TOWARDS PROBE-LESS AUGMENTED REALITY
A Position Paper

Claus B. Madsen and Michael Nielsen
Computer Vision and Media Technology Lab, Aalborg University, Aalborg, Denmark

Keywords: Augmented Reality, illumination estimation, rendering, shadow detection, shadow rendering.

Abstract: The main problem area for Augmented Reality is ensuring that the illumination of the virtual objects is continuously consistent with the illumination in the real scene. State of the art in the area typically requires the real scene illumination conditions to be captured as a High Dynamic Range environment map. The environment map is then used for shading and shadowing. Handling the real and the virtual shadows and their interaction is the single most difficult aspect. This paper presents a completely different approach to determining the illumination conditions in the real scene. Based on an assumption that the scene is outdoor we automatically detect shadows in the image and use this information to determine the ratio of sky irradiance to sun irradiance. We then present how to convert this information into radiance levels for both the sky and the sun. When combined with a computation of the Sun's position based on date, time and information about position on the Earth, we arrive at a full illumination model applicable for rendering virtual objects into real scenes.

1 INTRODUCTION

Without doubt Augmented Reality (AR) will become a widespread technology within few years. The proliferation of computer-based, portable imaging devices such as cell phones and PDAs makes it very attractive to develop AR techniques that can enable augmentation of images with credible renderings of virtual geometry,- for entertainment, education, and information purposes. To augment the real world with real-time renderings of monsters to combat in the street, to augment the real world with images of how ancient architecture looked or to augment the real world with route finding information, etc.

There are three technical problems areas to solve in order to accomplish photo-realistic AR: 1) camera registration, 2) occlusion handling, and 3) estimating the real world illumination. We conjecture that the first two problems will have a feasible technological solution within relatively few years. Regarding registration GPS combined with image-based tracking of features will allow for knowing the position and orientation of the camera in real time. Regarding occlusion handling, that is, determining whether a virtual object occludes a real object or vice versa, will also eventually be solved through combinations of laser range finding and multi-view 3D reconstruction. The third problem, illumination, still holds many complicated challenges, unless certain assumptions/restrictions are made.

In this paper we present an approach to estimating the illumination conditions in a real scene for use in AR applications. Our approach is quite different from main stream work in this area, which involve acquiring a complete High Dynamic Range omnidirectional environment map of the scene, e.g., acquired as a light probe using a reflective sphere. Our approach is to determine the illumination conditions directly from an image of the scene.

In general estimating illumination from an unknown scene is ill-posed and may never be completely solved. The things that make our approach work are: 1) we assume information about date, time and position on Earth is available for the image, 2) we assume the image is of an outdoor scene with only natural (sky and sun) illumination, and 3) we assume there is a predominant occurrence of approximately diffuse surfaces in the scene. The first assumption is very reasonable, since cameras in the near future will include GPS information into the image header, as well as date and time. The second assumption is reasonable, since there is a lot of outdoor world to photograph. The third assumption is also reasonable since urban scenes have a lot of road, pavement, brick, and concrete surfaces, which all are approximately Lambertian. Moreover, we conjecture that the diffuse surface assumption can be relaxed greatly in the future due to further research.

B. Madsen C. and Nielsen M. (2008).
TOWARDS PROBE-LESS AUGMENTED REALITY - A Position Paper.
In *Proceedings of the Third International Conference on Computer Graphics Theory and Applications*, pages 255-261
DOI: 10.5220/0001099502550261

Figure 1: Illustration of idea in this work with images. Left: input image acquired outside in sunshine. Center: real shadows automatically detected and removed to illustrate performance of detection. Right: information from shadow detection step has been used to estimate Sun and Sky illumination conditions and virtual objects have been rendered into the scene with credible shading and shadowing.

The general idea in the work presented here is that the outdoor (daylight) illumination conditions can be modeled quite accurately by a distant disk light source (the Sun) in combination with a Sky dome. We assume the radiance of the Sky dome is the same over the entire hemi-sphere. The respective radiances of the Sky and the Sun are estimated based on information from a automated shadow detection process. That is, the shadows already present in the image provide the *only source of information* for the parameters of the illumination model (a part from the position of the Sun, which is found procedurally from date, time and Earth location information). Figure 1 shows an example of application of the techniques presented here.

