Sparkle Vision: Seeing the World through Random Specular Microfacets

by

Zhengdong Zhang

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Computer Science and Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2014

Signature redacted

Author ... .
Department of Electrical Engineering and Computer Science
May 21, 2014

Signature redacted

Certified by ... .
Professor Edward H. Adelson
Thesis Supervisor

Signature redacted

Accepted by ....
Professor Leslie A. Kolodziejski
Chair, Department Committee on Graduate Theses

Sparkle Vision: Seeing the World through Random Specular Microfacets

by

Zhengdong Zhang

Submitted to the Department of Electrical Engineering and Computer Science
on May 21, 2014, in partial fulfillment of the
requirements for the degree of
Master of Science in Computer Science and Engineering

Abstract

This thesis studies the problem of reproducing the world lighting from a single image of an object covered with random specular microfacets on the surface. Such a reflector can be interpreted as a randomized mapping from the lighting to the image. This intrinsic randomness makes it challenging for humans to interpret the image of a specular surface. We propose a system to solve it algorithmically and demonstrate how a simple yet reliable method can calibrate the proposed system and do the inference. The success of such system relies on accurate exposure of the specular surfaces. However, such objects have very distinguished optical properties compared with both diffuse surfaces and smooth specular objects like metals. So we design a special imaging system to robustly and effectively photograph them. Finally we conduct experiments to verify the correctness of our model assumptions and prove the effectiveness of our pipeline.

Thesis Supervisor: Edward H. Adelson
Title: John and Dorothy Wilson Professor of Vision Science

Acknowledgments

I would like to acknowledge all the friends, lab mates, mentors and supervisors who guided me through the past two years at MIT. Without their company and help this thesis would not be possible.

First, I would like to express my sincere gratitude to my research supervisor Professor Edward H. Adelson for his patient guidance. I have learned a lot from his insightful and critical feedback during the discussion on both research projects and life.

Second, I would like to thank my lab mate Phillip Isola, who generously provided all kinds of technical instructions to me whenever I encountered a problem. He also demonstrated to me how to make the communication more efficient and effective, which benefits me a lot.

In addition, I would like to thank all of the genius lab mates and friends at MIT, who have been so amazing and friendly that all the moments I spent together with them shine in my memory.

Finally I would like to thank my Mom, my Dad and all of my families back in China who unconditionally offer the support whenever I need it.

Contents

1 Introduction **15**

1.1 Optical Arrangement of Sparkle Vision 15

1.2 Related Work 20

2 The notation and formulation of sparkle vision **25**

2.1 Notation 25

2.2 Imaging specular random facets through HDR 26

3 Calibration and Inference of Sparkle Vision System **29**

3.1 Calibration with Impulse Basis 29

3.2 Calibration with Overcomplete Basis to Reduce Noise 30

3.3 Reconstruction 32

3.4 Extensions and implementation details 32

4 Simulated Analysis of Sparkle Vision **35**

4.1 Setup of the Simulated Experiment 35

4.2 Simulated Experiment Results 37

4.3 Sensitivity to Noise 39

4.4 Impact of Spatial Resolution 42

5 Experiments and Results **47**

5.1 Experiment setup 47

5.2 Examine the assumption of the system 48

5.3 Experimental result 49

5.4 Stability to Noise 50
5.5 Instability to misalignement 50

6 Discussions and Conclusion **55**

List of Figures

1-1 Pictures of a piece of matte paper and Chicago cloud gate. The picture of the paper on the left tells few information of the surrounding light while the cloud gate acts as a slightly distorted mirror and its photo reveals so much information about the environment. 15

1-2 Reproducing the world from a single image of specular random facets: (a) shows the image of a surface covered with glitter illuminated by a screen showing the image of Obama. (b) gives a close up look of (a), highlighting both the bright spots and dark spots. (c) shows the lighting, i.e., the face of Obama the our algorithm constructs from (a). 16

1-3 Optical arrangement with flat mirror. 17

1-4 Optical arrangement with curved mirror. 18

1-5 Optical arrangement with smashed mirror. 18

1-6 Optical arrangement with irregular matte reflectors. 19

1-7 Optical arrangement with irregular mirror without pinhole camera. . 20

2-1 Optical properties of specular random micro facets: (a) shows an image of specular facet with scattered bright spots. (b) demonstrates its hsitogram, although the bright spots in the image catch the eyes, most of spots are actually dark. (c) the surface simultaneously illuminated by two impulses adjacent impulse lighting, one in red and one in green. The green lights and red lights seldom overlap because few spots in the image are yellow. 27

4-1 Synthetic photographs of specular objects under impulse light 38

4-2 Test light maps, photographs and results. The first row shows the images of the specular random surface illuminated by the test light. The second row presents the ground-truth test light on the screen. The third row shows the recovery by our algorithm. 39

4-3 The stability of our algorithm to noise added to the calibration stage and the test stage is different. The range of the image is in $[0, 1]$. . . . 40

4-4 The error of the recovery at noise level $\sigma = 0.01$. 41

4-5 Photography of a specular reflector with low spatial resolution. 42

4-6 Configuration of the mathematical model to compute the probability that p_i will be reflected by q_i to the camera. 43

4-7 Probability map of light from a block pixels getting reflected by the specular surface to the screen. 44

5-1 The left (a) shows the setup of the experiment in the lab. The right (b) shows the overlap graph between images from different impulses. It can be seen from the figure that only images from neighboring impulse have slight overlap. 49

5-2 Singular value distribution of the transformation matrix of different objects. Also the condition number is given. Note that the sparkling reflectors create systems with much lower conditional number compared with a diffuse object. 50

5-3 Sparkle vision through glitter board: Row 1 and 4 present the sensor output. Row 2 and 5 present the reconstruction, and Row 3 and 6 present the result of "sparkle vision" 51

5-4 Colored sparkle vision through glitter board: The first row presents the images of sparkling surfaces. The second row presents the ground-truth light map shown on the screen. The third row shows the reconstruction of sparkle-vision. The test images in the first two columns are synthetic scatter dots with random colors. The rest are real images from Internet. Although there is slight color distortion in the reconstruction, overall the inferred light is very close to the ground-truth. 52

5-5 Stability to noise: title of the subfigure represents noise level. They are pretty large considering the images are in $[0, 1]$ and only a few spots are bright. 52

5-6 Instability to misalignment: even if we shift the test image by one pixel horizontally, there is significant degrade in the output. 53

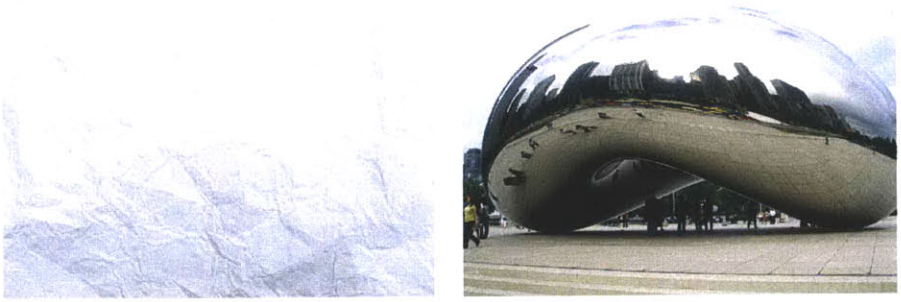
List of Tables

Chapter 1

Introduction

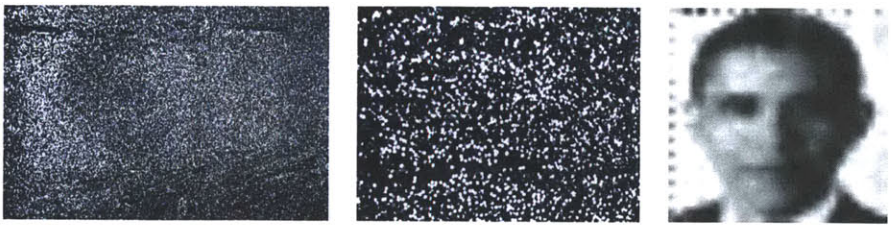
1.1 Optical Arrangement of Sparkle Vision

An object's 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 smooth and matte, we can tell rather little [3, 14, 13, 15]. The reason is that the surface of such objects uniformly reflects the incoming light into all of the directions. As a consequence the light map of the world gets blurred together. A simple example of such objects is a sheet of white paper whose appearance is always the same to human eye whatever lighting environment it is in. A picture of such



(a) Matte paper (b) Cloud gate

Figure 1-1: Pictures of a piece of matte paper and Chicago cloud gate. The picture of the paper on the left tells few information of the surrounding light while the cloud gate acts as a slightly distorted mirror and its photo reveals so much information about the environment.



(a) Specular microfacets (b) Closer look (c) Reconstructed lighting

Figure 1-2: Reproducing the world from a single image of specular random facets: (a) shows the image of a surface covered with glitter illuminated by a screen showing the image of Obama. (b) gives a close up look of (a), highlighting both the bright spots and dark spots. (c) shows the lighting, i.e., the face of Obama the our algorithm constructs from (a).

paper is shown in Figure 1-1(a). For such objects, it is not difficult to believe that we cannot tell much about the light from an photo of them. However, if the object is irregular and/or non-matte, there are more possibilities. For instance, an image of the Cloud Gate at Chicago in Figure 1-1(b) reveals an distorted version of the surrounding light, i.e., the appearance of the city and the sky.

