



# Detecting Specular Reflections and Cast Shadows to Estimate Reflectance and Illumination of Dynamic Indoor Scenes

Salma Jiddi, Philippe Robert, Eric Marchand

## ► To cite this version:

Salma Jiddi, Philippe Robert, Eric Marchand. Detecting Specular Reflections and Cast Shadows to Estimate Reflectance and Illumination of Dynamic Indoor Scenes. IEEE Transactions on Visualization and Computer Graphics, 2022, 28 (2), pp.1249 - 1260. 10.1109/TVCG.2020.2976986 . hal-02475059

**HAL Id: hal-02475059**

**<https://inria.hal.science/hal-02475059>**

Submitted on 11 Feb 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Detecting Specular Reflections and Cast Shadows to Estimate Reflectance and Illumination of Dynamic Indoor Scenes

Salma Jiddi, Philippe Robert and Eric Marchand

**Abstract**—The goal of Mixed Reality (MR) is to achieve a seamless and realistic blending between real and virtual worlds. This requires the estimation of reflectance properties and lighting characteristics of the real scene. One of the main challenges within this task consists in recovering such properties using a single RGB-D camera. In this paper, we introduce a novel framework to recover both the position and color of multiple light sources as well as the specular reflectance of real scene surfaces. This is achieved by detecting and incorporating information from both specular reflections and cast shadows. Our approach is capable of handling any textured surface and considers both static and dynamic light sources. Its effectiveness is demonstrated through a range of applications including visually-consistent mixed reality scenarios (e.g. correct real specular removal, coherent shadows in terms of shape and intensity) and retexturing where the texture of the scene is altered whereas the incident lighting is preserved.

**Index Terms**—Photometric registration, illumination, reflectance, diffuse, specular, shadow, texture, mixed reality, retexturing.

## 1 INTRODUCTION

THE overarching goal of Mixed Reality (MR) is to provide users with the illusion that virtual and real objects coexist indistinguishably in the same space. An effective illusion requires a geometrically and photometrically coherent registration between both worlds. In this paper, we focus on the photometric registration task [1] which consists in estimating surface reflectance properties and lighting characteristics of the real world using its geometric model and a set of photographs of its surfaces.

Usability is an important consideration for photometric registration approaches. Proposed methods must maintain independence from the real scene’s content (e.g. geometry, reflectance, illumination) as much as possible. In fact, constraining the MR user to have a single light source [2] or a textureless surface [3] reduces the range of possible scenarios. Also, even though approaches using light probes such as chrome spheres [4] or fisheye cameras [5] handle more generic scenes and deliver convincing augmentations, the use of such additional devices is neither practical nor adequate for wide-reaching MR systems. Consequently, several state-of-the-art methods aim to use a single RGB or RGB-D camera for photometric registration. Such approaches often rely on observed cues such as shading, shadows or specular effects to probelessly recover reflectance and illumination.

A shadow occurs when illumination coming from a light source is partially or totally obstructed by one or more objects. Since less illumination reaches these regions, both cast and self-shadows have a lower luminance in comparison with their surrounding regions (Figure 1). Nonetheless, one can not precisely tell if a surface is “dark” due

to its intrinsic color/texture or its shadowing/shading by only considering its local appearance. Consider Figure 1: the red boxes show that a patch located in a shadowed region retains -to some extent- the same color as a non-shadowed region. Another case where such regions can be misinterpreted occurs when the real scene is lit by a spot light. This type of light source has a cone of influence: points outside of this cone do not receive illumination from the light source. Consequently, their local appearance can be easily confused with shadowed regions, illustrated by the green-box patches in Figure 1. In presence of textured surfaces, such misinterpretations occur even more often.

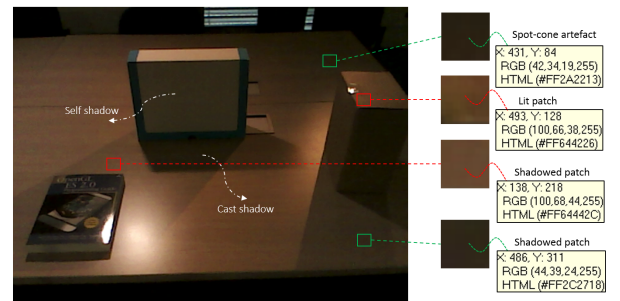


Fig. 1. Local analysis of shadows: the local appearance of surfaces can be ambiguous for shadow recognition: red-box patches correspond to points under different lighting conditions, yet they have similar colors. The upper green-box patch is often erroneously detected as a shadowed region due to spot light’s cone effect.

Specular reflections refer to bright pixels which can be observed in captured images. Such reflections are informative in terms of scene reflectance and illumination. In fact, they represent useful cues regarding the direction of the light source causing them. Hence, they should be taken into account for this purpose within the photometric registration task. Moreover, when cast shadows can not be

- Salma Jiddi is now with Geomagical Labs, Mountain View, USA.
- Philippe Robert is now with Interdigital, Rennes, France.
- Eric Marchand is with Univ Rennes, Inria, CNRS, IRISA, Rennes, France.

easily detected (e.g. weak shadows) due to the presence of spatially close specular effects, these cues can be efficiently incorporated to estimate illumination.

**Contributions.** In this paper, we propose a method to robustly detect and leverage both shadow and specular cues in a photometric registration framework. Specifically, we consider indoor real scenes composed of one or more objects with arbitrary shapes. Most importantly, scene surfaces can hold *arbitrary* textures and retain Lambertian and specular properties. Also, our framework handles multiple lights and their respective locations can be freely changed by the user overtime. Using an RGB-D camera, the proposed approach requires -at initialization only- a capture of the scene under a *near-ambient lighting*. Then, for each incoming frame, it jointly and robustly detects static and dynamic specular reflections and cast shadows to estimate the specular reflectance of scene surfaces and illumination characteristics (number of light sources, their respective 3D positions and colors). To summarize, our main contributions are:

- Detection of specular reflections and challenging cast shadows (e.g., weak shadows, overlapping shadows) on arbitrary textured surfaces.
- Estimation of the 3D position of static and/or dynamic light sources for each incoming frame.
- Estimation of the color of recovered light sources.
- Estimation of the specular component parameters of scene surfaces.
- Near real-time implementation of the proposed method in order to meet MR requirements.

The remainder of this paper is organized as follows: we first review related works which use a single sensor to estimate reflectance and/or illumination. Then, we present the inputs, assumptions and an overview of our proposed approach. We then describe how we jointly exploit specular reflections and cast shadows to estimate the reflectance and illumination of the scene. Finally, we present experimental results along with two applications: realistic mixed reality and retexturing.

## 2 RELATED WORK

In this section, we present related works which use a single sensor to estimate reflectance and illumination properties within real scenes. The considered sensor can be either an RGB camera (the 3D model is not required or reduced to basic geometry such as a plane) or an RGB-D camera. A broader overview of further approaches (e.g. using HDR image/video [6] [7], chrome spheres [4], fish-eye lens cameras [5]) is extensively discussed in [8].

**SPECULARITY-BASED APPROACHES.** Early work using specular cues was proposed by Nishino et al. [9] where they separate diffuse and specular reflection components using a sparse image set and a geometric model of the real scene. In [9], the scene is reduced to a single specular object assumed, in addition to the light sources, to be static. As the camera browses the object, specular reflections are brought out and used to derive both illumination and reflectance properties. Jachnik et al. [10] presented a similar method to [9] where surface light-field is captured from a single hand-held RGB

camera by moving it around a specular planar object (e.g., shiny book). Though this method achieves convincing MR renderings, it considers only a planar and small surface. Furthermore, the illumination is recovered using an environment map which implies distant light sources.

**SHADOW-BASED APPROACHES.** Early work using shadows to recover static illumination was proposed by Sato et al. [11] [12]. The authors proposed a method that estimates the illumination distribution using cast shadows by an object of known geometry. The main drawback consists in requiring extensive user intervention. In fact, in order to identify the shadowed regions, two captures of the scene are required: with and without occluding objects. Since then, extensive work has been carried in order to automatically detect shadows. For instance, by considering an object with simple and known geometry (e.g., cube), Arief et al. [3] analyze the shadows cast on a single-color surface. The algorithm first detects shadow contours and then, using corner features, it recovers the lines which relate shadow corners to their corresponding 3D points. The intersection of these lines corresponds to the 3D position of the light source. Panagopoulos et al. [13] [14] proposed an approach based on a Markov Random Field (MRF) model to detect shadows in less constraining environments. In fact, their approach handles textured surfaces and further recovers multiple light sources directions. Nonetheless, it only handles Lambertian surfaces and takes 3 to 5 minutes to process a single image. In [15], they proposed an approach to automatically detect shadows in Lambertian textured scenes using an RGB-D camera. The position of light sources is recovered within an iterative process where detected cast shadows are correlated with a set of candidate binary shadow maps. As shown in this **video**, the approach runs at an interactive framerate and handles both static and dynamic lighting.

**MULTIPLE CUES BASED APPROACHES.** The existence of a light source is more likely if it is supported by more than one cue. To our knowledge, there are only two related works which fall in this category. In [16], specular reflections, recovered as saturated regions, are used to estimate the light sources direction. Then, a search of the light source's 3D position is considered along each recovered direction. The final estimate of the light source position corresponds to the intersection of lines connecting detected corners of shadows (cast on a single-color and Lambertian planar surface) with their corresponding 3D points within the occluding object.

