Journal of the Optical Society of America, 15(7):1777-1786, 1988.

Range Estimation by Optical Differentiation

Hany Farid

Perceptual Science Group
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology
Cambridge, MA 02139

farid@persci.mit.edu

Eero P. Simoncelli

Center for Neural Science, and
Courant Inst. of Mathematical Sciences
New York University
New York, NY 10003-6603

eero.simoncelli@nyu.edu

Abstract

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. This method uses a single stationary camera and a pair of calibrated optical masks to directly measure this differential quantity. We also describe a variant based on changes in aperture size. The subsequent computation of the range image involves simple arithmetic operations, and is suitable for real-time implementation. We present the theory of this technique and show results from a prototype camera which we have constructed.

1 Introduction

Visual images are formed via the projection of light from the three-dimensional world onto a two dimensional sensor. In an idealized pinhole camera, all points lying on a ray passing through the pinhole will be projected onto the same image position. Thus, information about the distance to objects in the scene (i.e., *range*) is lost. Range information can be recovered by measuring the change in appearance of the world resulting from a change in viewing position (i.e., parallax). Traditionally, this is accomplished via simultaneous measurements with two cameras at different positions (binocular stereo), or via sequential measurements collected from a moving camera (structure from motion).

The recovery of range in these approaches frequently relies on the assumption of *brightness constancy* [1]: the brightness of the image of a point in the world is constant when seen from different viewpoints. Consider the formulation of this assumption in one dimension (the extension to two dimensions is straightforward). Let $f(x;v)$ describe the intensity function, as measured through a pinhole camera system. The variable v corresponds to the pinhole position (along an axis perpendicular to the optical axis). The variable x parameterizes the position on the sensor. This configuration is illustrated in Figure 1. According to the assumption, the intensity function $f(x;v)$ is of the form:

$$f(x;v) \quad = \quad I\left(x - \tfrac{vd}{Z(x)}\right), \qquad (1)$$

where d is the distance between the pinhole and the sensor and $Z(x)$ is the range (distance from the pinhole to a point in the world). That is, the image of a point in the world appears at a position that depends on both the pinhole position (v) and the range of the point, but always has the same intensity.

For the purpose of recovering range, we are interested in computing the change in the appearance of the world with respect to change in viewing position. It is thus natural to consider differential measurement techniques. Taking the partial derivative of $f(x;v)$ with respect to viewing position, and evaluating at $v = 0$ gives:

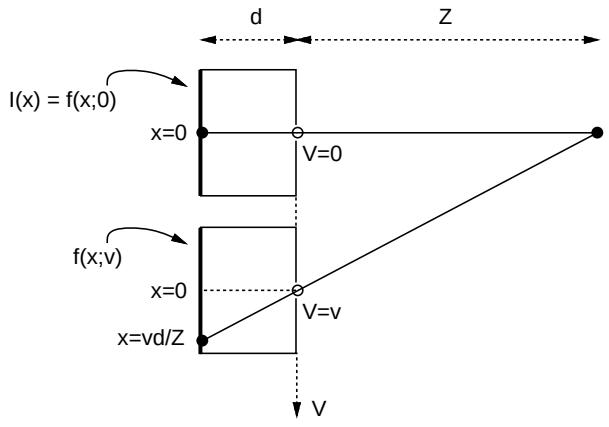

Figure 1: Geometry for a binocular stereo system with pinhole cameras. The variable V parameterizes the position of the camera pinholes. According to the brightness constancy constraint, the intensity of a point in the world, as recorded by the two pinhole cameras, should be the same.

$$\begin{aligned} I_v(x) \quad &\equiv \quad \tfrac{\partial f(x;v)}{\partial v} \mid_{v=0} \\ &= \quad -\tfrac{d}{Z(x)} I'(x), \end{aligned} \qquad (2)$$

where $I'(\cdot)$ indicates the derivative of $I(\cdot)$ with respect to its argument. Consider next the partial derivative with respect to the spatial parameter:

$$\begin{aligned} I_x(x) \quad &\equiv \quad \tfrac{\partial f(x;v)}{\partial x} \mid_{v=0} \\ &= \quad I'(x). \end{aligned} \qquad (3)$$

Combining these two differential measurements gives:

$$I_v(x) \quad = \quad -\tfrac{d}{Z(x)} I_x(x). \qquad (4)$$

Clearly, an estimate of the range, $Z(x)$, can be computed using this equation. With this formulation, the difficulty typically lies in obtaining an accurate and efficient measurement of the viewpoint derivative, $I_v(x)$.

Many traditional range algorithms may be framed in terms of this differential formalism. The viewpoint derivatives are typically computed by capturing images from a set of discrete viewpoints, either simultaneously (using two or more cameras),

or sequentially (using a single moving camera). In the case of binocular stereo (e.g., [2]), $I_v(x)$ is approximated as the difference of two images taken from cameras at different viewpoints. Similarly, differences of consecutive images are often used to compute viewpoint derivatives for structure from motion. However, these differences are generally a poor approximation to the viewpoint derivative. In addition, these approaches require precise knowledge of the camera position associated with each image.

In this paper, we describe a technique for direct measurement of the viewpoint derivative using a single camera at a fixed location. We begin by describing the conceptual framework for a simplified world consisting of a single point-source. These concepts are then generalized in Section 3, and are followed by a description of the implementation and results from a prototype camera.

