STATISTICAL COMPUTATION OF SALIENT ISO-VALUES

Shivaraj Tenginakai, Raghu Machiraju
The Ohio State University

Abstract

Detection of the salient iso-values in a volume dataset is often the first step towards its exploration. An error-and-trail approach is often used; new semi-automatic techniques either make assumptions about their data [4] or present multiple criteria for analysis. Determining if a dataset satisfies an algorithm's assumptions, or the criteria to be used in an analysis are both non-trivial tasks. The use of a dataset's statistical signatures, local higher order moments (LHOMs), to characterize its salient iso-values was presented in [10]. In this paper we propose a computational algorithm that uses LHOMs for expedient estimation of salient iso-values. As LHOMs are model independent statistical signatures our algorithm does not impose any assumptions on the data. Further, the algorithm has a single criterion for characterization of the salient iso-values, and the search for this criterion is easily automated. Examples from medical and computational domains are used to demonstrate the effectiveness of the proposed algorithm.

Categories and Subject Descriptors (according to ACM CCS): I.4.6 [Segmentation]: Edge and Feature Detection.

1. Introduction

Pat Hanrahan called data exploration (including transfer function design) one of the top ten problems in volume visualization in his 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][4][6][8][10]. In the panel session on transfer function design at the Visualization'00 conference [9] three classes of techniques were identified. 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 [4] or that computable signature functions be obtained [1]. Image-centric methods [6] 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. This year at the *Visualization'01* conference there were three papers on the topic of transfer function design [5] and salient iso-value determination [8][10]. Salient iso-values have to be determined first in any data-exploration activity. Suitable user interfaces and boundary emphasis functions can be placed over a salient iso-value to emphasize the surfaces in a semi-transparent way.

In an earlier paper [10], we proposed to employ data-signatures to explore data. These signatures are obtained from localized *k*-order central moments and cummulants including *kurtosis* and *skew*. There exist strong relationships between the various central moments that can be used to locate salient iso-values. A localized two-material mixture model was assumed and it was shown that the maximas and minimas of the moments and cummulants occur at spatial locations populated by material interfaces or boundaries. A scatter-plot was used to estimate the maximas and minimas in individual scatter - plots. Also, no attempt was made to correlate the information across the scatter-plots of various quantities. The scatter-plots, however, did yield useful characteristics for CT and fluid-dynamics data which allowed for the selection of the extremas and zero-crossings. Such plots are shown in Figure 2. In this paper, we present algorithms to estimate significant salient iso-values instead of requiring the user to guess the location of the value. We employ a 3D space of function value, and its *skew* and *kurtosis* values. However, unlike Kindlmann and Durkin [4] a pre-computed (histogram) volume is not employed. The resulting search allows in easier location of salient iso-values. In essence, this paper reports on computational methods in a statistical context to determine salient iso-values. Majority voting techniques are exploited to locate salient iso-values in our proposed algorithm.

Section 2 presents some of the existing techniques for salient iso-values detection. A brief introduction to LHOMs (Local Higher Order Moments) and their use in salient iso-values detection is presented in Section 3. The algorithm proposed in this paper is described in Section 4, and Section 5 presents its implementation. The results that validate the algorithm are presented in Section 6, and finally Section 7 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. This approach was taken by Kindlmann and Durkin [4]. 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. Based on analysis of this histogram, a distance function is created: a signed distance to the middle of the nearest boundary. The opacity is now defined as a function of position within a boundary. In [5] Kniss and Kindlmann devised multi-dimensional techniques based on the techniques proposed in [4].

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. 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. In [8] the contour spectrum computation was expedited through a Laplacian-weighted gray value histograms. Statistical techniques are gaining popularity in data-mining applications. Only a few reported statistics-based methods exist for volume data-analysis and exploration. In [2][3] statistical inference methods (e.g. Bayesian) are employed to determine the material density of each voxel. In [10] we explore the use of statistical signatures and show how salient iso-values can be determined. It turns out that [8] also exploits the histogram (the *first*-moment) for analysis.

