

Flexible and Accurate Transparent-object Matting and Compositing Using Refractive Vector Field

Qi Duan, Jianmin Zheng and Jianfei Cai

School of Computer Engineering, Nanyang Technological University, Singapore

Abstract

In digital image editing, environment matting and compositing are fundamental and interesting operations that can capture and simulate the refraction and reflection effects of light from an environment. The state-of-the-art real-time environment matting and compositing method is short of flexibility, in the sense that it has to repeat the entire complex matte acquisition process if the distance between the object and the background is different from that in the acquisition stage, and also lacks accuracy, in the sense that it can only remove noises but not errors. In this paper, we introduce the concept of refractive vector and propose to use a refractive vector field as a new representation for environment matte. Such refractive vector field provides great flexibility for transparent-object environment matting and compositing. Particularly, with only one process of the matte acquisition and the refractive vector field extraction, we are able to composite the transparent object into an arbitrary background at any distance. Furthermore, we introduce a piecewise vector field fitting algorithm to simultaneously remove both noises and errors contained in the extracted matte data. Experimental results show that our method is less sensitive to artifacts and can generate perceptually good composition results for more general scenarios.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [COMPUTER GRAPHICS]: Picture/Image Generation—I.2.10 [ARTIFICIAL INTELLIGENCE]: Vision and Scene Understanding—Modeling and recovery of physical attributes

1. Introduction

In digital image editing, matting and compositing are two fundamental techniques [SB96]. The process of matting extracts a foreground object of arbitrary shape and its related information from a source image, and the process of compositing places the foreground object over an arbitrary background using the matte to control the contribution. Usually image matting and compositing techniques can generate fantastic visual effects, but they have difficulty in handling transparent objects which often exhibit refraction and reflection phenomena. Environment matting and compositing algorithms were therefore developed, which generalize the conventional matting and compositing by incorporating the information of how a foreground object refracts and reflects light from the environment.

The concept of environment matting and compositing was first proposed by Zongker *et al* [ZWCS99]. They demonstrated the ability of environment matting and compositing in capturing and rendering the effects of reflection, refrac-

tion, and scatter of light from environment, which greatly improves the visual reality of resulting images. Since then, some environment matting and compositing techniques have been proposed [CZH^{*}00, MPZ^{*}02, PD03, WFZ02, ZY04]. Basically, these researches were conducted towards two goals: higher accuracy and real-time processing. Surprisingly, none of them has taken the distance between the foreground object and the background into consideration during the matting and compositing processes. Specifically, the existing environment matting and compositing techniques require the distance between the object and the background to remain the same during the matte extraction and the composition processes. If the distance is changed, the matte has to be re-extracted. This is very inconvenient since in practice, when a matte is extracted from one environment and used for composition in a new environment, the distance between the object and the new background could be different from that in the first environment. Perhaps one could scale the background to an appropriate size and then use the scaled back-

ground as the input for [CZH^{*}00] to handle the distance issue. Unfortunately, this simple extension does not work very well, as shown in Figures 9 and 10. Moreover, the extracted environment matte data often contains noises that are some small unwanted perturbations and errors that are salient outliers in a certain local region. Although filtering approaches have been employed in the existing environment matting and compositing methods to remove noises and have yielded acceptable composition results, they are incapable of removing errors.

To overcome these limitations of the existing environment matting and compositing methods, in this paper, we first introduce the concept of refractive vector and propose to use a refractive vector field as the matte. Similar to the work in [CZH^{*}00], here we assume that the transparent object is colorless and specularly refractive. This will simplify our discussion in elaborating our new idea and enable us to focus on simulating refraction property, which is the most typical feature of transparent objects. Note that for each point (or pixel) on the foreground transparent object, there is a ray originating from the background and hitting on the transparent object. It then enters the object and eventually exits the object at the point towards the camera. We reverse all these rays to form a refractive vector field. Such a refractive vector field can be regarded as an attribute of the transparent object and is independent of the background. Once the refractive vector field is obtained, it is possible to place the object over a new background with a changing distance to the object (see Figures 9 and 10). Second, we propose a piecewise vector field fitting algorithm to refine the refractive vector field, which can simultaneously eliminate both noises and errors in the extracted matte data. Experimental results show that, compared with the existing environment matting and compositing methods, our algorithm is less sensitive to artifacts and can generate perceptually good composition results in more general scenarios.

The contributions of this paper are twofold. First, although it is not new to study the light transportation route passing through a transparent medium, to the best of our knowledge, this is the first work to explicitly use the last transportation route as the matte data for environment matting. Second, our contribution lies in the proposed piecewise refractive vector field fitting, which takes the characteristics of the refractive vector field into consideration and eliminates both noises and errors effectively.

The rest of the paper is organized as follows. Section 2 reviews the related work. Section 3 introduces the new environment matte, i.e. the refractive vector field, and explains how it can be used to approximately describe refraction and reflection effects. The vector field fitting algorithm is presented in Section 4, which eliminates the artifacts in the extracted environment matte data. Some compositing experiments of both synthetic objects and real captured objects are

provided in Section 5. Section 6 discusses the limitations of the proposed method and Section 7 concludes the paper.

2. Related Work

While the conventional matting extracts the opacity value and color of a foreground object at each pixel, environment matting is more complicated because it needs to calculate the refraction and reflection information in addition to the opacity value and the foreground color. To recover the refraction and reflection information of a transparent object, a large number of image samples of the object under specially designed backgrounds are often needed. Zongker *et al* [ZWC99] first proposed to use a sequence of structured backgrounds to estimate the environment matte data. The structured backgrounds consist of a hierarchy of finer and finer horizontal and vertical square-wave stripes. For a background of $k \times k$ pixels, $O(\log k)$ images are needed. Later on, Chuang *et al* [CZH^{*}00] proposed two extensions to enhance the usability of the environment matting algorithm of [ZWC99]. The first extension aimed to improve the accuracy of the matting results by using 2D oriented Gaussian pattern instead of axis-aligned rectangle to recover light spatial variation and dispersion effects as well as multiple mappings of texture, which can better approximate the BRDF (bidirectional reflectance distribution function) model. The second extension was targeted to achieve fast/real-time environment matting by making certain assumptions and simplifying the matting process to the one with only one picture captured against a special backdrop. The real-time environment matting is more practical and has been widely used in many applications. However, this real-time method is sensitive to artifacts, for which the anisotropic filter [PM90] was used in [CZH^{*}00] to eliminate the artifacts caused by noises.

