On Refractive Optical Flow

Sameer Agarwal, Satya P. Mallick, David Kriegman, and Serge Belongie

University of California, San Diego, La Jolla CA 92093, USA,
{sagarwal@cs,spmallick@graphics,kriegman@cs,sjb@cs}.ucsd.edu,
http://vision.ucsd.edu/

Abstract. This paper presents a novel generalization of the optical flow equation to the case of refraction, and it describes a method for recovering the refractive structure of an object from a video sequence acquired as the background behind the refracting object moves. By structure here we mean a representation of how the object warps and attenuates (or amplifies) the light passing through it. We distinguish between the cases when the background motion is known and unknown. We show that when the motion is unknown, the refractive structure can only be estimated up to a six-parameter family of solutions without additional sources of information. Methods for solving for the refractive structure are described in both cases. The performance of the algorithm is demonstrated on real data, and results of applying the estimated refractive structure to the task of environment matting and compositing are presented.

1 Introduction

The human visual system is remarkable in its ability to look at a scene through a transparent refracting object and to deduce the structural properties of that object. For example, when cleaning a wine glass, imperfections or moisture may not be visible at first, but they become apparent when one holds the glass up and moves or rotates it. We believe that the primary cue here is the optical flow of the background image as observed through the refracting object, and our aims in this paper are to build a theory of how motion can be used for recovering the structure of a refracting object, to introduce algorithms for estimating this structure, and to empirically validate these algorithms. By structure here we mean a representation of how the object warps and attenuates (or amplifies) the light passing through it. Recall, as a light ray enters or exits the object, its direction changes according to Snell's Law (known as Descartes' Law in France). Furthermore, the emitted radiance may differ from the incident radiance due to the difference in solid angle caused by the geometry of the interfaces between the object and the air as well as absorption along the light ray's path through the object. The geometric shape of the object itself, while of independent interest, is not the focus of our inquiry here.

The primary contribution of this paper is to generalize the optical flow equation to account for the warping and attenuation caused by a refractive object, and to present algorithms for solving for the warp and attenuation using a sequence of images obtained as a planar background moves behind the refracting

T. Pajdla and J. Matas (Eds.): ECCV 2004, LNCS 3022, pp. 483–494, 2004.

object. Both the case where the background motion is known and where it is unknown are considered. We demonstrate the performance of these algorithms on both synthetic and real data.

While there is a vast literature on motion understanding including transparency, the warp induced by refraction appears to have been neglected hitherto [1, 2,3,4]. The ability to recover the refractive structure of an object has a number of applications. Recently, environment matting and compositing have been introduced as techniques for rendering images of scenes which contains foreground objects that refract and reflect light [5,6,7]. The proposed approach offers a new method for creating a matte of real objects without the need for extensive apparatus besides a video camera. Refractive optical flow may also be useful for visual inspection of transparent/translucent objects.

This paper was inspired by the work of Zongker et. al [7] on environment matting and its subsequent extension by Chuang et. al. [5]. We discuss this method in the next section. However the work that comes closest to ours in spirit is that of H. Murase [8]. Murase uses optical flow to recover the shape of a flexible transparent object, e.g., water waves that change over time. To make the problem tractable he makes a number of simplifying assumptions (a) The camera is orthographic, (b) The refracting object is in contact with the background plane which has a flat shape and a static unknown pattern on it, (c) the average shape over time of the refracting surface is known *a priori* to have zero slope. In this paper our interest is in scenes where the refracting object is rigid and stationary. Beyond that, we make no assumptions about the number of objects, their refractive indices, or the positioning of the refracting object in the scene.We do not address effects due to specular, Fresnel or total internal reflections in this work.

The rest of the paper is organized as follows. In the next section we begin by describing our image formation model and we then introduce the notion of an optical kernel and describe how the choice of a particular optical kernel leads to a new generalization of the optical flow equation. Section 3 describes algorithms for solving for the refractive structure using this equation. Section 4 demonstrates the performance of our algorithm on synthetic and real data, and its application to matting. We conclude in Section 5 with a discussion and directions for future work.

2 A Theory of Refractive Optical Flow

In this section we describe our image formation model and use it to derive the refractive optical flow equation. As illustrated in Figure 1(a), we assume that the scene consists of three entities.

