

Lighting Estimation in Indoor Environments from Low-Quality Images

Natalia Neverova, Damien Muselet, Alain Trémeau

Laboratoire Hubert Curien – UMR CNRS 5516, University Jean Monnet,
Rue du Professeur Benoît Luras 18, 42000 Saint-Étienne, France
natalia.neverova@etu.univ-st-etienne.fr
{[damien.muselet](mailto:damien.muselet@univ-st-etienne.fr),[alain.tremeau](mailto:alain.tremeau@univ-st-etienne.fr)}@univ-st-etienne.fr
<http://laboratoirehubertcurien.fr>

Abstract. Lighting conditions estimation is a crucial point in many applications. In this paper, we show that combining color images with corresponding depth maps (provided by modern depth sensors) allows to improve estimation of positions and colors of multiple lights in a scene. Since usually such devices provide low-quality images, for many steps of our framework we propose alternatives to classical algorithms that fail when the image quality is low. Our approach consists in decomposing an original image into specular shading, diffuse shading and albedo. The two shading images are used to render different versions of the original image by changing the light configuration. Then, using an optimization process, we find the lighting conditions allowing to minimize the difference between the original image and the rendered one.

Key words: light estimation, depth sensor, color constancy.

1 Introduction

Nowadays, there are growing demands in the context of augmented reality applications, which is a subject of enormous attention of researchers and engineers. The ability to augment real scenes with arbitrary objects and animations opens up broad prospects in the areas of design, entertainment and human-computer interaction. In this context, correct estimation of lighting conditions (3D positions and colors) inside the scene appears to be a crucial step in making the rendering realistic and convincing.

Today, there exist solutions that require complex hardware setup with high dynamic / high resolution cameras and light probes [1]. Instead, our goal is to design a system that can be used at home by any user owning a simple and cheap RGB-D sensor. In this context, a nice solution has been proposed in [2] but that solution requires the user to specify the geometry of the scene, the object interactions and the rough positions and colors of the light sources. Furthermore, that approach is adapted to simple scene geometries since the environment is represented as a cube.

In this paper, we show that using cheap depth sensors (such as Microsoft Kinect) allows to avoid these requirements of light probes, multiple user interactions and simple geometries. Indeed, from the rough geometric information provided by such a sensor, we can simulate different versions of an observed scene under different lighting conditions. Then, we can estimate the light conditions that minimize the difference between the rendered image (with estimated light) and the target image (i.e. the original one).

The contributions of our approach are threefold. First, we propose a new iterative algorithm allowing to estimate light colors in low-quality images. Second, unlike the classical approaches, we account the specular information in the rendering process and show in the experiments that this information improves the light estimation. Finally, we propose a rough light position estimation that is used for the initialization of the optimization process.

The rest of the paper is organized as follows: first we will provide a brief overview of state-of-the-art methods of light estimation and image decomposition. Then in Section 3 we will describe the main ideas of the proposed method and justify it from the physical point of view. In Section 4 we propose a way to initialize the optimization problem introduced in Section 3. Section 5 contains some experimental results and Section 6 concludes the paper and provides some details on future works.

2 Related work

Light estimation. Light estimation is one of the most challenging problems in computer vision, especially when it comes to indoor scenes. Presence of multiple light sources of different sizes and shapes, intensities and spectral characteristics is a typical situation for this kind of environments. The image based lighting approach described in [1] is one of the most advanced techniques of light modeling that allows to obtain high quality results but at cost of processing time. The main limitations of this approach are that it requires a complex hardware setup with additional cameras and/or light probes and is based on high dynamic and high resolution imaging. A modified approach proposed in [3] allows to directly estimate positions of light sources but is also based on using cumbersome hardware. One of the most popular alternatives to image-based lighting approaches aims on detection and direct analysis of shadings. These techniques are generally more suitable for outdoor environments with strong casts, directed light sources and simple geometry. An exhaustive survey of cast detection methods in different contexts is provided in [4], while [5] explores the possibility of their integration in real-time augmented reality systems. Finally, we must mention a recent work [2] exploiting an idea of light estimation and correction through rendering-based optimization procedure. This approach is the closest to the one proposed in this paper, therefore we will address to some parts of their work in the following sections.

