Visualization of three-dimensional data by ‘volume rendering’

N. Ramesh and G. Athithan

Advanced Numerical Research and Analysis Group,
P. O. Kanchehanbagh, Hyderabad 500 258, India

We report an implementation of a ray-tracing method for visualizing volumetric data. For evaluating the performance of the implementation, we consider the visualization of two classes of volumetric data. The first consists of the probability density functions of a subset of states of an electron in hydrogen atom and represents the case of visualizing data derived from analytical expressions. The second class of volumetric data is a set of three-dimensional fractals which are generated over regular three-dimensional cartesian grids of various sizes. The motivation for generating and visualizing the fractal data sets is two-fold. Firstly a fractal data set is a good test input to verify the correctness of the implementation. Secondly, we want to model stellar clouds and see how they can be visualized. Clouds are known to be akin to fractal data sets and we find that the volume rendered images of these data sets bear a close resemblance to stellar clouds. Ray-tracing is computationally expensive and we therefore report the CPU timings for a representative set of images. General comments on the utility of this method for visualizing in various disciplines conclude the paper.

Representation and visualization of data in a graphical form is a practice as old as experimental science. Whether it is an array of numbers or a mathematical function, a graphical plot of the same gives better insight. However, before the advent of computer graphics, one was limited to visualizing the relationship between just two variables in a graphical form. Visualization of any relationship among three variables usually called for highly skilled draftsmen and that among four variables was out of the question. With the application of computer graphics, the problem of visual-
ORIZING trivariate relationships has been solved in a number of ways. The key task in all the solutions is the hidden element (line, surface) removal. Currently, one of the main preoccupations in computer graphics is to represent and visualize relationships between four variables or what is more commonly known as volumetric data visualization. The data set to be visualized comes in the form of scalar values at each point of a regular or irregular, three-dimensional grid. Our discussion here is, however, confined to regular cartesian grids. The values at the grid points may come from an experiment such as magnetic resonance imaging of human bodies or may be calculated from a trivariate analytic function.

One of the standard methods of visualizing such a grid of data is to construct surfaces inside the volumetric grid, over which the scalar value remains invariant. Such surfaces, commonly known as 'iso-value surfaces' can then be visualized graphically by means of various surface display techniques. A single isovalue surface does not suffice to represent the entire volumetric data, and so one may construct many such surfaces simultaneously, for different isovalues so as to visualize them all together. Usually this approach works when the data vary somewhat smoothly over the grid. However, if the data are poor, noisy or vary rapidly from point to point, then one finds spurious additions or deletions from the isovalue surfaces. This happens due to the binary nature of the classification of the cells of the grid into those that are intercepted by the isovalue surface and those that are not. False negatives contribute to holes while false positives may cause spurious additions\(^1\). To circumvent this problem, researchers recommend the concept of 'volume rendering' wherein no intermediate geometric construction is carried out\(^2\). In this approach screen images are computed directly from the volumetric data by means of ray-tracing. The main advantage of this approach is that it generates a mechanism for extracting and displaying weak or fuzzy features out of the volumetric data. Here one can emulate the technique of isovalue surface visualization without the problems of spurious additions or deletions, as shown in ref. 1. Alternatively one can render the entire volume of data as we demonstrate below using two classes of volume data.

After a survey of the literature for a suitable algorithm for volume rendering we chose to implement the one suggested by Marc Levoy\(^3\). While he concentrates more on the isovalue surface visualization by means of volume rendering, our interest here is to carry out entire volume visualization. We give a brief outline of his method, modified to suit our need, in what follows. The interested reader may refer to his paper for further details.

