Synthesizing Reflections of Inserted Objects

Xiaochun Cao and Hassan Foroosh
Computational Imaging Lab, University of Central Florida
{xccao, foroosh}@cs.ucf.edu

Abstract

The aim of reflection synthesis of inserted objects is to generate reflections which would be seen by the same camera capturing the target scene and be reflected by the true reflective media in the target scene. This problem is inherently difficult because we are typically given only several, sometimes even just one, views of the target scene. In this paper, we explore the geometric constraints to synthesize geometrically correct reflections. We also demonstrate how to constraint the synthesized reflections to be photometrically consistent with those in the original target views. The proposed method, therefore, advances the current image-based compositing techniques one step further toward the scenes with variations in both lighting conditions and viewpoints. We demonstrate our approach for real scenes.

1 Introduction

Matting and compositing are important operations in the production of special effects. These techniques enable directors to embed actors in a world that exists only in imagination, or to revive creatures that have been extinct for millions of years. During matting, foreground objects are extracted from a single image or video sequence. During compositing, the extracted foreground objects are placed over novel background images. Recently, many approaches, that utilize the statistical framework [4, 7, 1], the graph-cut based optimization [11, 8] or Poisson equation [12], have been reported to estimate an optimal opacity, α, for each pixel of the foreground object, especially along the intricate boundaries such as hair strands and furs.

Although shadow effects are recently considered in video compositing applications [3, 9], another challenge in compositing applications is to generate realistic looking reflections of the inserted objects. Existing techniques either define beforehand a reflection model (e.g. [10]), or explicitly extract the reflection models from the images. The reflectance recovery algorithms, e.g. [2], typically either directly measure the reflectance on the object using a specific device or extract the reflectance from a set of images or a single image. While using a specific device is unlikely in many compositing applications, the reflectance recovery methods from images mostly limit to perfectly diffuse surfaces and require a 3D geometrical description of the surfaces of some object.

Generally, synthesizing such reflections for compositing applications is inherently difficult because we are typically given only limited input, i.e. one view or a short video clip of a target scene. For example, suppose we want to insert a new synthetic object on the top of a real anisotropic mirror inside a real scene. This operation clearly requires taking into account the interaction between the new object and its environment (especially the mirror). This is impossible to do, if we do not have an approximation of the reflectance properties of the real surfaces in the image.

In this work, we focus on a slightly easier situation, where the target scene contains a ground plane and some up-right vertical objects, e.g. walls, crowds, desks, street lamps, trees, etc., which are common in both indoor and outdoor environments. Basically, we divide the problem of synthesizing reflections of inserted objects into two subproblems. From the geometric point of view, we need to synthesize reflections which would be seen by the same camera used in capturing the target scene and be reflected by the true reflective media in the target scene (Section 2). Physically, we aim to infer the most likely rendered reflections given the set of known reflection patches in the target view. This problem is formulated as a Maximum *A Posteriori* (MAP) estimation problem in Section 3. We then illustrate the algorithm's performance in Section 4. Section 5 concludes this paper with observations and proposed areas of future work.

2 Geometric Constraints

In this work, we assume that the reflection surface is approximately planar, and the perturbation and curve are relatively small. Two typical example scenes are shown in Fig. 1. The rules for plane (flat surface) reflections are simple: everything is governed by lines perpendicular to the mirror

The 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00

(a) (b)

Figure 1. Two real scenes containing reflection effects.

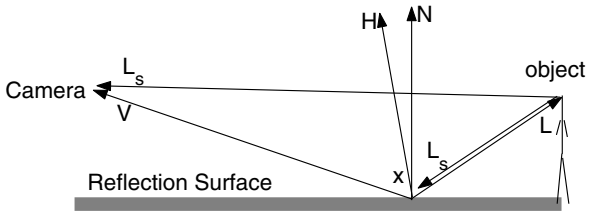


Figure 2. Geometric entities for Cook-Torrance model.

surface. That is, identical points in the object and its reflection lie on a single line perpendicular to the mirror surface, and the reflected object appears at the same distance behind the mirror as the actual object is in front of the mirror. Therefore, it is evident that a point t_i and its reflection r_i have the same perpendicular distances from the mirror surface. However, in perspective image, the object and its reflection will have an identical height only if the direction of view is nearly parallel to the water surface. In the example above (Fig. 1 (a)), however, the reflections of the standing persons are clearly viewed at a downward angle, which will make them appear foreshortened in comparison to the actual persons.