There exist other solutions for environment matting and compositing. Wexler *et al* [WFZ02] used a probabilistic model to extract the matte by assuming that each background pixel has a probability of contributing to the final color of a certain foreground pixel. Peers and Dutré [PD03] used wavelet background patterns and wavelet processes to calculate environment matting data. Zhu and Yang [ZY04] introduced a frequency-based method for environment matting. They calculated the frequency response of foreground pixels to the time-sequence backdrop based on Fourier analysis. All of these methods, especially the wavelet method [PD03] and frequency-based method [ZY04], require a large number of image samples, which make the capture process very complex and time-consuming.

There are also some interesting researches on transparent objects by combining environment matting with other techniques. For example, Matusik *et al* [MPZ^{*}02] built a complex light stage system, which consists of a turntable, a set of cameras and lights, and monitors. The system can capture environment matting under multiple viewpoints and reconstruct 3D visual hull of transparent objects and their surface

reflectance field. In this way, a transparent object can be synthesized under new viewpoints, background and illuminations. Moshe *et al* [BEN03] used a moving camera to capture a transparent object and tracked some feature points inside the object. Then, based on the pre-calculated camera motion and the refracted light direction, an approximate shape of the object can be estimated. Agarwal *et al* [AMKB04] used the optical flow method to track and simulate the refractive effect of a transparent object from a video sequence of the object in front of a moving background. All these methods involve 3D reconstruction of a transparent object and their computational complexity is very high. Other studies on the 3D reconstruction of transparent (or specular) objects include [MKI04, YIX07, KS08].

3. Refractive Vector Field For Environment Matting

The environment matting equation introduced in [ZWCS99] is

$$C = F + (1 - \alpha)B + \sum_{i=1}^m \rho_i \mathcal{M}(T_i, A_i) \quad (1)$$

where C , F and B are the colors of image, foreground, background respectively, α is the opacity value of the foreground object, ρ_i is the reflectance that describes the contribution of light emanating from the environment to the object, \mathcal{M} is the “texture-mapping operator” calculating the average color of region A_i in background texture T_i , and m is the number of the structured background textures. The region A_i is specified by the center $c = (c_x, c_y)$ and the size $w = (w_x, w_y)$ where w_x and w_y are the dimensions in x - and y -directions.

The approach in [CZH*00] simplifies the environment matting equation by making some assumptions with a focus on the refraction effects of the transparent object. With the assumption that the object is colorless and specularly refractive, the term $F + (1 - \alpha)B$ can be ignored. While only one specially designed background is used in the matte acquisition process, the summation and the script i can be dropped and $\mathcal{M}(T, A) \approx T(c)$. Thus the environment matting equation is simplified as

$$C = \rho \cdot T(c) \quad (2)$$

which implies that the refracted light seen from one point of the transparent object comes from only one point (or region) of the background, as illustrated in Figure 1. Based on (2), c and ρ are extracted, which are treated as the matte data. This approach is simple and it can capture environment matting data in real time with a one-shot picture of the object against a specially designed backdrop. Essentially, it establishes a point-to-point (or region) mapping between the object and the background plane. However, when the distance between the object and the background plane is changed, such mapping has to be rebuilt, which means that we have to repeat the complicated and time-consuming data acquisition process.

In this research, we propose a new idea to solve the above

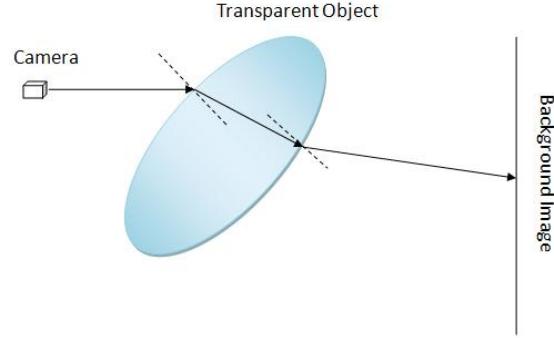


Figure 1: Illustration of a refractive light route.

problem, which is to calculate a new geometry element (rather than c) as the matte. Let us consider the ray originating from the camera and casting to a point on a transparent object (see Figure 1), which enters the object and eventually exits the object towards the background. This ray route is the reverse of an actual light originating from the background and finally entering into the camera after refraction. Here we are interested in the last transportation route (from the object to the background) of the ray, which we call a *refractive vector*. Note that by this definition the refractive vector can be specified by a point and a direction. It is not simply a free vector. For each point (or pixel) on the foreground transparent object, there exists a corresponding refractive vector. All these refractive vectors form a refractive vector field, which we choose as our new representation of environment matte. The refractive vector field can be regarded as an attribute of the foreground transparent object since it depends on the object and is independent of the background. In the following we describe how to represent, compute and extract the refractive vector field.

3.1. Refractive vector calculation

Suppose we have two samples which are the shots of a transparent object against the background at two different distances. Denote the distances between the object and the two backgrounds by d_1 and d_2 , respectively. Considering a foreground pixel (x, y) in the two pictures, if the corresponding regions in the first and second backgrounds that contribute to the color of this pixel are centered at coordinates (c_{1x}, c_{1y}) and (c_{2x}, c_{2y}) , then $(c_{2x} - c_{1x}, c_{2y} - c_{1y}, d_2 - d_1)$ is the direction of the refractive vector (see Figure 2).

Next we need to choose a point on the refractive vector. For simplicity, we choose the point at which the refractive vector hits on the foreground plane and call it a *starting point*. If we let the foreground plane have a z -coordinate of 0, then by some calculations we obtain the starting point $(x_0, y_0, 0) = (\frac{d_2 c_{1x} - d_1 c_{2x}}{d_2 - d_1}, \frac{d_2 c_{1y} - d_1 c_{2y}}{d_2 - d_1}, 0)$. The refractive vec-

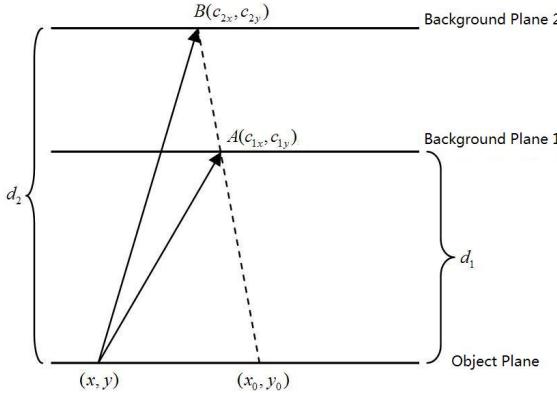


Figure 2: Computing a refractive vector shown in a dash line.

tor is thus described by parametric equation as

$$P(t) = (x_0, y_0, 0) + (c_{2x} - c_{1x}, c_{2y} - c_{1y}, d_2 - d_1)t$$

where t is a distance parameter.

