

# Observing Shape From Defocused Images\*

P. Favaro      A. Mennucci      S. Soatto

**Keywords:** Shape, surface geometry, low-level vision

## 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 (de-blurring) 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 establish the existence of a radiance that allows distinguishing any two surfaces (sufficient excitation) and finally (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.

## 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. In order to retrieve the spatial information lost in the imaging process, one can rely on prior assumptions on the scene and use 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 have it as their 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

---

\*P. Favaro is with Washington University in St.Louis, A. Mennucci is with the Scuola Normale Superiore, Pisa – Italy, S. Soatto is with the University of California, Los Angeles and Washington University in St.Louis. Address to be used for correspondence is S. Soatto, 4531E Boelter Hall, UCLA, Los Angeles - CA 90095-1596, USA, tel. (310)825-4840, email soatto@cs.ucla.edu. This research was supported by NSF grant IIS-9876145 and ARO grant DAAD19-99-1-0139.

properties of images which are associated with a changing viewpoint are called “parallax” cues (for instance stereo and motion)<sup>1</sup>.

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 which are associated with a change of the geometry of the imaging device.*

Accommodation, among other visual cues, is the one that has been studied the least from an analytical viewpoint, and even the most fundamental questions remain unanswered: is it a visual cue at all (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? While some of these questions would be ill-posed for pictorial cues, they can be answered rigorously for the case of accommodation.

Estimating shape from focus/defocus boils down to inverting certain integral equations, a problem known by different names in different communities: in signal processing it is “blind deconvolution” or “deblurring”, in communications and information theory “source separation”, in image processing “restoration”, in tomography “inverse scattering”. Since images depend both on the shape of the scene and on its reflectance properties – neither of which is known – estimating shape is tightly related to estimating reflectance. In this paper, we consider the two problems as one and the same, and discuss the simultaneous solution of both. We choose as criterion the minimization of the information divergence (I-divergence) between blurred images, motivated by the work of Csiszár [5]. The algorithm we propose is iterative, and we give a proof of its convergence to a (local) minimum. We present results on both real and simulated images.

## Main results and outline of the paper

It is often believed that accommodation is a “weak” cue, compared to pictorial or motion cues, in the sense that it provides little information on the depth of the scene. On the contrary, we prove that accommodation is an unambiguous cue that can be used to estimate the three-dimensional shape of any surface under suitable conditions<sup>2</sup>. We do so in three steps: (1) we characterize the observability of any surface in the presence of a controlled radiance (weak observability), (2)

---

<sup>1</sup>Note that it is still necessary to make a-priori assumptions, in order to solve the correspondence problem.

<sup>2</sup>There are deep relationships between accommodation and parallax cues. For one thing, accommodation can be interpreted as the limit of stereo, where the viewpoint is distributed in a continuum on the focal plane across the extension of the lens. One may be tempted to conjecture that all the known results on uniqueness of stereo reconstruction apply to accommodation cues. Unfortunately, that is not the case: the passage to the continuum significantly changes the analytical landscape, and new tools are necessary for the analysis. Furthermore, the analysis of stereo assumes that there is a finite number of points for which the correspondence can be assessed. Here, on the other hand, we give explicit conditions that the radiance of a surface must satisfy for us to be able to reconstruct its shape. Another important issue on the relationship between parallax and accommodation concerns *sensor fusion*, or how to combine information from the two cues. This is an extremely interesting topic of research, which we do not address in this paper; see [12, 15].

we establish the existence of a radiance that allows distinguishing any two surfaces (sufficient excitation) and finally (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).

This paper is concerned with theoretical issues, and is aimed at legitimizing shape from focus/defocus algorithms by giving specific conditions under which they can – or cannot – work. This issue has been largely neglected in most of the literature. Our goal is to provide analytical results which can be useful to anyone willing to venture into the design of algorithms to reconstruct shape from focus or defocus. Algorithms developed under the guidelines of this analysis are reported in Section 6.

## Relation to previous work

In the literature of computational vision a number of algorithms have been proposed to estimate depth from focus and accommodation information. The main assumption common to all algorithms available in the literature (overtly or covertly) is that the scene is a plane parallel to the focal plane (equifocal assumption)[7, 20, 1, 13, 6, 18, 7, 14, 9, 22, 21, 15]. There is also a vast body of related literature in the Signal Processing community, where the problem is known as “blind deconvolution” (or more generally “deblurring”). The equifocal assumption is equivalent to assuming a shift-invariant convolution kernel, which is also common to most of the literature. The interested reader can see the special issue [2] for references.

The capability to reconstruct the scene’s shape depends upon the energy distribution it irradiates. To our knowledge there is no literature on the conditions on the radiant energy that allows estimating shape.

A new analysis is therefore needed that can address the problem of estimating an unknown shape irradiating energy according to an unknown distribution. Hopefully this analysis will result in more reliable and general algorithms to estimate both the shape of the object and the energy distribution it irradiates. This paper extends and completes some preliminary results presented in [10]. We also present the first (locally) optimal algorithm for shape from defocus.

### 1.1 Notation and formalization of the problem

In order to introduce some of the concepts that we will address in this paper, consider a scene whose three-dimensional shape is represented symbolically by  $S$  - and described by a finite number of parameters  $\mathbf{s}$  - which emits energy with a radiance  $r$ . While these concepts will be defined formally in the next section, a basic physical intuition suffices for now. The image of the scene,  $I$ , can be represented as a function with support on a compact subset  $\Omega$  of the imaging surface (e.g. the retina or the CCD sensor) whose positive values represent the brightness at a particular point. Suppose further that the geometry of the imaging device is described by a finite set of parameters, for instance focal length, optical center, lens aperture etc. collected into the vector  $\mathbf{u}$ . Then we indicate the imaging process symbolically by

$$I = I_{\mathbf{u}}^{\mathbf{s}}(\mathbf{y}, r) \quad \mathbf{y} \in \Omega. \tag{1}$$

In [11] we have argued that a model of the imaging process that is suitable to study the accommodation cue is given by integral equations of the form

$$I_{\mathbf{u}}^{\mathbf{s}}(\mathbf{y}, r) = \int_{\mathbb{R}^2} h_{\mathbf{u}}^{\mathbf{s}}(\mathbf{y}, \mathbf{x}) dR(\mathbf{x}) \quad (2)$$

where  $\mathbf{y} \in \Omega \subset \mathbb{R}^2$ ,  $\mathbf{x} \in \mathbb{R}^2$ ,  $\mathbf{u} \in \mathcal{U} \subset \mathbb{R}^2$  and  $\mathbf{s} \in \mathbb{R}^s$ . We also introduce the density corresponding to the energy distribution  $R$  and denote it with the function  $r$  defined by

$$r(\mathbf{x}) d\mathbf{x} \doteq dR(\mathbf{x}). \quad (3)$$

Strictly speaking,  $r$  is the Radon-Nikodym derivative of  $R$ , and as such it may not be an ordinary function but, rather, a distribution of measures. We neglect such technicalities here, since they do not affect the derivation of our algorithm. The measure  $dR$  is a representation of the *energy distribution*, or *radiance*<sup>3</sup>, of the surface in  $\mathbf{x}$  coordinates;  $h = h_{\mathbf{u}}^{\mathbf{s}}$  is a parameterized family of positive maps which we call *kernels*; in general they can be either delta measures or bounded functions continuous inside a support, but possibly discontinuous across its border. By duality the function  $r$  will take values in a class  $\mathcal{L}_{loc}(\mathbb{R}^2)$ , which includes locally integrable positive functions as well as delta measures.

We assume that we can control the distance between the lens and the sensor,  $Z_I$ , and the lens aperture  $D$ , so that we have the control parameters

$$\mathbf{u} = \{Z_I, D\} \subset \mathbb{R}^2. \quad (4)$$

The shape parameters  $\mathbf{s}$  will describe a compact portion of the scene and will be left as general as possible. For the sake of example, in the case of a slanted plane we have that the surface  $S$  is described by the intercept with the optical axis,  $Z_S$ , and by a reduced normal  $\tilde{\mathbf{n}}$ , so that  $\mathbf{s} = \{Z_S, \tilde{\mathbf{n}}\} \subset \mathbb{R}^3$ . In the case of a simple occlusion, described by a half-plane with boundary at  $\mathbf{y} = 0$  parallel to the sensor at a distance  $Z_1$  occluding a plane, also parallel to the sensor, at distance  $Z_2$ , we have that  $\mathbf{s} = \{Z_1, Z_2\} \subset \mathbb{R}^2$ .

In equation (2), we assume that the values  $I_{\mathbf{u}}^{\mathbf{s}}(\mathbf{y})$  can be measured for each value of  $\mathbf{y} \in \Omega$  and control parameters  $\mathbf{u}$ . The energy distribution  $r$  is unknown, and  $h_{\mathbf{u}}^{\mathbf{s}}(\mathbf{y}, \mathbf{x})$  is known only *up to the shape parameters*  $\mathbf{s}$ , which are unknown. The scope of this paper can be summarized into the following question:

**Question 1** to what extent can the energy distribution  $r$  and shape parameters  $\mathbf{s}$  be reconstructed from measurements of  $I_{\mathbf{u}}^{\mathbf{s}}(\mathbf{y})$ ?

The next few sections are devoted to answering the above seemingly innocent question. We proceed in increasing order of generality, from assuming that the radiance  $r$  can be chosen purposefully (Section 2) to an arbitrary unknown variance (Section 5). Along the way, we point out some inconsistencies frequently encountered in the literature on shape from focus/defocus (Section 4).

