Fast freeform reflector generation using source-target maps

1-1-2010

Florian R. Fournier
University of Central Florida

William J. Cassarly

Jannick P. Rolland
University of Central Florida

Find similar works at: https://stars.library.ucf.edu/facultybib2010

University of Central Florida Libraries http://library.ucf.edu

Recommended Citation

Fournier, Florian R.; Cassarly, William J.; and Rolland, Jannick P., "Fast freeform reflector generation using source-target maps" (2010). Faculty Bibliography 2010s. 158.
https://stars.library.ucf.edu/facultybib2010/158

This Article is brought to you for free and open access by the Faculty Bibliography at STARS. It has been accepted for inclusion in Faculty Bibliography 2010s by an authorized administrator of STARS. For more information, please contact lee.dotson@ucf.edu.
Fast freeform reflector generation using source-target maps

Florian R. Fournier,1,* William J. Cassarly,2 and Jannick P. Rolland3,1

1 CREOL, The College of Optics and Photonics, University of Central Florida, Orlando, FL 32816, USA
2 Optical Research Associates, 1100 Hunt Club Drive, Wooster, OH 44691, USA
3 The Institute of Optics, University of Rochester, 275 Hutchison Rd, Rochester, NY 14627, USA
*florian.fournier@gmail.com

Abstract: We propose a freeform reflector design method based on the mapping of equi-flux grids between a point source and a target. This method imposes no restriction on the target distribution, the reflector collection angle or the source intensity pattern. Source-target maps are generated from a small number of target points using the Oliker algorithm. Such maps satisfy the surface integrability condition and can thus be used to quickly generate reflectors that produce continuous illuminance distributions.

©2010 Optical Society of America

OCIS codes: (080.4298) Nonimaging optics; (080.4228) Nonspherical mirror surfaces; (220.2945) Illumination design; (080.1752) Computation methods.

References and links
1. F. Fournier, and J. Rolland, “Design methodology for high brightness projectors,” J. Disp. Technol. 4(1), 86–91 (2008).
2. L. B. W. Jolley, J. M. Waldrum, and G. H. Wilson, The theory and design of illuminating engineering equipment (Chapman & Hall, London, 1930).
3. W. B. Elmer, The optical design of reflectors, 2d ed. (Wiley, New York, 1980).
4. V. I. Oliker, “Mathematical aspects of design of beam shaping surfaces in geometrical optics,” in Trends in Nonlinear Analysis (2002), pp. 191–222.
5. H. Ries, and J. Muschaweck, “Tailored freeform optical surfaces,” J. Opt. Soc. Am. A 19(3), 590–595 (2002).
6. W. A. Parkyn, “Segmented illumination lenses for step lighting and wall washing,” in Current Developments in Optical Design and Optical Engineering VIII, (SPIE, 1999), 363–370.
7. Y. Ding, X. Liu, Z. R. Zheng, and P. F. Gu, “Freeform LED lens for uniform illumination,” Opt. Express 16(17), 12958–12966 (2008).
8. L. Wang, K. Qian, and Y. Luo, “Discontinuous free-form lens design for prescribed irradiance,” Appl. Opt. 46(18), 3716–3723 (2007).
9. R. Frankot, and R. Chellappa, “A method for enforcing integrability in shape from shading algorithms,” IEEE Trans. Pattern Anal. Mach. Intell. 10(4), 439–451 (1988).
10. L. Noakes, and R. Kozeza, “2D leapfrog algorithm for optimal surface reconstruction,” in Vision Geometry VIII, (SPIE, 1999), 317–328.
11. F. R. Fournier, W. J. Cassarly, and J. P. Rolland, “Designing freeform reflectors for extended sources,” in Nonimaging Optics: Efficient Design for Illumination and Solar Concentration VI, (SPIE, 2009), 742302.
12. W. J. Cassarly, “Nonimaging Optics: Concentration and Illumination,” in Handbook of optics, 2nd ed. (McGraw-Hill, New York, 1995).
13. F. Fournier, and J. Rolland, “Optimization of freeform lightpipes for light-emitting-diode projectors,” Appl. Opt. 47(7), 957–966 (2008).