In this work we are considering a more complicated class of irregular, non-matte objects. Figure 1-2 shows a picture of a surface covered in glitter. The glitter is sparkly, and the image shows a scattering of bright specularities. We may think of the glitter as containing mirror facets randomly oriented. Each facet reflects light at a certain angle. If we knew the optical and geometrical properties of the facets, we could potentially decode the reflected scene.

Figure 1-3 to Figure 1-7 show a variety of optical arrangements in which light rays travel from a scene to a camera sensor by way of a reflector. For simplicity we assume the scene is planar; for example it could be a computer display screen showing a test image. A subset of rays are "seen" by the sensor in the camera. Here we show a pinhole camera for simplicity.

Fig. 1-3 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 one-to-one mapping between screen pixels and sensor pixels. The mapping does not distort the light and hence the light map can be inverted from the

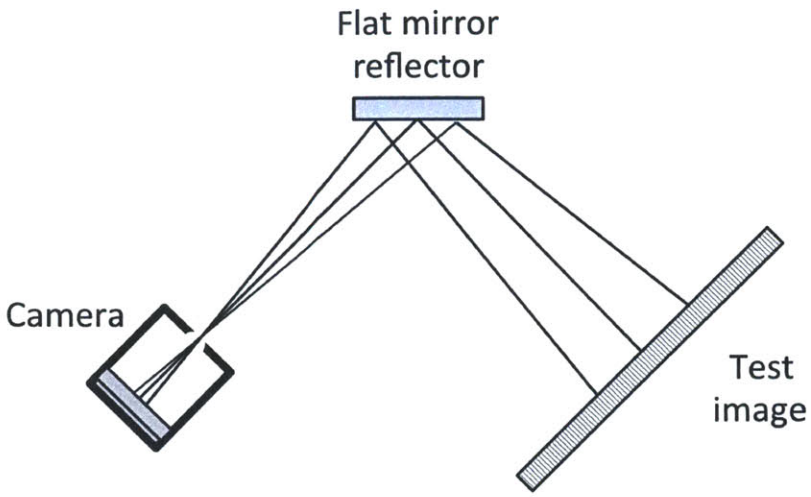


Figure 1-3: Optical arrangement with flat mirror.

image in a straight forward manner.

Fig. 1-4 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. However, the light is slightly distorted by the curved mirror and therefore to invert the lighting we need to first calibrate the mapping.

Fig. 1-5 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. Intuitively the mapping in this case is a random shuffling of the mappings discussed before. As a result, it is not obvious for human eye to establish the mapping. This is the situation we consider in the present work.

Figure 1-6 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 display screen. Unscrambling the resulting image is almost impossible, although there are cases where some information may be retrieved, as shown by [19] in their discussion of accidental pinhole cameras.

Fig. 1-7 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

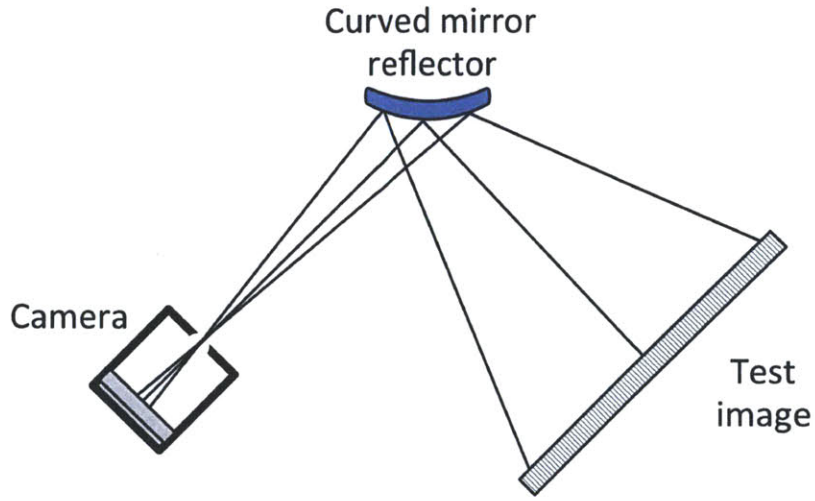


Figure 1-4: Optical arrangement with curved mirror.

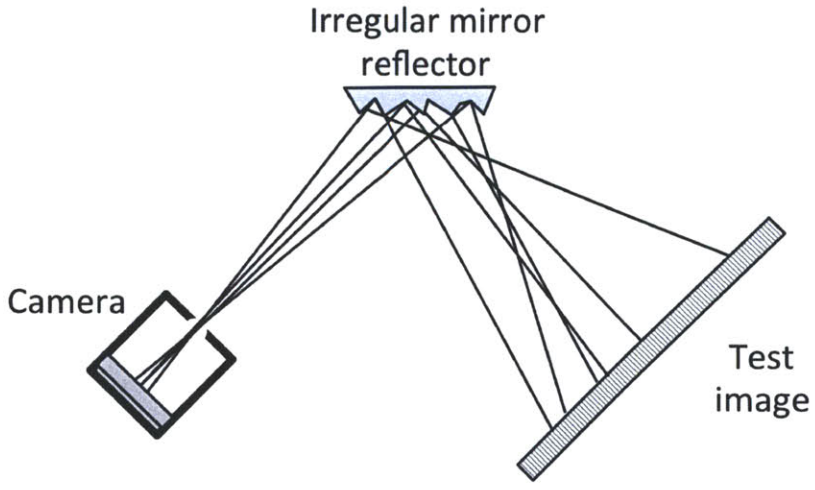


Figure 1-5: Optical arrangement with smashed mirror.

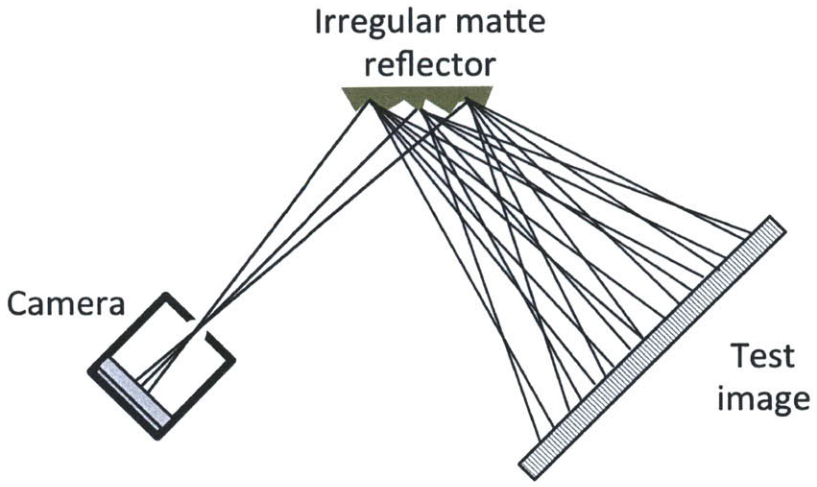


Figure 1-6: Optical arrangement with irregular matte reflectors.

camera" proposed by Fergus et al [8], 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 fig. 1-5, with a sparkly surface and a pinhole camera, deserves study. We call this case "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

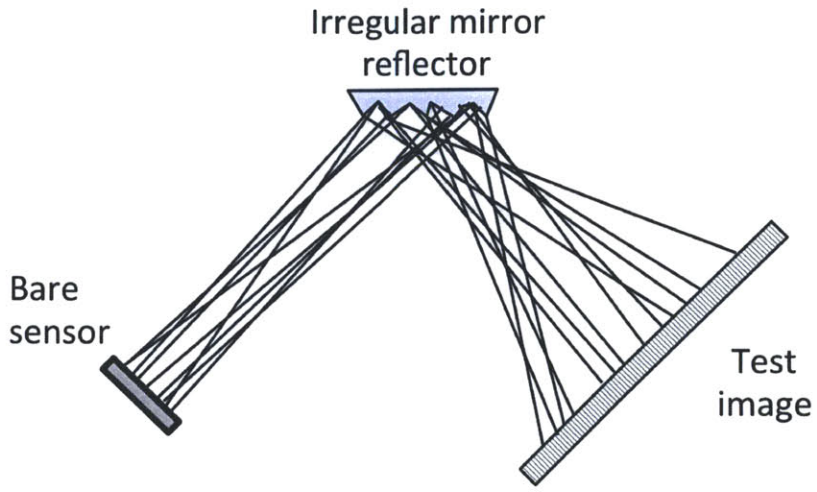


Figure 1-7: Optical arrangement with irregular mirror without pinhole camera.

determined the forward matrix we can compute its inverse.

All the optical systems shown in Fig. 1-5 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 fig. 1-5, sparkle vision, allows one to retrieve an image that is good enough to recognize objects and faces.

1.2 Related Work

The problem of estimating the environmental map or scene lighting from a single image of an object has a long history in computer vision. To analyze the relationship between the lighting and the appearance of the objects, researchers have built many mathematical models to characterize the reflectance property of the objects. The most classical model is rendering equation based on the bidirectional reflectance distribution function (BRDF).