2 Optical Differentiation

We now show a direct method for the measurement of the required derivatives (Equation (4)) from a single stationary camera and a pair of optical attenuation masks. Conceptually, this is made possible by replacing the pinhole camera model with a more realistic "thin lens" model, in which the camera collects light from a range of viewpoints. This viewpoint information is lost once the light is imaged on the sensor plane. Nonetheless, at the *front* of the lens, the information is available. It is precisely this information that we exploit.

Consider a world consisting of a single point light source and a lens-based imaging system with a variable-opacity optical mask, $M(u)$, placed directly in front of the lens (left side, Figure 2). The light striking the lens is attenuated by the value of the optical mask function at that particular spatial location (we assume that the values of such a mask function are real numbers in the range [0,1], and that the optical transfer function of the imaging system is constant). With such a configuration, the image of the point source will be a scaled and

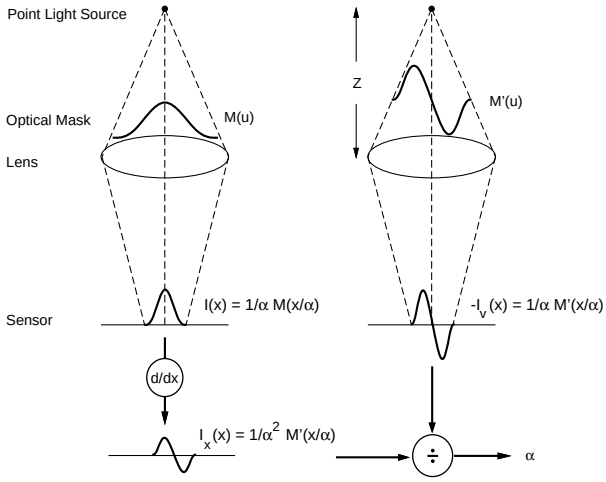

Figure 2: Illustration of direct differential range determination for a single point source. Images of a point light source are formed using two different optical masks, corresponding to the function $M(u)$ and its derivative, $M'(u)$. In each case, the image formed is a scaled and dilated copy of the mask function (by a factor α). Computing the spatial (image) derivative of the image formed under mask $M(u)$ produces an image that is identical to the image formed under the derivative mask, $M'(u)$, except for a scale factor α. Thus, α may be estimated as the ratio of the two images. Range is then computed from α using the relationship given in Equation (6).

dilated version of the optical mask function:

$$I(x) \quad = \quad \tfrac{1}{\alpha} M(\tfrac{x}{\alpha}), \tag{5}$$

as illustrated in Figure 2. The scale factor, α, is a monotonic function of the distance to the point source, Z, and is easily derived from the imaging geometry and the lens equation:

$$\alpha \quad = \quad 1 - \tfrac{d}{f} + \tfrac{d}{Z}, \tag{6}$$

where d is the distance between lens and sensor, and f is the focal length of the lens.

2.1 Optical Viewpoint Differentiation

In the system shown on the left side of Figure 2, the *effective* viewpoint may be altered by translating the mask, while leaving the lens and sensor

stationary. For example, consider a mask with a single pinhole; different views of the world are obtained by sliding the pinhole across the front of the lens. The generalized image intensity function for a mask centered at position v is written as:

$$f(x;v) \;=\; \tfrac{1}{\alpha} M\left(\tfrac{x}{\alpha} - v\right), \qquad (7)$$

assuming that the non-zero portion of the mask does not extend past the edge of the lens.

The *differential* change in the image with respect to a change in the mask position is obtained by taking the derivative of this equation with respect to the mask position, v, and evaluating at $v = 0$:

$$\begin{aligned} I_v(x) &\equiv \tfrac{\partial}{\partial v} f(x;v)|_{v=0} \\ &= -\tfrac{1}{\alpha} M'\left(\tfrac{x}{\alpha}\right), \end{aligned} \qquad (8)$$

where $M'(\cdot)$ is the derivative of the mask function $M(\cdot)$ with respect to its argument. Observe that this equation is of the same general form as Equation (5). Thus, the derivative with respect to viewing position, $I_v(x)$, *may be measured directly by imaging with the optical mask* $M'(\cdot)$. The spatial derivative of $I(x)$ may also be written in terms of $M'(\cdot)$:

$$\begin{aligned} I_x(x) &\equiv \tfrac{\partial}{\partial x} f(x;v)|_{v=0} \\ &= \tfrac{1}{\alpha^2} M'\left(\tfrac{x}{\alpha}\right) \end{aligned} \qquad (9)$$

Combining these two equations gives a relationship between the two derivatives:

$$I_x(x) \;=\; -\tfrac{1}{\alpha} I_v(x). \qquad (10)$$

From this relationship, the scaling parameter α may be computed as the ratio of the spatial derivative of the image formed through the optical mask $M(u)$, and the image formed through the derivative of that optical mask, $M'(u)$. This computation is illustrated in Figure 2. The distance to the point source can subsequently be computed from α using the monotonic relationship given in Equation (6). Note that the resulting equation for distance is identical to that of Equation (4) when $d = f$ (i.e., when the camera is focused at infinity). In practice, the mask $M'(u)$ cannot be directly used as an attenuation mask, since it contains negative values. This issue is addressed in Section 5.2.

