

A Geometric Approach to Blind Deconvolution with Application to Shape from Defocus*

Stefano Soatto and Paolo Favaro

Washington University, One Brookings dr. 1127, St.Louis - MO 63130.
{soatto,fava}@ee.wustl.edu

Abstract

We propose a solution to the generic “bilinear calibration-estimation problem” when using a quadratic cost function and restricting to (locally) translation-invariant imaging models. We apply the solution to the problem of reconstructing the three-dimensional shape and radiance of a scene from a number of defocused images. Since the imaging process maps the continuum of three-dimensional space onto the discrete pixel grid, rather than discretizing the continuum we exploit the structure of maps between (finite- and infinite-dimensional) Hilbert spaces and arrive at a principled algorithm that does not involve any choice of basis or discretization. Rather, these are uniquely determined by the data, and exploited in a functional singular value decomposition in order to obtain a regularized solution.

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 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, for instance, from different viewpoints (parallax), such as in *stereo* and *motion*¹. In addition to changing the position of the imaging

device, one can 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. There is a sizeable literature on algorithms to reconstruct shape from a number of images taken with different imaging geometry (shape from defocus) or from a controlled search over geometric parameters (shape from focus) [4].

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”, in communications and information theory “source separation”, in image processing “restoration” or “deblurring”, in tomography “inverse scattering”, in computer vision “generic calibration-estimation problem” [10]. 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 reconstruction of both².

The image formation process naturally takes place in the continuum of three-dimensional space, while image data are typically acquired on a discrete grid (e.g. the CCD array). Most algorithms in the literature entail a discretization in one way or another (sampling, decomposition of functions in the continuum into a combination of basis functions etc.), leaving the obvious problem of choosing the order of the discretization or the basis to the discretion of the user.

In this paper we propose a novel solution to the problem of reconstructing the shape and radiance of a scene when using a quadratic cost function and restricting to invariant integral kernels. Rather than approximating the continuum with a discretization, in our approach the size of the measurement array naturally imposes regularity in the solution, which is obtained in infinite-dimensional space using a functional SVD (singular value decomposition). We exploit the geometry of Hilbert spaces, which

*This research was supported by NSF grant IIS-9876145 and ARO grant DAAD19-99-1-0139. The authors wish to thank John C. Schotland and Joseph A. O’Sullivan for their enlightening suggestions, and Shree Nayar for kindly providing us with test data.

¹Note that we must still rely on prior assumptions on photometry in order to solve the correspondence problem.

²Since neither the light source nor the viewer move, we do not distinguish between *radiance* and *reflectance* of a surface.

makes the analysis simple and intuitive. Our solution results in a straightforward and efficient algorithm that is provably optimal and does not involve any choice of basis or discretization: all of these are determined by the data. We present results on real and simulated images that indicate the potential of our technique.

1.1 Statement of the problem

We are interested in inverting integral equations of the form³

$$I(x, y) = \int h(x, y) dR \quad (x, y) \in D \quad (1)$$

by finding a measure R and a kernel h that satisfy the equation where I is measured on a compact discrete lattice⁴ $D \subset \mathbb{R}^2$.

In order for the problem to admit a nontrivial solution, something needs to be known about the kernel h . We assume that it belongs to a parametric class of functions where some of the parameters – which we call u – are known while others – which we call σ – are not. We write this by indicizing the kernel h_u^σ with u a vector of known parameters. For any u , h_u^σ belongs to a family of kernels that we indicate with \mathcal{H}^σ : $\mathcal{H}^\sigma = \{h_u^\sigma \mid \sigma \in \Sigma\}$ where Σ is a compact set in \mathbb{R}^s for some s . Even so, the problem is well-known to be ill-posed. In general a solution does not exist, due to the fact that (1) is only an approximation of the model that generates the data. We will therefore look for solutions that minimize a suitable *optimization criterion*, for instance a regularized norm $\|\cdot\|$:

$$\hat{h}, \hat{R} \doteq \arg \min_{h_u^\sigma \in \mathcal{H}^\sigma, R} \|n\| \quad \text{subject to} \quad (2)$$

$$I(x, y) = \int h_u^\sigma(x, y) dR + n(x, y) \quad \forall (x, y) \in D. \quad (3)$$

Remark 1 (Choice of optimization criterion) *In his seminal paper [5], Csiszár presents a derivation of “sensible” optimization criteria for the problem above, and concludes that the only two that satisfy a set of consistency axioms are the two-norm – which we address in this paper – and the information-divergence – which we address in [8]. The two criteria result in radically different solutions, for in the first case we can exploit the geometry of Hilbert spaces, while in the second we have to resort to the tools of calculus of variations.*

1.2 Motivations

In digital images, the brightness value recorded at a pixel (x_i, y_j) is obtained by integrating the energy radiated by a certain region of space that depends upon the optical properties of the imaging system (h) as well as on the

physical properties of the environment (R). Typically, neither is known. While data are recorded in a discrete domain D (the pixel array), integration is naturally performed in the continuum \mathbb{R}^3 .

