Beyond Perspective Dual Photography with Illumination Masks

Sanjeev Koppal* *Member, IEEE* and Srinivasa Narasimhan**, *Member, IEEE*

***Abstract*—Scene appearance from the point of view of a light source is called a reciprocal or dual view. Since there exists a large diversity in illumination, these virtual views may be non-perspective and multi-viewpoint in nature. In this paper, we demonstrate the use of occluding masks to recover these dual views, which we term shadow cameras. We first show how to render a single reciprocal scene view by swapping the camera and light source positions. We then extend this technique for multiple views by both building a virtual shadow camera array and by exploiting area sources. We also capture non-perspective views such as orthographic, cross-slit and a pushbroom variant, while introducing novel applications such as converting between camera projections and removing refractive and catadioptric distortions. Finally, since a shadow camera is artificial, we can manipulate any of its intrinsic parameters, such as camera skew, to create perspective distortions. We demonstrate a variety of indoor and outdoor results and show a rendering application for capturing the light-field of a light-source.**

I. What are shadow cameras?

Of all the light rays emitted, scattered or reflected by a scene, only the light-field measured by a camera can be directly accessed. In addition, when a programmable source, such as a projector, illuminates the scene, the incident light-field can also be known and controlled. In this paper, we introduce a third set of computable rays: the shadow-field of a non-programmable light source. The shadow-field is captured by using an occluding mask to block rays from the light source. Therefore, our representation is purely geometric and not photometric, since incident light rays are detected by *not* measuring them. Ray reciprocity allows us to describe these rays as if they were viewed by a virtual camera, which we term a **shadow camera**.

Although specific instances of shadow cameras have been used for scene reconstruction (as in [4]), we develop a view synthesis framework for image-based rendering. Figure 1 lists traditional approaches to the image-based rendering (IBR) problem. Historically, view synthesis involves either camera motion or multiple cameras, while scene relighting requires illumination control (as in a light stage [31]). Dual methods, in contrast, exploit Helmholtz reciprocity and treat light sources as cameras and vice-versa. Shadow cameras utilize reciprocity to allow flexible control of camera pixels, rearranging them in geometries determined by the relative source-mask motion. In this sense, we extend the dual photography technique [27] beyond programmable light sources.

*S. Koppal is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611, USA

**S. Narasimhan is with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Research area	Traditional approaches	Dual approaches
View Synthesis	Move camera (or use multiple cameras)	Vary illumination
Scene Relighting	Vary illumination	Move camera

Fig. 1. **Dual approaches for view synthesis and relighting:** Most view synthesis approaches require camera motion or multiple cameras. Similarly, traditional scene relighting techniques involve varying illumination. However, there also exist dual approaches that exploit the Helmholtz reciprocity between viewing and illumination rays. This paper describes a dual technique termed **shadow cameras**, which allows virtual scene views to be obtained by varying illumination with an occluding mask.

The recovered light ray geometry depends on the mask shape, the light source type and the relative motion between the two. The shadow camera is located at either the mask or the source, depending on which is stationary. In this paper, we focus on the use of linear masks (approximated in practice by thin, rigid wires). Two such masks, perpendicular to each other and moving with uniform motion, produce two distinct intensity minima at each scene point. The times at which these minima occur represent the horizontal and vertical image coordinates of a shadow camera centered at the source.

If a distant light source is used, the shadow camera is **orthographic** in nature and if a near light source is used the shadow camera is **perspective**. Multiple pairs of such perpendicular masks allow the creation of a virtual shadow camera array. Varying the mask speed/angle controls the shadow camera's intrinsic parameters (skew and scaling). Depending on the type of viewing camera, it becomes possible to switch between orthographic and perspective views.

We can also create shadow cameras with multiple viewpoints. Traditional **cross-slit** cameras are created by imaging through two perpendicular thin slits separated by some distance. Similarly, a cross-slit shadow camera is created when the motion paths of the two linear masks do not intersect and are separated. When compared to mosaicing cross-slits from many perspective views (as in [39]), our technique requires only a single reciprocal pair.

Traditional **push-broom** cameras are created by imaging through a translating slit. Similarly, we concatenate scene points shadowed by a single, translating linear mask to create a "shadow-pushbroom" camera.

Finally, we demonstrate a variety of applications showing the utility of shadow cameras. We consider the case of a camera viewing an object reflected through an unknown mirror surface. If the scene is illuminated by a point source, the computed dual view removes any distortions in the non-single viewpoint catadioptric image. Similar results are shown for refractive surfaces. In outdoor scenes, changing illumination can produce shadows that can be exploited to render virtual viewpoints. Finally, our method can be used to obtain the light-field of a light-source, allowing digital capture of real-world light sources for use in computer graphics.

Programmable vs. Non-programmable In our discussion, we make a distinction between light-sources whose illumination profile can be changed electronically or in software, such as a projector. We call these programmable sources. However, since we use a moving occluding mask in front of non-programmable light-sources, this can also be considered as a rudimentary form of control. Our terminology is meant to indicate that there exists a huge variety of light-sources (lava lamps, fluorescent tubelights, car headlights, disco illuminations, etc.) whose illumination geometry is fascinating and yet inaccessible. The shadow camera method allows us to extract this information with simply performing linear scans of an illumination mask, without any additional electronics or spatial light modulators such as LCDs or LCoS.

The shadow-camera technique is simple, easy to implement and requires only a background calibration plane to estimate the relative source-mask motion. It is applicable to many scenes that contain a "diffuse" component to their BRDFs, since this allows reliable detection of the mask's shadow. In addition, it is not necessary to store the complete video and instead, we can efficiently detect the moving shadow edge using only a window of a few frames. Finally, the shadow camera resolution depends on the edge detector quality and not on the shadow width.

A. Related work

Illumination masking follows a 'less is more' trend where coded camera apertures capture additional light-field information [23]. Shadows can be used for spatio-temporal correspondences across cameras [6]. Other related work includes:

Light-field rendering and Mosaicing: Light-field rendering [19],[10] obtains new views by interpolating between densely sampled images, with applications such as view synthesis [36], all-focus images [21], seeing past obstacles [32] and mosaicing both multi-viewpoint images [26] and new cameras (cross-slit [39], pushbroom [35]). Although any shadow cameras could be rendered by capturing the entire light field, our approach is more efficient since we only use a single reciprocal pair. A related work has used ambient shadows to obtain "hidden" images from real-world scenes [30], although these are of lower resolution since they utilize fewer photographs than the technique in this paper.

Shadow-based approaches: These techniques have been popular since they are invariant to material properties [24], [25], [17]. Linear masks have been previously used for scene point triangulation [4], [5],[18], while shadowgram methods reconstruct intricate objects [34] and multi-flash approaches obtain occluding edges [33]. In [1] shadow cues are used to obtain a geometric constraint for web-camera calibration. Related work has also been done to exploit cloud shadow patterns for applications in vision [13], [15], [14]. Although we enjoy the advantages of shadows, we differ from these methods since our goal is not scene reconstruction.

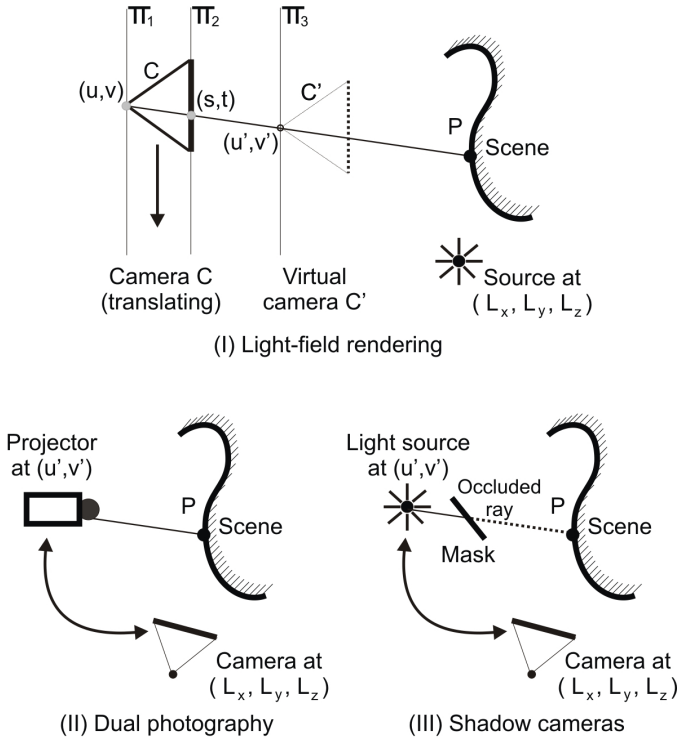


Fig. 2. **Dual views for non-programmable illumination:** Traditional image-based rendering techniques (IBR) seek to capture the entire light field so that any desired virtual view can be rendered, as in (I). Dual photography (II) reduces the number of images required by placing a projector at the desired viewing location. However, for a single projector position, dual photography is restricted to a perspective view. Our method (III) recovers the dual view by using an illumination mask. Depending on the relative source-mask motion, the shadow cameras can be non-perspective and multi-viewpoint in nature.

