A FAST INTERPOLATION METHOD TO REPRESENT THE BRDF IN GLOBAL ILLUMINATION

Pierre Y. Chatelier  
LAIC, Université d’Auvergne, Aubière, France

Rémy Malgouyres  
LAIC, Université d’Auvergne, Aubière, France

Keywords: Global illumination, BRDF, discrete representation, spherical harmonics, control-points, interpolation, rotation.

Abstract: Global illumination that simulates realistic lighting environments makes use of bidirectional reflectance distribution functions (BRDF). Such a function is a surface property, describing how light is re-emitted after hitting this surface. The present paper details a representation of BRDF and radiance distributions, based on control points, that aims at performance. It uses spherical triangles for interpolation, least square mean or linear programming for optimal representation, and very fast rotation by precomputing control points re-weightening. It has some more advantages for deforming an existing shape. This approach is compared to Spherical Harmonics, and outperforms this last one.

1 INTRODUCTION

Global illumination that simulates realistic lighting environments must be able to describe how light is re-emitted after hitting the surface of an object, depending on the optical properties of this surface: the material properties. The Lambertian model is one of the simplest: a fraction of light (depending on the zenithal incidence) is isotropically re-emitted above the surface. This is also known as the ideal diffuse case. In this case, the material property is embedded in a single factor, as a real number.

However, most objects, like mirrors, velvet, wet surfaces... should not use such a simple model. First, asymmetric materials like velvet must take in account not only the zenithal angle of an incoming ray, but also the azimuthal orientation of the ray according to the surface. Second, incoming light is anisotropically re-emitted. This is particularly obvious for a perfect mirror, which is said to be a perfect specular reflector. Implementing a material mixing up diffuse and specular properties implies a way to represent spatial anisotropy.

The notion of Bidirectional Reflectance Distribution Function (BRDF), and radiance distribution, is the common way to describe such a behaviour. This paper describes the well known problem of representing and using BRDFs (Neumann, 2001; Rusinkiewicz, 1997; Rusinkiewicz, 1998). The required operations for using them in a global illumination solver are presented, and we show how they are addressed with spherical harmonics. Then, we introduce a fast interpolation method, designed to perform these operations with maximal performance and a reasonable precision. Computational experiments have been performed to compare this method with the spherical harmonics.

2 USING A BRDF

2.1 Representation

A BRDF function is a function of two parameters: the direction of incoming light, and the direction of the outgoing ray for which we want to know the re-emission. Each direction is in fact a two-dimensional parameter, usually denoted by the two spherical angles $\theta$ and $\phi$, representing elevation and azimuth (see Fig. 1), so that a BRDF is rather seen as a 4-dimensional function. The returned value is the ratio,
of the reflected (or transmitted) radiance in the given outgoing direction, to the incoming energy flux from the incoming direction. The BRDF of an ideal-diffuse surface is constant, while an ideal reflector can be represented by a Dirac delta function. A simple BRDF may be stored as a mathematical function, but arbitrary BRDFs usually require a set of sample points that have to be interpolated. Many representations for BRDFs have been proposed in the computer graphics literature (Ward, 1992; Ashikhmin and Shirley, 2000; Kautz et al., 2002), generally based on a decomposition on a basis of spherical functions, like spherical harmonics or spherical wavelets (Schröder and Sweldens, 1995; Koenderink et al., 1996).

2.2 Models and Instanciations: Brdf and Radiance Distributions

By stating that we are using a BRDF to modelize light re-emission, we are only describing how we handle this material property. However, one must not forget that this material property is static and is different from the effective emission that appears in each element that constitutes the surface. Indeed, considering the 3D scene for which we are computing global illumination, it is usually discretized into several patches, each patch behaving like a light emitter. Patches from the same surface will share the same BRDF, that describes their behaviour. But the effective emission varies from patch to patch, according to the quantity of energy that each one is re-emitting, and that has been computed by the global illumination algorithm. This effective emission is the radiance distribution, which is a function of one parameter: the outgoing direction (see Fig. 2). Here again, it can be seen as a two-dimensional spherical function on $\theta$ and $\phi$. This function is not known in advance; it is what is computed by the global illumination algorithm. At first, only the light sources have non-zero radiance distributions. Then, progressively, light propagates and hits some patches, that affects their radiance distributions according to their local BRDF. Solving the global illumination problem means that BRDF are known 4-dimensional functions, while radiance distributions are the unknowns, 2-dimensional functions.

