

# Matting and Compositing of Transparent and Refractive Objects

SAI-KIT YEUNG

University of California, Los Angeles

CHI-KEUNG TANG

The Hong Kong University of Science and Technology

MICHAEL S. BROWN

National University of Singapore

and SING BING KANG

Microsoft Research Redmond

This paper introduces a new approach for matting and compositing transparent and refractive objects in photographs. The key to our work is an image-based matting model, termed the *attenuation-refraction matte* (ARM), that encodes plausible refractive properties of a transparent object along with its observed specularities and transmissive properties. We show that an object's ARM can be extracted directly from a photograph using simple user markup. Once extracted, the ARM is used to paste the object onto a new background with a variety of effects, including compound compositing, Fresnel effect, scene depth, and even caustic shadows. User studies find our results favorable to those obtained with Photoshop as well as perceptually valid in most cases. Our approach allows photo-editing of transparent and refractive objects in a manner that produces realistic effects previously only possible via 3D models or environment matting.

Categories and Subject Descriptors: I.3.3 [Computer Graphics]: Picture/Image Generation; I.4.8 [Image Processing And Computer Vision]: Scene Analysis; J.5 [Computer Applications]: Arts And Humanities

General Terms: Image Matting and Compositing

Additional Key Words and Phrases: Transparent objects

## ACM Reference Format:

Yeung, S.-K., Tang, C.-K., Brown, Michael S., Kang, S. B. 2010. Matting and Compositing of Transparent and Refractive Objects. To appear in ACM Trans. Graph., 13 pages.

---

Authors' addresses: land and/or email addresses.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or [permissions@acm.org](mailto:permissions@acm.org).

© 2010 ACM 0730-0301/2010/10-ARTN \$10.00

DOI 10.1145/1559755.1559763

<http://doi.acm.org/10.1145/1559755.1559763>

## 1. INTRODUCTION

The ability to cut and paste objects in photographs is a prerequisite for photo editing. While many effective approaches for segmenting and matting objects from a single image exist [Li et al. 2004; Rother et al. 2004; Wang and Cohen 2008], these approaches assume the foreground objects are opaque. In many of these approaches, a user marks a trimap that consists of definite foreground, definite background, and uncertain region. The opacity assumption simplifies the matting problem in the uncertain regions by reducing the recovery of the foreground object to an estimation of each pixel's fractional contribution of its color to the foreground (with the remainder being the background).

Transparent and refractive objects, on the other hand, have three properties that complicates their extraction and pasting. First, the fractional alpha associated with the transparent object is distributed over the *whole* object, as opposed to opaque objects where the fractional alpha values are mostly at the boundaries. Second, transparent objects commonly exhibit light attenuation through the object that affects each color channel differently. As a result, the conventional matting equation involving only a single scalar per pixel is insufficient. Finally, the refractive nature of these objects results in a warped appearance of the background. Since the object's 3D shape is typically unknown, this warping function is also unknown. To produce realistic composites, the refractive and transparent properties of these objects must be taken into consideration.

This paper describes a new approach for matting transparent and refractive objects from a photograph and compositing the extracted object into a new scene. To accomplish this task we have modified the opaque image matting and compositing equation to fuse refractive deformation, color attenuation, and foreground estimation. We term this extracted information the *attenuation-refraction matte* (ARM). In general, a single photograph is insufficient to extract accurate refractive properties of a transparent object. Our approach instead recovers *plausible* light-transport properties of the matted object, exploiting our visual system's tolerance to inaccuracies in refractive phenomena as previously demonstrated by work targeting image-based material editing [Khan et al. 2006].

We show how user markup can be used to extract each component of the ARM. By employing the unique properties of an object's ARM, the specular and attenuation components of the ARM can be optimized efficiently. Plausible refractive deformations are specified via control lines that are drawn on the image. One benefit of targeting *plausible* versus *accurate* light refraction is that our refractive markup need not be accurate.

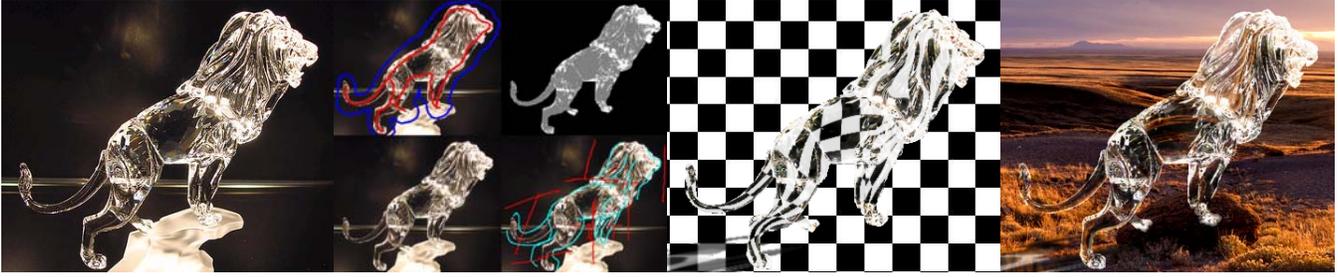


Fig. 1. **Attenuation-Refraction Matte (ARM)**. From left to right: input image, casual markup, ARM, and the composite completed with simulated Fresnel effect and caustic shadow.

Once the ARM has been extracted, we show how to produce a range of effects when compositing the object into a new background. These effects include a Fresnel effect to enhance boundary realism, scene depth, multiple-object pasting, and the simulation of caustic shadows. To our knowledge, this is the first approach to address the matting of transparent objects and their plausible refractive properties directly from a photograph. Figure 1 shows an example of a glass lion’s ARM and a composited result. Figure 2 shows a few examples that demonstrate compound compositing by overlapping multiple ARMs of the transparent and refractive objects we extracted. In both applications, the ARM is extracted from a single input image and no 3D models are used in the composited result.

Because no comparable systems are able to provide similar matting and compositing capabilities, we perform comparisons between transparent objects transferred using Photoshop and those using the ARM representation. To evaluate the quality of ARM composites compared with real images of the same transparent objects, we performed a user study on different object and background scenarios. Our results show that object extraction using ARM is faster than Photoshop and visually more favorable. In addition, our composite results are deemed perceptually valid in most cases when compared with the ground truth.

## 2. RELATED WORK

The basic challenge for opaque object image matting is to minimize user involvement without sacrificing output quality. Several techniques have been proposed and a survey can be found in [Wang and Cohen 2008]. Our task shares similar goals but targets the matting of specularities and foreground pixels, with the vast majority of the object being either transparent or attenuated background. In addition, the need for color attenuation is not considered in opaque matting.

The process of characterizing the light transport properties of transparent and refractive objects has been referred to as *environment matting*. Techniques for environment matting use either objects with known backgrounds [Zongker et al. 1999; Chuang et al. 2000; Peers and Dutre 2003], or multiple images [Wexler et al. 2002] to extract the light transport properties. For example, Matusik et al. [2002] used a turntable, multiple cameras, multiple lights, and monitors (as backgrounds) to capture an environment matte all around a transparent object. Our approach differs in that we target matting directly from an input photograph and use ‘plausible light transport’ to produce visually similar results.

Khan et al. [2006] demonstrated an effective material editing technique that approximated light transport through coarse 3D object

reconstruction. Using the object’s approximated 3D shape, the object’s material properties were changed, including the simulation of transparent and refractive effects. This technique, however, requires the original object to be opaque in order to approximate the 3D shape. Furthermore, material editing is applied on the original image itself; matting and compositing are not performed.

If the shape of the transparent object is known, rendering the object with refractive effects in a given scene is well studied [Foley et al. 1995]. However, obtaining a real object’s 3D shape and corresponding refractive properties is not a trivial task. Recent techniques for shape recovery of transparent objects are available, e.g., [Ben Ezra and Nayar 2003; Miyazaki and Ikeuchi 2007; Morris and Kutulakos 2007]. In addition, refractive deformation can be computed from a video [Agarwal et al. 2004]. However, these existing approaches operate under specific conditions, such as multi-camera capture, custom calibration, and often assume restrictions in 3D shape. This makes them inapplicable in our tool that is designed to operate entirely on photographs with no explicit 3D information.

Given a matted opaque object and a new background, multiple approaches exist for minimizing the object-background seam when compositing the object (e.g., [Pérez et al. 2003; Jia et al. 2006]). As previously mentioned, these existing matting and compositing approaches assume an opaque object with fractional pixel contributions at the object boundary. Compositing in our case requires more consideration due to the refractive and transparent nature of the object.

