Design of freeforms to uniformly illuminate polygonal targets from extended sources via edge ray mapping

Birch, Daniel A.; Brand, Matthew

TR2020-183  March 20, 2021

Abstract
We consider the design of a compact freeform optical surface that uniformly irradiates an arbitrary convex polygonal region from an extended light source while controlling spill. This problem has attracted a large literature, mainly treating highly symmetric special cases or cases where the solution is approximated by a zero-étendue design based on a point source. Practical versions of this illumination design problem will likely feature large asymmetric LEDs, compact lenses, and irregular targets on angled projection surfaces. For these settings, we develop a solution method based on an edge ray mapping that routes maximally off-axis rays from the edges of the source through the edge of the optic to the edges of the target polygon. This determines the sag and normals along the boundary of the freeform surface. A 'spill-free' surface is then interpolated from the boundary information, and optimized to uniformize the irradiance while preserving the polygonal boundary. Highly uniform irradiances (relative standard deviation < .01) can be attained with good control of spill, even when the exit surface is < 3 source diameters from the embedded source.

Applied Optics 2020
Design of freeforms to uniformly illuminate polygonal targets from extended sources via edge ray mapping

DANIEL A. BIRCH AND MATT BRAND

Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA 02139, USA
*brand@merl.com

Abstract: We consider the design of a compact freeform optical surface that uniformly irradiates an arbitrary convex polygonal region from an extended light source while controlling spill. This problem has attracted a large literature, mainly treating highly symmetric special cases or cases where the solution is approximated by a zero-étendue design based on a point source. Practical versions of this illumination design problem will likely feature large asymmetric LEDs, compact lenses, and irregular targets on angled projection surfaces. For these settings, we develop a solution method based on an edge ray mapping that routes maximally off-axis rays from the edges of the source through the edge of the optic to the edges of the target polygon. This determines the sag and normals along the boundary of the freeform surface. A "spill-free" surface is then interpolated from the boundary information, and optimized to uniformize the irradiance while preserving the polygonal boundary. Highly uniform irradiances (relative standard deviation < .01) can be attained with good control of spill, even when the exit surface is < 3 source diameters from the embedded source.

© 2020 Optical Society of America

1. Introduction

Illumination trends favor large LEDs for brightness and small lenses for compactness and cost. This has stimulated much interest in the problem of tailoring freeform optics to uniformly irradiate polygonal targets from extended sources. Early approaches ignored the geometry of the light source and numerically solved the Monge-Ampère problem, or equivalent formulation, for a point light source located on or near the true source [1, 2] (also see [3] for a review). This can be viable for low-étendue problems where the source is very small relative to the optic, but as is well known (and demonstrated in Figure 6, a significantly extended source will substantially degrade both the uniformity and perimeter of the irradiance. This can be somewhat ameliorated in a feedback loop where intensities in a point-source target distribution are repeatedly rescaled to compensate for extended-source irradiance defects as revealed by Monte Carlo simulation, e.g., [4–6]. However, significant nonuniformities and spills persist in the published results, which are also hard to assess because Fresnel losses are ignored.

For smaller optics or larger sources, it is necessary to incorporate the geometry and radiance of the extended source directly into the optimization problem. Some specialized methods have been proposed for highly symmetric problems, e.g., for circular sources, freeforms, and targets, tailoring can be treated as an essentially 1D problem with a correction for skew rays [7, 8]. Relatively few design principles have been proposed for low- or no-symmetry problems. A recent wavefront tailoring approach [9] raises the possibility of generating two-surface freeforms directly from constraints on wavefronts from the source corners, but the only worked example maps a square Lambertian emitter to a square target.

The most general approach is to directly model the full flux field by tabulating light arriving at each target point from each direction. This enormously expensive calculation can be approximated by backward ray tracing [6, 10] or by integrating flux through a thin cone [11] or tube [12] that
connects a region of the lens to a region of the target plane. These calculations, which can be faster and more informative than forward Monte Carlo, set the stage for optimization. For example, backward ray tracing with local surface perturbations was used to make finite-difference estimates of the irradiance error gradient error [10], and used to design lenses that produce blurry but recognizable irradiance patterns, with blur depending on the size of the source. More globally, an algebraic relation between surface curvature and irradiant intensity was characterized for general light fields in [12] and used to demonstrate that it is possible to obtain nontrivial sharp-edged irradiance patterns in very high-étendue settings.

