Overcoming Visual Reverberations

Yaron Diamant and Yoav Y. Schechner
Dept. of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, Israel
yarondi@rafael.co.il , yoav@ee.technion.ac.il

Abstract

An image acquired through a glass window is a superposition of two sources: a scene behind the window, and a reflection of a scene in front of the window. Light rays incident on the window are reflected back and forth inside the glass. Such internal reflections affect the radiance of both sources: a spatial effect is created of dimmed and shifted replications. Our work generalizes the treatment of transparent scenes to deal with this effect. First, we present a physical model of the image formation. It turns out that each of the transmitted and reflected scenes undergoes a convolution with a particular point spread function (PSF), composed of distinct delta functions. Therefore, scene recovery involves inversion of these PSFs. We analyze the fundamental limitations faced by any attempt to solve this inverse problem. We then present a solution approach. The approach is based on deconvolution by linear filtering and simple optimization. The input to the algorithm is a pair of frames, taken through a polarizing filter. The method is demonstrated experimentally.

1. Introduction

Transparent scenes pose a challenge to computer vision. They exist in setups having a semireflecting window, which superimposes the scene behind the window to a reflected scene. This creates confusing images. A range of methods were developed to attack this problem, based on motion [4, 9, 18, 25, 27] stereo [22], polarization [11, 12, 20, 21], focus [19, 23], illumination modulation [1] and image priors [14]. They successfully demonstrated separation of the scenes (layers). However, the prior studies ignored a spatial effect of internal reflections in such scenes, which we describe next. Therefore, the prior methods are valid only in the limit where this effect is negligible.

Fig. 1 demonstrates this effect in a real photograph taken via a window. In addition to the superimposed scenes (toys of a star vs. a tree in the sun), a shifted and weaker replica of the sun and tree are clearly seen. This is caused by internal reflections that take place inside a window. In addition to that clear replica, there is also a replica of the other scene (star). Additional higher order replicas exist for both objects, but are often too dim to see. Overall, the acquired photograph contains a superposition not only of the two original scenes, but also of those same scenes displaced to various distances and in different powers. The prior studies on transparent scenes did not account for this effect. There, the model and algorithms focused on the limit case, in which the displacement between the replicas was negligible. This is not a valid situation in general.

Figure 1. A real-world frame $S_{\perp}$ acquired through a transparent window. In addition to the superposition of two scenes, notice the secondary reflections (replications), e.g., of the sun and tree. **For clarity, please view the color images on the computer monitor.**

In this work we explicitly model this effect and deal with it, hence generalizing the study of transparent scenes. Optical reflections create visual spatial displacements. This is analogous to temporal displacement created by reflections of temporal signals of sound and radio-frequency. In analogy to the displaced replica in our study, a sound reflection creates a delayed echo. In the field of acoustics [3], this effect is generally referred to as *reverberations*. Hence, we use the term *visual reverberations* to describe the effect we deal with. A similar echo effect in radio frequency is affecting received television signals, creating shifted replicas. There, cancellation of the effect is termed *deghosting* [8].

We model the effect by physics-based expressions that account for properties of optical reflections, including polarization. Then, the paper translates the model to the language of signal-processing. It formulates the effect as

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

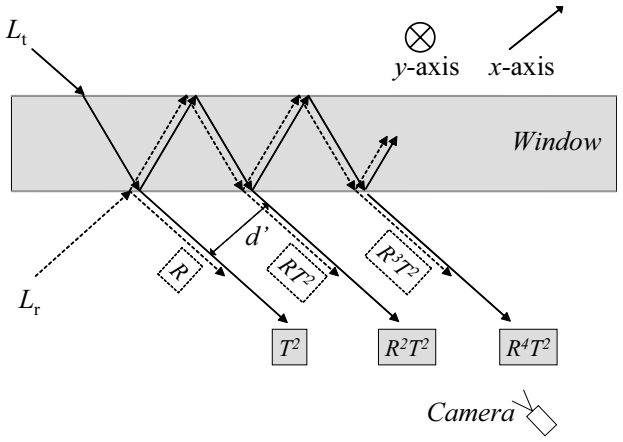

Figure 2. Primary and secondary reflections for $L_{\rm t}$ [solid] and $L_{\rm r}$ [dotted]. The distance between the emerging rays is d'.

convolution with point spread functions (PSFs), which are given in closed form. Transparent layers are thus a mixture of convolved scenes. The model used by prior studies is a special (limit) case. The fundamental limitations of this recovery problem are analyzed. Then, we present a physics-based method for inverting the image formation model. It recovers the separate scenes, while overcoming the spatial effect of visual reverberations. The parameters of the problem are derived from the raw images. The method is based on frames taken with a polarizing filter.[1] It is demonstrated in various examples, including real experiments.

2. Image Formation Model

A camera observes a scene via a semireflecting window. The object behind the window is *transmitted* through the window, thus variables associated with it are denoted by '`t`'. In addition, there is an object on the camera-side of the window. It is reflected, thus variables associated with it are denoted by '`r`'. Specifically, $L_{\rm t}$ is the radiance[2] of the object behind the window, as measured when there is no window. Similarly, $L_{\rm r}$ is the radiance of the reflected object, as measured if the window was replaced by a perfect mirror.

Consider Fig. 2. A light ray from the object $L_{\rm r}$ reaches the window. There, it undergoes a series of reflections and refractions. The internal reflections inside the window create a series of rays emerging from the window. Since the window is flat, all the rays are in the same plane, termed the *plane of incidence* (POI). The distance between successive emerging rays (secondary reflections) is d'. The power of successive secondary reflections rapidly tends to zero. A similar effect occurs with a ray from the object $L_{\rm t}$, as illustrated in Fig. 2. Prior studies neglected the shift d', hence spatial effects of secondary reflections were ignored. This assumption was valid as long as the window was thin and viewed in low spatial resolution, but it is not true in general.