Intrinsic images. In order to render the image with estimated lighting, it is recommended to decompose the color of each pixel into albedo and shading [6,

7]. Land and McCann proposed in 1975 the Retinex theory assuming that albedo is characterized by sharp edges while shading varies slowly [8]. Inspired by this work, some papers have tried to improve the decomposition results [9]. Most of the intrinsic image decompositions assume diffuse reflection and neglect the specular reflection. Using jointly segmentation and intrinsic image decomposition, Maxwell et al. [10] account the specularities in the decomposition but this kind of approach is not adapted to low-quality images acquired under uncontrolled conditions [7]. Thus, to overcome the presence of highlights, a preliminary step consists in separating specular and diffuse reflection [11] and then applying the intrinsic decomposition on the diffuse image. In order to decompose an image into diffuse and specular components we take advantage of the simplicity of the method proposed by Shen et al. [12]. However, since they assumed known illuminant, we propose to modify their approaches in order to estimate the light color during the process.

3 Light estimation through optimization

3.1 Assumptions and Workflow

In the first step, in order to simplify the estimation lighting process, we have considered the following assumptions. First, we assume that all light sources illuminating a scene have the same chromaticity. Second, we assume that the specular reflectance distribution (ρ_s in eq. 1) is the same over all the surfaces in a scene. This assumption which is not true from a theoretical point of view does not disturb the light estimation in practice. However, for the rendering of synthetic object, we can account different specular reflectance distributions. Moreover, we assume that the lights are planar and of negligible sizes. And last, we assume dichromatic reflection model as presented in Fig. 1.

As can be seen on Fig. 1, our approach consists in decomposing a color image into three images. First, we use a similar approach as [12] in order to separate diffuse and specular reflections. However, we modify the original process in order to evaluate the overall light color that we assume constant over the whole image. Then, from the diffuse reflection image, we apply a Retinex based decomposition [13] since Retinex has been shown to provide nice results in [6]. This intrinsic decomposition provides the shading and albedo images. The obtained specular A^0 and diffuse B^0 shading images are the inputs of the optimization process. They are independently compared with the rendered specular A and diffuse B shading images, which are obtained from geometric information provided by the Kinect and from the initial light condition estimation L^0 . Then the light conditions L are iteratively updated until the difference between real and rendered images is minimum.

3.2 Reflection model

Thanks to the depth information provided by the Kinect, we are able to account both diffuse and specular reflection in the rendering images used during the optimization. Consequently, we can consider the dichromatic reflection model [14]

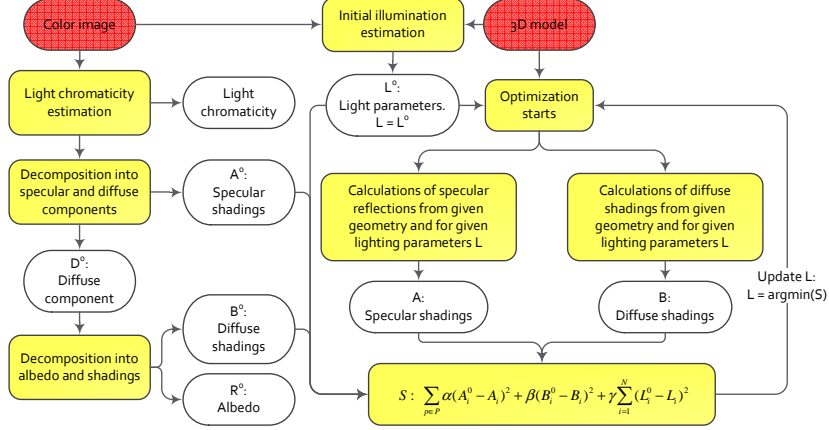


Fig. 1. Workflow of the proposed method

and more specifically, the Phong model [15] that estimates the spectral power distribution of the light reflected by a given surface point illuminated by N light sources $I_{0,i}(\lambda)$ as:

$$I(\lambda, p)_N = -\rho_d(\lambda, p) \sum_{i=1}^N I_{0,i}(\lambda) \frac{(\mathbf{n}_{s,i}, \mathbf{d}_i)(\mathbf{n}, \mathbf{d}_i)}{\|\mathbf{d}_i\|^4} + \quad (1)$$

$$\rho_s \sum_{i=1}^N I_{0,i}(\lambda) \frac{(\mathbf{n}_{s,i}, \mathbf{d}_i)}{\|\mathbf{d}_i\|^{k+1}} (\mathbf{v}, (\mathbf{d}_i - 2(\mathbf{n}, \mathbf{d}_i)\mathbf{n}))^k,$$

where $\rho_d(\lambda, p)$ is the diffuse reflectance of the considered surface point, the brackets model the dot product, k is a coefficient that can be defined and set experimentally (in our implementation we set $k = 55$), $\rho_s(\lambda, p)$ is the specular reflectance and all the other parameters $(\mathbf{n}_s, \mathbf{n}, \mathbf{d}, \mathbf{v})$ are introduced in Fig. 2. Assuming neutral specular reflection, as it is usually done, and constant maximum specular reflectance over the scene, $\rho_s(\lambda, p) = \rho_s$ is a constant in a given scene.