It should be noted that for extended sources, the irradiance tailoring problem is often infeasible and nonconvex, meaning that production of a target irradiance is physically impossible and iterative improvement will only converge to a local optimum of a merit function. Consequently, finding a good approximate solution depends on starting with a favorable initial surface.

Prior approaches for obtaining initial surfaces depend on obtaining a ray mapping (point to point correspondence) from the entire optical surface to the projection plane and then solving for an optical surface that (approximately) yields the desired mapping. The ray mapping may be an arbitrary map (e.g., [13–17]) or obtained via optimal transport algorithms (e.g., [1, 2, 18, 19]).

We find that an edge ray mapping — a smooth 1-to-1 mapping from points on the lens boundary to points on the irradiance boundary — is sufficient to determine the entire solution. By incorporating information about the geometry of the extended light source, the edge ray mapping determines a spill-free freeform boundary; the rest of the initial surface can be obtained via interpolation and then irradiance-uniformized by curvature adjustments. We will present the method in three steps:

1. Construct an edge ray mapping that determines sag and normals along the optical surface boundary to illuminate the polygon perimeter (§2.1).
2. Interpolate the boundary to produce an optical surface that smoothly illuminates the polygon interior via polyharmonic splines, extended to differential constraints (§2.2).
3. Uniformize the irradiance via light field tailoring [12], augmented to maintain the irradiance perimeter (§2.3).

The method is flowcharted in Figure 7. As demonstrated in §3, this method produces highly uniform and — when physically possible — sharp-edged illumination patterns, even when the source subtends a large angle from the point of view of the exit surface. No symmetries are assumed; Fresnel losses are correctly modelled; and spill can be strictly controlled. All these points are departures from the literature.

2. **Method**

For simplicity of exposition, we consider an optical system comprised of a square Lambertian LED embedded in a circular lens whose exit surface will be tailored to illuminate a polygonal region on a flat projection surface orthogonal to the optical axis. None of these assumptions are essential; the light source can have any geometry and radiance pattern, the lens can be polygonal, there can be an entrance surface to increase collection angle, the projection surface can be angled, *et cetera*.

2.1. **Determining the optical surface boundary**

First we determine sag and surface normals along the optical surface boundary, such that all rays from the extended source will refract through the optical surface boundary to illuminate the target boundary without spill, as depicted in Figure 1. This is feasible over most but not all of the target boundary, *e.g.*, acute corners of a triangular target cannot be illuminated from a square
source without spill, because the irradiance is ultimately a superposition of images of the source. It is, however, possible to guarantee a weaker condition: That the cone of light through any point on the optical surface boundary is touching and otherwise interior to at least one edge of the target boundary. In practice the interior guarantee can usually be upgraded to all edges.

The initial optical surface boundary is a simple ring of constant sag. Boundary design begins with a provisional mapping in which each point \( b \) on the optical surface boundary is assigned to a point \( t \) on the target irradiance boundary. Both boundaries are arc-length parameterized and put into correspondence by proportional arc-length. As can be seen in the left panel of Figure 1, this concentrates energy toward the corners of the target. This provides a favorable initial condition for uniformization (§2.3) which will redistribute flux along the boundary and interior.\(^1\)

To incorporate light source geometry, each optic boundary point \( b \) is then assigned a source point \( s \) on the light source boundary, and \( s \) and \( t \) are adjusted to make \( b \) “spill-free,” meaning that when the optic refracts a ray from \( s \) thru \( b \) to \( t \) on an edge of the irradiance target, all other extended-source rays refracted thru \( b \) will land on the irradiance-interior side of that edge. This can be accomplished in two steps: First, we pick a provisional source point \( s \) that lies farthest from \( b \) on a circle circumscribing the light source, and calculate a provisional surface normal \( n \) at \( b \) that will bend a ray from \( s \) through \( b \) toward \( t \) according to the vector refraction law

\[
\mathbf{n} \propto \frac{\mathbf{b} - \mathbf{s}}{||\mathbf{b} - \mathbf{s}||} - \frac{\mathbf{t} - \mathbf{b}}{||\mathbf{t} - \mathbf{b}||},
\]

with \( n \) being the ratio of refractive indices before and after \( b \). Second, rays from vertices of the light source boundary are refracted through \( b \) and projected to the target plane, giving the convex hull of the projected source. To prevent spill, these vertex projections need to be shifted so that their hull touches the irradiance target edge without crossing it. To do so, the projected vertex that is least interior to the irradiance polygon with respect to the target edge is orthogonally projected onto that edge, yielding an updated target point \( t \). The corresponding source vertex becomes the updated source point \( s \), and the normal \( n \) is recalculated. This shifts the projection of the source to the target edge but not along it.

