Normal Estimation of a Transparent Object Using a Video

Sai-Kit Yeung, Tai-Pang Wu, Chi-Keung Tang, Tony F. Chan, and Stanley J. Osher

Abstract—Reconstructing transparent objects is a challenging problem. While producing reasonable results for quite complex objects, existing approaches require custom calibration or somewhat expensive labor to achieve high precision. When an overall shape preserving salient and fine details is sufficient, we show in this paper a significant step toward solving the problem when the object's silhouette is available and simple user interaction is allowed, by using a video of a transparent object shot under varying illumination. Specifically, we estimate the normal map of the exterior surface of a given solid transparent object, from which the surface depth can be integrated. Our technical contribution lies in relating this normal estimation problem to one of graph-cut segmentation. Unlike conventional formulations, however, our graph is dual-layered, since we can see a transparent object's foreground as well as the background behind it. Quantitative and qualitative evaluation are performed to verify the efficacy of this practical solution.

Index Terms—Transparent object, normal estimation, graph-cuts, segmentation

1 Introduction

We address the problem of surface normal estimation for a transparent object, where the integrated surface is an overall shape of the target object that preserves salient features and fine structures if present on its exterior surface. Such detail-preserving exterior surface representation is adequate for vision and robotics applications where transparent objects such as glass bottles are to be grabbed by a robotic arm, or detected by a navigating robot for simultaneous localization and mapping in a cluttered scene. In our paper, our goal is different from photorealistic rendering or high-accuracy reconstruction of transparent objects, where custom equipment, calibrated and mechanical capture are often deemed necessary to achieve precision comparable to one of tracing complex refractive light-transport paths exhibited by the target object. On the other hand, when an adequate shape without this level of precision is sufficient, it is possible to propose a new approach that uses a simpler setup realizable using a smaller budget. Nowadays cameras are ubiquitous on mobile devices which further motivates our proposed lightweight approach.

Without expensive or complicated setup while still supporting a detail-preserving reconstruction, what visual cues concerning a transparent object can be utilized? Although some of us have had the unpleasant experience of smacking into a glass window without seeing it, we can still see a wide range of transparent objects despite their apparent transparency, because most of them *refract* and *reflect* incoming light. Tracing refractive light-transport paths using calibrated setup and capture had contributed to the success of techniques aiming at high-precision reconstruction. This paper on the other hand makes use of specularities *directly* reflected off an transparent object to produce an adequate normal estimation. Due to the low dynamic range of our inexpensive video camera, however, *indirect* reflection caused by complex light transport (e.g., caustics and total internal reflection) also produces strong highlights with intensity as strong as direct specular highlights. Thus, our main problem is to identify at each pixel the subset of collected highlights that are caused by direct specular reflection.

Technical overview. Our solution is based on the following working principle: given a fixed viewpoint video shot under dense lighting directions, a pixel location has a high probability to observe a specular highlight directly reflected off the surface point being observed, see Fig. 1. This can be used to obtain the surface normal from a specular object with known reference geometry. At first glance, given a dense collection of highlights at a pixel, direct application of orientation consistency similarly done in photometric stereo by examples [9] (where in our case each pixel receives a normal transferred from a known geometry) in a "winner-takes-all" or thresholded setting [7] for normal recovery would have solved our estimation problem. In fact, this example-based approach dates back to early works [13] and [27]. However, the problem proves to be challenging since our low-cost capture device (Fig. 2), which is similar to [7], is an off-the-shelf DV camera of limited dynamic range (Nikon D40), where highlights caused by direct or indirect reflections are likely be recorded with equally high intensities.

To make the problem tractable, we propose to use normal cues given by the object's silhouette, and sparse user markups for tracking true and rejecting false highlights for transferring normals from a reference shape [32]. The technical contribution consists of the optimal integration of these two normal cues for deriving the target normal map. It turns out that the optimization problem can be mapped into one similar to image segmentation and thus formulated into a graph-cut optimization. Our graph, on the other hand, is different from those in conventional graph-cut formulation: it is a dual-layered graph because we can observe a transparent object's foreground as well as the background behind it.

2 Related Work

For the state of the art in transparent and specular object reconstruction see the excellent survey [12] where most methods require precision or specialized capture and sophisticated ray tracing which are arguably difficult to repeat. Closely related methods mentioned in the survey are also reviewed here. Our graph-cut method is geared to reconstruct objects belonging to their class 3 and class 5, ideal or near ideal specular reflectance.

To employ specularities, in specular stereo [5] a two-camera configuration and image trajectory were used. A theory of specular surface was developed in [25], where the relationship between specular surface geometry and image trajectories was studied for classifying features as real or virtual. Virtual features, which are reflections by a specular surface not limited to highlights, contain useful information on the shape of the object. In [21], two views were used to model the surface of a transparent object, by making use of the optical phenomenon that the degree of polarization of the light reflected from the object surface depends on the reflection angle, which in turn depends on surface normal. This approach utilizing light polarization, where the light transport paths were ray-traced, was further explored in [19] where one camera was used. In [15], a theory was developed on refractive and specular 3D shape by studying the light transport paths, which are restricted to undergo no more than two refractions. Two views were used for dynamic refraction stereo [22] where the notion of refractive disparity was introduced. A reference pattern and refractive liquid were used. Scatter trace photography [23] was then proposed to reconstruct transparent objects made of inhomogeneous materials, by using

- *S.-K. Yeung is with the Pillar of Information Systems Technology and Design, Singapore University of Technology and Design, Singapore. E-mail: saikit@sutd.edu.sg.*
- *T.-P. Wu is with HK Jiuling Technology Co. Ltd. E-mail: tpwu@9ling.hk.*
- *C.-K. Tang and T.F. Chan are with the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong. E-mail: {cktang, tonyfchan}@ust.hk.*
- *S.J. Osher is with the Department of Mathematics, University of California, Los Angeles, CA. E-mail: sjo@math.ucla.edu.*

Manuscript received 12 Dec. 2012; revised 2 May 2014; accepted 27 May 2014. Date of publication 6 Aug. 2014; date of current version 3 Mar. 2015.
Recommended for acceptance by Y. Matsushita.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TPAMI.2014.2346195

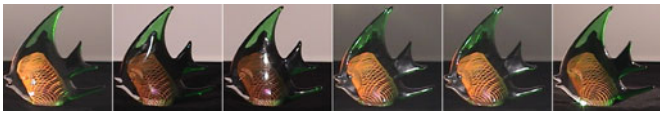

Fig. 1. Six images showing a typical input (example FISH).

a well-calibrated capture device to distinguish direct scatter trace from other indirect optical observations. In [10], a transparent object was immersed into fluorescent liquid. Computer-controlled laser projector and multiview scans were available for merging. A good review of various methods on reconstructing transparent objects can be found in [11]. While the recovered normal maps of mesostructures look good in [7] and were demonstrated to be useful for relighting, their assumption on specular highlight does not apply to general transparent objects that exhibit complex refraction phenomena (such as total internal reflection and caustics). Theories on shape from specular flow was developed [2] where differential equations were derived to relate the observed specular flow to the environment motion and smooth surface shape. Specular structure from motion [17] considers the geometric problem of estimating the poses of a reference plane in shape from specular correspondences. When imaged from a known background a height-field can be recovered from measured distortions in single or multiple images [28]. Our alternative solution makes use of rough initial shape (normals), sparse normal cues, and dense specular highlights, setting itself apart from the above approaches where mathematical theories were developed based on the physics of light transport, motion of sparse specularities, or simplifying assumptions were imposed on the transparent object.