1. Background

Reflectors are widely used for illumination tasks such as street lighting, automotive headlights, medical lighting, and projector light engines [1]. Designing such non-imaging reflectors consists in finding the surface shape that converts light emitted by a source into a desired target distribution. Historically, conics were the first reflector shapes used to intensify light emitted by candles or gas lamps. However, conics do not offer enough degrees of freedom to accurately control flux distribution over an extended area. In the 1930s, new techniques were derived to design non-imaging reflectors with rotational or translational
symmetry, where the source is assumed to be a point [2,3]. These symmetry considerations allow the system to be reduced to a 2D problem in a plane, thus simplifying calculations. Complexity increases greatly if we consider the general case where the system has no rotational symmetry. In that case the reflector shape is governed by a set of nonlinear partial differential equations [4]. Unfortunately, there are no known analytical solutions for this type of differential equation, and standard numerical solutions often fail to converge. Nonetheless, three approaches have so far been proposed to tackle the general reflector problem. Oliker solves the problem for a discrete target distribution with a finite set of ellipsoids using an iterative algorithm [4]. Ries and Muschaweck propose a new formulation of the problem but do not detail the actual numerical solution [5]. Finally, Parkyn, Ding and Wang use equi-flux grids that produce reflector surfaces with step discontinuities [6–8]. This latter approach breaks up the problem into two independent steps. First, the flux emitted by the source is distributed onto the target. Once a mapping between the source and the target has been derived, the reflector shape is obtained by numerical integration from an initial starting point. The mapping of the source flux onto the target is implicitly given by flux conservation:

\[ \int \int I_s(\theta, \phi) \sin \theta \, d\theta \, d\phi = \int \int E_t(x, y) \, dx \, dy \quad (1) \]

where \( I_s \) is the intensity distribution of the point source in spherical coordinates and \( E_t \) the desired illuminance distribution on a flat target perpendicular to the \( z \) axis, as shown in Fig. 1. Ideally, one would want to obtain from Eq. (1) relationships between the source emission angle and the target location, so that

\[ (x, y) = \left(f(\theta, \phi), g(\theta, \phi)\right) \quad (2) \]

where \( f \) and \( g \) are continuous functions that map uniquely each source ray direction \((\theta, \phi)\) to a target location \((x, y)\). Equation (2) cannot be solved directly in the general case, but mappings can be obtained analytically in some cases where the system geometry allows separation of variables.

Figure 1 shows an example where a square target must be illuminated uniformly with an isotropic source. In this case the reflector collects light over a hemisphere. The target map on the right was derived analytically using symmetry considerations, following the methodology presented in [6–8]. On the other hand, the target map on the left was derived using the method presented in this paper, which will be detailed in section 2. Both mappings achieve the same flux distribution on the target, but as we shall see only the mapping on the left can be achieved by a smooth continuous surface. In order to understand why, let’s have a look at how the map in Fig. 1b was created. The solid angle over which the light emitted by the source is collected can be partitioned into equal solid angles by making equal solid angle slices in both the \( \theta \) and \( \phi \) directions. Since the source is isotropic, the amount of flux is the same in each solid angle. The target can also be partitioned into equal area cells in the same fashion as the source by first slicing the target concentrically (concentric blue contours in Fig. 1b) and then radially (green lines in Fig. 1b). By comparing the source grid and the target grid, we can determine where each ray emitted by the source must hit the target. We can therefore compute the normals to the reflector surface required to aim source rays to their corresponding target locations. However, there is no guarantee that a continuous surface can fit all surface normals. In fact, a field of surface normals \( N \) cannot be compatible with a smooth continuous surface unless it satisfies the integrability condition

\[ \eta = N \cdot (\nabla \times N) = 0 \quad (3) \]

This smoothness constraint is often used for surface reconstruction algorithms in computer graphics [9]. The integrability condition comes from the fact that integrating the surface normals along any closed loop should always be zero (in other words, we should end up at the
same location where we originally started). Using Stokes’ theorem then yields Eq. (3), as detailed in the appendix.

Fig. 1. Source flux is mapped onto the target by matching equi-flux grids. In this example, the solid angle in red contains the same flux than the red area on the target. The equi-flux grid on the right was derived analytically based on the radial symmetry of the system but does not satisfy the integrability condition and thus leads to a discontinuous reflector surface. The equi-flux grid on the left also corresponds to a uniform square target, but it fulfills the integrability condition and can therefore be achieved with a smooth continuous surface.

Fig. 2. Magnitude of $\eta$ (residual projected curl) for the two target maps presented in Fig. 1. $\eta$ is plotted as a function of the target map location. The analytical target map on the right results in large residual curl as we move away from the axis. In comparison, the alternate mapping on the left minimizes $\eta$ over the entire target and can therefore be achieved with a smooth continuous reflector surface.