To produce a compelling composite the boundary of the matted object must be enhanced with the Fresnel effect. While this effect is inherently captured by environment matting approaches, in our case it needs to be imitated. In addition, compositing a transparent object without its caustic shadow will make the result look unrealistic. Traditional shadow matting techniques are not applicable here because a video is required [Chuang et al. 2003], or the single-image formulation does not account for caustic shadows [Wu et al. 2007]. In [Gutierrez et al. 2008], caustic shadows are simulated from a single image by detecting phase symmetry on the 3D depth map recovered from the image using the “dark is deep” assumption, which is not applicable to transparent objects.

## 3. ATTENUATION-REFRACTION MATTE (ARM)

We begin by defining a new matting and compositing equation that accounts for the appearance of the transparent and refractive object observed from an image. We start with a light scattering model similar to that proposed in [Matusik et al. 2002]. Assuming a distant



Fig. 2. Examples on compound compositing using multiple objects. All results are *entirely image-based*; in our photo-editing application, no 3D models are available. Input images for ARM extraction for the above and other examples are shown on the bottom row.

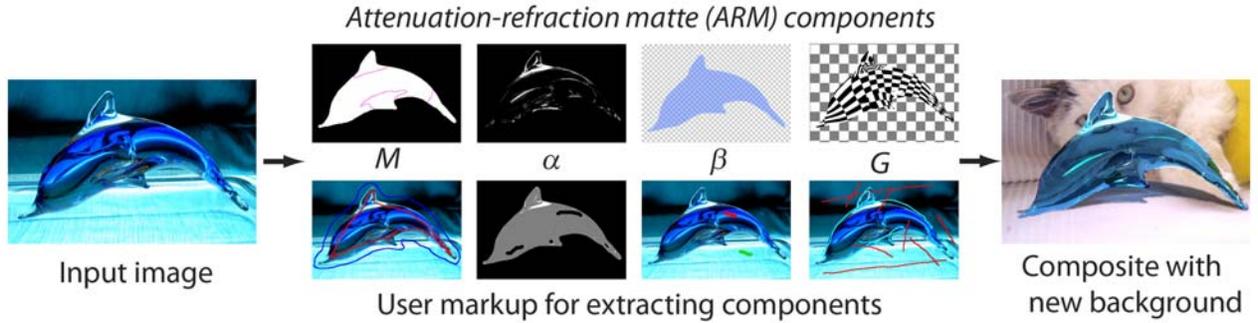


Fig. 3. Overview of attenuation-refraction matte (ARM) extraction.

scene, the amount of light recorded at the camera is

$$C = \int_{\Omega} W(\omega)L(\omega)d\omega, \quad (1)$$

where  $L(\omega)$  is the illumination from direction  $\omega$ ,  $W(\omega)$  is the contributing weight, and  $\Omega$  is the entire hemisphere. Note that  $\int_{\Omega} W(\omega)d\omega < 1$  because of transmission loss or attenuation due to material absorption of light.

Consider a transparent object in the image bounded by a mask  $M$ . Within this object, we make the simplifying assumption this object  $C_M$  consists of two items:

$$C_M = \int_I W(\mathbf{p})L(\mathbf{p})d\mathbf{p} + \int_{\Omega-I} W(\omega)L(\omega)d\omega = C_I + C_S, \quad (2)$$

where  $\mathbf{p}$  is the continuous 2D coordinate of the image. More specifically, we assume that the warped color seen through the transparent object comes from locations within the camera's field of view  $I$ . This is represented by the term  $C_I$ . Next, we assume that the specularities  $S$  are caused by illumination not directly visible to the camera (i.e.,  $\Omega - I$ ). The specularities on the object are represented by the term  $C_S$ .

For an object in the discrete image, we decompose these terms as

$$C_S = \alpha S, \quad C_I = (1 - \alpha)\beta B_G, \quad (3)$$

where  $\alpha \in [0, 1]$  is the relative contribution of the appearance of specularities to the object's image (specular matte),  $S$  is the 3-channel specularity,  $\beta$  is 3-channel color transmission factor ( $< 1$  in each channel because of attenuation), and  $B_G$  is the warped background. In discrete form, we write the image formation equation as

$$C_M(\mathbf{x}) = \alpha(\mathbf{x})S(\mathbf{x}) + (1 - \alpha(\mathbf{x}))\beta(\mathbf{x})B(G(\mathbf{x})). \quad (4)$$

Here,  $\mathbf{x}$  is pixel location within object  $M$ ,  $B(\mathbf{x})$  the appearance of the background without the object, and  $G(\mathbf{x})$  the warping function. The terms  $M$ ,  $\alpha$ ,  $\beta$ , and  $G$ , collectively make up an object's *attenuation-refraction matte*, or ARM. For an object in the input image, the terms of ARM are all unknowns.

#### 4. ARM EXTRACTION

The ARM extraction from a *single* image of a transparent object is an ill-posed problem. Even without considering the non-linear deformation  $G$ , we have seven unknowns (one for  $\alpha$ , three for  $S$ , and three for  $\beta$ ) to solve in the equation. We therefore make the following simplifying assumptions on our ARM extraction:

- (1) Strong, white specular highlights are observed on the transparent object. This simplifies the computation of  $\alpha$ .

Table I. Summary of user interaction (number of strokes or stroke pairs marked) and processing time for extracting the ARMs shown in this paper.

	image size	$M$		$\alpha$		$\beta$		$G$		total time (sec)
		#strokes	time (sec)	#strokes	time (sec)	#pairs	time (sec)	#pairs	time (sec)	
<i>glass</i>	602 × 400	2	1	0	6.3	1	10	12	14	31.3
<i>jug</i>	183 × 286	3	1	2	5.1	2	14	6	6	26.1
<i>elephant</i>	495 × 295	5	2	5	18	1	17	16	18	55
<i>chandelier</i>	339 × 451	2	1	0	7.3	1	15	13	15	38.3
<i>dolphin</i>	463 × 306	2	1	5	11.5	1	46	8	8	66.5
<i>lion</i>	359 × 294	3	1	5	3.8	0	0	14	16	20.8
<i>fish</i>	416 × 457	1	1	16	22.7	1	24.8	9	10	58.5
<i>swan</i>	266 × 512	2	1	2	2.6	0	0	6	6	9.6
<i>dragon</i>	353 × 269	4	2	5	3.5	0	0	9	10	15.5
<i>globe</i>	308 × 213	4	0.5	0	3.2	0	0	2	1	4.7

It typically takes 1–5 seconds to add a stroke. Short times are sufficient for specifying an acceptable object’s context  $M$ . In general, for  $M$  and  $G$ , processing after each interaction (stroke or curve pair markup) takes less than 1 second. For  $\alpha$  and  $\beta$ , processing is done once after all user markups are made. The program is run on a 3.6GHz PC with 2G RAM. The paper’s supplemental materials show the actual interaction of some examples recorded in real time.

- (2) The transmission factor distribution is piecewise smoothly-varying within a refractive medium. This simplifies the computation of  $\beta$ .
- (3) The refractive deformation  $G$  need only be plausible.

These assumptions are also consistent with Eqn (3), where the  $C_S$  component incorporates highlights which are largely white and opaque, while the  $C_T$  component mostly accounts for the transparency observed. Moreover, some ARM components are independent of others. As a result, it is possible to simplify the complex extraction problem by breaking it into several steps while achieving high-quality results.

Figure 3 shows the overview of ARM extraction. Specifically, we first extract  $M$  which defines the processing region for the extraction of  $\alpha$ ,  $\beta$  and  $G$ . Since  $\alpha$  and  $\beta$  are related to pixel color whereas  $G$  concerns with pixel movement, we assume that the extraction of  $(\alpha, \beta)$  is independent of that of  $G$ . Table I summarizes the user interaction and the processing time for all the ARM examples in this paper. In the following, we discuss the extraction of each component.

#### 4.1 $M$ -extraction

$M$  specifies the “footprint” of the object in the image as a binary mask. We use lazy-snapping [Li et al. 2004] to extract  $M$  (other techniques such as grab-cut [Rother et al. 2004] may be used as well). The user draws scribbles on the inside and outside of the object, shown respectively in red and blue in Figure 3. Although the color samples inside and outside the object can be similar due to transparency, color inconsistency at the object border allows the technique to work.  $M$  can be further partitioned to mask out non-transparent regions, and separate different refractive mediums or deformation regions. It typically takes from seconds to two minutes to specify a region mask; the region mask corresponding to each example can be found in supplementary material.