The bold statement in this paper is that photo-realistic Augmented Reality using no explicit illumination calibration, such as light probes, will be possible sometime in the not so distant future. The position taken in the paper is that the scene itself and images of it contain enough information to reconstruct the illumination conditions, and we present a specific technique which achieves exactly that on images within a well defined sub-set of outdoor scenery.

2 RELATED WORK

The most widely applied technique for obtaining illumination information from a real scene is environment maps, also sometimes referred to as light probes. Light probes are images taken of a reflective sphere placed in the scene. This image can be remapped to different omni-directional mappings which can be applied for shading virtual objects, either in a non-real-time or a real-time rendering system, (Debevec, 2005; Madsen et al., 2003; Havran et al., 2005; Barsi et al., 2005; Jensen et al., 2006; Madsen and Laursen, 2007; Debevec, 1998; Debevec, 2002).

This technique provides very precise information about the illumination conditions in terms of radiance and incoming direction, but there are several drawback. First of all the light probe image has to be in a High Dynamic Range (HDR) format, which as of yet requires multiple exposures of the same (static) scene. It is though fair to assume that consumer cameras in the future will provide HDR information in a single exposure. Secondly, the coordinate system of the probe has to be calibrated to the scene coordinate system, as well as the camera. Furthermore, the information contained in the light probe is only valid at the precise point where the probe image was acquired, so generally a single probe acquisition cannot be used to model the illuminations across a larger scene (the distant scene assumption). The latter problem can though be alleviated given a coarse 3D model of the scene onto which the light probe image can be back-projected, as in (Gibson et al., 2003). The single most annoying thing about the light probe approach, though, is that the probe information becomes invalid in dynamic scenes, with changing illumination conditions. For example in outdoor scenes. It is simply not realistic to place a reflective sphere in the scene every once in a while to capture the illumination conditions.

While the light probe approach is pre-dominant in the literature, there are a few other approaches. For a recent review of approaches to estimating illumination in Augmented Reality see (Jacobs and Loscos, 2004). The most promising frameworks employ some form of inverse rendering, where illumination and reflection properties of the surfaces of the scene are iteratively determined based on one or more images of the scene, (Yu et al., 1999; Boivin and Gagalowicz, 2001; Boivin and Gagalowicz, 2002). The drawback of these approaches is that they require perfectly modeled, complete 3D scene geometry and complete knowledge about light sources (position, radiance, radiation characteristics, etc.).

The final category of related work go in a direction toward trying to understand the illumination condition based on single images of the scene with as little as possible prior information and/or assumptions. Some work is based on a single image of a geometrically known 3D object casting shadow on a planar surface of known reflectance, (Sato et al., 1999). Although elegant this technique's primary shortcoming is the requirement that a 3D model is available of the shadow casting object, which makes the technique less suitable for general, automated real applications. Other work, (Cao et al., 2005), extracts quantitative illumination information from images of general shadows, in a manner related to what is presented in this paper, but Cao's work does not provide the illumination information required for *shading* a virtual object correctly, only for determining the color characteristics of virtual shadows.

The following issues have been a priority for the work presented in this paper: we aspire to develop techniques which have a potential for real-time performance on dynamic scenes responding to constantly changing illumination condition; we also wish for the technique to be based on information obtained directly from images without requiring the presence of special purpose objects for illumination estimation; we aim toward techniques which assume as little as possible concerning materials in the scene.

3 OVERVIEW OF APPROACH

The presented technique assumes the availability of the 3D geometry of the areas in the scene which must receive shadows from virtual objects. For most examples in this paper this simply means a ground plane. We strongly envisage that real-time dense stereo reconstruction in the future will provide this rough geometry. In this work we manually calibrate the camera position and orientation to the ground plane coordinate system, and estimate camera focal length, based on a small number (at least four) of known points in the scene. Similar information could in a real-world application be supplied by built-in inertia sensors in the camera, combined with automatic multi-view calibration and reconstruction. Additionally we assume knowledge of how the ground coordinate system is oriented relative to magnetic North. Finally, the approach requires that it is possible to compute the direction vector to the Sun, which requires knowledge of time of day, date, as well as the latitude and longitude of the position where the image is taken. A built-in solid state magnetic compass, a clock and a GPS receiver can provide this information for a consumer camera.