Among all possible mappings that generate the desired target flux distribution, we therefore need to find one that also satisfies the integrability condition. Figure 2 plots the magnitude of $\eta$, the residual projected curl of $\mathbf{N}$ as defined in Eq. (3), for the two target maps shown in Fig. 1. The analytical mapping presented in Fig. 2b, albeit intuitive, does not satisfy the integrability condition, and therefore cannot yield a smooth continuous surface. One possible solution is to introduce step discontinuities in the surface [6–8]. Some algorithms have also been devised to adjust surface normals in order to minimize $\eta$ [10], but a
satisfactory result may be difficult to obtain if the residual curl is too large or spreads unevenly. Alternatively, the plot in Fig. 2a shows that the mapping method we propose minimizes $\eta$ over the entire target, and can therefore yield smooth reflector surfaces. We will describe in section 2 how such maps are generated, and in section 3 how maps can be used to quickly generate reflector shapes.

2. Generating maps that satisfy the integrability condition

Vladimir Oliker has proposed an iterative technique to solve the reflector problem for a discrete target distribution and a point source [4]. An overview of the method is given in [11]. The strength of the algorithm is its flexibility, as it makes no assumption on the target shape, target distribution, source distribution or collection angle. It leverages the classic property of ellipsoids, which is that the flux emitted from one focus is reflected to the second focus. One ellipsoid is created for each point on the discrete target distribution. All ellipsoids have one common focus placed at the source location; the second focus coincides with each target point. Once the foci are defined, the only remaining degree of freedom is the focal parameter of each ellipsoid. This focal parameter can be used as a scaling parameter to control how much flux is collected by each ellipsoid, and therefore how much flux is reflected towards each target point. An iterative algorithm then scales each ellipsoid so that the desired amount of flux is obtained at each target point.

Reflectors generated this way are thus made of a collection of coincident ellipsoid facets (with slope discontinuities at the facet edges), as shown in Fig. 3a. Since the Oliker algorithm provides a discrete solution to the problem, it also indirectly provides a discrete solution to Eq. (2). In other words, we can use the reflector generated with the Oliker algorithm in order to a posteriori retrieve the mapping between the source and the target. Figure 3a describes the map retrieval process. In this example, the target is a square made of 25 target points. The center of each ellipsoid facet is associated to its corresponding target point, except for target edges where the middle of the facet outer edge is used instead of the center (red dots in Fig. 3a). Target values on the edge must thus be weighted appropriately to prevent edge effects. This way, we map the coordinates $(x_i, y_i)$ of each target point to a source direction $(\theta_i, \phi_i)$. In other words, we obtain discrete versions of the mapping relationships $x = f(\theta, \phi)$ and $y = g(\theta, \phi)$. This set of data points is then approximated by a surface in order to generate a continuous mapping function, as shown in Fig. 3b. The maps shown in Fig. 1, 4 and 5 can finally be generated by plotting the $(x,y)$ contours corresponding to constant $\theta$ values (green lines) or constant $\phi$ values (blue lines).

Figure 4 shows maps for three uniform target distributions: square, rectangle and off-axis square. These geometries were chosen because they are common in street lighting applications, but it is worth pointing out that the map retrieval method outlined before can be applied to any target distribution and any point source intensity distribution. Maps in Fig. 4 were generated from initial reflectors made of 400 ellipsoids. The number of ellipsoids was chosen as the result of a tradeoff between the map accuracy and the time it takes to generate the map. In this case, 400 ellipsoids correspond to a threshold after which increasing the number of ellipsoids has little effect on the maps. Using more ellipsoids would increase the computation time with negligible benefit. In this example all targets have uniform illuminance, which implies that all target cells have the same area. We see that for the square and rectangular targets, the iso-$\theta$ contours (blue lines) are nearly circular at the center of the target and gradually become squarer closer to the edges of the target. A good fit for these contours can be obtained with superellipses (also referred to as Lamé curves). Interestingly, it appears that the rectangular map matches closely a stretched version of the square map.
Fig. 3. (a) Target points are mapped to the center of their corresponding facets. In order to preserve edges, points at the edge of the target are mapped to the middle of the outer facet instead of the facet center. (b) The discrete mapping relationships \( x = f(\theta, \phi) \) and \( y = g(\theta, \phi) \) can then be approximated by a surface in order to get a continuous mapping function. (c) The generated map plots the contours corresponding to constant \( \theta \) and constant \( \phi \) values.

Fig. 4. Target maps for three different system geometries: square, rectangle and off-axis square. In all cases the target is uniform, the point source is isotropic, and the collection angle of the reflector is a hemisphere.

We looked at how maps evolve as we vary parameters such as the target distance \( z \), the target size, the reflector size, the source distribution, or the solid angle over which light is collected (collection angle). As a general rule, maps are invariant under moderate changes, although the corresponding reflector shapes may be dramatically different. For instance, the same map can be used if a Lambertian source is used instead of an isotropic source, as long as the source equi-flux grid is modified accordingly. The collection angle appears to be the most sensitive parameter, while target size, target distance and reflector size can sometimes be changed by an order of magnitude with minimal impact on the map.