Dual views from programmable light sources: Dual photography [27], [8] exploits Helmholtz reciprocity [12], [37] between every projector-camera pixel pair to render the scene from the projector's point-of-view. Since the full light transport is measured, any scene relighting technique can be applied. In contrast, our method only obtains a dual view of the scene. However, it has the advantage of allowing non-programmable illumination and can create multi-viewpoint cameras from a *single* reciprocal pair. To create similar images using [27] we would need a "dual video" created while the projector moves smoothly in a line or plane, which could be prohibitive.

II. Light-source centric image-based rendering

Let a scene be illuminated by a source located at $L = (L_x, L_y, L_z)$, and imaged by a pin-hole camera C moving along on a plane Π_1 whose location is given by (u, v) (Figure 2(I)). The moving camera samples the 4D light field at rays specified by their points of intersection, (u, v, s, t), on Π_1 and a parallel plane Π_2. To render any new scene view we select

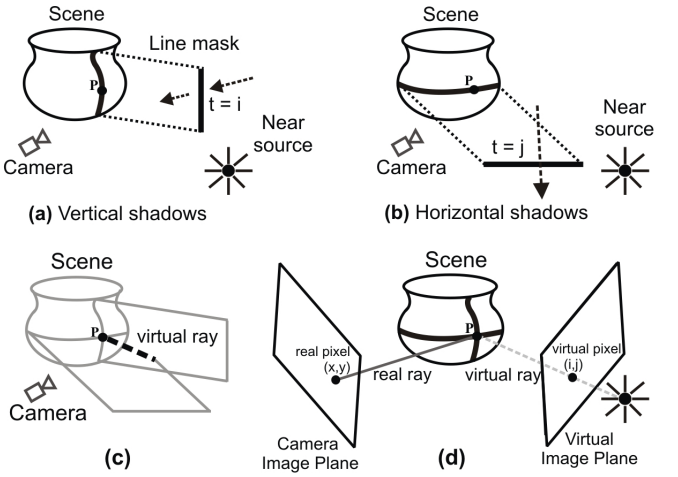


Fig. 3. **Virtual perspective view using linear masks:** In (I) we show a ray diagram for a perspective shadow camera. Using linear masks we obtain the input images, (III), which create a dual view in (IV). Note that foreshortening effects exist (shortened legs, extended scales), that specularity locations do not change and that shadows in (II) are occluded in (IV).

light rays that describe the virtual camera's *caustic*, which is a curve in space that all the light rays must be tangent to [29],[16],[11].

For the virtual pin-hole camera $C^{'}$ in the figure, the caustic degenerates to a point in space (the camera center) which lies on a third plane Π_3 parallel to Π_1, at position $(u^{'}, v^{'})$. The key advantage of traditional camera-centric IBR is that a desired caustic can be created without knowing scene shape. The trade-off is that the entire light field has to be captured, requiring many images. This can be reduced by making scene BRDF assumptions [28] or by finding correspondences across fewer samples by statistical modeling [38].

Dual photography solves these issues by replacing the source at (L_x, L_y, L_z) with a camera and placing a projector at the desired virtual camera location $(u^{'}, v^{'})$, as shown in Figure 2(II). If the camera image is given by $I(x, y)$ and a virtual image at the projector is denoted by $I^{'}(x^{'}, y^{'})$, then Helmholtz reciprocity relates these as $I^{'}(x^{'}, y^{'}) = I(x, y)$ where $(x, y) \leftrightarrow (x^{'}, y^{'})$ are corresponding projector-camera pixels. Surface normals, required in Helmholtz stereopsis, are not computed since the expression for pixel irradiance $I(x, y)$ contains both illumination and viewing foreshortening (see Appendix at [27]).

Dual photography's limitation is that, for a given projector position, it recovers a single *perspective* view of the scene. To create virtual views for other camera caustics, we have to imitate the traditional IBR setup in Figure 2(I), with a translating projector instead of a moving camera, capturing the virtual image at each projector location. Instead, we wish to more fully and efficiently exploit Helmholtz reciprocity by creating dual views directly for non-perspective, non-programmable light-sources. The major challenge is finding correspondences between the real view and the dual view without the control a projector affords.

A. Ray correspondence through shadows

Consider now a non-programmable point light source placed at $(u^{'}, v^{'})$, as in Figure 2(III). We move an opaque mask in front of the light source, creating a shadow that falls on the scene. A scene point P, located in the camera image I at pixel (x, y), displays a minima in its measured intensity at time t if it is occluded by the mask. Let R_P be the set of all time instances when such intensity minima occur at P. If, for every pair of scene points P and Q, $R_P \cap R_Q = \emptyset$, then R_P uniquely identifies a ray from the light source to P. Therefore each incident light ray at the scene is assigned a unique ID that corresponds to the locations in time when it was occluded.

If we know the caustic of the illuminant (defined similarly to a camera caustic) then we can map the identifier R_P to a pixel on the virtual image $I^{'}$, $R_P \rightarrow (x^{'}, y^{'})$. The intensity at this virtual pixel is found by exploiting Helmholtz reciprocity such that $I^{'}(x^{'}, y^{'}) = I(x, y)$ for the correspondence $(x, y) \leftrightarrow (x^{'}, y^{'})$ between real camera pixels and the virtual camera pixels, *exactly* as in dual photography.

Finding the mapping R_P can be done in two ways. The first is to design the mask motion such that the minima location gives the virtual image pixel directly. Since the pixel location has two degrees of freedom, we need at least two minima locations in time to do this. In Figure 3(I) we show a ray diagram of the shadow cast by a linear mask translating in front of a static point light source. Since the shadow hull of a line mask is a plane, the locations in time of two intersecting perpendicular plane shadow hulls specifies the horizontal and vertical coordinates of a virtual image pixel.

The second method to find R_P generalizes beyond linear masks but requires a simple calibration step. The experiment must be repeated twice, first with the actual scene and the second time with a plane placed at the location where we

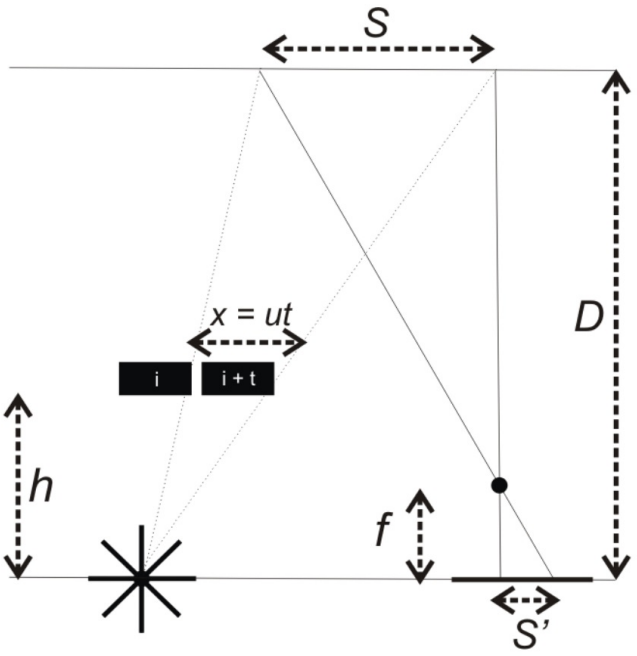


Fig. 4. **Effective resolution of the dual view:** The resolution of the rendered image from the light-source point-of-view depends on the speed of the mask and the greatest depth in the scene. In this case we assume the camera image plane is parallel to a plane passing through the scene point at depth D, but this can be relaxed with an extra camera skew term.

wish the virtual image plane to be. If the plane is visible to the camera and the motion of the mask is identical in each experiment, then we can map each scene point to a coordinate on calibration plane.

The minima detection is independent of both BRDF variation and intensity-fall off and allows cast/attached shadow disambiguation. The virtual camera resolution depends on the shadow edge detector and not the shadow width and is theoretically only limited by the camera resolution. Our method also handles multiple sources since these would create many intensity minima scene points, each of which gives a different view point. Similarly, we can detect a "minima interval" for area sources corresponding to a continuous set of viewpoints, and we show these results later.

B. Shadow camera resolution

In Figure 4 we show a 2D slice of a 3D scene where a shadow camera is being captured. We consider the sampled resolution by looking at the portion of the mask that is occluding the most distant part of the scene. This provides a bound on the resolution that can be obtained. The mask moves at a speed of u units per second. The camera exposure time is t, during which the mask moves a distance of $x = ut$. Since we only detect the shadow edge, S determines the scene points shadowed during a single second on a plane at depth D from the camera. Let D be the greatest depth that exists in the scene. Without losing generality, we can assume the camera image plane is parallel to the plane at D. This can be relaxed by adding a skew term to the following equations. Using similar triangles:

$$S = \frac{D\ x}{h}. \tag{1}$$

This is projected onto the camera image plane (of focal length f):

$$S' = \frac{f\ D\ x}{h\ (D - f)}. \tag{2}$$

If the width of a pixel (on the image plane) is given by W and replacing x by the mask speed and camera frame rate:

$$r = \frac{W\ h\ (D - f)}{f\ D\ u\ t}. \tag{3}$$

where r is the effective change in resolution in *each dimension* from the primal view to the dual view. Therefore if $r = 0.5$ the dual view is $\frac{1}{4}$ the resolution of the primal view. Note that r is inversely proportional to the speed of the mask (faster masks mean lower resolution) and directly proportional to the camera frame-rate (faster frame rate cameras mean higher resolution). Of course, the detectable resolution of the shadow camera is bound by the actual resolution of the real camera viewing the scene.

