

ESTIMATION OF POSITION AND INTENSITY OF DYNAMIC LIGHT SOURCES USING CAST SHADOWS ON TEXTURED REAL SURFACES

Salma Jiddi ^{*†} Philippe Robert ^{*} Eric Marchand [†]

^{*}Technicolor [†]Univ Rennes, Inria, CNRS, IRISA

ABSTRACT

In this paper, we consider the problem of estimating the 3D position and intensity of multiple light sources without using any light probe or user interaction. The proposed approach is twofold and relies on RGB-D data acquired with a low cost 3D sensor. First, we separate albedo/texture and illumination using lightness ratios between pairs of points with the same reflectance property but subject to different lighting conditions. Our selection algorithm is robust in presence of challenging textured surfaces. Then, estimated *illumination* ratios are integrated, at each frame, within an iterative process to recover position and intensity of light sources responsible of cast shadows. Estimated lighting characteristics are finally used to achieve realistic Augmented Reality (AR).

Index Terms— Illumination, Reflectance, Textures, Shadows, Augmented reality

1. INTRODUCTION

The image brightness of a 3D scene is a function of three components: scene geometry, surface reflectance and illumination distribution. Provided that these three components are accurately estimated, one is able to photo-realistically render a virtual scene. In this paper, we focus on estimating illumination characteristics. Specifically, we aim at recovering the 3D position and intensity of light sources in indoor scenes. Existing solutions often use additional devices such as light probes [1][2] or fisheye cameras [3][4]. Also, when only RGB-D data is considered, lighting is often assumed to be distant [5] or reduced to a single point light [6][7].

The proposed framework handles both static and dynamic lighting. It takes three inputs: (a) coarse 3D model of the scene acquired by an active sensor (Intel R200). (b) *near-ambient reference image* to which we will refer as *reference image*. The latter is acquired by simulating an ambient lighting which, in theory, does not generate shading or shadowing. In practice, it can be easily produced by considering a fairly uniform indirect lighting. (c) color images of the scene from which lighting will be recovered.

The assumptions we made for the proposed approach are: (i) scene geometry is assumed to be static and contains a main planar surface on which shadows are cast. (ii) scene

reflectance is described by the Lambertian reflection model. In this work, the viewpoint is fixed within the analysis part to avoid camera pose registration errors but can vary in AR.

The main contributions of our method are as follows:

Illumination ratio maps. we propose a novel approach to factor albedo/texture and illumination in arbitrary textured surfaces. Our algorithm relies on a fine selection of pairs of points with the same albedo but subject to different illumination conditions. Lightness ratios of selected pairs provide an *illumination ratio map* which is texture-free. This map contains mainly shadowing information and is recovered for each incoming frame as the lighting can be static and/or dynamic.

Position and intensity of multiple light sources. Illumination distribution is initially represented by a set of equally distributed point lights. Within an iterative process, we extract the 3D position of lights whose shadow maps correlate with the *illumination ratio map*. Finally, a robust estimator allows to recover their respective intensities.

Near real-time processing. On average, the implementation of our method is able to perform at approximately 4fps. Recovered illumination characteristics are further integrated in a real-time rendering pipeline for AR with virtual shadows that are visually coherent in terms of shape and attenuation.

2. RELATED WORK

Early work on illumination estimation using shadowing information was presented by Sato et al. [5]. They used cast shadows to recover the direction of multiple distant lighting. In [5], cast shadows are detected by manually removing occluding objects and capturing a *shadow-free image*. Panagopoulos et al. [8][9] used a graphical model and a voting procedure to detect shadows cast on planar textured surfaces under the assumption of distant lighting. Their framework takes about 3 minutes per image, which is critical for AR scenarios. In [6], Arief et al. analyzed the shadow cast by a cube to estimate the position of a single light source. The surface on which the shadow is cast must be uniform. Our methodology differs from previous shadow-based estimation work in three main points: first, our approach does not require any user intervention or reference objects to recover illumination. On the contrary, it targets middle scale scenes including various 3D objects with arbitrary geometry. Secondly, we do not make

the assumption of distant lighting and instead, estimate both 3D position and intensity of multiple light sources. Finally, our approach handles uniform and textured surfaces.