The second method proposed by [17] considers the critical case of textured surfaces. The approach determines the expected positions of shadow edges and specularities for hypothetical lighting directions sampled from a hemisphere. Then, it verifies the accuracy of recovered lighting by analysing shadows. For instance, shadow boundary points are computed using Canny edge filters and are compared to the generated hypothetical shadows using Euclidean distance. The use of the Canny edge filters is the downside of this method as it poorly performs in case of low-quality images or in presence of challenging textures and weak shadows.

**DATA-DRIVEN APPROACHES.** Several end-to-end learning based approaches have been recently proposed to tackle the challenge of estimating illumination from HDR [18] or LDR [19] [20] [21] [22] images. In [18], they propose

a method to infer high dynamic range illumination from a single panoramic photograph. The approach follows the work of Karsch et al [23] to detect light sources. Nonetheless, reflectance estimation is not addressed. On the other hand, in [19], a CNN based framework is proposed to estimate shape and spatially varying reflectance, represented as diffuse albedo and specular roughness, from a single mobile phone photograph. In the considered context, scenes must be captured under controlled lighting conditions. Marques et al [22] propose an approach to estimate the main light source in the scene from a capture of the user's hand which is exploited as a light probe of the 3D environment.

In this paper, we tackle several critical challenges (e.g. presence of specular reflections, spatially-varying textures, multiple shadows) by robustly incorporates both shadow and specular cues within a photometric registration framework. In the following, we provide an overview of the inputs, assumptions and proposed approach.

### 3 OUR PHOTOMETRIC REGISTRATION APPROACH

Our proposed method takes three inputs: (1) a coarse 3D model of the scene which can be acquired with an RGB-D sensor (e.g. Intel R200); (2) color images of the scene from which illumination will be recovered; and (3) a color image of the scene captured under a *near-ambient* lighting. This color image mainly contains the color/texture of scene surfaces and, ideally, does not exhibit any shadowing or specular effects. We will refer to this image as the *reference* image  $\mathbf{I}_{ref}$  (Figure 2). This *reference* image can be produced in two different manners: (i) using a floor lamp which is oriented towards the ceiling to create a uniform indirect illumination. This is well suited for indoor movie studios where the end-user would be able to easily achieve this step. (ii) using natural environment lighting where an inexperienced user can be asked to dim or turn off the light sources in the room. Then, he/she can capture the reference image and select a desired lighting conditions afterwards within MR. As nowadays sensors are getting better at capturing low-light images, the **video** submitted with this paper shows a panel of MR scenarios where this capture can be easily acquired.

We consider the case where we have a main planar surface on which arbitrary-shaped objects, with different reflective properties, cast shadows. Consequently, the color of a scene point  $p$  can be described using Phong model [24] as a combination of three components:

$$\mathbf{I}^p = \mathbf{I}_a^p + \mathbf{I}_d^p + \mathbf{I}_s^p \quad (1)$$

where  $\mathbf{I}^p$ ,  $\mathbf{I}_a^p$ ,  $\mathbf{I}_d^p$  and  $\mathbf{I}_s^p$  are respectively the color, the ambient, the diffuse and the specular components of point  $p$ . Using Phong model [24], the reflection components in equation (1) are described as follows:

$$\mathbf{I}^p = \mathbf{k}_d^p \mathbf{L}_a + \mathbf{k}_d^p \sum_{i=1}^M (\mathbf{n}^p \cdot \boldsymbol{\omega}_i^p) \mathbf{L}_i \mathbf{O}_i^p + \mathbf{k}_s^p \sum_{i=1}^M (\mathbf{r}_i^p \cdot \mathbf{v}^p)^{\alpha_p} \mathbf{L}_i \mathbf{O}_i^p \quad (2)$$

where  $\mathbf{L}_a$ ,  $\mathbf{L}_i$  are respectively the color vectors of ambient and light source  $i$ .  $\mathbf{k}_d^p$  and  $\mathbf{k}_s^p$  are respectively the diffuse and specular reflectances of point  $p$ ,  $\mathbf{n}^p$  is its normal vector,  $\mathbf{v}^p$  is its viewpoint vector, and  $\alpha_p$  is its shininess parameter.  $\mathbf{r}_i^p$

is the ideal reflection vector at point  $p$  with regard to light source  $i$  and  $\boldsymbol{\omega}_i^p$  is the direction of the light source  $i$  from point  $p$ .  $M$  is the number of light sources present in the scene.  $\mathbf{O}_i^p$  is a binary visibility term that is equal to 1 if light  $i$  is visible from the 3D point  $p$  and equal to 0 if occluded.

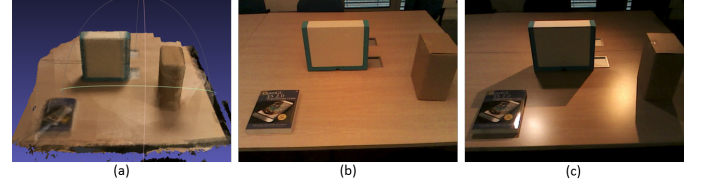


Fig. 2. Inputs of the proposed approach: (a) Acquired 3D model of the scene using the Intel R200 sensor. (b) *reference* image ( $\mathbf{I}_{ref}$ ) of the scene. One can notice the presence of mainly the surface's color/texture and hardly any shadowing or specular effects. (c) Scene's color image.

Given the color of a scene point  $p$ , its normal vector  $\mathbf{n}^p$  and its current view vector  $\mathbf{v}^p$  (all acquired or derived from the sensor's data), our goal is to estimate its reflectance properties ( $\mathbf{k}_d^p$ ,  $\mathbf{k}_s^p$ ,  $\alpha_p$ ) and the illumination in the scene ( $M$ ,  $\mathbf{L}_a$ ,  $\mathbf{L}_i$ ,  $\mathbf{r}_i^p$ ,  $\boldsymbol{\omega}_i^p$ ,  $\mathbf{O}_i^p$ ) under which the current color image is captured.

OVERVIEW: in order to robustly solve this inverse problem, we use specular reflections and cast shadows as illustrated in Figure 3: using the *reference* image ( $\mathbf{I}_{ref}$ ), we separate, for each incoming color image  $\mathbf{I}$ , surface texture from illumination effects (e.g., shading, shadowing and specular reflections). This step results in an image of the scene which mainly contains illumination-dependent variations and we will refer to it as the *illumination map*. From this *illumination map*, we detect the specular reflections and use them to estimate a rough direction of the light sources in the scene. Then, using detected specular effects and recovered lights directions, an adaptive and robust cast shadow detection is achieved and light sources positions and colors are estimated. Finally, using recovered scene illumination, specular reflectance parameters are estimated for all observed scene points.

In the following, we describe in detail the main components of our proposed photometric registration approach.

#### 3.1 Per-frame texture removal

In this section, our goal is to accurately separate texture/color variations from illumination-dependent effects such as shading, shadowing and specular reflections. This is of interest within our framework since it allows us to robustly detect and model these cues within captured images.

The proposed approach for texture and illumination separation is two-fold. To begin with, we are interested in recovering the diffuse reflectance  $\mathbf{k}_d^p$ , corresponding to the intrinsic color/texture, for all scene points. Let us consider a pixel  $p$  within the *reference* image ( $\mathbf{I}_{ref}$ ) shown in figure 4-a. The color of pixel  $p$ , corresponding to a 3D point in the scene, is described using Phong model [24] as:

$$\mathbf{I}_{ref}^p = \mathbf{k}_d^p \mathbf{L}'_a \quad (3)$$

where  $\mathbf{I}_{ref}^p$  and  $\mathbf{L}'_a$  are respectively the color vector of point  $p$  and *near-ambient* lighting (used to produce the *reference* color image). We are interested in recovering the



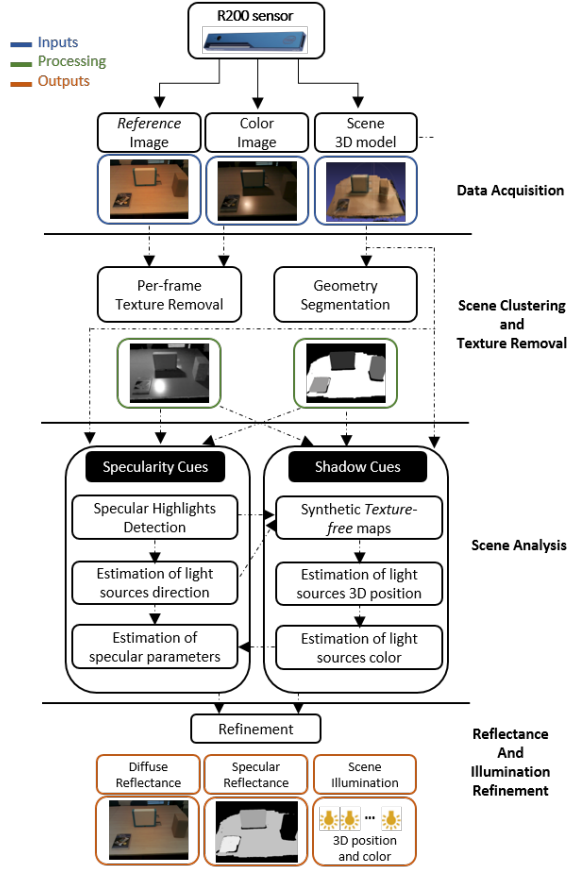


Fig. 3. Outline of the proposed photometric registration approach which jointly uses specular reflections and cast shadows to estimate the reflectance and illumination of the considered real scene.

texture/color of the scene which is independent from the lighting conditions under which the image is captured. To achieve this task, the unknown color of the *near-ambient* lighting  $\mathbf{L}'_a$  must be estimated. In this work, we use available white-color regions within the scene (red boxes in figure 4-(a,b)) to estimate this unknown as follows:

$$\mathbf{L}'_a = \frac{\sum_{p \in W} \mathbf{I}_{ref}^p}{\#W} \quad (4)$$

where  $W$  is a white-color region within the *reference* image and  $\#W$  is the pixels count within  $W$ . If no such region is available, one can use a color constancy algorithm, such as [25] to estimate  $\mathbf{L}'_a$ . Consequently, using equation (3), the albedo/texture  $\mathbf{k}_d^p$  of every scene points (Figure 4-c) can be recovered as:  $\mathbf{k}_d^p = \frac{\mathbf{I}_{ref}^p}{\mathbf{L}'_a}$ .

