Point Light Source Estimation based on Scenes Recorded by a RGB-D camera

Bastiaan J. Boom[1]
http://homepages.inf.ed.ac.uk/bboom/

Sergio Orts-Escolano[2]
http://www.dtic.ua.es/~sorts/

Xin X. Ning[1]
avener1446@hotmail.com

Steven McDonagh[1]
http://homepages.inf.ed.ac.uk/s0458953/

Peter Sandilands[1]
http://homepages.inf.ed.ac.uk/s0569500/

Robert B. Fisher[1]
http://homepages.inf.ed.ac.uk/rbf/

[1] Institute of Perception, Action and Behaviour
University of Edinburgh
Edinburgh, UK

[2] Computer Technology Department
University of Alicante
Alicante, Spain

Abstract

Estimation of the point light source position in the scene enhances the experience for augmented reality. The image and depth information from the RGB-D camera allows estimation of the point light source position in a scene, where our approach does not need any probe objects or other measuring devices. The approach uses the Lambertian reflectance model, where the RGB-D camera provides the image and the surface model and the remaining unknowns are the albedo and light parameters (light intensity and direction). In order to determine the light parameters, we assume that segments with a similar colour have the same albedo, which allows us to find the point light source that explains the illumination in the scene. The performance of this method is evaluated on multiple scenes, where a single light bulb is used to illuminate the scene. In this case, the average error in the angle between the true light position vector and our estimate is around 10 degrees. This allows realistic rendering of synthetic objects into the recorded scene, which is used to improve the experience of augmented reality.

1 Introduction

To render objects realistically in a scene requires the knowledge of the illumination in the scene. Our approach estimates the position of a point light source that explains the illumination observed in the scene. The advantage of a point light source is that it can be used easily by even simple rendering software. This allows us to render synthetic objects in real scenes (see Figure 1), improving the augmented reality experience. In this paper, the only sensor that is used for the estimation of the point light source is an RGB-D camera (Kinect sensor). No probe objects, extra fish-eye cameras or other devices are used to estimate the
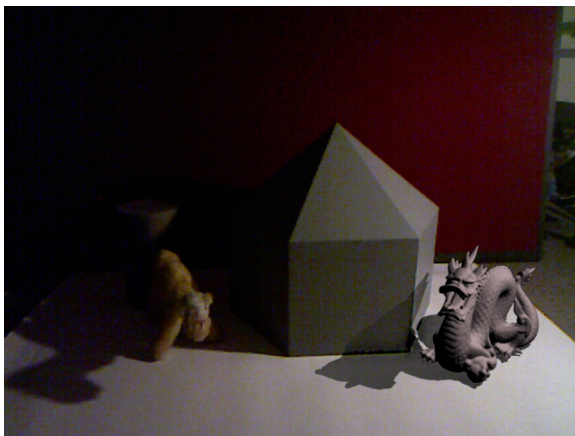
Figure 1: Rendering a synthetic object (dragon) into a real scene recorded with the Kinect sensor, where both the shading and shadows are based on an estimated light source position determined using only the information provided by the Kinect sensor.

illumination conditions.

In order to estimate the point light source position in the scene our method uses the following assumptions: 1) The Lambertian reflectance model describes roughly the appearance of objects in the scenes, where the normals of these objects can be obtained from the depth camera. The normals constrain the position of the point light source. 2) Segments in the RGB image with a similar colour that are contiguous tend to have the same albedo. Using these assumptions, we propose the following interactive search method where, using the surface normals, an image from a hypothesised light source position is rendered where each segment has a uniform albedo. The best hypothesised light source position is given by the rerendered scene that is closest to the original image. For every hypothesised light source position:

1. Given segmented image regions, we combine the individual pixel-based albedo estimates to obtain a single albedo value for each segment.
2. Synthesise an image of the scene given a single albedo value per segment, surface normals and hypothesised light source position.
3. Compute the error between the synthesised and captured images.

This allows us to search for the light source position by minimising this error.