2.1.1 Least-Squares Solution

The relationship in Equation (10) holds at all image positions, x, within the image. In order to compute a reliable estimate of the distance to the point source (and avoid problems with localized singularities where $I_x(x) = 0$ in Equation (10)), information can be combined over image position using a least-squares estimator (e.g., [2]). Specifically, an error measure is defined as follows:

$$E(\alpha) \;=\; \sum_{x \in P} (I_v(x) + \alpha I_x(x))^2, \qquad (11)$$

where P is a set of image positions in a local neighborhood around the origin. This assumes that the parameter α (and therefore range) is constant over this neighborhood. Taking the derivative with respect to α, setting equal to zero and solving for α yields the minimal solution:

$$\alpha \;=\; -\frac{\sum_{x \in P} I_v(x) I_x(x)}{\sum_{x \in P} I_x^2(x)}. \qquad (12)$$

By integrating over a small patch in the image, the least-squares solution avoids singularities at locations where the spatial derivative, $I_x(x)$, is zero. 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. This singularity may be avoided by considering a maximum a posteriori (MAP) estimator with a Gaussian prior on α (as in [3]). For a prior variance of σ^2, the resulting estimate is:

$$\alpha \;=\; -\frac{\sum_{x \in P} I_v(x) I_x(x)}{\sum_{x \in P} [I_x(x)]^2 + \sigma^2}. \qquad (13)$$

This algorithm extends to a two-dimensional image plane: we need only consider two-dimensional masks $M(u, w)$, and the horizontal partial derivative $M_u(u, w) \equiv \partial M(u, w)/\partial u$. For a more robust implementation, the vertical partial derivative mask, $M_v(u, w) \equiv \partial M(u, w)/\partial w$, may also be included. The least-squares error function becomes:

$$\begin{aligned} E(\alpha) \;=\; \sum_{(x,y) \in P} & [I_u(x,y) + \alpha I_x(x,y)]^2 \\ & + [I_w(x,y) + \alpha I_y(x,y)]^2. \end{aligned} \qquad (14)$$

As above, the MAP estimator gives:

$$\alpha = -\frac{\sum_{(x,y)\in P} I_u(x,y)I_x(x,y) + I_w(x,y)I_y(x,y)}{\sum_{(x,y)\in P}[I_x(x,y)]^2 + [I_y(x,y)]^2 + \sigma^2} \quad (15)$$

2.2 Optical Aperture Size Differentiation

Here we show a similar technique for estimating range that is based on the derivative with respect to aperture size. Discrete versions of this concept have been studied and are generally called "range from defocus" (e.g., [4, 5, 6, 7, 8, 9]). The generalized intensity function for a mask with aperture size A is written as:

$$f(x;A) = \frac{1}{A\alpha} M\left(\frac{x}{A\alpha}\right). \quad (16)$$

The $\frac{1}{A}$ factor is included to ensure that the changes in the aperture size do not change the mean intensity of the image.

The *differential* change in the image with respect to aperture size may be computed by taking the derivative of this equation with respect to aperture size, A, evaluated at $A = 1$:

$$\begin{aligned} I_A(x) &\equiv \frac{\partial}{\partial A} f(x;A)|_{A=1} \\ &= -\frac{1}{\alpha} M\left(\frac{x}{\alpha}\right) - \frac{x}{\alpha^2} M'\left(\frac{x}{\alpha}\right) \\ &= -\frac{1}{\alpha}\left[M\left(\frac{x}{\alpha}\right) + \frac{x}{\alpha} M'\left(\frac{x}{\alpha}\right)\right]. \end{aligned} \quad (17)$$

Since this equation is again of the general form of Equation 5, the derivative with respect to aperture size, $I_A(x)$, *may be computed directly by imaging with the optical mask* $-(M(u) + uM'(u))$. Similar to the viewpoint derivative, one can compute a related image by spatial differentiation. In particular, the image formed using the optimal mask $-uM(u)$ is:

$$J(x) = -\frac{x}{\alpha^2} M\left(\frac{x}{\alpha}\right), \quad (18)$$

and the spatial derivative of this image is:

$$J_x(x) = -\frac{1}{\alpha^2}\left[M\left(\frac{x}{\alpha}\right) + \frac{x}{\alpha} M'\left(\frac{x}{\alpha}\right)\right], \quad (19)$$

which is equal to Equation (17) with an additional factor of α. Thus, the ratio of these two image measurements could be used to compute α (and thus range).

A particularly interesting choice for a mask function is a Gaussian:

$$M(u;A) = \frac{1}{A}\exp\left(-\frac{u^2}{2A^2}\right),$$

for which

$$I_A(x) = \frac{1}{\alpha} M''\left(\frac{x}{\alpha}\right) \quad (20)$$

which may be related to the spatial second derivative:

$$\begin{aligned} I_{xx}(x) &= \frac{1}{\alpha^3} M''\left(\frac{x}{\alpha}\right) \\ &= \frac{1}{\alpha^2} I_A(x) \end{aligned} \quad (21)$$

Similar relationships are derived in [7, 9]. This equation may be used to compute the square of the scaling parameter α as the ratio of a spatial and aperture size derivative. Again, the important point is that the derivative with respect to aperture size is measured directly by imaging through the appropriate optical mask.

As in the previous section, this formulation can be extended to 2-D and a least-squares formulation can be used to combine information over spatial location. We write an error function of the form:

$$E(\alpha^2) = \sum_{(x,y)\in P} \left[I_A(x,y) - \alpha^2(I_{xx}(x,y) + I_{yy}(x,y))\right]^2 \quad (22)$$

and solving for the minimizing α^2 gives:

$$\alpha^2 = \frac{\sum_{(x,y)\in P} I_A(x,y)[I_{xx}(x,y) + I_{yy}(x,y)]}{\sum_{(x,y)\in P}[I_{xx}(x,y) + I_{yy}(x,y)]^2 + \sigma^2}. \quad (23)$$

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. Second, the ratio of the aperture size derivative and spatial derivative is proportional to the square of the parameter α. As such, only the absolute value of α can be determined. A second look at the optical masks reveals why this must be so. Whereas the viewpoint derivative mask is antisymmetric with respect to the center of the lens, the aperture size derivative mask is symmetric. As a result, points positioned on opposite sides of the focal plane will differ by a sign in the case of the

anti-symmetric mask, but could appear identical in the case of the symmetric mask. This ambiguity may be eliminated by focusing the camera at infinity, thus ensuring that $\alpha > 0$.

