Caustic Art

EPFL Technical Report, June 2012

Thomas Kiser
EPFL

Mark Pauly
EPFL

Abstract

The computation of caustics created by reflection or refraction of light is a well-studied problem in computer graphics. This paper investigates the inverse problem: Given a greyscale intensity image, find the shape of a surface that will cast a caustic that reproduces this image. We formulate this goal as an optimization problem on a 2D mesh representation on a specular surface. A normal field is obtained from this deformed mesh and subsequently integrated to a height field surface. We show with several examples that this height field surface reproduces the desired caustic images to very high accuracy, enabling new applications in sculptural art, product design, or architecture.

1 Introduction

Caustics are patterns of light created by reflection or refraction on curved surfaces (see Figure 2). Computing caustics in a given three-dimensional scene has been the subject of extensive research in computer graphics as it constitutes one of the main difficulties of photorealistic rendering algorithms. This paper investigates the inverse problem: Given an image, compute the shape of a reflective or refractive object that, when lit by a given light source, projects the image as a caustic onto a known diffuse surface (see Figure 1).

We introduce a solution, called *brightness warping*, where we consider a fixed, contiguous caustic represented as a quadrilateral mesh. This representation is computed from a known specular surface, most simply, a rectangle. The reflective or refractive surface is then divided into quadrangular patches which correspond to the faces in the caustic. By optimizing the area of these patches, the brightness of the corresponding caustic quads can be adjusted to a given distribution. In a second step, a normal field is computed from this mapping and subsequently converted to a height field. Since this conversion is not possible for arbitrary normal fields, a consistency term is introduced that accounts for the integrability of the normal field.

This procedure implies several restrictions for the algorithm. Most notably, it is limited to fixed, contiguous caustics represented by a mesh and requires a given brightness distribution for its faces. It furthermore ignores the actual shape of the reflector or refractor and operates only on surface normals. These simplifications result in a differentiable objective function, enabling the use of gradient-based optimization methods. In addition, they both remedy the problem of sub-optimal local minima and bypass costly caustic evaluations. Despite the simplifications, the approach is powerful enough to faithfully reproduce images.

Applications of this work are mainly of an artistic nature. Caustics are interesting from an artistic point of view, because the shape of the corresponding specular object does not directly reveal the caustic it will produce. Surfaces generated with the approach proposed in this paper could be used in architecture or interior design, for example as special windows that create interesting caustics when lit by sunlight.

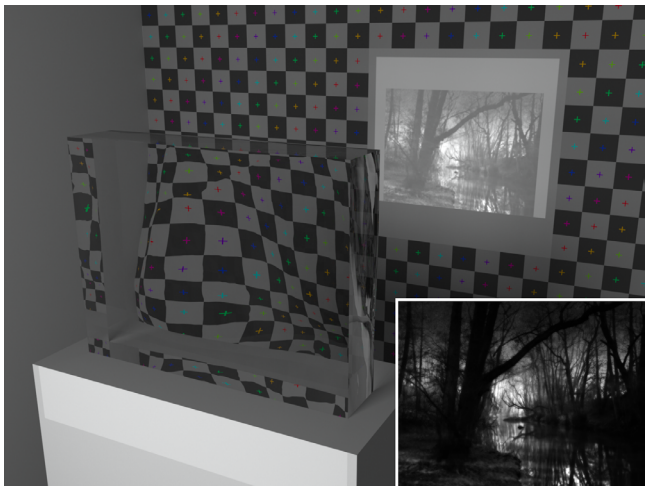


Figure 1: *Caustic art. Light shining onto the glass block creates a caustic on the wall that matches the desired input image (inset). The geometry of the glass block has been computed with our optimization algorithm.*

1.1 Related Work

The task of reproducing a prespecified light distribution by a specular surface also arises in the field of inverse reflector design, which concentrates on reflectors for lamps. A survey on inverse reflector design is given by Patow and Pueyo [2005]. There are two types of such light distributions, near-field and far-field distributions. Near-field distributions specify an irradiance distribution on a given surface (typically a plane) that is to be reproduced, which is also the goal in this paper. Far-field distributions can be considered as limit cases where the surface to be illuminated is infinitely far away from the reflector, so only the distribution of the ray directions matters.

Methods for inverse reflector design typically employ an analysis-by-synthesis approach: Some surface representation is chosen to parametrize the reflector, such as NURBS [Anson et al. 2008]. Then, the light distribution caused by a surface is evaluated and rated against the desired one, which is used to iteratively optimize the surface parameters. Various optimization strategies have been applied, ranging from frameworks that allow analytical differentiation, thereby enabling the use of the conjugate gradient method [Neubauer 1997], and methods that compute derivatives approximately [Finckh et al. 2010] to ones that employ no derivatives at all [Anson et al. 2008]. Examples using evolutionary optimization [Doyle et al. 1999] also belong to the latter category.

The simplifications imposed on the scene vary; assumptions of perfect specularity and only one bounce of light without interreflections or occlusions (as used in this paper) are common, though there are exceptions to both [Patow et al. 2007; Mas et al. 2009]. The restriction to rotationally symmetric reflectors is also commonly used, particularly in theoretical works [Westcott and Norris 1975].

Figure 2: *Reflective and refractive caustics caused by a plastic bottle filled with water, clearly distinguishable by their color.*

These works mostly focus on reflective surfaces, though many readily extend to refraction. One noteworthy example investigating the refractive problem is the work by Finckh et al. [2010]. They use GPU computations to speed up the caustic evaluation, and a stochastic approximation algorithm for the optimization, which is able to find a global optimum.

Concerning refractive objects, the field of lens design is also noteworthy, though the goals of these problems are different, e.g. aberration correction. These problems are often restricted to a small number of parameters such as radii of the underlying primitive shapes [Patow and Pueyo 2005]. Again, there are exceptions, e.g. the work by Loos et al. [1998], who use a NURBS-based representation to optimize progressive lenses.

Weyrich et al. [2009] have chosen a different approach to reproduce a prespecified far-field distribution: First, they generate a set of sloped, planar microfacets to realize the desired distribution of ray directions. Then, they arrange the microfacets in a regular array using simulated annealing to minimize the resulting discontinuities. Closely related to Weyrich et al.'s work is the system for near-fields proposed by Papas et al. [2011]. They extend the notion of microfacets to curved micropatches, which are used to produce specks of light with an anisotropic Gaussian distribution. Their paper focuses explicitly "on the more complex effect of refraction, while the reflective case follows by analogy". To demonstrate their results, they also physically manufactured refractive surfaces from acrylic glass.

There is an overlap between the brightness warping method proposed here and their work, in fact, the name was borrowed from a section titled "Irradiance Warp" [Papas et al. 2011, Section 5.1]. To compute the shape of the micropatches that produce a Gaussian irradiance distribution, Papas et al. "define a bijective mapping between points in the micropatch domain and points on the projection plane", "analytically compute the surface normals that refract/reflect the light in this way", and finally "integrate this normal field to arrive at the required micropatch surface". The approach proposed here works in the same way, with two main differences. First, the bijective mapping is inverted, i.e. the points on the projection plane are fixed, while the ones on the specular surface are moved. Second, due to the simple shape of the desired caustics, Papas et al. do not need to explicitly account for the integrability of their normal fields, which is a key ingredient in our method to reproduce arbitrary images.