The main contributions of this paper are:

Point light source estimation method: A method that is able to estimate a point light source that explains the illumination in the scene recorded by the Kinect. This method only uses the image and depth map provided by the Kinect sensor for the estimation, where all other methods use also other information, for instance a probe sphere to determine the illumination condition in a scene. **Simulation experiments to verify the method:** A simulation based on a rendered scene is performed to verify the methodology. Our methodology used the Lambertian reflectance model to estimate the illumination of a scene. In our simulation, we are able to verify what to expect in case of specular reflections and ambient lighting. **Scene with measured point light position:** In order to verify the performance of the methodology under real conditions, different scenes illuminated with a single light source are recorded by the Kinect sensor, where the position of the light source and the camera are known allowing us to estimate the accuracy of our methodology.

2   Related Work

Early work on estimating the illuminant based on image information and scene geometry is performed by [16, 17], where they take advantage of the cast shadows and specular reflections of known geometrical objects in the scene. Extensions to this work are performed in [10] and [2], where less information about the scene is used to obtain the illuminant. Stereo vision with a known geometrical object [18, 19], often a probe sphere (white shiny ball) is used to determine multiple light source areas. In [4], the illuminant of the scene is determined using a mirror surface probe sphere, where different exposed photographs of the scene are used. In augmented reality, these mirror surface probes are used by [1, 6, 8] to render objects under the same light condition as the scene. In [5], a double camera system is used for Augmented Reality, where the first camera films the actual scene while a fish-eye camera is used for filming the position of the light on the ceiling. Work that does not require a known geometrical object or second recording device usually assume an outdoor scene. Augmented reality in outdoor scenes is performed by [13, 14] using a 3D sensor for realistic shadow rendering and GPS and compass information to determine the sun position. Other work for estimating the illumination conditions in outdoor scenes is performed by [9, 11, 12], where known properties of sunlight make this an easier problem to solve.

Our methodology differs from previous illumination estimation work, where the main difference is that this method uses the depth information (an estimation of the normals) of the objects in the scene to estimate the position of the light source, and no probe sphere with known geometry and albedo or mirrored reflection is needed. In our case, the focus was indoor light sources because the Kinect sensor is usually used indoors. The method is described in Section 3, the simulation experiments to verify our method are performed in Section 4, while results on real scenes are given in Section 5. In Section 6, the potential of the method for rendering synthetic objects in a Kinect scene is explained and Section 7 gives some discussion.

3   Methodology

The method to determine the point light source position in a scene uses two assumptions: The first assumption is that we have a single point light source and the interaction between this point light source and the objects can be described by the Lambertian Reflectance model. The second assumption is that segments in the image with similar colours have the same albedo and variations of the intensity values within these segments are caused by the reflectance function. To clarify this point with an example, given a red billiard ball in an image, the RGB values of this ball in the image are not the same, however the albedo of this ball is the same where the variation is caused by the reflectance model. The Lambertian reflectance model can be defined as follows:

$$I_o(\mathbf{p}) = \rho(\mathbf{p}) \min(\mathbf{n}(\mathbf{p})\mathbf{s}(\mathbf{p})^T i, 0) \qquad (1)$$

The original image I_o at pixel $\mathbf{p} = \{x, y\}$ according to the Lambertian reflectance model consists of the albedo ρ, the surface normal $\mathbf{n}$ of the objects in the scene at $\mathbf{p}$ and the direction $\mathbf{s}$ (normalised vector) and intensity i of the light. Given the output of the Kinect sensor, both the image intensity I and the surface normals $\mathbf{n}$ are obtained, where the surface normals are computed from the depth map using [7]. The second assumption is that the albedo in segments of the image are the same: In this case, the set of all pixels $P = R_1 \bigcup ... \bigcup R_N\}$

consists of segments R_j that have the same albedo ρ at the pixels $\mathbf{p} \in R_j$. Using these two assumptions an iterative search can be performed on the light source parameters. For an arbitrary light source (with intensity i_r and a position which gives us the direction $\mathbf{s}_r(\mathbf{p})$) used to reconstruct image I_r:

1. Compute the albedo for each segment given the arbitrary light source.
2. Obtain a reconstructed image I_r using the uniform albedo for each segment.
3. Find the light source that minimises the error between the original image I_o and the reconstructed image I_r.

Section 3.1 explains how the reconstructed images are calculated, Section 3.2 gives the segmentation method and Section 3.3 describes how we search for the light parameters.

