Approximate svBRDF Capture From Uncalibrated Mobile Phone Video

by

Rachel Anastasia Albert

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Vision Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor James F. O'Brien, Chair
Professor Jitendra Malik
Professor Alexei Efros

Spring 2018

Approximate svBRDF Capture From Uncalibrated Mobile Phone Video

Abstract

Approximate svBRDF Capture From Uncalibrated Mobile Phone Video

by

Rachel Anastasia Albert

Doctor of Philosophy in Vision Science

University of California, Berkeley

Professor James F. O'Brien, Chair

I describe a new technique for obtaining a spatially varying BRDF (svBRDF) of a flat object using printed fiducial markers and a cell phone capable of continuous flash video. My homography-based video frame alignment method does not require the fiducial markers to be visible in every frame, thereby enabling me to capture larger areas at a closer distance and higher resolution than in previous work. Clusters of pixels in the resulting panorama that correspond to like materials are fit with a BRDF based on a recursive subdivision algorithm, utilizing all the light and view positions obtained from the video. I demonstrate the versatility of this method by capturing a variety of materials with both one- and two-camera input streams and rendering my results on 3D objects under complex illumination.

To my husband Chase,
who never let me give up — no matter what.

Contents

List of Figures

List of Tables

Acknowledgments

It is my firm belief that completing a PhD is one part knowledge acquisition and three parts perseverance. As such, this document would not be complete without acknowledging the many wonderful people who supported me, challenged me, and encouraged me to persevere during the past six years.

Through every confusing detour, monotonous slog, painful failure, and exhilarating success, my husband Chase has been there to cheer me on. I am forever grateful for the many hours he spent teaching me computer science, patiently helping me debug my code, and enthusiastically listening to me drone on and on about every detail of my work. Chase never wavered in his support for me completing my degree, and I couldn't have done it without him.

Certainly more than half the hours I worked to earn this degree were spent in a coffee shop next to Sahar Yousef. Through pomodoros and lists of lists, we wrung every ounce of productivity from every minute we spent together. Sahar always had an answer for my dilemmas and a system to solve my problems. She continues to inspire me with her fearlessness and incredible work ethic, and I'm so excited that we can now call each other "doctor".

I must also express my gratitude for the love and support of my family, who helped me remember that what I was endeavoring to do was truly monumental. My parents, Tim and Carlene Brown, kept me grounded and reminded me that the PhD is a marathon, not a sprint. My sister Elisabeth brightened my life with her beautiful family, Josh, Asher, Arabella, and Amos Reeve. My Aunt and Uncle, Jeff & Cindy Brown, frequently provided a home away from home for us. My in-laws, Casey and Ellen Albert, helped me to think creatively and cheered me on, especially when I felt stuck. My brother-in-law, Lee Albert, and his girlfriend Maggie Spear were also cheerful companions over the years. And finally, Grandma Mary Moulthrop inspired me to be strong but also stop and smell the roses. Her husband Bob is no longer with us, but I know he would be incredibly proud of me.

I would like to thank my advisor, Professor James O'Brien, whose sharp wit and intelligent critiques pushed me to do my very best. I was lucky to also have Professors Jitendra Malik and Alexei Efros on my committee, who were continually supportive and provided thoughtful feedback during the long tail of my dissertation. I am especially thankful to Alyosha for welcoming me to his lab meetings, which not only taught me quite a lot about computer vision but also provided me a strong academic community.

Several vision researchers had an important impact on my journey. Most importantly, I would like to express my deep gratitude to Professor Dennis Levi, who brought me into his lab and believed in my research potential when I was still very young. Along the way I had the great fortune to collaborate with many other talented researchers, including Michael Silver, Marty Banks, Greg Ward, Dan Goldman, Rahul Narain, and Marina Zannoli, each of whom taught me important lessons about research and about life.

Although I knew very few people in the Bay Area when I arrived, my support network has grown tremendously. Many thanks to all the folks who walked this path with me:

Hannah Naughton, Christina Gambacorta, Amy Pavel, Vasha Dutell, Jihyun "Jiji" Kim, Janos Botyanski, Arthur O'Dwyer, Samantha Keat, Matt Drescher, and Tenaya Rodewald, along with many others.

Finally, I would be remiss not to include my wonderful new colleagues at NVIDIA, who enthusiastically welcomed me to their group and waited patiently for me to finish my degree. In particular, I would like to thank Dave Luebke, Joohwan Kim, Ward Lopes, Kaan Aksit, Anjul Patney, Pete Shirley, and Turner Whitted for their advice and support.

Marcel Proust once wrote, "Let us be grateful to the people who make us happy; they are the charming gardeners who make our souls blossom." As I move on to the next chapter in my life, I am so incredibly grateful for all the gardeners who patiently nurtured me and helped me grow into the person I am today.

Chapter 1

Introduction

Artistic expressiveness in the creation of virtual objects has increased tremendously in recent years thanks to research in computer generated geometry, lighting, and materials. However, many real-world surfaces exhibit irregular variations in texture and reflectance that are difficult to reproduce algorithmically. Examples include organic materials such as specific pieces of wood or granite, hand-made surfaces such as paintings, and well-worn objects with particular patterns of dirt, scratches, and aging effects. High-quality results can be achieved when these missing details are filled in manually by artists, but doing so requires significant expertise, well-sourced input images, and hours of manual adjustment. Alternatively, it is also possible to obtain high-quality materials through direct capture, but the capture process is cumbersome due to the specialized equipment typically required.

There are several ways to represent opaque surface reflectance using data derived from the real world. The most common examples are artist-designed materials, direct measurements of real objects, and parametric reflectance models. Artist-designed materials are represented by a set of layers derived from images, wherein each layer describes a component of the reflectance such as the diffuse color, specular behavior, or normal displacement. The process for creating these materials typically involves sourcing a high-quality photograph of a nearly-flat object, and then recombining filtered versions of the photo with procedurally-generated noise layers [9, 10]. To obtain a realistic result, artists must expend significant time tweaking parameters via trial-and-error. Libraries of materials, called "material packs" are also widely available for purchase [45], demonstrating both the value of using realistic material models and the effort required to create them. Hand-designed material models generally do not accurately capture the actual reflectance behavior of the real-world material that they are based on. Rather they mimic the real material's appearance, which is sufficient for many rendering applications.

The most complex and complete representations of real materials come from direct measurement. The surface appearance of the object is measured over a densely sampled hemisphere of light and view angles, using a device such as a gonioreflectometer, and these data are interpolated at render time from a four-dimensional lookup table [40]. When variation across the surface is included, the measurements span a six dimensional space — azimuth

and elevation angles for both the camera and light source and 2D coordinates on the surface — called a spatially varying bi-directional reflectance distribution function (svBRDF) [43]. Obtaining a measured svBRDF is a time-consuming and memory-intensive process that requires a sample to be brought into a lab with controlled lighting and specialized equipment. Not only are there very few measured svBRDFs available, but this high level of physical accuracy is also generally excessive when only visually plausible images are required. Furthermore, the mechanical nature of most BRDF capture devices requires the physical extent of the sample to be quite small, making them better suited to recording a single uniform BRDF rather than an svBRDF with larger scale textural variation.

In many cases the physical plausibility of the material is important, but the reflectance behavior is simple enough it can be accurately represented by a parametric model with only a few parameters. In these cases a parametric BRDF model can be created either by choosing arbitrary parameter values, navigating the space of BRDFs with a user interface, or fitting a model to observations of a real object. Well-designed BRDF models adhere to physical limitations such as conservation of energy and reciprocity, and can be represented more compactly than a measured BRDF.

I propose a method for allowing artists to create physically plausible parametric svBRDF representations of nearly-flat surfaces using simple printed fiducial markers and video obtained from a mobile phone with a flash that can be constantly illuminated. My technique does not require any specialized equipment and delivers a realistic representation of the scanned material that is suitable for many artistic applications. To capture svBRDFs in this way, the fiducial markers are placed on the outside four corners of the captured area and the phone camera and flash are moved over the surface at a very close capture distance that maximizes the spatial and angular resolution of the light and view capture positions. I demonstrate the quality and versatility of my method by reproducing a variety of spatially varying materials including leather, fabric, metal, wood, paint, and tile. Figure 1.1 shows four example materials captured with my method and rendered using Mitsuba [32], an open-source physically based renderer that supports many different advanced rendering techniques commonly used by researchers.

My capture technique is limited to relatively flat surfaces that have some medium-scale, non-repeating color variation to provide sufficient features for the alignment process. Additionally, due to the collocation of the light and camera, I am able to fit a realistic svBRDF model to materials with reflectance properties that are complex only near the peak of the specular highlight and I cannot capture or show reflectance behavior such as Fresnel effects that are apparent at extremely glancing angles. Fortunately these restrictions still allow for capture of many interesting and useful materials with a simple, low-cost solution.

Figure 1.1: Four materials captured using my method and rendered with Mitsuba. Objects are illuminated by the Uffizi environment map (courtesy of the USC Vision & Graphics Lab) along with two point light sources. The left two materials were captured with one camera, and the two on the right were captured with two cameras. From left to right: green faux leather, blue damask fabric, red velvet lamé, and wrapping paper. Inset image shows the average color per pixel as described in section 3.2.

Chapter 2

Background

2.1 Light

Visible light has four physical properties in the real world: intensity, direction, spectral distribution, and polarization. In this work the spectral distribution and polarization are not taken into account. Light sources are therefore treated as either point lights illuminating equally in all directions or as infinitely distant ambient light that is uniformly illuminating from all directions.

The study of measuring light and other forms of electromagnetic radiation is called *radiometry*. In the context of surface reflectance, light intensity is described by the term *radiance* ($watts \cdot steridian^{-1} \cdot meter^{-2}$) for a given wavelength in nanometers, which signifies the power per unit solid angle per unit area that is either emitted by a light source, or transmitted or reflected from one object to another. The term *irradiance* ($W \cdot m^{-2}$) is commonly used to describe the amount of power per unit area arriving at a surface. The inverse square law states that the irradiance of light from a point light source is inversely proportional to the square of the radial distance to the light source; or equivalently, the irradiance at distance r is equal to $(1/r^2)$.

Photometry is a related area of study to radiometry which measures the perceived brightness of a light source as observed by the human eye or a camera sensor. The parellel terms to radiometry are *luminance* ($lumens \cdot sr^{-1} \cdot m^{-2}$ or $candelas \cdot m^{-2}$) for radiance and *illuminance* ($lumens \cdot m^{-2}$) for irradiance. While radiometry is concerned with the power output of a light source measured in watts, lumens are defined relative to both the power output and also human sensitivity to light at a given wavelength.

Dynamic Range

Real world light sources span an incredible range of measured luminance, from starlight at $10^{-3}\, cd \cdot m^{-2}$ to lamplight at $10\, cd \cdot m^{-2}$ to direct sunlight at $10^{5}\, cd \cdot m^{-2}$. However, given a particular adaptive state for the human eye or exposure level of a camera, it is usually only possible to observe a relatively narrow range of luminance values at one time. The ratio of

brightest to dimmest luminance values that an observer can resolve is called *dynamic range*. Observed values that are outside the dynamic range of an image are said to be *clipped* to the maximum and minimum values, and because of this the actual luminance values at clipped locations in the image are unknown. However, it is possible to reconstruct the physical luminance values by capturing the same scene with multiple images, each with different exposure, such that the set of images includes at least one unclipped value for each point in the image. The multiple exposures can then be combined into a single high dynamic range (HDR) image [17]. Viewing HDR images directly requires an HDR display, but it is also possible to *tone-map* an HDR image into an easily viewable low dynamic range (LDR) image by compressing some portions of the dynamic range to preserve the maximum level of detail in both highlights and shadows [33, 21, 22].

Representing Light in Computer Generated Scenes

The influx of light at every point and direction in a scene can be thought of as a Light Field [25, 35], a Global Radiance Function [49], or a Plenoptic Function [1]. However, representing the full light field is usually impractical for computer generated scenes. For example, the full spectral distribution that describes the color of incoming and reflected light is fully described by an intensity value for each wavelength of light, but this is typically represented by a single color triplet in a particular color space such as RGB. Surface reflectance properties are also approximated with a function such as a BRDF that describes both the color and intensity of reflected light in each direction relative to a given surface.