Let ω_i, ω_o be the direction of the incoming and outcoming light. The BRDF function $\rho(\omega_i, \omega_o)$ is a four-dimensional function that measures the amount of the light coming from ω_i reflected to ω_o. Let $L(x, \omega)$ be the irradiance of the light at spatial

location x with direction ω. Let $L_i(x, \omega_o)$ be the outcoming light and $L_o(x, \omega_i)$ be the incoming light. For simplicity, we assume that the object itself is not emitting any light rays. Therefore, the rendering equation can be written as

$$L_o(x, \omega_o) = \int_\Omega \alpha(x) L(x, \omega_i) \rho(\omega_i, \omega_o) V(x, \omega_i) \max(\omega \cdot n, 0) \, \mathrm{d}\omega_i \tag{1.1}$$

where x is a point on the surface, n is the surface normal, Ω is the hemisphere above the surface, $\alpha(x)$ is the albedo of the surface, $V(x, \omega_i)$ is the visibility mask for the specific light coming from ω_i.

Note that Eqn. (1.1) is linear with respect to $L(x, \omega_i)$, which respects the fact that any passive optical system is linear. Under this model, the problem of inverting the environmental map from the image can be rephrased as inverting the linear system determined from the rendering equation. Several important factors, like the geometric setup and the BRDF function $\rho(\cdot, \cdot)$, can affect how easily or difficult is it to invert such a system.

A diffuse object like ping pong ball tells us little about the details of the lighting. Such an object is usually modeled as a Lambertian object whose BRDF function is constant with respect to ω_o. If a Lambertian object is convex, its appearance approximately lies in a nine-dimensional subspace [3, 14, 13, 15], making it impossible to reconstruct an environment map with resolution better than a 3×3. For non-convex objects, thanks to the existence of strong shadows, many more images are needed to cover its appearance under all possible lighting conditions [21]. These additional images allow us to go beyond the 3×3 limitations and estimate a much better lighting map. Furthermore, special geometric setups like occluding geometry may help [9]. But in general, it is tough to work with these matte surfaces.

A smooth specular object like a curved mirror provides a distorted image of the lighting. Humans are capable of inferring the shape of the surface[4]. Inspired by humans, researchers have developed algorithms to enable computers to perform the same task[2, 5, 20]. All these success suggests that it is not infeasible to invert the light map for mirrors of reasonble shape.

However, for specular random facets, it is hard even for humans to perceive a coherent reflection of the world in the sparkling facets. This is because we add a layer of random shuffling on top of everything. This additional complexity fails most of the previous algorithms. We cannot rely on them to invert the optical arrangement that sparkle vision studies.

It is worthy noting that there is a class of techniques called inverse light transport[16, 17] that swaps the role between the incoming and the outcoming light and study their relationship accordingly. However, the theory developed for these techniques mostly focuses on the Lambertian case which is easy to analyze. Moreover, it is usually assumed that the BRDF is spatially invariant, which is far from true for the sparkle vision, where the specular surface is highly spatially discontinous, let alone invariant.

Randomness plays an important role in sparkle vision. We can view the specular microfacets scattered around the surface as a randomized mapping between the incident light and the image. This is related with the random measurement of the image that researchers have worked on in the field of compressive sensing. They find it actually possible to reconstruct the original image exactly from multiple random linear measurements of the lightfield. Based on this concept, they have built the "single pixel camera" [18], which employs specially designed optics to perform one randomized measurement over the world at a single pixel one at a time and use compressive sensing techniques [6] to reconstruct the light map from multiple measurements.

Many ideas in this thesis are inspired by previous work on random camera [8]. The random camera basically replaces the lens in the usual camera with randomly oriented reflector or refractor. Therefore, the mapping between the light and the sensor output become randomized. However, the key difference between our work and the random camera is that in [8] no lens is used and hence all the lights from all directions in the lightfield get mixed up, making it difficult to reconstruct the environment map from the sensor output. In our setup, we place a lens between the world and the camera sensor, which restricts the light that hits a sensor to be from only one particular direction. This makes the problem significantly easier and more tractable to solve.

In addition, researchers have applied micro-lens arrays to capture lightfields [1]. For a summary of the literature, please refer to [11]. To some extent, a specular reflector can also be considered as a coded aperture of a general camera system[10]. 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 array with known mapping whereas in our system the array is randomly distributed.

The idea that some daily objects can accidentally serve as a camera has been explored before. It is in shown in [12] that an photograph of a human eye reflects the environment in front of the eye, and this can be for relighting. In addition, a window or a door can act like a pinhole, in effect imaging the world outside the opening[19].

Chapter 2

The notation and formulation of sparkle vision

2.1 Notation

It is well known that any passive optical system is linear. Suppose the lightfield in a particular environment is denoted as $L^w(p)$ where $p \in \mathbb{R}^3$ is any arbitrary point 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 $L^c(q)$ be the sensor output and $\mathcal{A}(\cdot)$ be the linear mapping relating the lightfield $L^w(p)$ to the sensor output of the camera $L^c(p)$. To digitally represent the system, we discretize it by dividing both the lightfield and the image into blocks and then integrate over all the photons in each block. Let x be the discretized version of the original continuous $L^w(p)$ and y be the discretized version of $L^c(q)$. For ease of notation, we let x and y denote the stacked 1D vector of the light instead of the original 2D or 3D light. Doing this discretization will not hurt the linearity property of the system. Thus, the linear operator $\mathcal{A}$ can be represented with a matrix A so that $y = Ax$.

Note that all the above discussion makes no assumption on any material, albedo, smoothness or continuity properties of the objects in the scene. Therefore, our linear representation of the problem holds for any random specular microfacets we use. In

this notation, the task of *sparkle vision* falls into two categories:

1. Determine the matrix A, which is a calibration task.

2. Infer the light x from the image y, which is an inference task.

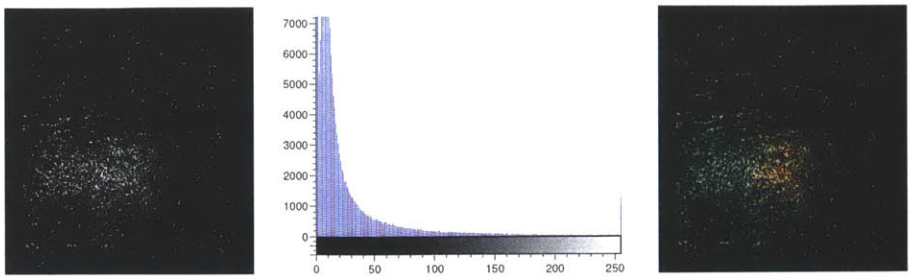
In the later discussion, we will use many pairs of (x, y) so we use the subscript (x_i, y_i) to denote the i-th pair. In addition, let e_i be the i-th unit vector of the identity basis, i.e., a vector whose entries are all zero except the i-th entry which is one. Similarly, let d_i represent the i-th unit vector of the bases of the Discrete Cosine Transform (DCT). Also let $A = [a_1, a_2, \ldots, a_N]$ with a_i as its i-th column.

2.2 Imaging specular random facets through HDR

In this section we examine the properties of sparkling objects with microfacets. Their special characteristics impose unique challenges to accurately capturing images of them. To deal with these challenges, we use High Dynamic Ranging (HDR) imaging, using multiple exposures of the same scene.

Specular random microfacets can be considered as a randomized mapping between the world 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 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 therefore appear bright. The rest of the facets will be unilluminated. This effect will tend to make the dynamic range of a photo of specular facets extremely high, creating a unique challenge to photographing them. Figure 2-1(a) and 2-1(b) plots the histogram of one such object.

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 narrow reflecting direction of each facet, this slight disturbance will cause a huge change of light patterns on the random facets. Provided that the orientations of the



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

Figure 2-1: Optical properties of specular random micro facets: (a) shows an image of specular facet with scattered bright spots. (b) demonstrates its hsitogram, although the bright spots in the image catch the eyes, most of spots are actually dark. (c) the surface simultaneously illuminated by two impulses adjacent impulse lighting, one in red and one in green. The green lights and red lights seldom overlap because few spots in the image are yellow.

facets are truly random across all the surfaces, then we should expect that the set of aligned facets will be significantly different. Figure 2-1(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.

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. To accurately capture them, long exposure is needed. Unfortunately, long exposure makes the bright spots saturated and therefore breaks the linearity assumption. If we adjust the exposure for the sparse bright spots, the exposure time would be too short to capture the dark spots, and we will suffer from noise in the dark regions. Therefore, it is not practical to capture both high and low intensity illumination with just a single shot with a commercial DSLR camera.

Our solution is to use multiple exposure imaging technique HDR [7]. Specifically 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. It is reasonable to assume that the intensities for different pixels are treated independently by the camera hardware. Therefore, we only need to develop a way to decide $I_0(x)$

from $I_k(x)$ for any arbitrary location x.

The Canon Rebel T2i camera that we use in our experiments has roughly linear response with respect to the exposure time for a fairly large range – roughly when the intensity ranges in $(0.1, 0.7)$. When the intensity goes beyond 0.7 the response function becomes curved and gradually saturated and hence the linearity assumption breaks down. When the intensity is lower than 0.1 the image is very noisy, and so we discard these intensities. Denote the remaining exposure time and intensity pairs $(t_i, I_i(r))$. The goal is to determine the value $I(r)$ independently for each location r. We solve this problem by fitting a least squares line to $(t_i, I_i(r))$:

$$I(r) = \text{argmin}_s \sum_i (s \cdot t_i - I_i(r))^2 \tag{2.1}$$

To get the optimal solution, we first compute the derivative of $\sum_i (s \cdot t_i - I_i(r))^2$ with respect to s,

$$\frac{\partial}{\partial r}\left(\sum_i (s \cdot t_i - I_i(r))^2\right) = \sum_i 2\,(s \cdot t_i - I_i(r))\,t_i = 2\sum_i t_i^2 s - 2\sum_i t_i I_i(r) \tag{2.2}$$

Setting the derivative to 0, we derive the closed form optimal solution to Eqn (2.1),

$$I(r) = \frac{\sum_i t_i I_i(r)}{\sum_i t_i^2} \tag{2.3}$$

This is sort of an average of intensities under different exposures weighted by the exposure time. Unlike classic HDR techniques where registration is an issue, we use a heavy tripod and conduct the experiments in a controlled lab environment to prevent even the slightest movement, thus avoiding the registration step.

