5) that the scene does not change between the acquisition of the image pair, I_1 and I_2. Further, computation of α, (monotonically related to range) assumes that (6) surfaces in the world are textured and locally frontal parallel. Finally, in order to compute range from α, it is assumed that (7) the intrinsic parameters of the imaging system are known (i.e., focal length, f, and lens to sensor distance, d_s).

In this section each of these assumptions and constraints are formulated precisely. The system's sensitivity to failures of these assumptions and to measurement noise is also described. Some of the details of this analysis are tedious, and so the reader uninterested in these details may wish to go directly to the summary in Section 3.3.8. At the conclusion of this chapter, the basic differential formulation and theoretical sensitivity analysis are validated in simulation and experimentation.

3.3.1 Brightness Constancy Assumption

Although not always explicit, the assumption of brightness constancy is made by virtually all range and motion estimation algorithms. [7] The brightness constancy assumption states that the brightness of a point in the world is constant when viewed from different positions. Note that in addition to constraining the photometric properties of surfaces,

[7] See [Negahdaripour 93] or [Gupta 95] for examples of an optical flow algorithm which relaxes the brightness constancy assumption.

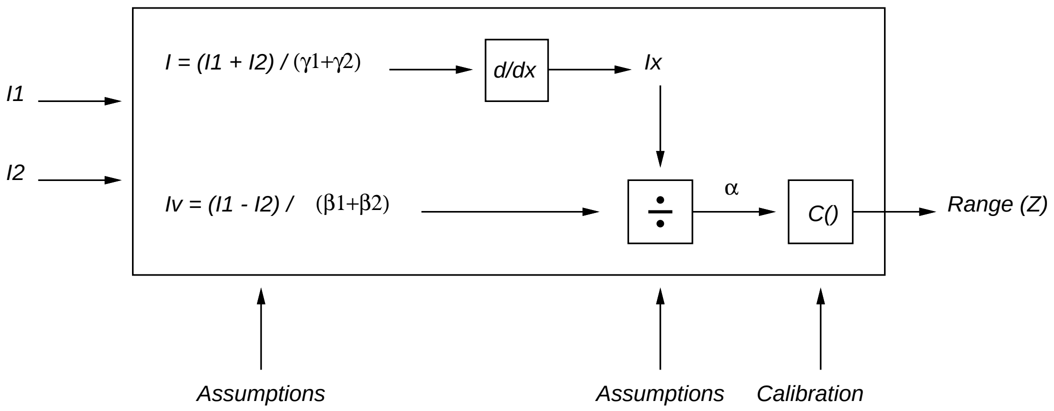


Figure 3.6: Assumptions and Constraints. Illustrated is a system diagram for computing depth from spatial and viewpoint derivatives (Section 3.2.1). As shown, and discussed in more detail in the text, several assumptions and constraints are introduced into various stages of the system. Section 3.3 formalizes these assumptions and analyses the effects of their failure on the overall system.

this assumption also implies that the scene cannot contain occlusions (Figure 3.7). For our purposes, this constraint need only hold across the diameter of the lens.

Denoting the brightness of a point as a function of viewing direction as $L(\cdot)$, the images of a single point light source under the required optical masks (Equation (3.4)) are:

$$I(x) = \frac{1}{\alpha} M\left(\frac{x}{\alpha}\right) L\left(\frac{x}{\alpha}\right) \quad \text{and} \quad I_v(x) = \frac{1}{\alpha} M'\left(\frac{x}{\alpha}\right) L\left(\frac{x}{\alpha}\right). \tag{3.40}$$

Note that the brightness variation, $L(\cdot)$, is introduced in a multiplicative fashion. In particular, the variation in light as a function of viewing position can be considered as simply projecting a *uniform* intensity point light source through an additional optical mask $L(\cdot)$, as illustrated in Figure 3.7. The combination of masks then acts in a multiplicative fashion. As before (Equation (3.5)) the spatial derivative of $I(x)$ is required; applying the chain rule gives:

$$I_x(x) \;=\; \frac{1}{\alpha^2} M'\left(\frac{x}{\alpha}\right) L\left(\frac{x}{\alpha}\right) + \frac{1}{\alpha^2} M\left(\frac{x}{\alpha}\right) L'\left(\frac{x}{\alpha}\right). \tag{3.41}$$

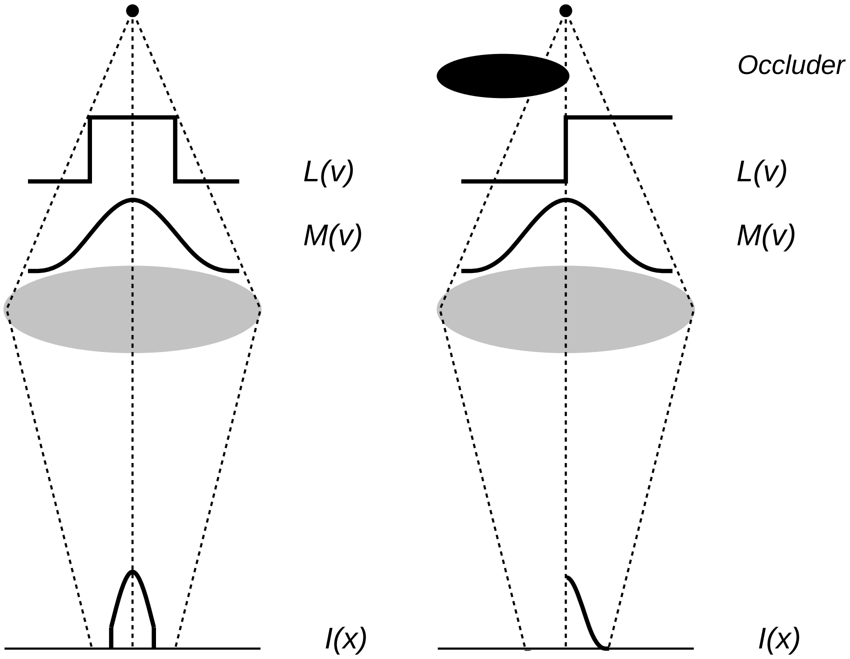


Figure 3.7: The Brightness Constancy Assumption. Illustrated are two examples of the failure of the brightness constancy assumption: the intensity of the point source is not uniform across the diameter of the lens, but varies with viewpoint, as specified by $L(v)$. In either case, the change in intensity can be modeled as a *uniform* light source passing through an additional optical attenuation mask. As a result, the image of the point source depends on the product of the light variation, $L(v)$, and optical mask, $M(v)$ (Equation (3.40)).

Combining with the viewpoint derivative gives:

$$\begin{aligned}
\alpha I_x(x) - I_v(x) &= \frac{1}{\alpha}M'\left(\frac{x}{\alpha}\right)L\left(\frac{x}{\alpha}\right) + \frac{1}{\alpha}M\left(\frac{x}{\alpha}\right)L'\left(\frac{x}{\alpha}\right) - \frac{1}{\alpha}M'\left(\frac{x}{\alpha}\right)L\left(\frac{x}{\alpha}\right) \\
&= \frac{1}{\alpha}M\left(\frac{x}{\alpha}\right)L'\left(\frac{x}{\alpha}\right) \\
\frac{I_v(x)}{I_x(x)} &= \alpha - \frac{M\left(\frac{x}{\alpha}\right)L'\left(\frac{x}{\alpha}\right)}{\alpha I_x(x)} \\
&= \alpha - \frac{I(x)L'\left(\frac{x}{\alpha}\right)}{I_x(x)L\left(\frac{x}{\alpha}\right)}. \qquad (3.42)
\end{aligned}$$

Where in the last step the definition of the image $I(x)$ was substituted for the mask function, $M(\cdot)$, to obtain an expression involving our measurement quantities. Note that the ratio of viewpoint and spatial derivatives no longer gives the desired parameter α (monotonically related to range, Equation (3.3)). However, if the brightness constancy assumption holds (i.e., $L(\cdot)$ is constant and its derivative, $L'(\cdot) = 0$) then the additional term in the above equation is zero and α can be solved for directly. If the brightness constancy assumption does not hold then our estimate of α, and hence range, will be either under or over estimated depending on the sign of the spatial and light intensity derivatives. [8]

Sensitivity of Brightness Constancy Assumption

It is difficult to precisely quantify the errors in range since they depend on several scene dependent factors. However, a perturbation analysis centered around $L'(\cdot) = 0$ (i.e., when the brightness constancy assumption holds) provides some insight into the sensitivity of the estimator. In particular, the estimate of α in Equation (3.42) is substituted into the definition of range (Equation (3.3)):

$$Z = \frac{d_s f}{d_s - f + f\left(\alpha - \frac{I(x)L'(\frac{x}{\alpha})}{I_x(x)L(\frac{x}{\alpha})}\right)}. \tag{3.43}$$

Taking the partial derivative with respect to L', and evaluating at $L' = 0$, gives:

$$\begin{aligned}
\frac{\partial Z}{\partial L'} &= \frac{-d_s f \frac{f I(x)}{I_x(x)L(\frac{x}{\alpha})}}{\left(d_s - f + f\left(\alpha - \frac{I(x)L'(\frac{x}{\alpha})}{I_x(x)L(\frac{x}{\alpha})}\right)\right)^2} \\
\frac{\partial Z}{\partial L'}|_{L'=0} &= \frac{-d_s f^2 \frac{I(x)}{I_x(x)L(\frac{x}{\alpha})}}{(d_s - f + f\alpha)^2} \\
&\propto Z^2 \frac{I(x)}{I_x(x)L\left(\frac{x}{\alpha}\right)} \qquad (3.44)
\end{aligned}$$

That is, errors due to the failure of the brightness constancy assumption scale with the square of the range. In addition, the errors are inversely proportional to both the spatial derivative, $I_x(x)$, and brightness, $L(\cdot)$, which intuitively makes sense since both may be considered as an indication of signal strength.

Relaxation of Brightness Constancy Assumption

The brightness constancy assumption may be restated in terms of a truncated Taylor series expansion on the light variation function, $L(\cdot)$. In particular, according to Taylor's

[8] We note an interesting (but probably not very useful) result from this analysis: if the light variation is symmetric about the lens center (e.g., $L(x) = x^2$) there is no effect on the estimate of range, even though, strictly speaking the brightness constancy assumption does not hold (i.e., $L'(x) \neq 0$). Intuitively this makes sense since given that the change in the light intensity on either side of the lens center is the same, the derivative with respect to viewpoint will not be affected. This result can also be seen when considering the above Equation (3.42) for a collection of point light sources. In this case, the derivative $L'(x/\alpha)$ in the numerator of the error term is replaced with the integral $\int dx_p \; L'((x - x_p)/\alpha)$, which integrates to zero for any symmetric function, $L(\cdot)$ (i.e., $L'(\cdot)$ is anti-symmetric).

theorem:

$$L(x) \quad = \quad L(c) + L'(c)(x-c) + \frac{L''(c)(x-c)^2}{2!} + \ldots + \frac{L^{(n)}(c)(x-c)^n}{n!} + \ldots . \tag{3.45}$$

The brightness constancy assumption then simply states that all higher-order terms of the Taylor series expansion are zero (i.e., $L(x) = L(c)$, a constant). One may imagine relaxing this assumption and allowing the first two terms of the expansion to be non-zero (i.e., $L(x) = L(c) + L'(c)(x-c)$). In other words, the light variation varies smoothly as a function of viewing position. In this case, the acquired pair of images takes the form:

$$\begin{aligned} I(x) &= \frac{1}{\alpha} M\left(\frac{x}{\alpha}\right)\left[L(c) + L'(c)\left(\frac{x}{\alpha} - c\right)\right] \quad \text{and} && (3.46) \\ I_v(x) &= \frac{1}{\alpha} M'\left(\frac{x}{\alpha}\right)\left[L(c) + L'(c)\left(\frac{x}{\alpha} - c\right)\right]. && (3.47) \end{aligned}$$

The above equation provides only two constraints in three unknowns, α, $L(\cdot)$ and $L'(\cdot)$, and is therefore underconstrained. However an additional measurement may be made with a second-order derivative mask:

$$I_{vv}(x) = \frac{1}{\alpha} M''\left(\frac{x}{\alpha}\right)\left[L(c) + L'(c)\left(\frac{x}{\alpha} - c\right)\right]. \tag{3.48}$$

Combined with the two previous measurements, $I(x)$ and $I_v(x)$, we now have a system of three equations in three unknowns. In order to relate the three measurements, the first and second-order spatial derivatives of $I_v(x)$ and $I(x)$ need to be considered (Figure 3.8). In particular, consider the second spatial derivative of $I(x)$, the first spatial derivative of $I_v(x)$ and the measurement $I_{vv}(x)$:

$$\begin{aligned} I_{xx}(x) &= \frac{1}{\alpha^3} M''\left(\frac{x}{\alpha}\right) L(c) \\ &+ \left[\left(\frac{1}{\alpha^3} M''\left(\frac{x}{\alpha}\right) L'(c)\left(\frac{x}{\alpha} - c\right) + \frac{1}{\alpha^3} M'\left(\frac{x}{\alpha}\right) L'(c)\right) + \frac{1}{\alpha^3} M'\left(\frac{x}{\alpha}\right) L'(c)\right] && (3.49) \\ I_{vx}(x) &= \frac{1}{\alpha^2} M''\left(\frac{x}{\alpha}\right) L(c) + \left[\frac{1}{\alpha^2} M''\left(\frac{x}{\alpha}\right) L'(c)\left(\frac{x}{\alpha} - c\right) + \frac{1}{\alpha^2} M'\left(\frac{x}{\alpha}\right) L'(c)\right] && (3.50) \\ I_{vv}(x) &= \frac{1}{\alpha} M''\left(\frac{x}{\alpha}\right)\left[L(c) + L'(c)\left(\frac{x}{\alpha} - c\right)\right]. && (3.51) \end{aligned}$$

With some algebraic manipulations, it is possible to show that:

$$\alpha^2 I_{xx}(x) - 2\alpha I_{vx}(x) + I_{vv}(x) \quad = \quad 0. \tag{3.52}$$

That is, α (monotonically related to range) can be solved for directly from the three measurements and their appropriate spatial derivatives. Solving for α now requires solving a second-order polynomial, and the two solutions of this polynomial will differ in their sign.

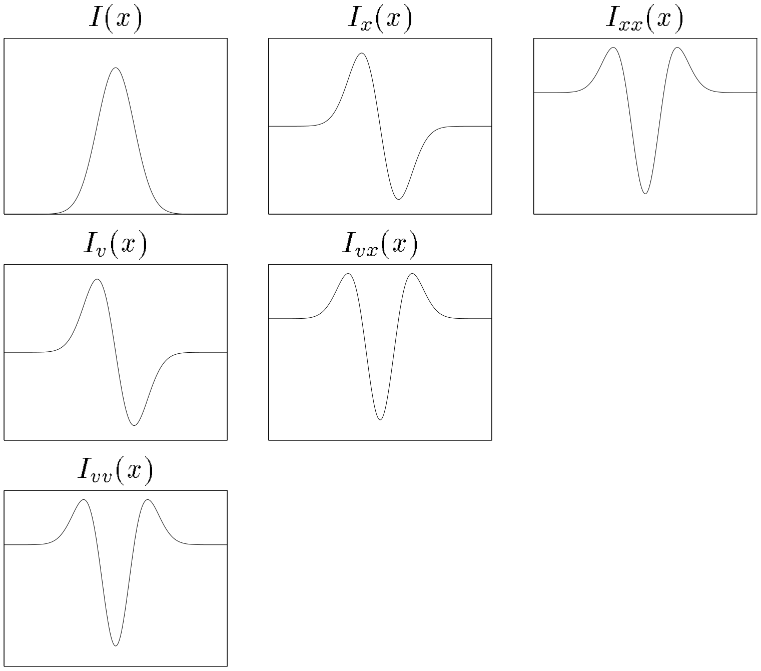


Figure 3.8: Higher-Order Derivative Masks. The first column of this figure illustrates the image of a uniform point light source imaged through an optical mask $M(\cdot)$, and its first- and second-order spatial derivative, $M'(\cdot)$ and $M''(\cdot)$. To the right of each plot are the spatial derivatives of these images. Note that, pictorially (and in fact mathematically, Equation 3.52), the spatial derivatives of the images formed under lower-order derivative masks are similar to the images formed under higher-order derivative masks (e.g., $I_x(x)$ is similar to $I_v(x)$, and $I_{xx}(x)$ is similar to $I_{vx}(x)$ and $I_{vv}(x)$).

Physically, the two solutions correspond to points equally spaced on either side of the focal plane. The reason for the ambiguity in this case is that, unlike the first derivative mask, the second derivate mask is symmetric about its origin (Figure 3.8). In general, even-order derivatives will have such an ambiguity, and odd-order derivatives will not.

In the original formulation ($\alpha I_x(x) - I_v(x) = 0$), it was assumed that the brightness variation was constant (i.e., $L'(\cdot) = 0$), while the above formulation assumes only that the second-order derivative is zero (i.e., $L''(\cdot) = 0$). It should be clear that by making additional measurements with increasingly higher-order derivative masks, the brightness constancy assumption can be further relaxed from a constraint on the first derivative of $L(\cdot)$ (with only two measurements) to a constraint on the n^{th} derivative (with $n+1$ measurements). In other words, additional measurements allow us to measure higher-order terms in the Taylor series expansion of $L(\cdot)$. The general form for approximating α from n measurements (under the optical mask $M(\cdot)$ and its first $n-1$ derivatives) is:

$$\begin{aligned} C(n-1,0)\alpha^{n-1}I_{v^{(0)}x^{(n-1)}} + C(n-1,1)\alpha^{n-2}I_{v^{(1)}x^{(n-2)}} + \ldots + & \\ C(n-1,n-2)\alpha^{1}I_{v^{(n-2)}x^{(1)}} + C(n-1,n-1)\alpha^{0}I_{v^{(n-1)}x^{(0)}} &= 0 \\ \sum_{i=0}^{n} C(n,i)\alpha^{n-i}I_{v^{(i)}x^{(n-i)}} &= 0, \qquad (3.53) \end{aligned}$$

where, $C(a,b) = \frac{a!}{b!(a-b)!}$, and $I_{v^{(a)}x^{(b)}}$ is the b^{th} spatial derivative of the image taken through the a^{th}-order derivative mask. Note that in the case of $n = 1$, the above reduces to the original equation, $\alpha I_x - I_v = 0$, under the general brightness constancy assumption.

Although in theory the brightness constancy assumption can be virtually eliminated, in practice this may not actually be practical. More specifically the number of non-zero terms in the Taylor series expansion of the lighting variation function ($L(x)$) will be large for regions where the failure of brightness constancy assumption is most severe (e.g. specularities and occlusion boundaries)

3.3.2 Locally Frontal-Parallel Assumption

As with the brightness constancy assumption, an assumption of locally frontal-parallel surface geometry is frequently made (often implicitly) by most range and motion estimation algorithms. Recall that this assumption was necessary in order to solve for α (monotonically related to range) embedded in the integral of Equation (3.8). Here, we quantify the effects of this assumption and the effects of its failure on the estimation of range.

For simplicity consider a world consisting of only two uniform intensity point light sources at different depths (Figure 3.9). These point light sources satisfy the brightness constancy assumption, but may have overall different individual brightnesses. The image of two such sources formed through an optical mask, $M(\cdot)$, is a superposition of scaled, dilated and translated (by different amounts) copies of $M(\cdot)$:

$$I(x) = \frac{1}{\alpha_1}M\left(\frac{x-x_1}{\alpha_1}\right)L_1 + \frac{1}{\alpha_2}M\left(\frac{x-x_2}{\alpha_2}\right)L_2. \qquad (3.54)$$

Similarly, the image through the derivative mask, $M'(\cdot)$ is:

$$I_v(x) = \frac{1}{\alpha_1}M'\left(\frac{x-x_1}{\alpha_1}\right)L_1 + \frac{1}{\alpha_2}M'\left(\frac{x-x_2}{\alpha_2}\right)L_2. \qquad (3.55)$$

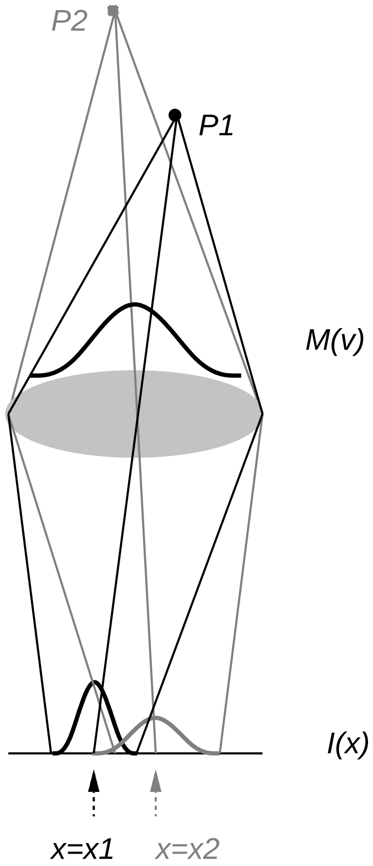


Figure 3.9: The Local Frontal-Parallel Assumption. Illustrated is an example of the failure of the local frontal-parallel assumption: two points at different depths are imaged onto the sensor plane. The resulting image is a superposition of scaled, dilated and translated copies (by different amounts) of the mask function. As a result, solving for α (monotonically related to range) becomes non-trivial (Equation (3.60)).

In both of these equations L_1 and L_2 are the brightnesses of the pair of points and are constant across viewpoint (i.e., the brightness constancy assumption of Section 3.3.1 holds). Computing the spatial derivative of $I(x)$ gives:

$$I_x(x) \quad = \quad \frac{1}{\alpha_1^2} M'\left(\frac{x - x_1}{\alpha_1}\right) L_1 + \frac{1}{\alpha_2^2} M'\left(\frac{x - x_2}{\alpha_2}\right) L_2. \tag{3.56}$$

As before, we consider the ratio of the viewpoint and spatial derivatives:

$$\frac{I_v(x)}{I_x(x)} \quad = \quad \frac{\frac{1}{\alpha_1} M'\left(\frac{x-x_1}{\alpha_1}\right) L_1}{\frac{1}{\alpha_1^2} M'\left(\frac{x-x_1}{\alpha_1}\right) L_1 + \frac{1}{\alpha_2^2} M'\left(\frac{x-x_2}{\alpha_2}\right) L_2} + \frac{\frac{1}{\alpha_2} M'\left(\frac{x-x_2}{\alpha_2}\right) L_2}{\frac{1}{\alpha_1^2} M'\left(\frac{x-x_1}{\alpha_1}\right) L_1 + \frac{1}{\alpha_2^2} M'\left(\frac{x-x_2}{\alpha_2}\right) L_2} \tag{3.57}$$

multiplying term i by $\frac{\alpha_i^2}{M'(x-x_i)L_i}$, with $i = 1, 2$ gives:

$$\frac{I_v(x)}{I_x(x)} = \frac{\alpha_1}{1 + \frac{\alpha_1^2}{\alpha_2^2}\frac{M'\left(\frac{x-x_2}{\alpha_2}\right)L_2}{M'\left(\frac{x-x_1}{\alpha_1}\right)L_1}} + \frac{\alpha_2}{1 + \frac{\alpha_2^2}{\alpha_1^2}\frac{M'\left(\frac{x-x_1}{\alpha_1}\right)L_1}{M'\left(\frac{x-x_2}{\alpha_2}\right)L_2}}. \tag{3.58}$$

Substituting in $I_x(\cdot) = \frac{1}{\alpha^2}M'(\cdot)L$ in the above expression gives:

$$\begin{aligned} \frac{I_v(x)}{I_x(x)} &= \frac{\alpha_1}{1 + \frac{I_x(x-x_2)}{I_x(x-x_1)}} + \frac{\alpha_2}{1 + \frac{I_x(x-x_1)}{I_x(x-x_2)}} \\ &= \frac{\alpha_1 I_x(x-x_1)}{I_x(x-x_1) + I_x(x-x_2)} + \frac{\alpha_2 I_x(x-x_2)}{I_x(x-x_2) + I_x(x-x_1)} \\ &= \frac{\alpha_1 I_x(x-x_1) + \alpha_2 I_x(x-x_2)}{I_x(x-x_1) + I_x(x-x_2)}. \end{aligned} \tag{3.59}$$

Note that if, according to our assumption, the two points have the same range, then $\alpha_1 = \alpha_2$, and the above equation reduces to α_1. The general form of Equation (3.59) for n points is given by:

$$\frac{I_v(x)}{I_x(x)} = \frac{\sum_{i=1}^{n} \alpha_i I_x(x-x_i)}{\sum_{j=1}^{n} I_x(x-x_j)}. \tag{3.60}$$

That is, from a local region in the image, the estimate of α is determined from an average weighted by the spatial derivative. Note once again that if $\alpha_i = \alpha$, $\forall i \in [1, n]$, then $\frac{I_v(x)}{I_x(x)} = \alpha$, and there is no bias in the estimate of range.