Reflection is sensitive to polarization.[3] The polarization component perpendicular to the POI is denoted by $\perp$, while $\|$ denotes the component parallel to the POI. The reflectance coefficients [21] from each interface of the window are

$$R_\perp = \frac{\sin^2(\phi-\phi_g)}{\sin^2(\phi+\phi_g)} \;\;,\;\; R_\| = \frac{\tan^2(\phi-\phi_g)}{\tan^2(\phi+\phi_g)}, \qquad (1)$$

where ϕ is the angle of incidence of light on the window (relative to the surface normal). Here ϕ_g is the angle of the refracted ray inside the glass. This angle is derived using *Snell's law* $\sin(\phi_g) = \sin(\phi)/n$, where n is the index of refraction of the window (for typical glass, $n \approx 1.5$). The transmittance coefficients [21] of light passing any one of the window interfaces are

$$T_\perp = 1 - R_\perp \;\;,\;\; T_\| = 1 - R_\|. \qquad (2)$$

The image coordinates are (*x*,*y*), where *x* is the horizontal coordinate. Here the *horizontal* direction in the image is defined as the projection of the POI on the detector plane.

It is clear from Fig. 2 that the secondary reflections create a spatial effect. Each object point is sensed simultaneously in different pixels, as its energy is dissipated among the different reflection orders. Hence the transmitted scene undergoes a *convolution* with a particular PSF: as seen in Fig. 2, the PSF of the transmitted scene is

$$h_{\rm t}^{\|} = T_\|^2[\delta(x) + R_\|^2\delta(x-d) + R_\|^4\delta(x-2d)\ldots] \;\;, \quad (3)$$

when measuring only the polarization component parallel to the POI, while $R_\|$ and $T_\|$ are given in Eqs. (1,2). Here d indicates the distance between successive visual echoes of $L_{\rm t}$, as received by the camera (in pixels). It is given by $d = \alpha d'$, where d' is the physical distance (in centimeters) between secondary reflections, depicted in Fig. 2, and α is the camera magnification.

Note that in this model, each object point corresponds to a parallel set of rays, which in turn correspond to a set equally interspaced pixels. This is consistent with orthographic projection,[4] which we use for simplicity. Similarly, the PSF of the reflected scene is

$$h_{\rm r}^{\|} = R_\|[\delta(x) + T_\|^2\delta(x-d) + T_\|^2R_\|^2\delta(x-2d)\ldots] \;\;, \quad (4)$$

when measuring only the parallel polarization component. The perpendicular components also undergo convolutions.

[1] Polarization has been used in studies of various computer vision issues [2, 6, 7, 16, 17, 24, 26, 28, 29].

[2] There is a constant proportion between the object radiance and the image irradiance. The proportion coefficient does not depend on the scenes or on the parameters of the problem. This coefficient depends only on the camera, and thus we disregard it in the context of our problem.

[3] Ref. [15] analyzed the polarization in internal reflections, in order to recover object shapes, including a flat slab.

[4] In perspective projection, chief rays are not parallel, but correspond to a slightly fanning beam emanating from the object point. In perspective, d and the PSFs are not spatially invariant. The consequences of the non-orthographic nature of the camera are discussed in [10].

Figure 3. Simulated objects. [Left] $L_{\rm t}$. [Right] $L_{\rm r}$.

Figure 4. A simulated acquired frame $S_\perp$. It exhibits both secondary reflections and superposition of a reflected scene with a transmitted scene. The separate scenes are shown in Fig. 3.

The corresponding PSFs $h_{\rm r}^\perp$ and $h_{\rm t}^\perp$ are derived analogously, by using $R_\perp$ and $T_\perp$ instead of $R_\parallel, T_\parallel$ in Eqs. (3,4).

The acquired image intensity is a linear superposition of the reflected and transmitted scenes. In Refs. [11, 21], this superposition was pointwise, since in the imaging conditions there, spatial effects were not seen. In contrast, here the superposition is of *convolved* scenes. Specifically, let us mount a polarizing filter on the camera, and orient the filter to pass only the parallel polarization component. Assuming as in Refs. [5, 11, 12, 21] that the objects $\{L_{\rm t}, L_{\rm r}\}$ are largely depolarized, the acquired frame is

$$S_\parallel = L_{\rm t} \star h_{\rm t}^\parallel + L_{\rm r} \star h_{\rm r}^\parallel \quad , \tag{5}$$

where $\star$ denotes convolution. Similarly, orienting the polarizer perpendicular to the POI yields

$$S_\perp = L_{\rm t} \star h_{\rm t}^\perp + L_{\rm r} \star h_{\rm r}^\perp \quad . \tag{6}$$

We can illustrate this using a simulation. Fig. 3 represents the original objects $L_{\rm t}$ and $L_{\rm r}$. Let $\phi = 70^o$ in a glass window and $d = 30$ pixels. Based on these values, the reflectance, transmittance and PSFs are given in closed form by the expressions above. Hence the simulated acquired images are given by Eqs. (5,6). Specifically, Fig. 4 shows $S_\perp$.