C. Shadow Camera Parameter Simulations

In Fig. 5 we show a variety of simulations using the PBRT rendering tool [22]. We simulated the Stanford bunny using a non-lambertian "acryl blue" BRDF available in PBRT. We placed the camera location at $(0.04, 0.13, 0.15)$ and translated the light source through $(0.1, 0.07, 0)$ from that location, which is approximately to the right and above the camera. In Fig. 5(I) we show linear mask renderings of size 0.0015 in PBRT units (which has a scale factor dependence to actual units) that generate a 100×100 shadow camera image of the Stanford bunny. We compare this to actually switching the camera and light source positions in PBRT. The images are geometrically similar but appearances vary slightly due to the near-lighting (inverse square fall-off) effect of the source. We used Euclidean distance measurements on the images directly as a image correlation metric, but other metrics may be used. We varied four different parameters and looked at how this error changes. Each graph in the figure has two curves, one using the entire image to calculate the error and the other using a mask provided by the ground truth reciprocal view. The four experiments are:

Varying Mask Speed: In Fig. 5(II) we render images as the mask speed changes. Assuming a constant camera frame rate and no motion blur, this translates into changing the number of frames that are used in the shadow camera algorithm, and therefore the resolution of the shadow camera image. The speed changes were done for ten sets of images by varying the image resolutions from 100×100 to 5×5 in constant steps. In (II) the increase in speed results in a decrease in resolution and subsequent increase in image error.

Varying Mask Thickness: In Fig. 5(III) we vary the mask thickness, and the graph has inset figures that show how mask variation occurs. The starting and ending locations of the masks were such that, even though the mask size changed, no scene point began or ended in shadow. We varied the mask size for ten cases from 0.0015 units to 0.015 units in constant steps. The image resolution was held constant at 50×50.

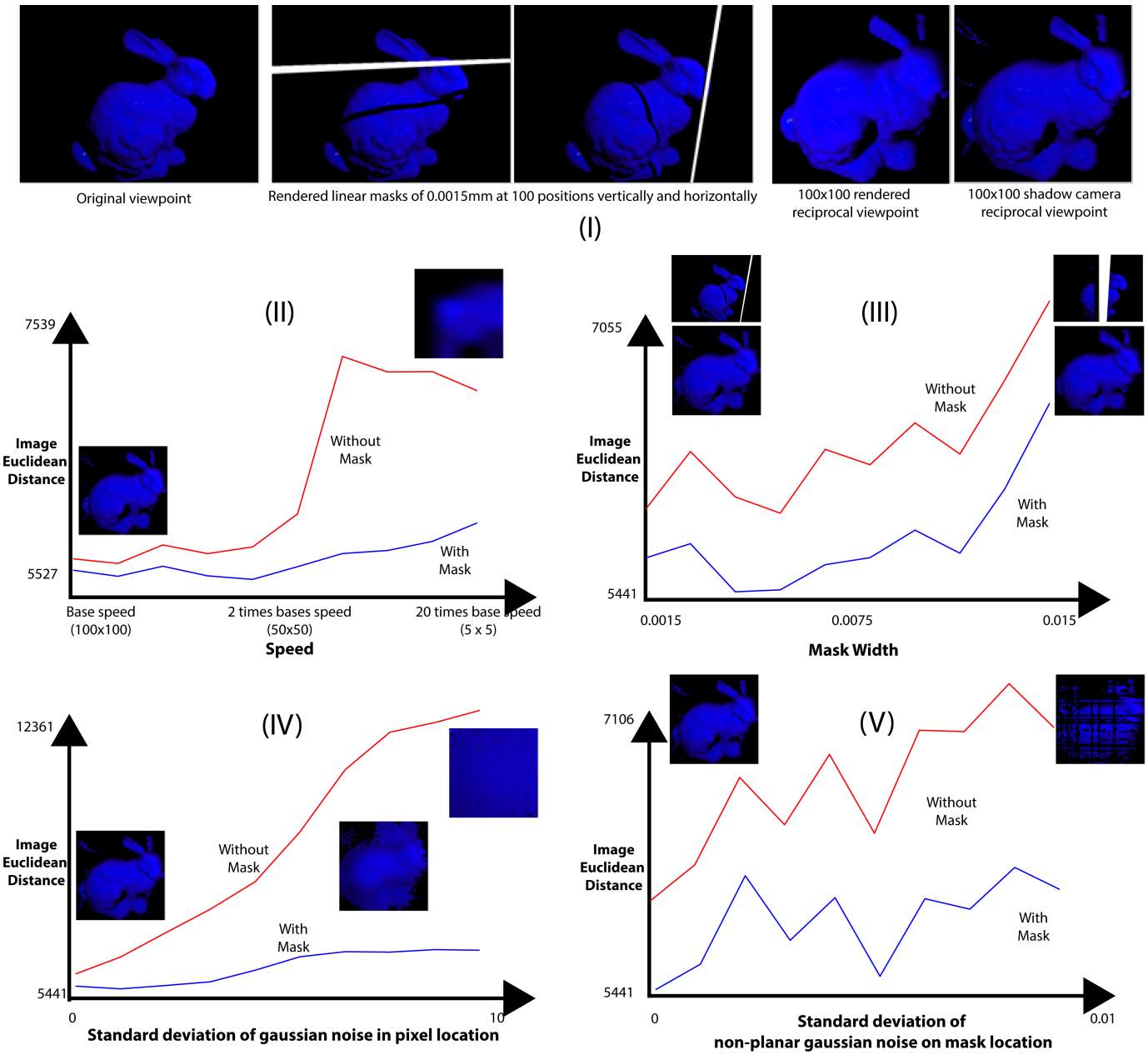


Fig. 5. **Quantitative Evaluation of Shadow Cameras:** Here we evaluate some shadow camera parameters using PBRT simulations [22]. In (I) we show our simulation setup, where the Stanford bunny is illuminated by a near light source that is to the right and upwards from the camera location. We render opaque linear masks moving in front of the light source. The right of (I) shows our shadow camera image and the reciprocal rendering obtained by switching the camera and light source in PBRT. (II) shows the change in quality of the result when the mask speed is increased. (III) shows the variation due to increase in mask thickness. (IV) shows what happens when noise is introduced into the shadow minima estimation. (V) demonstrates the variation due to noise in the mask motion that results in non-planar movement. In all cases, these effects reduce image resolution and increase error. Each graph has two curves, one using the entire image to calculate the error and the other using a mask provided by the ground truth reciprocal view.

The uncertainty in shadow locations results in a decrease in resolution and an increase in error.

Varying Shadow Location Accuracy: In Fig. 5(IV) we start with a 50×50 shadow image created by a 0.0015 unit mask. We add gaussian noise to the shadow minima locations in time. The standard deviation is varied from 0 (no noise) to 10 pixels. This creates noise in the rendered shadow camera image and the results demonstrate an increase in error.

Varying Mask Path (Non-planar motion): In Fig. 5(V) we investigate what happens when the mask does not move in a planar motion. We assume the mask has infinite length in one dimension, so noise along this dimension is not important. We assume the mask moves with constant velocity in the desired direction along a plane, and we do not add noise in this dimension. This leaves a single dimension with which to add noise. Note that even though noise is added along one vector, this results in change in the mask position in 3D, since this vector is from the light source to the mask and is not generally aligned with an axis. The results show that this non-planar noise results in increase in error.

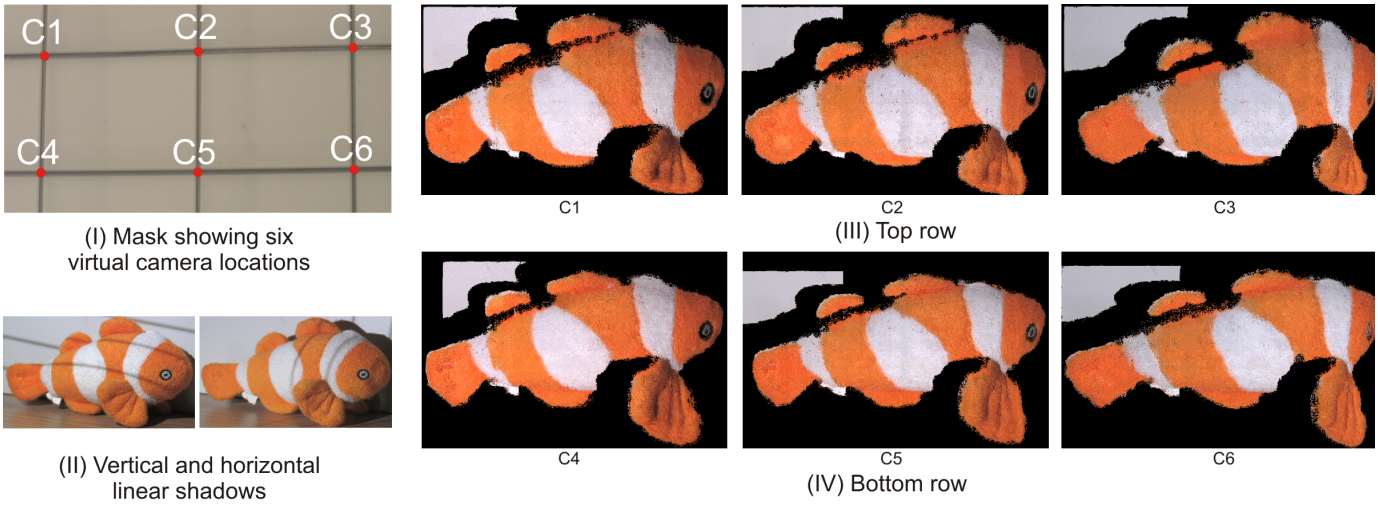


Fig. 6. **Virtual camera array from multiple linear masks:** In (I) we show a grid mask with six intersections. We collect data using the horizontal and vertical masks separately, as shown in (II). We then place light sources at the six intersections of the mask, capturing six images. We swap these images using Helmholtz reciprocity and the minima locations of the mask shadows. The result is a virtual camera array, and in (III) and (IV) we show the two rows of the array. Note the images are stereo rectified both horizontal and vertically.