Chapter 3

Calibration and Inference of Sparkle Vision System

In this chapter, we examine the algorithm to calibrate the system and reconstruct the environmental map x from y.

3.1 Calibration with Impulse Basis

The most straightforward method to determine the matrix A is to probe the system $y = Ax$ by illuminating the object with impulse lights e_i. Let a_i be the columns of A, i.e., $A = [a_1, a_2, \ldots, a_N]$. An impulse basis e_i is a vector of length N with only the i-th entry being 1 and all the other entries being 0. So we have

$$y_i = [a_1, a_2, \ldots, a_N] \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = a_i \tag{3.1}$$

Therefore, to get the whole A, we just have to enumerate all the basis vectors $e_i, 1 \leq i \leq N$, image them to get column vectors $a_i = y_i$, and then organize the a_i into A.

This simple method suffers from the noise present in the imaging system and hence will not perform well when we use it to invert the light map. But at least it provides us a functioning baseline to build upon.

3.2 Calibration with Overcomplete Basis to Reduce Noise

In practice we do not live in the ideal world where noise is not present. Every component of the imaging system is actually introducing noise. The commercial computer screen that we are using cannot emit light with constant luminance. The color shown on the screen can be biased. The camera sensor is not perfectly linear and even the actual shutter speed can be slightly different from the specification. We model all of the noise as zero-mean Gaussian noise and therefore the imaging system becomes $y = Ax + n$.

If we invert the system with a ground-truth transformation matrix A, the small noise n will not affect the quality of the recovered x too much. Analytically, let $\kappa(A)$ be the condition number of A, x' be the recovery. The error between relative error between x and x' is $A^{-1}n$. So the relative error can be bounded by

$$\frac{\|A^{-1}n\|/\|x\|}{\|n\|/\|y\|} \leq \kappa A \tag{3.2}$$

As we will show later, the system we work with in general has decent condition number, typically around 10. It is fair to assume that n is small compared with the intensity of the image and therefore the recovery should not be affected seriously.

However, we have no access to the ground-truth A. The matrix A we have is calibrated through the output of the imaging system when showing basis on the screen. Note that such images are exposed to noise, thus A is also noisy. As we will show in the experimental section, the noise in A causes way more trouble than the noise in y. To improve the quality of the reconstruction, we must reduce the noise in A. This can be done by display more than one set of basis vectors, i.e., overcomplete

basis on the screen, and try to learn a condensed dictionary from the images of them.

Therefore, in addition to the impulse basis e_i, we further probe the system with multiple different bases, for example the DCT basis d_i and the random basis 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 , $D \in \mathbb{R}^{N \times N}$ be the DCT basis matrix and $B_K \in \mathbb{R}^{N \times K}$ be the matrix of K random basis. This implies the following optimization to do the calibration:

$$\min_A \|Y_1 - AE\|_F^2 + \lambda \|Y_2 - AD\|_F^2 + \lambda \|Y_3 - AB_K\|_F^2 \tag{3.3}$$

λ 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 balance the two.

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 threshold during calibration. Specifically let Ω_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 Ω and throw the rest away. Let $P_\Omega(\cdot)$ represent such a projection. This turns the calibration into the following optimization:

$$\min_A \|P_\Omega(Y_1) - AE\|_F^2 + \lambda \|P_\Omega(Y_2) - AD\|_F^2 + \lambda \|P_\Omega(Y_3) - AB_K\|_F^2 \tag{3.4}$$

Note that the size of the output A from (3.5) is different from (3.3) due to the projection $\Omega(\cdot)$.

To solve the optimization in (3.5), we first rewrite it in matrix form

$$\min_A \left\| \begin{bmatrix} P_\Omega(Y_1) \\ \frac{1}{\sqrt{\lambda}} P_\Omega(Y_2) \\ \frac{1}{\sqrt{\lambda}} P_\Omega(Y_3) \end{bmatrix} - A \begin{bmatrix} E \\ \frac{1}{\sqrt{\lambda}} D \\ \frac{1}{\sqrt{\lambda}} B_k \end{bmatrix} \right\|_F^2 \tag{3.5}$$

With a substitute of variables, we can write the objective function above as $\|Y -$

$AX\|_F^2$, whose derivative is $-2(Y - AX)X^T$. Setting it to zero we get the solution $A = (XX^T)^{-1}YX^T$.

3.3 Reconstruction

Given A, reconstructing an environment map 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 try to solve a constrained least squares problem:

$$\min_x \|y - Ax\|_F^2, \quad s.t. \quad x \geq 0 \tag{3.6}$$

There exist several efficient solver to this problem, like the function *lsqnonneg* implemented in MATLAB. However, through experiments we find they are actually too slow for application. When the resolution of the screen is 20×20, i.e., $x \in \mathbb{R}^{400}$, solving the inequality constrained least square is approximately 100 times slower. Yet the improvement is minor. So we just solve the naive least square without non-negative constraints and then crop the out-ranged pixels back to $[0, 1]$.

In addition, we observe that there is room to further enhance the results by smoothing 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 sparkle vision, and hence we decide to stick to the most naive optimization (3.6).

3.4 Extensions and implementation details

In this section, we examine several techniques that we utilize to help the implementation of the sparkle vision system physically in practice. Also, some design choices in the actual implementation are discussed and elaborated in details.

Using raw from the camera We use RAW files from the camera. Any post-processing in other conventional image formats like JPEG and PNG is a non-linear transformation of the image and hence violate the linearity assumption of the system. In our experiments we observe that such non-linearity significantly hurts the performance. Therefore we avoid this issue by just using the original RAW files.

Subtraction of the leaking light In our experiments, no matter how carefully we control the experimental setup, it is inevitable that some ambient light will be present in the scene beside the basis lights we use. For example, when we use the screen to show the environmental light map the dark pixels are not perfectly dark. This problem is more serious when the impulse light we use to probe the system becomes small at a high resolution. To handle this, we assume that this leaking light is static across the whole timespan when the images are taken. Therefore we can model it as a constant vector e_0. Let $y_0 = Ae_0$ be the image of this background light. By default any image y_i we use is already the post-processed version after subtracting y_0 off.

Repeatedly taking image of one scene For some critical images like the image of the leaking light y_0, which is subtracted off all the other images, we want to suppress the noise level as much as possible. Since the camera noise is largely Poisson, we can reduce the noise level of it by repeatedly photographing the same scene and averaging the photos.

Handling Color Images The discussion above only applies to the gray-scale images. However, handling color images is no different in principle since the mapping from the colored environmental light to the colored images is still linear. The only difference is the size of the mapping. Suppose that the system maps a gray-scale lightfield with N blocks to a gray-scale image with M pixels. Then the transformation matrix of the system A is $\mathbb{R}^{M\times N}$. But for the colored system the matrix would be $\mathbb{R}^{3M\times 3N}$. Consequently, we need to stack all the three channels of either x or y into a vector. Furthermore, to probe this colored system, we need to use $3N$ impulses,

i.e., red, blue and green impulses for each block of the lightfield. Other than these adjustments, our system works in the same way for color images as for gray-scale images.

Chapter 4

Simulated Analysis of Sparkle Vision

The physical characteristics of the system have huge impact on the intrinsic invertibility of the optical system in sparkle vision. The position and the size of the screen with respect to the object determines the amount of the light that the sparkling surface can reflect. The amount of the glitter on the object surface changes the probability that a certain light ray from the screen can hit the camera. The pose of the camera constraints the light that can potentially come in from the screen. Besides them, a lot more other factors can potentially affect the performance of the system.

In this chapter, we conduct simulated experiment to study how those factors impact the invertibility of the system. Understanding them helps us interpret the experimental results from real data, and provides us guideline to tune the system to get better performance.

4.1 Setup of the Simulated Experiment

We consider the setup where we have a simulated screen showing the pictures in front of a sparkling surface. The microfacets on the surface randomly reflects the light to a virtual pinhole camera. The sensor plane of the camera captures the incoming lights and forms a photo. In order to render the image captured by the camera, we need to simulate the physical process that transforms the lightmap shown on the screen to the pinhole camera.

First let us consider the screen. We uniformly divide a rectangular screen with size $H \times W$ into $M_1 \times M_2$ pixels, thus the size of each pixel is $\frac{H}{M1} \times \frac{W}{M2}$. Without loss of generality, we assume that a point on this screen is uniformly emiting light in all directions. For simplicity we assume that there is no noise in the light from the screen, and we do not consider any attenuation effect.

