

A Statistical Method for SVBRDF Approximation from Video Sequences in General Lighting Conditions

Gianpaolo Palma, Marco Callieri, Matteo Dellepiane, Roberto Scopigno

Visual Computing Lab, ISTI-CNR, Pisa, Italy

Abstract

We present a statistical method for the estimation of the Spatially Varying Bidirectional Reflectance Distribution Function (SVBRDF) of an object with complex geometry, starting from video sequences acquired with fixed but general lighting conditions. The aim of this work is to define a method that simplifies the acquisition phase of the object surface appearance and allows to reconstruct an approximated SVBRDF. The final output is suitable to be used with a 3D model of the object to obtain accurate and photo-realistic renderings. The method is composed by three steps: the approximation of the environment map of the acquisition scene, using the same object as a probe; the estimation of the diffuse color of the object; the estimation of the specular components of the main materials of the object, by using a Phong model. All the steps are based on statistical analysis of the color samples projected by the video sequences on the surface of the object. Although the method presents some limitations, the trade-off between the easiness of acquisition and the obtained results makes it useful for practical applications.

Categories and Subject Descriptors (according to ACM CCS): Computer Graphics -Three-Dimensional Graphics and Realism [I.3.7]: Color, shading, shadowing, and texture—Image Processing and Computer Vision - Digitization and Image Capture [I.4.1]: Reflectance—

1. Introduction

The acquisition of the surface appearance, together with the reconstruction of the 3D geometry, is a fundamental step for the photo-realistic rendering of real objects. Although several methods have been proposed for the acquisition of the SVBRDF of objects with a complex geometry, they all have some limitations: the use of specialized instruments (dome, special setup for camera and light source, etc.); the difficulty to extend to objects of arbitrary shape and size; the need of highly controlled acquisition environments, such as a dark room. These are difficult to reproduce in the case the environment in which the object is placed cannot be modified, and the object cannot be moved (museums, outdoor locations, etc.).

We introduce a new method for an approximate SVBRDF reconstruction that simplifies the acquisition setup for two main aspects: it uses a single and standard camcorder to acquire video streams; it performs the acquisition in a general lighting environment without the need of special lighting conditions. Our method takes in input video sequences and estimates the surface appearance of the object of interest. The videos is acquired with fixed and unknown lighting con-

ditions, moving a Low Dynamic Range (LDR) video camera around the object. Starting from the alignment of the video streams on a previous acquired 3D geometry of the object, our target is to extract as much information as possible from the video frames to recover the illumination conditions and the SVBRDF. The estimation is based on statistical analysis of the redundant video color data which is projected on the geometry.

Our contribution is a new method to estimate the SVBRDF of the object with the following features:

- an easy acquisition step for fixed but general lighting conditions;
- the approximation of the acquisition environment map with a weighted accumulation approach using the object itself as a probe, that takes advantage from a temporal coherent multi-view acquisition;
- a statistical method for the estimation of SVBRDF from the video color data projected on the geometry, taking advantage of the temporal coherence and the data redundancy of a video sequence and using a straightforward solution without time-consuming iterative and optimization procedures.

2. Related Work

A criterion to classify the image-based methods proposed to acquire the SVBRDF is the density of the input data in term of angular sampling of the view and light directions.

The multi-view photometric stereo methods recover the reflectance using images captured from different view and light directions. The Bidirectional Texture Functions (BTF) [DNvGK97] are one of first solutions proposed. Their acquisition requires a complex dedicate device working in a highly controlled measurement environment, that allows a dense sampling of the light and view directions to estimate the reflectance of planar samples. This acquisition method is time consuming and data-intensive and, although several solutions have been proposed to reduce the acquisition time and the size of the final reflectance representation (see [MMS*04, FH09] for a complete overview), it cannot be extended in an easy way to objects with a complex shape. Debevec et al. [DHT*00] propose a specialized device setup for the relighting of the human face. Following extended by Wenger et al. [WGT*05] with the possibility to use time multiplexed illumination to speed-up the acquisition, the method requires a careful calibration and a data intensive acquisition. Schwartz et al. [SWRK11] use a complex dome setup to acquire the BTF of objects with a complex shape, that imposes strong limitations on the size of the object.

Several strategies have been proposed to reduce the amount of the data needed to acquire a SVBRDF. A class of methods is based on polarized light to separate specular and diffuse reflectance, by using spherical gradient illumination [MHP*07], second order spherical gradient illumination [GCP*09], or uniform spherical field of circularly polarized illumination [GCP*10]. A new process is presented in [GFT*11] for multi-view face capture. The key aspect is a new pair of linearly polarized lighting patterns which enables multi-view diffuse-specular separation under a given spherical illumination condition from just two photographs. In general, all these methods require a specialized light dome with an accurate orientation of the polarizing filter of the camera.

An alternative approach to reduce the data to acquire is to fit a BRDF model for each surface point using the redundancy of the points that share the same reflectance. Lensch et al. [LKG*03] propose an acquisition in a measurement lab with highly controllable lighting conditions. They apply a fitting process that clusters the acquired samples into groups of basis BRDFs that are used in a linear interpolation step to model the final BRDF of each texel. Zickler et al. [ZERB05] increase the angular accuracy of a spatially-varying reflectance function exploiting the high spatial resolution that images provide to fill the holes between sparsely observed view and lighting directions. It is assumed that reflectance varies smoothly in space. Holroyd et al. [HLZ10] present a novel optical setup for measuring

the 3D geometry and the SVBRDF. The basic building block is formed by a digital camera and a high frequency spatially-modulated sinusoidal light source aligned to share the focal point and the optical axis. Using at least two of these assemblies, a sequence of images of the object from different viewpoints is acquired under time-varying sinusoidal illumination. The spatially varying reflectance is modeled as a low-dimensional subspace spanned by a small set of basis BRDFs to share the data among different points that cover the same material. Haber et al. [HFB*09] propose an approach to recover the reflectance of a static scene with known geometry from collection of images taken under unknown and variable illumination. First, they use an all frequency relighting framework, based on a wavelet representation of the visibility and of the current estimation of illumination and scene reflectance, to render the scene using the triple-wavelet product integral. Therefore they employ an iterative optimization to estimate the illumination given the scene reflectance and vice versa.

