Accurate computation of single scattering in participating media with refractive triangle boundaries

Nicolas Holzschuch

Accurate computation of single scattering in participating media with refractive triangle boundaries

Nicolas Holzschuch

Accurate computation of single scattering in participating media with refractive triangle boundaries

Nicolas Holzschuch[*]

Project-Team Maverick

Abstract: Single scattering effects, also called volume caustics, are high-frequency effects, appearing in participating media with low opacity. They are more visible and interesting in the presence of refractive interfaces that focus light rays. Refraction effects also make them harder to compute, since screen locality does not correlate with spatial locality in the medium. In this paper we give a new method for accurate computation of single scattering effects in a participating media enclosed by refractive interfaces. Our algorithm is based on the observation that although radiance along each camera ray is irregular, contributions from individual triangles are smooth. It uses minimal information and does not require additional data structures.

---

[*] INRIA Grenoble Rhône-Alpes and LJK (CNRS and Université de Grenoble)

Calcul précis des événements de *single scatter* dans les matériaux translucides à frontière réfractive

Résumé : Nous nous intéressons ici aux matériaux translucides dont l'index de réfraction est différent de celui de l'environnement. La lumière interagit avec les particules qui composent le matériau, produisant des effets lumineux. Ces effets sont encore plus visibles en présence d'interfaces réfractives qui concentrent la lumière. La réfraction complique également les calculs: deux pixels proches à l'écran peuvent correspondre à des zones éloignées dans l'espace. Dans ce rapport, nous décrivons un nouvel algorithme pour le calcul des effets de scattering dans les milieux participants avec réfraction. Notre algorithme se base sur l'observation que, bien que la radiance le long d'un rayon soit très irrégulière, les contributions individuelles de chaque triangle sont très régulières, et il est possible d'en calculer les limites. Notre algorithme ne nécessite aucune structure de donnée annexe, et n'utilise que l'information déjà présente: la géométrie du triangle et la hiérarchie de boites englobantes.

Mots-clés : Simulation de l'éclairage, milieux participants, dispersion, réfraction

(a) Our algorithm (169 s) **(b)** Point samples [WZHB09] (27 samples, 170 s) **(c)** Photon mapping (650 000 photons, 167 s) **(d)** Bi-Directional Path-Tracing (520 samples per pixel, 171 s)
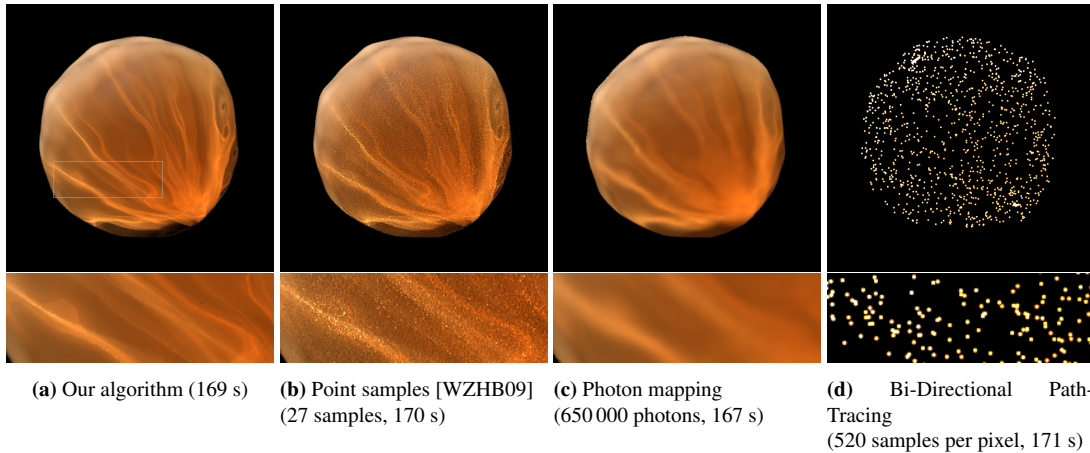
Figure 1: Single scattering: comparison between our algorithm and existing methods (equal computation time) on a translucent sphere illuminated from behind by a point light source. For Bi-Directional Path-Tracing, we replaced the point light source with a sphere so the algorithm actually produces a picture.

1   Introduction

Participating media interact with the light traversing them, scattering it and absorbing it. Combined with light concentration from occlusion or refractive interfaces, scattering results in high frequency effects, such as "god rays" or underwater caustics.

Observing these effects from outside the participating media adds another refractive interface between the camera and the media. This distorts the scattering effects, resulting in beautiful but challenging effets, visible in Figure 1. The two refractive interfaces between the light source and the camera make it difficult to find light paths going from the light source to the camera after exactly one scattering event.

In this paper, we focus on objects enclosed by a refractive interface defined by triangles with interpolated normals, a representation close to ubiquitous in computer graphics. For a given camera ray, refracted at the interface, the radiance is equal to the integral of radiance received along this ray and scattered in the right direction. This out-scattered radiance varies quickly along the ray, making it difficult to compute the integral accurately. We observe that the radiance caused by an individual triangle on the surface varies smoothly; discontinuities correspond to triangle edges.

We exploit this property in a new algorithm for computing single scattering effects with refractive interfaces. For each triangle on the surface, we quickly identify whether it is making a contribution to this specific camera ray, and on which segment of the ray. We then sample regularly inside this segment. Our method works for highly tesselated objects and is significantly faster than existing methods while providing better quality results. It does not require additional data structures or pre-processing, making it very easy to integrate inside existing renderers.

We review previous work on single and multiple scattering in the next section. We describe our algorithm in details in section 3. We study the behaviour of our algorithm and compare with previous work in section 4. Finally, we conclude and present avenues for future work in section 5.

2   Previous work

Pegoraro and Parker [PP09] provided an analytical solution for single scattering from a point light source in an homogeneous participating media. Their approach is useful for scenes filled with smoke or dust,

with no refractive interfaces.

Volume caustics are caused by light being refracted at the interface between the refracted media and air. The most striking examples are caustics in turbid water. Nishita and Nakamae [NN94] described an algorithm to compute these caustics by creating a light shaft for each triangle on the surface, then accumulating the contributions from these light shafts using the accumulation buffer. Iwasaki et al. [IDN01] improved the method by using color blending between vertex values inside the light shaft. Ernst et al. [EAMJ05] further improved the light shaft method using non-planar shaft boundaries and computing illumination inside the shaft with fragment shaders. All these work focused on a single refractive interface, displaying volume caustics as seen from inside the media.

Hu et al. [HDI*10] traced light rays inside the participating media, after up to two specular events (reflections or refractions). They accumulate radiance from these light rays in a second step. Their method produces volume caustics interactively, but with no refractive interface between the camera and the volume caustics.

Ihrke et al. [IZT*07] presented a method to compute light transport in refractive objects by modelling the propagation of the light wavefront inside the object. Their method is interactive and handles objects made of inhomogeneous materials. It relies on a voxelised representation of light distribution inside the object. Sun et al. [SZS*08] also used a voxelized representation of light inside the object, but computed scattered light using photon propagation. For both methods, the resolution of the voxel grid limits the spatial resolution of volume caustics.

Sun et al. [SZLG10] introduced the *line space gathering* method. They first trace rays from the light source and store them as points in Plücker coordinates. Then, for each camera ray, they find the closest light rays using Plücker space and compute their scattering contributions. Their technique produces high quality results, both inside and outside translucent objects, taking into account refraction. The data structure used to store camera and light rays in Plücker space uses approximately 400 MB.

Walter et al. [WZHB09] used a Newton-Raphson iterative method to find all paths connecting a sample point inside the participating media to the light source through a refractive interface. They place sample points along camera rays inside the translucent object to compute volume caustics. They used a position-normal tree to cull triangles in the search. Our method can be seen as an extension of theirs; we remove the need for additional data structure and place sample points optimally for each triangle.

