Single image based illumination estimation for lighting virtual object in real scene

Xiaowu Chen, Ke Wang and Xin Jin*
State Key Laboratory of Virtual Reality Technology and Systems
School of Computer Science and Engineering, Beihang University, Beijing, China
**Corresponding author (email: jinxin@vrlab.buaa.edu.cn)*

***Abstract*—Rendering virtual objects into real scenes with real illumination can greatly increase the realism of virtual objects and the consistency between the virtual and the real. The main challenge lies in illumination estimation from a single image. This article proposes a novel method of single image based illumination estimation for lighting virtual object in real scene. Only a single image, without any knowledge of the 3D geometry or reflectance, is needed, which greatly increases the applicability of the method. We first estimate coarse scene geometry and intrinsic components including shading image and reflectance image. Then the sparse radiance map of the scene is inferred based on the scene geometry and intrinsic components. Finally, the virtual objects are illuminated by the estimated sparse radiance map. Some experimental results show that this method can convincingly light virtual objects into a single real image, without any pre-recorded 3D geometry and reflectance, illumination acquisition equipments or imaging information of the image.**

***Keywords*-illumination estimation; geometry estimation; intrinsic image; sparse radiance map; lighting virtual object**

I. Introduction

Rendering virtual objects into real scenes has been widely employed in city planning, art design and film production. The realism of virtual objects and the consistency of virtual objects against real scene are determined to a great extent by the lighting effects. Lighting virtual objects with illumination from real scene is a hot topic in computer graphics community. The main challenge lies in illumination estimation from a single image.

In early works, pre-recorded scene geometry and reflectance are often required to infer the illumination distribution of the scene (e.g.[1]). Then scene illumination is recorded as a radiance map by various light probes located in real scenes (e.g.[2]). The requirement of pre-recorded scene geometry and reflectance or pre-located light probes limits the application of these methods.

Recently, researchers have been trying to estimate scene illumination from only a single image. Most of them use a sun and sky doom model to treat outdoor images [3], [4]. Most recently, Mei *et al.* [5] recover illumination from image with cast shadows via sparse representation. However, cast shadows must appear in their scenario.

We propose a novel method of single image based illumination estimation for lighting virtual object in real scene. We aim to use only a single image without any knowledge of the 3D geometry , reflectance or imaging information. Then we use the estimated illumination for lighting virtual objects in real scene of a single image (see Figure 1).

Lalonde *et al.* [4] directly use the three most evident appearance cues to estimate the illumination in a scene.

Figure 1. The objective. We estimate the illumination of the scene for lighting a virtual *chair* into a real scene. The virtual *chair* matches the existing image in lighting effects and cast convincing shadows on the real scene rendered with the estimated illumination.

In recent years, scene understanding from a single image has attracted plenty of researchers. The estimation of scene geometry (e.g. [6]) and intrinsic components (including shading and reflectance images, e.g. [7]) has achieved great success in some scenarios. Scene appearance is mostly determined by scene geometry, reflectance and illumination. We are thus inspired to estimate scene illumination through current scene understanding technologies.

Unlike [4], we first estimate coarse scene geometry and intrinsic components from various cues. Then we use a simple illumination model based on sparse radiance map to represent scene illumination via the estimated scene geometry and intrinsic components. The sparse radiance map contains several sparse and discrete directional light sources evenly distributed on the half sphere around the scene (see Figure 4). Finally, the virtual objects are illuminated by the estimated sparse radiance map. This can be regarded as a step-by-step way from appearance cues to estimate geometry and intrinsic components and then to estimate illumination of the scene. Some experiment results show that this method can convincingly light virtual objects into a single real image, without any illumination acquisition equipments, imaging information or human interactions.

The main contribution of this article is the usage of a simple sparse radiance map based illumination model to combine the estimated geometry and intrinsic components to estimate the scene illumination.

II. Related work

We will briefly review related work in four aspects: illumination estimation, geometry estimation, intrinsic component estimation and object relighting.