In practice, we can let the difference between d_1 and d_2 be fixed (say $d = d_2 - d_1$). Then we only need to store the first two components of the starting point and the direction of the refractive vector, which form a 4D vector $(x_0, y_0, c_{2x} - c_{1x}, c_{2y} - c_{1y})$. We denote this 4D vector by $\mathcal{R}(x, y)$, which is the representation of our proposed refractive vector.

Once the refractive vector is extracted, for a new background with a distance of d_3 to the object, we can easily find the corresponding mapping center (c_{3x}, c_{3y}) in the new background for foreground point (x, y) :

$$(c_{3x}, c_{3y}) = (x_0, y_0) + \frac{d_3}{d} (c_{2x} - c_{1x}, c_{2y} - c_{1y}) \quad (3)$$

3.2. Refractive vector field extraction

Now let us look at how we can extract the refractive vector field of a transparent object. During the data acquisition process, we fix the distance between the camera and the object while placing the background (a monitor) at two different distances. For each distance, we adopt the simplified environment matting model (2) proposed in [CZH*00] to calculate the corresponding mapping center c in the background and ρ for each pixel of the object, which requires only two images, with and without the transparent object in front of the specially designed background (a planar slice through the RGB cube). Once the two mapping centers are found, we can compute the refractive vector using the proposed method introduced in Section 3.1. In this way, we extract the entire refractive vector field for the foreground object, which requires only two pairs of images in total.

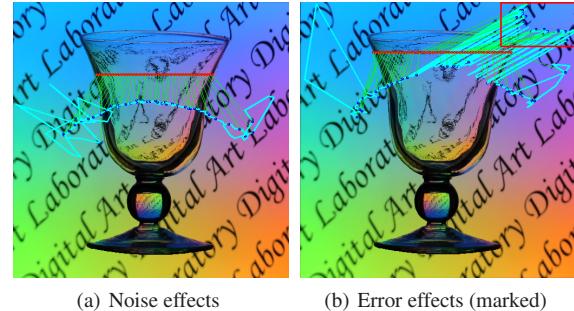


Figure 3: Artifacts in environment matting [DCZ09].

4. Piecewise Refractive Vector Field Fitting

Ideally, with the established refractive vector field, we should be able to composite the transparent object into any new background image. However, the pre-generated vector field usually contains a quite significant number of artifacts, which are brought in from the data acquisition and matte extraction process. Figure 3 shows examples of directly reusing the pre-generated refractive vector field for composition. In the figure, some marks are added to show the direct vectors from the red points on the transparent object to the corresponding mapping points in blue on the background.

Artifacts in the extracted refractive vector field can roughly be classified into noises that are small perturbation existing randomly in the whole refractive vector field and errors that are salient outliers in some local regions and are generated because of some assumptions of the approximate model or processing. Some approaches have been developed to remove such artifacts. For example, Chuang *et al* [CZH*00] used the anisotropic filtering method [PM90], which can remove noises after several iterations but is not suitable for removing errors. While most environmental matting methods handle those artifacts equally, Duan *et al* [DCZ09] proposed a method to treat them differently. The method identifies errors as the outliers in each horizontal line. A weighed curve fitting method is used to fit the direct vector field in each horizontal line with lower weight assigned to error vectors. The method can obtain a visually satisfactory result but it is limited to some symmetric situations.

In this section, we present a piecewise fitting method to remove both noises and errors of the refractive vector field. The method consists of region segmentation which decomposes the whole foreground area into some regions that have relatively consistent refractive vectors, and B-spline fitting which fits a 4D B-spline surface to the refractive vector field for each region. It is worth pointing out that in this method each region is independently fitted by a B-spline surface and there could be discontinuities around the region boundaries. Some feature-preserving boundary smoothing meth-

ods may be used for blending region boundaries. However, since the regions are obtained by the region segmentation process based on the refractive vector variance and the refractive vector field may not be continuous from one region to another, it is not very necessary to make a transition between regions to make the field continuous around the region boundaries. Our experiments also show that adding boundary smoothing does not render apparent visual improvement.

4.1. Region segmentation

The purpose of region decomposition is to segment the whole foreground object area into some regions such that each of the regions has consistent refractive vectors. The process of region segmentation is imperative because it affects the effectiveness of the subsequent B-spline surface fitting. In this research, we adopt MeanShift algorithm [CM02] to classify all refractive vectors and segment the whole refractive vector field into sub-regions based on a metric that considers both the distance of two pixels and the difference of their refractive vectors. To avoid over-segmentation, we also merge some small regions after the MeanShift segmentation by comparing the mean value and variance of the refractive vectors of small regions.

4.2. B-spline fitting

Consider a region $\Omega \in \mathbb{R}^2$ which is a part of the foreground area. The extracted refractive vector field over Ω is denoted by \mathcal{R}^* , which may contain noises and errors. Thus \mathcal{R}^* can be written as $\mathcal{R}^* = \mathcal{R} + \mathcal{E} + \mathcal{N}$, where \mathcal{R} , \mathcal{E} and \mathcal{N} stand for the ideal refractive vector field, errors and random noises, respectively. Our goal is to recover \mathcal{R} , the target refractive vector field that does not have artifacts. This is very difficult or even impossible if an analytical solution is sought. Therefore we consider this artifact removal task as an energy minimization problem, that is to minimize

$$\sum_{(x,y) \in \Omega} (\|\Delta\mathcal{R}(x,y)\|^2 + \lambda\omega(x,y)\|\mathcal{R}^*(x,y) - \mathcal{R}(x,y)\|^2) \quad (4)$$

where Δ is the Laplacian operator over the refractive vector field, the first item is for the smoothness of the refractive vector field, the second item is to maintain data fidelity during the minimization process, λ is a trade-off factor (in this work, we set $\lambda = 10$), and ω is a weight for each refractive vector. This is a typical weighted least square fitting [Die81, HH74]. If the weights are appropriately chosen, the solution of (4) could be a good approximation to the ideal \mathcal{R} .

The choice of weight $\omega(x,y)$ is crucial to the final solution. This is because the weight reflects the importance of a refractive vector. As verified in [DCZ09], it is better to give the highest weight, a lower weight, and the lowest weight to regular refractive vectors, noise vectors, and error vectors than to treat them equally in [CZH⁰⁰]. In this way, the error refractive vectors will be ignored in the fitting procedure,

and through the energy minimization, the corrupted error refractive vectors can be pulled back towards the regular ones.

