Creating Realistic Shadows of Composited Objects

Xiaochun Cao and Mubarak Shah
Computer Vision Lab
University of Central Florida
Orlando, FL, 32816-3262

Abstract

This paper presents an approach for creating realistic shadows of objects composited into a novel background. Existing techniques either do not deal with the shadows or require that the relationship of the light source, reference plane, and camera be the same in the source and target scenes. We observe that scenes containing a ground plane and some up-right objects, e.g. walls, crowds, desks, street lamps, trees, etc., are common in natural environments, and show that two vertical lines and their cast shadows are enough to recover primary light source, which makes it possible to construct shadows of inserted planar objects in the single target image. We also demonstrate a technique to calibrate the camera when two views are available, which enables us to create shadows of general $3D$ objects in a target scene. The proposed method therefore advances the image based techniques one step further toward the synthesis of the scenes with variation of both lighting and viewpoints. We demonstrate our proposed 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 elements are extracted from a single image or video sequence. During compositing, the extracted foreground elements are placed over novel background images. Many matting methods [13, 4, 8, 1, 12, 15] have been reported to estimate an optimal opacity for each pixel of the foreground objects especially along intricate boundaries.

Most of those works do not deal with the effects of shading. However, shadows provide important visual cues for depth, shape, contact, movement, and lighting in our perception of the world and thus increase scene realism. It is well known that a moving cast shadow contains more information for depth than a stereo disparity. Without some kind of shadow contact with the background, the objects composited over the new background may appear to float around or walk on the air (e.g. Figure 8 (c) and (f)).

Chuang et. al. [5] described a method for extracting shadows from one natural scene and inserting them into another under the assumption that a single primary point light source is responsible for dominant cast shadows within a scene. Assuming no interreflections, they develop physically-based shadow matting and compositing equations, and use those to extract a shadow matte from a source scene in which the shadow is cast onto an arbitrary planar background. One limitation of this approach is that they require that the relationship of the dominant light source, reference plane, and camera be the same in the source and target scenes.

In this paper, we present a method to create realistic shadows of objects matted from one natural scene and composited into novel background scene with different view and lighting condition. The process requires relatively little effort per target scene, and generates plausible cast shadows of even complex shapes.

At a high level, our approach works as follows. Starting from a target scenario where there are at least two vertical line segments casting shadows on the ground plane, the key light is firstly recovered based on planar homology, while the camera is calibrated when two views are available. In additional to the geometric information, shadow matte opacity for the object to be composited is assumed to be similar to those of shadows cast by existing objects. We then acquire the alpha matte and the silhouette set of edges for the object based on the obtained geometry of target scene. The silhouette edge set is the set of those edges that would appear in the silhouette of the shadow. Next, we composite the matted object into proper position in the target scene with user interaction using computed alpha map from the source scene. Finally, the cast shadow mattes are generated automatically according to the geometric constraints and shadow matte opacity.

Section 2 describes the details of our method. We demonstrate the results of our working system in Section 3. Section 4 concludes with observations and proposed areas of future work.

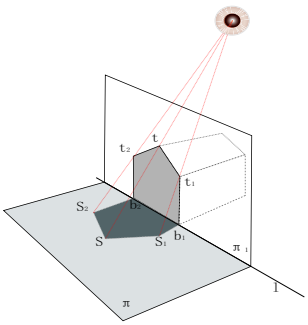


Figure 1: Computing the shadow positions for any points on the side wall, π_1, of a sample house. We first compute the planar homology, H, that casts shadows from plane, π_1, on the ground plane, π, and then decompose it to obtain primary light source, v, and characteristic invariant μ. Consequently, we are able to determine the shadow position of any point on plane π_1 using the characteristic invariant constraint. For instance, the roof corner t must cast shadow at s on ground plane based on the cross ratio μ along the line vt.

