Volume MLS Ray Casting

The Harvard community has made this article openly available. **Please share** how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Published Version</td>
<td>doi:10.1109/TVCG.2008.186</td>
</tr>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:4100249">http://nrs.harvard.edu/urn-3:HUL.InstRepos:4100249</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP</a></td>
</tr>
</tbody>
</table>
Volume MLS Ray Casting

Christian Ledergerber, Gaël Guennebaud, Miriah Meyer, Moritz Bächer and Hanspeter Pfister, Senior Member, IEEE

Abstract— The method of Moving Least Squares (MLS) is a popular framework for reconstructing continuous functions from scattered data due to its rich mathematical properties and well-understood theoretical foundations. This paper applies MLS to volume rendering, providing a unified mathematical framework for ray casting of scalar data stored over regular as well as irregular grids. We use the MLS reconstruction to render smooth isosurfaces and to compute accurate derivatives for high-quality shading effects. We also present a novel, adaptive preintegration scheme to improve the efficiency of the ray casting algorithm by reducing the overall number of function evaluations, and an efficient implementation of our framework exploiting modern graphics hardware. The resulting system enables high-quality volume integration and shaded isosurface rendering for regular and irregular volume data.

Index Terms— Volume Visualization, Unstructured Grids, Moving Least Squares Reconstruction, Adaptive Integration

1 INTRODUCTION

Volume visualization has become an integral component of the scientific discovery process and the simulation pipeline. Data coming from scanning devices such as MRI or ultrasound machines, as well as results from simulations in fields like computational fluid dynamics, geophysics, and biomedical computing, are often represented volumetrically. These volumes, however, can be defined over regular or irregular grids, giving rise to several distinct classes of rendering algorithms and a plethora of implementations.

In this paper we propose a unified framework for generating high-quality visualizations of arbitrary volumetric datasets using the method of Moving Least Squares (MLS). MLS is a popular scheme for reconstructing continuous functions from scattered data due to its well understood mathematical foundations. Furthermore, the MLS framework contains several distinct components that provide a high level of control over the function reconstruction, and allows for a tunable amount of data smoothing as well as interpolation.

Using MLS, we reconstruct continuous functions from data stored on regular as well as irregular grids. We introduce volume MLS ray casting to generate high-quality images with volume integration and isosurfaces, such as that shown in Figure 1.

The volume MLS framework enables computing analytic derivatives for high-quality isosurface shading with specularities. To decrease the over-all number of MLS function evaluations we propose a novel adaptive preintegration technique.

The main contributions of this work are: a unified MLS framework for reconstructing continuous functions and their derivatives from both regular grids and unstructured volume data; a novel adaptive preintegration scheme for ray casting continuous functions with transfer functions that include isosurfaces; and a system that combines high-quality isosurface rendering and volume integration. We provide details on the theoretical foundations of the MLS framework, along with an exploration of the various tunable parameters. We also present an implementation of volume MLS ray casting on modern graphics processor units (GPUs), provide details of our system implementation, and show results on several regular and irregular volume data sets.

2 PREVIOUS WORK

Direct Volume Rendering: The generation of high-quality images of volume data relies on methods that accurately solve the volume rendering integral [28], which models a volume as a medium that can emit, transmit, and absorb light. The standard method for solving this integral is ray casting, and was proposed by Levoy [21] for regular grids, and by Garrity [14] for irregular grids, specifically tetrahedral meshes. Splatting, an alternative to ray casting, was proposed by Westover [49] for regular grids, and has been extended to irregular grids [25, 51]. In either case, a continuous function must be reconstructed from the discrete volume data.

While there are well-established mathematical frameworks for reconstruction of continuous functions from regular volume data [10, 18, 27, 9], irregular volumes are more challenging as there is no implied spatial structure to the data points. Ray casting of tetrahedral mesh allows a direct interpolation of the data at ray-triangle intersections using homogeneous coordinates [14]. For curvilinear grids, interpolation is more complicated [45] and is typically performed at the intersection points with cell faces [6] or in computational space [13]. Another popular algorithm for rendering tetrahedral meshes is the cell projection method [42], in which tetrahedral cells are projected in back to front order onto the image plane [5, 11]. However, a correct depth-ordering of cells does not always exist, and great care must be taken to avoid visual artifacts [50]. Finally, another class of algorithms uses a sweep-plane through the volume followed by rendering of 2D cells [15, 43].

In the case where no explicit structure of the data is given, scattered data reconstruction schemes are applied. Arguably the most prominent method among these are radial basis functions (RBFs) that were introduced to volume rendering by Jang et al. [17]. While RBFs can be evaluated and rendered efficiently [31], the method requires a costly initial solution to a global system of equations, the size of which is tied to the number of data points. Other methods proposed in the context of volume visualization are inverse distance weighting [35], which is a special case of MLS [41], and discrete Sibson interpolation [36]. This latter method inherently relies on resampling the data onto a regular grid. In contrast to these approaches, volume MLS ray casting provides a unifying mathematical framework for the reconstruction of arbitrarily sampled volume data that allows the user to control the reconstruction and data smoothing in a principled manner.