3.1 Reconstructed Image

To compute the reconstructed image given the arbitrary light parameters $\mathbf{s}_r(\mathbf{p}), i_r$ (given by the search method), first all the albedo values ρ in the segment R_j are computed

$$\rho(\mathbf{p}) = \frac{I_o(\mathbf{p})}{\mathbf{n}(\mathbf{p})\mathbf{s}_r(\mathbf{p})^T i_r}, \mathbf{p} \in R_j \tag{2}$$

To obtain one albedo value for the entire segment ρ_{R_j}, the mean or median of the albedo values $\rho(\mathbf{p}), p \in R_j$ can be used. Because some segments contain normals $\mathbf{n}$ that are close to being perpendicular to the light direction $\mathbf{s}$, the median is a more stable function for obtaining the albedo of the segment ρ_{R_j}. The reconstructed image I_r for a given segment R_j can be computed using the following equation:

$$I_r(\mathbf{p}) = \rho_{R_j} \min(\mathbf{n}(\mathbf{p})\mathbf{s}(\mathbf{p})^T i, 0), \mathbf{p} \in R_j \tag{3}$$

$$E = \sum_{\mathbf{p} \in P} ||I_o(\mathbf{p}) - I_r(\mathbf{p})|| \tag{4}$$

In our case, we would like to minimise the error E between the original image I_o (given in Equation 1) and the reconstructed image I_r obtained in Equation 3, where in the case of an attached shadow $(\mathbf{n}(\mathbf{p})\mathbf{s}_r(\mathbf{p})^T i_r \leq 0)$, the image intensity values are set to zero. This error E indicates to which extent the reconstructed image I_r explains the observed scene I_o with respect to the gradient normally observed in object appearance due to the illumination and the surface of the object.

3.2 Segmentation

To find regions with the same albedo, a colour-based image segmentation method is used. Most colour-based segmentation methods will work for this problem. This paper uses the method described in [4]. Segments that contain less than $\lambda = 100$ pixels are removed for our error calculation, because small regions are often noisy and do not contain much of the gradient information on which the error is minimised. Although depth information can also be used in the segmentation, because it is often likely that different objects have a different albedo, it only contributes slightly and can not be used without also the colour information. Because objects often contain multiple colours and materials, only colour segmentation gives accurate multiple colour segments for objects in these cases.
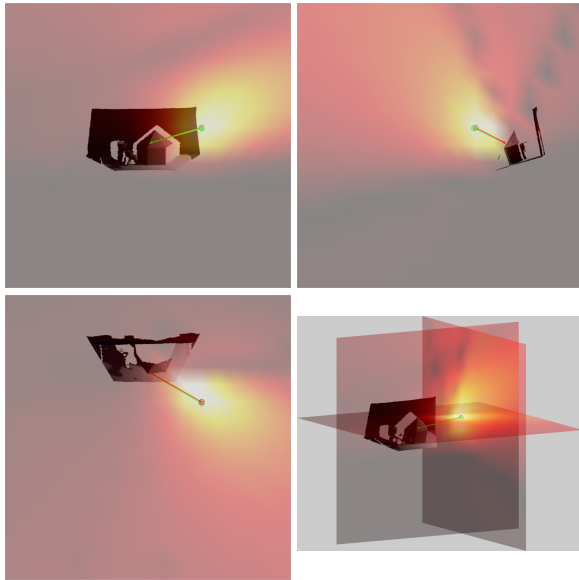
Figure 2: The interpolation of the error function is shown (brighter colour is lower error) computed from a grid of -2 to 2 meters with steps of 25 cm. The error function is very smooth, where the minimum of the error function (red dot) is located close to the true position of the light source (green dot). The red vector starts in the minimum of the error function and goes to the centroid of the scene, while the green vector starts from the groundtruth light positions.