Instead of tweaking input parameters another approach such as inverse design can be used for data-analysis [6]. Embodying this approach is the Design Galleries approach that was used to explore parameter spaces for a host of visualization, graphics and animation applications [6].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. We now define some our boundary model.

3. Local Higher Order Moments (LHOMS)

Higher Order Moments (HOMs), m_k, are statistical quantities that measure the tendency of a distribution S_N to cluster around a certain value [7], usually the distribution's mean, *M, and* are computed as*:*

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

where, *k* is the *order* of a *HOM*. *Local Higher Order Moments (LHOMs),* are *HOM*s that are calculated local to a sample point in a distribution. Consider a sample point x_i, such that $x_i \in S_N$, and let *W* be a neighborhood, of size *w*, around x_i. If M_i is the mean of the sample points in *W*, then *LHOM* at x_i is given by:

$$m_{k,i} = \frac{1}{w}\sum_{j=i-\left\lfloor \frac{w}{2} \right\rfloor}^{i+\left\lfloor \frac{w}{2} \right\rfloor}(x_j - M_i)^k \qquad (2)$$

where, *k* is the *order* of the LHOM, and $x_{i-\frac{w}{\lfloor 2 \rfloor}}, x_{i-\frac{w}{\lfloor 2 \rfloor}+1}, \ldots, x_i, \ldots, x_{i+\frac{w}{\lfloor 2 \rfloor}} \in W$.

Consider a neighborhood *W* of size $w \times w$ centered at a point *P,* in a two dimensional dataset, see Figure 1. Clearly, there are w^2 particles or sample points in *W.* (Since we are dealing with *number of points*, our analysis is not dependent on the dimensionality of the dataset; a two dimensional dataset is used only as an example.).

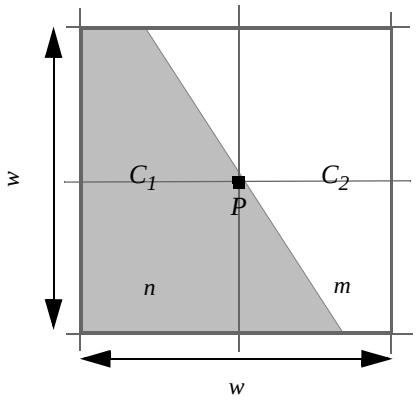
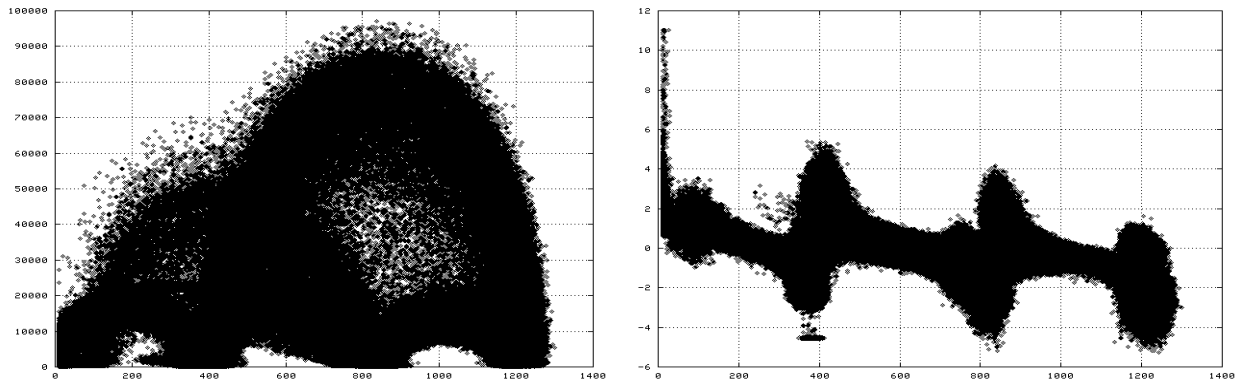


Figure 1: The boundary model (*w = 3)* for a small *2D* neighborhood. C_1 and C_2 are the two materials separated by a material interface.

