SALIENT ISO-SURFACE DETECTION WITH MODEL-INDEPENDENT STATISTICAL SIGNATURES

Shivaraj Tenginakai, Jinho Lee, Raghu Machiraju
Computer and Information Science
The Ohio State University

ABSTRACT

Volume graphics has not been accepted for widespread use. One of the inhibiting reasons is the lack of general methods for data-analysis and simple interfaces for data exploration. An error-and-trial iterative procedure is often used to select a desirable transfer function or mine the dataset for salient iso-values. New semi-automatic methods that are also data-centric have shown much promise [1][7]. However, general and robust methods are still needed for data-exploration and analysis. In this paper, we propose general model-independent statistical methods based on central moments of data. Using these techniques we show how salient iso-surfaces at material boundaries can be determined. We provide examples from the medical and computational domain to demonstrate the effectiveness of our methods.

CR Categories and Subject Descriptors: 1.3.3 [Computer Graphics]: Picture/Image Generation - Viewing Algorithms; 1.3.6 [Image Processing and Computer Vision]: Feature Measurement - Feature Representation.

Keywords: Iso-values, Transfer Functions, Surface Extraction, Direct Volume Rendering

1. INTRODUCTION

Direct volume rendering is a key technology for the visualization of large 3D datasets from scientific or medical applications. Consider the example of visualizing a 3D Computed Tomography (CT) dataset of the Visible Man from the National Library of Medicine. It is known for instance that the bowels can be extracted when the intensity value is set at *700*. The muscle can be displayed when intensity range includes the value *1010*. For many datasets, one is often ignorant of the salient iso-values; rather they have to be determined. In other words one needs to divide the voxel intensity space into segments that delineate homogenous materials, if at all possible. This is akin to feature mining in sample space. Similarly, one can employ transfer functions for opacity and color that best produce material interfaces and boundaries. Salient iso-value detection and transfer function design are related [7] and in this paper we will limit our discussion to the former.

Pat Hanrahan called data exploration (including transfer function design) one of the top ten problems in volume visualization in his inspiring keynote address at the *Symposium on Volume Visualization '92*. Recent research has focused on automatic and semi-automatic techniques for creating transfer functions and data exploration [1][2][4][5][7][8]. In the panel session on transfer function design at the Visualization'00 conference [10] three classes of techniques were identified.

Department of Computer and Information Science,
395 Dreese Laboratories, 2015 Neil Avenue Mall,
The Ohio State University, Columbus, Ohio.
([tenginak, leeji, raghu]@cis.ohio-state.edu).

One of the classes included techniques that required an error-and-trial approach. The other classes contained techniques that are either image- or data-centric. Of all the techniques, it was felt that data-centric techniques held most promise. These techniques required assumptions to be made about the data [7] or that computable signature functions be obtained [1]. Image-centric methods [8] on the other hand are based on searching a large space and offer little user control. The effectiveness of error-and-trial methods rests very heavily on the expertise and intuition of the user. In any case, the need for new work that offered general solutions was felt. It was also realized that these methods should be at least semi-automatic giving assistance to the practitioner rather than being completely automatic. Methods that eliminate the human from the exploration process are dangerous and should be avoided since they can generate images that may fulfill the expectations of the observer, but are not necessarily true to the nature of the data.

In this paper, we propose to employ data-signatures to explore data. These signatures are obtained from localized *k*-order central moments. The histogram is an example of the first-order central moment. There exists a strong relationship between the various central moments and this relationship is influenced by the presence of material interfaces. Thus, these signatures can be used to locate salient iso-values which occur at material interfaces. An attractive trait is that these methods are general and robust to noise. No assumption about the boundary's thickness or it's suitability is made for a given dataset. Also, the signatures based on moments are all related in a more comprehensible manner than those used in [1] and are relatively much easier to compute. Finally, our proposed method is not completely automatic; it provides ample cues to the presence of material interfaces.

Section 2 describes previous work in salient contour extraction and transfer function design. In Section 3, we introduce various mathematical concepts and derive and analyze our fundamental equation, the general moment equation. Section 4 describes the use of local higher order moments to detect boundaries in spatial domain and sample space. In Section 5, we present results that validate our analysis, while Section 6 offers a summary and describes future work.

2. PREVIOUS WORK

In general, the visualization process should be guided by information about the goal of the visualization, and specific information about the particular dataset in question. There exist techniques that employ a model of the desired feature, e.g., the boundary between homogeneous regions. This approach was taken by Kindlmann and Durkin [7]. An initial step in this process is the definition of a boundary. The boundary is essentially a Gaussian smoothed step function. The spatial component of the boundary is then removed by creating a 3D histogram of the data value and its first and second derivative. This histogram is very informative and the presence of sharp boundaries can be easily discerned from the projected plots of f vs. f' and f vs. f''. The number of zero-crossings in the second plot (or maximas in the first plot) essentially determines the number of interfaces. Based on analysis of

Presented at IEEE Visualization 2001
October 21 - October 26, 2001
San Diego Paradise Point Resort, San Diego

this histogram, a distance function is created which tries to bridge the unintuitive space of data values to a synthetic, but intuitive, spatial domain: a signed distance to the middle of the nearest boundary. The calculation of the distance function is largely automated. Defining opacity as a function of position within a boundary becomes a more intuitive task than defining opacity as a function of data value. The success of this method is limited when the boundary model is not Gaussian and far from being ideal. Other notable work has been conducted by Sato et al. [11]. They define a gradient based measure for certain shapes of human tissue.