Additionally, light sources have simplified representations including point lights, area lights, or environment lights. Point lights are represented as a single 3D position and radiance value, and they illuminate equally in all directions. Area lights also have positional representation that covers an extent in 3D space, and their radiance has directionality. The radiance of point lights follows the inverse square law, while the radiance of area lights is relative to the solid angle of the area light projected onto the surface. Environment or ambient lighting is a way to simulate a larger and more complex lighting environment without having to represent each individual emitting and reflecting light source in a scene. Environment lighting is typically represented as a hemisphere with varying radiant intensity, located at an effectively infinite distance from the object [16].

2.2 Camera

The pictorial representation of a scene that is formed by a camera is influenced by both the optical properties of the camera and also the sensor and post-processing settings that are used. Often both the optics and photometric representation of a camera are imperfect estimations of real world geometry and luminance and must therefore be calibrated using refrence objects with known physical properties.

Optics

Light enters a camera in the form of parallel rays corresponding to the same point in world space, and these rays are then refracted by the lens, restricted by the aperture, and finally imaged by the sensor. The distance between the lens and the sensor is called the *focal length*. An object is said to be *in focus* when the bundle of rays corresponding to a specific point on the object are all converged by the lens to meet at a single point at the exact depth of the sensor. When the convergence point is either in front of or behind the sensor, the spread of the rays on the sensor plane is called the *circle of confusion* and the radius of the spread is a measure of the defocus blur of the object.

Between the lens and the sensor there is also an aperture that restricts the maximum spread of rays from any given object. An ideal "pinhole" aperture restricts the rays from each point such that only a single ray passes through the aperture corresponding to only a single point on the sensor. Such an image is said to have an infinite *depth of field* (DOF), because all objects in the world are in focus regardless of their position relative to the focus distance. A wider aperture allows a larger maximum defocus blur, and will therefore have a narrower DOF since only the rays from objects near the same depth as the focus distance will converge to a single point on the sensor.

Sensor

A camera sensor is an array of *photosites* that collect red, green, or blue light and the resulting values from each triplet are combined to form a tri-color sample called a *pixel*. The resolution is determined by both the number of pixels per square inch and the size of the sensor.

Each photosite integrates all of the light that it receives over the period of time that the aperture is open, which is referred to as the *exposure time*. Additionally, it is possible to control the sensitivity of the sensor by adjusting the *ISO*. A high ISO causes the sensor to be less sensitive to the amount of incoming light, so more light is required to produce the same output pixel value. A low ISO causes the sensor to be more sensitive, which also means the output pixel values tend to have more noise. The size of the aperture, the shutter speed, and the ISO sensitivity, all combine together to produce a particular level of overall exposure for the image.

Because the sensor integrates light over time, any movement of the camera or objects may cause the same point to be imaged across multiple pixels in the same image and this effect is called motion blur. Motion blur may be reduced by decreasing either the speed of camera motion or exposure time.

Calibration

A photograph is a representation of the physical properties of the scene filtered through the physical properties of the camera. In order to recover an accurate estimation of the scene,

it is necessary to remove the distortions introduced by the camera via calibration.

The optics of the camera may cause distortions if there are aberrations in the shape of the lens or misalignment between the lens and the sensor. Nonlinear radial distortions are the most common, causing straight lines in a scene to appear curved either in a barrel, pincushion, or mustache shape (mustache distortion is a combination of both barrel and pincushion distortion that varies with eccentricity from the center of the image). The center of radial distortion curvature can also be displaced from the center of the image, known as tilt distortion. Finally, it is also sometimes common to observe a skew in the radial distortion that is caused by a horizontal or vertical tilt of the lens relative to the sensor.

These geometric distortions can be estimated by imaging an object of known size and shape (such as a checkerboard) from different points of view [59]. The calibration process relies on matching the corresponding checkerboard points in each image, which are then used to recover the relative orientation of the camera. The checkerboard points are then reprojected back into the images based on the geometry of the estimated camera pose, and any discrepancy between the actual position of the points in the image and the position of the reprojected points is attributed to lens distortion. A least squares optimization procedure is then used to fit a model that includes either two or three parameters for radial distortion and optionally includes two parameters for shift and two parameters for tilt. The resulting parameters describe the intrinsics of the camera and may be applied to other images captured with the same focal length. However, changes in focal length imply a change in the relative positions of the lens and the camera and therefore produce different distortions requiring a separate calibration. This change in lens distortion as a function of focal length is called lens "breathing". Breathing effects are most obvious at shorter focal lengths and they are particularly problematic in the case of free form camera motion due to the use of auto-focus.

In addition to geometric calibration, it is also sometimes necessary to determine the exact luminance and color balance of a scene relative to the particular exposure and white balance of the captured image. For this purpose it is necessary to introduce objects with known reflectance properties as reference points for calibration. For luminance it is common to use a gray card of known reflectance (such as 18%) and for color it is possible to either use an unclipped image of a white object to calibrate using the white-point method [12] or get a more complete spectral calibration using a Munsell color chart [13]. Both of these methods require the physical object to be present in the calibration image and the resulting adjustment values are only valid for the same lighting environment and camera sensor settings.

2.3 Geometry

Objects in a scene have a physical structure that may be broadly referred to as geometry, however, there are several levels of scale that may be implied by this term. The largest scale of geometry is the coarse shape of the object, represented by a set of polygons that dictate the overall surface orientation and curvature. The smallest scale of geometric variation is the microfacet normal distribution function that determines the microstructure of the surface

for each incoming light ray, which dictates the reflectance properties of the material. In between, the surface normal for each point on the surface can also be perturbed to achieve medium scale textural variation and shadowing effects. At each of these scales there are conventions for representing and measuring the geometry of objects for use in CG scenes.

Levels of Scale

The largest scale of object representation for rendered objects is typically an interpolated triangle mesh or bezier curve. An object is therefore composed of many smaller tiled units arranged in some geometric configuration relative to each other, wherein each component has some location and orientation. The triangles or rectangles that make up an object are typically represented by their vertices and, for bezier surfaces, control points.

The next smaller level of surface geometry is displacement, bump, and texture mapping. The surface of each geometric unit is also said to have an orientation called the *surface normal*. The orientation and position of the normal with respect to both artificial light sources and other parts of the object determine the amount of incident light on the surface. At this scale it is possible for parts of the object to both reflect light onto each other (*interreflections*) and also occlude incoming light from each other (*self-shadowing*). A *displacement map* is a perturbation of the surface geometry that is used for calculations of light interaction, including silhouettes, cast shadows, and self-shadowing. Displacement maps allow coarse object meshes to support finer geometric details. *Bump maps* are similar to displacement maps in that they perturb the surface normals, but they do not change the silhouette or cast shadows and they do not support self-shadowing. Color variation at this scale may also be added through the use of a *texture map*. This level of detail supports medium-scale variation that is obviously visible to the human eye even at viewing distances where the whole object is visible.

The smallest level of detail is surface roughness. Roughness is a descriptor of the *microfacet distribution* of a surface, where the scale of the normal variation is too small for the human eye to resolve, but the interaction of light with these small normal variations still produces larger scale effects in the reflectance behavior of the surface as a whole. This level of geometry is captured in the reflectance function, as described in the next section.

2.4 Surface Appearance

Generalized surface capture encompasses a wide variety of materials and techniques. The Bi-directional Texture Function (BTF) is used for materials with significant normal variation. Samples are captured in a series of images from various viewpoints and light positions to observe the shadowing and reflectance behavior of the material at a very fine scale [15]. The Bi-directional Scattering Distribution Function (BSDF) is used for materials with transparency or translucency which exhibit the property of subsurface scattering, wherein light may bounce around internally in a material eventually be re-emitted at another point on the

surface [6]. Finally, the Bi-directional Reflectance Distribution Function (BRDF) is used for homogeneous and relatively flat opaque materials [42]. In this work I focus on the spatially varying BRDF (svBRDF), a particular variant of the BRDF also introduced by Nicodemus et al. [43] that allows for different BRDF parameters at each point on the captured surface.

Sparse svBRDF Acquisition

A complete sampling of the six-dimensional svBRDF may be performed using a spatial gonioreflectometer [41], although complete capture is a lengthy and data-intensive task. Efforts have been made to simplify this process while still accurately capturing all the necessary variation in surface reflectance. Dong et al. [20] proposed a hand-held array of LEDs mounted in a circular array around a camera, and they interpolated an estimated svBRDF from a set of basis BRDFs obtained using manifold bootstrapping. Aittala et al. [3] used structured light from an LED panel display to show basis patterns of illumination, allowing them to fit a model to the observed reflectance behavior under these illumination patterns. Francken et al. [23] also use structured lighting varying in the spatial frequency domain to obtain high-quality surface normals and gloss estimation (rather than an svBRDF). Similarly, Ghosh et al. [27] used structured LEDs with polarizing filters to estimate the reflectance of spherical objects which contain a more complete sampling of surface normals. Zhou et al. [61] optimized a sparse blending of sparse basis BRDFs with a limited number of input views, as a way to minimize capture requirements. In another minimalist setup, Xu et al. [57] obtain interpolated uniform isotropic BRDFs from the MERL database with a two-shot capture system by using a dynamically determined error metric.

There have also been a variety of capture systems that employ polarized light to separate the diffuse and specular components. In 2010, Ghosh et al. extended their previous work to obtain spatially varying reflectance and refraction using a single lighting environment and circularly polarized spherical illumination [26]. Similarly, Ma et al. also used spherical objects but with linear and circular polarization to obtain surface normal and gloss levels [38]. Riviere et al. showed that high quality capture of surface normals and isotropic reflectance was possible under bright sunlight using calibration objects in conjunction with a linear polarizer attached to a DSLR camera [48]. Tunwattanapong et al. captured both reflectance and shape using continuous spherical harmonic illumination with very high spatial and angular resolution [52].

Linear light sources have also been used, such as by Chen et al., who placed the light source on a small electronic rig that traversed over the sample in a single dimension from which they estimated anisotropic BRDFs [14]. Ren et al. [46] proposed a portable setup involving a static mobile phone, a hand-held linear light source, and a collection of carefully selected physical material samples with known BRDFs. However, all of these methods still require expensive or highly specialized equipment for capture, such as multiple DSLR cameras or a complicated or bulky experimental LED hardware setup. Even the highly portable setup of Ren et al. still requires expensive reference material samples that would be difficult for an artist to obtain.

Appearance Matching

Another body of work focuses on tools to help artists match the appearance of a material through user input or by simplifying the material representation. Dong et al. [19] estimate a simplified model svBRDF for a single texture image and allow users to adjust the behavior of regions of similar appearance until they are satisfied. Di Renzo et al. [18] produce a layered BRDF plus texture image based on user edits in material space. Xuey et al. [58] create a static image with material weathering effects for a single lighting environment. Haro et al. [29] also produce a static image with a single light source "baked in" to the material appearance. All of these tools circumvent the need for capturing multiple lighting and viewing angles in favor of simplified appearance estimation.

Aittala et al. [4] combined texture synthesis from a no-flash photo with reflectance capture from a single flash photo to produce an svBRDF and normal map, however their technique was limited to highly regular, repeated textures. In subsequent work they replicated these results using a single flash image and deep learning techniques, but with less consistent results [2]. Most recently, Li et al. [36] also used deep learning to estimate the ambient lighting and thereby generate a diffuse, specular, and normal map decomposition of a single arbitrary image.

Image-Based Approximation

My work is most closely related to a group of approaches that approximate a full svBRDF model using a limited set of input images. Wang et al. [53] produce anisotropic svBRDFs by synthesizing a spatially varying Normal Distribution Function (NDF) from sparse light positions captured using a motorized LED array. Zickler et al. [62] estimate an isotropic svBRDF for objects with known pre-measured geometry using a fixed camera with a moving point light source in a controlled studio lighting environment. Goldman et al. [28] use a simplified studio setup to capture multiple high resolution photographs at various light and view angles and then used BRDF clustering along with linear blending of basis BRDFs to estimate an svBRDF for each scanned object. Lensch et al. [34] also used a studio setup for capture with metal spheres for light source tracking, and their iterative BRDF subclustering method is similar to the one presented here, although the capture setup is much more complex. Similarly, Zhou et al. [60] estimate an isotropic svBRDF for an arbitrary-shaped object by employing structure from motion (SFM) using a ring light source and a linear combination of basis BRDFs. Finally, two approaches have been proposed for estimating large scale geometry, specularity, and diffuse albedo based on input from a light field camera [54] and a small number of rendered input images [30], although the resulting parameterization of surface reflectance is less precise.