3. OUR APPROACH TO ILLUMINATION ESTIMATION

In the following, we choose Phong reflection model [10] to describe the way a diffuse point p in the scene reflects light:

$$\mathbf{I}^p = \mathbf{k}_d^p (\mathbf{L}_a + \sum_{i=1}^M (\mathbf{N}^p \cdot \omega_i^p) \mathbf{L}_i \mathbf{O}_i^p) \quad (1)$$

where \mathbf{I}^p is the color intensity of a 3D point p , \mathbf{k}_d^p is its albedo and \mathbf{N}^p is its normal vector. \mathbf{L}_a and \mathbf{L}_i are respectively the intensities of ambient lighting and light source i , ω_i^p is the incoming light direction vector of light source i , and M is the number of light sources present in the scene. \mathbf{O}_i^p is a binary visibility term, equal to 1 if point light i is visible from the 3D point corresponding to pixel p and equal to 0 if occluded.

3.1. Estimation of illumination ratio maps

Shadows are caused by the occlusion of incoming light, and thus contain various pieces of information about the illumination of the scene. For this reason, shadow detection has long been considered within the image processing research community [11][12][13][14][8][9]. If we consider only local texture features, we are not capable of accurately separating texture/albedo and illumination. Up to now, it remains an extremely challenging problem and no generic solution exists. In this section, our goal is to *implicitly* detect cast shadows for a given frame. Our approach is twofold:

The first step is achieved only once since geometry is static: we detect the main planar surface on which shadows are cast. Specifically, we compute surface normals \mathbf{N} and apply a region growing algorithm to cluster similarly oriented surfaces together (a deviation value of 3 degrees is typically allowed between normals). Then, we use a RANSAC estimator to fit each cluster to a planar surface model. Finally, the cluster including the largest number of inliers is considered as the main plane. Points above the detected plane are further grouped as belonging to the occluding objects.

Secondly, we aim at separating texture/albedo and illumination in the current frame. The analysis is limited within the 2D projection of the previously detected plane. Our factorization is achieved through a voting scheme using pairs of points (p, \hat{p}) with the same reflectance \mathbf{k}_d but subject to different lighting conditions. We propose a fine selection based on two features: (i) L^2 norm in the CIELAB color space within the *reference image R* as it provides accurate similarity measures compared to using only the current frame (Fig.1). In fact, the use of the *reference image R* makes our algorithm robust in presence of challenging textures, poor lighting conditions and/or sensor noise. (ii) the lightness channel L of the CIELAB color space within the current image enables

us to compare pairs illumination-wise. Hence, points with lower lightness values are prone to belong to shadowed regions. Using these two features, the voting scheme is applied as follows:

$$v(p) = \begin{cases} +1, & \text{if } \|R^p - R^{\hat{p}}\| \leq \epsilon_R \text{ and } L^{\hat{p}} \geq L^p + \epsilon_L \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where ϵ_R and ϵ_L are respectively thresholds with respect to color vectors distance in R and lightness difference in L . A value of 2.5 is chosen for ϵ_R as it corresponds to an *almost noticeable difference* [15]. ϵ_L is kept at a low value (typically 10) to handle both weak and strong shadows. Finally, pixels holding a significant voting value $v(p)$ are further considered to estimate their respective illumination ratio $\delta(p)$:

$$\delta(p) = \frac{L^p}{\bar{L}} \quad \text{with:} \quad \bar{L} = \frac{\sum_{\hat{p}} L^{\hat{p}}}{v(p)} \quad (3)$$

where \bar{L} is the mean lightness value of matched points \hat{p} with respect to p . Furthermore, pixels for which similar-reflectance pairs are found but received a low voting value hold an illumination ratio value equal to 1. In fact, these points are considered to be fully lit. Last but not least, pixels for which no match is found are discarded as we can not reliably estimate their illumination ratio (green pixels in Fig.1).

