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

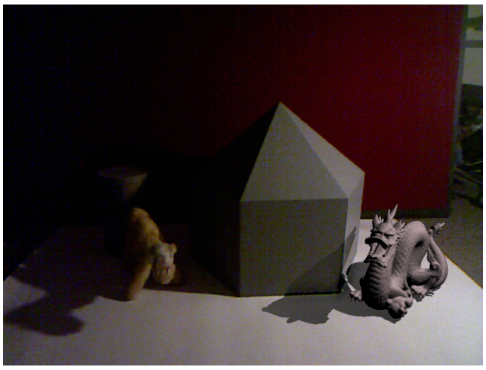

Figure 1: Rendering a synthetic object (dragon) into a real scene recorded with the Kinect sensor, where both the shading and shadows are based on an estimated light source position determined using only the information provided by the Kinect sensor.

Goal: To realistically render objects in a scene taking into account the illumination of the scene. Our approach achieves this by estimating 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 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 a single point light source. 2) Segments in the RGB image with a similar colour that are contiguous tend to have the same albedo.

Contributions: *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 the cases 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.

Method: Using the two assumptions mentioned above, the Lambertian reflectance model can be defined as follows:

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

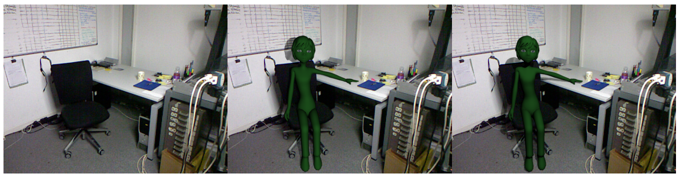

Figure 2: The rendering of a synthetic person in the scene where the shadows are different. Humans can however not observe what the correct light position was given the small difference in angle. (Left: original image, middle: true light position, right: estimated light position)

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. However, both the albedo $\rho(\mathbf{p})$ and light parameters $\mathbf{s}(\mathbf{p}), i$ are unknown. Using the assumption that the albedo in segments R_j of the image are the same: for instance, 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. Given the arbitrary light parameters $\mathbf{s}_r(\mathbf{p}), i_r$ (given by the search method), first the albedo values ρ in the segment R_j can be computed

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

To obtain the same albedo value for the positions in segment ρ_{R_j}, the mean or median of the albedo values $\rho(\mathbf{p}), \mathbf{p} \in R_j$ 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}. A reconstructed image I_r for a given segment R_j can be computed using the following equation:

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

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

By performing this for all segments, we obtain a reconstructed image I_r. An error E between the original image I_o (given in Equation 1) and the reconstructed image I_r obtained in Equation 3, can be calculated as follows: The error E (Equation 4) 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. This operation was performed for an arbitrary point light source (light parameters $\mathbf{s}_r(\mathbf{p}), i_r$), where search methods are used to find the point light source with that minimise the error E.

Discussion: In Figure 2, the scene is shown rendered using the true light source and the estimated light source position, where humans humans find it difficult to determine which is the more realistic rendering. In practise, the illumination in the scene can be very complex, however for the goal of rendering synthetic objects estimations of the illumination conditions are convincing.