Multi-valued Images and Their Separation⋆

Yoav Y. Schechner[1], Nahum Kiryati[2], and Joseph Shamir[3]

[1] Columbia University, New York NY 10027, USA,
yoav@cs.columbia.edu
[2] Tel-Aviv University, Ramat Aviv 69978, Israel,
nk@eng.tau.ac.il
[3] Technion - Israel Institute of Technology, Haifa 32000, Israel,
jsh@ee.technion.ac.il

Abstract. Consider scenes deteriorated by reflections off a semi-reflecting medium (e.g., a glass window) that lies between the observer and an object. We present two approaches to recover the superimposed scenes. The first one is based on a focus cue, and can be generalized to volumetric imaging with multiple layers. The second method, based on a polarization cue, can automatically label the reconstructed scenes as reflected/transmitted. It is also demonstrated how to blindly determine the imaging PSF or the orientation of the invisible (semi-reflecting) surface in space in such situations.

1 Introduction

This work deals with the situation in which the projection of the scene on the image plane is multi-valued due to the superposition of several contributions. This situation is encountered while looking through a window, where we see both the outside world (termed *real object* [12,13]), and a semi-reflection of the objects inside, termed *virtual objects*. It is also encountered in microscopy and tomography, where viewing a transparent object slice is disturbed by the superposition of adjacent defocused slices. Our goal is to clear the the disturbing crosstalk of the superimposing contributions, and gain information on the scene structure.

Previous treatment of this situation has been based mainly on motion [3,7,9, 13,23,26], and stereo [4,22]. Polarization cues have been used for such scenarios in Refs. [8,12]. Thorough polarization analysis, that enabled the labeling of the layers was done in [17,18,19,21]. This paper refers to these methods, and deals also with the use of depth of field (DOF) to analyze images. DOF has been utilized for analysis of transparent layers mainly in microscopy [1,5,6,11], but mainly in cases of opaque (and occluding) layers, as in [2,10].

Following [14] we show that the methods that rely on motion and stereo are closely related to approaches based on DOF. Then, we show how to recover the transparent layers using a focus cue. In the case of semireflections, we follow Refs. [15,16,20] to show that two raw images are adequate to recover the layers. The recovery is done in conjunction to a simple way to estimate the transfer function between the images, based on the raw images, yielding the optimal layer separation. We then show how to label each layer as reflected or transmitted using a polarization cue, which also indicates the orientation of the invisible semi-reflecting window in space. Following [17,18,19,21], our use of the polarization cue is effective also away from the Brewster angle.

2   Distance Cues

2.1   Defocus vs. Stereo or Motion

Depth cues are usually very important for the separation of transparent layers. In microscopy, each superimposing layer is at a different distance from the objective. Hence when one layer is focused (at a certain *image slice*) the others are defocus blurred, and this serves as the handle for removal of the inter-layer crosstalk. In case of semi-reflected scenes, the real object is unrelated to the virtual object. So, it is reasonable that also in this case the distance of each layer from the imaging system is different. Therefore, if the scene is viewed by a stereo system, each layer will have a different disparity; if the camera is moved, each layer will have different motion parameters; and again, if one layer is focused, the other is defocused.

Note that depth from defocus blur or focus are *realizations of triangulation* just as depth from stereo or motion are realizations of this principle [14]. Consider Fig. 1, where a stereo (or depth from motion) system is viewing the same scene as a monocular wide aperture camera of the *same physical dimensions*: the stereo baseline D is the the same as the lens aperture, the distance from the lens to the sensor, v, is the same, and the stereo system is fixated on a point at the same distance at which the wide-aperture system is focused. Then, the disparity is equal to the defocus blur diameter, under the geometric optics approximation [14].

Therefore, algorithms developed based on defocus blur can be applied to approaches based on stereo or motion, and vice-versa. Besides, these approaches will have similar fundamental limitations. Particularly for the scenarios treated in this work, seeking a separation of two transparent layers out of two raw images, in each of which the focus is different, is equivalent to separating the layers using two raw images in which the disparities are different. Blind estimation of the defocus blur kernels (or the transfer functions of the imaging system) is equivalent to seeking the motion parameters between the two stereo raw images.