#### 4.2 $(\alpha, \beta)$ -extraction

We extract  $\alpha$  and  $\beta$  within  $M$ . Eqn (3) shows that we have more unknowns than equations to solve.

However, note that  $S$  (specularities) is largely white and opaque, and  $\beta$  (attenuation) is largely transparent and homogeneous within each refractive medium.  $S$  contributes to the “foreground colors”

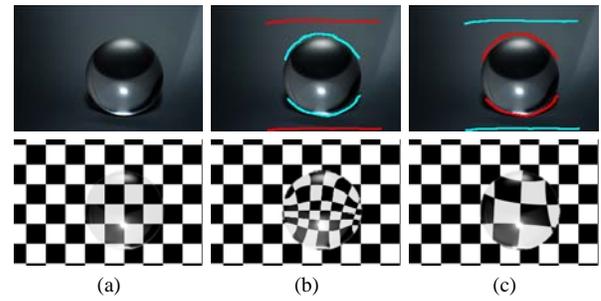


Fig. 4. Object cue. Three basic markups of refractive light-transport: (a) no markup, (b) markup for simulating light convergence, (c) markup for simulating light divergence.  $c_{\text{ref}}$  is in red and  $c_{\text{target}}$  is in cyan.  $c_{\text{target}}$  is where  $c_{\text{ref}}$  is perceived to distort to.

and  $\beta$  attenuates the “background colors” of the input image respectively. The extraction of  $\alpha$  and  $S$  is therefore relatively easier and largely insensitive to the extraction of  $\beta$ , which is smooth and transparent. The extraction of  $\beta$  can be improved after specular and opaque highlights are removed.

**Solving for  $\alpha$  and  $S$ .** To solve  $\alpha$  and  $S$ , a trimap is automatically generated.  $S$  is considered as “foreground” in the conventional matting problem and its contribution can be removed by multiplying it with the corresponding  $\alpha$ . Because the “definite foreground” consists of specular highlight, given the object mask  $M$ , definite foreground shown as white within  $M$  in Figure 3 can automatically be labeled by thresholding (we set the threshold as all the RGB values  $> 220$ , where input pixels range 0-255). “Definite background” is taken as pixel outside the mask (shown as black), with all remaining pixels within the mask labeled as “uncertain” (shown as gray). This automatic trimap construction is amenable to objects with a great deal of subtle structures that result in many small highlights. Additional strokes within  $M$  can be marked to specify definite background (black) or foreground (white) if necessary. Poisson matting [Sun et al. 2004] is performed using the trimap as input to extract the  $\alpha$ . Occasionally, strong highlights in the background are extracted in the  $\alpha$  component. Such unwanted artifacts usually come in blocks, and can be easily edited away. The mattes before and after editing of *swan*, *elephant*, *chandelier*, *jug* and *fish* are provided in the supplementary material, showing that the amount of additional editing of the specular mattes is acceptable.

**Solving for  $\beta$ .** To solve  $\beta$ ,  $S$ 's contribution will first be removed from the input image using  $\alpha$ . The user then marks up on the image to collect a number of attenuated and unattenuated background color samples, within and outside of the object's mask  $M$ . This is done using simple scribbles (red and green), as shown in Figure 3, where we assume that the scribbled background colors on the object are not severely affected by refractive deformation. To collect color statistics that are primarily affected by  $\beta$ -attenuation, the user should mark up sample pairs that are similar in texture. Here, the problem is translated into one similar to natural shadow matting [Wu et al. 2007], except that we do not need to handle hard boundaries.

Removing  $S$ 's contribution using  $\alpha$  leads to  $\bar{C}_I = (C_M - \alpha S)/(1 - \alpha) = \beta B$ . The term  $\bar{C}_I$  can be regarded as the color-attenuated version of the image background  $B$ . This can be solved using the following Bayesian optimization:

$$\beta^* = \arg \max_{\beta} P(\hat{B}|\beta) + P(\beta), \quad (5)$$

where  $P(\hat{B}|\beta)$  is the likelihood, and  $P(\beta)$  is the prior and  $\hat{B}$  is a rough estimation of the unattenuated background  $B$ , which can be estimated using the technique outlined in [Wu et al. 2007]. The likelihood is defined as

$$P(\hat{B}|\beta) = \exp\left(-\frac{\sum_{\mathbf{x} \in M} \|\bar{C}_I(\mathbf{x}) - \beta(\mathbf{x})\hat{B}(\mathbf{x})\|^2}{2\sigma_1^2}\right), \quad (6)$$

where  $\sigma_1^2$  is the variance of the measurement error ( $\sigma_1 = 1000$  for  $\beta = [0, 255]$  in our experiments). The smoothness prior of  $P(\beta)$  is defined by

$$P(\beta) = \exp\left(-\frac{\sum_{(\mathbf{x}, \mathbf{y}) \in N} \|\beta(\mathbf{x}) - \beta(\mathbf{y})\|^2}{2\sigma_2^2}\right), \quad (7)$$

where  $\sigma_2^2$  is the variance on the smoothness of  $\beta$ , and  $N$  is the set of first order pixel neighbors in  $M$ . ( $\sigma_2 = 5000$  in our experiments.)

### 4.3 $G$ -extraction

We describe how to specify the markup to simulate plausible refraction. A typical input image consists of a transparent object placed in front of a background with the refractive deformation observable within the object. When marking up the refractive deformation,  $G$ , visual cues from either the object or the background can be exploited. We refer to these as *object cue* and *background cue* respectively.

**Object cue.** When a background region is largely homogeneous or the deformed structure/texture is too complex to mark up, the shape of the transparent object itself provides the main cue for the user to mark up  $G$ . We first consider three basic cases of refractive light-transport (Figure 4):

- (1) No change (planar object requires no markup), Figure 4(a)
- (2) Convergence (convex object), Figure 4(b)
- (3) Divergence (concave object), Figure 4(c)

Depending on whether the perceived shape is convex or concave, the user first draws a rough line ( $c_{\text{ref}}$ ), then draws a curve ( $c_{\text{target}}$ ) that roughly follows the 2D shape of the object (refer to Figure 4). This object cue markup is quite effective in producing visually plausible deformation of the background. These basic cases when combined can be used to markup complex shapes. In addition, such

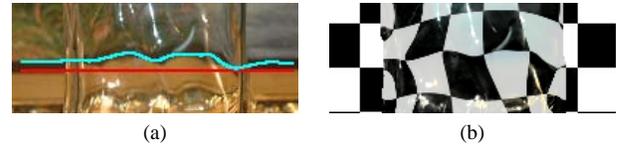


Fig. 5. Background cue. (a) Markup drawn using the background as a cue. As with object cue,  $c_{\text{target}}$  (cyan) is where  $c_{\text{ref}}$  (red) is perceived to distort to. (b) Example of the resulting deformation.



Fig. 6. Left: markups on *globe*, where the two corresponding curve pairs are labeled. Right: composite; note the inverted furry toy.

markup can be combined with markup specified based on background cues which is described next.

**Background cue.** Deformed background structure in the refractive object can also serve as a cue as shown in Figure 5. In this case, the user can draw  $c_{\text{ref}}$  to roughly indicate how this structure looks before deformation (e.g., a straight line), and then draw  $c_{\text{target}}$  to follow the deformed structure in the object as shown in Figure 5(a). Figure 5(b) shows the effect of this stroke pair on the deformation field. While automatic approaches such as deploying the object's mask to produce a simple blob-like 3D shape which refracts rays through it can model the deformation warp, in the following we describe an interactive approach that allows direct and easy user control.

### 4.4 Deformation Warping Implementation

When a  $c_{\text{ref}}$  and  $c_{\text{target}}$  curve pair is drawn, 2D points along the curves are sampled and serve as the input landmark-pairs for thin-plate-spline (TPS) warping [Bookstein 1989]. We use TPS because it is computationally efficient and is amenable to a 2D editing interface. Also, TPS requires no manual tuning and has a closed-form solution. TPS warping can easily simulate an image inversion effect based on the markup as shown in Figure 6. As demonstrated in the companion video, after each curve-pair markup, the deformation map (shown as a checkerboard pattern) is updated in real-time to provide instant visual feedback.

Depending on the input image, markup for  $G$  typically involves stroke pairs drawn using both object and background cues. The markup does not need to be that accurate. In fact, many different markups of an object can produce visually plausible refraction effects. This is attributed to our visual systems tolerance to errors in *complex* refractive light-transport as noted by Khan et al. [2006].

