Interactive High Dimensional Visualization
Challenges

• High dimensional data are popular
  – Digital libraries, bioinformatics, simulations, surveys

• Challenges of visualizing high dimensional datasets
  – Clutter
  – Difficult investigation and navigation in 2D visualization space
Example

• An OHSUMED dataset:
  – 215 dimensions
  – Parallel coordinates will have 215 axes
  – Hard to understand
Solution

• Dimension reduction
  – Project a high-dimensional dataset to a lower dimensional subspace
  – Visualize data items in the lower-dimensional subspace
Dimension Reduction

- Subspace projection from the data
- This process is not a pure visualization topic
- Existing methods
  - Principal Component Analysis
  - Multidimensional Scaling (MDS)
  - Self Organizing Map
Multidimensional Scaling

- **Multidimensional scaling (MDS) is a mathematical technique that maps the distances between subjects in a high dimensional space into a lower (for example, two) dimensional space.**
MDS Strategy

- Relocate \( n \) subjects in a 2D dimensional space as points
- So that the differences between pairs of points in this reduced space match the ordered differences between the subjects in original dimension
- Such “match” is optimally approximated and usually cannot be completely satisfied
Matching Distances in MDS

• Minimize the difference of distances in original and projection data space

\[ s = \sqrt{\frac{\sum (d_{ij} - \hat{d}_{ij})^2}{\sum d_{ij}^2}} \]

• \( \hat{d}_{ij} \) is the distance between two subjects in original space
• \( d_{ij} \) is the distance between two subjects in projected space
Problems

- Information loss
- Non intuitive meaning of generated dimensions
- Little user interaction allowed
Visual Hierarchical Dimension Reduction

- Use dimension hierarchy to convey dimension relationships
- Allow users to learn the dimension hierarchy
- Allow users to select dimensions or dimension clusters to form subspaces of interests

J. Yang, M.O. Ward, E.A. Rundensteiner and S. Huang, VisSym’03
Dimension Hierarchy

- Similar dimensions form cluster, clusters are grouped into larger clusters
- A dimension hierarchy of a 5-D dataset
Framework

- Step 1: build dimension hierarchy
- Step 2: navigate and manipulate dimension hierarchy
- Step 3: interactively select clusters from dimension hierarchy to form lower dimensional subspaces
Overview
Build Dimension Hierarchy

- Automatic dimension clustering
  - Cluster dimensions according to dissimilarities among them
- Manual hierarchy modification
Representative Dimension

- Dimension Cluster Representation
  - Representative Dimension - a dimension that represents a cluster of dimensions

- Approaches
  - Select a dimension from the cluster
  - Average all dimensions in the cluster
  - Use principal component analysis to generate weighted sum of dimensions within a cluster
Dissimilarity Representation

- Axis Width Method to show dissimilarity among a dimension cluster
Example

A 42 dimensional census data set

A 5 dimensional subspace visualization finding an interesting cluster
Interactive Hierarchical Dimension Exploration

- Large number of dimensions need to be managed
  - Ordering, spacing, filtering etc.

Jing Yang, Wei Peng, Matthew O. Ward and Elke A. Rundensteiner, InfoVis’03
Dimension Ordering

- Parallel coordinates with random axis ordering
Dimension Ordering

- Parallel coordinates with ordering by dimension similarity
Dimension Ordering

- Order dimensions
  - Similarity-oriented ordering: put similar dimensions close to each other
  - Importance-oriented ordering: map more important dimensions to more significant positions or attributes. The order of importance can be decided by Principal Component Analysis (PCA)
Dimension Spacing

- Convey dimension relationship information by varying the spacing between adjacent axes
- Similar dimensions are close to each other
Dimension Filtering

- Similar dimensions can be omitted
- Unimportant dimensions can be omitted
More challenges

- Can high dimensional datasets be visualized without dimension reduction to avoid information loss?
- Can dimension relationships be visualized in the same display as data values?
Visualization without Dimension Reduction

- Visualize SkyServer dataset (361 dimensions) using existing techniques
  - Parallel Coordinates: 361 axes
  - Scatterplot Matrix: 130,321 scatterplots
  - Pixel-Oriented techniques without overlaps: 50,000 data items: 18,050,000 pixels (23 times of number of pixels in a 1024*768 screen)
Dimension Relationship Visualization

- Sorting dimensions in a 1D or 2D grid
  - Not effective beyond hundreds of dimensions

Pixel-Oriented: Sort 50 dimensions in a 2D grid [Ankerst 98]
Value and Relation Display

Jing Yang, Anilkumar Patro, Shiping Huang, Nishant Mehta, Matthew O. Ward and Elke A. Rundensteiner, InfoVis’04
Features

• Conveys data values without dimension reduction
• Conveys dimension relationships

SkyServer dataset: 361 dimensions, 50,000 data items
Interactive Tool

- A rich set of interaction tools is necessary
- Scaling
Interaction

- Distortion
- Focus-within-context
- Enlarges clicked glyphs
Interaction

- Pixel-oriented techniques are critical
- Coloring by value difference of dimensions being compared
More: XRay Dimension Glyphs

- Each glyph: a scatterplot of two selected dimensions
  - X: a base dimension
  - Y: represented dimension
- Density based display
  - Bright: sparse and Dark: dense
  - Unoccupied area: semitransparent
A real dataset with 89 dimensions and 10,417 data items
Labeling

- All labels are shown compared with Labels of selected dimensions are shown
Case Study

• **Semantic Image Browser**
  – Interactive image exploration
  – Evaluate, monitor and improve analysis processes

• Applications: personal image management, satellite image analysis, ...
Annotation engine

- Automated semantic image classification process
  - Content-Based Image Annotation
- Semantic contents
  - high dimensional dataset
  - data items: images
  - dimensions: contents
Image View – MDS layout

1100 images, 20 contents
Content View

- MDS layout of content blocks
Major References

- Jing Yang, Lecture notes, UNCC
- Colin Ware. Information visualization, 2004
- Daniel Keim. Tutorial note in InfoVis 2000
- John Stasko. Course slides, Fall 2005