Observing Shape from Defocused Images*

PAOLO FAVARO

Electrical Engineering Department, Washington University in St. Louis, One Brookings Drive, St. Louis, MO 63130

fava@ee.wustl.edu

ANDREA MENNUCCI

Scuola Normale Superiore, Piazza Dei Cavalieri 7, Pisa 56100, Italy

mennucci@tonelli.sns.it

STEFANO SOATTO

Computer Science Department, University of California, Los Angeles, 405 Hilgard Avenue, Los Angeles, CA 90095

soatto@ucla.edu

Abstract. Accommodation cues are measurable properties of an image that are associated with a change in the geometry of the imaging device. To what extent can three-dimensional shape be reconstructed using accommodation cues alone? This question is fundamental to the problem of reconstructing shape from focus (SFF) and shape from defocus (SFD) for applications in inspection, microscopy, image restoration and visualization. We address it by studying the "observability" of accommodation cues in an analytical framework that reveals under what conditions shape can be reconstructed from defocused images. We do so in three steps: (1) we characterize the observability of any surface in the presence of a controlled radiance ("weak observability"), (2) we conjecture the existence of a radiance that allows distinguishing any two surfaces ("sufficient excitation") and (3) we show that in the absence of any prior knowledge on the radiance, two surfaces can be distinguished up to the degree of resolution determined by the complexity of the radiance ("strong observability"). We formulate the problem of reconstructing the shape and radiance of a scene as the minimization of the *information divergence* between blurred images, and propose an algorithm that is provably convergent and guarantees that the solution is admissible, in the sense of corresponding to a positive radiance and imaging kernel.

Keywords: shape, surface geometry, low-level vision, shape from defocus, shape from focus

1. Introduction

An imaging system, such as the eye or a video-camera, involves a map from the three-dimensional environment onto a two-dimensional surface. One can attempt

to retrieve the spatial information lost in the imaging process by relying on prior assumptions on the scene using pictorial information such as shading, texture, cast shadows, edge blur etc. All pictorial cues are intrinsically ambiguous in that the prior assumptions cannot be validated: given a photograph, it is always possible to construct (infinitely many) different three-dimensional scenes that produce that same image.

As an alternative to relying on prior assumptions, one can try to retrieve spatial information by looking at different images of the same scene taken with different imaging devices. Measurable properties of images which are associated with a changing viewpoint are called "parallax" cues (for instance stereo and motion).[1]

In addition to changing the position of the imaging device, one could change its *geometry*. For instance, one can take different photographs of the same scene with different lens apertures or focal lengths. Similarly, in the eye one can change the shape of the lens by acting on the lens muscles. *We call "accommodation cues" all measurable properties of images that are associated with a change of the geometry of the imaging device.*

The study of accommodation raises a number of questions. Is accommodation a visual cue (i.e. does it carry information about the shape of the scene)? What are the conditions under which two different scenes can be distinguished, if at all? Do such conditions depend upon the particular imaging device? If two scenes are distinguishable, is there an algorithm that provably distinguishes them? How does the human visual system make use of accommodation? Is it possible to render the accommodation cue, so as to create a "controlled illusion" in the same way photographs do for pictorial cues? In this manuscript we intend to answer some of these questions rigorously for an idealized imaging model, and test the resulting algorithms on realistic data sets.

1.1. Relation to Previous Work

Depth from defocus can be formalized as a "blind deconvolution" of certain integral operators, a problem common to many research fields. The same mathematical formulation appears in inverse filtering, in speech analysis, in image restoration (corrupted by noise or optical aberrations), in source separation, in inverse scattering, in computer tomography. These are problems studied in a number of fields such as signal processing, information theory, communication theory and computer vision. Consequently, this paper relates to a large body of literature.

Solving blind deconvolution can be posed as the minimization of a discrepancy measure between the input data and the model that generates the data. However, this approach inevitably leads to multiple solutions due to the ill-posedness of the blind deconvolution problem. A common procedure used to guarantee desirable prop-

erties of the solution and its uniqueness, is to add *regularization* terms. Most regularization schemes impose smoothness constraints that are not appropriate for all scenes. For example, Chan and Wong (1998) propose a total variation regularization that is effective in recovering edges of images. Alternatively, regularization with wavelets is explored by Kalifa et al. (1998), Namba and Ishida (1998), and Neelamani et al. (1999).

In the computer vision community, blind deconvolution is known as passive depth from defocus (e.g. Hwang et al., 1989; Schechner et al., 1994, 1998; Subbarao and Surya, 1994; Subbarao and Wei, 1992), which is different from active depth from defocus (Girod and Scherock, 1989; Nayar et al., 1995; Noguchi and Nayar, 1994; Scherock, 1981) where structured light is used. While in the passive technique both the radiance (or "texture" or "deblurred image") and the depth are unknown, in the active case the radiance is given as a pattern projected onto the scene.

The literature on active/passive depth from defocus can be divided into two main groups: one poses the problem within a statistical framework (Rajagopalan and Chaudhuri, 1995, 1997, 1998; Schechner and Kiryati, 1999), and one as a deterministic optimization (Pentland, 1997; Pentland et al., 1989, 1994; Hopkins, 1955). In the statistical approach, one choice is to use Markov random fields to describe both the images and the space-variant point spread function (see Section 1.2). This eliminates effectively the need for windowing, allowing for precise depth and radiance estimation also where the equifocal assumption is not fully satisfied. Other methods (Farid and Simoncelli, 1998; Simoncelli and Farid, 1996) work on differential variations in the image intensities using masks or exploiting aperture changes. In the deterministic approach, we find that many algorithms formulate the problem in the frequency domain. The advantages of such a representation are overshadowed by inaccuracies due to windowing effects, edge bleeding, feature shifts, image noise and field curvature (see Ens and Lawrence, 1993; Nair and Stewart, 1992). A solution is to use highly specialized filters that operate in narrow bands. The main problem is that narrowband filters have to be defined on a large support; therefore, the amount of computation grows considerably. Xiong and Shafer (1995) propose a set of 240 moment filters that select narrow bands of the radiance, and allow for precise depth estimation. Gokstorp (1994) uses a multiresolution local frequency representation of the input image pair. Many algorithms

operate in the spatial domain (see Watanabe and Nayar, 1996a; Nayar and Nakagawa, 1990; Noguchi and Nayar, 1994; Favaro and Soatto, 2000). One approach involves *rational* filters tuned for restricted frequency bands. For the active depth from defocus, optimal patterns are derived and projected onto the scene according to the chosen point spread function. When the optical point spread function is modeled with a 2D Laplacian, the best pattern is a checker board. A number of other spatial domain algorithms exist (see Subbarao and Surya, 1994; Ziou, 1998; Asada et al., 1998; Marshall et al., 1996) where the scene is approximated by a planar surface, or the analysis is concentrated on step edges, line edges, occluding edges and junctions (see Marshall et al., 1996; Asada et al., 1998).

In our paper we study depth from defocus from a deterministic point of view. We pose the problem in a mathematical framework, and derive conditions for a unique reconstruction of both radiance and surface. In particular, we give precise definitions of observability of radiance and shape for active and passive depth from defocus (see Sections 2 and 4), and identify which radiances are "unobservable" and to what degree reconstruction of radiance and shape is possible (see Section 3). A similar study, but restricted to finite impulse response filters, has been done in Harikumar and Bresler (1999).

In addition to answering observability issues, we propose a locally optimal algorithm which borrows mainly from blind deconvolution techniques (Kundur and Hatzinakos, 1998; Snyder et al., 1992). We choose as criterion the minimization of the information divergence (I-divergence) between blurred images, motivated by the work of Csiszár (1991). The algorithm we propose is iterative, and we give a proof of its convergence to a (local) minimum and test its performance on both real and simulated images.

1.2. Notation and Ideal Formalization of the Problem[2]

In order to introduce some of the concepts that we will address in this paper, consider a scene whose three-dimensional shape is represented by S and that emits energy with a radiance r. While we will make the meaning of S and r precise soon, an intuitive grasp suffices for now.

The image of the scene can be represented as a function with support on a compact subset $\Omega \subset \mathbb{R}^2$ of the imaging surface (e.g. the retina or the CCD sensor).

The brightness at a particular point is constrained to be a *positive*[3] value that depends on the radiance r, the surface S and the optics of the imaging device. Define the vector $\mathbf{u} \in \mathcal{U}$ where $\mathcal{U}$ is the set of parameters (for instance focal length, optical center, lens aperture etc.) that describe the geometry of the imaging device. We denote the image of the scene as $I_{\mathbf{u}}^S(\cdot, r)$ to emphasize its dependence on the radiance r, the surface S and the parameters $\mathbf{u}$ of the optics.

The shape S is represented as a piecewise smooth function. We consider the radiance[4] r to be a positive integrable function defined on S, and we also assume that the scene is populated by Lambertian[5] objects. Given these assumptions, it is possible to find a change of coordinates so that both the radiance r and the surface S are defined on the domain Ω.

Based on these assumptions, a model of the imaging process that is suitable to study the accommodation cue is given by integral equations of the form

$$I_{\mathbf{u}}^S(\mathbf{y}, r) = \int_{\Omega} h_{\mathbf{u}}^S(\mathbf{y}, \mathbf{x}) r(\mathbf{x}) \, d\mathbf{x} \qquad (1)$$

where $\mathbf{y} \in \Omega$ are the image plane coordinates, and $\mathbf{x} \in \Omega$ are the scene coordinates. The kernel h has two different interpretations: for any fixed $\mathbf{x} \in \Omega$, the function $\mathbf{y} \mapsto h_{\mathbf{u}}^S(\mathbf{y}, \mathbf{x})$ is the *point-spread function* of the optics; dually, for any fixed $\mathbf{y} \in \Omega$, the function $\mathbf{x} \mapsto h_{\mathbf{u}}^S(\mathbf{y}, \mathbf{x})$ is a function that has the normalization property

$$\int_{\mathbb{R}^2} h_{\mathbf{u}}^S(\mathbf{y}, \mathbf{x}) \, d\mathbf{x} = 1. \qquad (2)$$

Both functions can be either a delta measure or a bounded function (that is continuous inside a support, but possibly discontinuous across its border); they depend on $\mathbf{u} \in \mathcal{U}$, the camera parameters, and on the surface S.

Given the characteristics of the kernel, we note that, by duality, r belongs to $\mathcal{L}_{loc}(\mathbb{R}^2)$, which includes locally integrable positive functions as well as delta measures. This is why we call r the *radiance distribution*, or *energy distribution*.[6]

Since the model Eq. (1) is common to most of the literature on shape from defocus, we will not further motivate it here, and adopt it for analysis in the following sections (Chaudhuri and Rajagopalan (1999) for details).

In some particular cases the surfaces may be described by some parameters $\mathbf{s}$. For example, suppose that S consists of a single slanted plane: the scene S is then completely described by the intercept with the optical axis S_Z, and by a reduced normal $\tilde{\mathbf{n}}$, so that

the surface S can be represented by the parameters $\mathbf{s} = (S_Z, \tilde{\mathbf{n}}) \in \mathbb{R}^3$. In particular, in Appendix C we will study the case of *equifocal planes*. We will consider the (ideal) model of a thin lens camera that is imaging an equifocal plane. In this case the kernel $h_{\mathbf{u}}^S(\mathbf{y}, \mathbf{x})$ takes the form of a pillbox kernel. The scene is completely described by the depth $\mathbf{s} = z$ of the imaged plane, and the parameters $\mathbf{u}$ encode the geometry of the camera. The same surface representation will be used locally at each point in Section 6. This will considerably simplify the implementation of the proposed algorithm for estimating depth from defocus.

In Eq. (1), we assume that the values $I_{\mathbf{u}}^S(\mathbf{y}, r)$ can be measured for each $\mathbf{y} \in \Omega$ and control parameter $\mathbf{u} \in \mathcal{U}$. The energy distribution, or radiance, r is unknown, and $h_{\mathbf{u}}^S(\mathbf{y}, \mathbf{x})$ is known only up to the surface S, which is unknown. The scope of this paper can be summarized into the following question:

Question 1.   To what extent and under what conditions can the radiance r and the surface S be reconstructed from measurements of defocused images $I_{\mathbf{u}}^S(\mathbf{y}, r)$?