Jarosz et al. [JZJ08] introduced the *beam radiance estimate* for efficient rendering of participating media with photon mapping. It starts as conventional volumetric photon mapping, storing photons inside the participating media. In the gathering step, camera rays are represente as beams, and gathers contributions from all photons around the beam. The technique was later extended to *photon beams*, by Jarosz et al. [JNSJ11]: light entering the participating media is also stored as a beam of light. In the gathering step, they compute contributions from light beams to camera beams, as well as from photons to camera beams. *Progressive photon beams*, by Jarosz et al. [JNT*11], extends this idea by using beams of varying width in successive rendering steps.

3   Single scattering computations

We consider an object, filled with an homogenous participating media, of index of refraction η, mean free path $\ell = \sigma_T^{-1}$, albedo α and phase function p (see Figure 2a). The boundary of the translucent object is a smooth dielectric interface. We assume that it is modelled as triangles with interpolated normals.

This object is under illumination from a light source. In all that follows, $\mathbf{L}$ is a point sample on the light source: its position for a point light source, a random sample for an area light source.

We focus on single scattering effects. For each camera ray, we take the corresponding refracted ray inside the translucent object. Our goal is to compute the integral of all single scattering effects along this ray.
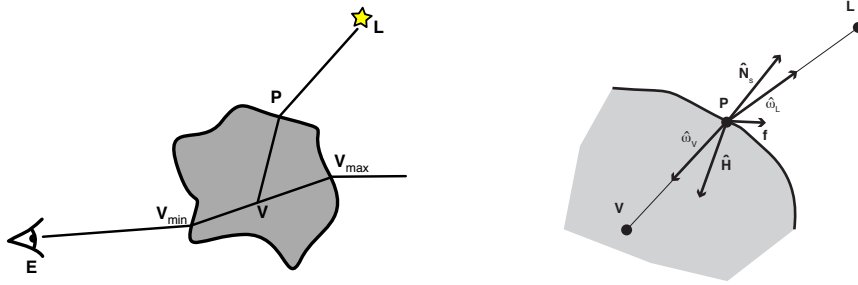
(a) We consider a translucent object, under illumination from a light source **L**.

(b) At point **P**, we define the half-vector $\widehat{\mathbf{H}}$, and compare it with the shading normal $\widehat{\mathbf{N}}_s$. Our goal is to find zeros of the function $\mathbf{f} = \widehat{\mathbf{H}} + \widehat{\mathbf{N}}_s$.

Figure 2: Our settings and notations.

3.1 Sample points algorithm

First, we briefly review the algorithm of Walter et al. [WZHB09]. They place several sample points **V** along each camera ray. For each sample point, they compute all paths connecting **V** to the light source **L** (see Section 3.1.1), then compute their contribution and sum them (see Section 3.1.2). For clarity, we reuse their notations. $\widehat{\mathbf{v}}$ is a normalized vector: $\widehat{\mathbf{v}} = \mathbf{v}/\|\mathbf{v}\|$.

3.1.1 For each individual triangle

Given an individual triangle with interpolated normals, a sample point **V** and a light source sample **L** the problem is to find all paths connecting **L** to **V** while following Snell's law. The key idea is to transform this into the search for zeros of a function **f**. Given a point **P** on the triangle, with shading normal $\widehat{\mathbf{N}}_S$, we define the normalized half-vector $\widehat{\mathbf{H}}$ and the function **f** (see Figure 2b):

$$
\begin{align}
d_V &= \|\mathbf{V} - \mathbf{P}\| \tag{1}\\
d_L &= \|\mathbf{L} - \mathbf{P}\| \tag{2}\\
\widehat{\omega}_V &= (\mathbf{V} - \mathbf{P})/d_V \tag{3}\\
\widehat{\omega}_L &= (\mathbf{L} - \mathbf{P})/d_L \tag{4}\\
\mathbf{H} &= \eta\widehat{\omega}_V + \widehat{\omega}_L \tag{5}\\
\widehat{\mathbf{H}} &= \mathbf{H}/\|\mathbf{H}\| \tag{6}\\
\mathbf{f}(\mathbf{P}) &= \widehat{\mathbf{H}} + \widehat{\mathbf{N}}_S \tag{7}
\end{align}
$$

All points **P** such that $\mathbf{f}(\mathbf{P}) = \mathbf{0}$ correspond to a path connecting **V** and **L**, with the refraction angle following Snell's law.

To find the zeros of **f**, Walter et al. [WZHB09] used a Newton-Raphson iterative method, with one small change: since point **P** on the triangle uses barycentric coordinates (a, b), **f** is a function from $\mathbf{R}^2$ into $\mathbf{R}^3$. The Jacobian J of **f** is a non-square matrix and thus not invertible. They used the pseudo-inverse of J, J^+:

$$J^+ = (J^\mathsf{T} J)^{-1} J^\mathsf{T} \tag{8}$$

The iterative Newton-Raphson method becomes:

$$\mathbf{P}_{n+1} = \mathbf{P}_n - J^+ \mathbf{f}(\mathbf{P}_n) \tag{9}$$

For this specific problem, it converges rapidly to a solution.

3.1.2   Contribution from each path

Once a path connecting $\mathbf{V}$ and $\mathbf{L}$ has been found, its contribution to the pixel is computed:

$$\text{contribution} = \frac{I_e F A p}{D} \tag{10}$$

where I_e is the intensity of the light source sample $\mathbf{L}$, F is the Fresnel factor at both interfaces, A is the volume attenuation (the integral of $e^{-\sigma_T x}$ along both segments inside the media), p is the phase function at $\mathbf{V}$ and D is the distance correction factor.

In the absence of participating media and refractive interface, D is the square of the distance between V and L, $D = \|\mathbf{L} - \mathbf{V}\|^2 = (d_V + d_L)^2$. With refractive media and constant geometric normal $\widehat{\mathbf{N}}_g$, D has a simple expression:

$$D = (d_V + \eta d_L) \left(\frac{|\widehat{\omega_L} \cdot \widehat{\mathbf{N}}_g|}{|\widehat{\omega_V} \cdot \widehat{\mathbf{N}}_g|} d_V + \frac{|\widehat{\omega_V} \cdot \widehat{\mathbf{N}}_g|}{|\widehat{\omega_L} \cdot \widehat{\mathbf{N}}_g|} d_L \right) \tag{11}$$

With interpolated shading normals, D has a more complicated expression, based on ray differentials at the interface [Ige99]. To compute D, Walter et al. take two vectors perpendicular to ω_V and each other: $\mathbf{u}_\perp$ and $\mathbf{u}_\parallel$, then compute how $\mathbf{L}$ changes when ω_V is perturbed along these vectors. D is the cross-product of these derivatives:

$$D = \left\| \frac{d\mathbf{L}}{d\mathbf{u}_\perp} \times \frac{d\mathbf{L}}{d\mathbf{u}_\parallel} \right\| \tag{12}$$

A complete expression of D can be found in [WZHB09]. We refer the interested readers to this article for more information.

3.1.3   Pruning triangles

The Newton-Raphson method converges with few iterations for each sample point $\mathbf{V}$ and each triangle. In theory, the search should be done for all triangles and all sample points, but that would be too costly. Walter et al. [WZHB09] used several methods to prune the triangles used in the search:

- Sidedness agreement: $\mathbf{V}$ must be in the negative half-space of the triangle, and $\mathbf{L}$ in the positive half-space, for both the geometric normal $\widehat{\mathbf{N}}_g$ and the shading normal $\widehat{\mathbf{N}}_s$.

- Spindle test: the angle between the incoming and the refracted ray is between $\frac{\pi}{2} + \arcsin(1/\eta)$ and π. For a given segment $[\mathbf{VL}]$, this restricts the potential solutions $\mathbf{P}$ to inside a surface of revolution around $[\mathbf{VL}]$, called the *spindle*.

The spindle test doesn't depend on triangle normals and can be used to prune large parts of the scene. For the sidedness agreement, Walter et al. [WZHB09] used a position-normal tree built from the triangles upward.

Together, these pruning methods reduced the computation time to something acceptable, a few minutes on a 8-core Xeon.

