Characterizing the Shadow Space of Camera-Light Pairs

Daniel A. Vaquero
UC Santa Barbara
daniel@cs.ucsb.edu

Rogerio S. Feris
IBM Research
rsferis@us.ibm.com

Matthew Turk
UC Santa Barbara
mturk@cs.ucsb.edu

Ramesh Raskar
MIT Media Lab
raskar@media.mit.edu

Abstract

We present a theoretical analysis for characterizing the shadows cast by a point light source given its relative position to the camera. In particular, we analyze the epipolar geometry of camera-light pairs, including unusual camera-light configurations such as light sources aligned with the camera's optical axis as well as convenient arrangements such as lights placed in the camera plane. A mathematical characterization of the shadows is derived to determine the orientations and locations of depth discontinuities when projected onto the image plane that could potentially be associated with cast shadows. The resulting theory is applied to compute a lower bound on the number of lights needed to extract all depth discontinuities from a general scene using a multiflash camera. We also provide a characterization of which discontinuities are missed and which are correctly detected by the algorithm, and a foundation for choosing an optimal light placement. Experiments with depth edges computed using two-flash setups and a four-flash setup illustrate the theory, and an additional configuration with a flash at the camera's center of projection is exploited as a solution for some degenerate cases.

1. Introduction

The analysis and interpretation of shadows is an important and challenging problem in computer vision. They often appear in real-world scenes, leading to failures in many vision algorithms such as segmentation, tracking, and recognition. On the other hand, shadows carry valuable 3D cues about surfaces in the scene and can be used as a positive source of information for many applications.

Shape from shadow (or darkness) techniques use a moving light source to build a continuous representation (also known as a shadowgram) from which depth estimates can be computed [3, 8, 11]. Good reviews of shadow-based shape analysis methods are available in [7, 16]. These methods offer advantages in the sense that they rely neither on correspondences nor on a model of the surface reflectance characteristics, and they may be implemented using inexpensive lighting and/or imaging equipment [12]. There is a growing interest in processing photos taken from a fixed view but under varying illumination [1, 2, 4, 15]. The analysis of shadows is also critical for microstructure analysis [9, 17]. Other uses include the estimation of heights of buildings from aerial images [6], interactive applications [13] and non-photorealistic rendering [10].

A common limitation of many existing shadow-based methods is that the light sources need to surround the object, in order to create significant shading and shadow variation from estimated or known 3D light positions. This requires a fixed lighting rig, which limits the application of these techniques to industrial settings and makes them unsuitable to be built into a self-contained camera.

Recently, this problem was addressed with a multiflash imaging method that exploits shadows created with small camera-light baselines [10]. This technique combines shadow information from a collection of images taken using flashes at different locations in order to detect depth edges (C0 discontinuities in the depth map, also known as depth discontinuities or occluding contours) in a scene. Depth edges and shadows were also exploited in [5] to improve stereo algorithms with small baseline illumination.

In this paper, we aim to characterize the space of depth edge orientations and locations, as projected onto the image plane, that can potentially be associated with cast shadows in a camera-light pair. Towards this goal, we analyze the epipolar geometry of camera-light pairs with general light placements, which comprise useful arrangements such as flashes in the camera plane (resulting in parallel light rays in the image plane) as well as unusual camera-light configurations such as lights aligned with the optical axis and in front of and behind the camera (resulting in light rays converging or diverging from a point in the image plane). We derive a mathematical representation of the space of shadows that can be potentially generated over all depth edge orientations and locations, providing a characterization of the shadows produced by a camera-light pair.

The usefulness of our theoretical analysis is then demonstrated on the problem of depth edge detection with multiflash imaging. A lower bound on the number of flash

lights that will cast shadows along the full range of depth discontinuities in a scene is derived. In addition, the theory characterizes the depth edges that are correctly detected or missed. Optimal light placement positions are presented and some failure cases inherent to the shadow detection algorithm are discussed. We then propose an improvement to multiflash imaging consisting of the capture of an additional image using a flash at the camera's center of projection, that addresses these failure cases by physically capturing a shadow-free image.

Summarizing, the main contributions of our work are:

- A theory that characterizes the shadows cast by a point light source in a camera-flash pair, with respect to the shadows generated along depth edges in all orientations and locations in the image plane;
- A lower bound on the number of lights necessary to capture all depth edges in a general scene using a multiflash imaging scheme;
- A theoretical foundation for the problem of placing light sources in a multiflash imaging setup for depth edge detection;
- An alternative way to obtain a shadow-free image in multiflash imaging, as a solution to some failure cases.

This paper is organized as follows: in Section 2, we characterize the shadow space of camera-light pairs. The proposed framework is then employed in Section 3 to infer a lower bound on the number of lights needed to find the full range of depth edges in a scene, using a multiflash camera; optimal light placements and degenerate cases are also presented. Section 4 illustrates the theory by showing experiments with two- and four-flash setups and addressing some failure cases using an alternative way to compute the shadow-free image.

2. The Shadow Space of a Camera-Light Pair

In this section, we introduce the mathematical concepts required for our analysis and derive a characterization of the shadows cast by a point light source in a camera-light pair.