In particular, since refractive vectors come from the refraction phenomena, the change of the refractive vectors in a smooth region should be consistent, which is mentioned as the criterion of "smooth change" in [DB03]. For point (i,j) in a certain smooth region, if the average of the refractive vectors in that region is denoted by $\bar{\mathcal{R}}(i,j)$, we measure the local consistency at (i,j) as

$$L(i,j) = \|\mathcal{R}(i,j) - \bar{\mathcal{R}}(i,j)\|. \quad (5)$$

The value of $L(i,j)$ can be used to detect whether there exist severe errors in the extracted refractive vector field. If $L(i,j)$ is above a certain threshold (say, $2 \times \|\bar{\mathcal{R}}(i,j)\|$), such a refractive vector is considered to be an error vector to which zero weights should be given. Moreover, considering that transparent objects such as glass are typically of piecewise smooth shape, it is reasonable to assume that the refractive vector field is locally smooth. We measure the smoothness of a refractive vector $\mathcal{R}(x,y)$ at coordinates (i,j) through calculating the difference of the 4D vector $\mathcal{R}(x,y)$ and its neighbors:

$$S(i,j) = \|\mathcal{R}(i,j) - \frac{1}{4} \sum_{(i^*,j^*) \in N_1(i,j)} \mathcal{R}(i^*,j^*)\| \quad (6)$$

where $N_1(i,j)$ denotes the 1-ring neighbors of point (i,j) and $\|\cdot\|$ is the \mathcal{L}_2 norm. The smaller $S(i,j)$ is, the smoother the refractive vector field is at point (i,j) . Thus $S(i,j)$ can be used to estimate the significance of the noises in the extracted refractive vector field and the weight $\omega(i,j)$ should be calculated based on it. Therefore we define the weight for each vector by

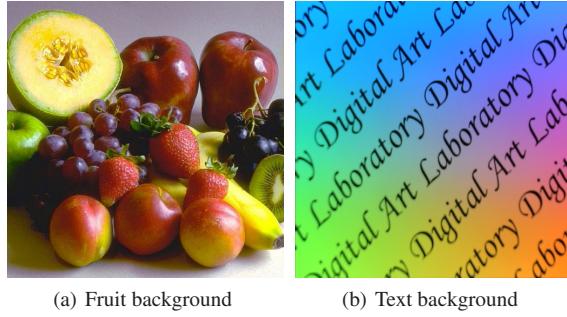
$$\omega(i,j) = \begin{cases} 0, & \text{if } L(i,j) > \text{threshold} \\ \exp(-\alpha S(i,j)), & \text{if } L(i,j) \leq \text{threshold} \end{cases} \quad (7)$$

where α is a pre-defined constant and we set $\alpha = 0.1$ in all our experiments.

To find a reasonable \mathcal{R} from the minimization problem, we let the solution space consist of bicubic B-spline surfaces. The mathematical equation of a parametric tensor product B-spline surface $\mathcal{R}(x,y)$ is

$$\mathcal{R}(x,y) = \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} r_{i,j} N_i(x) N_j(y) \quad (8)$$

where $r_{i,j}$ are the B-spline coefficients to be determined; n_x, n_y are the numbers of the B-spline coefficients in x - and y -directions, respectively, which are initially chosen to be the maximum of 4 and one-tenth of the dimension of the region and then are adaptively adjusted during the fitting process according to the fitting error [Die81]; $N_i(x)$ and $N_j(y)$ are normalized cubic B-spline functions [De 72]. Substituting (8) into (4), taking the partial derivatives of (4) with respect to $r_{i,j}$, and letting them equal zero lead to a system of linear equations. The solution to the linear system gives the



(a) Fruit background

(b) Text background

Figure 4: Two backgrounds used in the experiments.

optimal B-spline surface, from which the target refraction vector field \mathcal{R} can be recovered.

5. Experiments

We are now ready to summarize our transparent object matting and compositing method. First, we capture the images of the specially designed background at two different distances, with and without a transparent object in front of the background. Then we extract the refractive vector field as the matte for the transparent object. Next we apply piecewise B-spline fitting method to remove the artifacts of the refractive vector field. After that, we can perform compositing for any background at any distance away from the object. That is, given an arbitrary background at any distance, we can composite the transparent object into the background to synthesize a new image.

We compare our algorithm with the previous methods and also compare the compositing results with the ground truth. The comparisons are performed on both real captured transparent objects and a synthetic model. The synthetic transparent object is created using 3DS Max, whose material properties are set to be colorless and specularly refractive, which satisfy the assumptions and requirements of the real-time environment matting algorithm [CZH*00]. The object is then rendered in 3DS Max and the images generated are used as the inputs to our method or as the ground truth. To better illustrate the accuracy of the recovered refractive vector field, we choose two background images as the compositing background: the first one is a fruit background (Figure 4(a)) and the second one is a text background image with many letters (Figure 4(b)). Both background images contain much structure information, which helps to demonstrate the accuracy of the output refractive vector field. The more accurate refractive vector field we recover, the more structure details we can see in the area of the transparent object.

5.1. Accuracy

To verify the effectiveness and the accuracy of our proposed piecewise B-spline fitting method, we compare our method,

the anisotropic filter method [CZH*00], the vector fitting method [DCZ09], and the ground truth. Figures 5–8 are compositing examples of synthetic or real captured transparent objects with or without severe artifacts in the initially extracted environment matte data. In these figures, the first row shows the results of intermediate steps: the initially extracted refractive vector field, the segmentation result, and the weights in which the dark intensity indicates the severe noises or errors in the extracted environment matting data. The vector field is visualized by a color map in which the red and green channel values correspond to the x and y direction values after normalization. The second row shows the refined vector fields and the third row (or the fourth row) is for the compositing results. The results in the first, second, third, and fourth columns are generated by the anisotropic filter method [CZH*00], the vector fitting method [DCZ09], and the proposed method, and the ground truth, respectively.

Figure 5 shows the compositing results of a synthetic transparent glass with two different background images. Figures 6 and 7 are examples of a real transparent object captured in two different orientations. These three examples all contain significant amount of noises and errors in the initially captured refractive vector field data, which can be seen from the weights or the initial vector fields. For the anisotropic filter method, we can see that the result vector fields are still full of artifacts, especially in the regions with severe errors. This is because the anisotropic filter method can only eliminate noises but not errors. For the vector fitting method, it can remove both noises and errors simultaneously, but it only identifies and rectifies errors in each horizontal line. Moreover, the vector fitting method fails when the axial symmetry assumption is not satisfied (see images in Figure 7). Compared with these two methods, the proposed method can identify and eliminate artifacts globally and the compositing results are more accurate and clearer, even in error regions.

Figure 8 shows another real captured example, in which the extracted environment matte data contain only a small number of noises. In this case, all the three methods can produce visually satisfactory results, but in general our method still produces a more accurate and smoother result which is very close to the ground truth, especially around the edge regions.

