Significant Iso-values extraction for 3D Scalar Field Data Analysis

Neeraj Badal, Sai Sravan Reddy

December 7, 2019

Table of Contents

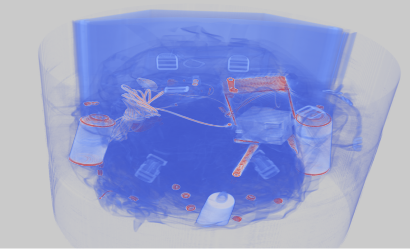
Problem Specification

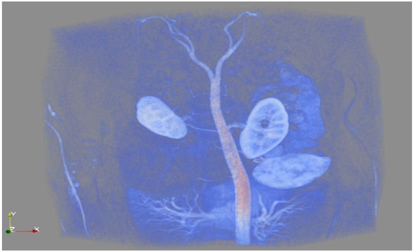
- Problem Statement : Identify a set of potentially **significant isovalue ranges** by analyzing measures computed on the scalar field data.
- Significant isovalues belonging to different range represents a different material surface / significant isosurface.
- Project Outcomes:
  - Exploration and Implementation of the measures contributing in making independent observation for the scalar field data.
  - Analyse the effectiveness of the techniques over the data sets related to an application domain.
- References :

1. *Wenger R.(2013), ISOSURFACES Geometry, Topology and Algorithms. CRC Press.*
2. *Carr, H., Duffy, B., and Barry, D. (2006a). On histograms and isosurface statistics. IEEE Transactions on Visualization and Computer Graphics, 12(5):1259-1266.*
3. *Scheidegger, C. E., Schreiner, J. M., Duffy, B., Carr, H., and Silva, C. T. (2008). Revisiting histograms and isosurface statistics. IEEE Transactions on Visualization and Computer Graphics, 14(6):1659-1666.*

Motivation

- Identifying significant iso-value corresponds to identifying a significant material/surface in the underlying scalar data.
    - Could be very important for Security Personnels for luggage scanning.
    - Identification of Tumors from the scalar data from the tissue being probed.





Project Break-Up

- Three major tasks identified.
  - Exploration and Implementation of Statistics Measures. (Phase 1)
  - Using the mentioned measure to perform the analysis on a subset of data related to medical domain. (Phase 1 & 2)
  - Optimization in terms of execution time. (Phase 2)
- The measures identified are divided into 3 categories.
- Surface Area
  - Edge Frequency (Phase 1)
  - Cube Frequency (Phase 1)
  - Triangle Count (Phase 1)
  - Iso-Surface Area (Phase 1)
- Gradient
  - Total Gradient (Phase 1)
  - Mean Gradient (Phase 1)
- Interval Volume
  - Scalar Frequency (Phase 1)
  - Edge Based (Phase 2)
  - Cube Based (Phase 2)

Data Sets Covered Phase 1 & 2 Inclusive

Data Set Name	Dimensions	Data Type	Category	Status
Tooth	103 x 94 x 61	uint8	Medical	
Foot	256 x 256 x 256	uint8	Medical	
Skull	256 x 256 x 256	uint8	Medical	
Kidney	384 x 384 x 96	uint8	Medical	
Knee	512 x 512 x 87	uint16	Medical	
Head Visible Male	128 x 256 x256	uint8	Medical	
Tumor DCIS	448 x 448 x 160	uint8	Medical	
Tumor Breast	448 x 448 x208	uint8	Medical	
Retrograde	320 x 320 x 72	uint16	Medical	
MRI Head	256 x 256 x 256	uint8	Medical	
DTI Scan	128 x 128 x 58	uint8	Medical	
CT-Knee Tibial	379 x 229 x 305	uint8	Medical	
Angio	384 x 512 x 80	uint8	Medical	
BackPack	512 x 512 x 373	uint8	Structural	
Engine	256 x 256 x 128	uint8	Structural	
Bonsai	256 x 256 x 256	uint8	Structural	
Tomato	256 x 256 x 64	uint8	Material	
Orange	256 x 256 x 64	uint8	Material	