3.3 Search Method

The estimated light parameters can be obtained by minimising the error E given by Equation 4. In the case of a point light source, we have the $(x, y, z$-position) of the light source and the intensity of light source. The light direction $\mathbf{s}(\mathbf{p})$ is a normalised vector between the position of the light source and the position of the pixel given by the depth map. The light intensity i is dependent on the albedo ρ in the scene, because the albedo ρ_{R_j} is computed by setting an arbitrary light intensity i_r (Equation 2). The search method does not have to search for the light intensity value and this is set to $i_r = 1$. For rerendering objects in the scene, the albedo values are divided by a single number to obtain normalised values between zero and one, which allows to compute a relative light intensity given the albedo distribution. In this paper, we used the downhill simplex method [15] or a grid search method to find the estimated position of the light source from the Kinect sensor output. For the downhill simplex method, the initial position of the light source is the viewpoint of the Kinect sensor and the initial stepsize used by this method is 10 cm. For the grid search, a stepsize of 50 cm is used in a grid of -4 by 4 meters around the Kinect sensor position. A finer interpolation of the error function is shown in Figure 2. The grid search is mainly used to visualise our error function but can be parallelised in the future. This function is smooth, which allows for good results with, for instance, a downhill simplex method. An important observation is that the light direction often does not change so computing the position of the light source can be performed only once.
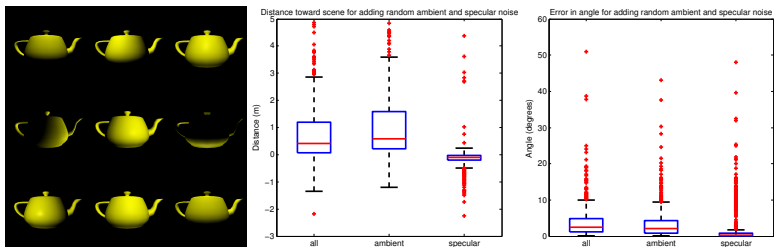
Figure 3: The teapot is rendered 1000 times (examples are shown left) with a light source at different positions, where we added both ambient and specular noise and only ambient and only specular noise. The boxplot in the middle shows the Euclidean distance in world coordinates between the estimated and true position of the light source, where the negative distance indicates that the estimated light source is closer to the scene than the true light source. The boxplot on the right gives the error in angle between the true and estimated light direction. Ambient light has a much larger effect on both the distance and angle than the specular reflections. Specular reflection cause the estimated light to be estimate closer to the scene while ambient light has the reverse effect. If both effects are taken into account, most of the error is caused by ambient light.

4 Verification of the methodology

To verify the correctness of the method, rendering software is used to verify our methodology in simulation. This rendering software also allows us to simulate more complex reflectance models like the Phong reflectance model, which in addition to the Lambertian reflectance model contains ambient and specular reflections. Notice that these terms are object specific, because the material of an object determines the ambient and specular terms of the object surface. Specular reflections are often used as a clue to determine the illumination condition in a scene given a known probe object in the scene. However, this is more difficult for unknown objects because if the specularity of the object is unknown this clue might not exist and is difficult to quantify. Figure 3 shows the results for our verification in simulation, where we varied both the ambient and specular components together and each component individually in order to determine their influence on the light position estimation. We added for both the ambient and specular component a random percentage to the diffuse component, sample from the normal distribution $|\mathcal{N}(0, 10\%)|$ and $|\mathcal{N}(20, 10)|$ for the specular power parameter. This verification shows that ambient light creates a larger error in the estimation of position than the specular reflections. The graph also shows that the error in angle is often not larger than 10 degrees, which gives realistic results for the application of augmented reality.

5 Experimental & Results

5.1 Experimental Setup

In this paper, 23 recordings with the Kinect sensor and a known light souce are used to measure the accuracy of this method. Six scenes are recorded with several different known illumination conditions. In the scenes, a single light source is used, where this was either a
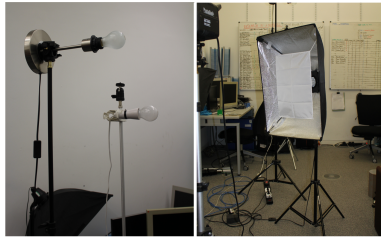
Figure 4: The light bulbs of 60 and 100 watts (left photo) and the spotlight (right photo) used to illuminate the scene.

light bulb of 60 or 100 watts or a spotlight (see Figure 4). To measure the position of the light source in the scene, the distance between the true light source position on the ground plane and two salient control points on the ground plane are measured. Afterwards, we measure the height of the light source above the ground plane. In the Kinect recordings, the ground plane is determined which allows us to calculate the position of the light source given the salient control points, which are also visible in the Kinect scene. Scenes that are recorded with the Kinect are shown in Figure 5. Given the known position of the light source, we are interested in the accuracy of our illumination estimation. The first measure to evaluate the accuracy is the **angle** between the estimated position and the true (measured) position. To calculate this angle, the mean for every pixel of the absolute difference between "the angle between camera position and estimated light position" and "the angle between camera position and true light position" is used. We also calculate the Euclidean **distance** between the estimated and true light position. However, the angle is more important because that has the largest influence on the appearance of objects that are rendered into the scene.