As with the object, we start with the simplest a planar surface filled with glitters. Note that in practice the object surface has infinite resolution. Moreover, since micro mirrors are randomly distributed and oriented, the light reflected by one mirror might need to go through several inter reflection before reaching the camera. In the worst case, one micro mirror may be completely occluded by another. Noticing that those interreflectance and occlusion factors do not change the linearity of the system, we skip them and model the specular facet in the following simplified way:

- The planar glitter is uniformly $K_1 \times K_2$ rectangle blocks.
- At each block we place a mirror facing a random orientation to cover the block.
- We ignore the depth variation of the mirror caused by its orientation. This is an approximation of a small randomly oriented mirror when no interreflection and occlusion is considered.
- The light reflected by one mirror cannot be blocked by other mirrors.

For each mirror, we assume that its orientation follows a distribution. Let $\theta \in [0, \pi/2]$ be the slant of the orientation and $\phi \in [0, 2\pi)$ be the its tilt. Then the normal of the mirror in the coordinate of the surface is $(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$. We model the tilt ϕ as uniformly distributed in $[0, 2\pi)$. Ideally, we want the slant to be also uniform. However, this does not match the specular surface in practice, which is more likely leaning towards 0. Therefore we model it as positive half of the Gaussian distribution with standard deviation σ_θ. Specifically, we have

$$\mathrm{P}_\theta(\theta_0) = \frac{2}{\sqrt{2\pi}\sigma_\theta} \exp\left(-\frac{\theta_0^2}{2\sigma_\theta^2}\right) \tag{4.1}$$

Note that the mean of θ is actually not 0 and hence the actual standar deviation is not σ_θ.

Therefore, the joint probability for (θ, ϕ) is

$$\mathrm{P}_{\theta,\phi}(\theta_0, \phi_0) = \frac{1}{\sqrt{2\pi}\sigma_\theta \pi} \exp\left(-\frac{\theta_0^2}{2\sigma_\theta^2}\right) \tag{4.2}$$

We further assume that the orientation of the different mirrors on the glittering surface is independently sampled. In this way, we create a simplified reflector that respects and the spirit of sparkle vision system.

For the rendering, we use the ray tracing algorithm to determine the amount of light received by a pixel on the sensor of the pinhole camera. Specifically we apply the following steps to determine the intensity at x_i.

Step 1 We intersect the light-ray $\vec{x_i o}$ with the object. If there is no interaction, just return black.

Step 2 Suppose the intersection is u_i with normal n_i. Judge if $\langle o\vec{x}_i, n_i \rangle > 0$. If not, return black.

Step 3 Compute the reflected light and intersect it with the screen. If there is no intersection, return black.

Step 4 Output the pixel intensity from the screen.

Based on this model, we can conduct simulated experiments to test the performance of the calibration and the reconstruction on physical setup with different parameters.

4.2 Simulated Experiment Results

We first conduct the experiments in the noise free case and show that we can get success. Specifically, we place a screen of size 100×100 on the plane $z = 0$. The top left corner of the screen is placed at the origin $(0, 0, 0)$ and the edges of the screen are

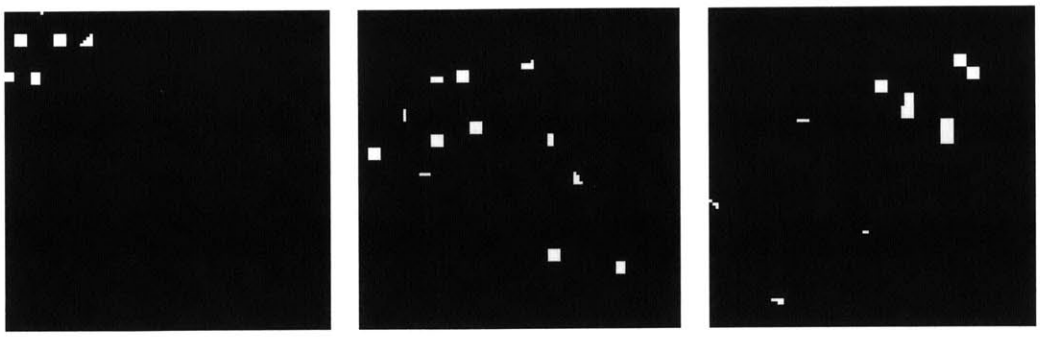
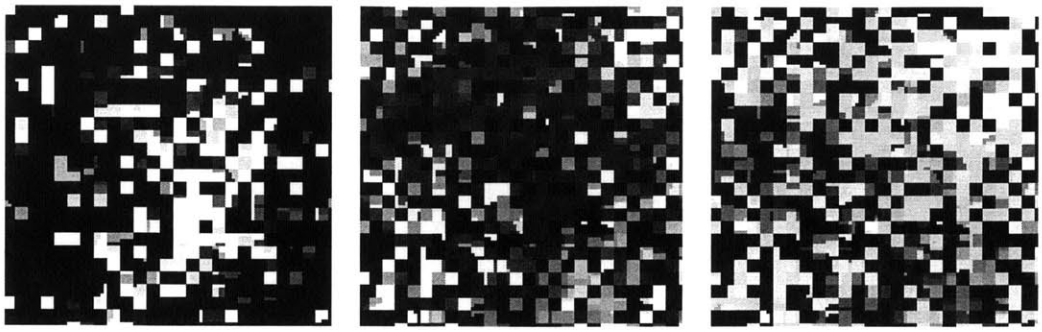


Figure 4-1: Synthetic photographs of specular objects under impulse light

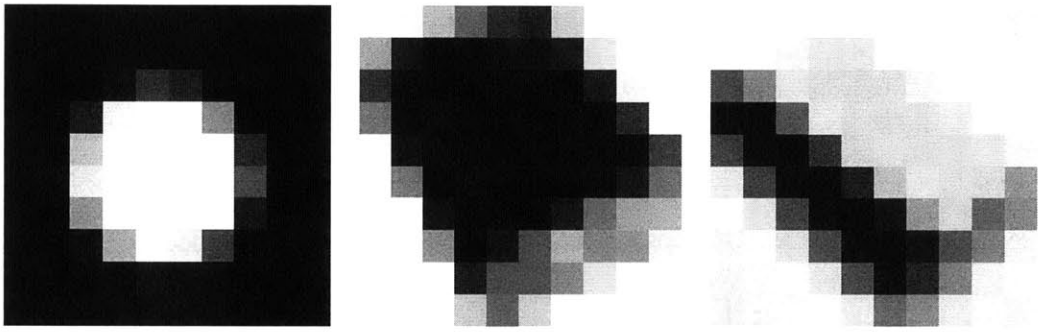
axis aligned. Then on the same scene, we place a rectangular planar random reflector also with size 100×100 on the plane $z = 25$. Similarly we set its top left corner to $(0, 0, 25)$ and let it aligned with the axis. The virtual camera is located right at the center of the screen, i.e., $(50, 50, 0)$ and it is looking towards the direction of the positive z-axis. The field of the view of the camera is chosen as $\pi/2$. The screen resolution is set to be 10×10. The size of each micro facet on the reflector is 2×2. Let the resolution of the camera be 20×20. Finally, let the standard deviation of the slant be $\frac{\pi}{6}$.

Figure 4-1 shows a few sample photographs of the reflector illuminated by different impulse light from the screen in the calibration stage. On each image, there are a few bright spots meaning that the corresponding micro facets are reflecting the impulse light from the screen to the camera sensor. Outside them most of the pixels are dark. These characteristics are fairly consistent with the real photographs of the sparkling surfaces. This validates the behavior of the synthetic model we develop.

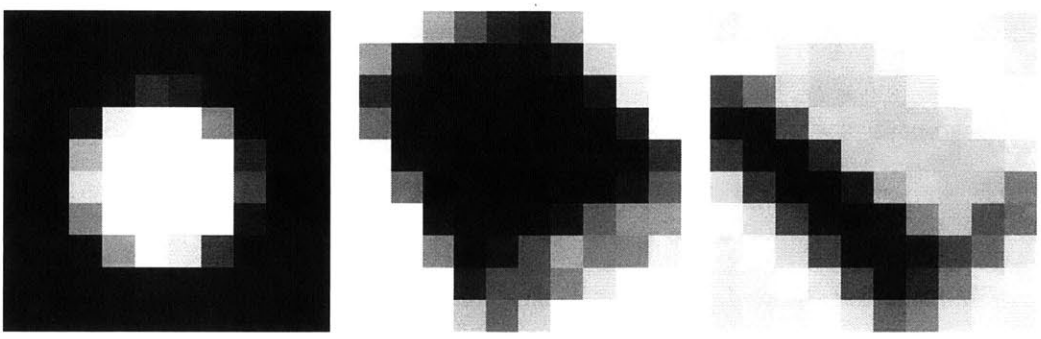
Figure 4-2 shows the test light maps on the screen, the images of the sparkling surface when reflecting those illuminations, and the reconstructed lights using the algorithm we propose. As the figure suggests, the recovery is perfect. Honestly, this success is largely due to the ideal setup and the lack of noise. But this at least gives us confidence that under well controlled conditions sparkle vision is feasible and the developed algorithm functions as is designed.



(a) Test photo 1 (b) Test photo 2 (c) Test photo 3



(d) Ground truth light map 1 (e) Ground truth light map 2 (f) Ground truth light map 3



(g) Recovered light map 1 (h) Recovered light map 2 (i) Recovered light map 3

Figure 4-2: Test light maps, photographs and results. The first row shows the images of the specular random surface illuminated by the test light. The second row presents the ground-truth test light on the screen. The third row shows the recovery by our algorithm.