Now that we have recovered the diffuse reflectance  $\mathbf{k}_d^p$  of scene points, the second step of the proposed texture/illumination separation consists in recovering an *illumination map*  $\delta$  for each incoming color image. This map must be texture-free and contain mainly shading, shadowing and specular reflections. The map  $\delta$  is recovered, for each point  $p$ , using the diffuse reflectance estimate as:

$$\delta^p = \frac{\mathbf{I}^p}{\mathbf{k}_d^p} = \frac{\mathbf{k}_d^p (\mathbf{L}_a + \sum_{i=1}^M (\mathbf{n}^p \cdot \boldsymbol{\omega}_i^p) \mathbf{L}_i \mathbf{O}_i^p)}{\mathbf{k}_d^p} + \frac{\mathbf{k}_s^p \sum_{i=1}^M (\mathbf{r}_i^p \cdot \mathbf{v}^p)^{\alpha_p} \mathbf{L}_i \mathbf{O}_i^p}{\mathbf{k}_d^p} \quad (5)$$

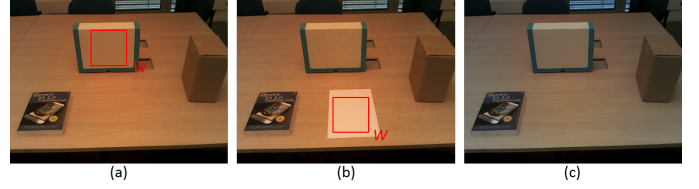


Fig. 4. Diffuse reflectance recovery: (a,b) are examples of *reference* images of the scene with available white-color regions ( $W$ ) from which the color of lighting can be estimated using equation (4). (c) estimated intrinsic texture/color of the scene which is independent from the lighting conditions under which it was produced (a,b).

Within equation (5),  $\mathbf{k}_d^p$  cancels out with regard to the ambient and diffuse components. Also, since specular reflections do not cover significantly large regions of the image, and for clarity reasons, the equation (5) is rewritten as follows:

$$\delta^p = \mathbf{L}_a + \sum_{i=1}^M (\mathbf{n}^p \cdot \boldsymbol{\omega}_i^p) \mathbf{L}_i \mathbf{O}_i^p + \epsilon_S^p \quad (6)$$

where  $\delta^p$  represents the *illumination* map value at pixel  $p$  and  $\epsilon_S^p$  corresponds to present specular highlights (Figure 5-b) at pixel  $p$ . As depicted in figure 5-b,  $\delta$  contains mainly shading effects (green boxes) which correspond to the scalar product  $(\mathbf{n}^p \cdot \boldsymbol{\omega}_i^p)$  in equation (6), cast shadows (blue boxes) results from the occlusion term  $\mathbf{O}_i^p$  and specular effects (red box) correspond to the term  $\epsilon_S^p$ . One can notice that in case of a Lambertian scene ( $\mathbf{k}_s^p = 0$ ), the  $\epsilon_S^p$  term is equal to 0 and no specular effect is recovered (second row in figure 5-b).

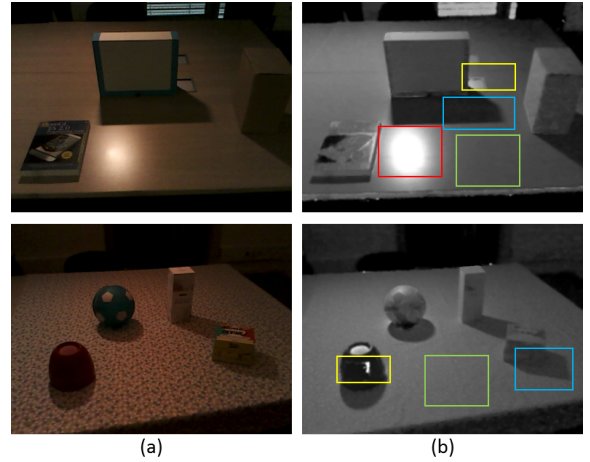


Fig. 5. Texture removal: (a) Input color image of captured scenes. (b) Recovered *illumination* maps for images in (a). Red, green, and blue boxes correspond respectively to specular, shading and shadowing variations in the scene. Yellow boxes are examples of noisy estimations due to low quality images or very shiny surfaces.

In the following sections, this recovered *illumination* map  $\delta$  is used to detect and model both specular reflections and cast shadows within the considered scene.

### 3.2 Specular highlights detection and lights direction estimation

Specular reflections represent view-dependent cues which are informative about the direction of the light source in the scene. In fact, these cues are observed when the camera

or the user's view direction is roughly aligned with the ideal specular reflection. In this section, our goal is to detect specularities within the recovered *illumination* map and use them to estimate the light sources direction.

To begin with, we detect specular highlights using [26]. By considering the Hue Saturation Value (HSV) color space, the approach recovers specular reflections at pixels where the color has high value (V) but low saturation (S). The value (V) corresponds to the maximum within the three-channel color vector (R,G,B) and the saturation (S) component is computed as follows:

$$S = \begin{cases} \frac{V - \min(R,G,B)}{V}, & \text{if } V \neq 0 \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

The chosen thresholds for minimum value and maximum saturation are respectively 0.8 and 0.2 (selected after extensive evaluation on our scene corpus). The results are written into a binary mask (H) with 1 where the specularity is detected and 0 otherwise. Because white surfaces may be misinterpreted as highlights, we make use of the *reference* image to improve our detection. This is achieved using the difference in terms of intensity (linear combination of the three color channels) between the reference  $I_{ref}^p$  and current frame  $I^p$  as follows:

$$H^p = \begin{cases} 1, & \text{if } I_{ref}^p < I^p + \epsilon \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

where  $\epsilon$  is a threshold over intensity difference. In fact, since the *reference* image  $I_{ref}$  mainly contains the intrinsic color of surfaces, when specular effects are captured within the current frame, point  $p$  holds a significantly more important brightness value in  $I^p$  than in  $I_{ref}^p$ . Both detected specularities and discarded bright regions are shown in figure 6.



Fig. 6. Specular reflections detection: (a) Input color image of the scene with two specular effects. (b) Detection of specular reflections: red pixels correspond to the detected specularities after discarding (magenta) pixels that do not check equation (8).

The second step consists in recovering the direction of light sources using the detected specularities (retrieved in the binary mask H). First, due to thresholding noise, small and/or isolated highlights can be detected. We handle these noisy regions using a simple blob detector: locally connected regions are initially recovered using [27] (Figure 7-a) and, each connected region is referred to as *blob* (Binary Large Object). Blobs with significantly low points count are discarded (yellow ellipse) and the center of each kept blob is computed along with its euclidean distance with regard to other groups. Close blobs are merged to form one specular effect (green and red ellipses near the specular book) and its center and radius are computed.

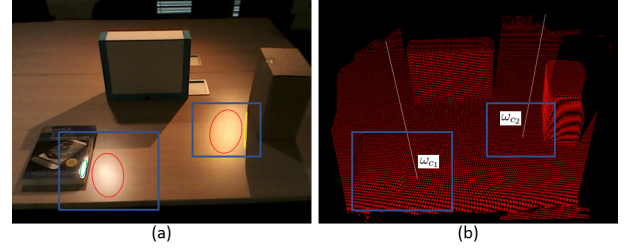


Fig. 7. Lights direction estimation: (a) Detected specularities: blobs with a small pixels count (yellow ellipse) are discarded. Spatially close detected regions are merged to form a single detected specularity (green ellipse with regard to the near red ellipse). (b) Recovered light sources directions using the recovered specular effects (blue boxes) in (a).

Finally, for each detected specularity, the ideal specular reflection direction  $\mathbf{r}^c$ , at the center  $c$  of each blob, is recovered as aligned with the view direction vector  $\mathbf{v}^c$  (computed using camera pose and coordinates of the corresponding 3D point to pixel  $c$ ). Consequently, an initial estimate of light sources direction, at the center of each blob  $c$ , is recovered as follows:

$$\omega_{c_k} = 2.(\mathbf{r}^{c_k} \cdot \mathbf{n}^{c_k}) \cdot \mathbf{n}^{c_k} - \mathbf{r}^{c_k} \quad (9)$$

where  $k$  iterates over the number of recovered blobs,  $\omega_{c_k}$  is the light direction vector estimated using the  $k^{th}$  blob,  $\mathbf{r}^{c_k}$  is its ideal specular reflection vector and  $\mathbf{n}^{c_k}$  is its normal vector. In figure 7-b, the scene is represented by a point cloud (red dots) and the recovered light directions are reported as white lines.

Unlike [16], we do not use recovered lights directions as search lines for their positions since it does not always deliver robust estimates. Our approach efficiently incorporates the information brought by the specularity analysis and is described in the next section.