Conclusions: Not surprisingly, each of the parameters affects the resolution of the rendered shadow camera image, increasing the error with respect to ground truth. What is interesting is the way in which each of these are affected. The location of the shadow minima and the (related) integrity of the planar motion of the mask seem very important. If these become problematic (and practically these are the ones that are most affected) then the quality of the results falls dramatically. The other two parameters, mask motion and thickness, seem to be those that should be set properly before capture, much like calibration. The non-planar noise seems to plateau after a point, indicating that a certain minimum accuracy is required. With a high enough frame rate camera and a thin enough mask relative to scene structures, these should not be a problem. Finally, we note that the high frequency effect in the graph shown in Fig. 5(V) is due to the stochastic nature of the experiment and the general trend is still an increase in error with decreased certainty in mask location.

III. Perspective Shadow Cameras

Our experiments were performed using a 12 bit Canon XL2 video camera running at 30fps. We selected a 1mm diameter piano wire for the mask, held by a Manfrotto tripod that allows controlled height adjustment and our scenes were illuminated by a Lamina ceramics DK4 LED. We placed a plane behind the scene, to estimate the shadow locations and uniformly interpolated these to get proper sampling for the virtual image.

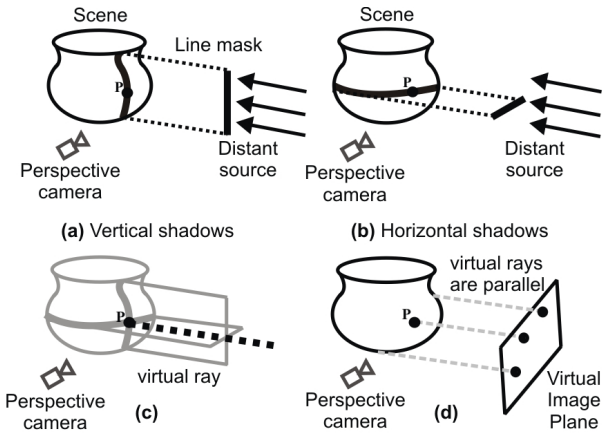
In Figure 3 (I) we show a ray diagram for a perspective shadow camera. I(a) and I(b) show two translating linear masks occluding P at times i and j respectively. The intersection of the shadows associated with the two masks is a virtual ray, shown at I(c). This shadow ray passes through the light source, and is associated with a pixel (i,j) in a virtual image at I(d). In (II) we show a non-convex plastic object. Using linear masks we obtain the input images, shown at (III). In Figure 3 (IV) we show a dual view for a non-convex plastic toy obtained using a set of two linear masks. Note the viewing foreshortening effects since the light source is higher than the camera position, such as shortening of the legs and the extension of the scales. In addition, the shadows of the dinosaur's tail are occluded in the dual image and vice-versa.

Multiple shadow cameras are possible by moving the source and performing experiments at the new light locations. We also demonstrate a different setup, where the mask is static and the light source moves. For example, in Figure 6 (I) we show a set of horizontal and vertical linear masks that are illuminated by a light source moving in a line. The shadows of the mask can be seen in Figure 6 (II) and horizontal and vertical experiments are performed separately to enable easy minima detection. By recording the times at which the horizontal and vertical masks occlude every scene point, we can recover virtual shadow cameras at the intersection of the horizontal and vertical masks.

Extra steps are required when moving the source instead of the mask. Since the scene intensities vary, shadow detection becomes harder. We also need additional images taken with the light source placed at each of the mask's grid intersections. However, the advantage is that a single experiment produces multiple viewpoints: more precisely, if n is the number of masks, we obtain an n^2 virtual array of shadow cameras. Six such images are shown in Figure 6 (III) and (IV), each of which are rectified horizontally and vertically since the wires in the mask array are perpendicular to each other.

IV. Multi-view and non-perspective shadow cameras

An indication of the potential of shadow cameras is that even for the limited case of linear masks and point light source discussed in this paper, we demonstrate many multi-view and non-perspective cameras. Figure 7 (I) shows that a ray diagram describing orthographic virtual view can be obtained from a perspective viewing camera by using a distant light source. In Figure 7 (II) we show a real view of plastic toy shark under distant (orthographic) illumination. Note the specularities on



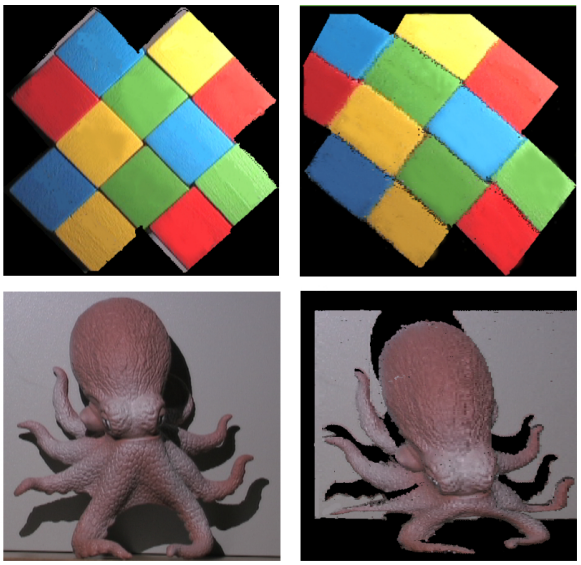
(I) Orthographic shadow camera from distant light source

(II) Original perspective view with orthographic light-source

(III) Dual orthographic view illuminated by perspective source

(IV) Real orthographic view by co-locating source and camera

Fig. 7. **Validating the dual view by perspective to orthographic conversion:** In (I) we show how the dual view due to a distant light source is orthographic. In (II) we show the perspective view of a plastic shark illuminated by a distant light source. Note the perspective effects since the shark's snout is close to the camera. In (III) we show the dual orthographic view. This compares reasonably with (IV) where we co-located the light source and the camera using a half-mirror (note the lack of shadows). The co-location does not move the light-source, explaining the difference in specularities between (III) and (IV).



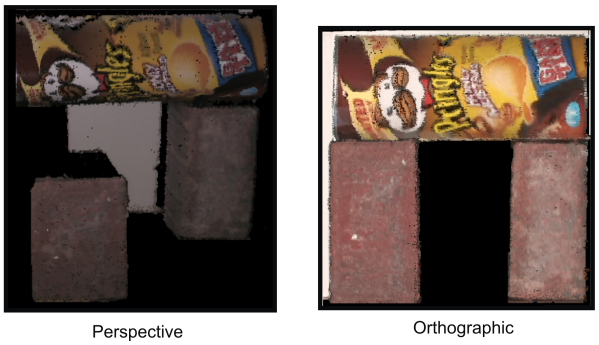
Original Dual

(I) Orthographic view/near light to perspective view/distant light

Fig. 8. **Orthographic to Perspective conversion:** In left column we show the original images viewed by an orthographic camera. In the right column, the dual views have foreshortening effects associated with perspective cameras.

Orthographic Perspective

Original images



Perspective Orthographic

Dual images

Fig. 9. **Orthographic/Perspective switching:** An illusion is created by orthographic viewing of two bricks and a can. Only one brick is actually holding up the can. This effect is broken when the camera comes near the scene due to perspective viewing. We are able to create and break this effect respectively, using the shadow camera rendering of the scene.

the shark and the shadows are correctly demonstrated in the dual view in Figure 7 (III). To validate the correctness of this view, we co-located the light-source and a camera using a half-mirror. Note the image in Figure 7 (IV) has no shadows. The shape of the shark is qualitatively similar in Figure 7 (III) and (IV). However, the appearance of the object is different since co-location is not the same as switching the source and camera positions. Despite this, the comparison demonstrates the correctness of the dual view.

Shadow cameras also allow us to demonstrate the opposite effect by switching from an orthographic view of the scene to a perspective views as shown in Figure 8. The perspective distortions are clear in the dual views, such as the change in angles of colored squares away from 90 degrees and the foreshortening when looking down at the octopus. The other aspect of the switch is that the illumination also changes: for example, in the dual view of the toy octopus the shadows are smaller, due to the orthographic source. An additional result shows how this type of renderings can break illusions that occur due to orthographic viewing. In Fig. 9 we show an example where two objects appear to touch, when in reality they do not. Switching to the reciprocal view allows the extraction of geometric information that was present in the original viewpoint and recovered by the motion of the illumination masks.

