An Information-theoretic Framework for Visualization

Min Chen and Heike Jänicke
Swansea University and Universität Heidelberg
1 Motivation: A Theoretic Framework

- Quantitative measurement and reasoning
- Explanation of facts and phenomena
- Laws and guidelines for optimization
- Falsifiable predictions

1 Motivation
2 Communication & Visualization
3 Quantifying Visual Info
4 Explanation - logarithmic plot & redundancy
5 Laws - interaction & user study
6 Prediction - DP inequality & ...
7 Conclusions & Future Work

Looking for theories?
2 Communication and Visualization

- Is a visualization system a communication system?

Example: a visual communication system image from http://chicagodesignintern.blogspot.com/
A General Visualization Pipeline (without interaction)

Source → Filtering → Visual Mapping → Rendering → Displaying → Optical Transmission

- raw data
- information
- geometry & labels
- image
- optical signal
- optical signal

Viewing → Perception → Cognition → Destination

- image
- information
- knowledge

A General Communication System

- message
- signal
- message

Source → Encoder (Transmitter) → Channel → Decoder (Receiver) → Destination
Three Visualization Subsystems

Source \rightarrow Filtering \rightarrow Visual Mapping \rightarrow Rendering
\text{vis-encoder}

Filtering \rightarrow Visual Mapping \rightarrow Rendering
\text{vis-channel}

Displaying \rightarrow Optical Transmission \rightarrow Viewing
\text{vis-decoder}

Perception \rightarrow Cognition \rightarrow Destination

A General Communication System
Three Visualization Subsystems

Source

vis-encoder

raw data

D

image

V

vis-channel

image

V'

vis-decoder

knowledge

K

Destination

Encoder (Transmitter)

Channel

Decoder (Receiver)

Destination

message

M

signal

S

signal

S'

message

M'

compactness

error detection

error correction

A General Communication System
<table>
<thead>
<tr>
<th>Information Theory Taxonomy</th>
<th>Relevance in Visualization</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fundamental Concepts</strong></td>
<td>A possible mathematical framework that underpins the subject of visualization.</td>
</tr>
<tr>
<td>Major Quantities and Properties</td>
<td>Quantitative measurements about the data and visualization space, and the relationship between input and output of a process or subsystem at different stages of a visualization pipeline.</td>
</tr>
<tr>
<td>Entropy</td>
<td>Measuring information content (see Section 5.1); salience in visualization.</td>
</tr>
<tr>
<td>Mutual Information</td>
<td>Uncertainty reduction in visualization (see Section 5.3); information-assisted visualization.</td>
</tr>
<tr>
<td><strong>Major Theorems</strong></td>
<td>Many theorems can be used to explain visualization phenomena and events.</td>
</tr>
<tr>
<td>Information balance (conservation law)</td>
<td>Given two visualizations, A and B, the amount of information about A contained in B is the same as that about B in A; overview + detail visualization; multi-view visualization.</td>
</tr>
<tr>
<td>Data processing inequality</td>
<td>After visual mapping, the visualization cannot contain more information than the original data (see Section 6.2); Information cannot be recovered after being degraded by some processes or subsystems in a visualization pipeline.</td>
</tr>
<tr>
<td><strong>Channel Types</strong></td>
<td>Providing a theoretical basis for classifying visualization subsystems (see Section 6).</td>
</tr>
<tr>
<td>Noiseless channel</td>
<td>Not common in practical visualization pipelines (see Section 3).</td>
</tr>
<tr>
<td>Noisy channel</td>
<td>Most visualization processes and subsystems can be affected by noise (see Section 3).</td>
</tr>
<tr>
<td><strong>Channel Capacity</strong></td>
<td>It can be adapted to define the maximum amount of information that can be visualized or displayed (see Sections 5.1 and 6).</td>
</tr>
<tr>
<td>Redundancy</td>
<td>Efficiency of a visual mapping; Error detection and correction (see Section 8).</td>
</tr>
<tr>
<td><strong>Source Coding (for Noiseless Channels)</strong></td>
<td>Inspiration for developing new data abstraction and visual encoding techniques.</td>
</tr>
<tr>
<td>Coding Schemes</td>
<td>Applicable, for example, to the following visualization algorithms:</td>
</tr>
<tr>
<td>Entropy coding (e.g., Huffman, arithmetic coding)</td>
<td>Logarithmic plots (see Section 7.1); importance-based visualization; information-assisted visualization; magic lens; illustrative deformation.</td>
</tr>
<tr>
<td>Dictionary-based coding</td>
<td>Legend design; icon design; visualization literacy; knowledge-assisted visualization.</td>
</tr>
<tr>
<td>Run-length encoding</td>
<td>Spatial clustering; cluttering reduction.</td>
</tr>
<tr>
<td>Differential encoding</td>
<td>Video visualization; time-varying data visualization.</td>
</tr>
<tr>
<td>Subband coding</td>
<td>Multi-resolution modeling; transfer function design in volume rendering; perceptually-based visual encoding, frequency-domain volume rendering.</td>
</tr>
<tr>
<td>Transform coding</td>
<td>Color mapping; multi-resolution modeling.</td>
</tr>
<tr>
<td>Multiplexing</td>
<td>Comparative visualization; volume rendering; multi-field visualization.</td>
</tr>
</tbody>
</table>
3 Quantifying Visual Information