From this equation, we propose to extract two terms that only depend on the geometry of the scene (viewing direction, surface orientation and light position) and not on the reflection properties of the surfaces:

$$B(p) = \left[\frac{(\mathbf{n}_{s,i}, \mathbf{d}_i)(\mathbf{n}, \mathbf{d}_i)}{\|\mathbf{d}_i\|^4} \right] \text{ called diffuse shading;} \quad (2)$$

$$A(p) = \left[\frac{(\mathbf{n}_{s,i}, \mathbf{d}_i)}{\|\mathbf{d}_i\|^{k+1}} (\mathbf{v}, (\mathbf{d}_i - 2(\mathbf{n}, \mathbf{d}_i)\mathbf{n}))^k \right] \text{ called specular shading.} \quad (3)$$

Starting from depth images, the only unknowns in these equations are the light positions and orientations. So, given an image provided by the Kinect, we

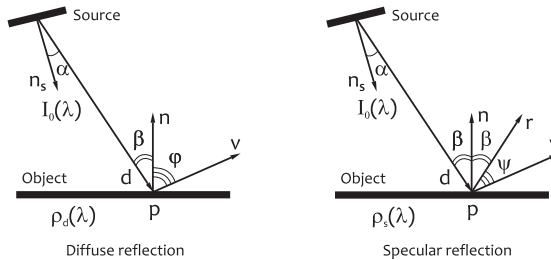


Fig. 2. Diffuse and specular reflections. \mathbf{n}_s – normal vector to the surface of the planar light source ($|\mathbf{n}_s| = 1$), \mathbf{d} – vector connecting the light source with the given object point ($|\mathbf{d}| = d$); α – angle between vectors \mathbf{n}_s and \mathbf{d} , \mathbf{n} – normal vector to the surface of the object ($|\mathbf{n}| = 1$), β – angle of incident light (between vectors \mathbf{n}_s and \mathbf{n}), \mathbf{v} – viewing direction ($|\mathbf{v}| = 1$), φ – angle between the surface normal and the viewing direction, $I_0(\lambda)$ – intensity of the light source in the direction perpendicular to its surface, $\rho_d(\lambda)$ – diffuse reflectance, $\rho_s(\lambda)$ – specular reflectance.

can render these two shading images (A and B) for all light geometries we want. The idea of the next step is to be able to extract these shading images (A^0 and B^0) from the original color image in order to find the best light geometries that minimize the differences between the rendered images A and B and their corresponding original images A^0 and B^0 (see Fig. 1).

3.3 Color image decomposition

To decompose an image into diffuse and specular components we consider and improve the method proposed in [12]. In this paper, the authors generated a specular-free image from a color image by subtracting from each pixel the minimum of its RGB values and adding a pixel-dependent offset. This simple approach provides good results when the light color is known and the image is normalized with respect to this color and rescaled to the range $[0, 255]$. In our case, the light color is unknown and we propose to estimate it during the process. Thus, in the first step, we assume white light and we run the algorithm on the original image. Then, once the specular component is separated from the diffuse one, the chromaticity of the illuminants is estimated as the mean chromaticity of the detected specular pixels. After that the original image can be normalized with respect to the "new" light color and specular component can be recalculated using the same formula. Consequently, we propose an iterative process that successively applies specular detection and light chromaticity estimation until convergence. Today we have no proof about the convergence properties but in practice, a maximum of 3 iterations are required to obtain stable specular image and light chromaticity on the tested images.