Illumination estimation. Several works estimate illumination with the help of pre-recorded 3D geometry model and reflectance such as [8]. Our work relies on only a single image without exact 3D such as geometry and specific reflectance models. Light probes such as light sphere [2] or fisheye cameras[9] are located in the real scene to record scene illumination directly. However, thousands of photos have been shot and one may have no chance to place light probes in the scene.

For illumination estimation from outdoor images, Madsen *et al.* [3] need to know the date, time and position on Earth of the image shooting. They make the assumption that there is a predominant occurrence of approximately diffuse surfaces and shadows in the scene, which limits their application. Moreover, not every photo has recorded the date or position of the image shooting. Unlike the above methods, our work tries to estimate scene illumination from only a single image without any pre-recorded 3D geometry and reflectance, illumination acquisition equipments or imaging information of the image.

Lalonde *et al.* [4] estimate scene illumination from only a single outdoor image. They use a dataset of 6 million images for training the illumination inference model and estimate a sun and sky doom model that is especially for outdoor images. The three most evident appearance cues(i.e. the sky, shadows on the ground and the varied intensities of the vertical surfaces to estimate the direction of light) are directly employed to estimate the illumination in a scene. However, with the great achievements in scene understanding (such as geometry and intrinsic component estimation), we believe that these scene understanding technologies can help estimate the scene illumination.

Geometry estimation. Recently, rich literatures have addressed the problem of geometry estimation from a single image. Hoiem *et al.* [10] use features of color, texture, edge, location etc. to recover surface layout (i.e. coarse surface orientation) from a single image. Saxena *et al.* [6] use similar features to directly estimate a 3D scene structure from a single image, and good performance is shown in various test images. Liu *et al.* [11] estimate single image depth with the help of predicted semantic labels. And Gupta *et al.* recover a 3D parse tree of a single image through physical reasoning. For our task, either [6] or [11] can be used to directly output a coarse 3D scene structure from a single image .

Intrinsic component estimation. The *intrinsic image* decomposition, which was first introduced by [12], decompose a photo into the pixel by pixel product of an illumination component and a reflectance component. This is an ill-posed problem and open challenge that has attracted lots of researchers such as several recent works [7], [13], [14], [15]. Due to the ill-posedness of the problem, automatic methods are challenged by complex natural images. So Bousseau *et al.*[14] propose a user-assisted approach to specify regions of constant reflectance or illumination for guiding the intrinsic images decomposition.

Lighting virtual objects. In the community of augmented reality, which renders 3D models into real scene, researchers use a simplified version of [2] to achieve real time merging [16], [17], [18]. They often use pre-recorded scene geometry or various light probe in the real scenes. Haber *et al.* [19] relight objects by recovering the reflectance of a static scene with known geometry from a collection of images taken under distant, unknown illumination. However, the geometry of the scene is estimated from lots of images containing nearly the same objects. By contrast, in our work the geometry is estimated from a pre-trained classifier. The virtual object used in our work is a 3D model with textures and is illuminated by the estimated sparse radiance map.

III. Illumination estimation for lighting objects

The workflow of the proposed method is illustrated in Figure 2. The geometry and intrinsic component estimation and lighting with sparse radiance map will be described in this section.