1. A background image plane B, where $B(\mathbf{u}, t)$ denotes the emitted radiance at the point $\mathbf{u} = (u, v)^\top$ and time t.
2. A foreground image plane I, where $I(\mathbf{x}, t)$ denotes the incident radiance at the point $\mathbf{x} = (x, y)^\top$ and time t.
3. A collection of one or more refracting objects between I and B.

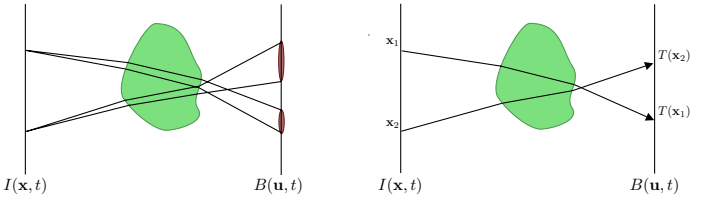


Fig. 1. The image formation model: (a) The general image formation model, where the incident irradiance at a point $\mathbf{x}$ in the foreground plane I is a result of the emitted radiance of some number of points in the background plane B. (b) The single ray image formation model, where the incident irradiance at $\mathbf{x}$ is linear function of the emitted radiance at the point $T(\mathbf{x})$ in the background plane.

The illumination in the scene is assumed to be ambient or coming from a directional light source, and the background plane is assumed to be an isotropic radiator. The background and the foreground planes are not constrained to be parallel. No assumptions are made about the shape of the refracting objects, or optical properties of their constituent materials. We treat the refracting objects as a "black-box" function that warps and attenuates the light rays passing through it. It is our aim to recover this function from a small set of images of the scene. Given the assumptions about scene illumination stated earlier, the incident radiance at a point $\mathbf{x}$ in the foreground is a result of the reflected and emitted radiance from the background plane. Now let the function that indicates what fraction of the light intensity observed at a point $\mathbf{x}$ in the foreground comes from the point $\mathbf{u}$ in the background be denoted by the optical kernel $\mathcal{K}(\mathbf{u},\mathbf{x})$. The total light intensity at $\mathbf{x}$ can now be expressed as an integral over the entire background plane,

$$I(\mathbf{x},t) = \int \mathcal{K}(\mathbf{u},\mathbf{x}) B(\mathbf{u},t) d\mathbf{u} \tag{1}$$

The problem of recovering the refractive structure of an object can now be restated as the problem of estimating the optical kernel associated with it.

The set of all functions of the form $\mathcal{K}(\mathbf{u},\mathbf{x})$ is huge. A large part of this set consists of functions that violate laws of physics. However the set of physically plausible optical kernels is still very big, and the reconstruction of $\mathcal{K}$ using a small number of images is an ill-posed problem. Additional constraints must be used to make the problem tractable.

One such direction of inquiry is to assume a low dimensional form for $\mathcal{K}$ by a small set of parameters. Zongker et al. in their work on environment matting, assume a parametric box form for $\mathcal{K}(\mathbf{u},\mathbf{x})$,

$$\mathcal{K}(\mathbf{u},\mathbf{x}) = \begin{cases} 1/\mu(\mathbf{x}) & \text{if } a(\mathbf{x}) \le u \le b(\mathbf{x}) \,\&\, c(\mathbf{x}) \le v \le d(\mathbf{x}) \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where, $a(\mathbf{x}), b(\mathbf{x}), c(\mathbf{x}), d(\mathbf{x})$ are functions of $\mathbf{x}$. $\mu(\mathbf{x})$ is the area of the rectangle enclosed by (a, c) and (b, d). The kernel maps the average intensity of a rectangular region in the background to a point in the foreground. The values of the parameters $\{a(\mathbf{x}), b(\mathbf{x}), c(\mathbf{x}), d(\mathbf{x})\}$ for each point $\mathbf{x}$ are determined using a set of calibrated background patterns and performing a combinatorial search on them so as to minimize the reconstruction error of the foreground image. Chuang et al. [5] generalize $\mathcal{K}$ to be from the class of oriented two-dimensional Gaussians. In both of these cases, knowledge of the background image is assumed.

In this paper we choose to pursue an alternate direction. We consider optical kernels of the form

$$\mathcal{K}(\mathbf{u}, \mathbf{x}) = \alpha(\mathbf{x})\delta(\mathbf{u} - T(\mathbf{x})) \tag{3}$$

where $\delta(\cdot)$ is Dirac's delta function, $T(\mathbf{x})$ is a piecewise differentiable function that serves as the parameter for the kernel indicating the position in the background plane where the kernel is placed when calculating the brightness at $\mathbf{x}$ in the foreground image plane. The function $\alpha(\mathbf{x})$ is a positive scalar function that accounts for the attenuation of light reaching $\mathbf{x}$. Figure 1(b) illustrates the setup.