2.1. Edge-Shadows

Consider a horizontal edge with a shadow cast above it, as in Figure 1(a). We will call such structures, composed of an edge and a shadow cast on one of its sides, as **edge-shadows**. For the purpose of this analysis, an edge-shadow can represent both straight and curved edges at an edge point (by considering the tangent line in the second case).

Let $e(\theta)$ be the result of rotating the edge-shadow in Figure 1(a) counterclockwise by θ radians, $0 \leq \theta < 2\pi$. Figure 1(b) depicts some examples of rotations by different angles. Now define $\mathcal{E}$ as the set of all edge-shadows resulting from rotating the horizontal edge-shadow above by angles in the $[0, 2\pi)$ interval, that is, $\mathcal{E} = \{e(\theta) \mid 0 \leq \theta < 2\pi\}$. $\mathcal{E}$

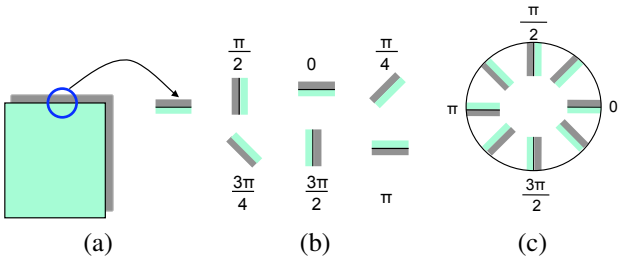


Figure 1. Edge-shadows: a) edge-shadow; b) edge-shadow rotations; c) space of edge-shadows.

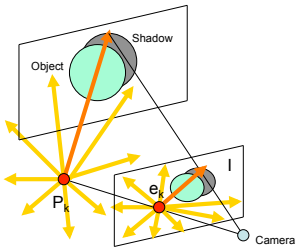


Figure 2. Imaging geometry.

is indexed by θ and is called the **space of edge-shadows**. It is illustrated in Figure 1(c). We may also refer to a single $e(\theta)$ as $e(\theta + 2k\pi)$, with $k \in \mathbb{Z}$.

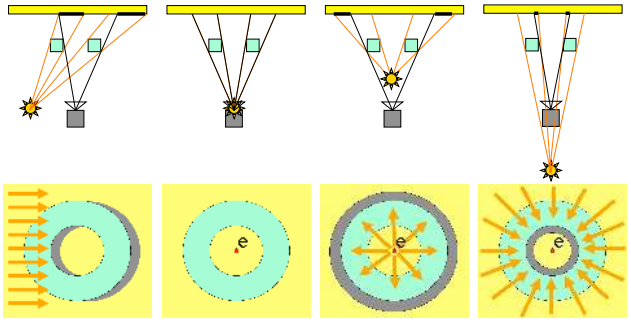
A simple observation is that $\mathcal{E}$ represents all possible edge orientations that might pass through a specific point in the image. Each orientation appears twice in the space $\mathcal{E}$, associated to angles α and $\alpha + \pi$, for $0 \leq \alpha < \pi$. The shadow is cast along one of the edge's sides for the first case and along the opposite side for the second case.

2.2. Epipolar Geometry for Camera-Light Pairs

In a camera-light setup, the camera captures an image of a scene illuminated by a single light source. Assuming a pinhole camera model and a small baseline point light source, Figure 2 illustrates the imaging geometry. The projection of the point light source P_k is at point e_k (called the *light epipole*) in the image plane. The images of the infinite set of light rays originating at P_k are called the *light epipolar rays*, which originate at e_k.

Let γ be the plane parallel to the image plane that contains the camera's center of projection. The geometry of the light epipolar rays can be classified into four classes, based on the relative position of the light in relation to γ:

Parallel rays: When the light is placed in γ, the light epipole is at infinity and the light epipolar rays are parallel, with all rays pointing at the same direction. The direction depends on the relative camera-light position. For example, if the light is above the camera, the rays go from the top to the bottom of the image; if the light is to the right, the rays go from right to left. Figure 3(a) shows an example with a torus-like object centered in the image. The epipolar rays go from left to right, as the light is placed to the left of the



(a) (b) (c) (d)

Figure 3. Epipolar geometry. Top row: top-down views of the image capture geometry viewing a torus-like object illuminated by a single light source. Bottom row: the resulting captured images, with light epipolar rays superimposed. a) Light to the left, in the camera plane; b) light in the camera's center of projection; c) light in front of the camera, aligned with the optical axis; d) light behind the camera, aligned with the optical axis.

camera, and the shadows are cast to the right of the object.

Co-located camera and light: When the light source and the camera's center of projection are at the same location, which can be physically generated using beamsplitters (see Section 4), the light epipole is a single point in the center of the image. This configuration does not generate any light epipolar rays, which means that no shadows are cast. Figure 3(b) illustrates this case.

Radial divergent rays: If the light is placed in front of γ (considering as "front" the half of the 3-D space that lies in front of the camera), the light epipole is a point e in the image plane and the light epipolar rays point away from e in the radial direction. The epipole e is called the *focus of expansion*. Figure 3(c) shows the divergent rays for a light source in front of the camera, aligned with the optical axis. This leads to shadows being cast over the external border of the object. As the light is aligned with the camera's optical axis, the epipole is at the center of the image.