What information about the dataset can one exploit? Bajaj and his associates devised the contour spectrum which can guide the selection of iso-values for contouring [1]. A 1D plot of geometric and topological characteristics or signatures (volume and gradient integral) are plotted against the function value. A viable selection of salient iso-values is thus facilitated. The contour spectrum technique does not employ any particular model for the boundary interface. Rather, they exploit geometrical and/or topological properties of the volume. Statistical techniques are gaining popularity in data-mining applications. Only a few reported statistics-based methods exist for volume data-analysis and exploration. In [3][6] statistical inference methods (e.g. Bayesian) are employed to determine the material density of each voxel. Multiscale statistical techniques have also been proposed by Yoo [13]. Histograms and local higher order moments have been used in computer vision and image analysis research to describe shapes of objects. The histogram certainly has some potential as described in [7]. However, not much has been explored with histograms and other statistical moments as a statistical entity to facilitate volumetric data-analysis.

Instead of tweaking input parameters another approach such as inverse design can be used for data-analysis. Inverse design is a general paradigm for computer-aided design of graphics, where the user supplies an objective function over the output values. This function will generate a high score (say) for desirable output values, and a low score for other output values. The computer then searches for a set of input parameters that will maximize the objective-function score. He et al. implemented a genetic algorithm to seek transfer functions with limited success [5]. A primary reason for failure is the difficulty of specifying a suitable objective function. A more viable approach entails a search that creates an ensemble of unique images which are then displayed in such a fashion that the differences are easily discerned. Embodying this approach is the Design Galleries approach that was used to explore parameter spaces for a host of graphics and animation applications [8]. Generated images are used in an evolutionary search algorithm in the parameter space. On termination, the chosen images are generated by a unique set of transfer functions. The problem with inverse design methods are that they are automatic in nature and do not allow for much control to be exercised on the process.

3. Statistical Signatures of Data

In this section we first define the concepts related to higher order moments. We then introduce our definition of a physical boundary, and proceed to define local higher order moments. This is followed by a derivation of the general moment equation, and its analysis. We use results of this section to present techniques for detecting boundaries in Section 4.

3.1 Higher Order Moments (HOMs)

Higher order moments (HOMs) are model-independent statistical estimators of central tendency of a data distribution, i.e. they measure the tendency of a distribution to cluster around some particular value [12]. Usually, the value around which clustering of a distribution, $x_1, x_2, x_3,\ldots, x_N$, is measured is the *mean (M)* of the distribution, given by:

$$M = \frac{1}{N}\sum_{j=1}^{N} x_j \tag{1}$$

The *higher order moment* of the same distribution is defined as:

$$m_k = \frac{1}{N}\sum_{j=1}^{N} (x_j - M)^k \tag{2}$$

where, m_k is the *higher order moment* of order *k*. We can see that there are infinitely many HOMs. The second moment evaluates variance, while the third and fourth moments evaluate skewness and kurtosis. For a symmetric set of samples, the skewness evaluates to zero, while, for positively skewed samples, the histogram of the sample will have a heavy tail on the right. A sample set obtained from the *Normal* distribution will have kurtosis equal to zero. When a sample is more dense than that derived from a normal distribution, then the sample's likely distribution is ascribed a negative kurtosis value. Otherwise it is positive. Thus, one can completely characterize the shape of any function if infinite number of central moments are known. HOMs are used in nonlinear signal processing for signal and image estimation [9].

3.2 Boundary Model and Local Higher Order Moments

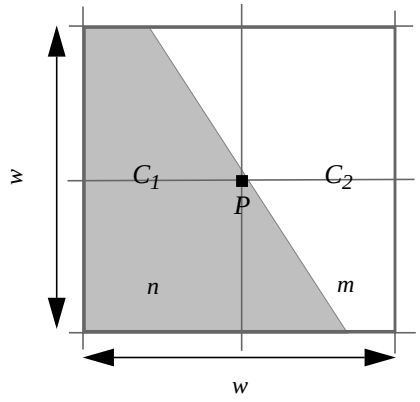


Figure 1: The boundary model ($w = 3$) for a small *2D* window. C_1 and C_2 are the two materials separated by a material interface.

The methods presented in this paper are for the following model of boundary or material interface (though, all the discussion below is for R^2, it generalizes easily to higher dimensions). Consider a window *W* of size $w \times w$ in a two-dimensional dataset, centered at a point *P*, (see Figure 1). At the boundary sample values of points in *W* change from C_1 to C_2. Here we assume that *W* is small enough to contain just one boundary. Similar mixture model was employed in [6]. However, we use much simpler analysis.

Since the size of *W* is $w \times w$, it follows that there are w^2 particles or sample points in *W*. We call the particles with sample value C_2 *boundary* particles, since they are introduced by the presence of the boundary. Further, let *m* be the number of such particles. Also, let *n* be the number of particles or sample points with sample value C_1. We call all such particles *non-boundary* particles. It follows

that, $n + m = w^2$. Assuming that our algorithm sweeps points left to right, we make the following observations:

- if $m = 0$ or $n = 0$, then there is no boundary in *W*, and we classify such a window as a *non-boundary* region.
- if $m > 0$ and $n > 0$, then there exists a boundary in *W*, and we classify such a window as a *boundary* region. Further,
  - if *n* is greater than *m*, then the majority of particles in *W* are *non-boundary* particles, and the point *P* lies *left* of the boundary. We classify such a window as a *left-boundary* region.
  - if *m* is greater than *n*, then the majority of particles in *W* are *boundary* particles, and the point *P* lies *right* of the boundary. We classify such a window as a *right-boundary* region.
  - if *n* equal to *m*, then the number of *boundary* particles and *non-boundary* particles in *W* are equal, and the point *P* lies *on* the boundary. We classify such a window as an *on-boundary* region.