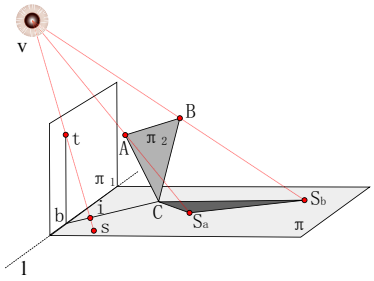


Figure 2: Computing the shadow positions for any points on the planar objects π_2. We first compute a point t on the intersection line bt between π_2 and π_1. Using the method described in Figure 1, we then determine the shadow, s, correspondent to point t, and thus the new characteristic invariant μ' for plane π_2 from the cross-ratio of four points: v, t, i and s. As a result, shadow position of any point on the new plane π_2 is computable from μ'.

2 Creating Shadows

In this section, we show how to generate a realistic shadow of an object matted from a natural scene and composited into a novel target background with different lighting condition. To accomplish this, we assume that a single light source is responsible for dominant cast shadows within a scene.

Two cases are examined: (i) The case when the object to be transferred is planar or distant from the camera. We refer this object as "2D object". In this case the "outline" of the shadow volume and the contour of alpha map can be modelled by a planar homography (Sec. 2.1). (ii) The case of a non-planar object. We refer to this object as "3D object". In this case we need to know the relative orientation and translation between the camera and principal light positions. We show this is possible when an extra image or a video sequence is also available(Sec. 2.2).

2.1 Planar or Distant (2D) objects

Before beginning the discussion of how to create shadows for matted objects, it is necessary to analyze the geometry of target scenes containing a ground plane and some up-right objects. Two vertical lines, for instance b_1t_1 and b_2t_2 in Figure 1, must be parallel and hence define a plane, π_1, that is perpendicular to the ground plane π. Therefore, there is a 3×3 planar transformation H between the image of the side wall plane π_1 and the image of its shadow on the ground plane π. Geometrically, H is a planar homology [14] since there are a line l of fixed points, which is the intersection b_1b_2 of the vertical plane π_1 and ground plane π, together with a fixed point vertex v, which in our case is the primary point light source, not on the axis l. Algebraically, H has two equal and one distinct eigenvalues and the axis is the line through the two eigenvectors (points) corresponding to the equal eigenvalues, and also that the ratio of the distinct eigenvalue to the repeated one is a characteristic invariant μ of the homology. Therefore, H has only five instead of eight degrees of freedom and hence in theory three matched points are sufficient to compute this planar homology.

We then parameterize H directly in terms of five freedoms, the homogeneous 3-vectors representing the axis l and the vertex v, and one characteristic invariant μ, as [7]:

$$H = I + (\mu - 1)\frac{vl^T}{v^T l}, \tag{1}$$

where I is the identity matrix. The eigenvectors of H are

$$e_1 = v, e_2 = l_1^\perp, e_3 = l_2^\perp, \tag{2}$$

with corresponding eigenvalues

$$\lambda_1 = \mu, \lambda_2 = 1, \lambda_3 = 1, \tag{3}$$

where $l_i^\perp$ are two vectors that span the space orthogonal to the l, i.e. $l = l_1^\perp \times l_2^\perp$.

From the eigenvalue decomposition, we obtain the dominant light source, v, the intersection, l, between the ground plane and the vertical plane, and also the ratio of the distinct eigenvalue to the repeated one which is characteristic invariant, μ, of the homology. One of the planar homology's most important properties is that the cross ratio defined by

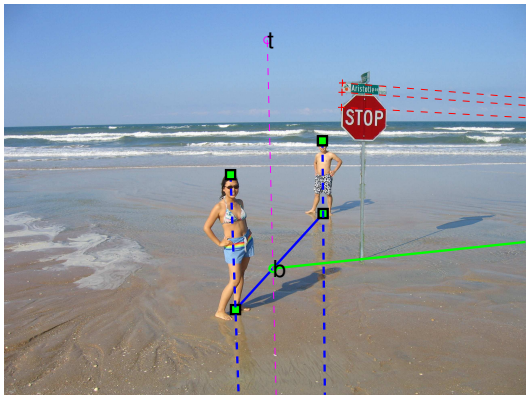


Figure 3: Creating the shadows using parallel lines.