Radial convergent rays: If the light is placed behind γ, the light epipole is a point e in the image plane and the light epipolar rays point towards e in the radial direction. The epipole e is called the *focus of contraction*. Figure 3(d) depicts the convergent rays for a light source behind the camera, aligned with the optical axis. The shadows are cast only over the internal border of the torus.

When the light is neither in the image plane nor along the optical axis, the projected rays will diverge or converge at the location of the light epipole in the image.

2.3. Camera-Light Shadow Characterization

One question that arises when building a camera-light setup is: if we place the light at a certain position, what kind of shadows may we have along depth edges in the captured image, without any prior knowledge about the location of these edges in the scene? Or, equivalently, for which depth edge orientations and locations in the image will a light source cast a thin sliver of shadow along the edge?

In order to answer this question, we aim to characterize the space of shadows that can be potentially cast over depth discontinuities by a single light source in a camera-light setup.

Assumptions: In order to isolate the problem of characterizing cast shadows from other related problems, we analyze it under a few assumptions that enable us to concentrate on the problem of interest while avoiding unrelated issues. In Figure 2, we can see that if we move the light back and forth along the line that connects P_k to the camera's center of projection, the light epipole e_k does not change. However, the size of the cast shadows change, and the shadows may become detached from small or narrow objects. We work under the assumption that the baseline (distance from the camera to the light along the aforementioned line) choice is ideal; that is, the camera-light distance is large enough to make a detectable shadow appear in the image, but is also small enough to prevent detached shadows. An analysis of baseline issues can be found in [10]. We also assume that the light source is a point light source and the light distribution over the scene is uniform.

The Space of Possible Shadows: Consider a light epipolar ray r oriented in a direction α. Then, for a given point (i, j) in the image plane that is crossed by r, the set of possible edge-shadows generated at (i, j) by the light source is given by $\mathcal{L} = \{e(\theta) \mid \alpha < \theta < \alpha + \pi\}$, due to the simple observation that shadows are cast on the opposite side of the edge with respect to the origin of the light epipolar ray. Notice that the inequalities are strict, which means that edges parallel to r are not included in $\mathcal{L}$.

Let us now define a space $\mathcal{S} = (i, j, \theta)$, where each triple corresponds to a point (i, j) in the image plane I and an orientation $\theta \in [0, 2\pi)$ which indexes an edge-shadow $e(\theta) \in \mathcal{E}$. The potential shadows cast by the light source can be represented by a function $\mathcal{C} : \mathcal{S} \rightarrow \mathbb{N}$ that is equal to one when the edge-shadow $e(\theta)$ can be generated at the position (i, j), and equal to zero otherwise. For a given point (i, j), the range of θ values which have $\mathcal{C}(i, j, \theta) = 1$ depends on the orientation α of the light epipolar ray which passes through (i, j), and corresponds to the set $\mathcal{L}$.

The function $\mathcal{C}$ is a representation of the shadows that can be cast by the light source over edges at all orientations for each point in the image plane. In other words, for each point $(i, j) \in I$, the function $\mathcal{C}$ has value 1 for edge orientations that will have a shadow attached when they appear at that point in the image, and has value 0 for orientations that will not have a shadow attached. In order to compute this representation, only the relative camera-light position is needed.

3. Multiflash Depth Edge Detection Analysis

When one builds a multiflash setup for depth edge detection, the following questions are relevant: how many flashes should be used, and where should they be placed such that no depth edges are missed by the detection algorithm?

In the previous section, we presented a characterization of the shadows cast by one point light source in a camera-light pair. Now we will use that representation to derive a lower bound on the number of light sources needed to compute all depth edges in a general scene using the multiflash scheme. The theoretical framework also enables us to provide a characterization of the missed and detected edges, as well as to present an optimal placement of lights.

3.1. Depth Edges with Multiflash Imaging

The multiflash technique [10] provides a way of computing depth edges in a scene without computing the depth map. It is based on the principle that when an image is taken from a scene illuminated by a light source close to the camera, thin slivers of shadow are cast along depth edges. The shadow position depends on the relative camera-flash position: for example, shadows will be to the left of objects when the flash is to the right of the camera. The technique combines the shadow information from multiple pictures taken using light sources (flashes) at different positions to compute the depth edges.

From Figure 2, two observations can be made regarding the geometry of the cast shadows: a shadow of a depth edge pixel is constrained to lie along the light epipolar ray passing through that pixel; and when a shadow is cast at a depth discontinuity, the shadow and the origin of the light epipolar ray will be on opposite sides of the depth edge. For each captured image, a ratio image is generated by dividing it by a shadow-free image (which is approximated by the maximum composite of all acquired images). The depth edge detection algorithm searches ratio images for negative jumps in intensity along the light epipolar rays and marks these points as depth edges. See [10] for more details.

