Revisiting Histograms and Isosurface Statistics

Carlos E. Scheidegger   John Schreiner   Brian Duffy   Hamish Carr   Cláudio T. Silva

Abstract—Recent results have shown a link between geometric properties of isosurfaces and statistical properties of the underlying sampled data. However, this has two defects: not all of the properties described converge to the same solution, and the statistics computed are not always invariant under isosurface-preserving transformations. We apply Federer's Coarea Formula from geometric measure theory to explain these discrepancies. We describe an improved substitute for histograms based on weighting with the inverse gradient magnitude, develop a statistical model that is invariant under isosurface-preserving transformations, and argue that this provides a consistent method for algorithm evaluation across multiple datasets based on histogram equalization. We use our corrected formulation to reevaluate recent results on average isosurface complexity, and show evidence that noise is one cause of the discrepancy between the expected figure and the observed one.

✦

1 INTRODUCTION

In scientific and medical visualization, we commonly represent physical quantities as continuous functions defined over a continuous domain. These are constructed by resampling observed points using some reconstruction kernel defined on the underlying space or on the connectivity of a geometric grid. When visualizing data, we work directly on the continuous function, which is mathematically and computationally convenient.

Since humans are not good at assimilating large quantities of numerical data, visualization seeks to map numerical properties of this continuous function to visual properties such as colour, brightness and saturation or to geometric properties such as boundaries and edges. Thus, one of the first steps in visualization is to define a mapping from the function to visual properties. Defining this mapping often depends on understanding the frequency and possibly the spatial distribution of the numerical values.

Historically, function distributions have been computed with *histograms*, which simply count the number of samples with each function value. However, recent work by Carr et al. [2] has shown that there are serious problems in using histograms as representations of function distributions. In particular, histogram computation assumes that the reconstruction uses a *box filter* (or *nearest neighbor interpolation*).

Using this observation, Carr et al. [2] proposes several alternative statistics that converge faster than histograms. These are based on interpreting the *isosurface areas* as measurements of higher-order interpolants on the domain. In addition, the authors related their statistics (and, implicitly, the histogram) to the algorithmic complexity of isosurface rendering, demonstrating a larger ($O(N^{0.82})$) experimental result than the $O(N^{2/3})$ previously predicted [6]. While the proposed statistics certainly converge faster than histograms, two problems can be identified. First, these seem to converge to a slightly different result than histograms. Second, the mathematics suggest some counterintuitive results about the average complexity of isosurfaces in the domain. Here, we use the term "convergence" to mean that as we use increasingly finer grids, the computed functions approach some limit function.

In this paper, we address these issues by revisiting the development of the isosurface statistics using a celebrated result in geometric measure theory: Federer's Coarea Formula [11]. This formula relates integrals over level sets of a function to integrals over the domain on which the function is defined. Crucially, it has the effect of normalizing isosurface statistics to the packing density of the isosurfaces, allowing us to correct the problems identified and propose improved solutions. Moreover, this formula gives us additional results. In particular, we show that histogram equalization can be used as a normalization to get the $O(N^{2/3})$ result originally predicted. Throughout the paper, we use the term "level set" interchangeably with isosurface, but will choose the term "level set" when we want to disregard the particular isovalue associated to the set of points in the preimage.

Our contributions are as follows. We build on the work of Carr et al. [2], and introduce a more mathematically grounded approach based on Federer's Coarea Formula (FCF). It clarifies the subtle relation between histograms and isosurface areas, and, crucially, shows the role of gradient magnitudes in that relationship. It also provides a well-founded way to compute expectations over all isosurfaces in a volume. Practically, we suggest using the cell span as an approximation for gradient magnitude, and provide experimental evidence for the validity of this approximation. Finally, we revisit Carr et al.'s study on average isosurface complexity and study the effect of volume noise in those estimates.

This paper is organized as follows. Section 2 briefly reviews relevant previous work, while Section 3 summarizes the contents of the previous result [2]. We then introduce the FCF in Section 4, and use this mathematical tool to develop an improved formulation of isosurface statistics in Section 5, based on dividing the statistics by the local gradient magnitude of the continuous function. In Section 6, we extend the application of the FCF to the computation of average isosurface complexity, and show that there exist transformations that preserve level sets, but change the original isosurface statistics (we will make this notion precise). We also show how to compute the average complexity in a way that implicitly accounts for any such transformations, and explain the discrepancy in the original results. In Section 7 we confirm that the average isosurface complexity of a function, when sampled increasingly finely in a domain, is $O(N^{2/3})$. We also revisit the experiments of Carr et al. [2] in light of the corrected integral formulation. In our experiments, we find an even higher average complexity, of $O(N^{0.96})$. Finally, we show that noise seems to largely explain this high figure, by performing a set of experiments on synthetic data. We then summarize our results in Section 8 and speculate on future directions of research in Section 9.

2 PREVIOUS WORK

Histograms are ubiquitous in plotting, and also in computer graphics and visualization. It is one of the oldest techniques available for displaying data [5], and is often used as the basis of techniques such as histogram equalization [4], which defines a non-linear transfer function that attempts to distribute the intensity in an image uniformly over