In the space of mobile phone capture, Thanikachalam et al. [50] estimated reflectance from video and optimized the sampling density and capture path, but with very low resolution output. The approach proposed by Hui and colleagues [31] requires capturing several images of a texture sample from different viewpoints with the flash providing illumination

and then using a dictionary-based approach to select a BRDF that matches the observations of each pixel. They provide an elegant proof showing that it is insufficient to fit a BRDF using only observations where the light is collocated with the camera, as is the case for a cellphone, but by using a dictionary they are able to still obtain plausible results for cases where the subject materials match an example stored in their library. I overcome this limitation in my approach by allowing the user to add a second cellphone camera to obtain observations from other perspectives that are not collocated with the light. This approach allows me to fit a BRDF model directly to the data so that each pixel's appearance is not restricted to an existing dictionary of materials. However, even when only one cell phone is used, my initialization strategy still allows the fitting process to obtain reasonable results.

Finally, Riviere et al. [47], also demonstrated svBRDF capture using mobile phone video. My proposed method improves on their capture system in two ways. First, my unique video frame alignment technique allows me to capture reflectance data from a much closer distance (10cm vs 50cm) and does not require the entire sample to be visible in each input image. By stitching together many partial observations from a closer view distance, I can obtain very high resolution results even for large samples. I have found that high resolution is generally required to obtain good results when rendering the svBRDFs on objects in 3D scenes, particularly for materials with fine-scale specular features such as gold thread or metallic flakes. The closer viewing distance also produces more oblique lighting and viewing angles and a brighter specular highlight, allowing me to accommodate capture under more varied ambient lighting conditions. I captured most of the data sets under approximately 400 LUX ambient illumination (compared to 40 LUX by Riviere et al.). Second, my method does not require either radiometric calibration of the device nor the inclusion of a specific color chart at capture time. The fiducial markers I used are less than 2cm square and can be printed anywhere and easily carried in a wallet. These differences expand the possible use cases of casual svBRDF estimation to more varied lighting environments and more accessible tools for capture.

Chapter 3

Approximate svBRDFs With Mobile Phone Video

My proposed capture and fitting technique requires only one or two cell phone cameras with continuous flash video capability and a set of four small fiducial markers which may be printed on standard copy paper. Using these commonly available tools, I am able to fit an svBRDF to mostly-flat, opaque surfaces that have spatially varying reflectance and uneven surface texture.

I first place the fiducial markers around the area of interest and capture a short, hand-held flash video at a relatively fixed distance over the object surface. I then align and warp the resulting video frame images into a single panorama in the global coordinate space with observations from multiple light and view locations for each pixel. In the second step, I cluster the pixels by similar appearance, fit a BRDF to the clusters, and then recursively sub-divide and fit a BRDF and normal vector displacement to each sub-cluster until the size of each cluster is sufficiently small.

My output is a high-resolution svBRDF based on the Ward model [55] that can be easily applied to 3D objects for rendering using standard rendering software. Additionally, because I do not require the fiducial markers to be visible in each frame, larger areas can be captured than could be contained in a single camera view, enabling me to obtain a high-resolution output with a more well-defined specular lobe in each sample image with maximum capture resolution.

3.1 Alignment and Pose Estimation

Each location in the svBRDF will correspond to pixels coming from many video frames. Each frame includes observations of some set of pixels from a particular light and view direction, which is the sample data from which we fit a BRDF model. Aligning the video frames is similar to a panorama stitching problem, but for fitting a BRDF the quality of the alignment must be very precise. Although a traditional panorama need only avoid noticeable seams

and distortions, in my case every pixel location needs to be correctly matched across all light and view positions to avoid spurious correlations between lighting and appearance.

The use of a mobile phone camera for this task creates several difficulties that must be overcome for good quality results. Mobile phone camera lenses are usually wide-angle, wide-aperture, and fixed focal length, with a very narrow depth of field (DOF) and significant barrel or moustache distortion. Traditionally, the lens distortion would be corrected using a checkerboard calibration technique [59], but such techniques require either a wide DOF or a relatively large viewing distance so that the entire checkerboard is in focus for all rotated camera positions. Furthermore, it is necessary to use auto-focus to accommodate the hand-held camera motion, but lens "breathing" effects are known to cause lens distortion to vary dramatically across different focus states. Stitching the panorama, therefore, requires solving for both the camera pose and the current lens distortion for every video frame independently.

One possible solution for correcting the distortion would be to use a parametric lens model in conjunction with a homography for the camera pose. However, in practice I found that typical low-order models with two or three parameters were not sufficiently accurate for my application, and higher-order models with up to seven parameters were cumbersome, slow to converge, and easily derailed by local minima. Another solution is to include fiducial markers to establish known physical locations from which one might compute a homography for each frame. To undistort the entire image, the markers would need to be placed on the outside edges of the captured area and all the markers would need to be visible in each video frame. Even for a relatively small sample of 10x10 cm, this arrangement requires a capture distance further than 30 cm. However, I found that a capture distance closer to 10-15 cm produces more oblique lighting and viewing angles, provides a brighter specular highlight, and allows for a higher resolution image of the surface for BRDF fitting. However, the closer capture distance necessitates that at most only one or even no markers are visible in each capture frame, and another alignment technique must be used.

In order to capture larger areas at a close distance, I therefore perform a feature-based alignment for each frame using multiple overlapping homographies to obtain a piece-wise linear approximation of the lens distortion, as explained in section 3.1. I establish a global coordinate system by placing four printed fiducial markers at the far edges of the captured region and obtaining a single reference image of the entire area taken from a larger distance than that of the video capture. Figure 3.1 shows a depiction of the capture setup for one and two cameras. The homography solution for each sub-frame is then calculated relative to this global space, allowing me to estimate the 3D world coordinate camera and light positions for all frames, even though 50-80% of frames have no fiducial markers visible.

Fiducial Markers and Reference Image

Fiducial markers are created and detected using the ArUco "6x6_250" predefined library [24]. The actual size of the printed markers is 1.6 cm square. Four markers are placed on the flat capture surface, one in each corner of the area to be captured. An additional reference image containing all four markers is also captured at a distance of approximately 30 cm

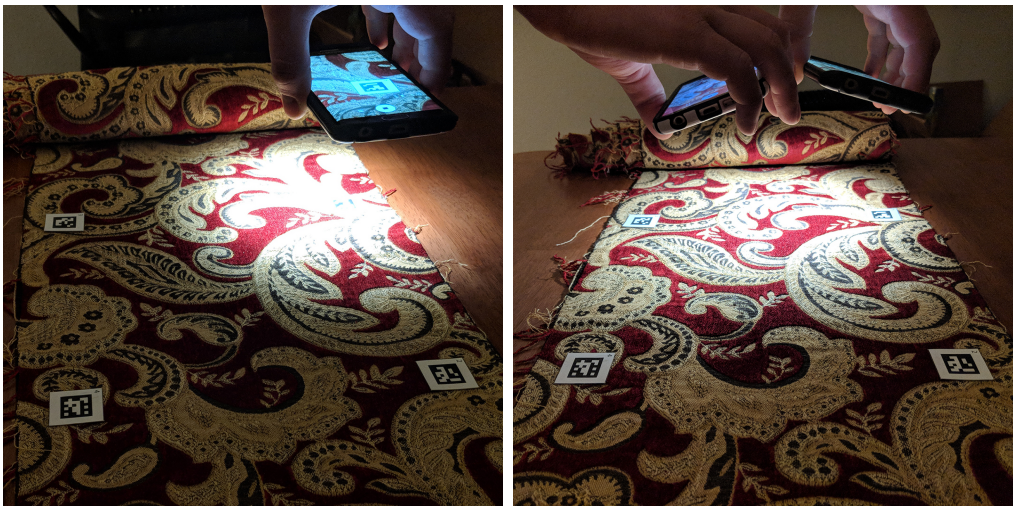


Figure 3.1: Example images showing the capture setup for one and two cameras. The capture distance from the surface is approximately 10-15 cm and the side-by-side arrangement of the two cameras allows for a larger range of relative light and camera angles.

perpendicular to the surface. The android app "Level Camera" is used to ensure the camera sensor was parallel to the surface for the reference image [56]. The locations of the fiducial marker corner points are recorded separately for both the reference image and in each video frame where the fiducials were visible.

Removing Blurry and Disconnected Frames

Because the camera is hand-held without any physical guide or apparatus, irregular camera motion can sometimes produce blurry frames as a result of intermittent defocus or motion blur. To detect blurry frames I compute a blur metric f based on the power spectrum for each image i such that

$$f(i) = \text{mean}(\log_{10}(0.001 + |\mathcal{F}(i)|) \tag{3.1}$$

where $\mathcal{F}(i)$ is the Fourier transform of the image, and the absolute value, log, and addition operators are applied competent-wise and the mean is taken across the result. Frames with a value of $f(i)$ less than 1.5σ below the mean across all frames are discarded.

When removing frames due to blur or insufficient feature matches, there is a potential for a small subset of frames to be well-connected to each other but lack sufficient connectivity to determine a unique homography transformation to other frames in the sequence. In that case it is impossible to determine an unambiguous projective transformation of that subset to the global space. At the end of the feature matching process I therefore obtain the connected sub-graph of the connectivity map with the most members and remove any frames not contained in that sub-graph.

Alignment Parameters

Scale	SURF	SIFT	Max. Dist.	Min. Inliers
0.25	1,200	3,000 (1,200)	20 pixels	20 points
1	4,800	12,000 (4,800)	60 pixels	60 points

Table 3.1: The parameters used for the coarse and fine alignment steps. Column 1 shows the scale of the input image (coarse or fine). Columns 2 and 3 show the number of SURF and SIFT features extracted from each image (the number of selected SIFT features for matching is shown in parentheses). Columns 4 and 5 show the maximum distance between matched inlier points and the minimum number of inlier points for the MLESAC procedure, respectively.

Coarse Alignment

Although each video frame may be trivially assumed to overlap with its immediate neighbors in time, accurate stitching of a full panorama also requires accurate loop closure for non-neighboring frames. However, feature matching across all pairs of frames at full resolution is costly and also likely to return many false matches for self-similar textures. I therefore first perform a coarse alignment step at a subsampled scale to determine overlapping frames, then repeat the process for the full resolution images to obtain the locations of matching features for the final homography estimation. Parameters for both alignment steps are shown in Table 3.1.

For the coarse alignment step, each frame is downsampled 4x, and a maximum of up to 1,200 uniformly distributed SURF features [7] and 3,000 SIFT features [37] are extracted from each frame. SIFT features are obtained and matched using the CudaSIFT library [8]. Features within a 75 pixel radius of the center of each image are discarded to avoid false matches of the specular highlight. During the feature matching process, all the SURF features and a random subset of 1200 SIFT features are uniquely matched (1 to 1) to all the features from each other frame. The matched feature points are used to estimate a similarity transformation between each pair of frames using MLESAC [51], with a maximum distance of 20 pixels between inlier feature locations. Any number of inliers greater than 20 is recorded as a potential match.

The resulting matrix of inlier counts (the connectivity map) is further thresholded and filtered to remove spurious matches. The threshold for the minimum number of inliers is determined by the 50th percentile of those frame pairs with some overlap. This ensured that no more than 50% of all frames can be overlapping and only the strongest connections remain. Finally, the connectivity map is smoothed using a 5x5 median filter to remove any non-continuous matches.

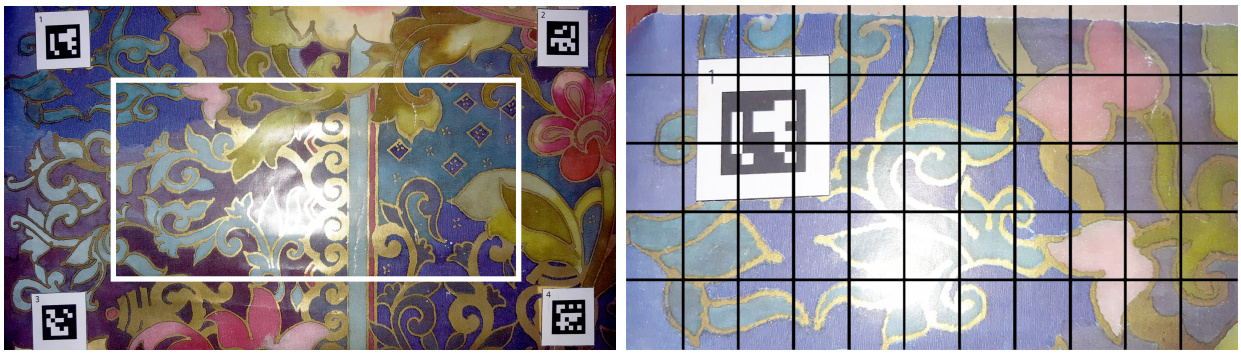


Figure 3.2: The left subfigure shows an example reference image with the user-selected output region outlined in white, while the right subfigure shows an example video frame from the same data set with sub-frame boundaries overlaid. The contrast of the example video frame has been decreased for visualization. The thickness of the black lines indicates the overlap between adjacent sub-frames.