3. Frequency Analysis of Conditioning

Sec. 2 showed that $L_{\rm t}$ and $L_{\rm r}$ are convolved by PSFs in the raw frames $S_\parallel$ and $S_\perp$. Therefore, in order to restore $L_{\rm t}$ and $L_{\rm r}$ we need to perform deconvolution. Let us examine the frequency response of these PSFs. The PSF $h_{\rm t}^\parallel$ is given in Eq. (3). Plotting its frequency response reveals that the response is rather flat, and typically ~ 1. The same applies to $h_{\rm t}^\perp$. Hence inverse-filtering of these PSFs is expected to be stable, making the recovery of $L_{\rm t}$ *well-conditioned*. The situation is different for $L_{\rm r}$. Its PSF $h_{\rm r}^\parallel$ is given in Eq. (4). The corresponding frequency response is plotted in Fig. 5. This response has values close to 0. Thus, some frequencies of the original $L_{\rm r}$ are greatly attenuated. Consequently, recovery of $L_{\rm r}$ is *ill-conditioned* around such frequencies.

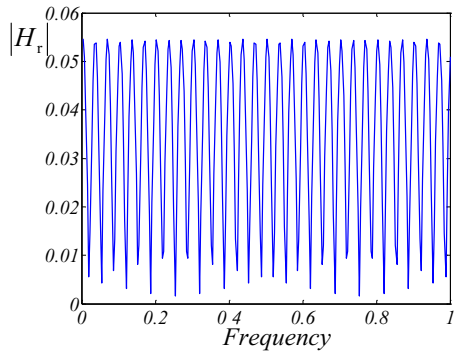

Figure 5. Frequency response of $h_{\rm r}$ for $d = 30$ pixels and $\phi = 27^o$. The frequency is normalized by the Nyquist frequency.

4. Recovery with Deconvolution

In this section we recover the source objects $\{L_{\rm t}, L_{\rm r}\}$. We show that this can be done using linear filters that are derived in closed-form, and invert the image formation model. Then, we point to a problem of simplistic inversion. Finally, we describe how recovery was performed by us in practice.

4.1. Linear Filtering

For the moment, assume that the values of the parameters d and ϕ are known. Their estimation is described in Sec. 5. Based on d and ϕ, the coefficients $R_\perp$, $R_\parallel$, $T_\perp$ and $T_\parallel$ are derived, thus the PSFs $h_{\rm t}^\parallel, h_{\rm r}^\parallel, h_{\rm t}^\perp, h_{\rm r}^\perp$ are known (Sec. 2).

First, we eliminate $L_{\rm t}$. It is easy to show that the filters

$$\tilde{h}_{\rm t}^\parallel(x) = \frac{1}{T_\parallel^2}[\delta(x) - R_\parallel^2 \delta(x-d)], \tag{7}$$

$$\tilde{h}_{\rm t}^\perp(x) = \frac{1}{T_\perp^2}[\delta(x) - R_\perp^2 \delta(x-d)] \tag{8}$$

satisfy

$$\tilde{h}_{\rm t}^\perp \star h_{\rm t}^\perp = \delta(x) \quad , \quad \tilde{h}_{\rm t}^\parallel \star h_{\rm t}^\parallel = \delta(x) \; . \tag{9}$$

Let us convolve Eq. (5) with Eq. (7). Then based on Eq. (9)

$$S_\parallel \star \tilde{h}_{\rm t}^\parallel = L_{\rm t} + L_{\rm r} \star h_{\rm r}^\parallel \star \tilde{h}_{\rm t}^\parallel . \tag{10}$$

Similarly, convolving Eq. (6) with Eq. (8) yields

$$S_\perp \star \tilde{h}_{\rm t}^\perp = L_{\rm t} + L_{\rm r} \star h_{\rm r}^\perp \star \tilde{h}_{\rm t}^\perp \quad . \tag{11}$$

Subtracting Eq. (10) from Eq. (11) yields

$$U = L_{\rm r} \star p, \tag{12}$$

where

$$U = S_{\perp} \star \tilde{h}_{\rm t}^{\perp} - S_{\parallel} \star \tilde{h}_{\rm t}^{\parallel}, \tag{13}$$

and

$$p \equiv h_{\rm r}^{\perp} \star \tilde{h}_{\rm t}^{\perp} - h_{\rm r}^{\parallel} \star \tilde{h}_{\rm t}^{\parallel}. \tag{14}$$

Eqs. (12) and (13) eliminate $L_{\rm t}$, thus isolating $L_{\rm r}$. However, $L_{\rm r}$ is still not recovered, since it is convolved with p in Eq. (12). Hence, we need to deconvolve the effect of p. In other words, we need the function v that satisfies

$$v \star p = \delta(x). \tag{15}$$

This function has a simple analytical form, directly in the spatial domain. We detail it in Sec. 4.2. Applying v on Eq. (12) yields an estimate

$$\hat{L}_{\rm r} = U \star v \ , \tag{16}$$

based on Eq. (15). Plugging Eq. (13) in Eq. (16) yields

$$\hat{L}_{\rm r} = S_{\perp} \star q_{\rm r}^{\perp} + S_{\parallel} \star q_{\rm r}^{\parallel}, \tag{17}$$

where

$$q_{\rm r}^{\perp} = \tilde{h}_{\rm t}^{\perp} \star v \quad , \quad q_{\rm r}^{\parallel} = -\tilde{h}_{\rm t}^{\parallel} \star v. \tag{18}$$

We now solve for $L_{\rm t}$. Using Eqs. (13,16) in (11) yields

$$\hat{L}_{\rm t} = S_{\perp} \star q_{\rm t}^{\perp} + S_{\parallel} \star q_{\rm t}^{\parallel} \ , \tag{19}$$

where

$$q_{\rm t}^{\perp} = \tilde{h}_{\rm t}^{\perp} - \tilde{h}_{\rm t}^{\perp} \star v \star h_{\rm r}^{\perp} \star \tilde{h}_{\rm t}^{\perp} \quad , \quad q_{\rm t}^{\parallel} = \tilde{h}_{\rm t}^{\parallel} \star v \star h_{\rm r}^{\perp} \star \tilde{h}_{\rm t}^{\perp}. \tag{20}$$