After running this algorithm, we obtain the light chromaticity, the specular shading image called A^0 and the diffuse image called D^0 . The diffuse image can be further decomposed into diffuse shading and albedo terms. It can be done

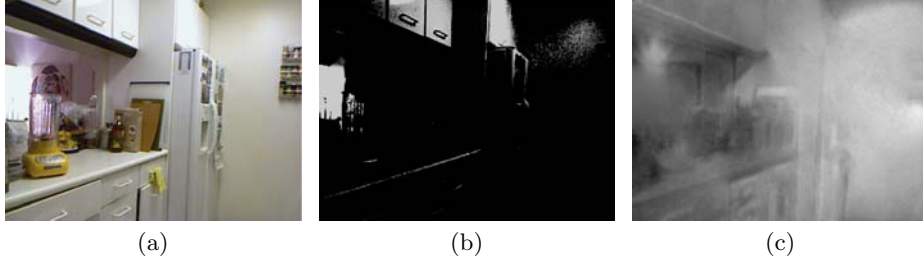


Fig. 3. Decomposition of a color image (a) into specular (b) and diffuse (c) components.

in different ways, but in this work we use the Retinex theory [8] that proved to be a state-of-the-art method of intrinsic image decomposition [6]. Here we use a fast implementation of Retinex proposed in [13]. The diffuse shading image is called B^0 . The decomposition is illustrated in Fig. 3.

3.4 Optimization

In the previous sections, we have explained how to render diffuse (B) and specular (A) shadings using the equations (2) and (3) respectively and how to obtain the corresponding original images B^0 and A^0 from the considered color image. The idea of the optimization step is to evaluate the light conditions that minimize the differences between these images:

$$L = \operatorname{argmin} \left\{ \sum_{p \in P} \alpha [A^0(p) - A(p)]^2 + \beta [B^0(p) - B(p)]^2 + \gamma \sum_{i=1}^N [L_i^0 - L_i]^2 \right\}, \quad (4)$$

where α , β and γ are coefficients set experimentally (in our implementation we set $\alpha = 1$, $\beta = 0.75$, $\gamma = 30M_xM_y$, where $M_x \times M_y$ is the size of the image).

The last term ($L_i^0 - L_i$) of the equation (4) constraints the process not to move far away from the initial light position estimation L_i^0 . Indeed, in a preprocessing step (detailed in the next section), we can roughly estimate the potential 3D position L_i^0 of all the lights i in the scene and the value of the coefficient γ depends on how confident we are in this first estimation. The next section explains different ideas on how to perform this estimation.

It is important to note here that the previous equation is used to optimize the light positions, but it could be easily extended to optimize both the positions and the colors of the lights. In this case, we would have to render the color image with equation (2) and compare it with the original color image.

4 Discussion about initialization

For initialization, we need to specify the number of light sources and their approximate positions. By detecting the areas of maximum intensities in specular



Fig. 4. Some images used for light color estimation.

Table 1. Mean angular error obtained on the images of Fig. 4.

Grey-world	MaxRGB	Shades of grey	Grey edge	our proposition
1.10	0.96	2.84	1.03	0.43

spots and knowing the surface orientation of these areas and the position of the camera, we can estimate the direction of reflected light. That gives an approximate direction of light sources locations. By specifying several points of specular reflections on different surfaces and finding the intersection points of corresponding lines, we can also find the distances to the sources. If there are several light sources illuminating the scene, different specular reflections will correspond to different positions. In this case all rays can be combined in several groups and number of sources and their directions can be roughly defined. We propose to do it with a greedy algorithm based on a voting scheme consisting of accumulation and search steps.

5 Experiments

5.1 Light color

In order to check the results of our iterative process that allows to estimate the light color, we have acquired a set of images, containing color target as ground truth (see Fig. 4). We have mentioned that the color images provided by the Kinect device are noisy and of low resolution. Therefore we wanted to assess the quality of the results provided by the classical color constancy algorithms in this context. We have tested the following algorithms [16]: Grey-world, MaxRGB, Shades of grey, Grey edge and our proposition. For each algorithm, we have evaluated the mean angular error as recommended in [16].

The results are displayed in table 1. We can see that algorithms based on the analysis of the edges do not perform well on this low-quality images. MaxRGB that is the nearest approach to our proposition provides good results but our approach outperforms all the tested methods. The advantage of our method compared to MaxRGB is that it is based on the detection of specular areas by using a pixel-dependent offset and this can help in the case of low-quality images [2]. The use of our iterative process also helps in this detection step.