Fine alignment and subdividing frames

In the fine alignment step, full-scale feature point locations are divided into sub-frame regions and I obtain a global least-squares solution for the homography transformation of each sub-frame.

Feature matching is only performed for overlapping image pairs from the coarse alignment connectivity map, with slightly modified parameters. The flash feature removal radius, maximum number of SURF features, and the max number of SIFT features are all scaled up by 4x. The maximum feature location distance for MLESAC is 60 pixels, and the minimum number of inliers is 60. The large allowable MLESAC distance error is a reflection of the amount of lens distortion. Although larger allowable error may cause incorrect matching, restricting the inliers to only precise matches causes only the undistorted portions of each frame to be matched, and this defeats the purpose of the alignment process completely. It is therefore much better to have a larger distance error and enforce precision by increasing the number of inliers. Ideally, any remaining false matches are greatly outnumbered in the final least squares minimization process.

The inliers from each frame are divided up into a 5x11 grid of uniformly sized sub-frames whose dimensions are determined by empirically examining the level of lens distortion in the phone cameras I used. An illustration of the sub-frame divisions is shown in Figure 3.2. For our setup, the size of each sub-frame is 448x370 pixels with an X and Y overlap of 25 and 22 pixels, respectively. Due to similarity of camera hardware specifications across mobile phones, it is likely that these values would be appropriate for other devices as well.

Linear approximation solution

Once I have obtained the corresponding feature locations, I solve for a homography transformation matrix for each sub-frame to the global space defined by the location of the fiducial markers in the reference image.

To obtain the transformation matrices, I perform a global least-squares fit simultaneously for all corresponding feature pairs across all overlapping frames. My solution is the set of homography matrices that minimizes the sum of squared differences between the projected global positions of each shared feature point p, as described in equation 3.2 below.

$$\min \sum \|F_{p_i} \cdot H_i - F_{p_j} \cdot H_j\|^2 \tag{3.2}$$

where F_{p_i} and F_{p_j} correspond to the $[x, y, w]$ homogeneous coordinates of feature point p in each pair of overlapping sub-frames i, j, and H_i and H_j are the corresponding homography matrices that project each image into the global space.

Unraveling and concatenating all homography matrices H_i into a single vector h allows me to construct a large sparse matrix F_{pij} where each column corresponds to one entry of h, and each row corresponds to $p_i - p_j$ in homogeneous coordinates. We then solve for the entries of h such that the error between p_i and p_j after they have been transformed to the global space by their corresponding entries is minimized. The minimization problem is therefore

$$F_{pij} \cdot h = 0. \tag{3.3}$$

Furthermore, since a homography is only precise up to a scale factor, I add the following constraints to define the global space:

$$H_i(3, 3) = 1 \tag{3.4}$$

such that the (3,3) entry of each homography matrix is defined to be one, and

$$F_m \cdot h_m + F_{\notin m} \cdot h_{\notin m} = 0 \tag{3.5}$$

$$F_{\notin m} \cdot h_{\notin m} = -k_m \tag{3.6}$$

where F_m is the set of rows in F_{pij} containing the m fiducial marker points, h_m is the corresponding entries of h, and $F_{\notin m}$ and $h_{\notin m}$ are the remaining entries of F_{pij} and h, respectively. The i entries of F_m are from the marker point locations in each sub-frame, while the j entries are from the marker point locations in the reference image. In (3.6) the product of the known entries is moved to the righthand side of the equation, yielding $-k_m$, so that $h_{\notin m}$ may be obtained via least squares.

Pose estimation

I determine the real world position of the camera using the homography of the center sub-frame of each input image. Each transformation matrix is decomposed into its component

rotation matrix, R_i, and translation vector, t_i according to the process described in Malis et al. [39]. I use these components to construct a set of 3D homogeneous matrices as shown in (3.7), wherein each matrix transforms from the reference image pose to the corresponding camera pose for each frame.

$$\begin{bmatrix} & R_i & & t_i \\ & & & \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.7}$$

The reference image camera pose and light position are determined as follows. The field of view (FOV) of the camera is calculated offline using a one-time calibration image with an object of known size at known distance. Both the rigid offset of the flash relative to the camera and the size of the fiducial markers are also measured offline. In my case the FOV was measured to be approximately 70°, the XYZ offset of the flash is [1.4, -0.3, 0] centimeters, and the fiducial markers are each 1.6 centimeters tall.

At capture time, the global XYZ origin point is defined to be the point on the captured surface corresponding to the center of the reference image. The reference camera pose is therefore located at [0, 0] in XY. The Z distance is triangulated from the FOV and the average size of the reference fiducial markers in pixels relative to their known physical size in centimeters. The reference light position is obtained by applying the known flash offset to the reference camera pose.

Finally, the world-to-image transformation matrix described in (3.7) is applied to both the reference camera and light positions to obtain the camera and light positions for each frame. Camera poses located outside the fiducial marker boundaries are not guaranteed to conform to the global constraints and so are discarded, along with their corresponding video frames.

3.2 Clustering and BRDF fitting

For each point on the surface, the sparse input samples typically cover only a very tiny sliver of the 4-dimensional hemisphere used to fit a BRDF. However, most materials are made up of many regions that share similar reflectance properties that would all be well-described by a single BRDF with minimal loss of accuracy. I take advantage of this similarity by clustering these pixels together and fitting a BRDF to each cluster.

Determining the number and size of the clusters presents a trade-off between generalizability and fidelity to the observed data. There are many ambiguous BRDF solutions that can produce the same appearance behavior. Larger clusters are likely to include a more complete sampling of the BRDF hemisphere and therefore converge to a more accurate representation, but they are also more likely to obscure the small details and variation which make spatially varying materials interesting. If the clusters are too small, however, it is probable that over-fitting will produce an incorrect result which does not generalize to novel light and view positions which were absent from the captured video.

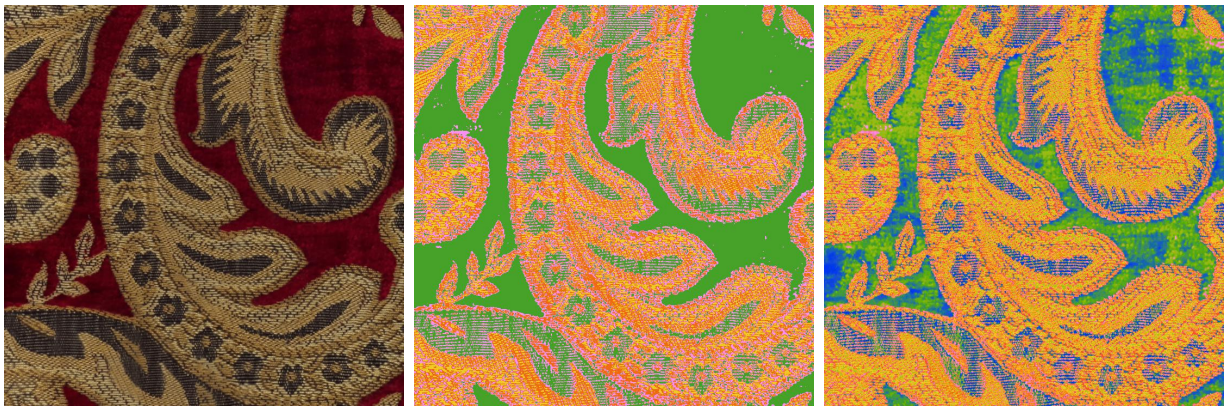


Figure 3.3: Initial and final sub-clusters for the two-camera red velvet lamé material. The left image shows the average color. Five clusters were obtained in the initial clustering (middle), and the final result included 5,152 sub-clusters (right).

Similar to Lensch et al. [34], my solution is to initialize the BRDF with very large clusters and a constrained BRDF model, and then recursively subdivide the clusters, initializing each sub-cluster with the fitting output of its parent. My initial clusters are grouped based on the average observed color of each pixel and then each cluster and sub-cluster is subdivided based on the per-pixel residual of the fitted BRDF. This encourages each smaller sub-cluster to find a solution in the neighborhood of solutions defined by the larger parent cluster, greatly reducing the likelihood of obtaining an incorrect ambiguous solution.

For each sub-cluster I fit an anisotropic Ward svBRDF [55] and a normal map. I am therefore able to fit opaque materials that do not have any Fresnel effects. Due to the feature-based homography alignment process, I also require the scanned material to be relatively flat and have at least some medium-scale albedo variation to align the input video frames.

Clustering and svBRDF Initialization

Using the aligned images, I coarsely approximate the diffuse albedo color by the average color of each pixel in the global coordinate space. This average color image is then converted to CIE 1976 L*a*b* color space. I then apply k-means clustering with k-means++ initial centroid positions [5] to the normalized albedo color values. The number of clusters, k, is chosen based on the linear bisection point of the summed squared Euclidean error across all values of k in the range $k = [2{:}20]$. For my data, typically k=4.

For each initial cluster, I fit an isotropic BRDF (see section 3.2) with a single normal vector for the cluster, constrained to be perpendicular to the surface (that is, $n = [0, 0, 1]$). This step initializes ρ_d, ρ_s, and α to reasonable values for the average normal vector orientation. The initial conditions for the isotropic fitting step are the average color over the entire cluster for the the diffuse component (ρ_d) and twice the average of the standard deviation across frames for the specular component (ρ_s). The roughness parameter (α) is initialized

to 0.1.

Once an isotropic BRDF has been fit to each initial cluster, I calculate the least squares fit error for each pixel in the cluster and recursively subdivide the pixels into two sub-clusters according to the threshold

$$t = \text{median}(E_{\text{px}}) + \text{mad}(E_{\text{px}}) \tag{3.8}$$

where mad is the median absolute deviation and E_{px} is the per-pixel fit error averaged over all observations for each pixel. Each sub-cluster is then fit with a full anisotropic BRDF and a normal offset, and the per-pixel fit error is calculated for the next iteration. I continue to subdivide clusters in this way until the size of each cluster reaches of a minimum of 50 pixels. Figure 3.3 shows an example of the progression from initial to final clusters for the red velvet lamé material.

Redundant Observations

Larger clusters tend to contain many redundant observations of similar materials with almost identical viewpoints and light locations. These extra observations dramatically increase the BRDF optimization runtime without improving the accuracy of the fit. To simplify the fitting process, I apply a binning and averaging step to obtain unique viewpoints and light locations. At each sub-clustering iteration, I group all observations for all pixels in the subcluster into 5° increments for each of θ_i, ϕ_i, θ_r, and ϕ_r, and 1 cm increments for the light radius, r. For each unique combination of these variables, all the BRDF input parameters (including light and view positions and observed color) are averaged together into a single unique observation for fitting. The contribution of each unique observation to the calculated fitting error is then weighted by the number of raw observations included in its average, according to Equation 3.11. To calculate the per-pixel fitting error, the fitted value for each unique viewpoint is applied to all the raw observations in its group.

Reflectance Modeling

I model the surface appearance for each point as the incident light from the camera flash multiplied by the BRDF and modulated by the solid angle of the light source as a function of the incident light angle.

The surface appearance is therefore described as

$$L_r(\theta_r, \phi_r) = \int_0^{2\pi} \int_0^{\frac{\pi}{2}} L_i \cdot (\theta_i, \phi_i) \cdot \rho_{bd}(\theta_i, \phi_i; \theta_r, \phi_r) \cdot \frac{\cos(\theta_i) \cdot dA}{r^2} \tag{3.9}$$

where

L_r is the reflected radiance to the camera

L_i is the incident radiance from the light

ρ_{bd} is the BRDF in RGB color space

θ_r and ϕ_r are the camera elevation and azimuth angles

θ_i and ϕ_i are the light elevation and azimuth angles

dA is the differential surface area of each pixel

r^2 is the radial distance to the light source

and all angles are relative to the normal vector. Similar to Aittala et al. [4] and many others, the ambient light is not explicitly modeled but rather implicitly incorporated into the BRDF.