Eqs. (17) and (19) are the basic recovery formulae. They show that $\{L_{\rm t}, L_{\rm r}\}$ can be recovered by operation of linear filters given in closed form directly in the spatial domain. These filters are given by Eqs. (7,8,18,20), and rely on v that is given in closed form in Sec. 4.2. To see the form of these filters, Fig. 6 plots $q_{\rm t}^{\perp}$ corresponding to $\phi = 27^o$ and $d = 5$ pixels. The filters $q_{\rm t}^{\parallel}, q_{\rm r}^{\perp}, q_{\rm r}^{\parallel}$ have a similar form.

4.2. The Filters p and v

We now derive the operator v. Define the coefficients

$$a \equiv \left(\frac{R_{\perp}}{T_{\perp}^2} - \frac{R_{\parallel}}{T_{\parallel}^2} \right) \quad , \quad b \equiv \left(\frac{R_{\perp} - 2R_{\perp}^2}{T_{\perp}^2} - \frac{R_{\parallel} - 2R_{\parallel}^2}{T_{\parallel}^2} \right). \tag{21}$$

Using Eq. (21) in Eqs. (3,4,7,8) and (14), it can be shown that p has a simple form

$$p(x) \equiv a\delta(x) + b\delta(x - d). \tag{22}$$

From Eqs. (15) and (22), v should satisfy

$$v \star [a\delta(x) + b\delta(x - d)] = \delta(x). \tag{23}$$

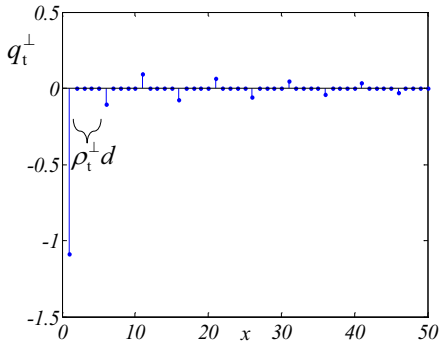

Figure 6. The filter $q_{\rm t}^{\perp}$ defined in Eq. (20). Here $\phi = 27^o$, $d = 5$ pixels. The effective width of $q_{\rm t}^{\perp}$ is $\rho_{\rm t}^{\perp} d$, based on Eq. (28).

To derive v, we use Fourier analysis. Following Eq. (22), the frequency response of p is

$$P(\omega) = a + be^{-j\omega d}. \tag{24}$$

where ω is the spatial frequency. From Eqs. (15,23), v is the inverse filter of p, hence the frequency response of v is

$$V(\omega) = \frac{1}{P(\omega)} = \frac{1}{a} \cdot \frac{1}{1 - \left(\frac{-b}{a}\right) e^{-j\omega d}}. \tag{25}$$

We note that b and a defined in Eq. (21) satisfy $|b/a| < 1$. Hence, the left hand side of (25) is a geometric series, i.e.,

$$V(\omega) = \frac{1}{a} \sum\nolimits_{m=0}^{\infty} \left(\frac{-b}{a} \right)^m e^{-j\omega md}. \tag{26}$$

The inverse Fourier transform of Eq. (26) is the sum

$$v(x) = \frac{1}{a} \left[\delta(x) + \sum\nolimits_{m=1}^{\infty} \left(\frac{-b}{a} \right)^m \delta(x - md) \right] \ . \tag{27}$$

This result can be verified in the spatial domain: convolving Eq. (27) with p (Eq. 22), the result can be shown to yield a δ function, as required by Eqs. (15,23).

4.3. Problems Caused by Boundary Conditions

The recovery problem seriously suffers from unknown boundary conditions. Actually, this problem is so serious, that sometimes it is preferable not to attempt inversion of the reverberations, unless the effect of unknown boundary conditions is addressed. Eqs. (5) and (6) are valid for images that have infinite support in the x-axis. In practice, the raw frames have a finite support $x \in [0, W]$. However, the convolution kernels $q_{\rm t}^{\perp}, q_{\rm t}^{\parallel}, q_{\rm r}^{\perp}$ and $q_{\rm r}^{\parallel}$ require values outside the boundaries of the raw frames.

Without loss of generality, let $d > 0$. Then, the support of $h_{\rm t}^{\parallel}, h_{\rm r}^{\parallel}, h_{\rm t}^{\perp}, h_{\rm r}^{\perp}$ (Eqs. 3,4) resides only in $x \geq 0$, i.e., these are causal filters. Consequently, only the unknown values in $x < 0$ cause a problem of boundary conditions. Unknown values at $x > W$ are not problematic.

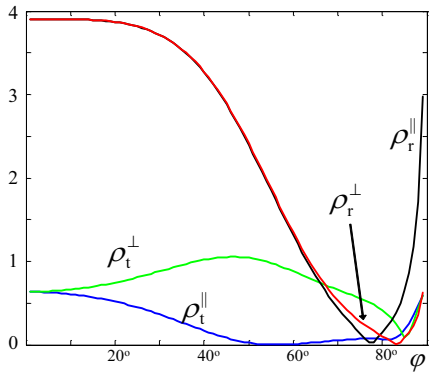

Figure 7. The effective widths $\rho_t^{\perp}, \rho_t^{\parallel}, \rho_r^{\perp}, \rho_r^{\parallel}$ as multiples of d. The plot corresponds to a glass window ($n = 1.5$). For most angles ϕ, the values of $\rho_r^{\perp}$ and $\rho_r^{\parallel}$ are significantly larger than $\rho_t^{\perp}, \rho_t^{\parallel}$.