3.2. The Coverage Problem

It is known that for the multiflash detection algorithm to work, the negative side of each depth edge (which corresponds to the background, as opposed to the foreground that is the object itself) must be shadowed by one of the light sources and not be shadowed by at least one other light source. We are now interested in the question: given a collection of point light sources $P_1, \ldots, P_N$, will the above condition hold for the full range of depth edges in a scene?

As we saw in Section 2.3, each light source P_k, $k \in \{1, \ldots, N\}$ induces a function $\mathcal{C}_k : \mathcal{S} \to \mathbb{N}$ (in that case, the function values are in $\{0, 1\}$, a subset of $\mathbb{N}$) which represents the depth edges shadowed by P_k. If we compute $\mathcal{C}^+ = \sum_{k=1}^{N} \mathcal{C}_k$, the result is a function that represents how many light sources can generate the edge-shadow $e(\theta)$ if a depth edge with orientation θ passes through the point (i, j) in the image.

Hence, one can determine whether an edge can be detected or not based on $\mathcal{C}^+$: if $0 < \mathcal{C}^+(i, j, \theta) < N$, then a depth edge with negative side corresponding to the edge-shadow $e(\theta)$ which passes through (i, j) will be detected; if $\mathcal{C}^+(i, j, \theta) = 0$ or $\mathcal{C}^+(i, j, \theta) = N$, that depth edge will be missed, because either no shadows are cast or a shadow is cast by all light sources. A *coverage map* $\mathcal{D} : \mathcal{S} \to \{0, 1\}$ can be computed from $\mathcal{C}^+$ by setting $\mathcal{D}(i, j, \theta)$ to 1 if $0 < \mathcal{C}^+(i, j, \theta) < N$, and setting it to 0 if $\mathcal{C}^+(i, j, \theta) = 0$ or $\mathcal{C}^+(i, j, \theta) = N$. The problem of deciding whether all edges are detected for a given set of light sources now reduces to the problem of verifying that the function $\mathcal{D}$ is equal to 1 for all triples $(i, j, \theta) \in \mathcal{S}$. The function $\mathcal{D}$ also provides a characterization of the detected and missed edges: the edges corresponding to triples (i, j, θ) with $\mathcal{D}(i, j, \theta) = 0$ are missed, while the edges such that $\mathcal{D}(i, j, \theta) = 1$ are detected.

3.3. Two Flashes

One might ask whether two flashes would suffice to capture all depth edges in general scenes. A setup like this is inexpensive and would have a short acquisition time when compared to the 4-flash setup presented in [10]. Intuitively, placing the flashes in configurations such as left-right (lights in the camera plane, one to the left and the other to the right) or above-below will lead to missed horizontal or vertical edges, respectively. Another possible solution would be to place the flashes in the camera's optical axis, one in front of and the other behind the camera. This setup would work well for the torus-like object from Figure 3, as the captured images would be the ones in Figures 3(c-d). However, in this section we will analyze general two-flash setups, and it follows that a front-behind setup also fails to capture other depth edges.

Consider a setup with two flashes and a camera. The following results play important roles in our analysis:

- It is optimal to place the lights in such a way that the light epipolar rays arriving at each point (i, j) in the image plane come from opposite directions;
- It is impossible to capture all depth edges in a general scene using only two light sources, no matter how they are placed in relation to the camera.

The veracity of the first result can be shown in the following way: consider a point $(i, j) \in I$. The objective is to maximize the range of θ values such that $\mathcal{D}(i, j, \theta) = 1$, or equivalently, $\mathcal{C}^+(i, j, \theta) = 1$. But consider the functions $\mathcal{C}_1(i, j, \theta)$ and $\mathcal{C}_2(i, j, \theta)$. If α_1 and α_2 are the orientations of the light epipolar rays generated by P_1 and P_2 that arrive at (i, j), $\mathcal{C}_k(i, j, \theta)$ is equal to 1 for $\theta \in \mathcal{L}_k = (\alpha_k, \alpha_k + \pi)$

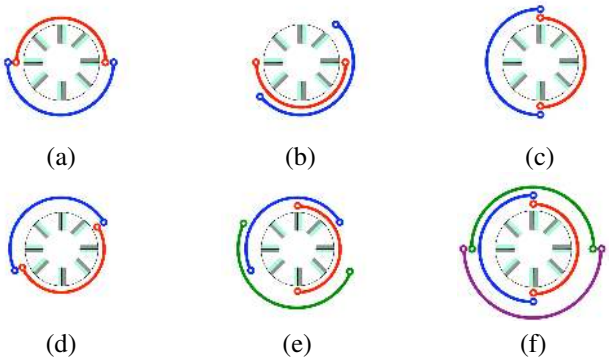


Figure 4. Coverage diagrams: a) two-flash optimal case; b) two-flash suboptimal case; c) two-flash above-below; d) two-flash diagonal; e) three flashes; f) four flashes.