Figure 7 shows several versions of an ARM with different refractive deformation,  $G$ . Figure 8 shows the rough scribbles marked on the input image produce the corresponding “fakes” shown in Figure 7. This markup is casually performed and uses both object and background cues. Although our simulated results are visually dissimilar, each result is sufficiently plausible to make it difficult to detect refractive inaccuracies due to the inherent complexity.

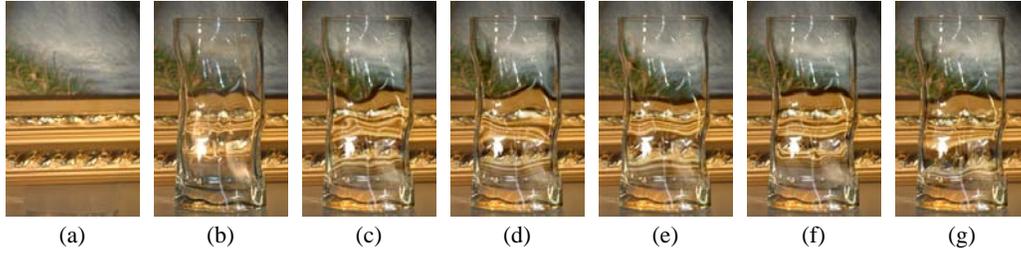


Fig. 7. *Can you spot the original?* We captured two images: one without the glass ( $I_0$ , (a)), and another with the glass ( $I_1$ , one of (b–g)). We extract the glass’ ARM directly from  $I_1$ . Several versions of the glass’ ARM are extracted using different user markup. We paste the ARMs onto  $I_0$  in (a) to produce the other ‘fake’ images in (b–g). All the edited examples look visually compelling and are not easy to distinguish from the real one. See Figure 8 to see which (b–g) is  $I_1$ .

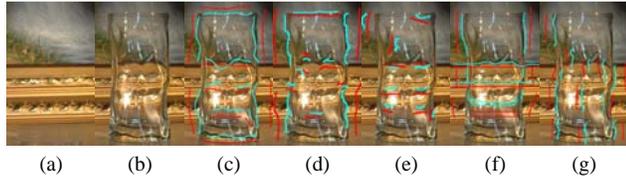


Fig. 8. Simple markups are sufficient for producing visually plausible results. Here, we show different scribbles marked on the original input image (b) and the corresponding ARM composites on the same background.

#### 4.5 Examples of Extracted ARMs

Figure 9 shows several example input images (column 1), markups (column 2), and the extracted ARMs (column 3). How to composite the ARMs onto a new background (column 4) will be explained in the next section. In Figure 9, the images *elephant* and *dragon* have very complex shapes. The *chandelier* has a complex surface with both opacity and transparency. A simple mask for separating opaque and transparent parts is available. The opaque part of the chandelier is extracted by conventional matting. In the *fish* example, the internal orange material is extracted as opaque colors. There is apparent color attenuation in the *fish* image and some in *elephant* and *chandelier* images. The images of colorless transparent objects (e.g., *dragon*) do not require the corresponding markup for  $\beta$ .

### 5. ARM COMPOSITING

A wide range of compositing effects can be produced from the extracted ARMs.

#### 5.1 Fresnel Effect

The occluding boundary of a transparent object produces a silhouette that exhibits rich reflection behavior. This phenomenon is characterized by the *Fresnel* effect. The Fresnel effect happens along the object’s silhouette, when the viewing angle to the transparent object becomes minimal and light glancing off this boundary makes it look as if it is opaque to the viewer [Hecht 1987]. Without the proper simulation of the Fresnel effect, the composited object often does not appear solid as shown in Figure 10(c).

In our image-based approach, the target background image can be regarded as the surrounding light. Hence, the resultant appearance due to the Fresnel effect depends on both the transparent object and the background image. To simulate the Fresnel effect, we apply

Poisson blending [Pérez et al. 2003] *along boundaries* to incorporate information from *both* the input and target background image.

Our approach amounts to solving a Poisson equation by taking into consideration: 1) the image gradient,  $\nabla C_{in}^{\partial M}$ , along the silhouette boundary ( $\partial M$ ) of the input image ( $C_{in}$ ) as the guidance field (Figure 10(a)), and 2) Dirichlet boundary condition derived from the pixel colors of the *composited image* surrounding  $\partial M$  (Figure 10(b)). In addition to the silhouette, this technique can also be applied along parts boundaries observed within the same object.

We define the pixel colors surrounding  $\partial M$  in the composite image to be  $C_{cp}^{\partial^2 M}$ , where  $\partial^2 M$  is the pixel location surrounding  $\partial M$ . The idea is to maintain the image structure in  $\partial M$  from the input image, and adapt the color from the composite image. Mathematically, it is equivalent to solving the minimization problem:

$$\min_{C_{cp}^{\partial M}} \int \int_{\partial M} |\nabla C_{cp}^{\partial M} - \nabla C_{in}^{\partial M}|^2 \quad \text{with} \quad C_{cp}|_{\partial^2 M} = C_{cp}^{\partial^2 M}|_{\partial^2 M} \quad (8)$$

where  $C_{cp}^{\partial M}$  is the color within  $\partial M$  in the composite image. The solution is the unique solution of the following Poisson equation with Dirichlet boundary conditions:

$$\Delta C_{cp}^{\partial M} = \text{div} \nabla C_{in}^{\partial M} \quad \text{over} \quad \partial M, \quad \text{with} \quad C_{cp}|_{\partial^2 M} = C_{cp}^{\partial^2 M}|_{\partial^2 M} \quad (9)$$

The width of the  $\partial M$  boundary can be adjusted given the image resolution and object shape. We found that widths from 4 to 6 pixels produced good results.

#### 5.2 Scene Depth

Although we have no 3D or depth information, we can simulate different scene depth when compositing the transparent object. Keeping the size of the object unchanged, we can make the background scene appear farther or closer to the object (shown in Figure 11(a) and (b)), by scaling the  $G$  component of the object’s ARM. Alternatively, if we want to animate the object by moving it towards the background (see Figure 11(c) and the accompanying video), a scale factor is applied to all components of the ARM ( $M, \alpha, \beta, G$ ) at its desired scene location. This scaling also simulates a change in the object’s apparent distance from the scene.

#### 5.3 Compound Compositing

To produce the effect of compound refraction involving multiple objects, the user assigns a depth order for each object. Each object’s ARM will be scaled as described in the previous section. Eqn (4)