The filters $q_t^{\perp}, q_t^{\parallel}, q_r^{\perp}$ and $q_r^{\parallel}$ are causal and have an infinite support. Their infinite support stems from the presence of v in them (Eq. 18,20), as seen by Eq. (27). However, practically, these filters decay fast, and we may define their *effective* width. The effective width of $q_t^{\perp}$ is defined by

$$\rho_t^{\perp} = \frac{1}{d}\sqrt{\frac{\sum\limits_x x^2 q_t^{\perp 2}(x)}{\sum\limits_x q_t^{\perp 2}(x)}}, \tag{28}$$

as multiples of d. Similarly, the effective widths $\rho_t^{\parallel}, \rho_r^{\perp}$ and $\rho_r^{\parallel}$ corresponding respectively to $q_t^{\parallel}, q_r^{\perp}$ and $q_r^{\parallel}$ can be calculated. Fig. 7 plots these values as a function of ϕ. For most angles ϕ, the values $\rho_r^{\perp}$ and $\rho_r^{\parallel}$ are significantly larger than $\rho_t^{\perp}$ and $\rho_t^{\parallel}$. Recall that $\hat{L}_r$ is affected by $\rho_r^{\perp}$ and $\rho_r^{\parallel}$ (Eq. 17), while $\hat{L}_t$ is affected by $\rho_t^{\perp}$ and $\rho_t^{\parallel}$ (Eq. 19). In other words, $\hat{L}_r$ is affected by filters with effectively much longer 'tails' than $\hat{L}_t$. Thus, problems associated with boundary conditions are generally expected to be more severe in $\hat{L}_r$.

Fig. 8 shows this severe effect, when reconstruction is based simply on zero-padding. In this simulation, $L_t = 0$, thus the only task of the recovery is elimination of the secondary reflections of L_r. In some cases, the created strips may be more disturbing (subjectively and objectively) than the original reverberations, undermining the recovery.

4.4. Solution in Practice

We now describe how we performed the recovery in practice. First, to significantly reduce problems associated with boundary conditions, we use mirror-padding at $x \leq 0$. The resulting images still have artifacts, but they decay with x. Then, reconstruction of L_t is done using Eq. (19). We found that practically $\hat{L}_t$ tolerates the unknown boundary conditions (provided that mirror padding is used). Moreover, as discussed in Sec. 3, the PSFs that act upon L_t are well conditioned. For this reason, the fast and simple linear filtering in Eq. (19) proved sufficient.

Figure 8. Inversion using zero padding yields sharp stripe artifacts.

The reconstruction of L_r is more difficult. It suffers more from unknown boundary conditions, due to the larger effective widths of $q_r^{\perp}$ and $q_r^{\parallel}$. Moreover, some frequency components of $\hat{L}_r$ are fundamentally ill conditioned, as discussed in Sec. 3. Therefore, we do not calculate $\hat{L}_r$ by the simplistic linear filtering described in Sec. 4.1. Rather, we pursue deconvolution of p as a solution to a regularized optimization problem. Based on Eq. (13), we solve

$$\hat{L}_r = \arg\min_{L_r} \left(\|U - L_r \star p\|^2 + \lambda \left\|\nabla^2 L_r\right\|^2 \right) \quad . \tag{29}$$

In Eq. (29), the term $\|U - L_r \star p\|$ is a *fitting term*. It is minimal when the data fits the model well. The term $\lambda \left\|\nabla^2 L_r\right\|$ introduces regularization. The parameter λ sets the relative weight between these two terms. Here regularization leads to a smooth image $\hat{L}_r$. However, other regularization terms from the literature [13] can be used.

The computational complexity of the filtering operation is $\mathcal{O}(\#\text{ of image pixels})$. The regularized solution is somewhat slower, since Eq. (29) is solved iteratively. Each iteration is $\mathcal{O}(\#\text{ of image pixels})$, and we observed that convergence effectively occurred within ≈ 25 normalized steepest descent iterations.

5. Estimation of Parameters

Up to this point, the parameters of the problem (d and ϕ) were assumed to be known. Their estimation is now detailed. The incidence angle ϕ at the window is independent of the wavelength. The displacement d is also practically insensitive to the wavelength. Hence, these parameters are estimated based on a grayscale (panchromatic) representation of the raw images, discarding the color.

Determining ϕ and the x axis: Estimation of ϕ is done in the same manner as in [21]. Furthermore, Ref. [21] describes how the axis corresponding to the POI is determined in the image plane, based on the polarization.[5] In our work, this determines the x axis, along which the displacement d of the visual reverberations occurs, as written in Sec. 2.

Determining d: A challenge raised by this study is the estimation of d. First, we estimate $|d|$. Then, $\text{sign}(d)$ is found.

[5] This is determined up to a 90^o ambiguity.

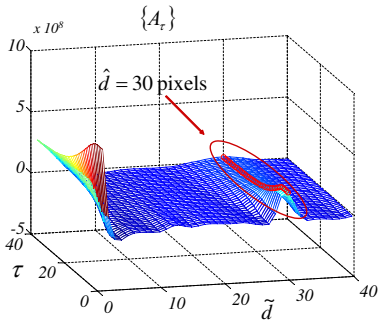

Figure 9. $\mathcal{A}_\tau$ in a simulation corresponding to images as appear in Figs. 3,4. The peaks at $\tilde{d} = 30$ are marked by small red circles.