The next few sections are devoted to answering the above question. We proceed in increasing order of generality, from assuming that the radiance r can be chosen purposefully (Section 2) to an arbitrary unknown variation (Section 4). Along the way, we point out some issues concerning the hypothesis on the radiance in the design of algorithms for reconstructing shape from focus/defocus (Section 3).

2.   Weak Observability

The concept of "weak observability", which we are about to define, is relevant to the problem of shape reconstruction from active usage of the control $\mathbf{u}$ and the radiance pattern r.

Consider two scenes with surfaces S and S' respectively.

Definition 1.   We say that a surface S is weakly indistinguishable from a surface S' if for all possible $r \in \mathcal{L}_{loc}(\mathbb{R}^2)$ there exists at least a radiance $r' \in \mathcal{L}_{loc}(\mathbb{R}^2)$ such that we have

$$I_{\mathbf{u}}^S(\mathbf{y}, r) = I_{\mathbf{u}}^{S'}(\mathbf{y}, r') \quad \forall \mathbf{y} \in \Omega \quad \forall \mathbf{u} \in \mathcal{U}. \quad (3)$$

Two surfaces are weakly distinguishable if they are not weakly indistinguishable. If a surface S is *weakly distinguishable* from any other surface, we say that it is *weakly observable*.

Piecewise smooth surfaces may have discontinuities (which form a set of measure zero). We do not distinguish between surfaces that differ on a set of measure zero.

The purpose of this section is to establish that piecewise smooth surfaces are weakly observable. In order to prove this statement we need to make explicit assumptions on the imaging system. In particular,

(1) For each scalar $z > 0$, and for each $\mathbf{y} \in \Omega$, there exists a setting $\bar{\mathbf{u}}$, which we call *focus setting*, and an open set $O \subset \Omega$, such that $\mathbf{y} \in O$ and for any surface S smooth in O with $S(\mathbf{y}) = z$, we have

$$h_{\bar{\mathbf{u}}}^S(\mathbf{y}, \mathbf{x}) = \delta(\mathbf{y} - \mathbf{x}) \quad \forall \mathbf{x} \in O$$

(2) Furthermore, given a scalar $z > 0$ and a $\mathbf{y} \in \Omega$, such a focus setting $\bar{\mathbf{u}}$ is unique in $\mathcal{U}$.

(3) For any surface S and for any open $O' \subset \Omega$ such that S is smooth in O', we have

$$h_{\mathbf{u}}^S(\mathbf{y}, \mathbf{x}) > 0 \quad \forall \mathbf{x}, \mathbf{y} \in O'$$

whenever $\mathbf{u}$ is not a focus setting.

The above statements formalize some of the properties of typical real aperture cameras. The first statement corresponds to guaranteeing the existence of a control $\mathbf{u}$ at each point $\mathbf{y}$ such that the resulting image of that point is in focus. Notice that the control $\mathbf{u}$ may change at each point $\mathbf{y}$ depending on the scene's surface. The second property states that such a control is unique. Finally, when the kernel is not a Dirac delta function, it is a strictly positive function over an open set.

We wish to emphasize that the existence of a focus setting is a mathematical idealization. In practice, diffraction and other optical effects prevent a kernel from approaching a delta distribution. Nevertheless, analysis based on an idealized model can shed light on the design of algorithms operating on real data. Under the above conditions we can state the following result:

Proposition 1.   *Any piecewise smooth surface S is weakly observable.*

Proof: See Appendix A.   □

Proposition 1 shows that it is possible—in principle—to distinguish the shape of a surface from that of any other surface by looking at images under

different camera settings **u** and radiances r. This, however, requires the active usage of the control **u** and of the radiance r. While it may be possible to change r in an active vision context, this is not the case in a general setting.

The definition of weak observability leaves the freedom to choose the radiance r to distinguish S from S'. As we have seen in the proposition above, this is indeed sufficient to render any piecewise smooth surface observable. However, a natural question is left open:

Question 2. Can a radiance r be found that allows distinguishing S from *any* other surface?

We address this issue in the next subsection.

2.1. Excitation

Let $\mathcal{I}(S \mid r)$ denote the set of surfaces that cannot be distinguished from S *given* the energy distribution r:

$$\mathcal{I}(S \mid r) = \left\{ \tilde{S} \mid I_{\mathbf{u}}^{\tilde{S}}(\mathbf{y}, r) = I_{\mathbf{u}}^{S}(\mathbf{y}, r) \ \forall \, \mathbf{y} \in \Omega, \mathbf{u} \in \mathcal{U} \right\} \tag{4}$$

Note that we are using the *same* radiance on both surfaces $S, \tilde{S}$. This is the case, for instance, when using *structured light*, so that r is a (static) pattern projected onto a surface. Clearly, not all radiances allow distinguishing two surfaces. For instance, $r = 0$ does not allow distinguishing any surface. We now discuss the existence of a "sufficiently exciting" radiance.

Definition 2. We say that a distribution r is *sufficiently exciting* for S if

$$\mathcal{I}(S \mid r) = \{S\}. \tag{5}$$

We conjecture that distributions with unbounded variation on any open subset of the image are sufficiently exciting for any surface, i.e. "universally exciting." A motivation for this conjecture can be extrapolated from the discussion relative to $\mathcal{L}_{loc}^2(\mathbb{R}^2)$ that follows in Section 3, as we will see in Remark 1.

3. Physically Realizable Radiances, Resolution and Harmonic Components

Energy distributions with unbounded variation cannot be physically realized. Therefore, in this section we are interested in introducing a notion of "bandwidth" on the space of energy distributions, which would allow

us to model it as $\mathcal{L}^2(\mathbb{R}^2)$. However, such a restriction is legitimate only if the radiance r does not have any harmonic component, which we will define shortly. Before doing so, we remind the reader of the basic properties of harmonic functions.

Definition 3. A function $r : \mathbb{R}^2 \mapsto \mathbb{R}$ is said to be *harmonic* in an open region $O \subset \mathbb{R}^2$ if

$$\Delta r(\mathbf{x}) \doteq \sum_{i=1}^{2} \frac{\partial^2}{\partial \mathbf{x}_i^2} r(\mathbf{x}) = 0$$

for all $\mathbf{x} \in O$.

Proposition 2 (*mean-value property*). *If $r : \mathbb{R}^2 \mapsto \mathbb{R}$ is harmonic, then, for any integrable function $F : \mathbb{R}^2 \mapsto \mathbb{R}$ that is rotationally symmetric,*

$$\int F(\mathbf{y} - \mathbf{x}) r(\mathbf{x}) \, d\mathbf{x} = r(\mathbf{y}) \int F(\mathbf{x}) \, d\mathbf{x}$$

whenever the integrals exist.

Proof: See Appendix B.   □

Corollary 1. *In general $\int h_{\mathbf{u}}^{S}(\mathbf{y}, \mathbf{x}) \, d\mathbf{x} = 1$; when we are imaging an equifocal plane, the kernel $h_{\mathbf{u}}^{S}(\mathbf{y}, \cdot)$ is indeed rotationally symmetric. Hence, the above property leads us to conclude that, in that case, if r is harmonic, then no information can be obtained from the accommodation cue; from Mennucci and Soatto*(1999), *we have that, in many cases, $h_{\mathbf{u}}^{S}(\mathbf{y}, \cdot)$ can be approximated by a circularly symmetric function; if r is a harmonic function, then*

$$\int h_{\mathbf{u}}^{S}(\mathbf{y}, \mathbf{x}) r(\mathbf{x}) \, d\mathbf{x} \cong r(\mathbf{y}) \tag{6}$$

so that we can state that harmonic radiances are negligible as carriers of shape information.

A simple example of a harmonic radiance is $r(\mathbf{x}) = a_0 \mathbf{x} + a_1$, where a_0 and a_1 are constants (i.e. a brightness gradient). It is easy to convince ourselves that when imaging an equifocal plane in real scenes, images generated by such a radiance do not change for any focal setting (i.e. they appear as always perfectly focused).

The above result says that if one decomposes a radiance into the sum of a harmonic function and a square integrable function, the harmonic component can (and

should) be neglected. Therefore, one can restrict the analysis to the non-harmonic component of the radiance, which we represent as a function in the space $\mathcal{L}^2(\mathbb{R}^2)$. This leads naturally to a notion of "bandwidth", as we describe below.

Definition 4. Let $\{\theta_i\}$ be an orthonormal basis of $\mathcal{L}^2(\mathbb{R}^2)$. We say that a radiance r has a *degree of definition* (or *degree of complexity*) k if there exists a set of coefficients α_i, $i = 1, \ldots, k$ such that

$$r(\mathbf{x}) = \sum_{i=1}^{k} \alpha_i \theta_i(\mathbf{x}). \tag{7}$$

When $k < \infty$ we say that the distribution r is *band-limited*.

Note that for practical purposes there is no loss of generality in assuming that $r \in \mathcal{L}^2(\mathbb{R}^2)$, since the energy emitted from a surface is necessarily finite and the definition is necessarily limited by the optics. If we also assume that the kernels $h_{\mathbf{u}}^S \in \mathcal{L}^2(\mathbb{R}^2 \times \mathbb{R}^2)$, we can define a *degree of resolution* for the surface S.

Definition 5. Let $h_{\mathbf{u}}^S \in \mathcal{L}^2(\mathbb{R}^2 \times \mathbb{R}^2)$. If there exists a positive real integer ρ and coefficients β_j, $j = 1, \ldots, \rho$ such that

$$h_{\mathbf{u}}^S(\mathbf{y}, \mathbf{x}) = \sum_{j=1}^{\rho} \beta_j(\mathbf{y}, \mathbf{u}, S)\theta_j(\mathbf{x}) \tag{8}$$

then we say that the surface S has a degree of resolution ρ.

The above definitions of degrees of complexity (or resolution) depend upon the choice of basis of $\mathcal{L}^2(\mathbb{R}^2)$. In the following, we will always assume that the two are defined relative to the same basis. There is a natural link between the degree of complexity of a distribution and the degree of resolution at which two surfaces can be distinguished.

Proposition 3. *Let r be a band-limited distribution with degree of complexity k. Then two surfaces S_1 and S_2 can only be distinguished up to the resolution determined by k, that is, if we write*

$$h_{\mathbf{u}}^S(\mathbf{y}, \mathbf{x}) = \sum_{j=1}^{\infty} \beta_j(\mathbf{y}, \mathbf{u}, S)\theta_j(\mathbf{x}) \tag{9}$$

then

$$\mathcal{I}(S_1 \mid r) \supset \{S_2 \mid \beta_j(\mathbf{y}, \mathbf{u}, S_1) = \beta_j(\mathbf{y}, \mathbf{u}, S_2)$$
$$\forall \, \mathbf{y} \in \Omega, \mathbf{u} \in \mathcal{U}, j = 1, \ldots, k\}. \tag{10}$$

Proof: Substituting the expression of r in terms of the basis θ_i, we have

$$I_{\mathbf{u}}^{S_1}(\mathbf{y}, r) = \sum_{i,j=1}^{k} \beta_j(\mathbf{y}, \mathbf{u}, S_1)\alpha_i \int \theta_j(\mathbf{x})\theta_i(\mathbf{x}) \, d\mathbf{x}$$
$$+ \int \sum_{i,j=k+1}^{\infty} \beta_j(\mathbf{y}, \mathbf{u}, S_1)\alpha_i \theta_i(\mathbf{x})\theta_j(\mathbf{x}) \, d\mathbf{x}. \tag{11}$$

From the orthogonality of the basis elements θ_i we are left with

$$I_{\mathbf{u}}^{S_1}(\mathbf{y}, r) = \sum_{i=1}^{k} \beta_i(\mathbf{y}, \mathbf{u}, S_1) \alpha_i \tag{12}$$

from which we see that, if $\beta_i(\mathbf{y}, \mathbf{u}, S_1) = \beta_i(\mathbf{y}, \mathbf{u}, S_2)$ for all $i = 1, \ldots, k$, then $I_{\mathbf{u}}^{S_1}(\mathbf{y}, r) = I_{\mathbf{u}}^{S_2}(\mathbf{y}, r)$ for all $\mathbf{y}, \mathbf{u}$. $\square$

Remark 1. The practical value of the last proposition is to state that, the more "irregular" the radiance, the more resolving power it has. In the case of structured light, the proposition establishes that irregular patterns should be used in conjunction with accommodation. An "infinitely irregular" pattern should therefore be universally exciting, as conjectured in the previous section.