In the following we will show how if we restrict ourselves to this class of optical kernels, we can recover the refractive structure without any knowledge of the background plane. We will also demonstrate with experiments how this subclass of kernels despite having a very simple description is capable of capturing refraction through a variety of objects.

The image formation equation can now be re-written as

$$I(\mathbf{x}, t) = \int \alpha(\mathbf{x})\delta(\mathbf{u} - T(\mathbf{x}))B(\mathbf{u}, t)d\mathbf{u} \tag{4}$$

$$= \alpha(\mathbf{x})B(T(\mathbf{x}), t) \tag{5}$$

We begin by differentiating the above equation w.r.t $\mathbf{x}$, to get

$$\nabla_{\mathbf{x}} I(\mathbf{x}, t) = (\nabla_{\mathbf{x}}\alpha(\mathbf{x}))B(T(\mathbf{x})) + \alpha(\mathbf{x})\mathbf{J}^{\top}(T(\mathbf{x}))(\nabla_{T(\mathbf{x})}B(T(\mathbf{x}), t)) \tag{6}$$

Here, $\mathbf{J}^{\top}(T(\mathbf{x}))$ is the transpose of the Jacobian of the transformation $T(\mathbf{x})$. Using Equation (5) the above equation can be written as

$$\nabla_{\mathbf{x}} I(\mathbf{x}, t) = I(\mathbf{x}, t)\frac{\nabla_{\mathbf{x}}\alpha(\mathbf{x})}{\alpha(\mathbf{x})} + \mathbf{J}^{\top}(T(\mathbf{x}))\left[\alpha(\mathbf{x})\nabla_{T(\mathbf{x})}B(T(\mathbf{x}))\right]$$

$$\mathbf{J}^{-\top}(T(\mathbf{x}))\left[\nabla_{\mathbf{x}} I(\mathbf{x}, t) - I(\mathbf{x}, t)\frac{\nabla_{\mathbf{x}}\alpha(\mathbf{x})}{\alpha(\mathbf{x})}\right] = \alpha(\mathbf{x})\nabla_{T(\mathbf{x})}B(T(\mathbf{x})) \tag{7}$$

Taking temporal derivatives of Equation (5) gives

$$I_t(\mathbf{x}, t) = \alpha(\mathbf{x})B_t(T(\mathbf{x}), t) \tag{8}$$

Now, let $\mathbf{c}(\mathbf{u}, t)$ denote the velocity of the point $\mathbf{u}$ at time t in the background plane. Then from the Brightness Constancy Equation [9] we know

$$\mathbf{c}(\mathbf{u}, t)^{\top}\nabla_{\mathbf{u}}B(\mathbf{u}, t) + B_t(\mathbf{u}, t) = 0 \tag{9}$$

Now assuming that the background image undergoes in-plane translation with velocity $\mathbf{c}(\mathbf{u},t) = \mathbf{c}(T(\mathbf{x}),t)$, we take the dot product of Equation (7) with $\mathbf{c}(T(\mathbf{x}),t)$ and add it to Equation (8) to get

$$\mathbf{c}(T(\mathbf{x}),t)^\top \mathbf{J}^{-\top}(T(\mathbf{x}))\left[\nabla_{\mathbf{x}} I(\mathbf{x},t) - I(\mathbf{x},t)\frac{\nabla_{\mathbf{x}}\alpha(\mathbf{x})}{\alpha(\mathbf{x})}\right] + I_t(\mathbf{x},t) = \alpha(\mathbf{x})\left[\mathbf{c}(T(\mathbf{x}),t)^\top \nabla_{T(\mathbf{x})} B(T(\mathbf{x})) + B_t(T(\mathbf{x}))\right] \tag{10}$$

From Equation (9) we know that the right hand side goes to zero everywhere, giving us

$$\mathbf{c}(T(\mathbf{x}),t)^\top \mathbf{J}^{-\top}(T(\mathbf{x}))\left[\nabla_{\mathbf{x}} I(\mathbf{x},t) - I(\mathbf{x},t)\frac{\nabla_{\mathbf{x}}\alpha(\mathbf{x})}{\alpha(\mathbf{x})}\right] + I_t(\mathbf{x},t) = 0 \tag{11}$$

Now for simplification's sake, let $\beta(\mathbf{x}) = \log\alpha(\mathbf{x})$. Dropping the subscript on ∇ we get

$$\boxed{\mathbf{c}(T(\mathbf{x}),t)^\top \mathbf{J}^{-\top}(T(\mathbf{x}))\left[\nabla I(\mathbf{x},t) - I(\mathbf{x},t)\nabla\beta(\mathbf{x})\right] + I_t(\mathbf{x},t) = 0} \tag{12}$$

This is the refractive optical flow equation.