MLS Function Reconstruction: MLS was introduced almost half a century ago by Shepard [41], and more recently shown by Levin as a powerful surface approximation technique [20]. This technique has been embraced by the computer graphics community for reconstructing point set surfaces [3, 4], where a number of extensions have been proposed, such as reconstructing surfaces with sharp features [12, 23] and using anisotropic support regions [1]. Unlike RBFs, MLS does not require a solution to a global system of equations.
Volume Integration: During ray casting, a transfer function (TF), which maps scalar values and (optionally) gradients to colors and opacities, is applied to the volume function. These color and opacity properties of the volume are integrated along rays to obtain the final pixel colors. The TF must be smooth in order to compute the numerical integral efficiently and accurately as no general, closed-form solutions are available. For the case of a smooth TF, well-known adaptive numerical schemes can be employed [32], and theory from Monte Carlo integration can guide their design [7]. Also, interval arithmetic has been proposed to make the above approaches robust [33].

In practice, however, smooth TFs limit the range of achievable rendering effects, e.g., isosurface rendering via thin spikes in the TF that, in theory, require an infinite sampling density. To circumvent this problem Roettger et al. [39] introduce preintegration, which avoids visual artifacts by precomputing the volume rendering integral for ray segments, assuming that the underlying scalar function is linear. Recent efforts focus on accelerating the precomputation [8, 37], faster approximation algorithms [24], multidimensional transfer functions [30], and preintegrated lighting [24, 29].

In this paper we propose an adaptive integration scheme based on preintegration. This scheme is similar in spirit to other numerical integration schemes, such as adaptive Simpson [32] and the second derivative refinement scheme of Roettger et al. [38].

3 Volume MLS

In order to generate high-quality images, a ray caster must evaluate a continuous representation, and the associated derivatives, of volumetric data. For volume data defined over a grid, two approaches can be taken: first, the connectivity of the grid can be used to define a reconstruction scheme over the locally connected data points; or second, the data points can be used independently of their grid connections in a scattered data reconstruction scheme. The first approach works well for regular grids, where higher order polynomials are often employed to reconstruct functions of arbitrary continuity. For irregular grids, however, reconstructions are challenging for anything other than piecewise linear interpolation. Thus, we advocate a unified reconstruction framework that takes the latter approach, allowing for continuous and differentiable representations of regular and irregular grids.

Scattered data reconstruction takes a set of n data points, \( \mathbf{p}_i = (x_i, f_i) \), defined over a volume \( V \), and computes an approximation (or interpolation) to the data points \( f(x) : V \rightarrow \mathbb{R} \). One such well-studied, meshless scheme is MLS, where \( f(x) \) is obtained by approximating the local neighborhood of \( x \) in a weighted least squares (WLS) sense. The MLS method has a high degree of flexibility, making it particularly appealing for the reconstruction of volumetric data, while avoiding the artifacts often incurred by reconstruction schemes that are constrained to arbitrarily poor connectivities (i.e., interpolation over poorly shaped mesh elements). This flexibility provides the capacity for (controllably) handling noise due to measurement errors in scanned data or numerical errors in simulated data. Furthermore, the MLS approximation is a continuous reconstruction with well-defined, smooth derivatives which allows for high-quality shading effects.

In this section, we first describe the MLS framework and mathematical foundations in the context of reconstructing continuous functions from scattered data. We then discuss and explore the various components of the MLS method as they relate to reconstructing data stored over regular and irregular grids specifically for generating high-quality visualizations by ray casting. Our notation is as follows: bold face lower-case variables denote column vectors, such as \( x = [x\ y\ z]^T \), while bold face upper-case variables denote matrices, such as the identity matrix \( I \). We note that the MLS discussion and mathematical definition holds for arbitrary dimensions, even though we restrict our examples and results to \( \mathbb{R}^3 \).