5.2. Comparison with the simple scaling method for distance-varying backgrounds

Now we consider composition with distance-changeable backgrounds, for which the state-of-the-art method needs to repeat the entire matte acquisition process for each new distance. In contrast, for our approach, the matte acquisition and the refractive vector field recovery process only need to be performed once. Based on the recovered refractive vector field, we can generate the compositing result for any background and any distance in a very convenient way. For com-

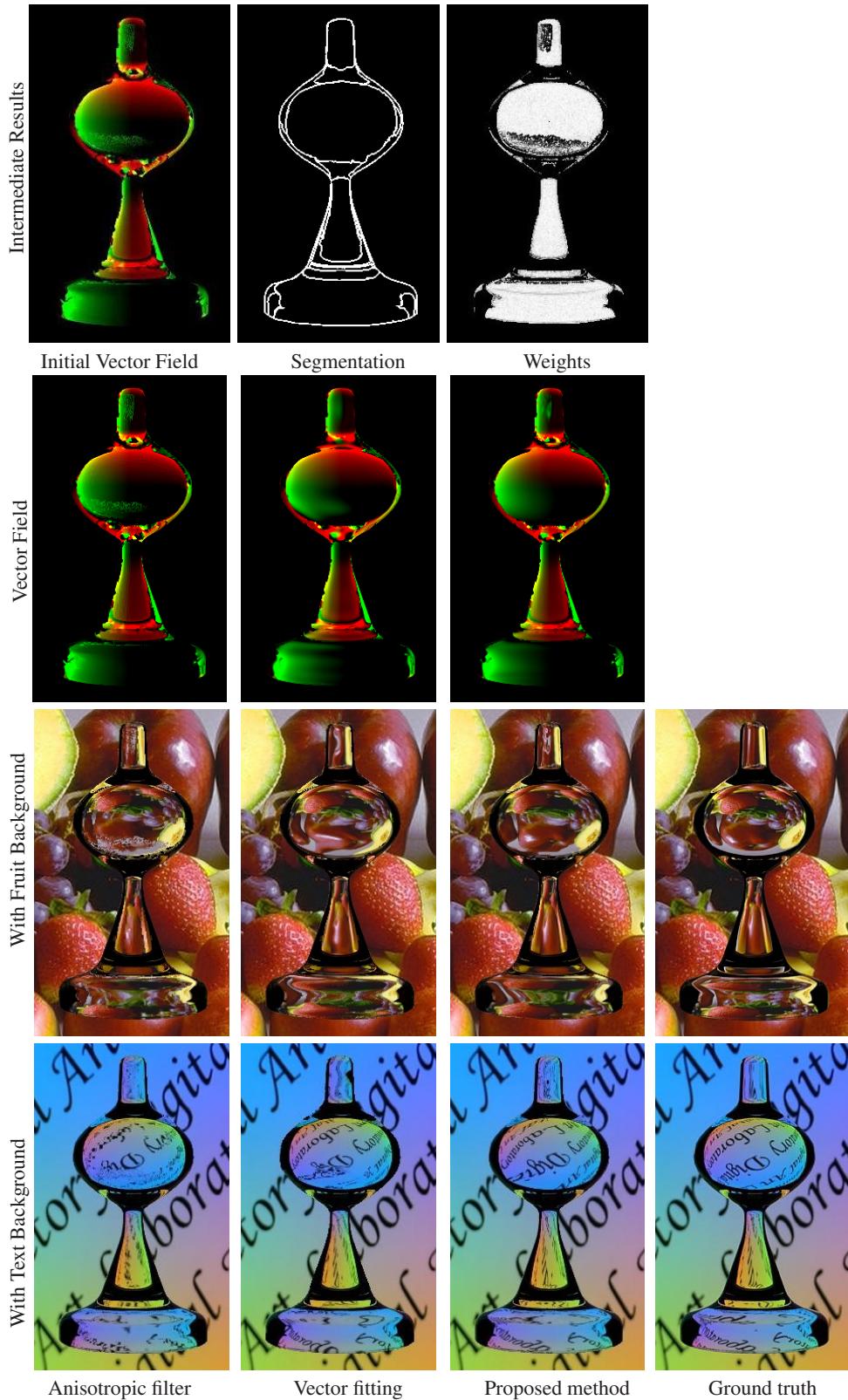


Figure 5: Compositing results of a synthetic transparent glass model with two different backgrounds.

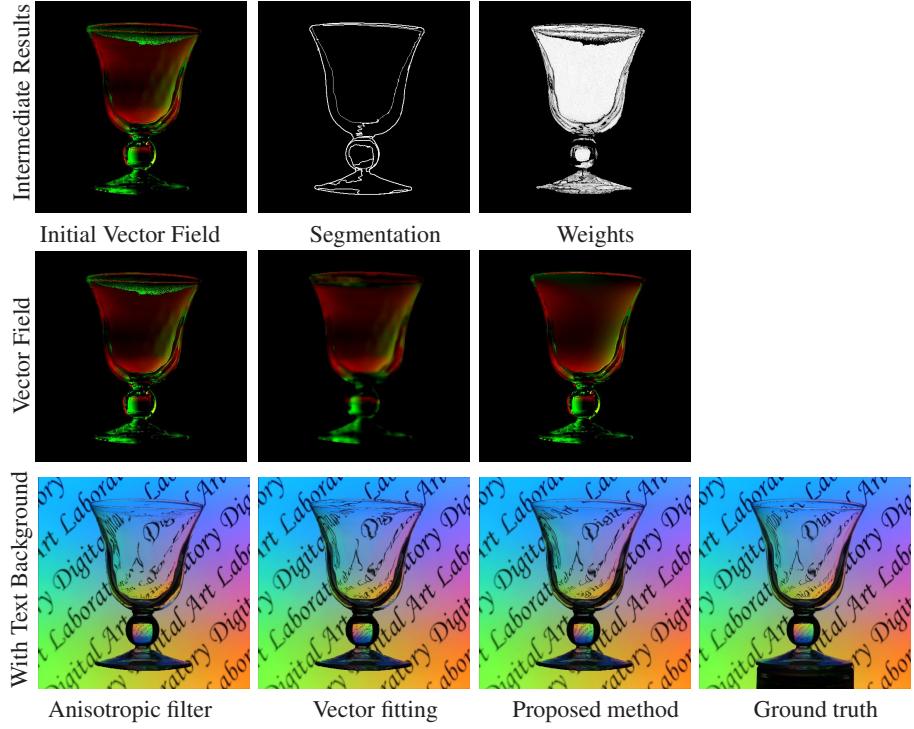


Figure 6: Compositing results of a real captured transparent goblet whose extracted matte data contain severe artifacts.