3.1.4   Numerical issues

The algorithm described by Walter et al. [WZHB09] converges quickly for objects made of flat triangles and moderately curved surfaces. The authors reported numerical issues for more complex scenes such as the bumpy sphere (see Figure 1):

- They have to place more samples $\mathbf{V}$ along each camera ray, up to 128 samples for the bumpy sphere,

- Some samples had extremely large values, which resulted in noise in the picture. These values had to be clamped in order to reduce the noise.
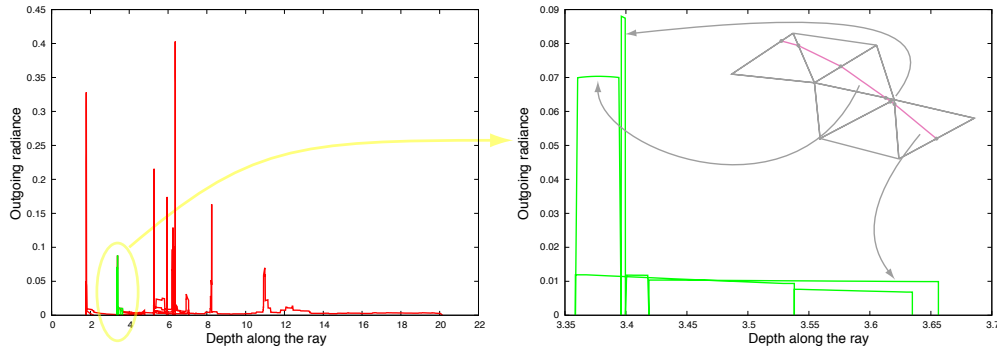
Figure 3: (left) Outgoing radiance, measured along a refracted ray inside the material. The radiance values are highly irregular: computing the integral accurately would requires many samples. If we isolate contributions from each triangle (right), each of them is smooth. Discontinuities occur only at the triangle boundaries.

3.2 Our algorithm

3.2.1 Motivation: study of a single camera ray

For a better understanding of single scattering effects with refraction, we focus on a single camera ray. It is refracted as it enters the translucent object, then travels in a straight line until it exits the object. We placed a large number of sample points **V** along the refracted camera ray. The set of corresponding points **P** defines a curve on the surface of the object (see Figure 4a). This curve is irregular and can be made of several disconnected components: it is piecewise continuous.

It is technically possible to construct a scene where the set of points **P** is of dimension 2 rather than 1, defining an area on the surface rather than a curve. This never happened in all our test scenes.

Figure 3, left, shows the radiance reaching the camera for all sample points along the ray, taking into account attenuation, phase function and ray differentials. It is highly irregular, with many spikes of high value, defined on a small interval.

Zooming in on a small part of this curve (Figure 3, right), we observe that it is a combination of slowly varying contributions, with sharp discontinuities. These discontinuities correspond to triangle boundaries. This our key observation: the contribution from a single triangle to the outgoing radiance is a smooth function; discontinuities occur only as the curve enters and exits the triangle.

Several triangles can be contributing to the same sample point on the refracted ray: there is not a one-to-one mapping between the ray and the curve. For all points on the interval on Figure 3, right, there are at least two triangles connecting this point.

Note that the spikes do *not* correspond to an infinite amount of light over an infinitely small sampling area, but to a very large, but finite, amount of light over a small, but measurable, area. Despite the name "volume caustics", these are not, technically, caustics, since they are not points of infinite energy.

The pixel value is the integral of the radiance values at all sample points, the area under the curve on Figure 3, left. Smooth parts of the curve (e.g. from depth = 12 to depth = 20) can contribute more than all the spikes together. But the spikes are making it difficult for sample-point methods: to avoid noise, the number of samples should be inversely proportional to the minimum width of the spikes.

3.2.2 Interval analysis along the ray

Based on our observation, our key idea is to start by identifying the limits of the contribution from each individual triangle, then sample only within these limits. This reduces the number of samples required
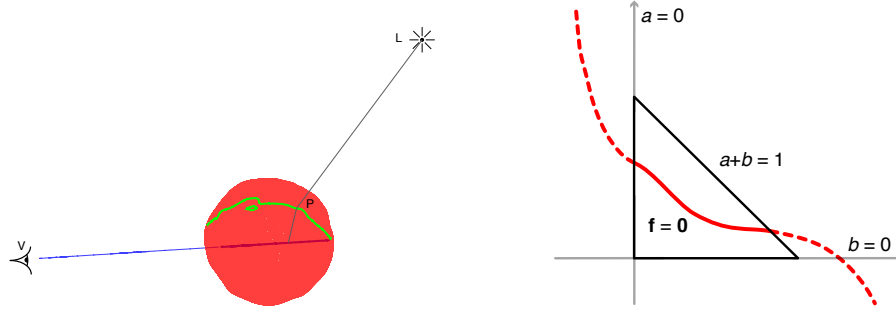
(a) Example of a curve. Notice it is piecewise continuous. (b) For every triangle, we search for intersections between this curve and the triangle in barycentric coordinates.

Figure 4: The set of points **P** corresponding to a refracted camera ray inside the object defines a curve on the surface.

along the ray.

We use the same function **f** defined in equations 1 to 7, but with a third parameter, t, the coordinate of point **V** along the refracted ray. t is limited in range by the entry and exit points, $\mathbf{V}_{\min}$ and $\mathbf{V}_{\max}$, corresponding to $t_{\min}$ and $t_{\max}$ respectively:

$$\mathbf{V} = \mathbf{V}_{\min} + t \frac{\mathbf{V}_{\max} - \mathbf{V}_{\min}}{\|\mathbf{V}_{\max} - \mathbf{V}_{\min}\|} \tag{13}$$

f is now a function of three parameters: t and the barycentric coordinates (a, b) of point **P** on the triangle plane. In the general case, $\mathbf{f} = \mathbf{0}$ is a polynomial equation of degree 6. **V**, **P** and $\mathbf{N}_s$ (before normalization) are linear combinations of the parameters; the equation $\mathbf{f} = \mathbf{0}$ is equivalent to:

$$\eta^2 (\mathbf{L} - \mathbf{P})^2 \|(\mathbf{V} - \mathbf{P}) \times \mathbf{N}_s\|^2 = (\mathbf{V} - \mathbf{P})^2 \|(\mathbf{L} - \mathbf{P}) \times \mathbf{N}_s\|^2 \tag{14}$$

In the plane (a, b), $\mathbf{f} = \mathbf{0}$ defines a curve parameterized by t. We need the intersection between this curve and the triangle defined by: $a \geq 0, b \geq 0, a + b \leq 1$ (see Figure 4b). The coordinates in t of these intersections will give us the support of function **f** for this triangle.

For any normalized vector $\widehat{\mathbf{v}} = \mathbf{v}/\|\mathbf{v}\|$, its derivative with respect to a parameter q is:

$$\frac{\partial \widehat{\mathbf{v}}}{\partial q} = \frac{1}{\|\mathbf{v}\|} \left(\frac{\partial \mathbf{v}}{\partial q} - \left(\widehat{\mathbf{v}} \cdot \frac{\partial \mathbf{v}}{\partial q} \right) \widehat{\mathbf{v}} \right) \tag{15}$$

Repeated applications of this rule give us the Jacobian of **f**:

$$J = \left(\frac{\partial \mathbf{f}}{\partial t} \frac{\partial \mathbf{f}}{\partial a} \frac{\partial \mathbf{f}}{\partial b} \right) \tag{16}$$

Note that ω_L and $\mathbf{N}_S$ are only functions of (a, b), thus their derivative with respect to t is null.

For each triangle edge, we find the intersection between the curve and this edge. This becomes again a two-parameter problem, with t and one barycentric coordinate being the parameters, and the second barycentric coordinate being either fixed to 0 or linked to the first by $a + b = 1$. We use the same technique as before for finding zeros of **f**, with the pseudo-inverse. The only difference is that we start with a 3×3 matrix, then convert it into a 3×2 matrix by removing one column, then compute the pseudo-inverse of this matrix.