For the sake of performance, these two kind of objects are usually represented with the same data structures. Instead of encoding a 4-dimensional function, the BRDF can be implemented as the interpolation of 2-dimensional functions. That is to say, $f(\theta_{in}, \phi_{in}, \theta_{out}, \phi_{out})$ is computed from some $g_{\theta_{in}, \phi_{in}}(\theta_{out}, \phi_{out})$. Hence, both BRDF and radiance distributions can be represented by 2D spherical functions, and we only need a way to encode efficiently such data.

2.3 Operations

As it is stated in Section 2.2, managing radiance distributions requires to encode 2D spherical functions. In order to use such data, we must ensure some operations to be easy to perform on the representation. These operations are querying value, scaling, adding, and rotating the distributions. Since they are to be performed very often, they must be very efficient.

2.3.1 Querying Value

The radiance distribution of a scene element evolves while this element receives light during global-illumination iterative solving. We must be able to query at any time the light leaving an element in a particular direction. This can be done using a mathematical model, which returns a value for some parameters. This can also be done with interpolation, which tries to guess a value according to some neighborhood. The Spherical Harmonics mathematical model is described in Section 2.4. We propose in Section 3 of this paper an interpolating method.

2.3.2 Scaling and Adding

When some light hits a scene element, its local BRDF is queried for the matching radiance distribution. Then this distribution is scaled by the effective incoming energy quantity, and added to the radiance distribution already held by the element. See Figure 3 for an illustration.

2.3.3 Rotating

Radiance distributions returned by BRDF queries are computed in a global system of coordinates. They have to be rotated to be added to a scene element, which uses a local system. Indeed, $\theta$ is related to the
normal vector of the surface, and the azimuthal angle is meaningful if the surface has an orientation.

2.4 Common Solution

A common method to fulfill all previous requirements is to use a decomposition of the radiance distribution function on a basis of spherical functions. The basis being known, the radiance distribution is just encoded by a set of coefficients. Adding and scaling distributions becomes as easy as adding and scaling the coefficients. Rotation is usually the stumbling block.

2.4.1 Spherical Harmonics

Spherical harmonics are a basis of functions that suits very well at representing spherical 2D functions (François X. Sillion and Claude Puech, 1994; Kautz et al., 2002). The spherical harmonics basis is denoted by \( Y_{l,m} \) and a function \( f \) is decomposed with \( C_{l,m} \) coefficients as

\[
f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} C_{l,m} Y_{l,m}(\theta, \phi)
\]

The definition of \( Y_{l,m} \) functions makes use of associated Legendre Polynomials \( P_{l,m}(x) \) and a normalizing constant \( N_{l,m}(x) \).

\[
N_{l,m} = \sqrt{\frac{2l+1}{2\pi} \frac{(l-|m|)!}{(l+|m|)!}}
\]

\[
P_{l,m}(x) = \begin{cases} 
(1-2m)\sqrt{1-x^2}P_{l-1,m-1}(x) & (l = m) \\
\frac{2m+1}{2l+2}P_{l,m} - \frac{1}{l-m}P_{l-2,m}(x) & (\text{otherw.}) 
\end{cases}
\]

2.4.2 Cost of the Representation

The representation with spherical harmonics requires to decide how many coefficients are kept for the decomposition. Since BRDFs are usually rather smooth, only a few dozens coefficients will be needed. It is important to keep the number of coefficients small, because since each scene element will hold a whole set of coefficients, memory becomes rapidly a critical resource. A good balance must be found between the expected precision and the required number of coefficients (see Fig. 4). For a real function, only real coefficients are needed.

Ideal specular reflectors are typical examples of failure of the assertion about smooth BRDFs. However, such reflectance should not be handled by a very high frequency distribution, but algorithmically (Sillion et al., 1991).

2.4.3 Rotation with Spherical Harmonics

The rotation with spherical harmonics is known to be the stumbling block of this representation (Krivánek et al., 2005). It is proved to be a linear operation (Öhrström; Blanco et al., 1997), by multiplying the vector of coefficients by a particular matrix.
However, such a matrix may be unpractical in the circumstances (Erdelyi, 1953).

3 CONTROL-POINTS BASED BRDF

In order to achieve maximal performance, we propose a method based on interpolation on spherical triangles. This method requires a reasonable amount of coefficients, is precise enough for smooth radiance distributions, achieves very good performance for common operations, and suits very well for a direction-based global illumination solver.

3.1 Interpolation

This simplest way to represent a spherical 2D function is to sample it on some points. If the polar angles of these points are fixed, the representation only needs to store the radius component of the spherical coordinates of these points. Then, two difficulties must be addressed: choose the set of control points, and choose an interpolation algorithm.

