Reflectance and Illumination Estimation for Realistic Augmentations of Real Scenes

Salma Jiddi, Philippe Robert, Eric Marchand

▶ To cite this version:

HAL Id: hal-01355581
https://hal.inria.fr/hal-01355581

Submitted on 23 Aug 2016

Reflectance and Illumination Estimation for Realistic Augmentations of Real Scenes

Salma Jiddi*
Technicolor R&D France
IRISA

Philippe Robert†
Technicolor R&D France

Eric Marchand‡
Université de Rennes 1
IRISA

Figure 1: **Left:** Original captured image of the scene. **Center:** Augmented scene with two virtual cubes. One can notice the effect of the estimated 3D position of the light source. **Right:** Augmented scene with a virtual sphere where the specular occlusion is correctly rendered.

ABSTRACT

The acquisition of surface material properties and lighting conditions is a fundamental step for photo-realistic Augmented Reality (AR). In this paper, we present a new method for the estimation of diffuse and specular reflectance properties of indoor real static scenes. Using an RGB-D sensor, we further estimate the 3D position of light sources responsible for specular phenomena and propose a novel photometry-based classification for all the 3D points. Our algorithm allows convincing AR results such as realistic virtual shadows as well as proper illumination and specularity occlusions.

Index Terms: I.4.1 [Image Processing and Computer Vision]: Digitization and Image Capture—Reflectance; I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Photometry;

1 INTRODUCTION

The Augmented Reality field is expected to exhibit a high growth in the years to come. The ability of augmenting real scenes with subjective virtual objects offers a large range of applications in various domains such as training, entertainment and user assistance. Therefore, many research streams focus on delivering highly accurate and robust techniques for the AR essential algorithms such as tracking, 3D reconstruction and photometry. As defined in [3], photometric reconstruction is the process of estimating the illumination and surface reflectance properties of an environment, given a geometric model of the scene and a set of photographs of its surfaces. Photometric and geometric reconstruction have often been considered separately. The existing solutions often addressed scenes with simple geometry (e.g. a plane) and reflectance (e.g. Lambertian or non-textured surfaces) as well as scenes reduced to an isolated single object. For complex lighting, a mirror ball [9] or a fisheye camera [5] is generally introduced in the scene to capture the lighting. The recent development of RGB-D sensors (Google Tango tablet, Intel RealSense sensors, Microsoft Kinect, etc.) have moved geom-

etry and photometry closer providing new possible approaches to the subject. In this paper, we are interested in estimating the diffuse and specular components which can spatially vary in textured real scenes. Also, we aim at recovering the 3D position of existing light sources without using any light probes or external assistance. We only consider as input the RGB-D data provided by a given 3D sensor. This paper makes four key contributions:

- Recovery of the spatially varying diffuse and specular properties for the 3D scene.
- Estimation of the 3D position of point light sources using only observed specular reflections in the scene.
- Photometry-based classification of all 3D points (shadowed areas, Lambertian and/or specular surfaces).
- Cast shadows detection and removal within the recovered diffuse component map.

2 RELATED WORK

Illumination and reflectance estimation is one of the most challenging problems in the computer vision field. It requires fitting a Bidirectional Reflectance Distribution Function (BRDF) to the color intensity in the input RGB images. In this task, also called Inverse Rendering, simple lighting and reflection models have been considered [10][13]. To address this problem, a wide variety of techniques have been developed. They are based on one or multiple images and scene geometry must be known (at least roughly). They can differ from each other through the type of addressed scenes and acquired data about the scene.

In [8], Nishino et al. proposed a method to separate the reflectance components in the input images based on the pixel's intensity variation using both sparse images and a geometric model. They considered only an isolated single object and assumed a uniform specular reflectance property as well as distant light sources. Sato et al. [11] proposed a method that estimates the illumination distribution of a simple scene with shadows cast by an object using a single color image of the scene. Practically, the image does not display specular effects, so the method is mainly based on the shadows and alternates reflectance estimation and illumination radiance

*e-mail: salma.jiddi@technicolor.com , salma.jiddi@irisa.fr
†e-mail:philippe.robert@technicolor.com
‡e-mail:eric.marchand@irisa.fr

estimation. The reconstructed data is only valid over a small region of the scene due to the hemispherical nature of the resulting illumination model. Gibson et al. [3] addressed the same problem in less restricted situations. They considered virtual light sources which mimic the effect of direct and indirect illumination. The intensity of these virtual light sources and the surface material properties are estimated using an iterative algorithm which attempts to match calculated radiance values to those observed in photographs. Jachnik et al. [4] presented an algorithm which is able to capture a surface light-field from a single hand-held camera by moving it around a specular planar object. In [4], the captured surface light-field was split into its diffuse and specular components and used the specular reflection to estimate an environment map.