---

<sup>3</sup>Since neither the light source nor the viewer moves, we do not make a distinction between radiance and reflectance in this paper. This corresponds to assuming that the appearance of a surface does not change when seen from different points on the lens. This is the case when the aperture is small relative to the distance from the scene.

## 2 Weak Observability

The concept of “weak observability”, which we are about to define, is relevant to the problem of reconstruction from structured light, where the radiance pattern  $r$  can be controlled.

**Definition 1** *We say that a surface  $S_1$  is weakly indistinguishable from a surface  $S_2$ , and write  $S_1 \mathcal{I} S_2$  if for all possible  $r_1 \in \mathcal{L}_{loc}(\mathbb{R})$  there exists at least a radiance  $r_2 \in \mathcal{L}_{loc}(\mathbb{R})$  such that we have*

$$I_{\mathbf{u}}^{\mathbf{s}_1}(\mathbf{y}, r_1) = I_{\mathbf{u}}^{\mathbf{s}_2}(\mathbf{y}, r_2) \quad \forall \mathbf{y} \in \Omega \quad \forall \mathbf{u} \in \mathcal{U}. \quad (5)$$

*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.*

The purpose of this section is to establish that weakly indistinguishable surfaces are equal up to a set of measure zero. In order to prove this, we need to make some assumptions on the imaging system.

**Hypothesis 1** *We will assume that there is always a unique point in focus, that is*

$$\forall \mathbf{y} \exists \mathbf{u} = \mathbf{u}^*(\mathbf{y}, \mathbf{s}) \quad | \quad h_{\mathbf{u}}^{\mathbf{s}}(\mathbf{y}, \mathbf{x}) = \delta(\mathbf{y} - \mathbf{x}) \quad (6)$$

*while for  $\mathbf{u} \neq \mathbf{u}^*(\mathbf{y}, \mathbf{s})$ ,  $\mathbf{x} \mapsto h_{\mathbf{u}}^{\mathbf{s}}(\mathbf{y}, \mathbf{x})$  is an ordinary function. We will also assume that the point in focus is bijectively related to the point on the surface, so that if  $S$  is described by the graph of a function  $S(\mathbf{y})$ , we have*

$$\mathbf{u}(\mathbf{y}, S) = \mathbf{u}(\mathbf{y}, S') \quad \Leftrightarrow \quad S(\mathbf{y}) = S'(\mathbf{y}). \quad (7)$$

This hypothesis only holds in the case of non self-intersecting surfaces. Under the above conditions we can state the following

**Proposition 1** *Any non self-intersecting smooth surface  $S$  is weakly observable.*

**Proof:** *Suppose on the contrary that  $S$  is weakly indistinguishable from  $S' \neq S$ ; from the definition we have that  $\forall r \exists r'$  such that  $\forall \mathbf{y}, \mathbf{u}$  we have  $I_{\mathbf{u}}^{\mathbf{s}}(\mathbf{y}, r) = I_{\mathbf{u}}^{\mathbf{s}'}(\mathbf{y}, r')$ . From the existence of a point in focus we have that, for every  $\mathbf{y}$  there exists  $\mathbf{u} = \mathbf{u}^*(\mathbf{y}, S)$  such that  $I_{\mathbf{u}}^{\mathbf{s}}(\mathbf{y}, r) = r(\mathbf{y})$ , and therefore*

$$r(\mathbf{y}) = \int h_{\mathbf{u}^*(\mathbf{y}, \mathbf{s})}^{\mathbf{s}'}(\mathbf{y}, \mathbf{x}) r'(\mathbf{x}) d\mathbf{x} \quad (8)$$

*for all  $\mathbf{y}$  and  $r'$ . We choose  $r = \delta_{\mathbf{y}}$ . In particular, we have that  $h_{\mathbf{u}}^{\mathbf{s}'} = \delta(\mathbf{y} - \mathbf{x})$  if and only if  $\mathbf{u} = \mathbf{u}(\mathbf{y}, \mathbf{s}')$ ; then the above implies that  $r' = \delta_{\mathbf{y}}$  and  $\mathbf{u}(\mathbf{y}, \mathbf{s}) = \mathbf{u}(\mathbf{y}, \mathbf{s}')$ .*

**Remark 1** *The above proposition 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  $\mathbf{u}$ . This, however, requires the active usage of the control  $u$  and of the radiance  $r$ . While this may be possible in an active vision context, it is not the case in a classical vision setting, where  $r$  is not free for the user to choose.*

The proof of the proposition above entails a choice of a particular distribution  $r$  - that can be a function of  $S$  but not of  $S'$  - that is “sufficiently exciting”, i.e. that allows distinguishing  $S$  from  $S'$ . It says that, for any given  $S$ , such  $r$  can be found. However, it leaves a natural question open:

**Question 2** *Can a radiance  $r$  be found, that allows distinguishing any two surfaces?*

We address this issue in the next section.

### 3 Excitation

Let us define  $\mathcal{I}(S|r)$  as the set of surfaces that cannot be distinguished from  $S$  given the energy distribution  $r$ :

**Definition 2**

$$\mathcal{I}(S|r) = \{\tilde{S} \mid I_{\mathbf{u}}^{\tilde{s}}(\mathbf{y}, r) = I_{\mathbf{u}}^s(\mathbf{y}, r) \quad \forall \mathbf{y}, \mathbf{u}\} \quad (9)$$

**Remark 2** *Note that we are using the same radiance on both surfaces. This is the case, for instance, when using structured light, so that  $r$  is a (static) pattern projected onto a surface.*

Clearly, not all energy distributions allow distinguishing two surfaces. For instance, the distribution  $r = 0$  does not allow distinguishing any surface. We now discuss the existence of a sufficiently exciting distribution.

**Definition 3** *We say that the distribution  $r$  is sufficiently exciting for  $S$  if*

$$\mathcal{I}(S|r) = \{S\} \quad (10)$$

Using the terminology above, we say that a shape  $S$  is *excitable* if it admits at least one sufficiently exciting distribution  $r$ .

We conjecture that distributions with unbounded variation are sufficiently exciting for any surface. A motivation for this conjecture can be extrapolated from the discussion relative to  $\mathcal{L}^2$  that follows in Section 4; in particular, see Remark 5.

**Conjecture 1** *Let  $r$  have unbounded variation on any open subset of the image. Then it is universally exciting, i.e. it is sufficiently exciting for any (piecewise smooth) surface  $S$ .*

### 4 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  $\mathcal{L}^2(\mathbb{R})$ . However, such a restriction is legitimate only if the radiance  $r$  does not have any harmonic component, since that would not belong to  $\mathcal{L}^2(\mathbb{R})$ . Our choice is justified by the fact that the harmonic component of the radiance does not carry any shape information.

Before venturing into the analysis, we remind the reader of the basic properties of harmonic functions.

**Definition 4** *A function  $r$  is said to be harmonic in an open region  $\Omega$  if*

$$\Delta r(\mathbf{y}) \doteq \sum_i \frac{\partial^2}{\partial \mathbf{y}_i^2} r(\mathbf{y}) = 0$$

for all  $\mathbf{y} \in \Omega$

**Proposition 2 (mean-value property)** *If  $r$  is harmonic, then, for any measurable function  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,*

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

*whenever the integral exists*

The above property leads us to conclude that harmonic functions are negligible as carriers of shape information.

**Corollary 1** *From [11], we have that  $h_{\mathbf{u}(\mathbf{y}, \mathbf{s})}^{\mathbf{s}'}$ ( $\mathbf{y}, \mathbf{x}$ ) can be approximated by a circularly symmetric function; if  $r'$  is a harmonic function, then*

$$\int h_{\mathbf{u}(\mathbf{y}, \mathbf{s})}^{\mathbf{s}'}$$
( $\mathbf{y}, \mathbf{x}$ ) $r'(\mathbf{x})d\mathbf{x} \cong r'(\mathbf{y}),$

*regardless of the surface  $\mathbf{s}'$  and the parameters of the imaging device  $\mathbf{u}$ .*

*In particular, if the surface is equifocal, then equality holds in the above equation, and shape parameters are unobservable.*

The above result has the consequence of allowing one to restrict the analysis to the non-harmonic component of the radiance, which we represent as a function in the space  $\mathcal{L}^2$ . This leads naturally to a notion of “bandwidth”, as we describe below.

**Definition 5** *Let  $\{\theta_i(\cdot)\}$  be an orthonormal basis of  $\mathcal{L}^2$ . We say that a radiance  $r$  has a degree of definition (or degree of complexity)  $k$  if there exists a set of coefficients  $\alpha_i$ ,  $i = 1 \dots k$  such that*

$$r(\mathbf{x}) = \sum_{i=0}^k \alpha_i \theta_i(\mathbf{x}). \tag{11}$$

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

**Remark 3** *Note that for practical purposes there is no loss of generality in assuming that  $r \in \mathcal{L}^2$ , 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  $h \in \mathcal{L}^2$ , we can define a *degree of resolution* for the surface  $S$ .

