Shadow Matting and Compositing

Yung-Yu Chuang[1] Dan B Goldman[1,3] Brian Curless[1] David H. Salesin[1,2] Richard Szeliski[2]

[1]University of Washington [2]Microsoft Research [3]Industrial Light and Magic

(a) Foreground scene (b) Background scene (c) Blue screen composite (d) Our method (e) Reference photograph

Figure 1 Sample result from our matting and compositing algorithm for shadows. Given a foreground element photographed against a natural background (a), we seek to matte the element and its shadow and then composite it over another background (b). Using a blue screen (not shown) to extract the shadow, followed by conventional matting and compositing, we obtain a result (c) with double darkening of the existing shadow and without proper warping of the cast shadow. The results of our new shadow matting and compositing method (d) compare favorably with an actual photograph (e). Note the correct dimming of the specular highlight, the convincing geometric deformation, and the seamless matte edges where the foreground and background shadows meet.

Abstract

In this paper, we describe a method for extracting shadows from one natural scene and inserting them into another. We develop physically-based shadow matting and compositing equations and use these to pull a *shadow matte* from a source scene in which the shadow is cast onto an arbitrary planar background. We then acquire the photometric and geometric properties of the target scene by sweeping oriented linear shadows (cast by a straight object) across it. From these shadow scans, we can construct a shadow displacement map without requiring camera or light source calibration. This map can then be used to deform the original shadow matte. We demonstrate our approach for both indoor scenes with controlled lighting and for outdoor scenes using natural lighting.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image Generation—Bitmap and framebuffer operations; I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Shading

Keywords: Blue-screen matting, displacement map, faux shadow, image-based rendering, layer extraction, shadow matte.

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 film or video sequence. During compositing, the extracted foreground elements are placed over novel background images.

Traditional approaches to matting include *blue-screen matting* [Smith and Blinn 1996] and *rotoscoping* [Wright 2001]. The former requires filming in front of an expensive blue screen under carefully controlled lighting, and the latter requires talent and intensive user interaction. Recently developed matting algorithms [Ruzon and Tomasi 2000; Chuang et al. 2001] can now pull alpha mattes of complex shapes from natural images. Chuang et al. [2002] extended their Bayesian approach to video by interpolating user-drawn keyframes using optical flow.

Shadows provide important visual cues for depth, shape, contact, movement, and lighting in our perception of the world [Petrovic et al. 2000], and thus are often essential in the construction of convincing composites. Shadow elements for compositing are typically created either by hand or by extracting them from blue screen plates.

The manual approach is commonly called *faux shadow* in the film industry [Wright 2001]. For this technique, artists use the foreground object's own alpha matte to create its shadow. By warping or displacement-mapping the shadow, it can be made to drape over simple objects in the background plate. However, this approach has several limitations. First, an alpha matte is a flat projection of the object from the point of view of the camera that filmed it. If the view from the light is too far from the camera's point of view, the silhouette of the alpha matte may be noticeably different from the silhouette of the correct shadow, and the resulting synthetic shadow will be unconvincing. Second, the shadow color characteristics are manually adjusted by the compositor and do not necessarily match the shadow characteristics of the real scene. Most importantly, this approach becomes unwieldy for casting shadows on backgrounds with highly complex geometry.

The second main approach is to extract shadows from the foreground plates using luma keying or blue-screen matting. These techniques provide a better approximation to the correct shadow characteristics. However, depending on the compositing model used, it may be difficult to obtain photometrically realistic results. For example, in Figure 1(c), note that the blue screen composite gives a noisy shadow with the wrong density, and it creates a double shadow where the ground plane was already in shadow.

Regardless of the shadow extraction method, target background

scenes with complex geometry present special compositing challenges. In many cases, a rough model must be built so the actors cast shadows onto the model as they would onto the target scene. This model may be a physical blue screen model onto which the actor casts his real shadow for extraction, or a computer-generated virtual model onto which faux shadows are cast using a renderer. In either case, it is often difficult to construct a model that matches the target object exactly, so additional manual warping and rotoscoping is required to align the transferred shadows to the receiving geometry.

In this paper, we introduce a new process for shadow matting and compositing that captures all of these effects realistically. We develop a physically-motivated shadow compositing equation, and design a matting and compositing process based on this equation. For shadow matting, we extract a shadow density map to describe the degree to which each pixel is in shadow. In contrast to previous approaches, our matting method works for natural backgrounds. For shadow compositing, we use an active illumination approach to extract an *illumination map* and a *displacement map* for the destination scene. These maps describe the shadow appearance and distortions over the novel target background. We recover the displacement map without requiring the calibration of the camera or the position of the light source, using an arbitrarily textured planar reference region. Using these acquired maps, we can realistically transfer shadows from one scene to another. Our method imposes certain restrictions on the lighting, camera placement, and at least some of the geometry in the source and target scenes, which we discuss when evaluating the merits and limitations of our method.