Consider for instance a piecewise smooth surface in space, parameterized by σ . Consider then an imaging system whose geometry can – to a certain extent – be modified by acting on some parameters $u \in \mathcal{U} \subset \mathbb{R}^k$ for some k . For instance, u could be the aperture radius of the lens and the focal length. Due to the additive nature of energy transport phenomena, the image is obtained by integrating the distribution R against a kernel h that depends upon σ and u . The generative model for the image (i.e. the model that generates the measurements I) is therefore of the form (1). We are interested in estimating the shape of the surface σ and the energy distribution R – also called *radiance* – to the extent possible, by measuring a number L of images obtained with different camera settings u_1, \dots, u_L . We want to exploit the fact that, while the energy distribution is naturally integrated in space, measurements are taken on a grid. As we will see, rather than using approximations, this will result in a natural way of enforcing regularity in the solution.

1.3 Relation to previous work

In the literature of computational vision a number of algorithms have been proposed to estimate depth from focus/defocus. The most common assumption is that the scene is a plane parallel to the focal plane (equifocal assumption)[2, 6, 7, 12, 14, 15, 16, 17, 18, 19, 20]. This paper is related to all of the above, since it also relies on the equifocal assumption.

The capability to reconstruct the scene’s shape depends upon the energy distribution it radiates. The conditions on the radiance distribution that allow a unique reconstruction of shape have been recently derived in [13]. Our method can be extended to solve a wider class of problems, as discussed by Koenderink and Van Doorn in [10].

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 [1] for references.

2 An operatorial solution

In this section we introduce the core of our algorithm. We work in function space and use the geometry of operators between Hilbert spaces. For basic results on operators

³See section 2.1 for more details on the notation.

⁴When R is finite-dimensional, such as in point-wise affine structure from motion, we have that $I = HR$ and the problem is known as “factorization”.

between finite and infinite-dimensional Hilbert spaces see, for instance, [11].

2.1 Notation

If we collect a number of images with different control parameters u_l and organize them into an array

$I \doteq [I_{u_1}, \dots, I_{u_L}]^T$, and so for the kernels h_{u_l} , we can get rid of the subscript u and write $I(x, y) = \int h^\sigma(x, y) dR$ for $(x, y) \in D$. The right-hand side can be interpreted as the “virtual image” of a given surface σ radiating energy with a given (spatial) distribution R ,

$R(X, Y, Z): \int h^\sigma(x, y, X, Y, Z) dR(X, Y, Z)$. For scenes made with opaque objects, the integral is restricted to their surface, and therefore it is to be interpreted in the Riemannian sense [3]. In coordinates we write the integral as $\int h^\sigma(x, y, \tilde{x}, \tilde{y}) r(\tilde{x}, \tilde{y}) d\tilde{x} d\tilde{y}$ for $(x, y) \in D$ and a suitably chosen parameterization $(\tilde{x}, \tilde{y}) \in \mathbb{R}^2$; we call r the radiant density⁵. 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

$$I(x_i, y_j) = \int h^\sigma(x_i, y_j, \tilde{x}, \tilde{y}) r(\tilde{x}, \tilde{y}) d\tilde{x} d\tilde{y} \quad (4)$$

for $i = 1 \dots N, j = 1 \dots M$. We now want to write the above equation in a more concise form. To this end, consider the Hilbert space $\mathcal{L}^2(\mathbb{R}^2)$, with inner product $\langle \cdot, \cdot \rangle : \mathcal{L}^2 \times \mathcal{L}^2 \rightarrow \mathbb{R}$ defined by

$$(f, g) \mapsto \langle \langle f, g \rangle \rangle \doteq \int f(x, y) g(x, y) dx dy \quad (5)$$

and norm $\|f\| \doteq \sqrt{\langle \langle f, f \rangle \rangle}$. Consider also the space $\mathbb{R}^{LN \times M} \sim \mathbb{R}^{LNM}$ with the inner product $\langle \cdot, \cdot \rangle : \mathbb{R}^{LN \times M} \times \mathbb{R}^{LN \times M} \rightarrow \mathbb{R}$ defined by

$$(A, B) \mapsto \langle A, B \rangle \doteq \text{Trace}\{AB^T\}. \quad (6)$$

and norm $|A| = \sqrt{\langle A, A \rangle}$. If we interpret points in $\mathbb{R}^{LN \times M}$ as LMN -dimensional vectors, then the inner product is the usual $\langle a, b \rangle \doteq a^T b$. We call the integer $LMN = K$.

2.2 Formalization of the problem

If we model⁶ the radiant density r as a point in $\mathcal{L}^2(\mathbb{R}^2)$, and the image I as a point in \mathbb{R}^K , then the imaging

⁵Strictly speaking, r is the Radon-Nikodym derivative of R and, as such, it is not an ordinary function but, rather, a distribution of measures. In what follows we will ignore such technicalities and assume that we can compute integrals and derivatives in the sense of distributions.

