# SparkleVision: Seeing the World through Random Specular Microfacets

Zhengdong Zhang, Phillip Isola, and Edward H. Adelson
Massachusetts Institute of Technology
{zhangzd, phillipi, eadelson}@mit.edu

## Abstract

*In this paper, we study the problem of reproducing the light from a single image of an object covered with random specular microfacets on the surface. We show that such reflectors can be interpreted as a randomized mapping from the lighting to the image. Such specular objects have very different optical properties from both diffuse surfaces and smooth specular objects like metals, so we design a special imaging system to robustly and effectively photograph them. We present simple yet reliable algorithms to calibrate the proposed system and do the inference. We conduct experiments to verify the correctness of our model assumptions and prove the effectiveness of our pipeline.*

## 1. Introduction

An objects appearance depends on the properties of the object itself as well as the surrounding light. How much can we tell about the light from looking at the object? If the object is a mirrored sphere, it reflects a complete picture of the surrounding scene. Specular reflections from naturally occurring objects, such as the human cornea, can be used as well [10]. On the other hand, diffuse reflectors mix lights from multiple directions, and they provide limited information[3, 12, 11].

Here we consider specular reflections from non-smooth surfaces. Figure 1 shows the example of a sequined dress. Each sequin is a tiny facet that reflects light from a particular direction. A dress with thousands of sequins can in principle tell us about light from thousands of directions. It is similar to a reflecting sphere except that the facet positions have been scrambled. A great many surfaces in the world provide specular reflections in multiple directions (e.g., wet pavement), and it is worthwhile asking how much information they offer. We call this the problem of Sparkle Vision.

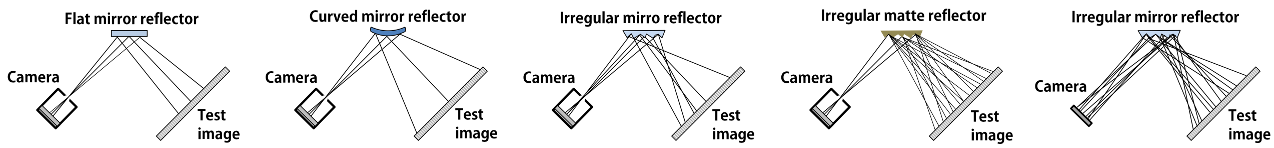
To demonstrate the similarity and difference of Sparkle Vision with conventional optical arrangements, we illustrate in Figure 2 how light rays travel from a scene to a camera sensor by way of a reflector in these optical systems. For simplicity we assume the scene is planar; for example it could be a computer display screen showing a test image.

Figure 1. How much information about the surrounding light can a dress with thousands of sequins tell us in principle? We call this the problem of Sparkle Vision.

A subset of rays are seen by the sensor in the camera. Here we show a pinhole camera for simplicity.

Figure 2(a) shows the case of an ordinary flat mirror reflector. The pinhole camera forms an image of the display screen (reflected in the mirror) in the ordinary way. There is a simple mapping between screen pixels and sensor pixels. Figure 2(b) shows the same arrangement with a curved mirror. Again there is a simple mapping between screen pixels and sensor pixels. The field of view is wider due to the mirror's curvature. Figure 2(c) shows the case of a smashed mirror, which forms an irregular array of mirror facets. The ray directions are scrambled, but the mapping between screen pixels and sensor pixels is still relatively simple. This is the situation we consider in the present work.

Figure 2(d) shows the case of an irregular matte reflector. Each sensor pixel sees a particular point on the matte reflector, but that point integrates light from a broad area of the




(a) flat mirror (b) curved mirror (c) smashed mirror, *sparkle-vision* (d) irregular matte reflector (e) irregular mirror without pinhole camera

Figure 2. Optical arrangements in which light rays travel from a scene to a camera sensor by way of a reflector: "SparkleVision" refers to the setup in (c).

display screen. Unscrambling the resulting image is almost impossible, although there are cases where some information may be retrieved, as shown by [16] in their discussion of accidental pinhole cameras. Figure 2(e) shows the case of an irregular mirror, but without benefit of a pinhole camera restricting the rays hitting the sensor. This case corresponds to the random camera proposed by Fergus et al [6], in which the reflector itself is the only imaging element. Since each pixel captures light from many directions, unscrambling is extremely difficult.

The case in Figure 2(c) is Sparkle Vision. It involves relatively little mixing of light rays, so unscrambling seems feasible. Moreover it could be of practical value, since irregular specular surfaces occur in the real world (e.g., with metals, certain fabrics, micaceous minerals, and the Fresnel reflections from foliage or wet surfaces).

For a surface covered in glitter, it is difficult to build a proper physical model. Instead of an explicit model, we can describe the sparkly surface plus camera as providing a linear transform on the test image. With a planar display screen, each sparkle provides information about some limited parts of the screen. Non-planar facets and limited optical resolution will lead to some mixture of light from multiple locations. However, the transform is still linear. There exists a forward scrambling matrix, and in principle we can find its inverse and unscramble the image.

