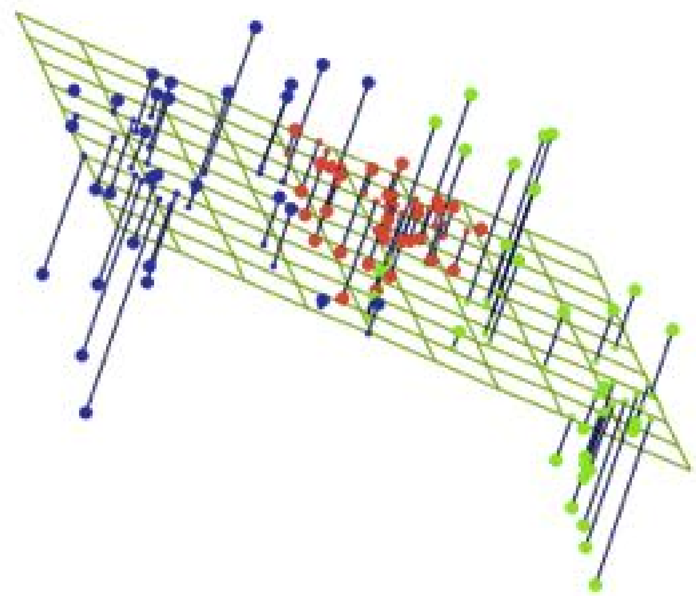
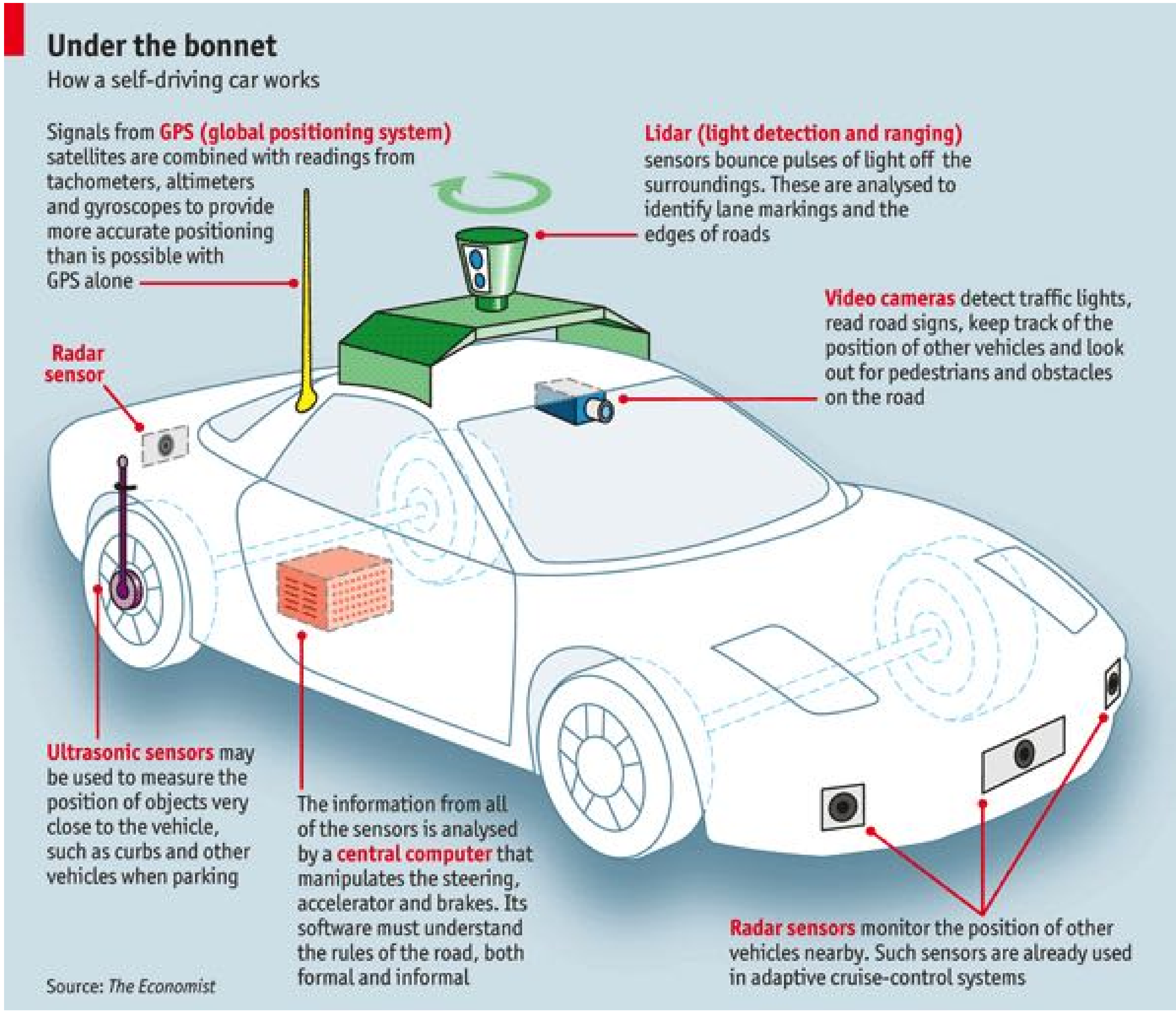
PCA, MDS, and SVD