**Definition 6** *Let  $h_{\mathbf{u}}^{\mathbf{s}}$ ( $\mathbf{y}, \mathbf{x}$ )  $\in \mathcal{L}^2$ . If there exists a positive real integer  $\rho$  and coefficients  $\beta_j$ ,  $j = 1 \dots \rho$  such that*

$$h_{\mathbf{u}}^{\mathbf{s}}$$
( $\mathbf{y}, \mathbf{x}$ ) =  $\sum_{j=0}^{\rho} \beta_j(\mathbf{y}, \mathbf{u}, \mathbf{s})\theta_j(\mathbf{x})$  \tag{12}

*then we say that the surface  $S$  has a degree of resolution  $\rho$ .*

**Remark 4** *The above definitions of degrees of complexity and resolution depend upon the choice of basis of  $\mathcal{L}^2$ . 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*

$$h_{\mathbf{u}}^{\mathbf{s}}(\mathbf{y}, \mathbf{x}) = \sum_{j=0}^{\infty} \beta_j(\mathbf{y}, \mathbf{u}, \mathbf{s}) \theta_j(\mathbf{x}) \quad (13)$$

then

$$\mathcal{I}(S_1) \supset \{S_2 \mid \beta_j(\mathbf{y}, \mathbf{u}, \mathbf{s}_1) = \beta_j(\mathbf{y}, \mathbf{u}, \mathbf{s}_2) \quad \forall \mathbf{y}, \mathbf{u}, j = 0 \dots k\}. \quad (14)$$

**Proof:** *Substituting the expression of  $r$  in terms of the basis  $\theta_i$ , we have*

$$\begin{aligned} I_{\mathbf{u}}^{\mathbf{s}_1}(\mathbf{y}, r) &= \sum_{i,j=0}^k \beta_j(\mathbf{y}, \mathbf{u}, \mathbf{s}_1) \alpha_i \int \theta_j(\mathbf{x}) \theta_i(\mathbf{x}) d\mathbf{x} + \\ &\int \sum_{i,j=k+1}^{\infty} \beta_j(\mathbf{y}, \mathbf{u}, \mathbf{s}_1) \alpha_i \theta_i(\mathbf{x}) \theta_j(\mathbf{x}) d\mathbf{x}. \end{aligned} \quad (15)$$

From the orthogonality of the basis elements  $\theta_i$  we are left with

$$I_{\mathbf{u}}^{\mathbf{s}_1}(\mathbf{y}, r) = \sum_{i=0}^k \beta_i(\mathbf{y}, \mathbf{u}, \mathbf{s}_1) \alpha_i \quad (16)$$

from which we see that, if  $\beta_i(\mathbf{y}, \mathbf{u}, \mathbf{s}_1) = \beta_i(\mathbf{y}, \mathbf{u}, \mathbf{s}_2)$  for all  $i = 1 \dots k$ , then  $I_{\mathbf{u}}^{\mathbf{s}_1}(\mathbf{y}, r) = I_{\mathbf{u}}^{\mathbf{s}_2}(\mathbf{y}, r)$  for all  $\mathbf{y}, \mathbf{u}$ .

**Remark 5** *The practical value of the last proposition is to state that, the more “irregular” the distribution, 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.*

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 this section we will consider the observability of a scene depending upon its particular energy distribution.

## 5 Strong Observability

**Definition 7** *We say that the pair  $(S_2, r_2)$  is indistinguishable from the pair  $(S_1, r_1)$ , and we write  $(S_1, r_1) \mathcal{I}(S_2, r_2)$  if*

$$\int h_{\mathbf{u}}^{\mathbf{s}_1}(\mathbf{y}, \mathbf{x}) r_1(\mathbf{x}) d\mathbf{x} = \int h_{\mathbf{u}}^{\mathbf{s}_2}(\mathbf{y}, \mathbf{x}) r_2(\mathbf{x}) d\mathbf{x} \quad \forall \mathbf{y} \in \Omega \quad \forall \mathbf{u} \in \mathcal{U}. \quad (17)$$

We define, as usual, the set of pairs  $(S_2, r_2)$  that are indistinguishable from  $(S_1, r_1)$  as

$$\mathcal{I}(S_1, r_1) \doteq \{(S_2, r_2) \mid (S_1, r_1)\mathcal{I}(S_2, r_2)\} \quad (18)$$

The following proposition characterizes the set of indistinguishable scenes in terms of their coefficients in the basis  $\{\theta\}$

**Proposition 4** *Let  $r_1$  have degree of definition  $k$ . 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 definition  $k$ , 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

$$\begin{aligned} \sum_{i,j=0}^k \beta_j(\mathbf{y}, \mathbf{u}, \mathbf{s}_1) \alpha_{1_i} \int \theta_i(\mathbf{x}) \theta_j \mathbf{x} d\mathbf{x} = \\ \sum_{i,j=0}^{\infty} \beta_j(\mathbf{y}, \mathbf{u}, \mathbf{s}_2) \alpha_{2_i} \int \theta_i(\mathbf{x}) \theta_j \mathbf{x} d\mathbf{x} \end{aligned} \quad (19)$$

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

$$\alpha_{1_i} = \alpha_{2_i} \quad \forall i = 1 \dots k \quad (20)$$

from which we have that  $\beta_j(\mathbf{y}, \mathbf{u}, \mathbf{s}_1) = \beta_j(\mathbf{y}, \mathbf{u}, \mathbf{s}_2)$  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 dependence is coded into the coefficients  $\beta_i$ . Any more 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 as described in [11], and concentrating on a class of surfaces, for instance slanted planes or occlusions. The calculations are straightforward but messy, and are therefore not reported here. Instead, in the following section we describe in detail the analysis for the case of a plane at constant depth, since this has consequences for practical applications and relates to the existing literature on SFF/SFD.

## 5.1 Equifocal approximation of physical scenes

If the scene being imaged is a plane parallel to the lens at a constant distance (an “equifocal plane”), then the imaging model (2) simplifies significantly since the kernel  $h$  is shift-invariant and, therefore, the integral can be represented as a convolution, as explained in detail in [11]:

$$I_{\mathbf{u}}^s(\mathbf{y}, r) = h_{\mathbf{u}}^s * r(\mathbf{y}). \quad (21)$$

We will consider two common models for defocusing, depending on the kernel  $h$ : a uniform model, and a Gaussian model. Before doing so, however, we warn the reader that using the equifocal approximation carries significant risks of designing algorithms that return biased results.

**Remark 6 (On the equifocal approximation)** *The equifocal assumption considerably simplifies the analysis and the design of estimation algorithms, which are commonly known as “blind deconvolution” algorithms. The basic idea consists in dividing the scene into a number of small “patches”, and approximating each patch with an equifocal plane. Often – overtly or covertly – transform techniques are applied, where one Fourier-transforms the radiance  $r$  in each window, implicitly assuming that  $r$  is periodic at the boundary of the window.*

*Any algorithm based upon the equifocal assumption, however, is bound to return inconsistent results. In fact, suppose that the radiance has a non-zero harmonic component, as discussed in Section 4. Then the following occurs:*

- *the results of any algorithm on the interior of a window are biased since depth is not observable from the harmonic component of the radiance (Section 4). Transform-based techniques return only information on the boundaries of each window. However,*
- *the results on the boundary of the window are also biased, since the boundary includes the periodic copy of the window, which is an artifact.*

Now that we are aware of the down-sides of equifocal assumption, we analyze the two common models in Appendix A. The conclusion reached there is that there cannot be full observability from Gaussian defocus; since – for small radii – a pillbox kernel can be approximated by a Gaussian kernel, this suggests that, in the presence of noise, recovering of a set of derived variables  $s_i$  (i.e. relative depth) will be more robust of recovering the radii ( $r_i$ ) themselves (i.e. absolute depth).

## 6 Optimal estimation

In this section we formulate the problem of optimally estimating depth from an equifocal imaging model. Recall the imaging model

$$I_{\mathbf{u}}(x, y) = \int h_{\mathbf{u}}^s(x, y) dR \quad (x, y) \in D. \quad (22)$$

where now we have made explicit the variable  $\mathbf{y} = \{x, y\}$ . Notice that, in the equation above, all quantities are constrained to be *positive*<sup>4</sup>:  $dR$  because it represents the radiant energy (which cannot be negative),  $h^s$  because it specifies the region of space over which energy is integrated, and  $I$  because it measures the photon-count on the surface of the sensor. We are interested in estimating the shape of the surface  $\mathbf{s}$  and the energy distribution  $R$  to the extent possible, by measuring a number  $l$  of images obtained with different camera settings  $\mathbf{u}_1, \dots, \mathbf{u}_l$ .

### 6.1 Formalization of the problem

If we collect a number of different images  $I_{\mathbf{u}_1}, \dots, I_{\mathbf{u}_l}$  and organize them into a vector  $I$  (and so for the kernels  $h$ ), we can write

$$I(x, y) = \int h^s(x, y) dR \quad (x, y) \in D. \quad (23)$$

---

<sup>4</sup>We use the term “positive” for a quantity  $x$  to indicate  $x \geq 0$ . When  $x > 0$  we say that  $x$  is “strictly positive”.

The right-hand side of the above equation can be interpreted as the “virtual image” of a given surface  $\mathbf{s}$  radiating energy with a given distribution  $R$ . We call such virtual image  $b$ :

$$b^{\mathbf{s}}(x, y, R) \doteq \int h^{\mathbf{s}}(x, y, X, Y, Z) dR(X, Y, Z) \quad (x, y) \in D. \quad (24)$$

Note that, for images of opaque objects, the integral is restricted to their surface, and therefore can be written in the Riemannian sense [3] as

$$b^{\mathbf{s}}(x, y, R) \doteq \int h^{\mathbf{s}}(x, y, \tilde{x}, \tilde{y}) dR(\tilde{x}, \tilde{y}) \quad (x, y) \in D \quad (25)$$

for a suitably chosen parameterization  $(\tilde{x}, \tilde{y}) \in \mathbb{R}^2$ . In either case, we write the integral in shorthand notation as  $b^{\mathbf{s}}(x, y, R) \doteq \int h^{\mathbf{s}}(x, y) dR$ . Since the image  $I$  is measured on the pixel grid, the domain  $D$  (i.e. a patch in the image) is  $D = [x_1, \dots, x_N] \times [y_1, \dots, y_M]$ , so that we have

$$I(x_i, y_j) = b^{\mathbf{s}}(x_i, y_j, R) \quad i = 1 \dots N, j = 1 \dots M. \quad (26)$$

We now want a “criterion”  $\phi$  to measure the discrepancy between the measured image  $I$  and the virtual one, so that we can formulate our problem as the minimization of the discrepancy between the measured image and the model (or virtual) image. Common choices of criteria include the least-squares distance between  $I$  and  $b^{\mathbf{s}}$ , or the integral of the absolute value (“total variation”) of their difference [4].