3 Range Map Estimation by Optical Differentiation

Equations (10) and (21) embody the fundamental relationships used for the optical differential computation of range for a single point light source. A world consisting of a collection of many such point sources imaged through an optical mask will produce an image consisting of a superposition of scaled and dilated versions of the masks. In particular, in the case of the viewpoint derivative, an expression for the image can be written by integrating over the images of the visible points, p, in the world:

$$f(x;v) \quad = \quad \int dx_p \, \frac{1}{\alpha(x_p)} \, M\left(\frac{x-x_p}{\alpha(x_p)} - v\right) \, L(x_p), \quad (24)$$

where the integral is performed over the variable x_p, the position in the sensor of a point p projected through the center of the lens. The intensity of the world point p is denoted as $L(x_p)$, and $\alpha(x_p)$ is monotonically related to the distance to p (as in Equation (6)). Note that each point source is assumed to produce a uniform light intensity across the optical mask (i.e., brightness constancy). Again, consider the derivatives of $f(x;v)$ with respect to viewing position, v, and image position, x, evaluated at $v = 0$:

$$\begin{aligned} I_v(x) \quad &\equiv \quad \frac{\partial}{\partial v} f(x;v) \mid_{v=0} \\ &= \quad -\int dx_p \, \frac{1}{\alpha(x_p)} \, M'\left(\frac{x-x_p}{\alpha(x_p)}\right) \, L(x_p), \quad (25) \end{aligned}$$

and

$$\begin{aligned} I_x(x) \quad &\equiv \quad \frac{\partial}{\partial x} f(x;v) \mid_{v=0} \\ &= \quad \int dx_p \, \frac{1}{\alpha^2(x_p)} \, M'\left(\frac{x-x_p}{\alpha(x_p)}\right) \, L(x_p). \quad (26) \end{aligned}$$

An exact solution for $\alpha(x_p)$ is nontrivial, since it is embedded in the integrand and depends on the integration variable. Nevertheless, the computation of Equation (10) gives an estimate:

$$\begin{aligned} \hat{\alpha}(x_p) \quad &\equiv \quad -\frac{I_v(x)}{I_x(x)} \\ &= \quad -\frac{\int dx_p \, \alpha(x_p) \frac{1}{\alpha^2(x_p)} \, M'\left(\frac{x-x_p}{\alpha(x_p)}\right) \, L(x_p)}{\int dx_p \, \frac{1}{\alpha^2(x_p)} \, M'\left(\frac{x-x_p}{\alpha(x_p)}\right) \, L(x_p)}. \quad (27) \end{aligned}$$

The estimate $\hat{\alpha}(x_p)$ is seen to be a local average of $\alpha(x_p)$, weighted by $\frac{1}{\alpha^2(x_p)} \, M'\left(\frac{x_p}{\alpha(x_p)}\right) \, L(x_p)$. In particular, the more out of focus the image, the larger the magnitude of α, and the larger the size of this averaging region. Of course, if the points in a local region lie on a frontal-parallel plane relative to the sensor, the estimate will be exact.

4 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.

The use of optical attenuation masks for range estimation has been considered in the work of Dowski and Cathey [10] and Jones and Lamb [11]. 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 [12] describe a single-lens, single-image range camera, termed the "plenoptic camera". This approach is based on the same underlying principles as our own, but is quite different in implementation. The authors place a lenticular array (a sheet of miniature lenses) over the sensor, allowing the camera to capture images from several viewpoints. More specifically, a group of

5×5 pixels under each lenticule (termed a macropixel) captures a unique viewpoint. From only a single image, a viewpoint derivative can be computed across several macropixels. An image derivative is computed on the mean values of the macropixels, and range is determined by the familiar ratio of these derivative measurements. Note however, that the viewpoint derivative is computed from a discrete set of viewpoints. The authors noted several technical difficulties with this approach, most notably, aliasing and the difficulty of precisely aligning the lenticular array with the CCD sensor.

A series of single-lens stereo systems have also been developed [13, 14, 15]. 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, 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.

5 Experiments

We have verified the principles of optical differentiation for range estimation both in simulation and experimentation [16]. This section discusses the construction of a prototype camera, and shows example range maps computed using this camera.