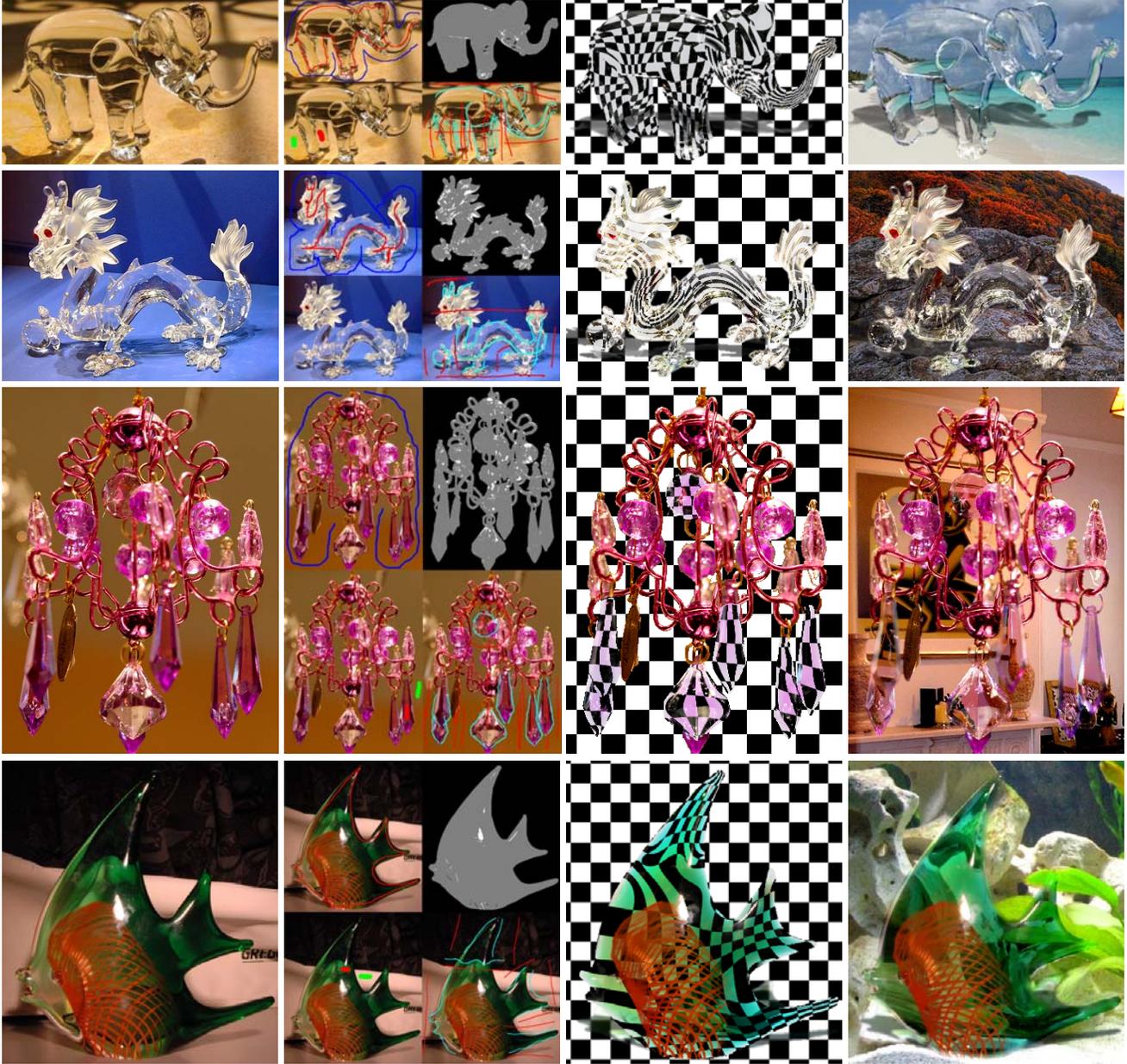


Fig. 9. ARM extraction and compositing. From left to right: Input, markup (from left in clockwise direction:  $M$ ,  $\alpha$ ,  $G$ , and  $\beta$ ), ARM, and composite. The examples are (from top to bottom): *elephant*, *dragon*, *chandelier*, *fish*. All the markup for  $M$ ,  $\alpha$ ,  $\beta$  and  $G$  shown here are complete. Refer to Table I for total number of stroke markups, and review the companion/supplemental videos for actual user interaction.

can be generalized to allow compound compositing of multiple and overlapping transparent objects, by writing it as a recurrence relation: Let  $S_n$ ,  $\alpha_n$ ,  $\beta_n$ , and  $G_n$  be respectively the  $S$ ,  $\alpha$ ,  $\beta$ , and  $G$  for object  $n$ ,  $n = 1, \dots, N$  and  $N$  is the total number of overlapping transparent objects, and  $M_n$  be the corresponding object mask. By letting  $B = C_{N+1}$  be the original background without any transparent object, we have

$$C_n(\mathbf{x}) = \alpha_n(\mathbf{x})S_n(\mathbf{x}) + \beta_n(\mathbf{x})C_{n+1}(G_n(\mathbf{x})) \quad (10)$$

for  $\mathbf{x} \in M_n$ . Figure 12 illustrates the recurrence relation where  $N = 3$ . The final composite is given by  $C_1$ . Figure 2 shows sev-

eral examples on compound compositing. Some of these examples show the extracted objects with different colors. This was achieved by rotating the hue of the input and the extracted  $\beta$ .

#### 5.4 Caustic Shadows

Since our ARM generation is non-geometric (no 3D model or explicit depth distribution), we exploit the availability of the refractive deformation  $G$  to aid in our shadow construction. Our method is intended only to add an additional touch of realism and cannot

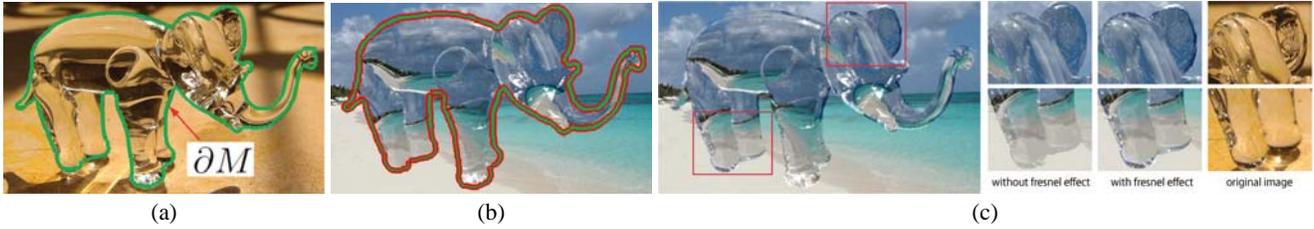


Fig. 10. Simulating Fresnel effect using *Poisson boundary blending*. (a) The image gradient in  $\partial M$  (green) in the input image serves as the guidance field. (b) Dirichlet boundary condition derived from the pixel colors of the composited image excluding  $\partial M$  (red). (c) Solving the Poisson equation on  $\partial M$  using (a) and (b) produces a Fresnel effect that helps make the composited object looks like a 3D transparent solid.



Fig. 11. Simulating scene depth without 3D. In (a) and (b), the object size is fixed. The  $G$  is scaled to make the background appear farther away and closer to the object, respectively. In (c), all the ARM's components are scaled to simulate object movement toward the background.

compete with accurate simulation of caustic shadows that use 3D models.

The procedure is as follows: at each pixel location inside  $M$ , the number of pixels mapped from the background based on refractive deformation  $G$  is accumulated. This estimates how light converges or diverges from the background to each pixel in the foreground object. The simulated shadow  $\mathcal{T}$  is given by

$$\mathcal{T} = p + \beta H(\mathbf{x}, G)q, \quad (11)$$

where  $p$  is a small constant serving as the ambient intensity for the shadow,  $H(\cdot)$  is a histogram function for tabulating at each pixel  $\mathbf{x}$  the number of pixels mapped to  $\mathbf{x}$  via the deformation warp  $G$ , and  $q$  is a user-supplied constant to control the apparent amount of light passing through the object. The  $\beta$  is the extracted attenuation map to give the shadow the same color as the transparent object.

In practice, the histogram  $H$  in Eqn (11) can be regarded as a high-dynamic-range (HDR) version of the caustic shadow. Directly using it will result in a shadow very bright at a few points but very dark overall. We find it useful to remap the histogram by applying a non-linear logarithmic compression (other re-mapping functions may be used as well) commonly used in tone reproduction methods.

While the original  $G$  marked up for the ARM can be used, the user can also specify a new  $G'$  for a more realistic caustic shadow effect, as shown respectively in Figure 13(a) and (b). In our examples, the *fish*, *jug*, *globe*, *lion* and *swan* have new  $G'$  specified for the caustic shadows. The simulated shadows are warped by projective transformation before compositing. The transformation is a simple heuristic involving 2D translation, scaling, and shearing. The shadow's contrast can also be adjusted to match the background image.

## 6. COMPARISON WITH PHOTOSHOP

We are not aware of any comparable matting and compositing system for editing transparent objects. The closest tool is Photoshop,

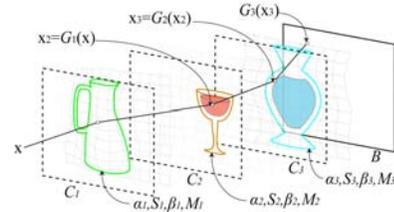


Fig. 12. Compositing three overlapping image-based transparent objects.



Fig. 13. Two results for simulating caustic shadows using our procedure. In one result (a), markup (left) is added to produce deformation  $G$  (middle) and the simulated shadow (right). In the result (b), a new deformation  $G'$  is created by simulating the convergent lens effect.

which is used to perform a comparison with our results. In addition, a user study is used to assess the preference for our results or those produced by a Photoshop expert.

Our ARM system is easy enough for a novice to be able to generate plausible results. Unfortunately, this is not so for Photoshop. As a result, we enlisted the help of a graphic artist professional. We contacted four professional agencies, with only two responding (they characterized our task as a 'challenge'). In the end, only one was able to complete our request. This particular person was an accomplished veteran digital artist in addition to a Photoshop expert. We also asked an intermediate Photoshop user (who has 5 years of experience) to generate results.