In order to get a “reasonable” result, the criterion  $\phi$  should satisfy a number of requirements. Csiszár makes this notion rigorous through the axiomatic derivation of cost functions that satisfy certain consistency conditions. He concludes that, when the quantities involved are constrained to be positive<sup>5</sup> (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

$$\phi(I, b^{\mathbf{s}}(R)) \doteq \sum_{i,j} \left\{ I(x_i, y_j) \log \frac{I(x_i, y_j)}{b^{\mathbf{s}}(x_i, y_j, R)} - I(x_i, y_j) + b^{\mathbf{s}}(x_i, y_j, R) \right\}. \quad (27)$$

In order to emphasize the dependency of the cost function  $\phi$  on the unknowns  $\mathbf{s}, R$ , we abuse the notation to write

$$\phi = \phi(\mathbf{s}, R). \quad (28)$$

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

$$\mathbf{s}, R = \arg \min_{\mathbf{s}, R} \phi(\mathbf{s}, R). \quad (29)$$

---

<sup>5</sup>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 [17].

## 6.2 Alternating minimization

In general, the problem in (29) 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

$$\begin{cases} \mathbf{s}_{k+1} \doteq \arg \min_{\mathbf{s}} \phi(\mathbf{s}, R_k) \\ R_{k+1} \doteq \arg \min_R \phi(\mathbf{s}_{k+1}, R) \end{cases} \quad (30)$$

leads to the (local) minimization of  $\phi$ , since

$$\phi_{k+1} \doteq \phi(\mathbf{s}_{k+1}, R_{k+1}) \leq \phi(\mathbf{s}_{k+1}, R_k) \leq \phi(\mathbf{s}_k, R_k) \doteq \phi_k \quad (31)$$

and the sequence  $\phi_k$  is bounded below by zero. However, solving the two optimization problems in (30) may be an overkill. In order to have a sequence  $\{\phi_k\}$  that monotonically converges it is sufficient that - at each step - we choose  $\mathbf{s}$  and  $R$  in such a way as to guarantee that equation (31) holds, that is

$$\begin{cases} \mathbf{s}_{k+1} \mid \phi(\mathbf{s}_{k+1}, R) \leq \phi(\mathbf{s}_k, R) & R = R_k \\ R_{k+1} \mid \phi(\mathbf{s}, R_{k+1}) \leq \phi(\mathbf{s}, R_k) & \mathbf{s} = \mathbf{s}_{k+1}. \end{cases} \quad (32)$$

In the imaging models described in Appendix A, the “shape” of the surface is trivial and represented by a positive scalar  $\mathbf{s}$  that depends upon  $Z$ , the depth of the patch  $U$ . Since the first step of the minimization depends only on this parameter, we can choose any of the known descent methods (e.g. Newton-Raphson). The choice is arbitrary and does not affect the considerations that follow. Therefore, we indicate this step generically as:

$$\mathbf{s}_{k+1} = \arg \min_{\mathbf{s}_k > 0} \phi(\mathbf{s}_k, r). \quad (33)$$

The second step is obtained from the Kuhn-Tucker conditions [8] associated with the problem of minimizing  $\phi$  for fixed  $\mathbf{s}$  under positivity constraints for  $r$ :

$$\sum_{i,j} \frac{h^{\mathbf{s}}(x_i, y_j, \tilde{x}, \tilde{y}) I(x_i, y_j)}{\int h^{\mathbf{s}}(x_i, y_j, \tilde{x}, \tilde{y}) r(\tilde{x}, \tilde{y}) d\tilde{x} d\tilde{y}} = \begin{cases} \sum_{i,j} h^{\mathbf{s}}(x_i, y_j, \tilde{x}, \tilde{y}) & \forall (\tilde{x}, \tilde{y}) \mid r(\tilde{x}, \tilde{y}) > 0 \\ \leq \sum_{i,j} h^{\mathbf{s}}(x_i, y_j, \tilde{x}, \tilde{y}) & \forall (\tilde{x}, \tilde{y}) \mid r(\tilde{x}, \tilde{y}) = 0. \end{cases} \quad (34)$$

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. [16], we choose

$$F(\mathbf{s}, r) \doteq \frac{1}{\sum_{i,j} h^{\mathbf{s}}(x_i, y_j, \tilde{x}, \tilde{y})} \sum_{i,j} \frac{h^{\mathbf{s}}(x_i, y_j, \tilde{x}, \tilde{y}) I(x_i, y_j)}{b^{\mathbf{s}}(x_i, y_j, r)} \quad (35)$$

and define the following iteration:

$$r_{k+1} = r_k F(\mathbf{s}, r_k). \quad (36)$$

It is important to point out that this iteration decreases the I-divergence  $\phi$  not only when we use the exact kernel  $h^{\mathbf{s}}$ , as it is showed in Snyder et al. [16], but also with any other kernel satisfying the positivity and smoothness constraint. This fact is proven by the following claim.

**Claim 1** 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 (36). Then  $\phi(\mathbf{s}, r_{k+1}) \leq \phi(\mathbf{s}, r_k) \forall k > 0, \forall \mathbf{s} > 0$ . Furthermore equality holds if and only if  $r_{k+1} = r_k$ .

**Proof:** The proof follows Snyder et al. [16]. From the definition of  $\phi$  in equation (27) we get

$$\phi(\mathbf{s}, r_{k+1}) - \phi(\mathbf{s}, r_k) = - \sum_{i,j} I(x_i, y_j) \log \frac{b^{\mathbf{s}}(x_i, y_j, r_{k+1})}{b^{\mathbf{s}}(x_i, y_j, r_k)} + \sum_{i,j} b^{\mathbf{s}}(x_i, y_j, r_{k+1}) - b^{\mathbf{s}}(x_i, y_j, r_k). \quad (37)$$

From the expression of  $r_{k+1}$  in (36) we have that the second sum in the above expression is given by

$$\begin{aligned} & \sum_{i,j} \int h^{\mathbf{s}}(x_i, y_j, x, y) r_{k+1}(x, y) dx dy - \sum_{i,j} \int h^{\mathbf{s}}(x_i, y_j, x, y) r_k(x, y) dx dy = \\ & = \int h_0^{\mathbf{s}}(x, y) r_{k+1}(x, y) dx dy - \int h_0^{\mathbf{s}}(x, y) r_k(x, y) dx dy \end{aligned}$$

where we have defined  $h_0^{\mathbf{s}}(x, y) = \sum_{i,j} h^{\mathbf{s}}(x_i, y_j, x, y)$ , while the ratio in the first sum is

$$\frac{b^{\mathbf{s}}(x_i, y_j, r_{k+1})}{b^{\mathbf{s}}(x_i, y_j, r_k)} = \int F(\mathbf{s}, r_k) \frac{h^{\mathbf{s}}(x_i, y_j, x, y) r_k(x, y)}{b^{\mathbf{s}}(x_i, y_j, r_k)} dx dy. \quad (38)$$

We next note that, from Jensen's inequality,

$$\log \left( \int F(\mathbf{s}, r_k) \frac{h^{\mathbf{s}}(x_i, y_j, x, y) r_k(x, y)}{b^{\mathbf{s}}(x_i, y_j, r_k)} dx dy \right) \geq \int \frac{h^{\mathbf{s}}(x_i, y_j, x, y) r_k(x, y)}{b^{\mathbf{s}}(x_i, y_j, r_k)} \log(F(\mathbf{s}, r_k)) dx dy \quad (39)$$

since the ratio  $\frac{h^{\mathbf{s}}(x_i, y_j, x, y) r_k(x, y)}{b^{\mathbf{s}}(x_i, y_j, r_k)}$  can be interpreted as a probability distribution dependent on the parameters  $\mathbf{s}$  and  $r_k$ , and therefore the expression in (37) is

$$\begin{aligned} \phi(\mathbf{s}, r_{k+1}) - \phi(\mathbf{s}, r_k) & \leq - \sum_{i,j} I(x_i, y_j) \int \log(F(\mathbf{s}, r_k)) \frac{h^{\mathbf{s}}(x_i, y_j, x, y) r_k(x, y)}{b^{\mathbf{s}}(x, y, r_k)} dx dy + \\ & \int h_0^{\mathbf{s}}(x, y) r_{k+1}(x, y) dx dy - \int h_0^{\mathbf{s}}(x, y) r_k(x, y) dx dy. \end{aligned}$$

The right-hand side of the above expression can be written as

$$\phi_c(h_0^{\mathbf{s}}(x, y) r_{k+1}(x, y), h_0^{\mathbf{s}}(x, y) r_k(x, y)) \quad (40)$$

where we define  $\phi_c(f(x, y), g(x, y))$  as  $\int f(x, y) \log \frac{f(x, y)}{g(x, y)} - f(x, y) + g(x, y) dx dy$ , which can be easily verified to be a positive function for any positive  $f, g$ . Therefore, we have

$$\phi(\mathbf{s}, r_{k+1}) - \phi(\mathbf{s}, r_k) \leq 0. \quad (41)$$

Note that Jensen's inequality holds with equality if and only if  $F(\mathbf{s}, r_k)$  is a constant; since the only admissible constant value is 1, then we have  $r_{k+1} = r_k$ , which concludes the proof.