---

- *Carlos E. Scheidegger, John Schreiner, and Cláudio T. Silva are with the Scientific Computing and Imaging Institute, University of Utah. E-mail: {cscheid, jmschrei, csilva}@sci.utah.edu.*
- *Brian Duffy and Hamish Carr are with the UCD School of Computer Science & Informatics, University College Dublin, Ireland. E-mail: {B.Duffy, hamish.carr}@ucd.ie.*
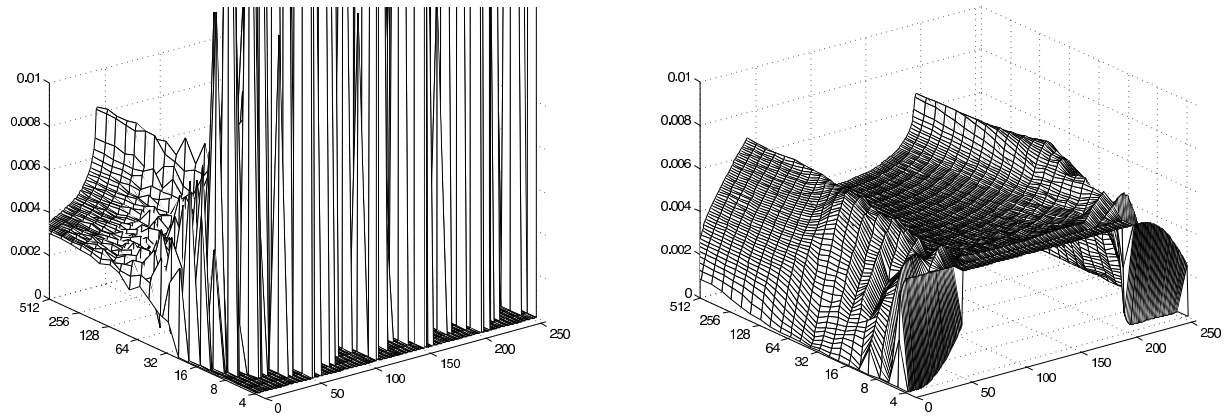
Fig. 1: Convergence properties of histograms (left) and isosurface area (right), from Carr et al. [2]. Note that the histogram's convergence is much poorer than the isosurface area; furthermore that the latter apparently converges to a slightly different result: at maximum resolution, the peaks are concave cusps for the histogram, but convex cusps for the isosurface area.

the range. Fundamentally, the histogram is a bar graph used to represent the distribution of function values in a population. In this bar graph, the independent variable represents the possible values of a set of observations, and the dependent variable represents the number of observations with a given value.

In computer graphics, histograms are often computed for sample values in a dataset. The histograms themselves are then used for equalization [4], transfer function construction [8] and detection of important isovalues [14, 15].

Isosurfaces are level sets of a function, and are commonly extracted and rendered using geometric approximations of the surface [9, 16]. Much subsequent work has gone into accelerating their construction—see [13] for a recent review. Crucially, most of the acceleration work has exploited the relationship between k, the output complexity (number of triangles) in the isosurface, and N, the input complexity (number of samples) in the data. A dimensionality argument has been used to estimate that $k \equiv O(N^{2/3})$ [6], but this estimate was not then tested rigorously.

Geometric statistics have also been used in visualization, principally in the form of the *contour spectrum* [1], which computes geometric and topological properties of isosurfaces and displays them in an interface to provide cues to interesting isovalues that can be used for guiding users during visualization.

Carr et al. [2] showed that histograms, isosurface complexity and geometric statistics were intimately related, and gave experimental results on the average complexity of isosurfaces in a range of freely available datasets. Since the present paper revises and extends these results, we defer full discussion of this work to the following section.

Our work depends heavily on a classic result in geometric measure theory—Federer's Coarea Formula, which describes the relationship between integrals defined on level sets and integrals on the space in which the level sets are defined. For an introduction to the topic, see Morgan's book [11]. To the best of our knowledge, the Coarea Formula has only very recently been presented to the computer graphics community by Mullen et al. [12]. We will return to this topic in Section 4.

3 REVIEW OF HISTOGRAMS AND ISOSURFACE STATISTICS

In the previous section, we identified the convergence of results on histograms, geometric isosurface statistics and algorithmic complexity. We now delve further into this convergence by reviewing the previously published results of Carr et al. [2].

Briefly, the paper points out that a histogram can be expressed as:

$$H(h) = \sum_{f(\mathbf{x}_i)=h} 1 \tag{1}$$

while the distribution of the function could be defined as:

$$\pi_f(h) = \int_{f^{-1}(h)} 1 \, dS \tag{2}$$

where $f : \mathscr{R} \subset \mathbb{R}^3 \to \mathbb{R}$ is a scalar field whose image is in the range $[h_0, h_1]$. A preimage $f^{-1}(h)$ is an *isosurface* of f, and h is an *isovalue*. As we will see shortly, Equation 2 is subtly incorrect, but the underlying point is clear: that the histogram is closely related to integrals defined over the isosurface at the function value h. It is this particular equation that we re-examine carefully, present a theoretical correction, and show experimental evidence for its superiority in terms of convergence to the histogram.

The Equation 2 result was then expanded to show that the histogram in fact assumes a nearest neighbour interpolant. Since isosurfaces explicitly represent the interpolant in the geometric extraction process, it follows that geometric statistics of the isosurfaces were better representations of function distributions. This was demonstrated by examining plots of histograms and isosurface statistics as the sampling density increased, in which it was clear that histograms had much poorer convergence properties. Figure 1 shows the convergence of histograms on the left and isosurface statistics on the right; it also shows one of the errors in the paper—*that the isosurface area in fact converges to a different result from the histogram.*

Having illustrated the convergence properties, the paper then illustrated the difference between the two sets of statistics for non-analytic data sets, describing consistent trends over a large number of freely available data sets. In this work, we show that using the isosurface area is not in fact the correct solution in that it does not converge to the same result as the histogram.

Finally, the paper measured the average isosurface complexity for all of the data sets on a log-log plot and demonstrated that the experimental complexity for uniform sampling in the range showed a $O(N^{0.82})$ relationship between k and N rather than the $O(N^{2/3})$ previously predicted [6]. With the machinery we introduce in this paper, we revise this result as well. For a single volume sampled increasingly densely, we can analytically show the $O(N^{2/3})$ result. However, for the same set of datasets collected on the internet as used in Carr et al. [2], we find a relationship closer to $O(N^{0.96})$.
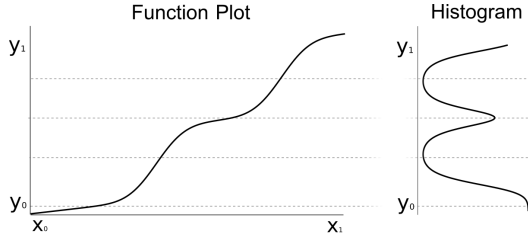
Fig. 2: A histogram of a continuous, one-to-one function. The histogram is sideways so that its relation to the original function is clear. Notice that the histogram peaks where the function derivative is low, and troughs where the derivative is high. If π_f didn't contain a term to account for the gradient magnitude, the histogram would be flat.
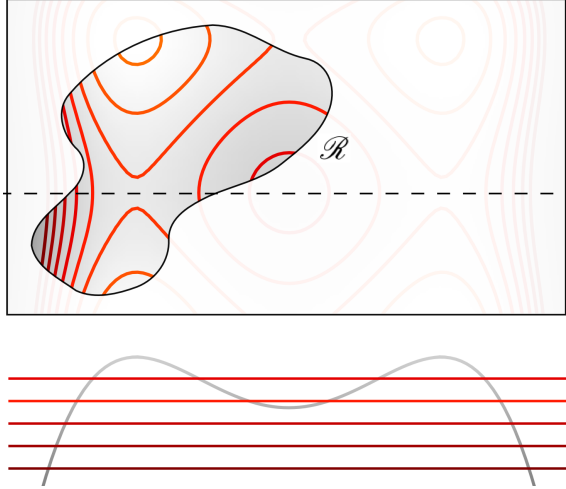