5.1 Prototype Camera

We have constructed a prototype camera for validating the differential approach to range estimation. As illustrated in Figure 3, the camera consists of an optical attenuation mask sandwiched between a pair of planar-convex lenses, and placed in front of a CCD camera. This arrangement places the mask at the center of the optical system, where the viewpoint information is isolated. The camera is a Sony model XC-77R, the lenses are 25mm in diameter, 50mm focal length, and were placed 31mm from the sensor. Thus, the camera was focused at a distance of 130mm. We have employed a liquid crystal spatial light modulator (LC SLM), purchashed from CRL Smectic Technology (Middlesex, UK), for use as an optical attenuation mask, also shown in Figure 3. This device is a fully programmable, fast-switching, twisted nematic liquid crystal display measuring 38mm (W) $\times$ 42mm (H) $\times$ 4.3mm (D), with a display area of 28.48mm (W) $\times$ 20.16mm (H). The spatial resolution is 640×480 pixels, with 4 possible transmittance values. The display is controlled through a standard VGA video interface, supplied by the manufacturer. The LC SLM refreshes its display at 30 Hz; when synchronized with the camera, the pair of images may be acquired by temporal interleaving at a rate of 15 Hz. Alternatively, a pair of images could be acquired simultaneously (i.e., 30 Hz), by employing an additional camera, some beam-splitting optics, and two fixed optical masks, as in [8]. The subsequent processing (i.e., convolutions and arithmetic combinations) can be performed in real-time on a general-purpose DSP chip or perhaps even on a fast RISC microprocesser.

5.2 Optical Masks

The most essential component of our range camera is the 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 full transmittance. The viewpoint and aperture size derivative masks both contain negative values, and thus may not be used directly. Furthermore, adding a positive constant to the derivative mask destroys the derivative relationship. Nonetheless a pair of non-negative masks can be constructed by taking the appropriate linear combination of the original masks. In particular, consider the fol-

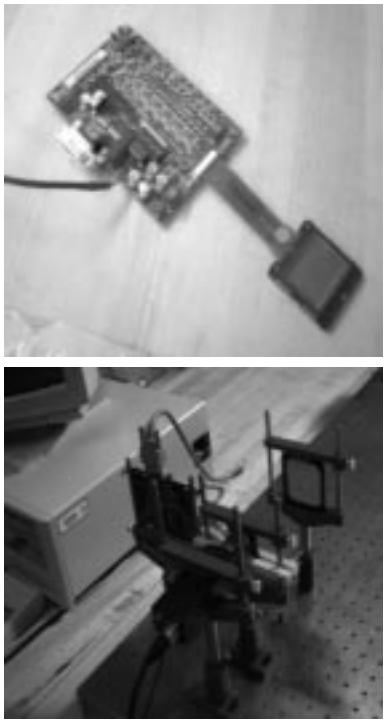

Figure 3: Prototype Camera. Illustrated on the top is a fast switching liquid crystal spatial light modulator (LC SLM) employed as an optical attenuation mask. Illustrated below right is our range camera consisting of an off-the-shelf CCD camera and the LC SLM sandwiched between a pair of planar-convex lenses. The target consists of a piece of paper with a random texture pattern.

lowing construction of a pair of non-negative masks:

$$\begin{aligned} M_1(u) &= \beta_1 M(u) + \gamma_1 M'(u) \\ M_2(u) &= \beta_2 M(u) - \gamma_2 M'(u), \end{aligned} \qquad (28)$$

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 then be reconstructed through a simple linear combinations of the non-negative masks, $M_1(u)$ and $M_2(u)$:

$$M(u) = \frac{\gamma_2 M_1(u) + \gamma_1 M_2(u)}{\gamma_2\beta_1 + \gamma_1\beta_2}$$

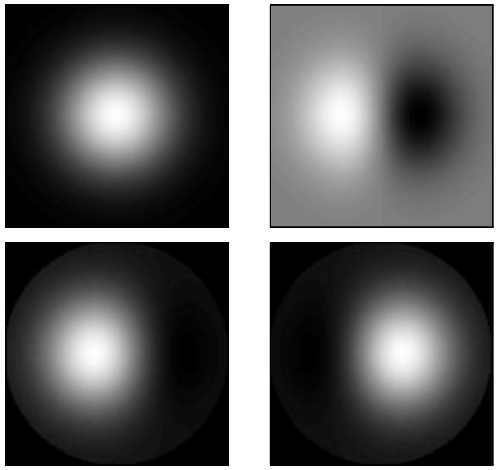

Figure 4: Gaussian aperture masks. Top: A two-dimensional Gaussian mask, $M(u, w)$ and its partial derivative, $M'(u, w)$. Bottom: Two non-negative aperture masks, $M_1(u, w)$ and $M_2(u, w)$. These are computed from the pair of left-most masks using Equation (28).

$$M'(u) = \frac{\beta_2 M_1(u) - \beta_1 M_2(u)}{\gamma_1\beta_2 + \gamma_2\beta_1}. \qquad (29)$$

If the imaging system is linear, 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 $M_2(u)$:

$$\begin{aligned} I(x) &= \frac{\gamma_2 I_1(x) + \gamma_1 I_2(x)}{\gamma_2\beta_1 + \gamma_1\beta_2} \\ I_v(x) &= \frac{\beta_2 I_1(x) - \beta_1 I_2(x)}{\gamma_1\beta_2 + \gamma_2\beta_1}. \end{aligned} \qquad (30)$$

where $I_1(x)$ and $I_2(x)$ are the images formed under the masks $M_1(u)$ and $M_2(u)$, respectively. Clearly, this construction extends to 2-D optical masks as well. Illustrated in Figure 4 is a 2-D Gaussian mask, $M(u, w) = \frac{1}{2\pi\sigma^2} e^{-(u^2+w^2)/2\sigma^2}$, and its partial derivatives with respect to u (viewpoint derivative). Also illustrated are the non-negative masks computed from Equation (28).

5.2.1 Dithering

The LC SLM described in the previous section is a 2-bit display (i.e., only 4 transmittance levels).