Equation (3.60) suggests that if the local frontal-parallel assumption does not hold, then the estimate of α is a weighted average (by the spatial derivative) of neighboring points. This effect is illustrated qualitatively in Figure 3.10. Shown are 1-D range maps with spatial position plotted along the horizontal axis and range along the vertical axis. Ground truth is a ramp in depth with varying slope (dotted line), and the estimated range map (solid line) is computed by a weighted sum of α over a local neighborhood. In each case the spatial derivative is a raised cosine (dotted line). Note that, as suggested by Equation (3.60), the periodic errors in range coincide with the underlying spatial derivative. That is, the estimate of range is unbiased at the points where the spatial derivative is maximal, and is otherwise under or overestimated. Although seemingly an overly restrictive constraint, we will show empirically that the frontal-parallel constraint poses little problem in the estimation of range even for significantly slanted or curved surfaces (Section 3.4.1).

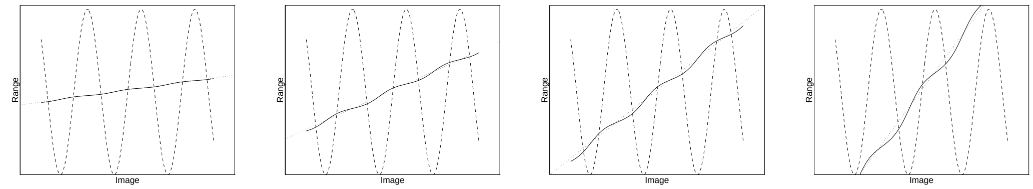


Figure 3.10: Failure of Local Frontal-Parallel Assumption. Illustrated are synthetic 1-D range maps, with spatial position plotted along the horizontal axis, and range along the vertical axis. Ground truth is a ramp oriented, from left to right, at 10, 30, 45, and 60 degrees (dotted line). The estimated range map is given by the solid line, and the spatial derivative is a raised cosine denoted by the dashed line. Range was computed from a weighted average (by the spatial derivative) of range over a local neighborhood, as in Equation (3.60). Note that the errors in range coincide with the underlying spatial derivative.

Sensitivity of Local Frontal-Parallel Assumption

The sensitivity to the failure of frontal-parallel assumption can be determined by a similar perturbation analysis as in the previous section. In particular, we examine the differential variation in the estimate of range for slanted surfaces. To simplify this analysis the failure of the frontal-parallel assumption is assumed to take the form of a planar surface with slope m. That is, the variation in range as a function of spatial position is of the form $mX + Z_0$, and the estimate of α is then proportional to $\sum_i \frac{1}{mX_i+Z_0}$. Substituting this estimate of α into the definition of range gives:

$$Z \quad = \quad \frac{d_s f}{d_s - f + f\sum_i \frac{1}{mX_i+Z_0}}. \qquad (3.61)$$

Differentiating this expression with respect to the slope of the surface, m, and evaluating at $m = 0$ (i.e., a frontal-parallel surface) gives:

$$\begin{aligned} \frac{\partial Z}{\partial m} &= \frac{d_s f^2 \sum_i \frac{X_i}{(mX_i+Z_0)^2}}{(d_s - f + f\sum_i \frac{1}{mX_i+Z_0})^2} \\ \frac{\partial Z}{\partial m}\Big|_{m=0} &= \frac{d_s f^2 \sum_i \frac{X_i}{Z_0^2}}{(d_s - f + f\alpha)^2} \\ &\propto Z^2 \sum_i \frac{X_i}{Z_0^2}. \end{aligned} \qquad (3.62)$$

That is, errors in range due to the failure of the frontal-parallel assumption scale with the square of the range. Also note that the errors will scale with the range of integration (i.e., the spatial extent of i in the above summation).

Correcting for Non-Planarity

We have found, empirically, that the range estimator performs quite well even in the presence of large deviations from the frontal-parallel assumption (see Section 3.4.1). Nonetheless, we sketch an iterative solution for solving for the α parameter embedded in the integral (Equation (3.6). That is, a solution for α when it cannot be pulled out of the integral due to a failure of the frontal-parallel assumption.

First, recall that for complex visual scenes the image formed through an optical mask, $M(\cdot)$, is a superposition of scaled, dilated, and translated copies of the mask function:

$$I(x) = \int dx_p \, \frac{1}{\alpha_p} M\left(\frac{x - x_p}{\alpha_p}\right) L\left(\frac{x_p}{\alpha_p}\right). \tag{3.63}$$

Consider then a discrete version of the image formed under the derivative mask $M'(\cdot)$ (we arbitrarily choose $I_v(\cdot)$, equivalently, $I_x(\cdot)$ may have been chosen):

$$I_v(x) \;=\; \sum_{x_p} \frac{1}{\alpha_p} M'\left(\frac{x - x_p}{\alpha_p}\right) L\left(\frac{x_p}{\alpha_p}\right), \tag{3.64}$$

which may be written in matrix notation as follows:

$$\begin{aligned} \left(\begin{array}{c} \\ \\ I_v(\cdot) \\ \\ \\ \end{array} \right) &= \left(\begin{array}{ccc} & & \\ & & \\ & L(\cdot)M'(\cdot) & \\ & & \\ & & \end{array} \right) \left(\begin{array}{c} \\ \\ \frac{1}{\alpha_p} \\ \\ \\ \end{array} \right) \\ \vec{I_v} &= A\vec{\alpha}, \end{aligned} \tag{3.65}$$

where the vector $\vec{I_v}$ contains the measured image, the rows of the matrix A consist of scaled, dilated (by an amount α_p) and translated copies of the mask M', and initially, the vector, $\vec{\alpha}$ contains the estimate of $\alpha_p = \frac{I_v(x_p)}{I_x(x_p)}$. The vector $\vec{\alpha}$ can be re-estimated by inverting the matrix A and pre-multiplying it by the vector $\vec{I_v}$. A new matrix A is then computed with this new estimate of $\vec{\alpha}$, and the process repeated until the process converges or the difference between successive iterations becomes sufficiently small.

3.3.3 Optical Mask Constraints

Due to the physical constraints of an optical attenuation mask, the functional form of any such mask must only contain values in the range $[0, 1]$, where a value of 0 corresponds to full attenuation and a value of 1 corresponds to no attenuation. The viewpoint and aperture size derivative masks both contain negative values, and as such may not be used directly. Furthermore, simply adding a positive constant to the derivative mask destroys the derivative relationship between them, and this destroys the simple relationship between range and the measured spatial and viewpoint derivatives. For example, consider the pair of non-negative 1-D masks, $M(u)$ and $M'(u) + c$. From Equation (3.2) the images of a uniform intensity point light source through these masks are given by:

$$I(x) = \frac{1}{\alpha} M\left(\frac{x}{\alpha}\right) L \quad \text{and} \quad I_v(x) = \frac{1}{\alpha}\left(M'\left(\frac{x}{\alpha}\right) + c\right) L. \tag{3.66}$$

As before, computing the spatial derivative of $I(x)$ gives:

$$I_x(x) = \frac{1}{\alpha^2} M'\left(\frac{x}{\alpha}\right) L. \tag{3.67}$$

But now, the ratio of $I_v(x)$ and $I_x(x)$ is no longer related by a simple scale factor, α; an additional additive term, $\frac{c\alpha}{M'(x/\alpha)}$, occurs in the equation.

Nonetheless a pair of non-negative masks can be constructed by taking the appropriate linear combination of the original masks. If the imaging system is linear (see Section 3.4.2), then the original masks can be reconstructed from the non-negative masks. In particular, consider the following construction of a pair of non-negative masks:

$$M_1(u) = \beta_1 M(u) + \gamma_1 M'(u) \quad \text{and} \quad M_2(u) = \beta_2 M(u) - \gamma_2 M'(u), \tag{3.68}$$

where the scaling parameters $\beta_{(1,2)}$ and $\gamma_{(1,2)}$ are chosen such that $M_1(u)$ and $M_2(u)$ lie in the range $[0, 1]$. The desired masks, $M(u)$ and $M'(u)$ can be reconstructed through a simple linear combinations of the non-negative masks, $M_1(u)$ and $M_2(u)$:

$$M(u) = \frac{M_1(u) + M_2(u)}{\beta_1 + \beta_2} \quad \text{and} \quad M'(u) = \frac{M_1(u) - M_2(u)}{\gamma_1 + \gamma_2}. \tag{3.69}$$

The scaling constants, $\beta_{(1,2)}$ and $\gamma_{(1,2)}$, may be determined in two stages: (1) add the appropriate amount of $M(u)$ to $M'(u)$ to make their sum/difference non-negative (if chosen properly, this will make the minimum value of the sum/difference equal to 0); (2) scale the sum/difference so that its maximum value is 1. These scaling constants will be denoted as $b_{(1,2)}$ and $c_{(1,2)}$, respectively. The former is the maximum ratio of $|M'(u)/M(u)|$ over the negative/positive values only, the latter is simply the maximum of the sum/difference computed after applying $b_{(1,2)}$. The non-negative masks then take the form:

$$M_1(u) = \frac{1}{c_1}\left(b_1 M(u) + M'(u)\right) \quad \text{and} \quad M_2(u) = \frac{1}{c_2}\left(b_2 M(u) - M'(u)\right), \tag{3.70}$$

where the desired scaling constants are simply:

$$\beta_1 = \frac{b_1}{c_1}, \quad \gamma_1 = \frac{1}{c_1} \qquad \text{and} \qquad \beta_2 = \frac{b_2}{c_2}, \quad \gamma_2 = \frac{1}{c_2} \tag{3.71}$$

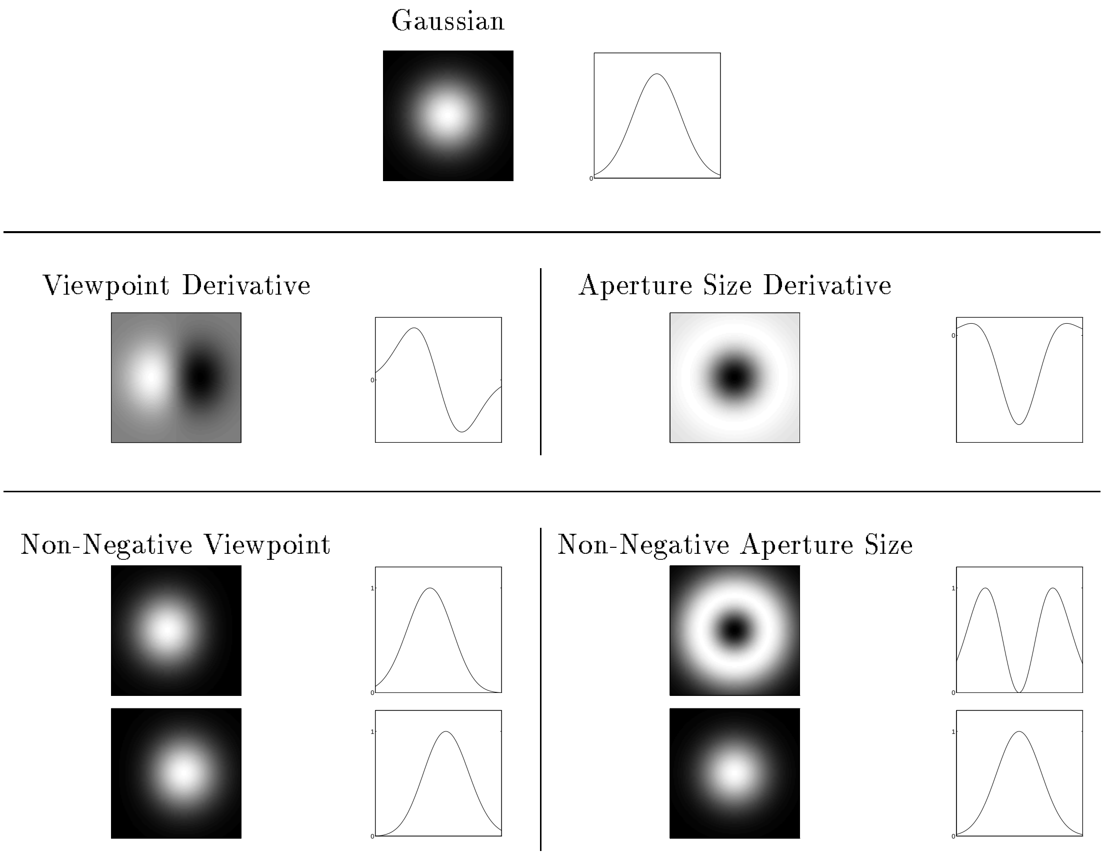
Now, according to our initial assumption of linearity, the desired *images* formed under the masks $M(u)$ and $M'(u)$ can be determined from the *images* formed under the masks $M_1(u)$ and $M_2(u)$:

$$I(x) = \frac{I_1(x) + I_2(x)}{\beta_1 + \beta_2} \quad \text{and} \quad I_{v,a}(x) = \frac{I_1(x) - I_2(x)}{\gamma_1 + \gamma_2}, \tag{3.72}$$

where $I_1(x)$ and $I_2(x)$ are the images formed under the masks $M_1(u)$ and $M_2(u)$, respectively, and $I_{v,a}$ corresponds to either the viewpoint or aperture size derivative. Clearly, this construction extends to 2-D optical masks as well. Illustrated in Figure 3.11 is a Gaussian mask, $M(u,w) = \frac{1}{2\pi\sigma^2}e^{-(u^2+w^2)/2\sigma^2}$, and it's partial derivatives with respect to u (viewpoint derivative) and σ (aperture size derivative). Also illustrated are the non-negative masks constructed according to Equation (3.68).

Calibrating for Camera's Point Spread Function

In describing the formation of an image through an optical attenuation mask we have been assuming that the camera's own point spread function (PSF) is constant across the lens diameter. In other words, it has been assumed that the image of a point light source with no attenuation mask is a scaled and dilated copy of a hard-edged rectangular function. This of course is generally not the case: the camera's PSF typically takes on a Gaussian-like shape. The combination of the PSF and the optical mask will then act in a multiplicative fashion. Whereas before, we employed a matched pair of masks, $M(v)$ and $M'(v)$, with



Gaussian

Viewpoint Derivative

Aperture Size Derivative

Non-Negative Viewpoint

Non-Negative Aperture Size

Figure 3.11: Non-Negative Gaussian-Based Optical Masks. Illustrated are a series of 2-D optical masks and 1-D horizontal slices through their vertical mid-point. In particular, illustrated is a Gaussian mask, $G(u, w, \sigma)$, (first row) and its derivative with respect to u (i.e., viewpoint derivative) (second row, left), and its derivative with respect to σ (i.e., aperture size derivative) (second row, right). Also shown are the non-negative masks constructed according to Equation (3.68). Note that, as required, the non-negative masks lie in the range $[0, 1]$.

$\frac{dM(v)}{dv} = M'(v)$, we now require a pair of masks that obey the following constraint:

$$\frac{d(M(v)H(v))}{dv} = \hat{M}(v)H(v), \qquad (3.73)$$

where $H(v)$ is the camera PSF. That is, the derivative relationship is no longer imposed on the optical masks, rather, this constraint is imposed on the product of the optical mask and the PSF. Although in our experiments (see Section 3.4.2), we have *not* calibrated for the camera's point spread function, it is most certain that doing so will improve the overall accuracy of this approach.

Aperture Size Masks

When considering the functional form of the viewpoint derivative mask, we required only that the mask to be C^1 differentiable. In the case of the aperture size mask (Section 3.2.2) we required the mask function to be C^2 differentiable and imposed the additional constraint that:

$$M''(u) \quad = \quad -(M(u) + uM'(u)) \tag{3.74}$$

It is straight-forward to show that, within a scale factor, a Gaussian satisfies this property:

$$\begin{aligned}
G(u) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-u^2/2\sigma^2} && (3.75)\\
G'(u) &= \frac{-u}{\sqrt{2\pi}\sigma^3} e^{-u^2/2\sigma^2} && (3.76)\\
G''(u) &= \frac{-1}{\sqrt{2\pi}\sigma^3} e^{-u^2/2\sigma^2} + \frac{u^2}{\sqrt{2\pi}\sigma^5} e^{-u^2/2\sigma^2} &&\\
&= \frac{-1}{\sigma^2}\left(\frac{1}{\sqrt{2\pi}\sigma} e^{-u^2/2\sigma^2} - \frac{u^2}{\sqrt{2\pi}\sigma^3} e^{-u^2/2\sigma^2}\right) &&\\
&= \frac{-1}{\sigma^2}\left(G(u) + uG'(u)\right) && (3.77)
\end{aligned}$$

The full family of permissible mask functions may be determined by noting that the constraint on the mask function is given by a second-order linear homogeneous differential equation that may be solved for explicitly. In particular, the constraint in Equation (3.74) is of the general form:

$$f''(x) + p(x)f'(x) + q(x)f(x) = 0 \tag{3.78}$$

We provide, without proof, the following theorem for differential equations of this form:

> **Theorem 3.3.1** *if f_1 and f_2 are solutions to a second-order linear homogeneous differential equation of the form in Equation (3.78), and $W(f_1, f_2)(x) = f_1 f_2' - f_1' f_2$, referred to as the Wronskian* [9]*, then:*

[9] in honor of the Polish mathematician Wronski (1778-1853). Also, note that the Wronskian, $W(f_1, f_2)(x) = f_1 f_2' - f_1' f_2$, is the determinant of the matrix $\begin{pmatrix} f_1 & f_2 \\ f_1' & f_2' \end{pmatrix}$, and the two cases in the above theorem amount to determining if these functions span the full space of possible solutions of the second-order differential equation (i.e., if the determinant is non-zero).

1. *if f_1 and f_2 are linearly dependent then* $W(f_1, f_2) = 0$,

2. *or, if f_1 and f_2 are linearly independent then* $W(f_1, f_2) = ce^{-\int dx\ p(x)} \neq 0$.

where two functions f_1 and f_2 are said to be *linearly dependent* if there exists two constants c_1 and c_2, not both zero, such that $c_1 f_1(x) + c_2 f_2(x) = 0$, $\forall x$. Two functions f_1 and f_2 are said to be *linearly independent* if $c_1 f_1(x) + c_2 f_2(x) = 0$, $\forall x$ only if $c_1 = c_2 = 0$. Working with a known solution, the Gaussian, the full family of permissible mask functions will be derived. We know, by inspection, that $f_1(x) = e^{-x^2/2}$ is a solution to the differential equation 3.78. Considering each case in the above theorem separately, let f_2 be another solution, then if f_1 and f_2 are linearly dependent:

$$\begin{aligned}
W(f_1, f_2) &= 0 \\
e^{-x^2/2} f_2' + x e^{-x^2/2} f_2 &= 0 \\
(e^{-x^2/2} f_2)' &= 0 \\
\int dx\ (e^{-x^2/2} f_2)' &= \int dx\ 0 \\
e^{-x^2/2} f_2 &= c \\
f_2 &= c e^{x^2/2} \qquad (3.79)
\end{aligned}$$

It is straight-forward to verify that functions of this form satisfy our constraint and thus provides an alternate mask function for the differential aperture size formulation. Now, if f_1 and f_2 are linearly independent then:

$$\begin{aligned}
W(f_1, f_2) &= ce^{-\int dx\ x} \\
e^{-x^2/2} f_2' + x e^{-x^2/2} f_2 &= ce^{-\int dx\ x} \\
(e^{-x^2/2} f_2)' &= ce^{-x^2/2} \\
\int dx\ (e^{-x^2/2} f_2)' &= \int dx\ ce^{-x^2/2} \\
e^{-x^2/2} f_2 &= erf(x) \\
f_2 &= erf(x) e^{x^2/2}, \qquad (3.80)
\end{aligned}$$

where $erf(x)$ is the error function, and although it cannot be solved for analytically, numeric approximations are available.

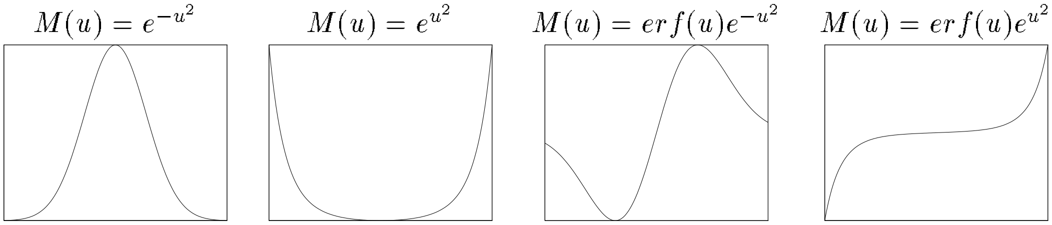


Figure 3.12: Family of Aperture Size Masks. The differential aperture size formulation requires that the functional form of the mask, $M(u)$ be C^2 differentiable and satisfy the constraint that $M''(u)+uM'(u)+M(u) = 0$. By solving this second-order differential equation, the full family of possible mask functions has been determined (see text). Illustrated above, are examples of each class of functions: the Gaussian, e^{-u^2}, the exponential, e^{u^2}, and the exponentially modulated error function, $erf(u)e^{u^2}$ and $erf(u)e^{-u^2}$. In practice, only the Gaussian mask is used because the other masks cannot be expressed analytically, or do not fall smoothly to zero, introducing artifacts when computing spatial derivatives of the image formed under them.

Additional solutions can be found by recursively considering each of the above solutions as a starting solution (i.e., f_1). In that case, a fourth and final solution of the form $f_2 = erf(x)e^{-x^2/2}$ is found. Illustrated in Figure 3.12 are examples of each of the possible mask functions: the Gaussian, e^{-u^2}, the exponential, e^{u^2}, and the exponentially modulated error function, $erf(u)e^{u^2}$ and $erf(u)e^{-u^2}$. For practical purposes, only the Gaussian is used since the other functions cannot be expressed analytically or because they do not fall smoothly to zero at the mask edge and will introduce artifacts in the spatial derivative computations.

Admittedly, this formulation was somewhat anticlimactic in that we finished at the same place in which we began, with the Gaussian mask and its derivative with respect to its dilation parameter as the only permissible and practical mask function for the differential aperture size formulation. Finally, note that the choice of a Gaussian is specific to the initial constraint Equation (3.74), which was somewhat arbitrary. There may be other constraints, with alternate solutions, that will allow us to solve for range from the spatial and aperture size derivatives. This may be an interesting avenue to explore, however we will not pursue this area further.

Sensitivity to Derivative Relationship

We have seen several times now that in order to estimate range, one mask must be the derivative of the other. In practice, this constraint may not be strictly met (e.g., see Section 3.4.2 on dithering), in which case we are interested in determining the resulting errors. Consider the pair of masks, $\tilde{M}(\cdot)$ and $M'(\cdot)$, where $M'(\cdot)$ is *not* necessarily the derivative $\tilde{M}(\cdot)$. The spatial derivative of the image formed under the mask $\tilde{M}(\cdot)$, and the image formed under the mask $M(\cdot)$ are:

$$I_x(x) = \frac{1}{\alpha^2}\tilde{M}'\left(\frac{x}{\alpha}\right) L \quad \text{and} \quad I_v(x) = \frac{1}{\alpha}M'\left(\frac{x}{\alpha}\right) L. \tag{3.81}$$

The estimate of α is the ratio of these images, $\frac{I_v(x)}{I_x(x)} = \alpha\frac{M'(x/\alpha)}{\tilde{M}'(x/\alpha)}$. Substituting this estimate of α into Equation (3.3) gives:

$$Z = \frac{d_s f}{d_s - f + \alpha\frac{\tilde{M}'}{M'}f} \tag{3.82}$$

The error in range, Z, as a function of the deviation from the derivative constraint on the masks can be quantified by taking the derivative of the above equation and evaluating at $\tilde{M}'(\cdot) = M'(\cdot)$ (i.e., when the masks are properly matched):

$$\begin{aligned}
\frac{\partial Z}{\partial \tilde{M}'} &= -\frac{\alpha d_s f^2}{M'\left(d_s - f + \alpha\frac{\tilde{M}'}{M'}f\right)^2} \\
\frac{\partial Z}{\partial \tilde{M}'}\Big|_{\tilde{M}'=M'} &= -\frac{\alpha d_s f^2}{M'(d_s - f + \alpha f)^2} \\
&\propto Z^2\frac{\alpha}{M'} \\
&\propto \frac{Z}{M'},
\end{aligned} \tag{3.83}$$

According to this perturbation analysis, the errors in range due to the failure of the matched mask constraint are proportional to range, Z, and inversely proportional to the magnitude of the derivative mask, $M'(\cdot)$. In the case of a Gaussian mask, the magnitude of the derivative is inversely proportional to the width of the Gaussian.

