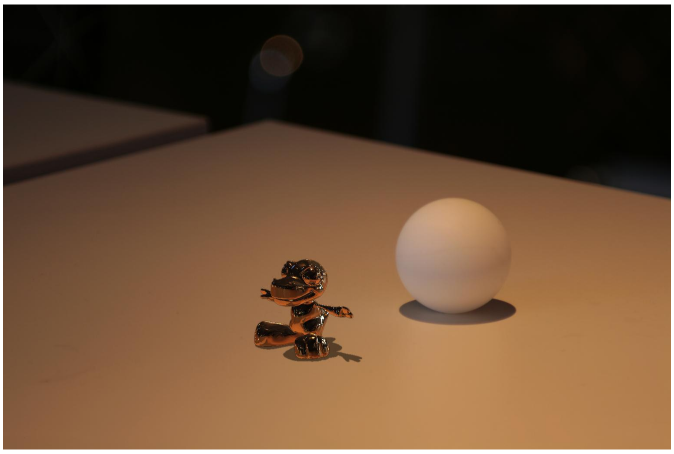
LIGHT SOURCE ESTIMATION USING DUAL LIGHT PROBES

IMAGING TECHNOLOGY

Dan Englesson Joel Ek

Thursday 20th October, 2011
18:17



Abstract

This report covers the implementation and testing of an image processing algorithm heavily based on the work done by [1]. It is de facto an image-based lighting algorithm which uses two reflective spheres as well as a calibration sphere for triangulating the positions of the different light sources of an environment. As opposed to conventional image-based lighting, this method is able to generate correct shadows as a result of the triangulation. The method works fairly well under good circumstances and can triangulate the light source position to a relative error of less than 10 percent when it is located at distances up to 150 cm. However, the method is also very sensitive to almost every parameter that can be modified when capturing the images. It should also be noted that when the light sources are moved further away from the subject, the method becomes obsolete as shadows can be approximated from parallel rays at an arbitrarily large distance.

1 Background

In modern computer graphics, the geometrical data of a computer generated object can be extremely detailed. It is not uncommon for applications to handle millions of triangles per frame. These triangles can be rendered in a straight forward manner on a modern graphics card within a few milliseconds, allowing for interactive frame rates. Though, there is one major drawback. The lighting model (especially for real-time purposes) has not changed drastically over the last 20 years. We are still pretty much limited to the basic Phong shading model or perhaps even Gouraud shading at worst. The problem lies in that light conditions are approximated by a few simultaneous discrete light source positions, located somewhere within the view of the observer. The awareness of this problem has started to grow over the last couple of years, introducing new techniques. Notable ones for real-time use include voxel-based lighting, spherical harmonics and... image-based lighting! But first, a brief introduction to high dynamic range imaging.

1.1 High dynamic range imaging

Most images generated by modern digital cameras have a low dynamic range (LDR). The dynamic range for LDR images is simply the number of possible values for each color channel, where each channel is usually permitted to occupy 8-bits of storage. This gives a dynamic range of 256:1 since it is possible to store 256 discrete values using 8 bits. Problems will arise when one tries to photograph a real environment where the brightest light can be millions of times brighter than the darkest spot in that scene. There is simply too much light detail to describe using only 8 bits. This is where high dynamic range imaging comes to use.

The true light conditions can be approximated from a set of LDR images, each letting the camera gather twice the amount of light compared to the previous image. Every image is referred to as an f-stop due to the fact that every increase in aperture size gives an opening which lets in roughly twice the amount of light. The images are combined using a HDR reconstruction technique which unfortunately is out of the scope for this report. When using this technique, the dynamic range of an HDR image is directly related to the number of f-stops. The relation is 2^n, where n is the number of f-stops. Thus, an image sequence with 15 f-stops would yield a contrast of 32768:1.

1.2 Image-based lighting

Image-based lighting is, simply put, the technique of capturing the lighting conditions of a real environment and introducing synthetic objects into it. The capture can be done using a fish-eye lens, multiple shots stitched into a panorama but is most commonly made using a so called light probe. A light probe is composed of a shiny metal sphere together with some kind of tripod or unipod for placing it in an environment. Common radii of the metal spheres for producing good results are 50 to 75 mm. Also, the better quality of the sphere, the better environment captures can be made. Typically, high dynamic range light probes are made by capturing a set of images where the amount of captured light differs by a factor of two. This can be achieved by altering either the ISO value, the aperture size, the exposure time or any combination of these. Both the ISO value and the aperture size affect other properties than the amount of captured light. Therefore, the best option to change the amount of captured light is given by altering the exposure time.