The rest of the paper is organized as follows: Section 2 first describes the assumptions and simplifications that our method is based on and explains the basic functionalities of the implementation. Section 3 then introduces the main algorithm before we present results in Section 4. We conclude with a discussion and some ideas for future work.

2 Framework

Throughout this paper, caustics are generated by a flat height field that is discretized as a regular grid. To further simplify the computations, the height field's actual displacement is disregarded, and a simple plane is used for raytracing. Only the reflected/refracted ray directions are computed from the height field's normals. In essence, the operations are based on a normal-mapped plane, although the physical reproduceability of the normal field must still be ensured (see Section 2.3).

The diffuse surface the caustic is projected on (hereafter referred to as the *receiver*) is also assumed to be a plane. The light is assumed to be emitted by a distant light source, so the rays are treated as parallel. The reflections and refractions are considered to be perfectly specular; an alternative is explained in Section 3.3. We assume that no shadowing, no interreflection and, for refraction, no total internal reflection occurs. Spectral effects like dispersion are also neglected. The caustic is furthermore assumed to be a contiguous pattern, which means that the normal field will be continuous and the corresponding height field continuously differentiable. This obviously limits the caustics that can possibly be generated, but makes the design of the algorithm considerably simpler.

Our method works for reflective as well as refractive setups. In the reflective case, there is only one height field. In the refractive case, two surfaces are involved, one where the light enters the medium, and one where it exits. The first of the two is simply assumed to be planar (keeping the incident light parallel), and the algorithm operates only on the second surface.

2.1 Forward Caustics

Our approach for generating caustics from a height field is related to photon mapping or beam tracing that starts from the light source. As mentioned above, we assume that and no interreflection and shadowing occurs, and the specular surfaces are assumed to be smooth, so the caustic will be contiguous. This allows using a two-dimensional triangle mesh that basically defines the cross section of the light beam (see Figure 3). We can then trace a ray for each vertex individually. In analogy to the photon mapping method, the projection of these vertices on the receiver will hereafter be referred to as *photon positions*, and the mesh they form as *photon mesh*.

In order to reflect or refract a ray, we need to obtain the normals from the height field. Figure 4 shows that the normal of a height

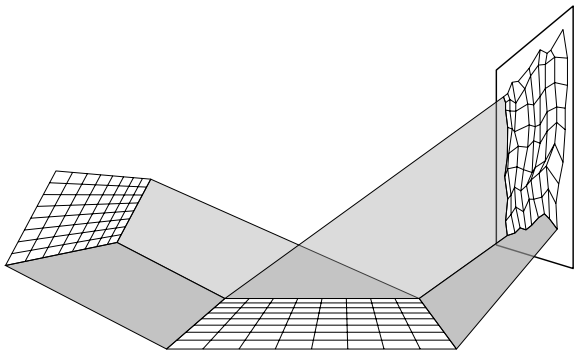


Figure 3: *Generating a caustic. On the left, a light beam is discretized using a mesh that matches the normal field in the middle. The field deforms the mesh as it is reflected and projected onto the receiver on the right.*

field $z(x, y)$ is collinear with the vector

$$\left(-\frac{\partial z}{\partial x}, -\frac{\partial z}{\partial y}, 1\right)^{\mathrm{T}}.$$

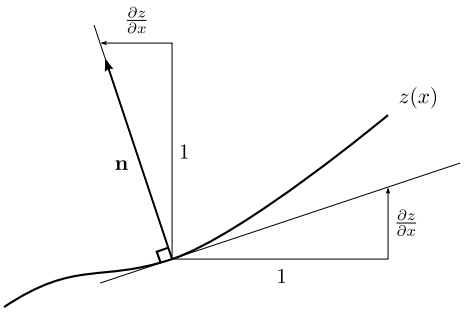


Figure 4: *Normal of a curve.*

As our height field is discretized as a regular grid, we can simply compute the normal at each node of the height field by using finite differences, and then use bilinear interpolation (along with a re-normalization) to obtain normals at arbitrary positions on the surface. Using the standard raytracing operations, we can now compute the photon mesh cast onto the receiver. The shape of the photon mesh defines the resulting brightness distribution in the caustic pattern.

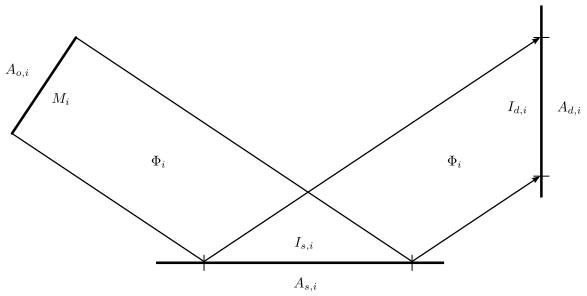


Figure 5: *Reflection of a light beam.*

Assuming perfect specularity, the radiant flux Φ_i (the total emitted or incident power) remains constant throughout each beam defined by a triangle i (see Figure 5). The flux is equal to the radiant exitance M_i (the emitted power per unit area) times the area in the original mesh $A_{o,i}$,

$$\Phi_i = M_i A_{o,i}; \tag{1}$$

it is equal to the irradiance $I_{s,i}$ (the incident power per unit area) times the area $A_{s,i}$ on the specular surface,

$$\Phi_i = I_{s,i} A_{s,i}; \tag{2}$$

and equal to the irradiance $I_{d,i}$ times the area $A_{d,i}$ on the receiver,

$$\Phi_i = I_{d,i} A_{d,i}. \tag{3}$$

Assuming a parallel light source and equally-sized triangles, M_i and $I_{s,i}$ are the same for each triangle. Consequently, the resulting irradiance $I_{d,i}$ on the receiver contributed by this triangle is proportional to the original area divided by the resulting area in the photon mesh. Basically, this is the resulting caustic brightness when a Lambertian (i.e. perfectly diffuse) reflectance model is assumed for the receiver.

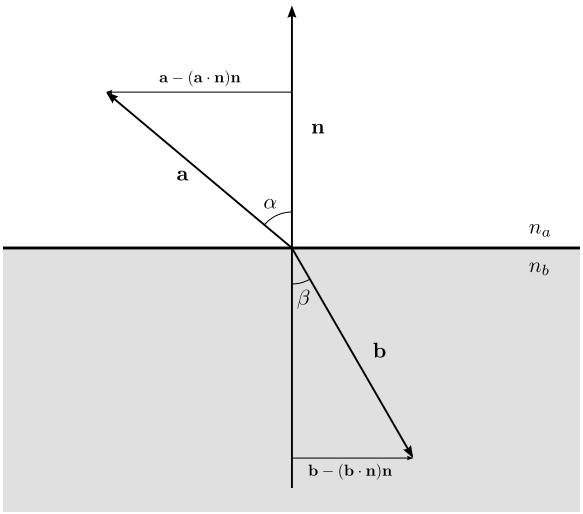


Figure 6: *Refraction.*

2.2 Backward Caustics

Considering that the caustic can be treated as a two-dimensional mesh of triangles (which may possibly overlap itself), it becomes apparent that one can modify the caustic directly by editing the photon mesh. In contrast to the previous section, where the caustic was computed from a height field, this is the inverse problem, i.e. a height field is to be inferred from a given caustic mesh, so this can be seen as "backward" caustic computation.