Fig. 3: Illustrating the coarea formula. The coarea formula allows us to go from an integral on the level sets of a (sub-)region $\mathcal{R}$ to an integral on the region itself. Even though isovalues of the level sets are uniformly spaced, changes in the gradient magnitude cause the level sets to have different local densities. It is exactly this change in density that is accounted for in the co-area formula by the $|\nabla f(x)|$ term. The dashed cross-section of the domain is plotted in the bottom.

We note that we are *not* advocating the use of a different range mapping for determining important isosurfaces, or that the distributions computed by Carr et al. [2] are more or less effective at informing the user about the volume. We are interested in finding histogram substitutes that converge as fast as previous proposals, while matching the histogram distribution in the limit. As we will show, this leads more generally to a mathematically sound way of determining integrals over all isosurfaces of a volume, when the particular isovalues of each isosurface are not important.

4 REVISITING HISTOGRAMS AS DISTRIBUTIONS

We begin by presenting Federer's Coarea Formula, the primary mathematical tool used in this work. The original result is extremely general in nature, and requires a level of sophistication that will be unnecessary for our purposes (see, e.g., Morgan's book [11]). We instead follow Mullen et al.'s presentation of the formula [12], who recently introduced the Coarea Formula to the computer graphics community as a useful geometry processing tool. At its core, the formula converts integrals over a range of level sets to integrals over the domain where the level sets are defined:

$$\int_{\mathbb{R}} \int_{f^{-1}(h) \cap \mathcal{R}} q(\mathbf{x}) \, dS \, dh = \int_{\mathcal{R}} q(\mathbf{x}) |\nabla f(\mathbf{x})| \, dV \qquad (3)$$
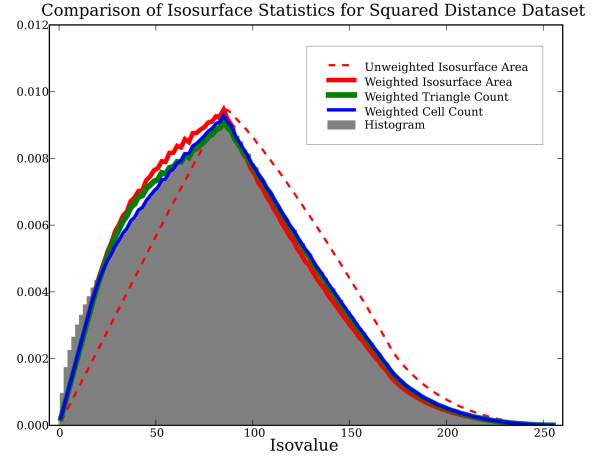

Fig. 4: Isosurface statistics for a spherical function distribution with non-constant gradient magnitude. Our improved isosurface statistics converge to the histogram even in the case of nonconstant gradient magnitudes in the domain. Notice although both weighted and unweighted triangle areas are smooth curves, indicating fast convergence, only the gradient weighted areas (and its approximations) converge to the histograms.

where $q : \mathcal{R} \to \mathbb{R}$ is any scalar function defined over the same domain as f, and f is a Lipschitz function. Note that in order to be consistent with the notation from Carr et al., the notation used in this paper is slightly different from Mullen et al.

In words, the formula says that if we want to integrate a certain function over a range of level sets of f, we can instead integrate the function directly over the domain of f, *provided we compensate for the local density of isosurfaces*. We will use this formula to provide an appropriate definition of the distribution function on a continuous domain. We would like to define a distribution function $\pi_f(h)$, that gives a measure of the "size" of the set of points for which $f(\mathbf{x}) = h$. We first define the *cumulative* distribution function:

$$
\begin{aligned}
C_f(h) &= P[f(\mathbf{x}) < h] \\
&= \frac{\int_{h_0 \leq f(\mathbf{x}) \leq h} 1 \, dV}{\int_{h0 \leq f(\mathbf{x}) \leq h_1} 1 \, dV} \\
&= \frac{1}{V} \int_{h_0 \leq f(\mathbf{x}) \leq h} 1 \, dV \qquad (4)
\end{aligned}
$$

We will assume without loss of generality that $V = 1$. Given this definition of C_f as an integral over a subset of the domain $\mathcal{R}$, we will use Federer's Coarea Formula to rewrite it in terms of integrals on the level sets. By defining $q(\mathbf{x}) = |\nabla f(\mathbf{x})|^{-1}$, and restricting the set of isovalues we are integrating over, we can write:

$$\int_{h_0}^{h} \int_{f^{-1}(\bar{h})} |\nabla f(\mathbf{x})|^{-1} \, dS \, d\bar{h} = \int_{h_0 \leq f(\mathbf{x}) \leq h} 1 \, dV \qquad (5)$$

This can be directly substituted into C_f:

$$
\begin{aligned}
C_f(h) &= \int_{h_0 \leq f(\mathbf{x}) \leq h} 1 \, dV \\
&= \int_{h_0}^{h} \int_{f^{-1}(\bar{h})} |\nabla f(\mathbf{x})|^{-1} \, dS \, d\bar{h} \qquad (6)
\end{aligned}
$$