Hanspeter Pfister
pfister@seas.harvard.edu

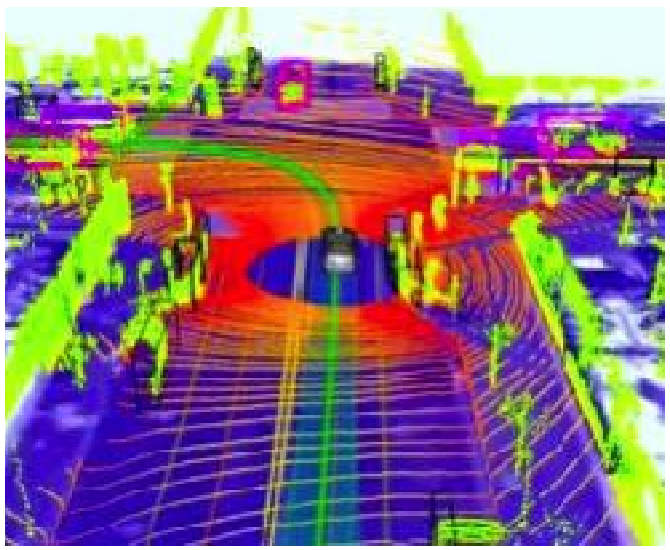


Self-Driving Cars



Car Features

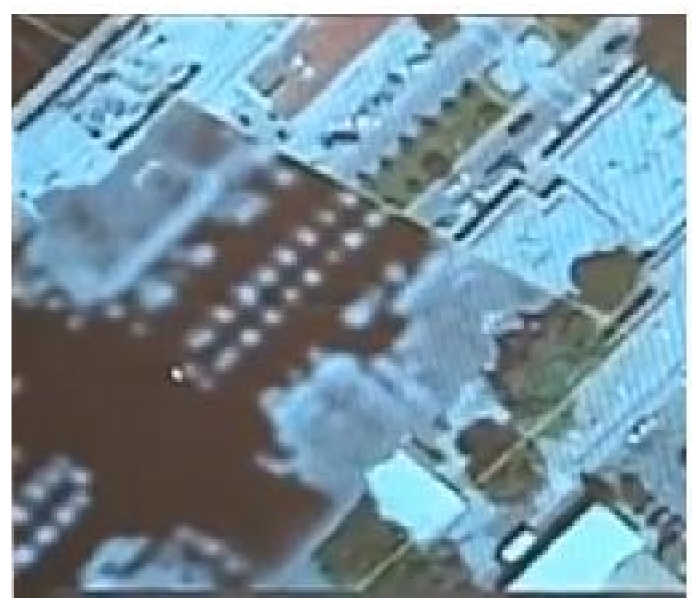
Laser scan



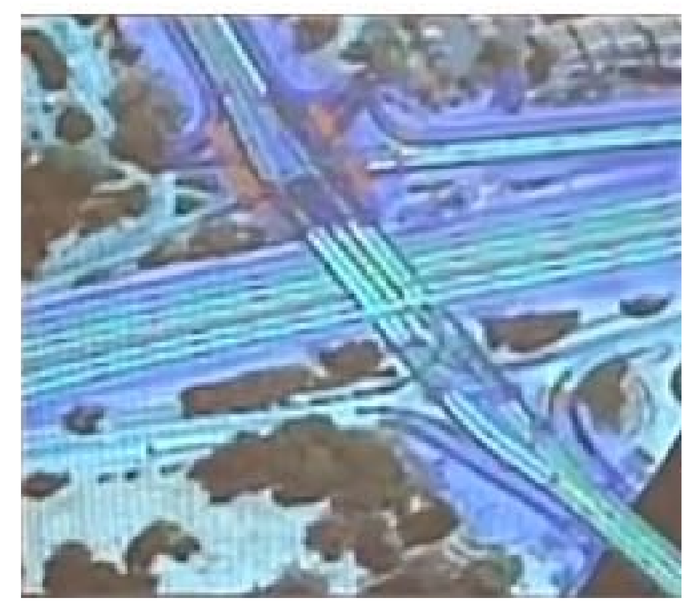
Intensity model

Elevation model

Camera vision



2D map



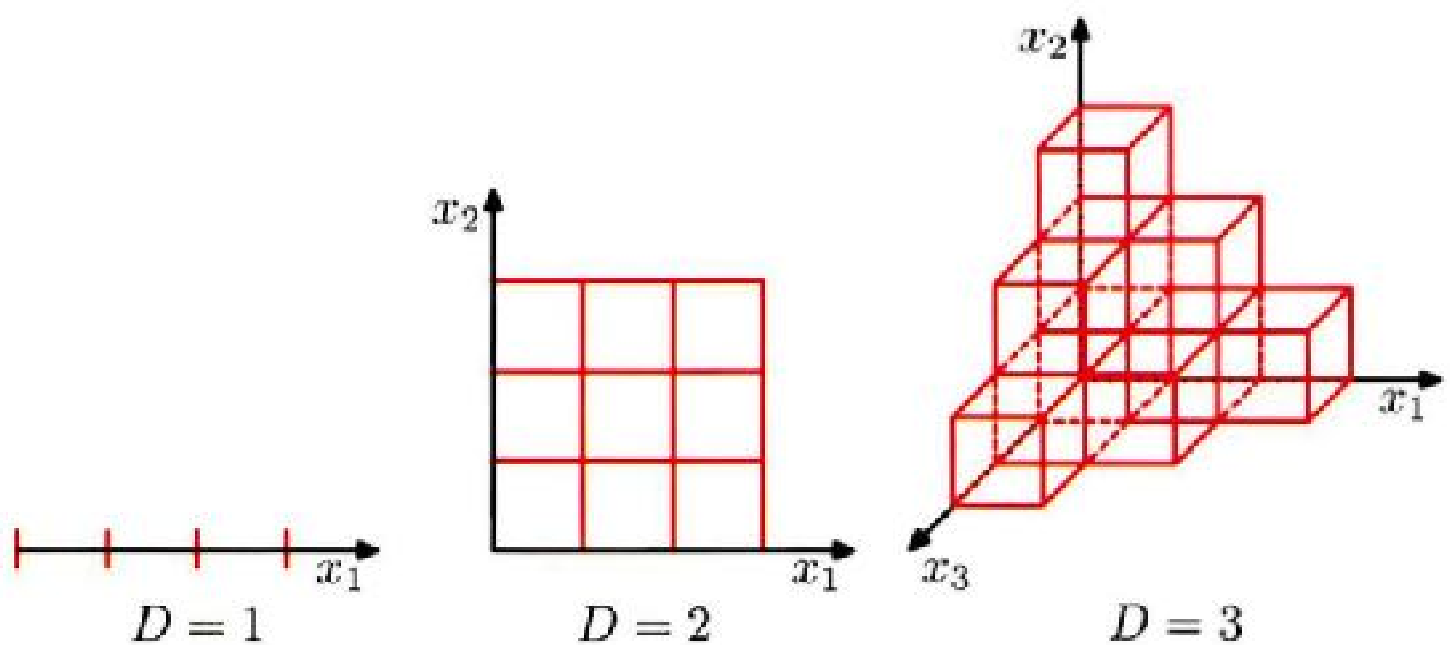
Lane model

Why don’t we just use more and more features?

Curse of Dimensionality

- When dimensionality increases, the volume of the space increases so fast that the available data becomes sparse

- Statistically sound results require the sample size N to grow exponentially with d



What are some examples of high-dimensional data?