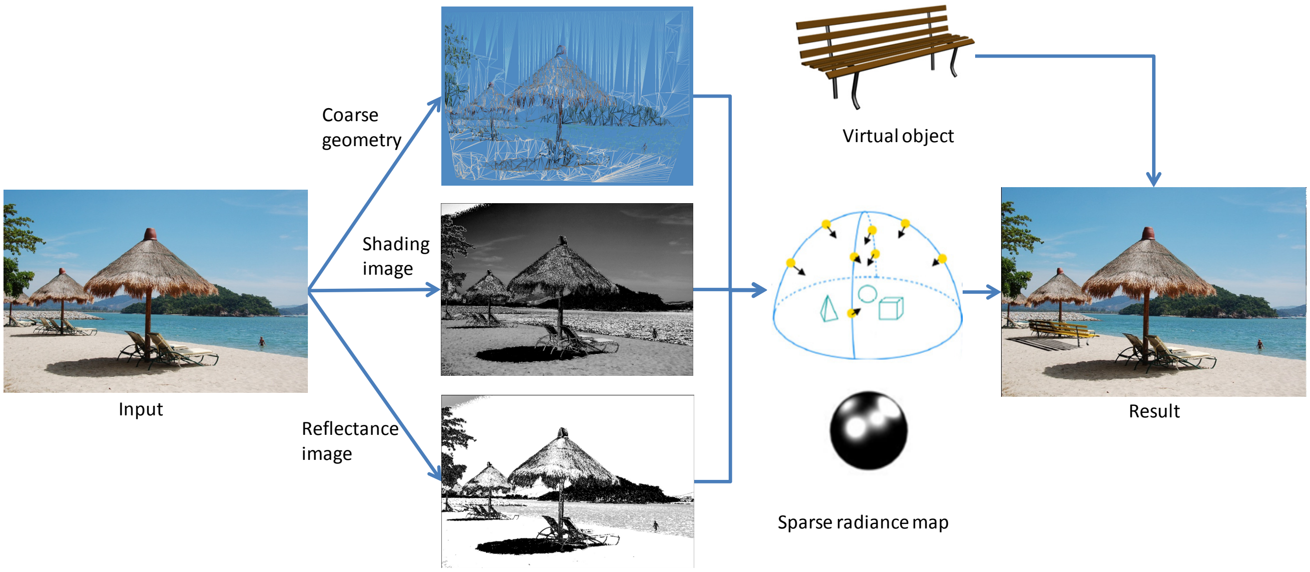

Figure 2. The workflow of our method. We estimate the coarse geometry model of the input image. The input image is decomposed into intrinsic components including a shading image and a reflectance image. Then we combine the coarse geometry model, the shading image and the reflectance image of the scene to estimate sparse radiance map. Finally the virtual object is illuminated with the estimated sparse radiance map. The virtual *chair* matches the input image in lighting effects and cast convincing shadows on the real ground. Although the estimated geometry and intrinsic components are not accurate in every pixel (see the shadow area under the umbrella in the estimated reflectance image), our estimated illumination is basically right via our illumination estimation algorithm.

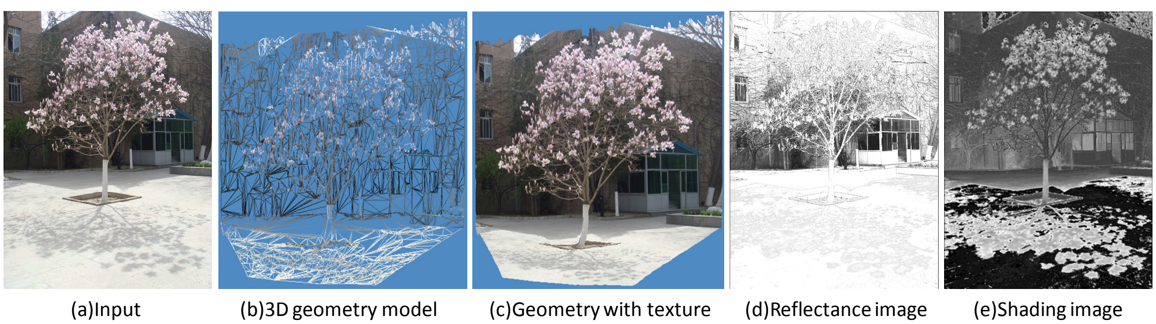

Figure 3. (a) is the input image. In this example, (b) is the 3D geometry model estimated by the method of [6]. (c) is the model with the input image as its texture. (d) and (e) are the shading image and reflectance image estimated by the method of [7].