Finally, we can conclude that the algorithm proposed generates a monotonically decreasing sequence of values of the cost function  $\phi$ . We say that the initial conditions  $\mathbf{s}_0, r_0$  are *admissible* if  $\mathbf{s}_0 > 0$  and  $r_0$  is a positive function defined on  $\mathbb{R}^2$ .

**Corollary 1** *Let  $\mathbf{s}_0, r_0$  be admissible initial conditions for the sequences  $\mathbf{s}_k$  and  $r_k$  defined from equations (33) and (36) respectively. Let  $\phi_k$  be defined as  $\phi(\mathbf{s}_k, r_k)$ , then the sequence  $\{\phi_k\}$  converges to a limit  $\phi^*$ :*

$$\lim_{k \rightarrow \infty} \phi_k = \phi^*. \quad (42)$$

**Proof:** *Follows directly from equation (33), (31) and Claim 1, together with the fact that the I-divergence is bounded from below by zero.*

Even if  $\phi_k$  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 (22), which has been analyzed in Section 5. In Appendix B we compute the Cramér-Rao lower bound for an unbiased estimator of shape.

## 7 Experiments

In this section we discuss some details that are important for implementing the algorithm just described on a digital computer, and test its performance on a set of experiments on real and simulated images.

### 7.1 Implementation

Since the algorithm we propose is iterative, we need to initialize it with a feasible radiance. We choose  $r = I_{u_1}$ , that is we choose the initial estimate of the radiance to be equal to the first image. This choice is guaranteed to be feasible since the image is positive. Since  $h^s(x_i, y_j, x, y)$  is discrete in the first two variables, one needs to exercise caution when performing numerical integration against kernels smaller than the unit step of the discretization  $(x_i, y_j)$ . This case cannot be discounted because it occurs whenever the patch on the image that we are observing is close to be in focus. In our implementation, integrals are computed with a first order (linear) approximation as a tradeoff between speed and accuracy.

Another important detail to bear in mind is that it is necessary to choose the appropriate integration domain. The fact that we use an equifocal imaging model allows us to use the same reference frame for the image and for space, which is represented locally by a plane parallel to it. However, the image in any given patch receives contributions from a region of space possibly bigger than the patch itself. Thus we write  $I(x_i, y_j) = \int h^s(x_i, y_j, x, y)(r_I(x, y) + r_o(x, y))dx dy$ , where  $r_o$  is the radiance outside the patch that contributes to the convolution with the kernel  $h^s$ . In the real and synthetic experiments we always use two images, with planes focused at 529 mm and 869 mm as in the data set provided to us by Watanabe and Nayar [20]; the lens diameter is such that the maximum kernel radius is around 2.3 pixels. With these values the kernel is well approximated by a Gaussian since the radius is small compared to the image patch dimension. Therefore, we define  $r_o$  on a domain that is 3 pixels wider than the domain of  $r_I$ , which we choose to be  $7 \times 7$ , ending up integrating on patches of dimension  $13 \times 13$ .

## 7.2 Experiments with synthetic images

In this set of experiments we investigate the robustness of the algorithm to noise. Even though we have not derived the algorithm based upon a particular noise model (all the discussion is strictly deterministic), it can be shown that minimizing the I-divergence can be cast into a stochastic framework by modeling the image noise as a Poisson process.

We have generated 10 noisy image pairs and considered patches of size  $7 \times 7$  pixels. Smaller patches result in greater sensitivity to noise, while larger ones challenge the equifocal approximation. We have considered additive Gaussian noise with a variance that ranges from the 1% to the 10% of the radiance magnitude, which guarantees that the positivity constraint is still satisfied with high probability. The results of the computed depths are summarized in Figure 1. We iterate the algorithm 5 times at each point.

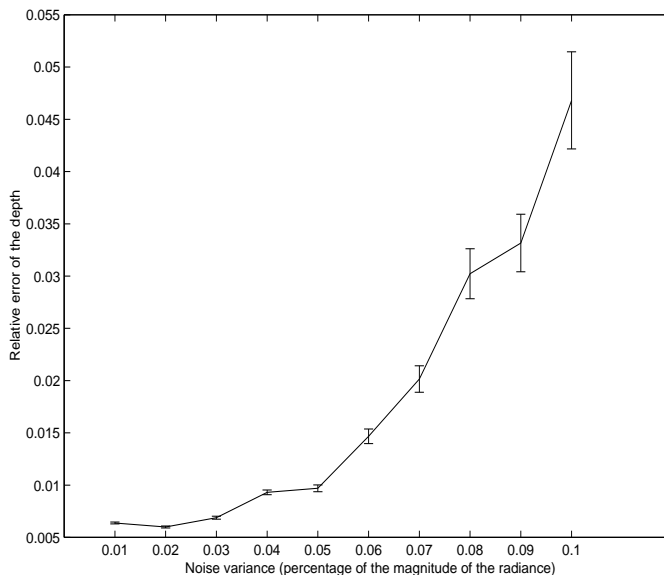


Figure 1: *Depth error as a function of image noise, mean and standard deviation.*

As it can be seen, the algorithm is quite robust to the additive noise, even though if the radiance is not sufficiently exciting (in the sense defined in Section 3) it will not converge. All this will be seen in the experiments with real images described below.

## 7.3 Experiments with real images

We have tested the algorithm on the two images in Figures 3 and 6 provided to us by M. Watanabe and S. Nayar. These images were generated by a telecentric optic (see [19] for more details) where there is no change in scale for different focus settings. A side effect is that now the real lens diameter is not constant, and therefore we need to correct our optical model according to Figure 2. More precisely, we substitute the diameter  $d$  with the new diameter  $D = \frac{2aZ_F}{Z_F - f}$  where  $a$  and  $f$  are indicated in Figure 2. For this experiment, in order to speed up the computation, we chose to iterate the algorithm for 5 iterations and to compute depth at every other pixel along both

coordinate axes. At points where the radiance is not rich enough, or where the local approximation with an equifocal plane is not valid, the algorithm fails to converge. This explains why in Figure 4 some points are visibly incorrect, and in Figure 7 the depth of the white background is poorly retrieved. A convergence test could be employed, although it would slow down the computations considerably.

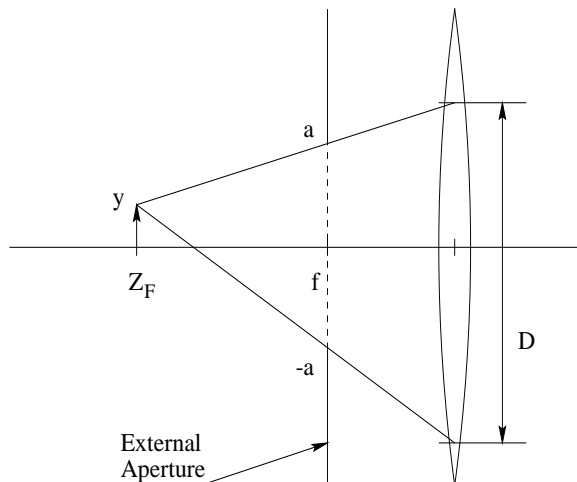


Figure 2: *The modified diameter in the telecentric lens model.*

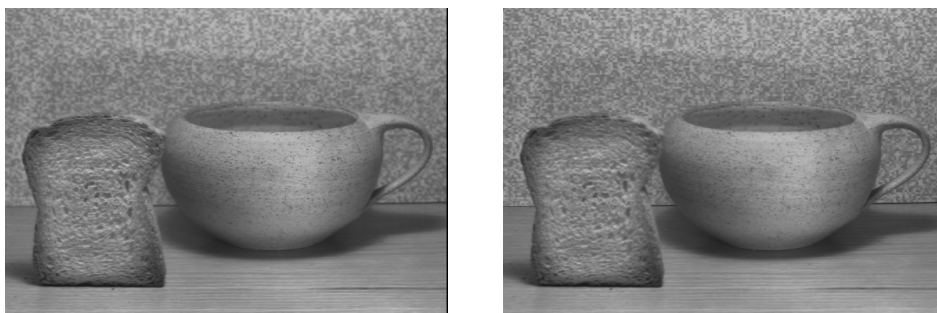


Figure 3: *Original images: near focused (left); far focused (right). The difference between the two images is barely perceivable since the two focal planes are only 340 mm apart.*

## Conclusions

It is often believed that accommodation is a “weak” cue for three-dimensional shape, compared for instance to stereo and motion. On the contrary, we have shown that (unlike all other cues) it is *unambiguous*. We have done so by defining two notions of observability: *weak observability* and *strong observability*. We show 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

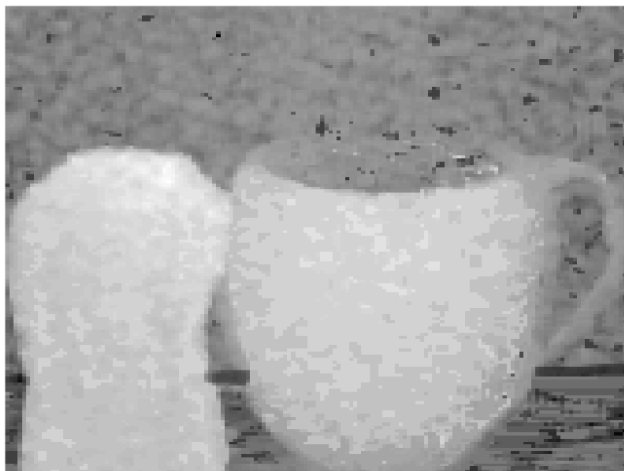


Figure 4: *Reconstructed depth for the scene in figure 3, coded in grayscale.*

to approximate to an arbitrary degree (although never realize exactly) structured patterns that allow distinguishing any two surfaces (i.e. that approximate a *universally exciting* pattern). In the absence of prior information on the radiance of the scene, we show that two surfaces can be distinguished up to a degree of resolution defined by the complexity of the radiance (strong observability).

