Are Textureless Scenes Recoverable?

Hari Sundaram   Shree Nayar

Dept. of Computer Science
Columbia University
New York New York 10027.
E-mail: {hari,nayar}@cs.columbia.edu

Abstract

It is widely accepted that textureless surfaces cannot be recovered using passive sensing techniques. The problem is approached by viewing image formation as a fully three-dimensional mapping. It is shown that the lens encodes structural information of the scene within a compact three-dimensional space behind it. After analyzing the information content of this space and by using its properties we derive necessary and sufficient conditions for the recovery of textureless scenes. Based on these conditions, a simple procedure for recovering textureless scenes is described. We experimentally demonstrate the recovery of three textureless surfaces, namely, a line, a plane, and a paraboloid. Since textureless surfaces represent the worst case recovery scenario, all the results and the recovery procedure are naturally applicable to scenes with texture.

1. Introduction

It is now widely accepted that passive vision techniques do not permit the depth recovery of textureless scenes. Methods such as stereo [Barnard and Fischler 82] and structure from motion [Ullman 79] rely on correspondence between points in different images. This process is known to be impossible in the absence of scene texture. Techniques based on depth from defocus [Pentland 87], [Subbarao and Surya 94], [Watanabe and Nayar 96] and depth from focus [Krotkov 87], [Nayar and Nagakawa 94] use the relative defocus in two or more images taken with different optical settings to compute depth. Both techniques make important assumptions about the presence of texture in the scene. Algorithms for shape from texture [Kender 78], [Stevens 79], as the name itself implies, rely explicitly on the presence of strong texture. The only approach that provides a partial solution to the problem of textureless surfaces is shape from shading [Horn and Brooks 89]. However, it is well-known that shape from shading requires the use of strong assumptions to make the recovery problem tractable.

So, is absolute depth recovery of a textureless scene truly impossible? To study this problem, we return to the first principles of image formation. All brightness images used in computational vision are measurements obtained in a compact three-dimensional space behind the imaging lens. We refer to this space as the *monocular visual space* (MVS). An image is but a two-dimensional slice (frontal, non-frontal, or even curved) through this space. Rather than asking how much more than what is already known can be recovered from an image, we explore what information about the three-dimensional scene is captured within the monocular space. Similar studies have been conducted earlier in the context of optical microscopy [Streibl 85]. However, the object illumination, the imaging optics, and the sensing methods differ significantly from those of a conventional imaging system used in computational vision.

If we image the world using a pinhole camera, the MVS is completely determined by any one slice (image) through it, since all images in the MVS are simply projective transformations of each other. It turns out that the lens used in all imaging systems to gather more light encodes the MVS with three-dimensional structural information of the scene. Based on this observation, we raise a series of questions: (a) Is all the information contained in the scene present in the MVS? Is there any information lost in the mapping process? (b) How must the MVS be sampled to ensure that the discrete images obtained preserve the information within the MVS? (c) Under what conditions are textureless scenes recoverable from the MVS?

Our investigations have resulted in quantitative answers to all of the above questions. In addition, it has lead us to a simple procedure for recovering the absolute depths of textureless three-dimensional surfaces in the scene. As examples, we experimentally demonstrate the recovery of textureless lines, planes, and conics. These results give us a good indication as to how the proposed method will behave with respect to arbitrary scenes. It is important to note that, since textureless scenes represent a worst case scenario, our results and methods can always be used to determine absolute depth maps of scenes with texture. We conclude by summarizing the important properties of the monocular visual space.

2. The Monocular Visual Space

Today, practically all vision and image processing algorithms are applied to brightness images. What additional information is present in the MVS, that is not captured by one or two brightness images?

To answer this question, we review the process of image formation, paying careful attention to the creation of the MVS rather than a single image. The resulting model of the MVS, will allow us to explore its information content and develop algorithms for scene recovery that go beyond existing ones.

2.1. The Optical System
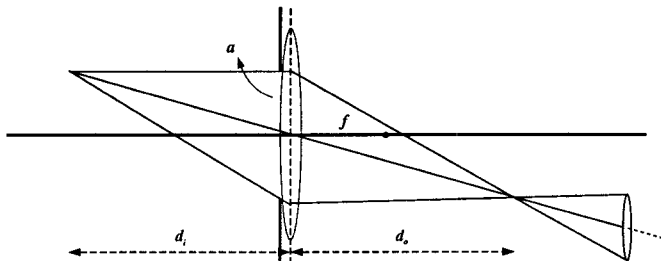
Figure 1. In a conventional imaging system, each scene point produces a light-cone. The orientation as well as the half-angle of the light-cone varies with the location of the scene point.

Traditionally, one views the lens system as transforming each point in the scene to a single focussed point behind the lens. As is well known, a simple relation exists for thin lenses, in the form of the lens law:

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f},$$ (1)

where, d_i is the distance of the scene point from the principal plane, d_o is the distance of the focussed point to the principal plane and f is the focal length of the lens. The scene therefore produces a focussed surface behind the lens, in accordance with equation 1.