The technique is based on first running a shadow detection process in the image with no prior information other than that the image is taken under outdoor daylight illumination conditions. Figure 1 shows the performance of the shadow detection process by its ability to remove shadow effects, although the removal of shadows naturally is not the objective. Based on the shadow detection process we get an estimate of the ratio between total irradiance in areas in direct Sun and the total irradiance in areas in Shadow.

From this ratio, combined with a simple white balancing assumption, this paper shows how it is possible to determine the values of all parameters of a complete illumination model consisting of a Sky dome of a certain radiance, and Sun disk light source of a certain radiance. Both the Sky and the Sun part of the model have scene consistent color balance.

Based on this illumination model, with its determined parameters, it is possible to render virtual objects with scene consistent shading and shadows, as illustrated through a number of examples in the paper. Here a software-based path tracing framework has been employed, but it is entirely straight forward to perform similar quality rendering in real-time.

4 SHADOW DETECTION

It is beyond the scope of the present paper to fully describe the applied shadow detection approach, which is documented in (Nielsen and Madsen, 2007b; Nielsen and Madsen, 2007a). In general terms the technique is based on using pixel statistics in the chromaticity plane to estimate the RGB values of an overlay, which when alpha blended on pixels in regions in direct light change these regions into shadow regions. Using a graph cut algorithm the method then determines the correct alpha values for all pixels in the image, such that an alpha value of 0 corresponds to full direct light, and a value of 1 corresponds to full shadow. Figure 2 illustrates the performance showing the technique's ability to deal with soft shadows.

The shadow detection is based on a general outdoor illumination model. The Achilles heel of the technique is the initialization where pixel statistics are used to determine the color of the shadow overlay. The technique is quite successful at initializing itself, but not flawless. We postulate that fully automated shadow detection will perform adequately at some point, given the research interests in the area and the current performance of known techniques. Additionally, it will be much simpler to achieve robust initialization in video streams of dynamic scenes, where e.g. moving people and vehicles, as well as the movement of shadows caused by the movement of the Sun, will make it much easier to hypothesize what regions in an image are shadow regions.

(a) Input image

(b) Detected shadow levels

Figure 2: Top: input image. Bottom: estimated levels of shadow at various regions in the image. Please refer to figure 1 to see shadows removed from this image.

5 ILLUMINATION ESTIMATION

The shadow detection overlay described in the preceding section actually has a concrete physical interpretation for diffuse surfaces. We use this information to drive the computation of all parameters of the scene illumination model.

As described previously the proposed illumination model consists of a Sky dome covering an entire hemisphere above the scene, combined with a distant disk source to model the Sun. The Sun subtends a 0.53 degree diameter disk viewed from Earth. The position of the Sun relative to the scene can be quite easily computed given the information listed in section 3 (time, date, latitude and longitude), see e.g. (Schlyter, 2007). Let $\vec{s}$ denote the direction vector in scene coordinates to the Sun at a given time, for a given location on Earth. The information concerning the illumination model we *do not* have are the respective radiances of the Sky dome and the Sun disk.

Subsequently, whenever a radiometric quantity is being used it is to be implicitly understood that the quantity has values for each color channel, i.e., has 3 components, Red, Green and Blue. For each color channel we therefore have two unknowns: the Sky radiance and the Sun radiance. The process of determining these values starts with the overlay produced by the shadow detection process. Let the color of the overlay be denoted by C_o. It is easy to prove that for diffuse surfaces the overlay value corresponds to the ratio of the irradiance in shadow to the irradiance in direct light for a given surface normal:

$$C_o = \frac{E_a}{E_a + E_s} \tag{1}$$

where E_s is the irradiance due to the Sun for a given surface normal, and E_a (a for atmosphere) is the irradiance due to the Sky. If somebody wishes to employ the ideas of this paper, but do not have access to a shadow detection system this quantity is easily found manually in images by taking the average pixel values in a shadow region and component-wise divide them by the average pixel values of the same surface in direct Sun light.