2.1 Properties of the Refractive Optical Flow Equation

Before we dive into the solution of Equation (12) for recovering the refractive structure, we comment on its form and some of its properties. The first observation is that if there is no distortion or attenuation, i.e.,

$$T(\mathbf{x}) = \mathbf{x}, \qquad \alpha(\mathbf{x}) = 1$$

the Jacobian of T reduces to the identity, and the gradient of $\beta(\mathbf{x})$ reduces to zero everywhere, giving us the familiar optical flow equation in the foreground image.

$$\mathbf{c}(\mathbf{x},t)^\top \nabla I(\mathbf{x},t) + I_t(\mathbf{x},t) = 0 \tag{13}$$

The second observation is that the equation is independent of $B(\mathbf{u},t)$. This means that we can solve for the refractive flow through an object just by observing a distorted version of the background image. Knowledge of the background image itself is not needed.

Finally we observe that Equation (12) is in terms of the Jacobian of T, i.e. any function $T'(\mathbf{x}) = T(\mathbf{x}) + \mathbf{u}_0$ will result in the same equation. This implies that T can only be solved up to a translation ambiguity. Visually this is equivalent to viewing the scene through a periscope. The visual system has no way of discerning whether or not an image was taken through a periscope. A second ambiguity is introduced into the solution when we note that the velocity $\mathbf{c}(\mathbf{u},t)$ is in the background plane and there is nothing that constrains the two coordinate systems from having different scales along each axis. Hence $T(\mathbf{x})$ can only be recovered up to a four parameter family of ambiguities corresponding to translation and scaling. The attenuation function $\beta(\mathbf{x})$ is not affected by the scaling in $\mathbf{c}(\mathbf{u},t)$ and hence is recovered up to a translation factor which in turn means that $\alpha(\mathbf{x})$ is recovered up to a scale factor.

3 Solving the Equation of Refractive Flow

In this section, we further analyze Equation (12) for the purposes of solving it. We begin by considering a further simplification of Equation (12). We assume that the background plane translates with in-plane velocity $\mathbf{c}(\mathbf{u}, t) = \mathbf{c}(t)$ that is constant over the entire background plane. We consider two cases, the calibrated case (when the motion of the background $\mathbf{c}(t)$ is known) and the uncalibrated case (when the motion of the background is unknown). In each case we describe methods for recovering $T(\mathbf{x})$ and $\alpha(\mathbf{x})$, and note the ambiguities in the resulting solution. Let $T(\mathbf{x})$ be denoted by

$$T(x, y) = (g(x, y), h(x, y))^\top \tag{14}$$

The Jacobian of T and its inverse transpose are

$$\mathbf{J}(T(\mathbf{x})) = \begin{bmatrix} g_x & g_y \\ h_x & h_y \end{bmatrix} \qquad \mathbf{J}^{-\top}(T(\mathbf{x})) = \frac{1}{g_x h_y - g_y h_x} \begin{bmatrix} h_y & -h_x \\ -g_y & g_x \end{bmatrix}$$

The translation velocity in the background plane is $\mathbf{c}(t) = (\xi(t), \eta(t))^\top$. Substituting these in Equation (12) we get

$$\begin{bmatrix} \xi & \eta \end{bmatrix} \frac{1}{g_x h_y - h_x g_y} \begin{bmatrix} h_y & -h_x \\ -g_y & g_x \end{bmatrix} \begin{bmatrix} I_x - \beta_x I \\ I_y - \beta_y I \end{bmatrix} + I_t = 0 \tag{15}$$

which rearranges to

$$\frac{\eta I_y g_x - \eta I_x g_y - \xi I_y h_x + \xi I_x h_y + \eta I(g_y \beta_x - g_x \beta_y) - \xi I(h_y \beta_x - h_x \beta_y)}{g_x h_y - g_y h_x} + I_t = 0 \tag{16}$$