We note that, as suggested by an anonymous reviewer, the degree of resolution with which two surfaces can be distinguished also depends upon the optics of the imaging system. In our analysis this is taken into account by the degree of complexity of the radiance, that can itself be limited, or it can be limited by the optics of the imaging system.

Besides the fact that delta distributions are a mathematical idealization and cannot be physically realized, it is most often the case that we cannot choose the distribution r at will. Rather, r is a property of the scene being viewed, over which we have no control. In the next section we will consider the observability of a scene depending upon its particular radiance.

4.   Strong Observability

Definition 6.   We say that the pair (S_2, r_2) is indistinguishable from the pair (S_1, r_1) if

$$I_{\mathbf{u}}^{S_1}(\mathbf{y}, r_1) = I_{\mathbf{u}}^{S_2}(\mathbf{y}, r_2) \quad \forall \, \mathbf{y} \in \Omega \quad \forall \, \mathbf{u} \in \mathcal{U}. \quad (13)$$

If the set of pairs that are indistinguishable from (S_1, r_1) is a singleton (i.e. it contains only (S_1, r_1)), we say (S_1, r_1) is *strongly distinguishable*. The following proposition characterizes the set of indistinguishable scenes in terms of their coefficients in the basis $\{\theta_i\}$.

Proposition 4.   *Let r_1 have degree of complexity k. The set of scenes (S_2, r_2) that are indistinguishable from (S_1, r_1) are those for which $r_2 = r_1$ up to a degree of complexity k, and $S_1 = S_2$ up to a resolution $\rho = k$.*

Proof: For (S_2, r_2) to be indistinguishable from (S_1, r_1) we must have

$$\sum_{i,j=1}^{k} \beta_j(\mathbf{y}, \mathbf{u}, S_1)\alpha_{1_i} \int \theta_i(\mathbf{x})\theta_j(\mathbf{x}) \, d\mathbf{x}$$

$$= \sum_{i,j=1}^{\infty} \beta_j(\mathbf{y}, \mathbf{u}, S_2)\alpha_{2_i} \int \theta_i(\mathbf{x})\theta_j(\mathbf{x}) \, d\mathbf{x} \quad (14)$$

for all $\mathbf{y}$ and $\mathbf{u}$. From the orthogonality of $\{\theta_i\}$ and the arbitrariness of $\mathbf{y}, \mathbf{u}$ we conclude that

$$\alpha_{1_i} = \alpha_{2_i} \quad \forall \, i = 1, \dots, k \quad (15)$$

from which we have that $\beta_j(\mathbf{y}, \mathbf{u}, S_1) = \beta_j(\mathbf{y}, \mathbf{u}, S_2)$ for almost all $\mathbf{y}, \mathbf{u}$ and for all $j = 1, \dots, k$, and therefore $S_1 = S_2$ almost everywhere up to the degree of definition k. □

The above proposition is independent of the particular imaging device used, in the sense that the dependency is coded into the coefficients β_i. Any explicit characterization will necessarily be dependent on the particular imaging device used. It is possible to give explicit examples using particular devices, for instance those with kernels described in Mennucci and Soatto (1999), and concentrating on a class of surfaces, for instance slanted planes or occluding boundaries. The calculations are straightforward but messy, and are therefore not reported here.

5.   Optimal Estimation

In this section we formulate the problem of optimally estimating depth from an equifocal imaging model. Recall that in the imaging model (1) all quantities are constrained to be positive: r because it represents the radiant energy (which cannot be negative), $h_{\mathbf{u}}^S$ because it specifies the region of space over which energy is integrated, and $I_{\mathbf{u}}^S(\mathbf{y}, r)$ because it measures the photon count on the surface of the sensor (CCD) corresponding to the pixel $\mathbf{y}$. We are interested in estimating the energy distribution r and the parameters $\mathbf{s}$ that describe the shape S, by measuring a finite number L of images obtained with different camera settings $\mathbf{u}_1, \dots, \mathbf{u}_L$. If we collect the corresponding images $I_{\mathbf{u}_1}^S, \dots, I_{\mathbf{u}_L}^S$ and organize them into a vector $I^S = [I_{\mathbf{u}_1}^S, \dots, I_{\mathbf{u}_L}^S]$ (and so for the kernels $h_{\mathbf{u}_i}^S$, $h^S = [h_{\mathbf{u}_1}^S, \dots, h_{\mathbf{u}_L}^S]$), we can write:

$$I^S(\mathbf{y}, r) = \int_{\Omega} h^S(\mathbf{x}, \mathbf{y}) r(\mathbf{x}) \, d\mathbf{x}. \quad (16)$$

The collection of images $I^S(\cdot, r)$ represents the observations (or measurements) from the scene. Together with it we define a collection of *estimated* images which we generate introducing our current estimates for radiance $\hat{r}$ and surface parameters $\hat{\mathbf{s}}$ (which encode the estimated surface $\hat{S}$) in the above equation. We call such an estimated image $B_{\mathbf{u}}^{\hat{S}}(\cdot, \hat{r})$, and the corresponding collection $B^{\hat{S}} = [B_{\mathbf{u}_1}^{\hat{S}}, \dots, B_{\mathbf{u}_L}^{\hat{S}}]$:

$$B^{\hat{S}}(\mathbf{y}, \hat{r}) \doteq \int_{\Omega} h^{\hat{S}}(\mathbf{x}, \mathbf{y}) \hat{r}(\mathbf{x}) \, d\mathbf{x}. \quad (17)$$

The collection $I^S(\cdot, r)$ is measured on the pixel grid and, hence, its domain Γ (and the corresponding domain of $B^{\hat{S}}(\cdot, \hat{r})$) is a finite lattice $\Gamma = [x_1, \dots, x_N] \times [y_1, \dots, y_M]$ for some integers N and M.

We now want a "criterion" Φ (or cost function) to measure the discrepancy between the collection of measured images $I^S(\cdot, r)$ and the collection of estimated images $B^{\hat{S}}(\cdot, \hat{r})$, so that we can formulate the problem of reconstructing both the radiance and the surface parameters in terms of the minimization of Φ. Common choices of criteria include, for example, the least-squares distance between $I^S(\cdot, r)$ and $B^{\hat{S}}(\cdot, \hat{r})$, or the integral of the absolute value of their difference (total variation) (Chan and Wong, 1998).

The choice of such a criterion is in principle arbitrary as long as the criterion satisfies a basic set of properties (i.e. it is always positive, and 0 if and only if $I^S(\cdot, r)$

and $B^{\hat{S}}(\cdot, \hat{r})$ are identical), and can lead to very different results. We follow Csiszár's approach (Csiszár, 1991) that poses the problem of defining an "optimal" discrepancy function within an axiomatic framework. Csiszár defines a set of desirable, but general, properties that a discrepancy measure should satisfy. He concludes that, when the quantities involved are constrained to be positive[7] (such as in our case), the only consistent choice of criterion is the so-called information divergence, or *I-divergence*, which generalizes the well-known Kullback-Leibler pseudo-metric and is defined as

$$\Phi(I^S(\cdot, r) \parallel B^{\hat{S}}(\cdot, \hat{r})) \doteq \sum_{\mathbf{y} \in \Gamma} \left[I^S(\mathbf{y}, r) \log \frac{I^S(\mathbf{y}, r)}{B^{\hat{S}}(\mathbf{y}, \hat{r})} \right.$$
$$\left. - I^S(\mathbf{y}, r) + B^{\hat{S}}(\mathbf{y}, \hat{r}) \right] \quad (18)$$

where $\log(\cdot)$ indicates the natural logarithm. Notice that the function defined above is always positive and is 0 if and only if $I^S(\cdot, r)$ coincides with $B^{\hat{S}}(\cdot, \hat{r})$. However, the I-divergence is not a true metric as it does not satisfy the triangular inequality.

The I-divergence is defined only for positive functions. We assume $\hat{s}$ represents a positive function (with respect to the chosen parameterization) and $\hat{r}$ is a positive function defined on $\mathbb{R}^2$. Furthermore, we assume that $\hat{s}$ and $\hat{r}$ are such that the estimated image $B^{\hat{S}}(\mathbf{y}, \hat{r})$ (where $\hat{S}$ is the surface generated with the parameters $\hat{s}$) has finite values for any $\mathbf{y} \in \Gamma$. In this case, we say that $\hat{s}$ and $\hat{r}$ are *admissible*.

In order to emphasize the dependency of the cost function Φ on the parameters $\hat{s}$ and the radiance $\hat{r}$, we define

$$\phi(\hat{s}, \hat{r}) = \Phi(I^S(\cdot, r) \parallel B^{\hat{S}}(\cdot, \hat{r})). \quad (19)$$

Therefore, we formulate the problem of simultaneously estimating the shape of a surface encoded by the parameters $\mathbf{s}$ and its radiance r as that of finding $\hat{s}$ and $\hat{r}$ that minimize the I-divergence:

$$\hat{s}, \hat{r} = \arg \min_{\tilde{s}, \tilde{r}} \phi(\tilde{s}, \tilde{r}). \quad (20)$$

5.1. Alternating Minimization

In general, the problem in (20) is nonlinear and infinite-dimensional. Therefore, we concentrate our attention at the outset to (local) iterative schemes that approximate

the optimal solution. To this end, suppose an initial estimate of r is given: r_0. Then, iteratively solving the two following optimization problems

$$\begin{cases} \mathbf{s}_{k+1} \doteq \arg \min_{\tilde{\mathbf{s}}} \phi(\tilde{\mathbf{s}}, r_k) \\ r_{k+1} \doteq \arg \min_{\tilde{r}} \phi(\mathbf{s}_{k+1}, \tilde{r}) \end{cases} \quad (21)$$

leads to the (local) minimization of ϕ, since

$$0 \le \phi(\mathbf{s}_{k+1}, r_{k+1}) \le \phi(\mathbf{s}_{k+1}, r_k) \le \phi(\mathbf{s}_k, r_k). \quad (22)$$

However, solving the two optimization problems in (21) may be an overkill. In order to have the sequence $\{\phi(\mathbf{s}_k, r_k)\}$ converging monotonically it suffices that, at each step, we choose $\mathbf{s}_k$ and r_k in such a way as to guarantee that Eq. (22) holds, that is

$$\begin{cases} \mathbf{s}_{k+1} \mid \phi(\mathbf{s}_{k+1}, \tilde{r}) \le \phi(\mathbf{s}_k, \tilde{r}) & \tilde{r} = r_k \\ r_{k+1} \mid \phi(\tilde{s}, r_{k+1}) \le \phi(\tilde{s}, r_k) & \tilde{s} = \mathbf{s}_{k+1}. \end{cases} \quad (23)$$

The iteration step of the surface parameters $\mathbf{s}_k$ can be realized in a number of ways.