Once we have an intersection, we compute the Jacobian of **f** at this point, to compute the local derivative of the curve (see section 3.2.3 for details). This tells us whether the value of t we found is an upper

```
Start with interval [t_min, t_max]
numIntersections = 0;
repeat
    for each triangle edge E
        Compute (t, a, b) intersection between E and curve
        Check if value of t is a min or a max (section 3.2.3)
        Update interval [t_min, t_max] accordingly
        If interval is empty: return 0
        If (a, b) inside triangle: numIntersections++;
    end for
until numIntersections = 2;
// Now, we sample the curve segment
Sum = 0
for t regularly sampled in [t_min, t_max]
    Find (a, b) such that f(t, a, b) = 0
    Sum += contribution from point (a, b)
end for
return Sum
```

Algorithm 1: computing single-scattering effects

or a lower bound, and we update the interval for t accordingly. We treat all boundaries of the triangle, in a loop, until we either have an empty interval for t or a segment where the curve enters and leaves the triangle. If we have an empty interval, we move to the next triangle. If we have a valid segment, we sample it regularly and compute the contribution from each sample point.

Our algorithm is summarized in Algorithm 1. A single loop over the triangle edges is usually sufficient to identify the entry and exit points. The curve occasionally enters and leaves the triangle on the same edge, requiring a second loop.

3.2.3 Tangent to the curve

Assuming we have found a point $(t, a, b)^\mathsf{T}$ on the curve of equation $\mathbf{f} = \mathbf{0}$ on the triangle plane, we want the tangent to the curve at this point.

By definition of the Jacobian,

$$\mathbf{f}(t + \delta t, a + \delta a, b + \delta b) \approx \mathbf{f}(t, a, b) + J \begin{pmatrix} \delta t \\ \delta a \\ \delta b \end{pmatrix} \qquad (17)$$

For a point on the curve, $\mathbf{f}(t, a, b) = 0$. Equation 17 defines the tangent to the curve:

$$J \begin{pmatrix} \delta t \\ \delta a \\ \delta b \end{pmatrix} = \mathbf{0} \qquad (18)$$

The Jacobian J at a point on the curve is of rank at most 2, and its kernel is the tangent to the curve. If J is of rank 2, there are at least two columns in J such that their cross-product is non-null. This cross-product $\mathbf{g}$ is the basis for the kernel of J and the direction of the tangent to the curve (see, for example, Reeder [Ree11]). We write its coordinates as $\mathbf{g} = (g_t, g_a, g_b)^\mathsf{T}$. If J was of rank 1 or 0, then locally the tangent would not be defined; this has never happened in all our test scenes.

$(t, a, b)^\intercal + \lambda\mathbf{g}$ ($\lambda \in \mathbf{R}$) is a straight line, tangent to the curve at point $(t, a, b)^\intercal$. We use it:

- to find better starting points for the other triangle edges. We compute the intersections between the tangent and the other triangle edges, and use them as starting points for the Newton-Raphson methods on these edges. This speeds up the computations.

- to decide whether the point corresponds to an upper or lower bound for t. Variations in t are connected to variations in a (resp. b) by the ratio g_t/g_a (resp. g_t/g_b). The sign of this ratio tells us whether we have a minimum or a maximum, depending on the edge:

    - For $a = 0$, t is a maximum if $g_t/g_a < 0$.
    - For $b = 0$, t is a maximum if $g_t/g_b < 0$.
    - For $a + b = 1$, t is a maximum if $g_t/(g_a + g_b) > 0$.

3.2.4   Pruning triangles and restricting the interval

In theory, we should run the search described in Algorithm 1 for all the triangles on the boundary of the object. This would be impractical, so have we developped two strategies to restrict the search by quickly rejecting triangles:

- One is based on the spindle test, uses only the geometric position of objects; it can be applied to any volume.

- The other is based on the shading normals; it can be applied only to triangles.

Both use the fact that the interval for $\mathbf{V}$ is restricted by the entry and exit points of the refracted camera ray, $[\mathbf{V}_{\min}, \mathbf{V}_{\max}]$.

We use the bounding volume hierarchy (BVH) already present in the renderer, in a hierarchical descent. Starting with the largest node containing the translucent object, we test whether it passes the first test. If yes, we iterate on its descendants. If not, we stop the search. We keep descending recursively until we reach triangles that have passed the first test. We apply to them the second test based on shading normals.

The output of the second test is both a boolean and a validity interval for t. If the boolean is true, we apply the extensive algorithm described in section 3.2.2.
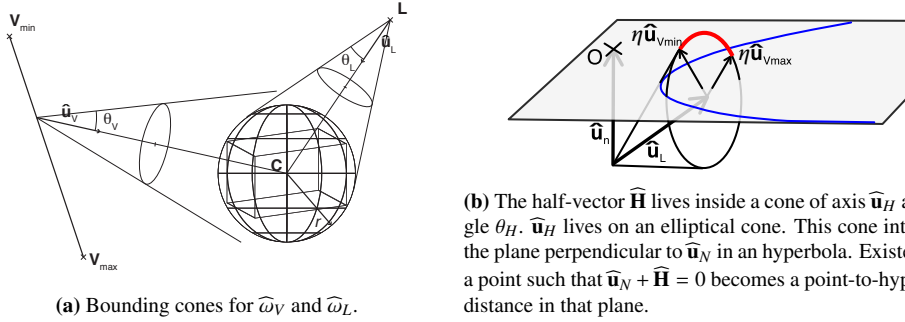
For both tests, we start by building the bounding sphere of the object (whether it is a triangle or a bounding volume), with center $\mathbf{C}$ and radius r. We build bounding cones for $\widehat{\omega}_L$ and $\widehat{\omega}_V$ (see Figure 5a):

- $\widehat{\omega}_L$ is inside a cone of axis $\widehat{\mathbf{u}}_L$ and angle θ_L:

$$
\begin{aligned}
d_{LC} &= \|\mathbf{L} - \mathbf{C}\| \\
\widehat{\mathbf{u}}_L &= (\mathbf{L} - \mathbf{C})/d_{LC} \\
\theta_L &= \arcsin(r/d_{LC})
\end{aligned}
$$

- $\widehat{\omega}_V$ is bounded by a sweeping cone, whose normalized axis moves from $\widehat{\mathbf{u}}_{V\min}$ to $\widehat{\mathbf{u}}_{V\max}$ and whose angle is θ_V:

$$
\begin{aligned}
\widehat{\mathbf{u}}_{V\min} &= \frac{\mathbf{V}_{\min} - \mathbf{C}}{\|\mathbf{V}_{\min} - \mathbf{C}\|} \\
\widehat{\mathbf{u}}_{V\max} &= \frac{\mathbf{V}_{\max} - \mathbf{C}}{\|\mathbf{V}_{\max} - \mathbf{C}\|} \\
\theta_V &= \arcsin(r/d_{\min}) \\
\text{where } d_{\min} &= \min\left(\|\mathbf{V} - \mathbf{C}\|\right) \text{ for } \mathbf{V} \in [\mathbf{V}_{\min}, \mathbf{V}_{\max}]
\end{aligned}
$$

(a) Bounding cones for $\widehat{\omega}_V$ and $\widehat{\omega}_L$.

(b) The half-vector $\widehat{\mathbf{H}}$ lives inside a cone of axis $\widehat{\mathbf{u}}_H$ and angle θ_H. $\widehat{\mathbf{u}}_H$ lives on an elliptical cone. This cone intersects the plane perpendicular to $\widehat{\mathbf{u}}_N$ in an hyperbola. Existence of a point such that $\widehat{\mathbf{u}}_N + \widehat{\mathbf{H}} = 0$ becomes a point-to-hyperbola distance in that plane.

Figure 5: Pruning triangles in the search: we start with the bounding sphere $(\mathbf{C}, r)$ of the current hierarchical level, and use bounding cones to restrict the search.