If *W* is small enough, such that it contains at most two distinct materials at a single boundary, it can be shown that the calculation of *LHOM* at *P,* using Equation 2, simplifies to, see [10]:

$$m_{k,p} = \frac{\Delta^k}{\left(w^2\right)^{k+1}}\left(mn^k + (-1)^k nm^k\right) \qquad (3)$$



(a) m_2 vs. sample value (b) *Skew, S,* vs. sample value

Figure 2: Large dynamic range, *0 - 100,000,* of the Second Order LHOM (*left),* and Small dynamic range, *-6 - 12,* of *Skew* (*right)*. The graphs were taken for a CT tooth dataset.

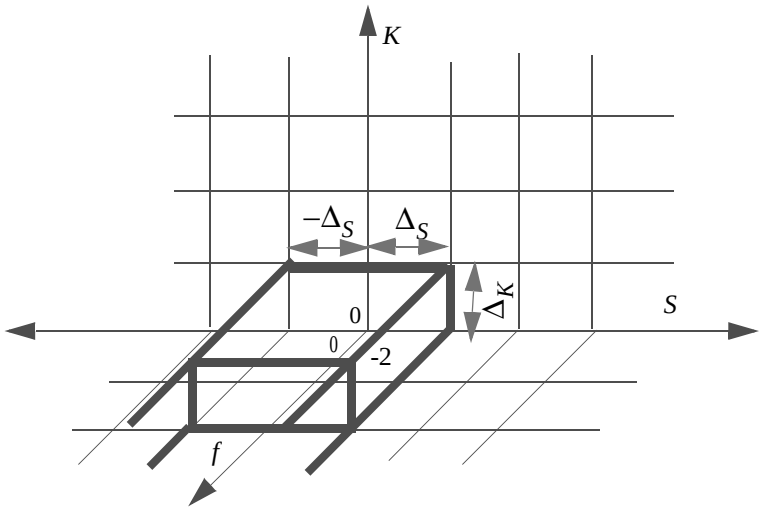


Figure 3: A three-dimensional histogram of *Skew (S), Kurtosis (K),* and sample value, *f.* A bin contains the number of points whose sample value, *Kurtosis*, and *Skew* are in the range represented by it. The region bounded by thick lines represents the region containing the *salient iso-value* bins.

where, Δ is the difference in the sample values of the two materials present in neighborhood *W,* and *m* is the number of sample points of the material C_1 and *n* is the number of sample points of material C_2. Further, if the sample value at *P* represents a salient iso-value, then it can be shown that at *P*:

- Second Order *LHOM*, $m_{2,p}$, has a maxima.
- Third Order *LHOM*, $m_{3,p}$, has a zero-crossing
- Fourth Order *LHOM*, $m_{4,p}$, has a local minima.
- *Skew, S,* a non-dimensional ratio of third and second order *LHOM*s has a zero-crossing.
- *Kurtosis, K,* a non-dimensional ratio of fourth and second order *LHOM*s has a minima of *-2*.

A method for manual determination of the salient iso-values through scatter plots, using above properties, was sketched in [10]. In the following section we describe an algorithm that uses the above characteristics to automate the task.

4. Determination of Salient Iso-Values

From the above description it is clear that we can use either of the five mentioned criteria for detecting the salient iso-values. However, the first three, $m_{2,p}$, $m_{3,p}$ and $m_{4,p}$, are direct powers, and hence have a very large dynamic range (Figure 2 (a)). Computation with large numbers is both slow, introduces errors such as overflow and expends much memory. *Skew* and *Kurtosis,* being ratios, have a relatively smaller dynamic range, (Figure 2(b)) and hence, are ideal candidates for fast salient iso-values detection. Though either *Skew* or *Kurtosis* can be used for detecting salient iso-values, a better approach involves using both these quantities. This approach reduces the possibility of detecting spurious salient iso-values. Kindlmann and Durkin in [4] demonstrated the use of a three-dimensional histogram in combining the information from two quantities, the *first order* gradient and the *second order* gradient.