A more insightful approach is to consider the following: Each point in the scene, which is not occluded, produces a double cone of light behind the lens. This double cone, which we call a *light-cone*, has its apex at the focussed image point. Now, consider the entire volume of light behind the lens. This volume is a linear superposition of light-cones due to the points in the scene.

In a conventional optical system like the one shown in Figure 1, the orientation as well as the half-angle of the light-cone varies with the location of the object point in front of the lens. Clearly, this is a space-variant system with a three-dimensional point spread function (the light-cone).

A simple modification to the imaging process renders the resultant optical system to be linear and shift-invariant. This system, which is shown in Figure 2 is telecentric on the image side of the lens [Kingslake 83], [Watanabe and Nayar 96].

The only difference with respect to Figure 1 is that an aperture has now been placed in the *front focal plane*, i.e at a distance f from the optical center of the lens. In contrast to the conventional system, the axis of each light-cone is *parallel* to the optical axis. Also, the half angle θ of the
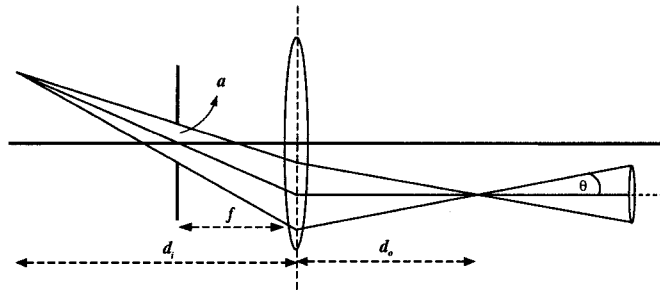
Figure 2. A telecentric system is obtained by placing a small aperture at the front focal plane. In this case, the axis of the light-cone is parallel to the optical axis for any scene point. The half-angle θ of the cone, is a system constant determined by the f-number of the system, $f_o = \frac{f}{a}$.

cone is a constant and is independent of the location of the scene point. The half angle is an intrinsic property of the system, dependent only on $f_o = \frac{f}{a}$, the f-number of the system:

$$\tan(\theta) = \frac{a}{2f}.$$ (2)

The telecentric system is linear and shift-invariant and is therefore, completely characterized by its impulse response, $h(x, y, z)$. Hence, the volume of light $v(x, y, z)$ behind the lens, can be expressed as a simple convolution between the focussed surface $o(x, y, z)$ and the system impulse response:

$$v(x, y, z) = h(x, y, z) * o(x, y, z).$$ (3)

As is well known, in a linear shift-invariant system, the Fourier transform of a convolution in the spatial domain is just the product of the corresponding Fourier transforms. Hence, equation 3 can easily be solved for $o(x, y, z)$ by inversion in the Fourier domain. Telecentricity is not a limiting assumption on our analysis, because, there exists an elegant relationship between telecentric systems and conventional lens systems [Sitter and Rhodes 90], [Sundaram and Nayar 96]. Hence all our results with telecentric systems are relevant to conventional systems as well.

3. The Light Cone

The telecentric system, being linear shift-invariant, is completely characterized by its impulse response, the light cone. The optical transfer function (OTF) of the impulse response imposes fundamental restrictions on the system resolution as well the accuracy of the recovered scene. Hence, an accurate model of the light cone is essential.

3.1. Diffraction Limited Optics

The geometric optics model[1] for the light cone is a first-order approximation to the effects of diffraction. The nature

---

[1] Details on the geometrical optics model of the light cone can be found in [Sundaram and Nayar 96].

of the OTF in the case of diffraction limited systems has been studied earlier [Born and Wolf 70], [Streibl 84a].

Consider a telecentric system, like the one in Figure 2, with a finite, circularly symmetric aperture a and focal length f. We assume the incoming light to be quasi-monochromatic and partially coherent. Taking diffraction into account, we get a telecentric system with a *diffraction limited* aperture $A(\rho)$, which is bandlimited:

$$A(\rho) = \left\{ \begin{array}{ll} 1 & \rho \leq \frac{2\pi\gamma}{\lambda}, \\ 0 & elsewhere, \end{array} \right. \quad (4)$$

where, $\gamma \equiv \frac{a}{2f}$, $\rho \equiv \sqrt{\omega_x^2 + \omega_y^2}$.

If we make an additional assumption that the incoming rays are paraxial, then the OTF for such a system is given by:

$$H(\rho, \omega_z) = \left\{ \begin{array}{ll} \frac{4\pi\gamma}{\lambda^2\rho}\sqrt{1 - \zeta^2} & \zeta \leq 1, \\ 0 & elsewhere, \end{array} \right. \quad (5)$$

where $\zeta \equiv (\frac{\lambda\rho}{4\pi\gamma} + \frac{|\omega_z|}{\gamma\rho})$.