3.1 The MLS Framework

The main idea behind the MLS reconstruction method is depicted for a 1D signal in Figure 2. Shown in red, a continuous approximation \( f(x) \) is reconstructed from a set of \( n \) data points \( (x_i, f_i) \) by computing

![Fig. 2. MLS reconstruction of a given set of scattered data points (x_i, f_i) in 1D. The local MLS approximation g_x at the point x is shown in blue. Its computation and evaluation at every point of the domain yields the complete MLS reconstruction shown in red.](image)

and evaluating a local approximation \( g_x \) at \( x \) (shown in blue):

\[
f(x) = g_x(x).
\]

(1)

The local approximation \( g_x \) is obtained by the following WLS minimization problem:

\[
g_x = \arg\min_{g} \sum_i w_i(x) |g(x_i) - f_i|^2,
\]

(2)

where \( w_i(x) \) is the weight of the sample \( p_i \) for the current evaluation point \( x \). These weights typically decrease with distance from \( x \) (see Section 3.1.2). Assuming \( g_x \) can be represented as a linear combination of \( k \) basis functions \( \mathbf{b}(x) = [b_1(x), \ldots, b_k(x)]^T \) with coefficients \( \mathbf{c}_x = [c_1, \ldots, c_k]^T \), Equation 2 can be rewritten in the following matrix form:

\[
\mathbf{c}_x = \min_{c \in \mathbb{R}^k} \| (\sqrt{W(x)} \mathbf{B} - f) \|^2,
\]

(3)

where \( W(x) = \text{diag}(w_1(x), \ldots, w_n(x)) \in \mathbb{R}^{n \times n} \) and \( B_{ij} = b_i(x_j), \mathbf{B} \in \mathbb{R}^{n \times k} \). It is well-known from linear algebra that the above least squares problem can be solved using the normal equation. Substituting the solution for \( \mathbf{c}_x \) into \( f(x) = \mathbf{b}(x)^T \mathbf{c}_x \) yields the following explicit formulation of the MLS approximating function:

\[
f(x) = \mathbf{b}(x)^T (\mathbf{B}^T W(x) \mathbf{B})^{-1} \mathbf{B}^T W(x) \mathbf{f}.
\]

(4)

This function is smooth, assuming the weights \( w_i(x) \) are smooth [19], and can be computed as long as the weighted covariance matrix \( \mathbf{B}^T W(x) \mathbf{B} \) is invertible, which is the case if \( \text{rank}(W(x) \mathbf{B}) \geq k \) — this
implies that at least \( k \) weights have to be nonzero. We emphasize that the matrix \( W(x) \) depends on the evaluation point, and hence the least squares system must be solved at every \( x \).

In the MLS framework, the smoothness of the reconstruction, as well as the tightness of the approximation to the data values (i.e., interpolation), is controlled through the choice of the basis and weighting functions. Each of these components are discussed in the following sections.

### 3.1.1 Basis Functions

When choosing the basis (i.e., \( b(x) \)), the main trade-offs are approximation power versus efficiency. Lower order polynomial bases are less accurate at approximating the data due to fewer degrees of freedom in the function. Higher order polynomial bases, however, are less efficient as a larger linear system must be solved, as well as the determination of a larger neighborhood to maintain stability in the linear system computation. Furthermore, a too-large basis can yield overfitting. We have experimented with several bases, and for ray casting system computation. Furthermore, a too-large basis can yield overfitting. We have experimented with several bases, and for ray casting MLS volumes, we advocate the use of linear \( b(x) = [1, x, y, z]^T \) basis functions. Indeed, as shown in Figure 3, a quadratic basis does not give much better results while increasing the computational burden by more than a factor of two.

#### 3.1.2 Weighting Functions

Once the basis function has been fixed, the quality of the reconstruction highly depends on the weights assigned to the neighboring samples. For uniform data sets, we employ the following simple weighting scheme:

\[
w_i(x) = \theta \left( \frac{\|x - x_i\|}{h} \right),
\]

where \( h \) is the support size, and \( \theta \) is a smooth, decreasing weight function for which several options have been proposed in the literature \([47]\). In this work, we use the following compactly supported polynomial:

\[
\theta(x) = \begin{cases} 
(1 - x^2)^\alpha & \text{if } x < 1 \\
0 & \text{otherwise}
\end{cases}
\]

where the power factor \( \alpha \) controls the speed at which the weighting function decreases. Although larger values of \( \alpha \) increases the continuity of the function, large values also require a wider support for stable computation. A good compromise is \( \alpha = 4 \), which guarantees a \( C^0 \) reconstruction. Note that in Equation 4, since \( \theta(x) \) has a finite support, only points within a distance \( h \) of \( x \) have to be taken into account, which allows for efficient computation while avoiding square roots in the distance computations. The effect of the support size in the reconstruction is depicted in Figure 3.

To reconstruct irregular, and possibly anisotropic, datasets we use the weighting scheme proposed by Adamson and Alexa \([1]\), and attach ellipsoidal weight supports \( H_i \) to each input sample:

\[
w_i(x) = \theta \left( \|H_i(x - x_i)\| \right).
\]

The weight transforms, \( H_i \), are obtained using a covariance analysis of the one-ring neighborhoods. Figure 6 depicts an example of such anisotropic weight supports. For irregular datasets that do not present significant anisotropy, it is sufficient to replace the expensive ellipsoids by balls, reducing the \( 3 \times 3 \) matrices \( H_i \) to a single scalar radius \( h_i \).