By the Fundamental Theorem of Calculus, $\pi_f = C'_f$:

$$\pi_f(h) = \int_{f^{-1}(h)} |\nabla f(\mathbf{x})|^{-1} \, dS \qquad (7)$$

Note that this is in contrast with how π_f is defined by [2] (in this paper, Equation 2). As we will show, when this inverse gradient magnitude
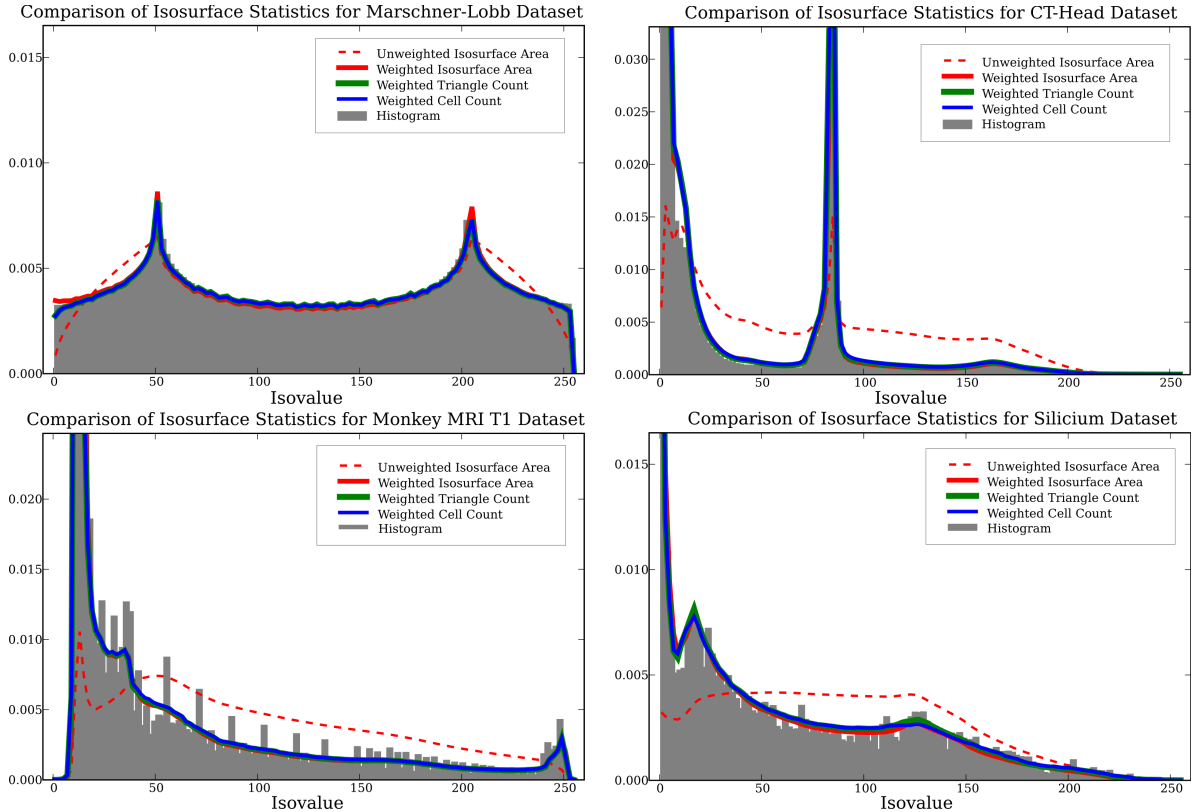
Fig. 5: Histograms and Corrected Isosurface Statistics. Notice that the unweighted isosurface areas show fast convergence, but produce different results from the histogram. The corrected term we introduce in this paper keeps the fast convergence, while still approaching the histogram distribution.

term is accounted for in the computations, the isosurface statistics investigated in that work converge to the same results as the histograms. One way of connecting the two results is to note that Carr et al.'s original formulation generates correct results whenever the volume satisfies a particular case of the Eikonal equation: $|\nabla f(x)| = k$ for some nonzero constant k.

To highlight the relationship between the gradient magnitude of a function and its histogram, we show two examples. In Figure 2 we show a simple one-dimensional example where we restrict ourselves to a one-to-one function. Notice that the histogram of a continuous function measures the density of values in the domain that attain a certain value in the image. Differentially, however, the density of the domain is inversely proportional to the derivative of the function at a point, since it is given by the "amount" of domain in an infinitesimally small slab of range values. If the derivative of the function (or more generally, its gradient magnitude) was not taken into account, we would expect the histogram of a one-to-one function to be constant — which cannot possibly be true in general. In Figure 3, we illustrate the co-area formula in two dimensions.

5 REVISITING ISOSURFACE STATISTICS AS DISTRIBUTIONS

With the correction presented above, we introduce an improved substitute for histograms which, unlike that presented in [2], converges to the histograms themselves. Instead of computing isosurface area directly, we need to weigh the area contributions by the inverse gradient magnitude of the level set. We also propose an approximation for the gradient magnitude at a level set that is easy to compute.

In Figure 4, we show the histogram of an analytic spherical distribution $f(\mathbf{x}) = |\mathbf{x}|^2$. Note that, unlike the one presented in [2], we are using *squared distances* from the origin, and that the gradient of this function is *not constant* across the volume. The previous formulation does not converge to the histogram distribution, while our im-

proved formulation does. The same approximations previously proposed, which use triangle count and active cell count as surrogates for isosurface area, also perform well when similarly corrected.

Carr et al. show that it is possible to use much cheaper alternatives such as active edge and active cell counts as a very good replacement for the potentially expensive isosurface extraction. In our case, there is an added complication that arises from the gradient magnitude in the correction term. In particular, we would like to compute the correction without having to refer to the original spatial distribution of values in the volume. If that were possible, we could improve the efficiency of the computations in the same way that using active cell counts can be used as an approximation of triangle counts. One such candidate is the *cell span*: the difference between maximum and minimum value. The cell span is an admittedly crude approximation of the gradient magnitude. Assuming cell width of 1, and depending on the geometric configuration of the values on the cell corners, the cell span might be off to up to a factor of $\sqrt{D}$, since the maximum and minimum values might be across a main diagonal of the D-dimensional cell. For $D = 3$, the cell span will overestimate the gradient magnitude in a cubic cell by ≈ 1.28 on average, if we consider all value configurations of extremal values are as likely to occur as one another. However, the histogram is by definition normalized, and since that factor is constant across all isovalues, the factor is canceled out. *We have found this potentially crude approximation to be essentially indistinguishable from using more sophisticated approximations.* Throughout the paper, we use this cell span approximation.