Placement of Optical Masks

Throughout our discussion we have claimed that in order to compute range by optical differentiation, the optical mask must be placed *directly* in front of the lens. It is straightforward to see why this must be the case. Consider the projection of the points P_1 and P_2 through an optical mask that has been displaced from the front of the lens (Figure 3.13). There are two things to notice about this situation. First, the image of each point is no longer a scaled and dilated copy of the mask function ($I(x) \neq \frac{1}{\alpha}M(x/\alpha)$), rather, the image is a scaled an dilated copy of only a portion of the entire mask function. Second, the portion of the mask function that is captured by each point depends on the spatial position of the point in the world. For these reasons, when the mask is not placed directly in front of the lens, the basic constraints (Equations (3.5) and (3.18)) will not hold, the required derivative relationship between the images will not hold, and estimates of range will be inaccurate.

In our experiments we avoid this problem by sandwiching the optical mask between a pair of planar-convex lenses (see Section 3.4.2).

Optimal Optical Masks

Up to this point we have discussed various constraints that must be imposed on the functional form of the optical masks. Since there are many functions that satisfy these constraints it is natural to ask whether certain masks are "better" than others. For example, it would be desirable to choose a mask that reduces sensitivity to noise or increases resolution. In this section we explore some of these constraints and show how to design optimal masks based upon them.

The first and simplest constraint is to choose a mask function that maximizes the amount of light throughput. Such a mask will clearly be beneficial in increasing the signal-to-noise ratio (SNR). Denoting the mask function as $y(x)$, we would like to maximize the integral of this function across the diameter of the lens:

$$I_1 \quad = \quad \int_{-a}^{a} dx \; y(x), \tag{3.84}$$

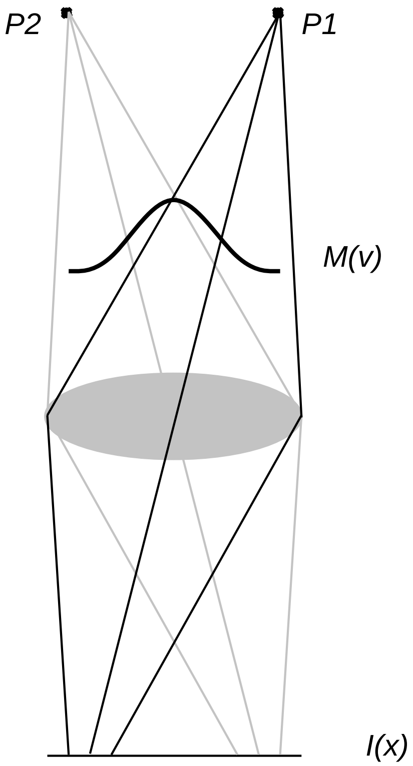


Figure 3.13: Placement of Optical Mask. Illustrated is the projection of two points through an optical mask that has been displaced from the front of the lens. Note that the basic constraint that the image of a point be a scaled and dilated copy of the mask function no longer holds (i.e., $I(x) \neq \frac{1}{\alpha}M(x/\alpha)$). As a result the derivative relationship between the pair of images imaged through the masks $M(v)$ and $M'(v)$ will not obey the necessary derivative relationship (e.g., Equation (3.5)), and range can not be determined from their ratio.

where a is the radius of the lens. The function $y(x)$ which maximizes this constraint can be determined by employing tools from the calculus of variation (see, for example, [Weinstock 52]). In particular, the above integral is of the general form of $I = \int_a^b dx\ F(x, y, y')$, where y' denotes the derivative of y with respect to its single argument x. Stable points for this function can be determined by solving the following Euler-Lagrange differential equation:

$$\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) \quad = \quad 0. \qquad (3.85)$$

Solving the Euler-Lagrange equation for the constraint I_1 gives the solution for the mask function:

$$y(x) = c \quad \text{and} \quad y'(x) = 0, \qquad (3.86)$$

where based on the physical constraints of our mask we choose the constant $c = 1$. This of course makes perfect sense: in order to maximize the light throughput, simply allow all the light to pass. But since the derivative of this mask function is identically zero everywhere, this choice of mask function is completely impractical.

Working towards a more practical solution, we may try to maximize the light throughput for the derivative mask, $y'(x)$.

$$I_2 = \int_{-a}^{a} dx \; y'(x). \tag{3.87}$$

Solving for the maximizing $y(x)$ (Equation (3.85)) gives:

$$y(x) = c_1 x + c_2 \quad \text{and} \quad y'(x) = c_1. \tag{3.88}$$

This solution holds more promise since neither mask function is identically zero. Illustrated in Figure 3.14 are examples of these masks and their non-negative counterparts (see Equation (3.68)). These masks seem quite reasonable, unfortunately, computing spatial derivatives of the image formed under the mask $y(x)$ will be problematic. In particular, the discontinuities of $y(x)$ at its boundaries (i.e., $x = a$ and $x = -a$) will lead to spurious derivative measurements at these points. In order to eliminate these discontinuities, we may impose the boundary conditions that $y(a) = y(-a) = 0$, and solve for the two degrees of freedom in $y(x)$ (i.e., the constants c_1 and c_2). Unfortunately the only solution to these boundary conditions is to have $y(x) = 0 \;, \forall x$ – again, a completely impractical solution

Taking a slightly different approach, we may choose to balance the light throughput for the mask function *and* its derivative. That is, minimize the difference of the means for these masks. This constraint can be expressed as:

$$I_3 = \int_{-a}^{a} dx \; y(x)^2 - y'(x)^2 \tag{3.89}$$

Solving the Euler-Lagrange equation gives the classical differential equation:

$$y(x) + y''(x) = 0. \tag{3.90}$$

The trivial solution to this differential equation is $y(x) = 0$, but of more interest is the solution of the form $y(x) = \cos(x)$ with $y'(x) = -\sin(x)$. Considering these functions in the range $[-\pi, \pi]$ poses a problem for the computation of the non-negative masks. That is,

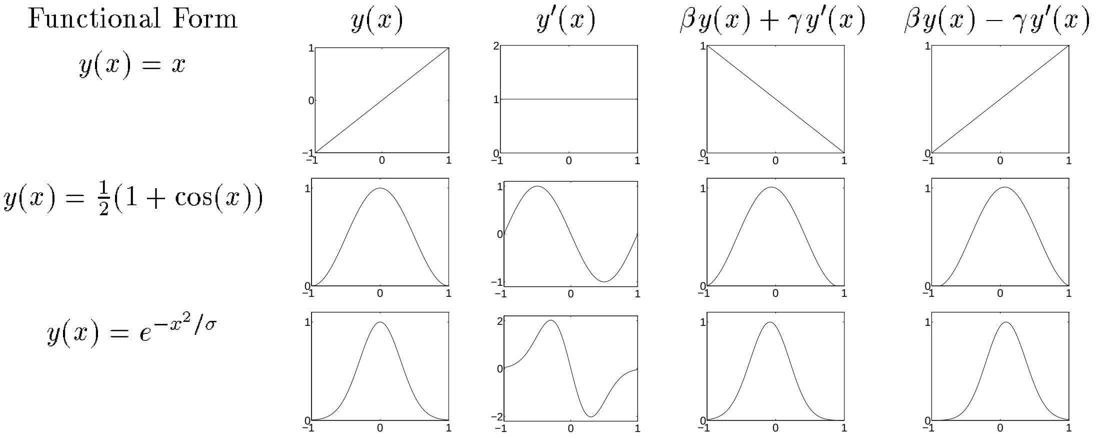


Figure 3.14: Optimal Masks. Shown, from top to bottom, are several optimal masks based on the constraints specified in Equations (3.87), and (3.89), with $a = 1$. For comparison the Gaussian-based mask is also shown (bottom row).

since $\cos(\pi/2) = 0$ and $-\sin(\pi/2) = -1$, it is impossible to generate a pair of non-negative masks from a simple linear combination of these masks functions (i.e., $\beta y(\pi/2) + \gamma y'(\pi/2) < 0,\ \forall\ \beta, \gamma$). Of course, we may consider these functions in the range $(-\pi/2, \pi/2)$, but then the mask $y(x)$ will have a discontinuity at the border leading to spurious derivative measurements in the image formed under this mask. Although not physically realizable, this solution leads to a feasible solution of the form:

$$y(x) = \tfrac{1}{2}(1 + \cos(x)) \quad \text{and} \quad y'(x) = -\tfrac{1}{2}\sin(x). \tag{3.91}$$

Illustrated in Figure 3.14 are examples of these masks and their non-negative counterparts. The mean light throughput of the non-negative mask is 0.5 and should be compared with a mean value of only 0.36 for the Gaussian-based mask (shown in the same figure). In terms of maximizing signal-to-noise ratio (SNR), the raised cosine function is clearly superior (although not optimal with respect to our constraint).

This section has just begun to touch on the issue of optimal mask design, and there are still many possible constraints and optimization techniques to be explored.

3.3.4 Sensitivity to Measurement Noise

Throughout this section we have examined the sensitivity of our range estimator to a variety of assumptions and constraints. Next, we examine the sensitivity to measurement noise, that is, with respect to Figure 3.6, noise in the initial measurements of the images $I_1(x)$ and $I_2(x)$. Recalling from Section 3.3.3 (Equation (3.72)) that the spatial and viewpoint derivatives are determined from these initial measurements as:

$$I_x(x) = \frac{\partial}{\partial x}\left(\frac{I_1(x)+I_2(x)}{\beta_1+\beta_2}\right) \quad \text{and} \quad I_v(x) = \frac{I_1(x)-I_2(x)}{\gamma_1+\gamma_2}, \tag{3.92}$$

we may now inject noise into $I_1(x)$ and $I_2(x)$ and determine its effect on the estimate of range:

$$\begin{aligned} I_x(x) &= \frac{\partial}{\partial x}\left(\frac{(I_1(x)+n_1(x))+(I_2(x)+n_2(x))}{\beta_1+\beta_2}\right) && (3.93)\\ I_v(x) &= \frac{(I_1(x)+n_1(x))-(I_2(x)+n_2(x))}{\gamma_1+\gamma_2}, && (3.94)\end{aligned}$$

and

$$\begin{aligned} Z &= \frac{d_s f}{d_s - f + f\alpha}\\ &= \frac{d_s f}{d_s - f + f\frac{\frac{(I_1(x)+n_1(x))-(I_2(x)+n_2(x))}{\gamma_1+\gamma_2}}{\frac{\partial}{\partial x}\left(\frac{(I_1(x)+n_1(x))+(I_2(x)+n_2(x))}{\beta_1+\beta_2}\right)}}. && (3.95)\end{aligned}$$

Computing the partial derivative of this expression with respect to $I_1(x)$ and evaluating at $n_1(x) = n_2(x) = 0$ (i.e. a noiseless system) gives:

$$\begin{aligned} \frac{\partial Z}{\partial I_1}\Big|_{n_1=n_2=0} &= -\frac{d_s f^2(\beta_1+\beta_2)}{(I_1'+I_2')(d_s-f+f\alpha)^2(\gamma_1+\gamma_2)}\\ &\propto \frac{Z^2(\beta_1+\beta_2)}{(I_1'+I_2')(\gamma_1+\gamma_2)} && (3.96)\end{aligned}$$

The partial derivative with respect to $I_2(x)$ yields a similar expression, differing only in the sign. According to this perturbation analysis, the errors in range due to the presence of noise in the initial measurements are proportional to the square of the range and inversely proportional to the spatial derivatives of the measurements. This is not surprising, we would expect the sensitivity to noise to increase with distance, and the spatial derivative provides a measure of signal strength. In addition, the errors are proportional to the scaling factors $\beta_{1,2}$ and inversely proportional to $\gamma_{1,2}$, which makes sense since they appear as a multiplicative and a divisive factor, respectively, in the computation of range.

In a manner similar to the above formulation the sensitivity of α to measurement errors can also be determined. Consider the case when the estimate is corrupted with noise:

$$Z = \frac{d_s f}{d_s - f + f(\alpha + n)}. \tag{3.97}$$

Computing the partial derivative with respect to α and evaluating at $n = 0$ (i.e., a noiseless estimate) gives:

$$\begin{aligned} \frac{\partial Z}{\partial \alpha}\Big|_{n=0} &= -\frac{d_s f^2}{(d_s - f + f\alpha)^2} \\ &\propto Z^2. \end{aligned} \tag{3.98}$$

That is, the errors are again proportional to the square of the range.

3.3.5 Calibration

In Equation (3.3) we showed that given an estimate of α (the ratio of a viewpoint or aperture size derivative to a spatial derivative) absolute range can be determined:

$$Z = \frac{d_s f}{d_s - f + f\alpha}, \tag{3.99}$$

where d_s is the lens to sensor distance, and f is the focal length. These intrinsic camera parameters need to be calibrated, and clearly errors in this calibration will lead to errors in the estimate of range. Here we determine the sensitivity of the estimator to errors in camera calibration. As in previous sections, a simple perturbation analysis is performed with respect to the calibration parameters, d_s and f, where the effect of adding noise to each of these parameters is considered independently. First, we consider the lens to sensor distance, d_s:

$$Z = \frac{\tilde{d}_s f}{\tilde{d}_s - f + f\alpha}, \tag{3.100}$$

where $\tilde{d}_s = d_s + n$, and n is the calibration noise. Computing the partial derivative with respect to $\tilde{d}_s$ and evaluating at $\tilde{d}_s = d_s$ (i.e., perfect calibration, $n = 0$) gives:

$$\begin{aligned}
\frac{\partial Z}{\partial \tilde{d}_s} &= \frac{f(\tilde{d}_s - f + f\alpha) - f\tilde{d}_s}{(\tilde{d}_s - f + f\alpha)^2} \\
\frac{\partial Z}{\partial \tilde{d}_s}\Big|_{\tilde{d}_s = d_s} &= \frac{f(d_s - f + f\alpha) - fd_s}{(d_s - f + f\alpha)^2} \\
&= \frac{f^2(\alpha - 1)}{(d_s - f + f\alpha)^2} \\
&\propto \frac{Z}{d_s^2}.
\end{aligned} \tag{3.101}$$

Errors in the calibration of the lens to sensor distance leads to errors in the determination of range that are proportional to the range. A similar analysis with respect to the focal length, f, gives:

$$Z = \frac{d_s\tilde{f}}{d_s - \tilde{f} + \tilde{f}\alpha}, \tag{3.102}$$

where $\tilde{f} = f + n$. Computing the partial derivative with respect to $\tilde{f}$ and evaluating at $\tilde{f} = f$ (i.e., perfect calibration, $n = 0$) gives:

$$\begin{aligned}
\frac{\partial Z}{\partial \tilde{f}} &= \frac{d_s(d_s - \tilde{f} + \tilde{f}\alpha) - d_s\tilde{f}(\alpha - 1)}{(d_s - \tilde{f} + \tilde{f}\alpha)^2} \\
\frac{\partial Z}{\partial \tilde{f}}\Big|_{\tilde{f} = f} &= \frac{d_s(d_s - f + f\alpha) - d_sf(\alpha - 1)}{(d_s - f + f\alpha)^2} \\
&= \frac{d_s^2}{(d_s - f + f\alpha)^2} \\
&\propto \frac{Z^2}{f^2}.
\end{aligned} \tag{3.103}$$

Whereas errors in the calibration of the lens to sensor distance are proportional to the range, the errors in the calibration of the focal length are proportional to the square of the range. Intuitively, this makes sense, since when computing range the estimated parameter α is scaled by the focal length but not by the lens to sensor distance.

3.3.6 Stationary World Assumption

Although more of an implementation detail, the final assumption made by our system is that both the camera and the world must remain stationary during the acquisition of the pair of images. If this constraint does not hold, then the images will not be related by the proper derivative relationship, and this will clearly lead to errors in the estimate of range. In this section we analyze the sensitivity of our estimator with respect to inter-frame motion. Consider a point light source that undergoes an arbitrary motion during image acquisition. The images of this point through the optical masks $M(u)$ and $M'(u)$ can be expressed in 2-D as:

$$I(x) = \frac{1}{\alpha} M\left(\frac{x}{\alpha}\right) L \quad \text{and} \quad I_v(x) = \frac{1}{\tilde{\alpha}} M'\left(\frac{\tilde{x}}{\tilde{\alpha}}\right) L, \tag{3.104}$$

where $\tilde{x} = x + \Delta_x$ and $\tilde{\alpha} = \alpha + \Delta_\alpha$ corresponds to the inter-frame motion, and L is the brightness of the point source. The quantity Δ_x corresponds to a translation parallel to the image plane and Δ_α corresponds to a translation in depth along the points principle ray. For simplicity, we will consider the effects of each of these motions independently.

Sensitivity of Stationary World Assumption

First, the case of a simple translation parallel to the sensor plane is considered; the pair of images of interest are expressed as:

$$I_x(x) = \frac{1}{\alpha^2} M'\left(\frac{x}{\alpha}\right) L \quad \text{and} \quad I_v(x) = \frac{1}{\alpha} M'\left(\frac{\tilde{x}}{\alpha}\right) L, \tag{3.105}$$

where $\tilde{x} = x + \Delta_x$ represents the displacement in the image due to the inter-frame motion. For non-zero displacements, $\Delta_x \neq 0$, the ratio of the viewpoint and spatial derivative yields the desired α parameter scaled by a ratio of the derivative mask and a translated copy of the same mask:

$$\frac{I_v(x)}{I_x(x)} = \alpha \frac{M'\left(\frac{\tilde{x}}{\alpha}\right)}{M'\left(\frac{x}{\alpha}\right)}. \tag{3.106}$$

Substituting into Equation (3.3) in order to compute range gives:

$$Z = \frac{d_s f}{d_s - f + f\alpha \frac{M'\left(\frac{\tilde{x}}{\alpha}\right)}{M'\left(\frac{x}{\alpha}\right)}} \tag{3.107}$$

Computing the partial derivative of Z with respect to $\tilde{x}$ and evaluating at $\tilde{x} = x$ (i.e., zero inter-frame displacement, $\Delta_x = 0$) gives:

$$
\begin{aligned}
\frac{\partial Z}{\partial \tilde{x}}\Big|_{\tilde{x}=x} &= -\frac{d_s f^2 \frac{M''\left(\frac{\tilde{x}}{\alpha}\right)}{M'\left(\frac{x}{\alpha}\right)}}{\left(d_s - f + f\alpha \frac{M'\left(\frac{\tilde{x}}{\alpha}\right)}{M'\left(\frac{x}{\alpha}\right)}\right)^2} \\
&= -\frac{d_s f^2 \frac{M''\left(\frac{x}{\alpha}\right)}{M'\left(\frac{x}{\alpha}\right)}}{(d_s - f + f\alpha)^2} \\
&\propto Z^2 \frac{M''\left(\frac{x}{\alpha}\right)}{M'\left(\frac{x}{\alpha}\right)} \\
&\propto Z \frac{I_{xx}(x)}{I_x(x)} \qquad (3.108)
\end{aligned}
$$

Where in the last step the definition of the image $I(x)$ was substituted for the derivatives of the mask function in order to obtain an expression involving measurable quantities. From this perturbation analysis, we see that the errors in range due to an inter-frame translation scale linearly with range.

Next, consider the case of inter-frame motion of the point source along the point's principle ray, a translation in depth with no change in the spatial position on the image plane:

$$
I_x(x) = \frac{1}{\alpha^2} M'\left(\frac{x}{\alpha}\right) L \quad \text{and} \quad I_v(x) = \frac{1}{\tilde{\alpha}} M'\left(\frac{x}{\tilde{\alpha}}\right) L, \qquad (3.109)
$$

where $\tilde{\alpha} = \alpha + \Delta_\alpha$ represents the dilation in the image due to inter-frame motion, and L the constant brightness of the point source. For non-zero dilation, $\Delta_\alpha \neq 0$, the ratio of the viewpoint and spatial derivative does *not* yield the desired α parameter, but rather:

$$
\frac{I_v(x)}{I_x(x)} = \frac{\frac{1}{\tilde{\alpha}} M'\left(\frac{x}{\tilde{\alpha}}\right)}{\frac{1}{\alpha^2} M'\left(\frac{x}{\alpha}\right)}. \qquad (3.110)
$$

Substituting into Equation (3.3) in order to compute range gives:

$$
Z = \frac{d_s f}{d_s - f + f \frac{\frac{1}{\tilde{\alpha}} M'\left(\frac{x}{\tilde{\alpha}}\right)}{\frac{1}{\alpha^2} M'\left(\frac{x}{\alpha}\right)}} \qquad (3.111)
$$

Computing the partial derivative of Z with respect to $\tilde{\alpha}$ and evaluating at $\tilde{\alpha} = \alpha$ (i.e., zero inter-frame displacement, $\Delta_\alpha = 0$) gives:

$$\begin{aligned}
\frac{\partial Z}{\partial \tilde{\alpha}} &= -\frac{d_s f^2 \left(\frac{\frac{1}{\tilde{\alpha}^2} M'(\frac{x}{\tilde{\alpha}}) + \frac{x}{\tilde{\alpha}^3} M''(\frac{x}{\tilde{\alpha}})}{\frac{1}{\alpha^2} M'(\frac{x}{\alpha})} \right)}{\left(d_s - f + f \frac{\frac{1}{\tilde{\alpha}} M'(\frac{x}{\tilde{\alpha}})}{\frac{1}{\alpha^2} M'(\frac{x}{\alpha})} \right)^2} \\
\frac{\partial Z}{\partial \tilde{\alpha}} \Big|_{\tilde{\alpha}=\alpha} &= -\frac{d_s f^2 \left(\frac{\frac{1}{\alpha^2} M'(\frac{x}{\alpha}) + \frac{x}{\alpha^3} M''(\frac{x}{\alpha})}{\frac{1}{\alpha^2} M'(\frac{x}{\alpha})} \right)}{(d_s - f + f\alpha)^2} \\
&\propto Z^2 \left(1 + \frac{x}{\alpha} \frac{M''\left(\frac{x}{\alpha}\right)}{M'\left(\frac{x}{\alpha}\right)} \right) \\
&\propto Z^2 \frac{I_{xx}(x)}{I_x(x)} \qquad (3.112)
\end{aligned}$$

From this perturbation analysis, we see that the errors in range due to an inter-frame translation in depth are proportional to the square of the distance.

Correcting for Non-Stationary Worlds

Perhaps the simplest solution to the problem of inter-frame motion is to *simultaneously* capture the pair of images from the same viewpoint. Such a system could be constructed from a pair of cameras, and beam splitting optics (e.g., [Nayar 95]). However, such a construction would certainly eliminate some of the benefits of a single-lens range sensor (e.g., calibration, physical size, and cost).

Alternatively, we may consider a sequential image capture system that estimates and corrects for the inter-frame motion. For example, the motion between every second frame (i.e., images taken through the same optical mask) can be estimated using techniques outlined in Section 2.3. Then the intermediate image can be warped by the appropriate amount, and range estimated from the arithmetic combination of the motion-compensated images. Of course such an implementation would require three images to compute range (although only two are used directly for estimating range), as well as additional computation time for estimating the inter-frame motion. Although its seems highly feasible, we consider the implementation of such an algorithm beyond the scope of this work.

3.3.7 Resolution

Of general interest in any range estimation technique is the achievable resolution. Since the optical differentiation range estimation relies, at least implicitly, on measuring the amount of blur in the image, we will look at the amount of blur as a function of range and determine a theoretical upper-bound for the achievable resolution in range.

