CS855: Data Visualization

Jaya Sreevalsan Nair

International Institute of Information Technology, Bangalore

Term I: 2014-15: Lecture 02

- Assignments 0 and 1 are due 23:59:59 hrs (IST) August 17, 2014 (Sunday).
- Assignment 7: Erratum in assignment booklet (pg 3):
  A single file in .pdf format: <RollNumber>_Assignment7.pdf, prepared in IEEE conference format using LaTeX style files, to be e-mailed to instructor by the due date for the assignment.
- List of papers for Assignment 7 will be on LMS by August 31, 2014; and the choice of paper must be emailed to instructor by September 7, 2014.

- Data Representation
- Graphics Pipeline
- Visualization Pipeline

- Continuous data comes from simulations; and is typically sampled to be discretized.
- Sampled data is discrete, but falls under the category of "intrinsically" continuous data.
- The large categories are: "intrinsically" continuous data and "intrinsically" discrete data.
- There are no equivalents of each other is each other's "domains".

- Continuous data comes from simulations; and is typically sampled to be discretized.
- Sampled data is discrete, but falls under the category of "intrinsically" continuous data.
- The large categories are: "intrinsically" continuous data and "intrinsically" discrete data.
- There are no equivalents of each other is each other's "domains".
- Continuous data may be modeled as a function $f : D \to C$, where $D \subset \mathbb{R}^d$ and $C \subset \mathbb{R}^c$; f is called a d-dimensional (or d-variate), c-value function. f or its sampled version is called a field.
- Cauchy criterion: f is continuous if for every point $p \in C$ the following holds:
  $\forall \epsilon > 0, \exists \delta > 0$ such that if $||x - p|| < \delta, x \in C$, then $||f(x) - f(p)|| < \epsilon$.
- Triplet $\mathcal{D} = (D, C, f)$ defines a continuous dataset.

- Geometrical (d) vs topological (s) dimensions: d is the dimensionality of the space $\mathbb{R}^d$ into which function domain D is embedded, and s is the dimensionality of the function domain itself (s$\leq$d).
- e.g. D is a plane or curved surface embedded in $\mathbb{R}^3$, then s=2, d=3.
- Co-dimension is the difference between geometric and topological dimensions (d-s).
- Usually geometrical dimension is 3 for most visualization applications; hence topological dimension is the variable and is also called dataset dimension. e.g. spatio-temporal datasets are 4-dimensional, and usually captured as sequences of spatial data.

- Function values are usually called attributes.
- Dimensionality c of co-domain also called attribute dimension.
- $1 \leq c \leq 4$ usually, e.g. scalar, vector, color, tensor attributes.

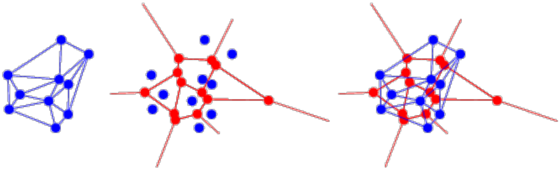
- Data acquistion provides *samples*.
- *Resampling* often occurs to get regular structured grids (or data).
- Understanding sampling is important to avoid *aliasing* (jagged artifacts) as well as *Moiré patterns*.
- Shannon sampling theorem: states that a band-limited signal (i.e., restricted to a particular frequency range), that is sampled with a frequency at least twice its highest frequency (Nyquist frequency), is completely determined by its samples.

We will cover in detail in the next class. For further reference:

- http://graphics.cs.ucdavis.edu/~okreylos/TAship/Seminar/SamplingTheory101.pdf
- http://www.pbrt.org/chapters/pbrt_chapter7.pdf
- http://graphics.cs.cmu.edu/courses/15-463/2011_fall/Lectures/SamplingReconstruction.pdf

Points without grid – however, part of continuous dataset; hence visualization using:

- sampled and reconstructed grids, point clouds, triangulation.