In Figure 5, we show the new histogram substitutes for a set of volumes of different modalities. These plots show the same general behavior as Figure 4: the new formulation converges faster than histograms, and it does so to the histogram distribution, unlike the previous proposal.
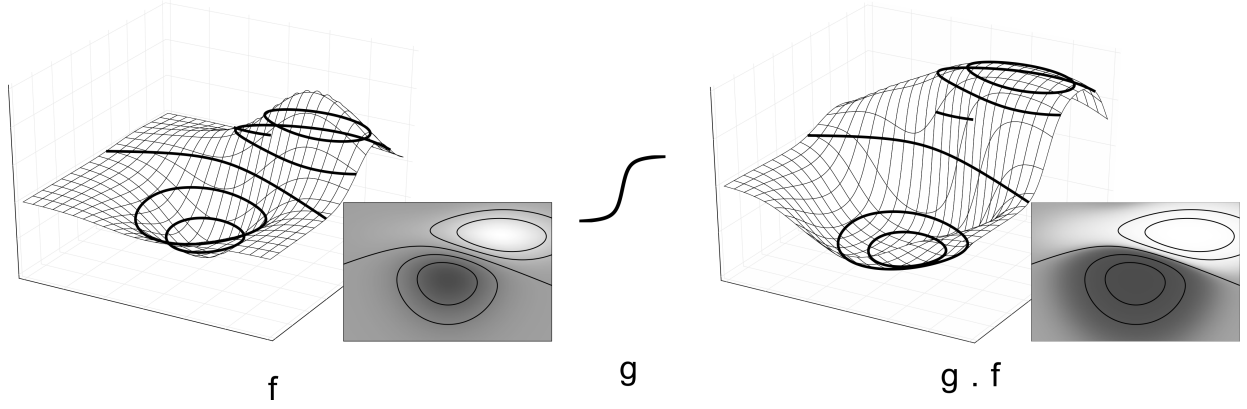
Fig. 6: Some transformations that preserve all level sets of a given volume change the value of $\tilde{A}$. The Coarea formula tells us that $\tilde{A}$ (defined in the text) is in essence the integral of the gradient magnitude. It is clear, then, that $\tilde{A}(f) \neq \tilde{A}(g \circ f)$, even though every level set in f also exists and $g \circ f$. In other words, the curves are generally shifted vertically (except for the fixed points of g), but each curve traces the same values in the domain.

6 AVERAGE ISOSURFACE COMPLEXITY

In this section, we investigate the average complexity of an isosurface in a volume. We use isosurface area as our measure of complexity, for the same reasons as Carr et al. [2]: it is mathematically convenient and, in most isosurface extraction algorithms, directly related to triangle count, which is arguably the most importantly computational unit to analyze in these algorithms.

The notion of average isosurface complexity is a delicate one. In particular, when computed naïvely, the average isosurface complexity might be different for two volumes that share *exactly* the same set of isosurfaces. In this section, we present a slightly different formulation that *is* invariant to these transformations. The straightforward way of defining the average isosurface complexity $\tilde{A}(f)$ is to simply evaluate the integral

$$\tilde{A}(f) \quad = \quad \frac{1}{h_1 - h_0} \int_{h_0}^{h_1} \int_{f^{-1}(h)} 1 \, dS \, dh \tag{8}$$

$$= \quad \frac{1}{h_1 - h_0} \int_{\mathbb{R}} |\nabla f(\mathbf{x})| \, dV, \tag{9}$$

where Equation 8 is the formulation used in [2], and Equation 9 is a direct application of the FCF. However, we can define $\bar{f}(\mathbf{x}) = g(f(\mathbf{x}))$, for some function $g : \mathbb{R} \to \mathbb{R}$ that is smooth and monotonically increasing. Notice that $|\nabla \bar{f}(\mathbf{x})| = g'(f(x))|\nabla f(x)|$, and so $\tilde{A}(f) \neq \tilde{A}(\bar{f})$, *even though f and $\bar{f}$ share the exact same set of isosurfaces ($f^{-1}(h) = \bar{f}^{-1}(g(h))$)*. If $\tilde{A}(f)$ is to be a measure of the average complexity of the level sets, we would like it to be invariant under transformations that preserve the level sets. These transformations preserve the level sets in the sense that every set of points in the preimage of the function f and value h (that is, the isosurface of f with isovalue h) exist in the function $\bar{f}$, *as the preimage of some value, namely $g(h)$* (that is, the isosurface of $\bar{f}$ with isovalue $g(h)$), and vice-versa. Figure 6 illustrates this preservation of level sets. In what follows, we describe how to change $\tilde{A}$ so that this is indeed the case.

6.1 Computing with histograms

In Section 4, we showed that the histogram of isosurface areas involves the integral of the inverse gradient magnitude over the level sets for it to be the real distribution function. We now show how this distribution function must be accounted for when computing averages over all isovalues. We use the standard definition of an average over a parameter with non-uniform probabilities:

$$A(f) \quad = \quad \int_{h_0}^{h_1} \pi_f(h) a_f(h) \, dh \tag{10}$$

$$a_f(h) \quad = \quad \int_{f^{-1}(h)} 1 \, dS \tag{11}$$

When the average is defined in this way, any transformation g applied to f will be exactly accounted for by $\pi_{g \circ f}$, so we will have $A(f) = A(\bar{f})$. Additionally, if g is specifically defined to be the histogram equalization function of f:

$$g(h) = \int_{h_0}^{h} \pi_f(\bar{h}) \, d\bar{h} \tag{12}$$

then the naïvely computed average area $\tilde{A}(f)$ will be exactly equal to the average area $A(f)$ that compensates for the density, since $\pi_{g \circ f}$ will be constant.