Recall from the thin-lens equation (Equation (1.7)) that the amount of blur is dependent on range, Z, the radius of the lens, R, the focal length, f, the lens to sensor distance, d_s, and the range:

$$r = \frac{R}{\frac{1}{f} - \frac{1}{Z}} \left| \left(\frac{1}{f} - \frac{1}{Z} \right) - d_s \right| . \tag{3.113}$$

Shown in Figure 3.15 is a plot of the amount of blur as a function of range (where the imaging parameters were chosen so that the focal plane was at 1 m, that is, the blur radius is 0 at 1m). Also shown in this figure is the resolution and percent resolution as a function of range, where the resolution is determined by the following ratio:

$$\text{Resolution} = \frac{\lambda}{\partial r / \partial Z}, \tag{3.114}$$

and where λ is the achievable sub-pixel resolution of the discrete spatial derivative operator (chosen to be 1/2 of a pixel). Of course, the resolution also depends on the radius of the lens, the focal length and the lens to sensor distance, each of which were fixed (for the purposes of obtaining the graphs in Figure 3.15 we used values of $R = 50mm$, $f = 32mm$, and $d_s = 32mm$). Figure 3.15 confirms our intuition that we should expect better resolution in our range estimation for nearby surfaces. Of course, this analysis does not take into account any of the other relevant parameters in the estimation of range, but these will be considered in the following sections.

3.3.8 Summary

In the preceding sections we formalized and studied the assumptions and constraints imposed on the differential range estimation formulation. The central results from this analysis have been collected and presented in Figure 3.16. Note that in all cases the sensitivity

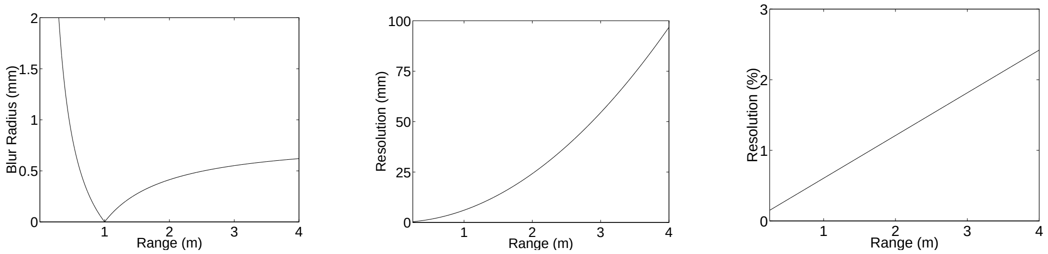


Figure 3.15: Resolution. Illustrated in the left-most panel is the amount of blur as a function of depth (where the imaging parameters were chosen so that the focal plane is at 1 m, i.e., blur radius is 0 at 1m). Illustrated in the middle and right-most panel are theoretical upper-limits on the achievable resolution and percent resolution (Equation (3.114)), for a fixed set of imaging parameters (i.e., lens radius ($R = 50mm$), focal length ($f = 32mm$), and lens to sensor distance ($d_s = 32mm$)).

of the system either scales linearly with range or is proportional to the square of the range. The latter result (which will of course dominate the errors) should not be surprising and amounts to a restatement of the basic laws of triangulation. In particular, recall that in the case of binocular stereo, range is given by the expression, $Z = \frac{d_s b}{\Delta}$ (Section 2.2). Applying the same type of perturbation analysis with respect to the disparity, Δ, gives:

$$\begin{aligned} \frac{\partial Z}{\partial \Delta} &= -\frac{d_s b}{\Delta^2} \\ &\propto \frac{Z^2}{d_s b}. \end{aligned} \tag{3.115}$$

That is, errors in range are proportional to the square of the range, and inversely proportional to the lens to sensor distance and baseline. Not surprisingly, these results are entirely consistent with our findings. As already mentioned, the sensitivity of our system is proportional to the square of the distance, and is also inversely proportional to the lens to sensor distance, d_s, and the width of the optical mask (analogous to the baseline, b).

3.4 Range Estimation

Up to this point we have presented the theory for the optical differential approach to range estimation, studied the various assumptions and constraints, and analyzed the overall sensitivity of the system. It is now time to validate the theory with experiments. The following two sections present results from computer simulations and from experiments with a prototype camera which we have constructed.

Assumption/Constraint	Formulation	Sensitivity
brightness constancy	$L'(\cdot) = 0$	Z^2, $\frac{1}{I_x}$
frontal-parallel	$\int 1/\alpha_p M(\cdot)L(\cdot) = 1/\alpha \int M(\cdot)L(\cdot)$	Z^2
derivative masks	$M_1(x) = M(x) \Rightarrow M_2(x) = \frac{\partial M_1(x)}{\partial x}$	Z
measurement noise	$\tilde{I}(x) = I(x) + n(x) \Rightarrow n(x) = 0$	Z^2, $\frac{1}{I_x}$, $\frac{1}{d_s}$
calibration: f	$Z = \frac{d_s f}{d_s - f + f\alpha}$	Z^2, $\frac{1}{f^2}$
calibration: d_s	$Z = \frac{d_s f}{d_s - f + f\alpha}$	Z $\frac{1}{d_s^2}$
stationary world	$I_1 = \frac{1}{\alpha} M_1\left(\frac{x}{\alpha}\right),\ I_2 = \frac{1}{\alpha} M_2\left(\frac{x+\Delta_x}{\alpha}\right) \Rightarrow \Delta_x = 0$	Z, $\frac{1}{I_x}$
stationary world	$I_1 = \frac{1}{\alpha} M_1\left(\frac{x}{\alpha}\right),\ I_2 = \frac{1}{\alpha+\Delta_\alpha} M_2\left(\frac{x}{\alpha+\Delta_\alpha}\right) \Rightarrow \Delta_\alpha = 0$	Z^2, $\frac{1}{I_x}$

Figure 3.16: Summary of Assumption and Constraints.

3.4.1 Simulations

To verify the basic differential formulation and the analysis presented in the previous sections we have developed a 2-D simulator (written in MATLAB). [10] In this section we describe the construction of the simulator, and present several examples of computed range maps under a variety of scene geometries.

We start with a scene specified in terms of its geometry, reflectance function, and the intensity of one or more light sources. The simulator then determines the appearance of the scene from a specified viewing position under a lens-based imaging system. As illustrated in Figure 3.17, an arbitrary scene is rendered by projecting multiple rays from each pixel at the sensor through an ideal thin-lens and onto the world (Equation (1.8)). The intensity at each pixel is then determined by simply averaging over the intensity of these rays. The image formed through an optical attenuation mask placed directly in front of the lens is determined by using the mask function to take a weighted average of the bundle of rays passing through the lens. In this fashion, the necessary pair of images can be generated and an estimate of range determined from the appropriate differentiation and arithmetic combinations. Note that this simulator differs from the more standard and simpler rendering process which involves only a single ray (i.e., a pinhole camera model).

[10] It was our feeling that a full-blown 3-D simulator would not provide additional insights and that due to the heavy computational cost of rendering through a thin-lens imaging system, a 3-D simulator would be computationally prohibitive.

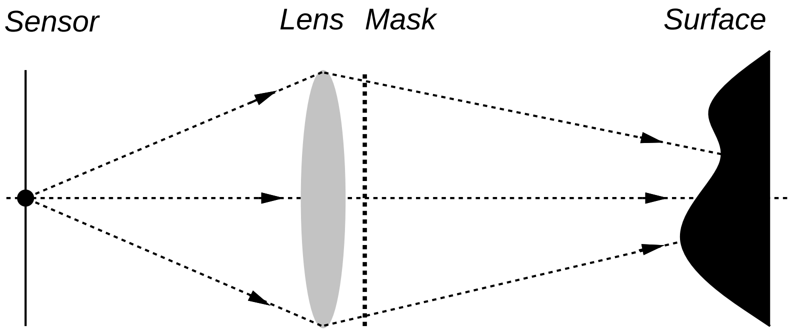


Figure 3.17: 2-D Simulator. This 2-D simulator renders simple geometric scenes through an ideal thin-lens imaging system. For each pixel in the sensor, a bundle of rays are projected through the lens and onto surfaces in the world. The intensity of each pixel is computed by averaging the intensity of this bundle of rays. The image formed under an optical attenuation mask is determined by computed a weighted average (as specified by the mask function) of the bundle of rays. Once the appropriate pair of image are formed, range is computed from the appropriate differentiation and arithmetic combinations.

The simulations presented here explore a variety of scene geometries while holding fixed most other scene and imaging parameters (see Figure 3.18).

In the first set of simulations a simple frontal-parallel surface with a $1/f$ (fractal) random texture pattern is considered. This surface was placed at depths of between 0.25 m and 4 m (with a focal plane near 1 m). Illustrated in the first column of Figure 3.19 are the recovered range maps. Note that as the surface is moved away from the focal plane in either direction the errors increase substantially. This result was predicted by the sensitivity analysis in Section 3.3 and is largely due to the loss of a strong spatial derivative signal due to blurring. These errors can be largely alleviated by blurring and subsampling the initial images before processing, as illustrated in the middle and right columns of the same figure. In this case, the errors are reduced since regions in the image having weak a spatial derivative are minimized. Also illustrated in Figure 3.20 are rms and percent rms plots, averaged over 10 runs with random fractal texture patterns. It is important to mention that subsampling requires the frontal-parallel assumption to hold over a larger spatial region. In these examples, the frontal-parallel assumption holds precisely, so no adverse side-effects of the subsampling are observed. Another point to notice about these examples is that there is a bias in the range estimate towards the focal plane at around 1 m. This is simply due to the fact that as the spatial derivative approaches 0 (i.e., $I_x(x) \rightarrow 0$), α approaches infinity, which is consistent with the value of α at the focal plane. Note

Lens:		
	diameter	50 mm
	focal length	50 mm
	sampling	0.1 mm
Optical Mask:		
	functional form	Gaussian $(G(x) = \frac{1}{\sigma} e^{-x^2/\sigma^2})$
Sensor:		
	lens to sensor distance (d_s)	52.63 mm
	width (w)	10 mm
	pixels (p)	500 pix
	resolution (w/p)	0.02 mm/pix
World:		
	geometry	$f(x) = c_1 + c_2 x + c_3 x^2$
	texture	fractal, $(\mathcal{F}(\omega) = 1/\lvert\omega\rvert)$
	lighting	uniform ambient

Figure 3.18: 2-D Simulator Parameters. Shown here are the variety of tunable parameters in the 2-D simulator described above. Most of these parameters were fixed throughout the simulations presented in this section.

also that there is an asymmetry in the percent rms errors with respect to the focal plane (Figure 3.20). This error has a minimum at the focal plane (1 m) and increases faster for near surfaces than for far surfaces. This is simply due to a similar asymmetry in the amount of blur as a function of distance from the focal plane, as illustrated in Figure 3.15

In the next set of simulations the same planar textured surface, placed at a depth of 1.2 m, is considered but now its orientation relative to the sensor plane was varied between 0° and 60°. Illustrated in the first column of Figure 3.21 are the recovered range maps. As before, the errors are substantially reduced by subsampling the initial measurements before processing. Note also that the errors increase only slightly as the surface is tilted away from frontal-parallel (see also Figure 3.22). Of course it is encouraging to see that the failure of the frontal-parallel assumption does not severely increase the errors. Because of the increased failure of the frontal-parallel assumption, we might expect the errors to increase substantially with subsampling rate, however Figures 3.21 and 3.22 clearly show that this is not the case. This is not entirely surprising given that the range at a given point is computed by averaging over points with larger and smaller range. Of course, if the variation in range were linear, then the over- and under-estimation would "cancel".

However, in our case the estimate of α is inversely proportional to range and therefore does not vary linearly. More specifically, if we approximate the variation in range as locally planar, then the estimate of α_0 at a fixed spatial position, x_0, may be approximated as $\sum_i \frac{1}{mX_i+Z_0}$, where m is the slope of the surface, and Z_0 is the range at x_0. Note that for a frontal-parallel surface (i.e., $m = 0$), $\alpha_0 = \frac{1}{Z_0}$, as desired. But, for non-frontal parallel surfaces (i.e., $m \neq 0$), the estimate of α_0 will be biased depending on the slope of the surface, m, and the spatial extent of integration. However note also that the term mX_i will typically be much smaller than Z_0, and so the average will not deviate substantially from the desired value of Z_0. This observation is substantiated in the rms error plots for increasingly non-frontal-parallel surfaces (Figure 3.22).

In the third set of simulations a concave quadratic surface with the same random fractal texture pattern is considered. The surface is placed at a depth of 1.2 m, with varying curvatures, with the maximum curvature varying from 0° to 60° (see figure caption of Figure 3.23). Illustrated in the first column of Figure 3.23 are the recovered range maps. As in the case of the slanted surfaces, the errors increase only slightly with increasing curvature (see also Figure 3.24). As before, these errors may be reduced by subsampling the initial measurements before processing.

In the fourth set of simulations a variety of occluded surfaces are considered. Again, the surfaces have the same random fractal texture as in the previous simulations. In these simulations, a single frontal-parallel surface was placed at 1.5 m, and a second partially occluding or occluded surface was placed at 1 m, 1.25 m, 1.75 m and 2 m. Illustrated in Figure 3.25 are the recovered range maps with and without subsampling. As before, subsampling reduces the errors, however note also that the sharp discontinuity at the occlusion boundary becomes increasingly broader as the subsampling rate increases. This is not desirable of course, and illustrates one of the drawbacks of subsampling – a loss of spatial resolution. To further illustrate this point consider the step structure in Figure 3.25 and the recovered range maps under varying subsampling rates. Note that as the subsampling rate increases the sharp transition between the step levels becomes increasingly more blurred.

We should also point out that for the frontal-parallel surfaces the errors decrease consistently as the subsampling rate is increased. But this is not the case for the slanted and quadratic surfaces. For these surfaces we see a reduction in error between no sub-sampling and a subsampling rate of 2, but little difference between a subsampling rate of 2 and 4. This effect is most likely due to an increase in the failure of the local frontal-parallel assumption that occurs with increased subsampling rates, which is not a concern in the case of frontal-parallel surfaces.

In the above simulations the feasibility of optical differentiation with respect to viewpoint was tested (Section 3.2.1). In the next set of simulations the optical differentiation with respect to aperture size formulation is examined (Section 3.2.2). In these simulations, the subsampling rate is set at 4, and the same set of frontal-parallel, slanted, and quadratic surfaces are explored. Illustrated in Figure 3.26 are the recovered range maps and shown in Figure 3.27 are the rms plots as a function of absolute range, orientation, and curvature. Note that in all of these examples, the rms errors are larger than in the viewpoint derivative simulations. This is most likely due to the need for a second-order spatial derivative as opposed to only a first-order spatial derivative in the viewpoint derivative formulation. Recall also that in the aperture size derivative formulation, surfaces on either side of the focal plane can not be distinguished (Equation (3.19)). In these simulations, this ambiguity was solved by manually adjusting the sign of the recovered α parameter (i.e., we cheated).

The above set of simulations were designed to illustrate the basic feasibility of range estimation by optical differentiation. In addition, the effect of some of the assumptions (e.g., the frontal-parallel assumption) were verified empirically. In the final set of simulations, the essential sensitivity results of the previous section (as summarized in Figure 3.16) are verified empirically. In particular, the sensitivity of the system to Z^2, $\frac{1}{d_s}$, and $\frac{1}{I_x}$ are explored.

In the first of these simulations we consider a single frontal-parallel surface placed at varying depths between 1 m and 4 m. The lens to sensor distance was held constant, and the focal length was adjusted so that the percent distance between the focal plane and surface remained constant. This ensures that the signal strength remains constant so that the effects of the absolute range can be determined. The initial measurements were not subsampled before processing. Uniform random noise was added to the initial measurements, with a signal-to-noise ratio (SNR) of approximately 20 db (i.e., SNR = $10\log_{10}(\text{std}(\text{signal})/\text{std}(\text{noise}))$, where std$(\cdot)$ denotes standard deviation). Illustrated in

Figure 3.28 are the rms errors in the range estimation averaged over 10 independent trials. Note that the errors are well fit to a second-order polynomial, that is, the errors scale approximately with Z^2.

In the next simulation a single frontal-parallel surface was placed at a fixed depth of 1.2 m (with the focal plane fixed at 1 m), and the lens to sensor distance, d_s, varied between 5 mm and 50 mm. As before, uniform random noise was added to the initial measurements with a SNR of approximately 20 db. Figure 3.28 shows the rms errors in the range estimation averaged over 10 independent trials. Note that the errors scale approximately with the inverse of the lens to sensor distance, $\frac{1}{d_s}$. Clearly a larger lens to sensor distance is desirable, but at the cost of a narrower field of view (i.e., field of view is proportional to $\tan^{-1}(1/d_s)$).

In the last simulation, a single frontal-parallel surface was again placed at a fixed depth of 1.2 m (with the focal plane fixed at 1 m). The initial signal strength was varied between 3 db and 20 db, where the signal strength was taken to be the standard deviation of the intensity image when imaged with no optical mask (converted to db as $10\log_{10}(\text{std})$). Uniform random noise was then added to the initial measurements with a SNR of between 10 db and 20 db. Illustrated in Figure 3.28 are the rms errors in the range estimation, averaged over 10 independent trials. Note that the errors scale approximately with the inverse of the strength of the spatial derivative, $\frac{1}{I_x}$.

3.4.2 Experiments

We have constructed a prototype camera for validating the differential approach to range estimation. As illustrated in Figure 3.29, the camera consists of an optical attenuation mask sandwiched between a pair planar-convex lenses, and placed in front of an off-the-shelf SONY CCD camera. We have employed a liquid crystal spatial light modulator (LC SLM) for use as an optical attenuation mask, also shown in Figure 3.29. In the following three sections, several of the technical details of the operation of the LC SLM are discussed, along with some issues specific to our particular LC SLM. The anxious (or uninterested) reader is encouraged to skip over these sections to the results section.

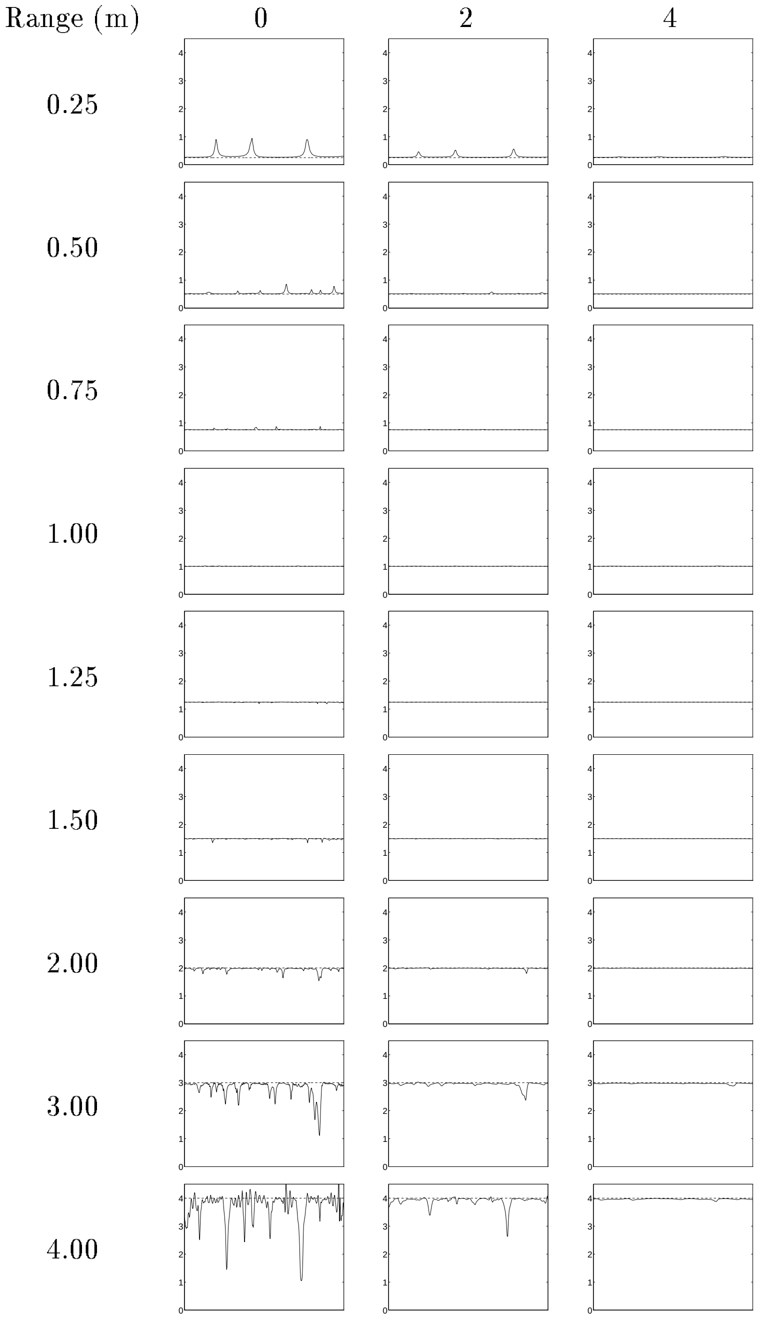


Figure 3.19: Simulation of Optical Viewpoint Differentiation. Illustrated are the estimated range maps (solid curve, with ground truth shown in dashed curve) for a frontal-parallel surface at varying depths, and varying subsampling rates. Note that as the surface is moved away from the focal plane (1 m), the errors increase, but that these errors are reduced by subsampling the initial measurements before processing (see also Figure 3.20).

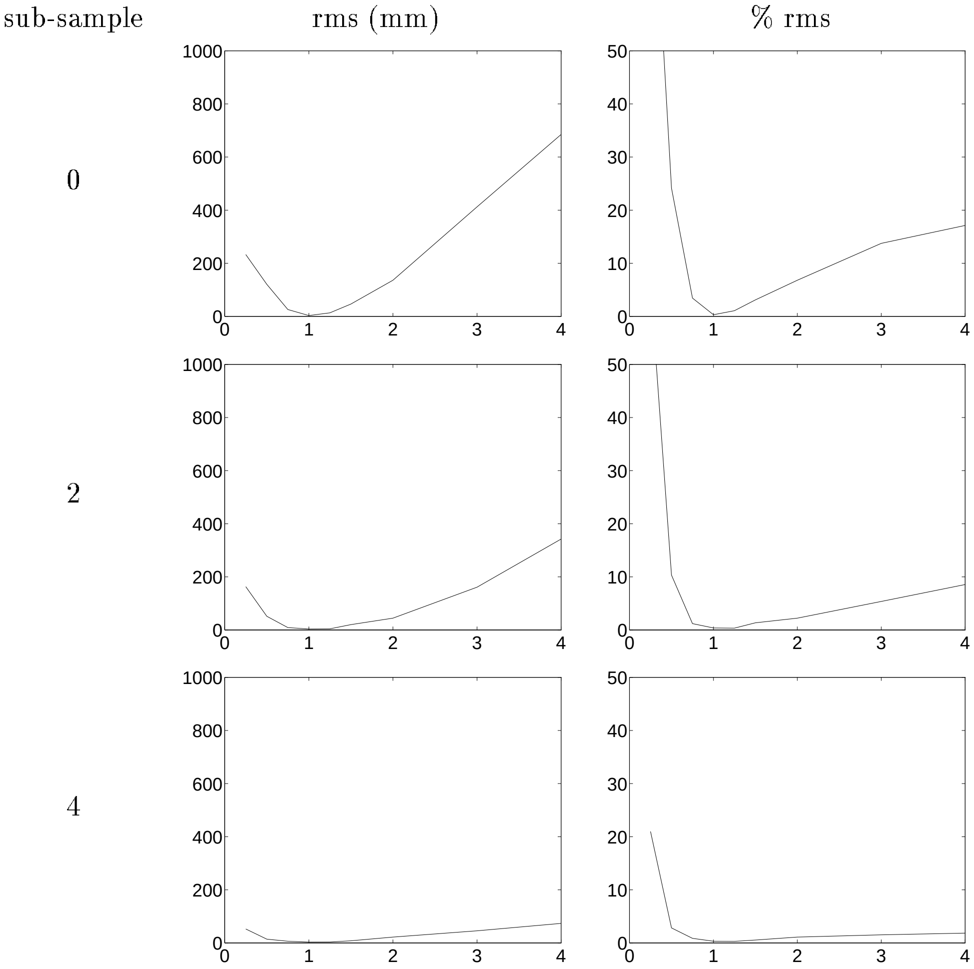


Figure 3.20: Simulation of Optical Viewpoint Differentiation. Illustrated are rms and percent rms errors for frontal-parallel surfaces at varying depths and subsampling rates (see also Figure 3.19). Rms errors were computed by averaging over 10 independent trials with random fractal texture patterns.

