EUROGRAPHICS 2005 / J. Marks and M. Alexa
(Guest Editors)

Volume 24 (2005), Number 3

Light Waving: Estimating Light Positions From Photographs Alone

Holger Winnemöller, Ankit Mohan, Jack Tumblin and Bruce Gooch

Northwestern University, Evanston, IL 60201, U.S.A. †

Abstract

We present an algorithm to automatically estimate three-dimensional light positions from an unordered set of images. We collect images using a single stationary camera while manually moving a light source around an object. Rather than measuring light positions directly, the algorithm extracts a three-dimensional manifold of positions from the images using dimensionality reduction techniques. This obviates the need for calibration and specialized equipment, making our approach inexpensive, portable and applicable to objects of almost any size. We demonstrate our results using image-based relighting applications.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism I.2.0 [Artificial Intelligence]: Vision and Scene Understanding I.4.1 [Image Processing and Computer Vision]: Digitization and Image Capture

1. Introduction

Image-based rendering and relighting can be used to create highly realistic images with arbitrary illumination without the need to compute light-transport from first principles. Most image-based relighting techniques require as input a set of basis images, each lit from a different *known* light position. Novel illumination conditions can then be created by computing a weighted sum of these basis images [NSD94].

This paper presents a light-position estimation algorithm that is simple, fast and reasonably robust. Our goal is to achieve a convincing relighting effect, rather than capture an accurate reflectance field. We are willing to trade some accuracy in positional estimates for increased portability and reduced data acquisition time.

Figure 1 shows the main steps of our algorithm. We take a set of digital images of an object with a fixed camera and varying light positions that form a roughly hemispherical pattern around the object. We perform dimensionality reduction on these images to determine a three-dimensional embedding of relative light-positions. To counteract embedding and scaling artifacts inherent in the reduction process, we project the resulting data-points onto a sphere using least-squares optimization.

2. Related Work

In recent years, image-based lighting has been used for various applications such as changing the incident light in images and videos [DHT*00, DWT*02], combining real and virtual worlds [Deb98], and creating realistic light-dependent textures [MGW01, DvGNK99].

Debevec et al. [DHT*00] introduced the 8D reflectance field of an object that couples the 4D incident light field with the 4D radiant light field of the object. They sampled a 4D slice of this field by capturing photographs of the object from a fixed viewpoint as a robotic arm moves a point-lightsource in a hemisphere around the object [DWT*02]. Malzbender et al. [MGW01] used a small geodesic dome of electronic flash units for capturing directional illumination effects. Matusik et al. [MPN*02] captured a 6D

† [holger|ankit|jet|bgooch]@cs.northwestern.edu

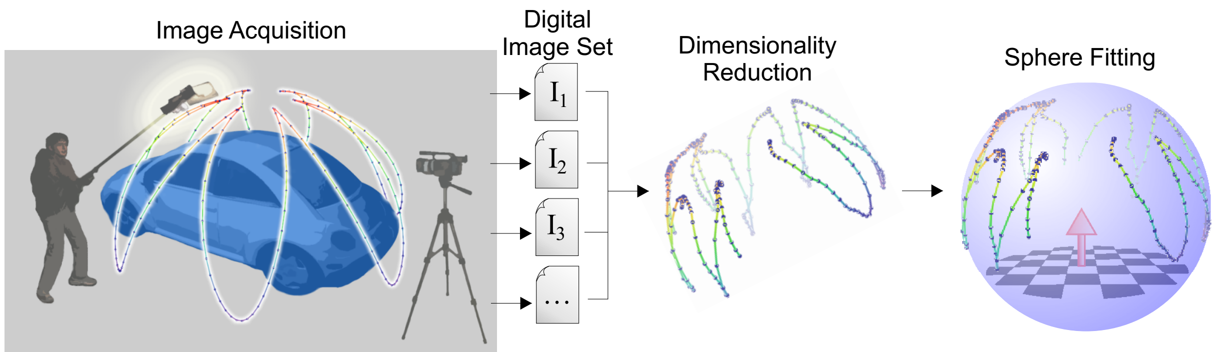

Figure 1: *System Overview: Images are acquired with an uncalibrated camera and a handheld light. Isomap reduces dimensionality of images to 3D. Resulting co-ordinates are projected onto the upper half of a sphere.*

slice by relaxing the fixed viewpoint constraint. Sen et al. [SCG*05] apply Helmholz reciprocity to interchange light source and camera positions in an attempt to speed up the image capture process. All of these methods require extensive calibration and precise measurements. Most require non-portable equipment and considerable setup and data acquisition time.