Both persons were given the original *elephant* and *glass* input image and the new backgrounds (same input used to produce our ARM composites). They were also shown our results merely to indicate the level of realism to achieve or exceed. Note that we did not ask them to replicate our results. The side-by-side comparisons are shown in Figure 14.

**Photoshop expert.** Interestingly, similar to our ARM representation, layers were used by the Photoshop expert to produce the transparent object transfer (Figure 15). However, unlike the ARM representation where each component has a semantic meaning (specular highlight, attenuation, deformation), the layers produced by the Photoshop expert are ad-hoc in nature, involving several lay-

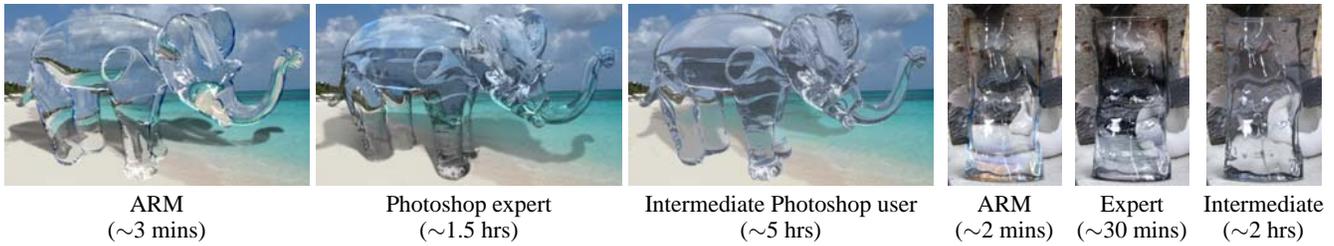


Fig. 14. An expert and an intermediate user of Photoshop were asked to produce comparable ARM results. The ARM extraction requires significantly less time (the times shown here include interaction time and overheads such as looking at the photo and deciding what to do). No caustic shadow and Fresnel effect were simulated by both Photoshop users. Notice that the glass produced by our intermediate Photoshop user is the poorest despite requiring the largest amount of effort.

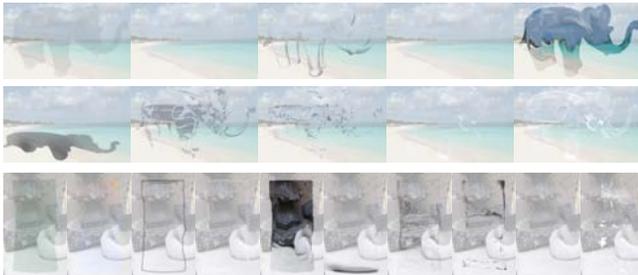


Fig. 15. Layers derived by the Photoshop expert to transfer the transparent object (the backgrounds have been made lighter to make the layers more apparent). As opposed to ARM extraction, the derived layers rely on the expert’s artistic sense and color perception on seeing the input and the background photo we provided; thus the approach may vary from other Photoshop experts.

ers of color range selection that were subsequently blended with the background. There were approximately 10 layers for each example as shown in Figure 15. The silhouette of the objects were drawn manually. Refraction was performed using the ‘liquify’ filter in both examples and has roughly the same effect for both results. The Fresnel effect is not apparent around the elephant legs and trunk. The Photoshop expert did not attempt caustics in the elephant shadow.

With the majority of time spent on the *elephant* example, it took the Photoshop expert approximately two hours to produce results comparable to our ARM results. In several places, our Photoshop expert painted in pixels that were not extracted from the original. In addition, the layer mixing is tuned for the given background. This is demonstrated in Figure 16 that shows the comparison of the Photoshop expert’s extracted *elephant* placed onto a different background. As opposed to the Photoshop expert’s layers, our ARM representation does not require fine tuning for a new background.

**Intermediate Photoshop user.** For our intermediate Photoshop user, 5 hours were spent on the *elephant* and 2 hours on the *glass*. This user’s photo-editing approach was first to remove pixels from the input image which were considered as background pixels using the eraser brush. The hues of the remaining pixels were then rotated to adapt to the new background. Because of the complex shape and illumination effects, a great deal of manual editing was necessary to make the result visually acceptable. Similar to the Photoshop expert, the liquify filter was applied to simulate the refractive deformation, however, the effect by the intermediate user is not as

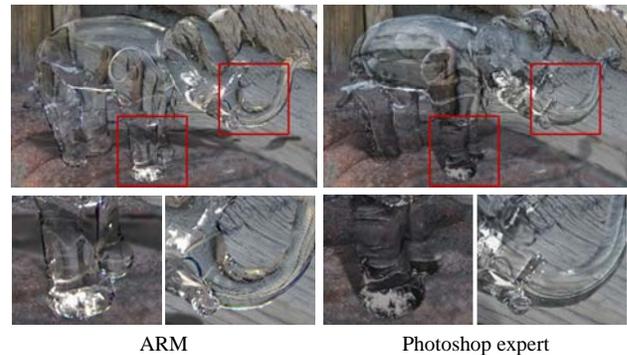


Fig. 16. Comparison of the ARM and Photoshop expert’s results when composited onto another background. Note in the ARM result the apparent Fresnel effects in the elephant legs and trunk, which are *automatically* generated when our ARM adapts to a new background image. The Photoshop expert’s layers do not blend naturally and requires manual fine-tuning.

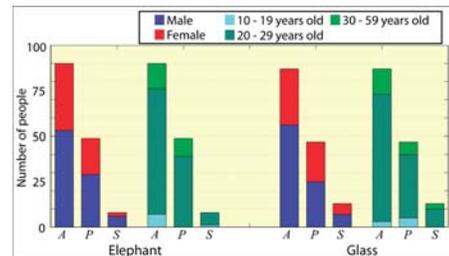


Fig. 17. Survey result on 147 subjects. Their preferences are indicated by A: ARM result, P: Photoshop expert’s result, and S: both look similar.

visually appealing as that done by the Photoshop expert. The caustic shadow was not attempted either.

**User study.** A user survey was performed to evaluate viewers preference between the composited results generated by our ARM approach or the Photoshop expert. In our study we did not include the results produced by the intermediate Photoshop user because they were not comparable in terms of visual quality as shown in Figure 14.

We sent a distribution list via email to invite people in our international campus community to participate in our survey. They were also invited to forward our invitation to their relatives and friends. A total of 147 persons in different age groups and genders responded to our online survey.

Table II. Input images  $I_A$  for ARM extraction.

	structured background			homogeneous background		
	simple	intermediate	complex	simple	intermediate	complex
unattenuated	 <i>martini</i>	 <i>wine</i>	 <i>horse</i>	 <i>martini</i>	 <i>wine</i>	 <i>horse</i>
attenuated	 <i>jug</i>	 <i>glassColor</i>	 <i>flask</i>	 <i>jug</i>	 <i>glassColor</i>	 <i>flask</i>

Six transparent objects over different backgrounds are tested for ground truth comparison. The transparent objects tested have different geometric complexities and attenuation properties.

In our online survey, the *elephant* and *glass* images produced by the ARM approach and the Photoshop expert were presented to a viewer. The images pair in each example were shown side by side for direct comparison, and we randomized the order shown for *elephant* to *glass* across different subjects to avoid bias. Each subject was asked to make a decision within a time limit of 30 seconds on each of the following questions: 1) “Please select your preferred image of a transparent and refractive elephant standing on a sandy beach.” 2) “Please select your preferred image of a glass placed in front of a background”. The user could also indicate that both results were visually similar, indicating no preference for either. Additional comments could be input in a text box available in the survey.

The statistics of our user study were collected and plotted in Figure 17, which indicates that our results are preferred. Comments by subjects indicate that the shadow and the refraction of the legs of the *elephant* in the Photoshop expert’s result look artificial. Some suggested that our results look clearer and are more pleasing. Some commented that results may depend on certain material properties and thus difficult to tell which one is better.

## 7. PERCEPTION VALIDATION

The Photoshop comparison serves to show our ARM matting and compositing approach outperforms current available image editing tools. To evaluate the realism of the ARM composites we performed a perception study. The experiment was conducted using a subjective two-alternative forced-choice preference approach similar to [Jimenez et al. 2009]. The research hypothesis was that an ARM-composited image produces as high a degree of visual realism as the real image of the same object against the same background.

ACM Transactions on Graphics, Vol. XX, No. Y, Article N, Publication date: ZZZ 2010.