For purpose of further analysis, we introduce the concept of *local higher order moments (LHOMs)*. LHOMs are HOMs calculated over the distribution of the sample values in a window *W*. Associated is also the concept of *local mean* which is the *mean* of sample values in *W*. The equation of *local mean (LM)* could be derived from Equation 1 as:

$$LM = \frac{1}{w^2}\sum x \qquad ,(\forall x \in W) \tag{3}$$

and the *LHOM* of order *k* is thus:

$$m_k = \frac{1}{w^2}\sum (x - LM)^k \quad ,(\forall x \in W) \tag{4}$$

It should be noted that for any *W* there are infinitely many LHOMs. We now show the derivation of the general moment equation based on the concepts introduced above.

3.3 Derivation of the General Moment Equation

In this section we proceed to simplify Equation 4, based on our model of boundary. Using our boundary model, the local mean is:

$$LM = \frac{(nC_1 + mC_2)}{(n+m)} \tag{5}$$

This model essentially assumes that any one material follows the binomial distribution in this small area. Further, the expression LHOM of *k-th* order reduces to:

$$\begin{aligned} m_k &= \frac{1}{(n+m)}\left\{ n[C_1 - LM]^k + m[C_2 - LM]^k \right\} \\ &= \frac{1}{(n+m)}\left\{ n\left[\frac{m(C_1 - C_2)}{(n+m)}\right]^k + m\left[\frac{n(C_2 - C_1)}{(n+m)}\right]^k \right\} \\ &= \frac{(C_2 - C_1)^k}{(n+m)^{k+1}}(mn^k + (-1)^k nm^k) \end{aligned} \tag{6}$$

Let $\Delta = C_2 - C_1$, and substitute $n + m = w^2$, in the above equation. This yields the general moment equation:

$$m_k = \frac{\Delta^k}{(w^2)^{k+1}}(mn^k + (-1)^k nm^k) \tag{7}$$

Equation 7 can be rewritten as:

$$m_k = \frac{\Delta^k}{(w^2)^{k+1}} f(m, n) \tag{8}$$

where *f(m,n)* is defined as the *distribution function*. It determines behavior of the *k-th* order LHOM in boundary regions. We now analyze Equation 7 for moments of various orders and examine various LHOMs for boundary presence.

3.4 Analysis of the General Moment Equation

If $\Delta \neq 0$ then *W* is a *boundary* region (else C_1 would be same as C_2, and Δ would be zero). From Equation 7, it is clear that for any non-zero Δ even-order LHOMs cannot be zero, i.e. a *boundary* region implies existence of non-zero even-order LHOMs. Conversely, if any even-order LHOMs is non-zero in a *W*, then Δ cannot be zero in that window *W*. Also, any non-zero, even-order LHOM implies that the associated region is a *boundary* region. The same argument holds for odd-order LHOMs, except that a odd-order LHOM will be zero if the associated region is an *on-boundary* region, and will be non-zero if the associated region is either a *left-boundary* region or a *right-boundary* region. Since both even-order and odd-order LHOMs are zero in *non-boundary* regions, we confine our interest to *boundary* regions.

Second Order LHOM: The equation for second-order LHOM is obtained as:

$$m_2 = \frac{\Delta^2}{(w^2)^2}(mn) \tag{9}$$

It can be shown that this equation has one maxima at $m=n=w^2/2$, i.e. when the associated region is an *on-boundary* region. The value of m_2 at the location of this maxima is $0.25\Delta^2$, which is independent of the size of the associated region and the actual distribution in the region.

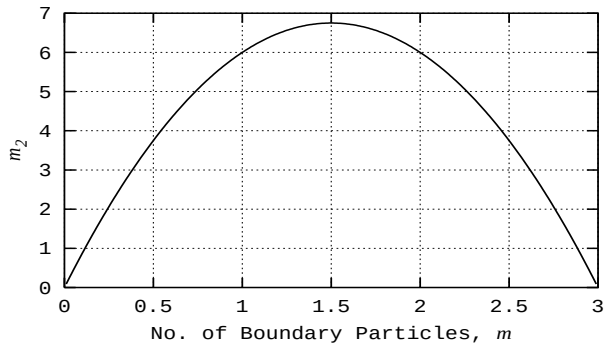
Third Order LHOM: The moment equation for the third-order LHOM reduces to:

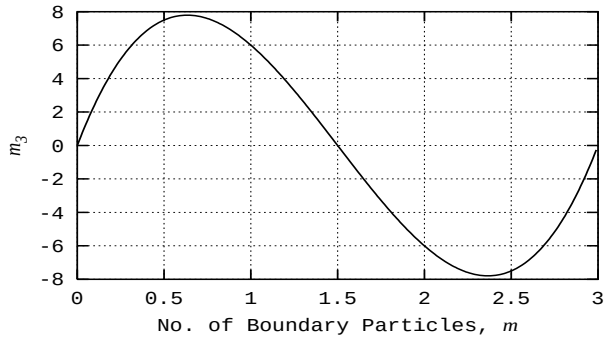
$$m_3 = \frac{\Delta^3}{(w^2)^3}mn(n-m) \tag{10}$$