The ρ_{bd} term in (3.9) is the Ward BRDF model, described by the following equation

$$\rho_{bd}(\theta_i, \phi_i; \theta_r, \phi_r) = \frac{\rho_d}{\pi} + \frac{\rho_s \cdot e^{-\tan^2(\theta_h)\cdot\left(\frac{\cos^2(\phi_h)}{\alpha_x^2}+\frac{\sin^2(\phi_h)}{\alpha_y^2}\right)}}{4\pi \cdot \alpha_x \cdot \alpha_y \cdot \sqrt{\cos(\theta_i)\cdot\cos(\theta_r)}} \tag{3.10}$$

where

ρ_d and ρ_s are the diffuse and specular albedo values in RGB color space

α_x and α_y are the roughness parameters in X and Y

θ_h and ϕ_h are the elevation and azimuthal angles of the half-vector between the light and camera

In the initial clustering step, an isotropic variant of this model is used wherein $\alpha_x = \alpha_y$. Subsequent subclustering iterations are fitted using the full anisotropic BRDF model and two normal vector offset angles, n_{θ_x} and n_{θ_z}, which describe the rotation of the normal vector about the X and Z axes respectively. In the final svBRDF and normal map, all the pixels in each sub-cluster are therefore represented by the eight BRDF parameters above (one per color channel for ρ_d and ρ_s) and two normal vector offset parameters.

The optimization problem is therefore

$$\begin{aligned} \text{minimize} \quad & \sum w \cdot \sum (L_f - L_o)^2 \\ \text{subject to} \quad & \{\rho_d, \rho_s\} \geq 0 \qquad 0^\circ \leq n_{\theta_x} \leq 45^\circ \\ & \{\alpha_x, \alpha_y\} > 0 \qquad 0^\circ \leq n_{\theta_z} \leq 180^\circ \\ & \rho_d + \rho_s \geq 1 \end{aligned} \tag{3.11}$$

where L_o is the observed color values, L_f is the fitted BRDF evaluated at corresponding angles to L_o, and w is the number of samples per unique viewpoint as described in section 3.2. I solve for (3.11) using a sequential quadratic programming (SQP) optimization function [44].

3.3 Joining Two Video Streams

Although the reflectance properties of some materials are well-described by observations using a single collocated camera and light source, incorporating a second simultaneous video stream allows me to also capture somewhat more complex materials without requiring other specialized tools. By capturing one video with the camera flash turned on while simultaneously capturing a second no-flash video, I can observe the behavior of the scanned material at more oblique light and view angles and thereby obtain a more complete sampling of the BRDF.

The majority of the pipeline is image-based and accepts a second video stream without any significant modification. The only requirement is that the two video streams be temporally synchronized at the first frame of each video, and that the length of the no-flash video be shorter than the flash video. This ensures that the position of the light source is known for all observed input frames.

To synchronize the time streams, I simply begin the no-flash recording first and then crop the start of the no-flash video to the frame where the light from the flash camera first appears. At the frame rates used in my capture setup the actual transition frame is typically highly visible because the rolling shutter effect produces an obvious transition line across the frame. This method afforded acceptable synchronization for my application where the hand held cameras are moving relatively slowly.

Chapter 4

Results

My capture data were obtained using a Samsung Galaxy S6 (or S7 for the second camera). The resolution of the reference images was 5312x2988, and the videos were captured at a rate of 30 frames per second (fps) and resolution of 2160x3840 pixels. I captured video in Pro-Mode with the flash turned on, using a shutter speed of 1/500 seconds and an appropriate ISO setting for the ambient light level, between 50 and 400. White balance was manually set at the beginning of each video to ensure consistency across frames.

The camera was moved over the surface by hand in a sinusoidal top-to-bottom and side-to-side fashion to achieve relatively even coverage of the entire captured area. Typical video capture distance was between 10-15 centimeters from the surface, and reference image capture distance was usually around 30 centimeters. Each video was 15-30 seconds in duration, covering an area of approximately 20x20 cm. From this sequence I extracted every 5th frame (6 fps) for the single-camera examples and every 10th frame (3 fps) for the two-camera examples. I found this sampling rate to be an acceptable trade-off between sampling density and data processing limitations.

I provide examples of seven materials captured with only a single camera (aged metal patina, blue damask fabric, buffalo leather, metallic embossed paper, orange ceramic tile, damask fabric reversed, and wood block), and seven materials captured with both one and two cameras for comparison (abstract oil painting, green faux leather, red velvet lamé, woven rattan mat, wrapping paper, corkboard, and shiny white tile).

Figure 4.1 shows a rendering of all the captured materials mapped onto spheres and illuminated by the Pisa environment map with a single additional point light source. The examples in the left column are captured with two cameras, the middle column depicts the same materials captured with a single camera, and the right column shows additional materials captured with only one camera. Figure 4.2 also shows several of the same materials used in a more natural scene under the the same illumination.

In Figures 4.3 and 4.4 I include a comparison to ground truth using a very oblique light position. This comparison is challenging because the lighting configuration is very different from anything in the input data for fitting the svBRDF, so the algorithm must rely on smoothness assumptions implicit in the Ward model. It is apparent that some high

frequency texture and corresponding specular highlights are missing for several materials. These highlights most likely occupy a very sharp peak of the BRDF, and are thus difficult for any method to accurately reproduce without direct observation. Nonetheless my method produces a plausible appearance for these samples. Additionally, in a video I also show a comparison between the input video frames and a rendering of the fitted svBRDF output using the input light and camera locations, for all materials.

Each of the svBRDF output channels is also included for more detailed analysis in Figures 4.5 through 4.10. The leftmost column is the average color as described in section 3.2. The remaining columns are the diffuse color (ρ_d), the specular color (ρ_s), the roughness parameter in the X direction (α_x), the roughness parameter in the Y direction (α_y), and the normal offset map. For materials captured with both one and two cameras, the results are shown side by side for comparison. A second video shows the materials textured onto a 3D object with animated illumination changes, as well as a comparison between the one- and two-camera results.

The differences between the quality of the single and dual camera results for the red velvet lamé and wrapping paper materials reveal the importance of broader sampling for more complex materials. The diffuse parameter color is slightly darker for the two-camera wrapping paper example, but the overall result is very similar to the one camera result. However, for the red velvet lamé, the single camera case has much more trouble separating and distinguishing reflectance behavior that changes quickly with direction, as predicted by Hui et al. [31]. I still get usable results with a single camera, but the algorithm is unable to disambiguate between a bright surface tilted away from the camera and a darker surface tilted toward the camera, resulting in over-fitting to the data. This problem could potentially be corrected manually, but given that it is relatively easy to use two cellphones, I feel that two cameras is the preferred option when accurate reproduction is desired.

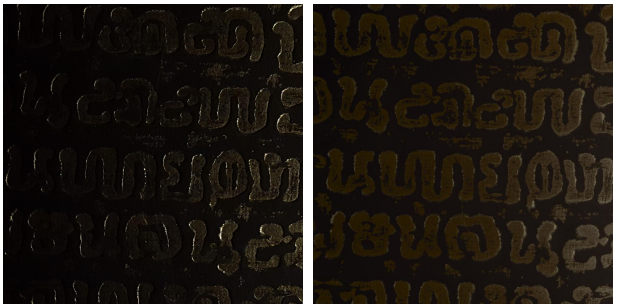
The corkboard and shiny tile materials illustrate failures for my method caused by over-fitting. In the case of the corkboard material, sloping of the overall surface normal at the top of the sample incorrectly bleeds into both the specular and roughness parameters. Although the resulting material can still be rendered with reasonable results, some applications such as selective editing of individual layers of the svBRDF would not be possible. In the shiny tile material the locations of the flash (dark circles) are segmented out into their own sub-clusters and fit with a darker specular value to offset the intense flash brightness. The manifestation of this is more severe, as specular highlights are only apparent at the locations they were originally observed. The shiny white tile was the only material for which I observed this behavior, and I speculate that it may be partly caused by a much darker auto-exposure setting on the camera that was incompatible with the estimated flash radiance value.

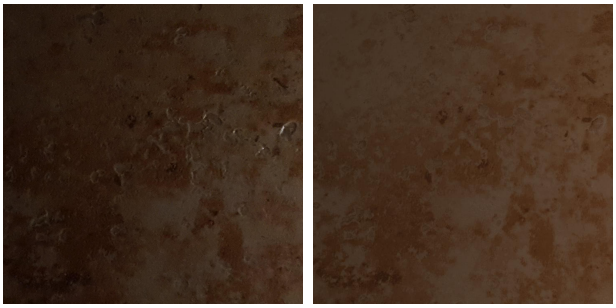
Figure 4.1: All of the scanned materials rendered onto spheres with Mitsuba and illuminated by the Pisa environment map and a single additional point light source. The right column is captured with two cameras, the middle column depicts the same materials captured with a single camera, and the left column shows additional materials captured with only one camera. From top to bottom and left to right: abstract oil painting, green faux leather, red velvet lamé, woven rattan mat, wrapping paper, corkboard, shiny white tile, aged metal patina, blue damask fabric, buffalo leather, metallic embossed paper, orange ceramic tile, damask fabric reversed, and wood block.

Figure 4.2: A natural scene with three scanned materials rendered with Mitsuba and illuminated by the Pisa environment map and a single additional point light source. The table surface is textured with the damask fabric reversed material, the teapot is textured with the faux green leather material, and the teacup is textured with the aged metal patina material.


aged metal patina


buffalo leather


metallic embossed paper


orange ceramic tile


damask fabric reversed


wood block


blue damask fabric

Figure 4.3: Comparison to a ground truth photo with an oblique light angle not included in the input fitting data. For each material shown, the first image is the ground truth and the second image is a rendering with the same light pose as the ground truth using the data captured with **one camera**. Images have been cropped square and resized to fit.