In the next section we will adopt the equifocal assumption, which in general holds only locally and where the surface is smooth. However, the derivation of the minimization algorithm that follows does not depend on this particular choice. We generically indicate the step on the surface parameters as:

$$\mathbf{s}_{k+1} = \arg \min_{\tilde{\mathbf{s}}} \phi(\tilde{\mathbf{s}}, r_k). \quad (24)$$

The second step is obtained from the Kuhn-Tucker conditions (Luemberger, 1968) associated with the problem of minimizing ϕ for fixed $\tilde{\mathbf{s}}$ under positivity constraints for r_{k+1}:

$$\sum_{\mathbf{y} \in \Gamma} \frac{h^{\tilde{S}}(\mathbf{y}, \mathbf{x}) I^S(\mathbf{y}, r)}{\int_\Omega h^{\tilde{S}}(\mathbf{y}, \bar{\mathbf{x}}) r_{k+1}(\bar{\mathbf{x}}) \, d\bar{\mathbf{x}}}$$
$$= \begin{cases} = \sum_{\mathbf{y} \in \Gamma} h^{\tilde{S}}(\mathbf{y}, \mathbf{x}) & \forall \mathbf{x} \mid r_{k+1}(\mathbf{x}) > 0 \\ \le \sum_{\mathbf{y} \in \Gamma} h^{\tilde{S}}(\mathbf{y}, \mathbf{x}) & \forall \mathbf{x} \mid r_{k+1}(\mathbf{x}) = 0. \end{cases} \quad (25)$$

Since such conditions cannot be solved in closed form, we look for an iterative procedure for r_k that will converge to a fixed point. Following Snyder et al. (1992), we choose

$$F_{r_k}^{\tilde{S}}(\mathbf{x}) \doteq \frac{1}{\sum_{\mathbf{y} \in \Gamma} h^{\tilde{S}}(\mathbf{y}, \mathbf{x})} \sum_{\mathbf{y} \in \Gamma} \frac{h^{\tilde{S}}(\mathbf{y}, \mathbf{x}) I^S(\mathbf{y}, r)}{B^{\tilde{S}}(\mathbf{y}, r_k)} \quad (26)$$

and define the following iteration:

$$r_{k+1}(\mathbf{x}) = r_k(\mathbf{x}) F_{r_k}^{\tilde{S}}(\mathbf{x}) \quad \forall \mathbf{x} \in \Omega. \qquad (27)$$

It is important to point out that this iteration decreases the I-divergence ϕ not only when we use the exact kernel h^S, as it is shown in Snyder et al. (1992), but also with any other kernel satisfying the positivity and smoothness constraint. This fact is proven by the following claim.

Proposition 5. *Let r_0 be a non-negative real-valued function defined on $\mathbb{R}^2$, and let the sequence r_k be defined according to (27). Then $\phi(\tilde{\mathbf{s}}, r_{k+1}) \leq \phi(\tilde{\mathbf{s}}, r_k) \forall k > 0$ and for all admissible surface parameters $\tilde{\mathbf{s}}$. Furthermore, equality holds if and only if $r_{k+1} = r_k$.*

Proof: The proof follows Snyder et al. (1992). From the definition of ϕ in Eq. (18) we get

$$\phi(\tilde{\mathbf{s}}, r_{k+1}) - \phi(\tilde{\mathbf{s}}, r_k) = -\sum_{\mathbf{y} \in \Gamma} I^S(\mathbf{y}, r) \log \frac{B^{\tilde{S}}(\mathbf{y}, r_{k+1})}{B^{\tilde{S}}(\mathbf{y}, r_k)} \\ + \sum_{\mathbf{y} \in \Gamma} B^{\tilde{S}}(\mathbf{y}, r_{k+1}) - B^{\tilde{S}}(\mathbf{y}, r_k). \qquad (28)$$

The second sum in the above expression is given by

$$\sum_{\mathbf{y} \in \Gamma} \int_\Omega h^{\tilde{S}}(\mathbf{y}, \mathbf{x}) r_{k+1}(\mathbf{x}) \, d\mathbf{x} - \sum_{\mathbf{y} \in \Gamma} \int_\Omega h^{\tilde{S}}(\mathbf{y}, \mathbf{x}) r_k(\mathbf{x}) \, d\mathbf{x} \\ = \int_\Omega h_0^{\tilde{S}}(\mathbf{x}) r_{k+1}(\mathbf{x}) \, d\mathbf{x} - \int_\Omega h_0^{\tilde{S}}(\mathbf{x}) r_k(\mathbf{x}) \, d\mathbf{x} \qquad (29)$$

where we have defined $h_0^{\tilde{S}}(\mathbf{x}) = \sum_{\mathbf{y} \in \Gamma} h^{\tilde{S}}(\mathbf{y}, \mathbf{x})$, while from the expression of r_{k+1} in (27) we have that the ratio in the first sum is

$$\frac{B^{\tilde{S}}(\mathbf{y}, r_{k+1})}{B^{\tilde{S}}(\mathbf{y}, r_k)} = \int_\Omega F_{r_k}^{\tilde{S}}(\mathbf{x}) \frac{h^{\tilde{S}}(\mathbf{y}, \mathbf{x}) r_k(\mathbf{x})}{B^{\tilde{S}}(\mathbf{y}, r_k)} \, d\mathbf{x}. \qquad (30)$$

We next note that, from Jensen's inequality (Cover and Thomas, 1991),

$$\log \left(\int_\Omega F_{r_k}^{\tilde{S}}(\mathbf{x}) \frac{h^{\tilde{S}}(\mathbf{y}, \mathbf{x}) r_k(\mathbf{x})}{B^{\tilde{S}}(\mathbf{y}, r_k)} \, d\mathbf{x} \right) \\ \geq \int_\Omega \frac{h^{\tilde{S}}(\mathbf{y}, \mathbf{x}) r_k(\mathbf{x})}{B^{\tilde{S}}(\mathbf{y}, r_k)} \log \left(F_{r_k}^{\tilde{S}}(\mathbf{x}) \right) d\mathbf{x} \qquad (31)$$

since the ratio $\frac{h^{\tilde{S}}(\mathbf{y}, \mathbf{x}) r_k(\mathbf{x})}{B^{\tilde{S}}(\mathbf{y}, r_k)}$ can be interpreted as a probability density on Ω (it integrates to 1 on Ω and it is always positive) dependent on the parameters $\tilde{\mathbf{s}}$ and r_k, and therefore the expression in (28) is

$$\phi(\tilde{\mathbf{s}}, r_{k+1}) - \phi(\tilde{\mathbf{s}}, r_k) \\ \leq -\sum_{\mathbf{y} \in \Gamma} I^S(\mathbf{y}, r) \int_\Omega \log \left(F_{r_k}^{\tilde{S}}(\mathbf{x}) \right) \frac{h^{\tilde{S}}(\mathbf{y}, \mathbf{x}) r_k(\mathbf{x})}{B^{\tilde{S}}(\mathbf{y}, r_k)} \, d\mathbf{x} \\ + \int_\Omega h_0^{\tilde{S}}(\mathbf{x}) r_{k+1}(\mathbf{x}) \, d\mathbf{x} - \int_\Omega h_0^{\tilde{S}}(\mathbf{x}) r_k(\mathbf{x}) \, d\mathbf{x} \\ = -\Phi \left(h_0^{\tilde{S}}(\cdot) r_{k+1}(\cdot) \, \| \, h_0^{\tilde{S}}(\cdot) r_k(\cdot) \right) \leq 0 \qquad (32)$$

Notice that the right-hand side of the last expression is still the I-divergence of two positive functions. Therefore, we have

$$\phi(\tilde{\mathbf{s}}, r_{k+1}) - \phi(\tilde{\mathbf{s}}, r_k) \leq 0. \qquad (33)$$

Note that Jensen's inequality becomes an equality if and only if $F_{r_k}^{\tilde{S}}$ is a constant; since the only admissible constant value is 1, because of the normalization constraint, we have $r_{k+1} = r_k$, which concludes the proof. $\qquad \square$

Finally, we can conclude that the proposed algorithm generates a monotonically decreasing sequence of values of the cost function ϕ.

Corollary 2. *Let $\mathbf{s}_0, r_0$ be admissible initial conditions for the sequences $\mathbf{s}_k$ and r_k defined from Eqs. (24) and (27) respectively. The sequence $\phi(\mathbf{s}_k, r_k)$ converges to a limit ϕ^*:*

$$\lim_{k \to \infty} \phi(\mathbf{s}_k, r_k) = \phi^*. \qquad (34)$$

Proof: Follows directly from Eqs. (24), (22) and Proposition 5, together with the fact that the I-divergence is bounded from below by zero. $\qquad \square$

Even if $\phi(\mathbf{s}_k, r_k)$ converges to a limit, it is not necessarily the case that $\mathbf{s}_k$ and r_k do. Whether this happens or not depends on the *observability* of the model (16), which has been analyzed in the previous sections. In Appendix D we compute the Cramér-Rao lower bound considering that the images are affected by Gaussian noise. The Cramér-Rao bound allows us to derive conclusions about the settings that yield the best estimates.

6. Experiments

In order to implement the iterative steps required in the algorithm just described, it is necessary to choose appropriate approximations for the radiance r and the surface parameters $\mathbf{s}$. As mentioned earlier, we adopt the equifocal assumption which is widely used in shape from defocus. However, as discussed above, our algorithm does not depend upon that choice, and other surface models can be considered as well. The equifocal assumption holds locally. Therefore, we consider each point in the images to be independent of each other, and restrict our attention to a window of fixed size around it.

6.1. Implementation

A necessary condition to start the iterative algorithm is to provide an initial admissible radiance r_0. We choose it to be equal to one of the two input images (i.e. as if it were generated by a plane in focus). This choice is guaranteed to be admissible since the image is positive and of finite values. Also, it is important to determine the appropriate local domain $\mathcal{W} \subset \mathbb{R}^2$ of the radiance r where the equifocal assumption provides a meaningful approximation of the local surface. The fact that we use an equifocal imaging model allows us to use the same reference frame both for the image plane and the surface. However, the image in any given patch is composed of contributions from a region of space possibly bigger than the one corresponding to the patch itself. Thus, we write $I_{\mathbf{u}}^S(\mathbf{y}, r) = \int_{\mathcal{W} \cup \mathcal{W}_0} h_{\mathbf{u}}^S(\mathbf{y}, \mathbf{x}) r(\mathbf{x}) \, d\mathbf{x}$, with $\mathbf{y} \in \Gamma_{\mathcal{W}}$ (the image patch domain corresponding to the local radiance domain $\mathcal{W}$), and where $\mathcal{W}_0$ is the domain corresponding to the contribution to the image patch $\Gamma_{\mathcal{W}}$ outside the region $\mathcal{W}$. The size of $\mathcal{W}_0$ clearly depends on the maximum amount of blurring that affects the patch. Hence, once we have chosen a finite dimensional representation of the regions $\Gamma_{\mathcal{W}}$, $\mathcal{W}$ and $\mathcal{W}_0$, we also have restricted the scenes for which we can reconstruct both the shape and the radiance.

Since the intensity value at each pixel is the (area) integration of all the photons hitting a single CCD cell, the smallest blurring radius is limited by the physical size of the CCD cell. This means that, when an image is in focus, we can at best retrieve a radiance defined on a grid whose unit elements depend on the pixel size (see also Murase and Nayar, 1995). Furthermore, diffraction and similar phenomena contribute to increase the minimum blurring radius. This motivates us to reduce the finite (local) representation of the radiance to a grid

whose coordinates correspond to the grid coordinates of the image domain. Hence, we will speak of window size in terms of pixels both for images and radiance, since there is a one-to-one correspondence between them.

We run experiments both on real and synthetic images. In the synthetic data set the overlapping effect of the contribution of the radiance from $\mathcal{W}_0$ is not considered since the images are generated according to the correct model. In the data set of real images provided to us by S.K. Nayar, the equifocal planes are at 529 mm and 869 mm, with a maximum blurring radius of 2.3 pixels. We found experimentally that a good tradeoff between accuracy and robustness in choosing $\mathcal{W}$ is to use a 7×7 pixels window. Hence, the total integration domain for each patch $\mathcal{W} \cup \mathcal{W}_0$ is of 13×13 pixels. In the second data set of real images the maximum blurring radius amounts to 5 pixels. In addition, the minimization step on the surface parameters includes a regularization term. This term guarantees that the recovered surface remains smooth while minimizing the cost function.