Finally, all boundary sags and normals are jointly updated to be mutually consistent with each other and with their associated source and target points. This is accomplished by alternately solving for each normal \( n \) given its corresponding boundary point \( b \) as in Eq. (1), and then for

\(^1\)We find this optimization is accelerated by further concentrating target points toward the polygon corners, e.g., if a target edge is arc-length parameterized \(-1 \leq x \leq 1\), target points can be moved according to a sigmoid map such as \( x \to \alpha(3x - x^3)/2 + (1 - \alpha)x \) with best results around \( \alpha = 1/2 \).
the sag components of all boundary points $\mathbf{b}$ given all normals $\mathbf{n}$, using the fact that boundary normals must be orthogonal to the boundary’s azimuthal derivative:

$$\mathbf{n}^T \frac{\mathbf{db}}{d\theta} = 0,$$

(2)

to yield a linear system of equations for the sag component of discrete control points $\mathbf{b}_i$ evenly spaced around the optical surface boundary. In the examples below ($\S$3), alternating solutions of Eq. (1) and Eq. (2) typically converged in 2–4 iterations.

2.2. Determining the optical surface interior

We then use polyharmonic splines [20] to determine an optical surface from its boundary sag curve and normals. A polyharmonic spline $f$ of order $m$ minimizes the deformation energy

$$U_m[f] = \int_{S \subset \mathbb{R}^d} |\nabla^2 f|^2 \, dx,$$

(3)
on region $S$ while interpolating a set of control points $\{(x_i, f(x_i))\}_i$. The $m=1$ case selects a minimum-variation surface and $m=2$ selects a minimum bending energy surface. Conveniently, Eq. (3) is minimized with exact interpolation of control points by an affine combination of radial basis functions $\psi(||\mathbf{x} - \mathbf{x}_i||)$:

$$f(\mathbf{x}) = \sum_i \psi(||\mathbf{x} - \mathbf{x}_i||)w_i + \mathbf{v}^T \mathbf{x} + c$$

with $\psi(r) = r^m \log(r)^{m+1 \text{ mod } 2}$

(4)

where the parameters $w_i$, $\mathbf{v}$, and $c$ are obtained by solving a system of linear equations, each an instance of Eq. (4) with $\mathbf{x}$ set to one control point $\mathbf{x}_j$. We note that one can also control derivatives $f_x, f_y, f_{xx}, f_{yy}, \ldots$ at additional points by differentiating some of these equalities before solving the linear system. Splines of order $m$ accommodate derivative constraints up to order $m−1$ and are optimally smooth with respect to the $m$th Laplacian off the control points.

To obtain an initial optical surface with the desired edge ray mapping, we seek to interpolate the sag values $f$ and surface gradients $\nabla f = (f_x, f_y)$ at the lens boundary as determined in $\S$2.1, therefore we typically use the biharmonic ($m=2$) spline. Figure 2(b,c) shows a spline surface that smoothly but nonuniformly illuminates a square, yielding the irradiance pattern in Figure 3(a).

It is also desirable (but not strictly necessary) that the interpolated surface does not produce any caustics (ray crossings) which might slow subsequent tailoring of the interior irradiance. The biharmonic solution is usually noncaustic, and a caustic-free surface can always be obtained as a linear combination of harmonic ($m=1$) and biharmonic splines. It is also possible to directly spline a “mostly uniform irradiance” freeform using triharmonic ($m=3$) interpolation with additional constraints that force some points on the interior of the surface to mimic the Laplacian curvatures $\nabla^2 f = f_{xx} + f_{yy}$ of a similar sized asphere that uniformly irradiates a circular spot from a Lambertian point source (explicit sag function given in [21]).
2.3. Uniformizing the irradiance

