International Journal of Computer Vision 39(1), 25–39, 2000

Separation of Transparent Layers using Focus

YOAV Y. SCHECHNER
Department of Computer Science, Columbia University, New York NY 10027, USA
yoav@cs.columbia.edu

NAHUM KIRYATI
Department of Electrical Engineering-Systems, Faculty of Engineering, Tel-Aviv University, Ramat-Aviv 69978, Israel
nk@eng.tau.ac.il

RONEN BASRI
Department of Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel
ronen@wisdom.weizmann.ac.il

Received December 2, 1999; Revised November 16, 1999; Accepted April 26, 2000

Abstract. Consider situations where the depth at each point in the scene is multi-valued, due to the presence of a virtual image semi-reflected by a transparent surface. The semi-reflected image is linearly superimposed on the image of an object that is behind the transparent surface. A novel approach is proposed for the separation of the superimposed layers. Focusing on either of the layers yields initial separation, but crosstalk remains. The separation is enhanced by mutual blurring of the perturbing components in the images. However, this blurring requires the estimation of the defocus blur kernels. We thus propose a method for self calibration of the blur kernels, given the raw images. The kernels are sought to minimize the mutual information of the recovered layers. Autofocusing and depth estimation in the presence of semi-reflections are also considered. Experimental results are presented.

Keywords: semireflections, depth from focus, blind deconvolution, blur estimation, enhancement, image reconstruction and recovery, inverse problems, optical sectioning, signal separation, decorrelation

1. Introduction

The situation in which several (typically two) linearly superimposed contributions exist is often encountered in real-world scenes. For example (Darrell and Simoncelli, 1993a; Fujikake et al., 1998), looking out of a car (or room) window, we see both the outside world (termed *real object* (Ohnishi et al., 1996; Oren and Nayar, 1995; Schechner et al., 1999a, 1999b, 1999c, 2000b)), and a semi-reflection of the objects inside, termed *virtual objects*. The treatment of such cases is important, since the combination of several unrelated images is likely to degrade the ability to analyze and understand them. The detection of the phenomenon is of importance itself, since it indicates the presence of a clear, transparent surface in front of the camera, at a distance closer than the imaged objects (Ohnishi et al., 1996; Schechner et al., 1999b, 2000b).

The term *transparent layers* has been used to describe situations in which a scene is semi-reflected from a transparent surface (Bergen et al., 1992; Darrell and Simoncelli, 1993a; Wang and Adelson, 1993). It means that the image is decomposed into depth ordered layers, each with an associated map describing its intensity (and, if applicable, its motion (Wang and Adelson, 1993)). We adopt this terminology, but stress the fact that this work *does not* deal with imaging through an object with variable opacity. Approaches to recovering each of the layers by nulling the others relied mainly on triangulation methods like motion (Bergen et al., 1992; Darrell and Simoncelli, 1993a, 1993b; Irani et al., 1994; Oren and Nayar, 1995; Shizawa and Mase, 1990), and stereo (Borga and Knutsson, 1999; Shizawa, 1993). Algorithms were developed to cope with multiple superimposed motion fields (Bergen et al., 1992; Shizawa and Mase, 1990) and ambiguities in the solutions were discovered (Shizawa, 1992; Weinshall, 1989). Another approach to the problem has been based on polarization cues (Farid and Adelson, 1999; Fujikake et al., 1998; Ohnishi et al., 1996; Schechner et al., 1999a, 1999b, 1999c, 2000b)). However, that approach needs a polarizing filter to be operated with the camera, may be unstable when the angle of incidence is very low, and is difficult to generalize to cases in which more than two layers exist.

In recent years, range imaging relying on the limited depth of field (DOF) of lenses has been gaining popularity. An approach for depth estimation using a monocular system based on focus sensing (Darrel and Wohn, 1988; Engelhardt and Hausler, 1988; Jarvis, 1983; Nair and Stewart, 1992; Nayar, 1992; Nayar et al., 1995; Noguchi and Nayar, 1994; Subbarao and Tyan, 1995; Sugimoto and Ichioka, 1985; Xiong and Shafer, 1993) is termed *Depth from Focus* (DFF) in the computer-vision literature. In that approach, the scene is imaged with different focus settings (e.g., by axially moving the sensor, the object or the lens), thus obtaining image *slices* of the scene. In each slice, a limited range of depth is in focus. Depth is extracted by a search for the slice that maximizes some focus criterion (Hausler and Korner, 1984; Jarvis, 1983; Nair and Stewart, 1992; Nayar, 1992; Noguchi and Nayar, 1994; Subbarao and Tyan, 1995; Torroba et al., 1994; Yeo et al., 1993) (usually related to the two dimensional intensity variations in the region), and corresponds to the plane of best focus. DFF and image-based rendering based on focused slices has usually been performed on opaque (and occluding) layers. In particular, just recently a method has been presented for generating arbitrarily focused images and other special effects performed separately on each occluding layer (Aizawa et al., 2000).

Physical modeling of DOF as applied to processing images of transparent objects has long been considered in the field of microscopy (Agard and Sedat, 1983; Agard, 1984; Castleman, 1979; Conchello and Hansen, 1990; Diaspro et al., 1990; Erhardt et al., 1985; Fay et al.,1983; Itoh et al., 1989; Marcias-Garza et al., 1988; McNally et al., 1994; Preza et al., 1992; Streibl, 1984), where the defocus effect is most pronounced. An algorithm for DFF was demonstrated (Itoh et al., 1989) on a layered microscopic object, but due to the very small depth of field used, the interfering layer was very blurred so no reconstruction process was necessary. Note that microscopic specimens usually contain detail in a continuum of depth, and there is correlation between adjacent layers, so their crosstalk is not as disturbing as in semi-reflections. Fundamental consequences of the imaging operation (e.g. the loss of biconic regions in the three dimensional frequency domain) that pose limits on the reconstruction ability, and the relation to tomography, were discovered (Chiu et al., 1979; Marcias-Garza et al., 1988; Streibl, 1984, 1985; Sundaram and Nayar, 1997). Some of the three dimensional reconstruction methods used in microscopy (Agard and Sedat, 1983; Agard, 1984; Conchello and Hansen, 1990) may be applicable to the case of discrete layers as well.

We study the possibility of exploiting the limited depth of field to detect, separate and recover the intensity distribution of transparent, multi-valued layers. Focusing yields an initial separation, but crosstalk remains. The layers are separated based on the focused images, or by changing the lens aperture. The crosstalk is attenuated by mutual blurring of the disturbing components in the images (Section 2). Proper blurring requires the point spread functions (PSF) in the images to be well estimated. A wrong PSF will leave each recovered layer contaminated by its complementary. We therefore study the effect of error in the PSFs. Then, we propose a method for estimating the PSFs from the raw images (Section 3). It is based on seeking the minimum of the mutual information between the recovered layers. Recovery experiments are described in Section 4. We also discuss the implication of semi-reflections on the focusing process and the depth extracted from it (Section 5). Preliminary and partial results were presented in Schechner et al. (1998, 2000a).

2. Recovery from Focused Slices

2.1. *Using Two Focused Slices*

Consider a two-layered scene. Suppose that either manually or by some automatic procedure (see Section 5), we acquire two images, such that in each image one of the layers is in focus. Assume for the moment that we also have an estimate of the blur kernel operating on each layer, when the camera is focused on the other one. This assumption may be satisfied if the imaging

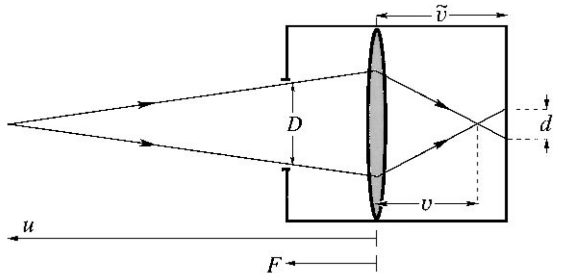

Figure 1. A telecentric imaging system (Watanabe and Nayar, 1996). An aperture D is situated at distance F (the focal length) in front of the lens. An object point at distance u is at best focus if the sensor is at v. If the sensor is at $\tilde{v}$, the image of the point is a blurred spot parameterized by its effective diameter d.

system is of our design, or by calibration. Due to the change of focus settings, the images may undergo a scale change. If a telecentric imaging system (Fig. 1) is used, this problem is avoided (Nayar et al., 1995; Watanabe and Nayar, 1996). Otherwise, we assume that the scale change[1] is corrected during preprocessing (Kubota et al., 1999).

Let layer f_1 be superimposed[2] on layer f_2. We consider only the slices g_a and g_b, in which either layer f_1 or layer f_2, respectively, is in focus. The other layer is blurred. Modeling the blur as convolution with blur kernels,

$$g_a = f_1 + f_2 * h_{2a} \qquad g_b = f_2 + f_1 * h_{1b} \tag{1}$$

(The assumption of a space-invariant response to constant depth objects is very common in analysis of defocused images, and is approximately true for paraxial systems or in systems corrected for aberrations). If a telecentric system is used, $h_{1b} = h_{2a} = h$.