In addition to data approximation, the MLS framework is also capable of data interpolation using weighting functions that contain a singularity \([41]\). One such function proposed by Adamson et al. \([2]\) is:

\[
\theta(x) = \begin{cases} 
\ln(x)^2 - x^2 + 2x - 1 & 0 \leq x \leq 1 \\
0 & x > 1
\end{cases}
\]

Compared with other interpolating weighting schemes, such as those based on inverse distances \([41, 35]\), Equation 8 has a finite support, which is important for efficient evaluation. One property of interpolating MLS which is noteworthy is the flat spot phenomenon \([41, 19]\). This implies that for a constant basis, the gradient of the approximation at any data point is zero \( \nabla f(x_i) = 0 \). To avoid this undesirable characteristic a higher order basis can be used.

### 3.2 Gradient Computation

The generation of high-quality shading effects requires accurate gradient information, as does the adaptive preintegration scheme discussed in Section 4. There are two established methods for computing the gradient within the MLS framework. The most accurate method is to analytically compute the gradient from the partial derivatives of the MLS scalar field formula (Equation 4):

\[
\begin{align*}
\frac{\partial f(x)}{\partial x_k} &= \frac{\partial b(x)^T}{\partial x_k} \epsilon_k \\
&= -b(x)^T A(x)^{-1} \left( \frac{\partial A(x)}{\partial x_k} \epsilon_k + B^T \frac{\partial W(x)}{\partial x_k} f \right)
\end{align*}
\]

where \( A(x) \) is the weighted covariance matrix \( B^T W(x) B \).
A more computationally efficient alternative is to compute the gradient of the local WLS approximation $g_k$, which has been shown to reasonably approximate the gradient of the underlying function $f$:

$$\frac{\partial f(x)}{\partial x_k} \approx \frac{\partial g_k(x)}{\partial x_k} = \frac{\partial b(x)^T}{\partial x_k} c_k. \quad (10)$$

In practice we find that images shaded using either the local gradient or the analytic formula most often lead to indistinguishable results, except for models presenting high frequencies, such as the result shown in Figure 4. Hence we advocate, in general, the use of the more efficient local gradient.

### 4 Adaptive Preintegration

The unified MLS framework proposed in this paper incurs a more expensive scalar field function evaluation than other lower order reconstruction schemes. As such, we propose an adaptive preintegration scheme that modifies the length of evaluation intervals along rays, focusing the computation along segments where the volume integral changes the fastest. We observe that all of the information needed to robustly determine an appropriate interval length for evaluating the volume integral is contained in the preintegration tables and the scalar function itself. First, because the preintegration tables ensure that isosurfaces will be composited as long as the scalar function is monotonic over an interval, the interval length can be increased with increasing composited opacity values. And second, because isosurfaces near extrema in the scalar function can be missed, the interval length must decrease near these features.

From these observations we propose a novel, adaptive preintegration method for computing the volume rendering integral. An illustrative result of the method is shown in Figure 5, where the adaptive sample locations along one ray are shown. The behavior of the scalar function along the ray is shown in red while the accumulated opacity is shown in blue. It can be seen that the sampling of the adaptive scheme (bold vertical lines) becomes more coarse as the accumulated opacity increases. Refinement occurs, however, around scalar function extrema to ensure that all isosurface intersections are detected. Finally, the sampling is terminated when the opacity is saturated. In contrast, standard preintegration methods [8] can only take advantage of early ray termination while choosing a conservative step size such that extrema in the scalar function are adequately sampled, producing oversampling along intervals where the volume integral changes slowly.

In the remainder of this section we first formulate an approximation of the volume rendering integral using arbitrary length evaluation intervals (Section 4.1), which is followed by a description of the interval length adaptation scheme (Sections 4.2 and 4.3). We then include details on the computation of the preintegration tables and a description of how to handle clipped ray segments incurred from intersecting rays with irregular grids (Section 5).

#### 4.1 Approximating the Volume Integral

In this section we derive an approximation of the volume rendering equation [28] for use with the proposed adaptive preintegration method. Specifically, the derivation allows for the integration along a ray to be approximated with a composition of colors and opacities computed over arbitrary-length subintervals of the ray. To begin, the volume rendering integral is given as:

$$I = \int_0^D \int_0^1 c(f(x(\lambda))) \tau(f(x(\lambda))) e^{-\int_0^\lambda \tau(f(x(\lambda')))d\lambda'} d\lambda d\lambda'$$