Faceted reflectors are commonly used to obtain uniform distributions and to reduce sensitivity to source tolerances [12]. However, designing faceted reflectors for non-rotationally symmetric distributions can be challenging. Figure 5 shows source-target maps for individual reflector facets centered on three different azimuth angles: 0°, 22.5° and 45°. In each case the elevation angle is \( \theta = 60^\circ \) and the collection angle is \( 20^\circ \times 20^\circ \). These maps were generated from 225 ellipsoids for each facet. We see that contours follow the facet...
orientation as the azimuth angle increases. The map is thus close to a standard rectangular grid when $\phi = 0^\circ$, and becomes diamond shaped when $\phi = 45^\circ$. It is worth noting that when $\phi$ is different from $0^\circ$, the corners of the facet do not match the corner of the target. Corners cannot be mapped to corners when the azimuth $\phi$ is different from $0^\circ$, because the resulting twist in the mapping would cause the curl to be non-zero and the surface to be non-integrable.

![Fig. 5. Target maps for reflector facets with three different azimuth angles $\phi$ ($0^\circ$, $22.5^\circ$ and $45^\circ$) and a uniform square target. In all cases the elevation $\theta$ is equal to $60^\circ$ and the collection angle subtended by each facet is $20^\circ$ by $20^\circ$. 225 ellipsoids were used to generate the maps.](image)

3. Generating smooth reflectors from maps

Since the target distribution produced by reflectors generated with the Oliker algorithm is discrete, we must find a way to modify the reflector shape in order to obtain the original continuous target distribution. Ideally, one would simply use a very large number of ellipsoids and then interpolate the resulting reflector surface into a smooth continuous surface. However, the algorithm complexity scales quadratically with the number of ellipsoids, and convergence becomes slow for large numbers of ellipsoids, so it becomes increasingly impractical to use more than a couple of hundreds target points with current computational capabilities. On the other hand, when the number of ellipsoids is small and convergence is fast, there is no easy way to convert the reflector surface made of ellipsoid facets into a smooth surface that will produce the desired continuous target distribution. Figure 6 shows illuminance distributions produced by directly interpolating the initial reflector made of ellipsoid facets with a NURBS surface. We used the same set of surface points that we use to generate maps (one point per facet located at the facet centroid). Direct surface fitting typically produces artifacts in the target distribution such as peaks at the edges, even when the total number of ellipsoids increases.
Fig. 6. Illuminance distributions produced by smooth reflectors obtained by direct interpolation of the initial reflectors made of ellipsoid facets, for the geometry shown in Fig. 4a. We use the same set of points that we used to generate maps. Artifacts typically remain at the edges, even for relatively large number of target points.

As we shall see, generating reflectors from maps circumvent some of the problems observed with direct surface fitting. The full reflector generation process is summarized in Fig. 7. Reflector shapes can be created from maps by using path integration along the map contours from an arbitrary starting point. We should keep in mind however that generated maps contain inaccuracies, so the generated surface is not completely independent of the integration path (this would only be true if the integrability condition was strictly satisfied). Retrieved maps do not rigorously fulfill the integrability condition since they are derived from a finite set of data points prone to error. The main sources of errors are related to the accuracy of the Oliker algorithm (based on Monte-Carlo ray tracing in our implementation) and the assumption that each target point is associated with the center of each corresponding ellipsoid. Nonetheless, retrieved maps are a good approximation of the ideal theoretical maps and can be used as a starting point before further refinements.

Fig. 7. Overview of the reflector generation process. The original desired target distribution is discretized into a small number of target points and a corresponding reflector shape is generated using the Oliker algorithm. A continuous target map is then created from the finite set of target points and their corresponding ellipsoid facets. Finally, a smooth reflector shape is generated by numerical integration along the contour lines of the map.