The OTF contains a δ function at the origin. The OTF (see Figure 3) is also bandlimited along ρ as well as ω_z with $\omega_{z\_max} \propto \frac{1}{f_o^2}$, $\rho_{max} \propto \frac{1}{f_o}$. We define the surface $\zeta = 1$ to be the bounding surface of the OTF, since $H(\rho, \omega_z)$ is zero whenever $\zeta \geq 1$.
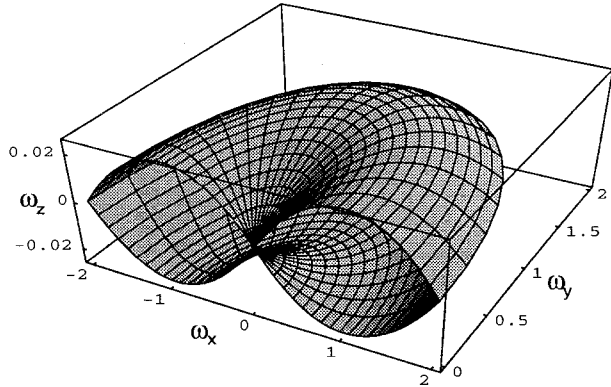
Figure 3. Plot of the bounding surface $\zeta = 1$ in the diffraction limited case; the OTF is non zero only within the torus. The bandwidth is limited by the f-number: $f_{z\_max} \propto \frac{1}{f_o^2}$, $\rho_{max} \propto \frac{1}{f_o}$. The unit length is 200π kHz along ω_x, ω_y, and ω_z.

4. Restrictions on Resolution

The light cone, being bandlimited, also bandlimits the monocular visual space. As a result, there are fundamental restrictions on the number of independent measurements within the MVS due to the Nyquist criterion for bandlimited signals. Nyquist's criterion establishes the minimum sampling frequency for a bandlimited signal in order for it to be

reconstructed exactly. It is this critical frequency (which is twice the bandwidth), that defines the resolution (distance between independent measurements) within the MVS.

It can be easily shown using equation 5 that the light cone is bandlimited along ρ with $\rho_{max} = \frac{4\pi\gamma}{\lambda}$. This establishes the lateral Nyquist sampling frequency, $\rho_s = 2\rho_{max}$. Then, the lateral resolution of the system is given by:

$$\delta x = \delta y = \frac{1}{\rho_s} = \frac{\lambda}{8\pi\gamma}. \quad (6)$$

Similarly, one can show that the system is bandlimited along ω_z with $\omega_{z\_max} = \frac{2\pi\gamma^2}{2\lambda}$. Hence, the Nyquist sampling rate in the longitudinal direction is given by $\omega_{z\_s} = 2\omega_{z\_max}$ and the longitudinal resolution of the system is given by:

$$\delta z = \frac{1}{\omega_{z\_s}} = \frac{\lambda}{2\pi\gamma^2}. \quad (7)$$

As noted earlier, the f-number limits the bandwidth of the light cone. It affects the system resolution as: $\delta x = \delta y \propto f_o$ and $\delta z \propto f_o^2$. An increase in the f-number will *decrease* the resolution.

5. The Road to Recovery

Now we are in a position to determine conditions wherein textureless scenes become recoverable. As was the case with resolution, the OTF of the light-cone plays a crucial role in determining conditions on the recoverability of textureless scenes. What is interesting is that the recoverability of a textureless surface turns out to be critically dependent upon the *shape* of the OTF.

5.1. Conjugate Spectra

Before we determine conditions on recoverability, we need a few definitions. Due to telecentricity and linear shift-invariance:

$$V = H \cdot O, \quad (8)$$

where, V, H and O are, respectively, the Fourier transforms of the MVS, the light-cone and the focussed surface.

A spectrum O is defined to be a *conjugate-spectrum* conjugate to H if it satisfies the following equation:

$$H(\omega_x, \omega_y, \omega_z) \, O(\omega_x, \omega_y, \omega_z) = k\delta(\omega_x, \omega_y, \omega_z), \quad (9)$$

where k is an arbitrary constant. The support of O is said to occupy the *conjugate-space*. Equation 9 essentially states that the supports of the Fourier transforms of H and O only intersect at the origin[2].

5.2. Scene Recovery

A way of dealing with the problem of recovery of textureless scenes is to first define a class of textureless scenes which are *irrecoverable*. Textureless scenes which do not

---

[2]The focussed surface being an intensity (positive) distribution, must have a d.c component in its spectrum. The δ function in equation 9 is due to the δ function in the Fourier transform H.

belong to this class are therefore recoverable. Intuitively, a surface is irrecoverable (in the sense of least squares) if the MVS contains no information about it. This can be readily formalized by using the idea of conjugate-spectra.

Consider an arbitrary focussed surface $o(x, y, z)$. A necessary and sufficient condition for it to be irrecoverable is that its spectrum O must be conjugate[3] to that of the OTF H.

Let us consider an arbitrary focussed surface $z = f(x, y)$ with the condition that there be no discontinuities[4]. This implies that the scene must have no occlusions or depth discontinuities. If either of them are present in the scene, they will produce discontinuities in the focussed surface $o(x, y, z)$.