Figure 1: An angular map created from two HDR sequences

The environment map is usually constructed by capturing two series of low dynamic range images from two different locations. These two locations should be separated 90 degrees apart to avoid bad sampling around the edges of the sphere and to give the possibility to remove the photographer from the final image [3]. This is shown in figure 1.

When the environment map has been created, it can be used to illuminate synthetic objects. For every illuminated fragment in the rendering system, at least one sample is made in the map, depending on the material properties of the fragment. The normal of the fragment is used to find a corresponding sample in the environment map. This sample represents the light coming from the normal direction and can be used as data in a standard Phong shading equation. Reflective and refractive materials make additional samples in the reflected and refracted view direction respectively.

1.3 Light source estimation

The image-based lighting technique comes with one major drawback. Only the light emitting in a certain direction can be determined, not at what distance along that ray the light source is located. By using dual light probes separated by a certain distance, the slight differences in how the environment is mapped onto the spheres can be used to triangulate the positions of the light sources. This enables an application to automatically determine how the shadows should be generated, be it using planar shadows, volumetric shadows or shadow mapping. One might think that the dual light probes would also create a good set of data for generating an environment map. This is not the case due to two main reasons. Firstly, the spheres are not located at the same position and thereby creates two different spatial projections. Secondly, the difference between the two projections is not nearly as big as the needed 90 degrees. If one were to create an environment map, the conventional method with a single light probe should be used.

2 Method

2.1 Equipment setup

For capturing the images, the following set of equipment was used.

- A Canon 400D digital camera
- A Canon 85 to 200 mm variable zoom lens
- Two tripods
- Two low-grade reflective spheres with a 25.5 mm radius
- One diffuse sphere with a radius of 20 mm
- Two rulers and some magnets
- Measuring tape and some string

The digital camera was mounted together with the zoom lens on one of the tripods and the spheres were arranged on the other tripod as shown in figure 2. The rulers were used to measure the distances between the three spheres for usage in the estimation application.

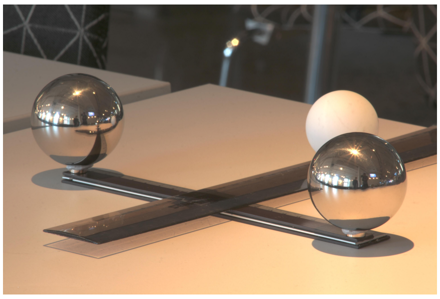


Figure 2: The setup of the dual light probes

2.2 Sphere detection

Spheres are shown as ellipses when projected onto an image by a camera. If the distance between a sphere and a camera is large enough, the projected ellipse will converge towards a circle. Circles are easier to describe mathematically than ellipses which is why a the camera was positioned as far away as permitted while still retaining focus on the three spheres.

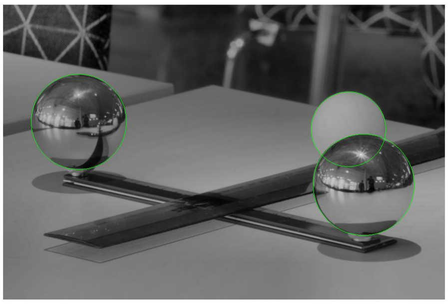


Figure 3: Hough transform for spheres

The Hough transform is a great method for finding mathematically definable shapes within an image. This method was used to detect the three most prominent circles, corresponding to the three spheres, see 3. However, only the image coordinates and the image radii of the spheres can be approximated. These are of no good use when trying to triangulate the real world position of a light source... yet.

2.3 Focal length correction

The lens used allowed for the focal length to be set between 85 and 200 mm. Using this information and the measured data about the arrangement of the spheres, one could estimate the inverse perspective projection of the camera as described in [1]. The focal length was supplied as meta data in the EXIF file produced by the camera and could also be read from the lens. However, this value is only a crude approximation as the real value can differ by several percent. In the implementation, three different estimations are made based on the method suggested in [1].

$$\begin{gathered} f = -\sqrt{\frac{d^2-(A_x-B_x)^2-(A_y-B_y)^2}{(\frac{A_R}{Ar}-\frac{B_R}{B_r})^2}} \\ (A_x, A_y) = (A_u\frac{A_R}{A_r}, A_v\frac{A_R}{A_r}) \\ (B_x, B_y) = (B_u\frac{B_R}{B_r}, B_v\frac{B_R}{B_r}) \end{gathered} \tag{1}$$

The method in equation 1 uses the known information about the relative distance between two of the spheres and the measured image coordinates (u, v) of the two center points as well as the radii, both physical (R) and projected (r). A virtual sphere located at one of the two real spheres is used for the estimation. Its radius is equal to the relative physical distance between the two spheres. This gives enough information to calculate the true focal length.