5.2   Results

Figure 5 shows the estimated light source location in several of the recorded scenes. It also shows that the recorded scenes are far from easy, having different kinds of variations that can be expected in real world scenes. The first two scenes in Figure 5 show the potential of this method. We observed that obtaining the direction (angle) of the light is easier than obtaining the exact position of the light source in this direction. In our experiment, the effects of three different parameters in our methodology are investigated.

Computation of the Albedo: To compute the albedo of the segment, in Section 3.1 the mean or median are used. In our experiments, we investigate the influence of the mean and median on both the angle and distance. Figure 6a shows that the median gives a more accurate angle estimation over all pixels, which is more important especially for rerendering of objects in the scene. Figure 6b shows however that in case of the median the absolute distance has a much larger standard deviation. After inspection we discovered that this is often caused by scene illuminate with spotlights, which are not true point light sources.

Segmentation Methods: To verify the dependency on segmentation method [4], we run this segmentation method with two different threshold settings ($K = 100$ and $K = 200$ (default)). In Figure 6c, the result of the different threshold settings do not give significant changes in the results of the methods. This means that our methods seems to be robust under different segmentation parameters. The segmentation is however still a determining factor in our methodology, because segmentation failures, that is two regions with different albedo values
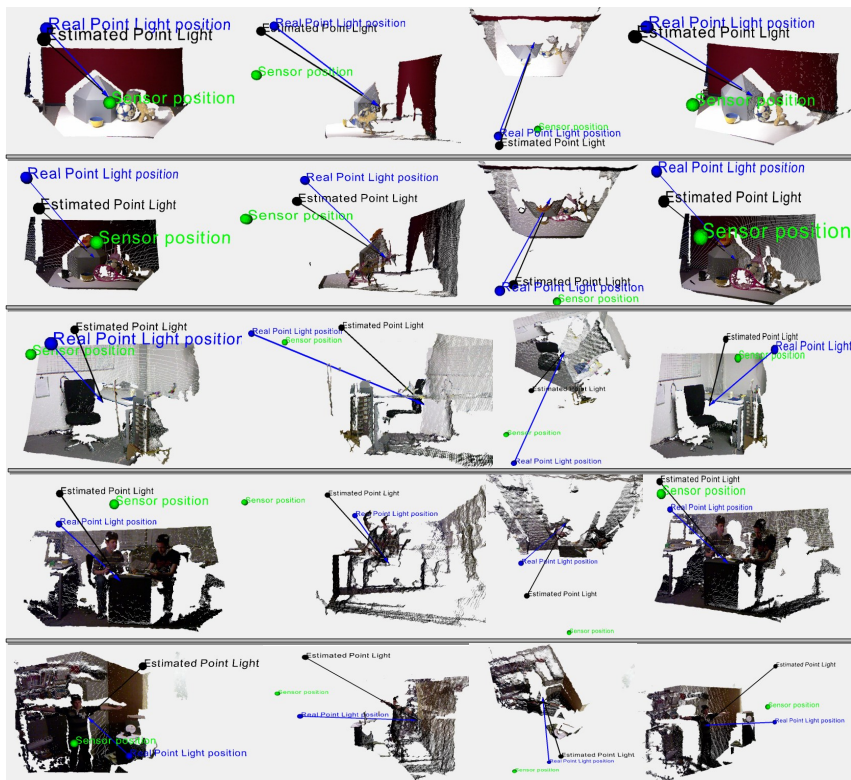
Figure 5: Different scenes in our dataset. The camera position, estimated point light source position and true point light source position are shown by respectively a green ball, black ball and blue ball. Each row shows four views (front, side, top, oblique). We used different angles to show the scene in order to visualise the 3D position and the angle the light sources correctly. In the last scene, we show that the estimated position is not always correct, because of mistakes in the segmentation

segmented as one region, often are the biggest cause of observed failure by our methodology. Improving segmentation by taking depth information into account might improve these failures.