- Medical Datasets : 11
- Structural : 3
- Material : 2

Surface Area Based Measures - Edge Frequency Cube Frequency

- For scalar value σ, the grid edges or grid cubes are included in it's frequency i.e. $\phi(\sigma)$ if the σ is included in their span.
  - Grid cube span is the interval from the minimum scalar value to the maximum scalar value of any cube vertices belonging to that cube.
  - Grid edge span is the interval from the minimum to the maximum scalar value of the two edge vertices.
- let $F_e(\sigma)$ and $F_c(\sigma)$,denote the edge frequency and cube frequency whose span contains the scalar value σ.
- Similar to scalar frequency, normalized edge and cube frequency can be defined as,$F_e^* = \dfrac{F_e(\sigma_0, \sigma_1)}{N_\tau^e * \alpha}$ and $F_c^* = \dfrac{F_c(\sigma_0, \sigma_1)}{N_\tau^c * \alpha}$, where α is the bin-width and $N_\tau^e = \#$ of grid edges, $N_\tau^c = \#$ of grid cubes.

Surface Area Based Measures - Triangle Counts and Iso-Surface Area

- The triangle count at a particular iso-value refers to the number of triangles that form the iso-surface and similarly the iso-surface area is obtained by summing the areas of the individual triangles. Since, the triangles generated are fairly uniform in size, so the area computed can also be approximated by counting the number of triangles in the iso-surface. Hence both the plots offer similar information.
- At each iso-value, marching cubes with trilinear interpolation is applied to get the iso-surface and triangle counts. Since marching cubes is computationally heavier than simple edge and vertex counting, so the plots are slightly computationally intensive.

Gradient Based Interest Iso-value Indicator

- For a scalar field $\phi : R^3 \rightarrow R$, the gradient $\nabla\phi = (\frac{\partial\phi}{\partial x}, \frac{\partial\phi}{\partial y}, \frac{\partial\phi}{\partial z})$, and its magnitude $|\phi| = \sqrt{\frac{\partial\phi}{\partial x}^2 + \frac{\partial\phi}{\partial y}^2 + \frac{\partial\phi}{\partial z}^2}$.
- High gradient magnitudes indicate transition in materials.
- Two ways to approximate measures of gradients over a level set -
  - Total Gradient of ϕ over a level set σ defined as, $TotalGrad_e(\sigma) = \sum_{(p,q)\in E(\sigma)} \frac{|s_q - s_p|}{3}$, where, $E(\sigma)$: set of grid edges (p,q) that intersects an isosurface.
  - Mean Gradient, defined as, $MeanGrad_e(\sigma) = \frac{TotalGrad_e(\sigma)}{F_e(\sigma)}$.
- The significant isosurfaces lie between the peaks. However some of the peaks in the total gradient plot correspond to significant isosurfaces as for engine data set, but others correspond to isosurfaces with large amounts of noise as in tooth data set.

Interval Volume Based Measures - Scalar Frequency

- Scalar Frequency Histogram, $F_v(\sigma)\ \forall\sigma \in \phi(x)$, $x \in R^3$.
- In case of $\phi(x) \in R$ frequency of scalar values are given in terms of ranges $[\sigma_0, \sigma_1)$ represented as $F_v(\sigma_0, \sigma_1)$ by binning σ in $\{\sigma : \sigma_0 \leq \sigma < \sigma_1\}$.
- Binning is also helpful if $F_v(\sigma)$ gives an noisy distribution over $\sigma \in Z$ for useful representation of data.
- Interest Portions:
  - High-Frequency scalar values other than zero represent transitions between two different material surfaces.
  - Better choice of iso-values are the scalar values in the regions between the high frequency peaks. Such regions represent the materials themselves.