In order to reflect or refract rays on the specular surface such that they intersect the receiver at the designated points, the normal field needs to be adjusted accordingly. First, a way to compute a normal $\mathbf{n}$ from the incident and exitant ray directions $\mathbf{a}$ and $\mathbf{b}$ is required, which is described below. Section 2.3 then explains how the normal field will be converted to a height field so it becomes physically meaningful. Unfortunately, this step can usually not reproduce the normal field accurately, making direct editing of a caustic in this way extremely cumbersome due to the arising distortions.

Ray Directions to Normals. As mentioned, we first need a way to compute a normal $\mathbf{n}$ from the incident and exitant ray directions $\mathbf{a}$ and $\mathbf{b}$. These vectors are assumed to be normalized and to point away from the specular surface. In the case of reflection, it is trivial to infer the normal from incident and exitant directions; it is collinear with their mean. For refraction, let us furthermore assume that the ray directions are physically meaningful[1] and that they describe an actual refraction and not a total internal reflection. Now consider Snell's law

$$\frac{\sin\alpha}{\sin\beta} = \frac{n_b}{n_a} =: \eta.$$

The vectors $\mathbf{a}-(\mathbf{a}\cdot\mathbf{n})\mathbf{n}$ and $\mathbf{b}-(\mathbf{b}\cdot\mathbf{n})\mathbf{n}$ (see Figure 6) are collinear with opposite directions, and have lengths $\sin\alpha$ and $\sin\beta$, respectively. Thus $\mathbf{a} - (\mathbf{a}\cdot\mathbf{n})\mathbf{n} = -\eta(\mathbf{b} - (\mathbf{b}\cdot\mathbf{n})\mathbf{n})$, which can be rearranged to $\mathbf{a}+\eta\mathbf{b} = ((\mathbf{a}+\eta\mathbf{b})\cdot\mathbf{n})\mathbf{n}$. It follows that the projection of

[1] If no total internal reflection occurs, the angle between refractive ray directions is always at least ninety degrees plus the critical angle θ_c of the corresponding material boundary; $\theta_c = \arcsin\tilde{\eta}$, where $\tilde{\eta}$ is either η or η^{-1}, whichever is less than 1.

$\mathbf{a} + \eta\mathbf{b}$ on $\mathbf{n}$ is the same as the original vector. For $\eta \neq 1$, this vector always has nonzero length, so $\mathbf{n}$ and $\mathbf{a} + \eta\mathbf{b}$ must be collinear. We can therefore compute the desired surface normal from a linear combination of the normalized ray directions, a much simpler method than the one described by Papas et al. [2011]. There is not even a need for a distinction between reflection and refraction, since setting $\eta = 1$ will yield the desired result in the case of reflection.

2.3 Normal Fields to Height Fields

From a given photon mesh we can now obtain the normal field using the insight from the previous section. In order to physically manufacture a reflective or refractive object that produces the desired caustic, the normal field must be converted to a height field.

As described in Section 2.1, a normal field $\mathbf{n}(x, y)$ can be computed from a height field $z(x, y)$ by normalizing the vector

$$\left(-\tfrac{\partial z}{\partial x}(x, y), -\tfrac{\partial z}{\partial y}(x, y), 1\right)^{\mathrm{T}}.$$

Conversely, we can easily obtain derivatives for the desired field from the normals:

$$\tfrac{\partial z}{\partial x}(x, y) = -\tfrac{n_x(x,y)}{n_z(x,y)} \tag{4a}$$

$$\tfrac{\partial z}{\partial y}(x, y) = -\tfrac{n_y(x,y)}{n_z(x,y)}. \tag{4b}$$

However, the resulting vector field may not be a gradient field and thus not be exactly reproduceable. Algorithms that generate a normal field therefore must account for its integrability in order to be practicable. From Equation (4), it seems natural to formulate the problem as a system of linear equations, which approximate the derivatives of height field values using finite differences. A slightly different, analytical approach to integrating gradient fields is used in the literature, e.g. by Fattal et al. [2002]. The simplest way is to compute the derivatives on the midpoints of the grid edges (see Figure 7). To obtain the values at these positions, the normals of the two adjacent grid nodes are averaged. It should be noted that this will lead to some blurring between the original normal field and the one reconstructed from the resulting height field.

In matrix notation, we obtain a linear system $\mathbf{A}\mathbf{x} \approx \mathbf{b}$. $\mathbf{A}$ and $\mathbf{b}$ consist of two parts; $(N_x - 1)N_y$ rows for the $\frac{\partial z}{\partial x}$, and $N_x(N_y - 1)$ rows for the $\frac{\partial z}{\partial y}$. The vector $\mathbf{x}$ contains the $N_x N_y$ unknowns. As

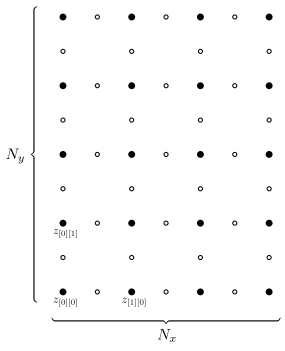


Figure 7: *Arrangement of the height field/normal field elements. The small points represent the locations of the derivatives. The unknowns are ordered from bottom to top and left to right, i.e.* $\{z_{[0][0]}, z_{[0][1]}, \ldots, z_{[1][0]}, \ldots\}$. *This determines the ordering of the matrix rows and columns in Section 2.3.*

expected, the possibility of inconsistent fields leads to a system that has more equations than unknowns for $N_x, N_y > 2$, which we solve by least-squares. The corresponding normal equations are

$$\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \mathbf{A}^{\mathrm{T}}\mathbf{b}. \tag{5}$$

Normal Equations Matrix. Let us take a closer look at $\mathbf{A}$, and derive $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ from Equation (5) to be used for an efficient, matrix-free implementation. As just mentioned, $\mathbf{A}$ consists of two parts:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_x \\ \mathbf{A}_y \end{bmatrix}. \tag{6}$$

$\mathbf{A}_x$ consists of $(N_x - 1) \times N_x$ blocks, each being an $N_y \times N_y$ submatrix, and each row of blocks corresponds to a column of $\frac{\partial z}{\partial x}$ derivatives. Analogously, $\mathbf{A}_y$ consists of $N_x \times N_x$ blocks, each being an $(N_y - 1) \times N_y$ submatrix, and each row of blocks represents a column of $\frac{\partial z}{\partial y}$ derivatives. For convenience, the factor $\frac{1}{h}$ from the finite differences is moved to the right-hand side, where h is the grid spacing. Thus, $\mathbf{b}$ contains the desired derivatives multiplied by h, and the only nonzero entries of $\mathbf{A}$ are ± 1.

