The Goal of Visualization

- The goal of visualization is to faithfully convey the maximal amount of information from the data through the display channel.

How do we measure the information content?

View Point Selection

- Due to 3D occlusion, images generated from different views will convey different amounts of information.
- How to choose views that can convey maximal amounts of information?

Multiresolution Visualization

- How do we measure and compare the quality of different LOD selections?
- Are the computation resources effectively utilized?

Information Theory

- Study the fundamental limits to reliably transmit messages through a noisy channel.
- Model the message as a random variable whose value is taken from a sequence of symbols.
- Information content of the message is measured by Shannon’s Information Entropy.

Shannon Entropy

- The random variable takes a sequence of symbols \(a_1, a_2, a_3, \ldots, a_n\) with probabilities \(p_1, p_2, p_3, \ldots, p_n\).
- The information contained in each symbol \(a_i\) is defined as:
  \[
  \log \left( \frac{1}{p_i} \right) = - \log p_i
  \]
- The average amount of information expressed by the random variable is the entropy:
  \[
  H(x) = - \sum_{i=1}^{n} p_i \log p_i
  \]
**Properties of Entropy**

- Entropy is to measure the average uncertainty of the random variable.
- Entropy is a concave function, which has a maximum value when all outcomes are equally possible:
  \[ p_1 = p_2 = p_3 = \ldots = p_n \]

An example of three dimensional Probability vector \([p_1, p_2, p_3]\).

**Information Theory and Visualization**

- A data set can be considered as a random variable.
- Each data point can be considered as an outcome for a random variable \(X\).
- We can measure the information content of the visualization output (image).

**Finding a Good View**

- Viewpoint entropy for a camera view:
  \[ H(x) = - \sum_{i=1}^{n} p_i \log p_i \]
- Sample the view sphere and evaluate from multiple sample views.

**Object Space Approach**

- Consider that the visualization is generated from 3D data.
- Evaluate the information content from each data point (voxel) to the screen.

**Information Content of a Voxel**

- Each voxel’s information content depends on its visibility and importance:
  - Visibility \(v_i\): transparency from the camera to the voxel
  - Importance \(w_i\): the voxel’s alpha value (defined by the transfer function)
- Visual probability for a voxel \(i\):
  \[ p_i = \frac{1}{\sigma} \frac{v_i}{w_i} \]
  where \(\sigma = \sum_{i=0}^{n-1} \frac{v_i}{w_i}\)
- Information Content for the voxel:
  \[ \log \left( \frac{1}{p_i} \right) = - \log p_i \]

**Example**

\[ H(x) = - \sum_{i=1}^{n} p_i \log p_i \]
Examples

View Partitioning

- Choose only one from the nearby similar view samples
- Use Jensen-Shannon (JS) divergence measure to estimate the distance between two view entropies \( q_1 \) and \( q_2 \)

\[
JS(q_1, q_2) = 2H \left( \frac{1}{2} q_1 + \frac{1}{2} q_2 \right) - H(q_1) - H(q_2)
\]

- Cluster the view samples based on the JS divergence

View Selection Criteria

- Larger projection size
  - more voxels to be visible
- Even opacity distribution
  - opaque voxels do not clump together
- Even color distribution
  - data in different ranges can be seen
- More salient geometric features (curvature)

Image Space Approach

- Consider only the visualization output, i.e., the images
- Evaluate the information content based on the pixel value distribution in the image

Projection & Opacity Distribution

- For an Imagine contains \( n \) pixels with accumulated opacities \( \{\alpha_0, \alpha_1, \alpha_2, ..., \alpha_{n-1}\} \), the probability of pixel \( i \) is defined as:

\[
p_i = \frac{\alpha_i}{\sum_{j=0}^{n-1} \alpha_j}
\]

- Background pixels are excluded
- The view entropy is defined in the same way:

\[
H(x) = -\sum_{i=1}^{n} p_i \log p_i
\]

Color Distribution

- A well designed transfer function should highlight salient features with attentive colors
- A good view should maximize the area of the salient colors while maintaining an even distribution
- Assuming there are \( n \) colors in the transfer function \( \{C_0, C_1, ..., C_{n-1}\} \), with \( C_0 \) being the background
- For each pixel, we measure the perceptual color distance (using CIELUV) to the salient colors and classify the pixel
Entropy Calculation

- Calculate the area $A_i$ for each color $C_i$.
- The probability for $C_i$ is defined as:
  $$p_i = \frac{A_i}{T}$$
  where $T = \sum_{i=0}^{n-1}(A_i)$
- The entropy for the image is calculated similarly:
  $$H(x) = -\sum_{i=1}^{n} p_i \log p_i$$
- The entropy of an image is a maximum when all salient colors are shown.