Delaunay triangulation *Voronoi diagram* *Delaunay and Voronoi*

Image courtesy: http://mathworld.wolfram.com/VoronoiDiagram.html

Cell types:

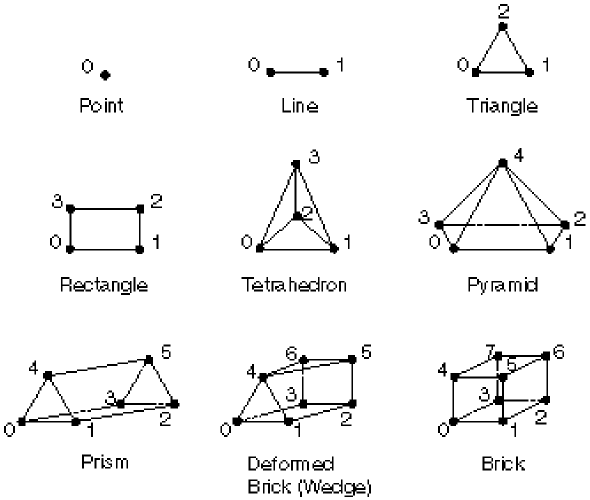
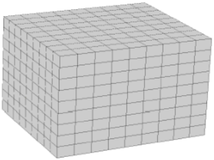
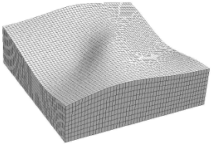


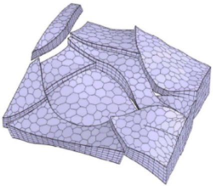
Image courtesy: SGI.

- Regular grid: position of vertex determined using indices in principal dimensions.
- Irregular grid: position stored explicitly.
- Structured grid: made up of topologically similar elements.
- Unstructured grid: made up of topologically dissimilar elements.
- Rectilinear grid: Axis-aligned matrix-like point ordering.
- Curvilinear grid: Edges need not be straight lines.

Grid Types


(a) Structured regular grid


(b) Structured irregular (stratigraphic) grid


(c) Unstructured 2.5D irregular grid

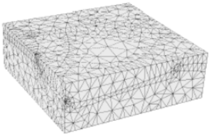
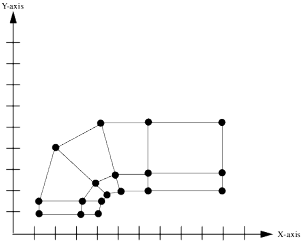
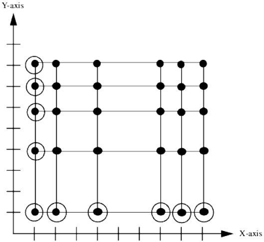

(d) Unstructured irregular (Fully unstructured) grid

Image courtesy: https://pangea.stanford.edu/ERE/pdf/pereports/PhD/Moog2013.pdf

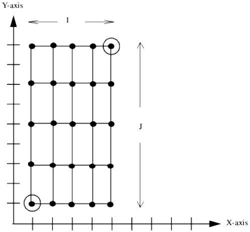
Structured Grid Types



Curvilinear



Rectilinear



Uniform

Image courtesy: https://geo-ide.noaa.gov/wiki/index.php?title=Structured_grids

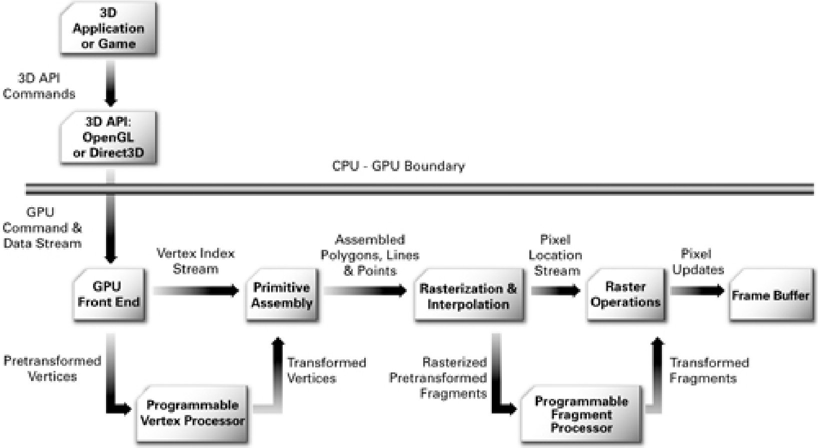


Image courtesy: NVIDIA CG

In short, can be described as *Modeling-Rendering-Displaying* paradigm.