Another class of solutions is represented by the multi-view methods that use images taken from different viewpoints with a fixed light direction. Nishino et al. [NZI01] propose a method to estimate the reflectance parameters and the illumination distribution from a set of sparse images. They compute three steps: the recovering of a view-independent reflectance map; the initial estimation of the illumination distribution by shooting the residual images along the perfect mirror direction; the iterative refinement of the illumination distribution and the surface reflectance parameters. This method assumes that the specular reflectance is uniform over all the object. Yu et al. [YWAC06] present a method to resolve the texture-illumination ambiguity using the view-dependent component of the surface reflectance. The output radiance of the object surface is arranged in a 3D tensor that can be decomposed in three unknown components: the illumination, expressed as combination of spherical harmonics; the light transport, that is the joint effect of the basis BRDFs and the surface geometry; the texture albedo. All the unknown variables are estimated at the same time by solving a system of bilinear equations obtained from a factorization of the radiance tensor. The reconstruction of the illumination environment map through spherical harmonics decomposition is used in other fields, as the multi-view stereo shape refinement ([WWM11, WVL*11]). In our method, for the estimation of the environment map, we extend the method proposed in [NZI01] with a weighed accumulation approach that takes advantage of the temporal coherent multi-view acquisition of a video sequence.

On the opposite side of the multi-view methods there are the photometric stereo solutions where the reflectance function is recovered using different images taken from the same viewpoint but with different light directions. The Polynomial Texture Maps [MGW01] are an example of image-based compact representation of photometric stereo data. Starting from the observation that many objects can be decomposed

into a small number of materials, Goldman et al. [GCHS05] present a method for the recovering of shape and SVBRDF. They set an iterative system initialized using the Lambertian photometric stereo data and then they compute the BRDF parameters of the basis materials, the per-textel normals and material weights, and the 3D surface using the normals. A similar approach is used in [AZK08] with a more general model based on a novel bi-variate approximations of isotropic BRDF.

A number of solutions have been proposed for the photometric stereo acquisition of flat target moving a linear light source [GTHD03, WZT*08]. Ren et al. [RWS*11] propose a portable setup device with an additional element: a BRDF chart, a small card with a set of known BRDF reference tiles. The light responses from the chart tiles as well as from the points on the target are acquired and then matched with dynamic time warping techniques to recover the target's appearance. The idea of using some reference objects of homogeneous reflectance and known geometry to reconstruct the appearance of a target object without any knowledge about the view and the light directions was introduced in [HS03]. Treuille et al. [THS04] extend the method to reconstruct the whole object from multi-view data. The main drawback of these methods is the need to have reference BRDFs which are similar to the target's reflectance. To relax this requirement, some methods perform a two-step acquisition on the object. To obtain the surface reflectance of large outdoor scenes, Debevec et al. [Deb04] measure a set of representative BRDFs from small regions of the scene using controlled lighting, as well as images of the entire scene under natural lighting. This approach fails with surfaces with an high specular reflectance. Dong et al. [DWT*10] present an another two-pass method. The first phase acquires some representative BRDF samples sparsely distributed over the surface with a dense angular sampling using a fast single-point BRDF measurement device. The second phase acquires a sparse angular sampling of the appearance of each target point with a set of images to constrain the final reconstruction of their reflectance as a linear combination of the nearest representative BRDFs.

3. System Overview

The input of our algorithm is a triangulated mesh of a real-world object with an associated parameterization and some LDR videos, which are acquired moving the camera around the object. We assume that the camera has manual settings, with fixed exposure and fixed white balance. The output is assumed to be in sRGB color space. The algorithm can work with complex illumination environments composed by different kind of light sources (e.g point/area light, positional/directional light), that have a fixed position and a nearly constant intensity during the acquisition. To model the surface appearance, we use a simple spatially varying Phong model, where for each texel (x, y) of the mesh parameterization we have the sum of a Lambertian diffuse compo-

nent $D_{x,y}$ with a specular lobe $S_{x,y}$:

$$\rho_{x,y}(\vec{v}, \vec{l}) = D_{x,y}(\vec{l}) + S_{x,y}(\vec{v}, \vec{l}) = \rho_d(\vec{l} \cdot \vec{n}) + \rho_s(\vec{l} \cdot \vec{r})^\alpha \quad (1)$$

where \vec{v} is the view direction, \vec{l} is the light direction, \vec{n} is the surface normal, \vec{r} is the mirror reflected vector of the view direction with respect to \vec{n} , the ρ_d is the color of diffuse component, ρ_s is the intensity of the specular highlight and α is the shininess of the highlight. The output of the algorithm is the diffuse texture map, the clustering of the texels in a set of basis materials and the specular parameters ρ_s and α for each basis material. The algorithm is composed by the following steps:

1. Registration of the videos over the mesh (Section 4)
2. Estimation of the environment map (Section 5)
3. Estimation of the diffuse color per texel (Section 6)
4. Estimation of the specular parameters per material (Section 7)

The assumptions of the algorithm are:

- the lights have almost the same color, not necessarily white (this is generally corrected by the white balance procedure available on the camera);
- the lights have almost the same intensity in terms of order of magnitude, with the consequence that the estimated diffuse color includes the light intensity;
- the possibility to cluster the object's materials with an user-assisted method starting from an automatic segmentation of the diffuse color. We use a generic growing regions method, but any other approach can be applied to solve more challenging cases.

Even if the first two assumptions are very strong, these conditions are quite common in the real world acquisition environments, like building rooms, museums or outdoor locations.

3.1. Visibility Approximation

In all the steps of the algorithm we need a fast way to compute the visibility along a given direction for each object surface point to take into account the effects of self-occlusion and self-shadowing. For this reason we precompute a spherical harmonics approximation of the visibility function $V : \vec{\omega} \in \Omega \rightarrow \{0, 1\}$. In the specific, we approximate the function for each vertex i with $\tilde{V}_i(\vec{\omega}) = \sum_{l=0}^5 \sum_{m=-l}^l k_i^{(l)(m)} Y^{(l)(m)}(\vec{\omega})$, using the first 6 bands of spherical harmonics $Y^{(l)(m)}$ (36 coefficients $k_i^{(l)(m)}$). The visibility value $\tilde{V}_{x,y}(\vec{\omega})$ of each texel (x, y) is obtained by barycentric interpolation among the values of the triangle vertices.