In the frequency domain Eq. (1) take the form

$$G_a = F_1 + H_{2a}F_2 \qquad G_b = F_2 + H_{1b}F_1. \tag{2}$$

Assuming that the kernels are symmetric, $\mathrm{Im}H_{2a} = 0$ and $\mathrm{Im}H_{1b} = 0$, so the real components of G_a and G_b are respectively

$$\begin{aligned} \mathrm{Re}G_a &= \mathrm{Re}F_1 + H_{2a} \cdot \mathrm{Re}F_2 \\ \mathrm{Re}G_b &= \mathrm{Re}F_2 + H_{1b} \cdot \mathrm{Re}F_1, \end{aligned} \tag{3}$$

with similar expressions for the imaginary components of the images. These equations can be visualized as two pairs of straight lines (see Fig. 2). The solution, which corresponds to the line intersection, uniquely exists for

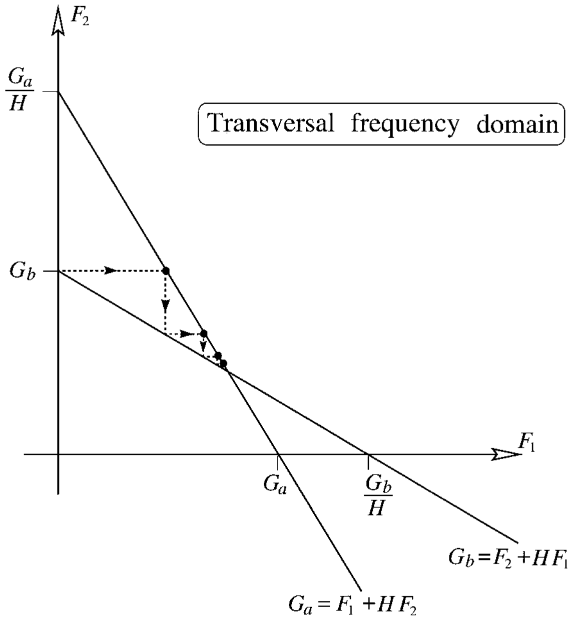

Figure 2. Visualization of the constraints on reconstruction from focused slices and the convergence of a suggested algorithm. For each frequency, the relations (3) between the real components of G_a, G_b, F_1 and F_2 take the form of two straight lines. The visualization of the imaginary parts is similar.

$H_{2a}H_{1b} \neq 1$. Since the imaging system cannot amplify any component ($H_{1b}, H_{2a} \leq 1$), a unique intersection exists unless $H_{2a} = H_{1b} = 1$.

To gain insight, consider a telecentric system (the generalization is straightforward). In this case, $H_{2a} = H_{1b} = H$, and the slopes of the lines in Fig. 2 (representing the constraints) are reciprocal to each other. As $H \to 1$ the slopes of the two lines become similar, hence the solution is more sensitive to noise in G_a and G_b. As the frequency decreases, $H \to 1$, hence at low frequencies the recovery is ill conditioned. Due to energy conservation, the average gray level (DC) is not affected by defocusing. Thus, at DC, $H = 1$. In the noiseless case the constraints on the DC component coincide into a single line, implying an infinite number of solutions. In the presence of noise the lines become parallel and there is no solution. The recovery of the DC component is thus ill posed. This phenomenon is also seen in the three dimensional frequency domain. The image space is band limited by a *missing cone of frequencies* (Chiu et al., 1979; Marcias-Garza et al., 1988), whose axis is in the axial frequency direction ν_v and its apex is at the origin. Recovery of the average intensity in each individual layer is impossible since the information about inter-layer variations of the average transversal intensity is in the missing cone (Sheppard and Gu, 1991). A similar conclusion may be derived from observing the three dimensional frequency domain support that relies on diffraction limited optics (Sundaram and Nayar, 1997).

In order to obtain another point of view on these difficulties, consider the naive inverse filtering approach to the problem given by Eq. (2). In the transversal spatial frequency domain, the reconstruction is

$$\hat{F}_1 = B(G_a - G_b H_{2a}) \qquad \hat{F}_2 = B(G_b - G_a H_{1b}) \tag{4}$$

where

$$B = (1 - H_{1b} H_{2a})^{-1} . \tag{5}$$

As $H \rightarrow 1$, $B \rightarrow \infty$ hence the solution is instable. Note, however, that *the problem is well posed and stable at the high frequencies*. Since H is a LPF, then $B \rightarrow 1$ at high frequencies. As seen in Eqs. (4), the high frequency contents of the slice in which a layer is in focus are retained, while those of the other slice are diminished. Even if high frequency noise is added during image acquisition, it is amplified only slightly in the reconstruction. This behavior is quite opposite to typical reconstruction problems, in which instability and noise amplification appear in the high frequencies.

Iterative solutions have been suggested to similar inversion problems in microscopy (Agard and Sedat, 1983; Agard, 1984; Diaspro et al., 1990) and in other fields. A similar approach was used in Aizawa et al. (2000) to generate special effects on occluding layers, when the inverse filtering needed special care in the low frequency components. The method that we consider can be visualized as progression along vectors in alternating directions parallel to the axes in Fig. 2. It converges to the solution from any initial hypothesis for $|H| < 1$. As $|H|$ decreases (roughly speaking, as the frequency increases), the constraint lines approach orthogonality, thus convergence is faster. A single iteration is described in Fig. 3. This is a version of the Van-Cittert restoration algorithm (Jansson et al., 1970). With slices g_a and g_b as the initial hypotheses for $\hat{f}_1$ and $\hat{f}_2$ respectively, at the l'th iteration

$$\begin{aligned} \hat{F}_1(m) &= \hat{B}(m)\,[G_a - G_b H_{2a}] \\ \hat{F}_2(m) &= \hat{B}(m)\,[G_b - G_a H_{1b}] \end{aligned} \tag{6}$$

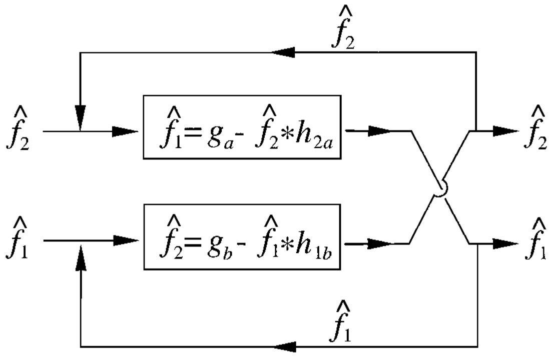
$\hat{f}_2$
$\hat{f}_2$
$\hat{f}_1 = g_a - \hat{f}_2 * h_{2a}$
$\hat{f}_2$
$\hat{f}_1$
$\hat{f}_2 = g_b - \hat{f}_1 * h_{1b}$
$\hat{f}_1$
$\hat{f}_1$

Figure 3. A step in the iterative process. Initial hypotheses for the layers serve as input images for a processing step, based on Eq. (1). The new estimates are fed back as input for the next iteration.

for odd l, where $m = (l + 1)/2$ and

$$\hat{B}(m) = \sum_{k=1}^{m} (H_{1b} H_{2a})^{k-1} . \tag{7}$$

$\hat{B}(m)$ has a major effect on the amplification of noise added to the raw images g_a and g_b (with the noise of the unfocused slice attenuated by H). Again, we see that at high frequencies the amplification of additive noise approaches 1. As the frequency decreases, noise amplification increases. The additive DC error increases linearly with m.

Let us define the *basic solution* as the result of using $m = 1$. Eq. (6) indicates that we can do the recovery directly, without iterations, by calculating the kernel (filter) beforehand. m is a parameter that controls how close the filter $\hat{B}(m)$ is to the inverse filter, and is analogous to regularization parameters in typical inversion methods.

In the spatial domain, Eq. (7) turns into a convolution kernel

$$\begin{aligned} \hat{b}_m(x, y) = \delta(x, y) &+ \underbrace{h_{1b} * h_{2a}}_{\text{once}} \\ &+ \underbrace{\underbrace{h_{1b} * h_{2a}} * \underbrace{h_{1b} * h_{2a}}}_{\text{twice}} + \cdots \\ &+ \underbrace{\underbrace{h_{1b} * h_{2a}} * h_{1b} * \cdots * h_{2a} * \underbrace{h_{1b} * h_{2a}}}_{m-1 \ \text{times}} . \end{aligned} \tag{8}$$

The spatial support of $\hat{b}_m$ is approximately $2dm$ pixels wide, where d is the blur diameter (assuming for a moment that both kernels have a similar support). Here, the finite support of the image has to be taken into account.

The larger m is, the larger the disturbing effect of the image boundaries. The unknown surroundings affect larger portions of the image. It is therefore preferable to limit m even in the absence of noise.

This difficulty seems to indicate at a basic limit to the ability to recover the layers. If the blur diameter d is very large, only a small m can be used, and the initial layer estimation achieved only by focusing cannot be improved much. In this case the initial slices already show a good separation of the individual layers, since in each of the two slices, one layer is very blurred and thus hardly disturbs the other one. On the other hand, if d is small, then in each slice one layer is focused, while the other is nearly focused—creating confusing images. But then, we are able to enhance the recovery using a larger m with only a small effect of the image boundaries. Using a larger m leads, however, to noise amplification and to greater sensitivity to errors in the assumed PSF (see Subsection 2.4).