### 3.3 Shadows analysis and lights position estimation

In this section, our goal is to efficiently incorporate the information brought by specular effects within an analysis approach of cast shadows in order to robustly estimate the position of light sources in the scene. This goal is achieved by considering two key steps: (i) the lighting in the scene is approximated by a set of equally distributed point lights ( $S_0$ ). By using recovered lights directions in section 3.2, we are able to consider a small set within ( $S_0$ ) which is more likely to contain actual real light sources. (ii) Using detected specular reflections along with the recovered *illumination* map  $\delta$  (section 3.1), we robustly recover the 3D position of light sources in the scene. The core idea consists in an iterative matching procedure between  $\delta$  and a set of synthetic *illumination* maps  $\tilde{\delta}$  proper to the hypothetical point lights. The approach to recover the light sources position is achieved in a three-pass procedure and described in detail in the rest of this section.

The first step consists in recovering a set ( $S$ ) of hypothetical point lights among which actual light sources will be identified later on. Within this task, we initially approximate the lighting in the scene by a set ( $S_0$ ) of point lights equally distributed in the scene (Figure 8). Then, for each recovered light direction  $\omega_{c_k}$  (section 3.2), we define a cone originating

from the detected specularity's center  $c_k$  and oriented using  $\omega_{c_k}$ . Finally, point lights located within the cone's volume constitute the set  $(S)$  as illustrated in figure 8. Beside the fact that, in comparison to the approach in [16], this method takes into account the inaccuracies which might exist within the roughly estimated lights direction, it also allows us to consider a smaller point lights set than the initial one  $(S_0)$ . This is of interest for MR scenarios where the processing time requirements must be considered as well. For instance, from an initial set  $(S_0)$  counting 1176 point lights, only 352 are comprised within the set  $(S)$  using a cone-angle  $\beta$  of  $10^\circ$ .

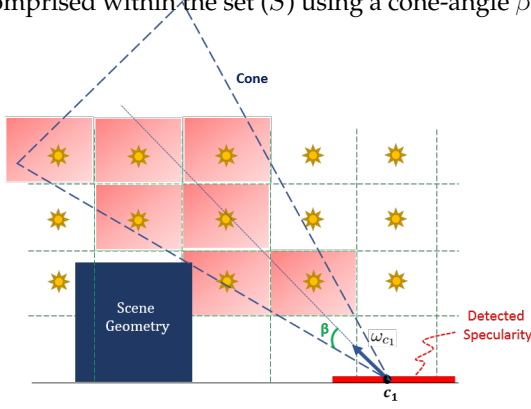


Fig. 8. Definition of a set  $(S)$  of hypothetical point lights (located at the center of red-color squares) from which actual real lights are recovered.

The second step consists in producing a synthetic *illumination* map  $\tilde{\delta}$  for every point light in the set  $(S)$ . Let us first consider equation (6), previously described to define the real *illumination* map  $\delta$ , in presence of one point light  $i$ :

$$\delta_i^p = L_a + (\mathbf{n}^p \cdot \omega_i^p) L_i O_i^p + \epsilon_S^p \quad (10)$$

where  $\delta_i^p$  represents the *illumination* map value at pixel  $p$  and  $\epsilon_S^p$  corresponds to present specular highlight at pixel  $p$ . The scalar product  $(\mathbf{n}^p \cdot \omega_i^p)$  represents the scene's shading and cast shadows result from the occlusion term  $O_i^p$  with regard to light source  $i$ . When a point  $p$  is geometrically occluded with regard to point light  $i$ , its *illumination* map value  $\delta_i^p$  is equal to  $L_a$  and it is located within a shadowed region ( $O_i^p = 0$  and  $\epsilon_S^p = 0$ ). On the contrary, when a point  $p$  is not occluded with regard to light  $i$ , its  $\delta_i^p$  value is defined by equation (10) and it does not belong to a cast shadow ( $O_i^p = 1$ ). By isolating the cast shadows effect, equation (10) can be rewritten as follows:

$$\begin{cases} \delta_i^p = L_a, & \text{if } p \text{ is light-occluded} \\ \delta_i^p = L_a + (\mathbf{n}^p \cdot \omega_i^p) L_i + \epsilon_S^p, & \text{otherwise} \end{cases} \quad (11)$$

In order to render a synthetic *illumination* map  $\tilde{\delta}$ , we must know all the parameters present in equation (11). However, since lighting characteristics  $(L_a, L_i, \omega_i^p)$  are not known in our case (our goal is to estimate them), we recover rough estimates of these parameters as follows:

$$\begin{cases} L_a = L_o, & \text{if } p \text{ is light-occluded} \\ (L_a + (\mathbf{n}^p \cdot \omega_i^p) L_i) + \epsilon_S^p = L_v + \sigma S_i^p, & \text{otherwise} \end{cases} \quad (12)$$

where:

- $\sigma$  is a binary term, equal to 1 if specular reflections are present and 0 otherwise. Specifically, specular

reflections are considered to be present if the detected pixels count within the recovered specularity mask  $(H)$  in section 3.2 is not null.

- $S_i^p$  is a synthetic specular map rendered, for a point light  $i$ , using Phong model [24]:

$$S_i^p = \mathbf{k}_s^p (\mathbf{r}_i^p \cdot \mathbf{v}^p)^{\alpha_p} L_i O_i^p \quad (13)$$

Because the 3D sensor delivers a coarse geometry of the scene, the rendering of  $S_i^p$  is achieved as follows: to begin with, the 3D model of the scene is clustered using [28] where the main planar surface is detected along with 3D objects lying on it (Figure 9-a). The rendering of the specular map  $S_i^p$  is then limited within the main planar surface, which is substituted by a perfect plane, in order to take account of the geometry's inaccuracies. To illustrate, in figure 9-(b,c), one can notice the effects of noisy geometric data in comparison with a perfectly modeled planar surface. The specular parameters value used to render the synthetic specular maps are respectively 1.0, 1.0 and 0.9 for  $\mathbf{k}_s^p$ ,  $L_i$  and  $\alpha_p$ .

- The terms  $L_v$  and  $L_o$  correspond respectively to the overall brightness in non-occluded/visible and occluded regions. Their computation is achieved in a three-step procedure: (i) for each clustered object, we define a proportional region of interested (ROI) recovered as the intersection of a sphere comprising the 3D object and the detected plane (Figure 10). (ii) *illumination* map values  $\delta^p$ , within the ROI, are increasingly sorted out and,  $L_o$  is recovered as the value at 25%. In fact, since these regions represent potential shadowed regions, the underlying assumption corresponds to having at least 25% of the ROI within a shadowed region. (iii)  $L_v$  is recovered as the mean of illumination values of pixels outside the ROI and detected highlights map  $H$ .

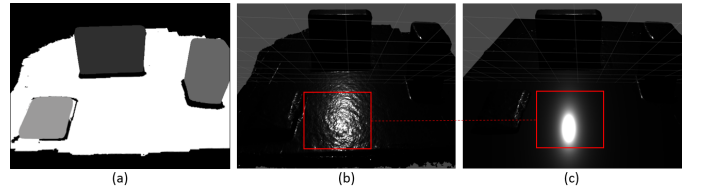


Fig. 9. 3D scene clustering and geometric inaccuracies: (a) Segmented 3D model of the scene: white-color pixels correspond to the detected plane, grayscale-color pixels represent 3D objects and black-color pixels represent background or noisy data. (b) Rendered specular map using the 3D model of the scene. (c) Rendered specular map using a perfect plane corresponding to the detected planar surface in the scene.

Finally, for every point light  $i$  in the set  $(S)$ , the rendering of its synthetic *illumination* map  $\tilde{\delta}_i$  (Figure 11-a) is achieved as follows:

$$\tilde{\delta}_i^p = \begin{cases} L_o, & \text{if } p \text{ is occluded} \\ L_v + \sigma S_i^p, & \text{otherwise} \end{cases} \quad (14)$$

In practice, since the geometry is static, occlusion  $O_i^p$  and specular maps  $S_i^p$  (Figure 11-(b,c)) are rendered only once. Only  $L_o$ ,  $L_v$  and  $\sigma$  are evaluated for each incoming frame since lighting conditions may change over time.



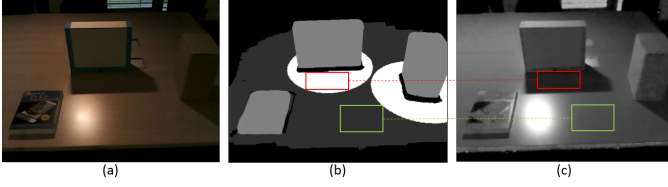


Fig. 10. Region Of Interest (ROI) for brightness analysis: (a) Color image of the scene. (b) Definition of the ROI: white-color pixels represent the ROI, grayscale-color pixels correspond to 3D objects in the scene and black-color pixels represent background or noisy data. (c) Recovered *Illumination map* with regard to (a).

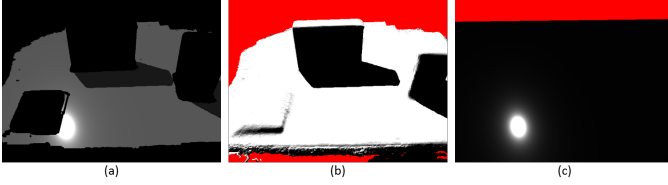


Fig. 11. Synthetic *illumination maps*: (a) Example of a synthetic *illumination map*  $\tilde{\delta}$  using rendered occlusion  $O_i^p$  (b) and specular  $S_i^p$  maps (c).