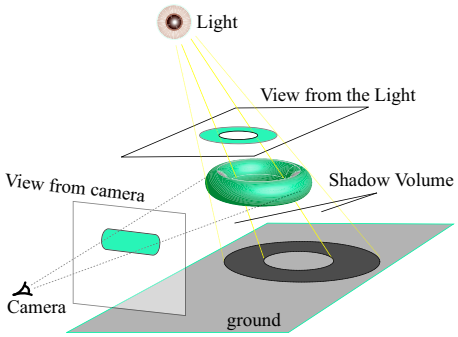


Figure 4: Obtaining the silhouette edges of shadow volume by taking an image from the lighting direction.

the vertex v, a pair of corresponding points, and the intersection of the line joining these points with the axis l, is the same as μ for all points related by the homology. Therefore, given any point in the vertical plane, it is easy to compute its shadow position on the ground plane using the characteristic invariant constraint.

In practice, however, it might be more interesting to generate shadows of objects which are not in the vertical plane π_1 as shown in Figure 2. The only extra information we need for this problem is the intersection line bt between planes π_2 and π_1. Fortunately, there is a simple but efficient strategy: paste deliberately the planar object ABC in such a position in the target background that the transferred object ABC intersects obviously with the vertical plane π_1 at some easily identified line. Once the line bt is computed, the process to create shadows of the object ABC which is not in the vertical plane π_1 is described in Figure 2. Otherwise, we need one extra target image to generally solve this problem as described in the next subsection 2.2.

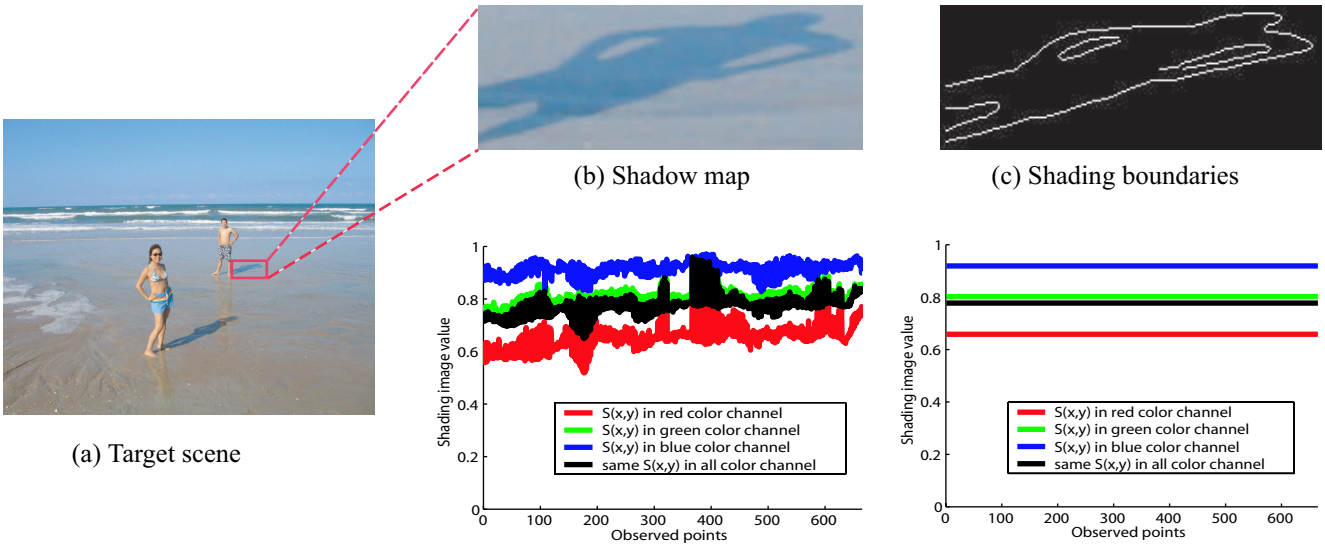
Vanishing points or parallel lines, however, are also useful features for this task. One example is shown in Figure 6 and elaborated in Figure 3. In this example, we were trying to generate the shadow of a stop-sign matted from one street scene (6 (a)) and composited into a totally different beach background (6 (b)) with two standing people. Since lines, l_i, on the stopsign and roadsign plotted in dash red in Figure 3 are parallel to the ground plane, we can compute the vanishing point v_x along the direction of the intersection (new axis l'_s in solid green) by minimizing the algebraic distance $\sum_i (v_x^T l_i)$. The line l'_s must thus pass v_x and the contact point of the stop-sign on the sand, and intersects old axis, l_s, in the solid blue at point b. We then fit the vertical vanishing point v_y using the vertical lines in dash blue, and pick up one point t on the vertical line, l_v, through b and the vertical vanishing point v_y. μ' and thus shadow of any point in the new stopsign plane is computable as described in Figure 2. The final computed shadow map is shown in figure 6(c). In addition to parallel lines on the object itself, we can also use the horizon line which is the apparent intersection of the earth and sky. Any line on the stopsign plane and parallel to the ground plane will intersect the horizon line at the vanishing point v_x.