$$\lambda \approx \int_0^\lambda \tau(f(x(\lambda')))d\lambda'$$

$$\int_0^\lambda \tau(f(x(\lambda')))d\lambda'$$

where $c$ is the emitted color and $\tau$ is the extinction which changes with the scalar function $f$ along the ray $x(\lambda)$. For use with preintegration, the domain $D$ of the integral can be split into the intervals [8] $(d_0, d_1, \ldots, d_{n-1}, d_n)$, where $d_0 = 0$ and $d_n = D$. The opacity ($\alpha$) and color ($C$) for arbitrary length intervals can then be defined as:

$$\alpha_{[d, d_{i+1}]} = 1 - \exp \left( - \int_{d_i}^{d_{i+1}} (\tau(f(x(\lambda))))d\lambda \right)$$

$$C_{[d, d_{i+1}]} = \int_{d_i}^{d_{i+1}} c(f(x(\lambda))) \tau(f(x(\lambda))) e^{-\int_0^\lambda \tau(f(x(\lambda')))d\lambda'} d\lambda d\lambda'$$

Using Equations 12 and 13 we arrive at the well-known recursive front-to-back composition formulas for opacity and color:

$$\alpha_{i+1} = \alpha_i + (1 - \alpha_i) \alpha_{[d, d_{i+1}]}$$

$$C_{i+1} = C_i + (1 - \alpha_i) C_{[d, d_{i+1}]}$$

where $I = C_n + \lambda_C \alpha_0$ and $d_0 = 0$. This reformulation is exact, and because there are no assumptions about the length of the intervals, the step size can be varied along the ray. Furthermore, as Equations 14 and 15 add the contribution of the interval $[d_i, d_{i+1}]$, this contribution is fully determined by $(1 - \alpha_i) \alpha_{[d, d_{i+1}]}$ and $(1 - \alpha_i) C_{[d, d_{i+1}]}$. This property is important in Section 4.2.

Assuming that $f(x(\lambda))$ is a linear function on the interval $[d_i, d_{i+1}]$, $\alpha_{[d, d_{i+1}]}$ and $C_{[d, d_{i+1}]}$ can be precomputed in a 3D table. This pre-computation depends on the value of the scalar function $f$ at the front $f(d_i)$ and the back $f(d_{i+1})$ of the interval, as well as on the length $l = d_{i+1} - d_i$ of the interval. We note that these table values are only an approximation to the true integral over the interval because the function is, in general, nonlinear and the table has a finite resolution.

For the remainder of this section let us denote the approximated color of an interval (stored in a preintegration table) by $\tilde{C}_{[d, d_{i+1}]}$ and similarly, the approximated opacity as $\tilde{\alpha}_{[d, d_{i+1}]}$.

#### 4.2 Interval Length Adaptation

The accuracy of the volume integral approximation along a interval can be increased by further subdivision along the interval. As such, the interval can be progressively refined until the difference of the volume integral approximation between any two successive subdivision levels is below some error threshold.

Starting with a coarse step size, the initial volume integral approximation along an interval for color (i.e., the contribution of the interval in Equation 14), $C_1$, is:

$$C_1 = \tilde{C}_{i,i+1}$$

$$\tilde{C}_{i,i+1}$$
To increase the accuracy of this approximation, the interval is subdivided, and the contribution, \( C_2 \), becomes a summation over the two subintervals:

\[
C_2 = C_{i,j+1/2} + (1 - \alpha_{i,j+1/2} ) C_{i+1/2,j+1/2} + (1 - \alpha_{i,j+1/2} ) C_{i+1/2,j+1/2}
\]

The subdivided contributions for opacity are defined similarly as:

\[
\alpha_{i,j+1/2} \approx (\alpha_{i,j+1/2} + (1 - \alpha_{i,j+1/2} ) \alpha_{i+1/2,j+1}).
\]

This subdivision continues recursively until the difference in the volume integral approximation over the entire interval changes by only a small amount.

Based on the volume integral approximations \( C_1 \) and \( C_2 \) we define the relative error of the approximation as \(|C_1 - C_2|/|C_1|\), where \(|C| = \sqrt{C_1^2 + C_2^2} - C_1 - C_2\) — note that this is not independent from the opacity (see Equations 15 and 16). To adapt the sampling as the opacity increases we multiply the relative error by \((1 - \alpha)\). The final termination criteria for the recursive subdivision is:

\[
(1 - \alpha) |C_1 - C_2| / |C_1| < \varepsilon.
\]

In the recursion it is important to first subdivide the interval \([d_i,d_{i+1/2}]\) such that \(\alpha_{i,j+1/2}\) is known when subdividing the interval \([d_{i+1/2},d_{i+1}]\). This is because \(\alpha_{i,j+1/2}\) is used in the computation of the termination criteria for the interval \([d_{i+1/2},d_{i+1}]\). Note that this scheme is automatically optimal for constant transfer functions and linear scalar fields, since \(C_1\) and \(C_2\) will be equal to within numerical errors.