Light field tailoring [12] is then adapted to uniformize the irradiance provided by the splined surface. The general method is motivated by the observation that, locally, the curvature of the optical surface changes the curvature of the wavefront, thereby controlling beam dilations that dilute the intensity of the irradiance. Therefore a field of irradiance errors can be related to a field of curvature corrections \( g(x, y) \), which is then applied to the optical surface by solving a Poisson problem \( \nabla^2 s = g \) for a field of sag adjustments \( s(x, y) \). For zero-étendue (point) sources, the relationship is pointwise and direct. For positive-étendue sources (extended light sources and more generally light fields), the relationship is indirect but can be determined by solving a sparse linear system which reveals how credit for irradiance errors should be distributed over the curvature correction field. Both cases allow for parallel implementations and can tailor lenses and mirrors to produce uniform irradiance as well as photograph-like irradiance patterns. However, the irradiance only approximates the target, because most extended-source irradiance tailoring problems do not admit exact solutions. Of the many possible approximate solutions, light field tailoring prioritizes fidelity to the interior the target irradiance rather than the boundary, simply because the interior accounts for most of the flux. This turns out to be a liability for the polygonal irradiance problems contemplated here, where the quality of the boundary weighs more heavily in the desirability of the result. Consequently the algorithm can be improved for this particular application.

Since the desired irradiance boundary is already provided by the boundary slopes of the polyharmonic spline, we modify the sag correction step to preserve this aspect of the optical surface geometry. To do so, we note that the solution of the Poisson problem \( \nabla^2 s = g \) is not unique; one may add any saddle surface of the form

\[
\bar{s}(x, y) = a_1(x^2 - y^2) + a_2xy + a_3x + a_4y
\]

to \( s \), because \( \nabla^2 (s + \bar{s}) = \nabla^2 s + \nabla^2 \bar{s} = \nabla^2 s + 0 = g \). On each iteration of curvature correction, we solve for coefficients \( a_1, a_2, a_3, a_4 \) of a saddle that cancels out unwanted boundary slope modifications, e.g., those that shift edge rays off the target perimeter. This is a small but overconstrained system of linear equations and so is solved in a minimum squared-error sense. Adding the saddle to the sag corrections effectively chooses, from a space of equivalent irradiance improvements, the one that least degrades the irradiance boundary. Together with the initial surface from §2.2, this consistently results in faster convergence to better results than obtained from the generic light field tailoring algorithm.

3. Examples

In this section we review several examples as a way of illustrating the trade-offs inherent in extended-source irradiance tailoring. Generally, the designer faces a choice between a target with "hard" edges, which will force ringing artifacts due to the bandlimited source, or "soft" edges, which ameliorate the ringing, but make more noticeable the asymmetric fall-offs that occur when some target edges are better aligned with the edges of the light source than others. The trade-off becomes less severe with larger lens apertures, e.g., compare Figure 3(b-c) below.

All examples are circular lenses with embedded sources, meaning there is a single refraction at the lens-air interface, where the index of refraction ratio is \( n = 1.5 \). The sources are 1 mm square Lambertian LEDs. The distance from the source to the projection plane is 1200 mm in all cases. The source-vertex distance is denoted \( v \); since the source is embedded this is also the lens thickness. \( S_{\text{in}} \) and \( S_{\text{out}} \) are the side lengths of the uniform interior and of the entire illumination pattern, respectively. The optimization goal is a uniform interior, a fall-off "skirt" of \( < 1/2 \) the LED spot size (except Figure 3(d)), and negligible spill outside the skirt. Compute time is typically a few seconds for construction of the edge ray mapping and splined surface, and a few minutes for uniformization using a \( 128 \times 128 \) grid of sag values to represent the surface.
Fig. 3. Uniformization with respect to a $1 \text{ mm}^2$ extended source (§2.3): (a) Initial irradiance due to a polyharmonic spline lens surface and (b) after uniformization. (c) Sharper-edged irradiance attained with a larger lens. (d) Sharp-edged irradiance due a very small lens where the target is an integer multiple of the LED spot size. (e) Irradiance due to a lens designed for an projection plane tilted $15.0^{\circ}$ from orthogonal.

Fig. 4. Irradiances obtained from targets with soft and hard edges.