The equifocal assumption considerably simplifies the analysis and the design of estimation algorithms, which is the reason why it is often encountered in the literature. Any algorithm based upon the equifocal assumption, however, is bound to return biased results, as discussed in Remark 6.

We have proposed a solution to the problem of reconstructing shape and radiance of a scene using I-divergence as a criterion in an infinite-dimensional optimization framework. The algorithm is iterative, and we give 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.

## Acknowledgments

This research was supported by NSF grant IIS-9876145 and ARO grant DAAD19-99-1-0139. We wish to thank S. Nayar for kindly providing us with experimental test data, and H. Jin for assisting with the experiments.

## References

- [1] N. Asada, H. Fujiwara, and T. Matsuyama. Edge and depth from focus. *Intl. J. of Comp. Vision*, 26(2):153–163, 1998.

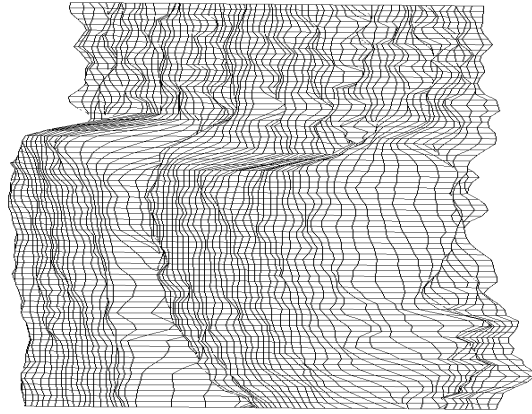


Figure 5: *Reconstructed depth for the scene in figure 3: smoothed mesh.*



Figure 6: *Original images: near focused (left); far focused (right). As in figure 3 the difference between the two is barely perceivable.*

- [2] Various Authors. Special issue on blind system identification and estimation. *Proceedings of the IEEE*, 86(10), 1998.
- [3] W. Boothby. *Introduction to Differentiable Manifolds and Riemannian Geometry*. Academic Press, 1986.
- [4] T. Chan, P. Blongren, P. Mulet, and C. Wong. Total variation blind deconvolution. *IEEE Intl. Conf. on Image Processing*, 1997.
- [5] I. Csiszár. Why least-squares and maximum entropy; an axiomatic approach to inverse problems. *Annals of statistics*, 19:2033–2066, 1991.
- [6] T. Darell and K. Wohn. Depth from focus using a pyramid architecture. *Pattern Recognition Letters*, 11(2):787–796, 1990.



Figure 7: *Reconstructed depth for the scene in figure 6, coded in grayscale. In the uniform region of the background, the radiance is not sufficiently exciting, in the sense defined in Section 3. Therefore, the algorithm cannot converge and the quality of the estimates, as it can be seen, is poor.*

- [7] J. Ens and P. Lawrence. An investigation of methods for determining depth from focus. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15:97–108, 1993.
- [8] D. Luenberger. *Optimization by vector space methods*. Wiley, 1968.
- [9] J. Marshall, C. Burbeck, and D. Ariely. Occlusion edge blur: a cue to relative visual depth. *Intl. J. Opt. Soc. Am. A*, 13:681–688, 1996.
- [10] A. Mennucci and S. Soatto. On observing shape from defocused images. In *Proc. of the Intl. Conf. on Image Analysis and Processing*, pages 550–555, 1999.
- [11] A. Mennucci and S. Soatto. The accommodation cue, part 1: modeling. Essrl technical report 99-001, Washington University, October 1999.
- [12] U. Mudenagudi and S. Chaudhuri. Depth estimation using defocused stereo image pairs. In *Proc. of the Intl. Conf. of Comp. Vision and Pat. Recogn.*, pages 4883–488, 1999.
- [13] S. Nayar and Y. Nakagawa. Shape from focus. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16(8):824–831, 1994.
- [14] A. Pentland. A new sense for depth of field. *IEEE Trans. Pattern Anal. Mach. Intell.*, 9:523–531, 1987.
- [15] Y. Schechner and N. Kiryati. The optimal axial interval in estimating depth from defocus. In *Proc. of the Intl. Conf. of Comp. Vision*, pages 843–848, 1993.

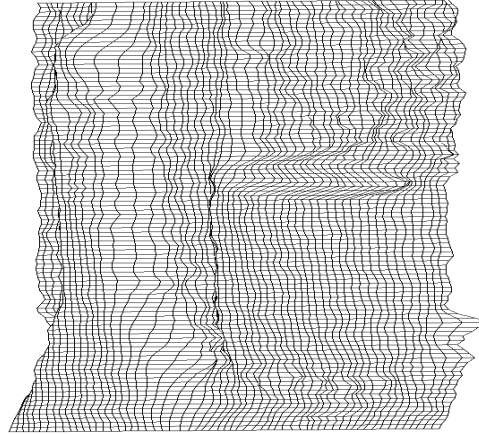


Figure 8: *Reconstructed depth for the scene in figure 6: smoothed mesh.*

- [16] D. Snyder, T. Schulz, and J. O’Sullivan. Deblurring subject to nonnegativity constraints. *IEEE Trans. on Signal Processing*, 40(5):1143–1150, 1992.
- [17] S. Soatto and P. Favaro. A geometric approach to blind deconvolution with application to shape from defocus. In *Intl. Conf. on Computer Vision and Pattern Recognition*, pages 10–17, June 2000.
- [18] M. Subbarao and G. Surya. Depth from defocus: a spatial domain approach. *Intl. J. of Computer Vision*, 13:271–294, 1994.
- [19] M. Watanabe and S. Nayar. Telecentric optics for constant-magnification imaging. Cucs-026-95, Columbia University, 1995.
- [20] M. Watanabe and S. Nayar. Rational filters for passive depth from defocus. *Intl. J. of Comp. Vision*, 27(3):203–225, 1998.
- [21] Y. Xiong and S. Shafer. Depth from focusing and defocusing. In *Proc. of the Intl. Conf. of Comp. Vision and Pat. Recogn.*, pages 68–73, 1993.
- [22] D. Ziou. Passive depth from defocus using a spatial domain approach. In *Proc. of the Intl. Conf. of Computer Vision*, pages 799–804, 1998.

## A Equifocal imaging models

### Uniform imaging model

Under conditions that are commonly accepted in the Vision community, and explained in detail in [11], we can approximate the kernel  $h_u^s(\mathbf{y}, \mathbf{x})$  in the equation (2) with a so-called “pillbox”

function:

$$h_{\mathbf{u}}^{\mathbf{s}}(\mathbf{y}, \mathbf{x}) = h(\mathbf{y} - \mathbf{x}, r(z, \mathbf{u})) \quad (43)$$

where  $\mathbf{s} = z$  and  $r = r(z, \mathbf{u})$  is the “radius” of the kernel, and

$$h(\mathbf{y}, r) = \frac{1}{\pi r^2} \chi_{\{|\mathbf{y}| \leq r\}}(\mathbf{y}) = \begin{cases} \frac{1}{\pi r^2} & |\mathbf{y}| \leq r \\ 0 & \text{elsewhere.} \end{cases} \quad (44)$$

Consider a finite number  $n \geq 1$  of different images  $I_1 \dots I_n$  of an equifocal infinite plane, generated according to the model

$$I_j(\mathbf{y}) = \int_{\mathbb{R}^2} h(\mathbf{y} - \mathbf{x}, r_j) r(\mathbf{x}) d\mathbf{x}. \quad (45)$$

Then, we would like to solve that system of equations for the variables  $(r_1 \dots r_n)$  (with  $r_i > 0$ ) and the radiance  $r$ , given the images; indeed, we will be able to recover the distance  $\mathbf{s} = z$ , given  $\mathbf{u}$  and  $(r_1 \dots r_n)$ <sup>6</sup>. We remark that we are considering an abstract, noiseless, model.

Assume that  $r \neq 0$ , and that  $r$  has finite energy (that is  $r \in \mathcal{L}^2(\mathbb{R}^2)$ ): then, by transforming to the Fourier domain,

$$\tilde{\nabla} f, \quad \hat{I}_j(f) = \hat{h}(f, r_j) \hat{r}(f) \quad (46)$$

so that we understand that the radiance is determined by the radius  $r_j$ . We recall that

$$\hat{h}(f, r_j) = \frac{2}{r|f|} J_1(|f|r) \quad (47)$$

where  $J_1$  is the Bessel function of the first kind.