Figure 1 shows an overview of our approach and compares to blue screen matting and compositing and to ground truth. Our method correctly occludes both diffuse illumination and specular highlights, retains soft shadow edges, warps shadows convincingly across arbitrary background geometry, and seamlessly blends newly introduced shadows with those already present in the background plate.

1.1 Related work

Matting and compositing have a long history in the film industry. Traditional approaches require a single-color background, hence the name blue screen matting [Smith and Blinn 1996]. Recently, several matting algorithms have been developed to extract mattes from natural backgrounds for images [Ruzon and Tomasi 2000; Chuang et al. 2001] and image sequences [Chuang et al. 2002]. Our method extends these approaches with a method to extract shadow mattes from natural image sequences.

Much research has been done on shadows in both the vision and graphics communities. For example, Finlayson et al. [2002] have attempted to remove shadows from a single image, and Pellacini et al. [2002] implemented a user interface for adjusting shadow placement using direct manipulation. However, none of this research deals directly with shadow matting and compositing.

The intrinsic image approach [Weiss 2001] attempts to decompose a video of an object under different illumination conditions into a reflectance map and a sequence of light maps. These light maps could be used as shadow mattes, though Weiss does not attempt to transfer them to other scenes. Matsushita et al. [2002] attempt to morph shadows for different lighting conditions using Weiss's approach. They capture the lightfield of a static scene with several light sources. With multiple cameras, they recover a view-dependent model using multiview stereo. A shadow mask is obtained by simply thresholding the illumination maps. Finally, the shadow for a novel light source is obtained by warping the shadow masks of neighboring samples and the estimated geometry.

Petrovic et al. [2000] also estimate 3D geometry for casting shadows, but with different goals and methods. Their method is intended to create shadow mattes for cel animation, so their estimates of 3D geometry can be somewhat more approximate. They create a 3D model for the scene by inflating the character and estimating simple geometry for the background from user gestures.

Researchers have developed a number of shape-from-shadow techniques. For example, Savarese et al. [2001] observe the self-shadowing of an object under different lighting conditions and carve a model to generate a plausible solution to match those observations. Our method builds on the shadow scanning approach of Bouguet et al. [1998]. However, we avoid explicitly reconstructing a 3D geometric model, which requires calibration of the camera and light sources. Instead, we estimate the displacement map in the image space directly.

Our compositing method is similar to Debevec's differential rendering approach for compositing synthetic objects into real scenes [1998]. Debevec records the differences between the rendered radiances with and without the synthetic objects in the scene. The differences are then added to the real destination background to make the composite. To avoid explicitly modeling the geometry and BRDF of the destination scene, we take a different approach.

Our approach also has some interesting similarities to environment matting and compositing [Zongker et al. 1999; Chuang et al. 2000]. Both lines of research attempt to capture lighting phenomena that are modeled incorrectly by the traditional compositing equation. Both use an active illumination approach to capture the information required for realistic composites. Our method casts oriented stick shadows (Section 4), whereas high-accuracy environment matting uses oriented Gaussian stripes. Our warping function plays a similar role to the warping function for single-frame environment matting.

1.2 Overview

In the following sections, we first develop our shadow matting equation and shadow matting algorithm for scenes with identical source and destination background geometry (Section 2). We then describe shadow compositing onto simple (planar) destination backgrounds (Section 3) and shadow warping for more geometrically complex backgrounds (Section 4). We present results using our technique in Section 5. Finally, we discuss the limitations, working range, and pros and cons of our approach (Section 6), and conclude with a summary and ideas for future research.

2 Shadow matting

In this section, we develop our shadow compositing equation and describe our algorithm to estimate shadow mattes.

Traditionally, matting and compositing operations are based on the compositing equation [Porter and Duff 1984],

$$C = \alpha F + (1 - \alpha)B. \tag{1}$$

The composite color C is a linear combination of the foreground color F and the background color B weighted by the opacity α of the foreground object. This color blending model is effective for capturing partial fill and motion blur effects. Some of the previous approaches for shadow matting assume this model and try to determine α and F for the shadow [Wright 2001]. This approach effectively represents the shadow as a translucent dark layer. However, the standard compositing model does not hold for shadows because shadows are caused by the occlusion of light sources rather than color blending.

2.1 The shadow compositing equation

To determine an appropriate model for shadow compositing, we treat the problem within a simplified lighting framework. In particular, we assume that a single, primary point light source (sometimes called a *key light*) is responsible for dominant cast shadows within

a scene. The remaining, secondary lighting is either dim or of such wide area (as in sky illumination) that its cast shadows are comparatively negligible. If we further assume no interreflections, then we can model the observed color C at a pixel as:

$$C = S + \beta I, \tag{2}$$

where S is the shadowed color, I is the reflected contribution of the primary light source, and β is the visibility to that light source. Let L be the color of a pixel when not in shadow. Substituting $I = L - S$ into Equation (2) and rearranging, we obtain the *shadow compositing equation*,

$$C = \beta L + (1 - \beta)S. \tag{3}$$

Equation (3) can be thought of in terms of images or layers. In this case, S is the *shadow image*, L is the *lit image*, and β is the *shadow density matte* (or just *shadow matte*) representing the per-pixel visibility of the light source casting the shadow. Note that β may be fractional. Such values represent the same phenomena as fractional α values, including motion blur and partial coverage. Fractional β also allows us to simulate penumbral effects. (However, for non-point light sources, our model is only an approximation. This limitation is discussed in more depth in Section 6.)