Search Methods: Section 3.3 describes two methods that have been used to search the point light source position. The first method is the downhill simplex method which has as initial starting position the camera position. We also performed a grid search, where a grid of -4 to 4 metres around the camera position is used with steps of 50 cm. The downhill simplex method gives slightly more accurate results (Figure 6d), where we observed that mistakes in the segmentation allow multiple explanations of the position of the light source, to which the grid search seems to be more sensitive.

The downhill simplex method is much faster than the grid search, where our current Matlab implementation takes around 25 second to estimate the light source position.
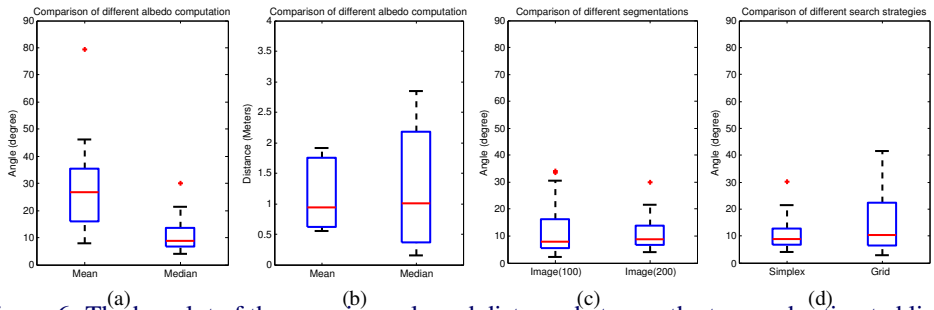
Figure 6: The boxplot of the error in angle and distance between the true and estimated light source position, for computing the albedo (a,b), segmentation method parameters (c) and the different search methods (d)
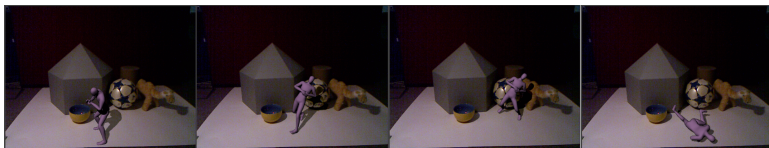

Figure 7: Multiple frames of a virtual person rendered into the scene showing the interaction between the shadows of the person and the background objects.

6 Rendering a synthetic object into the scene

Figure 7 shows that new objects can be rendered into the RGB image of the Kinect. These objects appear to be natural in terms of the illumination, due to the correct shadowing and shading of the objects. To achieve the rendering, the depth map of the scene is loaded into Autodesk Maya, together with the light position. We use a virtual camera that matches the Kinect's properties to record the rendered objects. First, a shadow pass is performed where the shadows on the virtual Kinect scene due to the new object are created, alongside any self-shadowing on the object. Second, the new object's diffuse and specular components are rendered onto the original RGB image using the light position given by our method. We then subtract the shadow map from this image to obtain the final image. This gives us a realistic looking result for both the shadows by keeping the original colour data from the Kinect (due to the shadow subtraction), and the object by having similar shading to the scene (Figure 7).

7 Discussion

A methodology for estimating the point light source position is described that enables realistic rendering of synthetic objects in a scene recorded by the Kinect. This method can estimate the point light source position based on only the RGB image and 3D depth data given by a RGB-D camera. In our experiments, an average error of 10 degrees between the estimated position and true position of the light source is achieved, which is good enough to render objects in a realistic manner into the scene. In Figure 8, the scene is shown rendered using the true light source and the estimated light source position, where humans can not determine which is the more accurate rendering. In this work, only a single point light
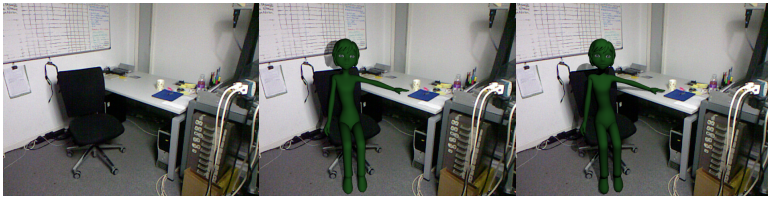
Figure 8: The rendering of a synthetic person in the scene where the shadows are different. Humans can however not observe what the correct light position was given the small difference in angle. (Left: original image, middle: true light position, right: estimated light position)

source is estimated by our method, where in practise the illumination in the scene can be much more complex. Although we experimented with more complex illumination conditions like fluorescent tubes, we observed that our method still performs, because it searches for a light source that gives an explanation for the scene. One possible solution is to extend the method to search for multiple light sources with different intensity, making it computationally more complex. Another domain that is interesting for this method is obtaining illumination invariant scene descriptions allowing recognition from different angles and with varying illumination.