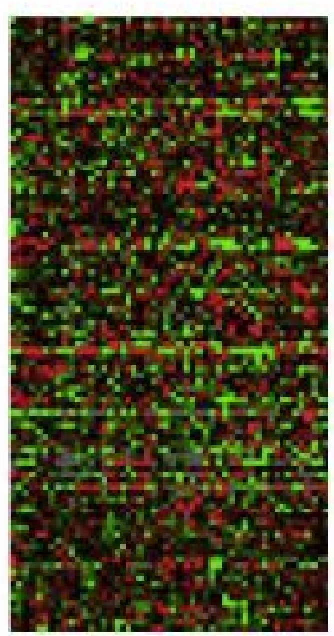
High-Dimensional Data

face images

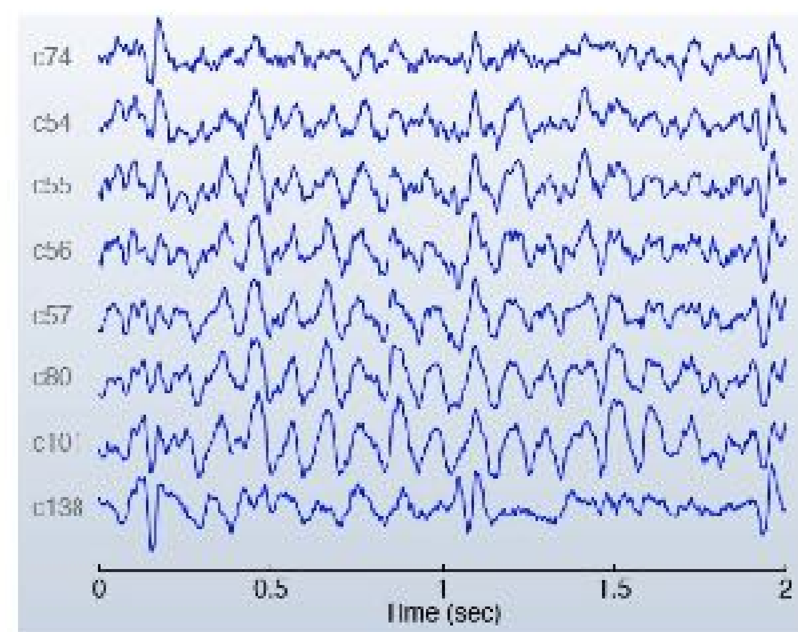
Zambian President Levy Mwanawasa has won a second term in office in an election his challenger Michael Sata accused him of rigging, official results showed on Monday.

According to media reports, a pair of hackers said on Saturday that the Firefox Web browser, commonly perceived as the safer and more customizable alternative to market leader Internet Explorer, is critically flawed. A presentation on the flaw was shown during the ToorCon hacker conference in San Diego.