If needed, maps can be optimized to minimize residual curl and adjust the flux in each target bin according to its theoretical value. Since residual errors accumulate during integration, it is often a good choice to start integration at the center of the reflector. For instance, the reflector in Fig. 8a was integrated starting at its apex ($\theta = 0^\circ$). From the starting point, reflector profiles are then generated for each $\phi$ by integrating the surface normals along $\theta$. Surface normals can be determined at each point using the law of reflection so that each source ray reaches its corresponding target location, in accordance with the mapping. The set of surface points created this way is finally interpolated with a NURBS surface and exported to illumination design software for analysis.
An interesting advantage of this method is that smooth reflectors can be generated from a small number of target points, hence very quickly. Figure 8 shows three examples of reflector surfaces generated from maps. The two reflectors on the left were created in about 10 seconds with a 2.5GHz Intel Quad core computer from initial reflectors made of 400 ellipsoids. The quadrant symmetry of the system was leveraged to speed up computations. Reflector 8a on the left has 1.3% RMS non-uniformities over the target area and 7% peak-to-valley. Similarly, reflector 7b for the rectangular target case produces 1.9% RMS non-uniformity and 8% peak-to-valley. In both cases the statistical noise inherent to Monte-Carlo ray tracing is about 0.7% [13]. Such uniformity levels are appropriate for most lighting tasks, but they may not be sufficient for demanding applications such as microlithography. In these cases lens arrays or mixing rods provide robust alternatives that are less sensitive to system tolerances. When using maps, higher uniformity can typically be obtained at the expense of computation time. Figure 9 shows that increasing the number of ellipsoids of the initial reflector improves target uniformity but increases the time it takes to generate the reflector. In this case a performance plateau is reached at about 600 ellipsoids where increasing further the number of ellipsoids brings no significant uniformity improvement.

The reflector facet in Fig. 8c proved to be more challenging to generate because of its singularities at both the target corners and the facet corners. The bulges at the edge of the square distribution correspond to the facet corners, which would require greater accuracy and
hence greater computation time to be rendered accurately. The facet surface in this case was generated from an initial reflector made of 225 ellipsoids in about 100 seconds. Nonetheless, the cross-section of the illuminance distribution shows that good uniformity can still be obtained in this case (7% RMS non-uniformities). Additionally, the relative invariance of target maps when system parameters are changed can be greatly leveraged when generating reflectors from maps. Reflectors with modified system parameters can be generated almost instantaneously, without having to compute new maps.

Figure 10 shows the sag difference between the shape of the initial reflector made of ellipsoid facets and the smooth reflector surface generated from a map. The system geometry is similar to the example in Fig. 8a except that the initial reflector is made of only 149 ellipsoids. The peaks in the plot correspond to the dips in the original reflector surface where multiple ellipsoid facets meet.

4. Concluding remarks

We propose in this paper a method to generate source-target mappings using the method of supporting ellipsoids developed by Oliker. As opposed to other existing methods using equi-
flux grids, this technique can be used for any system geometry, with no restriction on the desired target distribution, the collection angle, or the source intensity distribution. The only assumption made is the point source approximation. While many theoretical mappings can achieve the required flux distribution on the target, the mappings derived with this method also satisfy the integrability condition and can therefore be achieved with smooth reflector surfaces. Generating reflectors from maps presents two major advantages. First, since maps corresponding to smooth target distributions can be computed with sufficient accuracy from only a couple of hundred target points, reflectors can be generated quickly, typically in less than 15 seconds for systems with quadrant symmetry, and less than 2 minutes for systems with no symmetry on a 2.5GHz Intel Quad core computer. Second, because maps vary slowly when system parameters are changed, it is possible to generate instantaneously new reflectors for modified system geometries, without having to compute new maps. Finally, this design method is also applicable to the design of freeform refractive surfaces. In this case, the Oliker algorithm can use Cartesian ovals as a basis function instead of ellipsoids, and limitations imposed by total internal reflection must be taken into account.

Appendix

In order to ensure a smooth continuous surface, one can require that integration of the surface normals over any closed loop to be equal to 0. If there is a mismatch between the starting and ending points, then a step discontinuity exists and no continuous surface can fit the given set of surface normals. This constraint can be written as

$$\oint_C \mathbf{N} \cdot d\mathbf{l} = 0$$  \hspace{1cm} (4)

where $C$ is an arbitrary closed integration path within the domain over which the surface is defined (the collection angle of the reflector), $\mathbf{N}$ is the field of desired surface normals and $d\mathbf{l}$ is a differential displacement vector. Using Stokes' theorem, Eq. (4) can be rewritten as

$$\oint_C \mathbf{N} \cdot d\mathbf{l} = \iint_S (\nabla \times \mathbf{N}) \cdot \mathbf{N} \, ds = 0$$  \hspace{1cm} (5)

where $S$ is a surface capping contour $C$, and $ds$ is a differential surface area. Since this constraint applies for any arbitrary integration path, we must satisfy at any point on the surface

$$\mathbf{N} \cdot (\nabla \times \mathbf{N}) = 0$$  \hspace{1cm} (6)

Acknowledgments

The authors acknowledge Optical Research Associates for the educational license of LightTools. This work was funded under a fellowship from Optical Research Associates and the NYSTAR Foundation.