Note that it might be easier if the object to be composited has nontrivial contact area with the ground plane. In such cases (for example Fig 7), we can directly fit the new axis, l'_s, and compute the important characteristic invariant μ'. Notice that this approach described so far is also valid for point light sources which are not very distant to scenes.

2.2 General (3D) objects

In order to alleviate the planar limitations in the last subsection (2.1), we require at least two views of the target scene. We recover the camera information by extending the methods introduced in [2, 3]. After calibrating these images, we can compute the position of a primary light source in $3D$ by using triangulation [7], and thus determine the light source and camera position in $3D$ world. Consequently, we can acquire precise $3D$ model of the object to be transferred using opacity hull method [9] or structure from motion [7], and thus interactively create shadows [10, 11].

However, this model-based framework is expensive and also impractical in matting and compositing applications since the model is not always obtainable. In practice, we can capture two images as shown in Figure 4: one from the camera center while the other from the lighting direction, since we have already computed the camera and light positions. The view from camera position is used to matte the foreground object. The contour of the foreground object matted from the view taken from the lighting direction is approximately the silhouettes for shadow volume. We then extend the shadow volume technique [6] to render the shadow using the silhouette edges.



(a) Target scene (b) Shadow map (c) Shading boundaries

(d) *S(x,y)* for all observed points (e) Averaged *S(x,y)*

Figure 5: Estimating the shadow matte in a target scene. (b) shows the shadow map of a standing person in the target beach scene (a). (c) is the shadow boundary of the shadow in (b) detected by Canny edge detector. (d) and (e) plo the illumination image values $S(x, y)$ for all points (x, y) along the shadow boundaries.

2.3 Estimating the shadow matte

While the geometric information is useful for enforcing geometrical consistences of the constructed shadows, they are not enough to properly create realistic shadows. Denoting by $I(x, y)$ the input image and by $R(x, y)$ the reflectance image and $S(x, y)$ the shading (illumination) image, the three images are related by [17]:

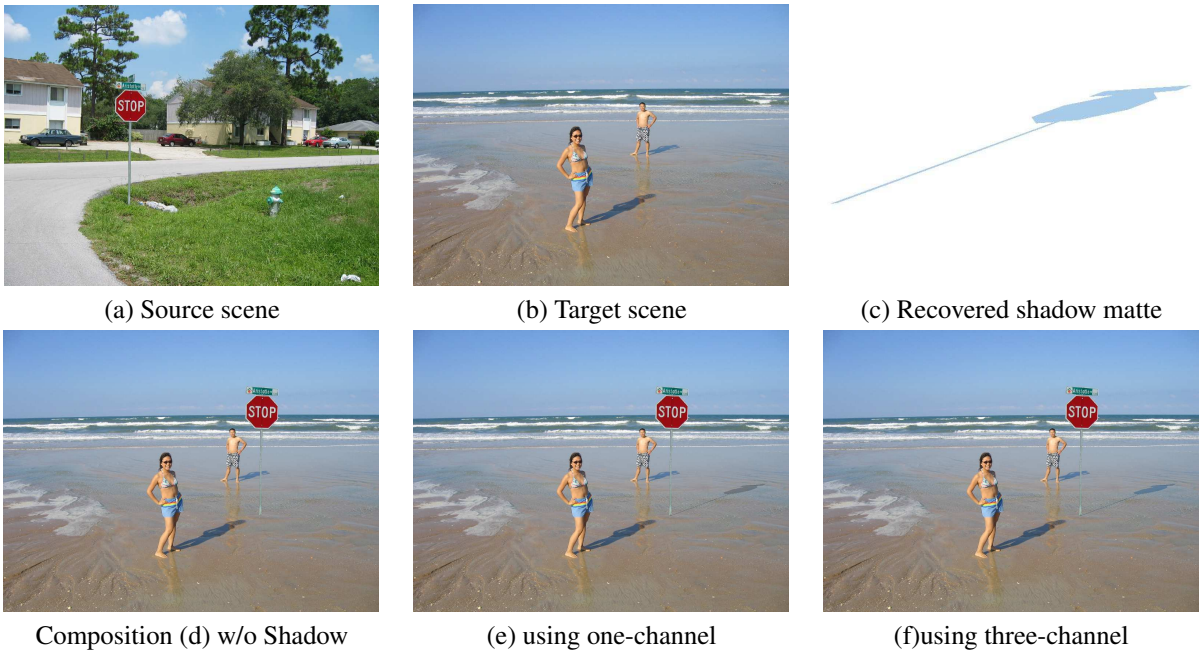
$$I(x, y) = S(x, y)R(x, y). \quad (4)$$

In order to create shadows over the target image, we assume the illumination image $S(x, y)$ of the inserted object to be same as that of shadows cast by the existing objects. Our approach takes advantage of the property that changes in color between pixels indicate either a reflectance change or shading effects. It is unlikely that significant shading boundaries and reflectance edges occur at the same point [16], thus we make the simplifying assumption that every color change along the shading edges is caused by shading only. In other words, the reflectance image colors across the shading boundaries should be the same. Let us consider the shadow map (figure 5 (b)) of one standing person from the target beach scene (figure 5 (a)). Shading boundaries (figure 5 (c)) are detected by canny edge detector. The interesting observation is that the changes in a color image due to shading affect three color channels disproportionally as shown in figures 5 (d,e). Obviously, the shading affects the red color channel the most and the blue color channel the least. While there are some complex models to model this effect, we simply use different shading image value along three color channels to approximate the effect. Our final result (figure 6(f)) is noticeably more realistic than the result (figure 6(e)) created by assuming that any changes in a color image due to shading should affect all three color channels proportionally.

3 Experimental Results

To demonstrate the proposed method, we applied it to different cases discussed in section 2. We first created shadows for planar (Figure 6) and Distant (Figure 7) objects. The process to generate shadows of stopsign and roadsign in Figure 6 needed some extra geometric constraints such as parallel lines or vanishing points, and is explained in section 2.1. For the shadow of Puss in Boots (a character in movie Shrek 2) in Figure 7, nonetheless, we do not need those extra constraints since the Puss in Boots has nontrivial contact area with the ground plane. Very impressive shadows are generated even for the sword of the character.