A multi-perspective cross-slit image of the scene can be created if the two linear masks do not intersect at a point, and are instead shifted by some amount, as in Figure 10 (II). Mathematically, this is identical to having an image plane intersected by rays passing through two slits if the light source is moved between experiments. In Figure 10 (III) we show such an image: note that the head of the octopus appears 'unwrapped' as if we are viewing all around it while the colored squares are curved.

We also introduce a new shadow camera view inspired by the real pushbroom camera, which is usually created by imaging a translating slit. In our case we take the scene points corresponding to the shadow of a single translating linear mask and concatenate them together, as illustrated in Figure 10 (IV). The image, shown in Figure 10 (V), appears to be lit with light sources at both the camera and the light source positions. Note the double shadows of each of the tentacles of the octopus. We call this camera the 'shadow pushbroom' camera. In Fig. 11 we show another example of this camera for a real-world illuminant, which is an overhead light-fixture. In this example, two tubelights create multiple shadows for a linear occluder. We use these to create two pushbroom images of a box. Note that relative orientation of the shadow and the fluorescent lights is important. If these are perpendicular to each other, the signal-to-noise ratio of the shadow will be low, resulting in errors in the shadow detection. Such scenarios are discussed further in the discussion and limitations section at the end of this paper.

Since shadow cameras are completely virtual, we can change their intrinsic parameters such as image skew or scaling. In Figure 12 we describe some of these perspective distortions applied to a scene. In Figure 12 (I) we show the dual view when the linear masks are perpendicular and move

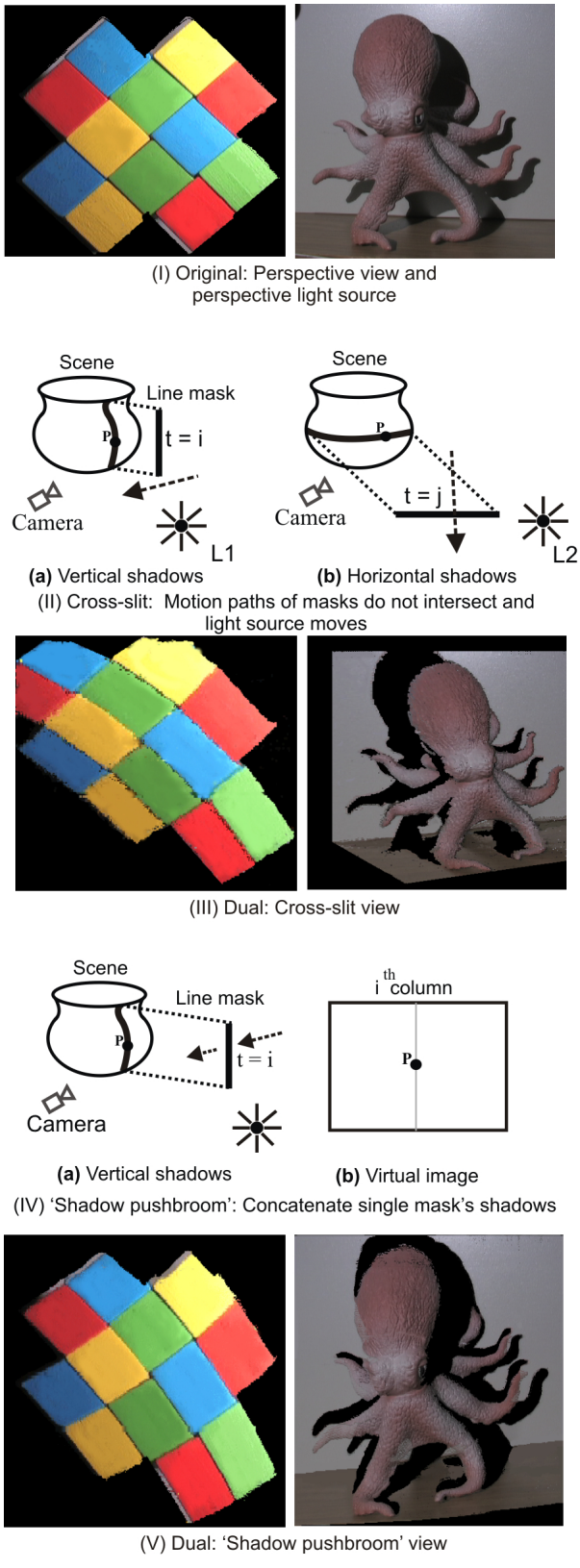


Fig. 10. **Non-perspective and multi-view cameras:** In (II) we show a cross-slit view created when the motion vectors of the linear masks do not intersect and the light source is moved between experiments. We demonstrate the distortion effects in (III). In (IV) we create a pushbroom-like view using the shadow from a single moving mask. Note the octopus in (V) appears illuminated by two light sources and each tentacle has two shadows.

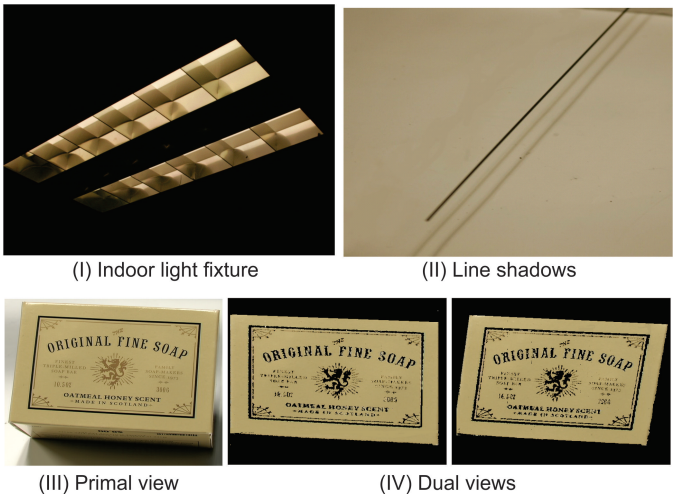


Fig. 11. **Shadow pushbroom cameras with fluorescent lighting in an indoor scene:** In (I) we show an image of a ceiling indoor light-fixture. The tubelights in the fixture act as line illuminants, producing a line-shadow (as in (II)). We use this to produce shadow pushbroom views of the scene shown in (IV).

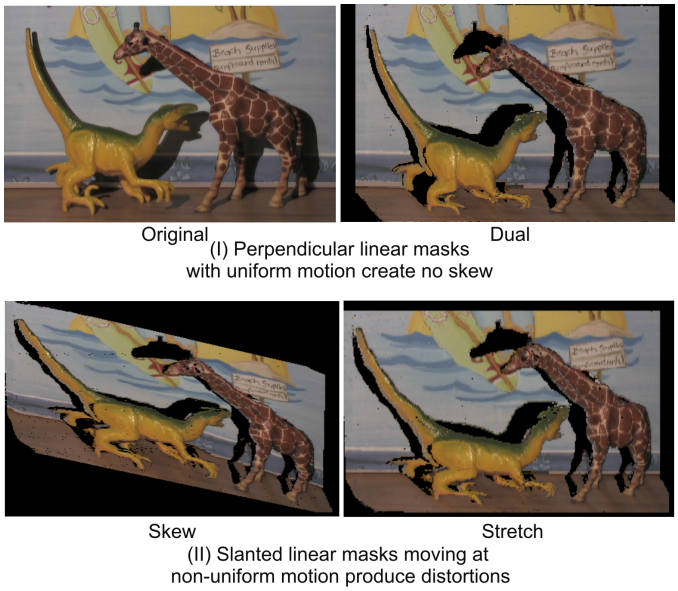


Fig. 12. **Controlling intrinsic parameters of shadow cameras:** Since shadow cameras are virtual, we have control over the locations of the pixels in space. In (I) we show a dual view created by two linear masks moving with uniform motion at 90 degrees to each other. Varying the speed of these two perpendicular masks or adding a third linear mask that is not perpendicular to either of the first two results in skewed/stretched images as seen in (II).

with uniform velocity, showing no distortions. We performed a third shadowing experiment with a slanted mask, which created virtual images with a non-zero pixel skew as shown at the left of Figure 12 (II). We also repeated the experiment with a horizontal mask moving slower than the vertical mask, resulting in higher horizontal sampling and a stretched image at the right of Figure 12 (II).

Finally, wide area and large masks can create ambiguity in the selection of a minima in intensity. Fig. 13 illustrates that arbitrary selection of the minima can result in different viewpoints of the scene. This illustrates the fact that shadows merely offer a geometric cue to recreate a viewpoint, and that any viewpoint can be recreated with similar or even much weaker cues.

V. An imaging application: Seeing through refraction and reflection

Various types of static optical elements exist whose material properties allow the bending of light. These include reflective surfaces, such as mirrors, as well as refractive solids such as glass and clear plastic. There are many situations where images are taken through these media: for example, catadioptric cameras ([29]) contain designed reflections that increase the camera field-of-view. However, unintended distortions are also possible due to lens imperfections and the presence of transparent occluders. Unwarping these effects in images without any prior knowledge of the geometry of the optical elements is almost impossible.