The Fourier transform of the surface is given by the Fourier transform of $\delta(z - f(x, y))$. Hence,

$$O(\vec{\omega}) = \iiint \delta(z - f(x,y)) e^{-j(\vec{\omega} \cdot \vec{x})} dv \qquad (10)$$

$$= \iint e^{-j(\omega_z f(x,y))} e^{-j(\omega_x x + \omega_y y)} dx \, dy, \qquad (11)$$

where, $\vec{\omega} = (\omega_x, \omega_y, \omega_z)$, $\vec{x} = (x, y, z)$ and $dv = dxdydz$. We now examine families of surfaces for various shapes and determine conditions on their recoverability.

5.2.1. Lines and Planes

Consider an arbitrary plane through the origin $ax + by + cz = 0$. It can be easily shown using equation 10, that the Fourier transform of the plane is:

$$O(\omega_x, \omega_y, \omega_z) = \delta(\omega_x - \omega_z \frac{a}{c}, \omega_y - \omega_z \frac{b}{c}). \qquad (12)$$

The δ function is only defined on the straight line

$$\frac{\omega_x}{\cos(\alpha)} = \frac{\omega_y}{\cos(\beta)} = \frac{\omega_z}{\cos(\phi)} = l, \qquad (13)$$

where, $\cos(\alpha)$, $\cos(\beta)$ and $\cos(\phi)$ are the direction cosines[5] of the plane. This is illustrated in the Figure 4.

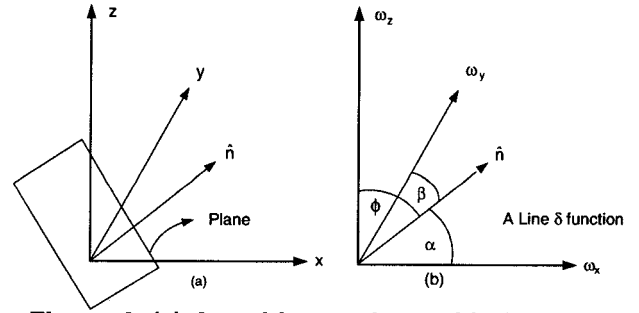Similarly, if we take the line $\frac{x}{\cos(\alpha)} = \frac{y}{\cos(\beta)} = \frac{z}{\cos(\phi)} = l$, then its Fourier transform is a plane δ function given by:

$$O(\omega_x, \omega_y, \omega_z) = \delta(\omega_x \frac{\cos(\alpha)}{\cos(\phi)} + \omega_y \frac{\cos(\beta)}{\cos(\phi)} + \omega_z). \qquad (14)$$

Figure 4. (a) An arbitrary plane with the normal vector $\hat{n}$ $(\cos(\alpha), \cos(\beta), \cos(\phi))$. (b) The Fourier transform of the plane is a line δ function with the same direction cosines as the normal to the plane.

This is simply the equation of a plane $\omega_x \cos(\alpha) + \omega_y \cos(\beta) + \omega_z \cos(\phi) = 0$. Hence the plane and the line are Fourier transform pairs.

Now, for a plane to be irrecoverable its Fourier transform must be conjugate to the OTF. A sufficient condition for this is

$$\phi \leq \pi/2 - \theta, \qquad (15)$$

where, $\tan(\theta) = \frac{a}{2f}$. This can be easily seen to be true by comparing figures 3 and 4. In fact, equation 15 shows that the line δ function must lie within the surface[6] $|\omega_z| = \rho \tan(\theta)$. Now, when equation 15 is *not* satisfied, i.e when $\phi > \pi/2 - \theta$, the line delta function will intersect the OTF and hence the plane will be recoverable in the least squares sense.

Like the plane, the Fourier transform of the line must be conjugate to the OTF for it to be irrecoverable. A sufficient condition is:

$$\phi \geq \pi/2 - \theta. \qquad (16)$$

Now, when this condition is not satisfied, i.e $\phi < \pi/2 - \theta$ the Fourier transform of the line will intersect the OTF, making it recoverable.

It is a widely accepted conjecture in computer vision that textureless planes parallel to the camera produce no information. Equation 15 is the quantitative explanation why this conjecture is indeed correct. Furthermore it shows the existence of a family of textureless planes which are irrecoverable. What is more intriguing is the fact that textureless planes become recoverable when the conjugacy-condition (equation 9) is not satisfied, a surprising result!

5.2.2. Cones and Quadratics

Let us look at more complex objects such as cones and higher degree polynomials. The equation of an infinite cone is $z = r \cot(\alpha)$, where α is the half angle of the cone. The Fourier transform of the cone is given by:

$$O(\omega_x, \omega_y, \omega_z) = 2j\pi \frac{|\omega_z| \tan^2(\alpha)}{(\rho^2 \tan^2(\alpha) - \omega_z^2)^{3/2}}. \qquad (17)$$

---

[3]We are not guaranteed an irrecoverable surface by taking the Fourier inverse of an arbitrary spectrum X conjugate to H, since both two-dimensional surfaces as well as three-dimensional densities have three-dimensional Fourier transforms. However, we can easily determine if a surface is irrecoverable, by computing its Fourier transform and then using equation 9.