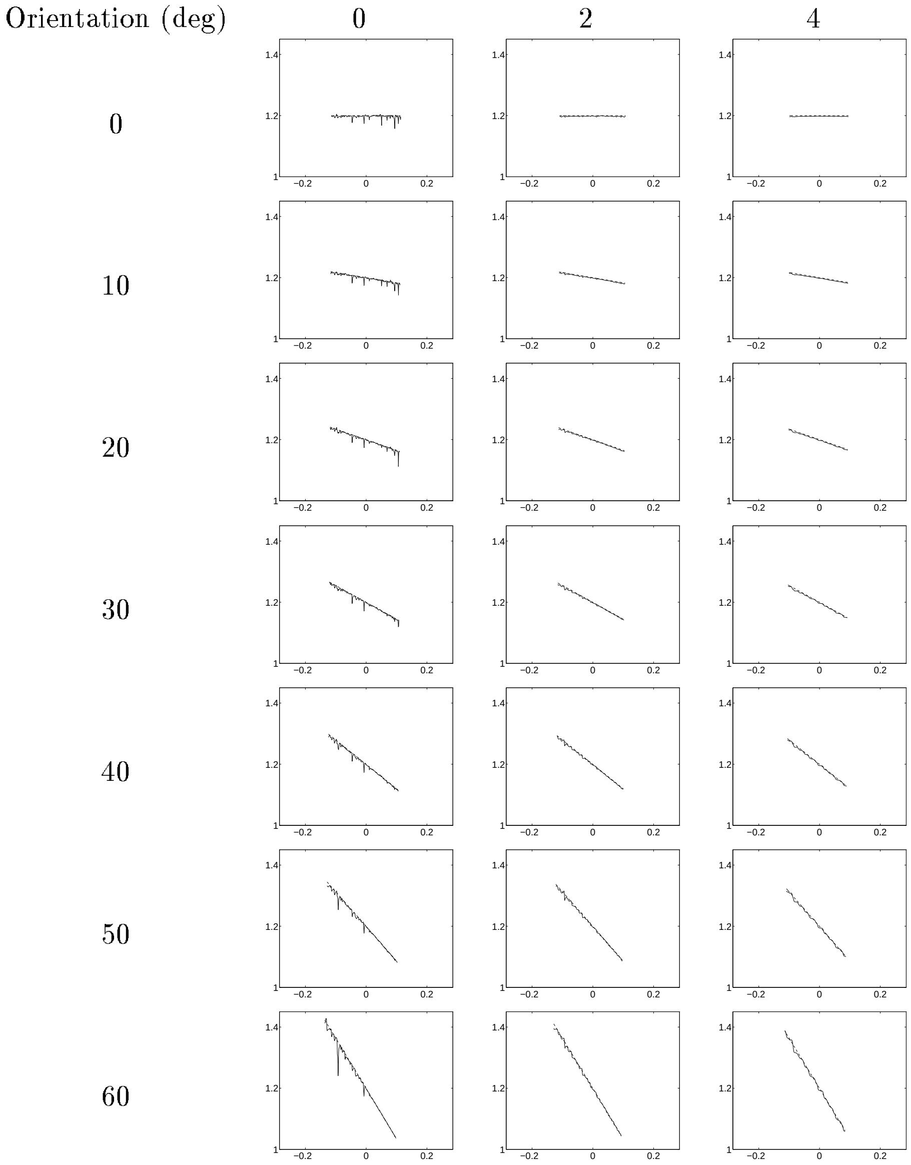


Figure 3.21: Simulation of Optical Viewpoint Differentiation. Illustrated are the estimated range maps (solid curve, with ground truth shown in dashed curve) for planar surfaces, centered at 1.2 m, at varying orientations, and varying subsampling rates. Note that the errors do not increase substantially as the surface becomes more slanted, and that subsampling the initial measurements before processing reduces these errors (see also Figure 3.22).

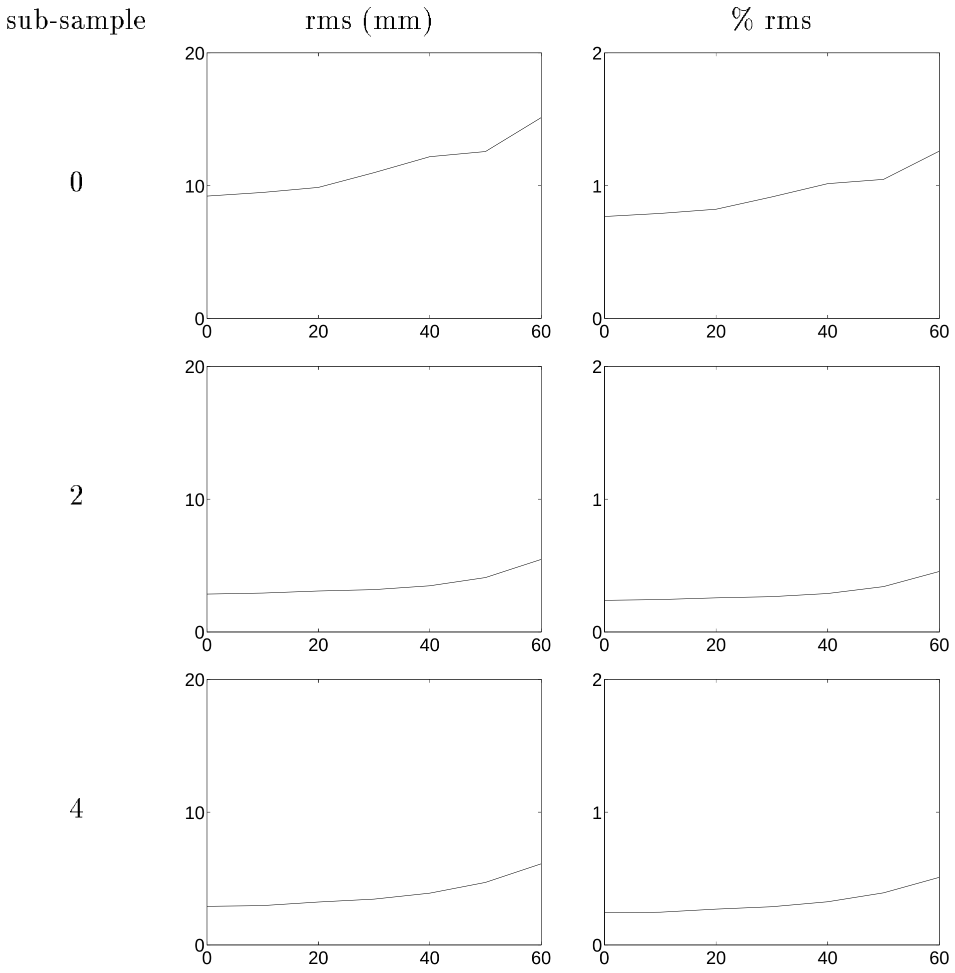


Figure 3.22: Simulation of Optical Viewpoint Differentiation. Illustrated are rms and percent rms errors for planar surfaces at varying orientations and subsampling rates (see also Figure 3.21). Rms errors were computed by averaging over 10 independent trials with random fractal texture patterns.

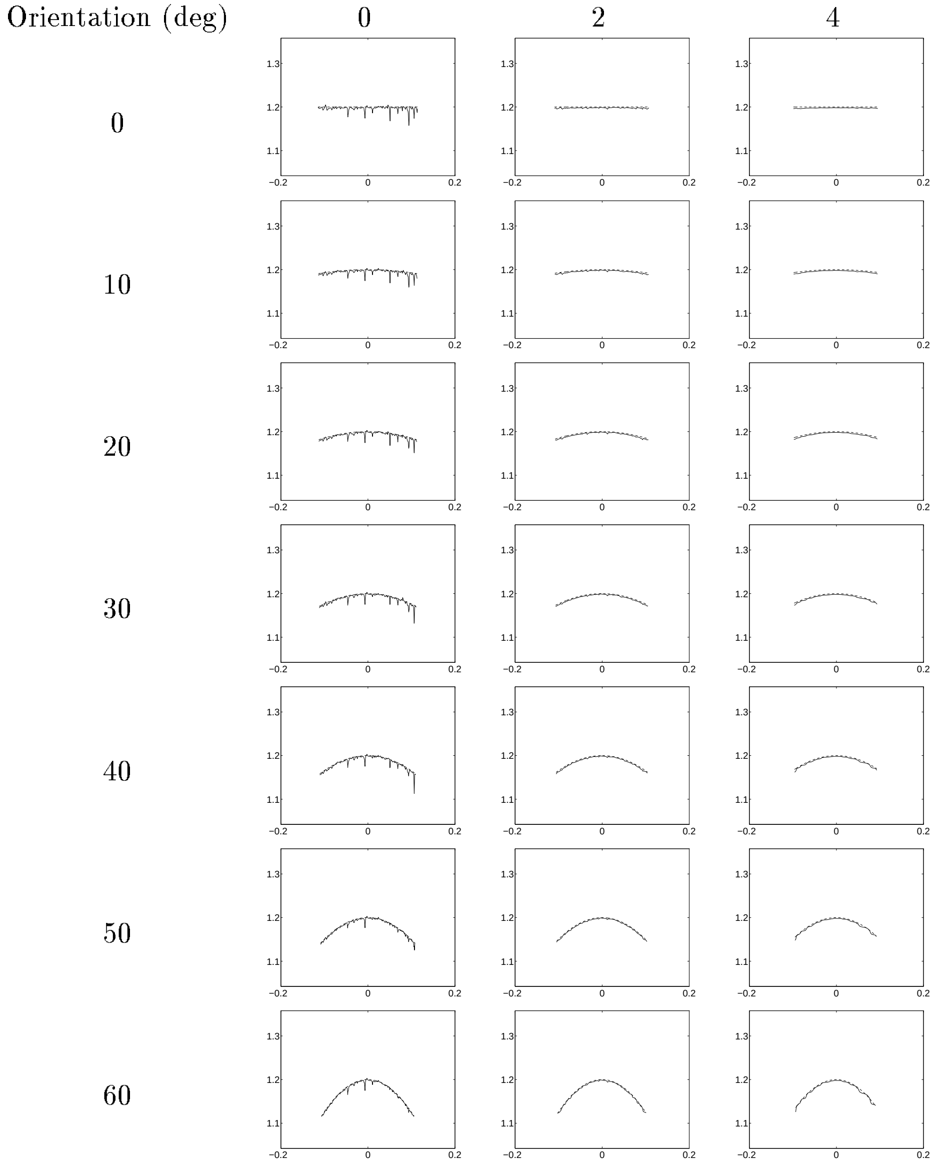


Figure 3.23: Simulation of Optical Viewpoint Differentiation. Illustrated are the estimated range maps (solid curve, with ground truth shown in dashed curve) for quadratic surfaces centered at 1.2 m with varying curvature and varying subsampling rates. The orientation refers to the tangent at the steepest point on the quadratic surface and ranges from 0 to 60 deg. Note that the errors do not increase substantially as the surface becomes more curved, and that subsampling the initial measurements before processing reduces these errors (see also Figure 3.24).

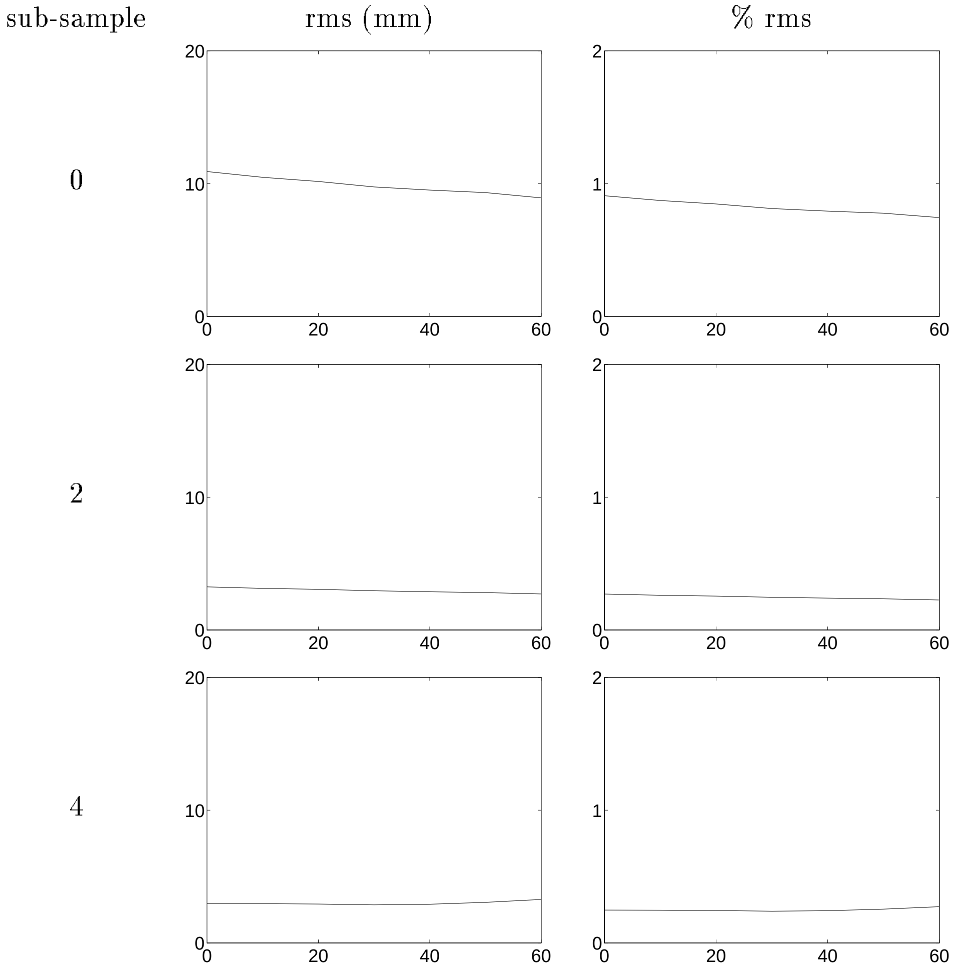


Figure 3.24: Simulation of Optical Viewpoint Differentiation. Illustrated are rms and percent rms errors for quadratic surfaces of varying curvature and subsampling rates (see also Figure 3.23). Rms errors were computed by averaging over 10 independent trials with random fractal texture patterns.

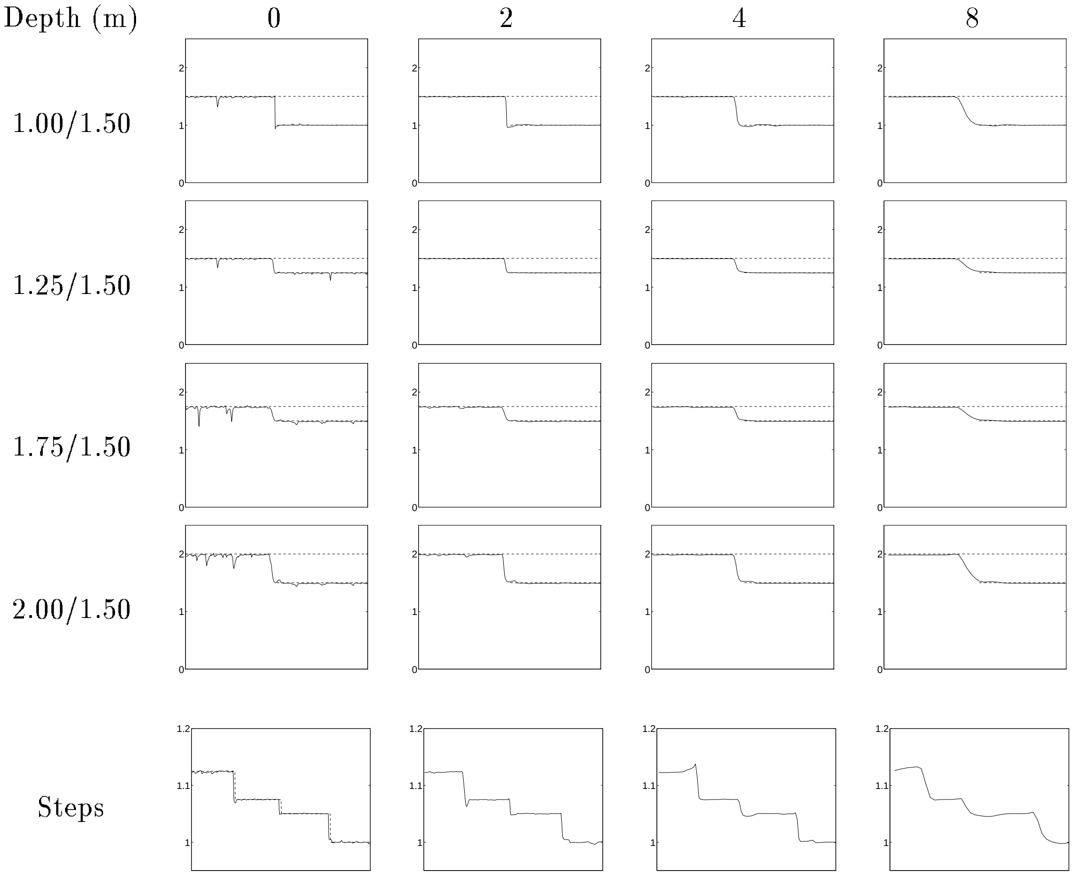


Figure 3.25: Simulation of Optical Viewpoint Differentiation. Illustrated are the estimated range maps (solid curve, with ground truth shown in dashed curve) for occluding frontal-parallel surfaces and varying subsampling rates. Note that as the subsampling rate increases, the errors are reduced, but that the sharp discontinuity between the surfaces becomes increasingly blurred.

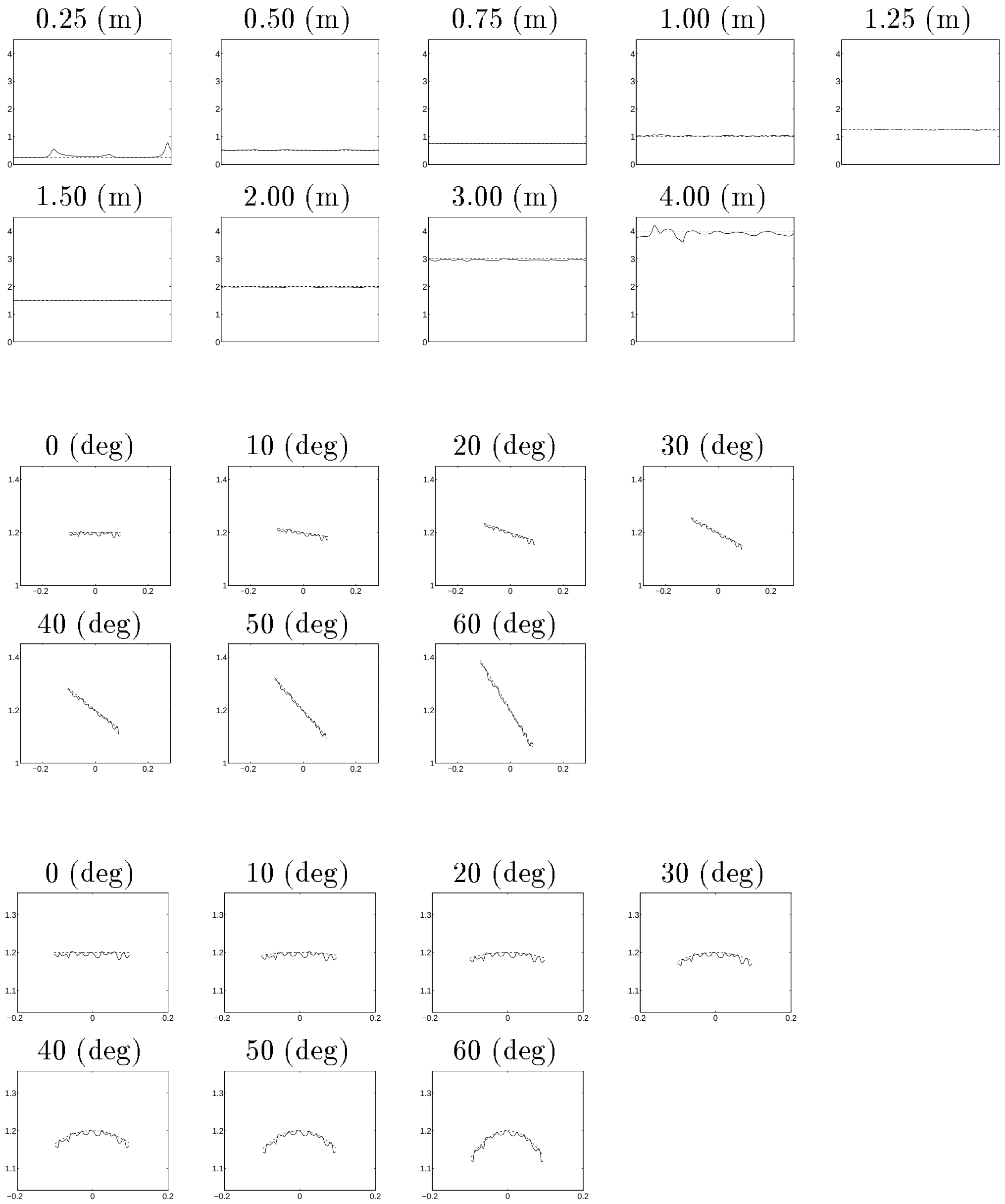


Figure 3.26: Simulation of Optical Aperture Size Differentiation. Illustrated are the estimated range maps (solid curve, with ground truth shown in dashed curve) for frontal-parallel surfaces at varying depths, planar surfaces at varying orientations, and quadratic surfaces of varying curvature. In each case, the subsampling rate was fixed at a factor of 4 (see also Figure 3.27).

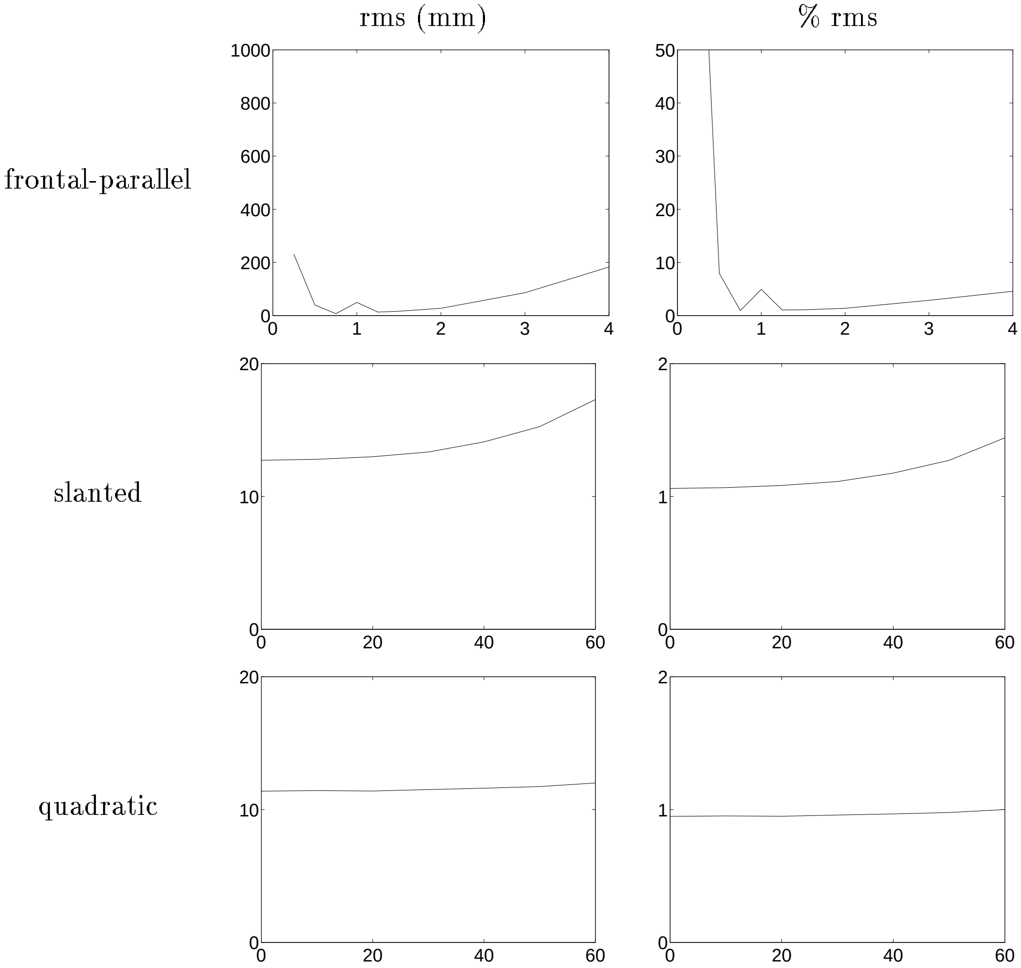


Figure 3.27: Simulation of Optical Aperture Size Differentiation. Illustrated are rms and percent rms errors for frontal-parallel surfaces at varying depths, planar surfaces at varying orientations, and quadratic surfaces of varying curvature (see also Figure 3.26). Rms errors were computed by averaging over 10 independent trials with random fractal texture patterns.