Although the ratios of lengths are now preserved under perspective projection, the following cross ratio constraint is valid,

$$\{v_x, t_i; b_i, r_i\} = \frac{(v_x - t_i)(b_i - r_i)}{(v_x - r_i)(t_i - b_i)} = 1, \qquad (1)$$

where v_x is the vertical vanishing point which is invisible in Fig. 1, b_i is the intersection of line $t_i r_i$ and the reflection surface. Notice that, line segments $t_i b_i$ and $b_i r_i$ in image plane have same distance only when v_x goes infinity, e.g. the direction of view is nearly parallel to the water surface. Our method first computes the vertical vanishing point by identify two pairs of t_i and r_i or three points t_i, b_i and r_i. Then we enforce the cross ratio property to find the correct positions of the reflections of each inserted object.

3 Photo-realistic Constraints

While the geometrical constraints help us to put reflections at correct positions, we also need to match the color characteristics of the reflections as those of the real scene. Statistically, this task is to infer the most likely rendered reflections given the set of known reflection patches.

In MAP estimate, we try to find the most likely estimates for the synthesized reflections, $\mathcal{V}$, of the composited object, given N observed reflection patches $\{\mathcal{P}_i\}_{i=1}^{N}$. In other words, we aim to maximize the posterior $P(\mathcal{V}|\mathcal{P}_1, ..., \mathcal{P}_N)$ in a Bayesian framework.

The light reflected on the object surface can be approximated as a linear combination of two reflection components: diffuse reflection component, I_d, and specular reflection component, I_s. While the diffuse reflection component I_d mostly remains fixed during the composition process of novel objects, the specular reflection can be modelled using the physically based Cook-Torrance illumination model [5]:

$$I_s(\lambda, x, L \to V) = L_s \frac{FGe^{-\alpha^2/m^2}}{\pi(N \cdot L)(N \cdot V)}, \qquad (2)$$

where L_s is the specular illumination, N is the surface normal at point x, V is the vector to viewer, L is vector to light source, α is the angle between N and the bisector H of V and L, m represents the surface roughness, G describes the geometrical attenuation caused by self-shadowing and masking of the micro-facets and F is the Frensel term.The geometric entities are depicted in Fig. 2. Note that I_s is wavelength λ dependent.

Obviously, the specular reflection component, I_s, at the wavelength, λ, depends on L_s, N, L, V, m, and η. For each position x, N, V, m and η are mostly the same. Under the assumption that the object to be reflected has a diffuse surface, i.e. the directions of the reflected lights from the surface are evenly distributed in all directions, the light captured by the camera would be the same as that to be reflected at point x. This basic idea is illustrated in Fig. 2 and also in Fig. 1 (b). The specular reflection component $I_s(x)$ at position x, therefore, mainly relies on L, which is different for each point on the object. The final observed specular reflection component I_s is given by

$$I_s(\lambda) = \int_x I_s(\lambda, x). \qquad (3)$$

In practice, to fully recover the illumination model parameters from several images remains a hard problem. In this work, the objects to be reflected is assumed to be relatively small and the camera is relatively far away. Under this assumption, the differences of the light directions L from x

to different positions on the object are ignored, and hence we use constant Fresnel terms, F_λ, for blending between diffuse I_d and specular I_s, similar to [6],

$$I_o(x,y) = F_\lambda \cdot I_s(x,y) + (1 - F_\lambda) I_d(x,y), \quad (4)$$

where I_o is the ingoing radiance received by the camera.

To simplify the problem, therefore, we assume that the Fresnel terms are constant for the original reflection patches and also the same for the reflections to be synthesized. This assumption is mostly valid in our cases because we are synthesizing reflections on a ground plane which typically have similar physical causes of image features. As a result, the posterior $P(\mathcal{V}|\mathcal{P}_1, ..., \mathcal{P}_N)$ simplifies to

$$P(\mathcal{V}|\mathcal{P}_1, ..., \mathcal{P}_N) = \prod_{i=1}^{N_r} \prod_{j=1}^{N_i^r} P(F_\lambda, I_d(i,j)|I_o(i,j), I_s(i,j)), \quad (5)$$

where $S_0 = diag([\beta_R, \ \beta_G, \ \beta_B])$ are the constant shading image values, and N_i^s is the number of pixels of the i^{th} shadow patch $\mathcal{P}_i^s$. Bayes' rule allows us to express the result as the minimization over a sum of negative log likelihoods:

$$\begin{aligned} &\arg\max_{F_\lambda, I_d} \prod_{i=1}^{N_r} \prod_{j=1}^{N_i^r} P(F_\lambda, I_d(i,j)|I_o(i,j), I_s(i,j)) \\ =&\arg\max_{F_\lambda, I_d} \prod_{i=1}^{N_r} \prod_{j=1}^{N_i^r} \frac{P(I_o, I_s|F_\lambda, I_d) P(I_d) P(F_\lambda)}{P(I_o) P(I_s)} \\ =&\arg\min_{F_\lambda, I_d} \sum_{i=1}^{N_r} \sum_{j=1}^{N_i^r} C(I_o, I_s|F_\lambda, I_d) + C(I_d) + C(F_\lambda), \quad (6) \end{aligned}$$