Defining the $(N - 1) \times N$ matrix

$$\mathbf{D}_N := \begin{bmatrix} -1 & +1 & & \\ & -1 & +1 & \\ & & \ddots & \ddots \\ & & & -1 \;\; +1 \end{bmatrix}$$

allows us to write $\mathbf{A}_x$ and $\mathbf{A}_y$ using the Kronecker product.

$$\mathbf{A}_x = \mathbf{D}_{N_x} \otimes \mathbf{I}_{N_y}$$
$$\mathbf{A}_y = \mathbf{I}_{N_x} \otimes \mathbf{D}_{N_y}$$

Using $(\mathbf{A} \otimes \mathbf{B})^{\mathrm{T}} = \mathbf{A}^{\mathrm{T}} \otimes \mathbf{B}^{\mathrm{T}}$ and the mixed-product property of the Kronecker product $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$, we obtain

$$\begin{aligned} \mathbf{A}^{\mathrm{T}}\mathbf{A} &= \mathbf{A}_x^{\mathrm{T}}\mathbf{A}_x + \mathbf{A}_y^{\mathrm{T}}\mathbf{A}_y \\ &= (\mathbf{D}_{N_x}^{\mathrm{T}} \otimes \mathbf{I}_{N_y})(\mathbf{D}_{N_x} \otimes \mathbf{I}_{N_y}) \\ &\quad + (\mathbf{I}_{N_x} \otimes \mathbf{D}_{N_y}^{\mathrm{T}})(\mathbf{I}_{N_x} \otimes \mathbf{D}_{N_y}) \\ &= (\mathbf{D}_{N_x}^{\mathrm{T}}\mathbf{D}_{N_x} \otimes \mathbf{I}_{N_y}) + (\mathbf{I}_{N_x} \otimes \mathbf{D}_{N_y}^{\mathrm{T}}\mathbf{D}_{N_y}), \end{aligned}$$

with the $N \times N$ matrix

$$\mathbf{D}_N^{\mathrm{T}}\mathbf{D}_N = \begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & \ddots & \ddots & \ddots \\ & & & -1 & 2 \;\; -1 \\ & & & & -1 \;\; 1 \end{bmatrix}.$$

Putting this together, one can see the format of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$, which consists of $N_x \times N_x$ blocks of $N_y \times N_y$ submatrices: it is the Laplacian matrix of the $N_x \times N_y$ grid graph. This allows for an efficient implementation of the product $\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{x}$. The matrix is similar to a five-point finite difference stencil, which corresponds to a two-dimensional Poisson equation on a regular grid. Indeed, the approach used by Fattal et al. [2002] leads to Poisson's equation; one slight difference is that the derivation shown here automatically implicates the boundary condition.

Matrix Singularity. It should be noted that if $z(x, y)$ is a solution to the optimization problem, then so is $z'(x, y) = z(x, y) + d$. In other words, when adding a constant to all unknowns, the result will remain unchanged; in matrix notation (using the vector of ones $\mathbf{1}$), this implies $\mathbf{A}\mathbf{1} = \mathbf{0}$, which is also evident from the fact

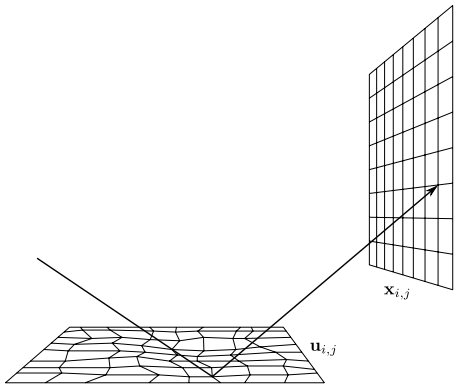


Figure 8: *Brightness warping: The mesh on the specular plane is deformed (vertices $\mathbf{u}_{i,j}$), while the mesh on the receiver (vertices $\mathbf{x}_{i,j}$) is fixed.*

that the sums in $\mathbf{A}$'s rows are 0. Consequently, the matrix $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ in the normal equations is singular and merely positive semi-definite. This problem can be avoided by fixing one of the unknowns to 0 by removing the corresponding row and column from $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ and $\mathbf{A}^{\mathrm{T}}\mathbf{b}$, which makes the matrix strictly positive definite and the solution unique. In fact, the effort for that additional handling is not necessary for both the Gauss-Seidel and the conjugate gradient method, which state symmetry and strict positive definiteness as sufficient (but not necessary) conditions for convergence [Golub and Van Loan 1996; Shewchuk 1994]. The former also converges for strictly diagonally dominant matrices, but it is easy to see that $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ is only weakly diagonally dominant. Despite this, the Gauss-Seidel as well as the the conjugate gradient method will work with the above matrix.

3 Brightness Warping

As explained in Section 2, a caustic can be generated by discretizing the light beam using a mesh, and tracing a ray for each vertex. The resulting projection on the receiver forms the photon mesh, and by deforming it, the caustic can be modified. For the optimization, we employ the reversed setup (see Figure 8): A fixed mesh is used to describe the caustic, and the points on the specular surface which cast the corresponding rays are moved. By deforming the latter mesh instead, the desired amounts of light can be allocated to the corresponding faces of the caustic; the larger the area of a face in this *warped mesh* on the specular plane, the more light is projected on the unchanged area in the caustic mesh, increasing the brightness. The boundary vertices of the warped mesh are confined to remain on the border. Once the deformation of this mesh is found, the normal field is obtained by interpolating the outgoing ray directions at the grid nodes using barycentric coordinates. The normal field can then be integrated to a height field.

It should be noted that this approach cannot introduce folds or otherwise change the overall shape of a caustic by design. However, there is no assumption that prevents the specified photon mesh from having folds—if there are folds, they will be preserved. In fact, this is another advantage over directly deforming the caustic mesh, since in that case it would be necessary to account for overlapping faces that make up the total resulting brightness. As a downside, the step where the warped mesh is mapped back to a regular grid may cause some degradation of the caustic's shape.

We will now set up a system of equations describing the brightness constraints (it will be seen that they are quadratic). Additionally, we want to ensure that the resulting normal field can be integrated to a height field. The corresponding equations will also be nonlinear.

The system of equations is posed as a least-squares problem, using a linear combination of the corresponding objective functions. This optimization problem can then, for example, be solved by steepest descent or the nonlinear conjugate gradient method.

3.1 Quad Brightness

We first determine the desired area of each quadrilateral in the warped mesh. As seen in equations (2) and (3), for a face i we have $I_{s,i}A_{s,i} = I_{d,i}A_{d,i}$, where $I_{s,i}$ is the irradiance on the reflector/refractor corresponding to the desired area $A_{s,i}$, and $I_{d,i}$ is the irradiance on the receiver corresponding to the area $A_{d,i}$. The value of $I_{s,i}$ is the same for all faces since a parallel light source is assumed; the $A_{d,i}$ are determined by the given caustic mesh. $I_{d,i}$ is not known but proportional to the desired brightness b_i of the caustic face, which is obtained from the desired image, hence $A_{s,i} \propto b_i A_{d,i}$. The right-hand side is known, and by summing over all faces, the proportionality constant can be eliminated because the total area on the specular surface is also fixed:

$$A_{s,i} = b_i A_{d,i} \cdot \frac{\sum_j A_{s,j}}{\sum_j b_j A_{d,j}}.$$

In Section 2.1, this was described using triangles. Our implementation uses a mesh of quadrilaterals, which should not be an issue since the algorithm does not seem to cause self-intersections in practice. To avoid singularities, a constant value can be added to the image, making all b_i strictly positive so the corresponding faces do not become arbitrarily small. The choice of this constant is a trade-off between stability and contrast—the implementation therefore allows the user to change it.