Visualization Pipeline: Simplified Overview

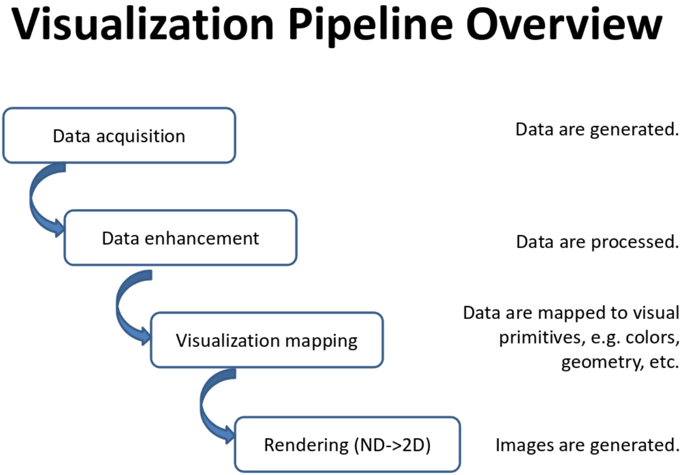


Image courtesy: http://www2.cs.uh.edu/~chengu/Teaching/Fall2013/Lecs/Lec2.pdf

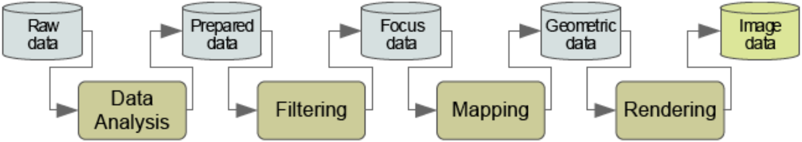


Image courtesy: http://www.infovis-wiki.net/index.php/Visualization_Pipeline

from: dos Santos, S. and Brodlie, K. Gaining understanding of multivariate and multidimensional data through visualization. Computers & Graphics, 28(3):311-325. 2004.

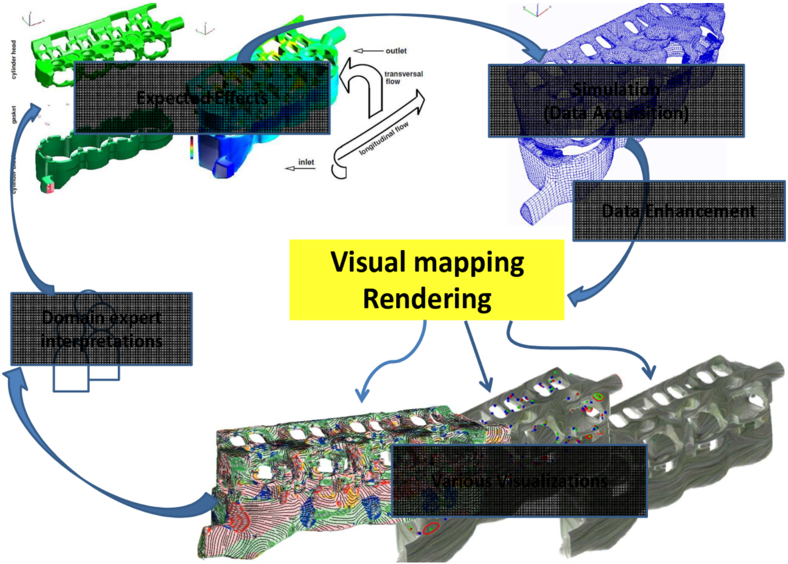


Image courtesy: http://www2.cs.uh.edu/~chengu/Teaching/Fall2013/Lecs/Lec2.pdf

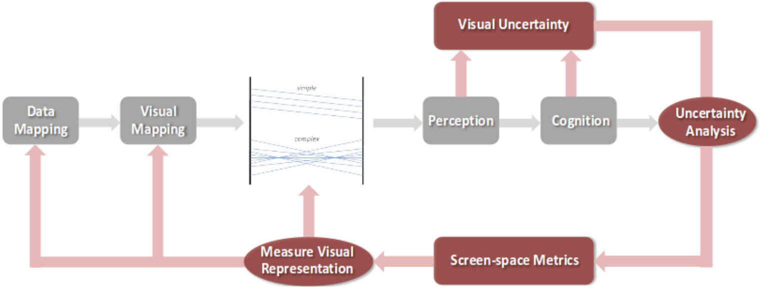


Figure 1: The conventional visualization pipeline (gray) augmented with a feedback loop (red) from the human-side to the machine-side for making visual representations better informed about visual uncertainty and reduce the number of *unknown unknowns* from a visualization designer's point-of-view.

Image courtesy: A. Dasgupta, and R. Kosara, *The importance of tracing data through the visualization pipeline*. In Proceedings of the 2012 BELIV Workshop: Beyond Time and Errors - Novel Evaluation Methods for Visualization (BELIV '12). ACM, New York, NY, USA, Article 9, 5 pages, 2012.