References

[1] Kusuma Agusanto, Li Li, Zhu Chuangui, and Ng Wan Sing. Photorealistic rendering for augmented reality using environment illumination. In *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, ISMAR '03, pages 208–216, Washington, DC, USA, 2003. IEEE Computer Society.

[2] Toby P. Breckon and Robert B. Fisher. Environment authentication through 3d structural analysis. In Aurélio Campilho and Mohamed Kamel, editors, *Image Analysis and Recognition*, volume 3211 of *Lecture Notes in Computer Science*, pages 680–687. Springer Berlin Heidelberg, 2004.

[3] Paul 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*, SIGGRAPH '98, pages 189–198, New York, NY, USA, 1998. ACM.

[4] Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Efficient graph-based image segmentation. *Int. J. Comput. Vision*, 59(2):167–181, September 2004.

[5] J. Frahm, K. Koeser, D. Grest, and R. Koch. Markerless augmented reality with light source estimation for direct illumination. In *Visual Media Production, 2005. CVMP 2005. The 2nd IEE European Conference on*, pages 211 – 220, 2005.

[6] S. Gibson, T.L.J. Howard, and R.J. Hubbold. Image-based photometric reconstruction for mixed reality. In *SIGGRAPH 2001 Sketches and Applications Program*, August 2001.

[7] Kenji Hara, Ko Nishino, and Katsushi Ikeuchi. Light source position and reflectance estimation from a single view without the distant illumination assumption. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(4):493–505, April 2005.

[8] S. Heymann, A. Smolic, K. Müller, and B. Froehlich. Illumination reconstruction from real-time video for interactive augmented reality. In *International Conference on Video and Image Processing*, 2005.

[9] Jean-François Lalonde, Alexei A. Efros, and Srinivasa G. Narasimhan. Estimating the natural illumination conditions from a single outdoor image. *International Journal of Computer Vision*, 98:123–145, 2012.

[10] Yuanzhen Li, S. Lin, Hanqing Lu, and Heung-Yeung Shum. Multiple-cue illumination estimation in textured scenes. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 1366 –1373 vol.2, oct. 2003.

[11] Yanli Liu and Xavier Granier. Online tracking of outdoor lighting variations for augmented reality with moving cameras. *IEEE Transactions on Visualization and Computer Graphics*, 18:573–580, 2012. ISSN 1077-2626. doi: http://doi. ieeecomputersociety.org/10.1109/TVCG.2012.53.

[12] Yanli Liu, Xueying Qin, Songhua Xu, Eihachiro Nakamae, and Qunsheng Peng. Light source estimation of outdoor scenes for mixed reality. *Vis. Comput.*, 25(5-7):637–646, April 2009.

[13] Claus B. Madsen and Brajesh Behari Lal. *Probeless Illumination Estimation for Outdoor Augmented Reality*. INTECH, 2010. ISBN 978-953-7619-69-5.

[14] Claus B. Madsen and Michael Nielsen. Towards probe-less augmented reality - a position paper. In *GRAPP*, pages 255–261, 2008.

[15] J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, 1965.

[16] Yang Wang and Dimitris Samaras. Estimation of multiple directional light sources for synthesis of mixed reality images. In *Proceedings of the 10th Pacific Conference on Computer Graphics and Applications*, PG '02, pages 38–47, Washington, DC, USA, 2002. IEEE Computer Society.

[17] Yang Wang and Dimitris Samaras. Estimation of multiple directional light sources for synthesis of augmented reality images. *Graphical Models (Special Issue on Pacific Graphics)*, 65(4):185–205, July 2003.

[18] Wei Zhou and Chandra Kambhamettu. Estimation of the size and location of multiple area light sources. In *Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 3 - Volume 03*, ICPR '04, pages 214–217, Washington, DC, USA, 2004. IEEE Computer Society.

[19] Wei Zhou and Chandra Kambhamettu. A unified framework for scene illuminant estimation. *Image Vision Comput.*, 26(3):415–429, March 2008.