Example. A simulated scene consists of the image of Lena, as the close object, seen reflected through a window out of which Mt. Shuksan[3] is seen. The original layers appear in the top of Fig. 4. While any of the layers is focused, the other is blurred by a Gaussian kernel with standard deviation (STD) of 2.5 pixels. The slices in which each of the layers is focused are shown in the second row of Fig. 4 (all the images in this work are presented contrast-stretched).

During reconstruction, "mirror" (Aghdasi and Ward, 1996) extrapolation was used for the surroundings of the image in order to reduce the effect of the boundaries. The basic solution ($m = 1$) removes the crosstalk between the images, but it lacks contrast due to the attenuation of the low frequencies. Using $m = 6$, which is equivalent to 13 iterations, improves the balance between the low frequency components to the high ones. With larger m's the results are similar.

2.2. Similarity to Motion-Based Separation

In separating transparent layers, the fact that the high frequencies can be easily recovered, while the low ones are noisy or lost, is not unique to this approach. It also appears in results obtained using motion. Note that, like focus changes, motion leaves the DC component unvaried. In Bergen et al. (1992), the results of motion-based recovery of semi-reflected scenes are clearly highpass filtered versions of the superimposing components. An algorithm presented in Irani et al. (1994) was demonstrated in a setup similar to Bergen et al. (1992). In Irani et al. (1994), one of the objects is "dominant". It can easily be seen there that even as the dominant object is faded out in the recovery, considerable low-frequency contamination remains.

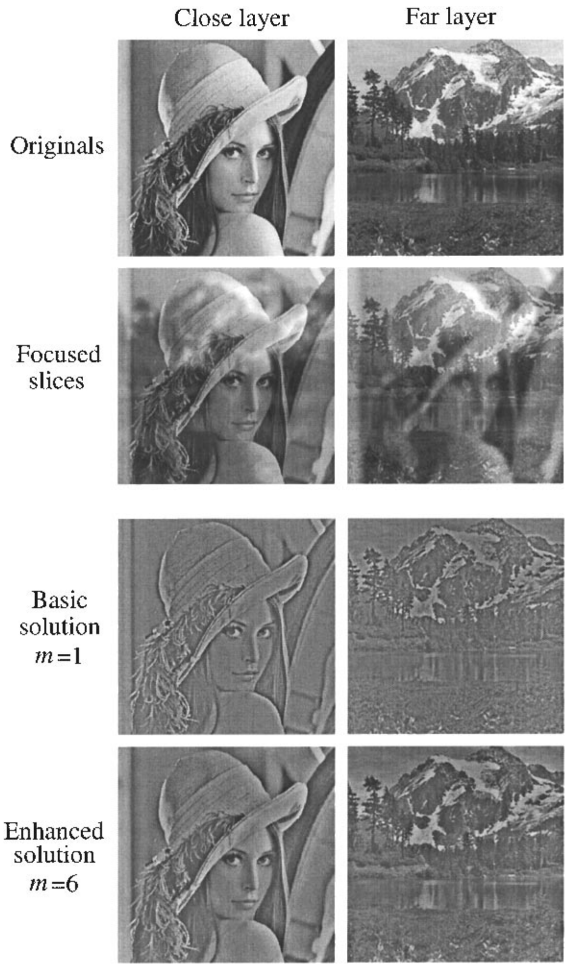

Figure 4. Simulation results. In the focused slices one of the original layers is focused while the other is defocus blurred. The basic solution with the correct kernel removes the crosstalk, but the low frequency content of the images is too low. Approximating the inverse filter with 6 terms ($m = 6$) amplifies the low frequency components.

Shizawa and Mase (1990) have shown that, in regions of translational motion, the spatiotemporal energy of each layer resides in a plane, which passes through the origin in the spatiotemporal frequency domain. This idea was used (Darrel and Simoncelli, 1993a, 1993b) to generate "nulling" filters to eliminate the contribution of layers, thus isolating a single one. However, any two of these frequency planes have a common frequency "line" passing through the origin (the DC), whose components are thus generally inseparable.

These similarities are examples of the unification of triangulation and DOF approaches discussed in Schechner and Kiryati (1998). In general, Schechner and Kiryati (1990) showed that the depth from focus or defocus approaches are manifestations of the geometric triangulation principle. For example, it was shown that for the same system dimensions, the depth sensitivity of stereo, motion blur and defocus blur systems are basically the same. Along these lines, the similarity of the inherent instabilities of separation based on motion and focus is not surprising.

2.3. *Using a Focused Slice and a Pinhole Image*

Another approach to layer separation is based on using as input a pinhole image and a focused slice, rather than two focused slices. Acquiring one image via a very small aperture ("pinhole camera") leads to a simpler algorithm, since just a single slice with one of the layers in focus is needed. The advantage is that the two images are taken without changing the axial positions of the system components, hence no geometric distortions arise. Acquisition of such images is practically impossible in microscopy (due to the significant diffraction effects associated with small objects) but is possible in systems inspecting macroscopic objects.

The "pinhole" image is described by

$$g_0 = \frac{(f_1 + f_2)}{a}, \tag{9}$$

where $1/a$ is the attenuation of the intensity due to contraction of the aperture. This image is used in conjunction with one of the focused slices of Eq. (1), for example g_a. The inverse filtering solution is

$$\hat{F}_1 = S(G_a - aG_0 H_{2a}) \qquad \hat{F}_2 = S(aG_0 - G_a) \tag{10}$$

where

$$S = (1 - H_{2a})^{-1} \ . \tag{11}$$

As in Subsection 2.1, S can be approximated by

$$\hat{S}(m) = \sum_{k=1}^{m} H_{2a}^{k-1} \ . \tag{12}$$

2.4. *Effect of Error in the PSF*

The algorithm suggested in Subsection 2.1 computes $\hat{F}_1(m) = \hat{B}(m)[G_a - G_b H_{2a}]$. We normally assume (Eq. (2)) that $G_a = F_1 + H_{2a}F_2$ and $G_b = F_2 + H_{1b}F_1$. If the assumption holds,

$$\hat{F}_1(m) = F_1(1 - H_{1b}H_{2a})\,\hat{B}(m). \tag{13}$$

Note that, regardless of the precise form of the PSFs, had the imaging PSFs and the PSFs used in the recovery been equal, the reconstruction would have converged to F_1 as $m \to \infty$ when $|H_{1b}|, |H_{2a}| < 1$. In practice, the imaging PSFs are slightly different, i.e., $G_a = F_1 + \tilde{H}_{2a}F_2$ and $G_b = F_2 + \tilde{H}_{1b}F_1$ where

$$\tilde{H}_{1b} = H_{1b} - E_{1b}, \qquad \tilde{H}_{2a} = H_{2a} - E_{2a}, \tag{14}$$

and E_{1b}, E_{2a} are some functions of the spatial frequency. This difference may be due to inaccurate prior modeling of the imaging PSFs or due to errors in depth estimation. The reconstruction process leads to

$$\begin{aligned}\tilde{F}_1 &= [F_1\,(1 - H_{1b}H_{2a}) + E_{1b}H_{2a}F_1 - E_{2a}F_2]\hat{B}(m)\\ &= \hat{F}_1(m) + \hat{B}(m)\,(E_{1b}H_{2a}F_1 - E_{2a}F_2)\,. \end{aligned} \tag{15}$$

A similar relation is obtained for the other layer.

An error in the PSF leads to contamination of the recovered layer by its complementary. The larger $\hat{B}$ is, the stronger is the amplification of this disturbance. Note that $\hat{B}(m)$ monotonically increases with m, within the support of the blur transfer function if $H_{1b}H_{2a} > 0$, as is the case when the recovery PSF's are Gaussians. Note that usually in the low frequencies (which is the regime of the crosstalk) $H_{1b}, H_{2a} > 0$. Thus, we may expect that the best sense of separation will be achieved using a small m, actually, one iteration should provide the least contamination. This is so although the uncontaminated solution obeys $\hat{F} \to F$ as m increases. In other words, decreasing the reconstruction error does not necessarily lead to less crosstalk.

Both H and $\tilde{H}$ (of any layer) are low-pass filters that conserve the average value of the images. Hence, $E \approx 0$ at the very low and at the very high frequencies, i.e., E is a bandpass filter. However, $\hat{B}(m)$ amplifies the low frequencies. At the low frequencies, their combined effect may have a finite or infinite limit as $m \to \infty$, depending on the PSF models used.