Spindle-based test   The spindle test tells us that the angle between $\widehat{\omega}_V$ and $\widehat{\omega}_L$ must be between $\frac{\pi}{2} +$ arcsin$(1/\eta)$ and π. Alternatively, the angle between $\widehat{\omega}_V$ and $-\widehat{\omega}_L$ must be smaller than $\frac{\pi}{2} -$ arcsin$(1/\eta)$.

We compute the minimum angle between $-\widehat{\mathbf{u}}_L$ and $\widehat{\mathbf{u}}_V$ and compare it with $\frac{\pi}{2} -$ arcsin$(1/\eta) + \theta_V + \theta_L$. If it is larger, the entire bounding sphere is outside of the spindle, and we can stop the search.

Shading-normal based test   When testing against a triangle, we start by computing a sweeping cone bounding the half-vector $\widehat{\mathbf{H}}$. We also build a cone bounding the shading normal, with axis $\widehat{\mathbf{u}}_N$ and angle θ_N and check for intersection between the cones.

The extremity of the normalized vector $\widehat{\omega}_L$ lies inside a sphere of radius $r_L = 2\sin(\theta_L/2)$, centered on the extremity of $\widehat{\mathbf{u}}_L$ (similarly for $\widehat{\omega}_V$). Thus, the extremity of $\mathbf{H} = \eta\widehat{\omega}_V + \widehat{\omega}_L$ is inside a sphere of radius $\eta r_V + r_L$, centered on a moving vector $\eta\widehat{\mathbf{u}}_V + \widehat{\mathbf{u}}_L$.
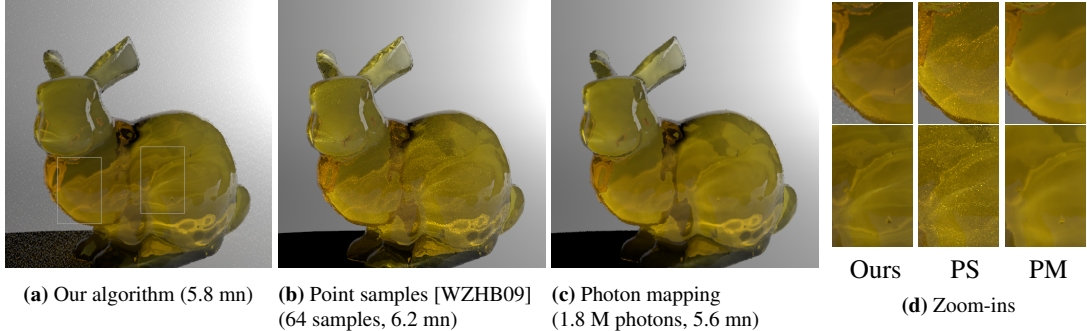
We denote $H_{\min}$ the minimal value of the norm of $\eta\widehat{\mathbf{u}}_V + \widehat{\mathbf{u}}_L$. The normalized half-vector $\widehat{\mathbf{H}}$ lies inside a cone of axis $\widehat{\mathbf{u}}_H$ and angle θ_H:

$$\begin{aligned}
\widehat{\mathbf{u}}_H &= \text{normalize}(\eta\widehat{\mathbf{u}}_V + \widehat{\mathbf{u}}_L) \\
\theta_H &= \arcsin\left((\eta r_V + r_L)/H_{\min}\right)
\end{aligned}$$

The axis $\widehat{\mathbf{u}}_H$ lies on a portion of an elliptic cone defined by $\widehat{\mathbf{u}}_L$ and the circle on which $\eta\widehat{\mathbf{u}}_V$ varies (see Figure 5b). It is possible to have $\widehat{\mathbf{H}} + \mathbf{N}_S = \mathbf{0}$ only if $-\mathbf{u}_H$ falls inside the cone of axis $\widehat{\mathbf{u}}_N$ and angle $\theta_N + \theta_H$.

Rather than having to compute the intersection of these two cones, we move to the plane perpendicular to $\widehat{\mathbf{u}}_N$, defined by $\mathbf{M} \cdot \widehat{\mathbf{u}}_N = 1$. The intersection between this plane and the cone $(\widehat{\mathbf{u}}_N, \theta_N + \theta_H)$ is a circle, of center $\mathbf{O}$ and radius $\tan(\theta_N + \theta_H)$. The intersection between this plane and the elliptic cone carrying $\mathbf{u}_H$ is a hyperbola. The two cones can intersect only if the curves intersect in the plane: if the minimum distance between $\mathbf{O}$ and the hyperbola is larger than $\tan(\theta_N + \theta_H)$ we discard the triangle.

Restricting the interval for t   A hyperbola defined as the intersection between a cone and a plane has two branches. Only one branch is relevant for us, the one corresponding to $\widehat{\mathbf{u}}_N \cdot \widehat{\mathbf{u}}_H < 0$. The point $\mathbf{V}$ such that $\widehat{\mathbf{u}}_N \cdot \widehat{\mathbf{u}}_H = 0$ defines the boundary between the two branches. We can discard the other part of the interval, which reduces the search interval for t.

(a) Our algorithm (5.8 mn)   **(b)** Point samples [WZHB09]   **(c)** Photon mapping
                                 (64 samples, 6.2 mn)   (1.8 M photons, 5.6 mn)

Ours   PS   PM
(d) Zoom-ins

Figure 6: Comparison between our algorithm for single-scattering effects and existing methods, for approximately equal time. Bunny model with 16 301 triangles.

3.3   Computing the integral

Once we have $t_{\min}$ and $t_{\max}$ for a given triangle, we sample this interval regularly in t. For K samples, we take:

$$t_i = t_{\min} + \frac{i + \frac{1}{2}}{K}(t_{\max} - t_{\min}) \tag{19}$$

For each t_i we compute $\mathbf{V}_i$ using equation 13, then $\mathbf{P}_i$ on the surface such that $\mathbf{V}_i\mathbf{P}_i\mathbf{L}$ is a valid path following Snell's law. We shoot a shadow ray from $\mathbf{P}_i$ to $\mathbf{L}$. If the points are visible from each other, we add the contribution associated to that path, C_i, to the contribution of the triangle:

$$C_{\text{triangle}} = \frac{t_{\max} - t_{\min}}{K} \sum_i C_i \tag{20}$$

The number K of samples in each interval is a parameter. In practice, we have used $K = 1$ for all our test scenes with no visible impact on quality.

4   Results and comparison

Unless otherwise specified, pictures and timings in this paper were generated on a quad-core Intel Xeon W3250 at 2.66 GHz with 6 GB of memory, running Windows 7. We used the Mitsuba renderer [Jak10] for our algorithm, photon mapping and bi-directional path-tracing, and the original implementation of Walter et al.'s algorithm [WZHB09].

There is no post-processing in any of the pictures: no clamping, no filtering. We compute single scattering effects, along with transmitted and reflected light, as well as internal reflections.

4.1   Equal time comparison

The key advantage of our algorithm is that it provides high quality pictures of single scattering effects quickly. Figures 1 and 6 show a side-by-side comparison between our algorithm and other algorithms for approximately the same computation time. Other algorithms do not provide the same quality:

- **Point samples**, the algorithm described in [WZHB09], computes exact solution at several points along the refracted camera ray. With equal time computation, there are too little samples, and point artefacts are clearly visible (see Figures 1b and 6b).
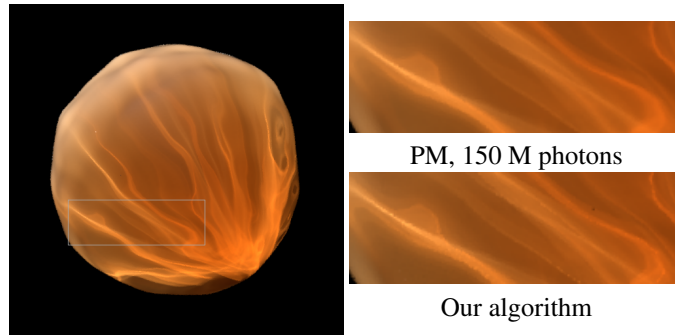
Figure 7: Comparison with a high quality solution: Photon mapping with 150 M photons (3.53 h on a dual Intel Xeon E5345, 2.33 GHz with 48 Gb RAM).