### 4.3 Handling Scalar Function Extrema

The subdivision scheme in the previous section works well as long as the scalar function is monotonic over an interval. If this property is violated, an isosurface may exist near an extremum of the scalar function which is never recovered during subdivision. The MLS reconstructions are not guaranteed to be monotonic over an arbitrary interval as they are not linear. Because of that, we allow extrema in the scalar function to induce subdivisions in the adaptive preintegration scheme.

We propose using the derivatives at the beginning and end of an interval to determine whether the interval contains an extremum. Thus, an interval \([d_i,d_{i+1}]\) will be further subdivided if

\[
D_k(f(x(d_i)))D_k(f(x(d_{i+1}))) < 0
\]

where \(D_k\) is the derivative of the scalar function \(f\) at \(d_i\) in the direction of the ray \(v\). To compute \(D_k\) we use the standard formulation:

\[
D_k(f(x)) = \nabla f(x) \cdot v.
\]

Equation 19 is efficient to compute in the proposed MLS framework as \(\nabla f(x)\) is already determined for shading and imposes very little overhead in an MLS function evaluation. It should be noted that the accuracy of the termination criteria heavily depends on the accuracy of the computed gradient \(\nabla f(x)\).

### 5 Implementation

The first step in the proposed volume MLS ray casting system is to compute the preintegration tables based on an input transfer function. We compute the table for opacity, as well as the packed tables for shading and material properties, using the incremental subrange preintegration algorithm proposed by Lum et al. [24]. In the adaptive preintegration scheme implementation we must choose two parameters, \(l_{\text{max}}\) and \(l_{\text{min}}\), which bound the length of an interval. Since we know that during the subdivision process only intervals of length \(l_{\text{max}}\) will be used, only those tables need to be computed.

We found that our adaptive preintegration scheme is sensitive to numerical errors in the preintegration tables. To be able to explicitly control the accuracy of the preintegration tables we improved the incremental subrange preintegration such that we can control the maximal step size of the numerical integration, i.e., we supersample intervals if necessary. While the table of length \(l_{\text{max}}\) can also be computed from the table of length \(l_{\text{min}}/2\) using the idea of incremental preintegration [46], we found it difficult to control the errors in such a scheme. The need for high accuracy preintegration also implies larger tables. All the results in this paper use tables with a resolution of 512^2. While these large tables increase the memory footprint of the system (the tables require approximately 72MB), they are relatively fast to compute (up to 10 seconds on the CPU) and could likely be generated on the GPU at faster speeds, enabling interactive user control of the TFS.

In the case where an interval is clipped against the bounding geometry of the volume we must compute the integral for a length that is not present in any of the tables. For this scenario, the integral can be determined by carefully combining values from different tables. Because this happens only twice per ray, this additional complexity is virtually unnoticeable.

Next, the ray caster intersects each ray with the boundary of the volume, finding all the intervals that lie within the (possibly nonconvex) hull of the grid. These intersection points are important as the MLS function approximation is ill-defined away from the data points. Along each of the intersected ray segments, integration intervals are determined using the adaptive scheme described in Section 4, and the MLS function and gradient are computed at the front and back of each interval (see next section). Finally, the volume integration along the interval is computed via lookups into the preintegration tables. The resulting color and opacity are composited with the stored values on the ray. Early ray termination occurs when the opacity value is greater than some threshold close to one (for the results in this paper we use 0.99).

### 5.1 Fast and Robust MLS Evaluation

The most time critical part of our rendering system is the evaluation of the MLS function and its gradient, which involves both a neighborhood query and solving a linear system. The complexity of the latter operation depends on the size of the neighborhood and the number of degrees of freedom in the basis. For a small basis, once the covariance matrix \(B^T W(x) B\) is computed, solving the normal equation itself is relatively cheap and can be done efficiently using a Cholesky decomposition without square roots of the covariance matrix. Given

\[
\begin{align*}
B^T W(x) B &= LDL^T, \\
B^T W(x) f &= L^{-1} D^{-1} L^{-1} \left( B^T W(x) f \right).
\end{align*}
\]

Not only does avoiding the computation of square roots improve performance, but it also increases the stability of the system. Moreover, both stability and efficiency can be further improved by shifting the origin of the basis to the the evaluation point, i.e., \(b(x) = b(x - \bar{x})\), which is mathematically equivalent and avoids large numbers when computing the function far from the origin. Since the basis function is only evaluated at \(x\), i.e., at the origin, it is therefore sufficient to compute the constant coefficient \(c_0\) of \(c_\mathbf{x}\), hence avoiding most of the calculation in Equation 21.