4. Video-to-3D Geometry Registration

The registration of the videos over the mesh is obtained using the algorithm presented in [PCD*10]. The output of the

registration algorithm is the perspective camera projection matrix of each video frame. These matrices allow to recover the set of color samples $C_{x,y} = \{I_{x,y}^{(j)} \in RGB\}$ projected by each frame j over the texel (x,y) . For each color sample, we compute a quality value $q_{x,y}^{(j)}$ (see figure in the additional materials) that is the product of three measures (normalized in the range $[0, 1]$): the distance in image space from the nearest depth discontinuity, to penalize the wrong color sample due to small misalignment; the depth of the texel in camera space, to assign a higher quality to the pixels that are sampled by closer views; the dot product between the view direction and the surface normal. Please refer to [CCCS08] for a formalization of the per-pixel quality function.

5. Environment Map Reconstruction

For the estimation of the environment map, we use an accumulation approach based on a statistic separation of the color samples that could have a specular behavior. Since we do not have ideal working conditions (residual geometric and registration errors, a camera with a non-linear response and with noise), that would allow to assign as diffuse color the sample with the minimum luminance, we compute a luminance threshold to separate, for each texel, the specular and the diffuse samples. The first ones are used in the estimation of the environment, the last ones in the computation of the diffuse color.

In a multi-view dataset, the computation of the median of color samples is a good choice for the separation between specular and diffuse components. Our choice is to compute for each texel (x,y) a luminance threshold $t_{x,y}$ that is an upper bound of the median (see Appendix A in the additional materials) and is equal to the sum of mean and absolute deviation of the luminance $Lum(I^{(j)})$ of the color samples $C_{x,y}$ projected on the texel. To be more robust in the computation of this threshold, we discard all the color samples with a quality lower than 0.1. All the color samples that have a luminance above the threshold $t_{x,y}$ are used in the estimation of the environment map. For each element in $A_{x,y} = \{I^{(j)} \in C_{x,y} \mid Lum(I^{(j)}) > t_{x,y}\}$, we compute the specular mirror direction \vec{r} of the view vector with respect to the texel surface normal and then we accumulate an amount $a_{x,y}$ in the environment map along this direction. The value $a_{x,y}$ is equal to the difference of the sample luminance from the threshold $t_{x,y}$ multiplied by the sample quality $q^{(j)}$ and the visibility along the direction \vec{r} .

$$a_{x,y} = (Lum(I^{(j)}) - t_{x,y}) \tilde{V}_{x,y}(\vec{r}) q^{(j)} \quad (2)$$

The computation of the value $a_{x,y}$ and the availability of a multi-view acquisition give us more robustness with respect to possible geometric and registration errors. Then, the map is normalized with the distribution of all color samples along the specular mirror direction in the environment. This distribution is computed as the total number of the color samples in $C_{x,y}$ that project in each specific pixel of the environment

map. This normalization gives us more robustness with respect to the camera movement, in term of temporal and spatial density of the acquisition. In this construction, we assume that a specularity is located where the object behaves like a mirror and, even if the mirror specular direction does not entirely match with the real specular direction, it is usually very close. Finally, the environment map is normalized in the range $[0, 1]$. The result is a probability map depicting the position of the lights in the environment. In this step the object itself is used as a probe without the need to put any other usual light probe objects (e.g. spheres and parabolic mirror) in the scene. Although the object does not have optimal features in term of angle coverage and uniform sampling of the space, the results obtained show that this approach is sufficient to describe the lighting environment with enough detail and precision (Figure 4).

6. Diffuse Color Estimation

The estimated environment map allows us to compute a new weight for each color sample, the specular weight, which represents the probability that the sample has a specular behavior. This weight is useful for the estimation of the diffuse color. It is computed by directional sampling of the environment map in a cone of directions along the specular mirror direction \vec{r} , followed by a normalization that depends on the luminance value of the samples. We use a cone of direction to take into account two problems: materials with a partially off-specular behavior; the geometric and registration errors. Given a set of 16 directions $B = \{\vec{\omega} \in \Omega \mid \vec{\omega} \cdot \vec{r} < \gamma\}$ distributed around \vec{r} with a stratified disk sampling strategy, the specular weight $s_{x,y}^{(j)}$ for each color sample is:

$$s_{x,y}^{(j)} = \frac{1}{w} (b_{x,y}^{(j)})^{w^2} \quad (3)$$

with:

$$b_{x,y}^{(j)} = \frac{\sum_{\vec{\omega} \in B} g(\vec{\omega} \cdot \vec{r}) \tilde{V}_{x,y}(\vec{\omega}) EnvMap(\vec{\omega})}{\sum_{\vec{\omega} \in B} g(\vec{\omega} \cdot \vec{r}) \tilde{V}_{x,y}(\vec{\omega})} \quad (4)$$

$$w = \frac{t_{x,y} - \min_{x,y}}{Lum(I_{x,y}^{(j)}) - \min_{x,y}} \quad (5)$$

where the function $EnvMap(\vec{\omega})$ returns the value of the environment map along the direction $\vec{\omega}$, g is a Gaussian function with standard deviation $\sigma = \gamma = \cos 5^\circ$ to weight the contribution of each direction with respect to the distance from the direction \vec{r} , $\min_{x,y}$ is the minimum luminance value projected on the texel and $Lum(I_{x,y}^{(j)})$ is the luminance of the current sample. The term $b_{x,y}^{(j)}$ is the weighed average of the values gathered from the environment map along the directions in B , and it represents the probability that there is a light source around the specular mirror direction. The term $\frac{1}{w}$ and the exponent w^2 are two texel-aware correction factors, which depend on the distribution of all color samples

around the threshold $t_{x,y}$ and the position of current sample in this distribution. These terms provide more robustness with respect to the real reflectance behavior of the sample, detecting the false positive cases (samples with a diffuse behavior that are geometrically aligned with an area of the environment map associated to a strong light source).