If the light from the source is not affected by the optical elements that cause the distortion then shadow cameras can be used to render an warp-free image of scene. This assumption is broad and is true for both catadioptric cameras and lens distortions. In Figure 14 (II), objects are reflected off a spherical mirror and the straight lines of the colored squares appear curved. Since the light source is perspective, these distortions are removed in the dual views in Figure 14 (III). In Figure 15 we show a planar scene with writing viewed through thick glass objects. Note the heavy distortion, especially in the glass on the right. Figure 15(II) shows the undistorted view, with readable text.

VI. A rendering application: Capturing the light-field of a light-source

So far we have presented methods to capture portions of the light-field that are incident on the scene. However, we can capture the light-field of the light-source itself, through the use of a novel imaging setup, described in Fig. 17. This can have application in graphics, where there is a need to digitally recreate complex, real-world sources, from LEDs and lava lamps to car-headlights and water caustics.

We place the camera and a light-source on opposite sides of a translucent screen. The source creates a pattern on the screen, and we capture such illumination patterns for two screen positions. Correspondences between the two illumination patterns are obtained using the horizontal and vertical shadows of linear mask motion, without changing the positions of the light-source. Our light-source light-field capture method uses four 1D sweeps of the linear opaque mask, which differs from other approaches that use time-intensive raster scans. For example, [2] requires obtaining many samples across the scene using a goniometer. In [9] the number of samples is reduced by using k basis filters to obtain MN samples (where M and N are the image rows and columns in the sampling camera). Since we scan with a linear mask, we need just $2M + 2N$ samples, which is much less than kMN samples in [9], especially when we note that the number of basis filters k increases as the geometry of the source becomes complex.

We do not smooth the illumination and recover high-frequency information when we render a new screen location.

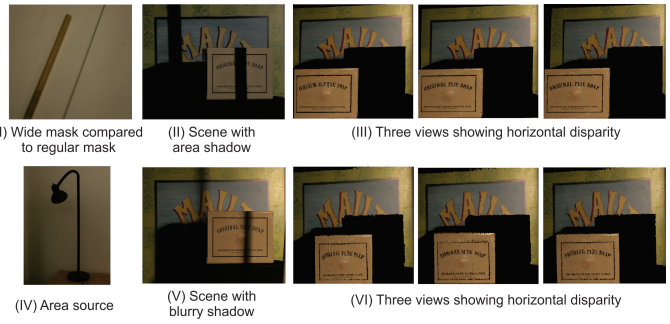


Fig. 13. **The advantages of wide area and blurry shadows:** We use both a large mask and an area light source to illustrate the situations where a sharp shadow peak does not happen. In these scenarios, the dual image that is created depends on the selection of the peak. We demonstrate that arbitrary selections of the peak give us multiple viewpoints.

This is because our filter is in fact a linear mask which has a high-spatial frequency itself. Finally, since we obtain correspondence between the two planes, measured by a camera, we can model the intensity fall-off along each ray given the distance between the planes. This allows us to extrapolate the light-field anywhere in the 3D space, and therefore the digital model captured can be used in any rendered scene.

In Fig. 17 we show the reconstruction of a flash-light using our method. Since the dynamic range of the display is much lower than that of the light source, we show an HDR set of four different exposures - overexposed pixels are not used in any rendering application. The first and third rows are inputs, whereas the second row is a rendered image. While a major weakness of this experiment is that we did not compare with the captured images at this location, we note that fine features in the center and periphery of the light-source are preserved. We utilized the digital model of the light-source to illuminate a rendered object in the last row of the figure, using a well-known rendering tool [22].

VII. Applications of Shadow Pushbroom Cameras for Outdoor Scenes

Of all the shadow cameras demonstrated in this paper, the shadow pushbroom camera is possibly the one most likely to appear in unstructured real-world scenes. This is because it simply requires either a linear mask or the edge of a planar mask to occlude the light-source.

In Fig. 18 we show samples from a time-lapse video of a mountain cliff as the sun sets. The shadow of another mountain, most likely the one on which the camera is placed, slowly engulfs the imaged cliffs, until all illumination is blocked. Unlike our experiments with thin piano wire and a point source, the shadow in these images is created with an area source (the sun) and the rough edge of natural rock.

We process the video a each frame to obtain the shadow edge. We use differences in subsequent frames to narrow the region where the shadow edge must be. This reduces the effect of sky illumination, since the images (on the whole) get darker as time goes on. To avoid temporal gaps in the estimated shadows between frames, we hallucinate edges in a keyframing manner, between estimated shadow edges.

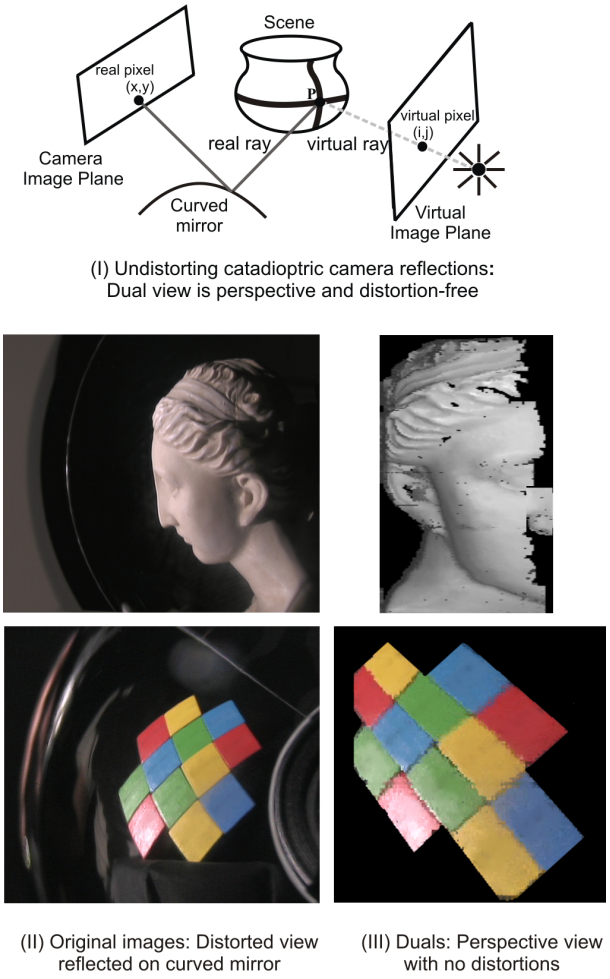
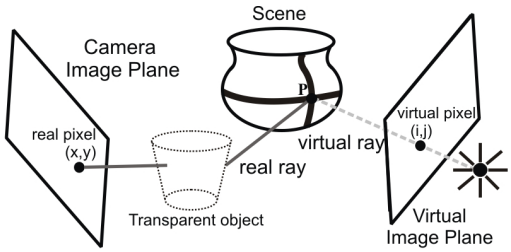


Fig. 14. **Unwarping catadioptric distortions:** In (I) we show how the dual view from a perspective light source avoids catadioptric distortions. In (III) we show unwarped results: note the colored squares' lines are straight. The non-shadow holes are due to stereo occlusion.

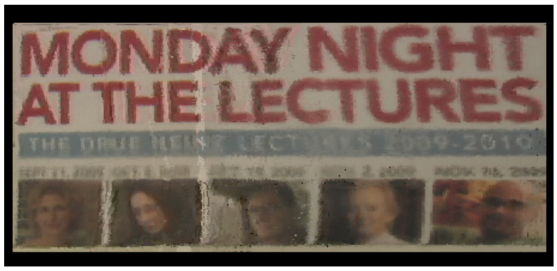
In Fig. 18 we show the shadow pushbroom image of the cliff. For better comparison, we have masked the sky with the same color as the original, first image of the mountain. We have outlined two areas with a red square that demonstrate the difference in the two images. These differences are due to the fact that the virtual viewpoint is high above the imaging camera, and therefore these give the impression of foreshortening from above. For example, the snowfields in the lower left of the images show a clear distortion, giving the effect (in the pushbroom image) of looking down on these. This is an expected effect, since the shadow-creating mountain edge is probably higher than the location of the camera. Another example is the sheer cliff on the extreme right, which shows foreshortening effects. Finally, we note that the image is not without artifacts - horizontal line effects can be seen where our keyframing does not work well.

Ridge pushbroom: Unlike our laboratory experiments, the shadow edge in our experiments here we rely on a naturally occurring shadow given by the ridge of a mountain. This ridge is not straight or flat, and therefore a non-uniform shadowing happens across the scene. This will result in distortions in the image.



(I) Undistorting transparent object refractions: Dual view is perspective and distortion-free

(II) Original image showing distortions



(III) Dual view with readable text

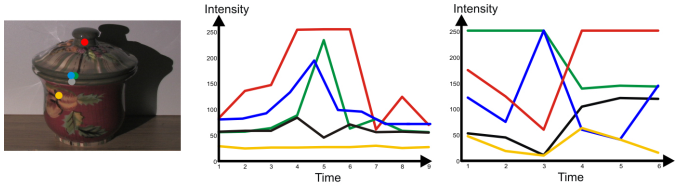
Fig. 15. **Unwarping refractive distortions:** In (I) we show how the dual view of an object looking through refractive elements can be warp-free. This occurs when the light-source illuminates the object without being occluded by the refractive elements. In (II) we show a planar poster through two thick glass objects. (III) shows the dual view, with undistorted, readable text.