Continuing with the example shown in Fig. 4, where the imaging PSF had an STD of $r = 2.5$ pixels, the

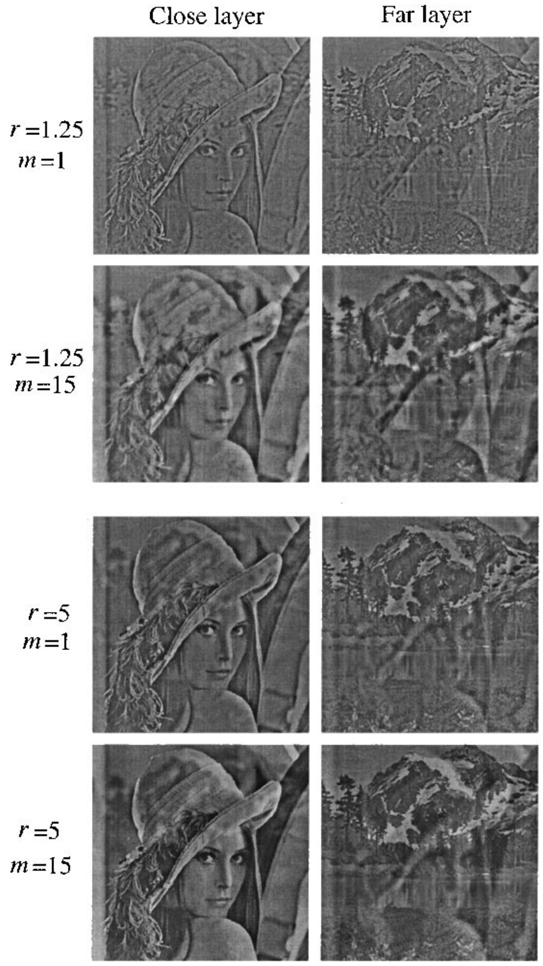

Figure 5. Simulated images when using the wrong PSF in the reconstruction. The original blur kernel had a STD of $r = 2.5$ pixels. Crosstalk between the recovered layers is seen clearly if the STD of the kernels used is 1.5 or 5 pixels. The contamination increases with m.

effects of using a wrong PSF in the reconstruction are demonstrated in Fig. 5. When the PSF used in the reconstruction has STD of 1.25 pixels, negative traces remain (i.e., brighter areas in one image appear as darker areas in the other). When the PSF used in the reconstruction has STD of 5 pixels, positive traces remain (i.e., brighter areas in one image appear brighter in the other). The contamination is slight in the basic solution ($m = 1$), but is more noticeable with larger m's, that is, when $\hat{B} \to B$. So, the separation seems worse, even though each of the images has a better balance (due to the enhancement of the low frequencies).

We can perform the same analysis for the method described in subsection 2.3. Now there is only one filter involved, H_{2a}, since the layer f_1 is focused. Suppose that, in addition to using H_{2a} in the reconstruction rather than the true imaging transfer function $\tilde{H}_{2a}$, we inaccurately use the scalar a rather than the true value $\tilde{a}$ used in the imaging process. Let e denote the relative error in this parameter, $e = (a - \tilde{a})/\tilde{a}$. We obtain that

$$\tilde{F}_1 = \hat{F}_1(m) - eH_{2a}\hat{S}(m)F_1 - (E_{2a} + eH_{2a})\hat{S}(m)F_2, \tag{16}$$

$$\tilde{F}_2 = \hat{F}_2(m) + (E_{2a} + e)\hat{S}(m)F_2 + e\hat{S}(m)F_1, \tag{17}$$

where here $\hat{F}_1(m)$ and $\hat{F}_2(m)$ are the results had the imaging defocus kernel been the same as the one used in the reconstruction and had $a = \tilde{a}$. Note the importance of the estimation of a: if $e = 0$ then $\tilde{F}_2$ (the defocused layer) is recovered uncontaminated by F_1. However, even in this case $\tilde{F}_1$ (the focused layer) will have a contamination of F_2, amplified by $\hat{S}(m)$ and E_{2a}.

3. Seeking the Blur Kernels

The recovery methods outlined in Section 2 are based on the use of known, or estimated blur kernels. If the imaging system is of our design, or if it is calibrated, and in addition we have depth estimates of the layers obtained during the focusing process (e.g., as will be described in Section 5), we may know the kernels a-priori. Generally, however, the kernels are unknown. Even a-priori knowledge is sometimes inaccurate. We thus wish to achieve self-calibration, i.e., to estimate the kernels out of the images themselves. This will enable blind separation and restoration of the layers.

To do that, we need a criterion for layer separation. Note that the method for estimating the blur kernels based on minimizing the fitting error in different layers as in Aizawa et al. (2000) may fail in this case as the layers are transparent and there is no unique blur kernel at each point. Moreover, the fitting error is not a criterion for separation. Assume that the statistical dependence of the real and virtual layers is small (even zero). This is reasonable since they usually originate from unrelated scenes. The Kullback-Leibler distance measures how far the images are from statistical independence, indicating their mutual information (Cover and Thomas, 1991). Let the probabilities for certain values $\check{f}_1$ and $\check{f}_2$ be $P(\check{f}_1)$ and $P(\check{f}_2)$, respectively. In practice these probabilities are estimated by the histograms of the recovered images. The joint probability is $P(\check{f}_1, \check{f}_2)$, which is in practice estimated by the joint histogram of the images, that is, the relative number of pixels in which $\tilde{f}_1$ has a certain value $\check{f}_1$ and $\tilde{f}_2$ has a certain value $\check{f}_2$ at corresponding pixels. The mutual information is then

$$\mathcal{I}(\tilde{f}_1, \tilde{f}_2) = \sum_{\check{f}_1, \check{f}_2} P(\check{f}_1, \check{f}_2) \log \frac{P(\check{f}_1, \check{f}_2)}{P(\check{f}_1)\, P(\check{f}_2)}. \qquad (18)$$

In this approach we assume that if the layers are correctly separated, each of their estimates contains *minimum information* about the other. Mutual information was suggested and used as a criterion for alignment in Thevenaz and Unser (1998), and Viola and Wells (1997), where its maximum was sought. We use this measure to look for the highest discrepancy between images, thus minimizing it. The distance (Eq. (18)) depends on the quantization of $\tilde{f}_1$ and $\tilde{f}_2$, and on their dynamic range, which in turn depends on the brightness of the individual layers f_1 and f_2. To decrease the dependence on these parameters, we performed two normalizations. First, each estimated layer was contrast-stretched to a standard dynamic range. Then, $\mathcal{I}$ was normalized by the mean entropy of the estimated layers, when treated as individual images. The self information (Cover and Thomas, 1991) (entropy) of $\tilde{f}_1$ is

$$\mathcal{H}(\tilde{f}_1) = -\sum_{\check{f}_1} P(\check{f}_1) \log P(\check{f}_1), \qquad (19)$$

and the expression for $\tilde{f}_2$ is similar. The measure we used is

$$\mathcal{I}_n(\tilde{f}_1, \tilde{f}_2) = \frac{\mathcal{I}(\tilde{f}_1, \tilde{f}_2)}{[\mathcal{H}(\tilde{f}_1) + \mathcal{H}(\tilde{f}_2)]/2}, \qquad (20)$$

indicating the ratio of mutual information to the self information of a layer.

The recovered layers depend on the kernels used. Therefore, the problem of seeking the kernels can be stated as a minimization problem:

$$[\hat{h}_{1b}, \hat{h}_{2a}] = \arg \min_{h_{1b}, h_{2a}} \mathcal{I}_n(\tilde{f}_1, \tilde{f}_2). \qquad (21)$$

According to Subsection 2.4, errors in the kernels lead to crosstalk (contamination) of the estimated layers, which is expected to increase their mutual information.

There are generally many degrees of freedom in the form of the kernels. On the other hand, the kernels are constrained: they are non-negative, they conserve energy etc. To simplify the problem, the kernels can be assumed to be Gaussians. Then, the kernels are parameterized only by their standard deviations (proportional to the blur radii). This limitation may lead to a solution that is suboptimal but easier to obtain.

Another possible criterion for separation is decorrelation. Decorrelation was a necessary condition for the recovery of semi-reflected layers by independent components analysis in Farid and Adelson (1999), and by polarization analysis in Schechner et al. (1999b, 1999c). Note that requiring decorrelation between the estimated layers is based on the assumption that the original layers are decorrelated: that assumption is usually only an approximation.

To illustrate the use of these criteria, we search for the optimal blur kernels to separate the images shown in the second row of Fig. 4. Here we simplified the calculations by restricting both kernels to be isotropic Gaussians of the same STD, as these were indeed the kernels used in the synthesis. Hence, the correlation and mutual information are functions of a single variable.[4] As seen in Fig. 6, using the correct kernel (with STD of 2.5 pixels) yields decorrelated basic solutions ($m = 1$), with minimal mutual information ($\mathcal{I}_n$ is plotted). The positive correlation for larger values of assumed STD, and the negative correlation for smaller values, is consistent with the visual appearance of positive and negative traces in Fig. 5. Observe that, as expected from the theory, in Fig. 5 the crosstalk was stronger for larger m. Indeed, in Fig. 6 the absolute correlation and mutual information are greater for $m = 6$ than for $m = 1$ when the wrong kernel is used.

In a different simulation, the focused slices corresponding to the original layers shown in the top of Fig. 4 were created using an exponential imaging kernel rather than a Gaussian, but the STD was still 2.5. The recovery was done with Gaussian kernels. The correlation and mutual information curves (as a function of the assumed STD) were similar to those seen in Fig. 6. The minimal mutual information was however at STD of $r = 2.2$ pixels. There was no visible crosstalk in the resulting images.