A. *Geometry and intrinsic component*

Saxena *et al.* [6] and Liu *et al.* [11] use different image features with similar Markov Random Field (MRF) model to infer the pixel wise depth map and the 3D geometry structure of scenes from a single image. Saxena *et al.* [6] use features of color, texture, edge, location etc. Liu *et al.* [11] adopt semantic and geometry constrains for simple linear regression method and use the same training set of [6]. In our implementation, one can choose either [6] or [11] to directly output a 3D scene structure from a single image. A example of estimated geometry model is shown in Figure 3.

Our aim is to automatic estimate illumination from a single image. So we adopt some automatic methods of intrinsic component estimation [7], [13], [15]. Each of them can be leveraged for our task of intrinsic component estimation . Then we will refine the estimation results based on the sparse radiance map. An example of estimated shading image and reflectance image is shown in Figure 3.

B. *Lighting with sparse radiance map*

Sparse radiance map. Light sources in the real scene may be of different shapes, directions and distances. Directly estimating real light sources may result in labour work of modelling different light sources. We choose light ray model to approximate the light sources in real scene with a radiance map, which can be considered as a half light ray sphere around the whole scene. For more efficient computing, we adopt a *sparse radiance map* (see Figure 4) instead of traditional dense radiance map to determine the main light rays from various light sources. One can regard the sparse radiance map as a simplified version of traditional radiance

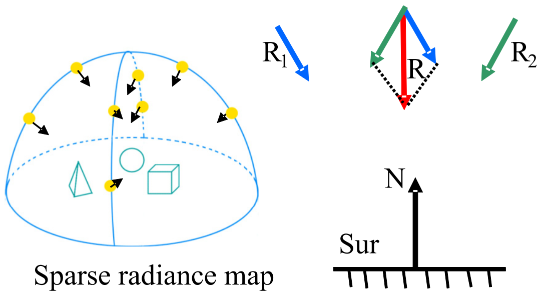

Figure 4. Sparse radiance map and ray combination (Eq. 1, 2, 3, 4.).The sparse radiance map contains m sparse and discrete directional light sources evenly distributed on the half sphere around the scene and directed to the center point of the ground circle.

map. We make the assumption that these sparse light sources can approximate the real scene by making a combination of the estimated light intensity. In the outdoor environment, the main light source is the sun. So a small m of the number of the virtual light sources is enough to estimate the sun direction. In the indoor environment, we can set a larger m for simulating multiple main light sources. We use the estimated ambient light value as the values of the remained points in the radiance map to simulate the sky outdoor and the other weak light sources indoor.

As shown in Figure 4, a light ray R in 3D space can be described as:

$$R = I_L L \tag{1}$$

where I_L is the intensity of the light ray R, L represents the unit normal vector along the ray direction. Suppose that the irradiance on surface sur caused by ray R can be a combination of the irradiance caused by ray $R_1 = I_{L_1} L_1$ and ray $R_2 = I_{L_2} L_2$:

$$I_{L_1} L_1 \cdot N + I_{L_2} L_2 \cdot N = (I_{L_1} + I_{L_2}) \cdot N = I_L L \cdot N \tag{2}$$

where N is the unit normal vector of the surface sur. However, not all the surface obey Eq. 2. The surface should be visible to R, R_1 and R_2 (Eq. 3). The cosine of the angles between the three rays and the surface normal N should be above zero (Eq. 4). Thus, for using sparse light sources to approximate the illumination in real scenes, we would mostly choose the surfaces obeying Eq. 3 and Eq. 4 to estimate the sparse radiance map.

$$\mathrm{Vis}(sur, L) = \mathrm{Vis}(sur, L_1) = \mathrm{Vis}(sur, L_2) = 1 \tag{3}$$

$$L_1 \cdot N > 0, L_1 \cdot N > 0, L \cdot N > 0 \tag{4}$$

Estimating illumination and lighting virtual objects. According to the intrinsic component decomposition, the intensity of an image scene pixel on a surface can be

