Assisted Color Acquisition for 3D Models

Daniel Coutinho, Ricardo Marroquim

Computer Graphics Lab, Federal University of Rio de Janeiro

Matteo Dellepiane, Roberto Scopigno

Visual Computing Lab, ISTI - CNR

Abstract

Capturing surface appearance precisely is paramount for modeling realistic materials. Nevertheless, the spatially varying nature of most materials is difficult to measure. State-of-the-art methods often rely on complex apparatus and controlled environments, and even if they are able to acquire reliable SVBRDFs, the whole process usually takes a long time and generates a large amount of data, that is often redundant.

In this work we propose a method for fast and assisted acquisition of material properties on-site. The system has a simple setup, requiring only a generic camera and a light source. Consequently, it is also very portable and appropriate for a broad range of object sizes and scenarios. The system guides the acquisition process, allowing for a fast capture session while at the same time producing high-quality per vertex diffuse colors. To help in achieving a complete coverage it suggests missing light directions, reducing the amount of necessary input images and the acquisition time. The system is designed to work *in situ*, therefore the whole acquisition process works with immediate feedback and interactive integration of new data.

We show results for a variety of objects differing in size and materials.

Keywords: Digitization and Image Capture, Reflectance

1. Introduction

The fast improvement of depth cameras and scanning devices allowed for faster geometric acquisition of objects and environments. Nevertheless, the issue of color quality of the objects has been only partially taken into account, and real-time (or near real-time) material acquisition has not received the adequate attention. A method that is robust and efficient enough for *in situ* acquisition, and that, at the same time, can handle a broad variety of materials and geometries, is still a great challenge in the field. In this work, we propose an acquisition method to narrow this gap.

Capturing the material appearance requires extra care with illumination conditions. Unfortunately, state-of-the-art acquisition methods rely on complex setups or impose severe ambient or geometric limitations in order to have a controllable light environment. Additionally, to achieve high quality results, often a massive amount of input data is acquired, and the result can only be properly visualized after a long post-processing step. To aggravate these issues, it is not uncommon during real acquisition sessions to notice only afterward that the input data is incomplete. Since no immediate feedback is usually available, there is no straightforward way to check coverage and completeness before arriving at the end of the process. In certain campaigns this may be a major problem, since another trip to the site may not be viable.

Another issue is the size of the produced dataset itself, that creates problems for distribution, operation, and storage of the data, specially when the goal is to digitize a collection of objects. Thus, compact and easy-to-use representations are more than desirable in this field.

In this work, we describe a general method for appearance acquisition that works with off-the-shelf equipments, and tackles the aforementioned issues. To use our system it is only necessary a camera, a portable light source, and the 3D geometric model of the object in question. Our assumptions are that the light source is predominant, and that the material is isotropic.

Briefly, the method works as follows. An initial camera position is chosen and calibrated in respect to the object (image to geometry alignment). Then, keeping the camera fixed, photos are acquired while placing the light source at different positions around the object. For each photo, the light direction is estimated and reflectance samples are stored per vertex. After the first few photos the system is able to suggest new light positions in order to achieve a complete coverage of the reflectance function per vertex in an efficient manner. When enough samples have been acquired for a viewpoint, the camera is moved or the object is rotated, recalibrated, and the process is repeated. At any given time it is possible to compute a fast polynomial approximation of the reflectance function per channel for each vertex, allowing for an immediate feedback about the overall acquisition progress. The main contributions and strong points of our method are:

Figure 1: Three datasets with different materials captured with our method.

The appearance acquisition process is simple and requires no special gear or setup.. This is important since many prior methods are not compatible with *in situ* digitization campaigns (such as large domes [1]). Methods that can work with off-the-shelf low cost equipments can also reach a broader audience, specially in contexts such as Cultural Heritage.

It is able to work with sparse data (only a few samples per vertex). Because the light is manually positioned, minimizing redundancy by acquiring the least amount of photos is important to reduce acquisition time.

It requires no previous knowledge about the target material. Fitting more complex BRDFs usually requires some knowledge of the target material, and often it is necessary to adjust at least a few parameters to achieve good results. A more generic, even though not as accurate, reflectance function that requires no previous knowledge or parameter setting offers an attractive advantage, specially for non-expert users. Furthermore, many BRDF models, in particularly those that model specular effects such as Fresnel, require a large amount of samples, which would go against our sparse data assumption. Some previous work, such as Lensch et al. [2] and Palma et al. [3], cluster BRDFs in order to handle this issue, but by doing so, lose the small intrinsic variations of more complex materials and significantly increase computational time.

Results are compact and can be visualized with very simple rendering shaders. Apart from a few extra per vertex parameters to represent the polynomial function per channel, no extra data is produced (such as textures). Compact representations are important in order to easily store, disseminate and share 3D models.

Immediate feedback during acquisition.. Most digitization campaigns are *in situ* and not in a lab environment. Immediate feedback is a very important feature since it allows the user to quickly evaluate the acquired data and take action where necessary, for example acquiring photos from a new view point or adding images with new light directions. Methods that rely on BRDF fitting usually require long processing times. For example, Lensch et al. [2] reported times in the scale of hours, while our polynomial fitting takes at the maximum a few seconds even for dense meshes.

We acquired the appearance of a variety of objects composed of different materials to show the robustness of the method. We believe the ensemble of the above points are unique and renders our method very useful for a broad audience of users that are not experts in appearance modeling and acquisition, that cannot afford expensive systems, that can greatly benefit from having immediate feedback of the acquisition process during the digitization campaign without long post-processing computations, and that must work in scenarios where the use of large apparatus may not be possible due to access restrictions.

The rest of the paper is organized in the following way. In Section 2 we present the works that inspired and that are most related to ours. In Section 3 we explain the required setup and give a brief overview to illustrate a typical acquisition session using our approach. Section 4 describes the main points of our method that allow us to achieve the mentioned contributions: a per vertex polynomial fit of the reflection function; and a guided acquisition approach that indicates optimal light positions to cover as many and as best as possible the vertices. To illustrate the robustness of our approach, we show and analyze results for a variety of different materials in Section 5, followed by some validations. Finally, we discuss the method's limitations in Section 6 and present our conclusions in Section 7.