- Random variable
  \( X \)
- It takes values
  \( x_1, x_2, \ldots, x_m \)
- Probability mass function
  \( p(x_i) \)
- Entropy
  \[ H(X) = -\sum_{i}^{m} p(x_i) \log_2 p(x_i) \]

Claude E. Shannon
(1916-2001)
Entropy: Example

- **Time Series**
  - 64 independent samples
  - Each sample is an integer in [0, 255]
  - Probability mass function is i.i.d. (independent and identically-distributed)

- **Entropy**
  \[
  H(Z) = - \sum_{t=0}^{64} \sum_{i=0}^{255} \frac{1}{256} \log_2 \left( \frac{1}{256} \right) = 512
  \]

- **Time Series Plot**
  - Minimal 256x64 pixels (\(2^{14}\) pixels)
The Most Compact, Why Not?

- The most compact:
  - 64 bytes (512 bits)

- One pixel per bit seems not enough for vision
  - 4x4 pixels per bit \(\rightarrow 2^{13}\) bits

- Not a counter example:
  - Sequential or parallel
  - Salient information
Three Visualization Subsystems

Source
→ vis-encoder
→ image V
→ vis-channel
→ image V'
→ vis-decoder
→ Destination

perceptual efficiency
intuitiveness
clarity

Encoder
(Transmitter)
Source
→ Encoder
message M
→ Channel
signal S
→ Decoder
(Receiver)
message M'
→ Destination

compactness
error detection
correction

A General Communication System
Three Measures for Visualization

- Entropy of Input Data Space: $H(X)$
- Visualization Capacity: $V(G)$
- Display Capacity: $D$

cf. (Yang-Peláez & Flowers, 2000)

Visual Mapping Ratio (VMR) = \[ \frac{V(G)}{H(X)} \]

Information Loss Ratio (ILR) = \[ \frac{\max(H(X) - V(G), 0)}{H(X)} \]

Display Space Utilization (DSU) = \[ \frac{V(G)}{D} \]
4 Explanation - Logarithmic Plot & Redundancy

- Logarithmic plot
  - Why is it useful in some situations?
  - In what situation?
  - What does it try to optimize?

- Redundancy in Visualization Design
  - Good and bad?
  - When is it useful?
  - Should this community pay more attention to “redundancy”?
Information Loss Ratio (ILR)

- Display Space Restriction
  - 64x64 pixels
- Evenly distributed probability mass function
- Linear visual mapping
- ILR is a probabilistic measure about
  - a data space $X$
  - not an instance $x_i$
Non-uniform Distribution

- Linear visual mapping

<table>
<thead>
<tr>
<th>probability</th>
<th>$Z$</th>
<th>linear</th>
<th>$Z'$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: $p = \frac{1}{2}$</td>
<td>32</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B: $p = \frac{1}{4}$</td>
<td>96</td>
<td>24</td>
<td>48</td>
</tr>
<tr>
<td>C: $p = \frac{1}{8}$</td>
<td>160</td>
<td>40</td>
<td>48</td>
</tr>
<tr>
<td>D: $p = \frac{1}{8}$</td>
<td>224</td>
<td>56</td>
<td>64</td>
</tr>
</tbody>
</table>

Information loss:
- Evenly distributed $p$: 25.0%
- Unevenly distributed $p$: 25.8%
Non-uniform Distribution

- Nonlinear visual mapping

Information loss:
- Evenly distributed $p$: 25.0%
- Unevenly distributed $p$: 25.8%
- 4 regional mappings: 22.6%

(a) evenly distributed $p$
(b) unevenly distributed $p$
(c) 4 regional mappings
Non-uniform Distribution

- Logarithmic visual mapping

<table>
<thead>
<tr>
<th>A:</th>
<th>p = 1/2</th>
</tr>
</thead>
<tbody>
<tr>
<td>B:</td>
<td>p = 1/4</td>
</tr>
<tr>
<td>C:</td>
<td>p = 1/8</td>
</tr>
<tr>
<td>D:</td>
<td>p = 1/8</td>
</tr>
</tbody>
</table>