3.1.1 Choosing the Control Points

The control points must be chosen regularly on the sphere, so that no region is arbitrarily finer than another. Once dispatched on the sphere, the control points are radially translated to represent spherical shapes. A good method to achieve this repartition is to recursively subdivide a regular polyhedron (Schröder and Sweldens, 1995). For instance, starting from an icosahedron (12 vertices), one can take the middle of each edge to divide each face into four new triangles. The initial polyhedron is an important parameter for the number of control points you wish to have. With an icosahedron, the number of points grows rapidly. At orders 0, 1, 2, 3 of subdivisions, this number is respectively 12, 42, 162, 642 (Fig. 5 and 6).

As stated in section 2.4.2, each scene element holds a whole set of coefficients, so that its cardinal should be kept as small as possible, while being big enough to allow precise radiance distributions. Our experiments has shown that 162 points are a good compromise (Fig 7). On these pictures, the dots are not the control points, but some samples interpolated by our method, using the control points.

3.1.2 Weightening

Let \( P \) be a point to be interpolated. If we project \( P \), and the control points, on the unit sphere, then the projection \( P' \) of \( P \) falls into a spherical triangle made by three of the projected control points on the sphere. \( P' \) can also lie on an edge or a vertex in degenerated cases, but this is not a problem.

Instead of using classical bi-linear interpolation, that is useful for rasterized triangles in computer graphics, we can instead weighten the vertices by the ratio of the triangles areas defined by \( P' \) and the control points of the neighborhood. These triangles are spherical triangles and their area is easy to compute with respect to the formula:

\[
\text{area}(ABC) = \hat{A} + \hat{B} + \hat{C} - \pi \quad (2)
\]

where \( \hat{A}, \hat{B}, \hat{C} \) are the angles at the vertices.

Let \( P \) be the point to be interpolated, its projection \( P' \)
falling into the spherical triangle defined by the projections $C_1^*, C_2^*, C_3^*$ of the control points $C_1, C_2$ and $C_3$. We denote the interpolation by writing
\[ P = \omega_1 C_1 + \omega_2 C_2 + \omega_3 C_3 \] (3)
the unknowns being the weights $\omega_i$. These weights can be computed according to the areas of the spherical triangles formed by the projections. Let these triangles be denoted by $P'C_1^* C_2^*$, $P'C_1^* C_3^*$, $P'C_2^* C_3^*$, their areas being respectively $\alpha_1, \alpha_2, \alpha_3$.

We can set
\[ \omega_i = \frac{\alpha_i}{\alpha_1 + \alpha_2 + \alpha_3} \] (4)
to get a normalized weightening system (Figure 8a).

### 3.1.3 Finding the Triangle

To determine the three control points to be used in interpolation, we have to quickly compute in which triangle a point is falling. If we are given the polar angles of this point, it is easy to determine some possible triangles where it is projected. This only requires the control points to be sorted by parallels and meridians. Once we have found the two parallels enclosing the projected point, we can determine which neighbour control points on these parallels surround most of the time, more than one triangle are candidate, but some cross products can then determine easily which is the good one (Figure 8b).

![Figure 8](image_url)

**Figure 8:** (a) Interpolation on spherical triangles. The weights $\omega_i$ are determined from the areas $\alpha_i$. (b) The parallels and meridians of the control points help finding candidates holding the projection.

### 3.2 Representing an Existing Function

We explained in Section 2.2 that even 4D-BRDFs are modeled using 2D-spherical functions, unifying BRDF and radiance distributions representations. Therefore, we must provide an efficient way to approximate a given 2D function, by finding an optimal set of control points regarding the interpolation algorithm.

To find such an optimal set, we can solve the problem using Least-Squares Method, or Linear Programming. Let $f$ be the 2D spherical function to interpolate, and $P_i(r_i, \theta_i, \phi_i), i \in [0; n]$ the samples of this function. Let $C_i(r_i, \theta_i, \phi_i), i \in [0; m]$ be the control points ($n > m$ is expected). The unknowns are the $r_i$. Our interpolation method states that $P_i$ is interpolated by at most three control points $C_1, C_2$ and $C_3$. We denote that by $r_i = \omega_1 r_1 + \omega_2 r_2 + \omega_3 r_3$.

Let $X$ be the unknown vector $(r_i)_{i \in [0; m]}$ of radial coefficients of the control points $C_i$. Let $Y$ be the vector $(r_i)_{i \in [0; n]}$ of radial coefficients of the samples $P_i$. Let $A$ be the matrix of the weights of interpolation. Each line of that matrix is the interpolation of one sample, and holds at most three non-null weights. The system to solve is $AX = Y$. Usually, this system is over-determined since we get as many samples as we want, while the number of control points remains limited. It has likely no solution and only an approximation is needed.