4.3 Sensitivity to Noise

As we have argued before, the noise present in the optical setup can significantly affect the performance of the algorithm. In this section, we verify this through synthetic experiments following the set up in Section 4.2. The only difference is that we add

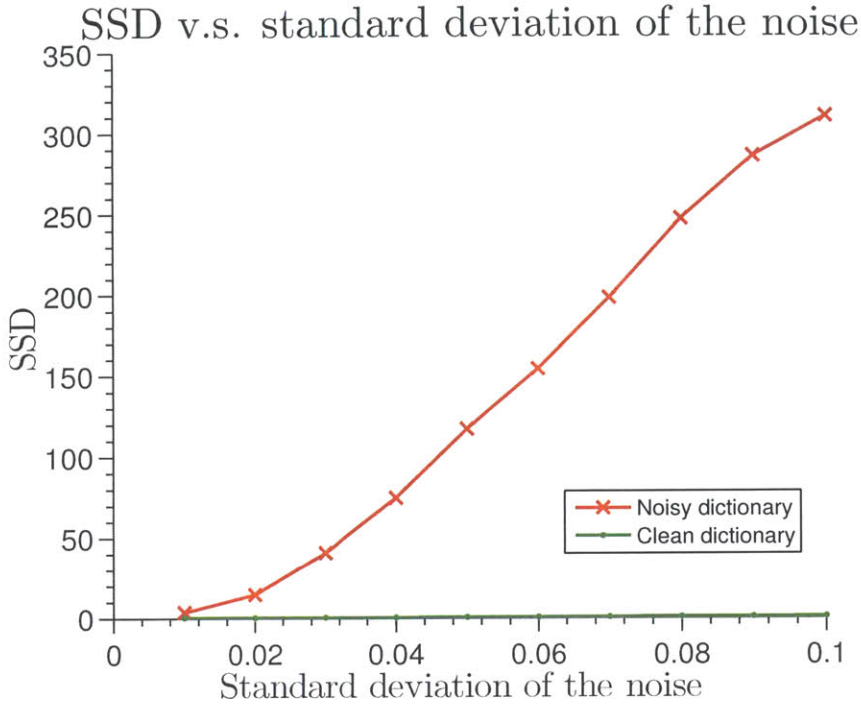


Figure 4-3: The stability of our algorithm to noise added to the calibration stage and the test stage is different. The range of the image is in $[0, 1]$.

noise to some of the photographs taken by the virtual camera.

In reality the noise level is dependent on the intensity of the pixel. However, to be simple, we assume that the noise added in our synthetic experiment is independent of the illumination. Therefore we model the noise as white gaussian with zero mean. Suppose we have N test images $I_i, 1 \leq i \leq N$ and the recovered images are $\widehat{I}_i$. Then the error of the recovery can be measured by the average sum of squared difference (SSD) between I_i and $\widehat{I}_i$, which is defined as

$$\text{SSD}\,(A, B) = \sum_{i,j} (A(i,j) - B(i,j))^2 \tag{4.3}$$

It is worth noting that the noise in the training images during the calibration stage will have different impact to the recovery compared with the noise present in the test image. To demonstrate it, we conduct two groups of experiments. In the first group, we add noise in both the calibration and the testing stage while in the second group

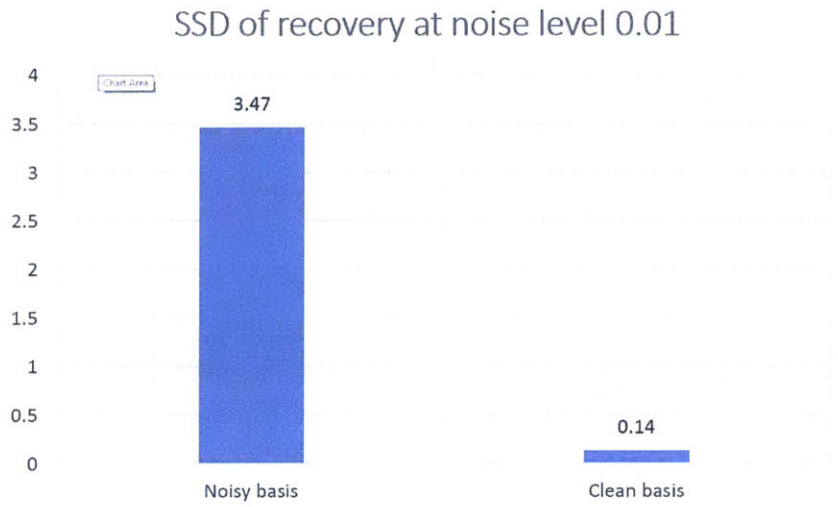
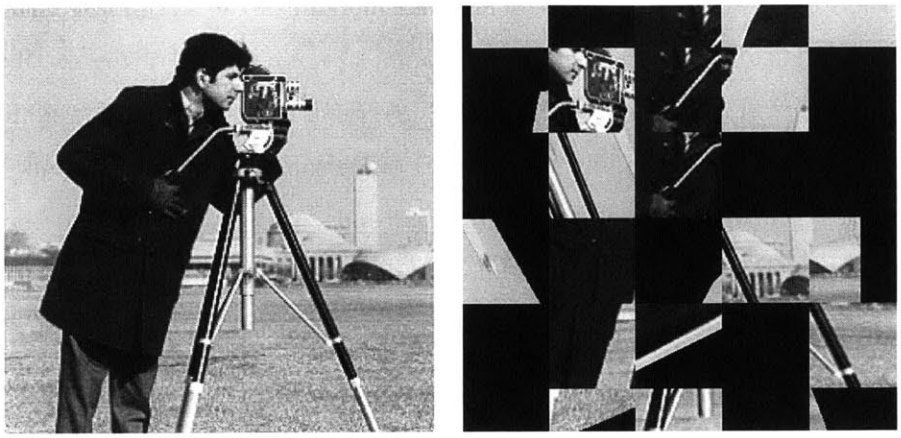


Figure 4-4: The error of the recovery at noise level $\sigma = 0.01$.

we only contaminate the testing images. Also, our system uses over-complete basis to reduce the noise in the dictionary. To expose the impact of the noise in the dictionary, we only use the impulse basis and skip the other basis, just for this section. Varying the noise standard deviation from 0.01 to 0.1, we get two curves shown in Figure 4-3.

As we can see, our system is much more robust to the noise in the test image than the noise in the images for calibration. In fact, when the standard deviation of the noise is 0.1 the recovery is still great with the clean dictionary. On the contrary, at the same noise level, the SSD on the recovery using polluted basis is three orders of magnitude higher. Even at negligible noise with standard deviation 0.01 the error with polluted basis is considerably higher than with clean basis, as is illustrated in Figure 4-4.

All these synthetic results suggest that keeping the dictionary clean is the key to the success of recovery. Thus it validates the need to use over-complete basis in our proposed pipeline.



(a) Image shown on the screen (b) Photography of a specular facet

Figure 4-5: Photography of a specular reflector with low spatial resolution.

4.4 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 micro facet 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 from 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.

Figure 4-5 demonstrates such phenomenon. Figure 4-7(d) shows a high-resolution image shown on the screen serving as the light map. The lightings are reflected by the micro facets to the camera sensor. However, the size of each micro facet is too large and hence the number of the facets is limited. As a consequence, we observe from Figure 4-5(b) that some blocks of the photo is dark and part of the light map is missing. So the spatial resolution of the specular surface can affect the invertibility of the system to a significant extend.

In this section, we develop a mathematical model to explicitly analyze the probability that a block of pixels on the screen will be reflected by at least one micro facet to the camera sensor. Based on the model, we will discuss how the spatial resolution of the reflector and the screen will affect the invertibility of the optical system.