The overlay values, C_o, provide one constraint on the two unknowns since it constrains the relative values of the Sun and the Sky radiances, although we shall stick with irradiances until at the very end. Let $E_s^\perp$ denote the irradiance, due to the Sun, for a normal pointing straight into the Sun. Furthermore, let V_a determine the fraction of the Sky dome visible for a given point in the scene. Then the fraction of shadow to direct light irradiance for a point in the scene can be expressed as:

$$\begin{aligned} C_o &= \frac{V_a \cdot E_a}{V_a \cdot E_a + E_s^\perp \cdot (\vec{N} \cdot \vec{s})} \\ &\Updownarrow \\ E_s^\perp &= E_a \frac{V_a \cdot (1 - C_o)}{(\vec{N} \cdot \vec{s}) \cdot C_o} \end{aligned} \tag{2}$$

where $\vec{N}$ is the surface normal of the point. Now the Sun's head-on irradiance is expressed in terms of the Sky irradiance times properties from the image and from partial 3D knowledge of the scene. If the only scene model consists of a ground plane (as in most of our examples) V_a is simply set to 1. This is the most information we can get from the overlay color, i.e., a relative constraint.

The next constraint, which enables us to determine the relative strengths of the RGB components, is based on an assumption that the camera has been white-balanced to the scene and the illumination conditions. If the camera is white-balanced to areas in direct sunlight the combined Sun and Sky irradiance at a white-balanced point is a constant for all color channels:

$$k = E_s^\perp \cdot (\vec{N}' \cdot \vec{s}) + E_a \cdot V_a' \tag{3}$$

Here we use $\vec{N}'$ and V_a' to indicate that the white-balance direction and its associated Sky dome visibility may be different than the direction for which the overlay is tuned to provide full shadow. For the examples given in the paper we have assumed that the camera is white-balanced for the ground plane. In practice we have in reality let the camera perform automatic white-balancing, but since the ground plane dominates in the example views this is roughly the same as groundplane white-balance. Note that we are talking about an illumination white-balancing which results in white paper lying on the ground appearing white (has balanced RGB values); we are not talking about assuming that the ground plane is grey/white!

Combining Eqs. 2 and 3 yields:

$$E_a = \frac{k}{V_a' + V_a \cdot ((\vec{N}' \cdot \vec{s})/(\vec{N} \cdot \vec{s})) \cdot (1/C_o - 1)} \quad (4)$$

Now the Sky irradiance is expressed solely in scene and image components, so E_a is computed first using Eq. 4, after which $E_s^{\perp}$ is computed by inserting into Eq. 2. Now the color balance and relative strengths of the Sky and the Sun irradiances are determined, and we only lack to determine the absolute levels. By arbitrarily setting the albedo, ρ, of some point in the scene to $1/3$ (Earth's average albedo) we get a final constraint allowing us to set the absolute values of the irradiances so as to be suitable for illuminating virtual objects into the image, in which the pixel values (scene radiances) are naturally subject to some unknown camera scale/gain factor. Let a pixel value for a surface with normal $\vec{N}_p$ be L_p. Then the unknown scaling factor, S, for the Sun and Sky irradiances can be found from:

$$L_p = 1/(2\pi) \cdot S \cdot (E_a \cdot V_{ap} + E_s^{\perp} \cdot (\vec{N}_p \cdot \vec{s})) \cdot \rho$$

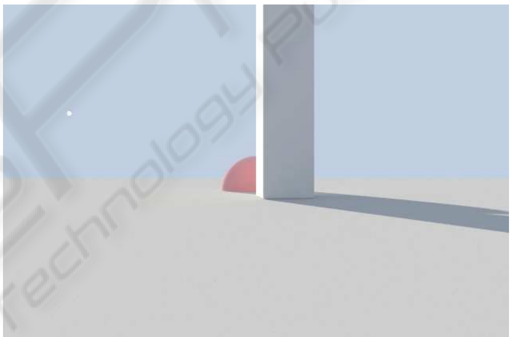
$$\Updownarrow$$

$$S = \frac{2\pi \cdot L_p}{(E_a \cdot V_{ap} + E_s^{\perp} \cdot (\vec{N}_p \cdot \vec{s})) \cdot \rho} \quad (5)$$

Both E_a and E_s must be scaled by S as determined by Eq. 5, setting irradiance values that are appropriate for a scene given the input image camera exposure. Now, all that remains is to convert from irradiance to radiance for the Sky dome and the Sun disk, respectively. The irradiance for a normal pointing straight into a hemi-spherical sky dome of radiance L is $\pi \cdot L$, therefore the Sky radiance is set to $L_a = E_a/\pi$. The irradiance produced on a normal pointing straight into a disk source of radius r (in radians) and radiance L is $2\pi \cdot (1 - cos(r)) \cdot L$ (for small disks), so the Sun radiance is set to $L_s = E_s^{\perp}/(2\pi \cdot (1 - cos(r)))$, which for a 0.53 degree diameter Sun disk corresponds to a scalefactor of 14880.