The final step consists in identifying the actual real light sources within the subset (S). The identification is carried within an iterative process as follows:

- We initially compute correlation values by matching recovered *illumination map*  $\delta$  and the rendered ones  $\tilde{\delta}$ . The light source whose synthetic *illumination map* has the best correlation value is selected.
- For each iteration, previously selected light sources are discarded. Also, previously matched pixels are not considered and point lights which are close to the previously selected ones are discarded. Both  $L_o$  and  $L_v$  are re-considered to be able to take account of shadows with different intensities (Figure 12).
- The process ends either when the currently selected  $\tilde{\delta}$  has a significantly low matching value or if the number of selected lights is higher than 4. The chosen correlation corresponds to Pearson's correlation coefficient ranging between 0 and 1.

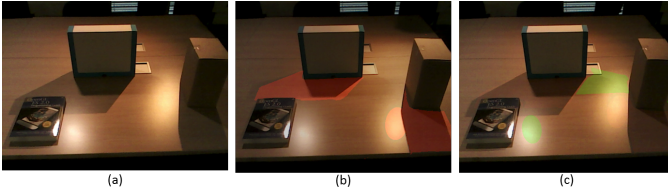


Fig. 12. Light sources 3D position estimation approach: (a) Color image of the scene. (b,c) Cast shadows and specular effects corresponding to the first (b) and second recovered light sources (c).

The proposed approach in this section recovers the 3D position of light sources in the scene. In the following section, we estimate their respective colors.

### 3.4 Light sources color estimation

Illuminating virtual objects within MR scenarios requires recovering the characteristics of light sources in the scene,

namely the 3D position and three-channel color vector (R,G,B). In the previous section, we used both specular highlights and cast shadows to recover the number of light sources along with their 3D positions. In this section, our goal is to estimate their respective colors. To achieve this task, we consider both the recovered illumination map  $\delta$  (Section 3.1) and subset (S) of point lights (Section 3.3). By considering equation (6) for a set of points  $p$  which belong to the main planar surface and are not detected as specular highlights (if the sensor saturates, their colors are not accurate), we obtain:

$$\begin{cases} \mathbf{L}_a + \sum_{i=1}^M (\mathbf{n}^{p_1} \cdot \omega_i^{p_1}) \mathbf{L}_i \mathbf{O}_i^{p_1} = \delta^{p_1} \\ \mathbf{L}_a + \sum_{i=1}^M (\mathbf{n}^{p_2} \cdot \omega_i^{p_2}) \mathbf{L}_i \mathbf{O}_i^{p_2} = \delta^{p_2} \\ \vdots \\ \mathbf{L}_a + \sum_{i=1}^M (\mathbf{n}^{p_N} \cdot \omega_i^{p_N}) \mathbf{L}_i \mathbf{O}_i^{p_N} = \delta^{p_N} \end{cases} \Rightarrow \mathbf{A} \mathbf{L} = \delta \quad (15)$$

where:

$$\mathbf{A} = \begin{pmatrix} 1 & (\mathbf{n}^{p_1} \cdot \omega_1^{p_1}) \mathbf{O}_1^{p_1} & \dots & (\mathbf{n}^{p_1} \cdot \omega_M^{p_1}) \mathbf{O}_M^{p_1} \\ 1 & (\mathbf{n}^{p_2} \cdot \omega_1^{p_2}) \mathbf{O}_1^{p_2} & \dots & (\mathbf{n}^{p_2} \cdot \omega_M^{p_2}) \mathbf{O}_M^{p_2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (\mathbf{n}^{p_N} \cdot \omega_1^{p_N}) \mathbf{O}_1^{p_N} & \dots & (\mathbf{n}^{p_N} \cdot \omega_M^{p_N}) \mathbf{O}_M^{p_N} \end{pmatrix}$$

$$\mathbf{L} = \begin{pmatrix} \mathbf{L}_a \\ \mathbf{L}_1 \\ \vdots \\ \mathbf{L}_M \end{pmatrix}; \quad \delta = \begin{pmatrix} \delta_1^p \\ \delta_2^p \\ \vdots \\ \delta_N^p \end{pmatrix}$$

where  $M$  is the number of recovered light sources and  $N$  is the number of considered pixels within equation (15). The linear system (15) is solved, for each color channel, using iterative Least Squares with bounds and equality constraints:

$$\hat{\mathbf{L}} = \arg \min_{\mathbf{L}} \left( \frac{1}{2} \|\mathbf{W}(\mathbf{A} \mathbf{L} - \delta)\|^2 \right) \quad \text{subject to:} \quad (16)$$

$$\begin{cases} 0 \leq \mathbf{L}_i \leq 1 \quad \text{and} \quad 0 \leq \mathbf{L}_a \leq 1 \\ \mathbf{L}_a + \sum_{i=1}^M \mathbf{L}_i = 1 \end{cases}$$

where  $\mathbf{W}$  is a diagonal matrix with weights computed using Tukey's bisquare loss function. Small weights are discarded throughout iterations and we recover color vectors for each light source in the scene.

### 3.5 Scene specular reflectance estimation

The estimation of scene reflectance is of interest for MR. For instance, if one considers adding a virtual light source in the real scene, real surfaces which retain a specular property must exhibit specular effects when viewed near the ideal specular reflection  $\mathbf{r}$  of the virtual lighting. Hence, in order to correctly render the virtual specularly, we must recover the surface's specular parameters, namely the specular reflectance  $\mathbf{k}_s^p$  and shininess coefficient  $\alpha_p$  for scene points  $p$ .

In section 3.2, we recovered a mask  $\mathbf{H}$  where detected specular reflections are retrieved. In the following, our goal is to estimate the specular reflection component within

detected specularities in  $H$  (points verifying  $H^p = 1$ ). By considering Phong model [24], the specular component  $I_s^p$  is described as follows:

$$I_s^p = I^p - k_d^p(L_a + \sum_{i=1}^M (n^p \cdot \omega_i^p) L_i O_i^p) \quad (17)$$

where  $I^p$  is the color of point  $p$  within the input color image and  $(k_d^p, L_a, \omega_i^p, L_i, O_i^p)$  are all parameters which we have estimated. Examples of recovered specular component  $I_s^p$  are shown in figure 13-b. In order to estimate the specular

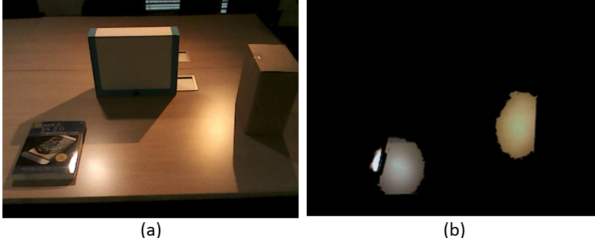


Fig. 13. Specular component of the scene: (a) Input images of the scene. (b) Recovered specular reflection component for input images in (a).

reflectance  $k_s^p$  of point  $p$ , we use Phong model [24] as follows:

$$k_s^p = \frac{I_s^p}{\sum_{i=1}^M (r_i^p \cdot v^p)^{\alpha_p} L_i O_i^p} \quad (18)$$

In fact, since specular reflections are viewed near the perfect specular reflection direction, vectors  $r_i^p$  and  $v^p$  are assumed to be roughly aligned. This assumption simplifies the denominator within equation (18) where the parameter  $\alpha_p$  is unknown. Consequently, equation (18) can be rewritten as:

$$k_s^p = \frac{I_s^p}{\sum_{i=1}^M L_i O_i^p} \quad (19)$$

Furthermore, because specular reflections are often observed only within parts of the scene, and we aim at estimating  $k_s^p$  for all scene points, we assume that each 3D object in the scene retains a unique specular reflectance. Consequently, the specular reflectance is recovered, for each object, as the maximum value of recovered  $k_s^p$  within points  $p$  belonging to the same object (Figure 14).

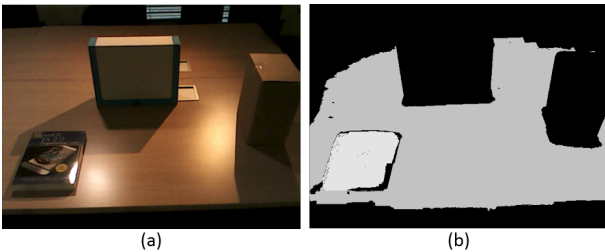


Fig. 14. Specular reflectance of the scene: (a) Input color image of the scene. (b) Estimated specular reflectance for each 3D object in the scene: the higher the brightness value, the more the surface is specular.

Finally, for each 3D object in the scene, the shininess parameter  $\alpha$  is recovered using the following loss function:

$$F(\alpha, k_s, \omega) = \sum_j (I^j - \tilde{I}^j(\alpha, k_s, \omega))^2 \quad (20)$$

where  $j$  iterates over pixels that belong to the considered object/cluster,  $I$  is the input color image and  $\tilde{I}$  is a rendered color image using Phong model (equation (2)). The optimization of function  $F$  is achieved using a Levenberg Marquardt algorithm where only the shininess coefficient  $\alpha$ , specular reflectance  $k_s$  and light sources positions  $\omega$  are varied by the solver. In fact, we refine  $k_s$  in order to take account for the approximation introduced in equation (19) as the reflection and view vectors may not be at optimal alignment within mask  $H$ . Also, since our light sources are recovered from a discrete set of hypothetical point lights (section 3.3), a trade-off between fine sampling and real-time constraints must be considered. Hence, we initially define a coarse sampling (1176 point lights with a sampling step of 20cm) and refine the positions using equation (20).