The area of a non-self-intersecting quadrilateral with counterclockwise labelling $ABCD$ is

$$\tfrac{1}{2}\left(\det\begin{bmatrix}\mathbf{u}_B-\mathbf{u}_A & \mathbf{u}_D-\mathbf{u}_A\end{bmatrix}+\det\begin{bmatrix}\mathbf{u}_C-\mathbf{u}_D & \mathbf{u}_C-\mathbf{u}_B\end{bmatrix}\right).$$

We turn this into the corresponding brightness condition, where $A_{s,ABCD}$ is the desired area from above. It is reasonable to use relative area deviations for the least-squares optimization, thus

$$\frac{\det\begin{bmatrix}\mathbf{u}_B-\mathbf{u}_A & \mathbf{u}_D-\mathbf{u}_A\end{bmatrix}+\det\begin{bmatrix}\mathbf{u}_C-\mathbf{u}_D & \mathbf{u}_C-\mathbf{u}_B\end{bmatrix}}{2A_{s,ABCD}} \stackrel{!}{=} 1.$$

We can rewrite this in matrix notation, where $\mathbf{u}$ is a vector containing all $\mathbf{u}_P$, and $\mathbf{M}_{ABCD}$ is symmetric:

$$f_{ABCD}(\mathbf{u}) := \mathbf{u}^{\mathrm{T}}\mathbf{M}_{ABCD}\mathbf{u} - 1 \stackrel{!}{=} 0.$$

Squaring and summing this quantity for all quadrilaterals, we obtain the objective function for the brightness part. The gradient of the squared expression,

$$\frac{\partial}{\partial\mathbf{u}}(f_{ABCD}(\mathbf{u})^2) = 2\left(\mathbf{u}^{\mathrm{T}}\mathbf{M}_{ABCD}\mathbf{u}-1\right)\cdot 2\mathbf{M}_{ABCD}\mathbf{u},$$

now allows us to use this brightness objective in a gradient-based optimization.

3.2 Integrability

As mentioned earlier, we also want to make sure that the resulting normal field can be faithfully reproduced by a height field. Let us briefly recap how the normal $\mathbf{n}_P$ at a vertex P on the reflective or refractive surface interlinks the height field and the warped mesh.

The normal is defined by the corresponding ingoing and outgoing light directions $\mathbf{i}$ and $\mathbf{o}_P$ and the quotient of refractive indices η (see Section 2.2, $\eta = 1$ for reflection). The incident direction $\mathbf{i}$ is constant and assumed to be normalized; $\mathbf{o}_P$ is simply the distance vector between the corresponding nodes on the specular and diffuse surface, and not normalized here. Without loss of generality, we consider the specular, normal-mapped surface to be located in the xy-plane. With a slight abuse of notation, we write $\mathbf{o}_P = \mathbf{x}_P - \mathbf{u}_P$, where the vertex position $\mathbf{u}_P$ is treated as three-dimensional vector with a z component of 0. The normal $\mathbf{n}_P$ is now linked to the derivatives of the unknown height field $z(\mathbf{u})$ as seen in Section 2.3:

$$\frac{\partial z}{\partial u} = -\frac{n_{Px}}{n_{Pz}} = -\frac{\|\mathbf{o}_P\| i_x + \eta o_{Px}}{\|\mathbf{o}_P\| i_z + \eta o_{Pz}}, \tag{8a}$$

$$\frac{\partial z}{\partial v} = -\frac{n_{Py}}{n_{Pz}} = -\frac{\|\mathbf{o}_P\| i_y + \eta o_{Py}}{\|\mathbf{o}_P\| i_z + \eta o_{Pz}}. \tag{8b}$$

Since the derivatives correspond to the normal divided by its z component, the normal $\mathbf{n}_P$ does not need to be normalized. Therefore, the modified definition $\mathbf{n}_P = \|\mathbf{o}_P\|\mathbf{i} + \eta\mathbf{o}_P$ instead of $\mathbf{n}_P = \mathbf{i} + \eta\frac{\mathbf{o}_P}{\|\mathbf{o}_P\|}$ is used here for convenience.

To ensure the existence of a height field with the same normals, the vector field defined by the normals must be conservative, i.e. the line integral around any closed loop must be zero,

$$\oint \nabla z(\mathbf{r}) \cdot d\mathbf{r} \stackrel{!}{=} 0.$$

This is a consequence of the gradient theorem. An alternative condition, which requires partial derivatives of the vector field defined by Equation (8), is stated by Fattal et al. [2002]. The approach described here operates on a quad mesh, where the vertices define the values of the vector field, so the computation of the field's partial derivatives would not be trivial. Moreover, we are also interested in the derivatives of the corresponding equations, which would complicate the computations even more. We therefore prefer the loop integral above to assert the normal field's integrability.

If that condition is satisfied for the edges of all quadrilaterals $ABCD$ of our mesh, it will hold for every closed loop along mesh edges, and be approximately satisfied for any loop in general. To formulate the loop integral, we linearly interpolate the vector field along the quad edges, which is equivalent to averaging the values of the edge end points. Again, the $\mathbf{u}_P$ are treated as three-dimensional vectors with a z component of 0.

$$\begin{aligned} g_{ABCD}(\mathbf{u}) := \frac{1}{2}\Bigg(&\left(\frac{\mathbf{n}_B}{n_{Bz}} - \frac{\mathbf{n}_D}{n_{Dz}}\right)\cdot(\mathbf{u}_C - \mathbf{u}_A) \\ &- \left(\frac{\mathbf{n}_C}{n_{Cz}} - \frac{\mathbf{n}_A}{n_{Az}}\right)\cdot(\mathbf{u}_B - \mathbf{u}_D)\Bigg) \\ \approx\ &-\oint_{ABCD} \nabla z(\mathbf{r}) \cdot d\mathbf{r} \end{aligned} \tag{9}$$

Squaring and summing this quantity for all quadrilaterals, we obtain the objective function for the integrability part. We exemplarily derive this term for u_{Ax}, to be used for steepest descent or the conjugate gradient method.

$$\begin{aligned} \frac{d}{du_{Ax}}\left(g_{ABCD}(\mathbf{u})^2\right) = g_{ABCD}(\mathbf{u})\Bigg(&-\left(\frac{n_{Bx}}{n_{Bz}} - \frac{n_{Dx}}{n_{Dz}}\right) \\ &+ \frac{d}{du_{Ax}}\left(\frac{\mathbf{n}_A}{n_{Az}}\cdot(\mathbf{u}_B - \mathbf{u}_D)\right)\Bigg) \end{aligned} \tag{10}$$