3 Observations and Assumptions

Given a dense set of images of a solid transparent object captured by a static video camera under variable illumination, the problem is to estimate the normal map by utilizing specular highlights directly reflected off the surface.

The directly reflected specular highlight corresponds to the observed normals of the exterior surface, after applying orientation consistency using a reference shape of known geometry, which is a chrome sphere in our case. Transparent objects are more challenging, since indirect reflections caused by complex light transport and caustics can also produce highlight as strong as direct specular reflection. Using our inexpensive capture system and under the orthographic camera assumption, we have the following key observations:

Data capture range. The light source must be facing toward the transparent object. Using a chrome sphere (Fig. 3) as a reference geometry, surface normals falling outside the 45-degree realm measured from the upward direction cannot be recovered, Fig. 4,

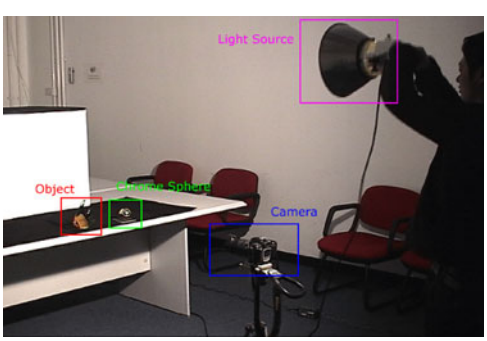

Fig. 2. Our budget setup. The spotlight is moved around to simulate a distant light source with constant radiance. The video camera is an off-the-shelf DV of limited dynamic range (Nikon D40). The transparent object and the reference chrome sphere are captured at the same time.

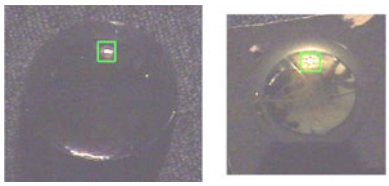

Fig. 3. Orientation consistency in the presence of transparency and indirect illumination effects. The reference sphere and the object have different material properties but they act like ideal specular objects when they reflect specular highlights (brightness and contrast were enhanced for visualization).

which can be explained using the law of specular reflection under orthographic projection.

Highlight appearance. Direct specular highlight is bright and concentrated (Fig. 5a). However, some non-specular indirect reflection observed will also be classified as highlight, which happens when caustics and total internal reflection behind the surface appear as bright as specular reflection, especially a low-dynamic-range digital video camera is used in image capture (Fig. 5b). Obvious true or false highlights can be marked up by user for disambiguation if needed. Given the limited data capture range, some pixels will have no highlight, or the observed intensity is not high enough to be considered a highlight (Fig. 5c).

Normal clusters. The normals transferred from the chrome sphere after applying orientation consistency will form distinctive clusters at a pixel over the time frames (Fig. 6), one of which corresponds to direct specular highlights whereas the rest are due to other types of illumination. While this paper investigates a limited number of (six) objects, in all of our experiments, we found that almost all pixel locations observe at most two *salient* normals clusters over all the frames captured. A third cluster is usually too weak to be detected as highlight. This two-cluster observation, however, is not critical: as long as the strongest is detected or identified as direct specular highlight while ignoring the non-salient ones, the exact number of clusters does not matter.

Thus, an implicit assumption is that if only one peak normal is present at a pixel, it must be due to direct transport. This assumption can be violated in practice, typically at caustics. While under a global optimization framework this problem can be ameliorated to some extent, we allow the user to mask off these false highlights which are arguably easy to mark.

4 Normal Optimization

We propose to solve the optimization problem by integrating the rough shape given by the normals on an object's silhouette as well as the normals transferred via orientation consistency. To resolve severe highlight ambiguity as stated in the previous section, a limited amount of user specification is employed. This improves not only the initialization but also the optimization process. The optimization can be mapped to one similar to image segmentation and solved via graph-cut optimization.

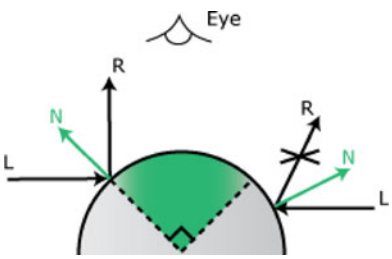

Fig. 4. Under the orthographic camera assumption, we can directly obtain the normal orientation from the reference sphere. Notice that given a single view, only a subset of normals can be computed. (L is light direction, R reflection direction, N is normal direction.)

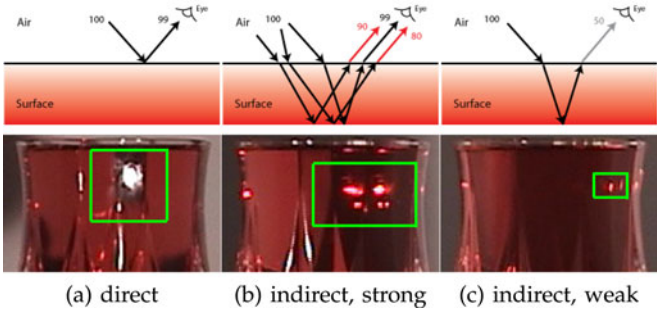

Fig. 5. Three typical scenarios on the observed light emanating from a point on the exterior surface of a transparent and refractive object. The numbers in the figure indicate intensity magnitudes. Orientation consistency should be applied to (a), but not to (b) or (c) because they are not specular reflections. True/false highlights can be marked easily.

4.1 Initialization

The limited data capture range due to the orthographic camera assumption does not allow normals outside the 90-degree realm to be transferred to regions around the object's silhouette. On the other hand, normals are assumed to be perpendicular to an object's silhouette, which can be readily available using [16] on any video frame (all of which are shot using a static video camera) that is reasonably lit. A rough overall shape can thus be derived using these silhouette normals, with the caveat that the true surface cannot deviate significantly from the resulting shape-from-gradients [24], where dense gradients (normals) are interpolated from the normals on the object's silhouette.