If we construct a three-dimensional histogram of *Skew, Kurtosis,* and sample values, see Figure 3, what do we expect to see? If a point's sample value is a salient iso-value, then at that point *Kurtosis* will be *-2*, and *Skew* will be *0*, we classify all bins in the histogram that correspond to these values as *salient iso-value* bins. We expect that the sample values that are salient iso-values would not only have a *majority* of their points in the *salient iso-value* bins, but also the number of their points in the non *salient iso-value* bins would be *minimum (*Figure 4). The cummulants are computed through out the volume. Hence, there will be several regions populated by voxels with same measured value *f* but with different values of skew and kurtosis. Hence, for detecting salient iso-values a robust algorithm is required to search for all those sample points that satisfy this *majority criteria.* Section 5 presents an efficient implementation of this algorithm.

5. Algorithm for Salient Iso-Value Detection

From Section 4 it is clear that we are searching for only those sample values that satisfy the following two criteria:

1. They have a majority of their points in the *salient iso-value* bins.
2. The number of their points in the non *salient iso-value* bins is minimum or relatively smaller.

Given the size of a bin, $(1, \Delta_K, \Delta_S)$ then for all sample points the size of the *salient iso-value* region is given by:

$$\begin{aligned} -\Delta_S \le S < \Delta_S \\ -2 \le K < \Delta_K \end{aligned} \qquad (4)$$

(Note: We take size of each bin along *f* to be 1, so that each bin represents exactly one sample value.). To evaluate the first criteria, we determine sample values for which the number of points that satisfy Equation 4 is greater than the number of points that don't. Further, to evaluate the second criteria, we consider all sample values that satisfy the first criteria, and compute the difference between the number of points that satisfy Equation 4 and the number of points that don't; if a sample value represents a salient iso-value we expect this difference, Δ, to be *maximum.* Our algorithm is described in Figure 5. Equation 4 essentially provides a majority voting

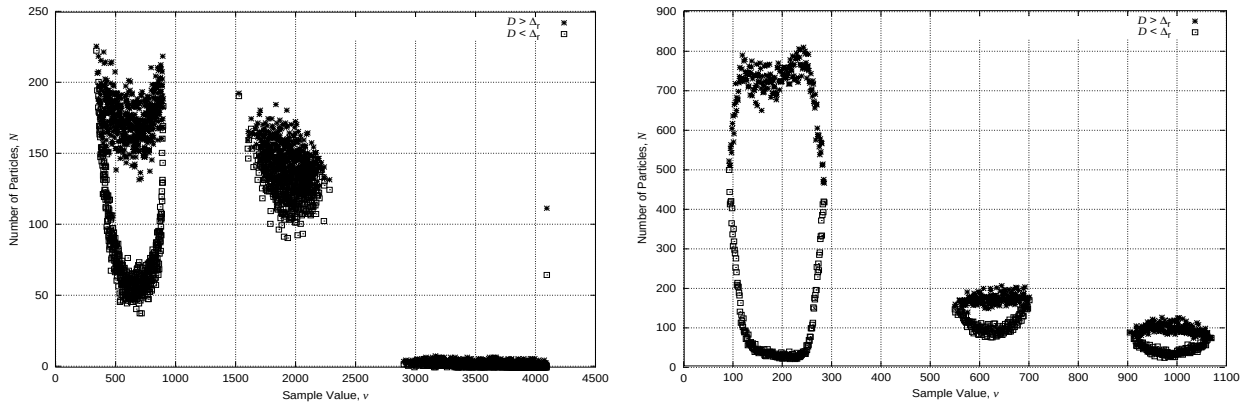


Figure 4: Shows the plots for the number of points in the salient iso-value bins ($D < \Delta_r$) and the number of points in the non salient iso-value bins $D > \Delta_r$ for those sample values that satisfy the first criteria. *(left)* CT Head dataset, and *(right)* CT Tooth dataset. (Note: Three salient iso-values are predicted for each of these datasets. A window of size *5* was used in calculation of LHOMS.)

criteria for sample values. The cummulants, in essence, vote that the sample values be considered as salient. The *maximas* are determined by examining the output of the algorithm.