where N_i^r is the number of pixels of the i^{th} reflection patch $\mathcal{P}_i^r$, I_o, I_d and I_s are the abbreviate forms of $I_o(i,j)$, $I_d(i,j)$ and $I_s(i,j)$ respectively, and the terms $P(I_o)$ and $P(I_s)$ are dropped because they are constant and thus can be ignored in the estimation process. The first term $C(I_o, I_s|F_\lambda, I_d)$ is of the form

$$C(I_o, I_s|F_\lambda, I_d) = \gamma_1 \|I_o - F_\lambda \cdot I_s - (1 - F_\lambda) I_d\|^2, \quad (7)$$

where γ_1 is the tuning parameter. To simplify the minimization process, we also assume that the Fresnel prior $C(F_\lambda)$ is constant, and is omitted from the minimization process.

For $C(I_d)$, we consider two cases. In the first case, where the reflection effects are sparse (e.g. Fig. 1 left), we utilize the spatial coherence of the diffuse reflection. That is, we model $C(I_d(x,y))$ as a sum of weighted distances between the estimated I_d and those pixels in a subset $\mathcal{N}_o'(x,y)$ of a small set of neighboring pixels of (x,y), with the subscript, o, implies that the pixels are outside of the reflection patches or have already been computed,

$$C(I_d) = \sum_{(u,v) \in \mathcal{N}_o'(x,y)} \gamma_2 w(u,v) \|V(u,v) - I_d\|^2, \quad (8)$$

where $w(u,v)$ is a spatial Gaussian falloff with $\sigma = 3$ to stress the contribution of nearby pixels over those that are far away,, γ_2 is a tuning parameter, and $V(u,v)$ is a pixel's value along only one color channel). In the second case, where the reflection effects are dense (e.g. Fig. 1 right), we assume that I_d is constant through the reflection surface and hence we ignore the second term $C(I_d)$.

4 Experimental Results

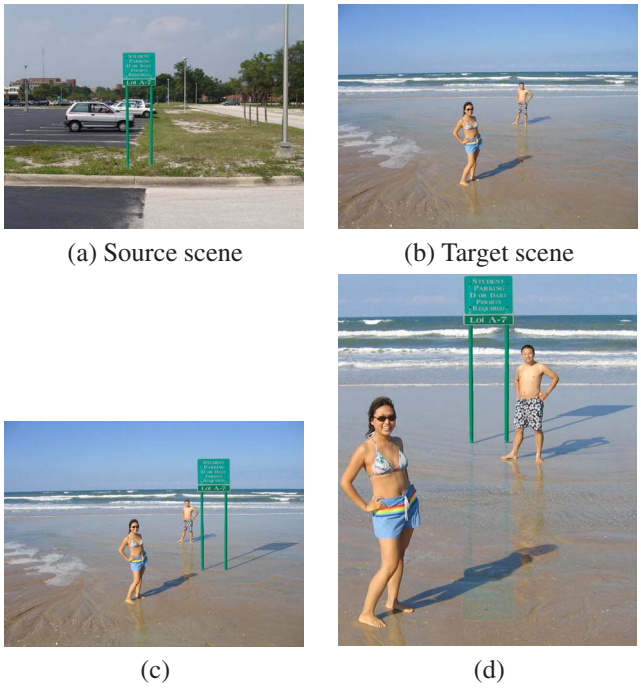
To illustrate the proposed method, we tried it on different cases. In the first example (Fig. 3), we insert a parking sign post (Fig. 3 (a)) into a target image of a beach scene (Fig. 3 (b)). First, we use our previous work [3] to create shadows for planar parking sign post as shown in Fig. 3 (c). However, in the target scene shown in Fig. 3 (b), we also need to synthesize the reflection effects to increase the realism. Fig. 3 (d) shows the composite with correct and realistic shadows and reflections, even the complicated interactions between the synthesized shadow and existing person in the original scene, and between the synthesized reflections and existing shadows in the original scene.

The target scene in Fig. 4, different from the sparse reflection in Fig. 3, introduces challenges both in estimating the Fresnel terms and in dealing with the distortion of the water surface. Our method synthesized promising reflections of the inserted *Puss in Boots* (a character in movie Shrek 2) as shown in (b) and zoomed in (c). In the result (d) using only α blending, is the cat touching on the water surface or floating in the air?