## 4 EXPERIMENTAL RESULTS

A calibrated RGB-D sensor browses the scene with a fixed aperture, shutter speed and gain. Using the acquired 3D model, we recover for each incoming color image, the reflectance and illumination of the scene. The framework runs at an interactive frame rate of 4fps. In the following, we evaluate the accuracy of the proposed framework within both synthetic and real scenes.

### Synthetic data

We consider synthetic scenes where ground-truth reflectance and illumination are available for comparison with our estimates. The synthetic dataset is composed of six scenes ('SynS1' to 'SynS6') with various shapes of 3D objects located on a main plane (Figure 15).

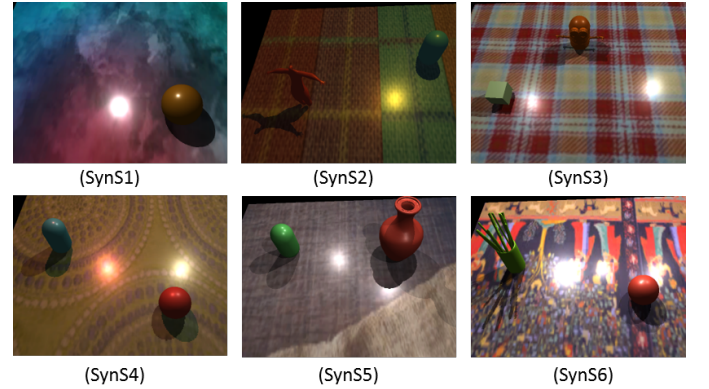


Fig. 15. Synthetic scenes: color image of six synthetic scenes ('SynS1' to 'SynS6') with various textures and geometries.

For each virtual 3D scene, the same inputs acquired or derived from the RGB-D sensor (R200) are rendered, namely the *reference* image (Fig-16) and color images captured under various lighting conditions. With regard to the scene's lighting, we have three lights which we freely move and turn on/off in the scene. Although our algorithm assumes point lights representation, we render the scenes using three different types of lights (point lights, spot lights and area lights). This allows us to better evaluate the limitations of the proposed method. Furthermore, we consider different lighting color vectors (R,G,B) to evaluate the accuracy of



our illumination estimation. The rendering is achieved in Unity engine using Phong model [24] where we generate 3017 color images using the 6 depicted virtual scenes. In the following, we evaluate both illumination and reflectance between the ground truth and estimated parameters for a set of rendered images.

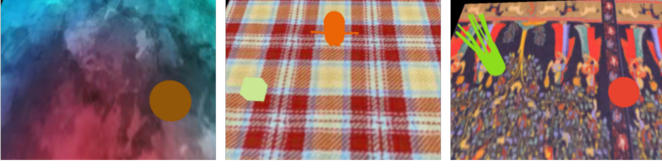


Fig. 16. Examples of rendered *reference* images with regard to scenes (SynS1, SynS3, SynS6) in Fig-16.

Our first test consists in evaluating the position of light sources. Table 1 shows a comparison between ground-truth and recovered positions for the six virtual scenes presented in figure 15. The results demonstrate the robustness of our approach in presence of challenging textures and lighting. In fact, the proposed algorithm recovers light sources position with an average error of 13cm for a mean distance of 3m to the light source. An interesting scenario is reported for SynS5: the scene is illuminated by three point lights which create overlapping cast shadows with different intensities. The approach recovers the correct number of light sources along with their 3D positions. Scene SynS6 is very challenging both in terms of texture and lighting: the cast shadows are weak due to a strong lighting in the scene. Although the 3D position of the second and third recovered light sources is not as accurate in comparison with the other scenes, we recover the correct number and orientations of light sources by exploiting the observed specular reflections.

Scene	GT pos. (x,y,z)			Estim. pos. (x,y,z)			Dist. error
SynS1	-0.78	3.89	2.54	-0.75	3.85	2.57	0.058
SynS2	0.92	2.01	2.23	0.94	2.04	2.21	0.041
SynS3	-1.38	2.71	2.62	-1.37	2.66	2.69	0.086
	1.57	2.65	2.16	1.62	2.67	2.11	0.073
SynS4	1.01	1.59	2.98	0.98	1.51	3.04	0.104
	-0.84	1.95	2.79	-0.86	1.89	2.71	0.101
SynS5	0.54	2.91	-1.62	0.59	2.87	-1.69	0.094
	0.67	1.59	1.98	0.72	1.51	2.04	0.111
	-0.87	1.12	2.79	-0.92	1.07	2.71	0.106
SynS6	-1.12	2.26	2.29	-1.19	2.21	2.37	0.117
	-1.32	2.41	2.47	-1.47	2.72	2.65	0.388
	-1.48	2.54	2.40	-1.65	2.69	2.57	0.283

TABLE 1

Comparison between ground truth and estimated 3D position of light sources in the virtual scenes ('SynS1' to 'SynS6').

The second evaluation concerns the color vector (R,G,B) of light sources as well as the specular reflectance of the scene. Within this task, we compare ground-truth images with rendered images using our estimates (Figure 17).

For the first three rows, scenes are lit respectively by one, two and three point light sources, their rmse are respectively 2.79%, 3.47% and 4.54%. Under point light sources assumption, we observe a general agreement between input and rendered views with a RMSE less than 5%. For instance, in row-1, the color of lighting as well as the specular reflectance are accurately recovered. The error is mainly present within the shadow contours resulting from

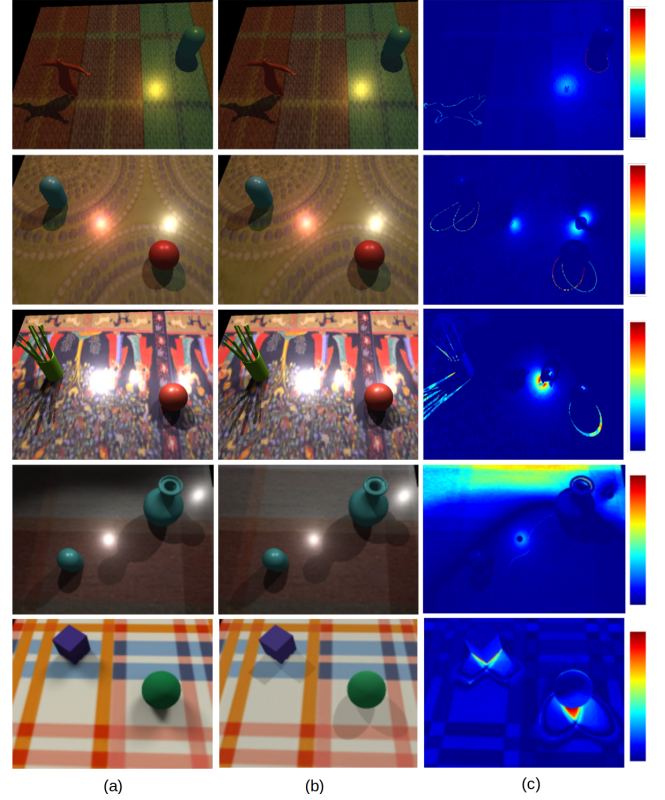


Fig. 17. Various virtual lighting conditions: (a) Input color images. (b) Rendered images using our reflectance and illumination estimates. (c) RMSE of the difference between images in (a) and (b): low and high values are represented respectively by blue and red color pixels.

the recovered 3D position of the light source. The most noticeable errors correspond to rows 4-5, where scenes are respectively lit by spot (9.32%) and area lights (8.65%). As can be noticed, the error is mainly due to the representation of light sources in the rendering pipeline. For instance, in row-4, errors are present in regions impacted by the spotlight cone effect. Similarly, in row-5, the main error is present in regions rendered using Ambient Occlusion (AO) within a global illumination pipeline. We also evaluate LAB color distance between ground-truth and recovered lights. Our algorithm recovers light sources color with an average distance metric of 3.7 - a barely noticeable color difference according to [29].

## Real data

In the following, we illustrate our results within a selection of five real scenes 'S1' to 'S5' grouped row-wise in figure 18. The scenes are composed of more than two objects located on a main plane. Both texture and reflectance properties vary within scene surfaces. For instance, scenes S1, S3 and S5 contain a planar surface with challenging textures. Also, scenes S2 and S4 exhibit specular reflections. Illumination-wise, scenes S1 and S2 are lit by one light source whereas scenes S3, S4 and S5 are lit by two light sources which create specular reflections if the scene surface is glossy (e.g. S4).

In figure 18-a, we present the captured *reference* images under a *near-ambient* lighting. These images are used to achieve the task of texture removal within color images (Figure 18-b) from which illumination is recovered. This results

in an *illumination* map  $\delta$  which mainly contains shading, shadowing and specular effects (Figure 18-c). Our algorithm recovers accurate *illumination* maps where the intrinsic texture/albedo is accurately separated from illumination-dependent effects.