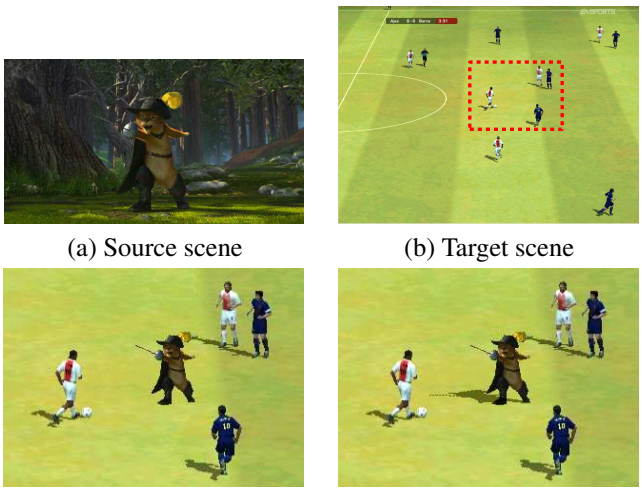
Figure 8 demonstrates the performance of our method to create a shadow of a $3D$ object. The shape of statue is complicated and challenging to render manually. Our method generated realistic shadows on correct image locations, no matter where the object is placed in space, either standing on the ground (Fig. 8(c)) or floating in the air (Fig. 8(e)).



(a) Source scene (b) Target scene (c) Recovered shadow matte

Composition (d) w/o Shadow (e) using one-channel (f)using three-channel

Figure 6: Shadow synthesis for planar objects. Starting from a source image (a), we first extract the stopsign and roadsign using image matting. Two standing people and their shadows in the target scene (b) are enough to recover the primary light source and thus it is possible to create geometrically consistent shadow maps (c) (a close view) for the stopsign and roadsign. Note that the shadow matte (c) is expressed in three rather than one color channels as discussed in section 2.3. We show in (d) what happens when α blending is used to composite the stop-sign in the beach image. While the result (d) is noticeably fake, the composite (e) remains unconvincing. Our proposed method creates a geometrically consistent and visually realistic shadows for the stop-sign and roadsign shown in (f).



(a) Source scene (b) Target scene

(c) w/o shadows (d) w/ shadows

Figure 7: Shadow synthesis for the distant object. The figure shows the composite of *PUSS IN BOOTS (a character in movie Shrek 2)* in game FIFA 2003. Both (c) and (d) are zoomed views of the marked rectangle area in (b).

4 Conclusion and Future Work

Synthesizing shadows for objects transferred from one natural scene into a novel scene with different lighting conditions remains a difficult problem for computer vision and computer graphics systems. Here 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 and simple algorithm for shadow construction, and demonstrated encouraging results for both planar or distant objects and $3D$ objects. As an image-based approach, it is 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. It would also be useful to have interactive tools for editing self-shadows Finally, we would like to learn the illumination condition from the target scene and apply it to the matted objects. This, for example, will make the character Puss in Boots