Figure 10. Simulated reconstructions corresponding to Fig. 3. [Left] $\hat{L}_{\rm r}$. [Right] $\hat{L}_{\rm t}$.

The reverberations in Figs. 1 and 4 create displaced replications of the image content. Apparently, this should create a secondary peak of the autocorrelation of a raw frame ($S_\perp$), as a function of a hypothesized displacement $\tilde{d}$. In practice, often such a peak at $|d|$ does not appear. Nevertheless, the anticipated peak at $|d|$ appears when autocorrelation is performed over the horizontal derivative of a raw frame, e.g., $|\partial_x S_\perp(x,y)|$. There is a still a problem in practice: local maxima of the autocorrelation function appear in additional values of $\tilde{d} \neq |d|$. These incorrect maxima change if $\partial_x S_\perp(x,y)$ is blurred (by a Gaussian filter of width τ). On the other hand, the correct peak at $\tilde{d} = |d|$ is consistent despite such blurring action.

This is seen, for instance, in Fig. 9. This plot is based on a simulated frame, similar to Fig. 4. The autocorrelation A_τ function is parameterized by the Gaussian width τ. The consistency of the correct peak at $|d|$ despite the change in τ is revealed by a simple voting process. This yields the final estimate $\hat{d}$. For example, in Fig. 9, this correctly yielded $\hat{d} = |d|$ (which was 30 pixels in this case).

The Sign of d

An autocorrelation function is symmetric around the origin. Hence, it gives no indication whether $d > 0$ or $d < 0$. To determine $\mathrm{sign}(d)$, a different criterion is developed, based on the following observation. Consider a single horizontal line profile of the images at $\tilde{y}$. There, let the source images $\{L_{\rm r}(x,\tilde{y}), L_{\rm t}(x,\tilde{y})\}$ be flat, except for a single edgel at $\tilde{x}$ in one of the sources. This edge appears also in the raw frames, e.g. in $S_\perp(\tilde{x},\tilde{y})$, with an absolute derivative $|\partial_x S_\perp(\tilde{x},\tilde{y})|$. Due to internal reflections, this edge reverberates and appears also in $(\tilde{x}+d,\tilde{y})$, $(\tilde{x}+2d,\tilde{y})$ etc. However, the strength of the edge weakens in each order, as the PSFs $h_{\rm t}^{\parallel}, h_{\rm r}^{\parallel}, h_{\rm t}^{\perp}, h_{\rm r}^{\perp}$. Hence,

$$|\partial_x S_\perp(\tilde{x},\tilde{y})| > |\partial_x S_\perp(\tilde{x}+|\hat{d}|,\tilde{y})| > |\partial_x S_\perp(\tilde{x}+2|\hat{d}|,\tilde{y})|. \quad (30)$$

if $0 < d$ and $\hat{d} = d$.

Consider a typical image, having a typical content and random noise, but *no* reverberations. Take a triplet of pixels $\{(\tilde{x},\tilde{y}), (\tilde{x}+|\hat{d}|,\tilde{y}), (\tilde{x}+2|\hat{d}|,\tilde{y})\}$. Eq. (30) should hold in some positions $(\tilde{x},\tilde{y})$, and be violated in other places in this image. Define $\mathcal{C}_+$ as the set of all pixels $(\tilde{x},\tilde{y})$ in the image that satisfy Eq. (30), for a specific $\hat{d}$. Similarly, define $\mathcal{C}_-$ as the set of all pixels $(\tilde{x},\tilde{y})$ in the image that satisfy

$$|\partial_x S_\perp(\tilde{x},\tilde{y})| < |\partial_x S_\perp(\tilde{x}+|\hat{d}|,\tilde{y})| < |\partial_x S_\perp(\tilde{x}+2|\hat{d}|,\tilde{y})|. \quad (31)$$

On average, in a typical image, Eq. (30) is expected to hold in a similar number of possible triplets as the number satisfying Eq. (31). For a random image having N pixels, $|\mathcal{C}_+| = |\mathcal{C}_-| = N/4$. However, the presence of reverberations creates a *bias* in this randomness. For instance, pixels satisfying Eq. (31) comply with a reverberation model in which $d < 0$ (orders decay leftwards). Hence, we set

$$\mathrm{sign}(\hat{d}) = \mathrm{sign}\left(|\mathcal{C}_+| - |\mathcal{C}_-|\right) . \quad (32)$$

We applied this criterion successfully in various simulations and in the experiments. The bias of $|\mathcal{C}_+|$ vs. $|\mathcal{C}_-|$ was $\approx 8\%$.

6. Validation

6.1. Simulation Example

The reflectance (Eq. 1) is typically much smaller than the transmittance (Eq. 2). Thus, to get a noticeable mixup of the two layers in the acquired images $S_\perp$ and $S_\parallel$, $L_{\rm r}$ should typically be very bright. We used $L_{\rm t}$ and $L_{\rm r}$ as in Fig. 3, where $L_{\rm t}(x) \in [0,113]$ and $L_{\rm r}(x) \in [0,513]$. The maximal value of $S_\perp$ and $S_\parallel$ was 255.