To learn the forward matrix we can probe the system by displaying a series of test images. These could be orthogonal bases, such as a set of impulses, or the DCT basis functions. They could also be non-orthogonal sets, and can be overcomplete. Having determined the forward matrix we can compute its inverse.

Any reflective surface, viewed by a camera, can be considered to perform a linear transform on the light field. Fully inverting this transform is impossible due to the noise and dimensionality reduction inherent in the imaging process. At the same time, much information can be retrieved. Our goal is to understand how difficult the problem is. Since no one has tried it before, no one knows whether it is trivially easy or hopelessly difficult.

We consider a simplified case: a planar surface covered with glitter, which is reflecting a planar scene displayed on a computer monitor. In addition we allow ourselves to calibrate the system by displaying test images on the monitor. These conditions provide a best case, that allows to determine the upper limit on the image information available in a single photograph of a sparkly object. In most real situations the task will be harder.

All the optical systems shown in Figure 2 implement linear transforms, and all can be characterized in the same manner. However, if a system is ill-conditioned, the inversion will be noisy and unreliable. The performance in practice is an empirical question. We will show that the case in Figure 2(c), SparkleVision, allows one to retrieve an image that is good enough to recognize objects and faces.

### 1.1. Related Work

A diffuse object like a ping pong ball tells us little about the light around it. If a Lambertian object is convex, its appearance approximately lies in a nine-dimensional subspace [3, 12, 11], making it impossible to reconstruct more than a $3 \times 3$ environment map. For non-convex objects, the image of the object under all possible lighting conditions lies in a much higher dimension space due to shadows and occlusions[19], enabling the reconstruction of light beyond $3 \times 3$ [7]. But in general, matte surfaces are tough to work with.

A smooth specular object like a curved mirror provides a distorted image of the light, which humans can recognize [4] and algorithms can process [2, 5, 17]. However, specular random facets are different. Typically they are highly irregular and discontinuous, making it hard even for humans to interpret what they reflect. We utilize a model similar to inverse light transport [13, 14] to analyze this new setup, and propose a novel pipeline to effectively reduce the noise and increase the stability of the system, both in calibration and reconstruction.

Researchers have applied micro-lens arrays to capture the light[1, 9]. To some extent, a specular reflector can also be considered as a coded aperture of a general camera system[8]. Our work differs from the previous work in the sense that our setup is randomized – to the best of our knowledge previous work in this domain mainly uses specially manufactured arrays with known mappings whereas in our system the array is randomly distributed.

Many ideas in this paper are inspired by previous work on Random Cameras [6]. However, the key difference between our paper and the previous work is that in [6] no lens is used and hence all light rays from all directions get mixed up, and this is difficult to invert. In our setup, we place a lens between the world and the camera sensor, which makes the problem significantly easier and more tractable to solve. Similar ideas also appear in the "single pixel camera" [15] where measurements of the light are randomized for compressed sensing.

The idea that some everyday objects can accidentally serve as a camera has been explored before. It is in shown in [10] that a photograph of a human eye reflects the environment in front of the eye, and this can be used for relighting. In addition, a window or a door can act like a pinhole, in effect imaging the world outside the opening[16].

# 2. The formulation of SparkleVision

We discuss the optical setup of SparkleVision in discretized settings. Suppose the light in a particular environment is denoted by a stacked, discrete vector $x$ in the space. We place a specular object $O$ with random specular microfacets into the environment. Further we use a camera $C$ with a focused lens to capture the intensity of the light reflected by $O$. Let the discrete vector $y$ be the sensor output. It is well known that any passive optical system is linear. So we use a matrix $\mathcal{A}(\cdot)$ to represent the linear mapping relating the world lighting $x$ to $y$. Therefore, $y = Ax$.

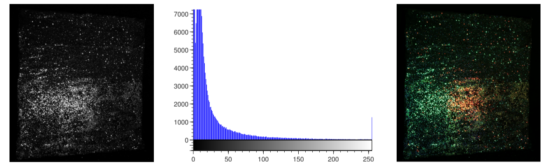
Note that all the above discussion makes no assumption on any material, albedo, smoothness or continuity properties of the objects in the scene. Therefore, this linear representation holds for any random specular microfacets. In this notation, the task of *SparkleVision* can be summarized as

- Accurately capture the image $y$ of a sparkling object.
- Determine the matrix $A$, which is a calibration task.
- Infer the light $x$ from the image $y$, which is an inference task.

In the later discussion, we will use many pairs of lightings and images so we use the subscript $(x_j, y_j)$ to denote the $j$-th pair of them. In addition, let $e_i$ be the $i$-th unit vector of the identity basis, i.e., a vector whose entries are all zero except the $i$-th entry which is one. Similarly, let $d_i$ represent the $i$-th unit vector of the bases of the Discrete Cosine Transform (DCT). We use $b_i$ to represent a random basis vector where all entries are i.i.d random variables. Also let $A = [a_1, a_2, \dots, a_N]$ with $a_i$ as its $i$-th column.