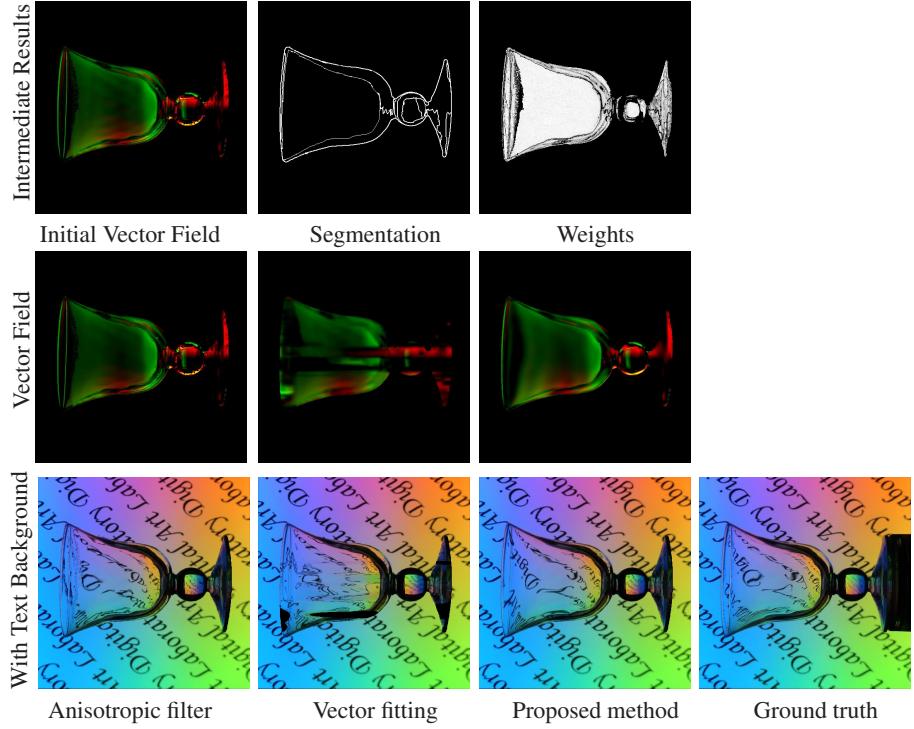


Figure 7: Compositing results of the transparent goblet captured in a different orientation. The extracted environment matte data also contain severe artifacts.

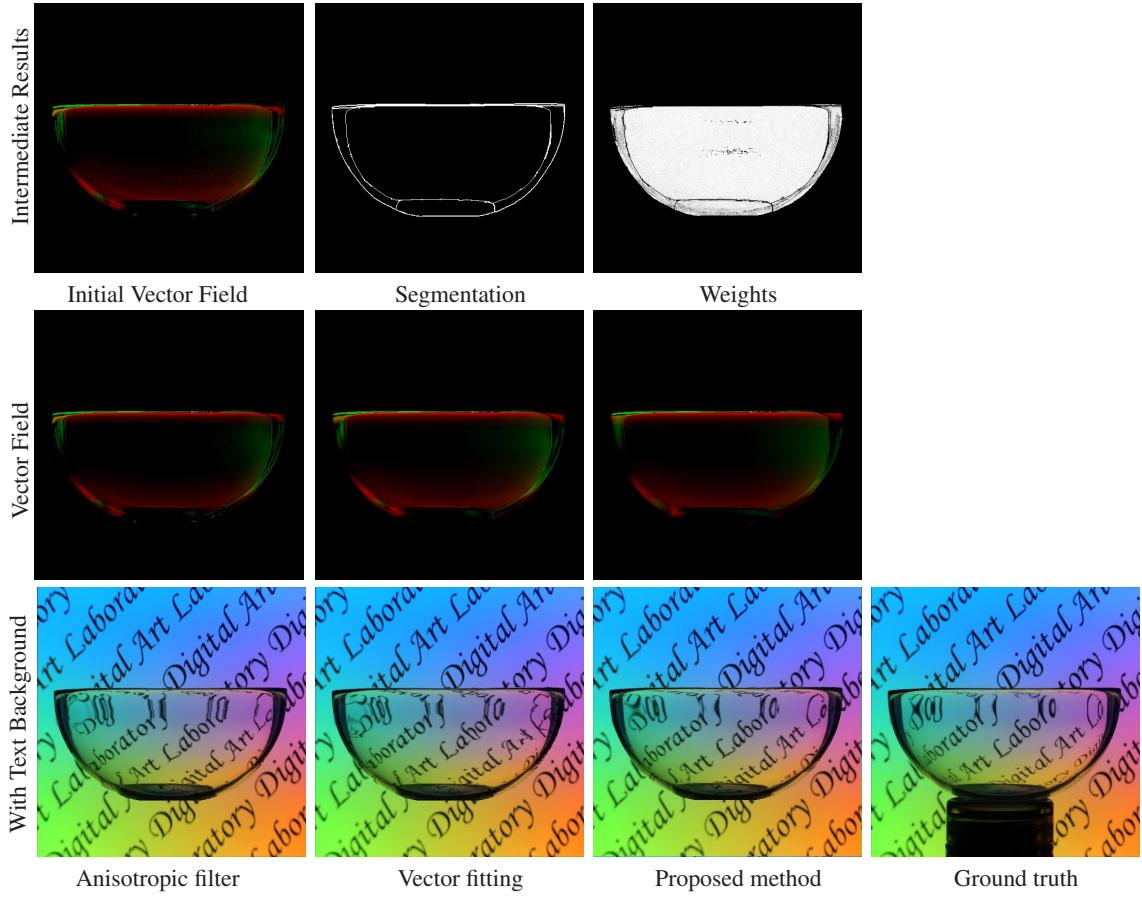


Figure 8: Compositing results of a real captured transparent bowl, for which the extracted environment matte data are quite clean.

parison, we implement the simple scaling method, which scales the background to an appropriate size as the input for the method of [CZH*00]. The simple scaling method implicitly assumes that the foreground pixel lies on the refractive vector, which is generally not true as shown in Figure 1. Therefore the simple scaling just gives an approximation of the refraction effect. Figures 9 and 10 show the composition results with different distances between the transparent object and the background plane. Compared with the ground truth rendered by 3DS Max, apparently our approach simulates the distance-changing effect better and produces more visually pleasing results.

6. Limitations

In our experiments, all objects are assumed to be colourless and specular refractive. This is because the real-time environment matting method [CZH*00] that we used to extract matte data under two background distances adopts a simplified environment matte model with such assumptions. More

general methods of extracting the initial environment matte may remove this limitation.

In addition, since our proposed method uses region segmentation and B-spline fitting to recover the refractive vector field, it is assumed that the refractive vector field should be piecewise smooth. In case that the field of a certain transparent object does not have such smoothness property or the field is over-segmented too much due to the very complicated geometry of the transparent object, the method may not work very well. For example, Figure 11 shows a transparent dragon model that has a lot of geometry variance in the surface. Our method produces a smooth composition result while the ground truth contains a lot of fine refractions.