Our algorithm is most similar to the *free-form light stage* [MDA02], where a light is manually moved around an object while a calibrated camera continuously takes pictures of the scene. Light directions for each photograph are computed from the shading of four diffuse spheres, which are placed at carefully chosen positions in the scene. Instead of a calibrated setup and specialized equipment we rely on dimensionality reduction techniques such as multi-dimensional scaling (MDS) and Isomap [TdSL00] to estimate the light-positions from the images.

Earlier work in inverse rendering tries to estimate material and illumination properties in photos and synthetic renderings, e.g. [MG97, YDMH99]. However, most of these approaches require at least an approximate 3D model of the scene. Various authors have suggested indirect, statistical approaches to estimating light direction in images by assuming purely Lambertian object surfaces [Pen82, LR85, CBG94]. We have found our algorithm to perform well even for objects with strong specular highlights.

Pless's work on estimating object rotation and camera elevation from an unordered set of images using Isomap dimensionality reduction [PS02a, PS02b] can be considered the dual of the light position estimation problem. Their initial two-dimensional embedding results sometimes exhibited non-intuitive deformations, which they corrected by applying external constraints. By analyzing our image-data in three dimensions we are able to minimize this problem.

3. Data Acquisition

Data acquisition for most previous image-based relighting work requires some combination of a carefully calibrated camera, accurate placement of light probes, complex lighting-domes of fixed size, high-dynamic range photography and a lengthy procedure that takes between 30 minutes and several hours to perform. While our algorithm results in physically inexact estimates of light positions, we gather input images in 2-5 minutes per object and obtain visually convincing results for various relighting applications.

As illustrated in Figure 1, all we require is a fixed camera framing the object to be photographed and a freely movable diffuse light source. The camera is then set to take pictures at regular intervals, while the user waves the light around the object in a hemispherical pattern. Users must take care to avoid shining the light directly into the camera as this may lead to an underexposed object and lens flare. Additionally, neither the light nor any of its attachments should obscure the object or cast visible shadows onto it. This means that slightly elevated camera positions that tightly frame the object from a distance outside the lighting hemisphere yield the best results.

4. Light Pose Estimation

The main contribution of our work is the estimation of relative light-positions from images alone, without the help of explicit light-probes, fiducials, shadow analysis or direct measurements. We achieve this by applying a dimensionality reduction technique, called Isomap, to the captured image set.

4.1. Dimensionality Reduction

To explain how a dimensionality reduction technique is able to extract positional estimates from our input images, consider the following illustrative example.

Imagine a black disk on which a white dot is painted. As the disk is rotated about its center, the center of the white dot moves through various positions $\mathcal{P} = \{P_1, P_2, \ldots, P_m\}, P_i \in \mathbb{R}^{n=2}$ (Figure 2, Left). At each configuration, P_i, an image, I_i, is taken, giving rise to the image set $\mathcal{I} = \{I_1, I_2, \ldots, I_m\}, I_i \in \mathbb{R}^N$. Each image can be regarded as an N-dimensional vector where $N = width \times height$ (each pixel in the image representing a dimension).

We can now ask the question: Given only $\mathcal{I}$, can we recover the relative positions of the white dots, $\mathcal{P}$? We obviously want to solve this problem as generically as possible without making any assumptions about the rotational axis of the disk, or any other geometry, i.e. we want to find a model-free solution. Strictly speaking, we then cannot even ask the above question, because we don't know about the existence of any dots. One way of addressing the problem is to simply compare pairs of images. Since the position of white dots is all that changes between two images, comparing pairs of images should give an estimate of the relative dot positions.