In this section we treat the transparent-layers problem using methods that rely on defocus blur. However, due to the close relationship of defocus to stereo and motion, the reader may generally interchange the "defocus parameters" with "motion parameters" and extend the conclusions to classical triangulation approaches.
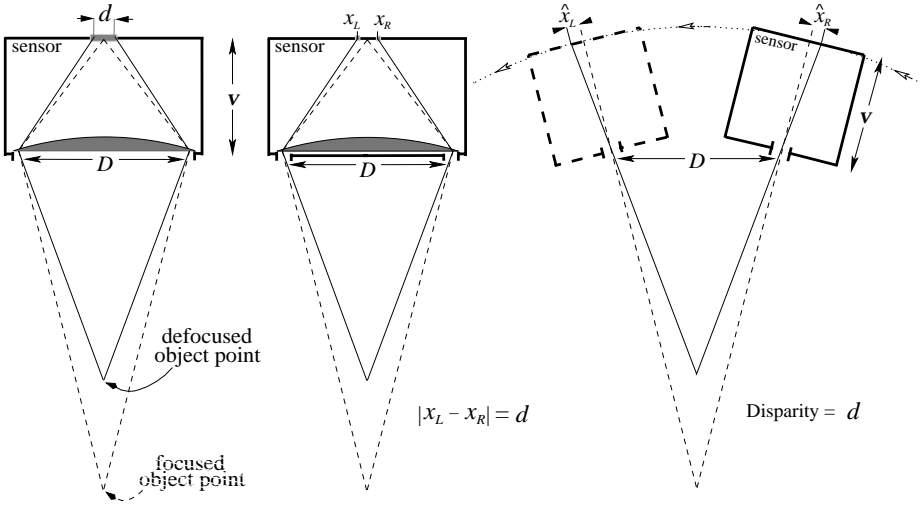
Fig. 1. [Left] The image of a defocused object point at a certain distance is a blur circle of diameter d. [Middle] Its image becomes two points separated by d, if the lens is blocked except for two pinholes at opposite sides on the lens perimeter. [Right] The disparity equals the same d in a stereo/motion system having the same physical dimensions and fixating on the point at the same distance as the point focused by the system on the left.

2.2   Recovery from Known Transfer Functions

Consider a two-layered scene. 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. Let layer f_1 be superimposed 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,

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

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

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

The naive inverse filtering solution of the problem is

$$\widehat{F}_1 = B(G_a - G_b H_{2a}) \qquad \widehat{F}_2 = B(G_b - G_a H_{1b}) \ , \tag{3}$$

where

$$B = (1 - H_{1b}H_{2a})^{-1} \ . \tag{4}$$

As the frequency decreases, $H_{2a}H_{1b} \to 1$, and then $B \to \infty$, hence the solution is unstable. Moreover, due to energy conservation, the average gray level (DC)

is not affected by defocusing ($H_{2a}H_{1b} = 1$), hence its recovery is ill posed. However, the problem is well posed and stable at the high frequencies. This behavior also exists in separation methods that rely on motion [15,16], as expected from the discussion in section 2.1. Note that this is quite opposite to typical reconstruction problems, in which instability and noise amplification appear in the high frequencies.

If $H_{2a}H_{1b} \neq 1$ (that is, except at the DC), B can be approximated by the series

$$\widehat{B}(m) = \sum_{k=1}^{m} (H_{1b}H_{2a})^{k-1} \quad . \tag{5}$$

The approximate solutions $\widehat{F}_1(m), \widehat{F}_2(m)$ are thus parameterized by m which controls how close the filter $\widehat{B}(m)$ is to the inverse filter, and is analogous to regularization parameters in typical inversion methods. We define the *basic solution* as the result of using $m = 1$.