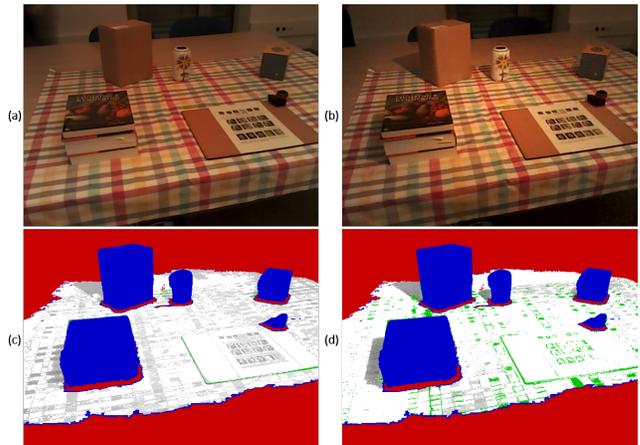


Fig. 1. (a): *reference image R* of the scene. (b): current scene capture. (c)(d): recovered *illumination ratio map* using respectively (b) and (a). (d) demonstrate a better separation of texture and lighting. Note the shadow of the front book in (d) compared to (c) as well as discarded pixels (green).

3.2. Estimation of light sources 3D position

Since surface materials are assumed to be opaque, illumination distribution is approximated by a set of point lights equally distributed in the 3D space above the detected plane. In this section, we aim at recovering the 3D position of light sources responsible of cast shadows in the scene. To achieve this, we extract a subset (S) of point lights whose shadow

maps correlate with the estimated *illumination ratio map* (Section 3.1) for each incoming frame. The identification of subset (S) is carried within an iterative process as follows:

1. We initially compute correlation values by matching the *illumination ratio map* with the shadow maps of the sampled light candidates. The light source whose shadow map has the best correlation value is selected.
2. For each iteration, previously selected light sources are discarded. Also, the matching operation is carried out by discarding previously matched pixels.
3. The process ends either when the currently selected shadow map has a significantly low matching value or if the number of selected lights is higher than N .

The chosen correlation corresponds to Pearson’s correlation coefficient ranging between 0 and 1. Perfectly matching maps have a coefficient equal to 1.

3.3. Estimation of light sources intensity

As we consider small to middle scale scenes, we assume that $((\mathbf{N}^p \cdot \omega^p) = \cos \theta)$ is equal across selected pairs, and use L_i to represent $L_i \cos \theta$. Subsequently, equation (1) becomes:

$$I^p = \mathbf{k}_d^p(L_a + \sum_{i=1}^M L_i O_i^p) \quad (4)$$

Our goal is to estimate the intensity of recovered light sources (Section 3.2). By considering the ratio of color intensities of a selected pair (p, \hat{p}) as described in Section 3.1, we have:

$$\frac{I^p}{I^{\hat{p}}} = \frac{\mathbf{k}_d(L_a + \sum_{i=1}^M (L_i O_i^p))}{\mathbf{k}_d(L_a + \sum_{i=1}^M L_i)} = \delta(p) \quad (5)$$

In fact, the albedo terms in the numerator and denominator cancel out. Subsequently, we obtain the *illumination ratio* value. As mentioned in Section 3.1, fully lit points hold a $\delta(p)$ equal to 1. Thus, we set $L_a + \sum_{i=1}^M L_i = 1$. Finally, we integrate illumination ratio values within a linear system:

$$\mathbf{A}\mathbf{L} = \delta \quad (6)$$

where:

$$\mathbf{A} = \begin{pmatrix} 1 & \mathbf{O}_1^{p_1} & \cdots & \mathbf{O}_M^{p_1} \\ 1 & \mathbf{O}_1^{p_2} & \cdots & \mathbf{O}_M^{p_2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \mathbf{O}_1^{p_N} & \cdots & \mathbf{O}_M^{p_N} \end{pmatrix}; \mathbf{L} = \begin{pmatrix} L_a \\ L_1 \\ \vdots \\ L_M \end{pmatrix}; \delta = \begin{pmatrix} \delta(p_1) \\ \delta(p_2) \\ \vdots \\ \delta(p_N) \end{pmatrix} \quad (7)$$