# 3. Imaging specular random facets through HDR

In this section we examine the properties of sparkling objects with microfacets. Their special characteristics enable the recovery of lighting from an image while imposing unique challenges to accurately capture images of them. To deal with these challenges, we use High Dynamic Ranging (HDR) imaging, using multiple exposures of the same scene.



(a) image (b) histogram (c) non-overlap

Figure 3. Optical properties of specular random microfacets: (a) shows an image of specular facets with scattered bright spots. (b) shows its histogram, although the bright spots in the image are shining, most of pixels are actually dark. (c) the surface simultaneously illuminated by two adjacent impulse lighting, one in red and one in green. The reflected green lights and red lights seldom overlap, which is shown here by the fact that few spots in the image are yellow.

## 3.1. Sparkling Pattern under Impulse Lightings

Specular random microfacets can be considered as a randomized mapping between the light and the camera. Each single facet faces a random orientation. It acts as a mirror reflecting all the incoming lights. However, because of the existence of a camera with a focused lens and the small size of each facet, only lights from a very narrow range of directions will be reflected into camera from any given facet. Therefore, given a single point light source, only a very small number of the facets will reflect light to the camera and appear bright. The rest of the facets will not be illuminated. This effect makes the dynamic range of a photo of specular facets extremely high, creating a unique challenge when photographing them. Figure 3(a) and 3(b) plots a photo of such a reflector as well as its histogram.

Now suppose we slightly change the location of the impulse light, generating a small disturbance to the direction of the incoming light to the facets. Thanks to the tiny sparkling facets, this slight disturbance will cause a huge change in the light pattern on the random facets. Provided that the orientations of the facets are random across all the surfaces, we should expect that the set of aligned facets will be significantly different. Figure 3(c) gives us an illustration of this phenomenon. Intuitively, if our task is just to decide whether a certain point light source is on or not, we could just count whether the corresponding set of facets for that light's position is active or not. This also suggests that our system is very sensitive to the alignment of the geometric setup, which we will address in Section 6.6.

## 3.2. HDR Imaging

As we have seen, the dynamic range of an image of a sparkling object is extremely high. Dark regions are noisy and numerous throughout the image while the sparse bright spots have very high luminance. Therefore we employ a standard high dynamic range imaging technique (HDR) to capture the scene. For completeness, we briefly explain how we apply this technique in our scenario. Suppose we take multiple shots of the same scene with $K$ different exposure time $t_k$. Let the resulting images be $I_1, I_2, \ldots, I_k$. We can then combine those $K$ images into a single image $I_0$ with much higher dynamic range than any of the original K images. Note that we use a heavy tripod in the experiment and therefore we assume all $I_i$ are already registered. Therefore, for any arbitrary location $x$ we can determine its combined value independently. Let the exposure time and pixel intensity value be $(t_i, I_i(x))$. The goal is to determine the value $I(x)$ for each location $x$. We solve this problem by fitting a line to $(t_i, I_i(x))$ by least squares :

$$I(r) = \text{argmin}_s \sum_i (s \cdot t_i - I_i(x))^2 \tag{1}$$

With simple algebra we can derive a closed form solution:

$$I(r) = \left(\sum_i t_i I_i(r)\right) / \left(\sum_i t_i^2\right) \tag{2}$$

# 4. Calibration and Inference of SparkleVision System

In this section, we examine the algorithm to calibrate the system and reconstruct the environmental lighting $x$ from $y$.

## 4.1. Calibration with overcomplete basis

We probe the system $y = Ax$ by illuminating the object with impulse lights $e_i$. Ideally, $y_i = A \cdot e_i = a_i$. So we can scan through all $e_i$ to get $A$. However, due to the presence of noise the measured $y_i$ will typically differ from the $a_i$ of an ideal system. As we will show later in the experimental sections, this noise on the calibrated matrix $A$ is lethal to the recovery of the light. Our system relies on a clean, accurate transformation matrix $A$ to succeed. Therefore, we further probe the system with multiple different bases. Specifically we use the DCT basis $d_i$ and a set of random bases $b_i$. Doing this we make the system over-complete and hence the estimated $A$ becomes more robust to noise. Let $E$ be the $N \times N$ impulse basis matrix, $D$ be the DCT basis matrix and $B_K \in \mathbb{R}^{N \times K}$ be the matrix of $K$ random basis vectors. This implies the following optimization to do the calibration:

$$\min_A \|Y_1 - AE\|_F^2 + \lambda \|Y_2 - AD\|_F^2 + \lambda \|Y_3 - AB_K\|_F^2 \tag{3}$$

$\lambda$ here is a weight to balance the error since the illumination from impulse lights tend to be much dimmer than the illumination from DCT and random lighting. In our experiments we set $\lambda = \frac{1}{N}$.

To further refine the quality of the calibrated $A$ against the noise in the dominant dark regions of $A$, we only retain intensities above a certain intensity during calibration. Specifically let $\Omega_i$ be the set of the 1% brightest pixels in $y_i$ illuminated by the impulse $e_i$. Let $\Omega = \bigcup_i \Omega_i$. We then only keep the pixels inside $\Omega$ and discard the rest. Let $P_\Omega(\cdot)$ represent such a projection. This turns the calibration into the following optimization:

$$\min_A \|P_\Omega(Y_1) - AE\|_F^2 + \lambda \|P_\Omega(Y_2) - AD\|_F^2 + \lambda \|P_\Omega(Y_3) - AB_K\|_F^2 \tag{4}$$

Note that the size of the output $A$ here is different from (3) due to the projection $P_\Omega(\cdot)$.