For the computation of the diffuse color, we adapt the same approach used in the estimation of the environment map. For each texel we compute a threshold $d_{x,y} = \mu_{x,y} + \bar{\sigma}_{x,y}$, where $\mu_{x,y}$ and $\bar{\sigma}_{x,y}$ are the weighted mean and the weighted absolute deviation with the specular weights $s_{x,y}^{(j)}$. The final diffuse color $D_{x,y}$ is equal to the weighted average of the color samples with a luminance lower than the threshold $d_{x,y}$:

$$D_{x,y} = \frac{\sum_{I^{(j)} \in A} q^{(j)} I^{(j)}}{\sum_{I^{(j)} \in A} q^{(j)}} \quad (6)$$

where $A_{x,y} = \{I^{(j)} \in C_{x,y} \mid Lum(I^{(j)}) < t_{x,y}\}$ and $q^{(j)}$ is the quality of the samples. The weighted average of the sample in $S_{x,y}$ is justified by the need to balance out the not ideal acquisition conditions (geometric and registration error, camera with noise and not linear response).

7. Specularity Estimation

The final step of the system is the estimation of the specular behavior of the object. Starting from the observation that the majority of the real objects have a specular component which is less spatially varying than the diffuse color, we assume that the object can be segmented in a set of basis materials, each one with a different specular reflectance. For this purpose, we complete the following steps:

1. Detection of the main light sources in the environment map;
2. Estimation of the specularity parameters per texel;
3. Clustering of the basis materials of the object based on the diffuse color;
4. Estimation of the specularity parameters per cluster.

The main idea under this step is that we can use the slope of a temporal coherent luminance peak to estimate the two specular parameters ρ_s and α (see Section 3) for each basis material.

The first task is the estimation of the position and shape of the main light sources in the scene by clustering. We start by approximating the environment map with a set of 4096 directional lights L , using a median cut algorithm [Deb05]. Then, we discard all the lights that cover an area of the environment greater than a tile of a sphere subdivided in 4096 uniform area regions ($4\pi/4096$). Finally, we cluster all the lights that create a connected component with their areas in the map. Each cluster represents a light sources in the scene. For each cluster, we compute the centroid and the local distribution around the centroid to have a rough estimation of two important features of the illuminant: the position \vec{l} and the shape.

The centroid is computed using the image moment of the area on the map covered by the cluster. For the local distribution, we compute the covariance matrix Σ_l of the directional lights in the cluster with respect to the centroid, using the projection on the tangent plane of the centroid. The covariance matrix is used to approximate the distance of a direction from a light source using the Mahalanobis distance. We assume that all the lights have the same intensity (Figure 3). This assumption is problematic only when the environment has a very large dynamic range that does not allow to observe all the main features of the object appearance.

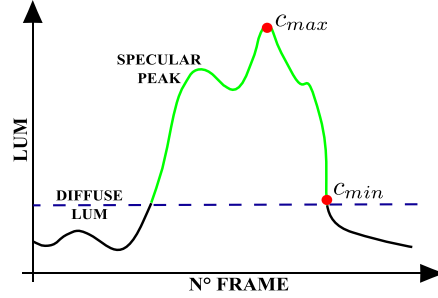


Figure 1: Temporal trend of the texel's luminance. The figure highlights the luminance of the diffuse color, the temporal coherent luminance peak (green line), and the color samples c_{min} and c_{max} used in the estimation of the specular parameters per texel.

The second step is the estimation of the specular parameters ρ_s and α for each texel. We model the specularity using only the luminance channel data, assuming that the lights are white. In the specific, for each texel we detect the temporal coherent luminance peaks, that is a consecutive sequence of color samples in the video that have a luminance greater than the luminance of the diffuse color $D_{x,y}$. Then, for each peak we select the two samples c_{min} and c_{max} with the minimum and the maximum luminance difference from the diffuse color (Figure 2). In this process, we discard all the samples with a luminance value higher than 98 (out of 100) to be more robust with respect to the saturation of camera CCD sensor. Following, we assign to each luminance peak the main light source $\vec{l} \in L$ that should produce it, which is the light source with the lower Mahalanobis distance from the reflected view direction \vec{r}_{max} of the maximum luminance sample c_{max} . The use of the Mahalanobis distance allows to take into account of the illuminant shape. With this data, we can compute the parameters ρ_s and α solving the following system of equations (see Appendix B in the additional material):

$$\begin{cases} \rho_s (\vec{l} \cdot \vec{p}_{max})^\alpha = Lum(c_{max}) - Lum(D_{x,y}) \\ \rho_s (\vec{l} \cdot \vec{p}_{min})^\alpha = Lum(c_{min}) - Lum(D_{x,y}) \end{cases} \quad (7)$$

where \vec{p}_{max} and \vec{p}_{min} are the directions that minimize the dot product with the light direction and have the same Maha-

lanobis distance of the mirror reflected view directions \vec{r}_{max} and \vec{r}_{min} (Figure 3):

$$\begin{aligned} \vec{p}_{max} &= \arg \min_{\vec{v} \in A} (\vec{l} \cdot \vec{v}) \\ \vec{p}_{min} &= \arg \min_{\vec{v} \in B} (\vec{l} \cdot \vec{v}) \end{aligned} \quad (8)$$

with:

$$\begin{aligned} A &= \{\vec{\omega} \in \Omega \mid d_{Mahal}(\vec{\omega}, \Sigma_l) = d_{Mahal}(\vec{r}_{max}, \Sigma_l)\} \\ B &= \{\vec{\omega} \in \Omega \mid d_{Mahal}(\vec{\omega}, \Sigma_l) = d_{Mahal}(\vec{r}_{min}, \Sigma_l)\} \end{aligned} \quad (9)$$

In this process, we discard all the peaks where the Mahalanobis distance d_{Mahal} of the sample c_{min} is less than the Mahalanobis distance of the sample c_{max} . If a texel has more than one peak, we choose the one with the minimum Mahalanobis distance from its main light.