The lit and shadow images L and S depend on the lighting conditions and albedos of the source scene. Hence, they are not directly transferable from scene to scene, unlike the shadow matte β, which is what we estimate during the matting process. For compositing, we therefore require two new images: the lit and shadow images, L' and S', of the new (destination) scene (Section 3).

2.2 Estimating the shadow matte

During the matting process, given the observed color C, we need to recover the shadow matte β. We therefore first need to estimate the shadow image S and the lit image L. We assume the image sequence is taken from a static camera with a static background. We can estimate the lit image and the shadow image using max and min compositing [Szeliski et al. 2000], i.e., finding the darkest and brightest value at each pixel. As Chuang et al. [2002] did for smoke, we first use the video matting algorithm to extract the mattes of the foreground objects and exclude them from the max/min compositing. For each pixel, we then compute

$$S = \min_f C_f \quad \text{and} \quad L = \max_f C_f, \tag{4}$$

where the min and max are computed across all frames f independently at each pixel. Given the color images C, L, and S, which can be thought of as 3-vectors at each pixel, we estimate the shadow matte β using

$$\beta = \frac{(C - S) \cdot (L - S)}{\|L - S\|^2}. \tag{5}$$

This equation simply projects the observed color C onto the color line between L and S and computes the parametric distance of the projection along that line. (It is also the least squares estimate of β given a noisy color C.)

The method works quite well where we have good estimates for both L and S. However, the β estimates are noisy where L and S are similar. This happens wherever some part of the background is never covered by the shadow or always lies in shadow. Where $L = S$ we cannot recover β at all, but this signifies that the pixel is completely unaffected by the presence of shadow, so we mark these pixels as unknown β. Small areas with unknown β can be filled in using inpainting [Bertalmio et al. 2000] or other hole-filling approaches, although such hole-filling was not necessary for the examples in this paper. Figure 3 illustrates the min and max composite and recovered β for a sample input. Note that our shadow compositing equation is derived in radiance space. In practice, we used the gamma-corrected pixel values directly but did not observe any visible artifacts.

3 Shadow compositing

To perform shadow compositing, we require the lit and shadow images L' and S' corresponding to the novel background scene. We assume that the novel background is captured with a static camera whose relation to the primary light source is the same as it was in the source scene. Equation (3) can then be used to calculate the composite color due to shadowing as a function of these two images as

$$C' = \beta L' + (1 - \beta)S'. \tag{6}$$

For a synthetic scene, it is easy to render both the lit and the shadowed versions of the scene. For a natural scene, we perform *photometric shadow scanning* by moving an object between the light and scene such that every part of the scene that we wish to composite shadows into is in shadow at some time. As in the shadow matting process previously described, we then use max/min compositing operations to recover the shadow and lit images corresponding to the scene.

With the recovered photometric parameters and source shadow matte β, we use the shadow compositing equation to make the composite as shown in Figure 4(a–c). The most noticeable flaw in this image is that the cast shadow does not conform to the geometry of the destination background scene as it should. We address this shortcoming in the following section.

4 Estimating shadow deformations

In this section, we show how to transfer a shadow cast on a source planar surface onto a target background with arbitrary geometry, assuming that some region of the target background has a planar region matching the source planar surface. To accomplish this, we construct a *displacement* or *warping map* W that places each pixel p in the target image in correspondence with some point $W[p]$ in the source image. Such a displacement map is sufficient because shadows cast from point light sources can be described using a 2D map and projection matrix, and the projected image of this map onto any geometry is some distortion of the map.

We use the same shadow compositing equation as in (6) but with a warped $\beta' = \beta[W[p]]$. Since the values of the displacement map $W[p]$ are not constrained to integer image coordinates, we use bilinear interpolation when resampling the shadow map β.

Our method for estimating the shadow displacement map is based on Bouguet's shadow scanning approach [1998]. Shadow scanning is an active illumination method, and as such its reconstructions contain gaps wherever surfaces are occluded from the view of either the light source or camera. Fortunately, in the context of shadow compositing, these gaps coincide exactly with regions of the image unaffected by shadows and with regions not visible to the camera. Therefore, shadow scanning is ideally suited for our application. Furthermore, we can avoid Bouguet's calibration of the camera and light source by not doing full 3D reconstruction. Instead, we perform multiple passes with different scan orientations in order to compute a 2D shadow warping function. We call this process *geometric shadow scanning* to distinguish it from the photometric shadow scanning described in the previous section, which recovers no geometric (deformation) properties of the scene.

