Visualization for High-dimensional Data: 

*VisHD*

Cheng-Chih Yang  
Cheng-Chieh Chiang  
Yi-Ping Hung  
Greg C. Lee

Image Processing and Pattern Recognition Lab.  
Dept. of Computer Science and Information Engineering  
National Taiwan University
Outline

- Introduction
- Dimension Reduction
- System Description
- Demonstration and Discussion
- Conclusion
Introduction

• Handling high-dimensional data is important in computer science
• Dimension reduction is a common practice
• In this work
  – Based on dimension reduction
  – Provide a visualization tool, VisHD
  – “Display” the spatial distribution of vector points in high dimensional feature space
Dimension Reduction

• Goal
  – reducing the data dimension with some criteria
  – for example, using LDA for getting best discriminability

• Methods adopted here
  – Principal Components Analysis (PCA)
  – Linear Discriminant Analysis (LDA)
  – Sammon Algorithm
Dimension Reduction

- Principal Components Analysis (PCA)
  - linear, unsupervised, maximises preserved variance
- Linear Discriminant Analysis (LDA)
  - linear, supervised, maximises class separability
Dimension Reduction

- **Sammon Algorithm**  
  - Non-linear, unsupervised  
  - preserving the pair-wise distances of data points

[Sammon 1969]
VisHD
Description

• Data
  – text file for feature vectors
  – images associated with the input vectors
  – labeled or unlabeled

• Visualization
  – projecting data into 2-D space
  – displaying labeled data in different tab-pages

• Auxiliary functions
  – zoom in/out
  – storing screen or result as images
  – details of individual data
Data Flow

- **Data set**
  - **Labeled**
    - PCA
    - LDA
    - Sammon
    - Manual Selection
    - Data visualization with labels in the reduced feature space
  - **Unlabeled**
    - c-mean
    - PCA
    - Sammon
    - Manual Selection
    - Data visualization without labels in the reduced feature space
Visualization
Visualization
Visualization
Visualization
Demonstration and Discussion

• Data Set
  – Face images
    • XM2VTS database
    • four poses for each person
    • 16 images for four persons
  – Corel Photos
    • four categories
    • 100 images for each category
    • 400 images total

International Conference on Information Visualization 2005
Demonstration and Discussion

Data: Corel Photo
Reduction: LDA
Feature: color moments (C) in Lab color space, co-occurrence texture (T)

Color moments is better for discriminating these images than co-occurrence texture
Demonstration and Discussion

Data: face images
Reduction: all of available methods in *VisHD*
Feature: color moments in Lab color space

We can observe the data characteristics under different criteria of dimension reduction

International Conference on Information Visualization 2005
Demonstration and Discussion

Data: face images
Reduction: None
Feature: two dimensions selected by the user in color moments

We can observe the data characteristics under different selections of two variables.
Conclusion

- *VisHD*
  - integrates different techniques of dimension reduction
  - displays the data in the reduced subspace
- By use of *VisHD*, we can realize the characteristics of data with:
  - different feature spaces
  - different techniques of dimension reduction
- **URL**
  http://ippr.csie.ntu.edu.tw:8080/demoDownload.html
Thank You!!