## 4.2. Reconstruction

Given $A$, reconstructing light from an image $y$ is a classic inverse problem. A straightforward approach to this problem is to solve it by least-squares. However, this unconstrained least square may produce entries less than 0, which is not physically meaningful. Instead we solve a constrained least squares problem:

$$\min_x \|y - Ax\|_F^2, \;\; s.t. \;\; x \geq 0 \tag{5}$$

Nevertheless, through experiments we find this optimization is too slow for practical application. When the resolution of the screen is $20 \times 20$, i.e., $x \in \mathbb{R}^{400}$, solving the inequality constrained least square is approximately 100 times slower than solving the unconstrained system. Yet the improvement is minor. So we just solve the unconstrained least squares system and then crop the out-of-range pixels back to $[0, 1]$.

We observe that in practice there is room for improvement to smooth the outcome of the above optimization. For example, we could impose stronger image priors to make the result more visually appealing. However, doing so would disguise some of the intrinsic physical behavior of SparkleVision, and therefore we decide to stick to the most naive optimization (5).

## 4.3. Extensions and implementation details

We use RAW files from the camera to avoid any nonlinearity post-processing in image formats like JPEG and PNG. In addition, we model the background light as $e_0$ and the photograph under only background lighting as $y_0 = Ae_0$ by turning all active light sources off. We subtract $y_0$ from every $y_i$ in the experiments by default. Since $y_0$ is used for all of $y_i$, we photograph it multiple times and take the average as the actual image $y_0$ to suppress the noise on $y_0$.

We can easily extend the pipeline to handle color images where the transformation matrix $A$ is $\mathbb{R}^{3M \times 3N}$ instead of $\mathbb{R}^{M \times N}$. For calibration, just use enumerate $e_i$ three times in red, blue and green.

# 5. Simulated Analysis of SparkleVision

In this section we conduct synthetic experiments to systematically analyze how noise affects the performance of

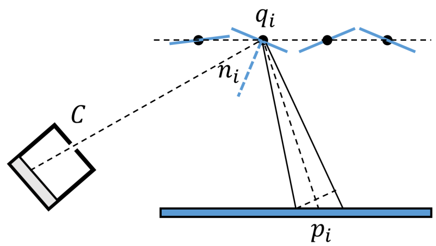



Figure 4. Configuration of the synthetic simulation. The pixel $p_i$ with certain width is reflected by the facet $q_i$ to the camera.

the proposed pipeline. In addition, we study how the size of the mirror facets and the spatial resolution of the sparkling surface influence the resolution of the light that the system can recover. These experiments improve our understanding on the limits of the geometric setup, provide guidance to tune the setup, and help us interpret the results.

### 5.1. Setup of the Simulated Experiment

The setup of the simulated experiment is shown in Figure 4, where a planar screen is reflected by a sparkling surface to the camera. The resolution of the screen is the resolution of the light. We model the sparkling plane as a rectangle with fixed size. We divide the plane into blocks and we place a mirror facing certain orientation in the block. We refer to the number of blocks in the unit area of the plane as its resolution. For simplicity, we do not consider inter-reflection or occlusion between the mirrors facets.

For each mirror facet, we assume that its orientation is sampled from a probability distribution. Let $\theta \in [0, \pi/2]$ be the slant of the orientation and $\phi \in [0, 2\pi)$ be the its tilt. We model the tilt $\phi$ as uniformly distributed in $[0, 2\pi)$. In practice the mirror facets are centered around 0. Therefore we model them as sampled from the positive half of the Gaussian distribution with standard deviation $\sigma_\theta$. Specifically, we have

$$\mathrm{P}_\theta(\theta_0) = \frac{2}{\sqrt{2\pi}\sigma_\theta} \exp\left(-\frac{\theta_0^2}{2\sigma_\theta^2}\right), \theta_0 \geq 0 \tag{6}$$

We assume that the orientation of each facet does not depend on any other so we can independently sample its value and create a random reflector. We use the classic ray-tracing algorithm to render the light reflected by the specular surface into the camera. We test the pipeline in this synthetic setup and find that in the ideal noise free case the recovery is perfect.