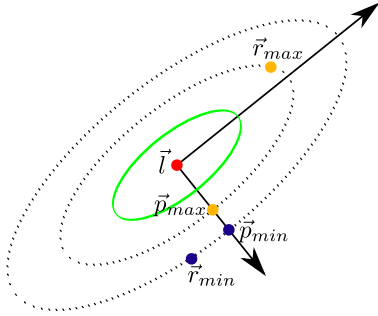


Figure 2: Characterization of a light source with the direction \vec{l} and the local distribution defined from the covariance matrix Σ_l (green ellipse). The characterization allows to project the mirror reflected view vectors \vec{r}_{max} and \vec{r}_{min} on the direction of minimum variance (\vec{p}_{max} and \vec{p}_{min}) defined by the covariance matrix Σ_l

Due to the fixed lighting conditions and the limited view directions covered with the video sequences, it is possible that some texels do not have enough data to estimate the specular parameters. To overcome this problem and to obtain a more consistent and robust estimation of the specular parameters, we employ a user-assisted method to create clusters of points with the same specular behavior. This clustering is based on the diffuse color of the texels. The main idea is to create a new cluster for each area of the object with a perceived uniform diffuse color that has a singular specular behavior. The clustering starts with an automatic step that creates an initial material segmentation. It distributes a set of seed points in the areas with the most uniform diffuse color and then applies a growing region for each of these seeds. A new point is added to the cluster if the distance of its color from the mean color of the cluster is below a threshold selected by the user. The distance between two colors is computed as euclidean distance in the CIE LAB color space. Starting from this initial segmentation, a simple interactive application allows to select a cluster by picking a point on the object and to apply two different operations: to merge a set of clusters; to

split a cluster along a stroke drawn by the user over the object. These operations allow to solve to different challenging cases: over-segmentation of an area with uniform material; separation of adjacent areas with the same diffuse color and different specular behavior.

The final step is the estimation of the specular parameters for each cluster, using the data computed for the texels. The most challenging parameter to estimate in this step is ρ_s due to several reasons: the clamped luminance signal returned by the LDR video camera; the highly spatially-varying nature of the diffuse color of the texel in the cluster; a bad clustering of the texel especially on the boundary among the clusters. To overcome this problem, we employ a statistical analysis of the ρ_s values of the texels for each cluster independently. We compute the probability density function of the ρ_s values in the cluster using a kernel density estimation density function where we look for the influence area of the higher mode, defined by its statistical bell. Finally, we compute the average of the specular parameters ρ_s and α for all the texels in the cluster that have a ρ_s value inside this statistical bell.

8. Results

For our experiments we used four different objects of different materials and different reflectance behaviors:

- the DWARF, a terracotta statue (30cm tall) that presents different types of specularities, in size and intensity: sharper and with a high-medium intensity on the dresses; wider on the face; almost completely absent on the beard;
- the GNOME, a ceramic statue (15cm tall) that has very sharp and high specularities on the hat and a near diffuse behavior on the body;
- the HEAD, a bronze copy of the head of the Arringatore Etruscan statue (30cm tall);
- the SLEEPING BUDDHA, an acrylic resin Buddha (10cm tall) with different types of coatings: a gold paint on the body; a reddish specular paint on the dress; a diffuse black paint on the hair.

The video sequences for the HEAD and the SLEEPING BUDDHA present a more uniform sampling of the view direction, while the sequences for the DWARF and the GNOME are more sparse. The 3D models were generated by 3D scanning using a Konica Minolta VI-910 laser scanner and then simplified to obtain a medium resolution model for the computation of the texture parameterization [PTC10]. In this process, we saved the normal map of the high resolution mesh to preserve the details lost during the simplification. The videos were acquired with a full HD video camera at the highest quality to reduce the compression artifact. To make the estimation of the appearance more robust, we set the camera in manual mode with a fixed white balance, defined with the automatic procedure available on the camera using a reference white object, and fixed exposure. We used

two different lighting environments for the acquisition, characterized from different type of lights: the first scenario (top-left Figure 4) is characterized by three near-circular halogen floodlights; the second scenario (bottom-left Figure 4) is characterized by six fluorescent tubes. The DWARF and the GNOME were acquired in the first scenario, while the HEAD and the SLEEPING BUDDHA were acquired in the second scenario.

8.1. Environment Map

Figure 4 shows a comparison between the real environment maps taken during the acquisition with a metal reflective sphere (left column) and the reconstruction obtained from the DWARF's and the HEAD's data (see the additional material for the environment map reconstructed from the other test cases). In general, the method does not reconstruct the entire environment map, but only the main light sources that produce a specularly on the object. More in details, in the first scenario we faithfully recovered the three main light sources, both in shape and position. In the second scenario, the method recovered only the two lights that were over the head, neglecting the other lights that are too far to produce highlights that can be recorded in the video during the acquisition.

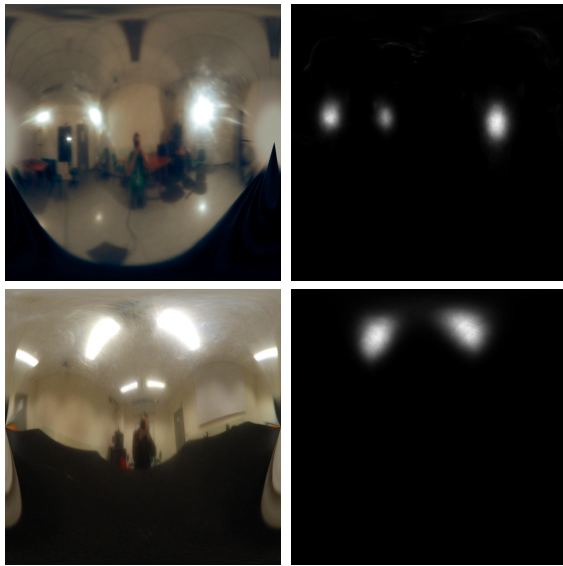


Figure 3: Comparison of environment maps: (Top-Left) real first scenario; (Top-Right) first scenario reconstructed from the DWARF's videos; (Bottom-Left) real second scenario; (Bottom-Right) second scenario reconstructed from the HEAD's videos.

8.2. Appearance

Figures 5 and 6 show a visual comparison between the rendering of the object with our SVBRDF reconstruction and



Figure 4: Comparison between an image of the object acquired in a different lighting environments (Left) and the rendering with our SVBRDF approximation (Right): (Top) HEAD; (Bottom) DWARF.