7. Conclusion

In this paper we have described a method for flexible and accurate transparent object matting and compositing. We have introduced a refractive vector field as a new representation of environment matte, which allows us to easily composite a

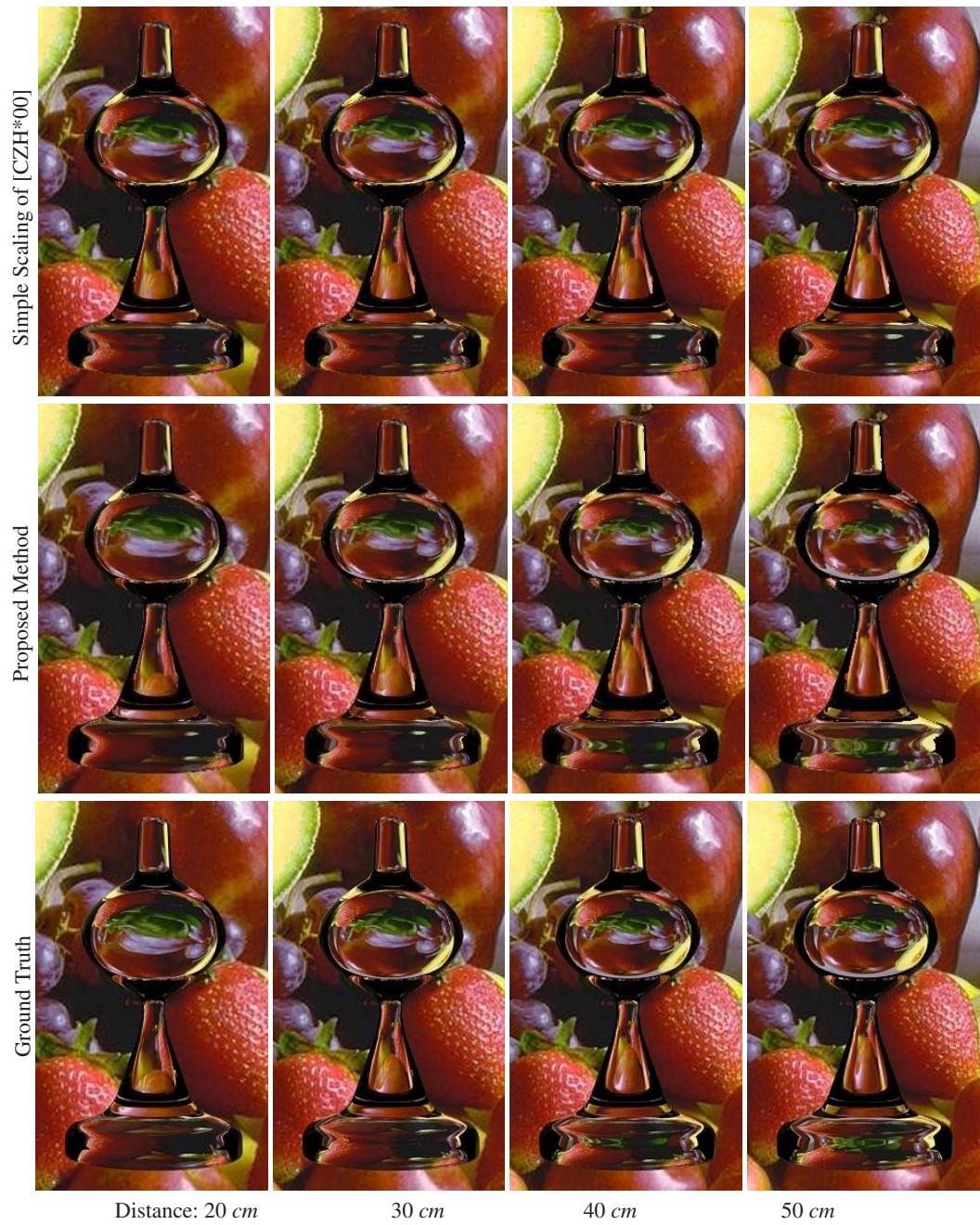


Figure 9: Compositing results with a fruit background placed at different distances.