2. Related work

The estimation of SVBRDFs and material properties in general became a topic of interest as soon as active acquisition devices (i.e. 3D scanners) started to reach a mature level. A complete overview of the literature in this field goes well beyond the scope of the paper. We mainly focus on approaches that share features with our proposed work. Please refer to [4] or [5] for an overview on acquisition and digital modeling of appearance. One of the first generic approaches was proposed by Lensch et al. [6]. They implemented a fitting process of the Lafortune BRDF model using only a professional digital camera, a reflecting sphere and a dark room. Instead of having a BRDF model per vertex, they segmented the mesh into clusters of BRDFs. Due to the clusterization, however, some areas may not be well represented. The authors extended the previous work by changing the calibration of the light source position and estimating normal maps in order to refine the geometric details [2]. The approach obtains accurate results, but the specificities of the setup and the amount of input needed make it impractical for a wide use. Therefore, especially for on-site acquisition, several efforts toward more applicable solutions have been made.

Simplifying the description. A first direction of research was the creation of simplified models to represent materials. Among them, the work of Malzbender et al. [7] allows for a much simpler and faster acquisition procedure. It fits a Polynomial Texture Map (PTM) by solving a linear system for N given images using singular value decomposition. They also show that it is possible to apply filter behaviors on the PTM and some lighting models such as anisotropic surfaces and Fresnel effects. Nevertheless, this approach was intended to acquire data from a single point of view in order to produce texture maps that can change their appearance with respect to the light direction. In a similar fashion, Goldman et al. [8] propose an approach based on photometric stereo to recover at the same time geometry and spatially-varying BRDFs using the isotropic Ward shading model. In the same line, the work of Alldrin et al. [9] acquires shape and BRDF simultaneously. The material is obtained from a bi-variate approximation of measured isotropic BRDFs, and the authors argue that it can represent a broader number of materials. These methods do achieve good results but rely on a very accurate acquisition, and are not trivially extensible to larger objects.

Controlling the acquisition. Another possible direction of work is to build ad-hoc devices to automatically acquire the massive amount of data needed for appearance properties estimation. Holroyd et al. [10] designed a complex coaxial optical scanner capable of synchronously acquiring shape and spatially varying reflectance using the Cook-Torrance model. Their device consists of a pair of assemblies, each containing a coaxial camera and a light source. Schwartz et al. [1] created a dome consisting of 151 DSLR cameras taking HDR sequences and one LED-Projector mounted on a tripod placed at five to eight different positions, projecting 38 different patterns. Instead of fitting BRDFs, it produces a *Bidirectional Texture Function* for the mesh. With a similar setup, the dome proposed by Nöll and colleagues [11] is also able to acquire the bottom side of objects since they are posed on a transparent surface. CultLab 3D [12] proposes an automatic modular digitization pipeline, which is able to acquire geometry and appearance at the same time. All the devices above present limitations due to their restricted portability and, more importantly, to the size of the object that they are able to handle (not more than a few tens cm).

Simplifying the acquisition. Given these limitation, research efforts have been placed in the direction of limiting the complexity of the acquisition setup and the amount of input data needed. Material properties can be inferred even from a single image, trying to classify the properties of types of materials [13], or by analyzing a single image with known geometry [14], but are limited to a single BRDF. Focusing on the acquisition of real objects, some recent works propose the acquisition of the SVBRDF of real object with very simple procedures, requiring only the light of mobile devices [15, 16] or even the screen of a laptop [17]. While they obtain very interesting results, these methods are limited to smaller and nearly planar objects.

The acquisition of larger and more demanding objects usually imply in slightly more complex acquisition setup and data. Palma et al. [3] propose a statistical method for estimating Spatially Varying BRDFs. The approach is based on video sequences with fixed but general lighting conditions. A user assisted clustering process is also performed, since in the video some regions may not have been appropriately specularly sampled. Some limitations are presented in this work due to the input data and the employed Phong model, and it may also present blur effects in some cases. In addition, the clustering step may sometimes require a significant amount of manual intervention. The work presented by Dong et al. [18] also tackles the unknown lighting conditions using a video, but in this case the object is rotated around its axis. As BRDF model they chose the isotropic microfacet model. The greatest limitation of their work is that it only handles convex geometry, and thus is not applicable to most real objects.

Ren et al. [19] proposed a portable acquisition setup that includes a BRDF chart to recover the object materials in a similar fashion as with color charts. Treuille et al. [20] extended the idea by using references of known BRDFs to reconstruct the object using multi-view stereo. These two methods depend and are limited, however, on the amount and quality of the reference BRDFs used.

Recently Wu et al. [21] presented AppFusion, an interactive material acquisition system that uses the Kinect to estimate the lighting environment and the material of the target object. The system also takes advantage of the infrared data provided by the Kinect device. While our proposed method follows a similar direction by aiming at a fast acquisition system with immediate feedback, AppFusion is unfortunately limited by the quality of the acquired data by the Kinect (especially the image resolution). The results shown are on small objects, composed by a single material.

Reducing the amount of input. A last and final interesting direction of research is to reduce the amount of input data needed to fit existing material models. The work by Ruiters et al. [22] proposes an initial effort to deal with SVBRDFs, but requires a very long processing time. In a more recent work, Nielsen et al. [23] analyze several examples of BRDFs and show that most of them can be acquired starting from a small number of samples (from five to twenty). This is a promising result, but has yet to be extended to the SVBRDF case. Our proposed work goes in the direction of acquiring an accurate material representation by minimizing the number of acquired samples. We also employ a more general model, instead of choosing a single BRDF as done by many of the related works. To the best of our knowledge, there is no other system that is as general as ours, i.e., does not impose severe restrictions to object's size, location and shape, and does not make assumptions about the material. In addition we provide immediate feedback that aids in reducing the acquisition time and avoids redundant data.

3. System setup and example of usage

The system setup, shown in Figure 2 is quite simple, and similar to the one used for the acquisition of RTI images. The camera is fixed, preferably using a tripod, in front of the object and the reflecting sphere. Regarding the illumination, the assumption is that the light must be ”directional” (same light direction for the whole object). To better approximate this effect, the light source used must cover a significant larger volume than the one occupied by the object.