#### 3.2.1 Least-Squares Method

To solve this system, we can use the Least-Squares Method and set $X = (A^t A)^{-1} Y$. This has shown very good results on smooth functions.

#### 3.2.2 Linear Programming

A common problem of the least-squares method is that there may be a high local error as long as the error is low elsewhere. In (Marzais et al., 2006), this problem is addressed using linear programming. By minimizing the infinite norm of the error, this method tries to output results with a locally limited error. For our problem, that linear programming method consists in solving Equation 5 to minimize the infinite norm of the error.

\[ -\|h\| \leq \left( (AX - Y)_i \right)_{i \in [0; n]} \leq \|h\| \] (5)

$h$ is the error that is to be minimized

However, better results are obtained in our case by minimizing the infinite norm of the relative error (Equation 6). This causes small scale areas to be less perturbed, at the cost of less precise big scale areas.

\[ -\|h\| \leq \left( \frac{(AX - Y)_i}{Y_i} \right)_{i \in [0; n]} \leq \|h\| \] (6)

As a compromise, we obtained best results by dividing with the square root of the value (Equation 7).

\[ -\|h\| \leq \left( \frac{(AX - Y)_i}{\sqrt{Y_i}} \right)_{i \in [0; n]} \leq \|h\| \] (7)

Indeed, as it can be seen on Figure 7 (c), some areas are very well approximated, while others are less precise. If the maximal error is found around the peaks, it may be a small error compared to the peak scale, but a big error compared to the small values of the shape.
Thus, if the error cannot be less than \(\varepsilon\), this may cause perturbations in the range \([-\varepsilon; \varepsilon]\). This would spoil small scale areas for which \(\varepsilon\) is not a small value. By taking the relative error, each area is constrained to be as good as possible. This is illustrated in Fig. 9, with a large peak as shape.

Figure 9: (a) Linear solving with absolute error may cause small scale areas to be perturbed. (b) Linear solving with relative error improves small scale areas but is less precise on big scale ones. (c) A compromise can be found with the relative error by dividing by the square root of the value instead of the value itself.

After many tests, the least-squares method kept performing a little better than the linear programming method. Both methods are easy to implement and may be of interest for complex shapes.

### 3.3 Deforming Existing Shapes

The set of coefficients characterizing the decomposition of a function, on a basis of functions, cannot be easily interpreted. Most of the time, the first coefficients characterize the shape globally, and additional coefficients introduce local details; but deforming the shape might have an impact on every coefficient. Therefore, it is not easy to deform an existing shape, just by acting on the coefficients, since the consequences are difficult to foresee.

Our interpolation method has the particularity that any point is characterized by at most three coefficients that are very easy to determine. Pulling or pushing a point away from the center of the shape has an effect only on these three coefficients. To change the radius of a point of an existing shape, it is only necessary to recompute the weights of the three associated control points. Let us consider \(P(r, \theta, \phi)\) interpolated by \(C_1(r_1, \theta_1, \phi_1), C_2(r_2, \theta_2, \phi_2)\) and \(C_3(r_3, \theta_3, \phi_3)\). This is denoted by \(r = \omega_1r_1 + \omega_2r_2 + \omega_3r_3\).

Let \(P' = (r + \delta r, \theta, \phi)\) be the new point \(P\). Then we must find \(\delta r_1\), \(\delta r_2\) and \(\delta r_3\) such that \(r + \delta r = \omega_1(r_1 + \delta r_1) + \omega_2(r_2 + \delta r_2) + \omega_3(r_3 + \delta r_3)\).

A trivial solution is \(\delta r = \delta r_1 = \delta r_2 = \delta r_3\), since

\[
\begin{align*}
\omega_1 (r_1 + \delta r) + \omega_2 (r_2 + \delta r) + \omega_3 (r_3 + \delta r) &= \\
\omega_1 r_1 + \omega_2 r_2 + \omega_3 r_3 + \delta r (\omega_1 + \omega_2 + \omega_3) &= \\
&=
\end{align*}
\]

However, it is a bad solution since it is just a shift of the spherical triangle in which \(P\) is falling. The risk is to perturbate other interpolated point too much. A better approach is to weighten \(\delta r\) by \(\omega_i\), so that a control point is shifted according to its importance in the interpolation of \(P\).