Curvature Distribution

- Low curvatures imply flat areas and high curvatures mean highly irregular shapes.
- We can represent curvatures in a volume by color coding each voxel based on its curvature.
- High luminance colors for high curvatures.

The Final Utility Function

- Decide the best view based on opacity, color, and curvature
  $$u(v) = \alpha \cdot \text{opacity}(v) + \beta \cdot \text{color}(v) + \gamma \cdot \text{curvature}(v)$$
  where $\alpha + \beta + \gamma = 1$
- Normalize each component to $[0,1]$
- User can decide a different combination of weights and/or introduce new factors.

Multiresolution Visualization

- How do we measure and compare the quality of different LOD selections?
- Are the computation resources effectively utilized?

Global LOD Quality Metric

- Measure the amount of information contained in the selected LOD
  - Compare LODs
  - Decide whether the computation resources are distributed evenly to render-worthy blocks
  - LOD adjustment
- Approach: Information theory
LOD Entropy

A LOD contains a sequence of blocks $B_i$ at particular resolutions.

$P_i$, the 'probability' of a data block $B_i$ at a particular resolution, is defined as:

$$P_i = \frac{C_i \times D_i}{S}$$

$S = \sum_{i=1}^{M} C_i \times D_i$

$C_i$ and $D_i$ are the block's contribution and distortion (if it is a low resolution block).

Contribution and Distortion

- Contribution: the block’s color ($\mu$), projection size ($a$), thickness ($t$), visibility ($v$)

$$C_i = \mu \cdot t \cdot a \cdot v$$

- Distortion: the difference between the block’s data values and those of a higher resolution block

$$d_{ij} = \frac{\sigma_i^2 + \mu_i^2 + C_1}{2\mu_j \cdot C_1} \cdot \frac{\sigma_j^2 + \sigma_i^2 + C_2}{2\sigma_j + C_2}$$

Contribution and Distortion

Maximize the entropy function when $P_i$ are all equal.

The entropy function prefers that the block's contribution matches its resolution.

$$H(X) = -\sum_i p_i \log p_i$$

LOD Comparisons using Entropy

A higher entropy value indicates a balanced probability distribution, thus a better overall quality.

Entropy vs. Quality

Entropy = 0.166 (34 blocks)  Entropy = 0.316 (259 blocks)

Entropy vs. # of Blocks

Graph showing entropy vs. number of blocks.
Visual Representation of LOD Quality

- An optimal selection of LOD is an NP complete problem
- Fine tuning of LOD selection is often necessary
- Can we visualize what are selected, and make adjustments if necessary?

LOD Map

- A visual user interface to visualize the LOD selection
- Allow the user to see individual block’s contribution vs. distortion, i.e., visualize the entropy terms

Treemap

- A space-filling method to visualize hierarchical information [Shneiderman et al. 1992]
  - Recursive subdivision of a given display area
  - Information of each individual node
  - Color and size of its bounding rectangle

LOD Map

- Display the blocks belong to the selected LOD in a tree-map like manner
- Color (blue to red) is used to encode the block’s distortion
- The contribution of the block \(\mu \cdot t.a.v\) is divided into two parts
  - The size of rectangle is to encode \(\mu \cdot t.a\)
  - The opacity of rectangle is to encode \(v\)

How Can LOD Map Help

Comparisons of different LOD selection schemes
How Can LOD Map Help

• Spot problematic regions in the current LOD
  • Large red rectangles – high contribution blocks rendered with low resolutions
    • Action: split the blocks and increase the resolutions
  • Small blue rectangles – low contribution blocks rendered with high resolutions
    • Action: join the blocks and reduce the resolutions
  • Dark rectangles – low visibility blocks
    • Action: join them and reduce the resolutions

How Can LOD Map Help

• View selection on the fly - High entropy and brighter LOD map for better views

LOD Adjustment

• Budget Control - Render fewer blocks, i.e., lower Resolutions in certain regions, for the same entropy

Conclusions

• Entropy can be used to quantify the information content in a visualization
• Applications: view selection & LOD selection
• The goal of visualization is to reduce the uncertainty perceived by the viewer
• More applications of information theory are expected to quantify the goodness of visualization

Acknowledgements

• US DOE SciDAC Ultra Scale Visualization Institute
• NSF Career Award CCF-0346883
• DOE Early Career Principal Investigator Award DE-FG02-03ER25572
• NSF ITR grant ACI-0325934