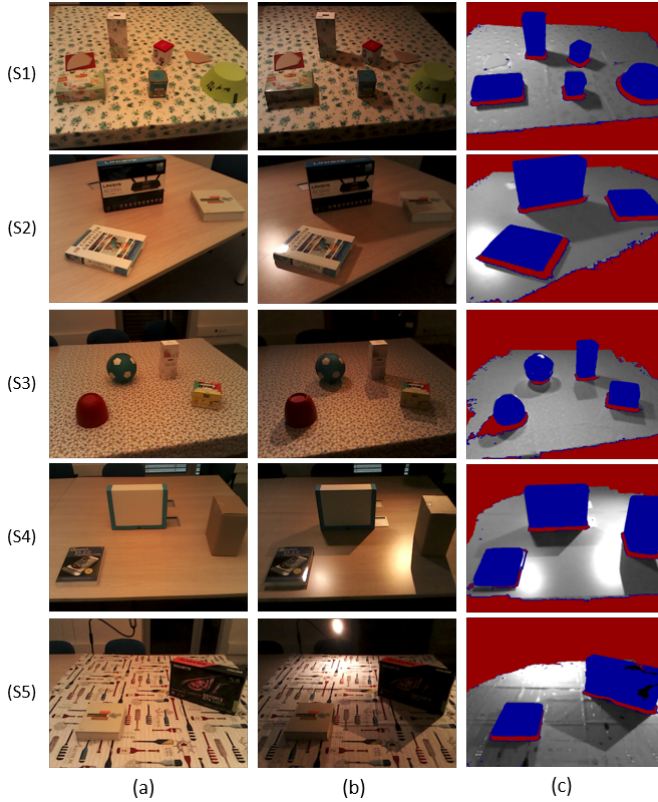


Fig. 18. Recovered *illumination* maps: (a) Reference image of the scene captured using an ambient lighting. (b) Input color image of the scene from which illumination is recovered. (c) Estimated *illumination* maps for uniform (S2,S4) and textured surfaces (S1,S3,S5): background/noise are represented by red-color pixels and occluding objects by blue-color pixels. Grayscale values correspond to the intensity of illumination values  $\delta$  (section 3.1).

Furthermore, recovered *illumination* maps are considered to estimate the 3D position of light sources in the scene. This is achieved by initially using specular reflections, when available, to estimate a rough direction of light sources. Then, within a matching process, we recover the position of light sources represented by point lights. In figure 19-b, we show detected specular effects (cyan pixels) for the five real scenes and overlay the shadow maps of recovered light sources (Figure 19-c) on the current color image. Our approach robustly recovers illumination in the scene in presence of specular effects (S2,S4) and challenging textures (S1,S3,S5). Moreover, within (S2, S4), the detected specular reflections allow us to consider less than 30% of the initial hypothetical point lights within the cast-shadow analysis. The proposed method handles overlapping shadows (S3), weak cast shadows (green pixels in S4) as well as shadows which do not retain a uniform intensity due to a strong near lighting (S5).

In order to evaluate the efficiency of incorporating specular effects within our framework, we compare our method with [15] where correlation is carried using binary shadow

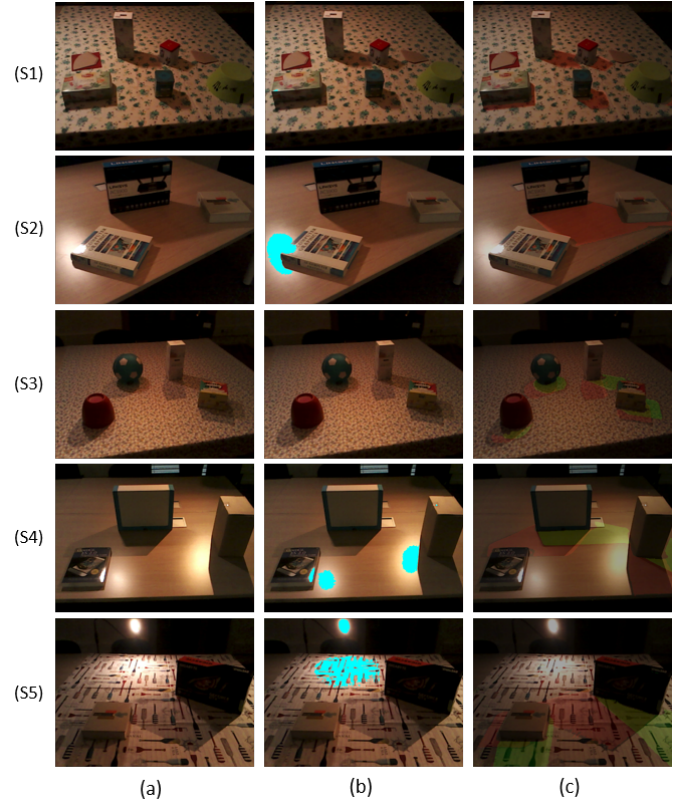


Fig. 19. Recovered specular reflections and cast shadows: (a) Color image of captured scenes. (b) Detected specular reflections are represented by cyan pixels. (c) Overlay of shadow maps corresponding to estimated light sources: first and second best matches are respectively represented by red and green color pixels.

maps. In figure 20-Scenario1, both methods achieve comparable results with regard to the recovered position of the light source. However, in more critical scenarios such as Scenario2 where the spot-cone effect is present, pixels belonging to this region are considered as potential shadows and erroneously contribute to the selection of matching shadow maps within the iterative correlation process. In contrast, by incorporating the specular effects, detected specular pixels allow to recover a light source whose synthetic cast shadow and specular reflection match best with the current input image. Furthermore, in Scenario3, green pixels correspond to the second recovered light source's shadow map. In fact, in presence of weak shadows, the proposed adaptive synthetic *illumination* map delivers better results.

In order to evaluate the precision of recovered light sources positions, we used a telemeter to measure the distance from a chosen world coordinate system to the light sources in the scene. Results for the five real scenes are shown in table 2. Our algorithm recovers light sources position with an average error of 9cm for a mean distance of 1.62m to the light source and a standard deviation of 3.2cm.

The goal of photometric registration is to achieve realistic mixed reality scenarios. In this context, we considered both static and moving camera scenarios. In figure 21, we show realistic augmentations of real scenes using a static



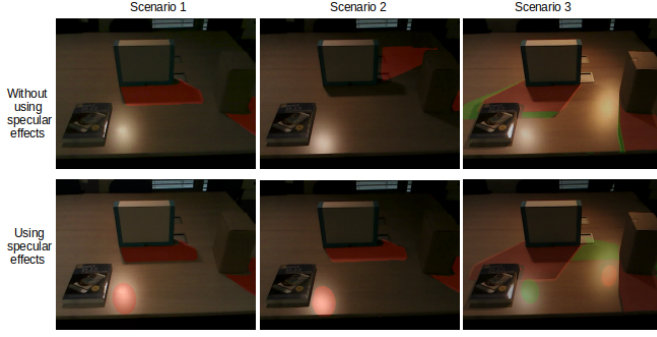


Fig. 20. Comparison within three different scenarios of the estimated light sources position results without (first row) and with (second row) incorporating specular effects analysis.

Scene	Measured (m)	Estimated (m)	Error (m)
S1	1.83	1.92	0.09
S2	1.68	1.61	0.07
S3	1.74	1.81	0.07
	1.72	1.83	0.11
S4	1.44	1.52	0.08
	1.63	1.75	0.12
S5	1.12	1.19	0.07
	1.52	1.61	0.09

TABLE 2

Comparison between measured and estimated distances using our proposed approach.

camera (*reference* and incoming color images are aligned). For instance, for scenes (S1,S3,S5), virtual shadows cast by a sphere are consistent with real shadows in terms of shape and color. Furthermore, within scene S2, we show a correct occlusion of a real specularly by a virtual object. One can notice the reconstructed texture within the specular area. In the submitted **video**, we further address the scenario of a moving camera by using a second fixed camera with a bird-eye view at the MR scene. This camera is connected to a server to which it sends the color stream. The lighting estimation is accomplished within the server and lighting estimates are sent to a mobile MR camera that the user can freely move to experiment MR under arbitrary lighting.

We further consider the scenario of retexturing the scene while preserving the current illumination. This is achieved in real-time and corresponds to the product of the *illumination* map  $\delta$  and a target texture  $\mathbf{T}$  (target diffuse reflectance):

$$\mathbf{I}_{retex}^p = \delta^p \mathbf{T}^p \quad (21)$$

where  $\mathbf{I}_{retex}^p$  is the color of the re-textured scene and  $p$  corresponds to points which belong to the main planar surface in the example shown in figure 22. Inaccuracies are mainly due to coarse or unavailable geometry (red-color pixels in figure 18-c).

## 5 LIMITATIONS AND FUTURE WORK

We presented a photometric registration approach which incorporates information brought by specular reflections and cast shadows to recover reflectance and illumination in the scene. Specifically, our method estimates both 3D position and color of dynamic light sources as well as the specular reflectance of surfaces. Our experimental results show satisfactory results on both synthetic and real data

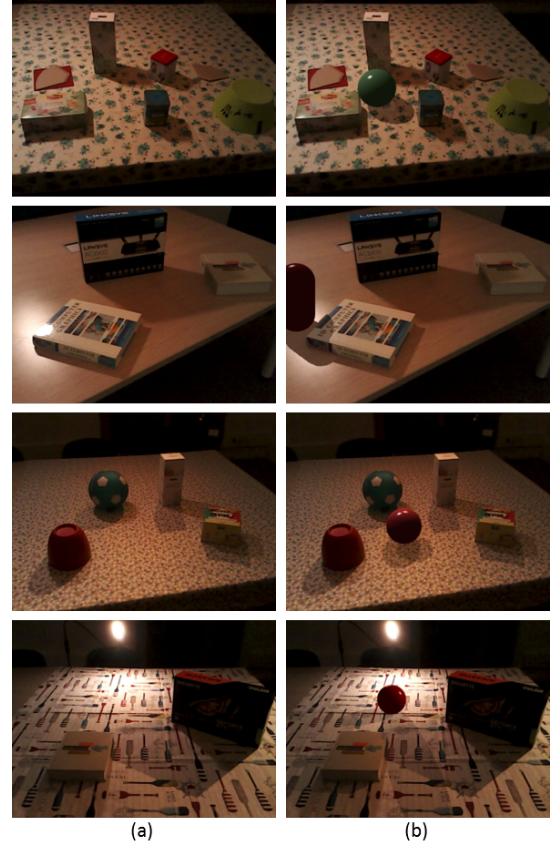


Fig. 21. Realistic augmentations of real scenes using our reflectance and illumination estimates. We demonstrate visually consistent virtual shadows in terms of shape and color with regard to real cast shadows. The second row is an example of a virtual object occluding a real specularity. One can notice a correct reconstruction of texture within the specular region.