To improve the quality of initial normals, the user may indicate true/false highlights on keyframes for automatic tracking. Example selections are depicted in Fig. 7. Normals from a chrome sphere are transferred to regions where true highlights are tracked. These normals, together with the silhouette normals, serve as hard constraints for sparse data interpolation: we compute the surface normal $n^{\mathcal{F}}$ for all pixels within the object's silhouette via a MRF formulation for which a stable implementation is available [24]. In essence this method solves Poisson equations with boundary conditions and compares favorably with known related methods reported in [3].

From Figs. 6 and 7, we see that it is quite easy for a user to identify and mark direct specular highlights on smooth surface, and that the subsequent automatic tracking is also stable. For complex shapes, marking up highlights is more tedious and tracking is more challenging, or there will be no markups at all—false highlight, and hence the corresponding normal, can be rejected if it deviates significantly (e.g. 90 degree) from the normal at that pixel obtained above.

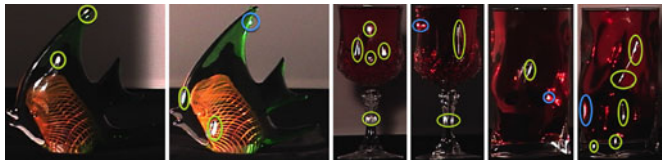

Fig. 7. User can select the regions for direct (green circles) or indirect highlights (blue circles). Pixels within the region with intensity higher than threshold will be labeled accordingly.

4.2 Normal Refinement by Graph Cuts

The next step is to integrate the initial normals with the information given by the dense highlight collected to optimize the normals. As discussed in the previous section, given a total of T image frames, the possible situations at a given pixel are: 1) no highlight, 2) highlights form a single normal cluster, 3) highlights form two or more normal clusters. So our problem is translated into a labeling problem, that is, given a normal cluster, determine if it corresponds to one that produces direct specular reflection on the exterior surface. Our idea is to utilize the data measurement given by applying orientation consistency to refine the initial surface normals. The initial shape also gives us relevant cues for automatically rejecting wrong measurements due to false highlights.

Normal clustering. The shape optimization problem can be posed as a binary labeling problem, by assigning every cluster as exterior surface normal or otherwise. Since each pixel can observe direct and indirect reflections, we adopt the following method to extract the two representative normals: Given T observations per pixel, we threshold the pixel intensities to discard weak intensities. It results in M usable observations, where $0 \leq M \leq T$. If $M = 0$, it means that this pixel contains no useful information and the initial normal will be used instead. Otherwise, we will perform clustering on the observed normals to find salient clusters. We set a simple threshold where the angle spanned by two normals larger than 15 degree are considered to belong to different clusters.

The process with two clusters is illustrated in Fig. 8. In practice, there may be more than two clusters at a given pixel. However, we find that such situation is very rare and in addition, they are not salient. Typically these non-salient clusters are only formed by observations from one to two time frames. Therefore, to simplify the main optimization process, we only pick the first two clusters which have the highest statistical significance, i.e., the two clusters with the most normal observations. The number of clusters formed using this threshold is shown in Fig. 9, showing that in most of the cases there are zero to two clusters. The situation with more than two clusters (shown in white) is relatively rare.

Graph formulation. Next, we construct a graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$, where $\mathcal{V}$ is the set of all nodes and $\mathcal{E}$ is the set of all edges connecting adjacent nodes. In our case, the nodes contain labels to the two most

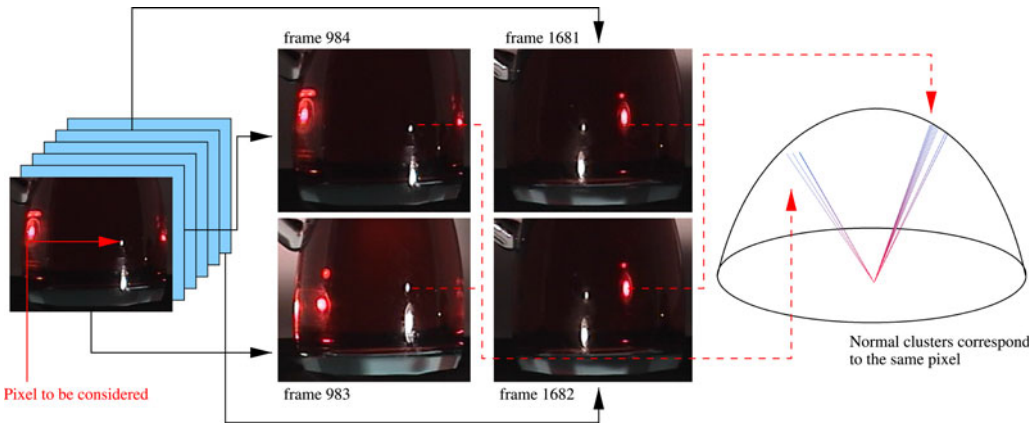

Fig. 6. The normal clusters at selected pixels over time.

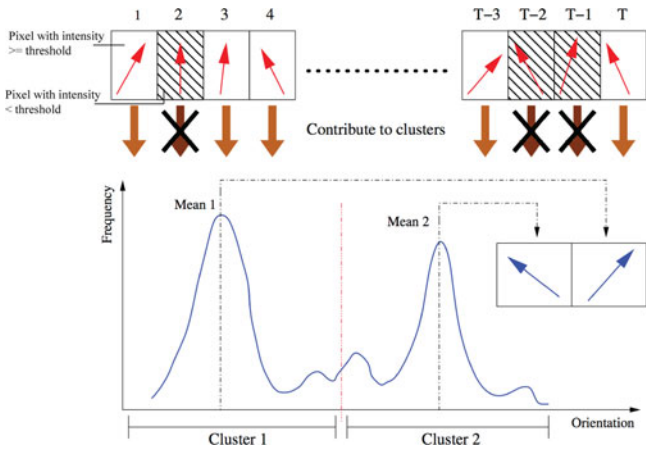

Fig. 8. For every pixel location, we cluster the transferred normals into two salient clusters given the T image frames by their intensity.

salient normals clustered, and the edges represent adjacency relationships. The graph can have up to $2N$ nodes for N processing pixels, and every node can have nine edges, as shown in Fig. 10a. For pixel locations with only one cluster, we duplicate the cluster in the other node to simplify the implementation (Fig. 10b).