There are several ways to numerically estimate the value of Equation 10 via Monte Carlo integration. It can most directly be written as:

$$A(f) \approx \frac{h_1 - h_0}{n} \sum_{i=1}^{n} \pi_f(h_i) a_f(h_i) \tag{13}$$

where n is the number samples h_i taken uniformly in the range $[h_0, h_1]$. However, π_f may be difficult to evaluate in practice for arbitrary functions f. Instead, we could first perform histogram equalization, to produce a function $\bar{f}$ with a constant $\pi_{\bar{f}}$. In this case, we can approximate $A(f)$ as:

$$A(f) = \tilde{A}(\bar{f}) \approx \frac{1}{n} \sum_{i=1}^{n} a_{\bar{f}}(h_i) \tag{14}$$

This is a much simpler computation, but still assumes that histogram equalization can be done easily and robustly. However, histogram equalization on quantized data will often produce an imperfectly equalized result. If we can create the samples h_i proportionally to π_f, we can avoid having to either evaluate π_f or perform histogram equalization. This is easy to do, and we approximate $A(f)$ as:

$$A(f) \approx \frac{1}{n} \sum_{i=1}^{n} a_f(f(\mathbf{x}_i)) \tag{15}$$

where $\mathbf{x}_i$ is a random point uniformly chosen from $\mathscr{R}$. That is, if we choose the values h to average by evaluating f at a random point, rather than choosing it randomly between h_0 and h_1, we can implicitly account for the non-uniform distribution of isosurfaces, and have the computation be invariant under any transformation g of the ranges.

Interestingly, this tells us how we should compute *any* average over isosurfaces. In the context of this paper, we have defined $a(h)$ to be the area of the isosurface with isovalue h. However, a could be any function of the isosurface, such as the execution time of some algorithm on

$f^{-1}(h)$, or the maximum gradient magnitude at points in $f^{-1}(h)$. Then when computing the averages of these functions *over all isovalues h*, the distribution function must be taken into account appropriately.

7 FORMAL COMPLEXITY

We now turn to the question of average isosurface complexity over the volume. For the moment, we will look at the average *isosurface area* as a convenient proxy for the correct notion, which involves a sampled volume and the average active cell count.

We have rigorously defined $A(f)$ in such a way that it is invariant to transformations of its range, so we can now investigate how it changes as a function of a uniform scale of s. We define a new function f_s : $\mathscr{R}_s \to \mathbb{R}$, where $f_s(\mathbf{x}) = f(\mathbf{x}/s)$ and $\mathscr{R}_s = \{\mathbf{x}|\mathbf{x}/s \in \mathscr{R}\}$. We note that $\pi_{f_s} = \pi_f$ since it is a simple uniform scaling. Additionally,

$$\int_{f_s^{-1}(h)} 1 \, dS = s^2 \int_{f^{-1}(h)} 1 \, dS \qquad (16)$$

since dS is a 2-form. Therefore, the average area of the isosurfaces of f_s is:

$$\begin{aligned} A(f_s) &= \int_{h_0}^{h_1} \pi_{f_s}(h) a_{f_s}(h) \, dh \\ &= \int_{h_0}^{h_1} \pi_f(h) s^2 a_f(h) \, dh \\ &= s^2 A(f) \qquad (17) \end{aligned}$$

Since the volume of the domain will increase by $N = s^3$, we have the scaled volume's average area growing exactly by $N^{2/3}$, as suggested by Itoh and Koyamada [6]. Note, however, that this is when considering a *single* function f. As Carr et al. [2] suggest, it is possible to find pathological cases where the average area relative to the volume increases as $O(N)$, when the functions are not just scaled versions of some base f. For example, this will happen when f is a checkerboard defined over all $\mathbb{R}^3$, and f_N is defined to be f (not scaled, but extended over a larger region).

7.1 Revisiting the experiments

In addition to these theoretical results, we would also like experimental confirmation of the formal complexity of isosurfaces. In practice, the scalar functions are typically represented by a finite set of samples, which are subject to quantization and aliasing artifacts, and the isosurfaces are often approximated piecewise linearly by triangles. To test the formal complexity that is seen in practice under these circumstances, we have revisited the experiments in [2] using the appropriate definition for π_f. We can use any of the previously described methods of accounting for π_f when numerically estimating the average isosurface complexity. To avoid numerical artifacts of histogram equalization of quantized volumes, we use Equation 15.

Figure 7 shows the results of the experiments performed. We have used approximately 80 different datasets for these experiments. We have found that the average triangle count across all isosurfaces, using the appropriately defined notion above, grows at around $N^{0.96}$, when N is the number of samples in the volume, close to the pathological $O(N)$ cases. This figure is even higher than the experiments of [2], which found the relationship to be approximately $N^{0.82}$. The difference in the results comes from compensating for the local density of isovalues when computing the average surface areas. Notice that the overall slope is bigger than the slope of the fits particular to any modality. A full investigation of the reason is not possible without a much bigger set of volumes than what we had available. However, we speculate that this is caused partly by a combination of outliers, a simplified growth model (the real data probably does not follow a power law exactly) and disparity of the volume sizes: the simulated data sets were typically smaller than the measured and medical data sets.
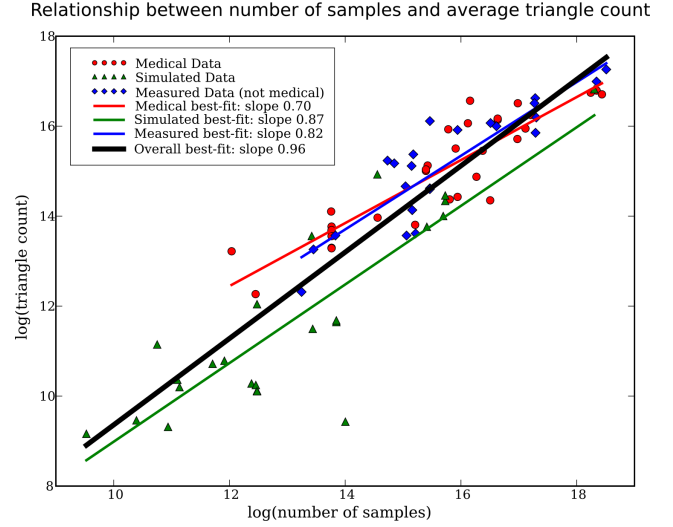

Fig. 7: Plot of input sample size versus average triangle count. Our experiments show an even higher power ($O(N^{0.96})$) than previously reported ($O(N^{0.82})$).