Once the setup is ready, calibration is performed using a single image, in two steps. First, since the light direction has to be estimated, the position of the reflecting sphere is automatically extracted from the image, using the standard method from RTI acquisition [24]. Then, the image is aligned to the 3D model by estimating the camera parameters using Corsini et al. [25] Mutual Information method. The amount of user interaction is limited to an initial rough alignment of the model, and usually does not require more than a few seconds. Optionally, a color chart can also be used for color calibration. Even though it is not mandatory, it may help to improve the final result and achieve more natural and faithful colors. The color chart process consists merely in detecting the squares with respective colors in the image, and applying a linear regression for each color channel to calibrate the photos.

Once the image is aligned with the 3D model, and the sphere has been identified, a set of photos is acquired by displacing the light source. For each photo, the light direction is automatically detected, and the pixel colors are directly mapped onto the 3D vertices.

Figure 2: An illustration of the system setup, containing the camera, a light source, a reflective sphere and the object.

After the first few photos have been acquired, the system is able to suggest new light directions to optimize the process. Figure 3 illustrates some steps of the acquisition process, and how the suggested lights improve the fitting after each new photo. Note that a vertex may be well represented from one light direction, but lacks samples from other angular ranges. For example, the second row in Figure 3 shows an improvement on the lions breast using an approximately frontal light direction. However, the same vertices appear with arbitrary colors (due to lack of samples) when illuminated from the side, so a new light is suggested to better cover these angles, as shown in the last row of the figure.

New photos are acquired using the suggested light directions, until a good coverage is achieved. This is mainly evaluated visually by the user after fitting the polynomials. Nevertheless, this evaluation is straightforward since poorly sampled vertices are rendered with arbitrary colors and can be easily spotted, as shown on the top rows of Figure 3. At this point, a new viewpoint is manually chosen by displacing the object or the camera, and the process is restarted taking into account the data acquired from the precedent view directions. Consequently, a lower number of light directions is usually needed to cover the new view point.

4. Method

Our approach aims at guaranteeing that at least the minimum amount of information per vertex has been acquired in order to reproduce its diffuse color and an approximate specular component for each color channel. Note that our goal is not to have a minimum number of samples per vertex but a good angular coverage for each one.

For each vertex we store all projected pixels from the acquired photos. Other than the pixel intensity per color channel, we also store the product $(R \cdot E)$ for each pixel sample, that is, the dot product between the reflected light direction, and the eye vector. Even though we only use the product $(R \cdot E)$, it implicitly contains information about the normal N and the light direction L.

The next sections explain the model fitting, the mechanism to suggest new light directions, and the neighbor expansion to assign a material model to vertices that lack enough samples.

4.1. Model fitting

We chose a very simple, yet efficient, model: a cubic polynomial for each color channel. This generates 12 parameters per vertex, i.e., 4 coefficients for each channel. See Figure 4 for some exemplary fitted curves.

The motivation for the cubic polynomial comes from our established goals of visualizing objects during digitization campaigns. Sometimes acquisition time may be short, and consequently there may be a low amount of data available. For that reason, we need a material appearance model that can be obtained efficiently, does not require previous knowledge about the reflectance function and does not need an initial solution. More traditional BRDF models could be used, but they usually require dense sampling and a nonlinear optimization, which falls on the pit of a good initial solution and slower execution times. Our fitting, on the other hand, takes only a few seconds using our GPU implementation even for meshes on the order of a million vertices. Since polynomials are good models for

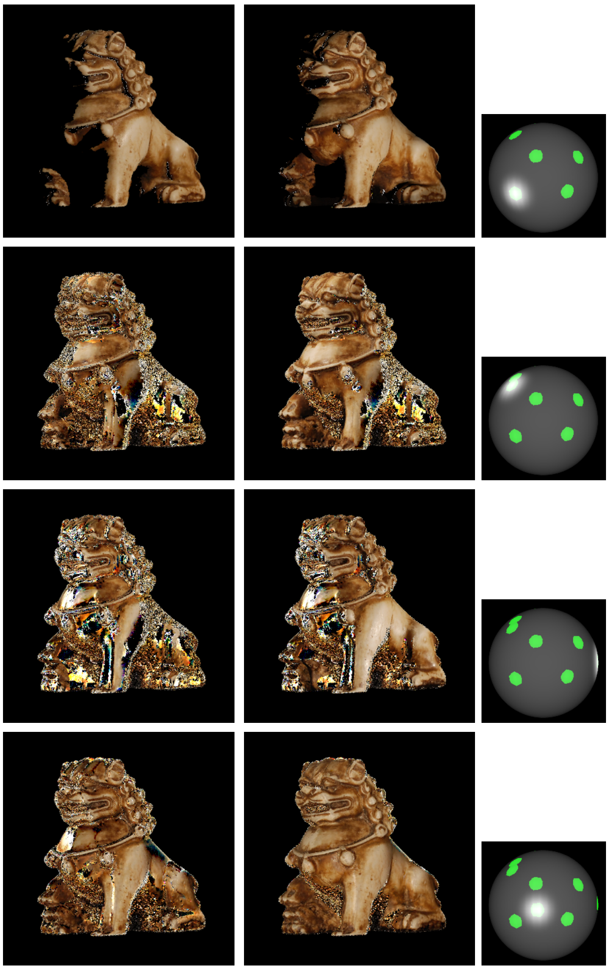

Figure 3: Top row: two input images projected directly onto the model before fitting the reflectance function. Bottom three rows: fitting the data before (left) and after (middle) a new photo is acquired from the suggested angle. Note how for each step the coverage improves significantly. Vertices with arbitrary colors do not yet have enough data, so their polynomial curve is unable to approximate the function well. The light direction is depicted on the right column, where the top row shows the five manually chosen light directions, and the bottom rows show the first three suggested light direction. The white spot is the current direction, the green spots are previously acquired directions.

low-frequency data [26], it suits our goal to guarantee a good diffuse color. As aforementioned, we are, however, restricted to isotropic materials.

An important point is that we do not divide our function into two components, such as diffuse and specular. Typically, a straightforward approach would be to plot each sample's intensity against two different products: (L · N) for the diffuse part and (R · E) for the specular part. However, as also noticed by Palma [3], specular information is particularly difficult to acquire. Only a small amount of photos cover a vertex in a specular manner since this behavior is usually concentrated around a small angular range. Fitting any interesting function with very scarce data is troublesome. Another option is to acquire more data until enough specular samples per vertex is obtained, but this would go against one of our main goals, which is to avoid an acquisition process that is excessively long. Thus, we treat all our input data equally, and try to make the most out of it in our fitting process.