5 Conclusion and Future Work

Synthesizing reflections for objects transferred from one natural scene into a novel scene with different lighting condition remains a hard problem. In this paper, we focused on a slightly easier situation, where the target scene has some up-right vertical objects. We showed that this assumption leads to a novel, simple algorithm for reflection synthesis, and showed encouraging results for different target scenes. Using an image-based approach, we are unlikely to estimate realism by using known geometry of the light, the object to be pasted, and the background objects. However, the proposed method advances the image based techniques one step further to improve the realism in applications of matting and compositing techniques.

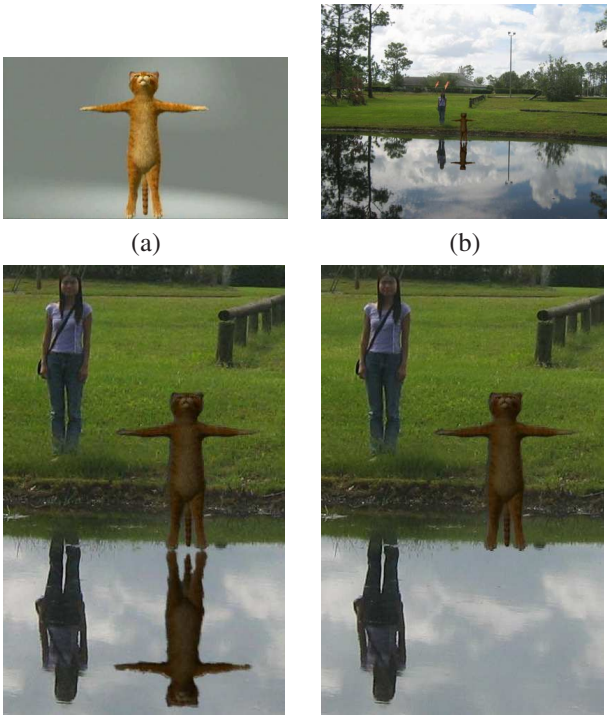
There are a number of ways in which our current approach can be extended. First we would like to relax the vertical object constraints for the target scene. Second, we currently use the tools in Photoshop to generate the ripple effects, and would like to reduce the interaction in the future work. Finally, we would also like to utilize richer models



(a) Source scene (b) Target scene

(c) (d)

Figure 3. Shadow and Reflection synthesis. (c): A zoomed view of a composite using our previous method [3]. The result (c) is noticeably fake. (d): A composite with geometrically correct and visually realistic reflections obtained by the proposed method.



(a) (b)

(c) (d)

Figure 4. Synthesizing the reflections. (a) is the source image and Fig. 1 (b) shows the target image. (b) is the composite with synthesized reflection. (c) is the zoomed version of (b). The result in (d) using only α blending, compared with (c), has no cues of the position where the object is placed in the space.

in learning the illumination conditions from the target scene and apply it to the composited objects.

Acknowledgment

We would like to thank Ms. Min Hu for her acting and help for preparing partial results presented in this paper.

References

[1] N. E. Apostoloff and A. Fitzgibbon. Bayesian video matting using learnt image priors. In *Proc. IEEE CVPR*, pages 407–414, 2004.

[2] S. Boivin and A. Gagalowicz. Image-based rendering of diffuse, specular and glossy surfaces from a single image. In *Proc. ACM SIGGRAPH*, pages 107–116, 2001.

[3] X. Cao, Y. Shen, M. Shah, and H. Foroosh. Single view compositing with shadows. *The Visual Computer*, 21(8):639–648, 2005.

[4] Y. Chuang, B. Curless, D. Salesin, and R. Szeliski. A bayesian approach to digital matting. In *Proc. IEEE CVPR*, volume 2, pages 264–271, 2001.

[5] R. L. Cook and K. E. Torrance. A reflectance model for computer graphics. *ACM Trans. Graph.*, 1(1):7, 1982.

[6] W. Heidrich. *High-quality Shading and Lighting for Hardware-accelerated Rendering*. PhD thesis, Univ. of Erlangen, 1999.

[7] P. Hillman, J. Hannah, and D. Renshaw. Alpha channel estimation in high resolution images and image sequences. In *Proc. IEEE CVPR*, pages 1063–1068, 2001.

[8] Y. Li, J. Sun, and H. Shum. Video object cut and paste. *ACM Trans. Graph.*, 24(3):595–600, 2005.

[9] H. Nicolas. Shadow synthesis for video postproduction. *IEEE Signal Processing Letters*, 12(4):321–324, 2005.

[10] A. Ostler. The primal seas: water on playstation 2. In *Proc. ACM SIGGRAPH on Sketches & applications*, pages 1–1, 2003.

[11] C. Rother, V. Kolmogorov, and A. Blake. "grabcut": interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, 23(3):309–314, 2004.

[12] J. Sun, J. Jia, C. Tang, and H. Shum. Poisson matting. *ACM Trans. Graph.*, 23(3):315–321, 2004.