Now let

$$p = \frac{g_x}{g_x h_y - g_y h_x} \qquad q = \frac{g_y}{g_x h_y - g_y h_x} \qquad a = \frac{g_y \beta_x - g_x \beta_y}{g_x h_y - g_y h_x} \tag{17}$$

$$r = \frac{h_x}{g_x h_y - g_y h_x} \qquad s = \frac{h_y}{g_x h_y - g_y h_x} \qquad b = \frac{h_y \beta_x - h_x \beta_y}{g_x h_y - g_y h_x} \tag{18}$$

we get

$$\eta I_y p - \eta I_x q - \xi I_y r + \xi I_x s + \eta I a - \xi I b + I_t = 0 \tag{19}$$

3.1 The Calibrated Case

In the case where the velocity of the background plane $\mathbf{c}(\mathbf{u}, t)$ is known, Equation (19) is linear in p, q, r, s, a, b. Given 7 or more successive frames of a video or 14 or more pairs of successive frames and the associated motion of the background plane, we can solve the equation point-wise over the entire image. Given

$n+1$ successive frames, we get n equations in 6 variables at each point in the foreground plane, giving us an over-constrained linear system of the form

$$\mathbf{A}\mathbf{p} = \mathbf{m} \tag{20}$$

here the i^{th} row of the matrix $\mathbf{A}$ is given by

$$\begin{aligned} \mathbf{A_i} &= \begin{bmatrix} \eta(i)I_y(i) & -\eta(i)I_x(i) & -\xi(i)I_y(i) & \xi(i)I_x(i) & \eta(i)I(i) & -\xi(i)I(i) \end{bmatrix} \\ \mathbf{m} &= \left[-I_t(1), \ldots, -I_t(n)\right]^\top \qquad \mathbf{p} = \begin{bmatrix} p & q & r & s & a & b \end{bmatrix}^\top \end{aligned} \tag{21}$$

The above system of equations is solved simply by the method of linear least squares. The corresponding values of $g_x, g_y, h_x, h_x, \beta_x, \beta_y$ can then be obtained as follows.

$$g_x = \frac{p}{ps - qr} \qquad g_y = \frac{q}{ps - qr} \qquad \beta_x = \frac{bp - ar}{ps - qr} \tag{22}$$

$$h_x = \frac{r}{ps - qr} \qquad h_y = \frac{s}{ps - qr} \qquad \beta_y = \frac{bq - as}{ps - qr} \tag{23}$$

3.2 The Uncalibrated Case

If the motion of the background image is not known, Equation (19) is bilinear in the variables p, q, r, s, a, b and $\xi(t), \eta(t)$. If we consider frames $i = 1, \ldots, n+1$ and pixels $j = 1, \ldots, m$ in the foreground image, then we can rewrite Equation (19) in the following form

$$\mathbf{c}_i^\top \mathbf{A}_{ij} \mathbf{p}_j = 1 \qquad i = 1, \ldots, n \quad j = 1, \ldots, m \tag{24}$$

where

$$\begin{aligned} \mathbf{c}_i &= \begin{bmatrix} \xi(i) & \eta(i) \end{bmatrix}^\top \qquad \mathbf{p}_j = \begin{bmatrix} p(j) & q(j) & r(j) & s(j) & a(j) & b(j) \end{bmatrix}^\top \\ \mathbf{A}_{ij} &= \frac{-1}{I_t(i,j)} \begin{bmatrix} 0 & 0 & -I_y(i,j) & I_x(i,j) & 0 & -I(i,j) \\ I_y(i,j) & -I_x(i,j) & 0 & 0 & I(i,j) & 0 \end{bmatrix} \end{aligned}$$

This gives us nm equations in $2n+6m$ variables, and we can solve them whenever $nm > 2n + 6m$.

Equation (24) is a system of bilinear equations, i.e. given $\mathbf{c}_i$, the system reduces to a linear system in $\mathbf{p}_j$ and vice versa. The overall problem however is highly non-linear. Using this observation a simple alternating procedure for solving the system can be used, which starting with a random initialization, alternates between updating $\mathbf{c}_i$ and $\mathbf{p}_j$ using linear least squares. Even with the fast iterative procedure, solving for the structure in this manner is not feasible. Hence we use a small slice through image stack and use it to solve for $\mathbf{c}_j$, which is then used as input to the calibrated algorithm described in the previous section to recover the flow over the entire foreground plane.