As in Bouguet's work, we require that our target background include some planar region, which is specified by the user via a hand-drawn garbage matte (Figure 4(d)). We refer to this region as the *reference plane region* and denote the 3D plane defined by the points

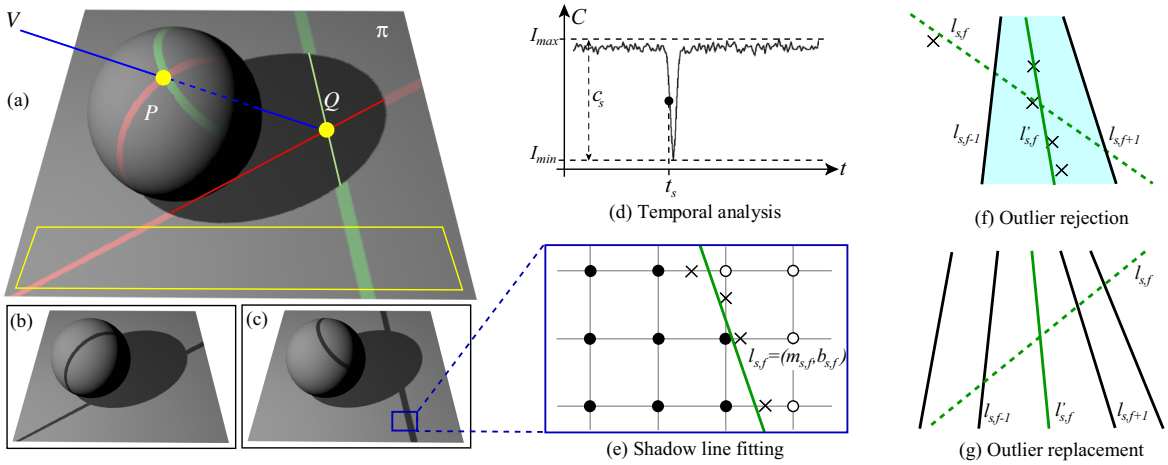


Figure 2 Illustration of the principle and details of geometric shadow scanning. Point Q is the point on the reference plane π that lies behind point P along ray V (a). It is found at the intersection of two shadow lines whose orientation is estimated from the reference planar region, outlined in light yellow. For reference, the pair of images below (b,c) shows 3D renderings of the two individual shadow lines. We estimate all shadow edge crossings using **temporal analysis** (d) to identify for each pixel p the shadow time $t_s[p]$, computed as the first time the pixel color goes below a threshold halfway between its min and max values, I_{min} and I_{max}. In the reference plane region, we can then determine the **shadow line** (e) for frame f of scan s, $l_{s,f} = (m_{s,f}, b_{s,f})$, by linearly interpolating between neighboring pixels with shadow times less than and greater than f (shown as black and white dots, respectively) and then fitting a line through the interpolated points. In some cases, line fits are poor due to spurious shadow time samples used in fitting. In the **outlier rejection** step (f), we identify lines with high error (the green dotted line $l_{s,f}$), discard samples outside the region defined by the nearest valid shadow lines on both sides (the cyan region between $l_{s,f-1}$ and $l_{s,f+1}$), and then re-fit the line (the green solid line $l'_{s,f}$). Inconsistent shadow lines may still occur when fitting to too few or closely spaced samples. In the **outlier replacement** step (g), we identify these lines (e.g. , the green dotted line $l_{s,f}$) by detecting rapid changes in line slopes between frames and then replace them by interpolating neighboring line coefficients. The solid green line $l'_{s,f}$ is the replacement computed by interpolating between $l_{s,f-1}$ and $l_{s,f+1}$.

in this region as π (Figure 2(a)). Consider a pixel p through which we can see a 3D point P in the target background. For such a point P, there is a projection Q on the reference plane that lies at the intersection of π and the shadow ray V that passes through P. Now consider a stick or rod that casts a shadow on P. By observing the shadow line on the reference plane, we know that Q must lie along this line. If we re-orient the stick such that the shadow still covers point P, we obtain a second shadow line on the reference plane that must also intersect Q. We can compute Q as simply the intersection of the two shadow lines on the reference plane. In our implementation, we solve directly for q, the image-space projection of Q, which is in fact the warping map $W[p]$, by performing all computations in image coordinates, as described in the rest of this section. (Note that the displacement between $W[p]$ and p is actually a measure of projective depth (parallax) in a plane-plus-parallax formulation of 3D reconstruction [Kumar et al. 1994].)

Acquisition. In practice, we use a series of directional scans, so that we can interpolate shadow lines for pixels that are not covered by a shadow edge at an integer frame time. Since the shadows may vary in width and the objects receiving the shadow may be very thin, we identify the leading temporal edge of the shadow lines, rather than trying to identify spatial shadow edges. Spatial edge detection is also less reliable in the presence of reflectance variations; the temporal analysis avoids this problem by focusing on a single pixel at a time.

To simplify the scanning process, a person moves the scanning stick by hand, potentially casting his or her own shadow onto the scene. Our algorithm is designed to ignore any additional moving shadows in the image frames, as long as the stick's shadow always appears to one side of the extra shadow. In principle, two passes suffice to find intersections, but we use three to five passes in order to improve the quality of the results and to cover more of the target background object. Figure 4(d) shows a frame from a geometric shadow scanning sequence. The user-specified reference plane region is shown with hatched blue lines.