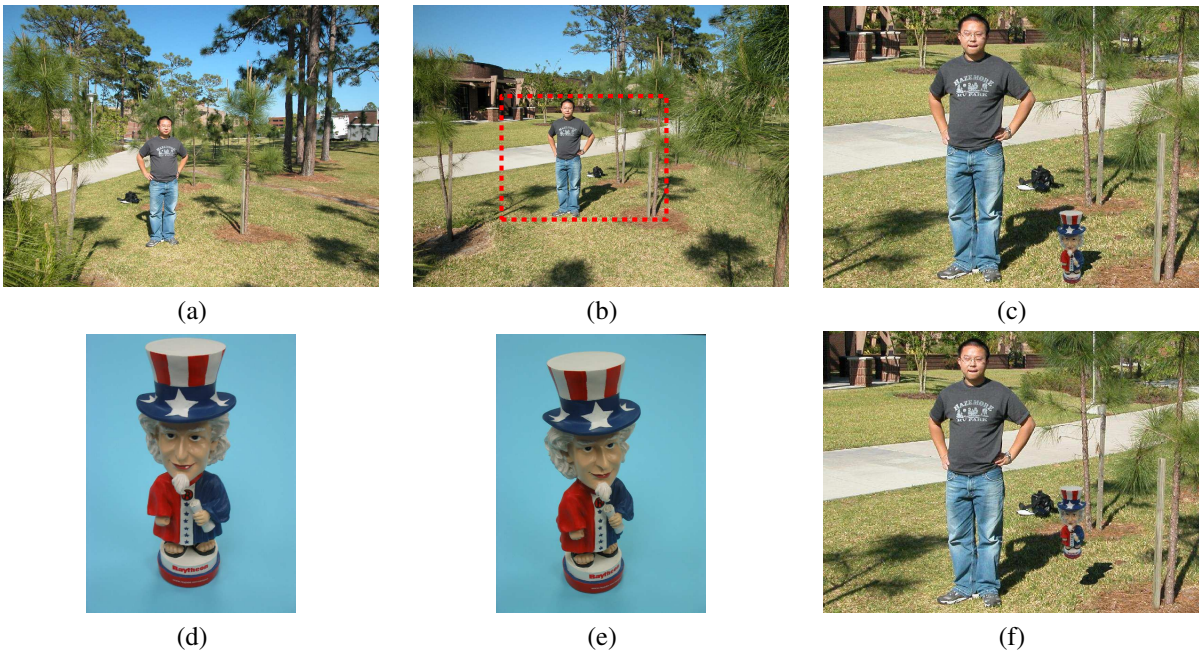


Figure 8: Shadow synthesis for a $3D$ object. Starting from two views (a) and (b) of the target scene, we first calibrate the camera and compute the light source orientation. Utilizing these computed geometric information, we then captured two images (d) and (e) of the statue. (d) is the view from the camera, while (e) is the view seen from the lighting direction. We next extracted the statue from (d) and generated the shadow of the statue using the contour of the statue in (e). Two results are shown: (i) the statue stands on the ground (c); (ii) the statue floats in the air (f). Both (c) and (f) are zoomed composite views of the marked rectangle area in (b).

in Figure 7 brighter and more realistic.

References

[1] N. E. Apostoloff and A.W. Fitzgibbon. Bayesian video matting using learnt image priors. In *IEEE CVPR*, 2004.

[2] Xiaochun Cao and Hassan Foroosh. Metrology from vertical objects. In *Proc. of BMVC*, 2004.

[3] Xiaochun Cao and Hassan Foroosh. Simple calibration without metric information using an isoceles trapezoid. In *Proc. of the 17th ICPR*, pages 104–107, 2004.

[4] Yung-Yu Chuang, Brian Curless, David H. Salesin, and Richard Szeliski. A bayesian approach to digital matting. In *IEEE CVPR*, volume 2, pages 264–271, December 2001.

[5] Yung-Yu Chuang, Dan B Goldman, Brian Curless, David H. Salesin, and Richard Szeliski. Shadow matting and compositing. *ACM Trans. on Graphics*, 22(3):494–500, 2003.

[6] Franklin C. Crow. Shadow algorithms for computer graphics. In *Proc. of SIGGRAPH*, pages 242–248, 1977.

[7] R.I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

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

[9] W. Matusik, H. Pfister, A. Ngan, P. Beardsley, R. Ziegler, and L. McMillan. Image-based 3d photography using opacity hulls. In *Proc. SIGGRAPH*, pages 427–437, 2002.

[10] Fabio Pellacini, Parag Tole, and Donald P. Greenberg. A user interface for interactive cinematic shadow design. In *Proc. SIGGRAPH*, pages 563–566, July 2002.

[11] Lena Petrovic, Brian Fujito, Lance Williams, and Adam Finkelstein. Shadows for cel animation. In *Proc. SIGGRAPH*, pages 511–516, July 2000.

[12] C. Rother, V. Kolmogorov, and A. Blake. Interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics*, August 2004.

[13] M. Ruzon and C. Tomasi. Alpha estimation in natural images. In *IEEE CVPR*, pages 18–25, 2000.

[14] C. E. Springer. *Geometry and Analysis of Projective Spaces*. Freeman, 1964.

[15] Jian Sun, Jiaya Jia, Chi-Keung Tang, and Heung-Yeung Shum. Poisson matting. *ACM Trans. on Graphics*, 2004.

[16] Marshall F. Tappen, William T. Freeman, and Edward H. Adelson. Recovering intrinsic images from a single image. In *Advances in Neural Information Processing Systems*.

[17] Yair Weiss. Deriving intrinsic images from image sequences. In *Proc. ICCV*, 2001.