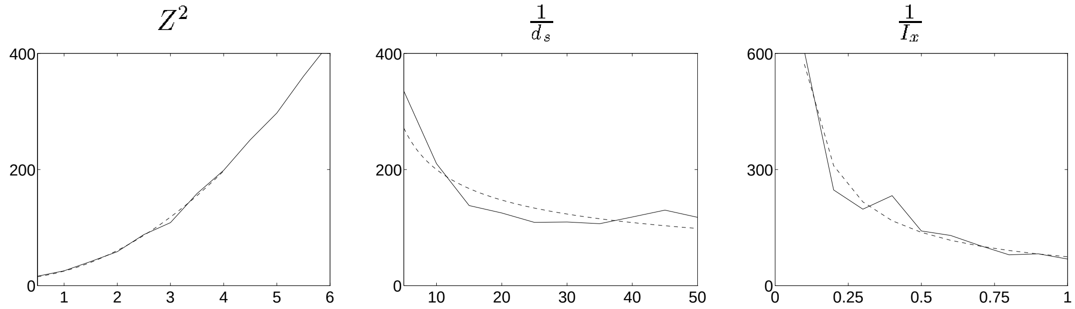


Figure 3.28: Sensitivity of Optical Differentiation. Illustrated are the errors in range estimation as a function of range, Z (in m), lens to sensor distance, d_s (in mm), and spatial derivative, I_x (in normalized SNR). Note that as predicted by our earlier sensitivity analysis, and summarized in Figure 3.16, the errors scale with Z^2, $\frac{1}{d_s}$, and $\frac{1}{I_x}$ (the dotted curve corresponds to a least-squares fit of the data to a second-order polynomial, and to the log of a first-order polynomial, i.e., $1/x$, respectively.

Polarization and Liquid Crystal Displays

In our experimental setup, we have employed a fast-switching, fully programmable liquid crystal display (LCD) for use as an optical attenuation mask. The principles underlying LCDs are discussed here (see [Collings 90] for a nice presentation). Since the polarization of light is central to the operation of these devices, we begin this discussion with a brief review of the electromagnetic wave properties of light and techniques for polarizing light. The latter of which will prove to be the key ingredient in LCDs.

Throughout our discussions light has been considered only in terms of geometric optics [11], and the wave properties of light have been ignored. This is a reasonable simplification when the wavelength of the light is much smaller than anything with which the light interacts, as has been the case so far. But now, in order to discuss the polarization of light, polarizing filters and liquid crystal displays, the electromagnetic wave properties of light need to be considered.

[11] All of the properties of geometric optics may be deduced from three simple rules: (1) light travels in straight lines in homogeneous media; (2) the angle at which light is reflected from a surface is equal to the angle at which it is incident; and (3) when light passes from one medium to another, its path is described by the equation $n_1 \sin(\theta_1) = n_2 \sin(\theta_2)$, where n_1 and n_2 are the indices of refraction of the different media.

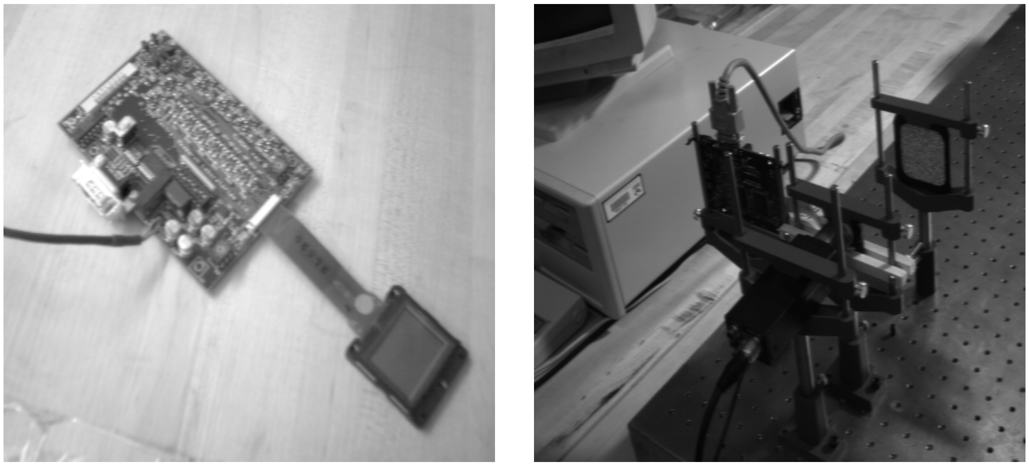


Figure 3.29: Prototype Camera. Illustrated on the left is a fast switching liquid crystal spatial light modulator (LC SLM) employed as an optical attenuation mask (see Section 3.4.2 for more details). Illustrated on the right is our range camera consisting of an off-the-shelf CCD camera and the LC SLM sandwiched between a pair of planar-convex lenses. The target consists of a piece of paper with a random texture pattern.

The electric field [12] of a light wave lies in a plane (i.e., light falls under the general category of plane waves) perpendicular to the direction of propagation. The electric field oscillates sinusoidally in the plane with a specified orientation(s), as illustrated in Figure 3.30. The polarization of light is specified by how the orientation of the electric fields changes as the wave propagates. For example, in an ideally monochromatic light, the electric field oscillates at a single frequency. Since the x- and y-components can oscillate independently, we need to first consider what effect varying the individual magnitudes has on the electric field. As illustrated in Figure 3.31, varying the magnitude of the x- and y-components changes the orientation of the electric field (in the xy-plane). But for a fixed relative magnitude, the orientation is constant. Under these conditions (i.e., when the x and y oscillations are in phase and vary only in magnitude), the light is said to be linearly polarized. On the other hand, if the x and y oscillations have the same magnitude but differ in phase by a multiple of $\frac{\pi}{2}$, then the orientation of the electric field traces out a circle as it propagates. Under these conditions the light is said to be circularly polarized (Figure 3.31). The more general case occurs when the phase difference is not a multiple of $\frac{\pi}{2}$ or when the magnitudes are different (and the phases are offset by any non-zero amount).

[12] Associated with a light wave are both an electric and magnetic field (i.e., light waves fall under the more general category of electromagnetic waves), however we need only consider the electric field explicitly, since the magnetic field is uniquely determined from the electric field and the direction of propagation.

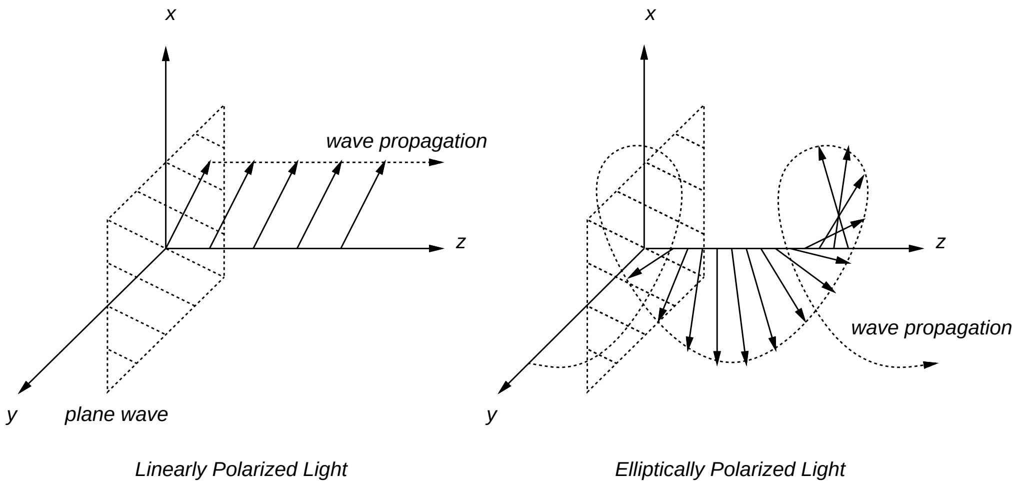


Figure 3.30: Light Wave Propagation. Illustrated is the propagation of the electric wave component of a plane light wave (lying in the xy-plane). This wave modulates sinusoidally in the plane, and the change in orientation of the modulation, as the wave propagates forward, defines the polarization of the light. In particular, if the orientation does not change, then the light is said to be linearly polarized. On the other hand, if the orientation changes as the wave propagates, the light is said to be elliptically polarized, that is, the tip of the vector defining the orientation traces out an ellipse as viewed down the axis of propagation (or in a special case, a circle).

Under these conditions the light is said to be elliptically polarized (Figure 3.31). Lastly, if the polarization changes more rapidly than can be detected, then the light is said to be unpolarized (e.g., sun light).

There are several techniques for polarizing light, these include scattering, birefringence, refraction, and polarizing filters. Here, we are only interested in the latter of these techniques, polarizing filters, the most common being the polaroid filter. Polaroid consists of a thin layer of small crystals of herapathite each aligned with their axis. These crystals absorb light when the oscillations of the electric field are in one direction, and do not absorb light when the oscillations are in the orthogonal direction. Illustrated in Figure 3.32 is an example of linearly polarized light, oriented at an angle θ, passing through a polaroid filter oriented perpendicular to the direction of propagation. The linearly polarized light can be decomposed into a horizontal component, proportional to $\sin(\theta)$, and a vertical component, proportional to $\cos(\theta)$. The amplitude of the light which passes through the polaroid is proportional to $\cos(\theta)$, while the $\sin(\theta)$ is absorbed. For example, if $\theta = 0$, (i.e., the light is oriented with the polaroid filter), then the light passes unattenuated, and if $\theta = \frac{\pi}{2}$,

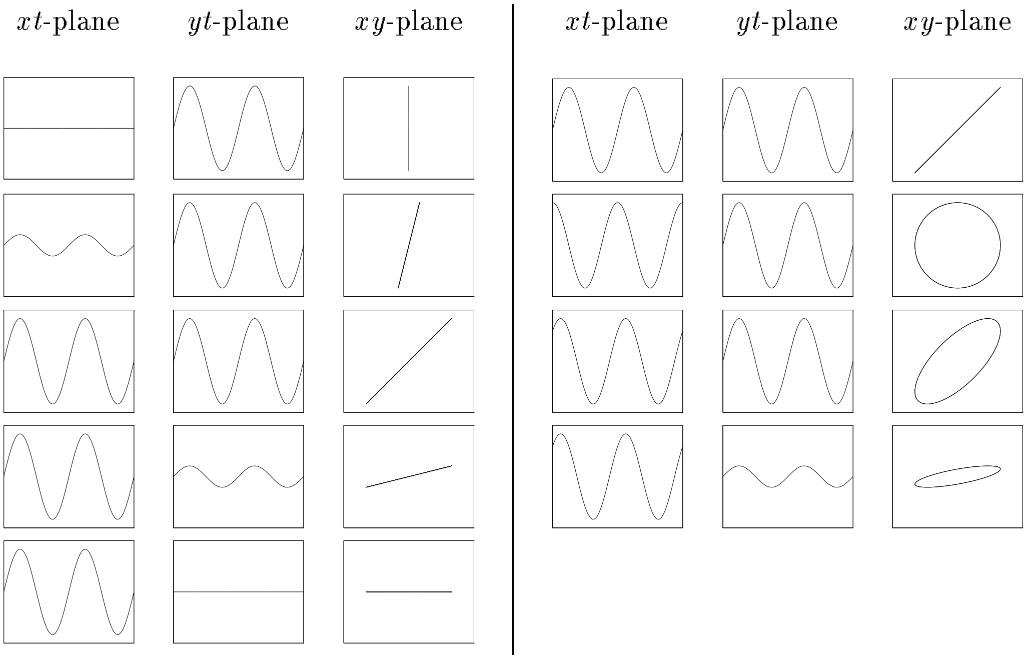


Figure 3.31: Linearly, Circularly and Elliptically Polarized Light. For a single frequency light wave, the x and y modulations may differ in both magnitude and/or phase. Depending on the differences, the resulting light wave is either linearly, circularly, or elliptically polarized. In particular, if the x- and y-components are in phase and differ only in magnitude (left column), then the light is linearly polarized, and the relative magnitudes define the orientation of the light wave. If the x- and y-components are of equal magnitude and differ in phase by $\frac{\pi}{2}$, then the light is circularly polarized (right column, second row). And finally, if the x- and y-components differ in phase by any amount, or in magnitude by any amount (and the phases are offset by any non-zero amount), then the light is elliptically polarized (right column, third and fourth rows), where the relative phases and magnitudes define the orientation of the ellipse.

(i.e., the light is perpendicular to the polaroid filter) then the light is fully absorbed. Note that the amplitude of the light which passes is reduced by a factor of $\cos(\theta)$, and so the intensity is reduced a factor of $\cos^2(\theta)$ (intensity is the square of the amplitude).

With a basic understanding of the wave properties and polarization of light, we are now prepared to discuss liquid crystal displays and there use as spatial light modulators (i.e., optical attenuation masks). Loosely speaking, liquid crystals have some properties characteristic of liquids and some characteristic of solids. Mechanically, liquid crystals resemble liquids (with varying degrees of viscosity), and optically, they resemble crystalline solids. Another way to describe liquid crystals is to consider the molecular difference between liquids, solids and liquid crystals. In liquids, molecules have six degrees of freedom:

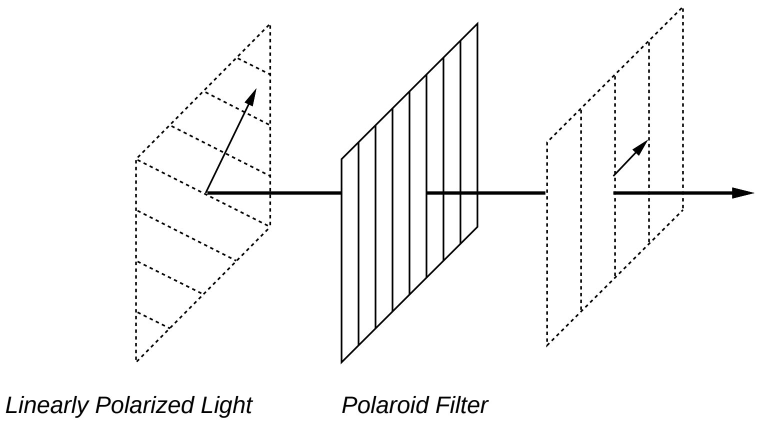


Figure 3.32: Polaroid Filter. Illustrated is a linearly polarized light passing through a polaroid filter. The resulting light is still linearly polarized, but at a different orientation, and lower intensity. The polaroid filter has the property that it absorbs light at some orientations and not at others. As a result, only the vertical component of the light is allowed to pass through the filter (see text for more details).

translation and rotation about all three axes. In a crystalline solid, molecules have zero degrees of freedom: they are fixed in space and cannot rotate. In liquid crystals, molecules have three to four degrees of freedom: translation in two directions and rotation about one axis, or translation in all three directions and rotation about one axis. Of particular interest to us is the fact that liquid crystals have many interesting optical properties including the ability to act as a polarizer.

Illustrated in Figure 3.33 is an example of a liquid crystal display (LCD). In this figure, the light entering the LCD is first linearly polarized. The relative orientation of the crystals through the medium either re-polarizes the light, or passes it unaltered. After passing through the liquid crystals, the light exits through a final polarizing filter that is oriented perpendicular to the first filter. The orientation of the liquid crystal molecules are controlled by a signal voltage applied to the glass substrates on either side of the medium. Depending on this voltage, the light passing through the first polarizing filter may be unaffected, in which case, no light exits through the second polarizing filter. On the other hand, if the light is re-polarized by 90°, then the maximum amount of light exits the LCD. The intensity of the light exiting the LCD may be controlled by varying the degree of re-polarization. Furthermore, the orientation of the liquid crystals may be controlled locally, as well as globally, allowing arbitrary text or images to be displayed.

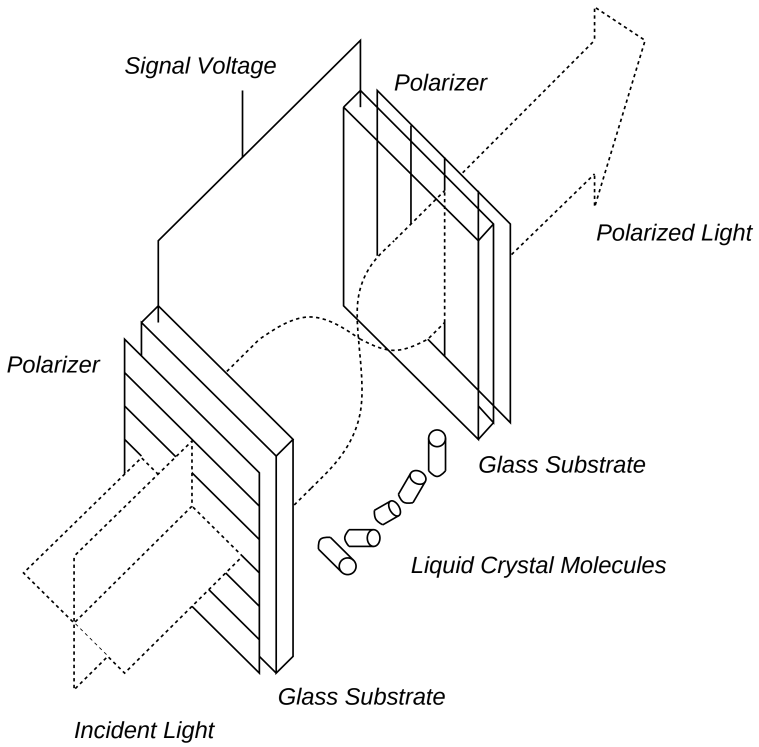


Figure 3.33: Liquid Crystal Display. Light entering the LCD first passes through a polarizing filter so that the light entering the liquid crystal molecules is linearly polarized. Depending on the relative orientations of the molecules (controlled by a signal voltage applied to the glass substrates), the light passes unaffected or is re-polarized. The light then passes through a final polarizing filter, perpendicular to the first filter.

We have purchased from CRL Smectic Technology (Middlesex, UK), a fully programmable, fast-switching, twisted nematic liquid crystal display for use as an optical attenuation mask (also referred to as a spatial light modulator). This device measures 38 mm (W) × 42 mm (H) × 4.3 mm (D), with a display area of 28.48 mm (W) × 20.16 mm (H). The spatial resolution is 640 × 480, with 4 possible grayscale values (some technical issues related to the low grayscale resolution are discussed below). The display is controlled through a PC VGA video interface, supplied by the manufacturer. The LCD refreshes its display at 30 Hz; when synchronized with the frame grabber, the required images taken through the pair of optical masks may be acquired at 15 Hz. [13] The subsequent processing (i.e., convolutions and arithmetic combinations) can be performed in real-time on any of a number of DSP chips or high-end workstations. This LCD display, integrated into the

[13] As in [Nayar 95], a pair of images may be acquired simultaneously (i.e., 30 Hz), by employing an additional camera, some beam splitting optics, and two fixed optical masks.

range sensor, is shown in Figure 3.29. Recently we have found a 3.5 cm, 800 × 480 pixel, 24-bit, fast switching SONY LC SLM that should provide us with a larger and higher resolution optical attenuation mask. We are encouraged that over the past year these devices have dramatically increased in quality and dropped in price (due primarily to the heavy demand produced by the virtual head-mount industry).

Dithering

A second experimental detail involves the resolution of the liquid crystal display described in the previous section. Although higher resolution LCDs exist, our display only allows 2-bit images (i.e., 4 gray levels) to be displayed. In order to minimize the quantization effects, standard dithering techniques were employed (see [Ulichney 88] for an extensive review of dithering and halftoning). Dithering is a process by which a digital image with a finite number of gray levels is made to appear as a continuous-tone image. Since there are numerous dithering algorithms (and even more variants within each algorithm), and few quantitative metrics for measuring their performance (beyond an RMS error metric), we have chosen a standard stochastic error diffusion algorithm based on the Floyd/Steinberg algorithm [Floyd 76].

The Floyd/Steinberg error diffusion dithering algorithm tries to exploit local image structure to reduce the effects of quantization. For simplicity, a 1-bit version of this algorithm is described here, the algorithm extends naturally to an arbitrary number of gray levels. The gray value of a pixel is first thresholded into “black” or “white”, the difference between the new pixel value and the original value is then computed and distributed in a weighted fashion to its neighbors. In [Floyd 76], the authors suggest distributing the error to four neighbors with the following weighting:

$$\tfrac{1}{16} \times \begin{pmatrix} & \bullet & 7 \\ 3 & 5 & 1 \end{pmatrix},$$

where the $\bullet$ represents the quantized pixel, and the position of the weights represent spatial position on a rectangular sampling lattice. Since this algorithm makes only a single pass through the image, the neighbors receiving a portion of the error must consist only of those pixels not already visited (i.e., the algorithm is casual). Note also that since the weights have unit sum, the error is neither amplified nor reduced. There are numerous algorithms which vary the choice of neighbors and the weighting function (see [Ulichney 88] for a review), these multitude of variations will not be considered.

Qualitatively, the error diffusion algorithm reduces the effects of quantization. However this algorithm does introduce correlated artifacts due to the deterministic nature of the algorithm and the scanning order. These problems may be partially alleviated by introducing a stochastic process. [14] A stochastic process may be introduced into the error diffusion algorithm in a variety of places. For example, in [Woo 84] the authors suggest randomizing the weighting function (making sure that the weights always have unit sum), and in [Billotet-Hoffman 83], the authors suggest randomizing the quantization threshold. Along these lines, we have taken the standard error diffusion algorithm and randomized the error (within 0.90% and 1.10%, before distributing it to its neighbors), and alternated the scanning direction (odd lines are scanned from left to right, and even lines are scanned from right to left). Illustrated in Figure 3.34 are dithered images of Richard P. Feynman based on this algorithm. Qualitatively, these images are a considerable improvement over those based on simple thresholding (Figure 3.35), especially for the low-bit images.

Since a quantitative analysis of dithering algorithms would require sophisticated models for measuring image quality, such algorithms are usually only evaluated at a qualitative level. However, for our purposes we can do slightly better. In particular, recall that the dithering procedure will be employed by the LCD to display a pair of optical attenuation masks, and for the purposes of range recovery we require that one mask be the derivative of the other (Section 3.3.3). To this end, the dithering algorithm should be evaluated by how well it preserves the derivative relationship between the pair of masks.

In order to begin to quantify the errors introduced by the dithering algorithm, several pairs of optical masks were dithered using the stochastic error diffusion algorithm described above at varying number of gray levels, and their derivative relationship evaluated in the frequency domain (Figure 3.36). In particular, a pair of non-negative, Gaussian-based optical masks were constructed (Equation (3.68)) and dithered to either 64, 16, 8, 4, or 2 gray levels. The appropriate linear combination of these masks (Equation (3.69)) produces

[14] There are several examples of where randomization has been introduced to reduce visual artifacts. Some of the earliest examples may be found in [Dippe 85], where the authors randomized the sampling grid to reduce the effects of spatial aliasing of undersampled images and in [Allebach 76], where the authors removed moire patterns from computer screens by randomized the centers of dot clusters.