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. The "pinhole" image is described by

$$g_0 = (f_1 + f_2)/a \quad , \tag{6}$$

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

$$\widehat{F}_1 = S(G_a - aG_0H_{2a}) \qquad \widehat{F}_2 = S(aG_0 - G_a) \quad , \tag{7}$$

where

$$S = (1 - H_{2a})^{-1} \quad . \tag{8}$$

Also in this method the filter S can be approximated by

$$\widehat{S}(m) = \sum_{k=1}^{m} H_{2a}^{k-1}. \tag{9}$$

2.3   Blind Estimation of the Transfer Functions

The imaging PSFs (and thus their corresponding transfer functions) may be different than the ones we estimate and use in the reconstruction algorithm. As shown in [16,20] 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 PSFs are Gaussians. Thus, we may expect that the best sense of

separation will be achieved in the basic solution, even though the low frequencies are less attenuated and better balanced with the high frequencies at higher m's.

We wish to achieve self-calibration, i.e., to estimate the kernels out of the images themselves. This enables blind separation and restoration of the layers. Thus, we need a criterion for layer separation. It is reasonable to assume that the statistical dependence of the real and virtual layers is small since they usually originate from unrelated scenes. Mutual information measures how far the images are from statistical independence [16]. We thus 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 [24,25], where its maximum was sought. We use this measure to look for the highest discrepancy between images, thus minimizing it. To decrease the dependence of the estimated mutual information on the dynamic range and brightness of the individual layers, it was normalized by the mean entropy of the estimated layers, when treated as individual images. This measure, denoted $\mathcal{I}_n$, indicates the ratio of mutual information to the self information of a layer. Additional details are given in [16,20].

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

$$[\tilde{h}_{1b}, \tilde{h}_{2a}] = \arg \min_{h_{1b},h_{2a}} \mathcal{I}_n(\hat{f}_1, \hat{f}_2) \ . \tag{10}$$

As noted above, errors in the kernels lead to crosstalk (contamination) of the estimated layers, which is expected to increase their mutual information. 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). The blurring along the sensor raster rows may be different than the blurring along the columns. So, generally we assigned a different blur "radius" to each axis. If two slices are used, there are two kernels, and the optimization is done over a total of four parameters. When a single focused slice is used in conjunction with a "pinhole" image, the problem is much simpler. We need to determine the parameters of only one kernel, and the factor a. a can be indicated from the ratio of the *f-numbers* of the camera in the two states, or from the ratio of the average values of the images.

2.4   Recovery Experiments

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 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 basic solution ($m = 1$) based on the optimal parameters is shown at the middle row of Fig. 2. It has $\mathcal{I}_n \approx 0.006$ (two orders of magnitude better than the raw images). Using $m = 5$ yields better a balance between the low and high frequency components, but $\mathcal{I}_n$ increased to about 0.02. As noted above, the theory [16,20] indicates that an error in the PSF