$k \to \infty$

$p = \frac{1}{2^k}$
$p = \frac{1}{2^{k-1}}$
$p = \frac{1}{2^3}$
$p = \frac{1}{2^2}$
$p = \frac{1}{2}$

Information loss:
- 25.0%
- 25.8%
- 22.6%
- 0%

(a) evenly distributed $p$
(b) unevenly distributed $p$
(c) 4 regional mappings
(d) logarithmic plot
Three Visualization Subsystems

A General Communication System
A General Communication System
Display Space Utilization (DSU) is generally very low

- 25-50 fps refresh rate can be utilized

- Redundancy
  
  - see Rheingans & Landreth, 1995
Law 3:
The information about an overview in one of its detailed view is the same as that about that detailed view in the overview.
Example 1

**Mutual Information**

\[
I(X,Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log_2 \left( \frac{p(x,y)}{p(x)p(y)} \right)
\]

- Example 1: \( I = 0.147 \)

<table>
<thead>
<tr>
<th>( p(x,y) )</th>
<th>Hint</th>
<th>No Hint</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vortex</td>
<td>25%</td>
<td>25%</td>
</tr>
<tr>
<td>No Vortex</td>
<td>5%</td>
<td>45%</td>
</tr>
</tbody>
</table>
Mutual Information

Example 1:

- Example 1: $I(X;Y) = 0.147$

Example 2:

- Example 2: $I(X;Y) = 0.278$
The Role of User Studies

- Do quantitative measurements make user studies less important?
If we can measure $p(x,y), p(x), p(y), ...$

- Example 1: $I(X;Y) = 0.147$
- Example 2: $I(X;Y) = 0.278$
6 Prediction – DP Inequality & a Prediction

Data processing inequality

- “No clever manipulation of data can improve the inferences that can be made from the data” [Cover and Thomas, 2006]

Markov chain conditions
- Closed coupling: (X, Y), (Y,Z)
- X and Z are conditionally independent

What if the condition is broken?

\[ I(X;Y) \geq I(X;Z) \]
Data, Information, and Knowledge in Visualization

In visualization, we use the terms data, information, and knowledge extensively, often in an interrelated context. In many cases, they indicate different levels of abstraction, understanding, or truthfulness. For example, “visualization is concerned with exploring data and information,” and “information visualization” is for “data mining and knowledge discovery.” In other cases, these three terms indicate data types, for instance, as adjectives in noun phrases, such as data visualization, information visualization, and knowledge visualization. These examples suggest that data, information, and knowledge could serve as both the input and output of a visualization process, raising questions about their exact role in visualization.

There are many competing definitions of data, information, and knowledge, in different aspects of computer science and engineering and in other disciplines such as psychology, management sciences, and epistemology (the theory of knowledge). The use of the three terms isn’t consistent and is often conflicting. For instance, data and information are often interchangeable in computing (for example, data processing and information processing or data management and information management). From a systems perspective, however, data is referred to as bits and bytes stored on or communicated via a digital medium. So any computerized representations, including knowledge representations, are types of data. On the other hand, from the perspective of knowledge-based systems, data is a simpler form of knowledge.

Researchers have attempted to clarify the taxonomy of terms used in the visualization community (for example, in the work of Ed H. Chi, Ben Shneiderman, and Melanie Tary and Torsten Möller). However, the terms data, information, and knowledge remain ambiguous. This article doesn’t attempt to offer a different taxonomy for visualization. Instead, we differentiate these three terms from the perspective of visualization processes. Furthermore, we examine the current and future role of information and knowledge in the development of visualization technology.

Definitions of Data, Information and Knowledge

Since we use data, data mining, and knowledge discovery in the computer, we must differentiate these terms in the perceptual and cognitive space. Because we can also store data, information, and knowledge in the computer, we must also differentiate them in the computational space.

Perceptual and Cognitive Space

The data-information-knowledge-wisdom (DIKW) hierarchy is a popular model for classifying human understanding in the perceptual and cognitive space. The origin of this hierarchy can be traced to the poet T.S. Eliot. Table 1 shows Russell Ackoff’s definitions of data, information, and knowledge.

Let \( D \) be the set of all possible explicit and implicit human memory. The former encompasses the memory of events, facts, and concepts, and the understanding of their meanings, context, and associations. The latter encompasses all unconscious forms of memory, such as emotional responses, skills, and habits. We can thus focus on three subsets of memory, \( D_{\text{data}}, D_{\text{info}}, \) and \( D_{\text{know}} \), where \( D_{\text{data}}, D_{\text{info}}, \) and \( D_{\text{know}} \) are the sets of all possible explicit and implicit memory about data, information, and knowledge, respectively.