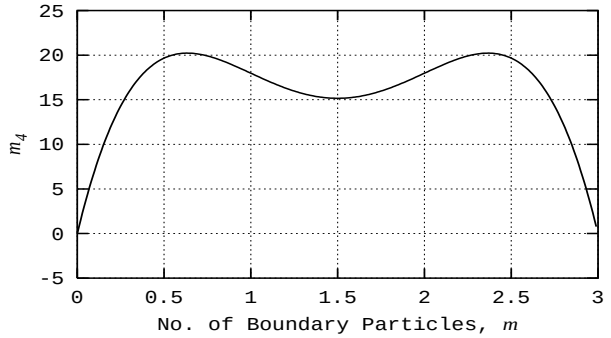
This equation has one maxima at $0.21w^2$ and a minima at $0.79w^2$ as measured from the center of the cell. The value of m_3 at the location of the maxima is $0.1\Delta^2$ and its value at the location of the minima is $-0.1\Delta^2$. Also, this equation is zero at $m=n=w^2/2$, i.e. when the associated region is an *on-boundary* region. Here again, it is to be noted that the value of m_3 at the location of maxima and minima is independent of the size of the associated region and the actual distribution in the region.

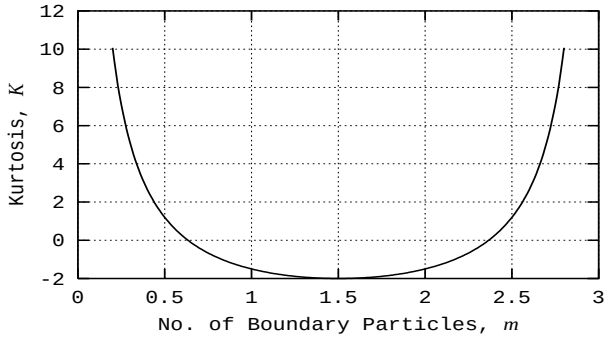
Fourth Order LHOM: The general moment equation for the fourth-order moment reduces to:

$$m_4 = \frac{\Delta^4}{(w^2)^4}mn(w^4 - 3mn) \tag{11}$$









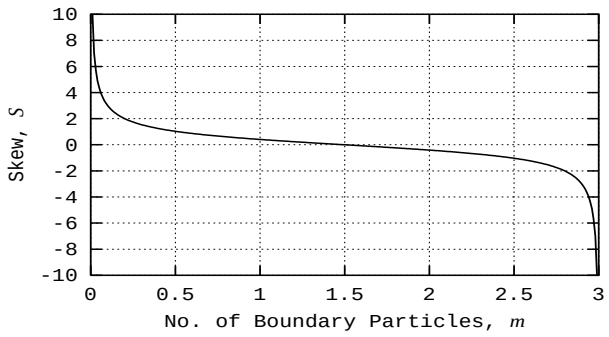


Figure 2: The plot of the distribution function for various LHOMs (*one dimensional window of size 3 is used*).

It can be shown that this equation has two maximas at locations $0.21w^2$ and $0.79w^2$ respectively. The value of m_4 at the location of these maximas is $0.083\Delta^4$. Also, this equation has one local minima at $m=n=w^2/2$, i.e. when the associated region is an *on-boundary* region. The value of m_4 at the location of this minima is $0.0625\Delta^4$.

Skew: This is a non dimensional quantity, and it characterizes the degree of asymmetry of a distribution around its mean. It is defined as:

$$S = \frac{m_3}{(m_2)^{3/2}} \tag{12}$$

It can be further shown that the equation for *skew* reduces to:

$$S = \frac{(n-m)}{\sqrt{nm}} \tag{13}$$

Also, it can be shown that skew is monotonically decreasing and is zero at $m=n=w^2/2$, i.e. when the associated region is an *on-boundary* region. It follows that *left-boundary* regions have positive *skew,* and *right-boundary* regions have negative *skew.*

Kurtosis: This another non-dimensional quantity which measures the relative peakedness or flatness of a distribution. It is defined as:

$$K = \frac{m_4 - 3m_2^2}{m_2^2} \tag{14}$$

It can be shown that the general moment equation for *kurtosis* in a region W reduces to:

$$K = \frac{w^4}{mn} - 6 \tag{15}$$

It can shown that this equation has a minima at $m=n=w^2/2$, i.e. when the associated region is an *on-boundary* region. The value of *kurtosis* at the location of this minima is *-2*. It is important that for the *on-boundary* regions the value of *kurtosis* is a constant. Figure 2 shows the plots of the *distribution function* corresponding to the various LHOMs. The plots confirm the analysis presented above for the second-, third-, and fourth-order LHOMs, and also for *skew* and *kurtosis.*