7.2 Noise and average complexity

The new figure seems to indicate that the average isosurface on datasets used in practice is close to what has been considered a pathological case, which is unexpected. In this section, we show that *noise* seems to be a main cause of this inflated complexity. We start by showing, in Figure 10a-c, the effect of increasing noise on a spherical isosurface. As the noise level increases, the surface essentially fills an increasingly thick region around the original isosurface. Figure 10d shows a typical random isosurface extracted during the computation of the results in Figure 7. Note that almost the entire volume is similarly punctuated by small individual connected surface components that arise from the presence of *noise* in f.

We know from Section 7 that a single volume, when sampled with increasingly dense points, will result in a $O(N^{2/3})$ average complexity. However, volume noise is usually modeled as being independent of the sampling resolution, so as we increase the sampling density, the volume is no longer the same. The results of our experiments are shown in Figure 8. In that experiment, we computed the average isosurface complexity for a series of synthetic sphere and Marschner-Lobb datasets, of increasing resolution, and examined the effects of additive gaussian noise on the complexity. As can be seen, a noiseless sphere has an average complexity close to $O(N^{2/3})$, while all of the volumes quickly approach complexity $O(N)$ as the noise increases.

Additionally, aliasing seems to play a similar role to noise in the observed complexity. By removing the Marschner-Lobb resolutions that are prone to aliasing (all volumes with less than 40^3 samples), the graph of the growth exponent as a function of the noise level is noisier due to fewer data points, but there is a trend that starts at $O(N^{2/3})$, and approaches $O(N)$ asymptotically. When volumes that have aliasing effects are used, the order of growth seems to be inflated because as the sample size increases, more of the higher frequency geometry is resolved, resulting in an increased number of triangles.

Since most of the datasets we collected were quantized to 8 or 16 bits, we also experimented with varying the quantization level of a particular dataset and computing the average order of growth. This experiment is presented in Figure 9. The experiment setup is similar to the previous one, but instead of increasing the noise level, we increase the quantization. If quantization by itself were a significant factor, then the plot would show a decreasing trend as we use more bits to store the data. We did not observe, however, such a trend. There are theoretical issues of examining severely quantized volumes with our approach based on the co-area formula (see Section 8), so a deeper investigation is necessary.
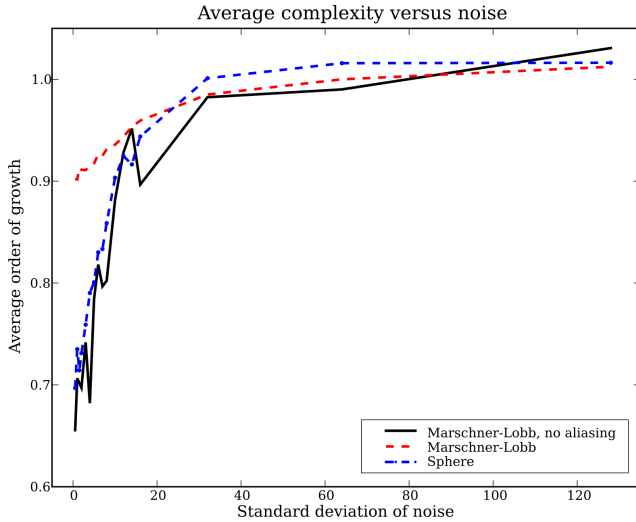
Average complexity versus noise

Fig. 8: Plot of noise level versus average complexity. Noiseless volumes have average complexity close to $O(N^{2/3})$, and noise moves the complexity towards the $O(N)$ asymptote. Notice that aliasing causes a higher initial complexity measurement.

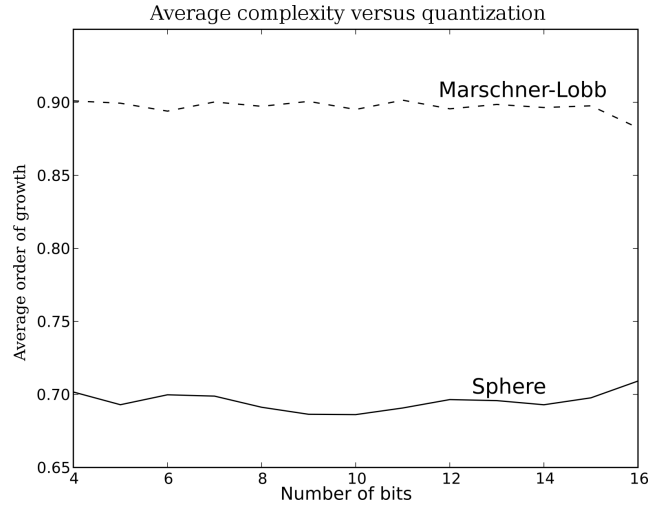Average complexity versus quantization

Fig. 9: Quantization level versus average complexity, with no noise. Quantization does not seem to influence the order of growth in either direction.

8 DISCUSSION

The formulation we propose in this work converges to the histogram distribution much faster than directly evaluating the histogram. Recently, there has been work in using *local* histograms for scientific visualization [10]. The results presented here should be attractive in these scenarios. Local histograms have less available data, and so are potentially much noisier than their global counterparts, making smoother histogram substitutes such as the ones we present an attractive possibility.

Section 6.1 focused on how to perform integrals of isosurface area and triangle count over all isosurfaces in the volume in a way that is invariant to range transformations. More generally, however, the techniques presented can be seen as a way to compute *expectations of random variables on isosurfaces* that is invariant to the transformations. The three methods we propose (Equations 13, 14 and 15) can be used to evaluate expectations of any random variables, by simply using a different $a_f(h)$. For example, comparisons of running times for isosurface extraction algorithms can use this formulation to sensibly measure the average time for a given volume.

The gradient magnitude correction we introduce must be treated carefully in situations where the volume gradient is zero. Volume quantization, in particular, tends to generate such volumes. While the Coarea formula ceases to be directly applicable in those cases, we believe that discrete counterparts of the formula (or volume perturbations in the spirit of Simulation of Simplicity [3]) could make computation of π_f meaningful in these scenarios. Additionally, some datasets include regions where there are clear discontinuities in the data (wave simulations with shocks are the canonical example). We have not experimented with these datasets, although it certainly warrants further investigation.