The blurring along the sensor raster rows may be different than the blurring along the columns. This is because blurring is caused not only by the optical processes, but also from interpixel crosstalk in the sensors, and the raster reading process in the CCD. Moreover, the inter-pixel spacing along the sensor rows is generally different than along the columns, thus even the optical blur may affect them differently. We assigned a different blur "radius" to each axis: r^{row} and r^{column}. When

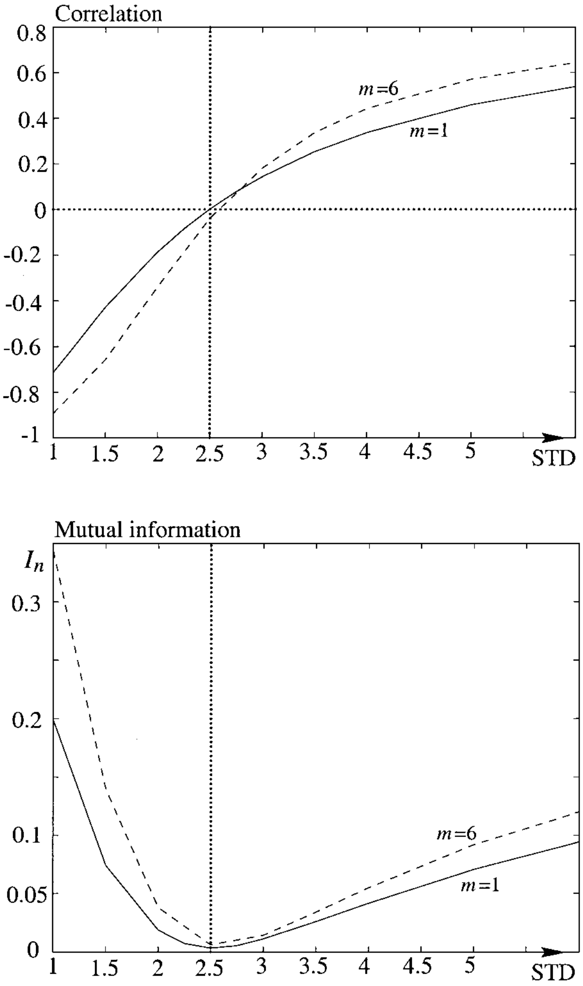

Figure 6. [Solid] At the assumed kernel STD of $r = 2.5$ pixels the basic solutions are decorrelated and have minimal mutual information (shown normalized), in consistency with the true STD. [Dashed] The absolute correlation and the mutual information are larger for a large value of m.

two slices are used, as in Subsection 2.1, there are two kernels, with a total of four parameters. Defining the parameter vector $\mathbf{p} \equiv (r_{1b}^{\text{row}}, r_{1b}^{\text{column}}, r_{2a}^{\text{row}}, r_{2a}^{\text{column}})$, the estimated vector $\hat{\mathbf{p}}$ is

$$\hat{\mathbf{p}} = \arg \min_{\mathbf{p}} \mathcal{I}_n[\tilde{f}_1(\mathbf{p}), \tilde{f}_2(\mathbf{p})]. \qquad (22)$$

When a single focused slice is used in conjunction with a “pinhole” image, as described in Subsection 2.3, the problem is much simpler. There are three parameters to determine: $r_{2a}^{\text{row}}, r_{2a}^{\text{column}}$ and a. The parameter a is easier to obtain as it indicates the ratio of the light energy in the wide-aperture image relative to the pinhole image. Ideally, it is the square of the reciprocal of the ratio of the *f-numbers* of the camera, in the two states. If, however, the optical system is not calibrated, or if there is automatic gain control in the sensor, this ratio is not an adequate estimator of a. a can then be estimated by the ratio of the average values of the images, for example. Such an approximation may serve as a starting point for better estimates.

When using the decorrelation criterion in the multi-parameter case, there may be numerous parameter combinations that lead to decorrelation, but will not all lead to the minimum mutual information, or to good separation. If $\mathbf{p}$ is N-dimensional, the zero-correlation constraint defines a $N - 1$ dimensional hypersurface in the parameter space. It is possible to use this criterion to obtain initial estimates of $\mathbf{p}$, and search for minimal mutual information within a lower dimensional manifold. For example, for each combination of r^{row} and r^{column}, a that leads to decorrelation can be found (near the rough estimate based on intensity ratios). Then the search for minimum mutual information can be limited to a subspace of only two parameters.

4. Recovery Experiments

4.1. Recovery from Two Focused Slices

A print of the “Portrait of Doctor Gachet” (by van-Gogh) was positioned closely behind a glass window. The window partly reflected a more distant picture, a part of a print of the “Parasol” (by Goya). The $f\#$ was 5.6. The two focused slices[5] are shown at the top of Fig. 7. The cross correlation between the raw (focused) images is 0.98. The normalized mutual information is $\mathcal{I}_n \approx 0.5$ indicating that significant separation is achieved by the focusing process, but that substantial crosstalk remains.

The optimal parameter vector $\hat{\mathbf{p}}$ in the sense of minimum mutual information is [1.9, 1.5, 1.5, 1.9] pixels, where r_{1b} corresponds to the blur of the close layer, and r_{2a} corresponds to the blur of the far layer. With these parameters, the basic solution ($m = 1$) shown at the middle row of Fig. 7 has $\mathcal{I}_n \approx 0.006$ (two orders of magnitude better than the raw images). Using $m = 5$ yields better balance between the low and high frequency components, but $\mathcal{I}_n$ increased to about 0.02. We believe that this is due to the error in the PSF model, as discussed above.

In another example, a print of the “Portrait of Armand Roulin” (by van-Gogh) was positioned closely

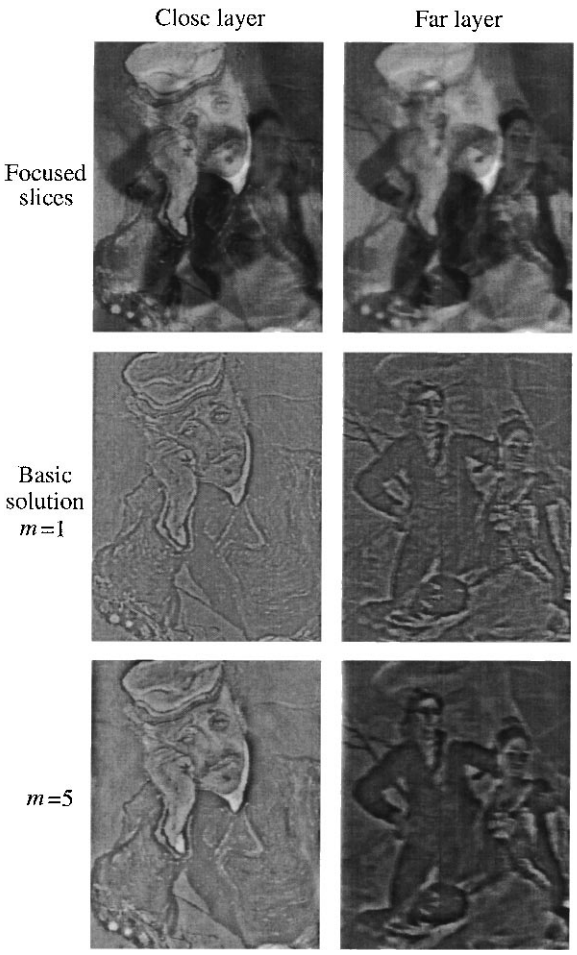

Figure 7. [Top] The slices in which either of the transparent layers is focused. [Middle row] The basic solution ($m = 1$). [Bottom row]: Recovery with $m = 5$.

behind a glass window. The window partly reflected a more distant picture, a print of a part of the "Miracle of San Antonio" (by Goya). As seen in Fig. 8, the "Portrait" is hardly visible in the raw images. The cross correlation between the raw (focused) images is 0.99, and the normalized mutual information is $\mathcal{I}_n \approx 0.6$. The optimal parameter vector $\hat{\mathbf{p}}$ here is [1.7, 2.4, 1.9, 2.1] pixels. With these parameters $\mathcal{I}_n \approx 0.004$ at the basic solution, rising to about 0.02 for $m = 5$.

In a third example, the scene consisted of a distant "vase" picture that was partly-reflected from the glass-cover of a closer "crab" picture. The imaging system was telecentric (Nayar et al., 1995; Watanabe and Nayar, 1996), so no magnification corrections were needed. The focused slices and the recovered layers are shown in Fig. 9. For the focused slices $\mathcal{I}_n \approx 0.4$, and

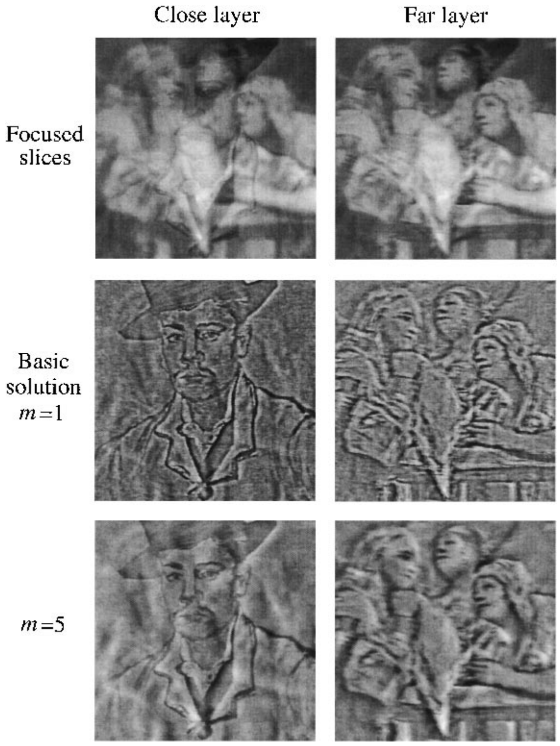

Figure 8. [Top] The slices in which either of the transparent layers is focused. [Middle row] The basic solution. [Bottom row]: Recovery with $m = 5$.