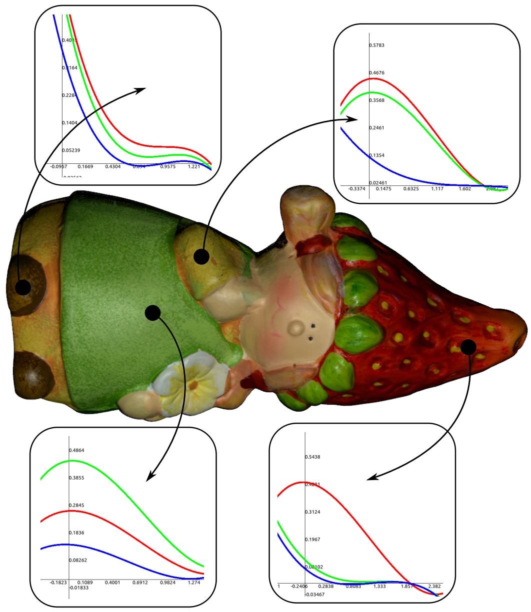

Figure 4: Exemplary fitted cubic polynomials from four different vertices of the Nana dataset. The vertical axis shows intensity, and the horizontal axis the product (R · E).

The fitting procedure is based on a simple Least-Squares polynomial fit [27], and can be applied at any time during the acquisition process, but usually at least five photos are needed to achieve any meaningful result. This can be very useful in understanding the amount and quality of coverage at any given moment during the acquisition session. If more data is needed, more photos can be acquired accordingly.

4.2. *Suggesting new light directions*

We try to reach the right balance between the amount of data gathered and the quality of the results. Hence, after acquiring the first few photos, the system suggests new light directions in order to optimize the vertex coverage. This initial number of photos is arbitrary, in the sense that during the acquisition process one can ask for a suggestion whenever needed. But in our tests we found that only around 3-5 initial photos are necessary. Furthermore, one can choose to mix the suggestions with a self-guided acquisition procedure in some cases.

4.2.1. *Vertex coverage*

To compute a vertex's coverage, we look at the distribution of its acquired samples, and analyze angular regions that are sub-sampled. As we do not know beforehand the true distribution of the function, we aim at sampling the domain as uniformly as possible. Thus, vertices with good coverage should have small maximum distance between two samples.

Note also that for each light direction, we check if the vertex is visible from the light source and discard samples that are in shadows.

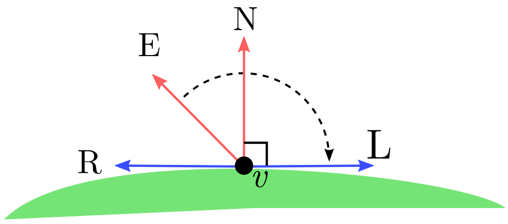

Figure 5: The maximum possible angle between L and E happens when R and N are perpendicular.

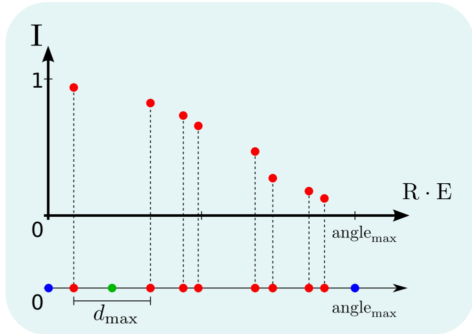

Figure 6: For each vertex we search for the largest interval between its acquired samples (red points). The green point represents the desired new optimal angle between R and E for this vertex. The blue points are anchor samples to create the first and last interval. The vertex weight to the containing bin is $d_{max}/2$.

4.2.2. *Optimal light direction*

To suggest a new light direction, vertices cast weighted votes on their preferred direction according to their current coverage. For each vertex, we search for the largest interval between two samples. First the samples are ordered by the value (R · E). We further include the two extremes $angle_{min} = 0$ and $angle_{max}$ as anchor points. $angle_{max}$ is the maximum angle between a possible light vector and the eye vector while still being visible from the light source: $\angle EN + \pi/4$, as illustrated in Figure 5.

Once the samples are ordered, we define the optimal angle α of the vertex as the medium point of the largest gap, i.e., where it most lacks samples. This process is depicted in Figure 6. Finally, to compute the actual light direction that would generate a sample at that position, the eye vector is rotated by α, and reflected back about the normal, as illustrated in Figure 7. The vertex votes on this direction with weight equal to half the length of this largest interval.

After all vertices have voted, we proceed to select the best global direction. The directions are clustered using regular angular bins. For our tests we used 36^2 bins. For each bin the weights of its contributing directions are accumulated. For the ten bins with highest total weights, we analyze their contribution to all visible vertices in the following manner. For each bin the average of the contributing light directions is used as its representative direction. The scene is rendered from the ten representative directions, and for each one, we sum the contribution of all visible vertices. Each vertex contribution is the distance from the product (R · E) using the representative direction, and the closest stored sample so far, as illustrated in Figure 8. Finally, the bin with the highest contribution is the suggested next optimal light direction.

4.3. *Sample propagation*

The coverage of the whole surface of a real object may be a difficult task, since it may be hard to frame all the portions of the surface due to self occlusions. Moreover, for some cases, it may not be possible to obtain light directions to cover the whole hemisphere around the object. Therefore, based on the assumption that the material properties are locally coherent, samples are shared between neighboring vertices in order to improve the coverage of those that lack samples.

Given a vertex, we use two metrics to spread the information around neighbors: ratio of covered photos, i.e., how many photos from the total set cover this vertex, and the standard deviation of the vertex's samples. The motivation for the first metric is very straightforward, if a vertex is covered by a small number of photos it requires more samples in order to achieve a good reflectance approximation. The second metric comes from the fact that, since we have no previous knowledge about the function, our best guess is to aim for a uniform distribution on the angle domain. Sparse sampling usually has a high standard deviation, therefore, we try to increase the standard deviation of each vertex with low ratio.