5.2 Light positions

We tested our method on some color images from NYU Depth V1 dataset [17]. This dataset contains 2284 VGA-resolution images of various indoor environments together with corresponding depth maps taken with the Kinect sensors. Since there are no other works trying to estimate light conditions from depth sensors, we can not show comparison results. Instead, we propose to show one convincing example illustrating how our approach is improving the state-of-the-art method [2], while not requiring any user interaction. Let us consider the top color image from Fig. 5. We have run different optimization procedures on this image, starting from the same initial 3D light positions:

- Method 1: we first apply our decomposition (specular vs. diffuse and then albedo vs. shading on the diffuse) and consider only the diffuse shading image B^0 for optimization, i.e. we optimize only $[B^0(p) - B(p)]^2$,
- Method 2: same as method 1 but considering only the specular shading image A^0 for optimization, i.e. we optimize only $[A^0(p) - A(p)]^2$,
- Method 3: same as method 1 but considering both the specular shading image A^0 and the diffuse shading image B^0 for optimization, i.e. we optimize equation (4),
- Method 4: we neglect the specular reflection and just decompose the original color image into albedo and shading and we optimize the shading part (similar to [2]).

On Fig. 5, for each method (from 1 to 4), we have plotted a cross corresponding to the center of the highlight that would be obtained if the light position was returned by this method. This is a good way to compare the returned position estimations from the different methods. On the image, we can see that crosses corresponding to methods 2 and 3 are the nearest to the real highlight. Since the cross 4 is very far from the real highlight, we can conclude that the specular reflection should not be neglected during the optimization process. Indeed, in this case, the highlight is considered as a diffuse spot and the algorithm try to optimize the light position so that we obtain this diffuse spot, leading to a high position error. This illustration validates the importance to apply first the multiple decompositions and to optimize both specular and diffuse shadings. Then, we show on the second and third rows of this figure the diffuse and specular rendering of each method (column j corresponds to method j). We can see that by only considering the specular component (column 2), the diffuse rendering is not correct because the distance between the light and the wall is hard to estimate from specularities only. Fortunately, the third column (proposed method) displays the best results, showing that both specular and diffuse components have to be used for the estimation.

Thus, this illustration shows that our approach is able to better estimate positions of light source present in the scene using color and depth data provided by a depth sensor.

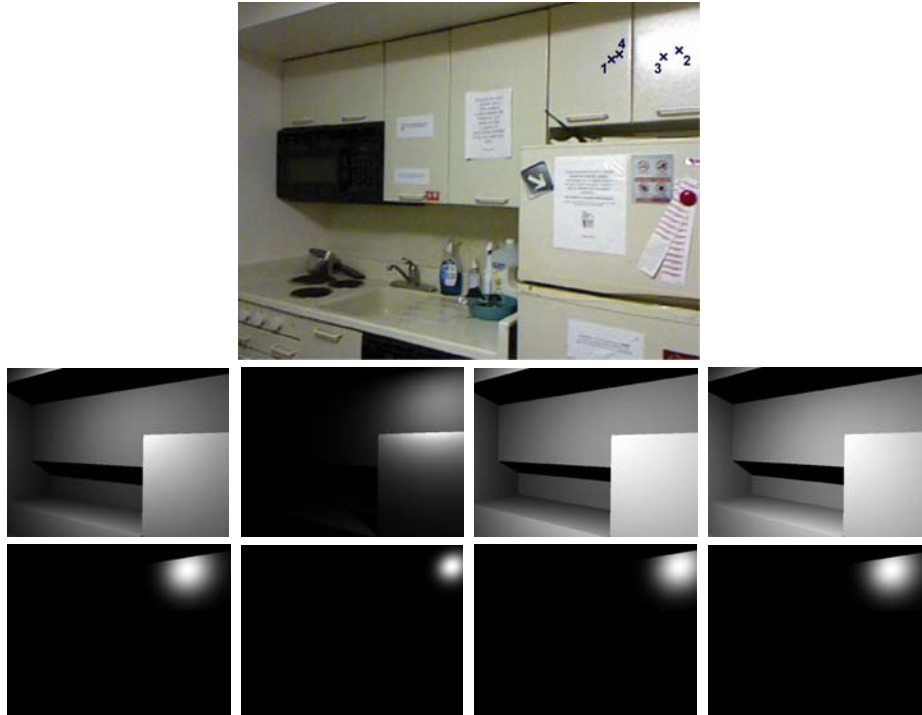


Fig. 5. First row: light position results in case of specular reflection. See text for details. Second row: diffuse rendering. Third row: specular rendering. Each column corresponds to one method from method 1 (left) to method 4 (right).