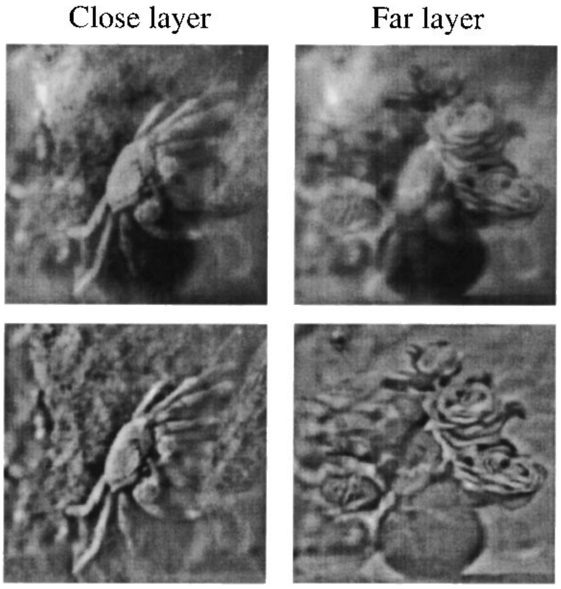

Figure 9. [Top] The slices in which either of the transparent layers is focused. [Bottom] The basic solution for the recovery of the "crab" (left) and "vase" (right) layers.

the cross correlation is 0.95. The optimal parameter vector $\hat{\mathbf{p}}$ in the sense of minimum mutual information is [4,4,11,1] pixels. The basic recovery, using $\hat{B}(1)$, are shown in the bottom of Fig. 9. The crosstalk is significantly reduced. The mutual information $\mathcal{I}_n$ and correlation decreased dramatically to 0.009 and 0.01, respectively.

4.2. *Recovery from a Focused Slice and a Pinhole Image*

The scene consisted of a print of the "Portrait of Armand Roulin" as the close layer and a print of a part of the "Miracle of San Antonio" as the far layer. The imaging system was not telecentric, leading to magnification changes during focusing. Thus, in such a system it may be preferable to use a fixed focus setting, and change the aperture between image acquisitions. The "pinhole" image was acquired using the state corresponding to the $f\# = 11$ mark on the lens, while the wide aperture image was acquired using the state corresponding to the $f\# = 4$ mark. We stress that we have not calibrated the lens, so these marks do not necessarily correspond to the true values. The slice in which the far layer is focused (using the wide aperture) is shown in the top left of Fig. 10. In the "pinhole" image (top right), the presence of the "Portrait" layer is more noticeable.

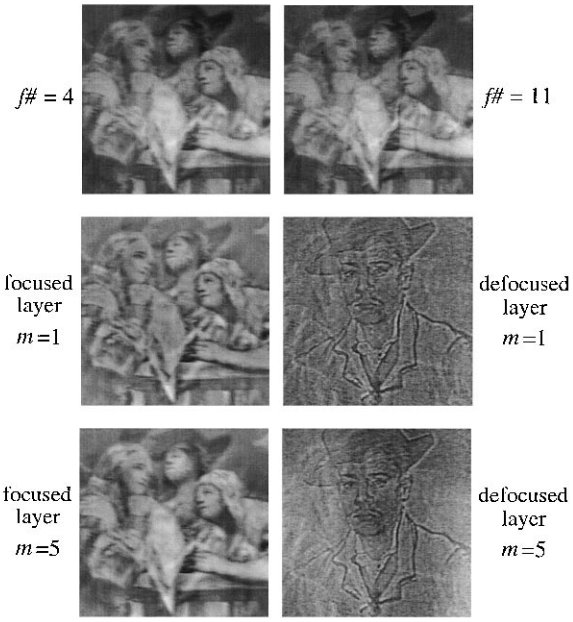

Figure 10. [Top left] The slice in which the far layer is focused, when viewed with the wide aperture. [Top right] The "pinhole" image. [Middle row]: The basic recovery. [Bottom row]: Recovery with $m = 5$.

According to the ratio of the $f\#$'s, the wide aperture image should have been brighter than the "pinhole" image by $(11/4)^2 \approx 7.6$. However, the ratio between the mean intensity of the wide aperture image to that of the pinhole image was 4.17, not 7.6. This could be due to poor calibration of the lens by its manufacturer, or because of some automatic gain control in the sensor. We added a to the set of parameters to be searched in the optimization process. In order to get additional cues for a, we calculated ratios of other statistical measures: the ratios of the STD, median, and mean absolute deviation were 4.07, 4.35 and 4.22, respectively. We thus let a assume values between 4.07 and 4.95. In this example we demonstrate the possibility of using decorrelation to limit the minimum mutual information search. First, for each hypothesized pair of blur diameters, the parameter a that led to decorrelation of the basic solution was sought. Then, the mutual information was calculated over the parameters that cause decorrelation. The blur diameters that led to minimal mutual information at $m = 1$ were $r^{\text{row}} = r^{\text{column}} = 11$ pixels, with the best parameter a being 4.28. The reconstruction results are shown in the middle row of Fig. 10. Their mutual information (normalized) is 0.004.

Using a larger m with these parameters increased the mutual information, so we looked for a better estimate, minimizing the mutual information after the application of $\hat{B}(m)$. For $m = 5$ the resulting parameters were different: $r^{\text{row}} = r^{\text{column}} = 17$ pixels, with $a = 4.24$. The recovered layers are shown in the bottom row of Fig. 10. Their mutual information (normalized) is 0.04. As discussed before, the increase is probably due to inaccurate modeling of the blur kernel.

5. Obtaining the Focused Slices

5.1. *Using a Standard Focusing Technique*

We have so far assumed that the focused slices are known. We now consider their acquisition using focusing as in Depth from Focus (DFF) algorithms. Depth is sampled by changing the focus settings, particularly the sensor plane. According to Abbott and Ahuja (1993), Krishnan and Ahuja (1996) and Schechner and Kiryati (1998, 1999), the sampling should be at depth of field intervals, for which $d \approx \Delta x$, where Δx is the inter-pixel period (similar to stereo (Schechner and Kiryati, 1998)). An imaging system telecentric on the image side (Nayar et al., 1995; Watanabe and Nayar, 1996) is a preferred configuration, since it ensures constant magnification as the sensor is put out of focus. For such a system it is easy to show that the geometrical-optics blur-kernel diameter is $d = D\Delta v/F$, where D is the aperture width, F is the focal length (see Fig. 1), and Δv is the distance of the sensor plane from the plane of best focus. The axial sampling period is therefore $\Delta v \approx F\Delta x/D$. The sampling period requirement can also be analyzed in the frequency domain, as in Sundaram and Nayar (1997).

Focus calculations are applied to the image slices acquired. The basic requirement from the focus criterion is that it will reach a maximum when the slice is in focus. Most criteria suggested in the literature (Itoh et al., 1989; Jarvis, 1983; Nayar, 1992; Noguchi and Nayar, 1994; Subbarao and Tyan, 1995; Torroba et al., 1994; Yeo et al., 1993) are sensitive to two dimensional variations in the slice.[6] Local focus operators yield "slices of local focus-measure", *FOCUS* $(x, y, \tilde{v})$, where $\tilde{v}$ is the axial position of the sensor (see Fig. 1). If we want to find the depth at a certain region (patch) (Nair and Stewart, 1992), and the scene is composed of a single layer, we can average *FOCUS* $(x, y, \tilde{v})$ over the region, to obtain *FOCUS* $(\tilde{v})$ from which a single valued depth can be estimated. This approach is inadequate in the presence of multiple layers. Ideally, each of them alone would lead to a main peak[7] in *FOCUS* $(\tilde{v})$. But, due to mutual interference, the peaks can move from their original positions, or even merge into a single peak in some "average" position, thus spoiling focus detection.

This phenomenon can be observed in experimental results. The scene, the focused slices of which are shown in Fig. 9, had the "crab"and the "vase" objects at distances of 2.8 m and 5.3 m from the lens, respectively. The details of the experimental imaging system are described in Schechner et al. (1998). Depth variations within these objects were negligible with respect to the depth of field. Extension of the STD of the PSF by about 0.5 pixels was accomplished by moving the sensor array 0.338 mm from the plane of best focus.[8] This extended the effective width of the kernel by about 1 pixel ($\Delta d \approx 1$ pixel), and was also consistent with our subjective sensation of DOF. The results of the focus search, shown by the dashed-dotted line in Fig. 11, indicate that the focus measure failed to detect the layers, as it yielded a single (merged) peak, somewhere between the focused states of the individual layers. This demonstrates the confusion of conventional autofocusing devices when applied to transparent scenes.