and equal to 0 otherwise, for $k \in \{1, 2\}$. $\mathcal{C}^+(i, j, \theta)$ is equal to 1 either if $\mathcal{C}_1(i, j, \theta) = 1$ or $\mathcal{C}_2(i, j, \theta) = 1$, but not both. This observation shows that $\mathcal{L}_1$ and $\mathcal{L}_2$ must be disjoint in order to maximize the range of θ values such that $\mathcal{D}(i, j, \theta) = 1$. As both $\mathcal{L}_1$ and $\mathcal{L}_2$ are π-sized intervals (excluding the extremes), it is possible for both $\mathcal{L}_1$ and $\mathcal{L}_2$ to be disjoint and contained in $[0, 2\pi)$. One case when this happens is when $\alpha_1 = \alpha_2 + \pi \pmod{2\pi}$. Therefore, if the light epipolar rays come from opposite directions, the coverage is optimal. Figure 4(a) depicts the coverage of $\mathcal{E}$ in a graphical way for a left-right setup (which is one of the optimal placements). The red and blue arcs represent the covered intervals of edge-shadows for each light source, and the uncovered points are depicted as open circles.

The second observation is a corollary from the optimality result: since in the optimal case $\mathcal{L}_1$ and $\mathcal{L}_2$ are disjoint and each has a length of π (not including the extremes), $\mathcal{L}_1 \cup \mathcal{L}_2$ has length 2π but also does not include the extremes. Hence, there are values of θ such that $\mathcal{D}(i, j, \theta) = 0$.

When the light epipolar rays arriving at a point (i, j) in the image plane do not come from opposite directions, as in Figure 4(b), we have a suboptimal situation: some edge-shadows cannot be generated by any of the lights and there are also edge-shadows generated by both lights, thus $\mathcal{D}(i, j, \theta) = 0$ in these cases.

Having made these observations, we now proceed to the characterization of some useful and unusual two-flash configurations.

Flashes in the camera plane: Placing the flashes in the camera plane on opposite sides of the camera is a very common practice in multiflash imaging. Such setups include: flashes to the left and right of the camera, which create horizontal light epipolar rays; flashes above and below the camera, which create vertical epipolar rays; and flashes in the diagonal, which create diagonal epipolar rays.

In all these cases, the epipolar rays are parallel, and the two rays that arrive at each point $(i, j) \in I$ come from opposite directions (and from the same directions for all points in the image). Thus, the optimal situation in terms of coverage is met, and only the edges which are parallel to the epipolar rays are missed. Figure 4(c-d) depicts the edge-shadows covered by the above-below and diagonal setups.

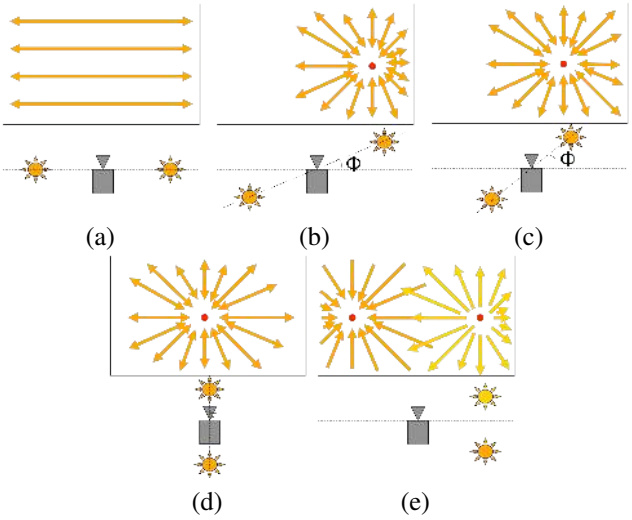


Figure 5. 2-flash light epipolar ray geometry. a) Left-right; b-c) intermediate configurations; d) front-behind; e) suboptimal configuration.

Placing the flashes in the camera plane but not on opposite sides of the camera (*e.g.*, one flash above and another one on the right hand side) is suboptimal, as previously shown; the light epipolar rays do not come from opposite directions.

Flashes outside of the camera plane: Consider the left-right setup with flashes placed in the camera plane. If we rotate the camera by ϕ radians around the "up" axis, with $0 \leq \phi \leq \frac{\pi}{2}$ (Figure 5), then we transition from a left-right setup to a setup with lights in front of and behind the camera and aligned with the optical axis as ϕ goes from 0 to $\frac{\pi}{2}$. All intermediate configurations consist of flashes outside of the camera plane.

Let us now analyze how the light epipolar rays change as ϕ progresses from 0 to $\frac{\pi}{2}$. The situation with $\phi = 0$ has light epipoles at infinity and parallel and horizontal light epipolar rays; as the camera is rotated, flashes are in front of and behind the camera, so the light epipoles are projected to points in the image plane along the line that joins the light source and the camera's center of projection. Light epipolar rays are convergent for the flash behind and divergent for the flash in front, and the light epipoles move from right to left as we rotate the camera from 0 to $\frac{\pi}{2}$ until both epipoles reach the center of the image (when $\phi = \frac{\pi}{2}$, which means that the lights are aligned with the optical axis). Figure 5(a-d) illustrates this transition.