[4]This is an assumption on the surface $z = f(x, y)$ made to simplify analysis. It has no bearing on the validity of equation 9, which is valid for arbitrary scenes.

[5]The direction cosines are components of the unit normal vector defining the plane. They are defined to be $\cos(\alpha) = \frac{a}{\sqrt{a^2+b^2+c^2}}$, $\cos(\beta) = \frac{b}{\sqrt{a^2+b^2+c^2}}$ and $\cos(\phi) = \frac{c}{\sqrt{a^2+b^2+c^2}}$.

[6]This is the bounding surface in the geometrical optics case [Sundaram and Nayar 96]. The conjugate space for the diffraction limited case, being bigger than the corresponding space for geometric optics, will contain this surface.

817

Figure 5 shows the cone and its Fourier transform. From
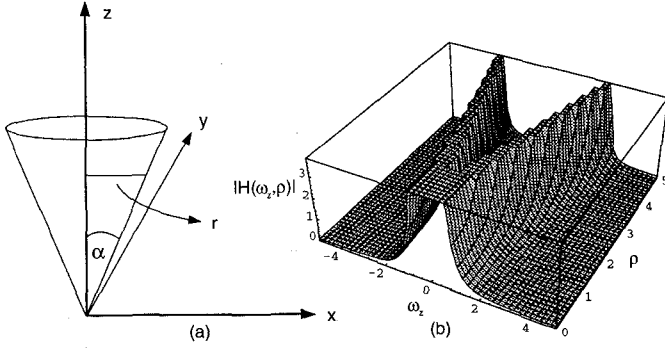
Figure 5. (a) An infinite cone with half angle α. (b) OTF of the cone with $\tan(\alpha) = 1/2$. Note that the energy is concentrated around the double cone $|\omega_z| = \rho \tan(\alpha)$. ω_z and ρ are in Hz.

Figure 5 it is seen that the energy in the Fourier transform is concentrated along a double cone $|\omega_z| = \rho \tan(\alpha)$ oriented along the ω_z axis with half-angle $\frac{\pi}{2} - \alpha$.

Here, the conjugacy condition is simply:

$$\alpha \geq \theta. \tag{18}$$

For the cone, unlike the line or the plane, there is no condition under which the Fourier transform becomes conjugate to the OTF. Hence, equation 18 is a weak conjugacy condition. The cone is partially recoverable under all circumstances. The recovery becomes significantly better when $\alpha < \theta$, since most of the energy lies on the cone $|\omega_z| = \rho \tan(\alpha)$.

We now consider paraboloids $z = ar^2$ where a is an arbitrary constant. It can be shown that the Fourier transform of the paraboloid is given by:

$$O(\omega_x, \omega_y, \omega_z) = \frac{-j\pi}{|a\omega_z|} \exp\left(\frac{j\rho^2}{|a\omega_z|}\right). \tag{19}$$

This is illustrated in Figure 6. Notice how the magnitude of the Fourier transform is independent of the radial frequency ρ.

As was with the cone, the paraboloid is partially recoverable in all cases. There is no case in which the Fourier transform is conjugate to the OTF. However, in the case of the parabola, the recovery improves significantly when $\phi < \theta$.

The Fourier transforms of higher powers of r i.e $z = ar^n$ can be derived using equation 10 but they are increasingly complex functions. In the case of most other surfaces, no analytical results exist for the Fourier transforms.

Textureless scenes are therefore completely recoverable, under the condition that their Fourier transforms lie within the support of the OTF. When the Fourier transform overlaps both the conjugate space as well as the support of the OTF, the scene is then only recoverable in the least squares sense.
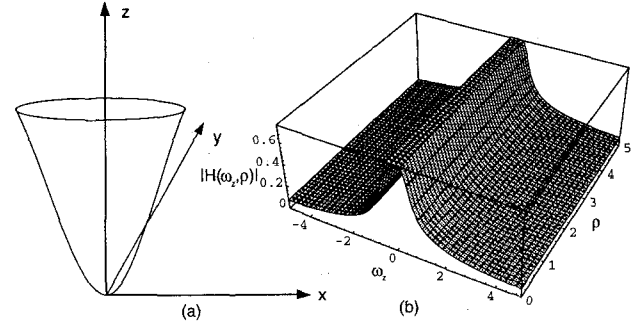
Figure 6. (a) A paraboloid. (b) The magnitude of the Fourier transform, which is seen to be independent of ρ. ω_z and ρ are in Hz.

5.3. Necessary Conditions on Shape

Since the MVS is bandlimited, the textureless scenes recovered from it cannot be arbitrary because the reconstructed focussed surface $o(x_i, y_i, z_i)$ must be bandlimited as well. In this section we shall see how the finite bandwidth of the MVS affects the gradient of the scene.