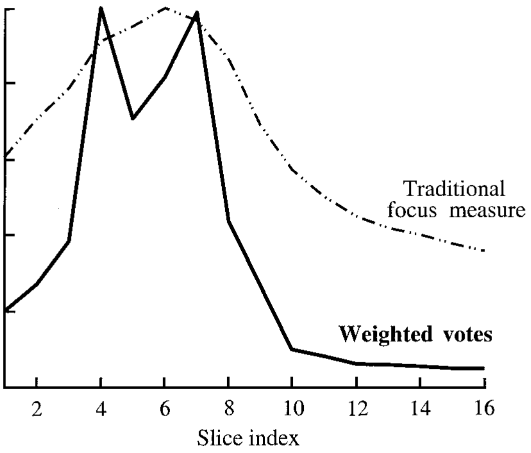

Figure 11. Experimental results. [Dashed-dotted line]: The conventional focus measure as a function of the slice index. It mistakenly detects a single focused state at the 6th slice. [Solid line]: The locations histogram of detected local maxima of the focus measure (the same scene). The highest numbers of votes (positions of local maxima) are correctly accumulated at the 4th and 7th slices—the true focused slices.

5.2. A Voting Scheme

Towards solving the merging problem, observe that the layers are generally unrelated and that edges are usually sparse. Thus, the positions of brightness edges in the two layers will only sporadically coincide. Since edges (and other feature-dense regions) are dominant contributors to the focus criterion, it would be wise not to mix them by brute averaging of the local focus measurements over the entire region. If point (x, y) is on an edge in one layer, but on a smooth region in the other layer, then the peak in *FOCUS* $(x, y, \tilde{v})$ corresponding to the edge will not be greatly affected by the contribution of the other layer.

The following approach is proposed. For each pixel (x, y) in the slices, the focus measure *FOCUS* $(x, y, \tilde{v})$ is analyzed as a function of $\tilde{v}$, to find its local maxima. The result is expressed as a binary vector of local maxima positions. Then, a vote table analogous to a histogram of maxima locations over all pixels is formed by summing all the "hits" in each slice-index. Each vote is given a weight that depends on the corresponding value of *FOCUS* $(x, y, \tilde{v})$, to enhance the contribution of high focus-measure values, such as those arising from edges, while reducing the random contribution of featureless areas. The results of the voting method are shown as a solid line in Fig. 11, and demonstrate its success in creating significant, separate peaks corresponding to the focused layers. Additional details can be found in Schechner et al. (1998). The estimated depths were correct, within the uncertainty imposed by the depth of field of the system. Optimal design and rigorous performance evaluation of DFF methods in the presence of transparencies remains an open research problem.

6. Conclusions

This paper presents an approach based on focusing to separate transparent layers, as appear in semi-reflected scenes. This approach is more stable with respect to perturbations (Schechner and Kiryati, 1998) and occlusions than separation methods that rely on stereo or motion. We also presented a method for self calibration of the defocus blur kernels given the raw images. It is based on minimizing the mutual information of the recovered layers. Note that defocus blur, motion blur, and stereo disparity have similar origins (Schechner and Kiryati, 1998) and differ mainly in the scale and shape of the kernels. Therefore, the method described here could possibly be adapted to finding the motion PSFs or stereo disparities in transparent scenes.

In some cases the methods presented here are also applicable to multiplicative layers (Shizawa and Mase, 1990): If the opacity variations within the close layer are small (a “weak” object), the transparency effect may be approximated as a linear superposition of the layers, as done in microscopy (Agard and Sedat, 1983; Conchello and Hansen, 1990; Marcias-Garza et al. 1988; Preza et al., 1992). In microscopy and in tomography, the suggested method for self calibration of the PSF can improve the removal of crosstalk between adjacent slices.

In the analysis and experiments, depth variations within each layer have been neglected. This approximation holds as long as these depth variations are small with respect to the depth of field. Extending our analysis and recovery methods to deal with space-varying depth and blur is an interesting topic for future research. A simplified interim approach could be based on application of the filtering to small domains in which the depth variations are sufficiently small. Note that the mutual information recovery criterion can still be applied globally, leading to a higher-dimensional optimization problem. We believe that fundamental properties, such as the inability to recover the DC of each layer, will hold in the general case. Other obvious improvements in the performance of the approach can be achieved by incorporating efficient search algorithms to solve the optimization problem (Luenberger, 1989), with efficient ways to estimate the mutual information (Thevenaz and Unser, 1998; Viola and Wells, 1997).

Semi-reflections can also be separated using polarization cues (Farid and Adelson, 1999; Schechner et al., 1999a, 1999b, 1999c, 2000b). It is interesting to note that polarization based recovery is typically sensitive to high frequency noise at low angles of incidence (Schechner et al., 2000b) On the other hand, DC recovery is generally possible and there are no particular difficulties at the low frequencies. This nicely complements the characteristics of focus-based layer separation, where the recovery of the high frequencies is stable but problems arise in the low frequencies. Fusion of focus and polarization cues for separating semi-reflections is thus a promising research direction.

The ability to separate transparent layers can be utilized to generate special effects. For example, in Aizawa et al. (2000) images were rendered with each of the occluding (opaque) layers defocused, moved and enhanced arbitrarily. The same effects, and possibly other interesting ones can now be generated in scenes containing semireflections.

Acknowledgments

This work was conducted while Yoav Y. Schechner was at the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa. We thank Joseph Shamir and Alex Bekker for their advice, support, and significant help, and for making the facilities of the Electrooptics Laboratory of the Electrical Engineering Department, Technion, available to us. We thank Bonnie Lorimer for the permission to use her photograph of Mt. Shuksan. This research was supported in part by the Eshkol Fellowship of the Israeli Ministry of Science, by the Ollendorff Center of the Department of Electrical Engineering, Technion, and by the Tel-Aviv University Research Fund. The vision group at the Weizmann Institute is supported in part by the Israeli Ministry of Science, Grant No. 8504. Ronen Basri is an incumbent of Arye Dissentshik Career Development Chair at the Weizmann Institute.

Notes

1. The depth dependence of the scale change can typically be neglected.
2. The superposition is linear, since the real/virtual layers are the images of the objects multiplied by the transmission/reflection coefficients of the semi-reflecting surface, and these coefficients

do not depend on the light intensities. The physical processes in transparent/semi-reflected scenes are described in Schechner et al. (1999a, 1999b, 1999c, 2000b). Nonlinear transmission and reflection effects (as appear in photorefractive crystals) are negligible at intensities and materials typical to imaging applications.

3. Courtesy of Bonnie Lorimer.
4. The STD was sampled on a grid in our demonstrations. A practical implementation will preferably use efficient search algorithms (Luenberger, 1989) to optimize the mutual information (Thevenaz and Unser, 1998; Viola and Wells, 1997).
5. The system was not telecentric, so there was slight magnification with change of focus settings. This was compensated for manually by resizing one of the images.
6. It is interesting to note that a mathematical proof exists (Hausler and Korner, 1984) for the validity of a focus criterion that is completely based on local calculations which do not depend on transversal neighbors: As a function of axial position, the intensity at each transversal point has an extremum at the plane of best focus.
7. There are secondary maxima, though, due to the unmonotonicity of the frequency response of the blur operator, and due to edge bleeding. However, the misleading maxima are usually much smaller than the maximum associated with the focusing on feature-dense regions, as edges.
8. Near the plane of best focus, the measured rate of increase of the STD as a function of defocus was much lower than expected from geometric considerations. We believe that this is due to noticeable diffraction and spherical aberration effects in that regime.

References

Abbott, A.L. and Ahuja, N. 1993. Active stereo: Integrating disparity, vergence, focus, aperture and calibration for surface estimation. *IEEE Trans. Pattern. Anal. Mach. Intell.*, 15(10):1007–1029.

Agard, D.A. 1984. Optical sectioning microscopy: Cellular architecture in three dimensions. *Ann. Rev. Biophys. Bioeng.*, 13:191–219.

Agard, D.A. and Sedat, J.W. 1983. Three-dimensional structure of a polytene nucleus. *Nature*, 302(5910):676–681.

Aghdasi, F. and Ward, R.K. 1996. Reduction of boundary artifacts in image restoration. *IEEE Trans. Image Processing*, 5(4):611–618.

Aizawa, K., Kodama, K., and Kubota, A. 2000. Producing object-based special effects by fusing multiple differently focused images. *IEEE Trans. on Circuits and Systems for Video Technology*, 10(2):323–330.

Bergen, J.R., Burt, P.J., Hingorani, R., and Peleg, S. 1992. A three-frame algorithm for estimating two-component image motion. *IEEE Trans. Pattern. Anal. Mach. Intell.*, 14(9):886–895.

Borga, M. and Knutsson, H. 1999. Estimating multiple depths in semi-transparent stereo images. In *Proc. Scandinavian Conf. on Image Analysis,* Kangerlussuaq, Greenland. Vol. I, pp. 127–133. Published by the Pattern Recognition Society of Denmark, Lyngby, Denmark.

Castleman, K.R. 1979. *Digital image processing*. Prentice-Hall, New Jersey, pp. 357–360.

Chiu, M.Y., Barrett, H.H., Simpson, R.G., Chou, C., Arendt, J.W., and Gindi, G.R. 1979. Three dimensional radiographic imaging with a restricted view angle. *J. Opt. Soc. Am. A*, 69(10):1323–1333.

Conchello, J.A. and Hansen E.W. 1990. Enhanced 3-D reconstruction from confocal scanning microscope images. I: Deterministic and maximum likelihood reconstructions. *App. Opt.*, 29(26):3795–3804.