Three different candidates are calculated as mentioned. They correspond to the three different pairs of two spheres possible to create out of the set of three. The one value which is closest to the value indicated by the EXIF meta data is used for the following steps.

2.4 Depth approximation of the spheres

Using the true focal length, the spatial depth information of the three spheres can be recovered. For every sphere, equation 2 uses the image coordinates (u, v), physical radius (R) and projected radius (r) to reconstruct the spatial depth information of the sphere.

$$\begin{gathered} A_z = f\frac{A_R}{A_r} \\ B_z = f\frac{B_R}{B_r} \end{gathered} \tag{2}$$

2.5 Segmenting the highlights

Highlights will appear on the two reflective spheres where bright light sources have been reflected through the normal towards the camera. By using the data extracted from the Hough transform, the two reflective spheres can be matted from the background. Since HDR imaging is used, a thresholding operation is made at 90 percent of the max intensity. This results in a binary image where ones correspond to pixels above the 90 percent limit. Morphological operators can be applied to extract the centroid coordinates (u, v) of the highlight S_i. In the implementation, two passes of erosion are made to filter out distant light sources which are of no interest since the algorithm focuses on light sources located nearby.

2.6 Calculating the surface normal

The image coordinates for the highlights can be inversely projected in the same manner as for the central positions of the spheres. The method described in [1] states that a surface normal can be found using the relations shown in equations 3 and 4. Equation 4 describes a quadratic equation with one unknown variable (N_z) which can be solved by conventional methods. In these two equations, (S_u, S_v) are the image coordinates of the highlights, (A_x, A_y, A_z) the estimated world position of the sphere, f the corrected focal length, A_R the actual radius and A_r is the projected radius of the sphere.

$$
\begin{gathered}
(S_u, S_v) = \left(f\frac{A_x + A_R N_x}{A_z + A_R N_z}, f\frac{A_y + A_R N_y}{A_z + A_R N_z}\right) \\
N_x = \frac{S_u A_R N_z + S_u A_z - f A_x}{f A_R} \\
N_y = \frac{S_v A_R N_z + S_v A_z - f A_y}{f A_R} \\
N_x^2 + N_y^2 + N_z^2 = 1
\end{gathered} \tag{3}
$$

$$
\begin{gathered}
A_R^2(S_u^2 + S_v^2 + f^2)N_z^2 + \\
2A_R(S_u(S_u A_z - f A_x) + \\
S_v(S_v A_z - f A_y))N_z + \\
(S_u A_z - f A_x)^2 + (S_v A_z - f A_y)^2 - \\
(A_R f)^2 = 0
\end{gathered} \tag{4}
$$

2.7 Approximation of the reflection vectors

By combining the approximated spatial information with the surface normals, simple vector reflection can be used to calculate two light direction vectors, each pointing from the highlight to the proximity of the light source. This is a normal raytracing procedure with a fundamental difference; the position of the light source is unknown. The entire raytracing procedure can be seen in figure 4.

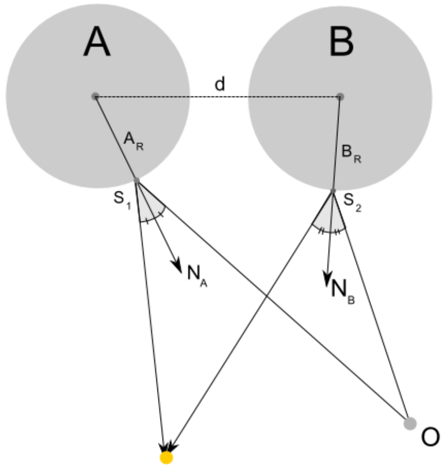


Figure 4: The raytracing procedure

2.8 Smallest perpendicular distance

Since two lines embedded in three dimensional space is most likely skew, a simple intersection point cannot be calculated. Instead a coordinate system is inferred from three, non-orthonormal, basis vectors. The basis vectors are the two light direction vectors and their cross product. By using a method detailed in [2], the scaling parameter for each light direction vector can be calculated. The two scaling parameters s and t correspond to the point where perpendicular distance is the smallest and are calculated as 5.