The labeling problem is to assign a unique label s_i for the clustered normal at each node $i \in \mathcal{V}$, i.e., $s_i \in$ {exterior surface normal $(=1)$, otherwise $(=0)$}. The solution $\mathbf{S} = \{s_i\}$ corresponding to the final frontal surface normals can be obtained by minimizing a Gibbs energy $E(\mathbf{S})$ [16]:

$$E(\mathbf{S}) = \sum_{i \in \mathcal{V}} E_1(s_i) + \lambda \sum_{\{i,j\} \in \mathcal{E}} E_2(s_i, s_j), \tag{1}$$

where $E_1(s_i)$ is the likelihood energy, denoting the cost when the label of node i is s_i, and $E_2(s_i, s_j)$ is the prior energy, encoding the cost when the labels of adjacent nodes i and j are s_i and s_j respectively. We set λ between 0.1-0.05 in our experiment. Our problem becomes a graph-cut energy minimization problem [14] and is very similar to the image segmentation problem using graph cuts [16], [26]. An analogy of our problem to image segmentation is shown in Table 1.

The main difference is that in image segmentation there is only one observation (color) per pixel. But in our case the exterior and non-exterior surface(s) can be observed at the same time as the object is transparent. While other choices may exist we believe modeling our graph as a dual-layered one as shown in Fig. 10a is a natural choice.

Not surprisingly, graph cuts approach is also adopted to solve the conventional photometric stereo problem [6], [20], [31] for opaque objects.

Likelihood energy and counter normal. We first summarize the major notations in Table 2. In Eq. (1), E_1 encodes at each node the normal similarity between the clustered normal observed from the collected highlights and the initial normal. However, simply using the initial normals from silhouette is not adequate to encode E_1, since the information on the non-exterior surface should also be

Fig. 9. [color figure] Color map showing the number of clusters for each example. Gray: No cluster. Blue: 1 cluster. Purple: 2 clusters. White: more than 2 clusters. Refer to the text for cluster formation.

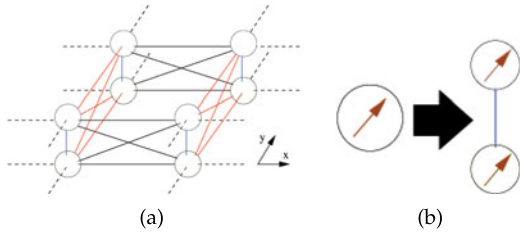

Fig. 10. (a) A dual-layered graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ is built where each pixel has two nodes. $\mathcal{V}$ is the set of all nodes representing the clustered normals, $\mathcal{E}$ is the set of all edges connecting adjacent nodes. (b) For pixels with only one cluster, one can duplicate the cluster to simplify the implementation. The final graph can have up to $2N$ nodes for N processing pixels. Every node can have up to nine edges connected to their neighboring nodes. Pixel locations with no observation will be completed by their initial estimation.

considered. This is analogous to image segmentation where color cues from foreground and *background* should be available for processing. We introduce the idea of *counter normal* which is a normal dissimilar to the initial normal, in order to define the affinity and hence the energy of assigning an observed normal to non-exterior surface.

The idea is as follows: If an observed (clustered) normal n_i does not belong to the exterior surface, it will not be similar to the initial normal $n_i^{\mathcal{F}}$. In other words, we may say n_i is similar to a normal that is *dissimilar* to the initial normal. Let us call it *counter* normal. Since our captured normals must be facing upward, we do not consider normals with negative z component. Therefore, we set the corresponding counter normal $n_i^{\mathcal{C}}$ by flipping the gradient of $n_i^{\mathcal{F}}$, i.e. $\{p_i^{\mathcal{C}}, q_i^{\mathcal{C}}\} = \{-p_i^{\mathcal{F}}, -q_i^{\mathcal{F}}\}$, where $p_i^{\mathcal{C}}$ and $q_i^{\mathcal{C}}$ are the x and y gradient of $n_i^{\mathcal{C}}$, and $p_i^{\mathcal{F}}$ and $q_i^{\mathcal{F}}$ are the x and y gradient of $n_i^{\mathcal{F}}$. Basically, this strategy adopts the most dissimilar surface normal (with z-axis as the reference) with upward direction as the counter normal.

With the initial normal $n_i^{\mathcal{F}}$ and the counter normal $n_i^{\mathcal{C}}$, we can now define our energy term E_1. For each node i, we compute the difference of the corresponding gradients with the frontal gradient and the counter (non-frontal) gradient by $d_i^{\mathcal{F}} = |p_i - p_i^{\mathcal{F}}| + |q_i - q_i^{\mathcal{F}}|$ and $d_i^{\mathcal{C}} = |p_i - p_i^{\mathcal{C}}| + |q_i - q_i^{\mathcal{C}}|$ respectively. Notice that for pixel locations with no highlight observation, we set the gradients at each of these pixels to be the initial $p_i^{\mathcal{F}}$ and $q_i^{\mathcal{F}}$, and set the data energy $E_1(0) = E_1(1) = 0.5$ so the optimal labeling problem will only be governed by the smoothness term but not the data term at these locations. It also applies to the situation when the initial gradient is zero, that is, the normal is pointing upward. Therefore, $E_1(s_i)$ can be defined as following:

$$\begin{cases} E_1(s_i=1) = 0 & E_1(s_i=0) = \infty & \forall i \in \mathcal{F}, \\ E_1(s_i=1) = \infty & E_1(s_i=0) = 0 & \forall i \in \mathcal{C}, \\ E_1(s_i=1) = \frac{d_i^{\mathcal{F}}}{d_i^{\mathcal{F}} + d_i^{\mathcal{C}}} & E_1(s_i=0) = \frac{d_i^{\mathcal{C}}}{d_i^{\mathcal{F}} + d_i^{\mathcal{C}}} & \forall i \in \mathcal{U}_1, \\ E_1(s_i=1) = 0.5 & E_1(s_i=0) = 0.5 & \forall i \in \mathcal{U}_2, \end{cases} \tag{2}$$

where $\mathcal{U}_1$ and $\mathcal{U}_2$ are the set of nodes from region with and without normal observation, respectively, and $\{\mathcal{U}_1 \cup \mathcal{U}_2\} = \mathcal{V} - \{\mathcal{F} \cup \mathcal{C}\}$ is the set of uncertain nodes to be labeled. Eq. (2) is similar to [16]

TABLE 1
Analogy of Our Problem to Image Segmentation (e.g., [16], [26])

	NORMAL OPTIMIZATION	IMAGE SEGMENTATION
Labels	Exterior/ Interior	Foreground/ Background
Nodes	Clustered normals	Pixels
Processing token	Gradients	Colors