⁶By choosing to work on \mathcal{L}^2 we exclude automatically all harmonic functions. We can do so because it has been proven in [13] that the harmonic component of the radiance does not carry shape information, and therefore our choice entails no loss of generality.

process, as understood in (1), can be represented by an operator H

$$H : \mathcal{L}^2 \rightarrow \mathbb{R}^K; \quad r \mapsto I = Hr. \quad (7)$$

In order to emphasize the dependence of H on σ , we write

$$I = H(\sigma)r. \quad (8)$$

This equation is just another way of writing (1). The original problem can therefore be stated, in more concise form, as

$$\hat{\sigma}, \hat{r} \doteq \arg \min_{\sigma \in \Sigma, r \in \mathcal{L}^2} |I - H(\sigma)r|^2 \quad (9)$$

for a suitable compact set Σ . This notation is not only elegant but also enlightening, for it will allow us to use the geometry of operators between Hilbert spaces to arrive at a principled solution of the blind deconvolution problem (3) that minimizes a quadratic cost function. Before doing so, we review some of the definitions that we will need in the sequel.

2.3 Adjoint and orthogonal projectors

The bounded operator $H : \mathcal{L}^2 \rightarrow \mathbb{R}^K$ admits an adjoint H^* defined by the equation

$$\langle Hr, I \rangle = \langle \langle r, H^*I \rangle \rangle \quad \forall r \in \mathcal{L}^2, I \in \mathbb{R}^K \quad (10)$$

from which we get that

$$H^* : \mathbb{R}^K \rightarrow \mathcal{L}^2; \quad I \mapsto h^T(x, y)I. \quad (11)$$

The (Moore-Penrose) pseudo-inverse $H^\dagger : \mathbb{R}^K \rightarrow \mathcal{L}^2$ is defined such that $r = H^\dagger I$ solves the equation

$$H^*Hr = H^*I \quad (12)$$

when it exists; with an abuse of notation⁷ we could write $H^\dagger = (H^*H)^{-1}H^*$. The orthogonal projector H^\perp is then defined as

$$H^\perp : \mathbb{R}^K \rightarrow \mathbb{R}^K; \quad I \mapsto H^\perp I = (I_d - HH^\dagger)I \quad (13)$$

where I_d is the identity in $\mathbb{R}^{K \times K}$. Note that this is a finite-dimensional linear operator, represented therefore by a matrix. The following proposition, which extends the results of Golub and Pereyra [9], is the key to our approach to blind deconvolution:

Proposition 1 *Let $\hat{\sigma}, \hat{r}$ be local extrema of the functional*

$$\phi(\sigma, r) \doteq |I - H(\sigma)r|^2 \quad (14)$$

⁷We have not defined the “inverse operator” $(\cdot)^{-1}$; however, in the next section we will give an explicit formula for the pseudo-inverse using the singular value decomposition.

and, assuming that H^\dagger exists, let $\tilde{\sigma}$ be a local extremum of the function

$$\psi(\sigma) \doteq |H^\perp(\sigma)I|^2. \quad (15)$$

Furthermore, let \tilde{r} be obtained from $\tilde{\sigma}$ by $\tilde{r} \doteq \chi(\tilde{\sigma})$, where χ is defined as

$$\chi(\sigma) \doteq H^\dagger(\sigma)I. \quad (16)$$

Then $\hat{\sigma}$ is also a local extremum of $\psi(\sigma)$, and $\tilde{\sigma}, \tilde{r}$ are also local extrema of $\phi(\sigma, r)$.

Proof: $\hat{\sigma}$ and \hat{r} are defined by the following coupled equations

$$\begin{cases} \frac{\partial \phi}{\partial \sigma}(\hat{\sigma}, \hat{r}) = 0 \\ D_r \phi(\hat{\sigma}, \hat{r}) = 0 \end{cases} \quad (17)$$

where $D_r \phi$ stands for the Fréchet functional derivative of ϕ with respect to r [11]. On the other hand, $\tilde{\sigma}$ and \tilde{r} are defined by

$$\begin{cases} \frac{d\psi}{d\sigma}(\tilde{\sigma}) = 0 \\ \tilde{r} \doteq \chi(\tilde{\sigma}). \end{cases} \quad (18)$$

Computing the derivatives explicitly, and indicating with a "dot" the derivative with respect to σ , we have that $\frac{\partial \phi}{\partial \sigma} = 2((H(\sigma)r)^T \dot{H}(\sigma)r - I^T \dot{H}(\sigma)r) = 0$, which leads to

$$(H(\hat{\sigma})\hat{r})^T \dot{H}(\hat{\sigma})\hat{r} = I^T \dot{H}(\hat{\sigma})\hat{r} \quad (19)$$

while $D_r \phi = 2(H^*(\sigma)H(\sigma)r - H^*(\sigma)I) = 0$ leads to

$$H^*(\hat{\sigma})H(\hat{\sigma})\hat{r} = H^*(\hat{\sigma})I. \quad (20)$$

Now, the last equation is what defines the pseudo-inverse H^\dagger (see (12)), and therefore it is satisfied, by construction, when

$$\hat{r} = H^\dagger(\hat{\sigma})I = \chi(\hat{\sigma}). \quad (21)$$

This shows that if $\hat{\sigma}$ is a stationary point of ϕ , its corresponding \hat{r} must be of the form $\chi(\hat{\sigma})$. The computation of $\frac{d\psi}{d\sigma} = 2I^T H^\perp \dot{H}^\perp(\sigma)I$ can be obtained from $H^\perp H r = 0 \quad \forall r$, which leads to $\dot{H}^\perp H r + H^\perp \dot{H} r = 0$, and hence to

$$\dot{H}^\perp(\sigma) = -H^\perp(\sigma)\dot{H}(\sigma)H^\dagger(\sigma). \quad (22)$$