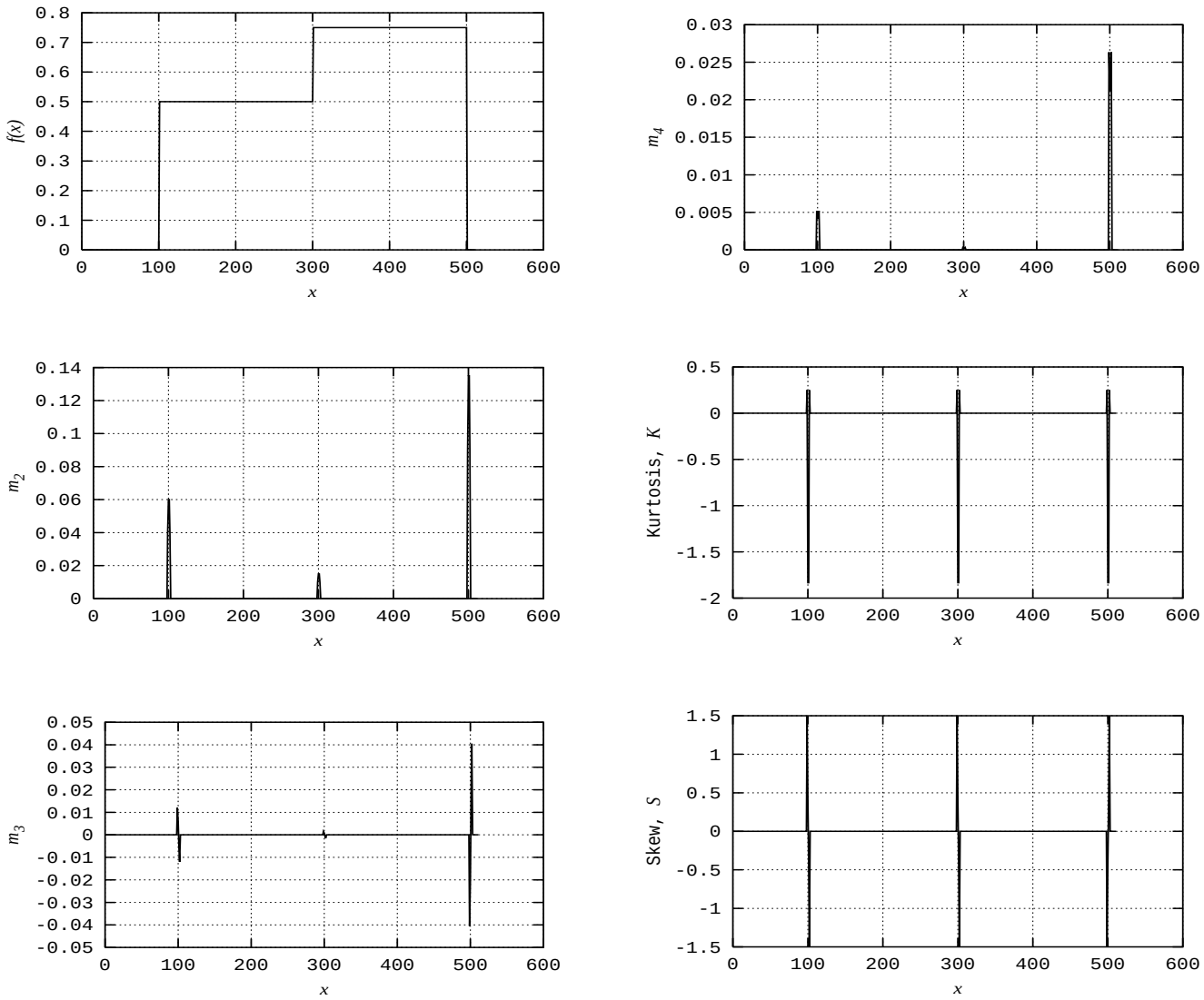


Figure 3: LHOMs plots for a step function (*one dimensional window of size 3 was used*).