One limitation of the presented formulation is that, in a precise sense, it regards every isosurface as equally important. We know, however, that many scientific visualization techniques exploit particular relationships between the isovalues and the underlying volume [7]. This means that while our method provides a mathematically sound method for computing integrals over all isosurfaces, the user might actually be interested in weighting them in a *semantically* meaningful way. How to integrate these two aspects remains an exciting avenue for future investigation. Additionally, we emphasize that we are not claiming that Carr et al. [2]'s formulation is more or less effective at informing the user. That is a very important question that merits careful future examination.

9 CONCLUSIONS AND FUTURE WORK

In this paper, we study the relationship between histograms and isosurfaces statistics following on the footsteps of Carr et al. [2]. We identify shortcomings of their theoretical framework, and using the Coarea formula from geometric measure theory, we are able to correct their method to correctly accounts for the non-constant gradient magnitude across a volume. This corrected formulation maintains its superior convergence rate over histograms, while now converging to the actual histogram. We have also shown that it is possible to define average isosurface complexity in a way that is invariant to transformations that change the gradient magnitude but preserve level sets. This formulation explains some of the unexpected results obtained experimentally, and provides an analytical demonstration that the complexity of average isosurfaces in a single volume scales as $O(N^{2/3})$, confirming, for a particular scenario, the previous estimate of Itoh and Koyamada [6].

In future work, we would like to investigate more deeply the influence of volume artifacts such as noise, quantization and aliasing in the average isosurface complexity. We also want to investigate possible extensions of this work for joint histograms of function and function derivatives, which are quite popular in scientific visualization applications.

REPRODUCIBILITY

ACKNOWLEDGMENTS

REFERENCES

[1] C. Bajaj, V. Pascucci, and D. Schikore. The contour spectrum. In *Proceedings of IEEE Visualization 1997*, pages 167–173, 1997.

[2] H. Carr, B. Duffy, and B. Denby. On histograms and isosurface statistics. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of IEEE Visualization 2006)*, 12(5):1259–1265, 2006.
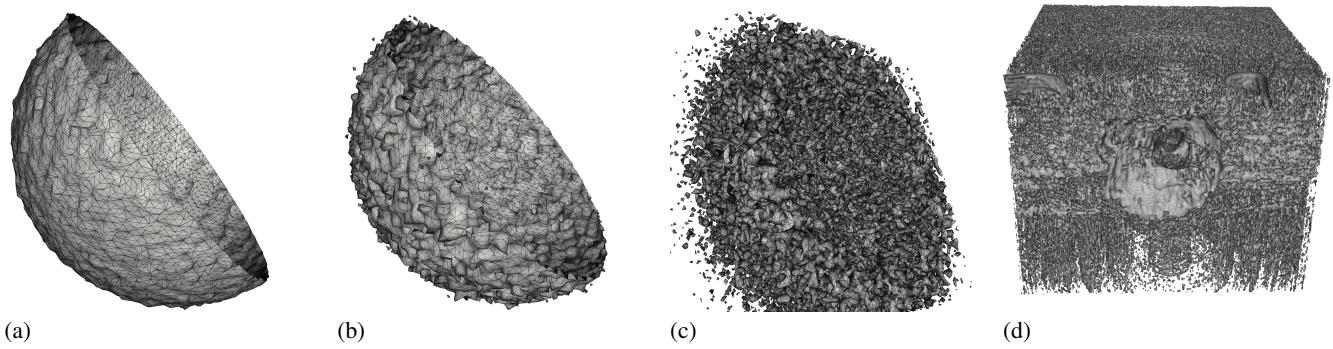
Fig. 10: Example isosurface extracted from a distance function $f(\mathbf{x}) = |\mathbf{x}|$ with added gaussian noise with standard deviation increasing from 1 (a), to 3 (b), and 30 (c). This effect can also be seen when extracting isosurfaces from typical publicly available experimental volumes (d). Notice that noise makes the volume essentially pathological.

[3] H. Edelsbrunner and E. P. Mucke. Simulation of simplicity: A technique to cope with degenerate cases in geometric algorithms. In *Symposium on Computational Geometry*, pages 118–133, 1988.

[4] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice-Hall Inc., Englewood-Cliffs, NJ, 2nd ed. edition, 2002.

[5] Y. Ioannidis. The history of histograms. In *Proceedings of Very Large Databases (VLDB) 2003*, 2003.

[6] T. Itoh and K. Koyamada. Automatic isosurface propagation using an extrema graph and sorted boundary lists. *IEEE Transactions on Visualization and Computer Graphics*, 1(4):319–327, 1995.

[7] G. Kindlmann and J. W. Durkin. Semi-automatic generation of transfer functions for direct volume rendering. In *IEEE Symposium on Volume Visualization*, pages 79–86, 1998.

[8] J. Kniss, G. Kindlmann, and C. Hansen. Interactive volume rendering using multi-dimensional transfer functions and direct manipulation widgets. In *Proceedings of IEEE Visualization 2001*, pages 255–262, 2001.

[9] W. E. Lorenson and H. E. Cline. Marching Cubes: A High Resolution 3D Surface Construction Algorithm. In *Proceedings of ACM SIGGRAPH 1987*, pages 163–169, 1987.

[10] C. Lundstrom, P. Ljung, and A. Ynnerman. Local histograms for design of transfer functions in direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 12(6):1570–1579, 2006.

[11] F. Morgan. *Geometric Measure Theory — A Beginner's Guide*. Elsevier, 2000.

[12] P. Mullen, A. McKenzie, Y. Tong, and M. Desbrun. A variational approach to eulerian surface processing. *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH 2007)*, 26(3):66, 2007.

[13] T. S. Newman and H. Yi. A survey of the marching cubes algorithm. *Computers And Graphics*, pages 854–879, 2006.

[14] V. Pekar, rafael Wiemker, and D. Hempel. Fast detection of meaningful isosurfaces for volume data visualization. In *Proceedings of IEEE Visualization 2001*, pages 223–230, 2001.

[15] S. Tenginakai, J. Lee, and R. Machiraju. Salient iso-surface detection with model-independent statistical signatures. In *Proceedings of IEEE Visualization 2001*, pages 231–238, 2001.

[16] G. Wyvill, C. McPheeters, and B. Wyvill. Data Structure for Soft Objects. *Visual Computer*, 2:227–234, 1986.