\Leftarrow) Let us now assume that $I^T H^\perp \dot{H}^\perp(\tilde{\sigma})I = 0$, and let $\tilde{r} = \chi(\tilde{\sigma})$. We want to show that $\frac{\partial \phi}{\partial \sigma} = 0$, that is (19) is satisfied with $\hat{\sigma} = \tilde{\sigma}$ (that (20) is satisfied follows directly from our choice of \tilde{r} from (21)). To this end, we write⁸

$$(Hr)^T \dot{H}r = I^T (HH^\dagger + H^\perp) \dot{H}r = I^T \dot{H}r \quad (23)$$

where the second term of the right hand side is zero from our assumption that $I^T H^\perp \dot{H}^\perp(\tilde{\sigma})I = 0$ and the expression of \dot{H} in (22).

\Rightarrow) Now let (19) and (20) hold for $\hat{\sigma}, \hat{r}$. We want to show that $I^T H^\perp \dot{H}^\perp(\hat{\sigma})I = 0$. To this end, we write (19) as⁹

$$(HH^\dagger I)^T \dot{H}r = I^T \dot{H}r \quad (24)$$

so that, after rearranging terms, we get that $I^T (I_d - HH^\dagger) \dot{H}H^\dagger I = 0$, but substituting the definition of \dot{H}^\perp , we get that $I^T H^\perp \dot{H}^\perp(\hat{\sigma})I = 0$, which allows us to conclude that (18) is satisfied with $\sigma = \hat{\sigma}$.

⁸In the following we omit the argument $\tilde{\sigma}$ in order to simplify the notation.

⁹For simplicity we omit the argument $\hat{\sigma}$.

Remark 2 The significance of the proposition above consists in the fact that, while (14) is an optimization problem on an infinite-dimensional space, (15) is on a finite-dimensional (and often small) space. Indeed, for the case of shape from focus/defocus that we consider, it is a one-dimensional space. Note also that the statement is non-trivial: in fact, (15) is obtained by multiplying on the left (1) by the singular matrix H^\perp . This can add spurious solutions to the problem, as we know by solving linear systems of equations¹⁰. The proposition shows that, in this specific case, this does not happen.

The conditions, however, impose the existence of the pseudo-inverse, which is equivalent to assuming r belongs to a finite-dimensional subspace of \mathcal{L}^2 of dimension less than K .

2.4 Invariant kernels and the SVD

In order to solve (15) we must be able to compute H^\perp . This, naturally, depends upon the operator H . A big help in the solution comes by assuming that H is shift-invariant, so that Hr can be represented as a convolution product $h * r$. In the case of depth from defocus, this is equivalent to approximating the scene (locally) by a planar patch parallel to the lens at depth σ . In this case, solving (15) reduces to a simple one-dimensional optimization problem, that can be solved in a variety of ways (Newton-Raphson, gradient descent, discrete search etc.)¹¹. The problem, therefore, boils down to computing H^\perp .

In order to do so, we want to express the operator H using its (infinite-dimensional) singular value decomposition. To this end, let $\{\lambda_k\}$, $k = 1, \dots, \infty$ be a sequence of positive scalars sorted in decreasing order, $\{I_k\}$ an orthonormal set in \mathbb{R}^K and $\{r_k\}$ an orthonormal set in \mathcal{L}^2 . We now look for the particular choice of such sets that allows us to express H as

$$H = \sum_{k=1}^K \lambda_k r_k I_k. \quad (25)$$

Note that H maps \mathcal{L}^2 onto \mathbb{R}^K as follows

$$r \mapsto Hr = \sum_{k=1}^K \lambda_k I_k \int r_k(x, y) r(x, y) dx dy. \quad (26)$$

Assuming that the pseudo-inverse exists, it is easy to verify that it is given by

$$H^\dagger = \sum_{k=1}^K \lambda_k^{-1} r_k I_k^T \quad (27)$$

¹⁰For instance, the solution of $Ax = 0$ is $\{x \in \text{Null}(A)\}$, while the solution to $BAx = 0$ is $\{x \in \text{Null}(a)\} \cup \{x \mid Ax \in \text{Null}(B)\}$.