Now, the OTF H bandlimits the recovered focussed surface to $(-\omega_{x_o}, \omega_{x_o})$, $(-\omega_{y_o}, \omega_{y_o})$ and $(-\omega_{z_o}, \omega_{z_o})$ along ω_x, ω_y and ω_z respectively. Then,

$$o(\vec{x_i}) = \frac{1}{(2\pi)^3} \int_{-\omega_{z_o}}^{\omega_{z_o}} \int_{-\omega_{x_o}}^{\omega_{x_o}} \int_{-\omega_{y_o}}^{\omega_{y_o}} O(\vec{\omega}e^{-j(\vec{\omega}\cdot\vec{x_i})} d\omega, \tag{20}$$

where $d\omega = d\omega_x d\omega_y d\omega_z$. The scene $s(x_o, y_o, z_o)$ and the focussed surface $o(x_i, y_i, z_i)$ are related via a non-linear transformation. The transformation equations are $x_o = \frac{-z_o x_i}{z_i}, y_o = \frac{-z_o y_i}{z_i}, z_o = \frac{-z_i f}{z_i - f}$. Due to the finite bandwidth of the system, the gradient of the recovered focussed surface $o(x_i, y_i, z_i)$ will be bounded uniformly in the MVS by a constant. However, since the scene $s(x_o, y_o, z_o)$ is related via a non-linear transformation to $o(x_i, y_i, z_i)$, its gradient will be bounded by a function that varies with depth. Now, it is easily shown from equation 20 and the transformation equations that:

$$\left|\frac{d}{dz_i} o(\vec{x_i})\right| \leq C\sqrt{\omega_{x_o}\omega_{y_o}\omega_{z_o}^3}, \tag{21}$$

$$\left|\frac{d}{dz_i} o(\vec{x_i})\right| = |\nabla s \cdot \vec{t}|, \tag{22}$$

where C is a fixed positive constant, $\vec{t} = \frac{x_o(z_o - f)}{z_o f}\hat{x} + \frac{y_o(z_o - f)}{z_o f}\hat{y} + \frac{(z_o - f)^2}{f^2}\hat{z}$. ∇s is the gradient of the scene and $\cdot$ denotes the usual vector dot product. Combining equations 22 and 21 we get,

$$|\nabla s \cdot \vec{t}| \leq C\sqrt{\omega_{x_o}\omega_{y_o}\omega_{z_o}^3}. \tag{23}$$

This is a necessary condition on the gradient of the scene. Using the familiar Cauchy-Schwartz inequality, we can get

a sufficient condition on ∇s, the scene gradient, so that equation 23 holds:

$$|\nabla s| \leq \frac{C\sqrt{\omega_{x_o}\omega_{y_o}\omega_{z_o}^3}}{\left(\frac{z_o-f}{f}\right)^2\sqrt{1+\left(\frac{f}{z_o}\right)^2\tan^2(\theta)}}. \quad (24)$$
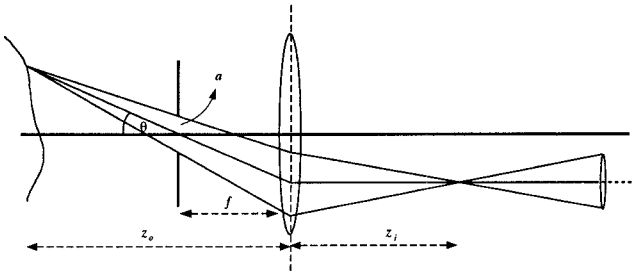
Figure 7. An arbitrary surface being imaged.

The upper bound on $|\nabla s|$ becomes loose as the surface approaches the front focal plane. This is easily understood as points near the front focal plane $z = f$, on the object side of the lens, will get transformed to distant points on the image side of the lens (i.e the shape will get stretched upon transformation). Therefore, large values in the surface gradient are permissible. But as z_o becomes large, the surface becomes compressed near f behind the lens. Hence, the gradient can only take small values in order to be faithfully captured by the system. Although equation 23 has been derived under the assumption that the scene is textureless, it holds true for any bandlimited surface with texture[7].

6. Recovered Scenes: Experiments

We now present results for three textureless surfaces: a plane, a line and a paraboloid (see Figures 8-10). In each case, the captured data was deconvolved in the Fourier domain using the diffraction-limited filter (see equation 5), and the final shape extracted after post processing of the deconvolved output. If we use a standard three-dimensional interpolating function (i.e the sinc function) to interpolate in the MVS, one obtains a continuous surface. Since the camera has been calibrated, we can map each computed point in the MVS to the the external world using the lens law.

We used an 8-bit Sony XC-77 CCD camera with $25mm$ lens in conjunction with a J300 frame grabber and a DEC Alpha workstation to grab the images. Each image was of size 640x480 pixels. A 3mm aperture was placed at the front focal plane of the lens, converting it into a telecentric system. The CCD sensor inside the camera was attached to a calibrated micrometer stage having a 10 micron accuracy. The nominal value of λ was set to be $500nm$.