Algorithm. Here is an outline of our algorithm, with details about each boldfaced step in the paragraphs that follow.

1. For each directional scan s:
   (a) Label each pixel location p with the continuous time $t_s[p]$ at which it is first crossed by the shadow edge using **temporal analysis**.
   (b) For each frame f, fit a **shadow line** equation $x = m_{s,f}y + b_{s,f}$ to the points r in the reference plane region labeled with time $t_s[r] = f$.
   (c) Perform **outlier rejection and replacement** on the line equation parameters $(m_{s,f}, b_{s,f})$ to reduce noise.
2. For each pixel location p:
   (a) For each scan s, interpolate the line equation parameters $(m_{s,\lfloor t_s[p] \rfloor}, b_{s,\lfloor t_s[p] \rfloor})$ and $(m_{s,\lceil t_s[p] \rceil}, b_{s,\lceil t_s[p] \rceil})$ from the nearest integer frames of the scan sequences to obtain the line equation parameters $(m_{s,t_s[p]}, b_{s,t_s[p]})$ for this pixel.
   (b) Compute the **intersection** q of lines for all scans s.
   (c) Store q as the value of the warp function for pixel p: $W[p] \leftarrow q$.
3. Optionally, perform **anisotropic diffusion** to smooth the displacement map $W[p]$ while maintaining its structure.

Details. We perform **temporal analysis** [Bouguet and Perona 1998] to identify the shadow time $t_s[p]$ (Figure 2(d)). The person holding the scanning stick stays behind the trailing edge of its shadow, so we find the first frame when the pixel color goes below a threshold halfway between its min and max values, I_{min} and I_{max}. We linearly interpolate from the previous frame time to compute the shadow edge time $t_s[p]$. Also, we define the *shadow contrast* $c_s[p]$ as $I_{max} - I_{min}$; this value is later used as a confidence measure.

For **shadow line** fitting, Bouguet [1998] used spatial analysis to find points on a planar region crossing the shadow edge and then fit those points to a line. However, Bouguet's method assumes that the reference plane is uniform in color and material. Since our goal is

to handle natural scenes with reflectance variations on the reference plane, we instead use temporal analysis to determine the shadow lines more accurately (Figure 2(e)). For each frame f, we check each pixel r within the reference planar region. For every neighboring pair of pixels for which the signs of $t_s[r] - f$ are different, we estimate the zero-crossing point and add it to a list of candidate points on the shadow edge. After finding all such zero-crossings, we fit them to a line using linear regression. This line is the shadow line for frame f of scan s, parameterized by the tuple $(m_{s,f}, b_{s,f})$. (This is similar to a 2D version of the marching cubes method [Lorensen and Cline 1987] for tracing isocurves, except we only collect the points, and need not connect them into a curve.)

Because of noise and bad shadow time estimates in the reference plane, some shadow lines may be poorly estimated. Accordingly, we perform **outlier rejection and replacement** to fix poorly fitted lines. For the lines whose fitting errors (as measured by the residual) are larger than some threshold, we reject all the points outside the image region defined by the nearest valid shadow lines on both sides and refit the line without those outliers (Figure 2(f)). Even after this refitting, some lines may still be poorly estimated, either because we have too few points to fit or because the points are too close together in the image. Since the orientation of the scanning stick varies smoothly across the sequence, we identify and reject poorly fit lines by noting rapid changes in the slopes m between frames. We then replace the rejected shadow line by interpolating the line coefficients of the neighboring frames (Figure 2(g)). (We have not found discontinuities in the intercept b values to be of use for outlier rejection, since the scanning may progress at widely varying rates.)

After interpolating the shadow lines from integer frames for each scan, we use weighted least squares to calculate the **intersection** point q of all the shadow lines corresponding to shadow edges crossing point p. We determine the weight for each line based on two factors. First, if the shadow contrast for the scan is low, the estimated shadow time $t_s[p]$ for that pixel will be less accurate. Second, if the shadow line was poorly fit, it might have a significant impact on the location of intersection. In practice, we weight each line by $w_c w_f$, where the first term w_c is defined as the square of the shadow contrast $c_s[p]^2$, and the second term, w_f, is defined as $\exp(-(E_{s,t_s})^2)$, where E_{s,t_s} is the interpolated line fitting error for scan s at time t_s.

In a final step, we apply an **anisotropic diffusion** algorithm to smooth the displacement map while maintaining its structure [Perona and Malik 1990]. Errors in the displacement map are most severe for pixels with low shadow contrast. Our diffusion process therefore flows data from high confidence areas to low confidence areas. Regions with very low contrast, e.g., areas that were already in shadow or were never shadowed, will be filled with nearby displacements that could be quite different from the true displacements. However, these areas will essentially be unaffected by shadows we want to transfer, so the values of the displacement map in those regions are, in principle, unimportant (see Section 7 for further discussion). In the end, the displacement map effectively places all pixels on the reference plane from which we extracted the source shadow matte (Figure 4(e)) to match the geometry of the target background (Figure 4(f)).