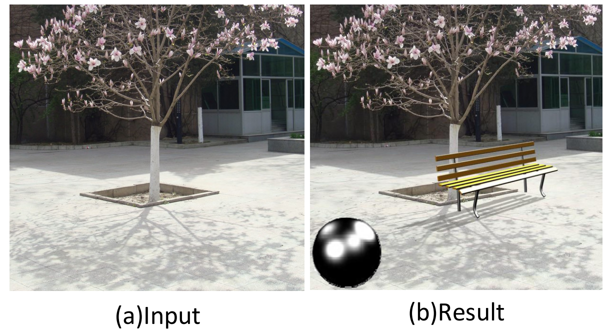

Figure 5. An example of the estimating and the rendering result.(a) is the input image. (b) A virtual object is rendered in (b). A black ball rendered with the estimated sparse radiance map is on the left corner of (b).

approximately decomposed into the irradiance collected by the surface in that point and the reflectance of the surface:

$$I = S * K \tag{5}$$

where I, S and K are the pixel values of the input scene image, the shading image and the reflectance image, respectively. For a Lambertian surface, the irradiance can be represented by [20] :

$$S = I_a + \sum_{i=1}^{m} I_{L_i} L_i \cdot N \tag{6}$$

where I_a is the ambient light of the scene. I_{L_i} and L_i are the intensity and direction of the i^{th} ray reaching surface sur. m is the number of the rays reaching surface sur. N and K are the normal and the reflectance estimated in Section III-A. Then we employ Levenberg-Marquardt algorithm [21] to obtain the solution of minimization between the shading image and the estimated irradiance:

$$\arg \max_{(I_a, I_1, I_2, \ldots, I_m)} \sum_{j=1}^{n_s} (S_j - (I_a + \sum_{i=1}^{m} I_{L_i} L_i \cdot N)) \tag{7}$$

where S_j is the value of the shading image of the j^{th} 3D triangle surface estimated in Section III-A. n_s is the number of the triangle surfaces used for illumination estimation. The parameters needed to be estimated of the sparse radiance map are :

$$SRM_{para} = I_a, I_1, I_2, \ldots, I_m \tag{8}$$

The estimated $I_1, I_2, \ldots, I_m$ form the sparse radiance map are considered as the main light sources. We use the estimated ambient light value as the values of the remained points in the sparse radiance map. With the estimated sparse radiance map, we light objects in the real scene using an off-the-shelf rendering software. An example of the estimated sparse radiance map and the rendering result is shown in Figure 5.

(a)Input (b)Result

Figure 6. An example of indoor result. The shadow of the virtual *bear* is quite similar to those of the real cups.

(a) Input (b) Madsen *et al.* (c) Our Result

Figure 7. Comparison with Madsen *et al.* [3]. Our estimated and rendered shadow are nearly the same as that of Madsen *et al.* with only a single image.

IV. Experiments

In this section, we will show the rendering results of outdoor and indoor images, and comparisons with the work of Madsen *et al.* [3] and Lalonde *et al.* [4].

Outdoor and indoor results. We test our method in both outdoor and indoor scenarios. As shown in Figure 1 and 5, we show examples of outdoor images. Unlike Madsen *et al.* [3] and Lalonde *et al.* [4], who only treat outdoor images with their sun and sky doom models of outdoor illumination, our sparse radiance map can also be used to estimate indoor illumination, as shown in Figure 6. Notice how the shadows on the ground or the desktop, and shading and reflections on the virtual models are consistent with the image.

Comparisons with related work. Madsen *et al.* [3] need to know the date, time and position on Earth of the image shooting. They use such information to infer the sun direction. Lots of consumer photos have not recorded the *EXIF* information containing the date or position of the image shooting. As shown in Figure 7, our work can estimate the scene illumination from only a single image without the imaging information.