Close layer   Far layer

Focused
slices
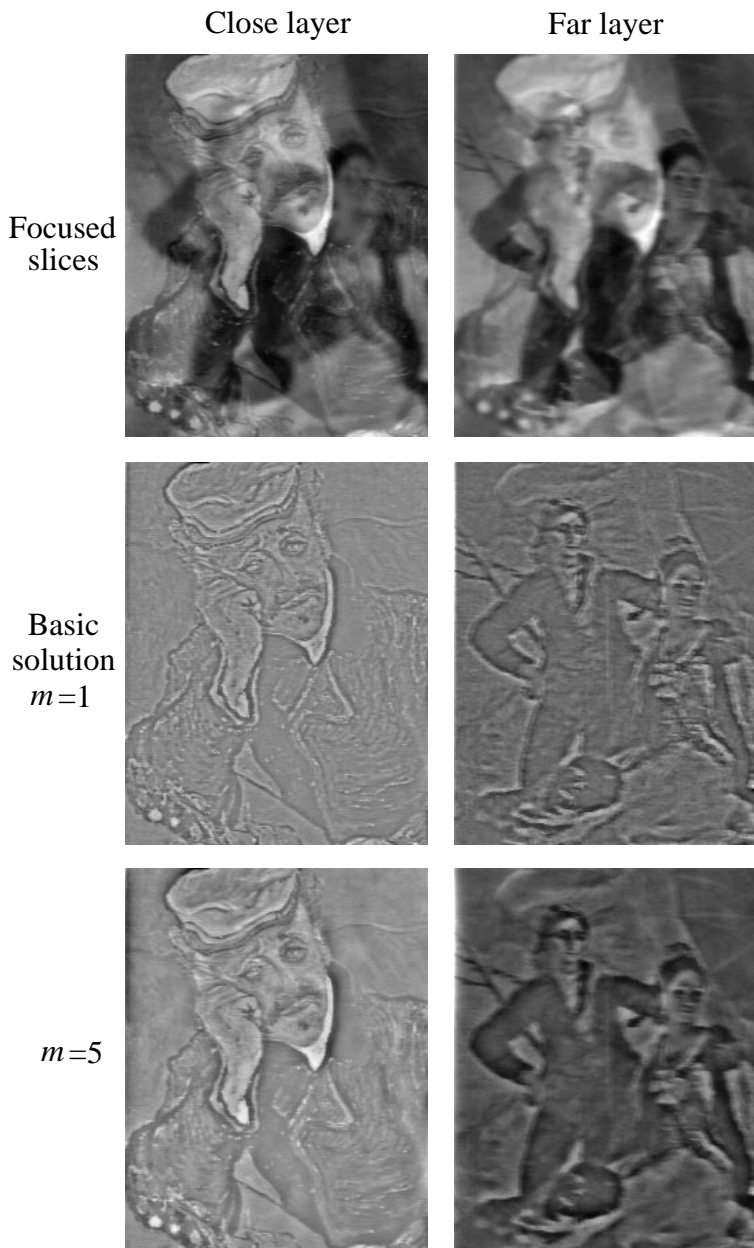

Basic
solution
$m=1$

$m=5$

Fig. 2. [Top]: The slices in which either of the transparent layers is focused. [Middle row]: The basic solution ($m = 1$). [Bottom row]: Recovery with $m = 5$.
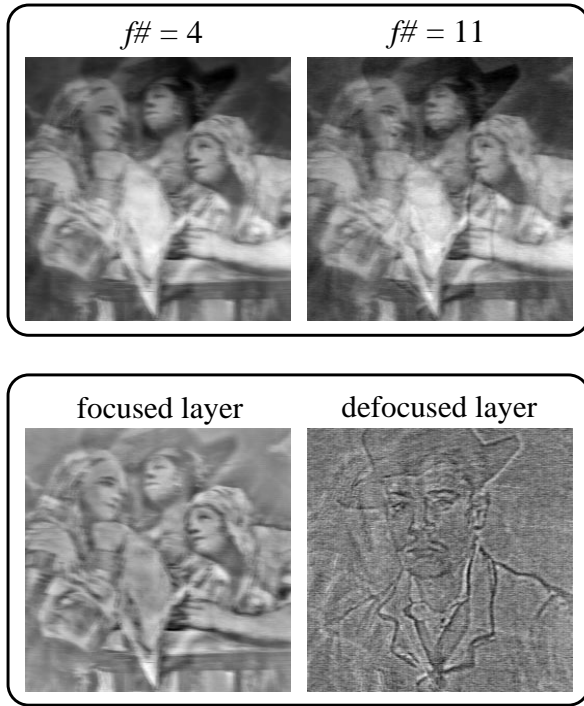
Fig. 3. [Top]: Raw images: the far layer is focused when viewed with a wide aperture, and with a "pinhole" setup. [Bottom]: The basic recovery.

model, yields a stronger crosstalk for larger m. Hence this increase in $\mathcal{I}_n$ may originate from the inaccuracy of our assumption of a Gaussian model.

In another example, 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. Here we used a fixed focus setting, and changed the aperture between image acquisitions. The slice in which the far layer is focused (using the wide aperture) is at the top-left of Fig. 3, and the "pinhole" image is to its right. The basic solution based on the optimal parameters is shown on the bottom of Fig. 3. The "Portrait" is revealed.