Figure 5: Rendering of the SLEEPING BUDDHA with two HDR environment maps: (Left) Uffizi Gallery; (Right) Dining room of the Ennis-Brown House.

a frame of the video used in the estimation process. From left to right we have the clustering of the basis materials, the diffuse component, the specular component (normalized in the range $[0, 1]$ to improve its visualization), the final rendering and the original video frame without background. For the rendering we used the environment maps estimated in Section 5, approximated with a set of 256 directional lights obtained through a median cut algorithm [Deb05]. The output of our algorithm is able to reproduce the main features of the object appearance maintaining the relative difference among the several materials of the object. We want to remark that our reconstruction is not an absolute SVBRDF measurement, but a simple and fast approximation to allow a photo-realistic rendering of the object. For this purpose, we assume that the video sequences have been acquired in a lighting en-

environment that permits to have a good global illumination on the object and to observe the main features of its appearance. This was obtained with limited input data, that can be acquired in a short time in an uncontrolled environment. These conditions are usually not taken into account in the context of most of the material properties acquisition methods proposed in literature.

In Figure ?? we show a comparison between a photo of the object and the rendering with our SVBRDF reconstruction. The photos were acquired in a lighting condition (a single halogen floodlight) different with respect to the acquisition one, shot with the same video camera. In Figure 4 we show the rendering of the SLEEPING BUDDHA in two different HDR environment maps using a Monte Carlo ray tracer.

Due to the input data and the Phong model, the method presents some limitations: the reconstruction of a shaded diffuse color, due to the limit of the data acquired with fixed lighting conditions; a blur effect on the very small and sharp specularity, for example on the GNOME's hat, due to small misalignment in the registration step; the different appearance or the loss of some highlights, for example over the DWARF's hat, due to the lack of the surface meso-structure, to imprecise normals, and to Fresnel effect at grazing angle that the Phong model is not able to reproduce; the final appearance is not independent from the lighting intensity of the acquisition environment. Anyway, even if there are some visible differences to ground truth, the results does not present major artifacts and complex multi-material objects can be taken into account. The final results are effective and realistic making the method useful for practical applications that need a simple way to acquire and reproduce the appearance of a real object in real-time and in a photo-realistic manner.

8.3. Performances

Table 1 shows some data about the test cases and the processing time. For each case the total number of frames, the size of the 3D model, the time of video-to-3D geometry registration obtained with the algorithm in [PCD*10], the time to reconstruct the environment map, the time to estimate the diffuse color and the specular parameters, and the number of basis specular materials are shown. We perform our test on a PC with an Intel Core i7 950 with 12GB of RAM and a NVIDIA GTX580 1536MB. The analysis of the table shows that most of the processing time stands in the registration step, while the rest of the steps is usually completed within minutes.

9. Conclusions and Future Work

We presented a statistical method for the acquisition of the SVBRDF of complex object starting from video sequences taken under uncontrolled lighting conditions. Given the video frames and a 3D model of the object, the method is able to estimate the environment map of the scene, using the

object itself as a probe and capturing enough lighting details for our SVBRDF estimation, a good approximation of the diffuse color, without view depended reflection effects, and the specular values of the basis materials.

Given the limited input data, and the very easy acquisition process, the results show that, even in the case of complex and multi-material objects, the reflectance properties are estimated with an accuracy that permits to produce very realistic renderings. Although the method presents some limitations (due to the type of input data and the specular model applied to describe the materials), the trade-off between the easiness of acquisition and the obtained results makes it extremely useful for practical applications. This is especially true when an on-the-field acquisition has been performed, and the interaction with the object and the surrounding environment is limited.

Regarding the future works, we would explore three possible directions: the possibility to use a more complex BRDF model to reproduce in a more accurate way the reflectance properties of the materials; the adaptation of our algorithm to use HDR videos acquired with a professional HDR camera, in order to obtain more data about the specular reflection; the study of the robustness of the algorithm with respect to the temporal sampling and the image resolution, testing it with photo sequences acquired with a lower frame-rate but higher resolution, moving a photo camera with continuous shooting around the object.

Acknowledgments

We thank Michael Goesele, Paolo Cignoni and Massimiliano Corsini for their useful advices and Francesco Banterle for the help with the renderings. The work presented in this paper was funded by the EC IST IP project "3D-COFORM" (IST-2008-231809).

References

- [AZK08] ALLDRIN N., ZICKLER T. E., KRIEGMAN D.: Photometric stereo with non-parametric and spatially-varying reflectance. In *CVPR* (2008), pp. 1–8. 3
- [CCCS08] CALLIERI M., CIGNONI P., CORSINI M., SCOPIGNO R.: Masked photo blending: mapping dense photographic dataset on high-resolution 3d models. *Computer & Graphics* 32, 4 (Aug 2008), 464–473. 4
- [Deb04] DEBEVEC P. E.: Digitizing the parthenon: Estimating surface reflectance properties of a complex scene under captured natural illumination. In *VMV 2004* (2004), p. 99. 3
- [Deb05] DEBEVEC P.: A median cut algorithm for light probe sampling. In *ACM SIGGRAPH 2005 Posters* (2005). 5, 7
- [DHT*00] DEBEVEC P., HAWKINS T., TCHOU C., DUIKER H.-P., SAROKIN W., SAGAR M.: Acquiring the reflectance field of a human face. In *ACM SIGGRAPH* (2000), pp. 145–156. 2
- [DNvGK97] DANA K. J., NAYAR S. K., VAN GINNEKEN B., KOENDERINK J. J.: Reflectance and texture of real-world surfaces authors. In *CVPR* (1997), pp. 151–157. 2