Dual view at t = 1 Dual view at t = 5 Dual view at t = 9

(a) Dual views for linearly translating light source

Dual view at t = 1 Dual view at t = 3 Dual view at t = 6

(b) Dual view for an approaching light source



Camera view at t = 1 EPI profiles in (a) EPI profiles in (b)

(c) EPI profiles for five scene points from (a) and (b)

Fig. 16. **Dual EPIs:** We look at epipolar plane images for two cases. In all these cases, we are looking at the dual shadow camera images created by our method. In (a) we show the dual images of a light source moving linearly. In (b) we show the images of a light source approaching the scene. In both these sets of images, the geometry of the renderings appear correct. However, plotting the intensity profiles gives unexpected results. The profiles contain change in surface normal illumination and near-lighting effects that would not be present if a passive camera sensor was moved and generated these images.

VIII. Limitations

We propose an imaging setup and an algorithm to render different views of a scene. We can exploit the relative light-source shadow geometry to render new and interesting views. However this work suffers from certain limitations that we outline here:

Static scenes: Our method requires illumination variation to occur while the scene is static relative to the change in shadow position. To apply our method to fast moving scenes, both a high-speed sensor would need to be used and a spatial light modulator such as an LCoS or LCD would be need to be used. This would affect the ease of use of the current system.

Signal-to-noise ratio for low albedo scenes, area light sources and low-light scenes: The SNR of the shadow minima is crucial for shadow detection. There are a variety of scenarios where this may be difficult. For example, scenes with low albedo and low-light scenes such as at night. Other examples include scenes that are brightly illuminated by a secondary source, such as the sun, or when the physical extent of the source is such that only small portion of it is occluded by the mask. This happened in the fluorescent light masking results in this document, and would happen for any area light source. Physically, the affect of the illumination mask is real, but the question is if enough photons (or lack of photons) are detected at the camera sensor. High-dynamic range imaging is one possible method to address this limitation.

Shadow detections: In scenes with cast shadows, attached shadows and other forms of illumination occlusion, an important problem is how shadows of the mask can be robustly detected when compared to other scene shadows. Currently, the near-uniform motion of the shadow provides some signal characteristics (such as sharp peaks) that are used to separate these from other shadows. However, this method can fail in a variety of scenarios. Proper shadow detection and classification remains an open problem.

Physically realistic dual views: In Fig. 16 we demonstrate the limits of dual imaging. We use a non-lambertian object where the light-source moves linearly. In one case (a), the light-source translates parallel to the imaging plane. In the other, (b), the light-source approaches the object. The shadow camera reciprocal views appear visually correct, since the geometry of our setup is correct. However, if we look at the intensity profiles of three points, we get, essentially, random collections of measurements. Consider how these profiles would look like for a translating camera and a static source - these would vary smoothly. The lack of such results show that

the appearances in shadow camera views change dramatically for some BRDFs. Understanding which BRDFs and which scenes are appropriate for reciprocal cameras would be a fascinating direction for future work.

Practicality: Our method may not have application in outdoor scenarios with moving objects. However, there are some real-world applications that we have shown. For example, capturing the light-field of a light-source does have application in lighting engineering and computer graphics and our approach has clear advantages over competing methods. Our goal is to generalize dual photography to light-sources that are not as easily controlled as conventional projectors, and this has been achieved.

IX. Discussion and Future work

In this paper we demonstrated a reciprocal approach by creating novel views from the relative motion between non-programmable sources and occluding masks. These views can be created for scenes with complex reflectance, without reconstruction and with no light-source assumptions.

Our method involves smooth motion between a light-source and an opaque line mask such as a thin wire. The mask shadow is easily detected as an intensity minima at each pixel. A minima location, say at some time instance, defines a shadow hull determined by the mask location at that time instance. For example, the shadow hull due to a thin wire mask is a shadow plane through the mask. Multiple minima detected at a scene point correspond to intersections of these shadow hulls. For example, minima created by two, different positions of the line mask define a line at the intersection of the two shadow planes. If the intersection of shadow hulls is a line (as in this case), then the line is a light-ray from the source to the scene point. Detecting two minima at each pixel, uniquely identifies all incident rays and defines the image coordinates of a virtual, dual scene view.

The input are images of a static scene taken either under smoothly moving illumination and fixed masks or smoothly moving masks and fixed illumination. The end result are mask shadows that fall onto the scene and are detected as intensity minima. The output is a dual view of the scene. The geometry of this dual view depends on whether the light-source or the mask is static. If the light-source is static, then this dual view is the point-of-view of the light-source.

Our first application unwarps images that have distortions due to refraction (such as imaging through thick glass or lens distortions) or due to reflection (such as catadioptric systems or other camera-mirror setups). This is done with no knowledge of the shape of the optical elements involved, and no calibration object in the scene. The second application uses two calibration planes to capture the light-field of a light-source. This is essentially a method to capture a digital model of any complex light-source for use in rendering.

Our approach shares aspects with Dual Photography and extends it beyond programmable light-sources (such as projectors) to all types of illumination. However, we are unable to recover the whole light-transport matrix and therefore are restricted to obtaining dual views.

Detecting the minima at pixels does not require storing the whole input data, but, instead, a sliding window can be used. Shadow detection determines the resolution and quality of the dual views, and good shadow detection is possible only if the mask moves slowly relative to the camera frame-rate. We now discuss future work in this area:

(a) Relighting scenes with camera motion: A translating camera can be used to generate epipolar plane images ([3]) with dense correspondences between scene points in each image. We can relight any image in the sequence by replacing its pixels with corresponding pixels in other frames. From Helmholtz reciprocity (as in [27]) such images are identical to those created by moving the light source.

(b) Direct-global components: Our method may perform badly when scenes have global effects such as interreflections. Fortunately, fast separation of global and direct components ([20]) allows using the intensity minima due to a thin shadow as the global component. We can apply our method on the extracted direct component.

(c) Frequency analysis: In [7] the occlusion of illumination masks represents a convolution with incoming light rays, making it possible to design shadow cameras in the frequency domain, which we will pursue in future work.

References

[1] A. Abrams, K. Miskell, and R. Pless. The episolar constraint: Monocular shape from shadow correspondence. *Computer Vision and Pattern Recognition*, 2013.
[2] I. Ashdown. Modelling complex 3D light sources. *SIGGRAPH*, 1993.
[3] R. Bolles, H. Baker, and D. Marimont. Epipolar-plane image analysis: An approach to determining structure from motion. *International Journal on Computer Vision*, 1987.
[4] J. Bouguet and P. Perona. 3d photography on your desk. *International Conference on Computer Vision*, 1998.
[5] Y. Caspi and M. Werman. Vertical parallax from moving shadows. *Computer Vision and Pattern Recognition*, 2006.
[6] J. Davis, D. Nehab, R. Ramamoorthi, and S. Rusinkiewicz. Spacetime stereo : A unifying framework for depth from triangulation. *Computer Vision and Pattern Recognition*, 2003.
[7] F. Durand, N. Holzschuch, C. Soler, E. Chan, and F. Sillion. A frequency analysis of light transport. *Transactions on Graphics*, 2005.
[8] G. Garg, E.-V. Talvala, M. Levoy, and H. P. A. Lensch. Symmetric photography: Exploiting data-sparseness in reflectance fields. *Eurographics Symposium on Rendering*, 2006.
[9] M. Goesele, X. Granier, W. Heidrich, and H. Seidel. Accurate light source acquisition and rendering. *Transactions on Graphics*, 2003.
[10] S. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen. The lumigraph. *SIGGRAPH*, 1996.
[11] M. D. Grossberg and S. K. Nayar. A general imaging model and a method for finding its parameters. *International Conference on Computer Vision*, 2001.
[12] H. V. Helmholtz. Treatise on physiological optics. *Dover*, 1925.
[13] N. Jacobs, B. Bies, and R. Pless. Using cloud shadows to infer scene structure and camera calibration. *Computer Vision and Pattern Recognition*, 2010.
[14] N. Jacobs, M. Islam, and S. Workman. Cloud motion as a calibration cue. *Computer Vision and Pattern Recognition*, 2013.
[15] N. Jacobs, J. King, D. Bowers, and R. Souvenir. Estimating cloud maps from outdoor image sequences. *Winter Conference on Applications of Computer Vision*, 2014.
[16] M. Langer and S. Zucker. What is a light source? *Computer Vision and Pattern Recognition*, 1997.
[17] M. S. Langer, G. Dudek, and S. W. Zucker. Space occupancy using multiple shadowimages. *International Conference on Intelligent Robots and Systems*, 2005.
[18] D. Lanman, R. Raskar, A. Agrawal, and G. Taubin. Shield fields: modeling and capturing 3d occluders. *Transactions on Graphics*, 2008.
[19] M. Levoy and P. Hanrahan. Light field rendering. *SIGGRAPH*, 1996.
[20] S. Nayar, G. Krishnan, M. D. Grossberg, and R. Raskar. Fast separation of direct and global components of a scene using high frequency illumination. *Transactions on Graphics*, 2006.