### 7.1 Participants

Twenty participants were recruited for the validation experiment. This number of participants is comparable with similar validations, [Jagnow et al. 2008; Jimenez et al. 2009], in which sixteen user were recruited. The participants consist of 15 males and 5 females whose ages range from 20 to 42. All subjects reported normal or corrected-to-normal vision with no color-blindness. The participants were volunteers and not aware of the purpose of the experiment. Fourteen subjects reported that they did not have any expertise in image processing beyond simple photo editing. It is expected that participants with knowledge in photorealistic rendering may be better at distinguishing the real images from the “fake” images.

### 7.2 Data

We took photographs of six transparent objects to serve as input images. These objects were photographed on two different type of backgrounds, one with structure and one that is homogeneous. Table II shows these images which we refer to as  $I_A$ . These images are used to extract the ARM for the objects. The different backgrounds represent the different type of “visual cues” that would be available for the refractive deformation markup.

For ground truth, these same objects were imaged on new backgrounds shown in Table III where the real images for each background is  $I_1$ . For each background, two fake images ( $I_2$  and  $I_3$ ) were generated by extracting the ARMs from two  $I_A$  images (one with a structured background and the other with a homogeneous background), and then compositing them respectively onto the new backgrounds. The resulting decoy images  $I_2$  and  $I_3$  are respectively compared to the real image  $I_1$  in the forced-choice comparison of the user study.  $I_1$  on background 3 is equivalent to  $I_A$  (structured background).

Table III. A total of 18 scenarios were used in our user study.

	Background 1			Background 2			Background 3		
	$I_1$	$I_2$	$I_3$	$I_1$	$I_2$	$I_3$	$I_1$	$I_2$	$I_3$
<i>martini</i>									
<i>wine</i>									
<i>jug</i>									
<i>glassColor</i>									
<i>horse</i>									
<i>flask</i>									

Comparison with ground-truth real images. Background 1 has dominating salient linear structures; background 2 is a patterned background with no dominating structures; background 3 exhibits the same background used in capturing one of the input images (that is,  $I_A$ ) for ARM extraction.  $I_1$  is real;  $I_2$  and  $I_3$  are ARM composites.

Transparent objects with a variety of shape complexities were tested. Some of the objects are relatively simple as exhibited by the *martini* glass and the *jug*. The shapes of the *wine* glass and *glassColor* are moderately complex. For complex objects, we have the *horse* and the *flask*. In particular, the *flask* has an internal structure which is colored and transparent. Recall that Table II shows the input images  $I_A$  where these objects were captured and their ARM components were extracted. The left column shows  $I_A$  with structured background, where both the object and background cues are available; the right column shows  $I_A$  with homogeneous background, where only the object cue will be used. It is therefore ex-

pected to be more difficult for participants to identify the real image  $I_1$  from an  $I_1/I_2$  pair than from an  $I_1/I_3$  pair.

Table III shows a total of 54 images, arranged into 18 object and background scenarios, consisting of 18 real images ( $I_1$ ; 6 objects on 3 backgrounds) and 36 ARM composites (the corresponding  $I_2$  and  $I_3$ ) used in our user study. Our extracted ARMs were composited onto different background images as revealed here, which can be structured (backgrounds 1 and 3) and patterned (background 2). Compositing onto homogeneous background is skipped here because it produces little apparent deformation. To make the comparison fair, the lighting condition was kept constant during all image captures. We do not compare caustic shadows in our experiments.

Table IV. Chi-square analysis (degree of freedom = 1, level of significance = 0.05)

	object	$I_1/I_2$		$I_1/I_3$	
		$\chi^2$ -value	p-value	$\chi^2$ -value	p-value
background 1	<i>martini</i>	<b>1.225</b>	<b>0.268</b>	9.025	0.003
	<i>wine</i>	7.225	0.007	<b>3.025</b>	<b>0.082</b>
	<i>jug</i>	<b>0.025</b>	<b>0.874</b>	<b>2.025</b>	<b>0.155</b>
	<i>glassColor</i>	4.225	0.040	5.625	0.018
	<i>horse</i>	<b>1.225</b>	<b>0.268</b>	7.225	0.007
	<i>flask</i>	<b>3.025</b>	<b>0.082</b>	<b>0.625</b>	<b>0.429</b>
background 2	<i>martini</i>	<b>0.025</b>	<b>0.874</b>	<b>0.225</b>	<b>0.635</b>
	<i>wine</i>	<b>3.025</b>	<b>0.082</b>	<b>0.025</b>	<b>0.874</b>
	<i>jug</i>	<b>0.025</b>	<b>0.874</b>	<b>0.025</b>	<b>0.874</b>
	<i>glassColor</i>	<b>3.025</b>	<b>0.082</b>	<b>0.225</b>	<b>0.635</b>
	<i>horse</i>	<b>0.025</b>	<b>0.874</b>	<b>2.025</b>	<b>0.155</b>
	<i>flask</i>	<b>0.025</b>	<b>0.874</b>	<b>0.025</b>	<b>0.874</b>
background 3	<i>martini</i>	<b>1.225</b>	<b>0.268</b>	<b>1.225</b>	<b>0.268</b>
	<i>wine</i>	9.025	0.003	11.025	0.001
	<i>jug</i>	<b>3.025</b>	<b>0.082</b>	4.225	0.040
	<i>glassColor</i>	<b>1.225</b>	<b>0.268</b>	<b>0.625</b>	<b>0.429</b>
	<i>horse</i>	<b>1.225</b>	<b>0.268</b>	13.225	0.000
	<i>flask</i>	<b>1.225</b>	<b>0.268</b>	<b>0.225</b>	<b>0.635</b>

$I_1$ ,  $I_2$  and  $I_3$  are the respective images in Table III. Values shown in boldface indicate no significant difference with ground truths.

### 7.3 Procedure

Each participant viewed 72 trials in total (2 paired comparisons x 18 scenarios x 2 trials). Participants were encouraged to ask any questions before the study. After filling in a consent form and questionnaire, they were given a sheet of the task description:

“This test is about selecting one image from an image pair, and there are 72 pairs in total. You will be shown the images side-by-side with a grey image displayed between each evaluation.

Your task is to select the image that looks more realistic in each evaluation (that is, most like a transparent object placed in front of a background) by clicking on the image. You can view the image pair for an unlimited amount of time, but we suggest that you spend around 10 seconds on each image before your selection.”

In each trial, the participants assessed the realism of the presented images. Since our goal is to evaluate how realistic our ARM decoy images compared to real images, a reference condition was not used (in contrast to [Jagnow et al. 2008]). This no reference condition was also adopted in [Jimenez et al. 2009]. The image pairs were presented in a different random order for each participant. Counterbalancing was used to avoid any order bias: each paired comparison was assessed twice by each participant, where in half of the trials the real image is displayed as the left image and in the other half as the right image.

### 7.4 Results and Analysis

The primary goal of the experiment is to validate the quality of our ARM compositing results against realism. If the real photograph is not a clear winner over our results, then our objective in producing visually plausible composites using the ARM model is considered successful.

We analyzed the results statistically to determine in any statistically significant trend exists. To find out whether the number of participants who selected the real image is what would be expected by

chance, or if there was a pattern of preference, we adopted the Chi-square nonparametric analysis technique. A one-sample Chi-square includes only one dimension, such as the case in our experiments.

The obtained ( $I_1/I_2$  and  $I_1/I_3$ ) frequencies were compared to an expected 20/20 (40 for each comparison) result to ascertain whether this difference would be significant. The Chi-square values were computed and then tested for significance. Table IV shows the results. Overall, the survey results indicate that the real image is not a clear preference over the other two fake images. For the  $I_1/I_2$  pairs, among the 18 scenarios, only 3 of them show significant difference ( $p < 0.05$ ), that is, most participants managed to identify the real image in these cases. It means that in most cases, there was no significant difference between our ARM composites and the real images ( $p > 0.05$ ).

**Background.** An interesting finding is that given the same ARM, the background plays an important role to the perception outcome. For ARMs composited on patterned background without salient or curvilinear structures (background 2 in Table III), the scores between the pair are relatively close (that is, Chi-square values are small) in all of the examples, or in other words, they look equally plausible. Shown in the  $I_1/I_2$  column of Table IV, for ARMs composited on a structured background (background 3) the real and the decoy are statistically indistinguishable except for the *wine* glass. In most cases users commented that it is not easy to select the real image from a given pair. On the other hand, when the ARM is extracted from a homogeneous background where no background cue is available for deformation markup, and then composited onto a background image with salient curvilinear structures, it is easier for the user to detect deformation error, as reflected by the particularly high Chi-square values (refer to the  $I_1/I_3$  column of Table IV: *martini*, *glassColor*, and *horse* on background 1, *wine*, *jug*, and *horse* on background 3).