All these cases are optimal with respect to the coverage of the edge space, since the light epipolar rays which arrive

at each image point come from opposite directions. The depth edges parallel to the light rays are missed, that is, the left-right setup misses horizontal edges while the other configurations miss edges in the radial direction starting from the light epipole position. Similar analyses can be done by setting the initial position to "flashes in the diagonal", and then rotating the camera in the same way from 0 to $\frac{\pi}{2}$ and from 0 to $-\frac{\pi}{2}$. The light epipoles will lie above or below the horizontal line in the center of the image.

If the flashes are not collinear with and on opposite sides of the camera's center of projection, a suboptimal situation exists, since for most points in the image plane the light rays do not come from opposite directions (Figure 5(e)).

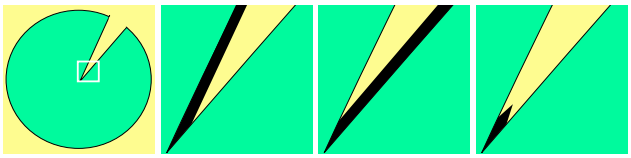
3.4. Three or More Flashes

Now consider three flashes in the camera plane positioned in such a way that they surround the camera and are vertices of an equilateral triangle. Figure 4(e) illustrates the edge-shadows covered by each light at a given point in the image (notice that the orientations of the arriving epipolar rays are the same for every point in the image). Each edge-shadow can be generated by at least one flash and no edge-shadows are simultaneously generated by all flashes. Therefore, the coverage map is such that $\mathcal{D}(i,j,\theta) = 1$ for all $(i,j,\theta) \in \mathcal{S}$. A similar analysis can be made for $N \geq 3$ flashes in the camera plane, by considering the lights to be vertices of a N-sided regular polygon. Figure 4(f) depicts the coverage map for $N = 4$.

The proposed coverage map is also useful to guide the light placement in the general case of three or more light sources outside of the camera plane, by providing an analytic framework to reason about the shadows potentially created by a given multiflash setup. For example, it is possible to show that, for the case of three light sources outside of the camera plane, if the three light sources are in front of the camera or the three light sources are behind the camera, and the three light epipoles are the vertices of a triangle, then the coverage map will be such that $\mathcal{D}(i,j,\theta) = 1$ for all (i,j,θ) with (i,j) inside the triangle. However, the full coverage is not attained outside of the triangle or on its boundaries, leading to missed edges. This has the practical implication that, in such setups, the image plane must be contained within the triangle formed by the light epipoles in order to warrant coverage of the space of edge-shadows for every point in the image plane.

3.5. Degenerate Cases

The condition $\mathcal{C}^+(i,j,\theta) = N$ for missing a depth edge, which means that all lights cast a shadow along the depth edge in question, is a consequence of the algorithm that calculates shadow regions by dividing each image by the maximum composite of all captured images. When all lights cast a shadow over a region in the image, the maximum image will contain a shadow at that region and the ratio images will have intensity values similar to unshadowed regions.



(a) (b) (c) (d)

Figure 6. Degenerate case: a) depth edge intersection; b) shadow from flashes above and to the left; c) shadow from flashes below and to the right; d) shadow intersection.

A degenerate case can happen when two edges intersect in such a way that there is an area that shares shadows cast by all lights. For example, consider a scene containing a disk without a 10-degree sector from say, 45 to 55 degrees counterclockwise. Suppose that the disk is centered in the image, as depicted in Figure 6(a). Both edges at the sector intersect at the center of the disk, and the four-flash setup will fail to accurately detect the edges near the intersection point, as there will be a region in shadow for all four flashes. Figure 6(b-d) illustrates this fact, with magnified views of the area of interest.

A solution for this failure case would be to add one additional light source which generates an epipolar ray that arrives in (i,j) with orientation between 45 and 55 degrees. However, complex scenes could require even more flashes. Another possible solution would be to capture the shadow-free image by placing a flash in the camera's center of projection, as was mentioned in Section 2.2. This requires only one additional flash.

Another issue is the problem of shadows of one object over another. If a given edge-shadow can only be generated by one of the light sources, but is located under the shadow cast by another depth edge, the algorithm will not detect it. This problem can be minimized by decreasing the camera-light baseline. Other limitations of the multiflash technique include specularities, lack of background and baseline issues (may cause undetectable or detached shadows), which were addressed in [10].

4. Experiments

In this section, we present a few experiments with a multiflash imaging setup to detect depth edges with the objective of illustrating the theory. Four configurations were built: a two-flash setup with flashes to the left and to the right of the camera, a two-flash setup with flashes above and below the camera, a two-flash setup with flashes in front of and behind the camera, aligned with the optical axis, and a four-flash setup combining the first two setups.

The first, second and fourth schemes place the flashes in the camera plane, so that the epipolar ray traversal is done in the horizontal direction for the first case and in the vertical

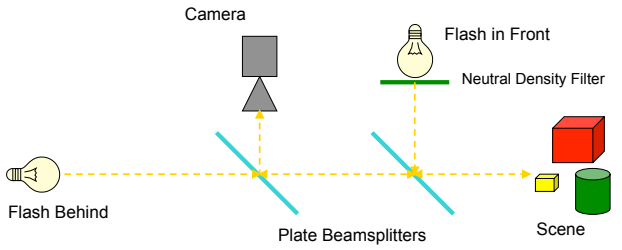


Figure 7. Front-behind setup. Notice the light travel paths before hitting the scene and on the way back to the camera.