To the best of our knowledge, the work with the most similar objective to ours is that of Lalonde *et al.* [4]. They use a dataset of 6 million images for training the illumination inference model and estimate a sun and sky doom model of outdoor illumination, which can be only used for outdoor images. They directly use the three most evident appearance cues to estimate the illumination in a scene. However, the experiments show that, with our sparse radiance map and the current scene understanding technologies of estimating scene geometry and intrinsic components, we can obtain the similar estimation and rendering results as [4], as shown in Figure 8, although our cues and the whole estimation process are quite different.

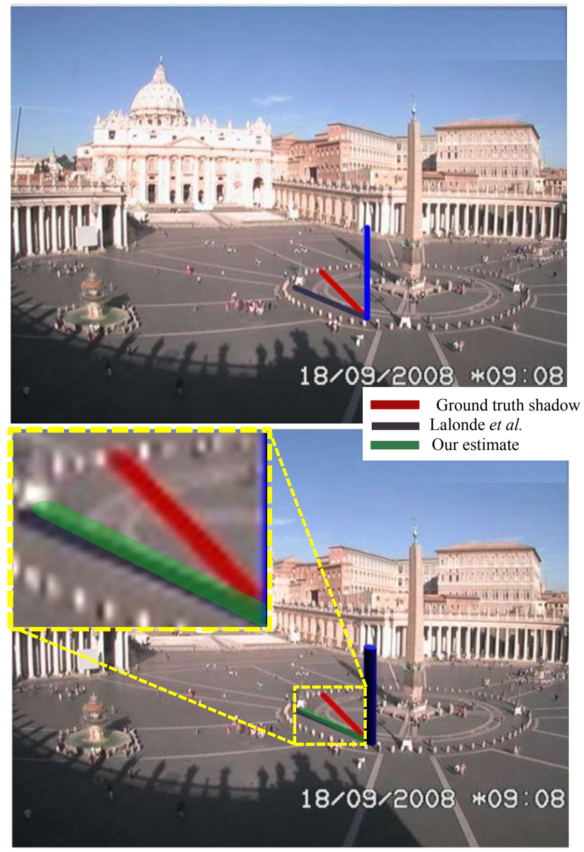

Figure 8. *Vatican sequence*: Comparison with Lalonde *et al.* [4]. We also render a blue cylinder into the real scene and obtain the similar shadow direction to that of [4]. The red shadow is the ground truth. The grey one and the green one are the estimated shadows of [4] and our method, respectively.

V. Conclusion and discussion

In this paper, we proposed a novel method of single image based illumination estimation for lighting virtual object in real scene. The main contribution of our work lies in using a simple sparse radiance map based illumination model to combine the estimated geometry and intrinsic components to estimate the scene illumination. With the help of current scene understanding technologies of estimating scene geometry and intrinsic components, we achieve convincing results with the state of art.

Discussion and future work. The illumination model used in this work is based on the classic Phong model [20] and merely a local illumination model. Although this simple model is enough for most outdoor and indoor consumer photos with less complex illumination by using current scene understanding technologies, for some indoor images with more complex illumination, this simple model is not enough to obtain plausible estimation and rendering results. Maybe the more realism global illumination model can be employed to estimate the complex illumination. However, more parameters such as materials, and more estimating accuracy are required by using future scene understanding technologies. But we believe this is a trend for future illumination estimation work. The estimated geometry model is not quite accuracy. The occlusion between the virtual shadow and the real objects, also the real shadow and the virtual objects, and the folding of the shadows are not simulated very well. And this may rely on more accurate 3D reconstruction in the future work.

Acknowledgment

This work was partially supported by Key Technology R&D Program of China (2008BAH29B02), National Natural Science Foundation of China (90818003&60933006), and Specialized Research Fund for the Doctoral Program of Higher Education (20091102110019).

References

[1] S. R. Marschner and D. P. Greenberg, "Inverse lighting for photography," in *In Fifth Color Imaging Conference*, 1997, pp. 262–265.

[2] 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*, ser. SIGGRAPH '98. New York, NY, USA: ACM, 1998, pp. 189–198.