$$
\begin{gathered}
L_1(t) = P_1 + tV_1 \\
L_2(s) = P_2 + sV_2 \\
\\
P_1 + tV_1 = P_2 + sV_2 \\
P_1 - P_1 + tV_1 = P_2 - P_1 + sV_2 \\
t(V_1 \times V_2) = (P_2 - P_1) \times V_2 + s(V_2 \times V_2) \\
t(V_1 \times V_2) \cdot (V_1 \times V_2) = (P_2 - P_1) \times V_2 \cdot (V_1 \times V_2) \\
\\
t = \frac{(P_2 - P_1) \times V_2 \cdot (V_1 \times V_2)}{(V_1 \times V_2) \cdot (V_1 \times V_2)} \\
s = \frac{(P_2 - P_1) \times V_1 \cdot (V_1 \times V_2)}{(V_1 \times V_2) \cdot (V_1 \times V_2)}
\end{gathered} \tag{5}
$$

3 Result

The implementation was tested using a test series composed of 12 images where the light source position was varied, both in incident direction and distance. Every test image was shot in a room where no other light sources were present. To verify the results, true distances were measured and written down for comparison. Only distances were measured since the environment was rather limited and provided no option to measure spatial coordinates in a more accurate way. The measured distances were in the range between 50 and 250 cm. As shown in figure 5, the accuracy of the method starts to diverge around 150 cm.

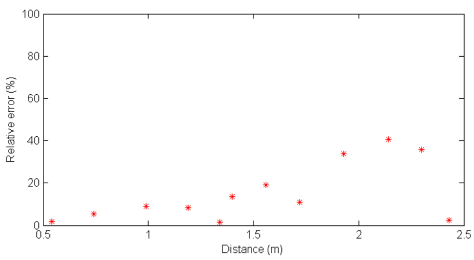


Figure 5: Relative error of the estimated distance

The figure shows an accuracy for the method of within 10 percent of the measured value. However, this is only true up to roughly 150 cm. The divergence is mainly due to the light direction vectors becoming more parallel as the distance grows. This in combination with the position of the camera can result in highlights which are projected at almost identical image coordinates relative to the two spheres, which is bad.

Figure 6 depicts the resulting triangulation of one of the test images giving a good result. The two light direction vectors clearly converge at a point roughly 150 cm above the spheres.

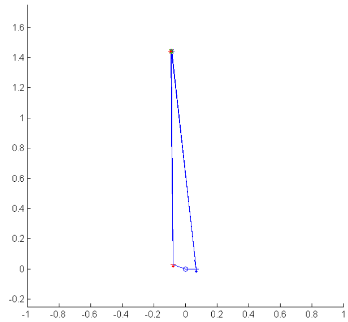


Figure 6: 3D plot of the estimated light position

4 Conclusion

The method is clearly useful for estimating the light source positions and from those the resulting shadows cast by the synthetic objects. However, since the relative error grows with distance, the method is rather limited to nearby light sources. It should once again be noted that image-based lighting cannot correctly reproduce shadows from the information in the environment map. The method can also be useful in SFX production of scenes where the light sources are close to the synthetic objects and close ups can be expected. In this case, this method can be used in conjunction with image-based lighting to model distant light sources as well as diffuse inter-reflection.

Another field of interest could be light-stage calibration where many close light sources positions needs to be spatially and spectrally calibrated. This seems to be untested as of 2011 and would be interesting and suitable for future work!

Furthermore, trying to estimate a single point in 3D space from two skew vectors can be a bad idea. If the method would incorporate area light sources, the method would be less dependent on the approximation of the surface normals. This is in general a good idea due to the fact that the errors in the normal estimation spread throughout the implementation causing inaccurate values. Such a technique could also be useful in offline rendering where the global illumination is approximated from image-based lighting. In contrast to the common method where the light is sampled over the entire hemisphere, the global illumination could use the extracted area light sources producing a significant speed up.

5 References

1. M. W. Powell, S. Sarkar, D. Goldgof. A Simple Strategy for Calibrating the Geometry of Light Sources, 2001.

2. A. S. Glassner. Intersection of two lines in three-space, 1998. Graphics gems, p 304.

3. ICT graphics lab. http://gl.ict.usc.edu/HDRShop /tutorial/tutorial5.html, fetched 2011-10-20.