TABLE 2
Summary of the Notations in This Paper

Notation	Definition
i, j	node index
s_i	label index (0 or 1)
n_i	clustered normals from video data
p_i	clustered gradients from video data
$n_i^{\mathcal{F}}$	initial surface normals
$n_i^{\mathcal{C}}$	initial counter normals (computed from $n_i^{\mathcal{F}}$)
$\mathcal{F}$	set of nodes labeled by user as exterior surface
$\mathcal{C}$	set of nodes labeled by user as non-exterior surface
$\mathcal{U}_1$	set of nodes with normal observation
$\mathcal{U}_2$	set of nodes without normal observation

except for $i \in \mathcal{U}_2$. $\mathcal{F}$ is the set of nodes labeled by the user as exterior surface normals, which are available in the initialization step (Section 4.1). $\mathcal{C}$ is the set of nodes corresponding to non-exterior normal observations, which are specified by the user in a similar manner as the exterior surface normals, for instance, by marking up false instead of true highlights. In other words, $\mathcal{F}$ and $\mathcal{C}$ correspond to the hard constraints. $\mathcal{F}$ and $\mathcal{C}$ are the bright pixels within the selected green and blue regions in Fig. 7 respectively. $\mathcal{U}_1$ corresponds to blue, purple and white pixels and $\mathcal{U}_2$ corresponds to gray pixels in Fig. 9.

Notice that since our graph is dual-layered, nodes in $\mathcal{F}$ and $\mathcal{C}$ can have the same pixel location. This is the main difference in the graph construction as compared to works in image segmentation [16], [26] where nodes in $\mathcal{F}$ and $\mathcal{C}$ must have different locations. The first two equations guarantee the nodes in $\mathcal{F}$ or $\mathcal{C}$ always have the label consistent with user inputs. The last two equations define the data energy for pixels with and without normal observations. When there is only one normal cluster at a particular pixel, both nodes at that pixel compute E_1 using the same normal cluster. When there is no normal observation, we assign the initial normal $n_i^{\mathcal{F}}$ to both nodes at that pixel and set $E_1(0) = E_1(1) = 0.5$. This means that for pixel with no normal observation, the node values will be contributed to the smoothness term E_2 only. See Table 3 for the node values at a pixel under different cases of normal observation.

Prior energy. We use E_2 to encode the smoothness constraint between neighboring nodes. Define the normal similarity function between two nodes i and j as an inverse function of the smoothness constraint:

$$E_2(s_i, s_j) = |s_i - s_j| \cdot g(|p_i - p_j| + |q_i - q_j|), \tag{3}$$

where $g(\xi) = (\xi + 1)^{-1}$. Note that $|s_i - s_j|$ captures the smoothness information when the adjacent nodes have different labels. In other words, E_2 is a penalty term when neighboring nodes are assigned with different labels. So if the neighboring normals are similar, assigning them with different labels will increase the energy of the graph and vice versa. This energy term encourages integrable normals to be grouped into the same surface. If we ignore the E_2 term in Eq. (1), minimizing the energy will produce a "winner-takes-all" labeling strategy based on normal similarity (E_1) only.

The graph cuts optimization is performed once. We use the max-flow algorithm [14] to minimize the energy $E(\mathbf{S})$ in Eq. (1). Readers may notice that we do not enforce the two nodes at the same pixel location (recall our graph is dual-layered) to have different labels in our graph formulation. Although in our examples, pixel originally having two clusters will always have one labeled as exterior surface normal, we cannot guarantee that the two clusters may both come from false highlights.

Final estimation. After graph cuts, we recompute the surface normals of the whole object by solving the Poisson equations with hard constraint from $\mathcal{F}$, object's silhouette and additional hard constraint from the labeled exterior normals. Pixel locations with no observation will thus receive improved surface normals by the additional hard constraint. See Table 4 for the hard constraints at different stages.

TABLE 3
Node Values at a Given Pixel (i and j Are the Pixel's Neighboring Nodes in the Top/Bottom Layers Shown in Fig. 10)

# of clustered normals	Node values
2	node i = normal cluster 1
	node j = normal cluster 2
1	node i = normal cluster 1
	node j = normal cluster 1
0	node i = node $j = n_i^{\mathcal{F}}$

TABLE 4
Hard Constraints for Estimating the Surface Normals at Different Stages

Stage	Hard Constraints
Initialization	1) Initial surface normals $n_i^{\mathcal{F}},\ \forall i \in \mathcal{F}$
	2) Normals along the object's silhouette
Final Estimation	1), 2) and labeled exterior normals in $\mathcal{U}_1$

5 Experimental Results

The data sets tested and running times of normal map estimation are summarized in Table 5. The running times shown exclude those of surface integration, where we again use the source codes from [24] (which enforces the integrability constraint in the continuous domain) to produce the final surfaces. While analytical methods can be used to solve the Poisson equations [30] which is possible in [24] since the solution is linear, we note the system is large where in [24] over-relaxation was adopted for speed-ups.

Setup. To capture a dense image set, similar in fashion to [7], [9], [31], we use an off-the-shelf DV camera with a fixed viewpoint to simultaneously capture the reference chrome sphere and the target object. A moving spotlight is used to mimic a distant light source at varying directions. The spotlight is swiped multiple times over the object to achieve a better coverage, similar to the procedure in [31].

Quantitative evaluation. To evaluate the performance of our system, we capture two real data sets whose analytical geometries are known (namely, a hemisphere and a cylinder). They can be served as the ground truth for quantitative comparison.

The spotlight used in our experiments is not a perfect directional source, and the classroom we used in conducting the experiments is not a dark room either. Note that the chrome sphere is captured simultaneously with the transparent object.

TABLE 5
Running Times Are Measured on a Desktop Computer with Dual-Core 2.6 GHz CPU and 3.0 GB RAM

	no. of images	image dimensions	total running time (sec)
Sphere	8,000	251 × 221	31.6
Cylinder	4,226	146 × 384	35.3
Jug	3,870	279 × 419	50.5
Fish	1,512	259 × 252	10.9
Wine Glass	3,133	153 × 324	13.2
Water Glass	3,383	153 × 279	15.1

TABLE 6
The Mean Errors of the Computed Surface Normals Using Our Method and [7] Are Shown (Where the Error Is Defined as Sum of the Squared Difference of Three Normal Components)

	Sphere		Cylinder	
	data region	whole region	data region	whole region
Our method	0.096	0.145	0.071	0.106
"winner-takes-all"	0.738	0.768	0.386	0.406

Data region consists of pixel locations with highlight observations. Whole region includes all processing pixels.