We need to establish the sampling rates along the x, y and z directions to ensure that we have captured the MVS accurately. The sampling rates are easily obtained from equations 6 and 7. Then, the Nyquist rates $\rho_s/2\pi$ and $\omega_{z\_s}/2\pi$ are 480kHz and 7.20kHz respectively.

Since the pixel locations within the CCD array are fixed, the sampling frequency along the x and y directions too are fixed. This turns out to be much less than $\rho_s/2\pi$, but this does not cause aliasing artifacts to appear because the *input* spectrum is extremely low pass[8] (being textureless). We are free to determine the sampling rate along the z direction as we have a precise, movable sensor. We thus sample the MVS along z obtaining 64 samples in all at a frequency slightly above $\omega_{z\_s}/2\pi$.

(a)

(b)   (c)
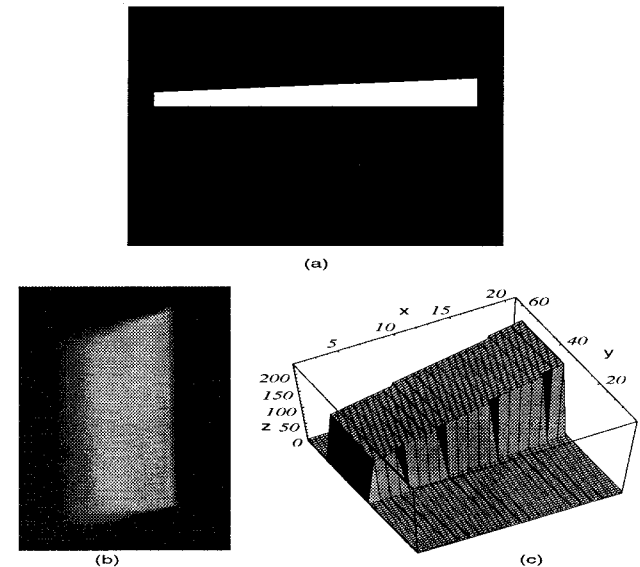
Figure 8. (a) A finite white plane is used as the input. (b) One of the samples (image) of the MVS exhibiting defocus. (c) The recovered plane in the MVS, after deconvolution and post-processing. Using a standard interpolating function on this data (e.g sinc) will yield a continuous plane. Grid spacing: $\Delta z = 26\mu m$, $\Delta x = 11\mu m$ and $\Delta y = 104\mu m$.

7. Summary

Traditional passive depth recovery methods such as stereo, structure from motion, depth from focus and defocus make strong assumptions about the presence of texture in the scene. Hence all these methods fail when attempting to recover textureless scenes. Shape from shading algorithms provide a partial solution but are limited by other assumptions.

We have studied the problem of textureless scene recovery by formulating image formation as a fully three-dimensional mapping characterized by the monocular visual space. We have shown how an imaging lens encodes structural information about the scene within the MVS. We also derived conditions needed to completely represent the MVS as a set of discrete images.

---

[7]If the bandwidth of the surface containing texture is less than that of the MVS, one simply changes the integration limits in equation 20.

[8]In the general case, when the scene contains texture, we must have a high sampling rate in the sensor in order to avoid aliasing problems.
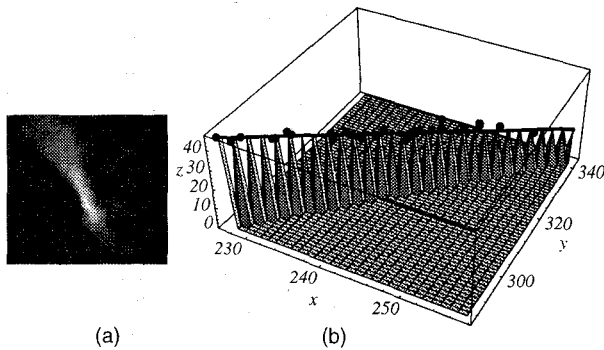
(a)   (b)

Figure 9. (a) One of the samples of the MVS in the case of the line. We generated a finite line by using a thread stretched in front of the camera. (b) The recovered line in the MVS, after deconvolution and post-processing. The spikes represent the least squares fitted data, while the bold dots represent the actual raw data. The bold line represents the least squares fit. Grid spacing: $\Delta z = 130\mu m$, $\Delta x = 11\mu m$ and $\Delta y = 13\mu m$.
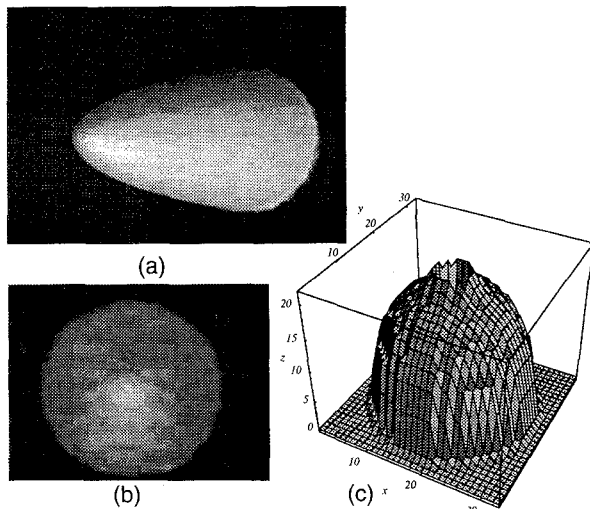