### 5.2. Sensitivity to Noise

For simplicity, we model the image noise as i.i.d. white noise. We perform three control experiments to tease apart the effects of noise during calibration versus during test time reconstruction. In the first test, we add noise to both calibration and test images. In the second test, we only add

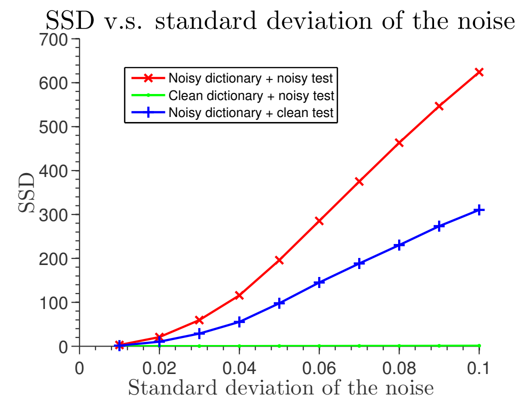



Figure 5. How noise in the calibration and the reconstruction stages affects the recovery accuracy. Our pipeline is stable to the noise in the reconstruction stage, but not stable to the noise in the calibration stage.

noise to the training dictionary while in the third we only add noise to the test images. Suppose we test our pipeline on $N$ test light $I_i, 1 \leq i \leq N$ and get the recovered $\widehat{I}_i$. We measure the error of the recovery by the average sum of squared difference (SSD) between $I_i$ and $\widehat{I}_i$. Varying the noise standard deviation from 0.01 to 0.1, we get three curves for the tests, which are shown in Figure 5.

These results indicate that our system is much more robust to the noise in the test image than to the noise in the images for calibration. In fact, when the standard deviation of the test image noise is 0.1 the recovery is still very good when using clean calibration images. In addition, when the noise std is as low as 0.01, the SSD with only calibration stage noise is 1.85 while the SSD with only testing stage noise is just 0.14. This comparison therefore validates the need to use an over-complete basis in our proposed pipeline to reduce the noise in the calibration stage.

### 5.3. Impact of Spatial Resolution

The spatial resolution of the random reflector determines the resolution of the light map that we can recover. Keep in mind that each microfacet in our model is a mirror and our system relies on the light from the screen reflected by the facet to the camera. If some part of the screen is never reflected to the camera, there is no hope to recover what that part of the screen is showing from the photograph taken by the camera. Since the facets are randomly oriented, this undesirable situation may well happen. Intuitively, if we have more facets, the chance that each part of the light is reflected to the camera will increase.

We develop a mathematical model to approximately compute the probability that a block of pixels on the screen will be reflected by at least one microfacets to the camera sensor. The model involves several approximations such as a small angle approximation, so the relative values in this analysis are more important than the raw values. Following the general setup in Figure 4, we first calculate the probability that a certain pixel $p_i$ on the screen gets reflected by the microfacet $q_i$ to the camera. Suppose the width of the pixel is $w$, then the foreshortened area of the pixel with respect to the incoming light direction $\overrightarrow{p_i q_i}$ is $w^2 \cos\theta$, where $\theta$ is the angle between $\overrightarrow{q_i p_i}$ and the screen. Then the solid angle of this foreshortened area with respect to $q_i$ is $\frac{w^2 \cos\theta}{\|\overrightarrow{p_i q_i}\|}$.

The normal $n$ that just reflects $\overrightarrow{p_i q_i}$ to the camera $C$ is the normalized bisector of $\overrightarrow{q_i p_i}$ and $\overrightarrow{q_i C}$. Since the incoming lights can vary in the solid angle of $\frac{w^2 \cos\theta}{\|\overrightarrow{p_i q_i}\|}$, $n$ can vary in $\frac{w^2 \cos\theta}{4\|\overrightarrow{p_i q_i}\|}$ and still the mirror can reflect some light from the pixel on the screen to the camera. Let $q_i \circ p_i$ be the event that the facet at $q_i$ will reflect some light emitted from $p_i$ to the camera $C$. Then its chance is the same as the probability for the orientation of the facet to be within that range, which is approximated by

$$\Pr(q_i \circ p_i) = \frac{2}{\sqrt{2\pi}\sigma_\theta} \exp\left(-\frac{\theta_0^2}{2\sigma_\theta^2}\right) \frac{w^2 \cos\theta}{4\|\overrightarrow{p_i q_i}\|} \tag{7}$$

Suppose there are $M$ microfacets in total and we compute $\Pr(q_i \circ p_i)$ for all $1 \leq i \leq M$. Then we can compute the probability that the light from pixel $p_i$ is reflected by at least one microfacet to the camera as follows.

$$\begin{aligned}\Pr(\exists j, q_j \circ p_i) &= 1 - \Pr(\forall j, q_j \not\circ p_i) = 1 - \prod_j \Pr(q_j \not\circ p_i) \\ &= 1 - \prod_j (1 - \Pr(q_j \circ p_i))\end{aligned}$$

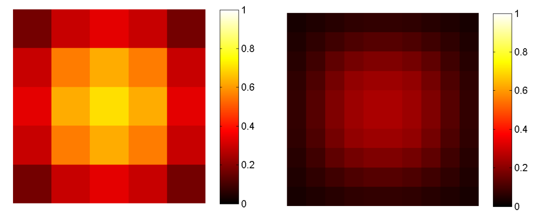
We visualize this probability in Figure 6 in four different configurations of the screen and specular surface resolutions.

From the results, we can see that overall higher resolution of the specular object and lower resolution of the screen will reduce the chance that some block of pixels on the screen are not reflected to the sensor. In addition, on the same screen, the chance to avoid such bad events are different for different blocks of pixels, which is due to the different distances and relative angles between different parts of the screen and the reflector. This suggests that for a specular object there will be a limit on the resolution of the light we can infer from it.