**Proposition 5** *Suppose that we are given  $n \geq 2$  (different) unfocused images. Suppose that  $r$  is non zero and has finite energy (or, equivalently, suppose that the images  $I_j$  are non zero and have finite energy); let  $(r_1 \dots r_n), r$  be<sup>7</sup> a solution to the equations*

$$\forall j, \mathbf{y} \quad I_j(\mathbf{y}) = \int_{\mathbb{R}^2} h(\mathbf{y} - \mathbf{x}, r_j) r(\mathbf{x}) d\mathbf{x} \quad (48)$$

*Then,  $(r_1 \dots r_n)$  is isolated in  $\mathbb{R}^n$ .*

**Proof:** *Suppose that  $(r_1^* \dots r_n^*), r^*$  is a solution; Let*

$$A = \{f \mid \hat{r}^*(f) = 0\} \cup \{0\}$$

$$D_0 \doteq \{f \mid \hat{I}_j(f) \neq \hat{h}(f, r_j^*) \hat{r}^*(f) \text{ for a } j\}$$

$$D_1 = \{f \mid \hat{h}(f, r_j^*) = 0 \text{ for a } j \leq n\} = \{f \mid J_1(|f|r_j^*) = 0 \text{ for a } j\}$$

---

<sup>6</sup>moreover, if the relationship  $r = r(z, \mathbf{u})$  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 auto-calibration of the model.

<sup>7</sup>By saying that the images are unfocused and different we mean  $r_i > 0$ , and  $r_i \neq r_j$  when  $i \neq j$ .

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  $\hat{I}_j(f) \neq 0$ , and then we define

$$g_j(f, r_j) \doteq \hat{h}(f, r_j) / \hat{I}_j(f)$$

so that  $g_i(f, r_i^*) = 1 / \hat{r}^*(f) \neq 0$  (for any  $i$ ).

Whenever  $f \notin A \cup D_0 \cup D_1$ , furthermore,  $\frac{g_i(f, r_i^*)}{g_j(f, r_j^*)} = 1$ , so  $\frac{g_i(f, r_i)}{g_j(f, r_j)}$  is a well defined real number for  $(r_i, r_j)$  in a neighborhood of  $(r_i^*, r_j^*)$ ; so we define

$$d_{i,j}(f, r_i, r_j) \doteq \log \left( \frac{g_i(f, r_i)}{g_j(f, r_j)} \right)$$

(for  $i \neq j$ ): then,  $d_{i,j}(f, r_i^*, r_j^*) = 0$ .

We can then take the derivative of  $g_i$  with respect to  $r_i$ : we introduce the auxiliary variables  $s_i = |f|r_i$ ; then

$$\frac{\partial}{\partial r_i} g_i(f, r_i) = \frac{1}{\hat{I}_j} 2|f| \frac{\partial}{\partial s_i} \left( \frac{1}{s_i} J_1(s_i) \right) = -\frac{2}{\hat{I}_i(f)r_i} J_2(|f|r_i).$$

We define the set

$$D_2 = \{f \mid J_2(|f|r_i^*) = 0 \text{ for a } i \leq n\}$$

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  $\frac{\partial}{\partial r_i} g_i(f, r_i^*) \neq 0$ <sup>8</sup>; the gradient of  $d$  with respect to  $r_i, r_j$  is then

$$\nabla d_{i,j} = \left( \frac{\frac{\partial}{\partial r_i} g_i}{g_i}, -\frac{\frac{\partial}{\partial r_j} g_j}{g_j} \right) = |f| \left( -\frac{J_2(|f|r_i)}{J_1(|f|r_i)}, \frac{J_2(|f|r_j)}{J_1(|f|r_j)} \right)$$

It is easy to see that  $\frac{\partial}{\partial r_i} d_{i,j}$  and  $\frac{\partial}{\partial r_j} d_{i,j}$  are non zero for  $f \notin A \cup D_0 \cup D_1 \cup D_2$ . Then, it is possible to express  $r_i$  in terms of  $r_j$  (and vice versa) in equation  $\{d_{i,j} = 0\}$ : this implies that, locally in a neighborhood of  $(r_1^* \dots r_n^*)$ , there is at most a parametric curve  $(r_1 \dots r_n) = (r_1(t) \dots r_n(t))$  of solutions. If there is such a curve, then, for any fixed  $i, j \neq i$ , all the gradients

$$\{\nabla d_{i,j}(f), \mid \text{for } f \notin A \cup D_0 \cup D_1 \cup D_2\}$$

would be parallel, that is, there would be constants vectors  $(a_{i,j}, b_{i,j})$  (not zero) such that  $a_{i,j} \frac{\partial}{\partial r_i} d_{i,j}(f, r_i^*, r_j^*) = b_{i,j} \frac{\partial}{\partial r_j} d_{i,j}(f, r_i^*, r_j^*)$ .

The equality

$$a_{i,j} \frac{J_2(|f|r_i^*)}{J_1(|f|r_i^*)} = b_{i,j} \frac{J_2(|f|r_j^*)}{J_1(|f|r_j^*)}$$

that is,

$$a_{i,j} J_2(|f|r_i^*) J_1(|f|r_j^*) = b_{i,j} J_2(|f|r_j^*) J_1(|f|r_i^*)$$

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

$$a_{i,j} J_2(yr_i^*) J_1(yr_j^*) = b_{i,j} J_2(yr_j^*) J_1(yr_i^*) \quad \forall y \in \mathbb{R} \quad (49)$$

---

<sup>8</sup>Incidentally, this tells us that, for many values of  $f$ , it is possible to make explicit the dependence of  $r_i$  on  $\hat{r}(f)$ : this will be mostly useful in studying how the noise affects the solution to (48).

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

$$\begin{aligned} & a_{i,j}((yr_i^*)^2/8 - (yr_i^*)^4/96)((yr_j^*)/2 - (yr_j^*)^3/16) - b_{i,j}((yr_i^*)^2/8 - (yr_i^*)^4/96)((yr_j^*)/2 - (yr_j^*)^3/16) + o(y^5) = 0 \\ & = y^3(a_{i,j}r_i^{*2}r_j^* - b_{i,j}r_i^*r_j^{*2})/16 + y^5(a_{i,j}(r_i^{*2}r_j^{*3}/128 + r_i^*r_j^*/192) - b_{i,j}(r_j^{*2}r_i^{*3}/128 + r_j^*r_i^*/192)) + o(y^5) = 0 \end{aligned}$$

By the theory of Taylor series,

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

or

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

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

$$\begin{aligned} & r_i^* = cr_j^* \\ & cr_j^{*3}(1/128 + 1/192) = c^3r_j^{*3}/128 + r_j^{*3}/192 \end{aligned}$$

or

$$c^3/128 - c(1/128 + 1/192) + 1/192 = (c - 1)(c^2/128 - 1/192) = 0$$

whose solutions are  $c \in \{1, \pm\sqrt{2/3}\}$ ; the first is to be excluded because otherwise  $r_i^* = r_j^*$ , the other two since they do not satisfy (49)

## Gaussian imaging model

A common model for the blur kernel is an isotropic Gaussian. It is more popular for the simplicity of the algorithms it results in, rather than for the accuracy with which it models the image formation process.

**Remark 7 (Heat equations)** Let  $u(x, t)$  be a function with  $x \in \mathbb{R}^n$ ,  $t \geq 0$ . Suppose we want to study the ‘‘heat evolution’’

$$\frac{\partial}{\partial t}u = \frac{1}{2}\Delta u$$

(where the Laplacian is calculated wrt to the  $x$  variable).

To solve the heat equation in  $\mathbb{R}^n$ , we will use the Fourier transform  $\widehat{u}(f, t)$  that is

$$\widehat{u}(f, t) = \int e^{if \cdot x} u(x, t) dx$$

then the heat equation becomes

$$\frac{\partial}{\partial t}\widehat{u} = -\frac{1}{2}|f|^2\widehat{u}$$

(indeed, formally,  $\widehat{\Delta} = -|f|^2$ ) so that the solution is obviously

$$\widehat{u}(f, t) = \widehat{u}(f, 0)e^{-t|f|^2/2}$$

or

$$u(x, t) = u(x, 0) * G(x, t) \quad (50)$$

where

$$G(x, t) \doteq \frac{1}{(t2\pi)^{n/2}} e^{-|x|^2/(2t)}$$

(Indeed,  $\widehat{G}(f, t) = e^{-t|f|^2/2}$ )

Note that if  $G$  is the density of a random vector  $Y$ , then the standard deviation is  $\sigma[Y] = \sqrt{t}$  i.e.

$$\sqrt{\int_{-\infty}^{\infty} |x|^2 G dx} = \sqrt{t}$$

We will then say that the “standard radius” of the kernel  $G(x, t)$  is  $\sqrt{t}$ . Note that the “standard radius” of the pillbox kernel  $h(x, r)$  is  $r/\sqrt{2}$ .

In many works on (de)focusing images it has been assumed that the kernel that models the defocusing  $I = r * G$  is the Gaussian  $G(x, r^2)$ , of “standard radius”  $r$ . We will show in the following what this implies.