The linear system (6) is solved using an iterative Least Squares with bounds and equality constraints:

$$\hat{L} = \min_L \left(\frac{1}{2} \|\omega \mathbf{A}\mathbf{L} - \omega \delta\|^2 \right) \quad \text{such as:} \quad (8)$$

$$\begin{cases} 0 \leq L_i \leq 1 & \text{and} & 0 \leq L_a \leq 1 \\ L_a + \sum_{i=1}^M L_i = 1 \end{cases}$$

The weights ω are computed using Tukey’s bisquare loss function. Small weights are discarded throughout iterations.

4. EXPERIMENTAL RESULTS

The proposed approach has been tested on various real scenes with varying lighting conditions. In Fig.2, we illustrate our results within a selection of four scenes (S1, S2, S3 and S4). Further results are shown in the video available at https://youtu.be/kj_dYNepRDg. In Fig.2, scenes are grouped row-wise. In the first column, we overlay the contours of shadow maps corresponding to recovered lights on the input color frame. For instance, red and green contours are used respectively for the first and second detected lights. In the second column, we demonstrate estimated *illumination ratio maps* where background/noise are represented by red pixels and occluding objects by blue pixels. On the detected planar surface, white pixels represent fully lit 3D points (their illumination ratio is equal to 1), green pixels represent 3D points for which no pairs are found. Finally, grayscale pixels are the illumination ratio value. These pixels correspond to 3D points partially or fully occluded with regard to lighting.

4.1. Illumination ratio maps

As illustrated in Fig.2, our algorithm delivers accurate *illumination ratio maps* where texture/albedo is accurately separated from illumination. The proposed framework handles both uniform surfaces (S1) and challenging textured surfaces (S2, S3 and S4). Also, we demonstrate robustness in presence of poor geometry, especially when the scene contains specular objects (S1 and S3). In fact, the Lambertian assumption must mainly hold on the principal plane. Finally, the use of the *reference image* enables us to be robust even when lighting and/or image quality lack (S4).

4.2. Recovered 3D position of light sources

In Fig.2, we overlay the contours of the selected shadow maps on the current color frame. For S2, the *illumination ratio map* contains shadow information but also illumination ratio values corresponding to the narrow spot light effect. Nonetheless, our algorithm succeeds in recovering an accurate light position. Furthermore, our approach demonstrates good results in the presence of overlapping shadows (S3 and S4).

In order to evaluate the precision of recovered lighting positions, we used a special setup for several experimental scenes. First, we choose a world frame on the main plane and measure distance to real lights using a telemeter. Our algorithm is tested on various scenes and recovers light sources position with an average error of 17cm for a mean distance of 2.55m to the light source and a standard deviation of 3.5cm.

As Augmented Reality is our target application, temporal stability with regard to recovered light sources is of paramount importance. When the lighting is static, recovered lighting properties (position and intensity) must be temporally stable, otherwise virtual shadows might suffer from apparent flickering. In Fig.3, we can notice that under the same lighting conditions, the selected light source (ID98)