Despite the lack of an agreeable set of the definitions of data, information, and knowledge, a consensus exists that data isn’t information, and information isn’t knowledge. Without diverting from this article’s scope, here we simply assume that \( D_{\text{data}} \subseteq D_{\text{info}} \) and \( D_{\text{info}} \subseteq D_{\text{know}} \) and none of them is a subset of another. Without losing generality, we can generalize \( D_{\text{know}} \) to include wisdom, and any other high level of understanding, in the context of the DIKW hierarchy.
The process of gaining insight can be improved by using numbers.

Information theory can explain many phenomena in visualization,
- but be careful with naive use.

The current limitations:
- assumption of memoryless,
- lack of real $p(x_i)$.

This talk is an overview of the paper.

Details of related work,
- in particular, Matt Ward’s suggestion (see Purchase et al. 2008)

“The purpose of computing is insight, not numbers.”

Richard W. Hamming
(1915-1998)
The Role of a Theoretic Framework

Facts

Wisdom

Theory

Theoretic Framework
<table>
<thead>
<tr>
<th>Information Theory Taxonomy</th>
<th>Relevance in Visualization</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fundamental Concepts</strong></td>
<td>A possible mathematical framework that underpins the subject of visualization.</td>
</tr>
<tr>
<td>Major Quantities and Properties</td>
<td>Quantitative measurements about the data and visualization space, and the relationship between input and output of a process or subsystem at different stages of a visualization pipeline.</td>
</tr>
<tr>
<td>Entropy</td>
<td>Measuring information content (see Section 5.1); salience in visualization.</td>
</tr>
<tr>
<td>Mutual Information</td>
<td>Uncertainty reduction in visualization (see Section 5.3); information-assisted visualization.</td>
</tr>
<tr>
<td>Major Theorems</td>
<td>Many theorems can be used to explain visualization phenomena and events.</td>
</tr>
<tr>
<td>Information balance (conservation law)</td>
<td>Given two visualizations, A and B, the amount of information about A contained in B is the same as that about B in A; overview + detail visualization; multi-view visualization.</td>
</tr>
<tr>
<td>Data processing inequality</td>
<td>After visual mapping, the visualization cannot contain more information than the original data (see Section 6.2); Information cannot be recovered after being degraded by some processes or subsystems in a visualization pipeline.</td>
</tr>
<tr>
<td>Channel Types</td>
<td>Providing a theoretical basis for classifying visualization subsystems (see Section 6).</td>
</tr>
<tr>
<td>Noiseless channel</td>
<td>Not common in practical visualization pipelines (see Section 3).</td>
</tr>
<tr>
<td>Noisy channel</td>
<td>Most visualization processes and subsystems can be affected by noise (see Section 3).</td>
</tr>
<tr>
<td>Channel Capacity</td>
<td>It can be adapted to define the maximum amount of information that can be visualized or displayed (see Sections 5.1 and 6).</td>
</tr>
<tr>
<td>Redundancy</td>
<td>Efficiency of a visual mapping; Error detection and correction (see Section 8).</td>
</tr>
<tr>
<td>Source Coding (for Noiseless Channels)</td>
<td>Inspiration for developing new data abstraction and visual encoding techniques.</td>
</tr>
<tr>
<td>Coding Schemes</td>
<td>Applicable, for example, to the following visualization algorithms:</td>
</tr>
<tr>
<td>Entropy coding (e.g., Huffman, arithmetic coding)</td>
<td>Logarithmic plots (see Section 7.1); importance-based visualization; information-assisted visualization; magic lens; illustrative deformation.</td>
</tr>
<tr>
<td>Dictionary-based coding</td>
<td>Legend design; icon design; visualization literacy; knowledge-assisted visualization.</td>
</tr>
<tr>
<td>Run-length encoding</td>
<td>Spatial clustering; cluttering reduction.</td>
</tr>
<tr>
<td>Differential encoding</td>
<td>Video visualization; time-varying data visualization.</td>
</tr>
<tr>
<td>Subband coding</td>
<td>Multi-resolution modeling; transfer function design in volume rendering.</td>
</tr>
<tr>
<td>Transform coding</td>
<td>Perceptually-based visual encoding, frequency-domain volume rendering.</td>
</tr>
<tr>
<td>Quantization</td>
<td>Color mapping; multi-resolution modeling.</td>
</tr>
<tr>
<td>Multiplexing</td>
<td>Comparative visualization; volume rendering; multi-field visualization.</td>
</tr>
</tbody>
</table>
An Information-theoretic Framework for Visualization

Min Chen and Heike Jänicke
Swansea University and Universität Heidelberg