Several methods are based on intrinsic image decomposition techniques. Their objective is to separate an image into its reflectance and shading components. The reflectance component contains the intrinsic color (albedo) of depicted surfaces while the shading component encodes the incident illumination in the scene, shadows and shading. Lee et al. [6] presented a technique to solve this under-constrained problem using an RGB-D sequence. They used both local and non-local, shading and temporal constraints, and obtained interesting intrinsic-images based decomposition results. Neverova et al. [7] presented an approach which consists in decomposing an original image into its specular shading, diffuse shading and albedo. The obtained specular and diffuse shadings represent the inputs of an optimization process aiming at finding the 3D position of the light sources. In [6] and [7], no further estimation of the specular parameters is proposed as they are either processed as an outlier or simply integrated in the shading.

Further methods have taken advantage of the RGB-D information available throughout single, sparse, or dense streams. Knecht et al. [5] proposed a method that reconstructs the surrounding environment and then estimates its BRDF properties. They used both a Kinect sensor to acquire the scene's geometry and a fisheye camera to capture the incident illumination. They assumed a uniform material property for the recovered scene clusters, detected specular effects as highlights in the image and did not consider shadows for the diffuse component estimation. Recently, Boom et al. [1] proposed a method to estimate surface reflectance and illumination properties for a real static scene using a 3D sensor. Following a color-based segmentation on the assumed Lambertian surfaces, the algorithm estimates a single albedo for each cluster and recovers the 3D position of one light source only.

3 Algorithm Overview

Our method is an offline photometric analysis of 3D real scenes. We use a calibrated Kinect sensor to capture the scene under various viewing angles in order to bring out specular and diffuse reflections. We estimate the camera positions using Mutual Information [2] and register the entire sequence with regard to a reference frame in order to continuously locate 3D points. For every tracked pixel, we recover a spatio-temporal color profile which retains the variations of the color intensity. We propose a statistical method to classify these profiles and recover both, view-independent (diffuse) reflectance and view-dependent (specular) reflectance components. Futhermore, we estimate the 3D position of light sources responsible for specular effects and provide a photometry-based classification of the 3D points of the scene to distinguish between surfaces (shadowed areas, Lambertian and/or specular surfaces). Our algorithm deals with a variety of scenes where the texture spatially varies and several geometry shapes can be present. We use our estimates to demonstrate specular occlusions between real and virtual objects as well as realistic virtual shadows (Fig. 1). In terms of assumptions, the scene and the lighting are supposed to be static. Only the sensor moves. The 3D model of the scene is retrieved using depth maps but no necessary reconstruction is needed within

our algorithm pipeline. The light sources are supposed to be sparse and are modeled as point lights. Therefore, only one light source creates a specular effect at a time.

4 Method

4.1 Sequence registration and spatio-temporal color profiles recovery

The sensor provides both RGB (Fig. 3-a) and depth streams at 30Hz with a 640×480 pixel resolution. Our first objective is to extract color profiles of the observed scene points through the sequence. This requires tracking all pixels over the sequence. Practically, this is achieved by registering the images with regard to a reference one using the camera poses and the depth maps. In order to obtain highly accurate results, we implemented a three-pass registration procedure: first, we use a robust homography estimator [2] (MI is used as a similarity metric due to the presence of specular effects), recover the camera's 3D position and then use it to register all RGB images.

In the following, we will equally talk about the pixels of the reference image and the 3D scene points they correspond to. Considering all the sequence frames, we are able to track and estimate the intensity variations of a given pixel. These variations consitute a spatio-temporal color profile to which we will refer as the Reflectance Profile (RP). Practically, our set of registered images provides for each 3D point/pixel observed in the reference frame the evolution of its color along the video, described by $I^p(x,t)$, where x is the constant 2D location of point p in the images, and t is the index of the image in the sequence. A particular curve is retrieved for each pixel p and corresponds to a linear combination of the three RGB color channels. Various and differently shaped profile curves can be obtained (Fig. 2). For instance, if the values of the profile significantly vary, the 3D point is bound to belong to a specular surface. We have observed three main cases where a 3D point p holds an invariant profile: **(1)** p is purely diffuse and never exhibits specular effects in spite of the presence of incident lighting; **(2)** the point is not subject to specular effects because no light source contribution was captured along the browsing trajectory or simply because it is geometrically occluded by another scene object; **(3)** a specular effect exists all along the acquired sequence.
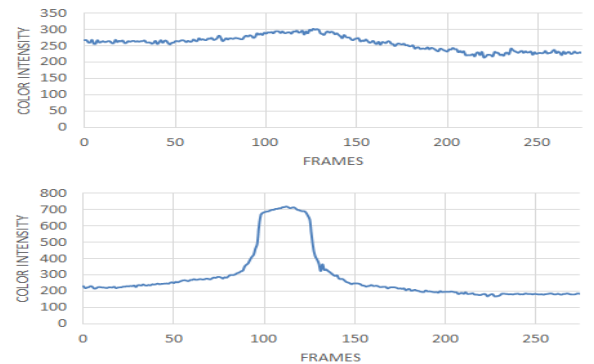

Figure 2: Reflectance Profiles. The upper figure corresponds to a point on the table that has not been impacted by the incident lighting. The lower figure demonstrates strong color intensity variations for a 3D point located on the specular black book.