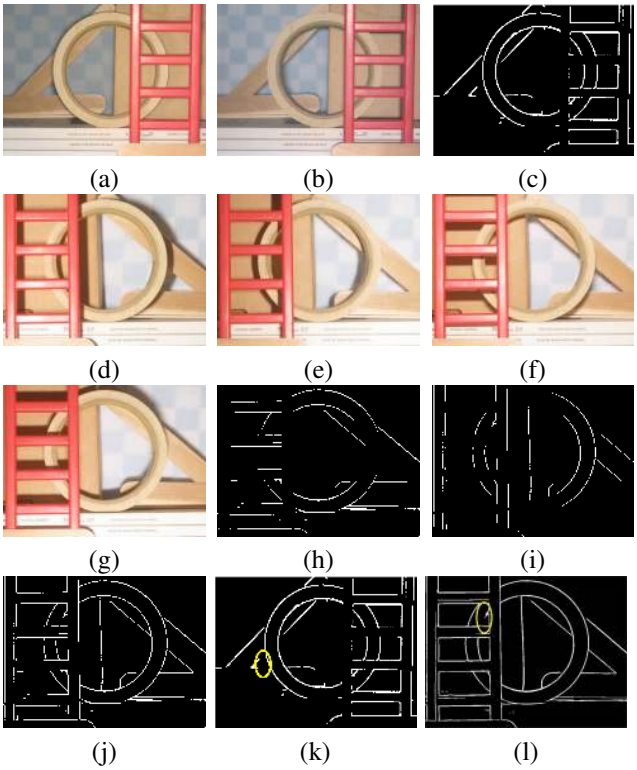
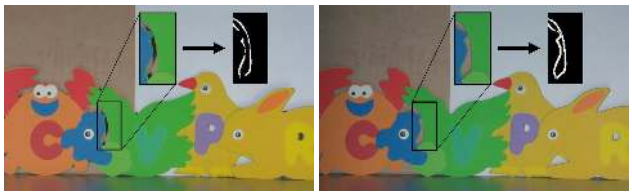


Figure 8. Depth edge detection. a) Flash in front; b) flash behind; c) depth edges for front-behind; d) flash to the left; e) flash to the right; f) flash above; g) flash below; h) depth edges for above-below; i) depth edges for left-right; j) depth edges for four-flash; k-l) degenerate cases.

direction for the second case. In the third case, we devised a scheme using two beamsplitters (half-silvered mirrors) to prevent one component from occluding the others. Figure 7 depicts the geometrical placement of the components. Notice that, since each beamsplitter deflects 50% of the light energy, the intensity of the light behind the camera is reduced by $\frac{3}{4}$ before hitting the scene, while the reduction for the flash in front is of $\frac{1}{2}$. In order to compensate for that difference, a neutral density filter that cuts the flash intensity in half was attached to the flash in front.

The images in Figure 8 show the obtained results. The left-right and above-below flash arrangements fail to detect



(a) (b)

Figure 9. Shadow-free image experiment. a) Computed maximum of the images taken with the left and right flashes; b) shadow-free image captured with the co-located flash. A problematic region is highlighted, and the detected depth edges are shown.

(a) (b)

Figure 10. Shadow-free image experiment (specularities). a) Maximum of the images taken with the left and right flashes; b) shadow-free image captured with the co-located flash.

horizontal and vertical depth edges, respectively; the front-behind flash setup fails to detect radial edges. Two examples of the degenerate case described in Section 3.5 can be seen in Figures 8(k-l): the circled region shows an area that is in shadow in all images.

As mentioned in Section 3.5, the degenerate case can be addressed by capturing a shadow-free image by placing the flash at the camera's center of projection. We built this configuration using a beamsplitter in a similar way to the front-behind setup (but placing a single light and the camera at equal distance from the beamsplitter). The direct shadow-free image was then used as the denominator for computing the ratio images in the depth edge detection algorithm (instead of the computed maximum image). In this experiment, only two flashes (left and right) were used due to constraints of our table-top configuration. Figure 9(a) shows the computed maximum of the images taken with the left and right flashes. The magnified view displays one of the problematic regions, where shadows are present. This leads to failures in the edge detection algorithm, as can be seen in the box. The same region has no shadows in the physically captured shadow-free image (Figure 9(b)), resulting in the improved depth map shown in the picture.

Computing the shadow-free image in this way has a few advantages as compared to the maximum composite image:

- The image is free of shadows, which is not always the case for the maximum image (*e.g.*, the case in Section 3.5 or setups with flashes with large baseline distance). This was explored in [18] to provide depth estimates in every pixel of the image;

- The shadow-free image is obtained in a single shot, which would be useful to generate shadow-free videos of dynamic scenes. This would be of interest to the medical community, as mentioned in [14];
- The shadow-free image will usually contain less specular reflection, as shown in Figure 10. In photography, the flash in the optical axis has the drawback of maximizing the red-eye effect, but, on the other hand, it enables the capture of a direct shadow-free image.