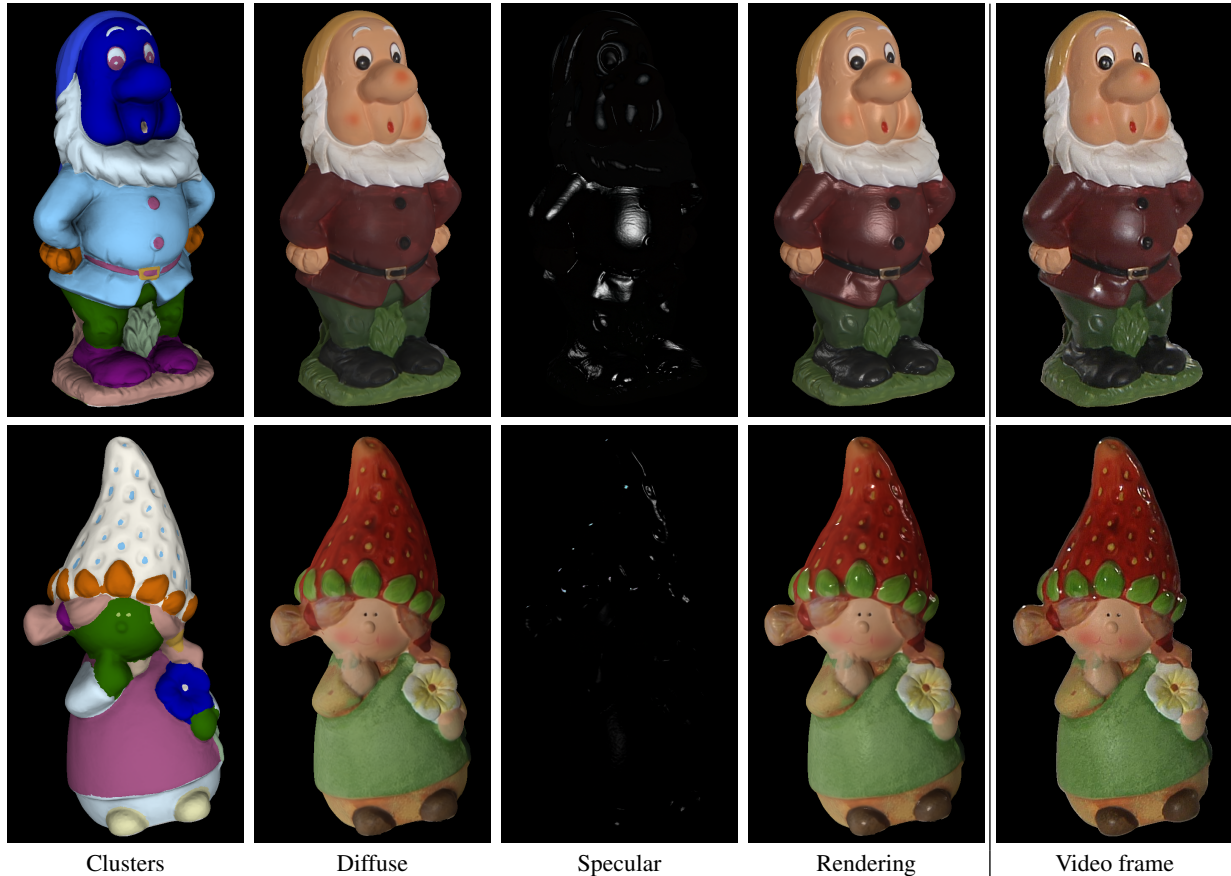


Figure 6: Results: (Top row) DWARF; (Bottom row) GNOME.

	Frames	Geometry (triangles)	Registration (mm:ss)	Env. Map (mm:ss)	Diffuse (mm:ss)	Specular (mm:ss)	N. Clusters
DWARF	3382	200k	113:00	8:11	8:05	3:05	12
GNOME	2092	135k	73:00	5:08	5:03	1:57	13
HEAD	8386	250k	265:00	25:24	25:15	8:32	1
SLEEPING BUDDHA	7240	205k	251:00	18:57	18:49	6:21	3

Table 1: Test case data

[DWT*10] DONG Y., WANG J., TONG X., SNYDER J., LAN Y., BEN-EZRA M., GUO B.: Manifold bootstrapping for SVBRDF capture. *ACM Trans. Graph* 29, 4 (2010). 3

[FH09] FILIP J., HAINDL M.: Bidirectional texture function modeling: A state of the art survey. *IEEE Trans. Pattern Analysis and Machine Intelligence* 31, 11 (Nov. 2009), 1921–1940. 2

[GCHS05] GOLDMAN D. B., CURLESS B., HERTZMANN A., SEITZ S. M.: Shape and spatially-varying BRDFs from photometric stereo. In *ICCV* (2005), pp. I: 341–348. 3

[GCP*09] GHOSH A., CHEN T., PEERS P., WILSON C. A., DEBEVEC P. E.: Estimating specular roughness and anisotropy from second order spherical gradient illumination. *Comput. Graph. Forum* 28, 4 (2009), 1161–1170. 2

[GCP*10] GHOSH A., CHEN T., PEERS P., WILSON C. A., DE-

BEVEC P.: Circularly polarized spherical illumination reflectometry. *ACM Transactions on Graphics* 29, 6 (Dec. 2010), 162:1–162:12. 2

[GFT*11] GHOSH A., FYFFE G., TUNWATTANAPONG B., BUSCH J., YU X., DEBEVEC P.: Multiview face capture using polarized spherical gradient illumination. In *ACM SIGGRAPH Asia 2011* (2011), pp. 129:1–129:10. 2

[GTHD03] GARDNER A., TCHOU C., HAWKINS T., DEBEVEC P.: Linear light source reflectometry. *ACM Transactions on Graphics* 22, 3 (July 2003), 749–758. 3

[HFB*09] HABER T., FUCHS C., BEKAER P., SEIDEL H. P., GOESELE M., LENSCH H. P. A.: Relighting objects from image collections. In *CVPR* (2009), pp. 627–634. 2