6 Conclusion

In this paper, we have proposed an approach to cope with the problem of lighting estimation from low-quality color images. First, we have used an iterative process that allows to well estimate the light color of the scene. Second, thanks to a multiple decompositions of the image, we have run an optimization framework that leads to fine estimation of light positions. In our experiments we used a depth sensor providing information exploited for the rendering of the different decompositions. We have shown that our light color estimation outperforms state-of-the-art methods and that accounting specular reflection during the optimization process improves the results over methods that just assume lambertian reflection. As future works, we propose to extend the approach to lights with different colors. In real indoor environments, the colors of the lights do not vary significantly within a scene, but it would help to detect even slight spatial variation and by this way refine the color of each individual light. Indeed, the correct estimation of the light color can also help in the specularity detection. Second, we could add one term in the final objective function that represents the final color rendering of the image. Thus, we could minimize the difference between this

image and the original color one and by this way also optimize the color of the light (instead of only the position). Finally, there is still a large possibility of improvement of the specularly detection if we consider the geometry information during this step. Until now, just pixel chromaticities were considered.

References

1. Debevec, P.: Image-based lighting: *IEEE Computer Graphics and Applications* 22, 26–34 (2002)
2. Karsch, K., Hedau, V., Forsyth, D., Hoiem, D.: Rendering synthetic objects into legacy photographs. In: *SIGGRAPH Asia Conference*, pp. 157:1–157:12. ACM Press, New York (2011)
3. Frahm, J.-M., Koeser, K., Grest, D., Koch, R.: Markerless augmented reality with light source estimation for direct illumination. In: *2nd IEEE European Conference on Visual Media Production (CVMP)*, pp. 211–220. IEEE Press, New York (2005)
4. Al-Najdawi, N., Bez, H. E., Singhai, J., Edirisinghe, E. A.: A survey of cast shadow detection algorithms. *Pattern Recognition Letters* 33, 752–764 (2012)
5. Jacobs, K., Loscos, C.: Classification of illumination methods for mixed reality. *Computer Graphics Forum* 25, 29–51 (2006)
6. Grosse, R., Johnson, M. K., Adelson, E. H., Freeman, W. T.: Ground truth dataset and baseline evaluations for intrinsic image algorithms. In: *12th IEEE International Conference on Computer Vision (ICCV)*, pp. 2335–2342. IEEE Press, New York (2009)
7. Shida B., van de Weijer, J. Object Recoloring based on Intrinsic Image Estimation. In: *13th IEEE International Conference on Computer Vision (ICCV)*, pp. 327–334. IEEE Press, New York (2011)
8. Land, E. H., McCann, J. J.: Lightness and retinex theory. *Journal of the Optical society of America* 61, 1–11 (1971)
9. Horn, B. K. P.: Determining lightness from an image. *Computer Graphics and Image Processing* 3, 277–299 (1974)
10. Maxwell, B. A., Shafer, S. A.: Segmentation and Interpretation of Multicolored Objects with Highlights. *Computer Vision and Image Understanding* 77, 1–24 (2000)
11. Artusi, A., Banterle, F., Chetverikov, D.: A Survey of Specularity Removal Methods. *Computer Graphics Forum* 30, 2208–2230 (2011)
12. Shen, H.-L., Cai, Q.-Y.: Simple and efficient method for specularly removal in an image. *Applied optics* 48, 2711–2719 (2009)
13. Limare, N., Petro, A. B., Sbert, C., Morel, J.-M.: Retinex Poisson Equation: a Model for Color Perception. *Image Processing On Line* (2011), http://www.ipol.im/pub/algo/lmps_retinex_poisson_equation/
14. Shafer, S. A.: Using color to separate reflection components. *Color Research and Application* 10, 210–218 (1984)
15. Phong, B. T.: Illumination for computer generated pictures. *Communications of the ACM* 18, 311–317 (1975)
16. Gijzenij, A., Gevers, T., van de Weijer, J.: Computational Color Constancy: Survey and Experiments. *IEEE transactions on image processing: a publication of the IEEE Signal Processing Society* 20, 2475–2489 (2011)
17. Silberman, N., Fergus, R.: Indoor scene segmentation using a structured light sensor. In: *13th IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pp. 601–608. IEEE Press, New York (2011)