3   Labeling by a Polarization Cue

3.1   Recovery from a Known Inclination Angle

Distance cues do not indicate which of the layers is the reflected (virtual) one, and which is the transmitted (real) one. However, polarization cues give us a simple way to achieve both the layer separation and their labeling as real/virtual. At the semi-reflecting medium (e.g, a glass window) the reflection coefficients are

different for each of the light polarization components, and are denoted by $\tilde{R}_\perp$ and $\tilde{R}_\parallel$ for the polarization components perpendicular and parallel to the plane of incidence, respectively. They can be analytically derived as functions of the surface inclination (angle of incidence, φ) from the Fresnel equations [17,21], taking into account the effect of internal reflections [17,19] in the medium. The transmission coefficient of each component is given by

$$\tilde{T} = 1 - \tilde{R} \ . \tag{11}$$

We denote the image due to the real layer (with no window) by I_T and the image due to the virtual layer (assuming a perfect mirror replacing the window) by I_R. The light coming from the real object is superimposed with the light coming from the virtual object. Let the scene be imaged through a linear polarizer, by which we can select to sense the perpendicular or the parallel components ($g_\perp$ and $g_\parallel$, respectively) of the observed light coming from the scene. Thus, the two raw images are:

$$g_\perp = I_R \tilde{R}_\perp/2 + I_T \tilde{T}_\perp/2 \qquad g_\parallel = I_R \tilde{R}_\parallel/2 + I_T \tilde{T}_\parallel/2 \ , \tag{12}$$

for initially unpolarized natural light. Solving these equations for the two images we obtain the estimated intensities of the layers as a function of an assumed angle of incidence, φ:

$$\hat{I}_T(\varphi) = \left[\frac{2\tilde{R}_\perp(\varphi)}{\tilde{R}_\perp(\varphi) - \tilde{R}_\parallel(\varphi)}\right] g_\parallel - \left[\frac{2\tilde{R}_\parallel(\varphi)}{\tilde{R}_\perp(\varphi) - \tilde{R}_\parallel(\varphi)}\right] g_\perp \tag{13}$$

$$\hat{I}_R(\varphi) = \left[\frac{2 - 2\tilde{R}_\parallel(\varphi)}{\tilde{R}_\perp(\varphi) - \tilde{R}_\parallel(\varphi)}\right] g_\perp - \left[\frac{2 - 2\tilde{R}_\perp(\varphi)}{\tilde{R}_\perp(\varphi) - \tilde{R}_\parallel(\varphi)}\right] g_\parallel \ . \tag{14}$$

Therefore, the layers are recovered by simple weighted subtractions of the raw images. Moreover, the equation for $\hat{I}_T$ is distinct from the equation for $\hat{I}_R$, so the separate layers are automatically *labeled* as reflected/transmitted (i.e., virtual/real). Note, however, that the weights in these subtractions are functions of the angle at which the invisible (but semireflecting) medium is inclined with respect to the viewer.

3.2   The Inclination of the Invisible Surface

In case the angle of incidence used in the reconstruction process is not the true inclination angle of the surface φ_{true}, each recovered layer will contain traces of the complimentary layer (crosstalk). In an experiment we performed [19,21], a scene composed of several objects was imaged through an upright glass window. The window semi-reflected another scene (virtual object). A linear polarizer was rotated in front of the camera between consecutive image acquisitions. The reflected layer is attenuated in $g_\parallel$ (Fig. 4) but its disturbance is still significant, since $\varphi_{\text{true}} = 27.5^o \pm 3^o$, was far from the Brewster angle 56° (at which the reflection disappears in $g_\parallel$). Having assumed that the true angle of incidence is

Fig. 4. The raw images. [Left]: $g_\perp$. [Right]: Although the reflected component is smaller in $g_\parallel$, the image is still unclear.

unknown, φ was guessed. As seen in Fig. 5, when the angle was underestimated negative traces appeared in the reconstruction (bright areas in $\hat{I}_R$ are darker than their surroundings in $\hat{I}_T$). When the angle was overestimated, the traces are positive (bright areas in $\hat{I}_R$ are brighter than their surroundings in $\hat{I}_T$). When the correct angle is used, the crosstalk is removed, and the labeled layers are well separated.