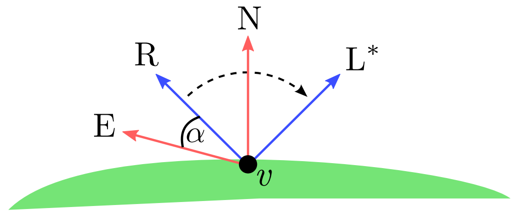

Figure 7: Once we find the candidate angle α for a vertex, we reflect back the R vector to get the corresponding optimal light direction L^*.

To propagate the samples to neighbors that lack information, the following steps are performed: for each photo that does not cover a vertex, we check if one or more neighbors are covered. If so, we choose the neighbor that most increases the standard deviation. The process is repeated until all vertices have been covered or convergence is achieved. We restrict the propagation to neighbors that have similar normals. In our tests we used a threshold of 30^o for the angle between normals, and 0.5 for the ratio metric. Figure 9 illustrates the result of the described propagation procedure.

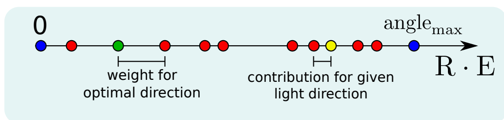

Figure 8: The sampling distribution of a given vertex. The vertex's optimal angle is depicted as the green point, which also defines its weighted contribution to the bin it falls into. The yellow point represents another bin's representative light direction, and in this case it will only add a small contribution to this vertex, since it already has a sample close to this direction.

Figure 9: The Pulcinella before (left) and after (right) the neighbor sample propagation process. Note how difficult to reach vertices inside the cloth folds had their coverage significantly improved.

5. Results and Discussions

We tested our approach on a variety of objects composed of different materials. In this section, we present our rendering system to illustrate the results, and briefly describe and comment each test case. We also provide some visual error validation method, and a limited comparison with existing methods, since only a few made their data available upon request.

5.1. Rendering

The resulting datasets were stored in the Stanford PLY format with normal and polynomial coefficients per vertex. No extra texture or color information is necessary. In comparison to a dataset with RGB diffuse and specular coefficients, plus a shininess exponent, that typically stores 6 extra floats per vertex, we store 12 extra floats but achieve a much more precise and general approximation of the reflectance function. Moreover, during our initial experiments, we noted (as expected) that fitting the Phong model using cosine functions generated an unsatisfactory fit for most datasets, specially regarding the specular component.

The figures in this paper were produced using a simple custom real-time viewer using OpenGL and GLSL shaders. Given light and view directions, for each vertex the product $(R \cdot E)$ is computed and the resulting value is used as the parameter for the three cubic polynomials. A simple shadow mapping algorithm was also incorporated only to convey a little more depth information, but apart from this, no extra effects were employed, i.e., no extra material description, global or local illumination effects, or shading functions. The wooden table that serves as a placeholder for the objects was digitized and its reflectance information was also acquired using our method. Note that we rescaled the table to better fit under each dataset, so it does not serve as a size comparison between the objects. The accompanying video was produced exclusively with this simple renderer.

5.2. Results

The geometry of the test models were acquired using different laser scanners. We also used different cameras during our tests, such as a Nikon D80 and a D5200. Simple and inexpensive spotlights were used to simulate the directional light. During an acquisition session the camera remains connected to a laptop, so the acquired photos are directly passed to our system and immediate feedback is provided. Apart from the reflection sphere and the color chart, no other extra apparatus was required.

During our tests we have simulated in the best way possible a common acquisition session, that is, we have acquired each dataset within a reasonable time, without aiming at a perfect vertex coverage, but at achieving a good trade-off between acquisition time, acquired data, and resulting dataset. Note that we could take an arbitrary large number of photos to achieve a near perfect acquisition for each model, but this would go against our main goals, as previously described. It would also mask our limitations that are discussed in the next section. As can be noted in the accompanying video some datasets have small holes or missing data in some regions, either due to lack of pixel samples or due to holes in the geometry. This may cause color artifacts or flickering in some cases, specially when the light hits the surface at grazing angles. These artifacts could actually be removed with some post-processing, but would show results that were not achieved solely during the online acquisition method, so we decided to leave them for a more fair illustration of the method capabilities and limitations.

Our method runs mostly in GPU with OpenGL shaders. Consequently, it has restrictions regarding the use of memory and processing time, especially in order to avoid GPU timeouts. Therefore, we try to balance the model size and the number of photos to avoid running into any hardware limitation. For the denser models, our limit was around 130 photos. Nevertheless, it did not severely impact the results.

Buddha. The Buddha is a small plastic statue composed of a highly specular golden paint representing the skin, and a less shiny, but still moderately specular surface composing its robe. The robe is painted in a dark red color, but there are a few physical painting artifacts, suggesting that it was actually hand painted. The hair is composed of a mostly dark diffuse surface with some golden spots. This is a particularly challenging scenario for appearance acquisition methods due to the highly specular materials. A few vertices were not sufficiently sampled, such as near the chin and its left foot. At these locations some artifacts can be noted, specially at grazing angles. Figure 10 shows some exemplary renders of the Buddha dataset.

Even though the golden painting is a rather hard material due to its intense shininess, our model was able to reliably replicate the location and general shape of the specular highlight. There is, however, a difference in specular intensity mainly due to the lack of highlight samples compared to diffuse ones, and due to the color saturation on the photos.

Book. Hardcover books usually have highly specular surfaces, as is the case with the dataset analyzed here. Furthermore, due to its predominantly flat geometry, aligning the model precisely is very challenging. In fact, due to some misalignments it is possible to note some degree of blurring in the results. On

Figure 10: Two renders of the Buddha plastic statue, painted with a challenging reflective material. Note that the red stain on its shoulder is present on the real object.

the positive side, it shows that the fitting is somewhat robust to small misalignments. Renders are shown in Figure 11.

Figure 11: A book cover made from a highly reflective material. Even with noticeable geometry-image misalignments responsible for ghosting artifacts, the reflectance functions are well approximated.

Cloth. Cloth is a challenging material to acquire properly. This small statue is basically made from cloth with small red stones attached, a black rigid mask and a dark wooden base. The main problem was covering small holes due to the cloth folds. Alignment with non-rigid objects is also challenging for obvious reasons. The scanners were not able to generate a very precise geometry for this case, aggravating the problem. An illustrative render is shown in Figure 12. Albeit these challenges, we were able to produce a very compelling appearance model for the Pulcinella statue. In this case, the information propagation was crucial to cover vertices that are hidden from most combinations of view and light directions.