Interval Volume Based Measures - Interval Volumes Edge & Cube Based

- Approximates the interval volume between two iso-surfaces corresponding to scalar values σ_0 & σ_1.
- It provides the same measure as that of scalar frequency but there are less spikes in the distribution obtained from it giving a good indication of long run of an iso-surface.
- $Vol_e(\sigma_0, \sigma_1) =$

$$\sum_{(p,q)\in E(\sigma_0,\sigma_1)|s_p - s_q| < |\sigma_0 - \sigma_1|} \frac{1}{3} + \sum_{(p,q)\in E(\sigma_0,\sigma_1)|s_p - s_q| > |\sigma_0 - \sigma_1|} \frac{|\sigma_0 - \sigma_1|}{3|s_q - s_p|}$$

Results

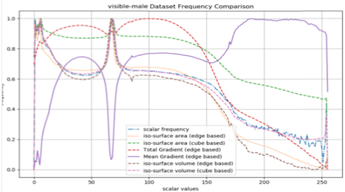
- The analysis using the help of measures were performed mostly on medical domain datasets.
- Apart from it, we have also included data from natural material i.e. orange fruit and data set containing metallic surfaces (Engine and Backpack objects).
- Based on the below observations a small recipe is followed during analysis through the different measures.
  - The Mean Gradient plot mostly stands-out in giving distinguishable features. The Total gradient for majority of cases gives false positives i.e. not showing significant surfaces around the valleys which is it's usual behaviour.
  - All the other measures, show significant surfaces around the flat portion of the curves. They signify transition between the materials around peaks observed which is opposite for the Mean Gradient measure.
  - The chances of missing very minutely observable features are very high in iso-surface area based, interval volume based measures.

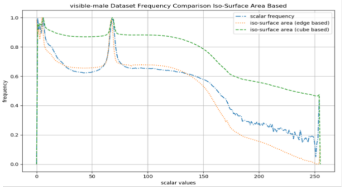
Analysis Recipe

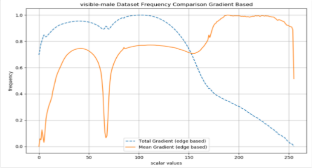
Algorithm 1: Scalar Data Analysis through Measures

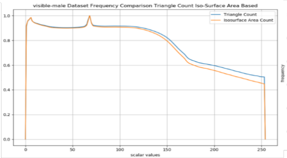
1. Perform Volume Rendering on the input data set;
2. Identify σ ranges belonging to stand-out peaks in Mean Gradient & troughs in total gradient and iso-surface area based approximations;
3. For the step 2. σ ranges, extract iso-surface area using Marching Cubes. These are the prominent occuring material surface in the scalar data;
4. Lookout for small half peaks in mean gradient for which the total gradient shows a half trough. These are the regions with sgnificant surface but with covering small portion in the scalar data;
5. Ignore the peaks in mean gradient, where all other measure shows region of near zero iso-value frequencies. These regions belongs to transition or noise in the data (False Positives);
6. Ignore the σ ranges where interval volume and surface area based measures show flat region but Mean gradient showing transition or valley.

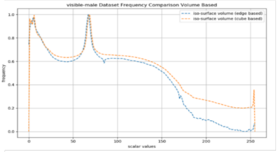
Head Visible Male





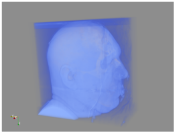




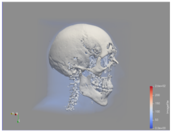


- Mean gradient plot shows existence of significance surface at iso-value ranges equal to 40, 110 and 200.

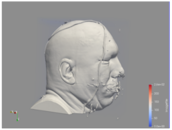
MRI Visible Male Contd.