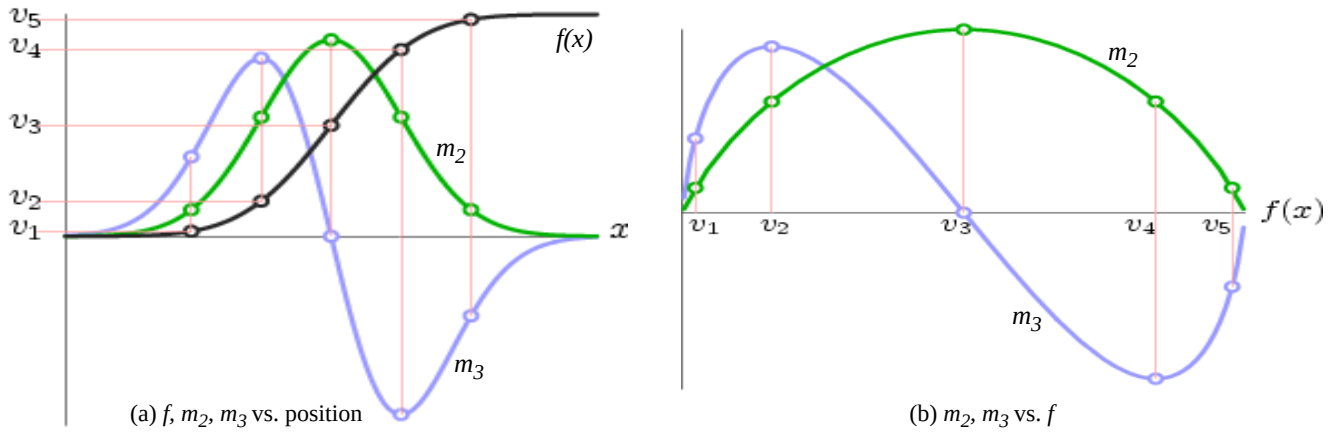
4. BOUNDARY DETECTION USING HOMS

From the above analysis and plots in Figure 2, we know that if a region for which we are calculating LHOMs is an *on-boundary region* then all of the following hold: second order LHOM is locally maximum, third order LHOM is locally zero, fourth-order LHOM is locally minimum, skew has a zero crossing, and *kurtosis* has a constant minima of *-2*. Using some or all of these criterions it is easy to detect all the *on-boundary* regions in a given dataset. Figure 3 shows the results of calculating LHOMs for a step function, from these plots it is clear that the *on-boundary regions* occur at the location of boundaries in the original signal. Since, we can detect *on-boundary* regions, we can indirectly obtain the spatial location of boundaries in a dataset. In the graph of fourth-order LHOM we see just a maxima, instead of a minima located between two maximas. This is likely due to the closeness of the maximas and the minima they appear merged when plotted.

For detection of salient iso-values in a dataset we need a method for obtaining boundaries in sample space rather than in spatial domain. From previous analysis we have observed the behavior of LHOMs in the spatial domain in the neighborhood of a boundary. Using the arguments offered by Kindlmann and Durkin in [7], if we plot the values of LHOMs *vs.* the sample values at their associated points, we expect to preserve the relation between *left*-and *right*-boundary and *on* the boundary itself. Figure 4 illustrates this point for LHOMs: m_2 and m_3.

5. RESULTS

Figure 5 shows the scatter plots for LHOMs *vs.* sample values for a CT Tooth dataset. The width *w* of each region *W* is five. Values corresponding to all the possible regions, i.e. regions associated with all the sample points in the dataset are considered. The shape of the scatter plots justifies the reasoning given in previous section for salient iso-surface detection using local central moments. We see from the plots that there are three distinct boundaries. These are in neighborhood of 200, 600, and 1100. The iso-surfaces corresponding to these boundaries are shown in Figure 6. Figure 7 shows *kurtosis* and *skew* scatter plots for the CT Head dataset, here again our methodology predicts three distinct boundaries in the neighborhood of 600, 1600, and 3500. The iso-surfaces corre-



(a) f, m_2, m_3 vs. position (b) m_2, m_3 vs. f

Figure 4: Relationship between f, m2, m3. (a) The plot on the left shows the dependence of the various LHOMs on the presence of an edge. (b) The plot shows the same relationship in an inverse manner.

sponding to these boundaries are shown in Figure 8. It is clear that the first iso-value represents the skin, the second skull, and the third teeth. Figure 9 shows the result of our analysis for computational dataset of flow over a delta wing. Here again we see that boundaries are in the range of 800, and 1000 (the original data was scaled and quantized). Iso-surface associated with an iso-value in this range is also shown in the same figure. The shape of the object, and the shape of the vortex and the shock in the flow are clearly visible.

6. Summary And Future work

We presented a method to explore volume data for salient iso-values. We employ statistical signatures based on localized central moments. We relate these localized moments to presence of boundaries. Later, we show the effectiveness of boundary and salient iso-value detection through examples. Future work includes MRI datasets which yield highly noisy LHOM plots. Denoising techniques are needed for further use. Also, we wish to exploit the relationship of moments to boundaries to create simple yet tangible interfaces for volume exploration.

7. Acknowledgments

The authors wish to thank Joe Marks, Hanspeter Pfister of Mitsubishi Electric Research Laboratories (MERL) for support. The work has been supported by grants from a NSF CAREER award, MERL and DoD. We would also like to thank Gordon Kindlmann for Figure 4.

8. References

[1] C. L. Bajaj, V. Pascucci, and D. R. Schikore, “The Contour Spectrum,” *Proceedings of Visualization '97*, pp. 167-175.

[2] Y.-K. Chang, A. P. Rockwood, and Q. He, “Direct rendering of freeform volumes,” *Computer-aided Design*, 27(7), 1995, pp. 553-558.

[3] R. Drebin, L. Carpenter, and P. Hanrahan, “Volume Rendering,” *Computer Graphics*, Vol. 22 No. 4, August 1988, pp. 65 - 74.

[4] S. Fang, T. Biddlecome, and M. Tuceryan, “Image-based transfer function design for data exploration in volume visualization,” *Proceedings of Visualization’98*, pp. 319-326.

[5] T. He, L. Hong, A. Kaufman, and H. Pfister, “Generation of transfer functions with stochastic search techniques,” *Proceedings of Visualization '96*, 1996, pp. 227-234.

[6] D. H. Laidlaw, K. W. Fleischer, and A. H. Barr, “Partial-volume Bayesian classification of material mixtures in MR volume data using voxel histograms,” *IEEE Transactions on Medical Imaging*, Vol. 17, No. 1, Feb. 1998, pp. 74-86.

[7] G. Kindlmann and J. W. Durkin, “Semi-automatic generation of transfer functions for direct volume rendering,” *Proceedings of the 1998 Symposium on Volume visualization*, 1998, pp. 79-86.

[8] J. Marks, B. Andalman, P.A. Beardsley, W. Freeman, S. Gibson, J. Hodgins, T. Kang, B. Mirtich, H. Pfister, W. Ruml, K. Ryall, J. Seims, and S. Shieber, “Design galleries: a general approach to setting parameters for computer graphics and animation,” *Proceedings of Siggraph’97*, pp. 389 - 400.