Vase. This vase has a simple geometry with some carved motif. The challenging part however, is the coat of specular paint, that is an issue for many color acquisition methods. An illustrative rendering can be seen in Figure 12.

Boomerang. The boomerang is a very simple object, basically a painted flat wood surface. Only a few photos are necessary to achieve a good approximation of the surfaces reflection function. Figure 12 illustrates the resulting dataset.

Thai Lion. This small souvenir Thai statue is made of some kind of composite material. Apart for some unreached vertices, specially between the front legs, the surface appearance was well captured (Figure 13).

Nana. The Nana is a small statue composed of a very specular head and a more predominantly diffuse body. Furthermore, groups of similar colors, such as the green belly, present high local color variation, *i.e.*, many close points reflect different shades of green, which renders our per-vertex approach specifically appropriate. The result can be seen in Figure 12.

In order to demonstrate the efficiency of our light suggestion algorithm, we did not acquire this dataset using our method, but used a dataset fabricated with a mini-dome, containing 114 light directions from a fixed view direction. We let our system automatically pick 40 photos from this set to produce the polynomial per-vertex fitting. For each suggested light direction the method selected the photo with closest light direction. We also produced a result with the full dataset. Figure 14 shows that our algorithm is able to faithfully capture the appearance using only approximately one third of the whole mini-dome dataset. We expect this disparity to increase even more if more viewpoints were available, as our method would profit from the sequential information.

Bas relief. To further show the versatility of our method, we also demonstrate how it behaves when dealing with data from a Reflectance Transformation Imaging (RTI) acquisition, plus the geometric model. This wall panel has dimensions 80x50cm, and is a typical case where RTI works well, since it is predominantly flat. We show how our method enhances the RTI paradigm by producing real 3D data. A rendering of the panel can be seen in Figure 12.

In Table 1 we list a few more details about each dataset, such as number of photos and views. The timing in the fourth column is relative to the last performed fitting, that is, with all acquired photos. Note that the first fitting procedures usually take much less time.

We did not time precisely the whole acquisition session, since it depends on the experience of the user and the time spent on making decision about the coverage, among other factors. For the largest datasets, such as the Thai Lion, it took around one hour. Datasets with less view points and photos took considerably lower times. Note however, that most of the time is spent moving the light source around and calibrating new viewpoints.

5.3. Validation

One way to verify the quality of the reflectance approximation is by not using some acquired photos, and comparing the result against these control group. Figures 15 and 16 show this analysis with a color-coded difference between the photo and the render, where the scale goes from light green (low error) to red (high error). As it can be observed, most of the surface is

Figure 12: From left to right: the Pulcinella miniature offers two significant challenges, the dark mask and the cloth, that is particularly difficult to cover entirely (note that the small black holes are due to missing geometry); a terracotta vase with a coat of shiny paint that makes it specifically challenging; a painted wood boomerang acquired with a reduced number of photos; the small Nana statue composed of different surface appearances; and a marble wall panel, produced from an originally RTI dataset.

Figure 13: The Thai lion statue from two view points.

Object	# photos	# views	time(s)	# vertices
Buddha	110	6	25.91	1.440M
Thai Lion	120	12	12.08	386K
Vase	114	6	11.19	758K
Nana	114	1	9.30	1M
Book	47	4	4.46	1M
Boomerang	22	2	1.58	442K
Cloth	50	3	2.27	789K
Bas relief	105	1	2.85	400k

Table 1: Description of the acquired data. The fourth column refers to the fitting time for all views combined.

Figure 14: The result with 40 photos (left) and 114 photos (middle) from the Nana dataset. The color difference image (right) confirms that our light suggestion approach efficiently produced practically the same result as with a more dense uniform distribution of the light directions.

well represented. Higher error can be noticed mainly on highlight regions, as expected, and near shadows, though this can be due to inconsistencies in the shadow map, since our light model is very approximative.

The low error on the diffuse color support our claim that we can reproduce a good base color with a reasonable specular approximation. This is expected since polynomials are good models for low-frequency data, such as diffuse color and some wide specular lobes. Note that since our method consists of a model per vertex, even with high-frequency materials, such as the golden painting, our system can approximately replicate the specular shape on the object, but not the intensity.

Figure 15: From left to right: original image, render, color-coded difference between the image and the render. We can notice how in general the error is very low, except on some highlight regions and near shadows. The color scale for the error is depicted below the image, where red represents a high error, and light green a low error.

We also provide a visual comparison with Palma's method [3]. Figure 17 shows a render from both methods and a real image at approximately the same position. As it is noticeable, our method better approximates the real image, since it is able to fit a per-vertex function independently, while Palma's method relies on clustering and thus tends to blur the result. Their method also uses the Phong model to approximate the BRDF, which in many cases does not fit well to the real data.

We have also used an available dataset acquired with the

Figure 16: The same pattern of error can be noted on the vase. In general the error is very low, except at the center of some highlight regions.

Figure 17: A comparison between our method (left), the real photo (middle), and Palma's method, at approximately the same camera and light configuration.

Bonn Dome [1] to further evaluate our method. Since it is very densely sampled - 151 cameras with 151 light directions for each one - we have chosen a subset of the photos for our comparison. For each suggested direction by our method, we pick the closest light direction from the dataset. For this test we have picked three viewpoints facing the front of the statue, and let the system choose the best 50 photos for each one, that is, a total of 150 photos. Figure 18 shows a comparison of a photo from the original dataset and the resulting model from our method rendered from the same viewpoint and with the same light direction. Even though our method is not able to reproduce all the fine specular details, the overall reflectance behavior is well captured with just a small fraction of the original dataset.

As a visual comparison, we have taken a frame from the BTF renderer demonstration video of Schwartz et al. [1] method. Figure 19 shows that even with a much simpler acquisition methodology we can achieve comparable results, since not even very dense sampling can capture all the fine specular details that can be seen in the original image.

Figure 18: Result using a dataset acquired with the Bonn Dome. We compare a photo (left) that was not used to produce the result using our system (center), and show the color scale difference (right). There is a more significant error around the neck due to a small variation on the rendered shadow.