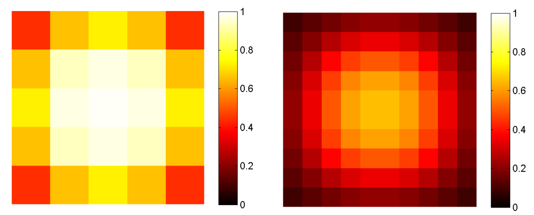
# 6. Experiments

## 6.1. Experiment setup

We place the sparkling object in front of a computer screen in a dark room and use a camera to photograph the object. The images displayed on the screen are considered as light. Figure 7(a) illustrates the setup. Specifically in this experiment, we use a 24-inches ACER screen, a CANON rebel T2i camera and a set of specular objects including a hair pin, skull, and painted glitter board. The camera is placed on a heavy tripod to prevent movement. We show



(a) Object $10 \times 10$, Screen $5 \times 5$ (b) Object $10 \times 10$, Screen $10 \times 10$



(c) Object $20 \times 20$, Screen $5 \times 5$ (d) Object $20 \times 20$, Screen $10 \times 10$

Figure 6. Probability map of light from a block pixels getting reflected by the specular surface to the screen.

that our system can reconstruct the world lighting at resolution up to $30 \times 30$. At this resolution many objects, such as faces, can be easily recognized.

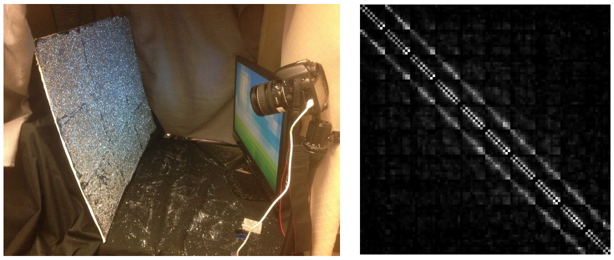
## 6.2. Examination of the assumptions of the system

**Overlapping of bright pixels** We measure quantitatively how the displacement of impulse lights will change the pattern of bright spots. Let $a_i$ be the image illuminated by an impulse light and $S_i$ be the set of bright pixels in $a_i$ with intensities larger than $1/10$ of the maximum. Then the overlap between $a_i$ and $a_j$, $i \neq j$ can be defined as

$$\mathrm{O}(i,j) = \frac{|S_i \cap S_j|}{\min(|S_i|, |S_j|)}, i \neq j \tag{8}$$

Here $|S|$ denotes set cardinality. At a light resolution of $10 \times 10$, there are 100 impulse basis images, and the overlap between each of them can be plotted in a $100 \times 100$ image with the entry at the $i$-th row and $j$-th column representing $\mathrm{O}(i,j)$, as is shown in Figure 7(b). As the figure suggests, most of the overlap happens between images from neighboring impulses. And the maximal overlap $\max_{i \neq j} \mathrm{O}(i,j) < 0.2$. This further validates the non-overlapping property of the SparkleVision system.

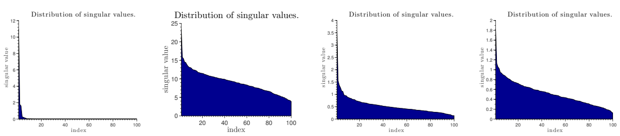
**Condition Number** The condition number of the transformation matrix $A$, $\kappa(A)$, determines the invertibility of a transformation matrix $A$. $\kappa(A)$ is defined as the ratio between the largest and the smallest singular values of $A$. For all the optical systems shown in Figure 8, we plot all the singular values of their $A$ in descending order. From the figure, we can see that the best $\kappa(A) \approx 4$, which is good in practice.



(a) Experiment Setup (b) Overlap graph

Figure 7. (a) shows the setup of the experiment in the lab. (b) shows the overlap graph between images from different impulses. It can be seen from the figure that only images from neighboring impulses have slight overlap.

(a) diffuse paper (b) glitter (c) hairpin (d) skull



(e) $\kappa = 1846.31$ (f) $\kappa = 6.02$ (g) $\kappa = 26.28$ (h) $\kappa = 14.72$

Figure 8. Singular value distribution of the transformation matrix of different objects. The condition number $\kappa$ is given below each graph. Note that the sparkling reflectors create systems with much lower condition numbers compared to those of diffuse objects.

### 6.3. Recovering images using SparkleVision

We first present SparkleVision using a board covered in glitter. We consider both grayscale and color imaging. For the grayscale setting, we push the resolution of the screen to $30 \times 30$. For the colored setting, we just present a few tests at a lower resolution of $15 \times 15$ to demonstrate that our system generalizes to color. The grayscale results are shown in Figure 9 and the color results are shown in Figure 10. They demonstrate the success of the pipeline at these resolutions.

Figure 11 shows SparkleVision using a 3D glittering skull. The recovery degrades as we increase the light resolution. But the artifacts are well predicted by our synthetic models.