Figure 3 visually illustrates the illumination estimation for a given scene. Notice how the color of the Sky dome corresponds to the actual Sky color in the image, although this information has not taken part at all. The only information used is the overlay color resulting from finding the intensity ratio between a region in direct light and a region in shadow.

(a) Input image



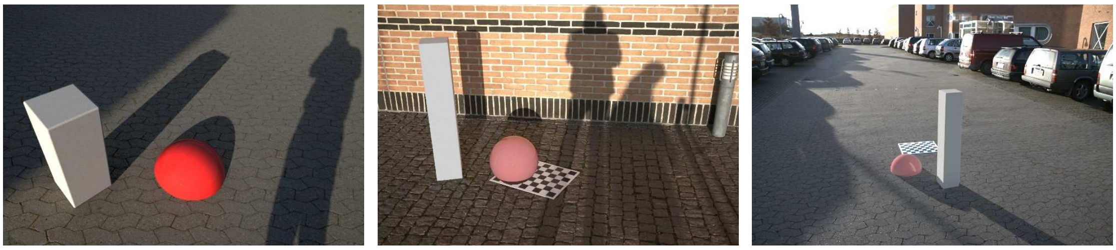
(b) Estimated sky radiance

Figure 3: Top: input image. Bottom: rendering showing local scene, sky and sun. The sky's and sun's radiances have been estimated directly from the shadow information in the input image. Figure 1 shows final composite of local scene into input image.

6 RENDERING PROCESS

In terms of actually rendering augmentation into the images there is a choice between two different overall approaches: 1) differential rendering as in (Debevec, 1998) or 2) a relighting approach similar to the one described in (Madsen and Laursen, 2007). We have chosen the latter as it is much easier to apply, due to the fact that it is in practice impossible to provide correct albedos for the so called local scene, i.e., that part of the real scene for which a 3D model exists.

We have used the free ray tracing package Radiance by Greg Ward for the renderings in this paper. Figure 4 shows a few more rendering examples. The



(a) Example 1 (b) Example 2 (c) Example 3

Figure 4: Different rendering examples based on images taken within a brief hour of sunlight during an otherwise very rainy and gray month of November 2007, in Denmark.

rendering process can be listed as follows (unfortunately we do not have space to illustrate with images):

1. render irradiance values of local scene without augmentation objects, e.g. just ground plane
2. render irradiance values of local scene including augmentation objects
3. render radiance values of local scene including augmentation objects, i.e., render the local scene with the estimated illumination conditions consisting of a Sky dome and a Sun disk
4. render augmentation mask (binary image, zero at pixels that correspond to augmentation objects, one elsewhere)
5. multiply input image with augmentation mask
6. divide masked input image from step 5 by irradiance image from step 1
7. multiply image from step 6 with irradiance image from step 2
8. multiply local scene rendering from step 3 with inverse of the augmentation mask from step 4
9. add relit image from step 7 with masked augmentation from step 8

7 DISCUSSIONS AND FUTURE WORK

There are naturally unresolved issues. One deals with how to avoid effectively creating double shadows when rendering a virtual shadow on top of a real one, resulting in a much too dark shadow. This we have a solution for, which is not described in this paper. Figure 5 shows that it is possible to render virtual shadows across real shadow without creating double shadow by simply using the shadow level mask created by the shadow detection module. Unfortunately overlapping shadows will not occur unless the virtual object casts a shadow also on the real object creating the real shadow, or vice versa. We have ideas for partially handling these interactions, but this is left for future research.

It would also be interesting to measure the performance of the technique presented in this paper against light probe images to establish the degree of absolute accuracy.

Figure 5: Shadow manually drawn into image. The pixel values inside the artificial shadow is automatically determined. Using this approach we intend to solve the shadow protection problem in order to avoid double shadows.

8 CONCLUSIONS

We have presented a method for determining all parameters of a complete outdoor illumination model based entirely on simple image measures from images with shadow. The illumination model consists of a Sky dome and a Sun disk. The main contribution is that the presented technique can continuously estimate the changing illumination conditions in a real outdoor scene bypassing the need for special purpose objects in the scene such as reflective spheres for light probe/environment map acquisition or known 3D objects in the scene for illumination estimation.