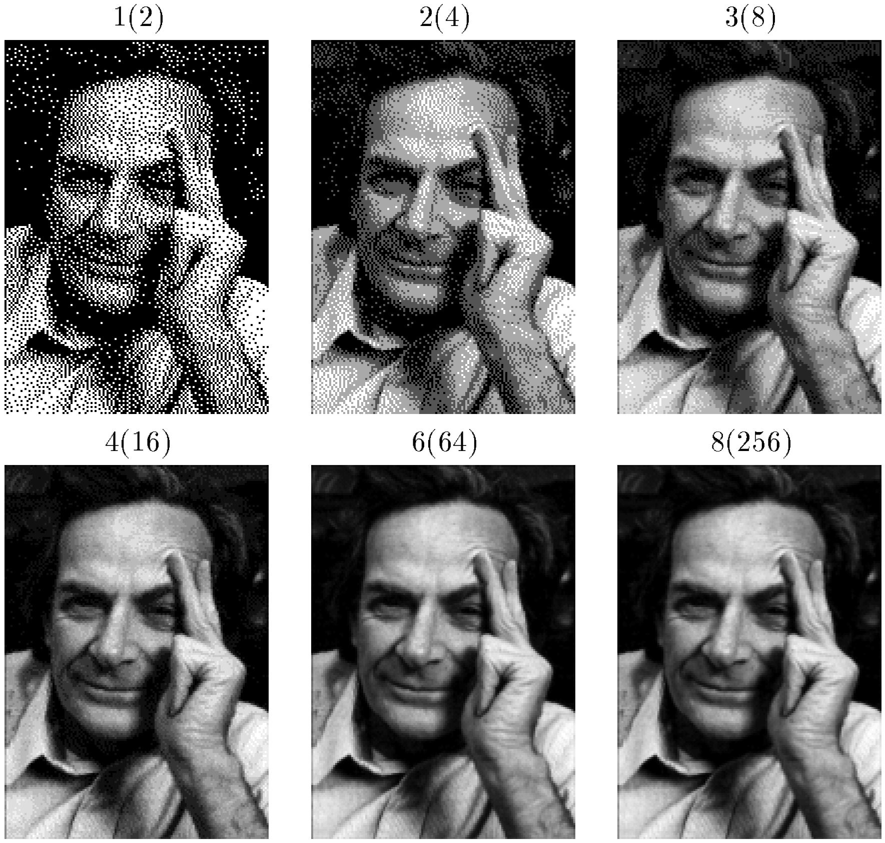


Figure 3.34: Dithering: Stochastic Error Diffusion. Illustrated, from left to right/top to bottom, are images of Richard P. Feynman dithered from 1 bit (2 gray levels) to 6 bits (64 gray levels). The original was an 8-bit image (bottom right). The dithering was accomplished using a stochastic error diffusion process based on the Floyd/Steinberg algorithm [Floyd 76] (see text for more details). These images should be compared with quantization based on simple intensity thresholding illustrated in Figure 3.35.

the required masks, $G(x,y)$ and $G'(x,y)$, where the latter should be the derivative of the former. These dithered masks are shown in the first two columns of Figure 3.36. The derivative relationship of these masks was then evaluated in the frequency domain. That is, if one mask is the derivative of the other, then the Fourier transform of one mask multiplied by a ramp should be equal to the Fourier transform of the other (i.e., $j\omega_x \mathcal{G}(\omega_x, \omega_y) = \mathcal{G}'(\omega_x, \omega_y)$). Illustrated in the third and fourth columns of Figure 3.36 are a ramp times the Fourier transform of the first mask and the Fourier transform of the second mask. For the purpose of display, the Fourier transforms were raised to the power 1/8, and only the central portion displayed ($[-\omega/2 : \omega/2]$). In the last column are 1-D horizontal slices of the

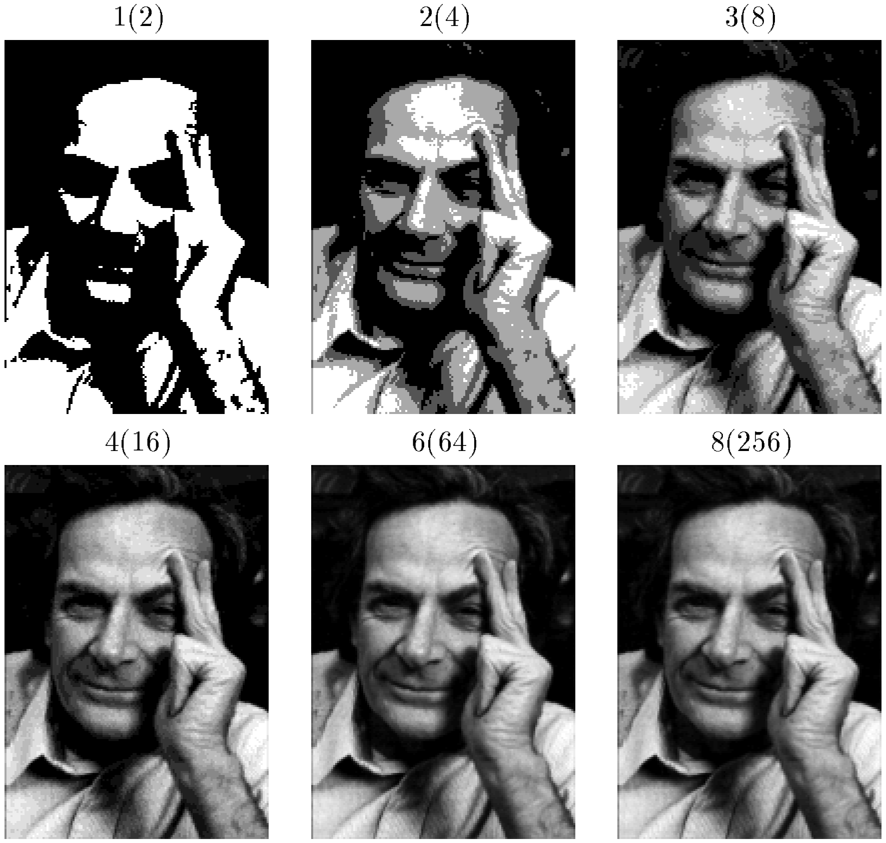


Figure 3.35: Quantization by Thresholding. Illustrated, from left to right/top to bottom, are images of Richard P. Feynman quantized from 1 bit (2 gray levels) to 6 bits (64 gray levels). The original was an 8-bit image (bottom right). The quantization was accomplished using simple thresholding, these images should be compared with those generated using a stochastic error diffusion dithering algorithm (Figure 3.34).

Fourier transforms, where the dashed line is the Fourier transform of $G(x, y)$ multiplied by a ramp, and the solid line is the Fourier transform of the derivative mask, $G'(x, y)$. Note that for the masks dithered at 16 levels and higher, the Fourier transforms are nearly identical, that is, the derivative relationship between the optical masks is preserved. Because our LCD dithers at 4 gray levels we can expect errors in the higher frequencies.

Although it is not possible to exhaustively test each dithering algorithm (and their numerous variants), empirically we have found that the stochastic error diffusion algorithm described above produces rather small artifacts in the low frequencies, and modest errors in the higher frequencies.

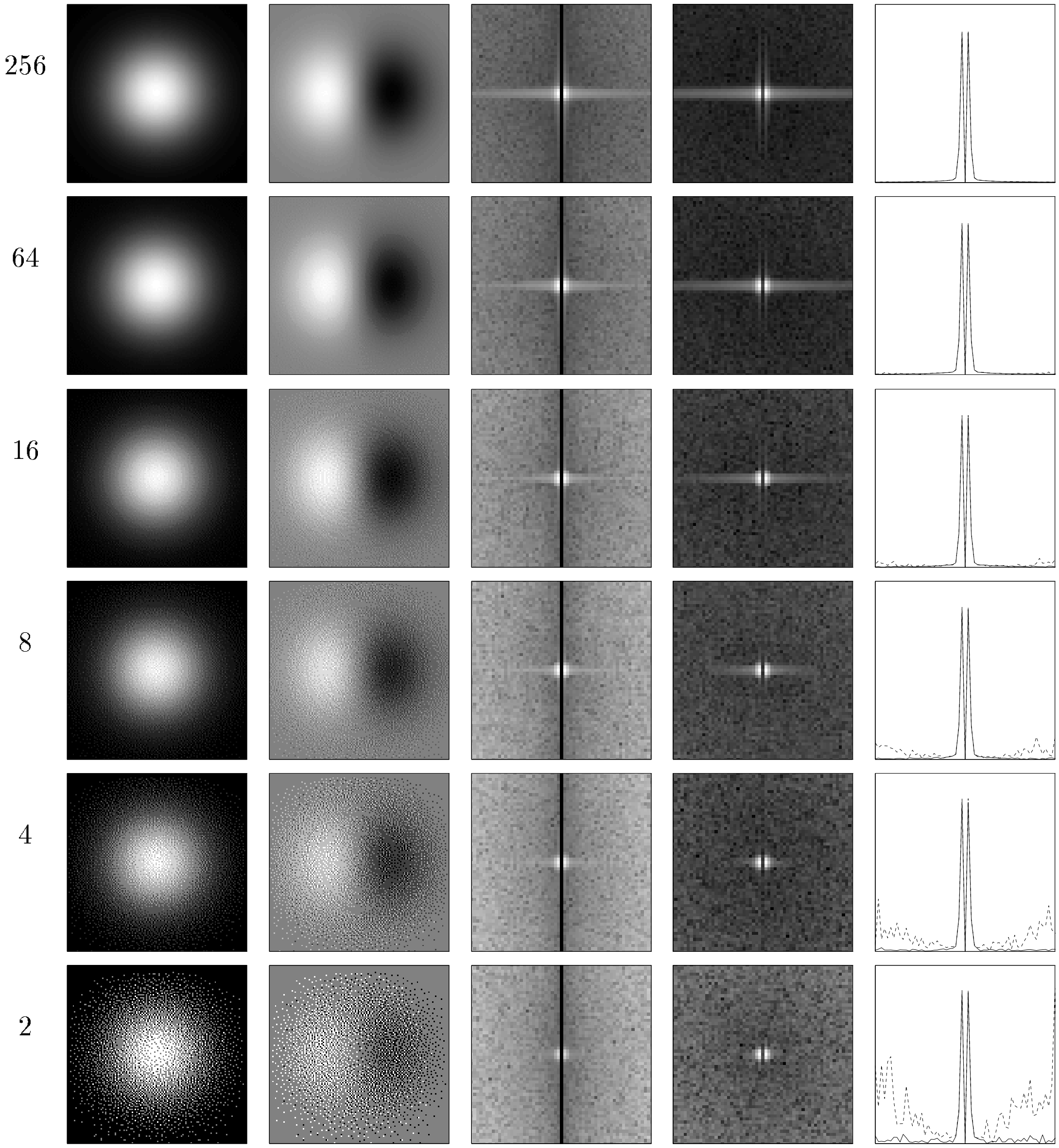


Figure 3.36: Dithering of Derivative Optical Attenuation Masks. Illustrated in the first two columns are a matched pair of dithered (to the specified number of gray levels) Gaussian-based optical attenuation masks, $G(x, y)$ and $G'(x, y)$ (see text for a full description of their construction). Illustrated in the third column is the frequency domain derivative of the first mask, $j\omega_x\mathcal{G}(\omega_x, \omega_y)$, and in the fourth column is the Fourier transform of the derivative mask, $\mathcal{G}'(\omega_x, \omega_y)$. For display purposes, the Fourier transforms were raised to the power 1/8, and only the central portion displayed ($[-\omega/2 : \omega/2]$). If the masks were perfectly matched (i.e., $G'(x, y)$ is the derivative of $G(x, y)$), then the Fourier transforms should be identical. Illustrated in the last column are 1-D horizontal slices of the Fourier transforms, where the dashed line is the Fourier transform of the mask $G(x, y)$ multiplied by a ramp, and the solid line is the Fourier transform of the derivative mask, $G'(x, y)$. Note that for dithering levels of 16 and higher, the filters are almost perfectly matched.

Non-Linearities

And finally, there are at least two possible non-linearities with which we need to contend. The first is the non-linearity in the light transmittance of the optical attenuation masks. A non-linearity at this stage will effect the derivative relationship of the optical masks. As discussed in the previous section, the optical masks were generated using a fully-programmable liquid crystal display (LCD). As illustrated in Figure 3.37, the LCD is highly non-linear. Shown in the first panel of this figure is the light transmittance (measured with a photometer) through one of four uniform gray mask (open circles) – if the LCD were linear, then these measurements would lie along a unit-slope line. The second panel of the same figure illustrates a similar plot for a brighter light source. The similarity of these two plots suggests that the non-linearity in the LCD is constant across this range of lighting conditions. Also shown in these two panels is a second-order polynomial fit to the data (dashed line), referred to as the gamma function, $\gamma(\cdot)$. The non-linearity can now be partially corrected by inverting the gamma function (shown in the third panel of Figure 3.37) and applying the inverse gamma function to the requested gray-value (i.e., when the gray value g is requested, the gray value $\gamma^{-1}(g)$ is displayed). To test how well the inverse gamma function linearizes the LCD a series of 32 uniform gray masks were piped through the inverse gamma function and then dithered (see Section 3.4.2 for more details on the dithering algorithm). Illustrated in the fourth panel of Figure 3.37 is the measured light transmittance across each each of these masks. If the corrected LCD were perfectly linear, then the data (open circles) would lie along a unit-slope line (dashed line). The corrected output is clearly not perfectly linear, but the improvement is substantial. The failure to perfectly linearize the LCD is most likely due to a relatively poor fit to the gamma function from only four data points, and to dithering artifacts.

An alternative approach to linearizing the LCD is to perform the gamma correction *within* the dithering algorithm. More specifically, in the stochastic error diffusion dithering algorithm the error between the quantized pixel value and the desired value is distributed to its neighbors (see Section 3.4.2). In order to correct for the non-linearity in the LCD the diffused error is now given by the difference between the desired value and the measured

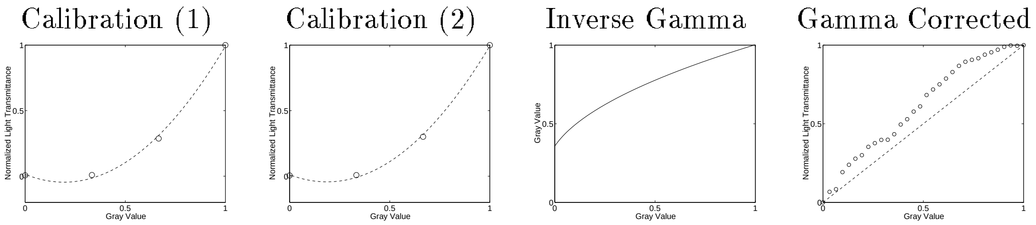


Figure 3.37: Calibration of LCD. Illustrated in this figure is the non-linearity of the LCD (i.e., optical mask). In particular, shown in the two left-most panels is the measured (with a photometer) normalized light transmittance (in cd/m^2) through one of four uniform gray masks, averaged over five trials (open circles). The two panels correspond to the same measurements under different lighting conditions. Also shown is the gamma function, a second-order polynomial fit to the data (dashed line). Illustrated in the third panel is the inverse gamma function used to linearize the LCD. Shown in the final panel is the normalized light transmittance measured through one of 32 uniform gamma-corrected and dithered gray masks, averaged over five trials. If the LCD were perfectly linear, the measurements (open circles) would lie along a unit-slope line (dashed line).

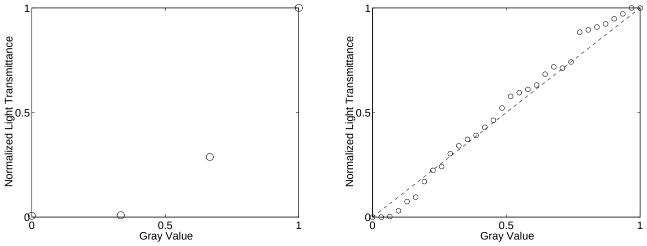


Figure 3.38: Calibration of LCD. Illustrated in this figure is the non-linearity of the LCD (i.e., optical mask). In particular, shown on the left is the measured (with a photometer) normalized light transmittance (in cd/m^2) through one of four uniform gray masks, averaged over five trials. Shown on the right is the normalized light transmittance measured through one of 32 uniform dithered *and* gamma-corrected gray masks, averaged over five trials. That is, the gamma correction is built into the dithering algorithm. If the LCD were perfectly linear, the measurements (open circles) would lie along a unit-slope line (dashed line).

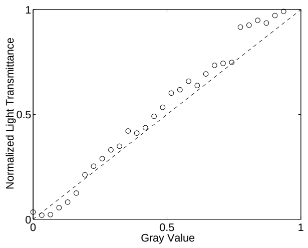


Figure 3.39: Calibration of Imaging Sensor. Illustrated in this figure is the normalized pixel intensity of a point light source as imaged through a series of 32 uniform, dithered optical masks (with gamma correction), averaged over five trials. If both the optical *and* imaging sensor were linear, then these measurements (open circles) would lie along a unit-slope line (dashed line). Although clearly not the case, it would appear that the deviation is due to a remaining non-linearity in the LCD (see Figure 3.38). As a result, we conclude that, for our purposes, the imaging sensor is linear.

light transmittance of the quantized pixel (i.e., the gamma function, γ, evaluated at either 1, 2, 3, or 4). Note that with this approach, we no longer require an explicit gamma function, only the measured light transmittance at the four gray settings. Illustrated in Figure 3.38 is the measured light transmittance through 32 uniform gray masks dithered according to this technique. If the corrected LCD were perfectly linear, then the data (open circles) would lie along a unit-slope line (dashed line). Because this non-linear correction outperforms the previous approach (Figure 3.37) it is employed by our range sensor.

The second possible non-linearity to consider is in the imaging sensor. A non-linearity at this stage will effect the recombination of the images $I(\cdot)$ and $I_{v,a}(\cdot)$ from the measured images $I_1(\cdot)$ and $I_2(\cdot)$ (Equation (3.72)). With the optical masks linearized, it is possible to test the linearity of the imaging sensor by measuring the pixel intensity of the image of a point light source imaged through a series of uniform gray masks. As shown in Figure 3.39, the imaging sensor is close to linear. If *both* the optical masks and imaging sensor were perfectly linear, then the measurements (open circles) would lie along a unit-slope line (dashed line). Clearly this is not the case, but it would appear that the deviations are due to a remaining non-linearity in the LCD. In particular, the data in this figure closely resembles the measured light transmittance of the LCD after gamma correction (Figure 3.38). For our purposes then, it is assumed that the imaging sensor is linear, and a second gamma correction on the initial measurements is not necessary.

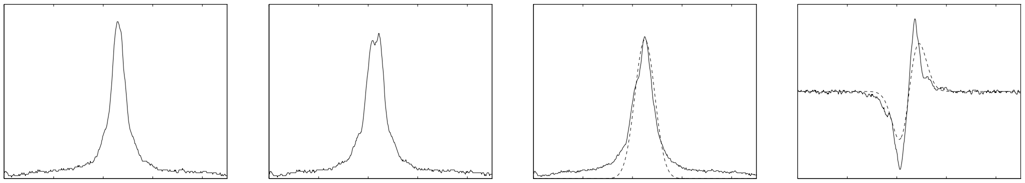


Figure 3.40: Illustrated in the first two panels are 1-D slices of the image of "point light source" taken through a pair of non-negative Gaussian-based masks, I_1 and I_2. Shown in the third and fourth panels are 1-D slices of the images I and I_v determined by the appropriate linear combination of these measurements (see Equation (3.72)). Note that, as expected, these images are reasonably well fit to a Gaussian and its derivative (dashed curve).

Results

In addition to the simulations presented earlier, we have verified the principles of range estimation by optical differentiation with a prototype camera which we have constructed (see Figure 3.29). The first experiment consisted of verifying the basic nature of the optical attenuation masks. In particular, according to our initial observation we expect that the image of a point light source to be a scaled and dilated copy of the mask function. Illustrated in Figure 3.40 is an example of this behavior: shown are 1-D slices of images taken through a pair of non-negative Gaussian-based optical masks, and the appropriate linear combination of these images (Equation (3.72)). Note that these 1-D slices are reasonably well fit to a Gaussian and its derivative (dashed curve). Errors in the fit are undoubtedly due to dithering artifacts in the optical masks, remaining non-linearities in the system, noise, the effects of the camera's intrinsic optical transfer function, and the inability to precisely construct a true point light source.

In the remaining experiments the target consisted of a sheet of paper with a random texture pattern and back illuminated with a desk lamp to help counter the low light transmittance of the LC-SLM optical mask (Section 3.4.2). The initial measurements were subsampled from their original size of 512×512 to 256×256. Spatial derivatives were computed using the 5-tap optimally matched filters described in Section 1.4. In regions with low spatial derivative (approximately 50% of the pixels) range was not computed directly but was instead determined by simple linear interpolation from the neighboring range estimates.

Illustrated in Figures 3.42 and 3.43 are a pair of recovered range maps where the target was a frontal-parallel surface relative to the sensor plane and placed at a distance of 11 and 17 cm from the camera. These figures illustrate the range maps computed using optical viewpoint and aperture size differentiation, respectively. In the case of the viewpoint differentiation, the recovered range maps had a mean of 11.01 and 17.07 cm, with a standard deviation of 0.08 and 0.06 cm, respectively. In the case of the aperture size differentiation, the recovered range maps had a mean of 10.99 and 16.80 cm, with a standard deviation of 0.005 and 0.01 cm, respectively. Statistics for these range maps are given in Figure 3.41. It was somewhat surprising to discover that the aperture size differentiation gave significantly better results than the viewpoint differentiation (in terms of standard deviation). We suspect that this is due to two possible reasons. First, the aperture size mask have a higher mean light throughput; for the Gaussian-based optical masks, the mean light throughput is 0.37, as compared to a mean of 0.20 for the viewpoint masks. The increased light throughput translates to a higher signal-to-noise ratio. Second, since the depth of field of our camera is narrow, the inherent sensitivity to changes in aperture size are likely to be larger than with respect to viewpoint. Although, the aperture size differentiation affords fewer errors in this example, it still suffers from a sign ambiguity (i.e., surfaces equally spaced on either side of the focal plane are indistinguishable).

Illustrated in Figure 3.44 is a recovered range map for a planer surface oriented approximately 30 degrees relative to the sensor plane, with the center of the plane 14 cm from the camera, and a pair of occluding surfaces placed at 11 and 17 cm. The recovered range maps in this figure were determined using the optical viewpoint differentiation formulation. Qualitatively, these range maps look quite reasonable.

3.5 Related Work

The idea of optical differentiation and its application to range estimation is novel to this work, however the concept of single-lens range imaging is not. Below are brief descriptions of several such systems, and shown in Figure 3.45 is a quantitative comparison of these techniques, several other standard range estimation techniques (e.g., range from motion,

Derivative	Range	Mean	Std	Min	Max
View.	11	11.01	0.08	9.42	12.91
View.	17	17.07	0.06	15.59	18.87
Ap. Size	11	10.99	0.005	10.85	11.20
Ap. Size	17	16.80	0.01	16.12	17.23

Figure 3.41: Range Estimation Statistics. Illustrated are the mean, standard deviation, minimum and maximum range values for the frontal parallel surfaces illustrated in Figures 3.42 and 3.43. All measurements are given in cm.

stereo, focus, and defocus) and our optical differentiation approach.

The use of optical attenuation masks for range estimation has been considered in the work of Dowski and Cathey [Dowski 94] and Jones and Lamb [Jones 93]. The former employs a sinusoidal aperture mask and computes range by searching for zero-crossings in the local frequency spectra. The latter system employs an aperture mask consisting of a pair of spatially offset pinholes. Imaging through such a mask produces a superimposed pair of images from different viewing positions. Range is determined using standard stereo matching or visual echo techniques. The masks used in both these systems are not based on differential operations. Furthermore, these systems operate on a single image and must therefore rely on assumptions regarding the spectral content of the scene.

Adelson and Wang [Adelson 91, Adelson 92] describe a clever single-lens, single-image range camera, termed the plenoptic camera [15]. The authors placed a lenticular array (a sheet of "miniature lenses") over the sensor, allowing the camera to capture images from several viewpoints. More specifically, each group of 5×5 pixels (termed macropixels) captured an image from a different viewpoint. From only a single image, a viewpoint derivative can be computed across the macropixels. By computing a spatial derivative within the macropixels, range is determined by the familiar ratio of these derivative measurements. Note however, that the viewpoint derivative is still being computed from a discrete set of viewpoints (i.e., across 5 macropixels). The authors noted several technical difficulties with this approach, most notably, aliasing and the alignment of the lenticular array with the CCD sensor. This approach is based on the same underlying principles as our own

[15] The term plenoptic is derived from *plenus*, complete or full, and *optic*, view.