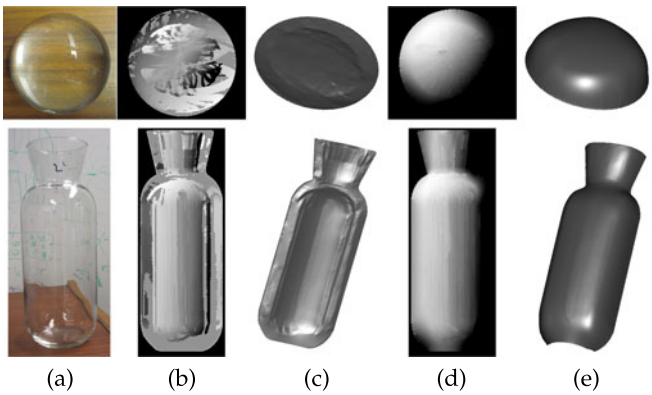

Fig. 11. Objects with known geometry (Sphere and Cylinder) for quantitative evaluation. (a) Photo of the object. (b)-(c) Lambertian-shaded normal map from [7] and integrated surface. (d)-(e) Our normal map and surface.

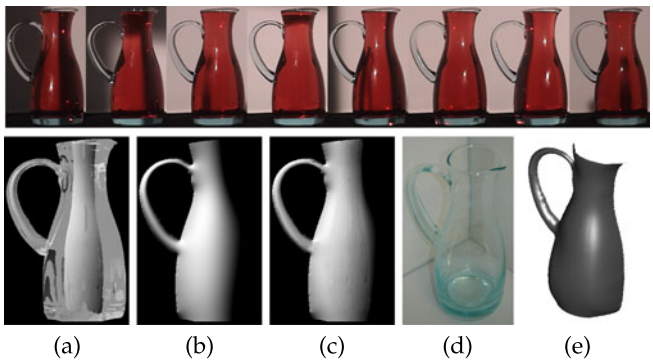

Fig. 12. Jug. The top shows eight captured images. The bottom shows the recovered normal maps $\mathbf{N}$, displayed as $\mathbf{N} \cdot \mathbf{L}$ with $\mathbf{L} = (-\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})^T$. (a) Normal map by [7]. (b) Initial normal map by [24] from the object silhouette. (c) Normal map using our method. (d) and (e) compare the reconstructed surfaces from our normal maps with the real object.

Given that the normals are transferred using an example-based approach similarly done in [9], surface normals with the same orientation appears to be the same under the same illumination condition which include global illumination effect as well.

We compute the average difference of our computed surface normals with Sphere and Cylinder respectively. The results are depicted in Table 6. We also generate the surface normals by using a winner-takes-all strategy such as [7]. We implemented this strategy as the sum of second order moments followed by eigen-decomposition. The corresponding visual results are shown in Fig. 11. From the results, we can see that our final reconstructed shape is more faithful compared with the one using the winner-takes-all strategy. Without proper labeling, the normals from the true highlight will be mixed up with the wrong normals transferred due to false highlights, making the final optimized normals fail to integrate into a reasonable surface as shown in Fig. 11c.

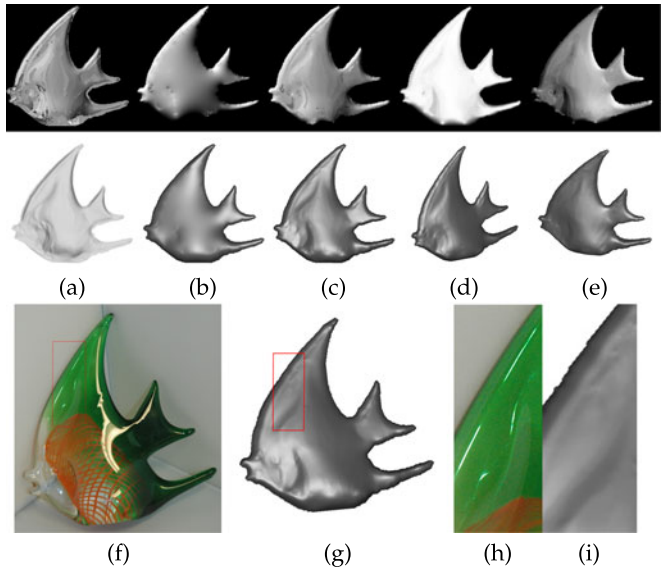

Fig. 13. Fish. The top shows the normal maps $\mathbf{N}$ displayed as $\mathbf{N} \cdot \mathbf{L}$. The lighting directions are: (a)-(c) $\mathbf{L} = (-\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})^T$, (d) $\mathbf{L} = (0, 0, 1)^T$, (e) $\mathbf{L} = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})^T$. The middle row shows the reconstructed surface. (a) "winner-takes-all" strategy [7]. (b) Initial normal map by [24] from the user-supplied normal cues and silhouette. (c) Our result in frontal view. (d)-(e) Our results in other views. The bottom shows the comparison of the reconstructed surface with the real object at similar viewpoint. (h) Zoom-in view of (f). (i) Zoom-in view of (g). The figurine is a solid transparent object with complex colors inside the object.

Qualitative evaluation. Fig. 12 shows the reconstruction result on a transparent glass Jug. This example is simple as the initial shape is already close to the final solution. We show one view of the surface reconstruction alongside with the real object in a similar view.

Fig. 13 shows the results on a glass figurine Fish, which is very similar to the one used in [23] and our result looks comparable. Note that the image sequence contains a lot of shadows and highlight. The figurine contains internal structures of various colors. However, it is highly specular which produces sufficient specular observations. The zoom-in views illustrate the details preserved in our reconstruction of the object's exterior surface. Although there are some errors due to shadows, overall the surface reconstruction result is quite robust to the complex color, texture and internal structure. The initial shape in (b) lacks the level of details in comparison with our final surface result shown in (c) and other views in the figure.

We tested our approach using two complex transparent objects Water Glass and Wine Glass. Fig. 14 shows the reconstruction result on a transparent Water Glass which has a complex shape. The result is acceptable in regions where salient highlights can be detected. Note the faithfulness of the recovered shape using our approach where true highlights are picked and false ones are suppressed. Fig. 15 shows another result on a transparent Wine Glass where a lot of orientation discontinuities are present. Note in the normal map the fine details of the wine glass preserved. Two views of the reconstructed surface, alongside with the real object in similar views are shown. Note that a similar glass was also used in fluorescent immersion range scanning [10]. The surface generated using our budget setup looks reasonable and comparable in many ways.