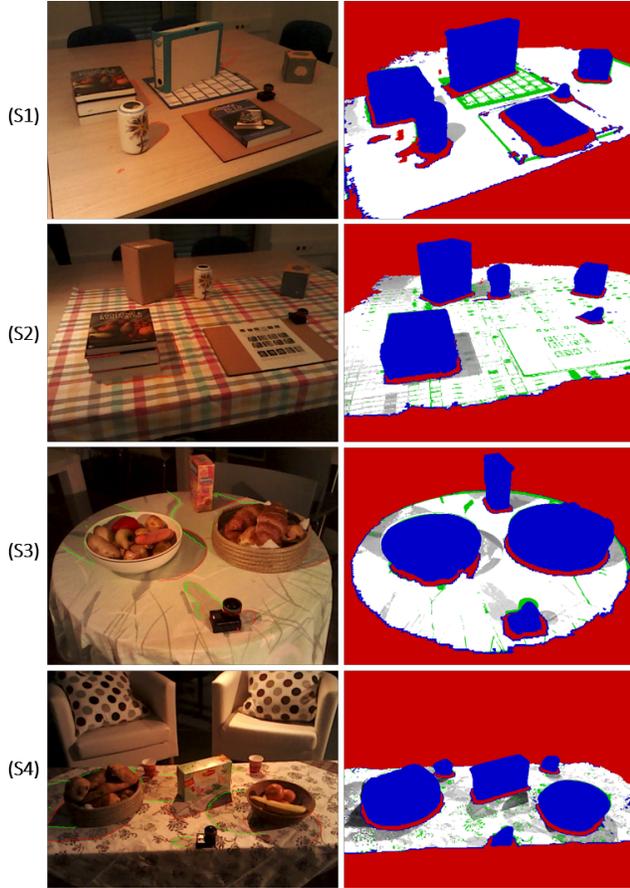


Fig. 2. Col.1: Overlay of selected shadow maps contours on the input color frame. **Col.2:** Estimated *illumination ratio maps* for uniform (S1) and textured surfaces (S2, S3 and S4).

remains the same throughout the sequence with slightly different correlation values. The second best shadow map holds a low correlation value and is thus discarded.

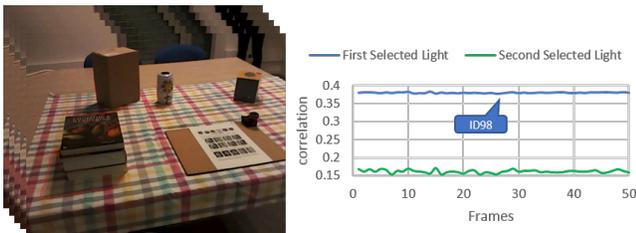


Fig. 3. First and second best correlation coefficients for a scene under static lighting.

4.3. Estimated light sources intensity

In this section, we demonstrate the effectiveness of our approach when estimating the intensity of light sources. First, we show the temporal stability of the intensity estimates throughout a sampled sequence for scenes S2 and S4 under

varying lighting (Fig.4). Furthermore, recovered lighting in-

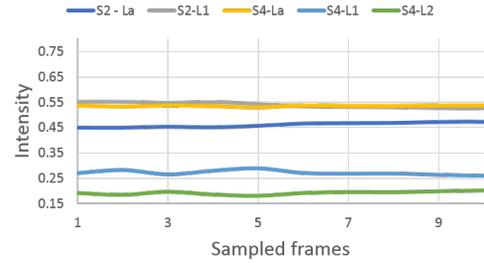


Fig. 4. Recovered ambient intensity L_a and point light sources intensity L_1 and L_2 for scenes S2 and S4.

tensities are used to render virtual shadows. In Fig.5, the first row represents reconstructed shading using scene geometry, lighting 3D position and intensity for scenes S3 and S4. A virtual sphere is introduced as well to demonstrate its interaction with the real scene. Furthermore, the second row in Fig.5 shows augmented scenes where virtual shadows are visually coherent with real shadows in terms of shape and intensity.