6. Results

From Section 5 we conclude that if we plot the output of the algorithm (Figure 5), we should expect maximas at the location of salient iso-values. The output of our algorithm include the counts of conforming and non-confirming points as dictated by our voting criteria. Figure 6 shows the graph obtained by plotting the output of the algorithm, and the corresponding salient iso-values for the CT Tooth dataset. Figure 7 shows the graph and salient iso-values for the CT Head dataset, while Figure 8 shows the results for a CFD shock wave dataset. It is clear that the salient iso-values predicted by our algorithm are indeed the salient iso-values in the input dataset.

7. Conclusion and Future Work

In this paper we presented an algorithm that uses statistical signatures of a dataset to accurately predict the salient iso-values in the dataset. Future work includes adopting the given algorithm for MRI datasets, and building an intuitive interface for transfer function design using the techniques presented in this paper.

8. Acknowledgments

The authors would to thank Joe Marks, Hanspeter Pisfter of Mitsubishi Electric Research Laboratories (MERL) for support. The work has been supported by grants from a NSF CAREER award, MERL, and DOD.

9. References

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

[2] R. Drebin, L. Carpenter, andP. Hanrahan, "Volume Rendering," *Computer Graphics*, Vol. 22 No. 4, August 1988, pp. 65 - 74.

[3] 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.

[4] 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.

[5] J. Kniss, G. Kindlmann, and C. Hansen, "Interactive Volume Rendering Using Mult-Dimensional Transfer Functions and Direct Manipulation Widgets," *Proceedings of IEEE Conference on Visualization '01*, pp. 255- 262.

[6] 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.

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

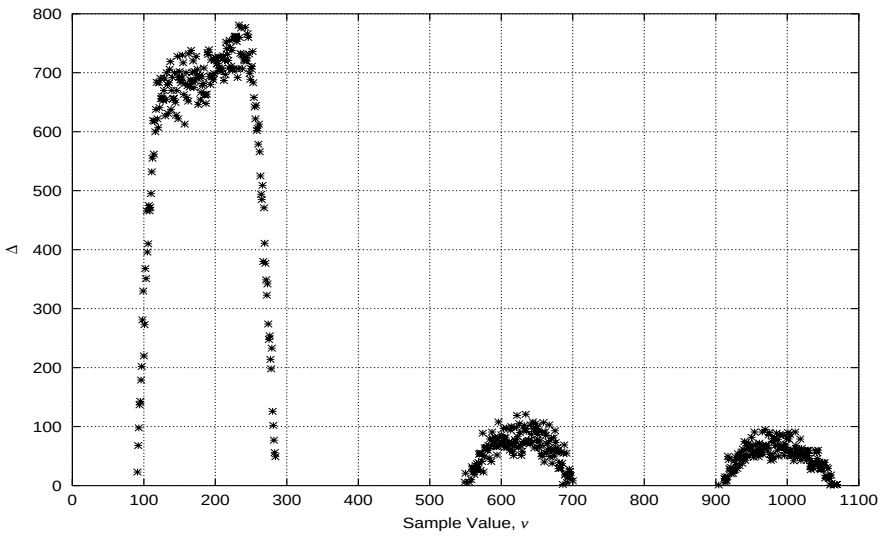
[8] V. Pekar, R. Wiemker, and D. Hempel, "Fast Detection of Meaningful Isosurfaces for Volume Data Visualization," *Proceedings of IEEE Conference on Visualization '01*, pp. 223- \230.

[9] 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.