Recalling the definitions $\mathbf{n}_A = \|\mathbf{o}_A\|\mathbf{i} + \eta\mathbf{o}_A$ and $\mathbf{o}_A = \mathbf{x}_A - \mathbf{u}_A$, we obtain:

$$\begin{aligned} &\frac{d}{du_{Ax}}\left(\frac{\mathbf{n}_A}{n_{Az}}\cdot(\mathbf{u}_B - \mathbf{u}_D)\right) \\ &= \frac{1}{n_{Az}}\left(\frac{d\mathbf{n}_A}{du_{Ax}}\cdot(\mathbf{u}_B - \mathbf{u}_D)\right) \\ &\quad - \frac{1}{n_{Az}^2}\frac{dn_{Az}}{du_{Ax}}(\mathbf{n}_A\cdot(\mathbf{u}_B - \mathbf{u}_D)) \\ &= \frac{1}{n_{Az}}\left(\frac{d\|\mathbf{o}_A\|}{du_{Ax}}(\mathbf{i}\cdot(\mathbf{u}_B - \mathbf{u}_D)) - \eta(u_{Bx} - u_{Dx})\right) \\ &\quad - \frac{1}{n_{Az}^2}\frac{d\|\mathbf{o}_A\|}{du_{Ax}} i_z(\mathbf{n}_A\cdot(\mathbf{u}_B - \mathbf{u}_D)) \\ &= \frac{1}{n_{Az}}\left(\frac{d\|\mathbf{o}_A\|}{du_{Ax}}\left(\left(\mathbf{i} - i_z\frac{\mathbf{n}_A}{n_{Az}}\right)\cdot(\mathbf{u}_B - \mathbf{u}_D)\right)\right) \\ &\quad - \frac{1}{n_{Az}}\eta(u_{Bx} - u_{Dx}). \end{aligned} \tag{11}$$

Finally,

$$\frac{d\|\mathbf{o}_A\|}{d\mathbf{u}_A} = \frac{1}{2\|\mathbf{o}_A\|}\cdot 2(\mathbf{u}_A - \mathbf{x}_A) = -\frac{\mathbf{o}_A}{\|\mathbf{o}_A\|}. \tag{12}$$

Putting everything together,

$$\begin{aligned} &\frac{d}{du_{Ax}}\left(g_{ABCD}(\mathbf{u})^2\right) \\ &\quad = g_{ABCD}(\mathbf{u})\Bigg(-\left(\frac{n_{Bx}}{n_{Bz}} - \frac{n_{Dx}}{n_{Dz}}\right) \\ &\qquad - \frac{1}{n_{Az}}\eta(u_{Bx} - u_{Dx}) \\ &\qquad - \frac{1}{n_{Az}}\frac{o_{Ax}}{\|\mathbf{o}_A\|}\left(\mathbf{i} - i_z\frac{\mathbf{n}_A}{n_{Az}}\right)\cdot(\mathbf{u}_B - \mathbf{u}_D)\Bigg), \end{aligned} \tag{13}$$

we are now able to use this integrability condition in an optimization. The resulting objective function is

$$\min_{\mathbf{u}} \sum_{ABCD} \left(\lambda\bar{A}_d \cdot f_{ABCD}(\mathbf{u})^2 + g_{ABCD}(\mathbf{u})^2\right), \tag{14}$$

where the average area term $\bar{A}_d$ ensures that the weighting coefficient λ is scale-independent. We solve this equation using a pyramid scheme reminiscent of multigrid solvers. The coarse problem can be obtained by joining four quads of the high-resolution warped mesh to one low-resolution quad. The desired face areas for the coarse mesh are then simply the sums of the four corresponding faces of the fine mesh. The integrability terms do not require any action for the coarsening step. After solving the problem on the lower resolution, the warped mesh is simply subdivided, and the nonlinear conjugate gradient method is applied on the higher resolution, reproducing finer details of the desired image.

3.3 Imperfect Specularity

Physical manufacturing of the computed surfaces will typically involve some polishing step to improve the specularity. Achieving perfect specularity, however, as assumed by the approaches devised in the previous section, is unrealistic. The obtained surfaces will not be perfectly smooth, resulting in some scattering, which we would

(a) rough glass (b) without deconvolution (c) with deconvolution

Figure 9: *Results using LuxRender's rough glass material for Lena, without and with deconvolution. To visualize the material, (a) shows a rendering of two glass blocks, one perfectly specular, and one whose top face uses the rough glass material.*

(b) without deconvolution (c) with deconvolution

Figure 10: *LuxRender results for Claude Monet's "Mohnblumen". The deconvolution was used to abolish the blur introduced by using an area light source, the corresponding point spread function is displayed in the inset in (c). Only the red color channel of the original was used so the contrast between the poppies and the grass is more noticeable than in a standard grayscale version.*

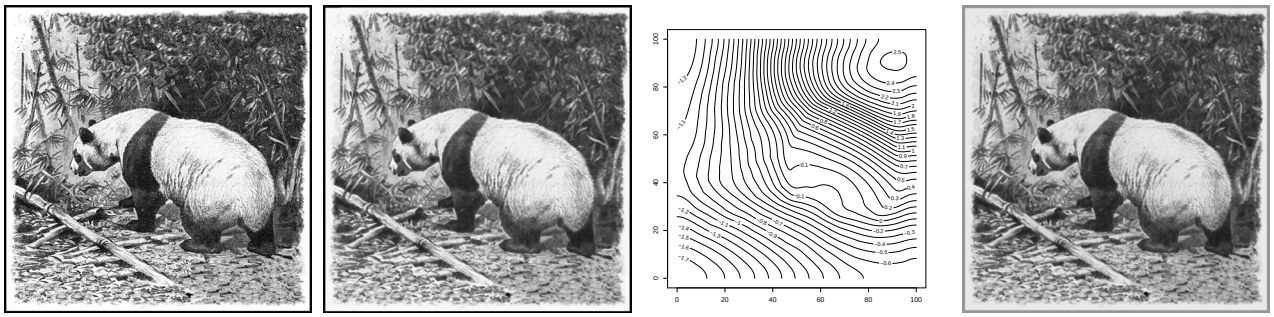
like to account for. Instead of working with a full-blown bidirectional scattering distribution function (BSDF), we resort to a simplification: we assume that we are given the point spread function (PSF) resulting on the receiver. This can be imagined as pointing a laser at the surface and looking at the speck of light it produces. We furthermore assume that its shape does not depend on the position where the laser hits the surface. In terms of a BSDF, this would mean a spatially homogeneous, isotropic distribution. Fresnel effects, i.e. varying reflectivity depending on the incidence angle on a surface, are neglected. Distortions of the PSF due to varying incidence angles on the receiver and due to varying distances between the positions on reflector/refractor and receiver are also disregarded.

Under these assumptions, the caustic resulting from a surface is simply the convolution of the point spread function with the corresponding caustic as it would look assuming perfect specularity. For the brightness warping approach, the problem of handling imperfect surfaces therefore turns into a deconvolution problem, which can be solved using the Richardson-Lucy algorithm [Richardson 1972]. The algorithm cannot produce negative values, but it can introduce ringing artifacts. The implementation we use extends the original image at the borders, so the deconvolution will be as large as the user-specified input and no cropping occurs. Furthermore, the values at the border are bevelled to prevent artifacts, as suggested by Richardson.