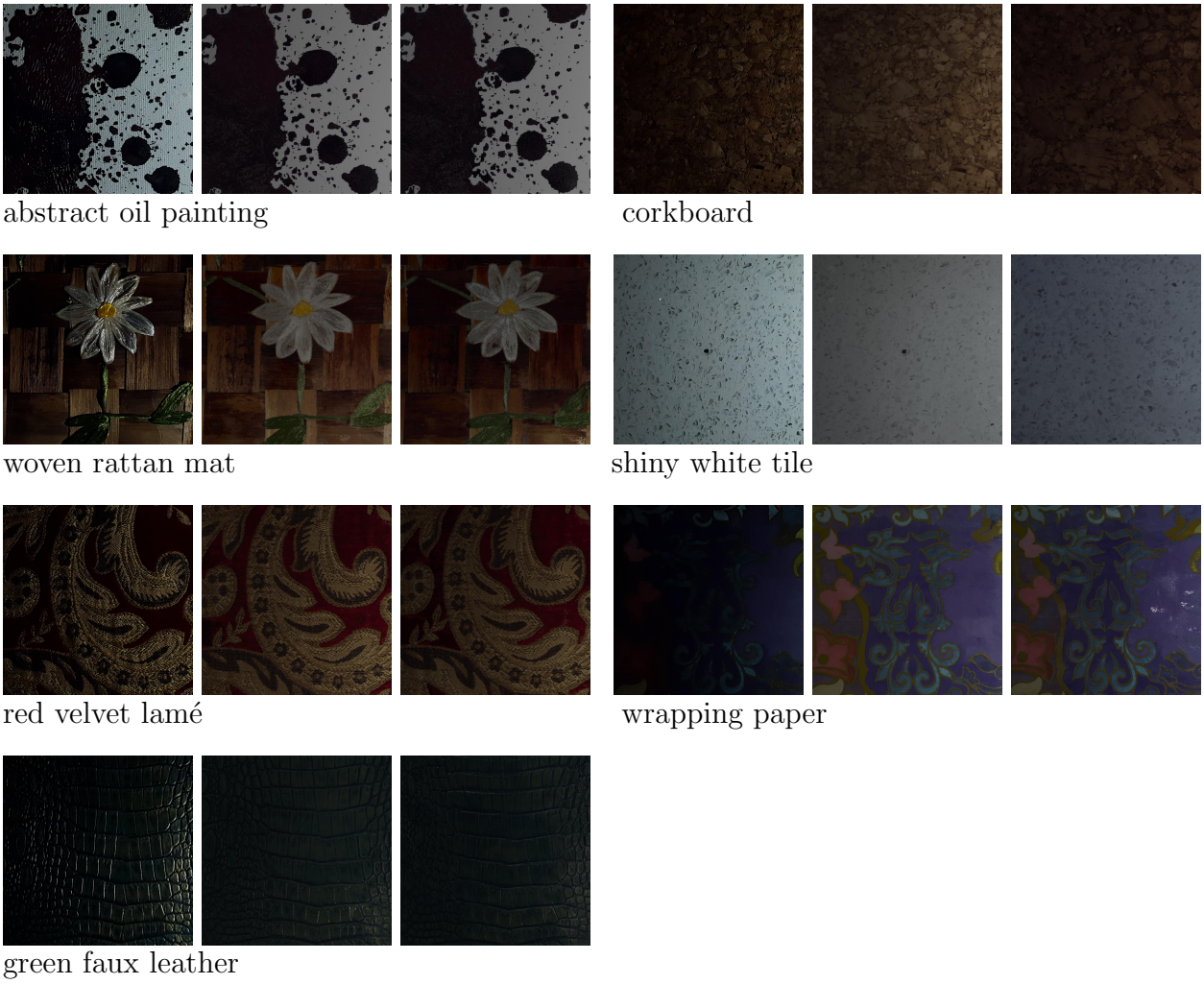
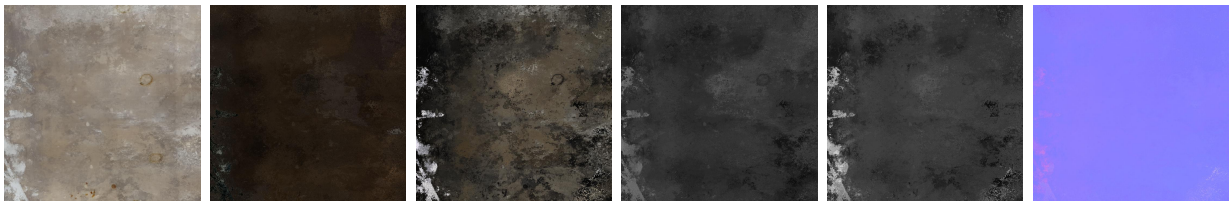
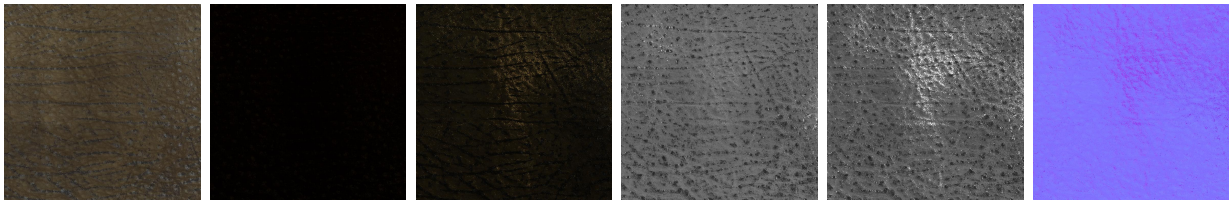


Figure 4.4: Comparison to a ground truth photo with an oblique light angle not included in the input fitting data. For each material shown, the first image is the ground truth, the second image is a rendering with the same light pose as the ground truth using the data captured with one camera, and the third image shows the same rendering using the data captured with two cameras. Images have been cropped square and resized to fit.

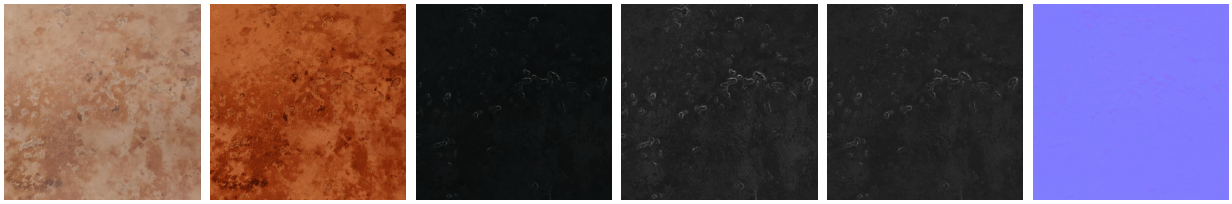


aged metal

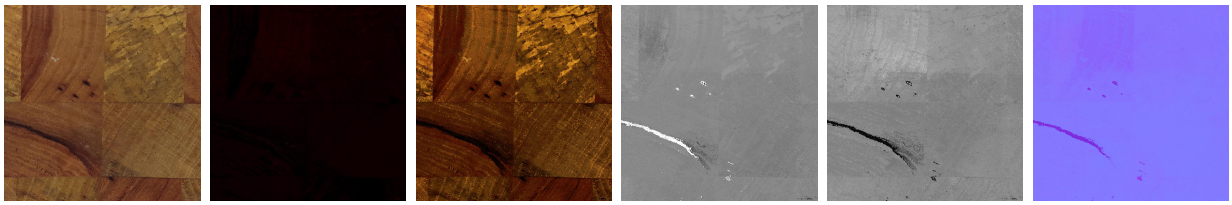


buffalo leather

metallic embossed paper

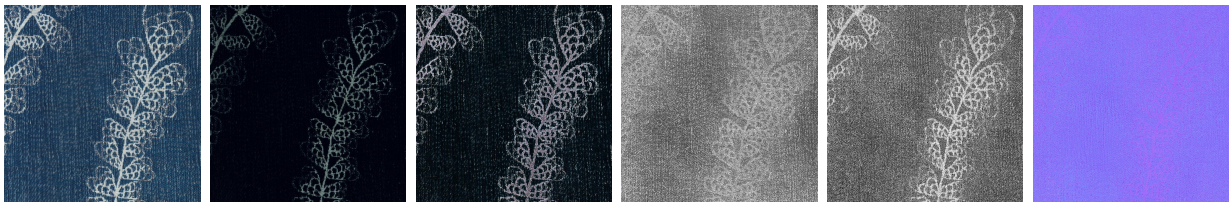


orange ceramic tile

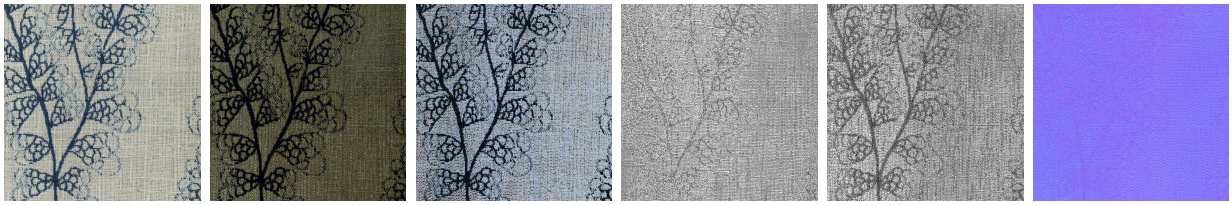


wood block

Figure 4.5: A sample of the results for five of the seven materials captured with **one camera**. Each row, from left to right: average color, ρ_d, ρ_s, α_x, α_y, and the normal offset map. Images have been cropped square and resized to fit.



blue damask fabric



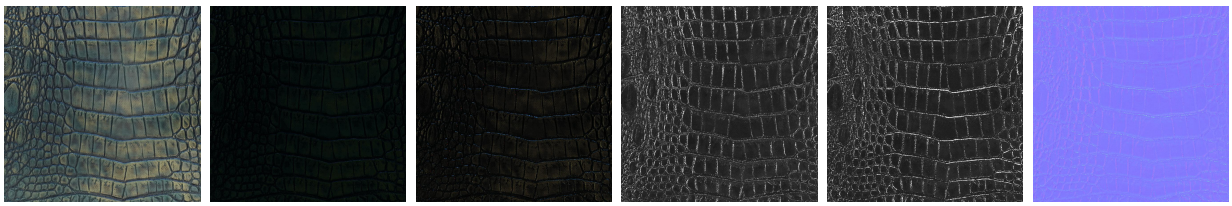
damask fabric reversed

Figure 4.6: A sample of the results for two of the seven materials captured with **one camera**. Each row, from left to right: average color, ρ_d, ρ_s, α_x, α_y, and the normal offset map. Images have been cropped square and resized to fit.

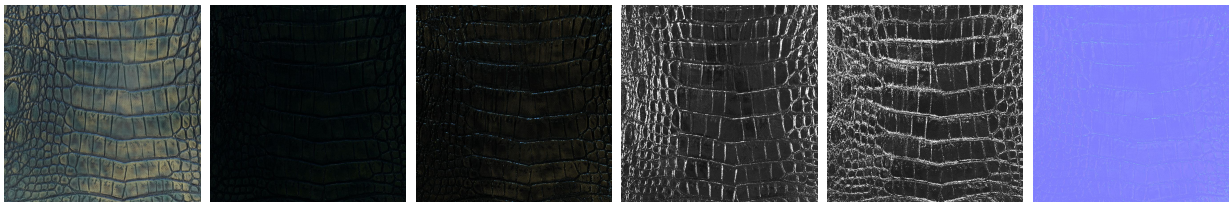
woven rattan mat (one camera)

woven rattan mat (two cameras)

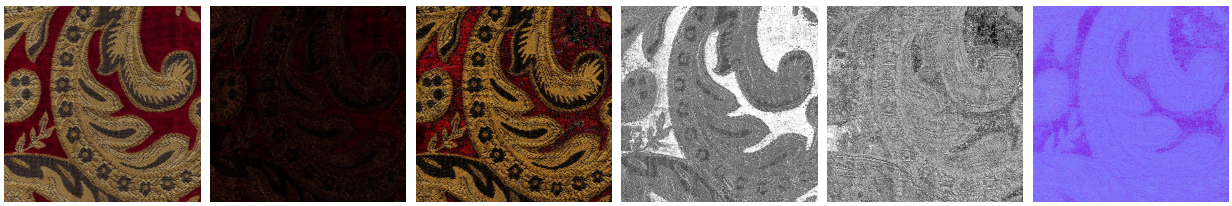
Figure 4.7: Example results showing the fitted svBRDF output. The top row of each material shows the results for one camera, while the bottom row shows the results for two cameras. Each row, from left to right: average color, ρ_d, ρ_s, α_x, α_y, and the normal offset map. Images have been cropped square and resized to fit.



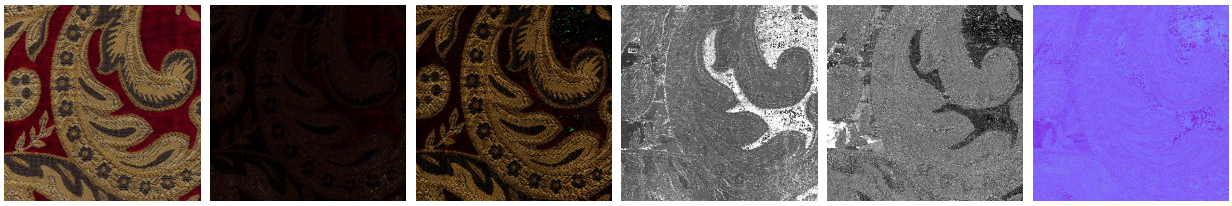
green faux leather (one camera)



green faux leather (two cameras)

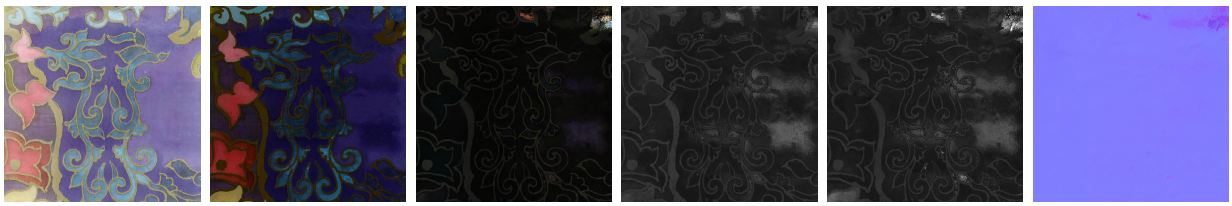


red velvet lamé (one camera)