[10] S. Tenginakai, J. Lee, and R. Machiraju, "Salient Iso-Surface Detection with Model-independent Statistical Signatures," *Proceedings of IEEE Conference on Visualization '01*, pp. 231-238.

```
// max_sample_value        the maximum sample value
// num_sample_points       the total number of sample ponits in a dataset
// deltas                  the size of a bin for skew
// deltak                  the size of a bin for kurosis

/* store the number of points in the salient iso-value bins for each sample value */
integer N_Salinet [0..max_sample_value]

/* store the number of points in the non salient iso-value bins for each sample value*/
integer N_Non_Salinet [0..max_sample_value]

/* Criteria 1: calculate the number of points in salient and non salient iso-value bins for each sample
   value. */
for(integer n = 0; n < num_sample_points; n++) {
  sample_value      = get_sample_value(n);    // sample value at sample point n
  skew              = get_skew(n);            // skew value at sample point n
  kurtosis          = get_kurtosis(n);        // kurtosis value at sample point n

  if((skew >= -deltas and skew < deltas) and  (kurtosis >=-2 and kurtosis < deltak))
    N_Salinet [sample_value]++;
  else
    N_Non_Salinet [sample_value]++;
}

/* Criteria 2: find the difference between the number of points that are in salient iso-value bins and the
   number of points that are not in salient iso-value bins for those sample values that satisfy Criteria 1.*/
for(integer s = 0; s < max_sample_value; s++) {
 if(N_Salinet [s] > N_Non_Salinet [s])
   delta = N_Salinet[s] - N_Non_Salient[s];
 else delta = 0;
 output delta;
}
```

Figure 5: The algorithm for salient iso-value detection. (Note: We can pre-compute the values for *Skew* and *Kurtosis*; in this case the complexity of the algorithm is linear with the number of sample points.)



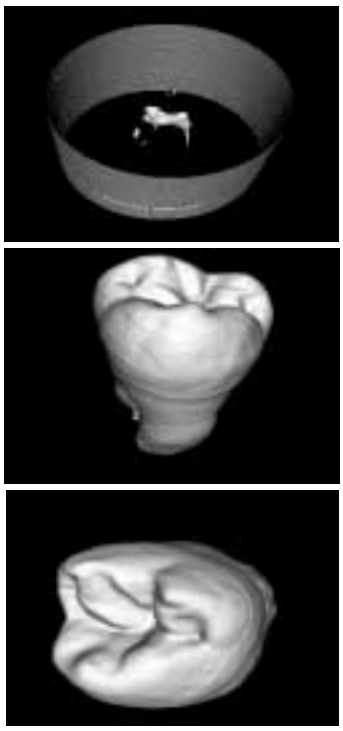


Figure 6: *(left)* Graph of the output from the algorithm for the CT Tooth dataset. Note there are three local maximas: *252, 622,* and *985;* hence, there are three salient iso-values in the dataset. The iso-surfaces corresponding to these salient iso-values are shown left. *(left-top)* iso-surface for *252*, *(left-center)* iso-surface for *622*, and *(left-bottom)* iso-surface for *985*.