### 6.4. Impact of number of basis for calibration

Figure 12 illustrates how increasing the number of random bases used in the calibration improves the recovery of the light $x$. Note that the resolution of the light in this setup is $20 \times 20$, hence the number of impulses and DCT bases are both $400$. It is worth noting that the benefit of more random bases gradually saturates so we only need to employ a limited number.

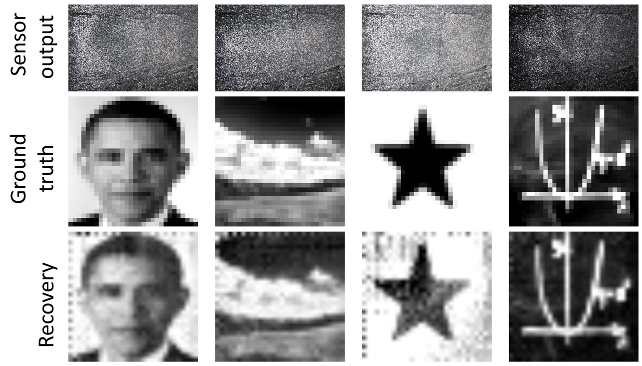



Figure 9. SparkleVision using board covered in glitter. Qualitatively the recovery is fairly close to the ground-truth light and a human can easily recognize the objects in the recovered image.

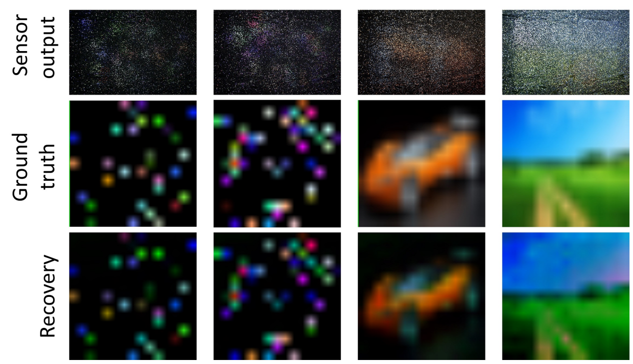



Figure 10. Color photography SparkleVision using board covered in glitter. Although there is slight color distortion, the overall quality of the recovery is good.

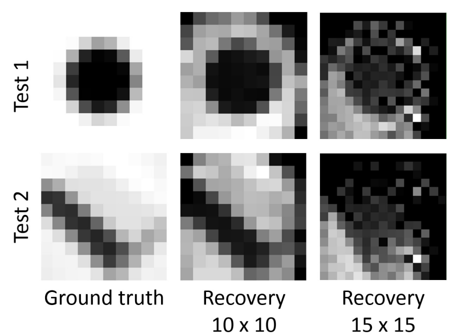



Figure 11. SparkleVision using a skull covered in glitter. The quality of the recovery degrades as the resolution increases. This observation matches the prediction from our synthetic modeling of light resolution.The dark spots indicate that little light is reflected by the skull to the camera.

### 6.5. Stability to noise

We perform synthetic experiments by adding noise to the real test image to understand how robust the calibrated transformation matrix is. We measure the robustness

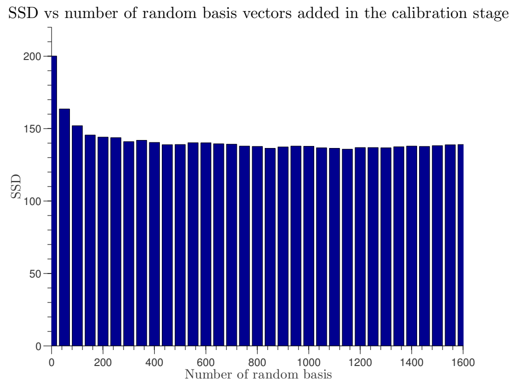


Figure 12. Adding more random basis vectors to the calibration helps to reduce the recovery error. But this benefit saturates.


(a) No noise (b) 0.04 (c) 0.08 (d) 0.12 (e) 0.16

Figure 13. Stability to noise on the test image: the title of the subfigure represents noise level. The noise is large considering the images are in $[0, 1]$ and only a few spots are bright.

by Root-mean-squared-error (RMSE) between the noisy recovery and the noise-less recovery. We plot how the reconstructed light changes as the noise level increases in Figure 13. This result validates the robustness of our system.

### 6.6. Sensitivity to misalignement and potential application

The success of SparkleVision relies largely on the sensitivity of the light pattern on a specular object to even a slight movement of the source light. However, this property simultaneously make the whole system extremely sensitive to subtle misalignment. To show this we perform synthetic experiments by shifting the test image $I$ by $\Delta x$ and examine the RMSE. Some representative results and the RMSE curve are shown in Figure 14. We could compensate for this misalignment by performing grid search over $\Delta x$ and pick out the recovery that has the minimum value of total variation $\sum_x \|\nabla I(x)\|_1$.

For the recovery of light, this phenomenon is harmful. But such sensitivity to even subpixel misalignment could enable the detection and magnification of tiny object motions, like in [18]. We leave this for future work.