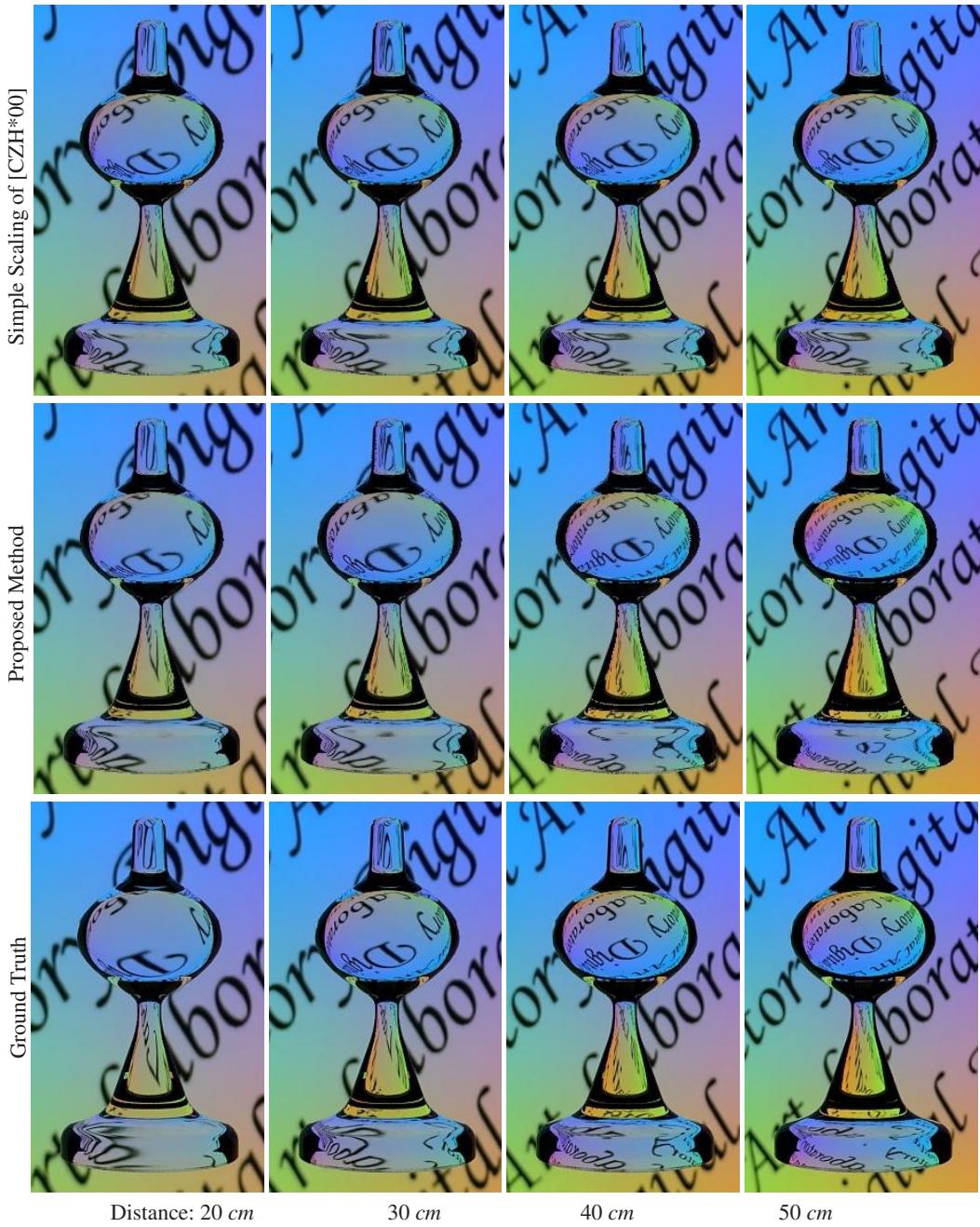


Figure 10: Compositing results with a text background placed at different distances.

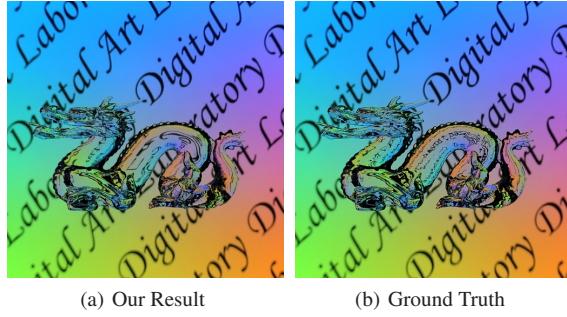


Figure 11: An example for which the proposed method does not work well.

transparent object into an arbitrary new background placed at any distance. Such flexibility cannot be provided by the existing environment matting algorithms, which require to repeat the whole complicated and time-consuming matting extraction process for any new distance. Moreover, to provide accurate environment matte data, we have also proposed a piecewise vector field fitting algorithm using bicubic B-splines, which can eliminate both noises and errors efficiently. Compared to the previous methods, our approach can produce more robust and accurate compositing results.

Note that the refractive vector field introduced in the paper describes the light interaction between objects and the environments. It would be interesting to extend this new representation to handle more general situations that include common transparent objects and other objects with light interaction properties such as light refraction, reflection and scattering. In addition, most environment matting algorithms judge the quality of the compositing results simply based on how the results visually look like the ground truth. When the ground truth (real captured image or rendered image) is available, existing objective metrics such as MSE and PSNR might be used to evaluate the matting and compositing performance. However, it is observed that the recovered structure information in the compositing results is usually crucial to the appearance, which is not well measured by those existing metrics. Therefore how to properly evaluate various environment matting and compositing methods warrants future investigation.

Acknowledgements

This work is supported by the ARC 9/09 Grant (MOE2008-T2-1-075) of Singapore.

References

- [AMKB04] AGARWAL S., MALLICK S. P., KRIEGMAN D. J., BELONGIE S.: On refractive optical flow. In *ECCV* (2) (2004), pp. 483–494.
- [BEN03] BEN-EZRA M., NAYAR S. K.: What does motion reveal about transparency? In *ICCV* (2003), pp. 1025–1032.
- [CM02] COMANICIU D., MEER P.: Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 24, 5 (2002), 603–619.
- [CZH*00] 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 *SIGGRAPH '00* (New York, NY, USA, 2000), ACM Press, pp. 121–130.
- [DB03] DANTE A., BROOKES M.: Precise real-time outlier removal from motion vector fields for 3D reconstruction. In *ICIP* (1) (2003), pp. 393–396.
- [DCZ09] DUAN Q., CAI J., ZHENG J.: Vector field fitting for real-time environment matting of transparent objects. In *ICIP* (1) (2009), pp. 393–396.
- [De 72] DE BOOR C.: On calculating with B-splines. *J. Approx. Theory* 6 (1972), 50–62.
- [Die81] DIERCKX P.: An algorithm for surface fitting with spline function. *IMA J. Numer. Anal.* 1 (1981), 267–283.
- [HH74] HAYES J. G., HALLIDAY J.: The least-squares fitting of cubic spline surface to general data sets. *J. Inst Math. Appl.* 14 (1974), 89–103.
- [KS08] KUTULAKOS K. N., STEGER E.: A theory of refractive and specular 3D shape by light-path triangulation. *Int. J. Comput. Vision* 76, 1 (2008), 13–29.
- [MKI04] MIYAZAKI D., KAGESAWA M., IKEUCHI K.: Transparent surface modeling from a pair of polarization images. *IEEE Trans. Pattern Anal. Mach. Intell.* 26, 1 (2004), 73–82.
- [MPZ*02] MATSIK W., PFISTER H., ZIEGLER R., NGAN A., McMILLAN L.: Acquisition and rendering of transparent and refractive objects. In *Rendering Techniques* (2002), pp. 267–278.
- [PD03] PEERS P., DUTRÉ P.: Wavelet environment matting. In *EGRW '03* (Aire-la-Ville, Switzerland, Switzerland, 2003), Eurographics Association, pp. 157–166.
- [PM90] PERONA P., MALIK J.: Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* 12, 7 (1990), 629–639.
- [SB96] SMITH A. R., BLINN J. F.: Blue screen matting. In *SIGGRAPH '96* (New York, NY, USA, 1996), ACM, pp. 259–268.
- [WFZ02] WEXLER Y., FITZGIBBON A. W., ZISSERMAN A.: Image-based environment matting. In *EGRW '02* (Aire-la-Ville, Switzerland, Switzerland, 2002), Eurographics Association, pp. 279–290.
- [YIX07] YAMAZAKI M., IWATA S., XU G.: Dense 3D reconstruction of specular and transparent objects using stereo cameras and phase-shift method. In *ACCV* (2) (2007), pp. 570–579.
- [ZWCS99] ZONGKER D. E., WERNER D. M., CURLESS B., SALESIN D. H.: Environment matting and compositing. In *SIGGRAPH '99* (New York, NY, USA, 1999), ACM Press, pp. 205–214.
- [ZY04] ZHU J., YANG Y.-H.: Frequency-based environment matting. In *PG '04* (Washington, DC, USA, 2004), IEEE Computer Society, pp. 402–410.