(a)

(b)   (c)

Figure 10. (a) A profile of the paraboloid used in the experiment. (b) One of the sample images of the MVS. (c) The recovered parabola after deconvolution and post-processing. Note that the tip of parabola is missing. This is because the tip is actually is a small planar patch parallel to the camera. This is irrecoverable (see subsection 5.2.1). As with the line and the plane, a standard interpolating function (e.g sinc) will give back a continuous paraboloid. Grid spacing: $\Delta z = 130\mu m$, $\Delta x = 208\mu m$ and $\Delta y = 208\mu m$.

We showed how the shape of the OTF of the MVS plays a critical role in the recovery of textureless scenes through a conjugacy condition. As specific examples, conditions for recovery of textureless planes, lines, cones and paraboloids were described. In addition, we derived a necessary condition on the gradient of the recovered scene that arises due to the finite bandwidth of the MVS.

Finally, we demonstrated the implications of our theory via experiments on a few textureless surfaces: a plane, a line and a paraboloid. It is worth reiterating that, since texture always aids in recovering depth, all the conditions that need to be satisfied by textureless surfaces are only significantly relaxed in the presence of texture.

References

[Barnard and Fischler 82] S. Barnard and M. Fischler *Computational Stereo*, ACM Surveys, Vol. 14 No. 4, 1982.

[Born and Wolf 70] Born,M. and Wolf,E., 1970 *The Principles of Optics* , 4e, Permagon Press.

[Dhond and Aggarwal 89] U. Dhond and J. K. Aggrawal *Structure from Stereo - A Review* IEEE Transactions on Systems, Man and Cybernetics, Vol. 19, No.6, 1489-510, 1989.

[Ens and Lawrence 91] J. Ens and P. Lawrence, *A Matrix Based Method for Determining Depth from Focus* Proc. CVPR, 600-609, 1991.

[Horn 88] B.K.P Horn *Robot Vision* The MIT Press, MIT, Cambridge MA-02142.

[Horn and Brooks 89] B.K.P Horn and M.J. Brooks *Shape from Shading* The MIT Press, MIT, Cambridge MA-02142.

[Kender 78] J.R Kender, *Shape from Texture: a brief overview and a new aggregation transform.* Proc. DARPA IU workshop, 79-84, 1978.

[Kingslake 83] R. Kingslake *Optical System Design.* Academic Press, 1983.

[Krotkov 87] E. Krotkov *Focusing* IJCV, Vol. 1, 223-237, 1987.

[Nayar and Nagakawa 94] S.K. Nayar, Y. Nagakawa *Shape from Focus* IEEE Trans. on PAMI Vol. 16, No. 8, 824-831, 1994.

[Pentland 87] A. P. Pentland *A New Sense of Depth of Field* IEEE Trans. on PAMI Vol. 9, No. 4, 523-531, 1987.

[Sitter and Rhodes 90] D. N. Sitter, W. T. Rhodes *Three dimensional Imaging : a space invariant model for space variant systems* Applied Optics, Vol. 29, No. 26, 3789-3794, 1990.

[Stevens 79] K.A Stevens *Representing and analyzing surface orientation* In Artificial Intelligence: An MIT Perspective, P.H Winston and R.H Brown (Eds.). The MIT Press, 1979, Cambridge MA.

[Streibl 84a] N. Streibl *Depth Transfer by an Imaging System* Optica Acta, Vol. 31, No. 11 1233-1241, 1984.

[Streibl 84b] N. Streibl *Fundamental restrictions for 3-D light distributions* Optik Vol. 66 No. 4, 341-354, 1984.

[Streibl 85] N. Streibl *Three-dimensional Imaging by a microscope* J. Optical Society of America, Vol. 2, No. 2, 121-127, 1985.

[Subbarao and Surya 94] M. Subbarao, G. Surya *Depth from Defocus: A Spatial Domain Approach* IJCV Vol. 13, 271-294, 1994.

[Sundaram and Nayar 96] H. Sundaram and S.K. Nayar, *On the recovery of textureless surfaces*, Tech Report, Columbia University, 1996.

[Ullman 79] S. Ullman *The Interpretation of Visual Motion* The MIT Press, 1979, MIT, Cambridge MA-02142.

[Watanabe and Nayar 96] M. Watanabe, S.K. Nayar *Minimal Operator Set for Passive Depth from Defocus*, Proc. CVPR, 431-438, 1996.

[Watanabe and Nayar 96] M. Watanabe, S.K. Nayar *Telecentric Optics for Computational Vision* Proc. ECCV, Vol. 2, 439-451, 1996.