6.2. Experiments with Synthetic Images

In this set of experiments, we investigate the robustness of the algorithm to noise. The radiance has been generated as a random pattern. The surface is an equifocal plane and the focal settings are chosen such as to generate blurring radii not bigger than 2 pixels. We consider additive Gaussian noise with a variance that ranges from 1% to 10% of the radiance intensity magnitude (see Fig. 1), which guarantees that the positivity constraint is still satisfied with high probability; however, we manually impose the positivity of the radiance (by taking its absolute value), since we need the radiance to be admissible for the algorithm to converge. We run 50 experiments changing the radiance being imaged, for each of the 10 noise levels. Of the generated noisy image pairs we consider patches of size 7×7 pixels. Smaller patches result in greater sensitivity to noise, while larger ones challenge the equifocal approximation. The results of the computed depths are summarized in Fig. 1. We iterate the algorithm 5 times at each point.

As can be seen, the algorithm is quite robust to additive noise, although if the radiance is not sufficiently exciting (in the sense defined in Section 2.1) it does not converge. This behavior is evident in the experiments with real images described below.
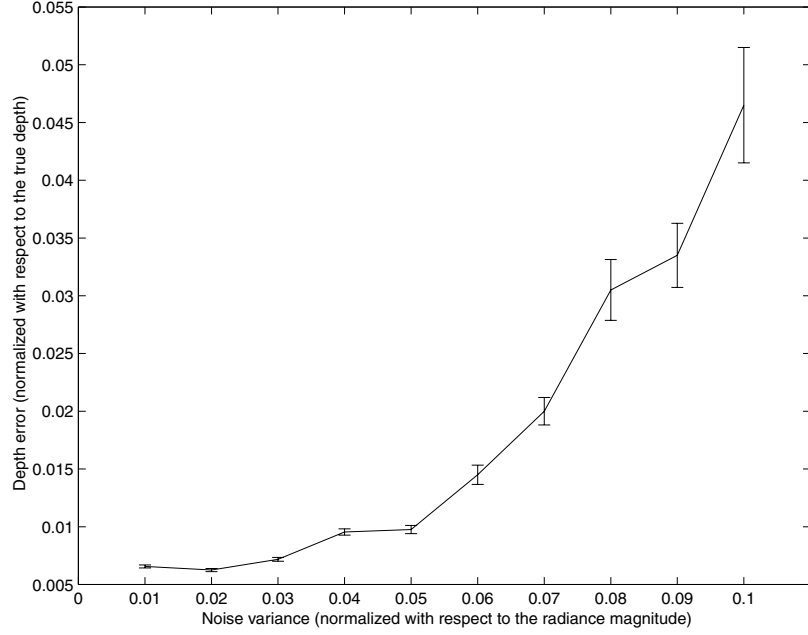
Figure 1. Experiments with synthetic data: mean and standard deviation of the depth error as a function of image noise variance. The mean of the depth error ϵ is normalized with respect to the true depth, i.e. $\epsilon = \frac{\hat{s}-s}{s}$ where $\hat{s}$ is the estimated surface parameter (a single parameter in the case of an equifocal plane) and s is the true surface parameter. The standard deviation of the depth error is also normalized with respect to the true surface parameter s. Similarly, the noise variance is normalized with respect to the radiance magnitude so that the value 0.01 in the abscissa means that the variance is $0.01 * M$ (i.e. 1% of M), where M is the maximum radiance magnitude.

6.3. Experiments with Real Images

We have tested the algorithm on the two images in Fig. 3 provided to us by S.K. Nayar, and on the two images in Fig. 5. These images were generated by a telecentric optical system(see Watanabe and Nayar (1996b) for more details) that compensates for magnification effects usually occurring as the focus settings are changing. A side effect is that the actual lens diameter needs to be adjusted according to the new model shown in Fig. 2. The external aperture is a plane from which a disc of radius a has been removed. The plane is equifocal and placed at a distance f (the focal length) from the lens. When defining the kernel parameters we substitute the true lens diameter with the modified diameter $D = \frac{2aZ_F}{Z_F - f}$, where Z_F is the distance from the lens which is in focus when the focal length is f. Notice that Z_F can be computed through the thin lens law once focal length and the distance between image plane and lens are known.

For this experiment, in order to speed up the computation, we choose to run the algorithm for 5 iterations only. At points where the radiance is not sufficiently exciting, or where the local approximation with an
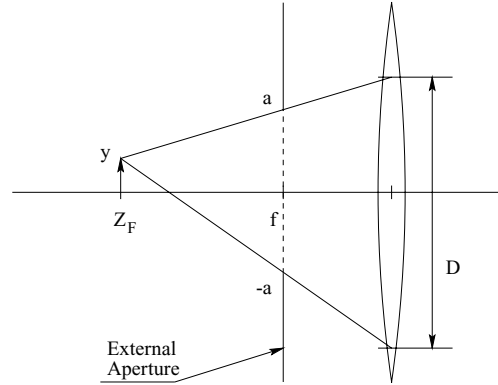

Figure 2. The modified diameter in the telecentric lens model.

equifocal plane is not valid, the algorithm fails to converge. This explains why in Fig. 4 some depth estimates are visibly incorrect. In the second experiment we explicitly impose a global smoothness constraint during the iteration step on the surface parameters and obtain the results shown in Fig. 6. Finally, in Fig. 7 we generate a novel pinhole view of the scene texture-mapping the estimated radiance onto the estimated surface.

Figure 3.   Near-focused original image (left); far-focused original image (right) (courtesy of S.K. Nayar). The difference between the two images is barely perceivable since the two focal planes are only 340 mm apart.
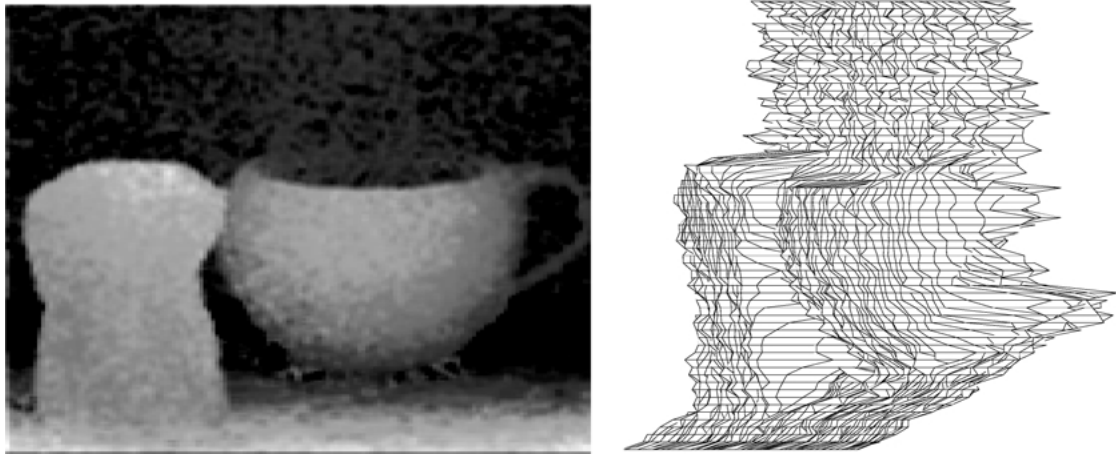

Figure 4.   Reconstructed depth for the scene in Fig. 3. On the left the depth map is rendered as gray levels (lighter corresponds to a smaller depth value), while on the right it is rendered as a smoothed mesh.

Figure 5.   Near-focused original image (left); far-focused original image (right). In this data set the maximum blurring radius is approximately of 5 pixels.
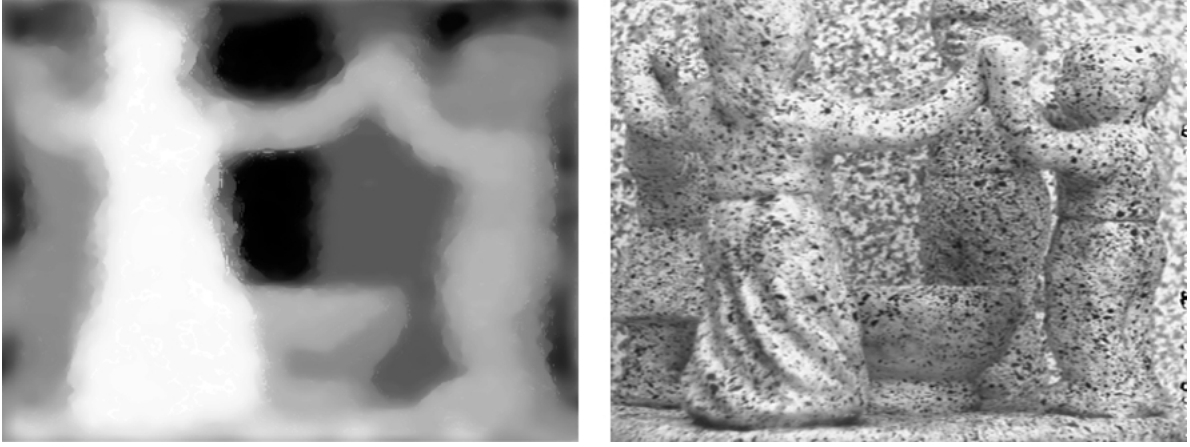
Figure 6. Reconstructed depth for the scene in Fig. 5 coded in grayscale (lighter corresponds to smaller depth values) on the left, and recovered radiance on the right. In this data set the algorithm is challenged with a larger blurring. During the iterative minimization of the I-divergence, a smoothing constraint is also imposed to the surface to regularize the reconstruction process.
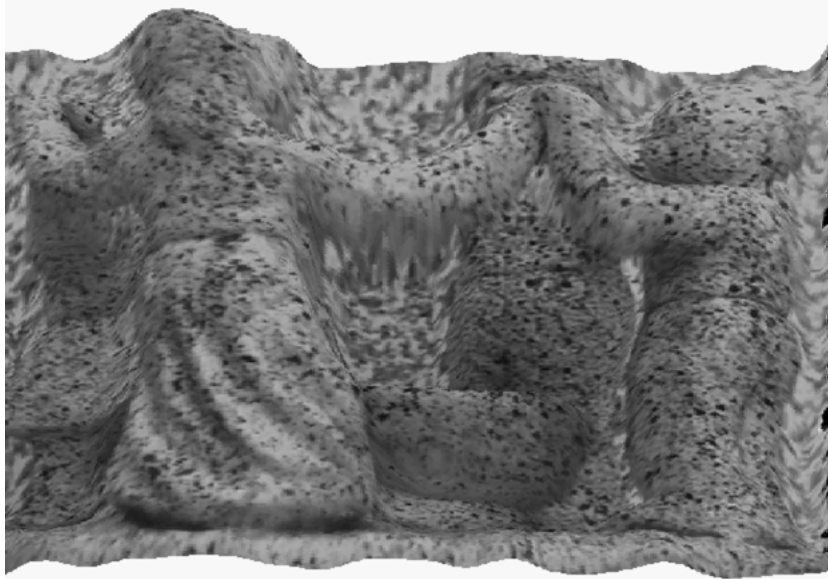

Figure 7. Texture mapping of the reconstructed radiance on the reconstructed surface of the scene in Fig. 5 as seen from a novel view.

7. Conclusions

We have shown that accommodation is an *unambiguous* visual cue. We have done so by defining two notions of observability: *weak observability* and *strong observability*. We showed that in the presence of a "scanning light" or a "structured light" it is possible to distinguish the shape of any surface (weak observability). It is also possible to approximate to an arbitrary degree (although never realize exactly) structured patterns that allow distinguishing any two surfaces (i.e. patterns that

approximate a *universally exciting* radiance). In the absence of prior information on the radiance of the scene, we showed that two surfaces can be distinguished up to a degree of resolution defined by the complexity of the radiance (strong observability) and by the optics of the imaging system.

We have proposed a solution to the problem of reconstructing shape and radiance of a scene using I-divergence as a criterion in an optimization framework. The algorithm is iterative, and we gave a proof of its convergence to a (local) minimum which, by

construction, is admissible in the sense of resulting in a positive radiance and imaging kernel.