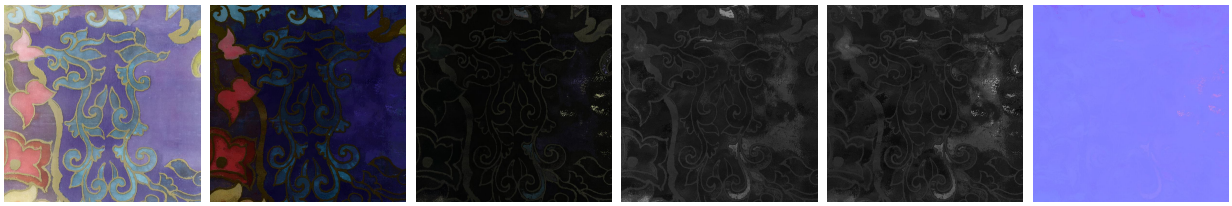


red velvet lamé (two cameras)

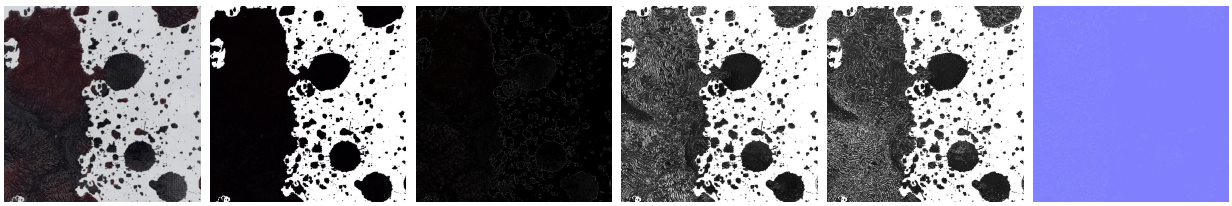
Figure 4.8: Example results showing the fitted svBRDF output. The top row of each material shows the results for one camera, while the bottom row shows the results for two cameras. Each row, from left to right: average color, ρ_d, ρ_s, α_x, α_y, and the normal offset map. Images have been cropped square and resized to fit.



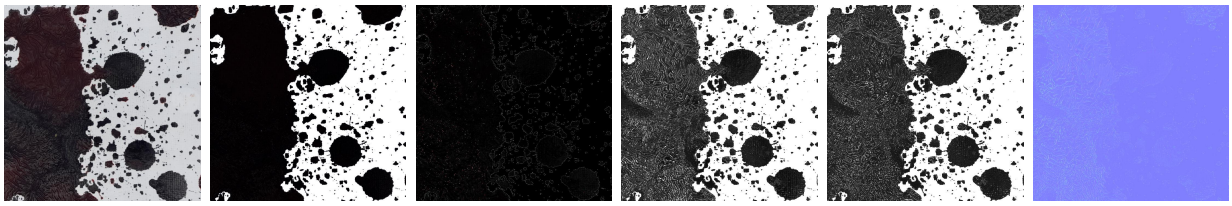
wrapping paper (one camera)



wrapping paper (two cameras)

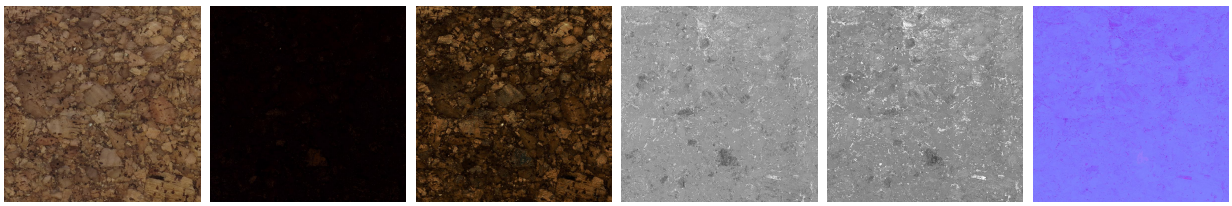


abstract oil painting (one camera)

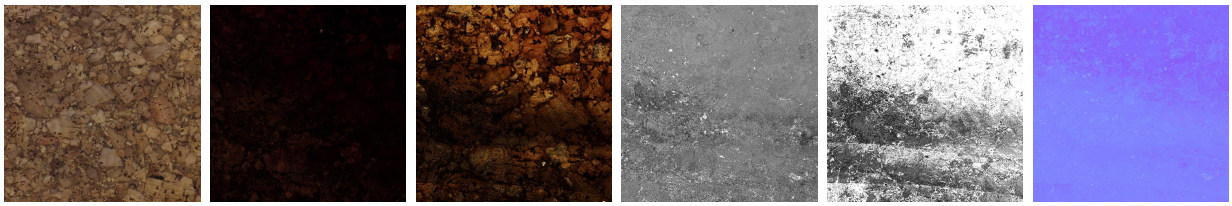


abstract oil painting (two cameras)

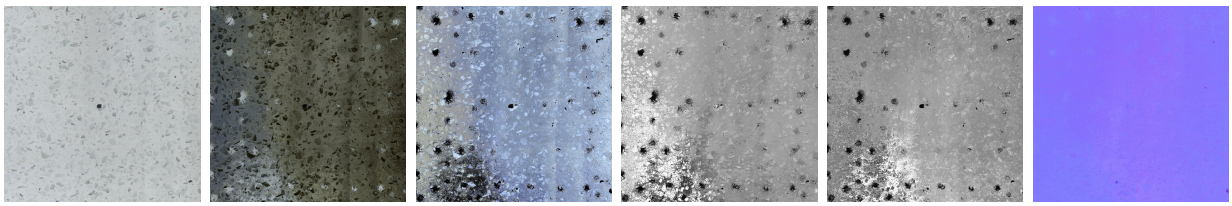
Figure 4.9: Example results showing the fitted svBRDF output. The top row of each material shows the results for one camera, while the bottom row shows the results for two cameras. Each row, from left to right: average color, ρ_d, ρ_s, α_x, α_y, and the normal offset map. Images have been cropped square and resized to fit.



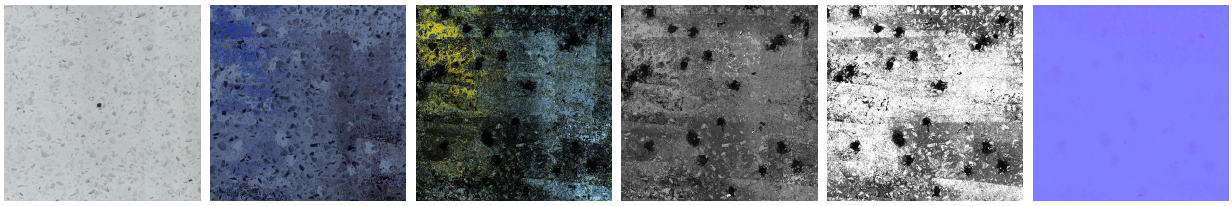
corkboard (one camera)



corkboard (two cameras)



shiny white tile (one camera)



shiny white tile (two cameras)

Figure 4.10: Example results showing two failure cases for the fitted svBRDF output. The top row of each material shows the results for one camera, while the bottom row shows the results for two cameras. Each row, from left to right: average color, ρ_d, ρ_s, α_x, α_y, and the normal offset map. Images have been cropped square and resized to fit.

Chapter 5

Conclusion and Future Work

I have demonstrated a new method for capturing and modeling the reflectance and small-scale normal displacement of nearly flat surfaces using printed fiducial markers and a mobile phone video with continuous flash. My technique employs a very simple capture process using off-the-shelf hardware, and the output of my system may be directly textured onto 3D objects using standard rendering software. I have also provided examples showing a variety of materials captured with both one and two cameras and rendered under complex lighting environments.

My technique has several limitations. First, because I align the video frame images using homographies I am only able to capture flat surfaces with relatively minimal surface relief. In practice, if the surface is smooth but unlevel, the alignment optimization will warp the unlevel surface to the canonical level surface defined in the reference coordinates and the camera poses will all be offset by a single homography transformation corresponding to the tilt of the surface plane. In the svBRDF fitting stage, this will produce reflectance behavior that is qualitatively similar to the original sample but not directly comparable for the same light and view positions. On the other hand, if the surface is level but has significant normal variation, then no correct homography solution can possibly exist and the strength and number of features on each side of the surface contours may bias the alignment toward different solutions for each video frame. The resulting svBRDF may therefore find a solution with correlated variation between reflectance behavior and light position (similar to the results for overall misaligned images) rather than the true normal vector variation. This is even more apparent for materials with strong self-shadowing effects.

The feature-based alignment also requires captured materials to have some irregular, medium-scale textural variation. Extremely repetitive textures may not be aligned with this method at all, and many materials with both strong specular highlights and repetitive textural motifs pose a significant challenge for alignment. It may be possible to create a better feature vector that is specifically tuned to capture similar geometric structure under varying lighting with greater robustness to a large number of close matches, and this could improve the robustness of my alignment method. However, Aittala and colleagues [4, 2] have already proposed two excellent solutions for capturing self-similar materials and I consider

this work to be complementary to theirs.

The results shown here are generated with very sparse sampling and a very simple BRDF model, and I am therefore unable to capture phenomena such as Fresnel effects. The methods I used do not place any restrictions on the BRDF model, and it has been suggested that some micro-facet models may be better suited to approximating more complex reflectance behavior [11]. However, it is unclear whether a BRDF model with more parameters might also require a greater number of constraints or a more precise initialization to achieve stable, well-behaved solutions. It could also be possible to combine the panorama stitching method with the dictionary approach proposed by Hui and colleagues [31] to obtain high-resolution models of complex materials that require sampling at very oblique light and camera angles. Alternatively, for materials with much more complicated reflectance, my implementation would allow the second no-flash camera to be placed on a tripod at an oblique angle to the surface to capture the entire flash sequence from the side. This might result in a non-uniform reduction in resolution because of the perspective distortion of the tripod camera, but the trade-off for more complete angular sampling could be worthwhile for some materials.

Finally, my research-quality code is not yet optimized and takes about 2 hours to align and fit an svBRDF from a 20 second video, half of which is taken up by feature extraction and matching. I speculate that using optic flow information for loop closure might produce a better estimation of overlap across frames without the need for feature matching in the coarse alignment step, providing a significant time savings.

There are many additional sensors that could be added to my pipeline to expand the possible applications and quality of the results. Depth cameras that are robust to unusual surface reflectance properties could be used in conjunction with robust point cloud optimization to fit an svBRDF of a non-flat object using a mobile phone. Accelerometer data from the phone could also be used to augment or even replace the pose estimation step. Recent advances in material estimation using deep learning networks could be incorporated into the BRDF fitting pipeline to augment material models without having to directly sample oblique light and viewing angles for each sample. Finally, the user could also be brought into the loop by selecting a family of materials such as wood or fabric to introduce a soft constraint on the reflectance properties of the final output.

However, even in its current form I believe that the reduced equipment requirements and simplicity of my capture methodology is a significant contribution to the state of the art. Any content creator with access to a printer and a mobile phone can quickly and easily obtain svBRDFs from a variety of interesting materials encountered in everyday life. I hope that more accessible capture techniques like this one will democratize realistic svBRDF capture for everyone.

Bibliography

[1] Edward H Adelson, James R Bergen, et al. "The plenoptic function and the elements of early vision". In: (1991).

[2] Miika Aittala, Timo Aila, and Jaakko Lehtinen. "Reflectance modeling by neural texture synthesis". In: *ACM Transactions on Graphics (TOG)* 35.4 (2016), p. 65.

[3] Miika Aittala, Tim Weyrich, and Jaakko Lehtinen. "Practical SVBRDF capture in the frequency domain." In: *ACM Trans. Graph.* 32.4 (2013), p. 110. URL: http://reality.cs.ucl.ac.uk/projects/fourier/fourier-lowres.pdf (visited on 05/27/2015).

[4] Miika Aittala, Tim Weyrich, and Jaakko Lehtinen. "Two-shot SVBRDF capture for stationary materials". In: *ACM Transactions on Graphics (TOG)* 34.4 (Aug. 2015), p. 110. URL: http://reality.cs.ucl.ac.uk/projects/two-shot-svbrdf/two-shot-svbrdf-lowres.pdf (visited on 09/06/2015).

[5] David Arthur and Sergei Vassilvitskii. "k-means++: The advantages of careful seeding". In: *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035.

[6] F O. Bartell, E. L. Dereniak, and W. L. Wolfe. "The Theory And Measurement Of Bidirectional Reflectance Distribution Function (Brdf) And Bidirectional Transmittance Distribution Function (BTDF)". In: vol. 0257. 1981, pp. 154–160. DOI: 10.1117/12.959611. URL: http://dx.doi.org/10.1117/12.959611 (visited on 04/24/2016).