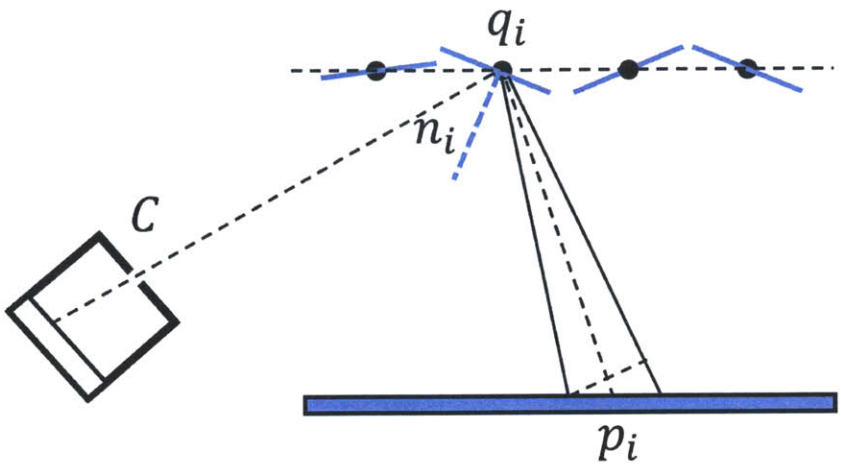


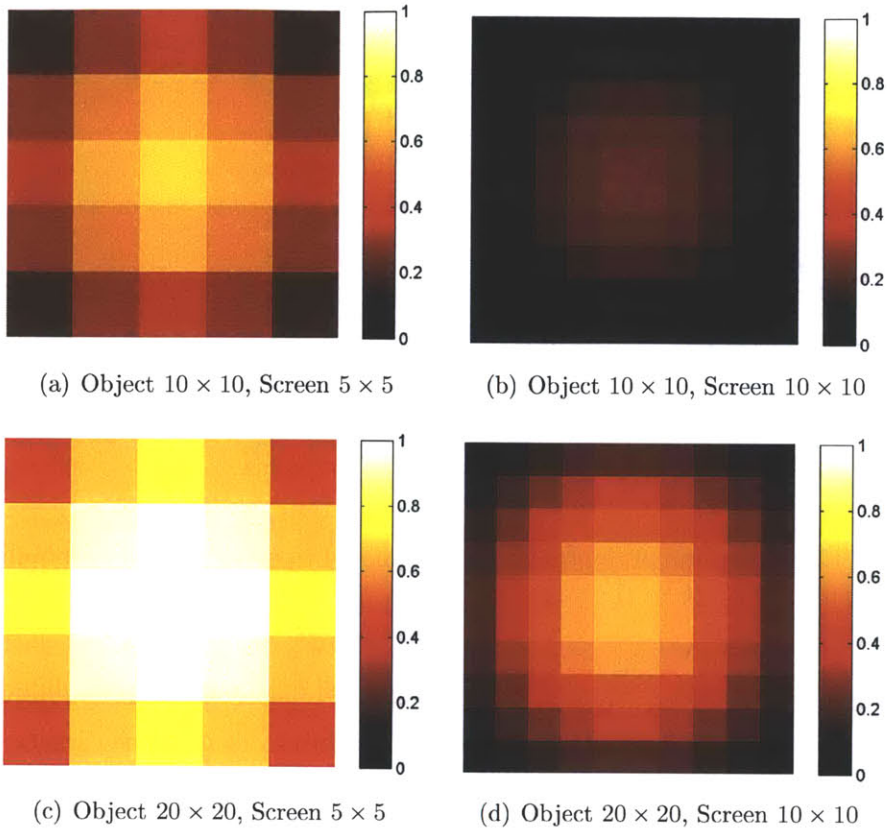
Figure 4-6: Configuration of the mathematical model to compute the probability that p_i will be reflected by q_i to the camera.

Following the general setup in this section, we first calculate the probability that a certain pixel p_i on the screen gets reflected by the micro facet q_i to the camera. Figure 4-6 elaborates the configuration visually. It is worth noting that we use small angle approximation throughout the modeling process. Suppose the width of the pixel is w, then the foreshortened area of the pixel with respect to the incoming light direction $\overrightarrow{p_iq_i}$ is $w^2 \cos\theta$, where θ is the angle between $\overrightarrow{q_ip_i}$ and the screen. Then the solid angle of this foreshortened area with respect to q_i is $\frac{w^2 \cos\theta}{\|\overrightarrow{p_iq_i}\|}$.

The normal n that just reflects $\overrightarrow{p_iq_i}$ to the camera C can be computed as the normalized bisector of $\overrightarrow{q_ip_i}$ and $\overrightarrow{q_iC}$. Since the incoming lights can vary in the solid angle of $\frac{w^2 \cos\theta}{\|\overrightarrow{p_iq_i}\|}$, n can vary in $\frac{w^2 \cos\theta}{4\|\overrightarrow{p_iq_i}\|}$ 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

$$\Pr(q_i \circ p_i) = \frac{2}{\sqrt{2\pi}\sigma_\theta} \exp\left(-\frac{\theta_0^2}{2\sigma_\theta^2}\right) \frac{w^2 \cos\theta}{4\|\overrightarrow{p_iq_i}\|} \tag{4.4}$$

Suppose there are M micro facets in total and we compute $\Pr(q_i \circ p_i)$ for all



(a) Object 10×10, Screen 5×5

(b) Object 10×10, Screen 10×10

(c) Object 20×20, Screen 5×5

(d) Object 20×20, Screen 10×10

Figure 4-7: Probability map of light from a block pixels getting reflected by the specular surface to the screen.

$1 \leq i \leq M$. Then we can compute the probability that the light from pixel q_i is reflected by at least one micro facet to the camera as follows.

$$\begin{aligned} \Pr\left(\exists j, q_j \circ p_i\right) &= 1 - \Pr\left(\forall j, q_j \not\circ p_i\right) = 1 - \prod_j \Pr\left(q_j \not\circ p_i\right) \\ &= 1 - \prod_j \left(1 - \Pr\left(q_j \circ p_i\right)\right) \end{aligned} \tag{4.5}$$

Then we can make an image $T(x, y)$ with the same resolution as the screen. The intensity of each pixel of $T(x, y)$ represents $\Pr\left(\exists j, q_j \circ p_i\right)$ for the corresponding pixel p_i. Figure 4-7 shows two such probability map when the resolution of the object is 10×10 / 20×20 and the resolution of the screen is 5×5 / 10×10.

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 distance and relative angle between different parts of the screen and the reflector. It suggests that for a specular object there will be a limit on the resolution of the light map we can infer from it.

Chapter 5

Experiments and Results

In this section, we conduct experiments to prove the effectiveness of sparkle vision under controlled conditions in a lab. First of all, we verify that the assumptions we make for specular random facets are actually satisfied in real life. Then we proceed to present the results of reconstructing different environmental maps using different specular objects with different algorithms. Most of the experiments are conducted with gray-scale environment maps. Nevertheless we present a few examples in color light field to demonstrate the capability of our system to handle them. Finally, we study the robustness of the system to noise and misalignment. section

5.1 Experiment setup

We need the ground-truth environmental map either to calibrate the system or to evaluate the quality of the reconstruction of the system. To achieve this, we place the sparkling object in front of the computer screen in a dark room and use a camera to photograph the object. The screen displays images that are used for both calibration and testing. The light from the screen illuminates the sparkling object. The task then is to reconstruct the displayed image on the screen from the object. Figure 5-1(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 band, skull, and painted

glitter board. The whole system is located in a dark room. The camera is placed on a heavy tripod to prevent even the slightest movement. To demonstrate the concept for most objects we only display the gray-scale image at resolution 20×20 pixels. However, we also show that the system can reconstruct the world lighting with higher resolution, showing results up to 30×30 pixels. At this resolution many objects, such as faces, can be easily recognized.

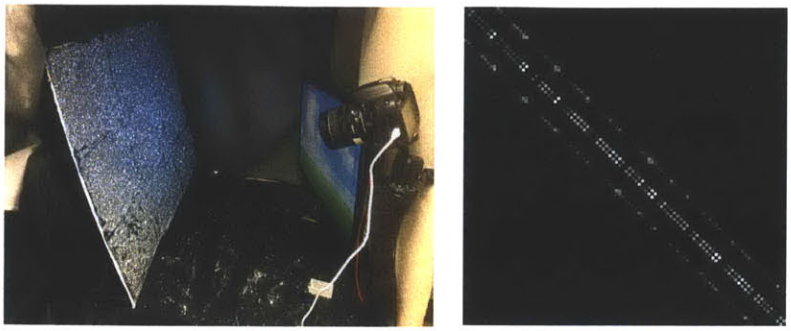
5.2 Examine the assumption of the system

Overlapping of bright pixels An important factor to the success of sparkle vision is that the location bright spots of an image changes drastically when the location of impulse shown on the screen changes slightly. Let a_i be the image of the object illuminated by the i-th impulse light. Let S_i be the set of bright pixels with intensities larger than $1/10$ of the brightest pixel in a_i. Then the overlap between a_i and a_j, $i \neq j$ can be defined as

$$\mathrm{O}(i,j) = \frac{|S_i \cap S_j|}{\min(|S_i|, |S_j|)} \tag{5.1}$$

Here $|S|$ denotes the number of elements in a set S. At resolution 10×10 pixels, there are 100 basis images, and the overlap between each of them can be plotted in a 100×100 figure where the entry at i-th row and j-th column representing $\mathrm{O}(i,j)$. Note that the diagonal of this plot should be 1 and dwarfs all the rest off-diagonal entries. Therefore we set those entries to 0 and thus emphasize only the overlap between different images.

It can be seen from Figure 5-1(b) that most of the overlap happens between images from neighboring impulses. For neighboring impulses, the maximal overlap $\max_{i \neq j} \mathrm{O}(i,j) < 0.2$, hence the overlap of bright spots even between adjacent impulses illuminated images is small. This implies that if what is shown on the screen is just a collection of bright dots, and the task is to only guess whether or not each dot is on or off, this system should have very good performance.



(a) Experiment Setup (b) Overlap graph

Figure 5-1: The left (a) shows the setup of the experiment in the lab. The right (b) shows the overlap graph between images from different impulses. It can be seen from the figure that only images from neighboring impulse have slight overlap.