The basic input to the algorithm is a regular three-dimensional grid of scalar values. In case the grid is irregular, it has to be transformed to a regular grid in a separate preprocessing step. In the subsequent discussion, the grid points are referred to as voxels (volume elements) and the scalar value is referred to as intensity. They are denoted by \(g_{ijk}\) and \(f(g_{ijk})\) respectively. The algorithm has two functional stages. In the first stage, all the grid points are assigned a colour value \(C_p\) (\(p = \text{red, green, blue}\) (shading) and an opacity value \(\alpha\) (classification). Both the shading and classification calculations depend heavily on the normal at that point. In the absence of an explicit geometry, the normal is computed from the intensity gradient using

\[
N(g_{ijk}) = \frac{\delta f(g_{ijk})}{|\delta f(g_{ijk})|},
\]

where the intensity gradient \(\delta f(g_{ijk})\) at \(g_{ijk}\) is computed as

\[
\delta f(g_{ijk}) = \left[ \frac{1}{2} (f(g_{i+1,j,k}) - f(g_{i-1,j,k})) + \frac{1}{2} (f(g_{i,j+1,k}) - f(g_{i,j-1,k})) + \frac{1}{2} (f(g_{i,j,k+1}) - f(g_{i,j,k-1})) \right].
\]

The colour value \(C_p\) at \(g_{ijk}\) is computed using the following colour model.

\[
C_p(g_{ijk}) = C_{s,p} k_{s,p} + C_{r,p} k_{r,p} (N(g_{ijk}) \cdot L) + k_{s,p} (N(g_{ijk}) \cdot H)^n,
\]

where, \(C_p(g_{ijk})\) is the \(p\)th colour component at voxel location \(g_{ijk}\); \(C_{s,p}\) the \(p\)th colour component of the light source at infinity; \(k_{s,p}\) the ambient reflection coefficient for \(p\)th colour component; \(k_{d,p}\) the diffuse reflection coefficient for \(p\)th colour component; \(k_{s,p}\) the specular reflection coefficient for \(p\)th colour component; \(N(g_{ijk})\) the normal vector at \(g_{ijk}\), \(L\) the normalized vector in the direction of light source; \(H\) the normalized vector in the direction of maximum highlight and \(n\) the exponent used in highlight approximation.

The highlight vector \(H\) is computed as

\[
H = \frac{V + L}{|V + L|},
\]

where \(V\) is the normalized vector in the direction of the observer.

The opacity value of a voxel is a measure of its transparency. A value of zero means the voxel is fully transparent (and hence invisible) and one means it is fully opaque (thus hiding everything behind it). Values between 0 and 1 mean the voxel is translucent. For
entire volume rendering the opacity value of a voxel is assigned in proportion to its intensity value. This makes the high intensity voxels more opaque (and hence seen more) and low intensity voxels more transparent. That is,

\[ \alpha(g_{ijk}) = \frac{f_{ijk} - f_{\text{min}}}{f_{\text{max}} - f_{\text{min}}} \]

Here, \( f_{\text{min}} \) and \( f_{\text{max}} \) are the minimum and maximum intensity values in the grid.

To avoid voxels having intensity value equal to \( f_{\text{min}} \) from becoming fully transparent (and hence not seen at all), instead of using the complete unit interval [0-1] for opacity value assignment, a sub interval within [0-1] may be used. That is,

\[ \alpha(g_{ijk}) = p + \frac{(f_{ijk} - f_{\text{min}})}{f_{\text{max}} - f_{\text{min}}} (q - p), \]

where \( 0 < p, q \leq 1, p < q \).

Once the colour and opacity values are assigned to all voxels in the volume, a beam of rays is sent parallel to the eye vector through the viewing plane into the volumetric grid. In this stage, a colour for each ray is computed and the pixel in the viewing plane through which the ray passes is assigned that colour. The viewing plane is analogous to the screen.

Along each ray, the volume is resampled at \( K \) equally spaced intervals. This interval is taken as the minimum distance inside a cell bounded by eight voxels. Colour and opacity value at a sample point are calculated by trilinearly interpolating from the eight nearby voxels that surround the point. Finally, a fully opaque background \( C_{\text{bg}} \) is draped behind the volume as the last sampled point of all the rays. These colour and opacity values along ray are denoted as \( C_g(R_{x,y}) \) and \( \alpha(R_{x,y}) \) and colour for the entire ray is denoted as \( C_g(R_{x,y}) \). \( R_{x,y} \) denotes the ray passing through \( (x, y) \) position of the viewing plane and \( R_{x,y} \) denotes the \( i \)th sample point on the ray \( R_{x,y} \). From this array of colour and opacity values, a single colour is computed per ray by a process of compositing. More specifically, the colour \( C_{\text{out,g}}(R_{x,y}) \) of the ray as it leaves sample point \( i \) is related to the colour \( C_{m,\beta}(R_{x,y}) \) of the ray as it enters that point and the colour and opacity values \( C_g(R_{x,y}) \), \( \alpha(R_{x,y}) \) at that sample point, by the transparency formula

\[ C_{\text{out,g}}(R_{x,y}) = C_{m,\beta}(R_{x,y})(1 - \alpha(R_{x,y}))) + C_g(R_{x,y})\alpha(R_{x,y}). \]