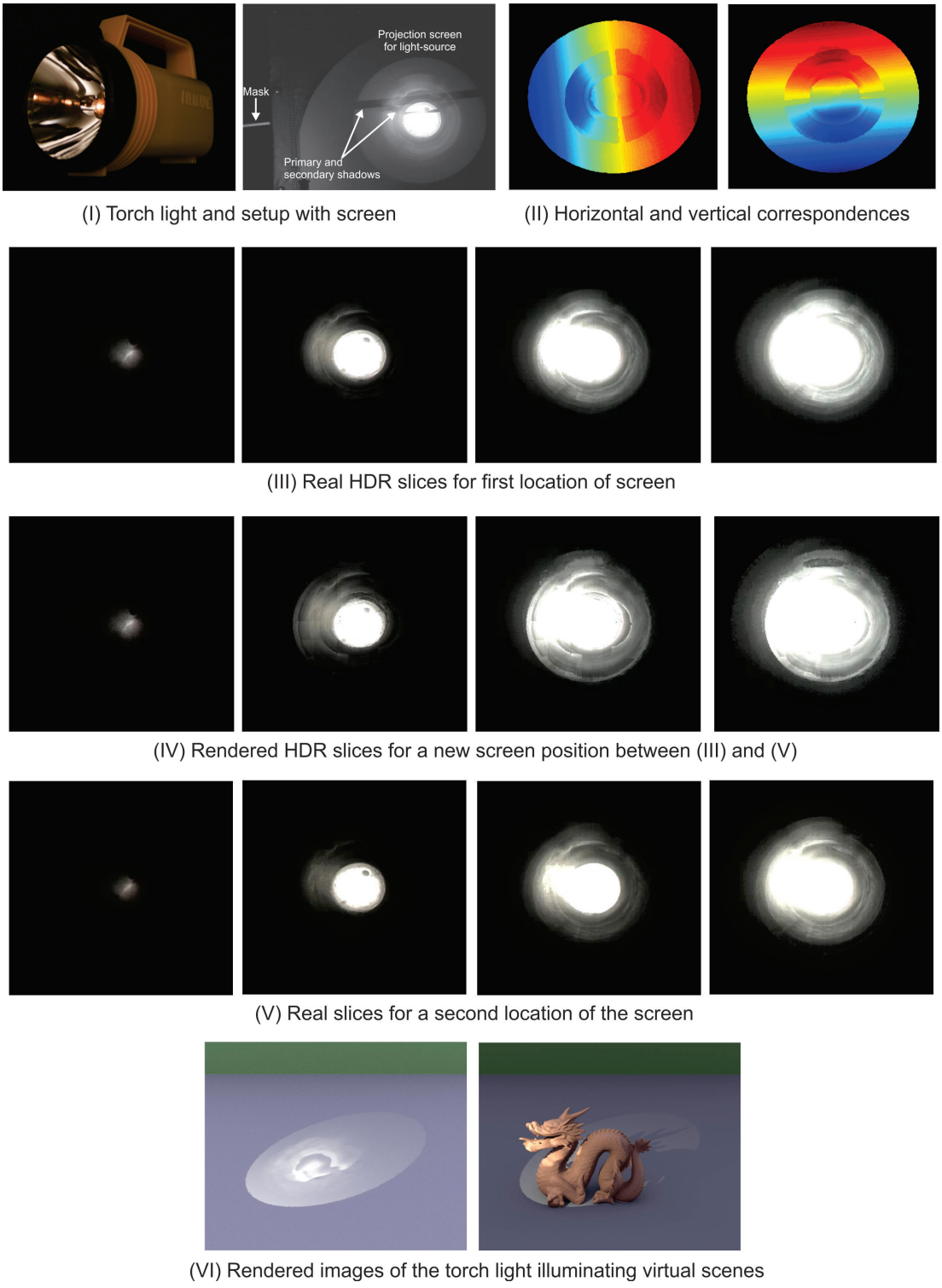


Fig. 17. **Capturing the light field of a light source:** In (I) a flash-light illuminates a screen in our setup. Note the complex radial patterns due to the reflective mirrors in the head of the flash-light. We move two linear masks, as shown in (I) horizontally and vertically, occluding the light-source. This casts a shadow on the screen. The horizontal and vertical minima locations in time for each pixel are shown as images in (II). In (III) and (V) we show HDR slices of the image of the torch, for two separate positions of the screen. Given the shadow correspondences, we capture the light-field of the light-source, allowing us to render how the torchs caustic would look at a virtual screen position between the two real positions. Note how the high-frequency information in the caustic is faithfully rendered. In (VI), we show illuminate a rendered scene using the captured light-field of the light-source.

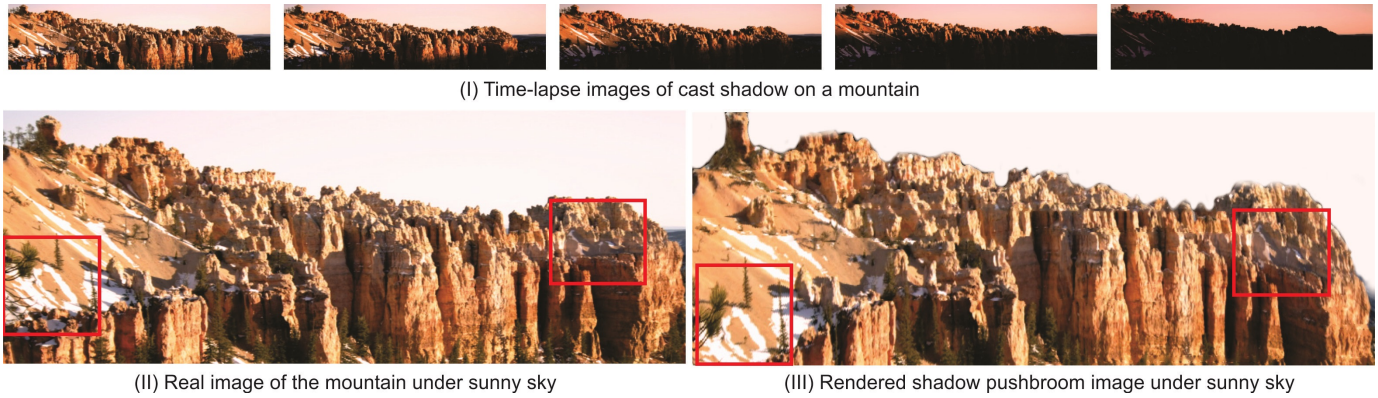


Fig. 18. **Shadow pushbroom camera image for an outdoor scene:** At the top of the image we show time-lapse photographs of a cliff occluded by the shadow of an (unseen) neighboring mountain. The edge of the mountain can be detected, and we can create a new second, shadow-pushbroom view of the scene. This view is from the top of the (unseen) neighboring mountain: for example, the snow-fields on the left appear to be viewed from above, as compared to the original image.

[21] R. Ng. Fourier slice photography. *Transactions on Graphics*, 2005.
[22] M. Pharr and G. Humphreys. Physically based rendering: From theory to implementation. *Elsevier*, 2004.
[23] R. Raskar, A. Agrawal, and J. Tumblin. Coded exposure photography: Motion deblurring using fluttered shutter. *TOG*, 2006.
[24] D. Raviv, Y. Pao, and K. Loparo. Reconstruction of three-dimensional surfaces from two-dimensional binary images. *Transactions on Robotics and Automation*, 1989.
[25] S. Savarese, M. Andreetto, H. Rushmeier, and F. Bernardini. 3D reconstruction by shadow carving: Theory and practical evaluation. *International Journal of Computer Vision*, 2005.
[26] S. Seitz. The space of all stereo images. *International Conference on Computer Vision*, 2001.
[27] P. Sen, B. Chen, G. Garg, S. R. Marschner, M. Horowitz, M. Levoy, and H. P. A. Lensch. Dual photography. *Transactions on Graphics*, 2005.
[28] H. Shum, S. Chan, and S. B. Kang. Image-based rendering. *Springer*, 2007.
[29] R. Swaminathan, M. Grossberg, and S. K. Nayar. Caustics of catadioptric cameras. *International Conference on Computer Vision*, 2001.
[30] A. Torralba and W. Freeman. Accidental pinhole and pinspeck cameras. *International Journal of Computer Vision*, in print.
[31] J. Unger, A. Wenger, T. Hawkins, A. Gardner, and P. Debevec. Capturing and rendering with incident light fields. *Eurographics Symposium on Rendering*, 2003.
[32] V. Vaish, R. Szeliski, C. L. Zitnick, S. B. Kang, and M. Levoy. Reconstructing occluded surfaces using synthetic apertures: Stereo, focus and robust measures. *Computer Vision and Pattern Recognition*, 2006.
[33] D. Vaquero, R. Feris, M. Turk, and R. Raskar. Characterizing the shadow space of camera-light pairs. *Computer Vision and Pattern Recognition*, 2008.
[34] S. Yamazaki, S. G. Narasimhan, S. Baker, and T. Kanade. Coplanar shadowgrams for acquiring visual hulls of intricate objects. *International Conference on Computer Vision*, 2007.
[35] J. Yu and L. McMillan. General linear cameras. *European Conference on Computer Vision*, 2004.
[36] C. Zhang and T. Chen. A self-reconfigurable camera array. *Eurographics Symposium on Rendering*, 2004.
[37] T. Zickler, P. Belhumeur, and D. Kriegman. Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. *European Conference on Computer Vision*, 2002.
[38] C. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High-quality video view interpolation using a layered representation. *Transactions on Graphics*, 2004.
[39] A. Zomet, D. Feldman, and S. Peleg. Mosaicing new views: The crossed-slits projection. *Pattern Analysis and Machine Intelligence*, 2003.