Properties of the solution. Observe that $\mathbf{Q}\mathbf{A}_{ij}\mathbf{R} = \mathbf{A}_{ij} \quad \forall i, j$ where

$$\mathbf{Q} = \begin{bmatrix} q_{12} & q_{12} \\ q_{21} & q_{22} \end{bmatrix} \quad \mathbf{R} = \frac{1}{q_{11}q_{22} - q_{12}q_{21}} \begin{bmatrix} q_{11} & 0 & q_{21} & 0 & 0 & 0 \\ 0 & q_{11} & 0 & q_{21} & 0 & 0 \\ q_{12} & 0 & q_{22} & 0 & 0 & 0 \\ 0 & q_{12} & 0 & q_{22} & 0 & 0 \\ 0 & 0 & 0 & 0 & q_{11} & q_{21} \\ 0 & 0 & 0 & 0 & q_{12} & q_{22} \end{bmatrix} \tag{25}$$

Hence, not knowing $\mathbf{c}(\mathbf{u}, t)$ gives rise to a solution which is ambiguous up to a 2×2 invertible transformation, and a corresponding ambiguity in the values of the various gradient estimates. Coupled with this is the translation ambiguity in g and h. This gives rise to a six parameter family of ambiguities in the overall solution.

3.3 Integration

The methods described in the previous two sections give us estimates of the partial derivatives of $g(x, y)$, $h(x, y)$ and $\beta(x, y)$. The final step is integrating the respective gradient fields to recover the actual values of the functions. Reconstruction of a function from its first partial derivatives is a widely studied problem. We use an iterative least squares optimization that minimizes the reconstruction error [10].

4 Experiments

All experiments were done with video sequences of 200 frames each. The synthetic data were generated using a combination of MATLAB and POV-Ray, a public domain ray tracer. The real data was captured by placing a refracting object in front of an LCD screen, and imaging the setup using a firewire camera. Figure 2 illustrates the data acquisition setup. Calculation of image derivatives is

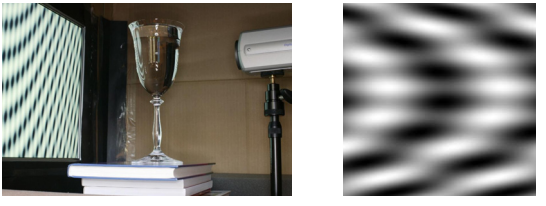


Fig. 2. (a) Shown above is a photograph of the data acquisition system used in our experiments. It consists of Samsung 19" LCD display screen on the left, a Sony DFW-VL500 firewire camera on the right and a refracting object between the screen and the camera. (b) shows a frame from the background image sequence for all our experiments.

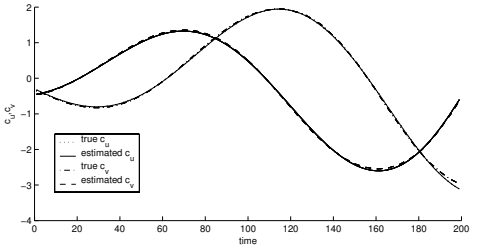


Fig. 3. Estimation of background motion in the uncalibrated case. This figure plots the true and estimated motion in the background plane. The two curves show motion along the $\xi(t)$ and $\eta(t)$ along the u and v axis respectively. As can be seen there is virtually no difference between the true and estimated values for $\xi(t)$ and $\eta(t)$.

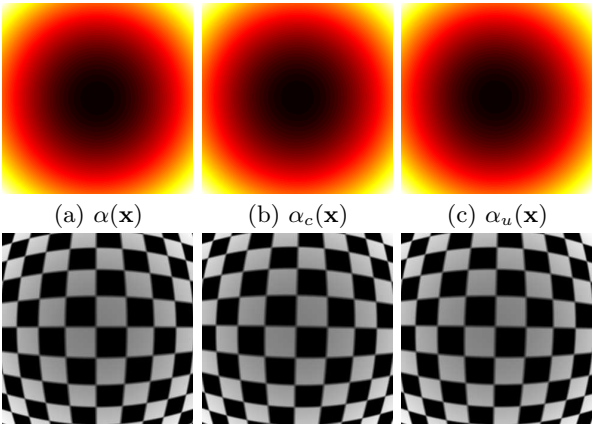
very sensitive to noise. Smoothing was performed on the images using anisotropic diffusion [11], which has superior behavior along sharp edges as compared to Gaussian filtering. This is important for objects that cause light rays to be inverted, which in turn causes the optical flow across the boundary to be opposite sign; a naive Gaussian based smoothing procedure will result in significant loss of signal. The least squares estimation step in the calibrated estimation algorithm was made robust by only considering equations for which the temporal gradient term I_t was within 85% of the maximum temporal gradient at that pixel over time. This choice results in only those constraints being active where some optical flow can be observed.