¹¹The equifocal assumption is very powerful, but equally dangerous, as we have pointed out in [13]. Here we will assume that the equifocal assumption is satisfied in a small patch of the image. This will allow us to resolve boundaries within a precision equal to the size of the patch.

while the orthogonal projector is

$$H^\perp = I_d - \sum_{k=1}^K I_k I_k^T. \quad (28)$$

In order for the pseudo-inverse to exist, we need to assume that the singular values λ_k are zero for k greater than an integer $\rho < K$. This is equivalent to assuming that the radiance belongs to a finite-dimensional subspace of \mathcal{L}^2 , which imposes a lower bound on the dimensionality of the data to be acquired (number of blurred images and their size).

The sequences $\{\lambda_k\}$, $\{r_k\}$ and $\{I_k\}$ are found by solving the *normal equations*:

$$\begin{cases} H^* H r_k = \lambda_k^2 r_k \\ H H^* I_k = \lambda_k^2 I_k \end{cases} \quad k = 1 \dots \rho \quad (29)$$

or, making the notation explicit

$$\begin{cases} \int h^T(x, y) h(\tilde{x}, \tilde{y}) r_k(\tilde{x}, \tilde{y}) d\tilde{x} d\tilde{y} = \lambda_k^2 r_k(x, y) \\ \int h(x, y) h^T(x, y) I_k dx dy = \lambda_k^2 I_k \end{cases} \quad (30)$$

for $k = 1 \dots K$. The second of the normal equations (30) can be written as

$$\mathcal{M} I_k = \lambda_k^2 I_k \quad k = 1 \dots \rho \quad (31)$$

where \mathcal{M} is the K -dimensional square symmetric matrix $\int h(x, y) h^T(x, y) dx dy$. Since this is a (finite-dimensional) symmetric eigenvalue problem, there exists a unique decomposition of \mathcal{M} of the form

$$\mathcal{M} = U \Lambda^2 U^T \quad (32)$$

with $U^T U = I_d$, $\Lambda^2 = \text{diag}\{\lambda_1^2 \dots \lambda_\rho^2\}$ and $U = [I_1, \dots, I_\rho]$. We are now left with the first equation in (30) in order to retrieve $r_k(x, y)$. However, instead of solving that directly, we use the adjoint operator H^* to map the basis of \mathbb{R}^K onto a basis of a ρ -dimensional subspace of \mathcal{L}^2 via $H^* I_k = \lambda_k r_k$. Making the notation explicit we have

$$r_k(x, y) = \frac{1}{\lambda_k} h^T(x, y) I_k \quad k = 1 \dots \rho. \quad (33)$$

Remark 3 (Regularization) *In the computation of H^\perp , the sum is effectively truncated at $k = \rho < K$, where the dimension K depends upon the amount of data acquired. As a consequence of the properties of the SVD, the solution obtained enjoys a number of regularity properties. Note that the solution is not the one that we would have obtained by first writing r using a truncated orthonormal expansion in \mathcal{L}^2 , then expanding the kernel h in (1) in series, and then applying the finite-dimensional version of the orthogonal projection theorem.*

Remark 4 (Dimensions) *Just to give the reader an idea on the dimensions at play, usually the normal equations are*

solved locally in a patch around each point in the image. In order for the invariance assumption on the kernel to hold, such windows are usually kept of sizes $M \times N$ in the order of 3×3 pixels to 10×10 pixels. Typically between $L = 2$ and $L = 4$ images are acquired. We choose $N = M = 5$ and $L = 2$, thereby having to compute the SVD of matrices of size 50. All these SVDs can be pre-computed.

We now have all the ingredients to write the recipe.

2.5 Blind deconvolution algorithm

1) **Construct** the matrix¹²

$$\mathcal{M}(\sigma) = \int h^\sigma(x, y) h^{\sigma T}(x, y) dx dy.$$

2) **Compute** its SVD: $\mathcal{M}(\sigma) = U \Lambda^2 U^T$. Let

$$H^\perp(\sigma) = I_d - U_\rho^T U_\rho$$

where U_ρ is the matrix built with the first ρ columns of U .

3) **Minimize** the norm¹³ $|H^\perp(\sigma)I|$ with respect to the depth of the patch σ . Call the minimizer $\hat{\sigma}$.

4) **Restore** the radiance¹⁴

$$\hat{r}(x, y) = h^{\hat{\sigma} T}(x, y) U \bar{\lambda}_{inv}$$

where $\bar{\lambda}_{inv} = [\lambda_1^{-1}, \dots, \lambda_\rho^{-1}]^T$.

Notice that the steps 1) – 2) can be pre-computed off-line for any given value of σ .

Remark 5 (Tradeoffs) *There is a tradeoff between memory and computational speed. Choosing a local descent algorithm in 3), one only needs to store one $K \times K$ matrix, but then needs to compute $H^\perp(\sigma)$ at each step of the iteration. Opting for a discrete search, instead, one needs to store $H^\perp(\sigma_i)$ for a number of depths σ_i , but then only the product $H^\perp(\sigma_i)I$ needs to be computed at each depth.*

Following the derivations in the previous section, as a consequence of proposition 1 and the properties of the SVD, we can conclude that

¹²For certain families of kernels, such as Gaussian ones, the integral can be computed in closed-form without therefore any approximation.