6 Conclusion, Discussion and Future Work

This paper presents a practical approach for estimating the normal map of the exterior surface of a transparent object. While inadequate for high-precision graphics rendering, our detail-preserving output is a faithful reconstruction of the transparent object, as demonstrated by our convincing results, and is potentially useful in a range of vision applications. Our approach makes use of an initial

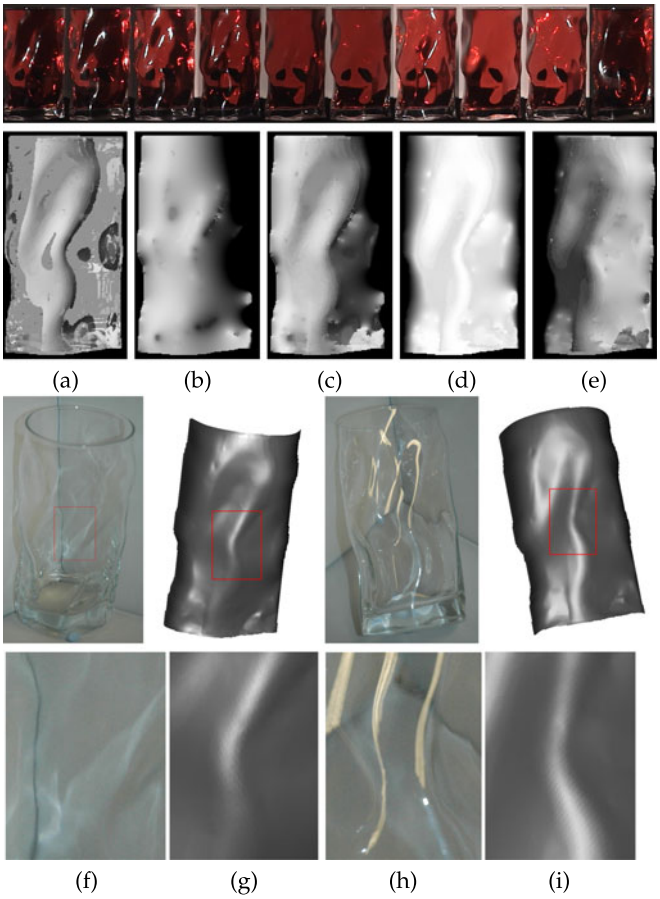

Fig. 14. WATER GLASS. The top shows 10 captured input images and normal maps $\mathbf{N}$ displayed as $\mathbf{N} \cdot \mathbf{L}$: The lighting directions respectively are: (a)-(c) $\mathbf{L} = (-\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})^T$, (d) $\mathbf{L} = (0, 0, 1)^T$, (e) $\mathbf{L} = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})^T$. (a) "winner-takes-all" strategy [7]. (b) Initial normal map by [24] from the user-supplied normal cues and silhouette. (c)-(e) our results. (f)-(i) show the comparison of our reconstructed surfaces with the real object at novel viewpoints with the corresponding zoom-in views.

shape derived from the object's silhouette, normals transferred using orientation consistencies, and sparse user markups if needed. The problem is translated into one similar to image segmentation, and the optimization can be formulated using graph cuts on a dual-layered graph. This alternative approach is desirable for quick 3D prototyping of existing transparent objects with details adequately preserved. Our current system produces a normal map for the exterior surface of a transparent object, from which a depth map can be integrated.

Our approach works for the objects which are largely convex and reflect light as ideal specular objects, for example, common transparent objects with no anti-reflective coating. For concave objects, we assume the concavity is shallow (e.g., 130 degree dent) so that the reflection of the light can be observed. Colored objects are easier for false highlight identification but they are not necessary for our approach. Interestingly, we found that for transparent objects with a complex shape, the number of ambiguous pixels is actually less than those with a simple shape such as a hemispherical lens. For complex objects, orientation consistency from direct highlights often suffices for obtaining good normals. For simple objects, the initial shape can serve as a good prior to reject false highlights. While the relationship between the number of ambiguous pixels and the surface shape is a topic for more investigation, a plausible explanation is that a simple transparent objects act as a lens: indirect highlights directly behind are as bright as direct highlights due to the lens convergent effect. We would also investigate the limit of our approach on the concavity from simple objects. As

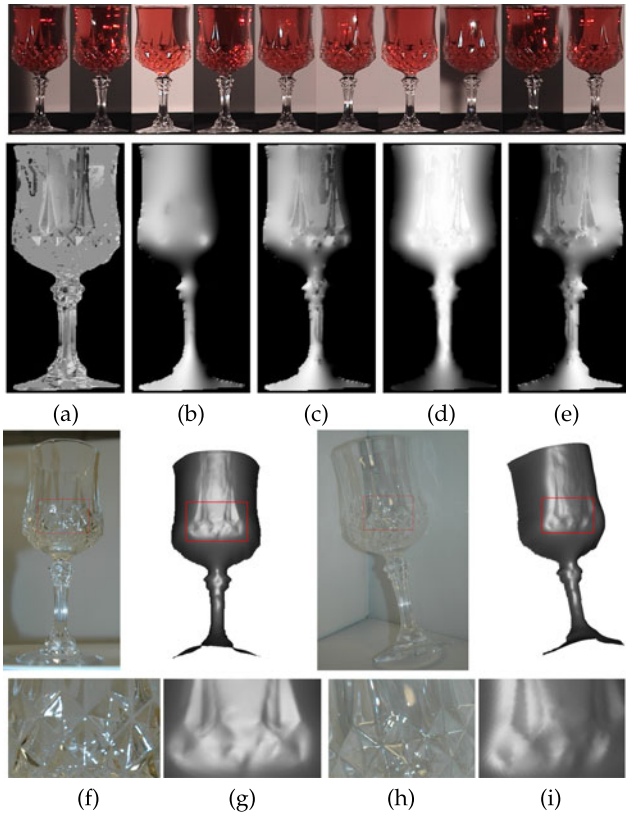

Fig. 15. WINE GLASS. See Fig. 14 caption.

discussed in [29], given sufficiently large number of lighting directions, a simple BRDF model could be derived for low-frequency component of general isotropic materials. Modeling the transparent reflectance on low-frequency domain might be an interesting direction as well. We are also interested in reconstructing transparent objects using outdoor photometric stereo approaches [1], [34] or using new sensing devices such as RGBD cameras for additional shape priors [4], [33].

To generate a full 3D reconstruction, a standard implementation consists of multiview capture of overlapping views of normal maps as described in the paper, followed by computing the visual hull using [18] which provides an overall 3D shape of the exterior surface. Finally, view dependent texture mapping [8] can be applied to project and map overlapping views of normal maps onto the visual hull.

ACKNOWLEDGMENTS

The authors would like to thank the Associate Editor and all of the anonymous reviewers. Special thanks to Reviewer 1 for his/her helpful and detailed comments throughout the review cycle. Sai-Kit Yeung is supported by Singapore University of Technology and Design (SUTD) StartUp Grant ISTD 2011 016, SUTD-ZJU Collaboration Research Grant 2012 SUTD-ZJU/RES/03/2012, SUTD-MIT International Design Centre Grant IDG31300106 and Singapore MOE Academic Research Fund MOE2013-T2-1-159. Chi-Keung Tang is supported by the Research Grant Council of the Hong Kong Special Administrative Region under grant no. 619112.