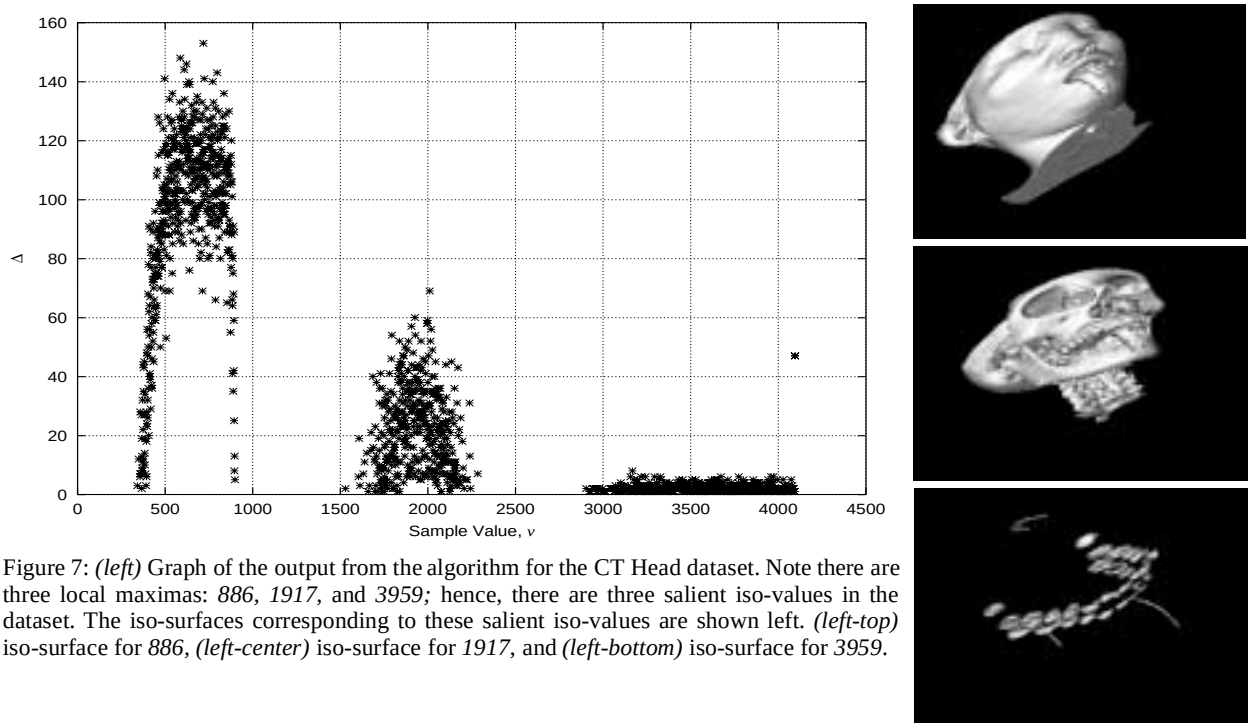


Figure 7: *(left)* Graph of the output from the algorithm for the CT Head dataset. Note there are three local maximas: *886, 1917,* and *3959;* hence, there are three salient iso-values in the dataset. The iso-surfaces corresponding to these salient iso-values are shown left. *(left-top)* iso-surface for *886*, *(left-center)* iso-surface for *1917*, and *(left-bottom)* iso-surface for *3959*.

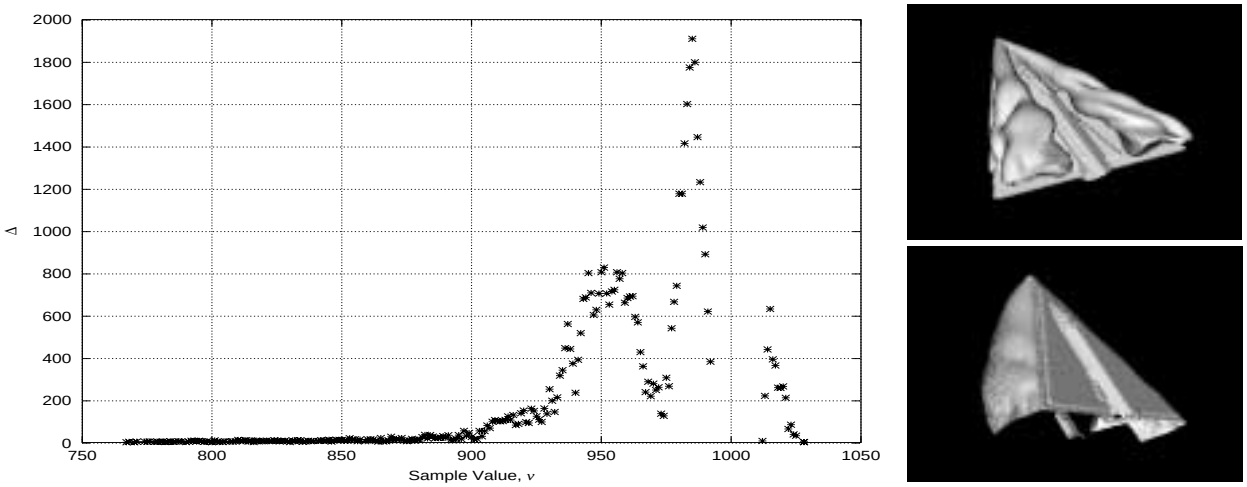


Figure 8: *(left)* Graph of the output from the algorithm for the a CFD dataset. Note there are two local maximas: *956*, and *985;* hence, there are two salient iso-values in the dataset. The iso-surfaces corresponding to these salient iso-values are shown left. *(left-top)* iso-surface for *956*, and *(left-bottom)* iso-surface for *985*.