- **Photon mapping** uses the implementation in the Mitsuba renderer, which relies on Beam-Radiance Estimates [JZJ08]. It provides a picture without artefacts, but the individual features characteristic of single scattering have been averaged (see Figures 1c and 6c).

- **Bi-directional path-tracing** also uses the implementation in the Mitsuba renderer. For scattering effects, we connect paths from the light source and paths from the camera inside the medium. For single scattering, the two paths must intersect exactly, which has a zero probability for a point light source. To produce a picture, we replaced it with a spherical light source (see Figure 1d). Picture quality is much lower than the other methods.

4.2 Validation: comparison with high-quality solution

Given the difference between our results and those of other methods, we need confirmation that we are actually computing the right solution. Photon mapping with a large number of photons provides this validation. The computer we used in our tests ran out of memory after 15 million photons. We used a different computer, with 48 GB of memory, to compute a picture with 150 million photons (see Figure 7). This high quality picture is identical to the one we compute in 169 s, up to the thinner details.
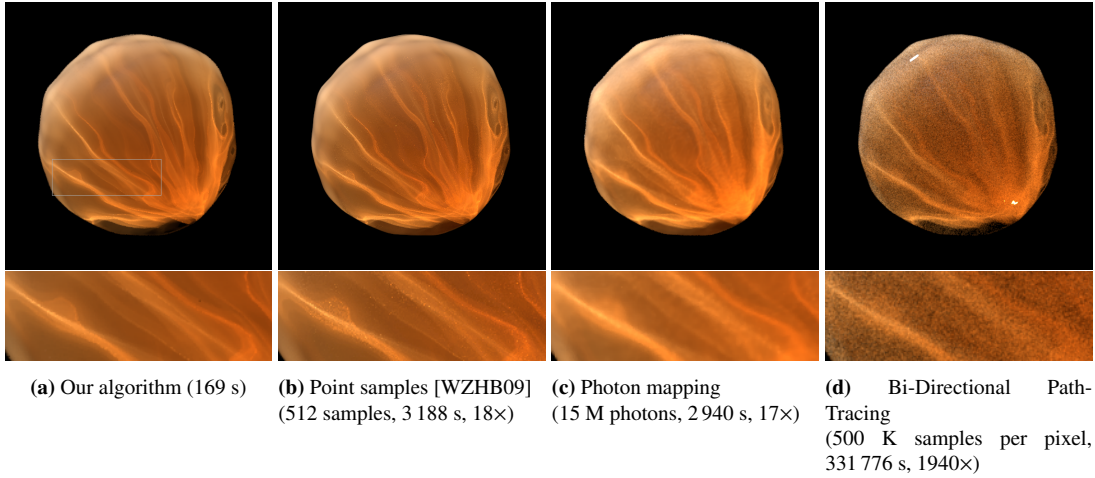
4.3 Equal-quality comparison

Increasing the number of samples or photons increases the quality of the other algorithms, along with their computation time. Figure 8 shows pictures with approximately the same quality. The computation time is at least 17 times larger, and the image quality is still not as good as ours. Note that bi-directional path tracing also converges to the same solution, but much slower than the other methods.

Equal quality comparisons with photon mapping is difficult because of the memory cost: on our testing computer, we cannot run tests with more than 15 million photons. We allocated all of them to the volume photon map, specifically for single scattering. In a more realistic scenario, part of the photons would be used for surface effects and for multiple scattering, reducing the quality of single scattering simulation. On the other hand, photon mapping can be used for rendering more complex scenarios, that are beyond the scope of our method.

A strong advantage of our algorithm is that it has no additional memory cost: we do not use any data structure beyond what is already present in the renderer: the geometry of the object and the bounding volume hierarchy.
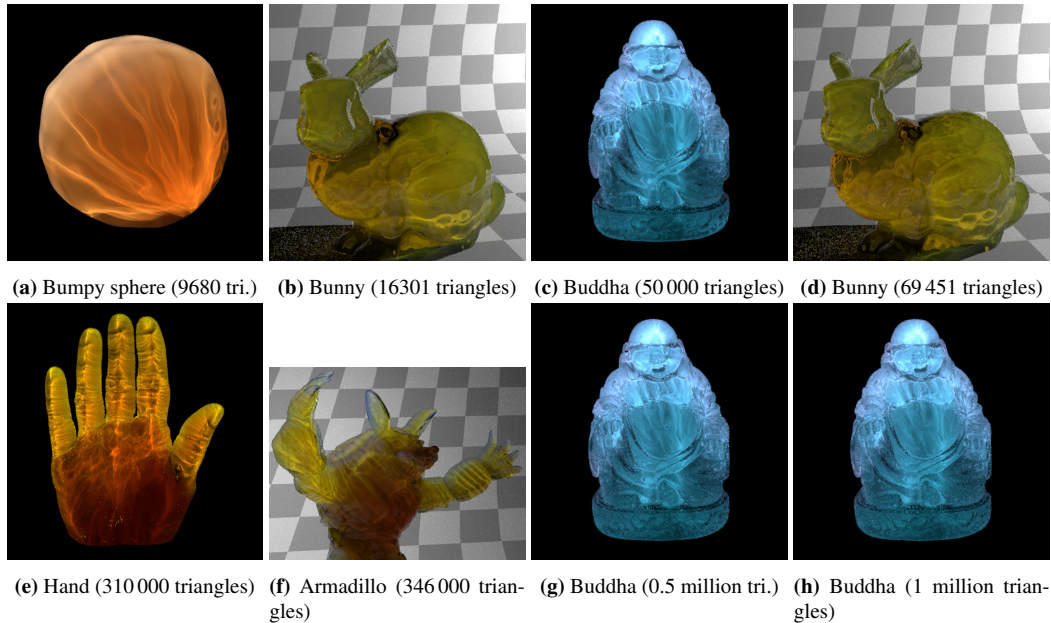
(a) Our algorithm (169 s)   **(b)** Point samples [WZHB09] (512 samples, 3 188 s, 18×)   **(c)** Photon mapping (15 M photons, 2 940 s, 17×)   **(d)** Bi-Directional Path-Tracing (500 K samples per pixel, 331 776 s, 1940×)

Figure 8: *Comparison between our algorithm for single-scattering effects and existing methods, for approximately equal quality (Photon Mapping: we used the maximum number of photons for our computer, 15 millions; Bi-Directional Path-Tracing: we use a sphere instead of a point light source).*

4.4   Computation time

We used five test scenes, for a total of eight different polygon counts: the bumpy sphere from [WZHB09], the Stanford bunny at two different resolutions, a Buddha statue at three different resolutions, the Stanford armadillo and a model of a hand (see Figure 9). Triangle count ranges from 10^4 to 10^6 triangles. For the Buddha and the hand, we show computation times for two different light positions: from the top and from behind. We also tried different light positions for the bumpy sphere and the bunny, but the difference in computation time was within the margin of error. The armadillo picture has a resolution of 683×512 pixels.

Figure 10a displays the average computation time for each camera ray that reached the translucent object. This measure is independent of the screen coverage of the model. We show the computation time with and without shadow rays for each sample point. We show separately the average time for internally reflected rays; 40 % to 55 % of the camera rays entering the object result in internal reflections (except for the bumpy sphere with only 20 %):

- The total computation time is connected to finding the zeros of **f**. Shooting shadow rays accounts for 3 % of computation time. This could be caused, in part, by the simplicity of our test scenes: few objects beyond the main one.

- The influence of light position depends on object shape. For compact objects, such as the sphere, the bunny and the Buddha, it has little influence (at most 10 %). For branching objects, such as the hand and armadillo, it has a strong influence: computation time is several times smaller than for a compact object with the same triangle count if each branch can be treated independently (armadillo, hand illuminated from behind). It is slower than for a comparable compact object when branches interfer with each other (hand from the top).