Appendix A: Proof of Proposition 1

Proof: We will prove the statement by contradiction. Suppose there exists a surface S' and an open $O \subset \Omega$ such that $S'(\mathbf{x}) \neq S(\mathbf{x}) \, \forall \mathbf{x} \in O$. Then, from the definition of weak indistinguishability, we have that for any radiance r there exists a radiance r' such that $I_\mathbf{u}^S(\mathbf{y}, r) = I_\mathbf{u}^{S'}(\mathbf{y}, r') \, \forall \mathbf{y} \in \Omega \; \forall \mathbf{u} \in \mathcal{U}$. By Property 1 in Section 2, for any $\mathbf{y} \in \Omega$ there exists a focus setting $\bar{\mathbf{u}}$ such that $I_{\bar{\mathbf{u}}}^S(\mathbf{y}, r) = r(\mathbf{y})$ and hence

$$I_{\bar{\mathbf{u}}}^{S'}(\mathbf{y}, r') \doteq \int h_{\bar{\mathbf{u}}}^{S'}(\mathbf{y}, \mathbf{x}) r'(\mathbf{x}) \, d\mathbf{x} = r(\mathbf{y}). \qquad (35)$$

Similarly, for any $\mathbf{y} \in \Omega$ there exists a focus setting $\bar{\mathbf{u}}'$ such that $I_{\bar{\mathbf{u}}'}^{S'}(\mathbf{y}, r') = r'(\mathbf{y})$ and

$$I_{\bar{\mathbf{u}}'}^S(\mathbf{y}, r) \doteq \int h_{\bar{\mathbf{u}}'}^S(\mathbf{y}, \mathbf{x}) r(\mathbf{x}) \, d\mathbf{x} = r'(\mathbf{y}). \qquad (36)$$

Notice that, for a given radiance r, we obtained an explicit expression of the radiance r' such that S and S' are indistinguishable. Substituting (36) in (35), we obtain

$$\iint h_{\bar{\mathbf{u}}}^{S'}(\mathbf{y}, \mathbf{x}) h_{\bar{\mathbf{u}}'}^S(\mathbf{x}, \tilde{\mathbf{x}}) r(\tilde{\mathbf{x}}) \, d\tilde{\mathbf{x}} \, d\mathbf{x} = r(\mathbf{y})$$

$$\forall \mathbf{y} \in \Omega \quad \forall r \in \mathcal{L}_{loc}^2(\mathbb{R}^2). \quad (37)$$

Since the above equation holds for any $r \in \mathcal{L}_{loc}^2(\mathbb{R}^2)$, it follows that

$$\int h_{\bar{\mathbf{u}}}^{S'}(\mathbf{y}, \mathbf{x}) h_{\bar{\mathbf{u}}'}^S(\mathbf{x}, \tilde{\mathbf{x}}) \, d\mathbf{x} = \delta(\mathbf{y} - \tilde{\mathbf{x}}) \quad \forall \mathbf{y}, \tilde{\mathbf{x}} \in O$$

$$(38)$$

However, since $S'(\mathbf{x}) \neq S(\mathbf{x})$ for $\mathbf{x} \in O \subset \Omega$, then $\bar{\mathbf{u}}$ is not a focus setting for S' in O and $\bar{\mathbf{u}}'$ is not a focus setting for S in O by Property 2. This means that, by Property 3, there exists an open set $\bar{O} \subset O$ such that $h_{\bar{\mathbf{u}}}^{S'}(\mathbf{y}, \mathbf{x}) > 0$ and $h_{\bar{\mathbf{u}}'}^S(\mathbf{x}, \tilde{\mathbf{x}}) > 0$ for any $\mathbf{y}, \mathbf{x}, \tilde{\mathbf{x}} \in \bar{O}$. Hence, if we choose $\mathbf{y}, \tilde{\mathbf{x}} \in \bar{O}$ with $\mathbf{y} \neq \tilde{\mathbf{x}}$, we have

$$\int h_{\bar{\mathbf{u}}}^{S'}(\mathbf{y}, \mathbf{x}) h_{\bar{\mathbf{u}}'}^S(\mathbf{x}, \tilde{\mathbf{x}}) \, d\mathbf{x} > 0 \neq \delta(\mathbf{y} - \tilde{\mathbf{x}}) = 0, \quad (39)$$

which is a contradiction.   □

Appendix B: Proof of Proposition 2

Proof: Let $f(|x|) = F(x)$, where $f : \mathbb{R}^+ \to \mathbb{R}$. Suppose that the above is true in the case $\mathbf{y} = 0$: then it is true in any case; indeed, we could otherwise substitute $r_\mathbf{y}(\mathbf{x}) = r(\mathbf{x} + \mathbf{y})$, and

$$r(\mathbf{y}) \int F(\mathbf{x}) \, d\mathbf{x} = r_\mathbf{y}(0) \int F(\mathbf{x}) \, d\mathbf{x}$$

$$= \int f(|\tilde{\mathbf{x}}|) r_\mathbf{y}(\tilde{\mathbf{x}}) \, d\tilde{\mathbf{x}}$$

$$= \int f(|-\tilde{\mathbf{x}}|) r(\tilde{\mathbf{x}} + \mathbf{y}) \, d\tilde{\mathbf{x}}$$

$$= \int f(|\mathbf{y} - \mathbf{x}|) r(\mathbf{x}) \, d\mathbf{x}. \qquad (40)$$

Now, we are left with proving the proposition for $\mathbf{y} = 0$. We change coordinates by

$$\int f(|\mathbf{x}|) r(\mathbf{x}) \, d\mathbf{x} = \int_0^\infty \int_{C_\rho} f(\rho) r(z) \, d\rho \, d\mu_\rho(z)$$

$$= \int_0^\infty f(\rho) \int_{C_\rho} r(z) \, d\mu_\rho(z) \, d\rho$$

$$(41)$$

where C_ρ is the circumference of radius ρ centered in 0, $z \in C_\rho$, and μ_ρ is the measure on C_ρ (that is, the restriction of the Hausdorff measure $\mathcal{H}^1$ to C_ρ). Then, by the mean value theorem, (see Taylor, 1996, Proposition 2.4 in Section 3.2)

$$\int_{C_\rho} r(z) \, d\mu_\rho(z) = 2\pi\rho r(0)$$

and then

$$\int f(|\mathbf{x}|) r(\mathbf{x}) \, dx = r(0) \int_0^\infty f(\rho) 2\pi\rho \, d\rho$$

$$= r(0) \int F(\mathbf{x}) \, d\mathbf{x}. \qquad \square$$

Appendix C: Equifocal Imaging Models

C.1. Pillbox Imaging Model

Under conditions that are commonly accepted in the computer vision community, and explained in detail in Mennucci and Soatto (1999), we can approximate the kernel $h_\mathbf{u}^S(\mathbf{y}, \mathbf{x})$ in Eq. (1) with a so-called "pillbox"

function $p_\rho(\mathbf{y})$:

$$h_{\mathbf{u}}^S(\mathbf{y}, \mathbf{x}) = p_{\rho(z,\mathbf{u})}(\mathbf{y} - \mathbf{x}) \qquad (42)$$

where the surface S is represented by the parameter $\mathbf{s} = z$, the depth of the imaged plane, $\rho = \rho(z, \mathbf{u})$ is the radius of the kernel, and

$$p_{\rho(z,\mathbf{u})}(\mathbf{y}) = \begin{cases} \dfrac{1}{\pi\rho^2} & |\mathbf{y}| \leq \rho(z, \mathbf{u}) \\ 0 & \text{elsewhere.} \end{cases} \qquad (43)$$

Consider a finite number $n \geq 2$ of different images $I_{u_1}^S, \ldots, I_{u_n}^S$ of an equifocal infinite plane, generated according to the model

$$I_{u_j}^S(\mathbf{y}) = \int_{\mathbb{R}^2} p_{\rho_j}(\mathbf{y} - \mathbf{x}) r(\mathbf{x}) \, d\mathbf{x} \qquad (44)$$

where $\rho_j = \rho(z, u_j)$. Then, we would like to solve that system of equations for the variables $(\rho_1, \ldots, \rho_n)$ (with $\rho_i > 0$, $i = 1, \ldots, n$) and the radiance r, given the images; indeed, we will be able to recover the distance $\mathbf{s} = z$, given $\mathbf{u}$ and $(\rho_1, \ldots, \rho_n)$.[8] We remark that we are considering an abstract, noiseless, model.

Assume that $r \not\equiv 0$, and that r has finite energy (that is $r \in \mathcal{L}^2(\mathbb{R}^2)$): then, by transforming the image $I_u^S(\mathbf{y})$ to the Fourier domain, we have

$$\forall f, \quad \hat{I}_{u_j}^S(f) = \hat{p}_{\rho_j}(f)\hat{r}(f) \qquad (45)$$

where $\hat{p}_{\rho_j}(f)$ is the Fourier transform of the pillbox function $p_{\rho_j}(\mathbf{y})$, and $\hat{r}(f)$ is the Fourier transform of the radiance $r(\mathbf{x})$. Thus, we can infer that the radiance is well determined once one of the radii ρ_j is known. We recall that

$$\hat{p}_{\rho_j}(f) = \frac{2}{\rho_j|f|} J_1(|f|\rho_j) \qquad (46)$$

where J_1 is the Bessel function of the first kind.

Proposition 6. *Suppose that we are given $n \geq 2$ (different) defocused images.[9] Suppose that r is non zero and has finite energy (or, equivalently, suppose that the images $I_{u_j}^S$ are non zero and have finite energy); let $(\rho_1, \ldots, \rho_n), r(\mathbf{x})$ be a solution to the equations*

$$\forall j, \mathbf{y} \quad I_{u_j}^S(\mathbf{y}) = \int_{\mathbb{R}^2} p_{\rho_j}(\mathbf{y} - \mathbf{x}) r(\mathbf{x}) \, d\mathbf{x} \qquad (47)$$

Then, $(\rho_1, \ldots, \rho_n)$ is isolated in $\mathbb{R}^n$.

This proposition states that, in this simplified model, the accommodation cue is "unambiguous", that is, the shape and camera parameters are "observable", given the images: it is indeed possible to solve Eq. (47) for the radiance and the distances of the planes, and to obtain a finite number of isolated solutions.

Proof: Suppose that $(\rho_1^*, \ldots, \rho_n^*), r^*(\mathbf{x})$ is a solution. Let

$$\begin{aligned} A &= \{f \mid \hat{r}^*(f) = 0\} \cup \{0\} \\ D_0 &\doteq \{f \mid \hat{I}_{u_j}^S(f) \neq \hat{p}_{\rho_j^*}(f)\hat{r}^*(f) \text{ for a } j\} \\ D_1 &= \{f \mid \hat{p}_{\rho_j^*}(f) = 0 \text{ for a } j \leq n\} \\ &= \{f \mid J_1(|f|\rho_j^*) = 0 \text{ for a } j\} \end{aligned} \qquad (48)$$

the set D_1 is union of a countable number of circles, so D_0, D_1 are sets of measure zero. For $f \notin A \cup D_0 \cup D_1$, we get $\hat{I}_{u_j}^S(f) \neq 0$, and then we define $g_{\rho_j}(f) \doteq \frac{\hat{p}_{\rho_j}(f)}{\hat{I}_{u_j}^S(f)}$ so that $g_{\rho_i^*}(f) = \frac{1}{\hat{r}^*(f)}$ (for any i).

Whenever $f \notin A \cup D_0 \cup D_1$, furthermore, $\frac{g_{\rho_i^*}(f)}{g_{\rho_j^*}(f)} = 1$, so $\frac{g_{\rho_i}(f)}{g_{\rho_j}(f)}$ is a well defined real number for (ρ_i, ρ_j) in a neighborhood of (ρ_i^*, ρ_j^*); so we define

$$d_{\rho_i,\rho_j}(f) \doteq \log\left(\frac{g_{\rho_i}(f)}{g_{\rho_j}(f)}\right) \qquad (49)$$

(for $i \neq j$): then, $d_{\rho_i^*,\rho_j^*}(f) = 0 \,\forall\, f$; more in general, if $(\rho_1, \ldots, \rho_n)$ is a solution to (47), then $d_{\rho_i,\rho_j}(f) = 0 \,\forall\, f$. We will prove that the solution $(\rho_1^*, \ldots, \rho_n^*)$ is isolated, by studying the derivatives of d_{ρ_i,ρ_j} w.r.t. ρ_i, ρ_j.