**Proposition 6** *Suppose that we are given  $n \geq 2$  (different) unfocused images  $I_1..I_n$ . Suppose that  $r$  is non-zero and has finite energy (or, equivalently, suppose that the images  $I_j$  are non-zero and have finite energy); then, there is a unique vector  $(s_2...s_n)$  in  $\mathbb{R}^{n-1}$ , such that, if  $(r_1...r_n), r$  is a solution to to the equations*

$$\forall j, \mathbf{y} \quad I_j(\mathbf{y}) = \int_{\mathbb{R}^2} G(\mathbf{y} - \mathbf{x}, r_j^2) r(\mathbf{x}) d\mathbf{x} \quad (51)$$

then  $s_i = r_i^2 - r_1^2$

**Proof:** *Indeed, the quantity*

$$d_{i,j}(f, r_i, r_j) \doteq \log\left(\frac{g_i(f, r_i)}{g_j(f, r_j)}\right)$$

*found in the previous proof, takes the value*

$$d_{i,j}(f, r_i, r_j) = (s_j - s_i)|f|^2 + \log\left(\frac{\hat{I}_j(f)}{\hat{I}_i(f)}\right)$$

*so that, for any solution  $(r_1...r_n), r$ , for any  $f$  such that  $\hat{r}(f) \neq 0$ ,*

$$s_j = -\frac{1}{|f|^2} \log\left(\frac{\hat{I}_j(f)}{\hat{I}_1(f)}\right).$$

## B Cramér-Rao lower bound

As for any sensor, camera measurements are subject to noise. Therefore, it is important to consider how this affects the estimation process. In what follows we restrict our analysis to unbiased estimators and compute the Cramér-Rao lower bound (CRLB) in the case of a Poisson imaging model. The main motivation for such a choice is that the process of photons hitting the CCD can be well approximated by a Poisson process. As a consequence, the unknowns involved in the estimation are only non-negative functions, that is a hypothesis we use in developing our algorithm (see Section 6). The parametric model of the image  $b^{\mathbf{s}}(x_i, y_j, r)$  is defined as the convolution between the kernel  $h^{\mathbf{s}}$ , depending on the optical settings and the shape  $\mathbf{s}$ , and the radiance  $r$ . Recalling:

$$b^{\mathbf{s}}(x_i, y_j, r) = \iint h^{\mathbf{s}}(x_i - x, y_j - y)r(x, y)dxdy. \quad (52)$$

The image  $I_0(x_i, y_j)$  is defined as  $b^{\bar{\mathbf{s}}}(x_i, y_j, \bar{r})$  where  $\bar{\mathbf{s}}$  and  $\bar{r}$  are the exact parameters of the scene. Thus, we define:

$$I(x_i, y_j) \sim \mathcal{P}(I_0(x_i, y_j)) \quad (53)$$

as the Poisson process of mean  $I_0$ , and the distribution  $p_{\mathbf{I}}$  as:

$$p_{\mathbf{I}}(I) = \prod_{(x_i, y_j)} e^{-I_0(x_i, y_j)} \frac{I_0(x_i, y_j)^{I(x_i, y_j)}}{I(x_i, y_j)!}. \quad (54)$$

The Cramér-Rao lower bound is defined as:

$$\text{var}[\theta] - I_F^{-1} \geq 0 \quad (55)$$

with

$$[I_F]_{k,m} = -E \left[ \frac{\partial^2 \log(p_{\mathbf{I}}(I))}{\partial \theta_k \partial \theta_m} \right] \quad (56)$$

where  $\text{var}[\cdot]$  is the variance,  $E[\cdot]$  is the expectation and  $I_F$  is the Fisher information matrix. We define  $\theta = [\mathbf{s} \ r]^T$  as the vector of the estimators and refer to its  $k$ -th component as  $\theta_k$ . The inequality is in the sense of positive semi-definiteness. Hence, we have:

$$\log(p_{\mathbf{I}}(I)) = \sum_{(x_i, y_j)} \left[ -I_0(x_i, y_j) + I(x_i, y_j) \log(I_0(x_i, y_j)) - \sum_{i=1}^{I(x_i, y_j)} \log(i) \right]. \quad (57)$$

Taking derivatives the last expression we have:

$$\frac{\partial \log(p_{\mathbf{I}}(I))}{\partial \theta_k} = \sum_{(x_i, y_j)} \frac{\partial I_0(x_i, y_j)}{\partial \theta_k} \left( \frac{I(x_i, y_j)}{I_0(x_i, y_j)} - 1 \right) \quad (58)$$

and it is easy to verify that it satisfies the “regularity” condition:

$$E \left[ \frac{\partial \log(p_{\mathbf{I}}(I))}{\partial \theta_k} \right] = 0 \quad \forall \theta_k \quad (59)$$

that is a necessary condition for the CRLB to hold. Taking derivatives again with respect to  $\theta_m$  we obtain:

$$\frac{\partial^2 \log(p_{\mathbf{I}}(I))}{\partial \theta_k \partial \theta_m} = \sum_{(x_i, y_j)} \left[ \frac{\partial^2 I_0(x_i, y_j)}{\partial \theta_k \partial \theta_m} \left( \frac{I(x_i, y_j)}{I_0(x_i, y_j)} - 1 \right) - \frac{\partial I_0(x_i, y_j)}{\partial \theta_k} \frac{\partial I_0(x_i, y_j)}{\partial \theta_m} \frac{I(x_i, y_j)}{I_0^2(x_i, y_j)} \right]. \quad (60)$$

We are left with computing the expectation of the above expression, that is:

$$E \left[ \frac{\partial^2 \log(p_{\mathbf{I}}(I))}{\partial \theta_k \partial \theta_m} \right] = \sum_{(x_i, y_j)} - \frac{\partial I_0(x_i, y_j)}{\partial \theta_k} \frac{\partial I_0(x_i, y_j)}{\partial \theta_m} \frac{1}{I_0(x_i, y_j)}. \quad (61)$$

Since the vector  $\theta = [\mathbf{s} \ r]^T$  has only two components, the Fisher information matrix  $I_F$  is:

$$I_F = \begin{bmatrix} \sum_{(x_i, y_j)} \left( \frac{\partial I_0(x_i, y_j)}{\partial \mathbf{s}} \right)^2 \frac{1}{I_0(x_i, y_j)} & \sum_{(x_i, y_j)} \frac{\partial I_0(x_i, y_j)}{\partial r} \frac{\partial I_0(x_i, y_j)}{\partial \mathbf{s}} \frac{1}{I_0(x_i, y_j)} \\ \sum_{(x_i, y_j)} \frac{\partial I_0(x_i, y_j)}{\partial \mathbf{s}} \frac{\partial I_0(x_i, y_j)}{\partial r} \frac{1}{I_0(x_i, y_j)} & \sum_{(x_i, y_j)} \left( \frac{\partial I_0(x_i, y_j)}{\partial r} \right)^2 \frac{1}{I_0(x_i, y_j)} \end{bmatrix}. \quad (62)$$

In order to invert the Fisher information matrix we need to compute its determinant:

$$\det(I_F) = \sum_{(x_i, y_j, \bar{x}_i, \bar{y}_j)} \left( \left( \frac{\partial I_0(x_i, y_j)}{\partial r} \right)^2 \left( \frac{\partial I_0(\bar{x}_i, \bar{y}_j)}{\partial \mathbf{s}} \right)^2 - \frac{\partial I_0(x_i, y_j)}{\partial r} \frac{\partial I_0(\bar{x}_i, \bar{y}_j)}{\partial r} \frac{\partial I_0(x_i, y_j)}{\partial \mathbf{s}} \frac{\partial I_0(\bar{x}_i, \bar{y}_j)}{\partial \mathbf{s}} \right) \frac{1}{I_0(x_i, y_j) I_0(\bar{x}_i, \bar{y}_j)} \quad (63)$$

and finally, substituting all in (55):

$$\text{var} \begin{bmatrix} \mathbf{s} \\ r \end{bmatrix} \geq \frac{1}{\det(I_F)} \begin{bmatrix} \sum_{(x_i, y_j)} \left( \frac{\partial I_0(x_i, y_j)}{\partial r} \right)^2 \frac{1}{I_0(x_i, y_j)} & - \sum_{(x_i, y_j)} \frac{\partial I_0(x_i, y_j)}{\partial r} \frac{\partial I_0(x_i, y_j)}{\partial \mathbf{s}} \frac{1}{I_0(x_i, y_j)} \\ - \sum_{(x_i, y_j)} \frac{\partial I_0(x_i, y_j)}{\partial \mathbf{s}} \frac{\partial I_0(x_i, y_j)}{\partial r} \frac{1}{I_0(x_i, y_j)} & \sum_{(x_i, y_j)} \left( \frac{\partial I_0(x_i, y_j)}{\partial \mathbf{s}} \right)^2 \frac{1}{I_0(x_i, y_j)} \end{bmatrix} \quad (64)$$

The partial derivatives of the noise mean  $I_0(x_i, y_j)$  depend on the kernel we assume for the image formation process. Under the usual assumption of Gaussian map we have:

$$h^{\mathbf{s}}(x_i - x, y_j - y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x_i-x)^2 + (y_j-y)^2}{2\sigma^2}} \quad \text{where } \sigma = k \left| 1 - \frac{\mathbf{s}}{u} \right| \text{ for some constant } k \quad (65)$$

where  $u$  is the focal setting and the derivatives are:

$$\frac{\partial I_0(x_i, y_j)}{\partial \mathbf{s}} = \iint -\frac{k \text{sign}(\mathbf{s})}{u} \left( -\frac{2k}{2\pi\sigma^3} + \frac{k((x_i-x)^2 + (y_j-y)^2)}{2\pi\sigma^5} \right) e^{-\frac{(x_i-x)^2 + (y_j-y)^2}{2\sigma^2}} r(x, y) dx dy \quad (66)$$

and

$$\frac{\partial I_0(x_i, y_j)}{\partial r} \Big|_{\nu} = \iint \frac{1}{2\pi\sigma^2} e^{-\frac{(x_i-x)^2 + (y_j-y)^2}{2\sigma^2}} \nu(x, y) dx dy \quad (67)$$

where the partial derivative of  $I_0$  with respect to  $r$  is to be intended as the Fréchet derivative since  $r$  is a continuous function.