5 Results

To illustrate our method, we filmed some outdoor scenes in uncontrolled settings (Figure 3) and transferred the shadows to different target backgrounds (Figure 4(g,h)). Note that the source scenes were captured with complexly color-textured backgrounds. Video matting [Chuang et al. 2002] was employed to extract the matte of the foreground actor. Our outdoor destination backgrounds were scanned with three to five passes of a $96'' \times 3'' \times 1''$ wooden board (Figure 4(d)). We took care to sweep the scanning stick so as to cover all the regions where the replacement shadow might be cast, while simultaneously covering a reasonable part of the reference plane. In all our examples, we used a section of the ground plane as our reference plane, and our outdoor scenes were filmed at hours of the day when the sun lay low on the horizon, so that an 8-foot stick sufficed to scan an adequate region. In addition, we matched the camera pose and lighting simply by looking at a source image with a shadow on the ground plane, and then "eye-balling" how the camera should be placed in the target sequence by looking at another cast shadow.

Figure 5 demonstrates the advantages of our method over previous methods on a target scene with complex geometry. The surface of the ivy is bumpy and challenging to model manually. There is also severe self-shadowing in this scene (Figure 5(b)). Using previous methods, the composite would have double-shadowing in these regions. With our method, the shadows deform naturally to match the background geometry. Furthermore, there is no double-shadowing (Figure 5(c)). In Figure 6, we transferred an outdoor foreground to an indoor background scene scanned with wooden dowels.

6 Discussion

Our present matting and compositing method has a number of important restrictions. First, our shadow compositing equation (3) is strictly only valid for scenes with one dominant, pointlike light source, and it does not model potentially complex effects arising from interreflections. Second, to estimate the lit and shadow images, we require a static camera. Third, in order to construct the shadow displacement map, we require the source background to be planar and the target background to contain a planar reference region. Finally, we require that the relationship of the dominant light source, reference plane, and camera be matched in the source and target scenes.

Despite these restrictions, we believe that in many settings our approach provides a less restrictive capture mechanism compared to previous shadow extraction and compositing methods. Existing techniques typically share our requirement for a single matched key light source and matched cameras in order to avoid the difficult "foreground relighting" problem. However, they also require the construction and calibration of matching geometry (physical or virtual sets) and the use of painstakingly lit bluescreens for source capture. Our technique requires neither matching geometry nor blue screens.

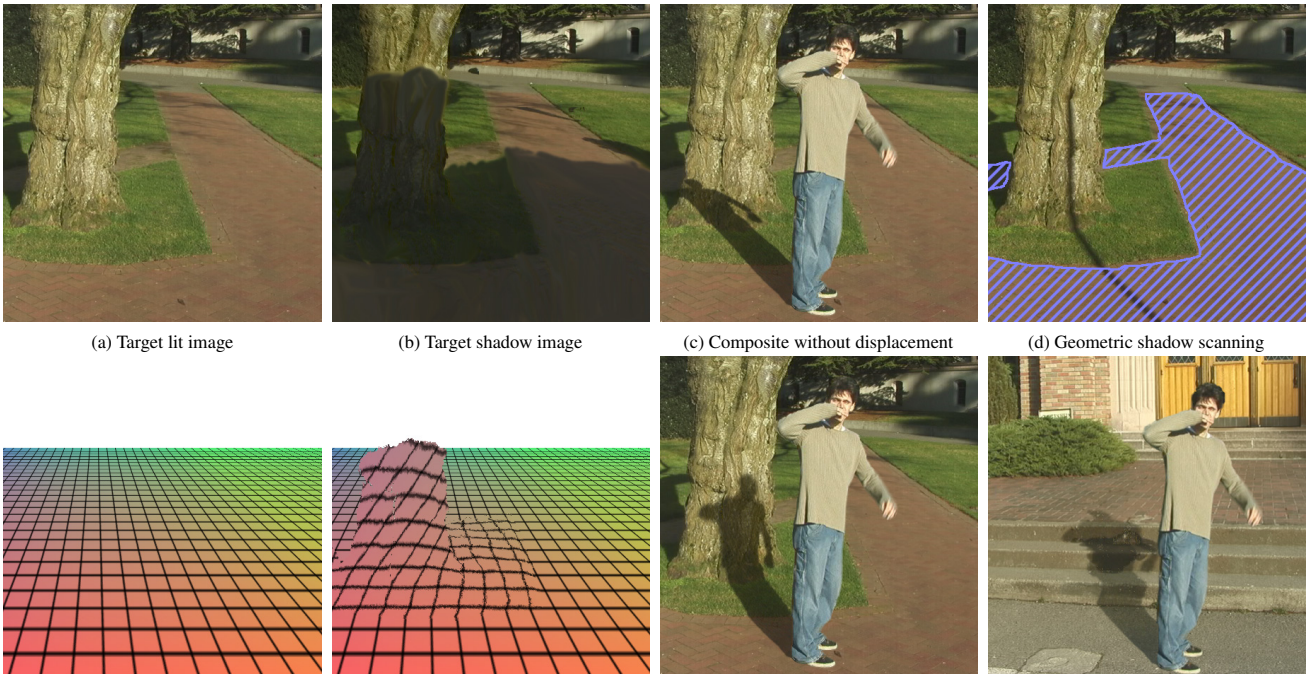
Furthermore, some of the theoretical restrictions on our technique can be relaxed in practice. For instance, the camera and light directions need not match precisely. As shown in the previous section, approximate matches were enough to create convincing composites. In addition, the dominant light sources in our scenes were not perfect point lights, but no objectionable artifacts were evident. For less point-like sources, we can transfer approximate penumbrae as long as the source and target backgrounds are at a similar distance to the casting object. We could potentially even blur or sharpen the shadow matte to create a *faux shadow* with a softer or harder penumbra.