An automatic way to detect this angle is by seeking the reconstruction that minimizes the mutual information between the estimated layers, in a similar manner to the procedure of section 2.3:

$$\hat{\varphi} = \arg\min_{\varphi} \mathcal{I}_n[\hat{I}_T(\varphi), \hat{I}_R(\varphi)] \ . \tag{15}$$

In this experiment $\mathcal{I}_n$ is minimized at $\varphi = 25.5^o$. The angle at which the estimated layers are decorrelated is $\varphi = 27^o$ (Fig. 5). Both these values are within the experimental error of the physical measurement of φ_{true}. Note that the correlation sign is consistent with the "positive/negative" traces when the assumed angle is wrong.

The reconstruction of the real layer ($\hat{I}_T$) in the experiment is more sensitive to an error in the angle of incidence, than the reconstruction of the virtual layer ($\hat{I}_R$). In Fig. 5 the contamination in the estimated $\hat{I}_R$ by I_T is hardly visible, if at all. On the other hand, the contamination in the estimated $\hat{I}_T$ by I_R is more apparent. This result is consistent with a theoretical prediction published in Ref. [21]. In that work, the relative contamination of each recovered layer by its complimentary was derived. It was normalized by the "signal to noise ratio" of the layers when no polarizer is used. We outline here just the final result of the first-order derivation from Ref. [21]. Fig. 6 plots (solid line) the first order of c_T, which is the theoretical relative contamination of the transmitted layer by the virtual layer. It is much larger than the relative contamination c_R of the recovered reflected layer by the real layer (dashed line).

According to Fig. 6, at $\varphi_{\text{true}} = 27^o$, $\hat{I}_T$ will be $\approx 8\%$ contaminated by I_R per 1° error in φ. Thus for the 10° error of Fig. 5 we get $\approx 80\%$ contamination (if the first order approximation still holds), when the reflected contribution is as bright as the transmitted one. If the reflection is weaker, the contamination will
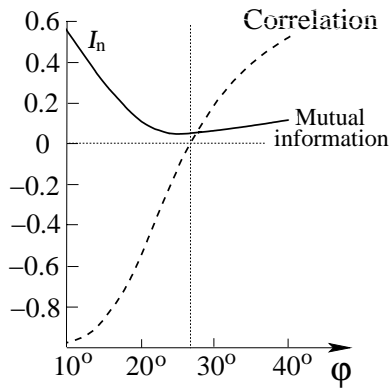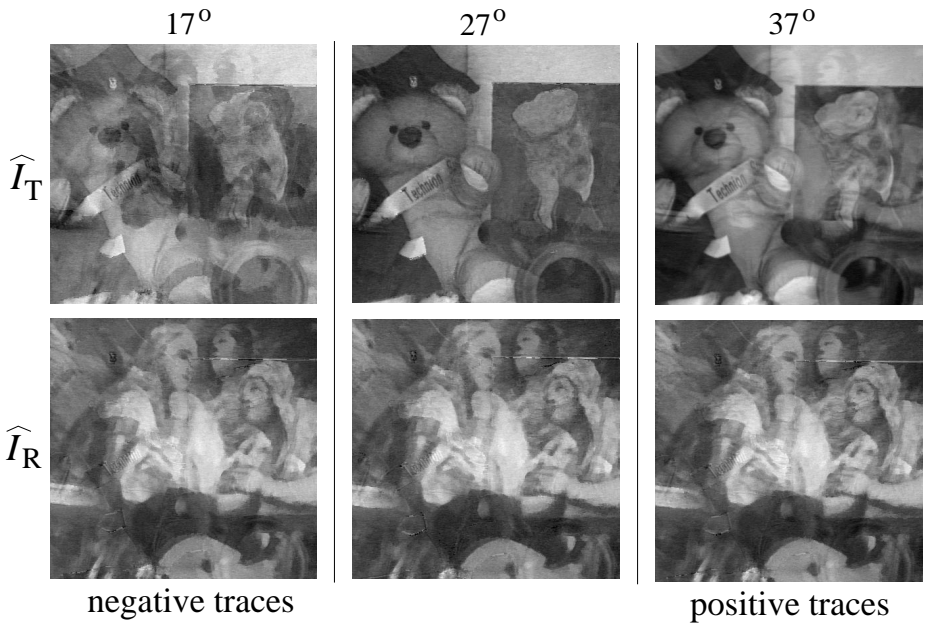
Fig. 5. Experimental results. When the assumed angle of incidence is correct ($\varphi = \varphi_{\text{true}} = 27^o$), the separation is good. In cases of under-estimation or over-estimation of the angle, negative or positive traces of the complementary layer appear, respectively. This is also seen in the increase of mutual information and in the correlation of the estimated layers. The traces are stronger in $\hat{I}_T$ than in $\hat{I}_R$, in consistency with Fig. 6.
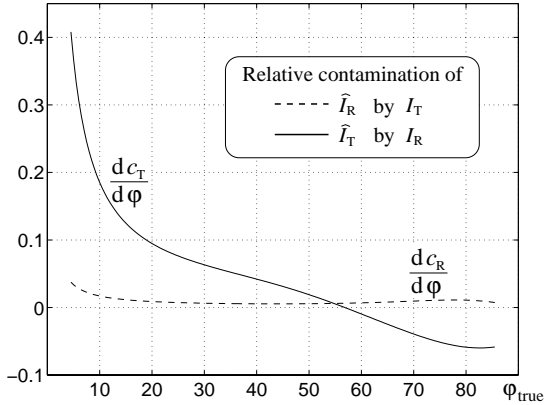
Fig. 6. The relative contamination of the each layer, per 1° of error in the angle of incidence, if the reflected contribution is as bright as the transmitted one (after the incidence on a glass window).