This type of problem is commonly addressed using Multidimensional Scaling [CC01]. Metric MDS takes as input a dissimilarity matrix, $D(i, j)$, representing the pairwise distances between all N-dimensional element pairs (i, j) and yields equivalent co-ordinates in an alternative co-ordinate system, such that the original distance relationships are preserved. This new co-ordinate system has the important property that the first axis contains the largest error/variance, the second a little less, and so on. This means that the contributions of higher-order axes become decreasingly significant so that in practice only the first n dimensions need be considered, where $n < N$. This effectively reduces the dimensionality of the data representation while introducing a well-defined error.

For MDS to work well, we need a distance metric which faithfully represents the dissimilarity of the changes in configuration that we wish to preserve through the dimensionality reduction, i.e.

$$D(i, j) \propto |P_i - P_j| \; \forall \, (i, j) \in (\{1 \ldots m\}, \{1 \ldots m\}) \quad (1)$$

where $D(i, j)$ is the difference between I_i and I_j using some difference metric.

Lacking a model on which to base a suitable difference metric, the L^2 (Euclidean) norm is commonly used and simple to compute. In our example, the L^2

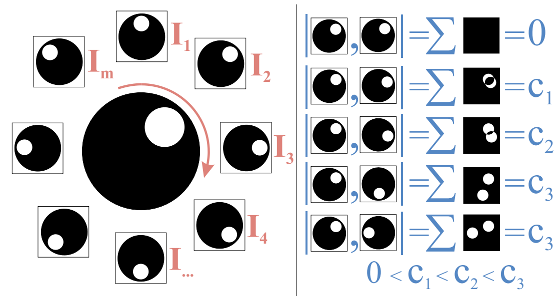

Figure 2: *Dimensionality Reduction Example: (Left) A black disk containing a white dot is rotated around its center. For the i^{th} position of the white dot, P_i, we take a snapshot image, I_i. (Right) Example distances: two identical configurations have zero distance. Distances increase as the two configurations diverge until they become disjoint. In this case, the distance remains constant for any pair of disjoint configurations.*

distance gives a measure of the dot-overlap in two images (Figure 2, Right). If the dots are in the exact same position in two images then the total image difference is zero. As the overlap area decreases, the L^2 distance between images increases monotonically. However, if the dots do not overlap, the image distance equals twice the area of the dot, and is *independent of the actual position of the dots*. This means that L^2 only satisfies Equation 1 for small image distances.

Tenenbaum et al [TdSL00] devised a dimensionality reduction scheme, called Isomap, which addresses the problem that most generic dissimilarity measures are reliable only when differences are small, but unreliable when differences are large. The basic Isomap algorithm works as follows:

1. Compute distance matrix, $D(i, j)$ for all data-points using some locally reliable dissimilarity measure (such as the L^2 norm).
2. For each I_i, find its k closest neighbors based on the distance matrix, $D(i, j)$, and call these sets $\mathcal{N}_k(I_i)$. (Note that although $I_j \in \mathcal{N}_k(I_i) \not\Leftrightarrow I_i \in \mathcal{N}_k(I_j)$, such symmetry is enforced in the Isomap implementation.)
3. Construct a modified distance matrix that only contains reliable distance measures as follows:
$$D_G(i, j) = \begin{cases} D(i, j), & \text{if } I_j \in \mathcal{N}_k(I_i) \\ \infty, & \text{otherwise.} \end{cases}$$
4. Use a shortest path algorithm such as Floyd-Warshall's or Dijkstra's algorithm [CLRS01] to solve the all-pairs shortest path problem on the

distance matrix $D_G(i,j)$. The matrix $D_G(i,j)$ now contains pairwise distances derived from reliable distance information.