Iso-Value 200



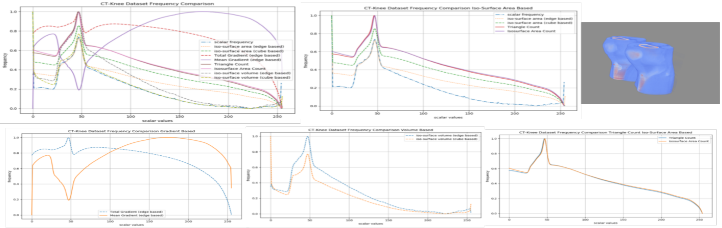
Iso-Value 110



Iso-Value 40

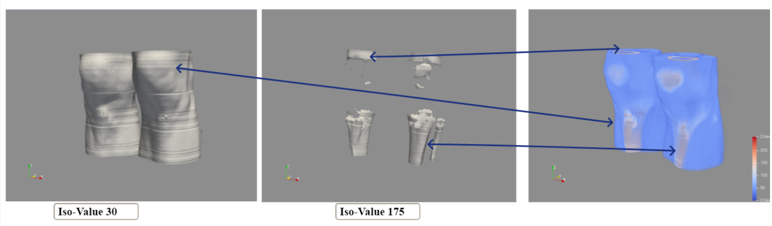
- Total Gradient plot has shown reverse, i.e. instead of valleys it shows existence of distinguishable iso-surfaces at 40 and 110 as peaks, but fails to do it for teeth surface.

CT Knee



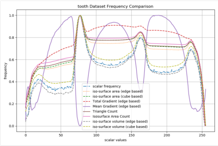
- The Total Gradient plot shows the existence of two peaks at iso-values 25 and 175.
- This flat peaks are an indication of a consistent iso-surface in the scalar data.
- iso-value 25 corresponds to flesh part, whereas, iso-value 175 corresponds bones inside the flesh.

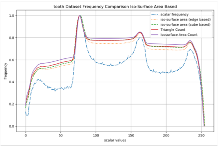
CT-Knee Contd.

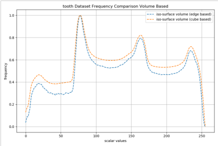


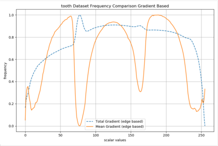
- The other remaining plots gives an indication that iso-surface will be continously in transition state. This is not true always as can be seen in the Fig. for iso-value range we actually have existence of bone structures which is not identifiable from the other measures.
- However, other measures are correct at identifying noisy region around the iso-value 50.

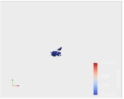
Tooth : Case were most Measures perform

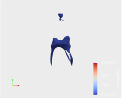


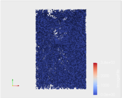


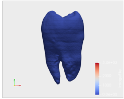


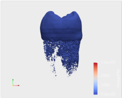


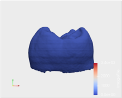

Isovalue 10


Isovalue 65


Isovalue 75

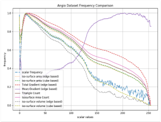

Isovalue 125

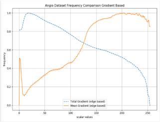

Isovalue 165

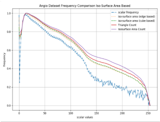

Isovalue 190

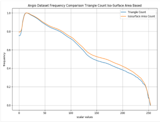
- Peaks of interval volume plots(75,165,235) corresponding to high noise regions and significant isosurfaces lie between the peaks (65,125,190).
- Mean gradient plots correctly identify the significant iso-surfaces as peaks (65,125,190) and transition regions as troughs (75,165,235) whereas the total gradient plots peak instead.

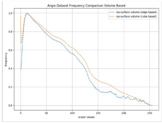
Angio

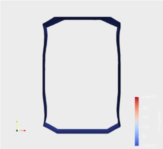




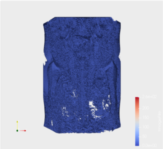




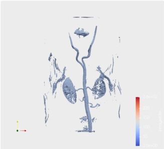




Isovalue 1



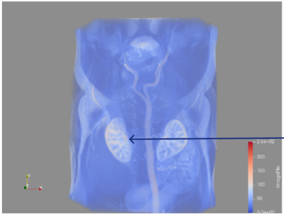
Isovalue 15

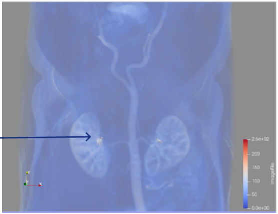


Isovalue 90

- Zero gradient points of mean gradient plot at 60-100 range indicate significant isosurfaces correctly and they correspond to the blood vessels.