¹³There are a number of ways in which this can be done. Although there exists no closed-form solution, local gradient methods, tangent methods, Newton methods are all viable possibilities. Another alternative consists in pre-computing $H^\perp(\sigma)$ for a number of σ s (however many are necessary in order to achieve the desired resolution in depth), and then simply compute $|H^\perp(\sigma_i)I|$ for all i . Choose the i that leads to the smallest norm, call $\hat{\sigma} \doteq \sigma_i$.

¹⁴If all we are concerned with is the depth σ , this step can be omitted.

Proposition 2 *The algorithm described in 1)–4) converges to a local extremum of the problem (3) for an invariant kernel and a quadratic cost function.*

Remark 6 (Observability) *The conditions under which shape can be uniquely reconstructed from blurred images depend upon the radiance of the scene. As it stands, the algorithm described in section 2.5 seems to return an answer at every point, regardless for the radiance. However, in the presence of radiances which are not “sufficiently exciting” (see [13] for rigorous definitions and characterizations), step 3) of the algorithm 2.5 will return a flat profile that is independent of σ . This is easy to test, and it is possible to associate the local curvature of the function $\phi(\sigma) \doteq |H^\perp(\sigma)I|$ with a reliability measure for the localization of depth, as in figure 6.*

It would be desirable if such conditions could be stated directly in terms of the data I , so as to avoid useless computations at points where depth cannot be recovered (for instance where the radiance is harmonic). A thorough analysis of this aspect of the algorithm is still under way.

3 Experiments

In this section we describe an implementation of the algorithm presented in section 2.5 for the case of Gaussian kernels. We choose Gaussians not because they are a good model of the imaging process, but because they make the analysis and the implementation of the algorithm straightforward. The algorithm does not depend upon this choice, and indeed we are in the process of building realistic models for the kernels of commercial cameras.

3.1 Gaussian kernels

One of the simplest families of kernels are the Gaussians, which have the property of being invariant with respect to convolution. It turns out that this choice is often used in the literature. We recall that the kernel h is a (column) vector obtained by stacking the kernels $h_{u_l}^\sigma(x_i, y_j)$ on top of each other, so we only need to specify the generic kernel, which is

$$h_{u_l}(x_i, y_j, x, y) = \frac{1}{\sqrt{2\pi\kappa(u_l, \sigma)}} e^{-\frac{1}{2} \frac{(x_i - x)^2 + (y_j - y)^2}{\kappa(u_l, \sigma)^2}} \quad (34)$$

$$\doteq h_0(x_i - x, y_j - y, u_l) \quad (35)$$

where the “blur radius” κ depends both on the focal length of image l , u_l , and on the depth of the patch being considered, σ . We organize the kernels into a vector, compute the matrix $\mathcal{M}(\sigma)$ in closed form and evaluate it off-line for 200 values of σ . We then compute their SVD and follow the steps of the algorithm 2.5.

Remark 7 (User’s choices) *Of course, real scenes do not satisfy the equifocal assumption if not locally away from discontinuities. Therefore, we run the algorithm on small patches*

around each point on the image, content with having consistent results only away from discontinuities¹⁵. Since the choice of a family of kernels comes from a model of the optics of the camera, the size of the patches is the only choice involved in implementing our algorithm. We choose it to be 5×5 pixels, as a tradeoff between the validity of the equifocal assumption and compensation for noise.

Having pre-computed $H^\perp(\sigma)$ at 200 values of σ , the algorithm requires 510KFlops at each point. Since these operations are all independent, and therefore highly parallelizable, there is potential for real-time operation on commercial hardware.

A detailed experimental analysis of the performance of the algorithm is best carried out on carefully controlled simulations. However, for the purpose of illustration of the functioning of the algorithm on real images, we show the results on sets of images provided to us by Watanabe and Nayar (figures 1 and 6). Although no rigorous

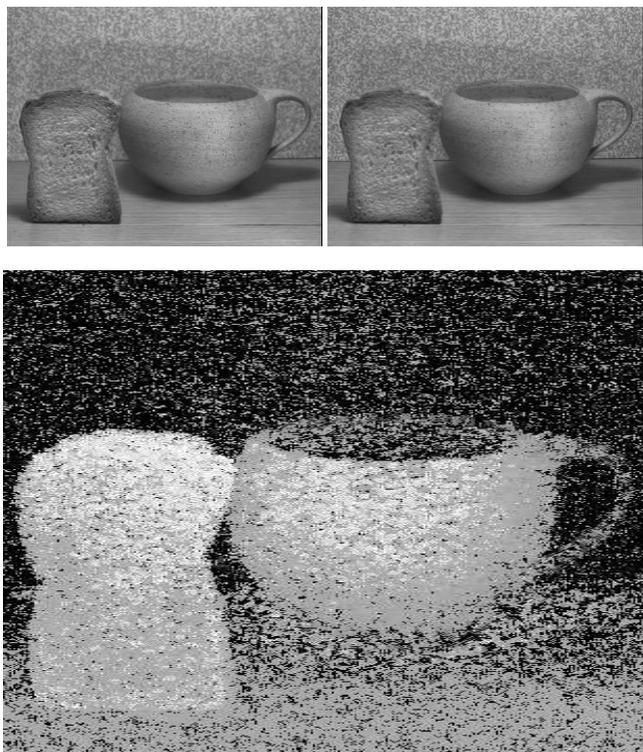


Figure 1: (Top) Two images taken with different focal lengths, courtesy of Watanabe and Nayar [18]. The two focal planes are very close so the difference between near (left) and far (right) is barely visible. (Bottom) Unaltered depth profile as estimated by the algorithm. Grayscale values are proportional to the depth of the scene relative to the first focal plane. Although we have no rigorous evaluation of the estimation error, the qualitative shape of the scene is visible.