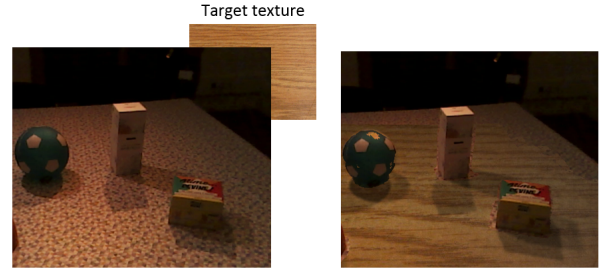


Fig. 22. Retexturing scenario: (a) Input color image of the scene with the target texture  $\mathbf{T}$  (top right). (b) Retextured main planar surface using the *illumination* map  $\delta$  and texture  $\mathbf{T}$ .

where challenging textures are correctly handled and the presence of specular effects is efficiently handled.

Although the assumption of having a main planar surface is not very constraining for MR scenarios (e.g., table, desk, playground, floor, etc.), one may encounter configurations where this assumption does not hold. In fact, the main reason behind this assumption is related to the coarse geometry provided by common RGB-D sensors. We are therefore interested in handling more generic 3D models. Finally, the proposed approach requires to capture the scene under an near-ambient lighting to accurately separate texture from illumination in color images. An interesting and challenging research direction corresponds to achieving this task using

only the color image and 3D model of the scene.

## REFERENCES

- [1] S. Gibson, T. Howard, and R. J. Hubbard, "Flexible image-based photometric reconstruction using virtual light sources," *Comput. Graph. Forum*, vol. 20, 2001.
- [2] B. Boom, S. Orts, X. Ning, S. McDonagh, P. Sandilands, and R. Fisher, "Point light source estimation based on scenes recorded by a rgb-d camera," in *British Machine Vision Conference*, 2013.
- [3] I. Arief, S. McCallum, and J.Y. Hardeberg, "Realtime estimation of illumination direction for augmented reality on mobile devices," *Color and Imaging Conference*, pp. 111–116, 2012.
- [4] P. Debevec, "Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography," in *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*, 1998, pp. 189–198.
- [5] K. Rohmer, W. Baschel, R. Dachsel, and T. Grosch, "Interactive near-field illumination for photorealistic augmented reality on mobile devices," in *IEEE Int. Symp. on Mixed and Augmented Reality (ISMAR)*, 2014, pp. 29–38.
- [6] M. Meilland, C. Barat, and A. Comport, "3d high dynamic range dense visual slam and its application to real-time object re-lighting," 10 2013, pp. 143–152.
- [7] J. Unger, J. Kronander, P. Larsson, S. Gustavson, J. Low, and A. Ynnerman, "Spatially varying image based lighting using hdr-video," 2013, vol. 37, pp. 923 – 934.
- [8] J. Kronander, F. Banterle, A. Gardner, E. Miandji, and J. Unger, "Photorealistic rendering of mixed reality scenes," 05 2015, vol. 34, pp. 643–665.
- [9] 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 *IEEE Int. Conf. on Computer Vision*, 2001, vol. 1, pp. 599–606 vol.1.
- [10] J. Jachnik, R. Newcombe, and A. Davison, "Real-time surface light-field capture for augmentation of planar specular surfaces," in *IEEE Int. Symp. on Mixed and Augmented Reality*, 2012, pp. 91–97.
- [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," *Proceedings of the ICCV*, vol. 2, 12 1999.
- [12] I. Sato, Y. Sato, and K. Ikeuchi, "Illumination from shadows," *IEEE on Pattern Analysis and Machine Intelligence*, vol. 25, no. 3, pp. 290–300, 2003.
- [13] A. Panagopoulos, D. Samaras, and N. Paragios, "Robust shadow and illumination estimation using a mixture model," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2009, pp. 651–658.
- [14] A. Panagopoulos, C. Wang, D. Samaras, and N. Paragios, "Illumination estimation and cast shadow detection through a higher-order graphical model," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2011, pp. 673–680.
- [15] S. Jiddi, P. Robert, and E. Marchand, "Estimation of position and intensity of dynamic light sources using cast shadows on textured real surfaces," in *IEEE Int. Conf. on Image Processing (ICIP)*, 2018, pp. 1063–1067.
- [16] A. Bunteong and N. Chotikakamthorn, "Light source estimation using feature points from specular highlights and cast shadows," in *Int. Journal of Physical Sciences*, 2016, vol. 11, pp. 168–177.
- [17] Y. Li, S. Lin, H. Lu, and H.Y. Shum, "Multiple-cue illumination estimation in textured scenes," in *IEEE Int. Conf. on Computer Vision*, 2003.
- [18] M.A. Gardner, K. Sunkavalli, E. Yumer, X. Shen, E. Gambaretto, C. Gagne, and J.F. Lalonde, "Learning to predict indoor illumination from a single image," 03 2017, vol. 36.
- [19] Z. Li, Z. Xu, R. Ramamoorthi, K. Sunkavalli, and M. Chandraker, "Learning to reconstruct shape and spatially-varying reflectance from a single image," 12 2018, vol. 37, pp. 1–11.
- [20] M. Garon, K. Sunkavalli, S. Hadap, N. Carr, and J.F. Lalonde, "Fast spatially-varying indoor lighting estimation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
- [21] S. Song and T.A. Funkhouser, "Neural illumination: Lighting prediction for indoor environments," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
- [22] B. Marques, R. Drumond, C. Vasconcelos, and E. Gonzalez Clua, "Deep light source estimation for mixed reality," in *VISIGRAPP*, 2018.
- [23] K. Karsch, K. Sunkavalli, S. Hadap, N. Carr, H. Jin, R. Fonte, M. Sittig, and D. Forsyth, "Automatic scene inference for 3d object compositing," 05 2014, vol. 33, pp. 1–15.
- [24] B.T Phong, "Illumination for computer generated pictures," *Commun. ACM*, vol. 18, no. 6, pp. 311–317, 1975.
- [25] B. Thai, G. Deng, and R. Ross, "A fast white balance algorithm based on pixel greyness," *Signal, Image and Video Processing*, vol. 11, no. 3, pp. 525–532, 2017.
- [26] F. Ortiz and F. Torres, "Automatic detection and elimination of specular reflectance in color images by means of ms diagram and vector connected filters," *IEEE on Systems, Man, and Cybernetics*, vol. 36, pp. 681–687, 2006.
- [27] C. H. Teh and R. T. Chin, "On the detection of dominant points on digital curves," *IEEE on Pattern Analysis and Machine Intelligence*, vol. 11, no. 8, pp. 859–872, 1989.
- [28] R. Rusu, *Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments*, Ph.D. thesis, Computer Science department, Technische Universitaet Muenchen, Germany, 2009.
- [29] Gaurav Sharma, *Digital Color Imaging Handbook*, CRC Press, Inc., Boca Raton, FL, USA, 2002.



**Salma Jiddi** is now a research scientist at Geomagical Labs. She received her Ph.D in Computer Science from Université de Rennes 1. Her research focuses on computer vision and computer graphics related subjects such as reflectance and illumination estimation, 3D scene reconstruction, camera pose estimation and realistic rendering. Her doctoral work mainly concerned photometric registration approaches using a single RGB-D sensor. In 2015, she obtained her engineering degree from the ENSEIRB-MATMECA school in Bordeaux where she studied electronics with a major in signal and image processing.



**Philippe Robert** received his Ph.D from Université de Rennes 1 and joined Thomson/Technicolor in 1986 as a research engineer. He is now with Interdigital. His main activities have concerned the development of algorithms of image sequence processing in particular motion estimation and frame interpolation, detection, segmentation and tracking of objects, color correction, inpainting, shape from motion. They have also concerned stereo, multiview and multi-sensor processing and fusion for 3D reconstruction and modeling, and 2D to 3D conversion. He is author or co-author of 50 papers and about 80 patents.



**Eric Marchand** is a professor of computer science at Université de Rennes 1 in France and a member of the Inria/IRISA Rainbow team. He received a Ph.D degree and a "Habilitation Diriger des Recherches" in Computer Science from the Université de Rennes 1 in 1996 and 2004 respectively. He has been an INRIA research scientist at INRIA Rennes Bretagne Atlantique from 1997 to 2009. His research interests include robotics, visual servoing, real-time object tracking and augmented reality. He received a Best application paper award at IEEE IROS 2007 and Best paper runner-up award at IEEE ISMAR 2010 and IEEE 3DUI 2011. He has been an associate editor for IEEE Trans. on Robotics (2010-2014), for IEEE Robotics and Automation Letters (2015-2018) and guest co-editor of The Int. J. of Robotics Research (IJRR) special issue on "Robot Vision" (april 2015). From June 2018, he is a senior editor for IEEE Robotics and Automation Letters (RA-L).