As in [13, 15, 16, 18], uniformity is scored as $u = 1 - \sigma / \mu$ — one minus the coefficient of variation (also known as relative standard deviation) of the irradiance in the polygonal target region (where irradiance should be constant). Here $\sigma$ and $\mu$ are the standard deviation and mean of the irradiance.

**Squares:** Figure 3 shows several irradiances obtained when the alignment of the LED edges and the target edges makes the trade-off between uniformity and sharpness more favorable. Very high uniformities are achieved with no spill outside the anticipated fall-off skirt. Figure 3(d), a special case where a very small lens provides an "unskirted" irradiance, is discussed at the end of this section. Figure 3(e) demonstrates an off-axis design which also achieves high uniformity, however fall-off varies along the boundary because the spot size image of the LED varies significantly.

**Other regular polygons:** Figure 4 shows hexagonal, pentagonal, and triangular irradiances obtained with different constraints on the edge sharpness. The hexagonal and triangular targets were designed with sharp edges, resulting in irradiances (a,c) somewhat lower uniformity than obtained (b) for a soft-edged target such as the pentagon; ringing is apparent in the intensity transects (d,f). Note that because none of the edges are aligned with the LED edges, the attained fall-off varies between edges of different orientations and at the vertices.
Irregular polygon: Figure 5 shows the target and obtained irradiance for an irregular polygon. The target was translated to place its centroid on-axis before constructing the edge ray mapping. For this example, we found that uniformization converged significantly faster if the splined surface was first tailored for a point light source.

Very small lenses: Figure 6 demonstrates the difference between point-source tailoring and the method presented herein. We optimized a small (Ø = 4.0 mm, v = 3.88 mm) pentagon-producing lens to produce a \( u > 0.999 \) uniform pentagonal irradiance from a point source, then ray-traced from a 1.00 mm square extended source to find that the interior uniformity of the irradiance is substantially degraded (arched blue line in Figure 6(b)). We then uniformized the irradiance as in all previous examples, restoring uniformity to \( u = 0.995 \). Note that the vertex of this lens is less than 3 source diameters from the source.

There is one notable exception to uniformity-versus-sharpness trade-off: In some special cases the target irradiance can be approximated by tiling multiple focused images of the light source. Figure 3(d) shows such an irradiance, where \( 2 \times 2 \) LED spots are partially focused to form a sharp irradiance boundary, and partially defocused to uniformize the interior irradiance. The
lens is not completely focusing because the tiled images are radially distorted due to off-axis projection and therefore cannot pack together neatly to cover the target (this also prevents perfect uniformization). Nonetheless, a union of focusing surfaces provides a good initial surface, and this can be accomplished by constraining the initial splined surface to reproduce local curvatures taken from four tilted Cartesian ovals. Despite its small size – the vertex is less than 4 source diameters from the source – the uniformized lens provides one of the sharpest irradiance fall-offs in our experiments.

4. Conclusion

We have shown that an edge ray mapping suffices to determine a freeform lens that uniformly illuminates a target polygon from an extended light source. The key tools are surface interpolation (§2.2) and extended-source irradiance tailoring methods (§2.3) that preserve the edge ray mapping (§2.1). We demonstrated several results with very high uniformity, including challenging problems such as targets with no symmetry, irradiance patterns with sharp edges, and lenses not much bigger than their encapsulated LEDs. For every example above, the edge ray mapping and saddle correction provided faster convergence to a visibly better result than the generic light field tailoring algorithm.

One limitation of this approach is that the construction of the lens perimeter from the edge ray mapping is based on the assumption that rays from the far side of the source can be always refracted to the target. This limits the collection angle, but also eliminates the possibility of TIR bounces within the lens, which would contaminate the irradiance pattern.

The next step is to incorporate an entry surface into calculations so that collection angles can be increased without risking TIR. Since light-field tailoring can accommodate arbitrary radiance functions, this may reduce to a matter of characterizing the radiance through a plane after the entry surface. Future research will consider mixed TIR/refractive optics and nonconvex target polygons (whose feasibility is uncharacterized).

A. Flowchart

Fig. 7. Flowchart of the tailoring process. The left hand column outlines construction of the edge ray mapping; the right hand, interpolation and optimization of the surface.
Disclosures

DAB and MB: Mitsubishi Electric Research Laboratories (E, P).

Acknowledgements

We thank the reviewers for many helpful comments and suggestions.

References