documents



gene expression data



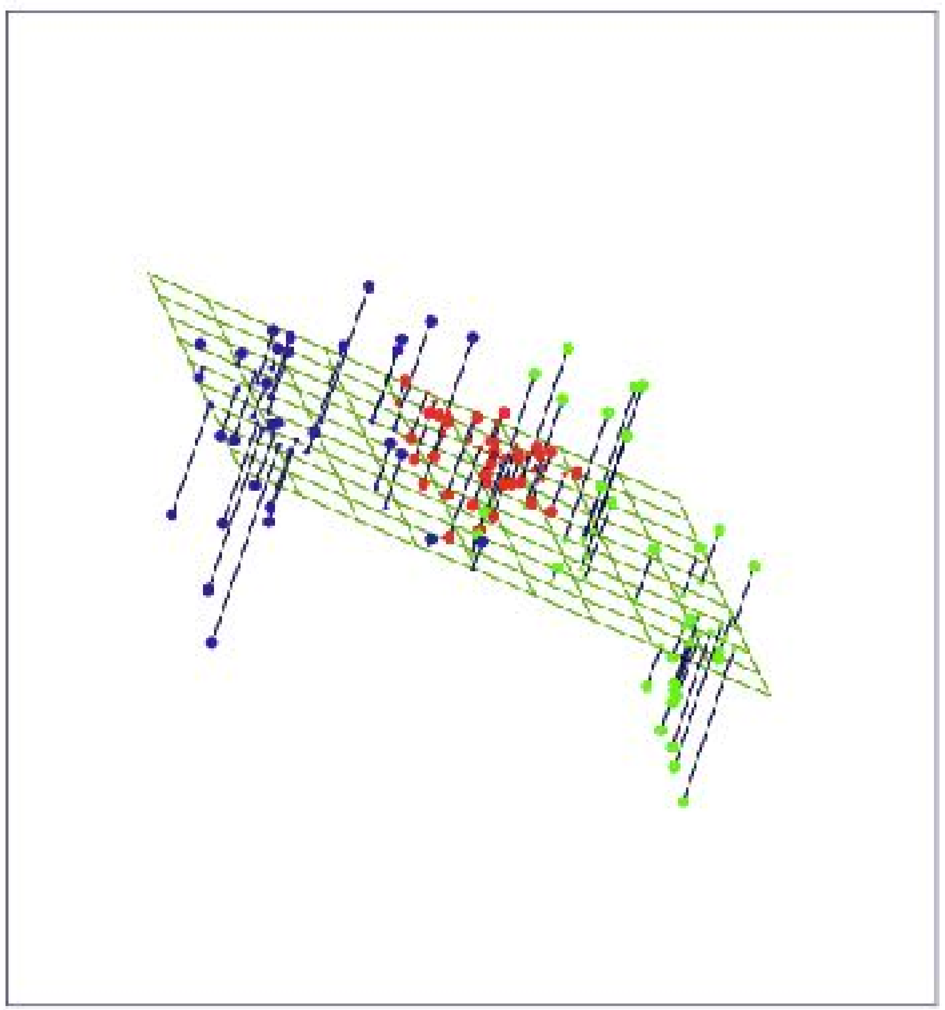
MEG readings

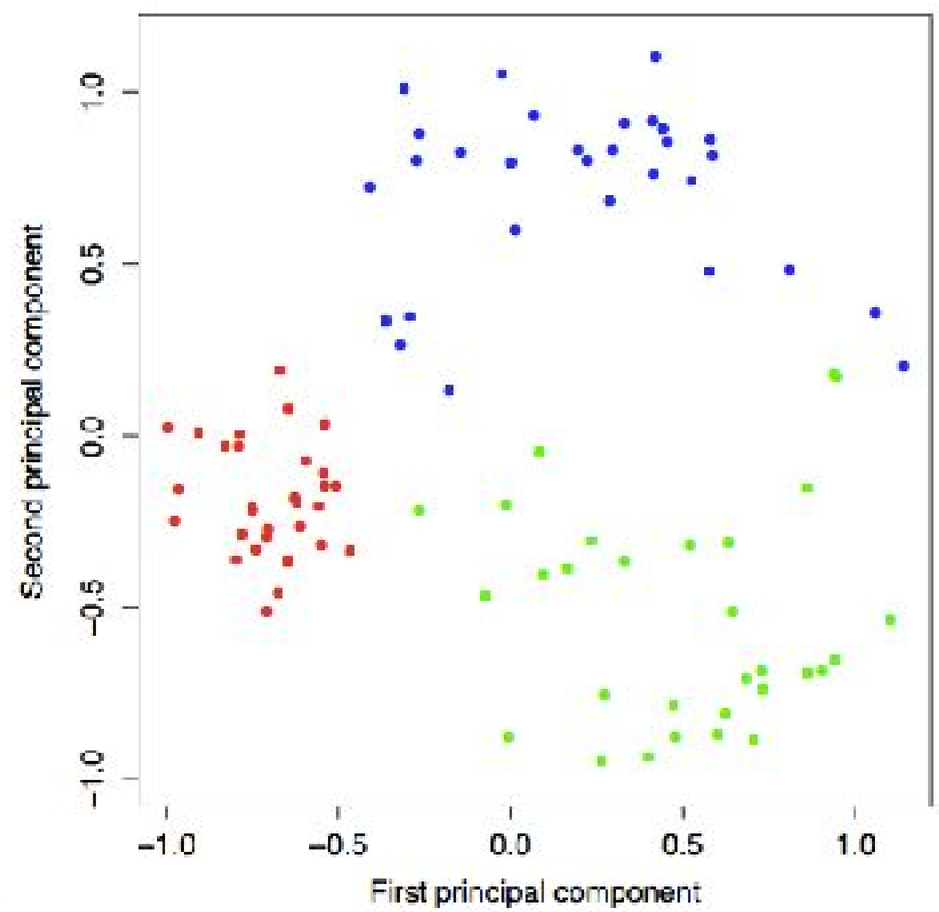
Based on slide from P. Liang

Dimensionality Reduction

Dimensionality Reduction

Project the high-dimensional data onto a lower-dimensional subspace that best “fits” the data