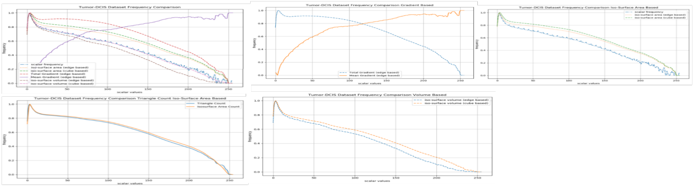
Angio Contd.

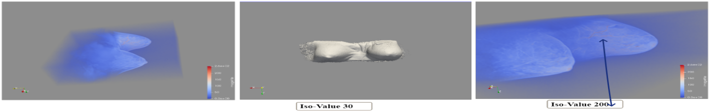




- Peak of mean gradient plot at 0-5 and troughs of all other plots indicate a significant iso-surface but is actually the material used during the scanning.
- Trough of mean gradient and peaks of other plots at 10-20 indicate transition region.
- Peak of the mean gradient plot occurs at 200 also, but since all other plots are tending to zero, this range indicates noise in the data.
- Mean Gradient, though showing some peak where there is minor trough in total gradient plot at iso-value 160, belongs to tumor like substance.

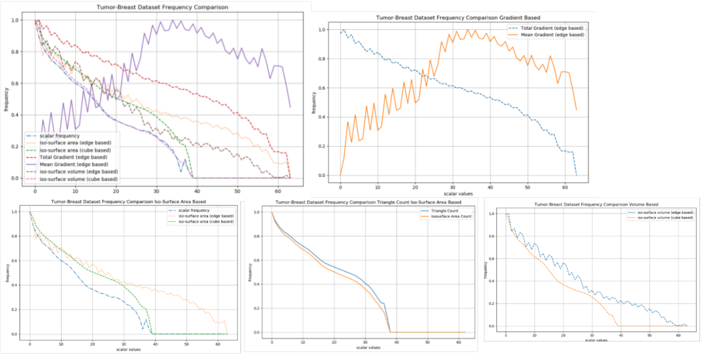
Breast Tumor





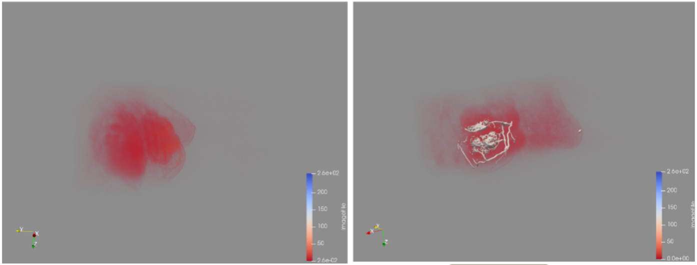
- Total gradient and other plots shows existence of significance surface at iso-value ranges equal to 20-40 (skin tissues).
- Mean Gradient plot has shown a minute observation by giving a peak at iso-value range 200-220 which belongs to veins inside the breast tissue showing some abnormal color than their counterparts in the volume surface.

Failure Case : Breast Tumor



- All the plots failed to give any clue of significant tissue in the breast data.
- The flat line belonging iso-values range from 40-60 are actually region having no iso-surfaces. So we see a flat line in majority of the plot.

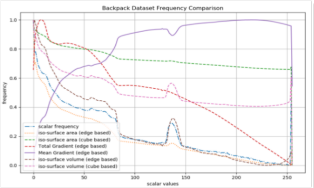
Failure Case : Breast Tumor Contd.

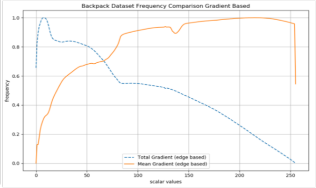


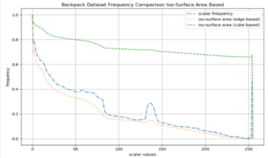
Iso-Value 13

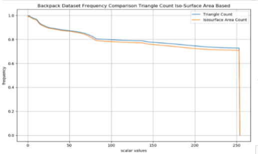
- The presence of abnormality could only be reported by looking at iso-value 13 where it covers a lump of tissue.