- Scattering computations for internal reflections can be either slower or faster than for direct rays, depending on light position. The behavior evolves consistently with object resolution. Internal reflections are focusing computations on part of the object, which can be more or less complex than the average.

(a) Bumpy sphere (9680 tri.) **(b)** Bunny (16301 triangles) **(c)** Buddha (50 000 triangles) **(d)** Bunny (69 451 triangles)

(e) Hand (310 000 triangles) **(f)** Armadillo (346 000 triangles) **(g)** Buddha (0.5 million tri.) **(h)** Buddha (1 million triangles)

Figure 9: Our test scenes (by increasing order of complexity).

Figure 10b displays the same results in log-log space. For compact objects and identical lighting conditions, computation time for each sample evolves roughly as $O(n^{3/4})$, where n is the number of polygons. A smooth curve drawn on a spherical object tesselated into n triangles crosses $O(n^{1/2})$ triangles. This is our lower bound. The upper bound is exploring all triangles: $O(n)$. Complexity is half-way between upper and lower bounds.

Comparing Figures 9b and 9d shows the effect of changing object complexity: caustics become more convoluted. The effect is less visible on the Buddha model (Figures 9c, 9g and 9h) probably because the first model is already of high accuracy.

5 Conclusion

We have described a new algorithm for accurate computation of single scattering effects inside a translucent object with refractive boundary. It is an extension of Walter et al.'s algorithm [WZHB09]. Our main idea is to compute the limits of the influence of each triangle over each refracted ray, then sample only inside these limits. Our algorithm proves to be both faster and more accurate than existing methods, with a smaller memory footprint.

Our method does not require any additional memory structure beyond what is already available in a renderer: the geometry of each triangle, and a bounding volume hierarchy. This makes it easier to implement inside existing renderers, but impacts performance: with adjacency information between triangles, we could reuse computations from neighboring triangles for shared edges. This should reduce computation time by 50 %. Using the results from neighboring pixels as a starting point for the Newton-Raphson iteration would also speed up computations. This information would also help for adaptive anti-aliasing.

Our algorithm does not allow a speed-quality tradeoff. In future work we want to introduce this possibility, based on a hierarchical representation of surface details. We also want to port the algorithm to the GPU. Since our computations are highly parallel, we expect significant gains.

(a) Linear representation

(b) Log-log space.
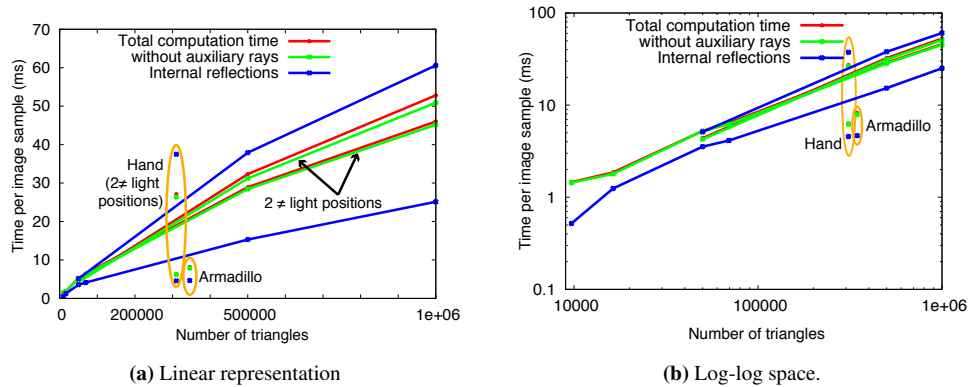
Figure 10: *Average computation time for each camera ray reaching the object, as a function of scene complexity.*

Finally, we want to extend our algorithm to more complex configurations. We are currently limited to two refractive interfaces, and would like to handle more complex cases, with multiple refractions.

Acknowledgments

References

[EAMJ05] ERNST M., AKENINE-MÖLLER T., JENSEN H. W.: Interactive rendering of caustics using interpolated warped volumes. In *Graphics Interface* (2005), pp. 87–96.

[HDI*10] HU W., DONG Z., IHRKE I., GROSCH T., YUAN G., SEIDEL H.-P.: Interactive volume caustics in single-scattering media. In *Symposium on Interactive 3D Graphics and Games* (2010), ACM, pp. 109–117.

[IDN01] IWASAKI K., DOBASHI Y., NISHITA T.: Efficient rendering of optical effects within water using graphics hardware. In *Pacific Conference on Computer Graphics and Applications* (2001), pp. 374–383.

[Ige99] IGEHY H.: Tracing ray differentials. In *SIGGRAPH '99* (1999), ACM, pp. 179–186.

[IZT*07] IHRKE I., ZIEGLER G., TEVS A., THEOBALT C., MAGNOR M., SEIDEL H.-P.: Eikonal rendering: Efficient light transport in refractive objects. *ACM Transactions on Graphics (proc. Siggraph) 26*, 3 (July 2007), 59:1 – 59:9.

[Jak10] JAKOB W.: Mitsuba renderer, 2010. http://www.mitsuba-renderer.org.

[JMLH01] JENSEN H. W., MARSCHNER S., LEVOY M., HANRAHAN P.: A practical model for subsurface light transport. In *Computer Graphics Proceedings, Annual Conference Series (SIGGRAPH 2001)* (August 2001).

[JNSJ11] JAROSZ W., NOWROUZEZAHRAI D., SADEGHI I., JENSEN H. W.: A comprehensive theory of volumetric radiance estimation using photon points and beams. *ACM Transactions on Graphics 30*, 1 (January 2011), 5:1–5:19.

[JNT*11] JAROSZ W., NOWROUZEZAHRAI D., THOMAS R., SLOAN P.-P., ZWICKER M.: Progressive photon beams. *ACM Transactions on Graphics (proc. Siggraph Asia) 30*, 6 (December 2011), 181:1–181:12.

[JZJ08] JAROSZ W., ZWICKER M., JENSEN H. W.: The beam radiance estimate for volumetric photon mapping. *Computer Graphics Forum (Proc. Eurographics 2008) 27*, 2 (2008), 557 – 566.

[NN94] NISHITA T., NAKAMAE E.: Method of displaying optical effects within water using accumulation buffer. In *SIGGRAPH '94* (1994), ACM, pp. 373–379.

[PP09] PEGORARO V., PARKER S. G.: An Analytical Solution to Single Scattering in Homogeneous Participating Media. *Computer Graphics Forum (Proc. Eurographics 2009) 28*, 2 (2009), 329–335.

[Ree11] REEDER M.: The kernel of a three-by-three matrix. `https://www2.bc.edu/~reederma/Linalg15.pdf`, 2011.

[SZLG10] SUN X., ZHOU K., LIN S., GUO B.: Line space gathering for single scattering in large scenes. *ACM Transactions on Graphics (proc. Siggraph 2010) 29*, 4 (July 2010), 54:1–54:8.

[SZS*08] SUN X., ZHOU K., STOLLNITZ E., SHI J., GUO B.: Interactive relighting of dynamic refractive objects. *ACM Transactions on Graphics (proc. Siggraph 2008) 27*, 3 (Aug. 2008), 35:1–35:9.

[WZHB09] WALTER B., ZHAO S., HOLZSCHUCH N., BALA K.: Single scattering in refractive media with triangle mesh boundaries. *ACM Transactions on Graphics (proc. Siggraph 2009) 28*, 3 (July 2009), 92:1–92:8.

A   Implementation details

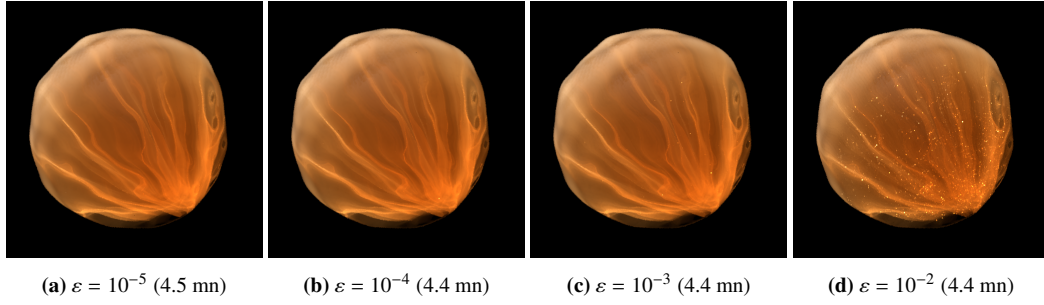In this section, we describe several details about practical implementation of our algorithm.