5. Compute MDS using $D_G(i,j)$ as input.

By linking up local neighborhoods through shared neighbors, Isomap is able to consider a globally consistent embedding even without a globally valid difference metric. Therefore, Isomap relaxes Equation 1 to:

$$D_G(i,j) \propto |P_i - P_j| \; \forall \; i \in \{1 \ldots m\}, I_j \in \mathcal{N}_k(I_i) \quad (2)$$

In summary, by applying Isomap to a set of raw input images we can directly compute co-ordinates for the changing parameters that gave rise to the images. To emphasize, this does not imply that Isomap is able to identify an object and track its progress through space (no temporal or other data ordering is assumed), but that relative distances between all data-pairs are preserved via an appropriate positional arrangement in the low-dimensional embedding.

4.2. Application to Lightwaving

We can now proceed to apply the principles discussed in the previous section to the problem of estimating relative light positions for Lightwaving. As the light moves around an object through positions $\mathcal{P}$, we take a set of images, $\mathcal{I}$, of the object. We then apply Isomap to $\mathcal{I}$ using the L^2 distance metric between pairs of images. If Equation 2 holds, the result is an embedding in 3D space equivalent to $\mathcal{P}$.

Unfortunately, the notion of *overlap*, is more difficult to define for Lightwaving than for our simple example, because image intensities are a complicated function of scene geometry, object material properties, light source parameters, camera optics and camera response curves. Nonetheless, we can take certain measures to maximize coherence as the light is moved. We employ a light source with a large diffuse reflector, and move it to ensure significant overlap in the area covered by the light in consecutive frames. We also use auto-exposure on the camera and linearize images [DM97] before computing distances.

Figure 3 shows some of the patterns we found to work well. A symmetric and uniform pattern generally works better than ad hoc patterns where large regions may be undersampled. In particular, we found it helpful to have a good sampling of light positions vertically over the object.

Isomap with a very small neighborhood count, k, often results in poor mappings. This is not surprising as Isomap recalculates the initial distance-matrix between all image pairs by pruning all but the closest k neighbors. Intuitively, this means that for a star pattern like those used in Figure 3, *Hoberman* and *Telephone* we would compute the distance of a point on one star-arm to a point on the next by going via the central axis. This artificially increases the estimated distances between those points. In such cases Isomap allows us to increase the k value to include adjacent arms in the light waving pattern. The value for parameter k is currently found using a binary search approach to minimize the residual error of the Isomap operation.

4.3. Spherical Fitting

As Isomap results contain their largest statistical spread in the first co-ordinate, a little less in the second, etc., recovered light-positions may be non-uniformly scaled across all dimensions. To address this issue, we take advantage of the fact that during data-gathering the light was moved in a hemispherical pattern and fit the data to a sphere using least-squares optimization. This yields a center for the sphere, a radius, as well as a normal defining the relevant hemi-part of the sphere. Additionally, it also helps to remove outliers and other noise-related artifacts, which may exist in the embedding.

Additionally, the relative ordering of light-positions is invariant with respect to rotation and reflection. The user has the option of fixing these degrees of freedom by interactively choosing two known light-positions, such as opposite and perpendicular to the camera.

5. Results

We used a *Nikon D70* digital camera in continuous mode (approximately 3 fps) for the *Teapot* dataset, and a *Canon GL2* digital camcorder for all others. We found a diffuse table lamp to be adequate for most of the smaller objects, and a professional 250W hand-held light attached to a diffuse foam core reflector was enough for larger objects such as the *Beetle car*. As our method does not require any calibration or precise measurement, the setup time was approximately the same as that required for a single photograph, and we gathered about 1–3 minutes of light waving video, discarding all but a few (about 300–500) evenly spaced frames. We also discard frames where the light moved into the camera's view or caused lens flare. We then resized the frames to around 240×160 before using them for Isomap computation, which took approximately a minute on a modern workstation.