ground truth is available, the qualitative shape of the

¹⁵Extending the algorithm to discontinuities is part of our research agenda.

scene seems to have been captured. In this experiment, the algorithm runs on 2 images, with patches of dimension 5×5 . These conditions are challenging for the algorithm, since it forces the rank of the orthogonal projector H^\perp to be at most 50. However, the behavior of the algo-

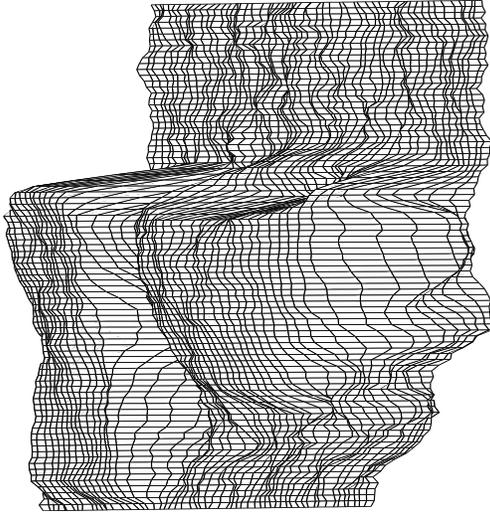


Figure 2: Smoothed mesh of estimated depth for the scene in figure 1. Two images have been used.

gorithm substantially improves when more than two images are available. As shown in figure 3, the average error in a sequence of 50 trials decreases significantly with the number of input images. Furthermore, for a constant number

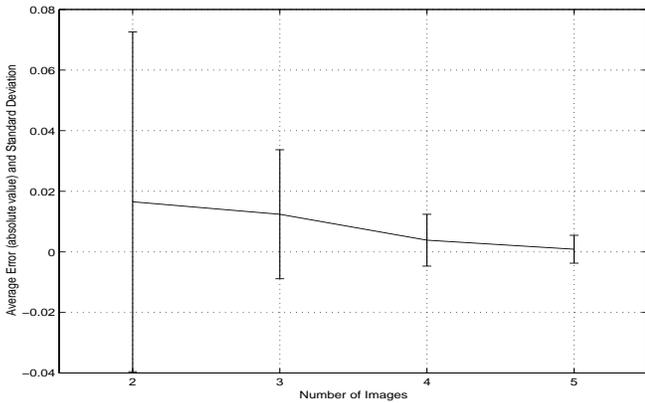


Figure 3: Average reconstruction error (in absolute value) as a function of the number of input images for 50 trials of the same experiment. Performance improves dramatically with more than 2 images.

of images, the average reconstruction error is not uniform across the depth field, as shown in figure 4. Just to give the reader an idea on the profile of the residual cost function, which is minimized with respect to the depth of the patch at each step, we report an example in figure 5, where it can be seen that the residual is neither smooth

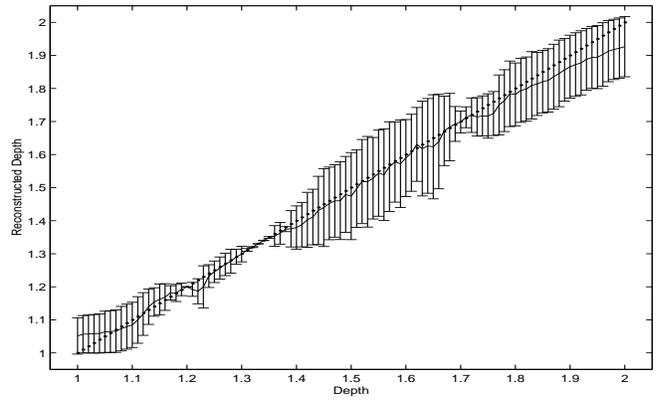


Figure 4: Average reconstruction error as a function of depth for 50 trials of the same experiment. Three blurred images have been used.

nor convex. In figure 6 we show the reconstruction of a real scene together with a measure of its reliability.