be proportionally smaller. On the other hand, $\hat{I}_R$ will be just $\approx 3\%$ contaminated by I_T at 10° error from φ_{true}, when the reflected contribution is as bright as the transmitted one. This is the reason why in this experiment $\hat{I}_R$ appears to be much more robust to the angle error than $\hat{I}_T$.

4   Discussion

This paper concentrated on the analysis of multi-valued images that occur when looking through a semi-reflecting surface, such as a glass window. We have shown that two raw images suffice to separate the two contributing layers. Polarization cues enable the labeling of each of the layers as real or virtual. They also enable the extraction of information on the clear semi-reflecting surface itself (its inclination in space). However, it seems to be very difficult to use this approach if the problem is scaled to more than two contributing layers. Another shortcoming of this approach is that it is not applicable if the transparent layers do not involve reflections (as occur in volumetric specimens).

The distance cue can easily be scaled for cases where there are more than two layers. Actually, it is used in volumetric specimens (which may have a continuum of layers), based on the focus cue. Our demonstration is definitely not limited to the focus/defocus cue, since defocus blur, motion blur, and stereo disparity have similar origins [14] and differ mainly in the scale and shape of the kernels. Therefore, the success of the methods that are based on defocus blur is an encouraging step towards understanding and demonstrating the estimation of the motion PSFs or stereo disparities in transparent scenes from as few as 2 images, and recovering the layers from them. However, if a small baseline suffices to separate the layers, then a method based on defocus blur may be preferable since depth from defocus is more stable with respect to perturbations and occlusions

than methods that rely on stereo or motion, for the same physical dimensions of the setup [14]. On the other hand, methods that rely on distance cues seem to have an inherent ill-conditioning at the low frequency components, and a labeling ambiguity in cases of semi-reflections.

In microscopy and in tomography, the suggested method for self calibration of the PSF can improve the removal of crosstalk between adjacent slices. However, in these cases significant correlation exists between adjacent layers, so the correlation criterion may not be adequate. This is therefore a possible direction for future research.

Since methods relying of distance cues somewhat complement methods that rely of polarization cues, fusion of these cues for separating semi-reflections is a promising research direction. The ability to separate transparent layers can be utilized to generate special effects. For example, in Ref. [2,10] images were rendered with each of the occluding (opaque) layers defocused, moved and enhanced independently. The same effects, and possibly other interesting ones can now be generated in scenes containing semireflections.