Figure 3 shows some of the objects and scanning patterns we tested. While we have no tracking equipment to gather ground truth light-position values for comparison, the recovered positions faithfully resemble the

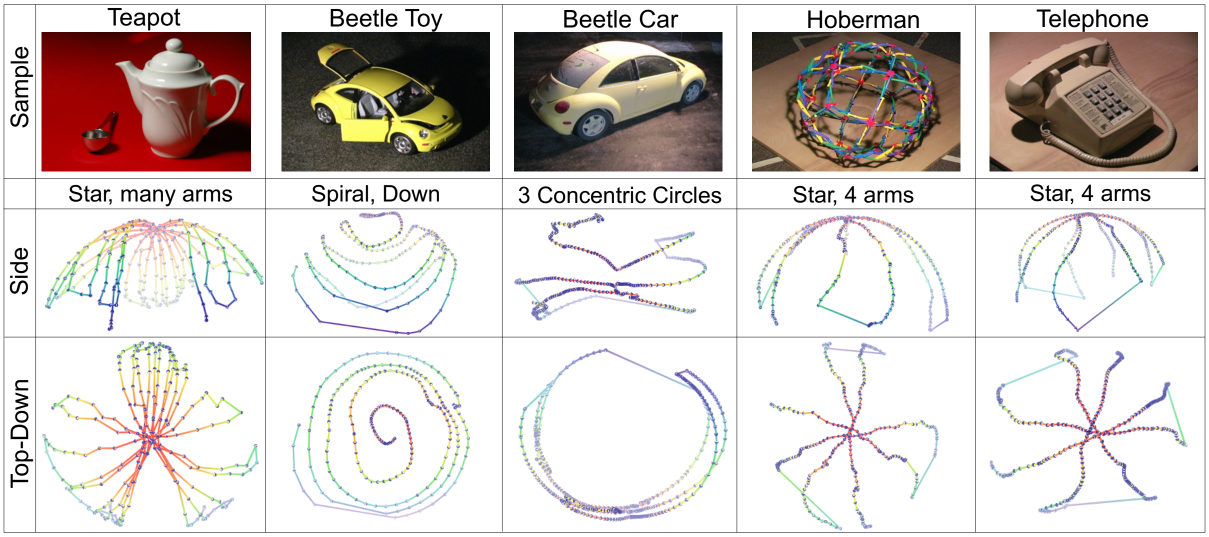

Figure 3: *Recovery results for a series of objects and light waving patterns: (Top) Sample images with arbitrary light position. (Bottom) 3D Light-waving patterns recovered by Isomap, shown in top & side views.*

scanning patterns that we used during data capture. Even samples that were eliminated due to the light passing in front of the camera or shining into it leave a corresponding gap in the reconstructed light-waving pattern (see supplemental video for an example).

We applied our results to generate polynomial texture maps (PTMs) as described by Malzbender et al. [MGW01]. A PTM encodes the effect of lighting on a texture. Each pixel of a PTM stores coefficients of a 2D quadratic polynomial that best approximates pixel color changes with changing illumination direction. Due to the low degree polynomial function, illumination varies smoothly, but is not always an accurate model for sharp specular highlights and occlusions. Using our approximate light positions to construct PTMs results in very little visible difference. Please see accompanying video for examples of PTMs obtained by our algorithm.

We also demonstrate relighting of objects with arbitrary environment maps captured using light probe images as shown in Figure 4. For each basis image, the light positions are determined with our technique. We use the methods described in [DHT*00, MDA02] on these light positions to compute the weights of the basis images and sum the weighted images to obtain the final result.

6. Discussion and Future Work

Apart from the shape of the light-waving pattern and number of samples obtained, the neighborhood count, k, is the only parameter in our approach. As the quality of light-position estimates depends on selecting k from an appropriate range we are currently investigating how neighborhood size is related to scene properties, such as object shape and material, lighting, and camera placement.

Since our data acquisition requires no calibration, we anticipate that correlating results from one or more additional cameras will improve the accuracy of our results without significantly increasing the complexity of the algorithm. Most of these cameras can be of webcam quality, while a main camera captures hi-resolution or even high dynamic range images.

We envision our algorithm being used in settings where ease and speed of use are paramount to physical accuracy. One such setting would be fast and easy acquisition for a relightable image-stock database. Images of requested objects could be downloaded by artists and relit for compositing with an existing scene.

7. Acknowledgements

We would like to thank the computer graphics group at Northwestern University for their support and feedback and the anonymous reviewers for their helpful comments and suggestions. This material is based upon work supported by the National Science Foundation under Grant No. 0415083. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors.