7 Conclusions

In this paper, we have introduced a physically-based shadow matting and compositing method. Our approach has many advantages over previous approaches to shadow extraction and compositing. First, it can extract both the photometric and geometric information required for shadow compositing from natural (planar) scenes. Second, it can cast shadows onto scenes that already contain shadows, self-shadowing, and specularities. Third, it can cast shadows onto complex geometry without manually modeling the scene (at the price of having to do a shadow scan). Because we use the same camera to capture our warping function as we do to capture the images of the scene itself, we avoid the difficulties of having to accu-

(a) Source frame (b) Source lit image (c) Source shadow image (d) Recovered β

Figure 3 Shadow matting. Starting from a source image sequence (one frame shown in (a)), we first remove the foreground character using video matting. Our shadow matting algorithm recovers lit and shadow images (b,c) using max/min compositing. It then estimates β by projecting observed pixel colors onto the color lines between them (d).



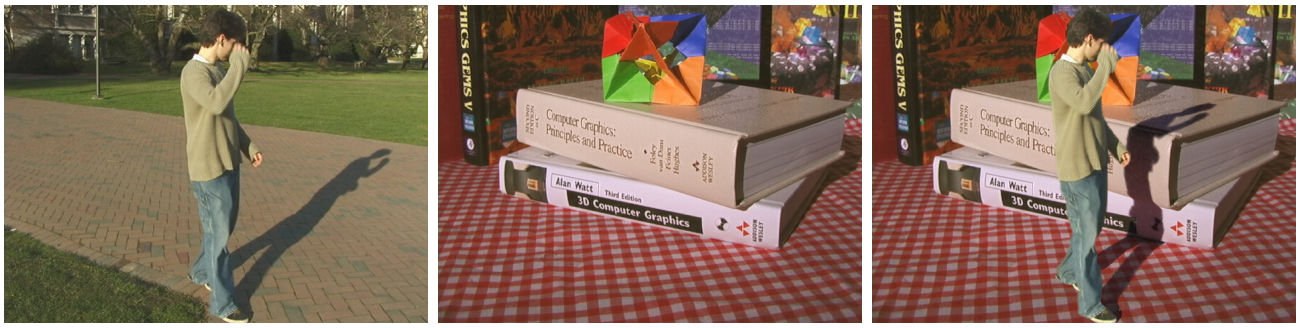
(a) Target lit image (b) Target shadow image (c) Composite without displacement (d) Geometric shadow scanning

(e) Regular grid on the reference plane (f) Displacement map visualization (g) Composite with displacement (h) Using another background

Figure 4 Shadow compositing. Lit and shadow images are recovered for target geometry as well (a,b). Our composite remains unconvincing because the shadow does not conform to the background geometry (c). We acquire multiple, oriented scans of a straight line shadow (d), and compute line equations in the user-specified reference plane region, shown here with hatched blue lines. We then recover a displacement map (f) from the target scene to the source reference plane (e). This map distorts the shadow into the correct shape (g). The results using a second background are shown in (h).

(a) Source frame (b) Target frame (c) Composite

Figure 5 An example of more complicated background geometry. The foreground element and shadow are extracted from a source frame (a) and transferred to a target frame (b) containing ivy to obtain a composite (c).



(a) Source frame (b) Target frame (c) Composite

Figure 6 *Honey, I Shrunk the Grad Student!* A source (a) and target (b) for a miniature composite (c).

rately register an independently reconstructed model to our image.

There are a number of ways in which our current approach can be extended. Due to sensor noise, min/max compositing can introduce artifacts, particularly in areas that are always in shadow in the target scene. In these areas, the brightness will vary with β in a perceptually objectionable way, since the underlying color is dark. Statistical methods could be employed to better estimate the mean shadow and lit colors.

We would like to relax the planarity constraints for the source background. For a source scene that is not fully planar, but has at least a planar segment, we could simply shadow scan the source scene to get its displacement map relative to its own ground plane, and use this to produce an unwarped (planar) shadow matte.

It would also be useful to have interactive editing tools for adjusting shadow direction and shape. Such tools could extend the operating range of our technique to cases in which lights or reference planes are somewhat misaligned between the source and target scene.