The computations of the neighbor samples, on the other hand, requires efficient spatial-data structures. More precisely, given an evaluation point \(x\), the challenge is to find all samples \(p_i\) having strictly positive weights. Therefore, the choice of both the data structure and search algorithm is related to the weighting scheme. For the weighting function defined in Section 3.1.2, we need to find all the ellipsoids (or balls) containing \(x\). In our CPU implementation this is accomplished using a kd-tree partitioning of the space where a sample is referenced by all leaves intersecting its corresponding ellipsoid. Our kd-tree is built recursively from top to bottom until the size of the node is smaller than the average radius of the node’s samples.
Table 1. This table lists the data sets presented in this paper, along with the timings of our GPU implementation for the standard and adaptive integration schemes as well as the number of function evaluations needed by both schemes to satisfy the RMSE shown. We also show the step size parameters \( l_{\text{min}}, l_{\text{max}} \) and the termination criterion \( \varepsilon \) (see Section 4.2) for the adaptive scheme which were used to generate these results.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Grid Type</th>
<th>Size</th>
<th>FPS Standard</th>
<th>FPS Adaptive</th>
<th># Func Eval Standard</th>
<th># Func Eval Adaptive</th>
<th>Min / Max Interval Length</th>
<th>( \varepsilon )</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marschner-Lobb</td>
<td>regular</td>
<td>41x41x41</td>
<td>0.28</td>
<td>0.46</td>
<td>22M</td>
<td>9.4M</td>
<td>0.16 / 7.6</td>
<td>0.03</td>
<td>0.044</td>
</tr>
<tr>
<td>Heptane Plume</td>
<td>regular</td>
<td>302x302x302</td>
<td>0.72</td>
<td>0.69</td>
<td>44.6M</td>
<td>18.8M</td>
<td>0.16 / 2.6</td>
<td>0.03</td>
<td>0.044</td>
</tr>
<tr>
<td>Blunt Fin</td>
<td>irregular</td>
<td>180K tets</td>
<td>0.45</td>
<td>0.64</td>
<td>26M</td>
<td>10M</td>
<td>0.25 / 4</td>
<td>0.03</td>
<td>0.017</td>
</tr>
<tr>
<td>Bucky Ball</td>
<td>irregular</td>
<td>1.2M tets</td>
<td>1.9</td>
<td>3.2</td>
<td>5.4M</td>
<td>1.9M</td>
<td>0.25 / 4</td>
<td>0.05</td>
<td>0.043</td>
</tr>
<tr>
<td>Torso</td>
<td>irregular</td>
<td>1.1M tets</td>
<td>1.9</td>
<td>3.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.2 CUDA Implementation

In order to achieve nearly interactive performance we implemented a GPU-based version of the raycaster using CUDA [34]. In our current implementation, each thread processes a single ray, and the ray traversal is entirely decoupled from the volume data structure. For the ray traversal we implemented both a constant step size strategy and the adaptive algorithm described in Section 4. In the latter case, the recursion is managed via a small stack stored in local memory. To efficiently find neighbor samples, we use the \textit{dynamic redundant octree} of Guennebaud \textit{et al.} [16]. This spatial search data structure is particularly efficient for GPU applications, allowing for dynamic updates as well as intuitive control of the memory footprint versus performance. We did not yet implement the anisotropic weights discussed in Section 3.1.2 on the GPU, even though that is possible.

6 RESULTS

In this section we present results of the proposed MLS framework for several well-known regular and irregular datasets. Except where noted otherwise, all of the MLS images were generated using a linear basis with approximate gradients and the proposed adaptive preintegration scheme with our GPU-based implementation.

Table 1 provides information on each data set, including the relevant parameters used to generate the images and performance results. The results were computed on a 2.6GHz Intel Core 2 Duo with 8 GB of RAM and an NVIDIA Quadro FX5600 GPU with 1.5 GB of video memory. The parameters in Table 1 lead to a root mean square error (RMSE) below 5% compared to ground truth images computed using a very small step size. We found that this error tolerance leads to visually indistinguishable results from the ground truth images. Of all the parameters \( \varepsilon \) (see Section 4.2) is the least sensitive for the resulting image quality, while a too-large value for \( l_{\text{max}} \) (see Section 5) can cause the adaptive integration scheme to entirely miss extrema in the volume integral by inducing an early termination of the automatic refinement scheme. This early termination results in visible discontinuities in the final image. A too-large segment length for the standard integration scheme also introduces artifacts, but those are harder to detect because the images is less likely to exhibit discontinuities as the sampling positions on neighboring rays are very similar.

The capabilities of the MLS system for regular grids are shown in the renderings of the heptane plume dataset in Figure 1 and the Marschner-Lobb dataset in Figure 7. To illustrate MLS ray casting of curvilinear grids we show renderings of the highly anisotropic NASA blunt fin dataset in Figure 6. We show performance numbers for rendering the blunt fin on the CPU because we did not yet implement anisotropic weights on the GPU. Note that the CPU implementation is not optimized and takes orders of magnitude longer than the GPU implementation. Irregular grid renderings of tetrahedral meshes are shown in Figure 8 with the bucky ball dataset (left) and the Utah torso model (right). For the torso model we estimated the sampling density at the input samples using the \( k \)-nearest neighbor radius and rendered the reconstructed function. In all of these images, the high-quality shading of smooth isosurfaces confirms the quality of the MLS reconstruction, while the range of datasets and grid-types illustrates the versatility of the MLS scheme.