[HLZ10] HOLROYD M., LAWRENCE J., ZICKLER T.: A coaxial

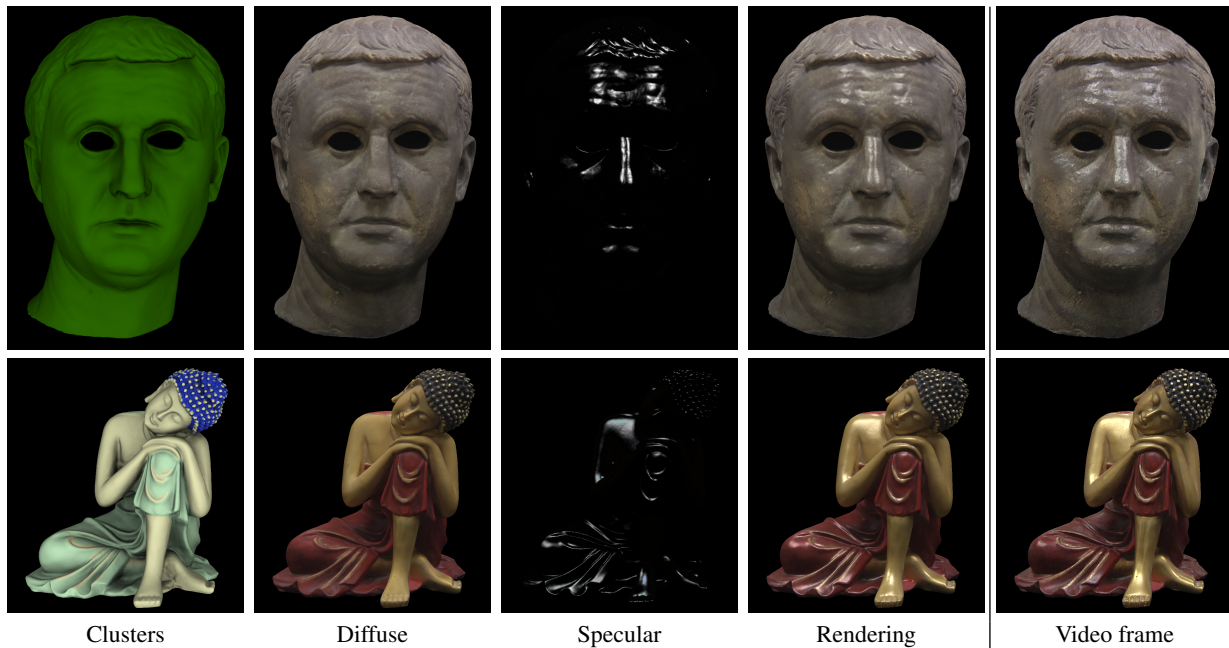


Figure 7: Results: (Top row) HEAD; (Bottom row) SLEEPING BUDDHA

- optical scanner for synchronous acquisition of 3D geometry and surface reflectance. *ACM Trans. Graph* 29, 4 (2010). 2
- [HS03] HERTZMANN A., SEITZ S. M.: Shape and materials by example: A photometric stereo approach. In *CVPR* (2003), pp. 533–540. 3
- [LKG*03] LENSCH H. P. A., KAUTZ J., GOESELE M., HEIDRICH W., SEIDEL H.-P.: Image-based reconstruction of spatial appearance and geometric detail. *ACM Transactions on Graphics* 22, 2 (Apr. 2003), 234–257. 2
- [MGW01] MALZBENDER T., GELB D., WOLTERS H.: Polynomial texture maps. In *ACM SIGGRAPH 2001* (2001), pp. 519–528. 2
- [MHP*07] MA W.-C., HAWKINS T., PEERS P., CHABERT C.-F., WEISS M., DEBEVEC P.: Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. In *EGSR* (2007), pp. 183–194. 2
- [MMS*04] MÜLLER G., MESETH J., SATTLER M., SARLETTE R., KLEIN R.: Acquisition, synthesis and rendering of bidirectional texture functions. In *Eurographics 2004, State of the Art Reports* (2004), pp. 69–94. 2
- [NZI01] NISHINO K., ZHANG Z., IKEUCHI K.: Determining reflectance parameters and illumination distribution from a sparse set of images for view-dependent image synthesis. In *ICCV* (2001), pp. 599–606. 2
- [PCD*10] PALMA G., CALLIERI M., DELLEPIANE M., CORSINI M., SCOPIGNO R.: Geometry-aware video registration. In *VMV* (2010), pp. 107–114. 3, 8
- [PTC10] PIETRONI N., TARINI M., CIGNONI P.: Almost isometric mesh parameterization through abstract domains. *IEEE TVCG* 16, 4 (July/August 2010), 621–635. 7
- [Ros56] ROSENBLATT M.: Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics* 27, 3 (Sept. 1956), 832–837. 6
- [RWS*11] REN P., WANG J., SNYDER J., TONG X., GUO B.: Pocket reflectometry. *ACM Trans. Graph* 30, 4 (2011), 45. 3
- [SWRK11] SCHWARTZ C., WEINMANN M., RUITERS R., KLEIN R.: Integrated high-quality acquisition of geometry and appearance for cultural heritage. In *VAST 2011* (Oct. 2011), pp. 25–32. 2
- [THS04] TREUILLE A., HERTZMANN A., SEITZ S. M.: Example-based stereo with general BRDFs. In *ECCV* (2004), pp. Vol II: 457–469. 3
- [WGT*05] WENGER A., GARDNER A., TCHOU C., UNGER J., HAWKINS T., DEBEVEC P.: Performance relighting and reflectance transformation with time-multiplexed illumination. *ACM Transactions on Graphics* 24, 3 (July 2005), 756–764. 2
- [WVL*11] WU C., VARANASI K., LIU Y., SEIDEL H.-P., THEOBALT C.: Shading-based dynamic shape refinement from multi-view video under general illumination. In *ICCV* (2011), pp. 1108–1115. 2
- [WWMT11] WU C., WILBURN B., MATSUSHITA Y., THEOBALT C.: High-quality shape from multi-view stereo and shading under general illumination. In *CVPR* (2011), pp. 969–976. 2
- [WZT*08] WANG J., ZHAO S., TONG X., SNYDER J., GUO B.: Modeling anisotropic surface reflectance with example-based microfacet synthesis. *ACM Trans. Graph* 27, 3 (2008). 3
- [YWAC06] YU T., WANG H., AHUJA N., CHEN W.-C.: Sparse lumigraph relighting by illumination and reflectance estimation from multi-view images. In *EGSR* (2006), pp. 41–50. 2
- [ZERB05] ZICKLER T., ENRIQUE S., RAMAMOORTHY R., BELHUMEUR P.: Reflectance sharing: Image-based rendering from a sparse set of images. In *EGSR* (2005), pp. 253–264. 2