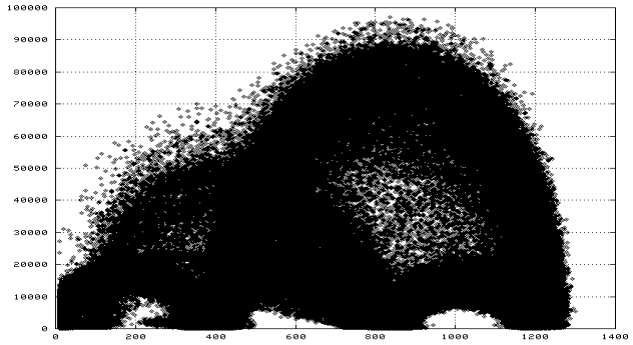
[9] C. L. Nikias, and A. P. Petropulu, *Higher-Order Spectra Analysis: A Nonlinear Signal Processing Framework*, Prentice Hall, New Jersey, 1993.

[10] H. Pfister, B. Lorensen, C. L. Bajaj, G. Kindlmann, W. Schroeder, L.Sobierajski-Avila, K. Martin, R. Machiraju, and J. Lee, “Visualization Viewpoints,” *IEEE Computer Graphics and Applications*, Vol. 21, No. 3, pp.16-22.

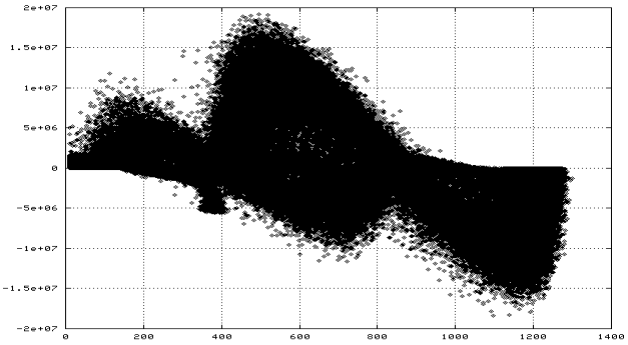
[11] Y. Sato, C.-F. Westin, A. Bhalerao, S. Nakajima, N. Shiraga, S. Tamura, and R. Kikinis, “Tissue Classification Based on 3D Local Intensity Structures for Volume Rendering,” *IEEE Transactions on Visualization and Computer Graphics*, Vol. 6, No. 2, April/June 2000, pp. 160-180.

[12] A. Stuart, J.K. Ord, and S. Arnold, *Kendall’s Advanced Theory of Statistics*, 5th ed. Griffin and Co., London., 1987.

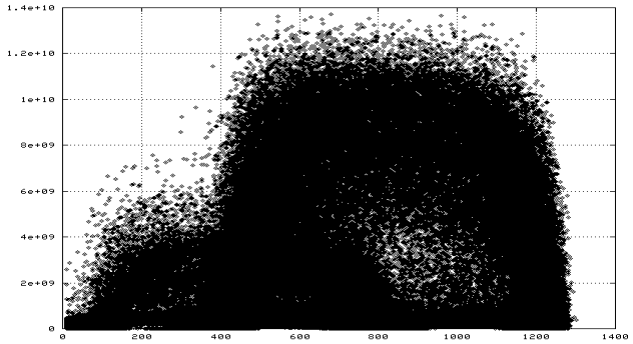
[13] T. S. Yoo, Image Geometry Through Multiscale Statistics. *Ph.D. Dissertation*, University of North Carolina at Chapel Hill, Department of Computer Science, 1996.



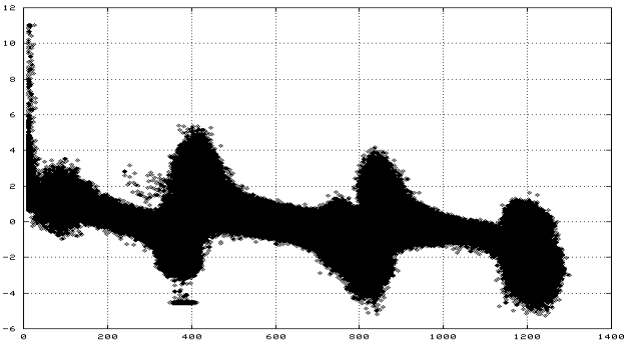
(a) m_2 vs. sample value



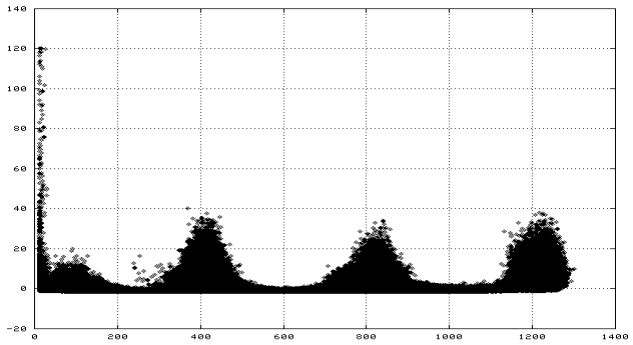
(b) m_3 vs. sample value



(c) m_4 vs. sample value



(d) *Skew* vs. sample value



(e) *Kurtosis* vs. sample value

Figure 5: LHOM scatter plots for the tooth dataset (*a window of dimensions* $5 \times 5 \times 5$ *was used).*