The main disadvantage of this method is the increased complexity of the setup: an additional light and a beamsplitter are needed, and they have to be accurately aligned. The image can have very small shadows due to imprecisions in the alignment and the use of non-point light sources. However, for depth edge detection those will at worst result in small errors in the position of the detected depth edges or missing edges in very small holes.

5. Conclusions and Future Directions

We have analyzed the space of shadows that can be created over depth edges given a camera-light pair. Our formulation is based on the geometry of the light epipolar rays in the image plane. The theory is applied to show that it is impossible to capture all depth edges in a general scene using fewer than three flashes in a multiflash imaging setup. It also provides a characterization of the detected and missed edges. For example, while horizontal edges are difficult to detect using left-right pairs, and vertical edges are not captured by above-below setups, we show that radial edges are problematic for front-behind pairs. The proposed theory provides a foundation for the choice of the illuminant positions in a multiflash setup for detection of depth edges.

We hope that our work will inspire investigations on a wide range of light sources. We have considered omnidirectional point light sources, but further analysis is required to take into account the off-axis intensity fall-off of spotlight sources, the distance between the light source and the depth edge (so that there is sufficient signal to noise ratio to distinguish lit regions from shadowed unlit regions), and non-point light sources such as area light sources. We have only looked at the binary problem of lit/unlit classification, but our work could be extended to an analysis of epipolar constraints on continuous variations in shading for light sources moving along parameterized paths.

Acknowledgements

This work was supported in part by the National Science Foundation under award number IIS-0535293.

References

[1] D. Akers, F. Losasso, J. Klingner, M. Agrawala, J. Rick, and P. Hanrahan. Conveying shape and features with image-based relighting. In *IEEE Visualization*, 2003. 1

[2] D. Crispell, D. Lanman, P. Sibley, Y. Zhao, and G. Taubin. Beyond sillhouettes: Surface reconstruction using multi-flash photography. In *International Symposium on 3D Data Processing, Visualization and Transmission*, 2006. 1

[3] M. Daum and G. Dudek. On 3-D surface reconstruction using shape from shadows. In *Conference on Computer Vision and Pattern Recognition*, pages 461–468, June 1998. 1

[4] R. Fattal, M. Agrawala, and S. Rusinkiewicz. Multiscale shape and detail enhancement from multi-light image collections. *SIGGRAPH*, 2007. 1

[5] R. Feris, R. Raskar, L. Chen, K. Tan, and M. Turk. Discontinuity preserving stereo with small baseline multi-flash illumination. In *International Conference on Computer Vision*, Beijing, China, 2005. 1

[6] R. Irvin and D. McKeown. Methods for exploiting the relationship between buildings and their shadows in aerial imagery. *IEEE Systems, Man, and Cybernetics*, 19(6):1564–1575, 1989. 1

[7] D. Kriegman and P. Belhumeur. What shadows reveal about object structure. *Journal of the Optical Society of America*, pages 1804–1813, 2001. 1

[8] M. Langer, G. Dudek, and S. Zucker. Space occupancy using multiple shadow images. *International Conference on Intelligent Robots and Systems*, pages 390–396, 1995. 1

[9] R. Ramamoorthi, M. Koudelka, and P. Belhumeur. A Fourier theory for cast shadows. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(2):288–295, 2005. 1

[10] R. Raskar, K. Tan, R. Feris, J. Yu, and M. Turk. A non-photorealistic camera: depth edge detection and stylized rendering using multi-flash imaging. *SIGGRAPH*, 2004. 1, 3, 4, 6

[11] D. Raviv, Y. Pao, and K. A. Loparo. Reconstruction of three-dimensional surfaces from two-dimensional binary images. In *IEEE Transactions on Robotics and Automation*, volume 5(5), pages 701–710, Oct 1989. 1

[12] S. Savarese, H. Rushmeier, F. Bernardini, and P. Perona. Shadow carving. In *International Conference on Computer Vision*, Vancouver, Canada, 2001. 1

[13] J. Segen and S. Kumar. Shadow gestures: 3D hand pose estimation using a single camera. In *Conference on Computer Vision and Pattern Recognition*, pages 479–485, 1999. 1

[14] K. Tan, J. Kobler, P. Dietz, R. Feris, and R. Raskar. Shape-enhanced surgical visualizations and medical illustrations with multi-flash imaging. In *Intl. Conf. on Medical Imaging Computing and Comp. Assist. Intervention*, France, 2004. 8

[15] S. Yamazaki, S. Narasimhan, S. Baker, and T. Kanade. Coplanar shadowgrams for acquiring visual hulls of intricate objects. In *International Conference on Computer Vision*, Rio de Janeiro, Brazil, 2007. 1

[16] D. K.-M. Yang. *Shape from darkness under error*. PhD thesis, Columbia University, 1996. 1

[17] Y. Yu and J. Chang. Shadow graphs and 3D texture reconstruction. *Int. J. Comput. Vision*, 62(1-2):35–60, 2005. 1

[18] L. Zhang and S. Nayar. Projection defocus analysis for scene capture and image display. *SIGGRAPH*, 2006. 7