Since no experiments with physical objects were carried out to evaluate this method, a simulation was performed using LuxRender's rough glass material, see Figure 9. To determine the point spread function of this surface, a screen with a small hole in it was used to reduce the light to a narrow beam before it passed through the glass. The resulting image of the PSF was then cropped, scaled[2], and used for the deconvolution of the target image.

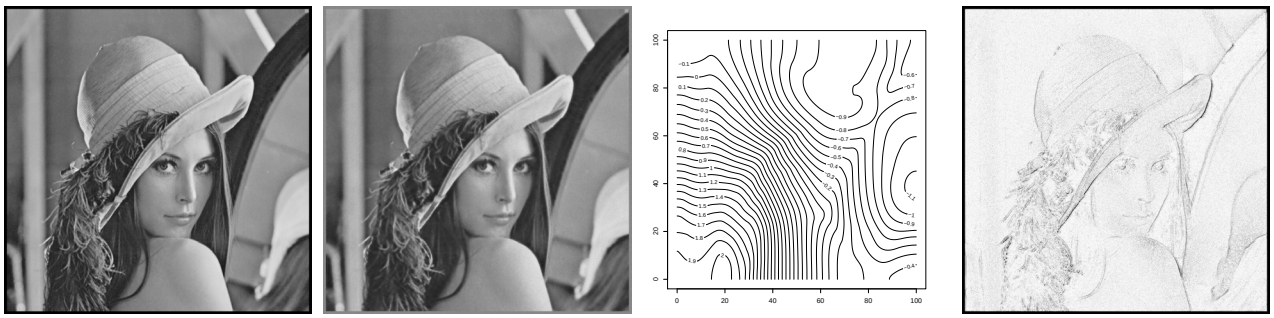
Similar to imperfect specularity, an area light source also results in a cone of light emanating from a single point on the surface instead of a single ray. Consequently, the resulting blur of a caustic can likewise be treated as a convolution, and the same approach may be used to account for it. This time, the PSF can obviously not be determined using a laser, instead a screen with a small hole can be placed directly at the specular surface. Figure 10 shows the result of a simulation carried out in this way. Since the PSF is a disk, the original image was blurred to avoid severe ringing artifacts in the deconvolved image. The blurred version was also used for the reference example without deconvolution. In practice, this obviously defeats the purpose of deconvolving, so depending on the input image, it is only of limited use for this kind of PSF. The alternative is to simply accept the resulting ringing in the deconvolved image, though it might lead to considerable problems due to inaccuracies when the surface is manufactured.

[2]The size of one pixel of the PSF must match the size of one pixel of the original image as it is projected onto the receiver. For example, the configuration used for Figures 9(b) and 9(c), the setup using a planar surface results in a caustic that is 911 pixels wide. The same configuration was used to render the PSF. Since the Lena image is 512 pixels wide, the image of the PSF produced by LuxRender was scaled by a factor of $\frac{512}{911}$.



(a) template image (b) after normals to height field optimization (c) height field isocontours (d) LuxRender result

Figure 12: *Brightness warping result for the panda template (originally by Friedrich W. Kuhnert, 1865–1926), using a 10×10 cm field of 641×641 nodes, with a distance of 25 cm to the receiver and a refractive index of 1.5. The measurements of the height field isocontours ((c)) are given in millimeters.*



(a) template image (b) LuxRender result (c) height field isocontours (d) enhanced difference

Figure 13: *Brightness warping result for the Lena test image. The setup was the same as in Figure 12. The difference image (d) was obtained by computing the difference from (a) to (b). The result was multiplied by 4 for better visibility and then inverted.*

4 Results and Discussion

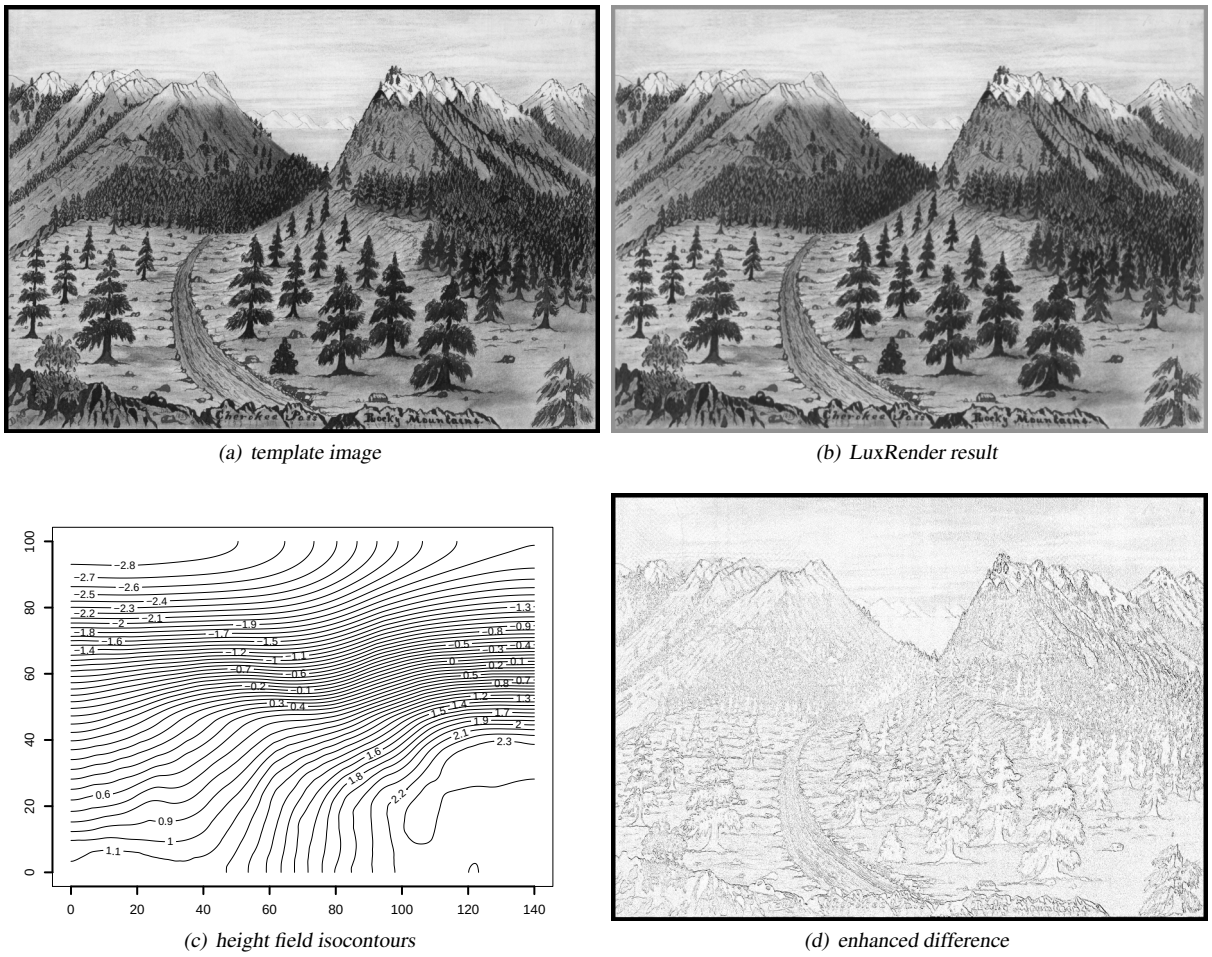
Figures 11–14 show some examples generated by our method. To verify the results, the physically-based software LuxRender was used to generate output images. For these images the actual height field was used to render the caustics and not just a normal field. For each image except Figure 14, the optimization for a 641×641 field (limited to 200 iterations per multigrid level, and excluding the normal field to height field conversion) took between three and four minutes. The normal field to height field conversion took just under twenty seconds at this resolution. For the 1793×1281 field in Figure 14, the optimization took about eighteen minutes, plus two minutes for the height field computation. The tests were run on an Intel Core 2 Quad Processor Q6600 at 2.4GHz; the implementation is single-threaded and hardly optimized. The brightness of the resulting images was adjusted so the brightest areas became roughly white.