We used $x \in [-d, W]$, where $W = 226$ pixels, d=30 pixels, and $\phi = 27^o$. Gaussian noise having standard deviation of 3 gray levels was added independently to every pixel in $S_\perp$ and $S_\parallel$. The simulated acquired images $S_\perp$ and $S_\parallel$ look similar to Fig. 4. Then, the unavailability of boundary values is simulated by chopping off the whole part corresponding to $x < 0$ in the frames, leaving their support to be $x \in [0, W]$. Now, we simulated the reconstruction. We used mirror extrapolation as described in Sec. 4.3. Fig. 10 depicts the reconstructions obtained as described in Sec. 4.4, using

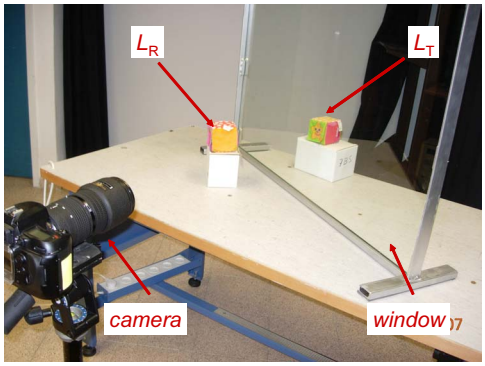

Figure 11. The setup used in the experiments.

$\lambda = 0.01$. The scenes are separated, and the reverberations are eliminated, making the result more visually pleasing.

Quantitative Assessment

To quantitatively measure the recovery, we use the mean squared error (MSE) in simulations, where we have access to the ground truth. First, let us ignore the spatial effect of the visual reverberations, as in the state of the art. In this case, we simply run the pointwise method of [21]. Here we obtained a value $\mathrm{MSE}_{\mathrm{r}}^{\mathrm{pointwise}} = 161$. When accounting for the spatial effect of reverberations using our method, we obtained $\mathrm{MSE}_{\mathrm{r}}^{\mathrm{recovery}} = 53$. Hence, quantitatively, the method greatly improved the MSE. This quantitative improvement is evident, since reverberations in the raw data were significant quantitatively and subjectively (visually).

6.2. Experimenting with Real-World Objects

We applied the method on a real setup. We used a Nikon D100 camera, to obtain data which is linearly related to the scene radiance (no γ correction). A $200mm$ lens and a polarizer were fitted to it. The camera was set in front of a glass window, similarly to the way depicted in Fig. 11. We acquired a few frames in various polarizer orientations. Based on them, we derived $S_\perp$ and $S_\parallel$, as described in [17]. The image $S_\perp$ is shown in Fig. 1. It clearly demonstrates the secondary reflection (reverberation), as well as the confusion caused by the superposition of the reflected and transmitted scenes. The image $S_\parallel$ looks similar to it.

The estimation of the parameters was performed as described in Sec. 5. The automatic estimation of d yielded $\hat{d}$=36 pixels, which was consistent with manual measurement. The estimated ϕ is 41^o. Consequently, the recovery described in Sec. 4.4 was applied separately to each color band. Finally, all the processed color bands were combined to the resulting output color images. The final[6] reconstructed $\hat{L}_{\mathrm{t}}$ is depicted in Fig. 12 while the final $\hat{L}_{\mathrm{r}}$ is depicted in Fig. 13. The reconstructions are visually pleasing. They have no crosstalk (good separation) and the visual reverberations are eliminated. In another experiment, the acquired image $S_\parallel$ appears on the left side of Fig. 14. The acquired image $S_\perp$ looks similar. The reconstructed $\hat{L}_{\mathrm{r}}$ in this experiment appears on the right side of Fig. 14.

[6]The reconstructed images contain residual edge artifacts. An explanation hypothesis and the way we overcame them are described in [10].

Figure 12. The reconstructed $\hat{L}_{\mathrm{t}}$ in the experiment corresponding to Fig. 1. It has neither visual reverberations nor apparent trace of the complementary scene L_{r}, whose estimate is shown in Fig. 13.

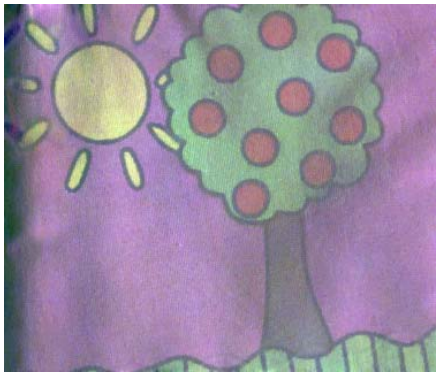

Figure 13. The reconstructed $\hat{L}_{\mathrm{r}}$ in the experiment corresponding to Fig. 1. It has neither visual reverberations nor apparent trace of the complementary scene L_{t}, whose estimate is shown in Fig. 12.

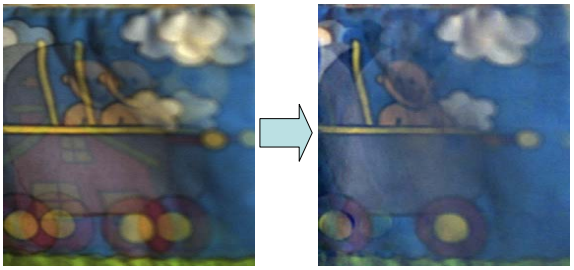

Figure 14. [Left] A real-world image $S_\parallel$. It contains a superposition of two scenes. The visual reverberation, e.g., of the baby, can clearly be seen. [Right] The reconstructed $\hat{L}_{\mathrm{r}}$ is separated while the reverberation is eliminated.

7. Discussion

The presented closed form physical model elucidates the fundamental limitations of the problem (conditioning and boundary conditions), for each source scene. Furthermore, it creates the basis for future, improved recovery algorithms. The task is essentially one of solving a convolutive mixture (see [23]). The work can be extended to methods that do not rely on a polarizer, as has been done in other studies that dealt with transparent scenes. Note that the true d may be non-integer. This creates residual errors that may need to be assessed. Moreover, this aspect may be incorporated explicitly into the algorithm. The analysis may also be generalized to non-planar windows.