Finally, we would like to extend the operating range of the shadow matting and compositing equations, and preliminary experiments do suggest that at least some extension is possible. For instance, assuming the source and target backgrounds are Lambertian and geometrically similar, our method could still matte and composite plausible shadows cast by multiple light sources without taking separate images for each light source. In addition, our shadow density mattes are currently single-channel mattes bounded between 0 and 1. By using unbounded, multichannel β values, our method could also be modified to transfer approximate color-filtered shadows and caustics.

Acknowledgments

The authors would like to thank Benjamin Stewart for his acting. This work was supported by the University of Washington Animation Research Labs, NSF grant CCR-987365, and an industrial gift from Microsoft.

References

BERTALMIO, M., SAPIRO, G., CASELLES, V., AND BALLESTER, C. 2000. Image inpainting. In *Proceedings of SIGGRAPH 2000*, 417–424.

BOUGUET, J.-Y., AND PERONA, P. 1998. 3D photography on your desk. In *Proceedings of IEEE International Conference on Computer Vision (ICCV 98)*, 43–50.

CHUANG, Y.-Y., ZONGKER, D. E., HINDORFF, J., CURLESS, B., SALESIN, D. H., AND SZELISKI, R. 2000. Environment matting extensions: Towards higher accuracy and real-time capture. In *Proceedings of ACM SIGGRAPH 2000*, 121–130.

CHUANG, Y.-Y., CURLESS, B., SALESIN, D. H., AND SZELISKI, R. 2001. A Bayesian approach to digital matting. In *Proceedings of Computer Vision and Pattern Recognition (CVPR 2001)*, vol. II, 264–271.

CHUANG, Y.-Y., AGARWALA, A., CURLESS, B., SALESIN, D. H., AND SZELISKI, R. 2002. Video matting of complex scenes. *ACM Transactions on Graphics 21*, 3, 243–248.

DEBEVEC, P. 1998. Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proceedings of SIGGRAPH 98*, 189–198.

FINLAYSON, G. D., HORDLEY, S. D., AND DREW, M. S. 2002. Removing shadows from images. In *Proceedings of European Conference on Computer Vision (ECCV 2002)*, vol. 2353 of *LNCS*, 823–836.

KUMAR, R., ANANDAN, P., AND HANNA, K. 1994. Direct recovery of shape from multiple views: A parallax based approach. In *Twelfth International Conference on Pattern Recognition (ICPR'94)*, 685–688.

LORENSEN, W. E., AND CLINE, H. E. 1987. Marching cubes: A high resolution 3D surface construction algorithm. In *Computer Graphics (Proceedings of ACM SIGGRAPH '90)*, 163–169.

MATSUSHITA, Y., KANG, S. B., LIN, S., SHUM, H.-Y., AND TONG, X. 2002. Lighting interpolation by shadow morphing using intrinsic lumigraphs. In *Proceedings of Pacific Graphics 2002*, 58–65.

PELLACINI, F., TOLE, P., AND GREENBERG, D. P. 2002. A user interface for interactive cinematic shadow design. *ACM Transactions on Graphics 21*, 3, 563–566.

PERONA, P., AND MALIK, J. 1990. Scale space and edge detection using anisotropic diffusion. *IEEE Trans. on Pattern Analysis and Machine Intelligence 12*, 7 (July), 629–639.

PETROVIC, L., FUJITO, B., WILLIAMS, L., AND FINKELSTEIN, A. 2000. Shadows for cel animation. In *Siggraph 2000, Computer Graphics Proceedings*, K. Akeley, Ed., Annual Conference Series, 511–516.

PORTER, T., AND DUFF, T. 1984. Compositing digital images. In *Computer Graphics (Proceedings of ACM SIGGRAPH '84)*, 253–259.

RUZON, M. A., AND TOMASI, C. 2000. Alpha estimation in natural images. In *Proceedings of Computer Vision and Pattern Recognition (CVPR 2000)*, 18–25.

SAVARESE, S., RUSHMEIER, H., BERNARDINI, F., AND PERONA, P. 2001. Shadow carving. In *Proceedings of IEEE International Conference on Computer Vision (ICCV 2001)*, 190–197.

SMITH, A. R., AND BLINN, J. F. 1996. Blue screen matting. In *Proceedings of ACM SIGGRAPH 96*, 259–268.

SZELISKI, R., AVIDAN, S., AND ANANDAN, P. 2000. Layer extraction from multiple images containing reflections and transparency. In *Proceedings of Computer Vision and Pattern Recognition (CVPR 2000)*, 246–253.

WEISS, Y. 2001. Deriving intrinsic images from image sequences. In *Proceedings of IEEE International Conference on Computer Vision (ICCV 2001)*, 68–75.

WRIGHT, S. 2001. *Digital Compositing for Film and Video*. Focal Press.

ZONGKER, D. E., WERNER, D. M., CURLESS, B., AND SALESIN, D. H. 1999. Environment matting and compositing. In *Proceedings of ACM SIGGRAPH 99*, 205–214.