To compare the MLS system with existing ray casting methods, we implemented a standard piecewise-linear scheme with weighted normals [26] for tetrahedral meshes (see Figure 6 (left)). The image shows significant visual artifacts, even though we are using weighted normals. Figure 7 (middle) shows the result of cubic B-spline convolution for the reconstruction of a continuous function over a regular grid. The B-spline reconstruction has been shown to be optimal for the Marschner-Lobb dataset [27]. The comparison shows that MLS reconstruction obtains competitive results. Using a linear MLS basis on a regular grid does not reduce approximation errors at the grid points [44], which explains why the MLS approximation is not tighter. However, we note that convolution is equivalent to MLS with a constant basis function [44], making MLS a more general framework that subsumes convolution.

6.1 Performance Analysis

In theory, the complexity of our algorithm is \( O(m \cdot p \cdot (k + \log(n))) \) for irregular data sets where \( m \) is the number of pixels, \( p \) is the number of function evaluations per ray, \( n \) is the size of the dataset, and \( k \) is the number of neighbors used. As pointed out in Section 4.2, the number of function evaluations per ray depends on the complexity of the scalar function and transfer function. The runtime of a function evaluation is \( O(k + \log(n)) \), where \( O(\log(n)) \) represents the traversal complexity of the hierarchical data structure — in practice the size of the dataset has very little impact on the evaluation cost that is highly dominated by the size of the neighborhoods. Using an adaptive scheme, the total number of function evaluations can be significantly reduced (see Table 1). However, the limited flexibility of current GPU architectures makes it very challenging to take full advantage of this gain. In our experiments we observed that more complex datasets (\textit{i.e.}, those that contain more high-frequency features) benefited the most from the adaptive scheme.

To give further insight into the complexity of a function evaluation we compare the number of floating point operations (flops) used for rendering the blunt fin dataset in Figure 6. W e show performance numbers for rendering the blunt fin on the CPU because we did not yet implement anisotropic weights on the GPU. Note that the CPU implementation is not optimized and takes orders of magnitude longer than the GPU implementation. Irregular grid renderings of tetrahedral meshes are shown in Figure 8 with the bucky ball dataset (left) and the Utah torso model (right). The torso model displays the sampling density of the sample points estimated using the \( k \)-nearest neighbor radius.
6.2 Reconstruction Quality Analysis

In practice, function reconstruction is an ill-posed problem for which there are many reasonable solutions. Furthermore, for visualization there is a tradeoff between approximation power and visually pleasing results. The latter is subjective, while the former is well-studied for MLS. Convergence theory [48, 22] yields error bounds of the MLS reconstruction, and multivaried nonparametric regression in statistics [40] has lead to a thorough understanding of its bias and variance.

7 CONCLUSIONS & FUTURE WORK

Volume MLS ray casting produces high-quality images and provides an easy to understand mathematical framework with intuitive parameters. It is a first step towards creating an efficient framework which is able to handle any kind of data regardless of its topology. A next step in this direction would be to remove the necessary boundary information and compute a (possibly) nonconvex hull of the data purely based on the sampling. Our volume MLS framework ties in nicely with a wealth of existing tools for MLS reconstruction. For example, one could extend this work to include exact reconstruction of sharp features.

The proposed system lends itself well to streaming data. Since an additional data point only affects the scalar function in a bounded region, a local update of the image could be computed and no global system has to be solved. The ray casting approach is attractive due to its simplicity and because it can easily be extended to include complex cut and region of interest geometry. Finally, our adaptive preintegration scheme is immediately useful for other ray casting frameworks. We believe that progress on the theory of the convergence of preintegration could improve the efficiency and robustness of our scheme.

ACKNOWLEDGEMENTS

The authors sincerely thank Markus Gross, ETH Zurich, for inspiring this work during stimulating discussions at the outset of this project. The authors wish to thank David Luebke and NVIDIA for generous donation of hardware. Christian Ledergerber, Miriah Meyer, and Moritz Bächer were supported by funding from the Initiative in Innovative Computing (IIC) at Harvard. Gael Guennebaud was supported by the ERCIM “Alain Bensoussan” Fellowship Programme. The Utah torso model is courtesy of the NCRR Center for Integrative Biomedical Computing at the University of Utah, and heptane plume dataset is courtesy of the Center for the Simulation of Accidental Fires and Explosions, also at the University of Utah.