The boundary of refracting objects typically have little or no optical flow visible. This results in the refractive optical flow constraint breaking down along the boundary as well as certain medium interfaces. We mask these pixels out by considering the average absolute temporal gradient at each point in the foreground image plane and ignoring those pixels that fall below a chosen threshold. This results in a decomposition of the image into a number of connected components. All subsequent calculations are carried out separately for each connected component.

4.1 Results

We begin by considering a synthetic example to illustrate the performance of our algorithm in the calibrated and the uncalibrated case. The warping function used was $T(x, y) = (xe^{-(x^2+y^2)}, ye^{-(x^2+y^2)})^\top$ and the attenuation factor was $\alpha(x, y) = e^{-(x^2+y^2)}$. Figure 3 shows a comparison between the estimated and the true motion in the uncalibrated case. Figure 4 illustrates the results of the experiment. The estimated warp and attenuation functions are virtually identical to the true warp and attenuation functions. In the uncalibrated case, the estimated motion was disambiguated by doing a least squares fit to the ground truth for the purposes of illustration.

Figure 5 illustrates the result of applying the refractive structure of a glass with and without water inside it to the task of environment matting. Note that the region along the rim of the glass is missing. This is more so in case of the glass filled with water. These are the regions where the refractive optical flow equation breaks down. The black band in the case of the filled glass is due to a combination of the breakdown of the refractive optical flow equation along the air/water/glass interface and the finite vertical extent of the image. More results can be seen at `http://vision.ucsd.edu/papers/rof/`.



(a) $\alpha(\mathbf{x})$ (b) $\alpha_c(\mathbf{x})$ (c) $\alpha_u(\mathbf{x})$

(d) $T(\mathbf{x})$ (e) $T_c(\mathbf{x})$ (f) $T_u(\mathbf{x})$

Fig. 4. This figure shows compares the performance of refractive structure estimation in the calibrated and the uncalibrated case. (a) and (d) show the true warp $T(\mathbf{x})$ applied to a checkerboard pattern and the attenuation factor $\alpha(\mathbf{x})$. (b) and (e) show the estimated warp and the alpha for the calibrated case, and (c) and (f) show the estimated warp and alpha for the uncalibrated case.

5 Discussion

We have introduced a generalization of the optical flow constraint, described methods for solving for the refractive structure of objects in the scene, and shown that this can be readily computed from images. We now comment on the limitations of our work and directions of future work.

First, our method does not address Fresnel and total internal reflection. This places a limit on our analysis to the case where all the illumination comes from behind the object being observed. Methods for recovering surface geometry from specular reflections are an active area of research and are better suited for this task [12,13,14].

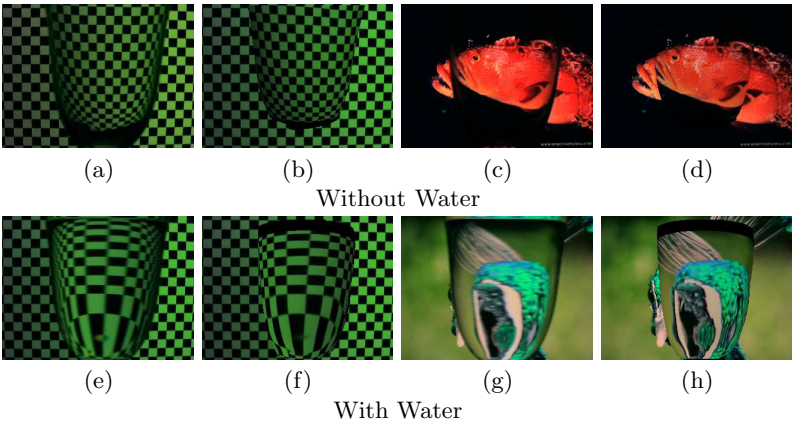


Fig. 5. Results of using the refractive structure for environment matting. (a), (c) show the true warping of a background image when an empty glass is placed in front of it, (b) and (d) show the estimated refractive structure applied to the same images. (e) and (g) show the true warping of a background image when a glass filled with water is placed in front of it, (f) and (h) show the estimated refractive structure applied to the same images.