Figure 19: Comparison between: (left) a frame from Schwartz et al. [1] demonstration video; (middle) rendering with the model produced with our method using a reduced dataset; (right) one of the original photos from the dataset with approximately same viewpoint and light direction.

6. Limitations

Our method still depends on a reasonably fine and accurate geometric representation of the model, and it is somewhat sensitive to the camera alignment. In some cases, such as the book, the misalignment is noticeable, even though it did not greatly affect the appearance result. Nevertheless, if necessary, this issue can be corrected in a post-processing stage, followed by a simple refitting of the data.

Another issue comes from the fact that small regions not covered by enough light direction may not be included during the light directions suggestion. Nevertheless, these vertices can be spotted by visual inspection in most cases, and a new light direction can be intuitively set to complete the data. This is one of the great advantages of having a system with immediate feedback.

Our current implementation is very GPU intensive, and assumes that all data can fit in GPU memory. It thus restrains the maximum number of photos that can be used simultaneously. This restriction can be removed by using a multi-pass strategy to fit the data, where each pass would process a limited number of vertices.

The polynomial fit has its advantages, but it's probably not the best choice to model all materials. We cannot treat complex specular behaviors such as Fresnel effects, the main reason for this is the lack of high sampling density and the inability of the model to fit data with high variance. We are also limited to isotropic materials. Nevertheless, it is important to stress that the data acquired during the photographic session may be used, in a second stage, to fit other material models, possibly by clustering the main materials of the objects (like in Palma et al. [3]), or by taking advantage of the recent studies on BRDF sampling [28].

Finally, even though there is no hard restriction about the size and geometry, the object has to be illuminated by a predominant light source for the method to work properly. The manual placement of the light source may also not fit exactly with the proposed direction by the system, but we noted that this was not an issue during the acquisitions, that is, the method is able to work with a close enough direction. Occasionally the system would suggest a new light direction very close to the previous when the manually placement was too far off, but this did not impose any significant overhead to the acquisition time.

7. Conclusion and Future Works

In this work we have presented a system for interactively acquiring the material properties of an object using only a camera and a light source. Apart from having a simple setup, the method guides the acquisition by suggesting new light directions to minimize the capture effort, and avoid long acquisition sessions. Moreover, also due to the suggestion system the model achieves overall better per vertex coverage, and the sample distribution can faithfully approximate the reflection function. We have shown results for a variety of materials, such as plastic, plaster, ceramic, cloth, clay, marble and wood, and different geometric complexities.

Even without a very sophisticated model, we are able to achieve a good per vertex approximation of the reflectance function by fitting a third degree polynomial to each color channel. With an efficient implementation we provide immediate feedback, which is crucial for *in situ* digitization campaigns. We believe that our method can considerably ease the burden of quality reflectance function acquisition while keeping the acquisition system simple and portable.

In order to have immediate feedback we have also avoided costly computational processes, such as clustering as proposed by Lensch et al. [2], or more complex BRDFs non-linear fits. Recent methods such as Nielsen et al. [23] only requires a few samples to properly fit a reflectance function. Nevertheless, their method works well for acquisition of homogeneous materials since it requires sampling the surface from specific angles. This is incompatible with the acquisition of complex spatially-varying BRDFs for heterogeneous materials, where we have to deal with thousands of vertices simultaneously.

As future work, we would like to improve a few points of the system, that are not all directly related to the fitting method. The alignment, for example, could be further refined using approaches such as Dellepiane et al's [29] to avoid blurring artifacts in the final results, even if, as previously explained, this can also be corrected in a post-processing stage. Another direction is to try to reduce even more the system setup, and allow for acquisitions using a webcam for example, but our initial tests show that it produces very unreliable colors, making a precise fit an ever greater challenge.

We would also like to improve the acquisition itself by introducing more sophisticated calibrations, such as modeling the light source, and new ways to improve the vertex coverage. We have also not addressed the view point selection, which is a task left to the user in our system. Even though selecting view points to cover the object is much more intuitive than selecting light directions, an optimized set of view points would probably reduce even more acquisition times. Finally, we would like to test our system against other types of materials and its scalability with larger objects.

8. Acknowledgments

We would like to acknowledge Brazilian funding agencies CNPq for the grant of the first author and FAPERJ for the project grant of the second author. We would also like to thank Gianpaolo Palma for providing data for methods comparison.

References

[1] Schwartz C, Weinmann M, Ruiters R, Klein R. Integrated high-quality acquisition of geometry and appearance for cultural heritage. In: The 12th International Symposium on Virtual Reality, Archeology and Cultural Heritage VAST 2011. Eurographics Association; Eurographics Association. ISBN 978-3-905674-34-7; 2011, p. 25–32. URL: http://diglib.eg.org/EG/DL/WS/VAST/VAST11/025-032.pdf. doi:10.2312/VAST/VAST11/025-032.

[2] Lensch HPA, Kautz J, Goesele M, Heidrich W, Seidel HP. Image-based reconstruction of spatial appearance and geometric detail. ACM Trans Graph 2003;22(2):234–57. URL: http://doi.acm.org/10.1145/636886.636891. doi:10.1145/636886.636891.

[3] Palma G, Callieri M, Dellepiane M, Scopigno R. A statistical method for svbrdf approximation from video sequences in general lighting conditions. Comp Graph Forum 2012;31(4):1491–500. URL: http://dx.doi.org/10.1111/j.1467-8659.2012.03145.x. doi:10.1111/j.1467-8659.2012.03145.x.

[4] Dorsey J, Rushmeier H, Sillion F. Digital Modeling of Material Appearance. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 2008. ISBN 9780080556710, 9780122211812.

[5] Weyrich T, Lawrence J, Lensch HP, Rusinkiewicz S, Zickler T. Principles of appearance acquisition and representation. Foundations and Trends in Computer Graphics and Vision 2009;4(2):75–191.

[6] Lensch H, Kautz J, Goesele M, Heidrich W, Seidel HP. Image-based reconstruction of spatially varying materials. In: Gortler S, Myszkowski K, editors. Rendering Techniques 2001. Eurographics; Springer Vienna. ISBN 978-3-211-83709-2; 2001, p. 103–14. URL: http://dx.doi.org/10.1007/978-3-7091-6242-2_10. doi:10.1007/978-3-7091-6242-2_10.