[7] Herbert Bay et al. "Speeded-up robust features (SURF)". In: *Computer vision and image understanding* 110.3 (2008), pp. 346–359.

[8] Mårten Björkman, Niklas Bergström, and Danica Kragic. "Detecting, segmenting and tracking unknown objects using multi-label MRF inference". In: *Computer Vision and Image Understanding* 118 (2014), pp. 111–127.

[9] Neil Blevins. *Layering Materials*. Sept. 2013. URL: http://www.neilblevins.com/cg_education/layering_materials/layering_materials.htm (visited on 12/12/2017).

[10] Neil Blevins. *Leather Material*. June 2005. URL: http://www.neilblevins.com/cg_education/leather_material/leather_material.htm (visited on 12/12/2017).

[11] Adam Brady et al. "genBRDF: discovering new analytic BRDFs with genetic programming". In: *ACM Trans. Graph.* 33 (2014), 114:1–114:11.

[12] Vlad C Cardei, Brian Funt, and Kobus Barnard. "White point estimation for uncalibrated images". In: *Color and Imaging Conference*. Vol. 1999. 1. Society for Imaging Science and Technology. 1999, pp. 97–100.

[13] Munsell Soil Color Charts. "Munsell color". In: *Macbeth Division of Kollmorgen Corporation, Baltimore, Maryland, USA* (1975).

[14] Guojun Chen et al. "Reflectance scanning: estimating shading frame and BRDF with generalized linear light sources". In: *ACM Transactions on Graphics (TOG)* 33.4 (2014), p. 117.

[15] Kristin J. Dana et al. "Reflectance and texture of real-world surfaces". In: *ACM Transactions on Graphics (TOG)* 18.1 (1999), pp. 1–34. URL: `http://dl.acm.org/citation.cfm?id=300778` (visited on 04/24/2016).

[16] Paul Debevec. "Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography". In: *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*. ACM. 1998, pp. 189–198.

[17] Paul E Debevec and Jitendra Malik. "Recovering high dynamic range radiance maps from photographs". In: *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co. 1997, pp. 369–378.

[18] Francesco Di Renzo, Claudio Calabrese, and Fabio Pellacini. "AppIm: linear spaces for image-based appearance editing". In: *ACM Transactions on Graphics (TOG)* 33.6 (2014), p. 194. URL: `http://dl.acm.org/citation.cfm?id=2661282` (visited on 06/02/2015).

[19] Yue Dong et al. "AppGen: interactive material modeling from a single image". In: *ACM Transactions on Graphics (TOG)*. Vol. 30. ACM, 2011, p. 146. URL: `http://dl.acm.org/citation.cfm?id=2024180` (visited on 06/02/2015).

[20] Yue Dong et al. "Manifold bootstrapping for SVBRDF capture". In: *ACM Transactions on Graphics (TOG)*. Vol. 29. ACM, 2010, p. 98. URL: `http://dl.acm.org/citation.cfm?id=1778835` (visited on 09/08/2015).

[21] Frédo Durand and Julie Dorsey. "Interactive tone mapping". In: *Rendering Techniques 2000*. Springer, 2000, pp. 219–230.

[22] James A Ferwerda et al. "A model of visual adaptation for realistic image synthesis". In: *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM. 1996, pp. 249–258.

[23] Yannick Francken et al. "Gloss and normal map acquisition of mesostructures using gray codes". In: *International Symposium on Visual Computing*. Springer, 2009, pp. 788–798.

[24] Sergio Garrido-Jurado et al. "Automatic generation and detection of highly reliable fiducial markers under occlusion". In: *Pattern Recognition* 47.6 (2014), pp. 2280–2292.

[25] Arun Gershun. "The light field. Moscow". In: *Journal of Mathematics and Physics* 18 (1936).

[26] Abhijeet Ghosh et al. "Circularly polarized spherical illumination reflectometry". In: *ACM Transactions on Graphics (TOG)*. Vol. 29. ACM, 2010, p. 162. URL: http://dl.acm.org/citation.cfm?id=1866163 (visited on 04/26/2016).

[27] Abhijeet Ghosh et al. "Estimating specular roughness from polarized second order spherical gradient illumination". In: *SIGGRAPH 2009: Talks*. ACM, 2009, p. 30.

[28] Dan B Goldman et al. "Shape and Spatially-Varying BRDFs from Photometric Stereo". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32.6 (June 2010), pp. 1060–1071. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2009.102. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4912219 (visited on 05/27/2015).

[29] Antonio Haro and Irfan A. Essa. "Exemplar-Based Surface Texture." In: 2003, pp. 95–101. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.398.581&rep=rep1&type=pdf (visited on 06/03/2015).

[30] Z. Hui and A. C. Sankaranarayanan. "Shape and Spatially-Varying Reflectance Estimation from Virtual Exemplars". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.10 (Oct. 2017), pp. 2060–2073. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2016.2623613.

[31] Zhuo Hui et al. "Reflectance Capture Using Univariate Sampling of BRDFs". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 5362–5370.

[32] Wenzel Jakob. *Mitsuba renderer*. http://www.mitsuba-renderer.org. 2010.

[33] Adam G Kirk and James F O'Brien. "Perceptually based tone mapping for low-light conditions". In: *ACM Trans. Graph.* 30.4 (2011), pp. 42–1.

[34] Hendrik Lensch et al. "Image-based reconstruction of spatial appearance and geometric detail". In: *ACM Transactions on Graphics (TOG)* 22.2 (2003), pp. 234–257. URL: http://dl.acm.org/citation.cfm?id=636891 (visited on 05/27/2015).

[35] Marc Levoy and Pat Hanrahan. "Light field rendering". In: *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM. 1996, pp. 31–42.

[36] Xiao Li et al. "Modeling Surface Appearance from a Single Photograph Using Self-augmented Convolutional Neural Networks". In: *ACM Trans. Graph.* 36.4 (July 2017), 45:1–45:11. ISSN: 0730-0301. DOI: 10.1145/3072959.3073641. URL: http://doi.acm.org/10.1145/3072959.3073641.

[37] David G. Lowe. "Distinctive image features from scale-invariant keypoints". In: *International journal of computer vision* 60.2 (2004), pp. 91–110.

[38] Wan-Chun Ma et al. "Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination". In: *Proceedings of the 18th Eurographics conference on Rendering Techniques*. Eurographics Association, 2007, pp. 183–194. URL: http://dl.acm.org/citation.cfm?id=2383873 (visited on 04/26/2016).

[39] Ezio Malis and Manuel Vargas. "Deeper understanding of the homography decomposition for vision-based control". PhD Thesis. INRIA, 2007.

[40] Wojciech Matusik. "A data-driven reflectance model". PhD thesis. Massachusetts Institute of Technology, 2003.

[41] David K. McAllister, Anselmo Lastra, and Wolfgang Heidrich. "Efficient rendering of spatial bi-directional reflectance distribution functions". In: *Proceedings of the ACM SIGGRAPH/EUROGRAPHICS conference on Graphics hardware*. Eurographics Association, 2002, pp. 79–88. URL: http://dl.acm.org/citation.cfm?id=569057 (visited on 04/26/2016).

[42] Fred E. Nicodemus. "Directional Reflectance and Emissivity of an Opaque Surface". en. In: *Applied Optics* 4.7 (July 1965), p. 767. ISSN: 0003-6935, 1539-4522. DOI: 10.1364/AO.4.000767. URL: https://www.osapublishing.org/abstract.cfm?URI=ao-4-7-767 (visited on 04/24/2016).

[43] Fred E. Nicodemus et al. *Geometrical considerations and nomenclature for reflectance*. Vol. 160. US Department of Commerce, National Bureau of Standards Washington, DC, USA, 1977. URL: http://graphics.stanford.edu/courses/cs448-05-winter/papers/nicodemus-brdf-nist.pdf (visited on 04/24/2016).

[44] Jorge Nocedal and Stephen Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

[45] Andrew Price. *Introducing Poliigon - our new texture site!* May 2016. URL: https://www.blenderguru.com/articles/introducing-poliigon-new-texture-site (visited on 12/12/2017).

[46] Peiran Ren et al. "Pocket reflectometry". en. In: *ACM Transactions on Graphics (TOG)* 30.4 (Aug. 2011), p. 45. DOI: 10.1145/1964921.1964940. URL: http://portal.acm.org/citation.cfm?doid=1964921.1964940 (visited on 05/27/2015).

[47] Jérémy Riviere, Pieter Peers, and Abhijeet Ghosh. "Mobile surface reflectometry". In: *Computer Graphics Forum*. Vol. 35. Wiley Online Library, 2016, pp. 191–202.

[48] Jérémy Riviere et al. "Polarization imaging reflectometry in the wild". In: *ACM Transactions on Graphics (TOG)*. Vol. 36. ACM, Nov. 2017, p. 206. URL: https://doi.org/10.1145/3130800.3130894.

[49] Peter S Shirley. "Physically based lighting calculations for computer graphics". PhD thesis. University of Illinois at Urbana-Champaign, 1991.

[50] N. Thanikachalam et al. "Handheld reflectance acquisition of paintings". In: *IEEE Transactions on Computational Imaging* PP.99 (2017), pp. 1–1. DOI: 10.1109/TCI.2017.2749182.

[51] Philip HS Torr and Andrew Zisserman. "MLESAC: A new robust estimator with application to estimating image geometry". In: *Computer Vision and Image Understanding* 78.1 (2000), pp. 138–156.

[52] Borom Tunwattanapong et al. "Acquiring reflectance and shape from continuous spherical harmonic illumination". In: *ACM Transactions on graphics (TOG)* 32.4 (2013), p. 109.

[53] Jiaping Wang et al. "Modeling anisotropic surface reflectance with example-based microfacet synthesis". In: *ACM Transactions on Graphics (TOG)*. Vol. 27. ACM, 2008, p. 41. URL: http://dl.acm.org/citation.cfm?id=1360640 (visited on 09/08/2015).

[54] Ting-Chun Wang et al. "SVBRDF-Invariant Shape and Reflectance Estimation From Light-Field Cameras". In: 2016, pp. 5451–5459. URL: https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Wang_SVBRDF-Invariant_Shape_and_CVPR_2016_paper.html (visited on 10/28/2017).

[55] Gregory J. Ward. "Measuring and modeling anisotropic reflection". In: *ACM SIGGRAPH Computer Graphics* 26.2 (1992), pp. 265–272.

[56] Sylvia Wenzl. *Level Camera - Picture Series - Crunchy ByteBox*. 2013. URL: http://www.crunchybytebox.de/app.php?id=levelcamera (visited on 01/13/2018).

[57] Zexiang Xu et al. "Minimal BRDF Sampling for Two-shot Near-field Reflectance Acquisition". In: *ACM Trans. Graph.* 35.6 (Nov. 2016), 188:1–188:12. ISSN: 0730-0301. DOI: 10.1145/2980179.2982396. URL: http://doi.acm.org/10.1145/2980179.2982396.

[58] Su Xuey et al. "Image-based Material Weathering". In: *Computer Graphics Forum*. Vol. 27. Wiley Online Library, 2008, pp. 617–626. URL: http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8659.2008.01159.x/abstract (visited on 08/17/2015).

[59] Zhengyou Zhang. "A flexible new technique for camera calibration". In: *IEEE Transactions on pattern analysis and machine intelligence* 22.11 (2000), pp. 1330–1334.

[60] Zhenglong Zhou, Zhe Wu, and Ping Tan. "Multi-view photometric stereo with spatially varying isotropic materials". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2013, pp. 1482–1489. URL: http://www.cv-foundation.org/openaccess/content_cvpr_2013/html/Zhou_Multi-view_Photometric_Stereo_2013_CVPR_paper.html (visited on 02/09/2016).

[61] Zhiming Zhou et al. "Sparse-as-possible SVBRDF Acquisition". In: *ACM Trans. Graph.* 35.6 (Nov. 2016), 189:1–189:12. ISSN: 0730-0301. DOI: 10.1145/2980179.2980247. URL: http://doi.acm.org/10.1145/2980179.2980247.

[62] T. Zickler et al. "Reflectance sharing: predicting appearance from a sparse set of images of a known shape". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28.8 (Aug. 2006), pp. 1287–1302. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2006.170.