In order to minimize the quantization effects due to this low resolution, a standard stochastic error diffusion dithering algorithm based on the Floyd / Steinberg algorithm [17] was employed. The value of a pixel is first quantized to one of the four available levels. The difference between the quantized pixel value and the desired value is computed and distributed in a weighted fashion to its neighbors. In [17], the authors suggest distributing the error to four neighbors with the following weighting:

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

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 should be causal). In order to avoid some of the visual artifacts due to the deterministic nature of this algorithm, stochastic variations may be introduced. Along these lines, we have taken the standard error diffusion algorithm and randomized the error (by a factor distributed uniformly between 90% and 110%) 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).

5.2.2 Calibration

In our system, there are at least two potential non-linearities that need to be eliminated. The first is the non-linearity in the light transmittance of the optical attenuation masks. A non-linearity at this stage will affect the derivative relationship of the optical masks. As illustrated in Figure 5, the LC SLM used to generate the optical masks is highly non-linear. Shown in the first panel of this figure is the light transmittance (measured with a photometer) through a uniform mask set to each of the four LCD levels. If this device were linear, then these measurements would lie along a unit-slope line. Clearly they do not.

This non-linearity may be corrected in the dithering process. More specifically, in the dithering algorithm the error between the quantized pixel value and the desired value is distributed to its neighbors. In order to correct for the non-linearity in the LC SLM, the diffused error may be assigned the difference between the desired and measured light transmittances of the quantized pixel. Note that we are assuming that the LC SLM pixels are independent (i.e., there are no interactions between neighboring transmittance values). Illustrated in Figure 5 is the measured light transmittance through 32 constant masks, dithered using to this technique. The data are seen to be much more linear than the original measurements.

The second potential non-linearity to consider is in the imaging sensor. With the optical mask 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 (dithered) uniform gray masks. As shown in Figure 5, the imaging sensor is fairly linear. In particular, the data in this figure closely resembles the measured light transmittance of the LC SLM after linearization (Figure 5). Thus, it is assumed that the imaging sensor is linear, and a second linearization correction is not necessary.

The final calibration that needs to be considered is that of the camera's intrinsic point spread function (PSF). More specifically, in describing the formation of an image through an optical attenuation mask we have been assuming that the camera's PSF is constant across the lens diameter (i.e., the image of a point light source is assumed to be a hard-edged rectangular function). This is generally not the case in a real camera: the PSF typically takes on a Gaussian-like shape. The PSF and the optical mask will be combined in a multiplicative fashion. Whereas before, we employed a matched pair of masks, $M(u)$ and $M'(u)$, with $\frac{dM(u)}{du} = M'(u)$, we now require a pair of masks, $M(u)$ and $\hat{M}(u)$, that obey the following constraint:

$$\hat{M}(u)H(u) \quad = \quad \frac{d(M(u)H(u))}{du}, \qquad (31)$$

where $H(u)$ is the camera PSF. That is, the deriva-

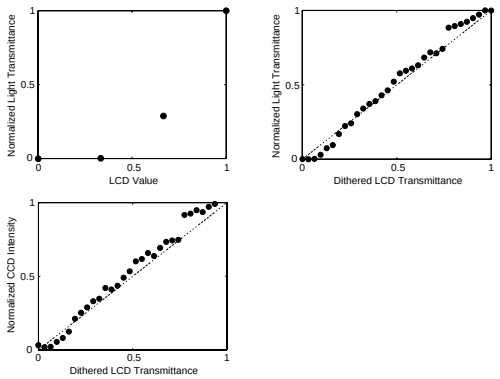

Figure 5: Calibration of LCD mask and CCD sensor. Shown in the top left panel is the normalized light transmittance (in cd/m^2, as measured with a photometer) through constant masks set to each of the four LCD values. Shown in top right panel is the normalized light transmittance measured through each of 32 uniform dithered and gamma-corrected masks, averaged over five trials. If our dithering and gamma-correction were perfect, the measurements (circles) would lie along a unit-slope line (dashed line). Shown in the lower panel is the normalized CCD 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, and spatially integrated over a 5×5 pixel neighborhood. If both the optical mask *and* imaging sensor were linear, then these measurements (circles) would lie along a unit-slope line (dashed line).

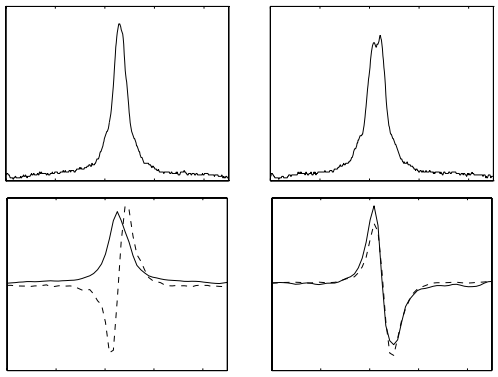

Figure 6: Illustrated in the top two panels are 1-D slices of the image of a "point light source" taken through a pair of non-negative Gaussian-based masks, I_1 and I_2. Shown in the bottom left panel are 1-D slices of the linear combination of the measurements (see Equation (30)). Shown in the bottom right panel are 1-D slices of the resulting images I_x (solid) and $-I_v$ (dashed). These images should be related to each other by a scale factor of α (see Equation (10)).

tive relationship should be imposed on the product of the optical mask and the PSF. We have not included this calibration in our experiments.

5.3 Results

We have verified the principles of range estimation by optical differentiation with a prototype camera which we have constructed (see Figure 3). 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 6 is an example of this behavior: shown are 1-D slices of images taken through a pair of non-negative Gaussian-based optical masks printed onto a sheet of transparent plastic (later experiments utilized the LC SLM described above). The appropriate linear combination of these slices (Equation (30)), and the resulting slices of $I_x(x)$ and $I_v(x)$.