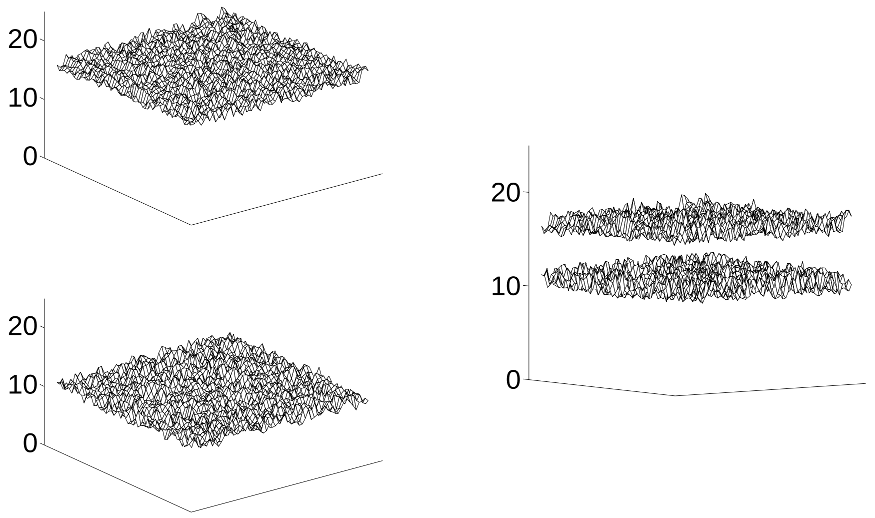


Figure 3.42: Results. Illustrated are the recovered range maps using optical viewpoint differentiation for a pair of frontal-parallel surfaces at a distance of 11 and 17 cm from the camera. The computed range maps have a mean of 11.01 and 17.07 cm with a standard deviation of 0.08 and 0.06 cm, respectively.

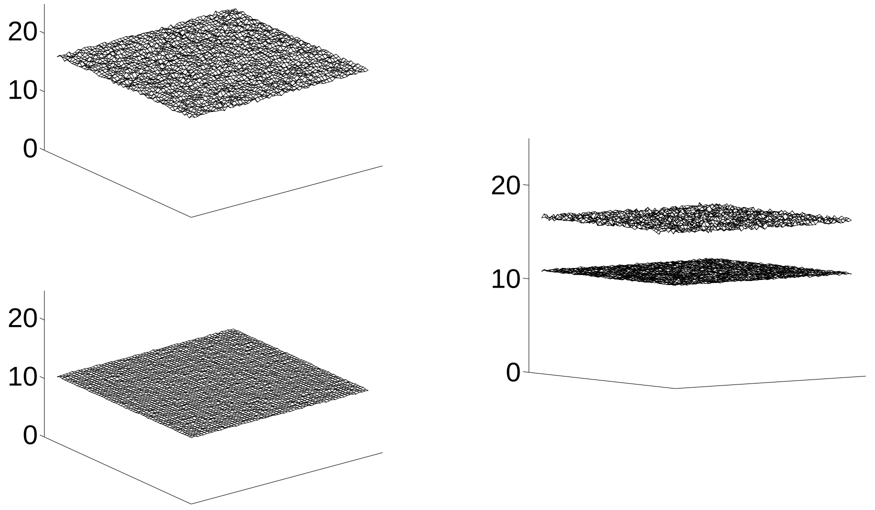


Figure 3.43: Results. Illustrated are the recovered range maps computed using optical aperture size differentiation for a pair of frontal-parallel surfaces at a distance of 11 and 17 cm from the camera. The computed range maps have a mean of 10.99 and 16.80 cm with a standard deviation of 0.005 and 0.01 cm, respectively.

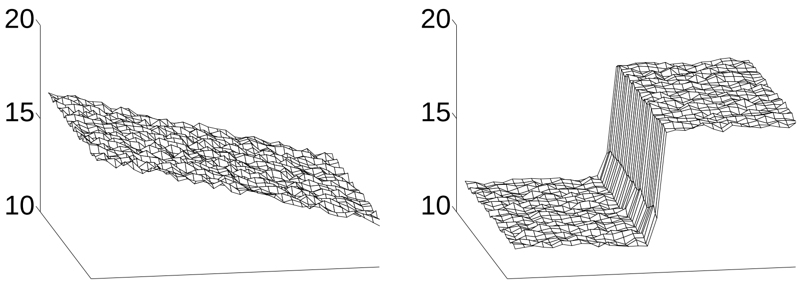


Figure 3.44: Results. Illustrated on the left is the recovered range map computed using optical viewpoint differentiation for a slanted surface oriented approximately 30 degrees relative to the sensor plane with the center of the plane at a depth of 14 cm. Illustrated on the right is the recovered range map for a pair of occluding surfaces at a depth of 11 and 17 cm.

(and inspired many aspects of our approach), but is entirely different in implementation and, most importantly, the method of calculating viewpoint derivatives.

A series of single-lens stereo systems have also been developed [Teoh 84, Nishimoto 87, Goshtasby 93]. In the first of these systems, a stereo pair of images is generated by two fixed mirrors, at a 45° angle with the camera's optical axis and a rotating mirror made parallel to each of the fixed mirrors. In the second system, a rotating glass plate placed in front of the main lens, shifts the optical axis slightly, simulating two cameras with parallel axis. The last system places two angled mirrors in front of a camera producing an image where the left and right half of the image correspond to the view from a pair of verged virtual cameras. In each case, range is calculated using standard stereo matching algorithms. The benefit of these approaches is that they eliminate the need for extrinsic camera calibration (i.e., determination of the relative positions of two or more cameras), but they do require slightly more complicated intrinsic calibration of the camera optics.

We have compared the performance of our technique for range estimation with the classical approaches outlined in Chapter 2 as well as the more exotic approaches described above. Figure 3.45 shows a comparison of the mean errors for various range estimation techniques. This list is not intended to be exhaustive, and only a few examples from each technique are given. The choice of techniques was based primarily on the availability of quantitative error measurements, but we also have tried to give an appropriate sampling of early and current work in each area. Whenever possible, the following information is provided: (1) mean range of surfaces in the world, (2) camera focal length, (3) frame rate, (4) density of range image, (5) whether structured illumination was employed, (6) mean error as a percent of range, and (7) normalized error. The normalized error compensates for differences in range and focal length (see figure caption for more details). However, this is only intended as a first approximation, and ignores many other factors. In addition, we give the benefit of the doubt to the range from motion algorithms and assume that they are able to exactly recover the camera motion, that is, the errors are given with respect to the velocity measurement (see Section 2.3). This comparison shows that our approach is competitive with other techniques, where we benefit over others in terms of physical size, cost, calibration, and computation.

3.6 Summary

In this chapter, we have presented the theory, analysis, and implementation of a novel technique for estimating range from a single stationary camera. The computation of range is determined from a pair of images taken through one of two optical attenuation masks. The subsequent processing of these images is simple and analytic involving only a few 1D convolutions and arithmetic operations. We have shown that the errors and sensitivity of this approach are proportional to the square of the range, which is in line with most other range estimation techniques.

The simplicity of this technique has some clear advantages. In particular, the use of a single stationary camera reduces the cost, size and calibration of the overall system, and the simple and fast computations required to estimate range makes this technique amenable to a real-time implementation. In comparison to classical stereo approaches, our approach completely avoids the difficult and computationally demanding "correspondence" problem. In addition, with only a single stationary camera, we avoid the need for extrinsic camera calibration. Of course there are also some clear disadvantages as well: most notably, the construction of a non-standard imaging system, the limited resolution due to the width of the lens, and the requirement of two sequentially acquired images.

It is unlikely that this technique will supplant the multitude of existing range estimation techniques, rather, we expect that this approach will prove to be well suited for certain domains, and poorly suited for others. We are hopeful that with time and advancements in the liquid crystal technology employed by the optical masks, the construction, cost, and performance of our system will improve.

			Focal	Frame		Structured		Normalized
Classification	Reference	Range	Length	Rate	Density	Illum.	Error	Error
		(mm)	(mm)	(frame/s)	(pixels)		(%range)	
stereo	[Cochran 92]	-	-	0.002	128×128	No	0.76	-
stereo	[Kanade 95]	750	50	30	256×240	No	0.8	1.42
diff. stereo	[Gokstorp 95]	-	-	-	144×144	No	3.5	-
motion	[Lucas 81]	-	-	-	128×128 (8.8%)	No	3.55	-
motion	[Simoncelli 93]	-	-	-	128×128	No	3.81	-
motion	[Fleet 90]	-	-	-	128×128 (34.1%)	No	4.17	-
motion	[Horn 81]	-	-	-	128×128 (33.3%)	No	7.12	-
motion	[Uras 88]	-	-	-	128×128	No	10.44	-
motion	[Horn 81]	-	-	-	128×128	No	16.23	-
motion	[Lucas 81]	-	-	-	128×128	No	17.93	-
focus	[Xiong 93]	1000	130	-	-	No	0.5	1.30
focus	[Subbarao 95]	600	35	-	256×256	No	2.07	4.02
focus	[Subbarao 95]	5000	35	-	256×256	No	17.27	0.48
defocus	[Nayar 95]	300	12.5	30	512×480	Yes	0.25	0.69
defocus	[Pentland 94]	45	-	4	64×64	Yes	0.5	-
defocus	[Xiong 93]	1000	130	-	-	No	1.3	3.38
defocus	[Pentland 94]	2000	-	4	64×64	No	2.5	-
sinusoid mask	[Dowski 94]	1400	50	-	256×256 (<1%)	No	1.4	0.71
pinhole mask	[Jones 93]	-	16	-	-	No	-	-
plenoptic	[Adelson 92]	-	35	-	100×100	No	-	-
active	[Kramer 93]	310	45	50	128×128	Yes	1.2	11.23
optical diff. (sim)	-	4000	50	-	1×128	No	0.56	0.04
optical diff. (sim)	-	2000	50	-	1×128	No	0.19	0.05
optical diff. (sim)	-	500	50	-	1×128	No	0.36	1.44
optical diff. (view)	-	110	-	-	256×256 (50%)	No	<<1%	0.75
optical diff. (view)	-	170	-	-	256×256 (50%)	No	<<1%	3.41
optical diff. (ap)	-	110	-	-	256×256 (50%)	No	<< 1%	0.75
optical diff. (ap)	-	170	-	-	256×256 (50%)	No	< 1%	9.7

Figure 3.45: Comparison of Range Recovery Techniques. *Classification:* general categorization of technique (see Chapter 2 and Section 3.5 for descriptions). *Reference:* citation from which results are taken. *Range:* mean distance to surface(s) in the world (mm). *Focal Length:* camera focal length (mm). *Frame Rate:* number of range images computed per second (frames/second); note that this rate is *not* normalized over the size of the range images. *Density:* size of computed range image (pixels); the parenthesesized number indicates the percentage of pixels for which an estimate was available, no specified value indicates 100%. *Structured Illumination:* indicates whether the technique employed structured illumination. *Error:* mean percent error with respect to range (e.g. 1% error at 1000mm, reflects a mean error of 10mm in the recovered range image). *Normalized Error:* Error normalized to a 1000mm viewing distance with a 50mm focal length $\left(\text{normalized error} = \frac{\%\text{error}}{(\text{range}/1000)^2} \times \frac{\text{focal length}}{50}\right)$; the normalized error metric is intended only as a first approximation and ignores many other factors which may influence errors. All the structure from motion values are for the "Yosemite" sequence [Barron 92]. The range maps for the optical differentiation from simulation (sim) are illustrated in Figure 3.19; for the viewpoint differentiation (view) see Figure 3.42; and for the aperture size differentiation (ap) see Figure 3.43.

Bibliography

[Adelson 91] Adelson, E.H. and Wang, J.Y.A. A Stereoscopic Camera Employing a Single Main Lens. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 619–624, Maui, HI, 1991.

[Adelson 92] Adelson, E.H. and Wang, J.Y.A. Single Lens Stereo with a Plenoptic Camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):99–106, 1992.

[Allebach 76] Allebach, J.P. and Liu, B. Random Quasi-Periodic Halftone Process. *Journal of Optical Society of America*, 66:909–917, 1976.

[Barron 92] Barron, J.L., Fleet, D.J., and Beauchemin, S.S. *Performance of Optical Flow Techniques.* Technical Report RPL-TR-9107, Department of Computing and Information Science, Queen's University, Kingston, Ontario, 1992.

[Beil 94] Beil, W. Steerable Filters and Invariance Theory. *Pattern Recognition Letters*, 15:453–460, 1994.

[Billotet-Hoffman 83] Billotet-Hoffman, C. and Bryngdahl, O. On the Error Diffusion Technique for Electronic Halftoning. *Proc. SID*, 24:253–258, 1983.

[Boyle 70] Boyle, W.S. and Smith, G.E. Charge-Coupled Semiconductor Devices. *Bell Systems Technical Journal*, 49:587, 1970.

[Cochran 92] Cochran, S.D. and Medioni, G. 3-D Surface Description from Binocular Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(10):981–994, 1992.

[Collings 90] Collings, P.J. *Liquid Crystals: Nature's Delicate Phase of Matter.* Princeton University Press, Princeton, NJ, 1990.

[Costeira 95] Costeira, J. and Kanade, T. A Multi-Body Factorization Method for Motion Analysis. In *Proceedings of the International Conference on Computer Vision*, pages 1071–1076, Cambridge, MA, 1995.

[Danielsson 80] Danielsson, Per-Erik. Rotation-Invariant Linear Operators with Directional Response. In *5th Int'l Conf. Patt. Rec.*, Miami, FL, December 1980.

[Dhond 89] Dhond, U.R. and Aggarwal, J.K. Structure from Stereo - A Review. *IEEE Transactions on Systems, Man and Cybernetics*, 19(6):1489–1510, 1989.

[Dippe 85] Dippe, M.A. and Wold, E.H. Antialiasing Through Stochastic Sampling. *Computer Graphics*, 19(3):69–78, 1985.

[Dowski 94] Dowski, E.R. and Cathey, W.T. Single-Lens Single-Image Incoherent Passive-Ranging Systems. *Applied Optics*, 33(29):6762–6773, 1994.

[Farid 96a] Farid, H. The Design of Digital Derivative Filters. 1996.

[Farid 96b] Farid, H. The Fourier Transform. 1996.

[Farid 96c] Farid, H. The Principles of Steerability. 1996.

[Farid 96d] Farid, H. and Simoncelli, E.P. A Differential Optical Range Camera. In *Proceedings of the Annual Meeting of the Optical Society of America*, 1996.

[Farid 97] Farid, H. and Simoncelli, E. P. The Design of Multi-Dimensional, Higher-Order Derivative Filters. *IEEE Transactions on Image Processing*, 1997.

[Feynman 77] Feynman, R. P., Leighton, R. B., and Sands, M. *The Feynman Lectures on Physics.* Addison-Wesley Publishing Company, 1977.

[Fleet 90] Fleet, D.J. and Jepson, A.D. Computation of Component Image Velocity from Local Phase Information. *International Journal of Computer Vision*, 5:77–104, 1990.

[Floyd 76] Floyd, R.W. and Steinberg, L. An Adaptive Algorithm for Spatial Grey Scale. *Proceedings of the Society for Information Display*, 17(2):75–77, 1976.

[Freeman 91] Freeman, W T and Adelson, E H. The Design and Use of Steerable Filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9):891–906, 1991.

[Gokstorp 95] Gokstorp, M. and Westelius, C.J. Multiresolution Differential-Based Disparity Estimation. In *9th Scandinavian Conference on Image Analysis*, pages 67–76, Uppsala, Sweden, 1995.

[Goshtasby 93] Goshtasby, A. and Gruver, W.A. Design of a Single-Lens Stereo Camera System. *Pattern Recognition*, 26(6):923–937, 1993.

[Grimson 81] Grimson, W.E.L. A Computer Implementation of a Theory of Human Stereo Vision. *Proceedings of the Royal Society of London*, B292:217–253, 1981.

[Grossman 87] Grossman, P. Depth from Focus. *Pattern Recognition Letters*, 5:63–69, 1987.

[Gupta 95] Gupta, S.N. and Prince, J.L. On Variable Brightness Optical Flow for Tagged MRI. In *Information Processing in Medical Imaging: 14th Int'l Conf.* Kluwer, 1995.

[Heeger 92] Heeger, D.J. and Jepson, A.D. Subspace Methods for Recovering Rigid Motion I: Algorithm and Implementation. *International Journal of Computer Vision*, 7(2):95–117, 1992.

[Hel-Or 96] Hel-Or, Y. and Teo, P. Canonical Decomposition of Steerable Functions. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 809–816, San Francisco, CA, 1996.

[Hoff 89] Hoff, W. and Ahuja, N. Surfaces from Stereo: Integrating Feature Matching, Disparity Estimation and Contour Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(2):121–136, 1989.

[Horn 81] Horn, B.K.P. and Schunck, B.G. Determining Optical Flow. *Artificial Intelligence*, 17, 1981.

[Horn 86] Horn, B.K.P. *Robot Vision.* MIT Press, Cambridge, MA, 1986.

[Jones 93] Jones, D.G. and Lamb, D.G. *Analyzing the Visual Echo: Passive 3-D Imaging with a Multiple Aperture Camera.* Technical Report CIM-93-3, Department of Electrical Engineering, McGill University, 1993.

[Kanade 95] Kanade, T., Kano, H., Kimura, S., Yoshida, A., and Oda, K. Development of a Video-Rate Stereo Machine. In *Proceedings of the International Conference on Computer Vision*, pages 95–100, Cambridge, MA, 1995.

[Knutsson 93] Knutsson, H. and Granlund, G. H. Texture Analysis Using Two-Dimensional Quadrature Filters. *IEEE Comput. Soc. Workshop Comp. Architecture Patt. Anal. Image Database Mgmt.*, pages 388–397, 1993.

[Koenderink 87] Koenderink, J. J. and van Doorn, A. J. Representation of local geometry in the visual system. *Biological Cybernetics*, 55:367–375, 1987.

[Koenderink 91] Koenderink, J.J. and van Doorn, A.J. Affine Structure from Motion. *Journal of Optical Society of America*, 8(2):377–385, 1991.

[Koschan 93] Koschan, A. *What is New in Computational Stereo Since 1989: A Survey on Current Stereo Papers.* Technical Report 93-22, Technische Universitat Berlin, 1993.

[Kramer 93] Kramer, J., Seitz, P., and Baltes, H. Inexpensive Range Camera Operating at Video Speed. *Applied Optics*, 32(13):2323–2330, 1993.

[Krotkov 87] Krotkov, E. Focusing. *International Journal of Computer Vision*, 1:223–237, 1987.

[Lim 87] Lim, H.S. and Binford, T.O. Stereo Correspondence: A Hierarchical Approach. In *Proceedings of the DARPA Image Understanding Workshop*, Los Angeles, CA, February 1987.

[Lucas 81] Lucas, B.D. and Kanade, T. An Iterative Image Registration Technique with an Application to Stereo Vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pages 674–679, Vancouver, 1981.

[Marr 79] Marr, D. and Poggio, T. A Computational Theory of Human Stereo Vision. *Proceedings of the Royal Society of London*, B204:301–328, 1979.

[Menard 96] Menard, C. and Leonardis, A. Robust Stereo on Multiple Resolutions. In *Proceedings of the International Conference on Pattern Recognition*, pages 910–914, Vienna, Austria, August 1996.

[Nayar 95] Nayar, S.K., Watanabe, M., and Noguchi, M. Real-Time Focus Range Sensor. In *Proceedings of the International Conference on Computer Vision*, pages 995–1001, Cambridge, MA, 1995.

[Negahdaripour 93] Negahdaripour, S. and Yu, C. A Generalized Brightness Change Model for Computing Optical Flow. In *Proceedings of the International Conference on Computer Vision*, pages 2–11, Berlin, Germany, 1993.

[Nishimoto 87] Nishimoto, Y. and Shirai, Y. A Feature-Based Stereo Model Using Small Disparities. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 192–196, 1987.

[Oppenheim 83] Oppenheim, A. V., Willsky, A. S., and Young, I. T. *Signals and Systems.* Prentice Hall, 1983.

[Oppenheim 89] Oppenheim, A. V. and Schafer, R. W. *Discrete-Time Signal Processing.* Prentice Hall, 1989.

[Ozanian 95] Ozanian, T. Approaches for Stereo Matching. *Modeling Identification and Control*, 16(2):65–94, 1995.

[Pentland 87] Pentland, A.P. A New Sense for Depth of Field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(4):523–531, 1987.

[Pentland 94] Pentland, A., Scherock, S., Darrell, T., and Girod, B. Simple Range Cameras Based on Focal Error. *Journal of Optical Society of America*, 11(11):2925–2934, 1994.

[Perona 92] Perona, P. Steerable-Scalable Kernels for Edge Detection and Junction Analysis. *Image and Vision Computing*, 10(10):663–672, 1992.

[Perona 95] Perona, P. Deformable Kernels for Early Vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(5):488–499, 1995.

[Sangster 69] Sangster, F.L.J. and Teer, K. Bucket-Brigade Electronics. *IEEE Journal Solid-State Circuits*, SC-4:131, 1969.

[Simoncelli 92] Simoncelli, E. P., Freeman, W. T., Adelson, E. H., and Heeger, D. J. Shiftable Multi-scale Transforms. *IEEE Trans. Information Theory*, 38(2):587–607, March 1992. Special Issue on Wavelets.

[Simoncelli 93] Simoncelli, E.P. *Distributed Analysis and Representation of Visual Motion.* PhD Dissertation, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, Cambridge, MA, January 1993. Also available as MIT Media Laboratory Vision and Modeling Technical Report #209.

[Simoncelli 94] Simoncelli, E.P. Design of Multi-Dimensional Derivative Filters. In *First International Conference on Image Processing*, Austin, TX, 1994.

[Simoncelli 95] Simoncelli, E.P. and Farid, H. Single Lens Range Imaging from Aperture Derivative. 1995. Patent Pending.

[Simoncelli 96a] Simoncelli, E P and Farid, H. Steerable Wedge Filters for Local Orientation Analysis. *IEEE Transactions on Image Processing*, 1996.

[Simoncelli 96b] Simoncelli, E.P. and Farid, H. Direct Differential Range Estimation from Aperture Derivatives. In *Proceedings of the European Conference on Computer Vision*, pages 82–93 (volume II), Cambridge, England, 1996.

[Stevenson 95] Stevenson, D.E. and Fleck, M.M. *Nonparametric Calibration of Distortion.* Technical Report TR95-07, Department of Computer Science, University of Iowa, Iowa City, Iowa, 1995.

[Strang 88] Strang, G. *Linear Algebra and its Applications.* Saunders College Publishing, 1988.

[Subbarao 88] Subbarao, M. Parallel Depth Recovery by Changing Camera Parameters. In *Proceedings of the International Conference on Computer Vision*, pages 149–155, 1988.

[Subbarao 95] Subbarao, M. and Choi, T. Accurate Recovery of Three-Dimensional Shape from Image Focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(3):266–274, 1995.

[Teoh 84] Teoh, W. and Zhang, X.D. An Inexpensive Stereoscopic Vision System for Robots. In *International Conference on Robotics*, 1984.

[Terzopoulos 85] Terzopoulos, D. Concurrent Multilevel Relaxation. In *Proceedings of the DARPA Image Understanding Workshop*, Miami Beach, FL, December 1985.

[Tomasi 92] Tomasi, C. and Kanade, T. Shape and Motion from Image Streams Under Orthography: a Factorization Method. *International Journal of Computer Vision*, 9:137–154, 1992.

[Ulichney 88] Ulichney, R. *Digital Halftoning.* MIT Press, 1988.

[Uras 88] Uras, S., Girosi, F., Verri, A., and Torre, V. A Computational Approach to Motion. *Biological Cybernetics*, 60:79–97, 1988.

[Vleeschauwer 93] Vleeschauwer, D. De. An Intensity-Based, Coarse-to-Fine Approach to Reliably Measure Binocular Disparity. *CVGIP: Image Understanding*, 57(2):204–218, 1993.

[Weinstock 52] Weinstock, R. *Calculus of Variation With Applications to Physics and Engineering.* McGraw-Hill Book Company Inc., 1952.

[Woo 84] Woo, B. *A Survey of Halftoning Algorithms and Investigation of the Error Diffusion Technique.* Technical Report S.B. Thesis, Massachusetts Institute of Technology, 1984.

[Xiong 93] Xiong, Y. and Shafer, S. Depth from Focusing and Defocusing. In *Proceedings of the DARPA Image Understanding Workshop*, pages 967–976, 1993.

Index