[3] C. B. Madsen and M. Nielsen, "Towards probe-less augmented reality - a position paper," in *GRAPP*, 2008, pp. 255–261.

[4] J.-F. Lalonde, A. Efros, and S. Narasimhan, "Estimating natural illumination from a single outdoor image," in *IEEE International Conference on Computer Vision*, October 2009.

[5] X. Mei, H. Ling, and D. W. Jacobs, "Illumination recovery from image with cast shadows via sparse representation," *IEEE Trans. on Image Processing (TIP)*, to appear, 2011.

[6] A. Saxena, M. Sun, and A. Y. Ng, "Make3d: Learning 3d scene structure from a single still image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, pp. 824–840, May 2009.

[7] M. F. Tappen, E. H. Adelson, and W. T. Freeman, "Estimating intrinsic component images using non-linear regression," *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, vol. 2, pp. 1992–1999, 2006.

[8] R. Basri, D. Jacobs, and I. Kemelmacher, "Photometric stereo with general, unknown lighting," *Int. J. Comput. Vision*, vol. 72, pp. 239–257, May 2007.

[9] J. michael Frahm, K. Koeser, D. Grest, and R. Koch, "Markerless augmented reality with light source estimation for direct illumination," in *In Conference on Visual Media Production CVMP*, 2005.

[10] D. Hoiem, A. A. Efros, and M. Hebert, "Recovering surface layout from an image," *Int. J. Comput. Vision*, vol. 75, pp. 151–172, October 2007.

[11] B. Liu, S. Gould, and D. Koller, "Single image depth estimation from predicted semantic labels," in *CVPR*, 2010, pp. 1253–1260.

[12] H. Barrow and J. Tenenbaum, "Recovering intrinsic scene characteristics from images," in *Computer Vision systems'78*, 1978, pp. 3–26.

[13] L. Shen, P. Tan, and S. Lin, "Intrinsic image decomposition with non-local texture cues," *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, vol. 0, pp. 1–7, 2008.

[14] A. Bousseau, S. Paris, and F. Durand, "User assisted intrinsic images," 2009.

[15] X. Jiang, A. J. Schofield, and J. L. Wyatt, "Correlation-based intrinsic image extraction from a single image," in *Proceedings of the 11th European conference on Computer vision: Part IV*, ser. ECCV'10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 58–71.

[16] P. Supan, I. Stuppacher, and M. Haller, "Image based shadowing in real-time augmented reality," *IJVR*, vol. 5, no. 3, pp. 1–7, 2006.

[17] J. Pilet, A. Geiger, P. Lagger, V. Lepetit, and P. Fua, "An all-in-one solution to geometric and photometric calibration," in *Proceedings of the 5th IEEE and ACM International Symposium on Mixed and Augmented Reality*, ser. ISMAR '06. Washington, DC, USA: IEEE Computer Society, 2006, pp. 69–78.

[18] T. Jensen, M. S. Andersen, and C. B. Madsen, "Real-time image based lighting for outdoor augmented reality under dynamically changing illumination conditions," in *GRAPP*, 2006, pp. 364–371.

[19] T. Haber, C. Fuchs, P. Bekaert, H.-P. Seidel, M. Goesele, and H. P. A. Lensch, "Relighting objects from image collections," in *CVPR*, 2009, pp. 627–634.

[20] B. T. Phong, "Illumination for computer generated pictures," *Commun. ACM*, vol. 18, pp. 311–317, June 1975.

[21] J. J. Moré, "The Levenberg-Marquardt algorithm: Implementation and theory," in *Numerical Analysis*, G. A. Watson, Ed. Berlin: Springer, 1977, pp. 105–116.

[22] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, pp. 381–395, June 1981.

[23] A. Gupta, A. A. Efros, and M. Hebert, "Blocks world revisited: Image understanding using qualitative geometry and mechanics," in *European Conference on Computer Vision(ECCV)*, 2010.