Cover, T.M. and Thomas, J.A. 1991. *Elements of information theory*. John Wiley & Sons, New York, pp. 12–21.

Darrell, T. and Simoncelli, E. 1993a. Separation of transparent motion into layers using velocity-tuned mechanisms. M.I.T Media Lab., Massachusetts Institute of Technology, Cambridge, MA. Media-Lab TR-244.

Darrell, T. and Simoncelli, E. 1993b. 'Nulling' filters and the separation of transparent motions. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, New York, pp. 738–739.

Darrell, T. and Wohn K. 1988. Pyramid based depth from focus. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Ann Arbor, MI, pp. 504–509.

Diaspro, A., Sartore M., and Nicolini, C. 1990. 3D representation of biostructures imaged with an optical microscope. *Imag. and Vis. Comp.*, 8(2):130–141.

Engelhardt, K. and Hausler, G. 1988. Acquisition of 3-D data by focus sensing. *App. Opt.*, 27(22):4684–4689.

Erhardt, A., Zinger, G., Komitowski, D., and Bille, J. 1985. Reconstructing 3-D light-microscopic images by digital image processing. *App. Opt.*, 24(2):194–200.

Farid, H. and Adelson, E.H. 1999. Separating reflections from images by use of independent components analysis. *J. Opt. Soc. Am. A*, 16(9):2136–2145.

Fay, F.S., Fujiwara, K., Rees, D.D., and Fogarty, K.E. 1983. Distribution of actinin in single isolated smooth muscle cells, *J. of Cell Biology*, 96:783–795.

Fujikake, H., Takizawa, K., Aida, T., Kikuchi, H., Fujii, T., and Kawakita, M. 1998. Electrically-controllable liquid crystal polarizing filter for eliminating reflected light. *Optical Review*, 5(2):93–98.

Hausler, G. and Korner, E. 1984. Simple focusing criterion. *App. Opt*, 23(15):2468–2469.

Irani, M., Rousso, B., and Peleg, S. 1994. Computing occluding and transparent motions. *Int. J. Comp. Vis.*, 12(1):5–16.

Itoh, K., Hayashi, A., and Ichioka, Y. 1989. Digitized optical microscopy with extended depth of field. *App. Opt.*, 28(16):3487–3493.

Jansson, P.A., Hunt, R.H., and Plyler, E.K. 1970. Resolution enhancement of spectra. *J. Opt. Soc. Am.*, 60(5):596–599.

Jarvis, R.A. 1983. A perspective on range-finding techniques for computer vision. *IEEE Trans. Pattern Anal. Mach. Intell.*, 5(2):122–139.

Krishnan, A. and Ahuja, N. 1996. Panoramic image acquisition. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, pp. 379–384.

Kubota, A., Kodama, K., and Aizawa, K. 1999. Registration and blur estimation methods for multiple differently focused images. In *Proc. International Conference on Image Processing*, Kobe, Japan, Vol. 2, pp. 447–451.

Luenberger, D.G. 1989. *Linear and Nonlinear Programming,* 2nd ed., Adisson-Wesley, London.

Marcias-Garza, F., Bovik, A.C., Diller, K.R., Aggarwal, S.J., and Aggarwal, J.K. 1988. The missing cone problem and low-pass distortion in optical serial sectioning microscopy. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, New York, Vol-II, pp. 890–893.

McNally, J.G., Preza, C., Conchello, J.A., and Thomas, L.J., Jr. 1994. Artifacts in computational optical-sectioning microscopy. *J. Opt. Soc. Am. A*, 11(3):1056–1067.

Nair, H.N. and Stewart, C.V. 1992. Robust focus ranging. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Champaign, IL, pp. 309–314.

Nayar, S.K. 1992. Shape from focus system. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Champaign, IL, pp. 302–308.

Nayar, S.K., Watanabe, M., and Nogouchi, M. 1995. Real time focus range sensor. In *Proc. IEEE International Conf. on Computer Vision*, Cambridge, MA, pp. 995–1001.

Noguchi, M. and Nayar, S.K. 1994. Microscopic shape from focus using active illumination. In *Proc. IAPR International Conference on Image Processing*, Jerusalem, Israel, Vol-1, pp. 147–152.

Ohnishi, N., Kumaki, K., Yamamura, T., and Tanaka, T. 1996. Separating real and virtual objects from their overlapping images. In *Proc. European Conf. on Computer Vision* Vol. II, Cambridge, UK. Springer, New York, pp. 636–646. Lecture notes in Computer Science 1065.

Oren M. and Nayar, S.K. 1995. A theory of specular surface geometry. In *Proc. IEEE International Conf. on Computer Vision*, Cambridge, MA, pp. 740–747.

Preza, C., Miller, M.I., Thomas, L.J., Jr., and McNally, J.G. 1992. Regularized linear method for reconstruction of three-dimensional microscopic objects from optical sections. *J. Opt. Soc. Am. A*, 9(2):219–228.

Schechner, Y.Y. and Kiryati, N. 1998. Depth from defocus vs. Stereo: How different really are they? In *Proc. IEEE Computer Society International Conference on Pattern Recognition*, Brisbane, Australia, Vol. 2, pp. 1784–1786. To be published in the Int. J. Computer Vision.

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

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

Schechner, Y.Y., Kiryati, N., and Shamir, J. 1999a. Separation of transparent layers by polarization analysis. In *Proc. Scandinavian Conf. on Image Analysis*, Kangerlussuaq, Greenland, Vol. I, pp. 235–242. Published by the Pattern Recognition Society of Denmark, Lyngby, Denmark.

Schechner, Y.Y., Shamir, J., and Kiryati, N. 1999b. Vision through semireflecting media: Polarization analysis. *Optics Letters*, 24(16):1088–1090.

Schechner, Y.Y., Shamir, J., and Kiryati, N., 1999c. Polarization-based decorrelation of transparent layers: The inclination angle of of an invisible surface. In *Proc. IEEE International Conf. on Computer Vision*, Kerkyra, Greece, Vol. II, pp. 814–819.

Schechner, Y.Y., Kiryati, N., and Shamir, J. 2000a. Blind recovery of transparent and semireflected scenes. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Hilton Head Island, SC, Vol. 1, pp. 38–43.

Schechner, Y.Y., Shamir, J., and Kiryati, N. 2000b. Polarization and statistical analysis of scenes containing a semireflector. *J. Opt. Soc. Am. A*, 17(2):276–284.

Sheppard, C.J.R. and Gu, M. 1991. Three dimensional optical transfer function for an annular lens. *Optics Communications*, 81(5):276–280.

Shizawa, M. 1992. On visual ambiguities due to transparency in motion and stereo. In *Proc. European Conf. on Computer Vision*, Santa Margherita Ligure, Italy, Springer-Verlag, New York, pp. 411–419. Lecture notes in Computer Science 588.

Shizawa, M. 1993. Direct estimation of multiple disparities for transparent multiple surfaces in binocular stereo. In *Proc. IEEE International Conf. on Computer Vision*, Berlin, pp. 447–454.

Shizawa, M. and Mase, K. 1990. Simultaneous multiple optical flow estimation. In *Proc. International Conference on Pattern Recognition*, Atlantic City, NJ, Vol. 1, pp. 274–278.

Streibl, N. 1984. Fundamental restrictions for 3-D light distributions. *Optik*, 66(4):341–354.

Streibl, N. 1985. Three-dimensional imaging by a microscope. *J. Opt. Soc. Am. A*, 2(2):121–127.

Subbarao, M. and Jenn-Kwei Tyan. 1995. The optimal focus measure for passive autofocusing and depth from focus. In *Proc. SPIE 2598—Videometrics VI*, Philadelphia, PA, pp. 89–99.

Sugimoto, S.A. and Ichioka, Y. 1985. Digital composition of images with increased depth of focus considering depth information. *App. Opt.*, 24(14):2076–2080.

Sundaram, H. and Nayar, S. 1997. Are textureless scenes recoverable? In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, pp. 814–820.

Torroba, P., Cap, N., and Rabal, H. 1994. Defocus detection using a visibility criterion. *Journal of Modern Optics*, 41(1):111–117.

Thevenaz, P. and Unser, M. 1998. An efficient mutual information optimizer for multiresolution image registration. In *Proc. IEEE Computer Society International Conference on Image Processing*, Chicago, IL, Vol. I, pp. 833–837.

Viola, P. and Wells, W.M. III, 1997. Alignment by maximization of mutual information. *Int. J. of Computer Vision*, 24(2):137–154.

Wang, J.Y.A. and Adelson, E.H. 1993. Layered representation for motion analysis. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, New York, pp. 361–365.

Watanabe, M. and Nayar, S.K. 1996. Telecentric optics for computational vision. In *Proc. European Conf. on Computer Vision*, Cambridge, UK. Springer, New York, Vol. II, pp. 439–451. Lecture notes in Computer Science 1065.

Weinshall, D. 1989. Perception of multiple transparent planes in stereo vision. *Nature*, 341(6244):737–739.

Xiong, Y. and Shafer, S.A. 1993. Depth from focusing and defocusing. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, New York, pp. 68–73.

Yeo, T.T.E., Ong, S.H., Jayasooriah, and Sinniah, R. 1993. Autofocusing for tissue microscopy. *Image and Vision Computing*, 11(10):629–639.