In the remaining experiments the target consisted of a sheet of paper with a random texture pattern and back-illuminated with an incandescent lamp to help counter the low light transmittance of the LC SLM optical mask. Spatial derivatives were computed using a pair of 5-tap filter kernels described in [18]. For example, the x-derivative is computed via separable convolution with the one-dimensional derivative kernel in the x direction, and with a one-dimensional blurring kernel in the y direction. The viewpoint derivative was filtered with the blurring kernel in both spatial directions. Range was estimated using the least-squares formulation (Equations (15) or (23)), with a spatial integration neighborhood of 31×31 pixels.

Illustrated in Figures 7 and 8 are a pair of recovered range maps for frontal-parallel surfaces placed at distances of 11 and 17 cm from the camera. These figures illustrate the range maps computed using optical viewpoint and aperture size differentiation, respectively. The camera is focused at a distance of 13 cm. In the case of the viewpoint differentiation, the recovered range maps had a mean of 10.9 and 17.0 cm, with a standard deviation of 0.27 and 0.75 cm, and a minimum / maximum estimate of 10.1 / 11.8 cm and 15.1 / 19.4 cm, respectively. In the case of the aperture size differentiation, the recovered range maps had a mean of 11.0 and 17.0 cm, with a standard deviation of 0.06 and 0.16 cm and a minimum / maximum estimate of 10.8 / 11.2 cm and 16.5 / 17.5 cm, respectively. 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 one possible reason for this is that the aperture size masks have a higher total light transmittance: for the Gaussian-based optical masks, the mean light transmittance is 0.37, as compared to a mean of 0.20 for the viewpoint masks. Increased light transmittance produces a higher signal-to-noise ratio in the measurements. Although the aperture size differentiation has smaller errors in this example, it suffers from a sign ambiguity (i.e., surfaces on either side of the focal plane can be equally defocused). Illustrated in Figure 9 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 viewpoint differentiation formulation. Qualitatively, these range maps look quite reasonable.

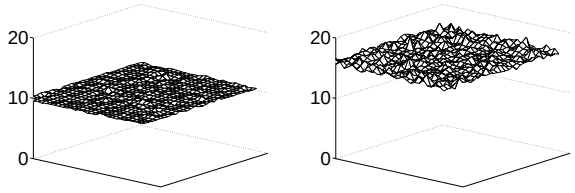

Figure 7: Illustrated are the recovered range maps using optical viewpoint differentiation for a pair of frontal-parallel surfaces at a distance of 11 and 17 cm from the camera. The computed range maps have a mean of 10.9 and 17.0 cm with a standard deviation of 0.27 and 0.75 cm, respectively.

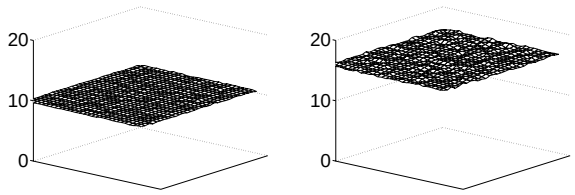

Figure 8: Illustrated are the recovered range maps computed using optical aperture size differentiation for a pair of frontal-parallel surfaces at a distance of 11 and 17 cm from the camera. The computed range maps have a mean of 11.0 and 17.0 cm with a standard deviation of 0.06 and 0.16 cm, respectively.

5.4 Sensitivity

As with most other techniques, the inherent sensitivity of our method of range estimation is dependent on the basic rules of triangulation. In particular, from classical binocular stereo we know that range is inversely proportional to disparity: $Z = \frac{db}{D}$, where d is the distance between lens and sensor, b is the distance between the stereo camera pair (i.e., the "baseline"), and D is the measured disparity. The sensitivity in estimating range with respect to errors in the disparity estimate can be determined by differentiating with respect to this parameter:

$$\left|\frac{\partial Z}{\partial D}\right| = \left|\frac{db}{D^2}\right| \propto \frac{Z^2}{db}. \quad (32)$$

In our system, α plays the role of disparity, and effective baseline is proportional to lens diameter

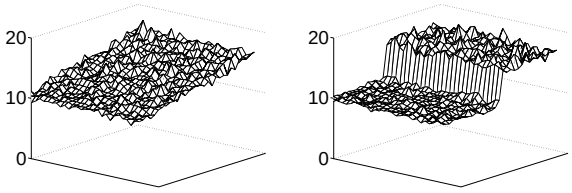

Figure 9: Illustrated on the left is the recovered range map computed using optical viewpoint differentiation for a slanted surface oriented approximately 30 degrees relative to the sensor plane with the center of the plane at a depth of 14 cm. Illustrated on the right is the recovered range map for a pair of occluding surfaces at a depth of 11 and 17 cm.

and dependent on the choice of optical masks.

In addition, errors in estimating α will be proportional to α^2. More specifically, we consider the effects of additive noise in the differential measurements:

$$\begin{aligned} \hat{\alpha} &= -\frac{I_v + \Delta_v}{I_x + \Delta_x} \\ &= -\frac{\frac{1}{\alpha}M' + \Delta_v}{\frac{1}{\alpha^2}M' + \Delta_x} \\ &= -\frac{\alpha + \frac{\alpha^2 \Delta_v}{M'}}{1 + \frac{\alpha^2 \Delta_x}{M'}}. \end{aligned} \tag{33}$$

Expanding the denominator as a polynomial series in Δ_x, the leading error terms of the full expression are seen to scale as the square of α. Rewriting in terms of depth (using Equation (6)) and incoporating with the above, we expect errors in the differential measurements to produce range errors of the form:

$$\Delta Z \propto \frac{Z^2}{db}\left(1 - \frac{d}{f} + \frac{d}{Z}\right)^2 \Delta_{\{v,x\}}. \tag{34}$$

That is, measurement errors lead to range errors that scale as the square of the distance from the focal plane.