The amount of information brought by these Reflectance Profiles is essential to our reflectance and illumination estimation process. Thus, in order to make the best use of it, we propose to separate these data into two main categories: Constant Reflectance Profiles (CRP) and Variable Reflectance Profiles (VRP). The former represents 3D points with weak intensity variations. The latter represents
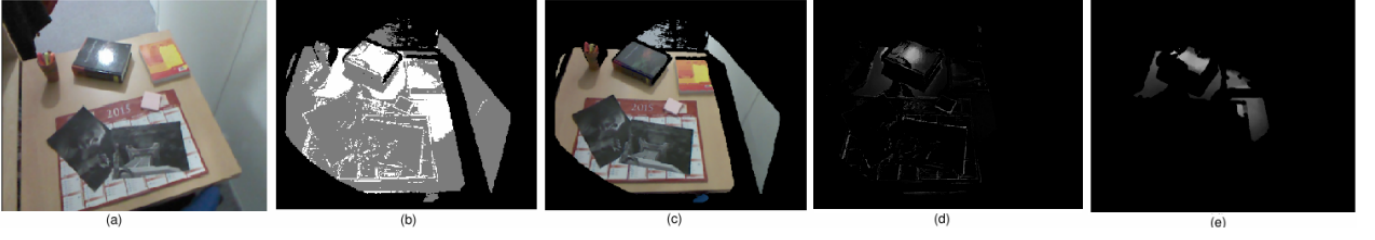
Figure 3: Estimation of the scene's photometric properties. **(a)**: RGB image of the scene. **(b)**: Reflectance Profile based classification. The black points represent the discarded pixels, the grey ones correspond to constant color profiles and the white ones hold variable color profiles. **(c)**: Recovered diffuse reflectance component. **(d)**: specular intensity at a given frame t. **(e)**:Recovered specular intensity for VRP points.

3D points which have been impacted by the light sources during the acquisition, their color intensity variations hold both, diffuse and specular reflectance components. In order to correctly apply this RP-clustering, we propose a simple and efficient statistical analysis. First, we consider the profiles (RP) which length is above half the total of frames (some points may be visible only on a part of the sequence). If a pixel's RP satisfies this first condition, we apply a gaussian filter in order to smooth the reflectance curve. For all the selected color profiles, we compute the minimum m_p and maximum M_p color intensity, the mean MN_p and median MD_p values of all stored variations, and the standard deviation of the distribution SD_p. Based on our own data analysis and observations, we propose a simple separation criterion to distinguish variable profiles from constant ones:

$$\begin{cases} |MD_{\mathbf{p}} - MN_{\mathbf{p}}| \geq \xi_1 \text{ or } |SD_{\mathbf{p}}| \geq \xi_2 & , p \in VRP \\ \text{else} & , p \in CRP \end{cases} \quad (1)$$

Using these critera, we quantify the amount of color intensity variations and dispersion throughout the entire sequence. Thus, when a profile holds significant variations, the difference between the mean and median values is expected to be significant as well. Furthermore, a Variable Reflectance Profile corresponds to a curve that demonstrates dispersion with regard to the mean value. As far as the VRP points are concerned, we are mainly interested in estimating their specular reflectance parameters which we will further use to recover the 3D position of the scene illumination. For CRP-points, the specular component will be set to 0 provided that their constancy comes from their purely diffuse property or constant geometry occlusion. In fact, as the light sources are sparse and the camera's trajectory is supposed to significantly cover the scene, we do not observe the case of a specular reflection all along the sequence. The results of our RP-based classification (Fig. 3-b) match our initial observations. The 3D points which are classified as VRP ones, have actually been impacted by the light sources during our scene capture. On the other hand, points which are constantly occluded by an object or that simply were not in the direction of incident lighting appear in the CRP group.

4.2 Reflectance estimation for Variable Reflectance Profiles

Observed colors are a combination of both, the illuminant color and the object color. They firmly depend on several dynamics such as the angle of incident lighting, specular reflectance and viewpoints. In order to approximate these light reflections and surfaces properties, several empirical models have been proposed. We refer to them as Reflection Models (RM). In our work, we have chosen Phong Reflection Model [10], to describe the way a point p on a surface reflects light as a combination of diffuse and specular reflections:

$$I^p = I^p_d + I^p_s \quad (2)$$

where I^p, I^p_d and I^p_s are respectively the color, the diffuse and the specular intensities. Diffuse reflectance is the property that defines an ideal "matte" surface, also called Lambertian surface. Its apparent brightness to an observer is the same regardless of his angle of view. On the other hand, specular reflection is the mirror-like reflection of light from a surface, in which light from a single incoming direction is reflected into a single outgoing direction. Using Phong model [10], diffuse and specular intensities can be described as:

$$I^p_d = \sum_{m=1}^{l} \mathbf{k}^p_d i_m (\mathbf{N}^p \cdot \mathbf{L}^p_m) \quad (3)$$

$$I^p_s = \sum_{m=1}^{l} \mathbf{k}^p_s i_m (\mathbf{R}^p_m \cdot \mathbf{V}^p)^{\alpha_p} \quad (4)$$

where $\mathbf{k}^p_d$ is the diffuse reflectance parameter of point p, $\mathbf{k}^p_s$ is its specular reflectance parameter, $\mathbf{N}^p$ is its normal vector, $\mathbf{V}^p$ is its viewpoint vector, α_p is its shininess parameter, i_m is the intensity of the light source m, $\mathbf{L}^p_m$ is its light direction vector, $\mathbf{R}^p_m$ is its reflection vector, and l is the number of light sources present in the scene. A recovered Reflectance Profile contains the color intensity of its corresponding pixel p throughout the sequence. Using Phong model, we approximate the profile $I^p(t)$ at frame t as a linear combination of both reflectance components:

$$I^p(t) = I^p_d + I^p_s(t) \quad (5)$$

where t refers to the image index, I^p_d to diffuse reflectance intensity and $I^p_s(t)$ to specular reflectance intensity at frame t. As we have previously pointed out, the lighting and scene are supposed to be static. Thus, the intensity variations present in the RP can only originate from the specular component, e.g. the viewpoint vector estimated at each frame t. In [4] and [14], the median of the observed values is used to compute I^p_d. Although this value is robust in presence of shadows or bad registration errors, it gives an over-estimation of the view-independent component. In our work, we have chosen as in [8], to estimate the diffuse intensity as the minimum observed value in the RP since it should be closer to the correct diffuse reflection value (Fig. 3-c):

$$I^p_d = m_p \quad (6)$$

The specular intensity is retrieved, at each frame t, as the difference between the diffuse intensity estimate I^p_d and the observed color intensity $I^p(t)$ (Fig. 3-d):

$$I^p_s(t) = I^p(t) - I^p_d \quad (7)$$

The light sources are supposed to be sparse and modeled as point lights. Subsequently, only one lighting creates specular effects at a given point at a time. If the moving camera crosses a specular effect at a given 3D point along its trajectory, the profile of the point will resemble a complete or partial lobe. Furthermore, if the trajectory is long, the profile may have more than one lobe. We separate these lobes at the minimal values between them and independently process them. Hence, for a given light source in the scene, the specular reflectance intensity can be rewritten as follows:

$$I^p_s(t) = \mathbf{k}^p_s i_S (\mathbf{R}^p_S \cdot \mathbf{V}^p(t))^{\alpha_p} \quad (8)$$

The unknown parameters are the combination of both, the specular parameter and the intensity of the light source S ($\mathbf{k}_s^p \mathbf{i}_S$), the reflection vector $\mathbf{R}_S^p$ and the shininess coefficient α_p. If the pixel's intensity at the peak of the RP is not saturated, the product ($\mathbf{k}_s^p \mathbf{i}_S$) can be retrieved and is equal to the lobe's peak intensity value. We refer to the index of the frame where the maximum is reached as t_m. In fact, when the specular effect occurs, the viewpoint vector and the reflection vector are initially supposed to be aligned. Subsequently, for every 3D point of the scene that holds a variable profile, we have:

$$\begin{cases} \mathbf{k}_s^p \mathbf{i}_S = I_s^p(t_m) \\ \mathbf{R}_S^p = \mathbf{V}^p(t_m) \end{cases} \tag{9}$$

Estimated specular components can contain errors due to registration misalignments. Hence, we apply a morphological erosion and recover the specular intensities $\mathbf{k}_s^p \mathbf{i}_S$ and reflection vectors $\mathbf{R}_S^p$ (Fig. 3-e). Since we have no guarantee that the profile has reached its maximum possible value, we will be refining this estimation further in this work. The shininess parameter α_p is proper to each material and is known to be larger for surfaces that are smoother and more mirror-like: the larger α_p, the narrower the RP lobe is. We estimate α_p using a non-linear Gauss-Newton optimization method with the following cost function:

$$\widehat{\alpha_p} = \arg\min_{\alpha_p} \sum_{t=1}^{N} (I_s^p(t) - \mathbf{k}_s^p \mathbf{i}_S(\mathbf{R}_S^p \cdot \mathbf{V}^p(t))^{\alpha_p})^2 \tag{10}$$

where N is the number of total frames. As described in equation (10), we estimate the variable α_p that minimizes the difference between the observed specular intensities $I_s^p(t)$, and the intensities approximated by Phong model throughout the entire sequence.