For instance, if we require \(\delta r_i\) to be proportional to \(\omega_i r_i\), we just have to divide by \(\sum \omega_j\) to normalise. Hence, the solution we chose is:

\[
\delta r_i = \frac{\omega_i}{\sum \omega_j} \delta r \quad (8)
\]

### 3.4 Rotation

To perform an efficient interpolation, our method does not allow the angles of the control points to change. Therefore, the rotation cannot be applied to the control points. It is possible, but too expensive, to recompute the control points by resampling the function, rotating the samples, and reapplying Least-Squares Method (or Linear Programming) to get the new control points. The rotation being a critical function, it should be more efficient than that.

#### 3.4.1 Permutation Approach

In our control points system, a rotation can be approximated by a permutation of the control points, by assigning to each virtually rotated control point, one (another) of the (non-rotated) control points, most likely the closest (see Figure 10).

Such a process is very fast since the rotation is then reduced to a set of pre-computed permutations for relevant angles. The drawback is that the rotation is no more a continuous process, can be transformed into the Identity for small angles, and that much error can be accumulated in a sequence of rotations.

To compute the permutation associated to a rotation, a naive method would be to match each virtually rotated control point to the closest control point, starting from the best matching (with minimal error) and ending with the worst one. The problem is that it can
end with catastrophic matchings since less and less control points remain free while the matchings occur (see Figure 11). To overcome the problem, another strategy can be adopted: first, the best matching is realized with a control point $C$. Then each matching is done the best possible, starting from the point which is the closest to $C$ and ending with the furthest one. That prevents a point to become isolated, with all its neighbours having realized a match (monopolizing the control points around). This would be the cause of a poor matching.

3.4.2 A Continuous Approach

The best representation of the rotated shape would be found by resampling it, and perform LSM again to solve the constrained system. It is not affordable, but makes it obvious that the weights of the control points should be reconsidered rather than merely permuted. As a matter of fact, it is rather easy to perform, by interpolating the virtually rotated control points on the original, non-rotated, control points. Each control point, when rotated, can be seen as a sample, that is to be interpolated by three “real” control points. Thus, a “real” control point can be involved in the interpolation of many “virtually rotated” control points. It is given one weight per interpolation. Then, its new radial distance is set according to a weighted sum of the interpolations for which it is participating. Its new position is a compromise and does not necessarily satisfy each interpolation, but this simplified processus gives us a simple, continuous rotation method.

3.5 Direction-Driven Global Illumination Solver

Global illumination algorithms usually use patches as scene elements, and gathers for each patch the quantity of light received by other patches visible from there. By iterating this algorithm, light propagates in the scene and converges towards a radiosity solution. Another approach presented in (Chatelier and Malgouyres, 2006) uses a special method to compute visibility between volumetric scene elements, and the radiosity algorithm is then parametrized by a set of directions in space to consider. Light is transported only on these directions. This comes from the fact that visibility computations are factorized along some rays (Figure 12).

A consequence of this direction-driven radiosity solver is that it is also possible to factorize computations that are done along a ray. Indeed, the interpolation method described in the present paper computes weights for a given direction $(\theta, \phi)$. Combined with the direction-driven solver, this computation can be factorized for all the scene elements that lie along a ray. It results in an additional optimization that our method brings to global illumination solving.

This is not a marginal optimization. We have also used our BRDF representation in the radiative transfer problem (Lenoble, 1993); in this problem, a cloud is illuminated and the energy propagations occur according to some distributions that are called phase functions. These phase functions can be represented by 2D spherical functions. Only one global coordinate system is used, since the volumic elements that are representing the cloud are not oriented. So, no rotation is needed. In this context, if $N$ voxels are encoding the cloud, each one is queried for its radiance distribution in the current direction $(\theta, \phi)$. Then, the computation of the weights for $(\theta, \phi)$ can be factorized for the $N$ voxels.

4 EXPERIMENTAL RESULTS

To compare our implementation of BRDFs using control points, with an implementation using Spherical Harmonics, we have used our radiosity solver (Chatelier and Malgouyres, 2006). The spherical harmonics relies on the SpharmonicKit (Kostele and Rockmore, 2004), which performs very well, so that the performance gap should not be a defect of our implementation of the spherical harmonics. Figure 13 shows that our method largely outperforms the spherical harmonics approach. Moreover, we did not implement the rotation of the spherical harmonics.
harmonics, so that it is not even slowed down by their known bottleneck. The scene that has been used is a very simple assemblage of cubes, with diffuse BRDFs. It has been discretized into 13000 voxels for the solver.

Control Points (162 coeffs) Spherical Harmonics (64 coeffs)

13000 voxels, 5 itérations

Figure 13: The performance of the control points method (plain stroke) is far better than the spherical harmonics.

(a) (b)

Figure 14: (a) Simple scene. (b) Raw data per voxel after radiosity solving.

REFERENCES