## 7. Discussions and Conclusion

In this paper we have shown that it is possible to infer an image of the world around an object that is covered in random specular facets. This class of objects actually provide rich information about the environmental map and is significantly different from smooth objects with either Lamber-


(a) 1 pixel (b) 0.8 pixel (c) 0.4 pixel (d) 0.2 pixel (e) No shift

Figure 14. Instability to misalignment: even if we shift the test image by one pixel horizontally, there is significant degrade in the output. One way to compensate for this would be to grid search over shifts and choose the best reconstruction according to an image prior.

tian or specular surfaces, which researchers in the field of shape-from-X have worked on.

The main contributions of the paper are twofold. First, we have presented the phenomenon that specular random microfacets can encode a large amount of information about the surrounding light. This property may seem mysterious at the first sight but indeed is intuitive and simple once we understand it. We also analyzed the factors that affect the optical limits of these reflectors. Second, we proposed and analyzed a physical system that can efficiently perform the calibration and inference of the surrounding light map based on these sparkling surfaces.

Currently our approach only reconstructs a single image of the scene facing the sparkling object. Such an image corresponds to a slice of the lightfield around the object. Using similar setup, it should be possible to reconstruct other slices of the lightfield. Thus, our system could be naturally extended to work as a lightfield camera. In addition, this new reflector has the ability of reveal subtle motions of the optical setup. We leave all these exciting directions for future exploration.

## References


[1] E. H. Adelson and J. Y. Wang. Single lens stereo with a plenoptic camera. *IEEE transactions on pattern analysis and machine intelligence*, 14(2):99–106, 1992. 2

[2] Y. Adto, Y. Vasilyev, O. Ben-Shahar, and T. Zickler. Toward a theory of shape from specular flow. In *11th IEEE International Conference on Computer Vision (ICCV 2007)*. 2

[3] R. Basri and D.W.Jacobs. Lambertian reflectance and linear subspaces. *IEEE Transaction on Pattern Analysis and Machine Intelligence (TPAMI)*, 25:218 – 233, 2003. 1, 2

[4] A. Blake. Does the brain know the physics of specular reflection? *Nature*, 343(6254):165–168, 1990. 2

[5] G. D. Canas, Y. Vasilyev, Y. Adato, T. Zickler, S. Gortler, and O. Ben-Shahar. A linear formulation of shape from specular flow. In *12th IEEE International Conference on Computer Vision (ICCV 2009)*. 2

[6] R. Ferbus, A. Torralba, and W. T. Freeman. Random lens imaging. *MIT CSAIL Tecnical Report*, (058), 2006. 2, 3

[7] S. W. Hasinoff, A. Levin, P. R. Goode, and W. T. Freeman. Diffuse reflectance imaging with astronomical applications.

In *13th IEEE International Conference on Computer Vision (ICCV 2011)*. 2

[8] A. Levin, R. Fergus, F. Durand, and W. T. Freeman. Image and depth from a conventional camera with a coded aperture. *ACM Transactions on Graphics (TOG)*, 26(3):70, 2007. 2

[9] M. Levoy. Light fields and computational imaging. *IEEE Computer*, 39(8):46–55, 2006. 2

[10] K. Nishino and S. K. Nayar. Eyes for relighting. In *SIGGRAPH 2004*, pages 704 – 701. 1, 3

[11] R. Ramamoorthi and P. Hanrahan. A signal-processing framework for inverse rendering. In *SIGGRAPH 2001*, pages 117 – 128. 1, 2

[12] R. Ramamoorthi and P. Hanrahan. On the relationship between radiance and irradiance: determining the illumination from images of a convex lambertian object. *JOSA A*, 18:2448 – 2459, 2001. 1, 2

[13] S. M. Seitz, Y. Matsusita, and K. N. Kutulakos. A theory of inverse light transport. In *10th IEEE International Conference on Computer Vision (ICCV 2005)*. 2

[14] P. Sen, B. Chen, G. Garg, S. R. Marschner, M. Horowitz, M. Levoy, and H.P.A.Lensch. Dual photography. In *SIGGRAPH 2005*, pages 745 – 755. 2

[15] D. Takhar, J. Laska, M. Wakin, M. Duarte, D. Baron, S. Sarvotham, K. Kelly, and R. Baraniuk. A new compressive imaging camera architecture using optical-domain compression. In *Computational Imaging IV at SPIE Eletronic Imaging*, 2006. 3

[16] A. Torralba and W. T. Freeman. Accidental pinhole and pinspeck cameras: Revealing the scene outside the picture. In *25th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2012)*. 2, 3

[17] Y. Vasilyev, T. Zickler, S. Gortler, and O. Ben-Shahar. Shape from specular flow: Is one flow enough? In *24th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2011)*. 2

[18] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. T. Freeman. Eulerian video magnification for revealing subtle changes in the world. *ACM Transactions on Graphics (Proc. SIGGRAPH 2012)*, 31(4), 2012. 8

[19] Y. Zhang, C. Mu, H.-W. Kuo, and J. Wright. Toward guaranteed illumination models for non-convex objects. In *14th International Conference on Computer Vision (ICCV 2013)*. 2