Figure 4: *Relighting examples: Images of a telephone and a plant (original samples in top-right corner) re-lit using an incident light map measured by a light probe (bottom-left circle).* [Light probe image courtesy Paul Debevec.]

References

[CBG94] Chojnacki W., Brooks M., Gibbins D.: Revisiting Pentland's estimator of light source direction. *Journal of the Optical Society of America A 11* (Jan. 1994), 118–124. 2

[CC01] Cox T. F., Cox M. A. A.: *Multidimensional Scaling*. Chapman & Hall/CRC, 2001. 3

[CLRS01] Cormen T. H., Leiserson C. E., Rivest R. L., Stein C.: *Introduction to Algorithms (Second Edition)*. The MIT Press, 2001. 3

[Deb98] Debevec P.: Rendering synthetic objects into real scenes. In *SIGGRAPH '98* (1998), ACM Press, pp. 189–198. 1

[DHT*00] Debevec P., Hawkins T., Tchou C., Duiker H.-P., Sarokin W., Sagar M.: Acquiring the reflectance field of a human face. In *SIGGRAPH '00* (2000), pp. 145–156. 1, 5

[DM97] Debevec P. E., Malik J.: Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH '97* (1997), vol. 31, pp. 369–378. 4

[DvGNK99] Dana K. J., van Ginneken B., Nayar S. K., Koenderink J. J.: Reflectance and texture of real-world surfaces. *ACM Trans. Graph. 18*, 1 (1999), 1–34. 1

[DWT*02] Debevec P., Wenger A., Tchou C., Gardner A., Waese J., Hawkins T.: A lighting reproduction approach to live-action compositing. In *SIGGRAPH '02* (2002), pp. 547–556. 1

[LR85] Lee C.-H., Rosenfeld A.: Improved methods of estimating shape from shading using the light source coordinate system. *Artificial Intelligence 26*, 2 (1985), 125–143. 2

[MDA02] Masselus V., Dutré P., Anrys F.: The free-form light stage. In *Eurographics workshop on Rendering* (2002), pp. 247–256. 2, 5

[MG97] Marschner S. R., Greenberg D. P.: Inverse lighting for photography. In *Proceedings of IS&T/SID Fifth Color Imaging Conference* (1997), pp. 262–265. 2

[MGW01] Malzbender T., Gelb D., Wolters H.: Polynomial texture maps. In *SIGGRAPH '01* (2001), ACM Press, pp. 519–528. 1, 5

[MPN*02] Matusik W., Pfister H., Ngan A., Beardsley P., Ziegler R., McMillan L.: Image-based 3D photography using opacity hulls. In *SIGGRAPH '02* (2002), pp. 427–437. 1

[NSD94] Nimeroff J. S., Simoncelli E., Dorsey J.: Efficient re-rendering of naturally illuminated environments. In *Fifth Eurographics Workshop on Rendering* (1994), pp. 359–373. 1

[Pen82] Pentland A. P.: Finding the illumination direction. *Journal of the Optical Society of America 72*, 4 (April 1982), 448–455. 2

[PS02a] Pless R., Simon I.: Embedding images in non-flat spaces. *Conference on Image Science Systems and Technology* (2002), 182–188. 2

[PS02b] Pless R., Simon I.: Using thousands of image of an object. In *Proceedings of the 6th Joint Conference on Information Science* (March 2002), pp. 684–687. 2

[SCG*05] Sen P., Chen B., Garg G., Marschner S. R., Horowitz M., Levoy M., Lensch H. P. A.: Dual photography. In *SIGGRAPH '05* (2005). 2

[TdSL00] Tenenbaum J. B., de Silva V., Langford J. C.: A global framework for nonlinear dimensionality reduction. *Science 290*, 5500 (December 2000), 2319–2323. 2, 3

[YDMH99] Yu Y., Debevec P., Malik J., Hawkins T.: Inverse global illumination: recovering reflectance models of real scenes from photographs, 1999. 2