6 Discussion

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, involving only a few 1D convolutions and arithmetic operations.

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. There are some disadvantages as well. Most notably, the construction of a non-standard imaging system, and the limited range accuracy due to the small effective baseline.

A counterintuitive aspect of our technique is that it relies on the defocus of the image. In particular, a perfectly focused image corresponds to $\alpha = 0$, leading to a singularity in Equation (10). We have partially overcome this problem by imposing a prior density on α that biases solutions toward the focal plane. But in general, accuracy will be best for surfaces outside of the focal plane.

The results presented here can be improved in a number of ways. A better mask design, which includes the effects of the PSF of the camera optics and optimizes light transmittance while satisfying the desired derivative relationship could have a large effect on the quality of the estimator. In the proposed camera, a pair of images are acquired in a temporally interleaved fashion, so that motion in the scene will be misinterpreted as false range information. A more sophisticated algorithm should be developed that compensates for any inter-frame motion. Alternatively, the technique could be modified to measure the two images simultaneously

(using a beamsplitter, as in [8]). Finally, as with all intensity-based range imaging approaches, the results may be improved by illuminating the scene with structured light.

Acknowledgments

This research was performed while HF was in the GRASP Laboratory at the University of Pennsylvania, where he was supported by ARO DAAH04-96-1-0007, DARPA N00014-92-J-1647, and NSF SBR89-20230. HF is currently at MIT where he is supported by NIH Grant EY11005-04 and MURI Grant N00014-95-1-0699. This research was performed while EPS was in the GRASP Laboratory at the University of Pennsylvania, where he was partially supported by ARO / MURI DAAH04-96-1-0007. EPS is currently at NYU, where he is partially supported by NSF CAREER grant 9624855. Portions of this work have appeared in [19, 20, 21, 16].

References

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

[2] B.D. Lucas and T. Kanade. 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.

[3] E.P. Simoncelli, E.H. Adelson, and D.J. Heeger. Probability distributions of optical flow. In *Proc Conf on Computer Vision and Pattern Recognition*, pages 310–315, Mauii, Hawaii, June 1991. IEEE Computer Society.

[4] A.P. Pentland. A new sense for depth of field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(4):523–531, 1987.

[5] M. Subbarao. Parallel depth recovery by changing camera parameters. In *Proceedings of the International Conference on Computer Vision*, pages 149–155, Tampa, FL, 1988. IEEE, New York, NY.

[6] Y. Xiong and S. Shafer. Depth from focusing and defocusing. In *Proceedings of the DARPA Image Understanding Workshop*, pages 967–976, New York, NY, 1993. IEEE, Piscataway, NJ.

[7] A.M. Subbarao and G. Surya. Depth from defocus: A spatial domain approach. *International Journal of Computer Vision*, 13(3):271–294, 1994.

[8] S.K. Nayar, M. Watanabe, and M. Noguchi. Real-time focus range sensor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(12):1186–1198, 1995.

[9] M. Watanabe and S. Nayar. Minimal operator set for passive depth from defocus. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 431–438, San Francisco, CA, 1996. IEEE, Los Alasmitos, CA.

[10] E.R. Dowski and W.T. Cathey. Single-lens single-image incoherent passive-ranging systems. *Applied Optics*, 33(29):6762–6773, 1994.

[11] D.G. Jones and D.G. Lamb. 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.

[12] E.H. Adelson and J.Y.A. Wang. Single lens stereo with a plenoptic camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):99–106, 1992.

[13] W. Teoh and X.D. Zhang. An inexpensive stereoscopic vision system for robots. In *International Conference on Robotics*, pages 186–189, 1984.

[14] Y. Nishimoto and Y. Shirai. A feature-based stereo model using small disparities. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 192–196, Tokyo, 1987.

[15] A. Goshtasby and W.A. Gruver. Design of a single-lens stereo camera system. *Pattern Recognition*, 26(6):923–937, 1993.

[16] H. Farid. *Range Estimation by Optical Differentiation*. PhD thesis, Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA., 1997.

[17] R.W. Floyd and L. Steinberg. An adaptive algorithm for spatial grey scale. *Proceedings of the Society for Information Display*, 17(2):75–77, 1976.

[18] H. Farid and E.P. Simoncelli. Optimally rotation-equivariant directional derivative kernels. In *Computer Analysis of Images and Patterns*, Kiel, Germany, 1997.

[19] E.P. Simoncelli and H. Farid. Direct differential range estimation from aperture derivatives. In *Proceedings of the European Conference on Computer Vision*, pages 82–93 (volume II), Cambridge, England, 1996.

[20] E.P. Simoncelli and H. Farid. Single-lens range imaging using optical derivative masks, 1995. U.S. Patent Pending (filed 14 Nov 1995). International Patent Pending (filed 13 Nov 1996).

[21] H. Farid and E.P. Simoncelli. A differential optical range camera. In *Proceedings of the Annual Meeting of the Optical Society of America*, Rochester, NY, 1996.