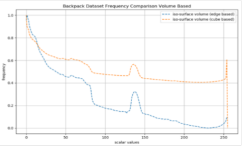
Back Pack





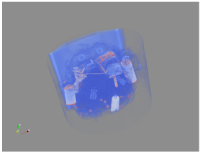


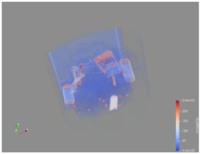




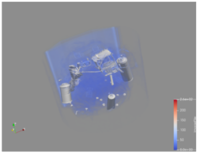
- The mean gradient, interval volume and iso surface area approximation measures all have provided with the significant clue of existence of different surface corresponding to iso-values 230, 100 and 30.

Back Pack Contd.

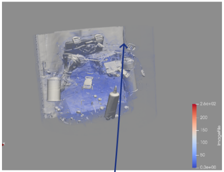




Iso-Value 230



Iso-Value 100



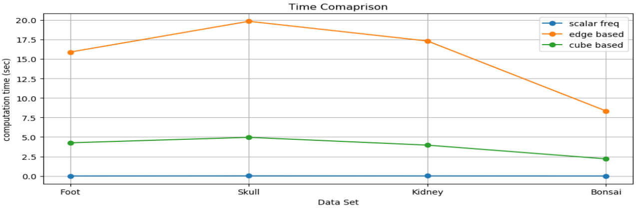
Iso-Value 30

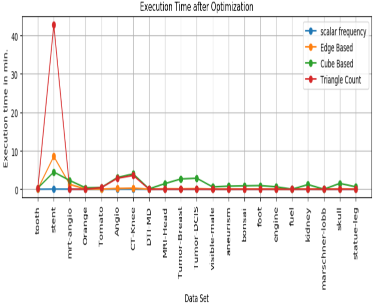
- The iso-value 230 corresponds to metallic surface, while 100 is plastic covering around the cylindrical objects. The iso-value range around 30 however covers all the objects inside the backpack.
- However, none of the measure was able to separate these objects out.

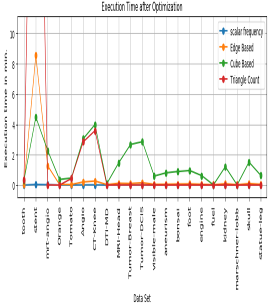
Time Optimization

- Reason for Optimization Effort : The measures requiring computation through edge and cube based marching took around 15-20 minutes for a single data set of size 256 x 256 x 256.
- For these measures it was also not trivially possible to parallelize the function which required us to implement the same functionality using some better approach suitable to the underlying programming platform (Python in our case).
- The measures especially Iso-surface area frequency relative to scalar values using Triangle Count approach were directly possible to parallelize.

Time Optimization Contd.







Challenges Faced

- Naive Implementation of the measures were in-efficient in terms of time complexity.
- Able to improve this time, but was only for $(\sigma_i, \sigma_j) = (0, 255)$, for higher range of values, speed is still slow.
- For some data sets, measures were not much useful for the significant analysis, specifically for the Tumor data sets.
- Selection of data sets from the vast resources available.

Concluding Remark on Measures

1. Iso-surface area Estimation by edge and cube:
   - It is faster to compute. Edge version includes spikes in the plot in some cases where there could be consistent iso-surfaces.
2. Accurate estimation by marching cubes:
   - It is slower to compute. It generally gives smoother plots than edge and cube based estimations.
3. Interval volume:
   - Scalar frequency:
     - Fast to compute but gives false peaks for data at low resolutions.
   - Edge based and cube based estimation:
     - Smoother than the scalar frequency and converges to scalar frequency.
4. Gradient Plots - Total Gradient Plot:
   - It cannot relied upon solely. However works in identifying minute features in combination with the Mean Gradient measures.However for Level sets with large gradients but a small area can have the same total gradient as level sets with small gradients but a large area.
5. Gradient Plots - Mean Gradient Plot:
   - It can be relied upon solely. Can be used as a goto measure.