Second, the presented approach, like [5,7], formulates the objective as determining a plane-to-plane mapping, i.e., it only informs us about how the object distorts a plane in 3-space. A more satisfactory solution will be to recover how the object distorts arbitrary light rays. We believe this is solvable using two or more views of the background plane and is the subject of our ongoing work.

The choice of the optical kernel that resulted in the single ray model was made for reasons of tractability. This however leaves the possibility of other optical kernels that account for multiple ray models.

Our experiments were carried out using a 19 inch LCD screen as the background plane. For most refracting objects, as their surface curvature increases, the area of the background plane that they project onto a small region in the foreground image can increase rapidly. If one continues to use a flat background one would ideally require an infinite background plane to be able to capture all the optical flow. Obviously that is not practical. An alternate approach is to use a curved surface as the background, perhaps a mirror which reflects an image projected onto it.

The current work only considers gray scale images; extensions to color or multi-spectral images is straightforward. There are two cases here: if the distortion T is assumed to be the same across spectral bands, the generalization can be obtained by modeling $\alpha(\mathbf{x})$ as a vector valued function that accounts for attenuation in each spectral band. In case T is dependent on the wavelength of light, each spectral band results is an independent version of Equation (12) and can be solved using the methods described.

Finally, the refractive optical flow equation is a very general equation describing optical flow through a distortion function. This allows us to address distortion not due to transmission through transparent objects, but also due to reflection from non-planar mirrored and specular surfaces. We believe that the problem of specular surface geometry can be addressed using this formalism, and this is also a subject of our future work.

Acknowledgements. The authors would like to thank Josh Wills and Kristin Branson for helpful discussions. This work was partially supported under the auspices of the U.S. DOE by the LLNL under contract No. W-7405-ENG-48 to S.B. and National Science Foundation grant IIS-0308185 to D.K.

References

1. Irani, M., Rousso, B., Peleg, S.: Computing occluding and transparent motion. International Journal of Computer Vision **12** (1994) 5–16
2. Ju, S.X., Black, M.J., Jepson, A.D.: Skin and bones: Multilayer, locally affine, optical flow and regularization with transparency. In: CVPR 1996. (1996) 307–314
3. Levin, A., Zomet, A., Weiss, Y.: Learning to perceive transparency from the statistics of natural scenes. In: Neural Information Processing Systems, 2002. (2002)
4. Schechner, Y., Kiryati, N., Shamir, J.: Blind recovery of transparent and semireflected scenes. In: CVPR 2000. Volume 1., IEEE Computer Society (2000) 38–43
5. Chuang, Y.Y., Zongker, D.E., Hindorff, J., Curless, B., Salesin, D.H., Szeliski, R.: Environment matting extensions: Towards higher accuracy and real-time capture. In: Proc. of ACM SIGGRAPH 2000, ACM Press (2000) 121–130
6. Wexler, Y., Fitzgibbon, A.W., Zisserman, A.: Image-based environment matting. In: Proc. of the 13th Eurographics workshop on Rendering. (2002)
7. Zongker, D.E., Werner, D.M., Curless, B., Salesin, D.H.: Environment matting and compositing. In: Proc. of ACM SIGGRAPH 99, ACM Press (1999) 205–214
8. Murase, H.: Surface shape reconstruction of a nonrigid transparent object using refraction and motion. IEEE. Trans. on Pattern Analysis and Machine Intelligence **14** (1992) 1045–1052
9. Horn, B.K.P., Schunck, B.G.: Determining optical flow. Artificial Intelligence **17** (1981)
10. Horn, B.: Robot Vision. MIT Press (1986)
11. Perona, P., Malik, J., Shiota, T.: Anisotropic Diffusion. In: Geometry-Driven Diffusion in Computer Vision. Kluwer, Amsterdam (1995) 73–92
12. Savarese, S., Perona, P.: Local analysis for 3D reconstruction of specular surfaces. In: CVPR 2001. (2001) 738–745
13. Savarese, S., Perona, P.: Local analysis for 3D reconstruction of specular surfaces: Part II. In: ECCV 2002. (2002)
14. Oren, M., Nayar, S.: A theory of specular surface geometry. IJCV **24** (1997) 105–124