References

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

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

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

[4] Borga, M., Knutsson, H.: Estimating multiple depths in semi-transparent stereo images. Proc. Scandinavian Conf. on Image Analysis Vol. I (1999) 127–133. Kangerlussuaq, Greenland. Published by the Pattern Recognition Society of Denmark, Lyngby, Denmark.

[5] Castleman, K. R.: Digital image processing (1979) 357–360. Prentice-Hall, New Jersey.

[6] Conchello, J. A., Hansen E. W.: Enhanced 3-D reconstruction from confocal scanning microscope images. I: Deterministic and maximum likelihood reconstructions. Applied Optics bf 29 (1990) 3795–3804.

[7] Darrell, T., Simoncelli, E.: 'Nulling' filters and the separation of transparent motions. Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (1993) 738–739,

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

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

[10] Kubota, A., Aizawa, K.: Inverse filters for generation of arbitrarily focused images. Proc. SPIE **4067** Visual Communications and Image Processing (2000) 1633–1641.

[11] Marcias-Garza, F., Bovik, A. C., Diller, K. R., Aggarwal, S. J., Aggarwal, J. K.: The missing cone problem and low-pass distortion in optical serial sectioning microscopy. Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol-II (1988) 890–893.

[12] Ohnishi, N. Kumaki, K., Yamamura T., Tanaka, T.: Separating real and virtual objects from their overlapping images. Proc. European Conf. on Computer Vision Vol. II, Lecture notes in Computer Science **1065** (1996) 636-646.

[13] Oren M., Nayar, S. K.: A theory of specular surface geometry. Proc. IEEE International Conf. on Computer Vision (1995) 740–747

[14] Schechner, Y. Y., Kiryati, N.: Depth from defocus vs. Stereo: How different really are they? Int. J. Computer Vision **39** (2000) 141–162.

[15] Schechner, Y. Y., Kiryati N., Basri, R.: Separation of transparent layers using focus. Proc. IEEE International Conf. on Computer Vision 1061–1066 (1998) Mumbai, India.

[16] Schechner, Y. Y., Kiryati N., Basri, R.: Separation of transparent layers using focus. International J. on Computer Vision, **39** (2000) 25–39.

[17] Schechner, Y. Y., Kiryati N., Shamir, J.: Separation of transparent layers by polarization analysis. Proc. Scandinavian Conf. on Image Analysis, Vol. I (1999) 235–242. Published by the Pattern Recognition Society of Denmark, Lyngby, Denmark.

[18] Schechner, Y. Y., Shamir, J., Kiryati, N.: Vision through semireflecting media: Polarization analysis. Optics Letters **24** (1999) 1088-1090.

[19] Schechner, Y. Y., Shamir, J., Kiryati, N.: Polarization-based decorrelation of transparent layers: The inclination angle of of an invisible surface. Proc. IEEE International Conf. on Computer Vision, Vol. II (1999) 814–819.

[20] Schechner, Y. Y., Kiryati, N., Shamir, J.: Blind recovery of transparent and semireflected scenes. Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. I (2000) 38–43,

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

[22] Shizawa, M.: On visual ambiguities due to transparency in motion and stereo. Proc. European Conf. on Computer Vision. Lecture notes in Computer Science **588** (1992) 411–419.

[23] Shizawa, M., Mase, K.: Simultaneous multiple optical flow estimation. Proc. International Conference on Pattern Recognition, Vol. 1 (1990) 274–278.

[24] Thevenaz, P., Unser, M.: An efficient mutual information optimizer for multiresolution image registration. Proc. IEEE Computer Society International Conference on Image Processing, Vol. I (1998) 833–837.

[25] Viola, P., Wells, W. M. III: Alignment by maximization of mutual information. Int. J. of Computer Vision **24** (1997) 137-154.

[26] Wang, J. Y. A., Adelson, E. H.: Layered representation for motion analysis. Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (1993) 361–365.