4.3 Light Source Position Estimation

In this section, we aim at estimating the 3D position of the light source responsible for specular effects int the scene. To begin with, we compute for each VRP-point p the light direction vector $\mathbf{L}_S^p$ using the estimate of the reflection vector $\mathbf{R}_S^p$:

$$\mathbf{L}_S^p = 2.(\mathbf{R}_S^p \cdot \mathbf{N}^p) \cdot \mathbf{N}^p - \mathbf{R}_S^p \tag{11}$$

Using the Point Cloud Library (PCL), we proceed to surface normal estimation for every 3D point in the scene. The method estimates the normal of a plane tangent to the surface and uses a least-square plane fitting solution. The problem of finding the 3D position of the light source is similar to computing the intersection point of multiple 3D lines. Hence, we estimate the position of the light source $\mathbf{S} = [x_S, y_S, z_S]^T$ by using a least square method. Based on the rough estimation of the VRP specular components, we propose an iterative process to refine these estimates and improve the localization of the light source. Using both the 3D coordinates $\mathbf{P}$ of the scene points and the 3D position of the light source $\mathbf{S}$, we are able to update the light direction vector $\mathbf{L}_S^p$ and refine the reflection vector $\mathbf{R}_S^p$ for all 3D points:

$$\mathbf{L}_S^p = \frac{\mathbf{S} - \mathbf{P}}{||\mathbf{S} - \mathbf{P}||} \tag{12}$$

Finally, using a Gauss-Newton minimization method, we alternatively refine the specular components and the lighting position until they both converge:

$$\widehat{(\mathbf{k}_s^p \mathbf{i}_S}, \widehat{\alpha_p}) = \arg\min_{(\mathbf{k}_s^p \mathbf{i}_S, \alpha_p)} \sum_{t=1}^{N} (I_s^p(t) - \mathbf{k}_s^p \mathbf{i}_S(\mathbf{R}_S^p \cdot \mathbf{V}^p(t))^{\alpha_p})^2 \tag{13}$$

$$\widehat{\mathbf{R}_S^p} = \arg\min_{\mathbf{R}_S^p} \sum_{t=1}^{N} (I_s^p(t) - \mathbf{k}_s^p \mathbf{i}_S(\mathbf{R}_S^p \cdot \mathbf{V}^p(t))^{\alpha_p})^2 \tag{14}$$

In Fig. 4, the Euclidean distance between the estimated 3D position and the true (measured) position of the light source is equal to 11cms for a mean distance of the light to the scene of 3.5 ms.

Figure 4: 3D position of the light source. The left image is a captured picture of scene and lighting. The right image describes the point cloud of the scene (white pixels) and shows the estimated position of the light source (center of the yellow sphere).

4.4 Photometry-based classification of the scene

Unlike [8], we further consider 3D points which hold a constant spatio-temporal color profile, we have been referring to them as CRP points. These profiles contain exploitable information that allow to recover their reflectance properties. In order to make the best use of these data, we first need to refine our RP-based classification. In section 4.1, we proposed a first classification within the entire color profiles dataset that separates Reflectance Profiles into VRP and CRP groups. Within the CRP class, we can encounter different points with regard to their reflectance properties. Hence, they need to be processed differently in order to obtain accurate reflectance estimates. Phong model [10] is rewritten as:

$$I^p(t) = I_d^p + O_S^p \mathbf{k}_s^p \mathbf{i}_S(\mathbf{R}_S^p \cdot \mathbf{V}^p(t))^{\alpha_p} \tag{15}$$

where O_S^p is an occlusion parameter, equal to 1 if light source S is visible from point p, and 0 otherwise. We added this parameter to take account of visibility within the scene. According to equation (15), a profile's constancy is observed when the specular component is equal to 0. This can be due to four possible reasons: the 3D point is not visible from the specular light source S because an object occludes it ($O_S^p = 0$). The 3D point belongs to a purely diffuse surface ($\mathbf{k}_s^p = 0$). The trajectory of the camera with respect to point p and to the light source S is such that the reflection vector $\mathbf{R}_S^p$ and the viewpoint vector $\mathbf{V}^p(t)$ have significantly different directions ($\mathbf{R}_S^p \cdot \mathbf{V}^p(t)$ is always equal or close to 0). Last, the 3D point has a constant specular effect all along the sequence, however as our light sources are sparse and the camera trajectory is significantly varying, this case is not met. Based on these observations, we categorise the CRP points into three main classes: Constantly Diffuse Points (CDP) such that ($\mathbf{R}_S^p \cdot \mathbf{V}^p(t)$) is equal or close to 0 all along the sequence, Pure Diffuse Points (PDP) which are 3D points lit by all light sources and showing no specular reflection ($O_S^p = 1$ and $\mathbf{k}_s^p = 0$). Constantly Occluded Points (COP) are 3D points which are constantly occluded with respect to the estimated light source S ($O_S^p = 0$). In the following, our objective is to fully classify CRP-points with regard to the previously defined subgroups (CDP, PDP and COP). Our classification is a two-pass procedure. The first pass consists in computing a visibility map. We have chosen the *shadow mapping* technique, a standard computer graphics algorithm. We refer to the shadow map as O_S^p where:

$$\begin{cases} O_S^p = 1 & \text{, p is a visible point} \\ O_S^p = 0 & \text{, p is an occluded point} \end{cases} \tag{16}$$