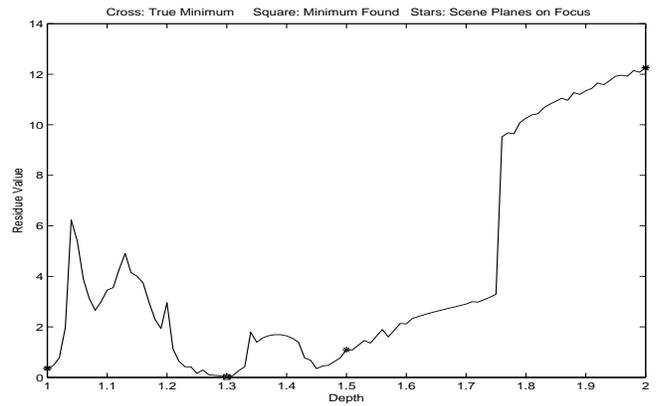


Figure 5: Example of residual to be minimized $|H^\perp(\sigma)I|$: notice that it is non-smooth and non-convex. The true minimum - indicated by a cross - and the estimated one - indicated by a square - coincide. Asterisks indicate the focal planes of image data.

4 Conclusions

We have proposed a solution to the problem of reconstructing the shape and radiance of a scene when using a quadratic cost function under invariant integral imaging models. Rather than approximating the continuum with a discretization, in our approach the size of the measurement array naturally imposes regularity in the solution, which is obtained in infinite-dimensional space using a functional singular value decomposition. We use the structure of maps between (finite and infinite-dimensional) Hilbert spaces, which makes the analysis simple and intuitive. Our solution results in a straightforward and efficient algorithm that does not involve any tuning, choice

of basis, or discretization: all of these are determined by the data.

References

- [1] Various Authors. Special issue on blind system identification and estimation. *Proceedings of the IEEE*, 86(10), 1998.
- [2] N. Asada, H. Fujiwara, and T. Matsuyama. Edge and depth from focus. *Intl. J. of Comp. Vision*, 26(2):153-163, 1998.
- [3] W. Boothby. *Introduction to Differentiable Manifolds and Riemannian Geometry*. Academic Press, 1986.
- [4] S. Chaudhuri and A. Rajagopalan. *Depth from defocus: a real aperture imaging approach*, Springer Verlag, 1999.
- [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.
- [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] P. Favaro and S. Soatto. Shape and reflectance estimation from the information-divergence of blurred images. In *Proc. of the Eur. Conf. of Computer Vision* (in press), June 2000.
- [9] G. Golub and V. Pereyra. The differentiation of pseudo-inverses and nonlinear least-squares problems whose variables separate. *SIAM J. Numer. Anal.*, 10(2):413-532, 1973.
- [10] J. J. Koenderink and A. J. vanDoorn. The generic bilinear calibration-estimation problem. *Intl. J. of Computer Vision*, 23(3): 217-234, 1997.
- [11] D. Luenberger. *Optimization by vector space methods*. Wiley, 1968.
- [12] J. Marshall, C. Burbeck, and D. Ariely. Occlusion edge blur: a cue to relative visual depth. *J. Opt. Soc. Am. A*, 13:681-688, 1996.
- [13] 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.
- [14] S. Nayar and Y. Nakagawa. Shape from focus. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16(8):824-831, 1994.
- [15] A. Pentland. A new sense for depth of field. *IEEE Trans. Pattern Anal. Mach. Intell.*, 9:523-531, 1987.
- [16] 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.
- [17] M. Subbarao and G. Surya. Depth from defocus: a spatial domain approach. *Intl. J. of Computer Vision*, 13:271-294, 1994.
- [18] M. Watanabe and S. Nayar. Rational filters for passive depth from defocus. *Intl. J. of Comp. Vision*, 27(3):203-225, 1998.
- [19] 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.
- [20] D. Ziou. Passive depth from defocus using a spatial domain approach. In *Proc. of the Intl. Conf. of Computer Vision*, pages 799-804, 1998.

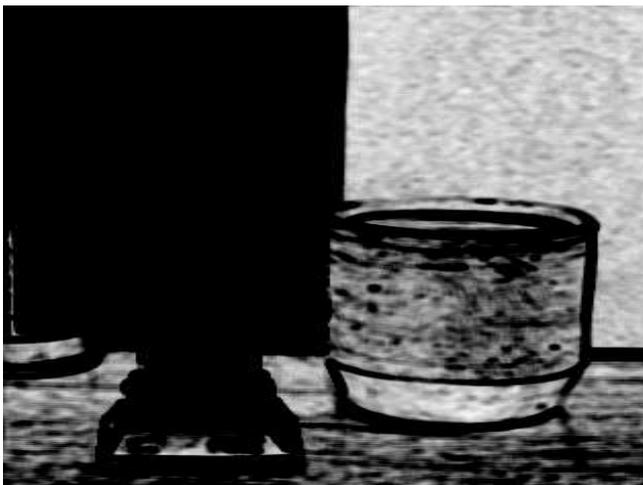


Figure 6: (Top) Two images taken with different focal lengths, courtesy of Watanabe and Nayar [18]. The difference between the two images is barely visible, since the two focal planes are very close. (Center) Reconstructed (relative) unaltered depth profile. (Bottom) reliability parameter computed from the local curvature of the residual function around its minimum. As it can be seen, the estimates of depth corresponding to the uniform region of the background have a high uncertainty associated to them. Note that at occluding boundaries uncertainty is judged to be low by the algorithm, although the actual estimates are unreliable due to the violation of the equifocal assumption.