**Object.** Concerning object attenuation, we found that it does not produce noticeable difference on user’s preference. On the other hand, the object’s shape and its corresponding cues for ARM extraction do matter. In general, our ARM composites are preferred if both the object cue and the background cue are available for marking up deformation. This is evidenced by the fact that only 3 out of 18  $I_1/I_2$  pairs show significant difference, while the number rises to 7 out of 18 for  $I_1/I_3$  pairs. Referring to the three  $I_2$  images shown in Table III, they are quite acceptable except that the deformed structures look a bit unnatural, which become more apparent when the images are placed side by side with real photos. As for the relatively poor performance of  $I_3$  images, recall that they were composited using ARMs extracted from homogeneous background where only object cues are available for markup. Among all of the scenarios, the ARMs of the *horse* and the *wine* glass extracted from a homogeneous background produced the worst results with Chi-square values 13.225 and 11.025, respectively, where we are in lack of salient background cues for marking up the complex shape deformation.

## 8. DISCUSSION

While 3D object reconstruction requires measurement against ground-truth for validation, our ARM approach targets plausible refraction effects. This can be a subjective matter; in this paper we showed numerous examples and conducted user evaluations to demonstrate that our system is successful in achieving this goal. Our efficient system provides interactive photo-editing capability with almost instant feedback for each stroke the user marks up on

the image, thereby allowing the user to easily experiment with the effects of different deformations.

Because we make some simplifying assumptions in our single-image scenario to facilitate the ARM extraction, our technique cannot be readily applied to extract complex transparent objects from photos. Examples include multi-colored transparent objects (that is, violation of the smooth  $\beta$  assumption). In addition, specularities on the transparent object extracted cannot be adapted to a different lighting environment where the background image was captured. This limitation is inherent in all matting and compositing techniques using single images.

Our ARM cannot handle translucent objects which scatter light (e.g., jade). For such objects, they can be regarded as largely reflective and conventional matting is sufficient for their extraction if their apparent foreground colors are known. For objects with both transparency and opacity (e.g., *chandelier*) or consisting of different refractive mediums (e.g., *jug*), a rough segmentation on  $M$  (object mask) is needed to separate the processing regions.

While transparent object transfer can be performed by an experienced Photoshop user, our ARM representation has several advantages. First, its extraction is significantly faster and geared towards novice users. In addition, the ARM produces a meaningful and complete transparent object representation that can be used to paste onto a new background without fine tuning. The explicit ARM representation also allows for effects such as multiple object compositing, depth manipulation, and a procedural mechanism for producing caustic shadows.

Finally, knowing the 3D shape of the object offers much more opportunities for producing more realistic compositing effects, however, we emphasize that our work operates directly on a single image without 3D shape information and demonstrates a significant first attempt at photo-editing for transparent and refractive objects.

## 9. SUMMARY AND CONCLUDING REMARKS

We have introduced the attenuation-refraction matte (ARM) which provides an image-based model to encode the visual effects associated with transparent and refractive objects. In this paper, we described how the ARM can be extracted from a single image, and how to use the ARM to paste the object into a new background. We show that plausible refractive deformation suffices in producing visually compelling results. We believe our work is the first to allow photo-editing of transparent and refractive objects in cases where only a single image is available. In addition, we have shown a variety of compositing results that cannot be easily replicated using existing single-image editing tools. Our perception validation experiment indicates that ARM works very well with more structured backgrounds, and these should be preferred to homogeneous backgrounds if physical realism is desirable. Our future directions include investigating issues described in Section 8.

## Acknowledgments

We thank anonymous reviewers for their constructive comments. Special thanks to Holly Rushmeier for pointing out [Jagnow et al. 2008; Jimenez et al. 2009] and helpful suggestions on the perception study and Thomas Chow for his technical advice on professional photography. The research was supported by the Hong Kong Research Grant Council under grant numbers 620207 and 619208.

## REFERENCES

- AGARWAL, S., MALICK, S. P., KRIEGMAN, D. J., AND BELONGIE, S. 2004. On refractive optical flow. In *ECCV (2)*. 483–494.
- BEN EZRA, M. AND NAYAR, S. 2003. What does motion reveal about transparency? In *ICCV03*. 1025–1032.
- BOOKSTEIN, F. 1989. Principal warps: Thin-plate splines and the decomposition of deformations. *PAMI* 11, 6 (June), 567–585.
- CHUANG, Y.-Y., GOLDMAN, D. B., CURLESS, B., SALESIN, D. H., AND SZELISKI, R. 2003. Shadow matting and compositing. *ACM Trans. Graph.* 22, 3, 494–500.
- CHUANG, Y.-Y., ZONGKER, D. E., HINDORFF, J., CURLESS, B., SALESIN, D. H., AND SZELISKI, R. 2000. Environment matting extensions: towards higher accuracy and real-time capture. In *SIGGRAPH '00*. 121–130.
- FOLEY, J. D., VAN DAM, A., FEINER, S. K., AND HUGHES, J. F. 1995. *Computer Graphics: Principles and Practices (2nd edition in C)*. Addison-Wesley.
- GUTIERREZ, D., LOPEZ-MORENO, J., FANDOS, J., SERON, F., SANCHEZ, M. P., AND REINHARD, E. 2008. Depicting procedural caustics in single images. *ACM Trans. Graph.* 27, 5.
- HECHT, E. 1987. *Optics (2nd ed.)*. Addison Wesley.
- JAGNOW, R., DORSEY, J., AND RUSHMEIER, H. 2008. Evaluation of methods for approximating shapes used to synthesize 3d solid textures. *ACM Trans. Appl. Percept.* 4, 4, 1–27.
- JIA, J., SUN, J., TANG, C.-K., AND SHUM, H.-Y. 2006. Drag-and-drop pasting. *ACM Trans. Graph.* 25, 3, 631–637.
- JIMENEZ, J., SUNDTSTEDT, V., AND GUTIERREZ, D. 2009. Screen-space perceptual rendering of human skin. *ACM Trans. Appl. Percept.* 6, 4, 1–15.
- KHAN, E. A., REINHARD, E., FLEMING, R. W., AND BULTHOFF, H. H. 2006. Image-based material editing. *ACM Trans. Graph.* 25, 3, 654–663.
- LI, Y., SUN, J., TANG, C.-K., AND SHUM, H.-Y. 2004. Lazy snapping. *ACM Trans. Graph.* 23, 3, 303–308.
- MATUSIK, W., PFISTER, H., ZIEGLER, R., NGAN, A., AND McMILLAN, L. 2002. Acquisition and rendering of transparent and refractive objects. In *Eurographics Workshop on Rendering*.
- MIYAZAKI, D. AND IKEUCHI, K. 2007. Shape estimation of transparent objects by using inverse polarization ray tracing. *PAMI* 29, 11 (November), 2018–2030.
- MORRIS, N. AND KUTULAKOS, K. 2007. Reconstructing the surface of inhomogeneous transparent scenes by scatter trace photography. In *ICCV07*.
- PEERS, P. AND DUTRE, P. 2003. Wavelet environment matting. In *14th Eurographics Workshop on Rendering*. 157–166.
- PÉREZ, P., GANGNET, M., AND BLAKE, A. 2003. Poisson image editing. *ACM Trans. Graph.* 22, 3, 313–318.
- ROTHER, C., KOLMOGOROV, V., AND BLAKE, A. 2004. “GrabCut”: interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.* 23, 3, 309–314.
- SUN, J., JIA, J., TANG, C.-K., AND SHUM, H.-Y. 2004. Poisson matting. *ACM Trans. Graph.* 23, 3, 315–321.
- WANG, J. AND COHEN, M. F. 2008. *Image and Video Matting*. Now Publishers Inc., Hanover, MA, USA.
- WEXLER, Y., FITZGIBBON, A., AND ZISSERMAN, A. 2002. Image-based environment matting. In *EuroGraphics Workshop on Rendering*. 289–299.
- WU, T.-P., TANG, C.-K., BROWN, M., AND SHUM, H.-Y. 2007. Natural shadow matting. *ACM Trans. Graph.* 26, 2.
- ZONGKER, D. E., WERNER, D. M., CURLESS, B., AND SALESIN, D. H. 1999. Environment matting and compositing. In *SIGGRAPH '99*. 205–214.