(a) before normals to height field optimization (b) after normals to height field optimization

Figure 11: *Brightness warping output for the Mona Lisa template on a 641×641 field, using a multigrid approach with at most 200 nonlinear conjugate gradient iterations per level. Besides some minor blurring (as discussed in Section 2.3), the difference introduced by converting the normal field to a height field is hardly noticeable.*

It can be seen that the proposed approach is able to accurately reproduce images. Since it produces a smooth surface, the results do not suffer from the quantization artifacts inherent to systems like the one proposed by Papas et al. [2011]. There are also no small discontinuities that may introduce artifacts, and there is less danger of damaging surface features while polishing a milled prototype. As seen in Figure 14, high resolutions can be realized in a practicable amount of time, generating results of virtually arbitrary precision. Thus, the achievable accuracy of caustics using this method is mainly limited by the manufacturing process, namely how accurately a height field can be produced, and how well the surface can be polished to near-perfect specularity.

The Lena test image (Figure 13) was also used by Papas et al. as one of their examples; in addition, they have asked Finck et al. [2010] to run their algorithm on it. A simulation of the resulting caustic



(a) template image

(b) LuxRender result

(c) height field isocontours

(d) enhanced difference

Figure 14: *Brightness warping result for the "Cherokee Pass, Rocky Mountains" image (a 1859 drawing by Daniel A. Jenks). The image was generated using a deliberately abundant field resolution of 1793×1281 over 14×10 cm, the same setup as in Figure 12 was used otherwise. Again, (d) was obtained by adjusting (a) according to the constant added to the brightness (see Section 3.1), computing the difference to (b), multiplying by 4 and inverting the result.*

can be seen in [Papas et al. 2011, Figure 11] for the approach by Finckh et al. (which apparently generated a fold-free caustic), and in [Papas et al. 2011, Figure 12] for the one that Papas et al. have proposed. Their setup, a 10×10 cm field at a distance of 25 cm to the receiver, was also used in the examples shown here.

Considering the achievable accuracy, it is not surprising that both the methods by Papas et al. and Finckh et al. are outperformed by the approach proposed in this paper. For a more meaningful and fair comparison, however, it would be interesting to physically manufacture the field as Papas et al. have done.

We notice that the height fields computed using brightness warping tend to be rather flat, ranging between 3–9 mm for a piece of area 10×10 cm . Directly related to the field's amplitude is the amount of distortion when comparing caustics computed from a normal-mapped plane and actually displaced surfaces (as used by the LuxRender simulations shown in Figures 9–14). Putting the low depth range in relation to the distance to the receiver for reasonable setups, this justifies the simplification of using a normal-mapped plane.

5 Conclusion

We have introduced an optimization method for computing height field surfaces that create a pre-defined reflected or refracted caustic image. Our experiments show that high accuracy can be achieved and that complex images can be reproduced faithfully.

Future work will be directed towards lifting some of the limitations of our current approach. Most notably, directly incorporating the height distribution of the computed surface into the optimization and extending the approach to general 3D surfaces offers challenging problems for future research. In addition, we plan to extend the formulation to deal with dis-contiguous caustics and folds, multiple caustics from different input illumination, more realistic light sources, and dynamic caustics.

References

ANSON, O., SERON, F. J., AND GUTIERREZ, D. 2008. NURBS-based inverse reflector design. In *Proceedings of CEIG 2008*, 65–74.

DOYLE, S., CORCORAN, D., AND CONNELL, J. 1999. Automated mirror design using an evolution strategy. *Optical Engineering 38*, 2, 323–333.

FATTAL, R., LISCHINSKI, D., AND WERMAN, M. 2002. Gradient domain high dynamic range compression. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, ACM, New York, NY, USA, SIGGRAPH '02, 249–256.

FINCKH, M., DAMMERTZ, H., AND LENSCH, H. P. 2010. Geometry construction from caustic images. In *Computer Vision – ECCV 2010*, Springer, K. Daniilidis, P. Maragos, and N. Paragios, Eds., vol. 6315 of *Lecture Notes in Computer Science*, 464–477.

GOLUB, G., AND VAN LOAN, C. 1996. *Matrix computations*, 3rd ed. Johns Hopkins University Press.

LOOS, J., SLUSALLEK, P., AND SEIDEL, H.-P. 1998. Using wavefront tracing for the visualization and optimization of progressive lenses. *Computer Graphics Forum 17*, 3, 255–265.

MAS, A., MARTN, I., AND PATOW, G. 2009. Fast inverse reflector design (FIRD). *Computer Graphics Forum 28*, 8, 2046–2056.

NEUBAUER, A. 1997. Design of 3D-reflectors for near field and far field problems. *Institute for Mathematics and its Applications 92*, 101–118.

PAPAS, M., JAROSZ, W., JAKOB, W., RUSINKIEWICZ, S., MATUSIK, W., AND WEYRICH, T. 2011. Goal-based caustics. *Computer Graphics Forum (Proceedings of Eurographics '11) 30*, 2 (June), 503–511.

PATOW, G., AND PUEYO, X. 2005. A survey of inverse surface design from light transport behavior specification. *Computer Graphics Forum 24*, 4, 773–789.

PATOW, G., PUEYO, X., AND VINACUA, A. 2007. User-guided inverse reflector design. *Computers & Graphics 31*, 3, 501–515.

RICHARDSON, W. H. 1972. Bayesian-based iterative method of image restoration. *Journal of the Optical Society of America 62*, 1 (Jan.), 55–59.

SHEWCHUK, J. R. 1994. An introduction to the conjugate gradient method without the agonizing pain. Tech. rep., Carnegie Mellon University, Pittsburgh, PA, USA.

WESTCOTT, B. S., AND NORRIS, A. P. 1975. Reflector synthesis for generalized far-fields. *Journal of Physics A: Mathematical and General 8*, 4, 521.

WEYRICH, T., PEERS, P., MATUSIK, W., AND RUSINKIEWICZ, S. 2009. Fabricating microgeometry for custom surface reflectance. *ACM Transactions on Graphics (Proc. SIGGRAPH) 28*, 3 (Aug.).