We compute the derivative of g_{ρ_i} with respect to ρ_i: we introduce the auxiliary variables $\sigma_{\rho_i} = |f|\rho_i$; then

$$\begin{aligned} \frac{\partial}{\partial\rho_i} g_{\rho_i}(f) &= \frac{1}{\hat{I}_{u_j}^S} 2|f| \frac{\partial}{\partial\sigma_{\rho_i}}\left(\frac{1}{\sigma_{\rho_i}} J_1(\sigma_{\rho_i})\right) \\ &= -\frac{2}{\hat{I}_{u_i}^S(f)\rho_i} J_2(|f|\rho_i). \end{aligned} \qquad (50)$$

We define the set $D_2 = \{f \mid J_2(|f|\rho_i^*) = 0 \text{ for a } i \leq n\}$ the set D_2 is again union of a countable number of circles, so it is a set of measure zero. Whenever $f \notin A \cup D_0 \cup D_1 \cup D_2$, we have that the derivative $\frac{\partial}{\partial\rho_i} g_{\rho_i^*}(f) \neq 0$;[10] the gradient of d_{ρ_i,ρ_j} with respect to ρ_i, ρ_j is then

$$\begin{aligned} \nabla d_{\rho_i,\rho_j} &= \left(\frac{\frac{\partial}{\partial\rho_i} g_{\rho_i}}{g_{\rho_i}}, -\frac{\frac{\partial}{\partial\rho_j} g_{\rho_j}}{g_{\rho_j}}\right) \\ &= |f|\left(-\frac{J_2(|f|\rho_i)}{J_1(|f|\rho_i)}, \frac{J_2(|f|\rho_j)}{J_1(|f|\rho_j)}\right) \end{aligned} \qquad (51)$$

It is easy to see that $\frac{\partial}{\partial \rho_i} d_{\rho_i,\rho_j}$ and $\frac{\partial}{\partial \rho_j} d_{\rho_i,\rho_j}$ are non zero for $f \notin A \cup D_0 \cup D_1 \cup D_2$. Then, it is possible to express ρ_i in terms of ρ_j (and vice versa) in equation $\{d_{\rho_i,\rho_j} = 0\}$: this implies that, locally in a neighborhood of $(\rho_1^*, \ldots, \rho_n^*)$, there is *at most* a parametric curve $(\rho_1, \ldots, \rho_n) = (\rho_1(t), \ldots, \rho_n(t))$ of solutions. If there was such a curve, then, for any fixed i, $j \neq i$, all the gradients $\{\nabla d_{\rho_i,\rho_j}(f) \mid \forall f \notin A \cup D_0 \cup D_1 \cup D_2\}$ would be parallel, that is, there would be constants vectors $(a_{i,j}, b_{i,j})$ (not zero) such that

$$a_{i,j} \frac{\partial}{\partial \rho_i} d_{\rho_i^*,\rho_j^*}(f) = b_{i,j} \frac{\partial}{\partial \rho_j} d_{\rho_i^*,\rho_j^*}(f). \quad (52)$$

We wish to show that this is not the case: we will show that the above cannot be an equality: this will prove the proposition.

The equality

$$a_{i,j} \frac{J_2(|f|\rho_i^*)}{J_1(|f|\rho_i^*)} = b_{i,j} \frac{J_2(|f|\rho_j^*)}{J_1(|f|\rho_j^*)} \quad (53)$$

that is,

$$a_{i,j} J_2(|f|\rho_i^*) J_1(|f|\rho_j^*) = b_{i,j} J_2(|f|\rho_j^*) J_1(|f|\rho_i^*) \quad (54)$$

is an equality between analytic functions, valid for $f \notin A \cup D_0 \cup D_1 \cup D_2$. Therefore, it is valid for all f: we obtain that

$$a_{i,j} J_2(t\rho_i^*) J_1(t\rho_j^*) = b_{i,j} J_2(t\rho_j^*) J_1(t\rho_i^*) \quad \forall t \in \mathbb{R} \quad (55)$$

(where t has been substituted for $|f|$). Since $J_1(t) = t/2 - t^3/16 + o(t^3)$ while $J_2(t) = t^2/8 - t^4/96 + o(t^4)$, by substituting,

$$\begin{aligned}
& a_{i,j}((t\rho_i^*)^2/8 - (t\rho_i^*)^4/96)((t\rho_j^*)/2 - (t\rho_j^*)^3/16) \\
& \quad - b_{i,j}((t\rho_j^*)^2/8 - (t\rho_j^*)^4/96)((t\rho_i^*)/2 \\
& \quad - (t\rho_i^*)^3/16) + o(t^5) = 0 \\
& = t^3(a_{i,j}\rho_i^{*2}\rho_j^* - b_{i,j}\rho_i^*\rho_j^{*2})/16 \\
& \quad + t^5(a_{i,j}(\rho_i^{*2}\rho_j^{*3}/128 + \rho_i^{*4}\rho_j^*/192) \\
& \quad - b_{i,j}(\rho_j^{*2}\rho_i^{*3}/128 + \rho_j^{*4}\rho_i^*/192)) + o(t^5) = 0
\end{aligned} \quad (56)$$

By the theory of Taylor series,

$$\begin{aligned}
& a_{i,j}\rho_i^{*2}\rho_j^* = b_{i,j}\rho_i^*\rho_j^{*2} \\
& a_{i,j}(\rho_i^{*2}\rho_j^{*3}/128 + \rho_i^{*4}\rho_j^*/192) \\
& \quad = b_{i,j}(\rho_j^{*2}\rho_i^{*3}/128 + \rho_j^{*4}\rho_i^*/192)
\end{aligned} \quad (57)$$

or

$$\begin{aligned}
& a_{i,j}\rho_i^* = b_{i,j}\rho_j^* \\
& a_{i,j}(\rho_i^*\rho_j^{*2}/128 + \rho_i^{*3}/192) \\
& \quad = b_{i,j}(\rho_j^*\rho_i^{*2}/128 + \rho_j^{*3}/192)
\end{aligned} \quad (58)$$

or (having $c = a_{i,j}/b_{i,j}$)

$$\rho_i^* = c\rho_j^* \quad (59)$$
$$c\rho_j^{*3}(1/128 + 1/192) = c^3\rho_j^{*3}/128 + \rho_j^{*3}/192$$

or

$$\begin{aligned}
& c^3/128 - c(1/128 + 1/192) + 1/192 \\
& \quad = (c-1)(c^2/128 - 1/192) = 0
\end{aligned} \quad (60)$$

whose solutions are $c \in \{1, \pm\sqrt{2/3}\}$; the first is to be excluded because otherwise $\rho_i^* = \rho_j^*$, the other two since they do not satisfy (55). $\qquad\square$

Appendix D: Cramér-Rao Lower Bound

In the previous sections we studied the observability and proposed an optimal algorithm for depth from defocus from a deterministic point of view. In what follows we take a stochastic approach to the problem. We consider images affected by Gaussian noise. Then, we compute an approximation of the Cramér-Rao lower bound of the radiance step in Eq. (27) proposed in Section 5.

The implementation of the algorithm requires that we represent the kernel family, the radiance and the surface of the scene with finite dimensional vectors. Recall the discrete lattice $\Gamma \doteq [x_1, \ldots, x_N] \times [y_1, \ldots, y_M]$ for some integers N and M. We substitute the continuous coordinates $\mathbf{x}$ and $\mathbf{y}$ with their discrete versions $\mathbf{x}_i$ and $\mathbf{y}_n$, where $i \in [1, \ldots, N]$ and $n \in [1, \ldots, M]$. Also, to simplify the notation, we drop the arguments of the functions defined in Eq. (16) and denote images as $I_n \doteq I^s(\mathbf{y}_n, r)$, kernels as $h_{n,i} \doteq h^s(\mathbf{y}_n, \mathbf{x}_i)$ and radiances as $r_i \doteq r(\mathbf{x}_i)$. In this analysis we approximate the image model with

$$I_n = \sum_{i=1}^{N} h_{n,i} r_i \quad (61)$$

where the kernel satisfies

$$\sum_{i=1}^{N} h_{n,i} = 1 \quad \forall\, i \in [1, \dots, N]. \qquad (62)$$

We define $I_n^{\mathcal{N}}$ (where the superscript $\mathcal{N}$ denotes that it is affected by noise), as a Gaussian random variable with mean I_n and covariance σ^2; the components of the random variable vector $I^{\mathcal{N}} \doteq \{I_n^{\mathcal{N}}\}_{n=1..M}$ are also independent from each other.

The probability density distribution of $I^{\mathcal{N}}$ is therefore

$$p_{I^{\mathcal{N}}}(I^{\mathcal{N}}) = \prod_{n=1}^{M} \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(I_n^{\mathcal{N}} - I_n)^2}{2\sigma^2}}. \qquad (63)$$

The radiance step defined in Eq. (27) is recursive and nonlinear, and thus it is not suitable for an evaluation of its covariance. However, we are interested to analyze the "stability" of the estimator around the solution. In this case we can approximate the estimated radiance $\hat{r}$ with the true radiance r and plug it in the right hand side of Eq. (27). Hence, with the notation just introduced, we obtain

$$\hat{r}_i = r_i \frac{1}{\sum_{n=1}^{M} h_{n,i}} \sum_{n=1}^{M} \frac{h_{n,i}\, I_n^{\mathcal{N}}}{I_n}. \qquad (64)$$

This estimator is unbiased, as we have

$$E[\hat{r}_i] = r_i \frac{1}{\sum_{n=1}^{M} h_{n,i}} \sum_{n=1}^{M} \frac{h_{n,i}\, E[I_n^{\mathcal{N}}]}{I_n} = r_i \qquad (65)$$

where $E[\cdot]$ denotes the expectation of a random variable.

Computing the Cramér-Rao lower bound consists in evaluating the *Fisher information* matrix

$$[I_F]_{i,j} = -E\left[\frac{\partial^2 \log(p_{I^{\mathcal{N}}}(I^{\mathcal{N}}))}{\partial r_i \partial r_j}\right] \quad i,j \in [1, \dots, N]. \qquad (66)$$

Once I_F is known, the bound is expressed as

$$\mathrm{var}\{\hat{r}\} - I_F^{-1} \geq 0 \qquad (67)$$

where the vector $\hat{r} \doteq [\hat{r}_i]_{i=[1,\dots,N]}$ and the inequality is interpreted in the sense of positive semi-definite matrices.

Now, we proceed with computing I_F:

$$\log(p_{I^{\mathcal{N}}}(I^{\mathcal{N}})) = \sum_{n=1}^{M} -\frac{(I_n^{\mathcal{N}} - I_n)^2}{2\sigma^2} - \log(\sqrt{2\pi\sigma^2}). \qquad (68)$$

Taking the derivative of the last expression with respect to r_i we have

$$\frac{\partial \log(p_{I^{\mathcal{N}}}(I^{\mathcal{N}}))}{\partial r_i} = \sum_{n=1}^{M} \frac{I_n^{\mathcal{N}} - I_n}{\sigma^2}\, h_{n,i} \qquad (69)$$

and it is easy to verify that it satisfies the *regularity condition* (a necessary condition to apply the Cramér-Rao lower bound):

$$E\left[\frac{\partial \log(p_{I^{\mathcal{N}}}(I^{\mathcal{N}}))}{\partial r_i}\right] = 0 \quad \forall\, i \in [1, \dots, N]. \qquad (70)$$

Now, taking derivatives with respect to r_j we obtain:

$$\frac{\partial^2 \log(p_{I^{\mathcal{N}}}(I^{\mathcal{N}}))}{\partial r_i \partial r_j} = -\sum_{n=1}^{M} \frac{h_{n,i}\, h_{n,j}}{\sigma^2}. \qquad (71)$$

Since the above equation is not dependent on any random variable, taking the expectation does not change the expression, and we have:

$$[I_F]_{i,j} = \sum_{n=1}^{M} \frac{h_{n,i}\, h_{n,j}}{\sigma^2}. \qquad (72)$$

To make the notation more compact, we denote with H_n the row vector $[h_{n,1}, \dots, h_{n,N}]$. Hence, the bound is

$$\mathrm{var}\{\hat{r}\} \geq \sigma^2 \left(\sum_{n=1}^{M} H_n^T H_n\right)^{-1}. \qquad (73)$$

As a last step we need to evaluate $\mathrm{var}\{\hat{r}\}$. This can be immediately determined and it results in:

$$[\mathrm{var}\{\hat{r}\}]_{i,j} = E[(\hat{r}_i - r_i)(\hat{r}_j - r_j)]$$
$$= \frac{r_i r_j}{\sum_{m=1}^{M} h_{m,i} \sum_{m=1}^{M} h_{m,j}} \sum_{n=1}^{M} h_{n,i}\, h_{n,j} \frac{\sigma^2}{I_n^2}. \qquad (74)$$

Notice that the variance σ^2 appears on both sides of the inequality, and unless σ^2 is zero (absence of noise), we can remove it from both sides of Eq. (73).