A.1   Numerical computations and importance of the D factor

The most *difficult* part in the computation is finding the point $\mathbf{P}$ that connects $\mathbf{V}$ and $\mathbf{L}$. But computing D based on ray differentials is the most *important* for an accurate result. Figure 11b shows a picture of the bumpy sphere test scene, with all paths computed, but without the ray differentials. The differences with the reference picture (Figure 11a) are striking. The visual effects associated with single scattering come more from light concentration by the refractive interface than from the number of paths connecting to the sample point.

To get the right value for D, you need to have the right position for $\mathbf{P}$. Even minor inaccuracies over the position of $\mathbf{P}$ have a strong impact on picture quality. Figure 12 shows the result of increasing the ε value used to stop the search for $\mathbf{P}$ in the Newton-Raphson method (if the distance between $\mathbf{P}$ in two consecutive iterations is smaller than ε, we stop the search). Having a larger epsilon value results in inaccurate values for D, and thus noise on the picture. Because Newton-Raphson method has quadratic convergence, decreasing ε has little influence on computation time (timings for this figure were recorded on a Intel Core2 at 2.66 GHz).

(a) Full *D* factor
(based on shading normals).

(b) Without the *D* factor.
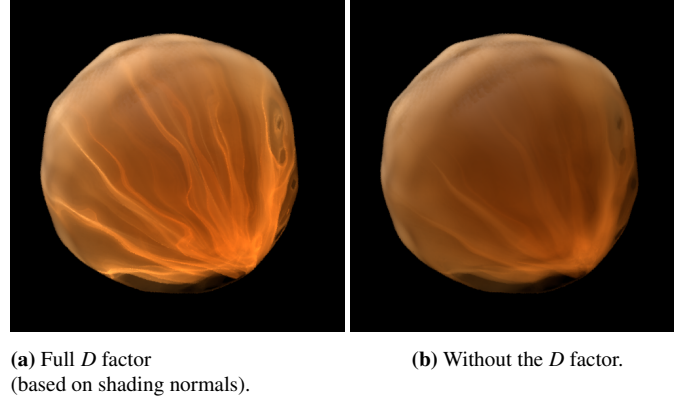
Figure 11: *Importance of the D factor in single scattering computations. Both pictures have computed the same paths connecting to the light source.*
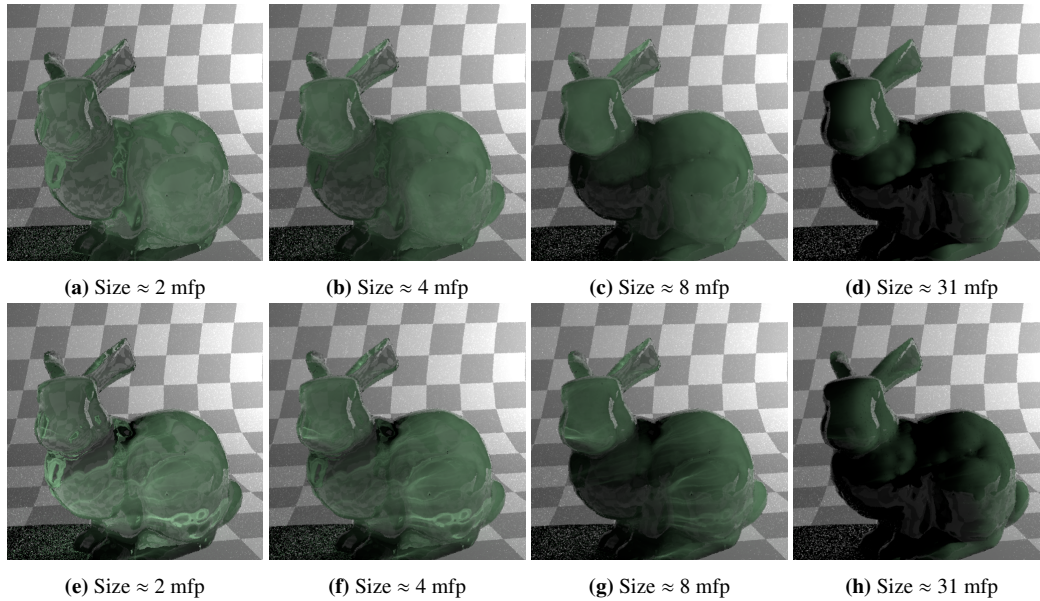

(a) $\varepsilon = 10^{-5}$ (4.5 mn)   **(b)** $\varepsilon = 10^{-4}$ (4.4 mn)   **(c)** $\varepsilon = 10^{-3}$ (4.4 mn)   **(d)** $\varepsilon = 10^{-2}$ (4.4 mn)

Figure 12: *Reducing the accuracy of the computations increases the noise with little benefit on computation time.*

A.2 Point to hyperbola distance

In order to cull an entire triangle from a segment point of view, we have to compute the distance between a point and an hyperbola (see Section 3.2.4 and Figure 5b). More precisely, we have to test whether the distance between point **O** and the hyperbola defined by the intersection of an elliptic cone and a plane is larger than $\tan(\theta_N + \theta_H)$. If yes, we can discard the entire triangle. Our problem is that there are no easy to compute formula to give the distance between a point and a quadric.

We are only interested in a subset of the hyperbola, corresponding to $\widehat{\mathbf{u}}_V$ going from $\widehat{\mathbf{u}}_{V\min}$ to $\widehat{\mathbf{u}}_{V\min}$, and on a single branch. The two extremities define a segment on the hyperbola, $[\mathbf{H}_{\min}, \mathbf{H}_{\max}]$. Because each branch of the hyperbola is convex, this segment is included inside the triangle defined by the line joining the two extremities and the tangents to the hyperbola at $\mathbf{H}_{\min}$ and $\mathbf{H}_{\max}$.

The distance between point **O** and this triangle is always larger than the distance between **O** and the hyperbola. If it is larger than the threshold we reject the triangle. If it is smaller, we proceed iteratively: subdivide the segment on the hyperbola in two halves, compute a bounding triangle for each, then the distance between **O** and each triangle. If it is larger for both, we can reject the triangle. Otherwise, we keep subdividing until we reach a minimum size for our hyperbola segments.

(a) Size ≈ 2 mfp	**(b)** Size ≈ 4 mfp	**(c)** Size ≈ 8 mfp	**(d)** Size ≈ 31 mfp
(e) Size ≈ 2 mfp	**(f)** Size ≈ 4 mfp	**(g)** Size ≈ 8 mfp	**(h)** Size ≈ 31 mfp

Figure 13: Influence of the scale parameter on single scattering effects. Top row: straight line approximation (9 s). Bottom row: our algorithm (5.6 mn). Figures include transparency and reflections.

A.3   Comparison with straight line approximation

Many implementations of scattering effects in participating media use the dipole approximation [JMLH01]. This approximation is usually combined with a straight line approximation for single scattering. The dipole approximation is valid when the object dimensions are much larger than the mean free path of the material. Our method, by contrast, gives more interesting results when object dimensions are close to the mean free path.

Figure 13 of single scattering effects on a translucent bunny, while changing the mean free path of the material (all other parameters, including albedo, remain constant). The straight line approximation is close to the reference (our method) when object size is larger than 30 mfp. When the ratio of object size over mean free path is smaller, differences between the straight line approximation and our method become visible.

The position of the volume caustics inside the object does not change as we change the object parameters, as they only depend on the object geometry. Their intensity changes, as light attenuation becomes stronger.