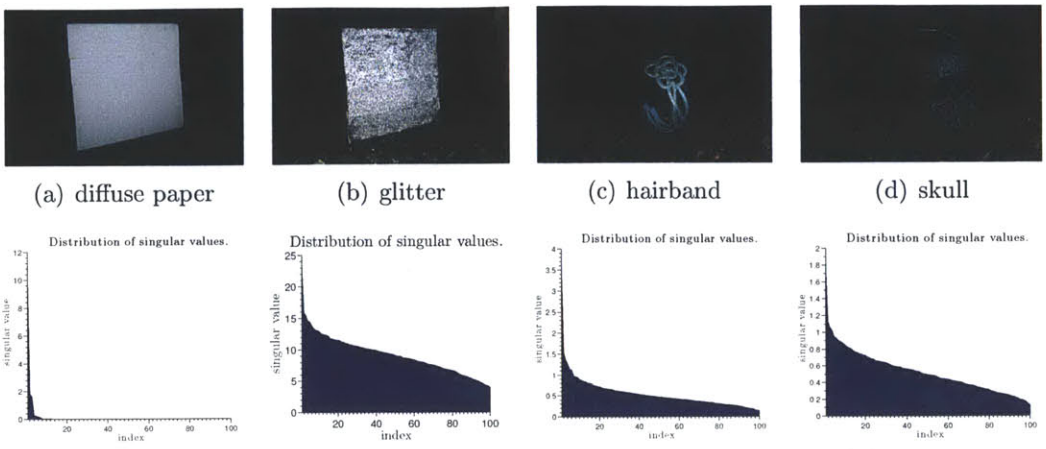
Condition Number To analyze the capabilities of our system, we measure the condition number of the transformation matrix A, $\kappa(A)$, for a piece of diffuse paper and for many different sparkling objects. In general the condition number $\kappa(A)$ of the matrix A measures the stability of the system respect to the noise. Systems with smaller condition number are more robust to noise. Mathematically $\kappa(A)$ is defined as

$$\kappa(A) = \frac{\sigma_{max}}{\sigma_{min}} \tag{5.2}$$

where σ_{max} and σ_{min} are the smallest singular values of the matrix A. For all the optical systems shown in Figure 5-2, 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 pretty good.

5.3 Experimental result

Here we show results of the glitters facing different images shown on the screen. For the gray-scale setting, we push the resolution of the screen to 30×30. For the colored setting, we just show presents a few test at a lower resolution 15×15 to demonstrate that our system naturally generalizes to color ones. The gray scale results are shown in Figure 5-3 and the color ones are shown in Figure 5-4.



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

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

Figure 5-2: Singular value distribution of the transformation matrix of different objects. Also the condition number is given. Note that the sparkling reflectors create systems with much lower conditional number compared with a diffuse object.

From the results one can easily perceive what the faces and the objects are. This demonstrates the success of of our sparkle vision system.

5.4 Stability to Noise

As we have discussed before the system is substantially stable to noise thanks to the good condition number. To demonstrate that, we perform synthetic experiments by adding noise to the test image and compute the Root-mean-squared-error (RMSE) between the noisy recovery and the non-noisy recovery. The reason that we do not capture real noisy data is that we rely on clean real data to calibrate the system well. We plot hwow the reconstructed lighting change as the noise level increases in in Figure 5-5.

From the results, we conclude that our proposed method is indeed stable to noise.

5.5 Instability to misalignement

The success of sparkle vision relies largely the sensitivity of light pattern on a specular object to even a slight movement of the source light. However, this property

Figure 5-3: Sparkle vision through glitter board: Row 1 and 4 present the sensor output. Row 2 and 5 present the reconstruction, and Row 3 and 6 present the result of "sparkle vision"

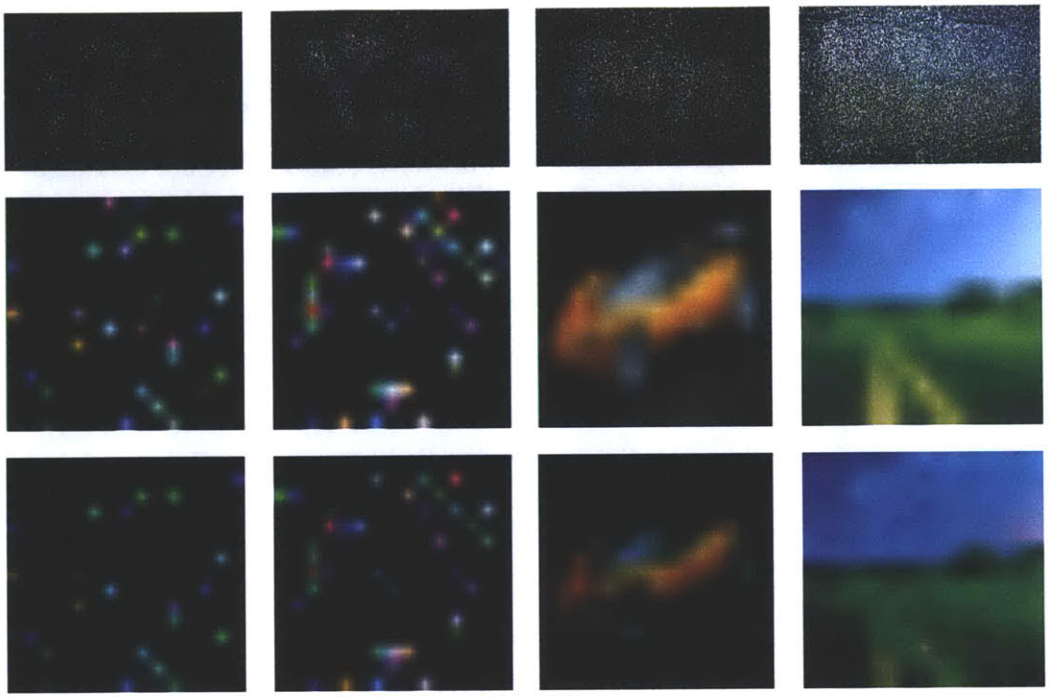
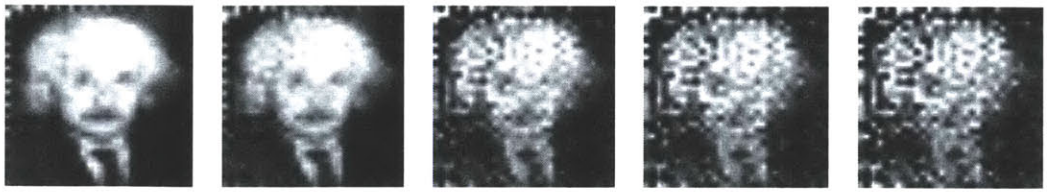


Figure 5-4: Colored sparkle vision through glitter board: The first row presents the images of sparkling surfaces. The second row presents the ground-truth light map shown on the screen. The third row shows the reconstruction of sparkle-vision. The test images in the first two columns are synthetic scatter dots with random colors. The rest are real images from Internet. Although there is slight color distortion in the reconstruction, overall the inferred light is very close to the ground-truth.

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

Figure 5-5: Stability to noise: title of the subfigure represents noise level. They are pretty large considering the images are in $[0, 1]$ and only a few spots are bright.



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

Figure 5-6: Instability to misalignment: even if we shift the test image by one pixel horizontally, there is significant degrade in the output.

simultaneously make the whole system extremely sensitive to subtle misalignment. To show this we perform synthetic experiments by shifting the test image I by δx and examine the RMSE. Some representative results and the RMSE curve are shown in Figure 5-6.

Chapter 6

Discussions and Conclusion

In this paper we show 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 the objects with Lambertian surfaces and specular metals, which researchers in the field of shape-from-X have worked on. The experiments verify that the information in the random facets is enough to reconstruct the environmental map at reasonable resolution. In particular, from the reconstructed images in our experiments, it is generally easy to recognize the object that was displayed on the screen.

The contribution of the paper is 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 surprising at first sight but indeed is intuitive and simple once we understand it. Second, we proposed a system and algorithm that can efficiently perform the calibration and inference necessary to reconstruct images of the world from the reflections on a sparkling surface.

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 an identical 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. We leave this as an exciting direction for future work.

Bibliography

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

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

[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.

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

[5] Guillermo D. Canas, Yuriy Vasilyev, Yair Adato, Todd Zickler, Steven Gortler, and Ohad Ben-Shahar. A linear formulation of shape from specular flow. In *12th IEEE International Conference on Computer Vision (ICCV 2009)*.

[6] E. J. Candes and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51:4203–4215, 2004.

[7] Paul E Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH 2008 classes*, page 31. ACM, 2008.

[8] Rob Ferbus, Antonio Torralba, and William T. Freeman. Random lens imaging. *MIT CSAIL Tecnical Report*, (058), 2006.

[9] Samuel W. Hasinoff, Anat Levin, Philip R. Goode, and William T. Freeman. Diffuse reflectance imaging with astronomical applications. In *13th IEEE International Conference on Computer Vision (ICCV 2011)*.

[10] Anat Levin, Rob Fergus, Frédo Durand, and William T Freeman. Image and depth from a conventional camera with a coded aperture. *ACM Transactions on Graphics (TOG)*, 26(3):70, 2007.

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

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

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

[14] 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.

[15] R. Ramamoorthi and P. Hanrahan. A signal-processing framework for reflection. *ACM Transaction on Graphics*, 2:104 – 1042, 2004.

[16] 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)*.

[17] 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.

[18] 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.

[19] Antonio Torralba and William 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)*.

[20] Yuriy Vasilyev, Todd Zickler, Steven Gortler, and Ohad Ben-Shahar. Shape from specular flow: Is one flow enough? In *24th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2011)*.

[21] Yuqian Zhang, Cun Mu, Han-Wen Kuo, and John Wright. Toward guaranteed illumination models for non-convex objects. In *14th International Conference on Computer Vision (ICCV 2013)*.