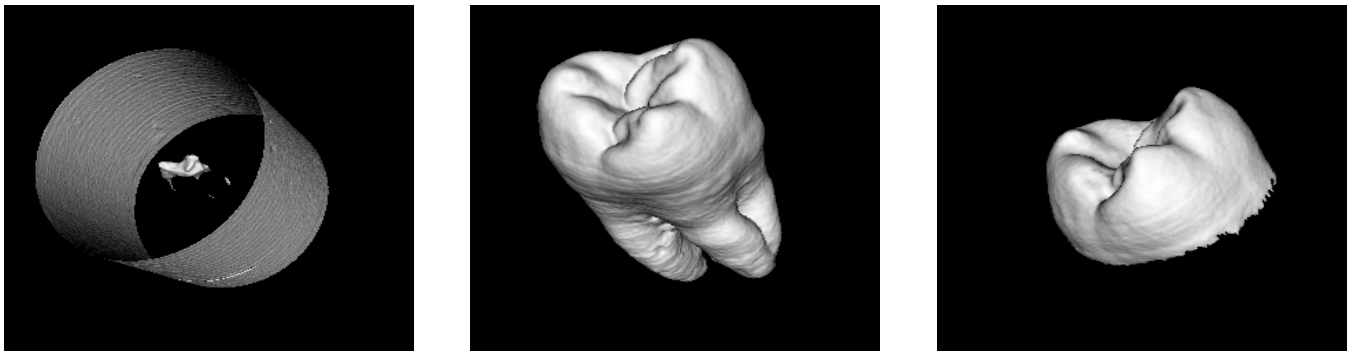
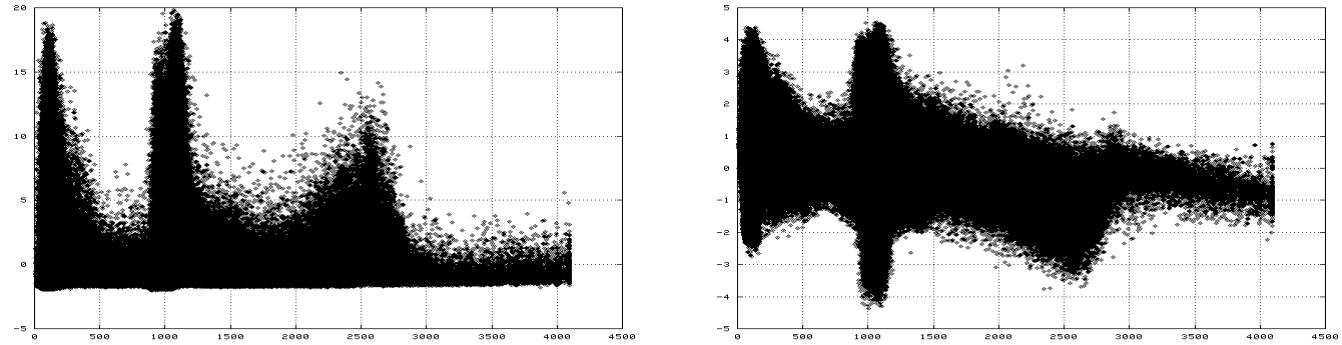


Figure 6: Iso-surfaces associated with the three salient iso-values predicted by the scatter plots in Figure 5: *(left)* 200, *(center)* 600, and *(right)* 1100.



(a) *Kurtosis* vs. sample value

(b) *Skew* vs. sample value

Figure 7: LHOMs scatter plots for the CT head dataset (*a window of dimensions* $3 \times 3 \times 3$ *was used).*

Figure 8: Iso-surfaces associated with the three salient iso-values predicted by scatter plots in Figure 7: *(left)* 600, *(center)* 1600, and *(right)* 3500.

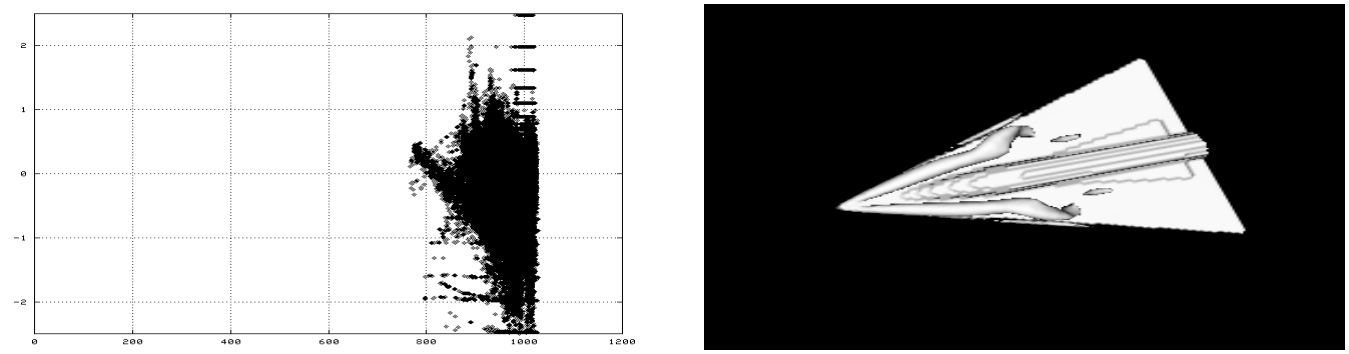


Figure 9: *(left)* Scatter plot of *Skew vs.* sample value for the delta wing dataset, and *(right)* iso-surface for iso-value of 900 (*a window of dimensions* $3 \times 3 \times 3$ *was used).*