REFERENCES

[1] J. Ackermann, F. Langguth, S. Fuhrmann, and M. Goesele, "Photometric stereo for outdoor webcams," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2012, pp. 262–269.

[2] Y. Adato, Y. Vasilyev, T. Zickler, and O. Ben-Shahar, "Shape from specular flow," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 2054–2070, Nov. 2010.

[3] A. K. Agrawal, R. Raskar, and R. Chellappa, "What is the range of surface reconstructions from a gradient field?" in *Proc. Eur. Conf. Comput. Vis.*, 2006, vol. 1, pp. 578–591.

[4] J. T. Barron and J. Malik, "Intrinsic scene properties from a single RGB-D image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2013, pp. 17–24.

[5] A. Blake and G. Brelstaff, "Specular stereo," in *Proc. 9th Int. Joint Conf. Artif. Intell.*, 1985, pp. 973–976.

[6] M. Chandraker, S. Agarwal, and D. Kriegman, "Shadowcuts: Photometric stereo with shadows," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Minneapolis, MN, USA, 2007, pp. 1–8.

[7] T. Chen, M. Goesele, and H.-P. Seidel, "Mesostructure from specularity," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2006, vol. 2, pp. 1825–1832.

[8] P. E. Debevec, "Modeling and rendering architecture from photographs," Ph.D. dissertation, Comput. Sci. Division, Univ. California at Berkeley, Berkeley, CA, USA, 1996.

[9] A. Hertzmann and S. Seitz, "Example-based photometric stereo: Shape reconstruction with general, varying BRDFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1254–1264, Aug. 2005.

[10] M. B. Hullin, M. Fuchs, I. Ihrke, H.-P. Seidel, and H. P. A. Lensch, "Fluorescent immersion range scanning," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 87:1–87:10, Aug. 2008.

[11] I. Ihrke, K. Kutulakos, H. Lensch, M. Magnor, and W. Heidrich, "Transparent and specular object reconstruction," *Comput. Graph. Forum*, vol. 29, no. 8, pp. 2400–2426, Dec. 2010.

[12] I. Ihrke, K. N. Kutulakos, H. P. A. Lensch, M. Magnor, and W. Heidrich, "State of the art in transparent and specular object reconstruction," in *Proc. STAR Proc. Eurographics*, 2008, pp. 87–108.

[13] K. Ikeuchi, "Determining surface orientations of specular surfaces by using the photometric stereo method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. TPAMI-3, no. 6, pp. 661–669, Nov. 1981.

[14] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 147–159, Feb. 2004.

[15] K. Kutulakos and E. Steger, "A theory of refractive and specular 3d shape by light-path triangulation," in *Proc. Int. Conf. Comput. Vis.*, 2005, pp. 1448–1455.

[16] Y. Li, J. Sun, C.-K. Tang, and H.-Y. Shum, "Lazy snapping," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 303–308, 2004.

[17] M. Liu, K.-Y. K. Wong, Z. Dai, and Z. Chen, "Pose estimation from reflections for specular surface recovery," in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 579–586.

[18] W. Matusik, C. Buehler, R. Raskar, S. J. Gortler, and L. McMillan, "Image-based visual hulls," in *Proc. SIGGRAPH*, 2000, pp. 369–374.

[19] D. Miyazaki and K. Ikeuchi, "Shape estimation of transparent objects by using inverse polarization ray tracing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 11, pp. 2018–2030, Nov. 2007.

[20] D. Miyazaki and K. Ikeuchi, "Photometric stereo using graph cut and m-estimation for a virtual tumulus in the presence of highlights and shadows," *IEEE Comput. Society Conf. Comput. Vision Pattern Recog. Workshops (CVPRW)*, pp. 70–77, Jun. 2010.

[21] D. Miyazaki, M. Kagesawa, and K. Ikeuchi, "Transparent surface modeling from a pair of polarization images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 73–82, Jan. 2004.

[22] N. Morris and K. Kutulakos, "Dynamic refraction stereo," in *Proc. Int. Conf. Comput. Vis.*, 2005, pp. 1573–1580.

[23] N. Morris and K. Kutulakos, "Reconstructing the surface of inhomogeneous transparent scenes by scatter trace photography," in *Proc. Int. Conf. Comput. Vis.*, 2007, pp. 1–8.

[24] H.-S. Ng, T.-P. Wu, and C.-K. Tang, "Surface-from-gradients without discrete integrability enforcement: A Gaussian kernel approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 2085–2099, Nov. 2010.

[25] M. Oren and S. Nayar, "A theory of specular surface geometry," *Int. J. Comput. Vis.*, vol. 24, no. 2, pp. 105–124, Sept. 1997.

[26] C. Rother, V. Kolmogorov, and A. Blake, ""GrabCut": Interactive foreground extraction using iterated graph cuts," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 309–314, 2004.

[27] A. Sanderson, L. Weiss, and S. Nayar, "Structured highlight inspection of specular surfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 10, no. 1, pp. 44–55, Jan. 1988.

[28] Q. Shan, S. Agarwal, and B. Curless, "Refractive height fields from single and multiple images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2012, pp. 286–293.

[29] B. Shi, P. Tan, Y. Matsushita, and K. Ikeuchi, "Bi-polynomial modeling of low-frequency reflectances," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 36, no. 6, pp. 1078–1091, Jun. 2014.

[30] T. Simchony, R. Chellappa, and M. Shao, "Direct analytical methods for solving poisson equations in computer vision problems," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 5, pp. 435–446, May 1990.

[31] T. Wu, K. Tang, C. Tang, and T. Wong, "Dense photometric stereo: A markov random field approach," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 28, no. 11, pp. 1830–1846, Nov. 2006.

[32] S.-K. Yeung, T.-P. Wu, C.-K. Tang, T. F. Chan, and S. Osher, "Adequate reconstruction of transparent objects on a shoestring budget," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2011, pp. 2513–2520.

[33] L.-F. Yu, S.-K. Yeung, Y.-W. Tai, and S. Lin, "Shading-based shape refinement of RGB-D images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2013, pp. 1415–1422.

[34] L.-F. Yu, S.-K. Yeung, Y.-W. Tai, D. Terzopoulos, and T. F. Chan, "Outdoor photometric stereo," in *Proc. IEEE Int. Conf. Comput. Photography*, 2013, pp. 1–8.