For an array of sample values, compositing is done by repeated use of the above formula. That is

\[ C_g(R_{x,y}) = \sum_{i}^{\alpha} C_g(R_{x,y}) \alpha(R_{x,y}) \left( (1 - \alpha(R_{x,y})). \right) \]

Note that the compositing is done in back-to-front order (i.e. from the background screen towards the eye).

After the implementation of the above algorithm, we tried its utility in visualizing two classes of data. The first one was the class of probability density functions of an electron in hydrogen atom. The second class, a set of fractals, was generated artificially over a regular grid using the mid-point displacement method. While the former represents a multivariate relationship whose form is analytically known, the latter is a case of discrete samples of data akin to what one obtains from numerical computations or experimental measurements.

It is well-known that the probability density function of an electron in hydrogen atom is given by the square of the modulus of the corresponding wave function \( \psi_{\text{nlm}}(r, \theta, \phi) \). Here \( nlm \) denote the quantum numbers of the state of the electron. In analytical form (ignoring scale factors)

\[ \psi_{\text{nlm}}(r, \theta, \phi) = C e^{-r/2} L_n^{1/2} (r) P_l^m (\cos \theta) e^{i m \phi}, \]

where \( L_n^{1/2} \) and \( P_l^m \) are the associated Laguerre polynomials and the associated Legendre polynomials respectively. \( C \) is the normalization constant which is in general a function of the quantum numbers. The factors \( P_l^m (\cos \theta) e^{i m \phi} \) together constitute what are known as the spherical harmonics \( Y_{lm} (\theta, \phi) \) which express the angular dependence of the corresponding wavefunction.

For visualizing the electron probability density function, the input data are obtained by computing \( |\psi|^2 \) over a regular cartesian grid of points for a given triplet of quantum numbers. Using the series expansions given in Morse and Feshbach we could evaluate both \( L_n^{1/2} \) and \( P_l^m \) without the need for computing large factorials to compose the summands. The associated Laguerre polynomials are given by

\[ L_n^{1/2}(r) = C \cdot F(-p|q + 1| r), \]

where \( F(a|b|r) \), the confluent hypergeometric function is evaluated using

\[ F(a|b|r) = 1 + \frac{a}{b} r + \frac{a(a+1)}{2! b(b+1)} r^2 \]

\[ + \frac{a(a+1)(a+2)}{3! b(b+1)(b+2)} r^3 + ... \]

For the associated Legendre polynomials we used

\[ P_l^m(z) = (1-z^2)^{m/2} T_{l-m}^m(z). \]
where

\[ T_n^m(z) = C_2 \left[ z^n - \frac{n(n-1)}{2(2n+2m-1)} z^{n-2} + \frac{n(n-1)(n-2)(n-3)}{24(2n+2m-1)(2n+2m-3)} z^{n-4} \ldots \right] \]

\( C_1 \) and \( C_2 \) are constants of proportionality which can be safely ignored. With the help of the above expressions, the probability density was computed over a grid of size \( 50 \times 50 \times 50 \). The exact computation of the overall normalization constant was omitted. Instead, a suitable scale factor was found out by trial and error, so that a substantial part of the density function would be enclosed within the chosen volumetric grid. The resulting volumetric data were rendered by the ray tracing algorithm over a screen window of size \( 120 \times 120 \) picture elements, or pixels for short. Figure 1a shows the set of probability density functions for all possible quantum states when \( n=6 \) and \( m \) is positive. The view point is on the x-axis and the origin is the view reference. In Figure 1b, the same set of functions have been volume rendered from a different viewpoint (150, 150, 500). For the purpose of a structural comparison, we also created images of the angular part of the probability density functions, i.e. \( |Y_{\phi_m}(\theta, \phi)|^2 \) with the aid of a surface rendering program. These are shown in Figure 1c.