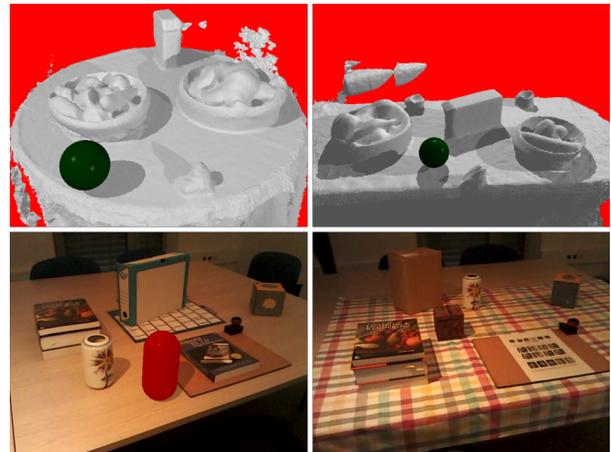


Fig. 5. Row 1: Reconstructed shading using scene geometry and recovered lighting properties (position and intensity) for S3 (left) and S4 (right). **Row 2:** AR scenarios with visually coherent virtual shadows (red capsule in S1 (left) and brown cube in S2 (right)).

5. CONCLUSION AND FUTURE WORK

We presented a new approach to estimate both position and intensity of multiple light sources. The algorithm succeeds in accurately factoring texture/albedo and illumination on challenging textured surfaces. For future work, we are interested in relaxing the Lambertian assumption. Furthermore, achieving a full GPU implementation would allow us to tackle on-line analysis and rendering.

6. REFERENCES

- [1] P. Debevec, "Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography," in *ICCV*, 1999.
- [2] D. Nowrouzezahrai, S. Geiger, K. Mitchell, R. Sumner, W. Jarosz, and M. Gross, "Light factorization for mixed-frequency shadows in augmented reality," in *ISMAR*, 2011.
- [3] K. Rohmer, W. Buschel, D. Raimund, and T. Grosch, "Interactive near-field illumination for photorealistic augmented reality on mobile devices," in *VGTC*, 2014.
- [4] M. Knecht, G. Tanzmeister, C. Traxler, and M. Wimmer, "Interactive brdf estimation for mixed-reality applications," in *WSCG*, 2012.
- [5] I. Sato, Y. Sato, and K. Ikeuchi, "Illumination distribution from brightness in shadows: Adaptive estimation of illumination distribution with unknown reflectance properties in shadow regions," in *ICCV*, 1999.
- [6] I. Arief, S. McCallum, and J.Y. Hardeberg, "Realtime estimation of illumination direction for augmented reality on mobile devices," in *CIC*, 2012.
- [7] B.J. Boom, S. Orts-Escolano, X. Ning, S. McDonagh, P. Sandilands, and R.B. Fisher, "Point light source estimation based on scenes recorded by a rgb-d camera," in *BMVC*, 2013.
- [8] A. Panagopoulos, C. Wang, D. Samaras, and N. Paragios, "Illumination estimation and cast shadow detection through a higherorder graphical model," in *CVPR*, 2011.
- [9] A. Panagopoulos, D. Samaras, and N. Paragios, "Robust shadow and illumination estimation using a mixture model," in *CVPR*, 2009.
- [10] B.T. Phong, "Illumination for computer generated pictures," in *CACM*, 1975.
- [11] G.D. Finlayson, M.S. Drew, and C. Lu, "Entropy minimization for shadow removal," in *IJCV*, 2009.
- [12] R. Guo, Q. Dai, and D. Hoiem, "Paired regions for shadow detection and removal," in *PAMI*, 2012.
- [13] Y. Shor and D. Lischinski, "The shadow meets the mask: Pyramid-based shadow removal," in *EUROGRAPHICS*, 2008.
- [14] J. Zhu, K.G.G. Samuel, S.Z. Masood, and M.F. Tappen, "Learning to recognize shadows in monochromatic natural images," in *CVPR*, 2010.
- [15] Gaurav Sharma, "Digital color imaging handbook (1.7.2 ed.)," in *CRC Press*, 2003.