[7] Malzbender T, Gelb D, Wolters H. Polynomial texture maps. In: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques. SIGGRAPH '01; New York, NY, USA: ACM. ISBN 1-58113-374-X; 2001, p. 519–28. URL: http://doi.acm.org/10.1145/383259.383320. doi:10.1145/383259.383320.

[8] Goldman D, Curless B, Hertzmann A, Seitz S. Shape and spatially-varying brdfs from photometric stereo. In: Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on; vol. 1. 2005, p. 341–348 Vol. 1. doi:10.1109/ICCV.2005.219.

[9] Alldrin NG, Zickler T, Kriegman DJ. Photometric stereo with non-parametric and spatially-varying reflectance. In: CVPR. IEEE Computer Society; 2008,URL: http://dblp.uni-trier.de/db/conf/cvpr/cvpr2008.html.

[10] Holroyd M, Lawrence J, Zickler T. A coaxial optical scanner for synchronous acquisition of 3d geometry and surface reflectance. In: ACM SIGGRAPH 2010 Papers. SIGGRAPH '10; New York, NY, USA: ACM. ISBN 978-1-4503-0210-4; 2010, p. 99:1–99:12. URL: http://doi.acm.org/10.1145/1833349.1778836. doi:10.1145/1833349.1778836.

[11] Nll T, Khler J, Reis G, Stricker D. Fully automatic, omnidirectional acquisition of geometry and appearance in the context of cultural heritage preservation. In: Scopigno R, editor. Journal on Computing and Cultural Heritage (JOCCH); 8 ed. ACM; 2015,.

[12] Santos P, Ritz M, Tausch R, Schmedt H, Monroy R, Stefano AD, et al. CultLab3D - On the Verge of 3D Mass Digitization. In: Klein R, Santos P, editors. Eurographics Workshop on Graphics and Cultural

Heritage. The Eurographics Association. ISBN 978-3-905674-63-7; 2014,doi:10.2312/gch.20141305.

[13] Schwartz G, Nishino K. Automatically discovering local visual material attributes. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015,.

[14] Romeiro F, Zickler T. Computer Vision – ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part I; chap. Blind Reflectometry. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN 978-3-642-15549-9; 2010, p. 45–58. URL: http://dx.doi.org/10.1007/978-3-642-15549-9_4. doi:10.1007/978-3-642-15549-9_4.

[15] Aittala M, Weyrich T, Lehtinen J. Two-shot svbrdf capture for stationary materials. ACM Trans Graph 2015;34(4):110:1–110:13. URL: http://doi.acm.org/10.1145/2766967. doi:10.1145/2766967.

[16] Riviere J, Peers P, Ghosh A. Mobile Surface Reflectometry. Computer Graphics Forum 2016;doi:10.1111/cgf.12719.

[17] Aittala M, Weyrich T, Lehtinen J. Practical svbrdf capture in the frequency domain. ACM Trans Graph 2013;32(4):110:1–110:12. URL: http://doi.acm.org/10.1145/2461912.2461978. doi:10.1145/2461912.2461978.

[18] Dong Y, Chen G, Peers P, Zhang J, Tong X. Appearance-from-motion: Recovering spatially varying surface reflectance under unknown lighting. ACM Trans Graph 2014;33(6):193:1–193:12. URL: http://doi.acm.org/10.1145/2661229.2661283. doi:10.1145/2661229.2661283.

[19] Ren P, Wang J, Snyder J, Tong X, Guo B. Pocket reflectometry. ACM Trans Graph 2011;30(4):45:1–45:10. URL: http://doi.acm.org/10.1145/2010324.1964940. doi:10.1145/2010324.1964940.

[20] Treuille A, Hertzmann A, Seitz SM. Example-based stereo with general brdfs. In: In European Conference on Computer Vision. 2004, p. 457–69.

[21] Wu H, Zhou K. Appfusion: Interactive appearance acquisition using a kinect sensor. Computer Graphics Forum 2015;34(6):289–98. URL: http://dx.doi.org/10.1111/cgf.12600. doi:10.1111/cgf.12600.

[22] Ruiters R, Schwartz C, Klein R. Data driven surface reflectance from sparse and irregular samples. Computer Graphics Forum (Proc of Eurographics) 2012;31(2):315–24.

[23] Nielsen JB, Jensen HW, Ramamoorthi R. On optimal, minimal brdf sampling for reflectance acquisition. ACM Trans Graph 2015;34(6):186:1–186:11. URL: http://doi.acm.org/10.1145/2816795.2818085. doi:10.1145/2816795.2818085.

[24] Mudge M, Schroer C, Earl G, Martinez K, Pagi H, Toler-Franklin C, et al. Principles and practices of robust, photography-based digital imaging techniques for museums. In: The 11th International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage. Paris, France: Eurographics Association (VAST10). ISBN 978-3-905673-76-0; 2010,.

[25] Corsini M, Dellepiane M, Ponchio F, Scopigno R. Image-to-geometry registration: a mutual information method exploiting illumination-related geometric properties. Computer Graphics Forum 2009;28(7):1755–64. URL: http://vcg.isti.cnr.it/Publications/2009/CDPS09.

[26] Shi B, Tan P, Matshushita Y, Ikeuchi K. Bi-polynomial Modeling of Low-frequency reflectances. IEEE Transactions on Pattern Analysis and Machine Intelligence 2014;36(6):1078–91.

[27] Cohen-Or D, Greif C, Ju T, Mitra NJ, Zhang HR. A Sampler of Useful Computational Tools for Applied Geometry, Computer Graphics, and Image Processing. CRC Press; 2015. ISBN 978-1-498-70628-5.

[28] Nascimento FM, de Carvalho AB, ao Andrade BT. Improving the selection of bases of brdfs for appearance preservation. In: 2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). 2016, p. 440–7. doi:10.1109/SIBGRAPI.2016.067.

[29] Dellepiane M, Marroquim R, Callieri M, Cignoni P, Scopigno R. Flow-based local optimization for image-to-geometry projection. IEEE Transactions on Visualization and Computer Graphics 2012;18(3):463–74. doi:http://doi.ieeecomputersociety.org/10.1109/TVCG.2011.75.