Acknowledgments

We thank Einav Namer for helping in the experiments and Arie Yeredor for useful comments. Yoav Schechner is a Landau Fellow - supported by the Taub Foundation. The work was supported by the Israeli Ministry of Science, Culture and Sport (Grant 3-3426). It was conducted in the Ollendorff Minerva Center. Minerva is funded through the BMBF.

References

[1] A. Agrawal, R. Raskar, S. K. Nayar, and Y. Li. Removing photography artifacts using gradient projection and flash-exposure sampling. *ACM TOG*, 24:828–835, 2005.

[2] G. A. Atkinson and E. R. Hancock. Shape estimation using polarization and shading from two views. *IEEE Trans. PAMI*, 29:2001–2017, 2007.

[3] F. Basbug, K. Swaminathan, and S. Nandkumar. Noise reduction and echo cancellation front-end for speech codecs. *IEEE Trans. Speech & Audio Process.*, 11(1):1–13, 2003.

[4] E. Be'ery and A. Yeredor. Blind separation of superimposed shifted images using parameterized joint diagonalization. *IEEE Trans. IP*, 17:340–353, 2008.

[5] A. M. Bronstein, M. M. Bronstein, M. Zibulevsky, and Y. Y. Zeevi. Sparse ICA for blind separation of transmitted and reflected images. *Int. J. Imag. Sys. Tech.*, 15:84–91, 2005.

[6] T. Chen, H. P. A. Lensch, C. Fuchs, and H. Seidel. Polarization and phase-shifting for 3d scanning of translucent objects. *Proc. IEEE CVPR*, 1–8, 2007.

[7] O. G. Cula, K. J. Dana, D. K. Pai, and D. Wang. Polarization multiplexing and demultiplexing for appearance-based modeling. *IEEE Trans. PAMI*, 29:362–367, 2007.

[8] M. Dadic. Influence of channel transfer function on LMS algorithm basednonrecursive deghosting in analog TV. *Int. Conf. Trends in Comm., EUROCON*, 2, 2001.

[9] T. Darrell and E. Simoncelli. Nulling filters and the separation of transparent motions. In *Proc. IEEE CVPR*, 738–739, 1993.

[10] Y. Diamant and Y. Y. Schechner. Eliminating artifacts when inverting visual reverebrations. *Tech. Rep.* CCIT-692, Dept. Electrical Eng., Technion, March 2008.

[11] H. Farid and E. H. Adelson. Separating reflections from images by use of independent component analysis. *JOSA A*, 16:2136–2145, 1999.

[12] Hermanto, A. K. Barros, T. Yamamura, and N. Ohnishi. Separating real and virtual objects using independent component analysis. *IEICE Trans. Inf. & Syst.*, E84-D:1–9, 2001.

[13] R. Kaftory, Y. Y. Schechner and Y. Y. Zeevi Variational distance-dependent image restoration *Proc. IEEE CVPR* 2007.

[14] A. Levin and Y. Weiss. User assisted separation of reflections from a single image using a sparsity prior. *Proc. ECCV*, 602–613, 2004.

[15] D. Miyazaki and K. Ikeuchi. Shape estimation of tranparent objects by using inverse polarization ray tracing. *IEEE Trans. PAMI*, 29:2018–2029, 2007.

[16] D. Miyazaki, M. Saito, Y. Sato, and K. Ikeuchi. Determining surface orientations of transparent objects based on polarization degrees in visible and infrared wavelengths. *JOSA A*, 19:687–694, 2002.

[17] S. K. Nayar, X. S. Fang, and T. Boult. Separation of reflection components using color and polarization. *IJCV*, 21:163–186, 1997.

[18] B. Sarel and M. Irani. Separating transparent layers through layer information exchange. *Proc. ECCV*, 4:328–341, 2004.

[19] Y. Y. Schechner, N. Kiryati, and R. Basri. Separation of transparent layers using focus. *IJCV*, 39:25–39, 2000.

[20] Y. Y. Schechner and S. K. Nayar. Generalized mosaicing: Polarization panorama. *IEEE Trans. PAMI*, 27:631–626, 2005.

[21] Y. Y. Schechner, J. Shamir, and N. Kiryati. Polarization and statistical analysis of scenes containing a semi-reflector. *JOSA A*, 17:276–284, 2000.

[22] M. Shizawa. Direct estimation of multiple disparities for transparent multiple surfaces in binocular stereo. In *Proc. IEEE ICCV*, 447–454., 1993.

[23] S. Shwartz, Y. Y. Schechner and M. Zibulevsky. Efficient separation of convolutive image mixtures. In *Proc. ICA (LNCS 3889)*, 246-253, 2006.

[24] S. Shwartz, E. Namer, and Y. Y. Schechner. Blind haze separation. In *Proc. IEEE CVPR*, 1984-1991, 2006.

[25] R. Szeliski, S. Avidan, and P. Anandan. Layer extraction from multiple images containing reflections and transparency. *Proc. IEEE CVPR*, 1:246–253, 2000.

[26] T. Treibitz and Y. Y. Schechner. Instant *3D*escatter. In *Proc. IEEE CVPR*, vol. 2, 1861–1868, 2006.

[27] Y. Tsin, S. B. Kang, and R. Szeliski. Stereo matching with linear superposition of layers. *IEEE Trans. PAMI*, 28:290–301, 2006.

[28] S. Umeyama and G. Godin. Separation of diffuse and specular components of surface reflection by use of polarization and statistical analysis of images. *IEEE Trans. PAMI*, 26:639–647, 2004.

[29] L. B. Wolff. Polarization vision: a new sensory approach to image understanding. *Image & Vision Comp.*, 15:81–93, 1997.