We have demonstrated on real images that we can render credible augmentations into the images including global illumination effects such as contact shadows and color bleeding from virtual objects into real objects.

The technique assumes the availability of time, date, compass heading and Earth location information, all of which represent information which it is quite feasible can be produced automatically in consumer cameras.

ACKNOWLEDGEMENTS

This research is funded by the CoSPE project (26-04-0171) under the Danish Research Agency. This support is gratefully acknowledged.

REFERENCES

Barsi, A., Szirmay-Kalos, L., and Szécsi, L. (2005). Image-based illumination on the gpu. *Machine Graphics and Vision*, 14(2):159 – 169.

Boivin, S. and Gagalowicz, A. (2001). Image-based rendering of diffuse, specular and glossy surfaces from a single image. In *Proceedings: ACM SIGGRAPH 2001*, Computer Graphics Proceedings, Annual Conference Series, pages 107–116.

Boivin, S. and Gagalowicz, A. (2002). Inverse rendering from a single image. In *Proceedings: First European Conference on Color in Graphics, Images and Vision, Poitiers, France*, pages 268–277.

Cao, X., Shen, Y., Shah, M., and Foroosh, H. (2005). Single view compositing with shadows. *The Visual Computer*, pages 639 – 648.

Debevec, P. (1998). Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proceedings: SIGGRAPH 1998, Orlando, Florida, USA*.

Debevec, P. (2002). Tutorial: Image-based lighting. *IEEE Computer Graphics and Applications*, pages 26 – 34.

Debevec, P. (2005). A median cut algorithm for light probe sampling. In *Proceedings: SIGGRAPH 2005, Los Angeles, California, USA*. Poster abstract.

Gibson, S., Cook, J., Howard, T., and Hubbold, R. (2003). Rapic shadow generation in real-world lighting environments. In *Proceedings: EuroGraphics Symposium on Rendering, Leuwen, Belgium*.

Havran, V., Smyk, M., Krawczyk, G., Myszkowski, K., and Seidel, H.-P. (2005). Importance Sampling for Video Environment Maps. In Bala, K. and Dutré, P., editors, *Eurographics Symposium on Rendering 2005*, pages 31–42,311, Konstanz, Germany. ACM SIGGRAPH.

Jacobs, K. and Loscos, C. (2004). State of the art report on classification of illumination methods for mixed reality. In *EUROGRAPHICS*, Grenoble, France.

Jensen, T., Andersen, M., and Madsen, C. B. (2006). Real-time image-based lighting for outdoor augmented reality under dynamically changing illumination conditions. In *Proceedings: International Conference on Graphics Theory and Applications, Setúbal, Portugal*, pages 364–371.

Madsen, C. B. and Laursen, R. (2007). A scalable gpu-based approach to shading and shadowing for photo-realistic real-time augmented reality. In *Proceedings: International Conference on Graphics Theory and Applications, Barcelona, Spain*, pages 252 – 261.

Madsen, C. B., Sørensen, M. K. D., and Vittrup, M. (2003). Estimating positions and radiances of a small number of light sources for real-time image-based lighting. In *Proceedings: Annual Conference of the European Association for Computer Graphics, EUROGRAPHICS 2003, Granada, Spain*, pages 37 – 44.

Nielsen, M. and Madsen, C. (2007a). Graph cut based segmentation of soft shadows for seemless removal and augmentation. In *Proceedings: Scandinavian Conference on Image Analysis, Aalborg, Denmark*, pages 918 – 927.

Nielsen, M. and Madsen, C. (2007b). Segmentation of soft shadows based on a daylight and penumbra model. In *Proceedings: MIRAGE 2007*, pages 341 – 352.

Sato, I., Sato, Y., and Ikeuchi, K. (1999). Illumination distribution from brightness in shadows: Adaptive estimation of of illumination distribution with unknown reflectance properties in shadow regions. In *Proceedings: International Conference on Computer Vision*, pages 875 – 882.

Schlyter, P. (2007). *Computing Planetary Positions - a Tutorial With Worked Examples*. www.stjarnhimlen.se/comp/tutorial.html.

Yu, Y., Debevec, P., Malik, J., and Hawkins, T. (1999). Inverse global illumination: Recovering reflectance models of real scenes from photographs. In *Proceedings: SIGGRAPH 1999, Los Angeles, California, USA*, pages 215 – 224.