If point p is detected as an occluded point with regard to light source S, it is then classified as a COP point. Otherwise, it can belong to either CDP or PDP subgroups. The second pass consists in separating these two sub-classes by detecting 3D points which might be subject to a specular effect but still conserve a constant

profile, e.g. purely diffuse surfaces (PDP). To begin with, using the estimated reflection vector $\mathbf{R}_S^p$ (Section 4.2) and the viewpoint vector $\mathbf{V}^p(t)$, we retrieve a novel profile $\varepsilon_S^p(t)$ such as:

$$\varepsilon_S^p(t) = \mathbf{R}_S^p \cdot \mathbf{V}^p(t) \tag{17}$$

We use a method similar to the one described in section 4.1 (Eq. 1), with different thresholds, to distinguish variable and constant intensity profiles. If $\varepsilon_S^p(t)$ is significantly variable, then we conclude that $\mathbf{k}_s^p = 0$ (e.g. the current point intensity shows a constant diffuse value), point p belongs then to the PDP subgroup. If $\varepsilon_S^p(t)$ is a constant intensity profile, the point belongs to the CDP subgroup. The different steps of the classification are reported in (Fig. 5). The results of our 3D classification are shown in (Fig. 6-a).
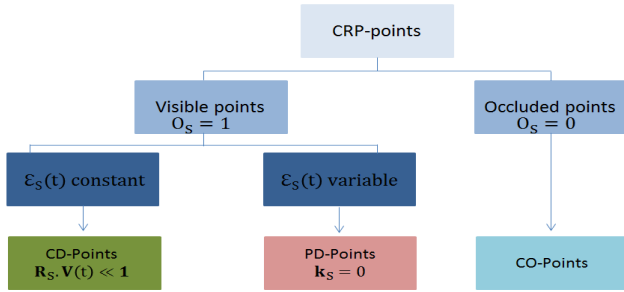
Figure 5: 3D-classification pipeline. It separates CRP-points to three main classes: Constantly Diffuse Points (CDP) which are not observed under the impact of incident lighting. Pure Diffuse Points (PDP) which demonstrate a pure diffuse property (their specular intensity is equal to 0), and Constantly Occluded Points (COP) which are constantly occluded by another surface/point with regard to the light source.
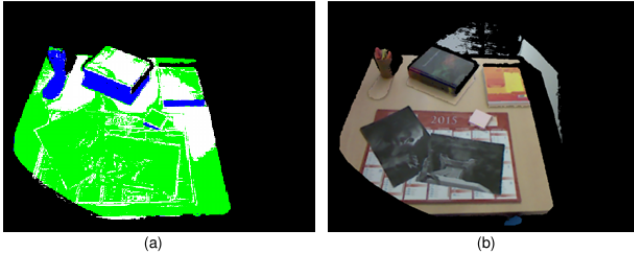
Figure 6: **(a)**: Photometry-based classification. *White* pixels are 3D points with variable profiles, *green* pixels are diffuse points that belong to the CDP subgroup and *blue* pixels are occluded points with regard to lighting (COP). **(b)**: Recovered diffuse component for Constantly Occluded Points (COPs).

4.5 Diffuse reflectance estimation for Constant Reflectance Profiles

In this section, our objective is to estimate surface diffuse reflectance properties for all points with Constant Reflectance Profiles (CRP). Our approach is based on cast shadows that are due to occlusion of identified light sources. These areas corresponding to the COP points are described in the occlusion map. The method is based on the comparison between points with the same diffuse reflectance that differ through lighting: CRP-visible points (e.g. CDP and PDP) and CRP-occluded points (COP). The color intensity of these points differs as follows:

$$\begin{cases} I^p = I_d^p = I_{d,O}^p + I_{d,V}^p & \text{, if p} \in \text{(CDP or PDP)} \\ I^p = I_d^p = I_{d,O}^p & \text{, if p} \in \text{COP} \end{cases} \tag{18}$$

where $I_{d,V}^p$ corresponds to the diffuse color intensity due to the light sources that are visible from CDP and PDP points only, and $I_{d,O}^p$ corresponds to the diffuse color intensity due to the light sources that are visible from all points. To begin with, we suppose -for now- that we are able to identify two points p_1 and p_2 with the same unknown diffuse parameter ($\mathbf{k}_d^p = \mathbf{k}_d^{p_1} = \mathbf{k}_d^{p_2}$). Point p_1 is supposed to belong to the CRP-visible subgroups whereas p_2 belongs to the COP subgroup. Since p_1 and p_2 are assumed to have the same diffuse parameter, we can write:

$$\begin{cases} I_d^{p_1} = \mathbf{k}_d^p l_O^p + \mathbf{k}_d^p l_V^{p_1} \\ I_d^{p_2} = \mathbf{k}_d^p l_O^p \end{cases} \tag{19}$$

where $l_V^{p_1}$ refers to the n light sources occluded from COP points and l_O^p refers to light sources visible from both types of points:

$$l_V^{p_1} = \sum_n (\mathbf{N}^p \cdot \mathbf{L}_n^p) i_n \tag{20}$$

l_O^p in both equations (19) is supposed to be same since the distance between CRP-visible and COP points is considered to be small with regard to the distance of the light source. The n light sources in equation (20) are supposed to correspond to the ones identified through their specular effect. Hence, both $\mathbf{k}_d^p$ and l_O^p can be estimated for the selected points:

$$\begin{cases} \mathbf{k}_d^p = \frac{I_d^{p_1} - I_d^{p_2}}{l_V^{p_1}} \\ l_O^p = \frac{I_d^{p_2}}{\mathbf{k}_d^p} \end{cases} \tag{21}$$

Except close to shadow edges, lighting is locally constant and there is interest to consider a group of points as far as they are identified as having the same diffuse parameter $\mathbf{k}_d^p$:

$$\widehat{\mathbf{k}}_d^p = \frac{\bar{I}_{d,V}^p - \bar{I}_{d,O}^p}{\bar{l}_V^p} \tag{22}$$

where $\bar{I}_{d,V}^p$ and $\bar{l}_V^p$ correspond respectively to average diffuse component and visible light sources intensities computed over all points belonging to the CRP-visible subgroups, and $\bar{I}_{d,O}^p$ is the average diffuse component intensity computed over all points belonging to the COP group (e.g. constantly occluded points). l_O^p is then estimated as:

$$\widehat{l_O^p} = \frac{\sum_p \bar{I}_{d,COP}^p}{\sum_p \widehat{\mathbf{k}}_d^p} \tag{23}$$

Using the estimated $\widehat{\mathbf{k}}_d^p$, we are able to recover the contribution of the occluded light sources with regard to COP points (Fig. 6-b).

Diffuse-reflectance clustering

The main challenge in estimating the parameter $\mathbf{k}_d^p$ remains in identifying visible and occluded surfaces corresponding to points with the same unknown diffuse intensity. To achieve this goal, we propose to group points with regard to several strong similarity metrics: chromaticity values, color intensities, normal vectors and 3D locations. Chromaticity values are computed using the Modified Specular Free (MSF) image that is more robust than Specular Free images [12]. In order to compute the average diffuse values $\bar{I}_{d,V}^p$ and $\bar{I}_{d,O}^p$, we propose a feature-weighted filter defined as follows:

$$\bar{I}_d^p = \frac{\sum_q (\omega_{p,q} I_d^q)}{\sum_q \omega_{p,q}} \tag{24}$$
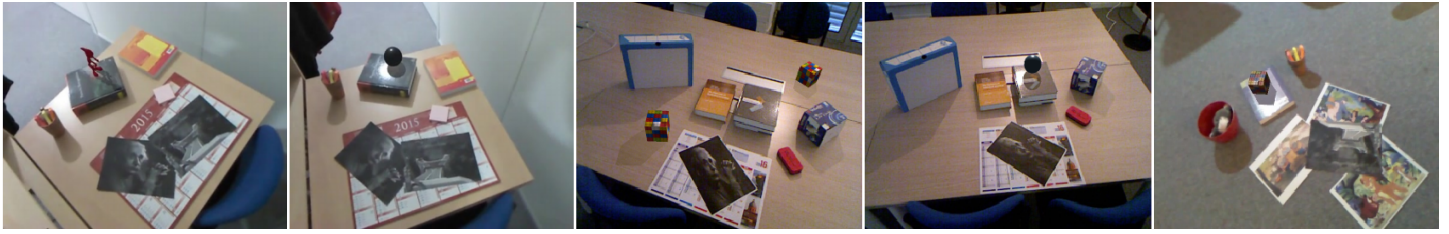
Figure 7: Augmented scenes, with different reflectance and illumination conditions. We demonstrate correctly rendered virtual objects as they occlude real specular effects (note the presence of the recovered diffuse component in the occluded region) and show realistic shadows.

$\bar{I}^p_{d,V}$ and $\bar{I}^p_{d,O}$ use a set of points q located respectively in visible and occluded areas. These areas are known thanks to our photometry-based classification results. The weights $\omega_{p,q}$ consider all the previously mentioned similarity features:

$$\omega_{p,q} = e^{-\Sigma_f \, cost(f)} \qquad (25)$$

where index f refers to a feature and cost(f) refers to the cost of dissimilarity between features attached to points p and q.