The class of volumetric data that we tried to visualize next, was generated by means of the mid-point displacement technique. Employed for creating fractal data sets, this technique is described by Barnsley et al.

Figure 1. a, Entire volume rendering of all possible probability density functions \( |Y_{\phi_m}|^2 \) for \( n=6 \) and positive \( m \). While \( l \) varies from 0 to \( n-l \) along the columns, \( m \) varies from 0 to \( l \) along each row. The axis of symmetry is \( z \) and the viewpoint is on the x-axis. b, Entire volume rendering of the same functions as in a, seen from a different viewpoint (150, 150, 500). c, Surface rendering of the angular part \( |Y_{\phi_m}(\theta, \phi)|^2 \) of the \( |Y_{\phi_m}|^2 \) functions shown in a and b seen from (150, 150, 500).
for the case of one and two dimensions. We extended their algorithm in a natural manner for the case of three dimensions. One of our motivations to generate fractal data sets over a three-dimensional grid was to subject our implementation of Mark Levoy's algorithm to a good test. Besides, we were also hoping to create cloud-like images as a result of volume rendering a fractal volumetric data set. After experimenting with various values for the fractal dimensions we settled for a value of 3.7. For the standard deviation of the Gaussian random additions to the mid-points we chose a value of 10, while the mean was set to be 0. Use of different seeds for the random number generator was employed to generate different fractal data sets. With a view to creating some empty space in the grid, a suitable threshold value was subtracted uniformly from the values at every point and among the resulting values the negative ones were replaced by zeros. To ward off some undesirable boundary effects, the data were modulated by a decreasing ramp before thresholding. Figure 2a was created using one set of fractal data while Figure 2b using another. One cannot fail to notice the resemblance between these images and stellar clouds.

All the images were generated using the IRIS workstation 4D/20. The demands on the computing resources for rendering volumetric data are enormous. The summary of our observations with regard to the computing times required for various cases is given in Table 1. It may be relevant to mention here that the LINPACK rating of the IRIS 4D/20 workstation is about 0.9 MFLOPS and the computations to be performed in our implementation of ray tracing are LINPACK-like. It may also be noted that it is the floating-point computation speed of the system which is crucial as against the graphic performance. One sure way of speeding up the rendering is to use a parallel computer. Since the ray tracing algorithm is easily parallelizable, we are in the process of exploiting PACE³ type of parallel computers for speeding up our implementation; results will be reported later.

The need to visualize volumetric data, is met with in a number of disciplines, crystallography and radiology being two examples. In the former, one obtains the electron density distribution within the unit cell of a crystal to be studied. The volume rendering method implemented by us will be useful in this case. Likewise, the electron density distributions in crystals as obtained from band structure calculations can also be visualized. Indeed, such a visualization would be particularly useful when one repeats the calculation for various values of the unit cell parameters and tries to see

<table>
<thead>
<tr>
<th>Type of data set</th>
<th>Grid size</th>
<th>Window size</th>
<th>Time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability density functions</td>
<td>50 × 50 × 50</td>
<td>120 × 120</td>
<td>3 35</td>
</tr>
<tr>
<td>Fractal set-I</td>
<td>65 × 65 × 65</td>
<td>240 × 240</td>
<td>18 7</td>
</tr>
<tr>
<td>Fractal set-II</td>
<td>65 × 65 × 65</td>
<td>240 × 240</td>
<td>16 9</td>
</tr>
<tr>
<td>Fractal set-III</td>
<td>129 × 129 × 129</td>
<td>480 × 480</td>
<td>149 5</td>
</tr>
</tbody>
</table>

Figure 2 a. Image obtained by entire volume rendering of a threedimensional fractal data set, generated by the mid-point displacement method. The fractal dimension is 3.7. The size of the volumetric grid is 129 × 129 × 129 and the window is 480 × 480 pixels wide. b. Image obtained for another fractal data set of the same dimension and size.