Again, to simplify the notation we define the modified kernel $\tilde{h}_{n,i} \doteq \frac{h_{n,i}}{\sum_{m=1}^{M} h_{m,i}}$ and its corresponding vector $\tilde{H}_n \doteq [\tilde{h}_{n,1}, \ldots, \tilde{h}_{n,N}]$. Finally, we introduce the vectors $r \doteq [r_1, \ldots, r_N]^T$ and $R_n \doteq [r_1 \tilde{h}_{n,1}, \ldots, r_N \tilde{h}_{n,N}]$. The bound is now expressed as

$$\sum_{n=1}^{M} \frac{R_n^T R_n}{r^T \tilde{H}_n^T \tilde{H}_n r} \geq \left(\sum_{n=1}^{M} \tilde{H}_n^T \tilde{H}_n \right)^{-1}. \quad (75)$$

We notice that the bound depends on the radiance being imaged and the kernel of the optics, that in turn depends on the surface of the scene.

Consider the simplifying case of having an almost constant radiance vector r; recalling the earlier discussions about observability of shape and radiance, this case yields a poor reconstruction and can be analyzed as the worst case. The equation above becomes approximately

$$\sum_{n=1}^{M} \tilde{H}_n^T \tilde{H}_n \geq \left(\sum_{n=1}^{M} \tilde{H}_n^T \tilde{H}_n \right)^{-1} \quad (76)$$

which is satisfied if and only if $\sum_{n=1}^{M} \tilde{H}_n^T \tilde{H}_n$ is the identity matrix (given the constraints on $\tilde{H}_n$). This corresponds to having an image focused everywhere. On the other hand, as our setting is such that the scene is more and more defocused, the bound is less and less tight, and the estimation of the radiance worsens as expected.

Acknowledgments

Notes

1. Note that it is still necessary to make a-priori assumptions in order to solve the correspondence problem.
2. This section introduces the problem of depth from defocus and our notation. The reader who is familiar with the literature may skip to Section 2.
3. We use the term "positive" for a quantity x to indicate $x \geq 0$. When $x > 0$ we say that x is "strictly positive".
4. Note that, since neither the light source nor the viewer move, we do not make a distinction between radiance and reflectance in this paper. This corresponds to assuming that the appearance of the surface does not change when seen from different points on the lens. This is the case when the aperture is negligible relative to the distance from the scene.
5. A surface is called *Lambertian* if its bidirectional reflectance distribution function is independent of the outgoing direction (and, by the reciprocity principle, of the incoming direction as well). An intuitive notion is that a point on a Lambertian surface appears equally bright from all viewing directions.
6. The term distribution is used in the sense of the theory of distributions (see for instance Friedlander and Joshi, 1998), and not in the sense of probability.
7. When there are no positivity constraints, Csiszár argues that the only consistent choice of discrepancy criterion is the $\mathcal{L}^2$ norm, which we have addressed in Soatto and Favaro (2000).
8. Moreover, if the relationship $\rho_j = \rho(z, \mathbf{u}_j)$ is known only up to certain number of parameters of the optics, then, given a large enough number n of images, we may be able to recover the distance of the plane alongside the parameters of the optics: this could lead to an auto-calibration of the model.
9. By saying that the images are defocused and different we mean $\rho_i > 0$, and $\rho_i \neq \rho_j$ when $i \neq j$.
10. Incidentally, this tells us that, for many values of f, it is possible to make explicit the dependence of ρ_i on $\hat{r}(f)$: this will be mostly useful in studying how the noise affects the solution to (47).

References

Asada, N., Fujiwara, H., and Matsuyama, T. 1998. Seeing behind the scene: Analysis of photometric properties of occluding edges by reversed projection blurring model. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20:155–167.

Chan, T.F. and Wong, C.K. 1998. Total variation blind deconvolution. *IEEE Transactions on Image Processing*, 7(3):370–375.

Chaudhuri, S. and Rajagopalan, A.N. 1999. *Depth from Defocus: A Real Aperture Imaging Approach*. Springer-Verlag: Berlin.

Cover, T.M. and Thomas, J.A. 1991. *Elements of Information Theory*. Wiley Interscience.

Csiszár, I. 1991. Why least squares and maximum entropy? An axiomatic approach to inverse problems. *Ann. of Stat.*, 19:2033–2066.

Ens, J. and Lawrence, P. 1993. An investigation of methods for determining depth from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(2):97–108.

Farid, H. and Simoncelli, E.P. 1998. Range estimation by optical differentiation. *Journal of the Optical Society of America A*, 15(7):1777–1786.

Favaro, P. and Soatto, S. 2000. Shape and radiance estimation from the information divergence of blurred images. In *Proc. European Conference on Computer Vision*, 1:755–768.

Friedlander, G. and Joshi, M. 1998. *Introduction to the Theory of Distributions*. Cambridge University Press: Cambridge.

Girod, B. and Scherock, S. 1989. Depth from focus of structured light. In *SPIE*, pp. 209–215.

Gokstorp, M. 1994. Computing depth from out-of-focus blur using a local frequency representation. In *International Conference on Pattern Recognition*, Vol. A, pp. 153–158.

Harikumar, G. and Bresler, Y. 1999. Perfect blind restoration of images blurred by multiple filters: Theory and efficient algorithms. *IEEE Transactions on Image Processing*, 8(2):202–219.

Hopkins, H.H. 1955. The frequency response of a defocused optical system. *Proc. R. Soc. London Ser. A*, 231:91–103.

Hwang, T.L., Clark, J.J., and Yuille, A.L. 1989. A depth recovery algorithm using defocus information. In *Computer Vision and Pattern Recognition*, pp. 476–482.

Kalifa, J., Mallat, S., and Rouge, B. 1998. Image deconvolution in mirror wavelet bases. In *Proc. International Conference on Image Processing*, p. 98.

Kundur, D. and Hatzinakos, D. 1998. A novel blind deconvolution scheme for image restoration using recursive filtering. *IEEE Transactions on Signal Processing*, 46(2):375–390.

Luemberger, D. 1968. *Optimization by Vector Space Methods*. Wiley: New York.

Marshall, J.A., Burbeck, C.A., Ariely, D., Rolland, J.P., and Martin, K.E. 1996. Occlusion edge blur: A cue to relative visual depth. *Journal of the Optical Society of America A*, 13(4):681–688.

Mennucci, A. and Soatto, S. 1999. The accommodation cue, part 1: Modeling. Essrl Technical Report 99-001, Washington University.

Murase, H. and Nayar, S. 1995. Visual learning and recognition or 3d object from appearance. *Intl. J. Computer Vision*, 14(1):5–24.

Nair, H.N. and Stewart, C.V. 1992. Robust focus ranging. In *Computer Vision and Pattern Recognition*, pp. 309–314.

Namba, M. and Ishida, Y. 1998. Wavelet transform domain blind deconvolution. *Signal Processing*, 68(1):119–124.

Nayar, S.K. and Nakagawa, Y. 1990. Shape from focus: An effective approach for rough surfaces. In *IEEE International Conference on Robotics and Automation*, pp. 218–225.

Nayar, S.K., Watanabe, M., and Noguchi, M. 1995. Real-time focus range sensor. In *Proc. International Conference on Computer Vision*, pp. 995–1001.

Neelamani, R., Choi, H., and Baraniuk, R. 1999. Wavelet-domain regularized deconvolution for ill-conditioned systems. In *Proc. International Conference on Image Processing*, pp. 58–72.

Noguchi, M. and Nayar, S.K. 1994. Microscopic shape from focus using active illumination. In *International Conference on Pattern Recognition*, pp. 147–152.

Pentland, A. 1987. A new sense for depth of field. *IEEE Trans. Pattern Anal. Mach. Intell.*, 9:523–531.

Pentland, A., Darrell, T., Turk, M., and Huang, W. 1989. A simple, real-time range camera. In *Computer Vision and Pattern Recognition*, pp. 256–261.

Pentland, A., Scherock, S., Darrell, T., and Girod, B. 1994. Simple range cameras based on focal error. *Journal of the Optical Society of America A*, 11(11):2925–2934.

Rajagopalan, A.N. and Chaudhuri, S. 1995. A block shift-variant blur model for recovering depth from defocused images. In *Proc. International Conference on Image Processing*, pp. 636–639.

Rajagopalan, A.N. and Chaudhuri, S. 1997. Optimal selection of camera parameters for recovery of depth from defocused images. In *Computer Vision and Pattern Recognition*, pp. 219–224.

Rajagopalan, A.N. and Chaudhuri, S. 1998. Optimal recovery of depth from defocused images using an mrf model. In *Proc. International Conference on Computer Vision*, pp. 1047–1052.

Schechner, Y.Y. and Kiryati, N. 1999. The optimal axial interval in estimating depth from defocus. In *IEEE Proc. International Conference on Computer Vision*, Vol. II, pp. 843–848.

Schechner, Y.Y., Kiryati, N., and Basri, R. 1998. Separation of transparent layers using focus. In *IEEE Proc. International Conference on Computer Vision*, pp. 1061–1066.

Scherock, S. 1981. Depth from focus of structured light. In Technical Report-167, Media-Lab, MIT.

Schneider, G., Heit, B., Honig, J., and Bremont, J. 1994. Monocular depth perception by evaluation of the blur in defocused images. In *Proc. International Conference on Image Processing*, Vol. 2, pp. 116–119.

Simoncelli, E.P. and Farid, H. 1996. Direct differential range estimation using optical masks. In *European Conference on Computer Vision*, Vol. II, pp. 82–93.

Snyder, D.L., Schulz, T.J., and O'Sullivan, J.A. 1992. Deblurring subject to nonnegativity constraints. *IEEE Transactions on Signal Processing*, 40(5):1142–1150.

Soatto, S. and Favaro, P. 2000. A geometric approach to blind deconvolution with application to shape from defocus. *Proc. IEEE Computer Vision and Pattern Recognition*, 2:10–17.

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

Subbarao, M. and Wei, T.C. 1992. Depth from defocus and rapid autofocusing: A practical approach. In *Computer Vision and Pattern Recognition*, pp. 773–776,

Taylor, M. 1996. *Partial Differential Equations* (volume i: Basic Theory). Springer Verlag: Berlin.

Watanabe, M. and Nayar, S.K. 1996a. Minimal operator set for passive depth from defocus. In *Computer Vision and Pattern Recognition*, pp. 431–438.

Watanabe, M. and Nayar, S.K. 1996b. Telecentric optics for computational vision. In *European Conference on Computer Vision*, Vol. II, pp. 439–445.

Xiong, Y. and Shafer, S.A. 1995. Moment filters for high precision computation of focus and stereo. In *Proc. of International Conference on Intelligent Robots and Systems*, pp. 108–113.

Ziou, D. 1998. Passive depth from defocus using a spatial domain approach. In *Proc. International Conference on Computer Vision*, pp. 799–804.