4.6 Rendering

Using our estimates, we handle the way the color in the real scene is modified by the presence of a virtual object while being under different reflections (diffuse and specular). This is achieved in a two-step procedure. First, using the estimated average diffuse components in both visible and occluded surfaces, we compute an attenuation coefficient β as the ratio of both recovered values. The second step consists in detecting, with regard to our photometry-classification, the corresponding group/subgroup of the surface points that the virtual object occludes (VRP, CDP, PDP or COP). Hence, if the virtual object occludes a COP point, no shadow is rendered. On the contrary, if the virtual object occludes another type of point, we render the virtual shadow by multiplying its diffuse value by the attenuation factor β.

5 RESULTS

RGB-D sequences have been captured for eleven indoor static scenes with differently shaped and textured objects under various indoor lightings (e.g spot lights, led lights). In these scenes, we used salient control points with regard to our world frame in order to measure the accuracy of this method. In our experiments, an average error of 16cm (for an average distance of 3m to the light source) between the estimated position and true position of the light source is achieved, with a standard deviation of 3.8cm. In many indoor scenes, where the assumption of distant lighting may not hold, the recovery of the lighting's 3D position has its importance to realistically render virtual objects.

5.1 AR Experiments

Figure 7 shows a variety of augmented scenes where virtual shadows and specular occlusions are correctly rendered. The most important aspects of realism are the synthetic shadows (same attenuation as the corresponding real shadows) and the occlusion of specularities by virtual objects to observe the recovered diffuse component. Our algorithm deals with textured surfaces and recovers convincing diffuse maps for 3D real scenes where multiple objects with different colors and shapes are present. Also, our proposed classification (Sec. 4.4) allows us to remove shadows from our diffuse estimate. Subsequently, the latter can be used as a texture for our recovered 3D model. Further results and rendering comparisons are shown in: *https://youtu.be/CzxB2yA7kGw*.

6 CONCLUSION

We presented a method for recovering a diffuse map of the scene from a sequence with specular reflections and cast shadows. We estimated the reflectance parameters of the scene and recovered the 3D position of light sources that create specular effects at the observed viewpoints. Specifically, in this paper we make use of the photometric properties of a real scene to correctly insert virtual objects for realistic Augmented Reality applications. Our method only deals with point light sources which produce hard shadows, and does not consider indirect illumination. For future work, we are interested in considering the soft shadow aspect that real surfaces can exhibit due to the presence of area lights. A clear step is then to extend our model and integrate it in our pipeline for more realistic indoor AR applications.

REFERENCES

[1] B. Boom, S. Orts-Escolano, X. Ning, S. McDonagh, P. Sandilands, and R. Fisher. Point light source estimation based on scenes recorded by a rgb-d camera. *BMVC, 2013*.

[2] A. Dame and E. Marchand. Second order optimization of mutual information for real-time image registration. *IEEE Trans. on Image Processing*, 2012.

[3] S. Gibson, T. Howard, and R. Hubbold. Flexible image-based photometric reconstruction using virtual light sources. In *Eurographics, 2001*.

[4] J. Jachnik, R. A. Newcombe, and A. Davison. Real-time surface light-field capture for augmentation of planar specular surfaces. In *ISMAR, 2012*.

[5] M. Knecht, G. Tanzmeister, C. Traxler, and M. Wimmer. Interactive brdf estimation for mixed-reality applications. *WSCG, 2012*.

[6] K. Lee, Q. Zhao, X. Tong, M. Gong, S. Izadi, S. Lee, P. Tan, and S. Lin. Estimation of intrinsic image sequences from image+depth video. In *ECCV, 2012*.

[7] N. Neverova, D. Muselet, and A. Tremeau. Lighting estimation in indoor environments from low-quality images. In *ECCV, 2012*.

[8] K. Nishino, Z. Zhang, and K. Ikeuchi. Determining reflectance parameters and illumination distribution from a sparse set of images for view-dependent image synthesis. In *ICCV, 2001*.

[9] D. Nowrouzezahrai, S. Geiger, K. Mitchell, R. Sumner, W. Jarosz, and M. Gross. Light factorization for mixed-frequency shadows in augmented reality. In *ISMAR, 2011*.

[10] B. Phong. Illumination for computer generated pictures. *Communication of the ACM*, 1975.

[11] I. Sato, Y. Sato, and K. Ikeuchi. Illumination distribution from brightness in shadows: Adaptive estimation of illumination distribution with unknown reflectance properties in shadow regions. In *ICCV, 1999*.

[12] H.-L. Shen and Q.-Y. Cai. Simple and efficient method for specularity removal in an image. *Applied optics*, 2009.

[13] K. Torrance and E. Sparrow. Theory for off-specular reflection from roughened surfaces. *JOSA*, 1967.

[14] D. Wood, D. Azuma, W. Aldinger, B. Curless, T. Duchamp, D. Salesin, and W. Stuetzle. Surface light fields for 3d photography. In *SIGGRAPH*, 2000.