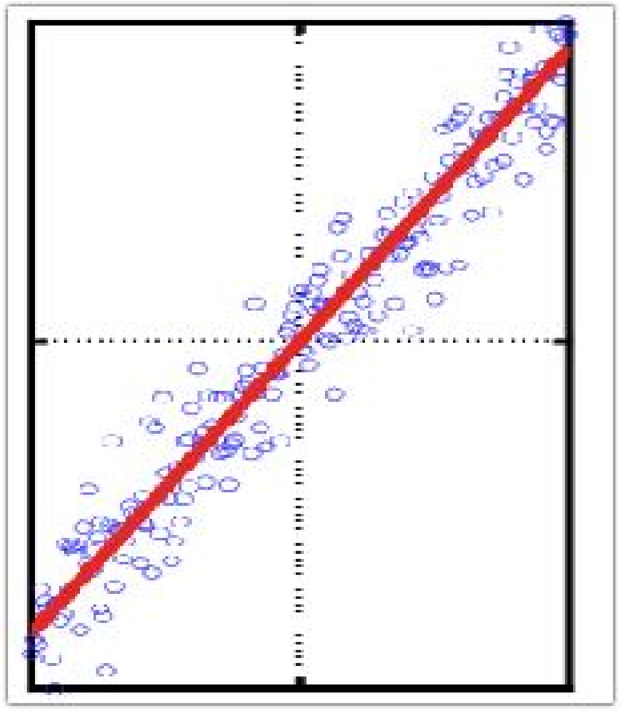


Hastie et al., ELSL

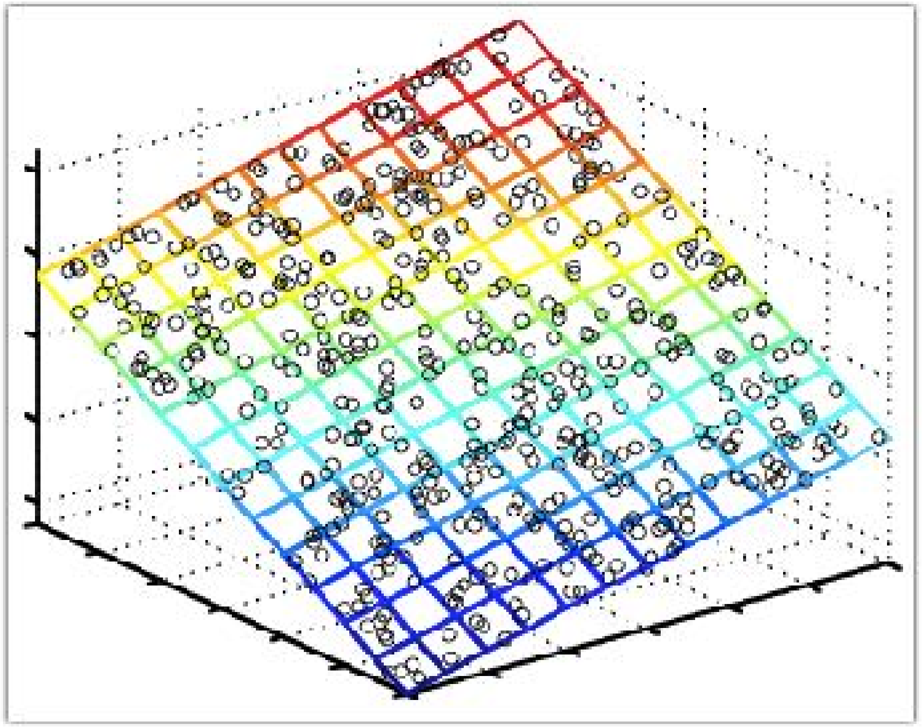
Linear Methods

- Does the data lie mostly in a hyperplane?
- If so, what is its *intrinsic* dimensionality d?

D = 2
d = 1



D = 3
d = 2

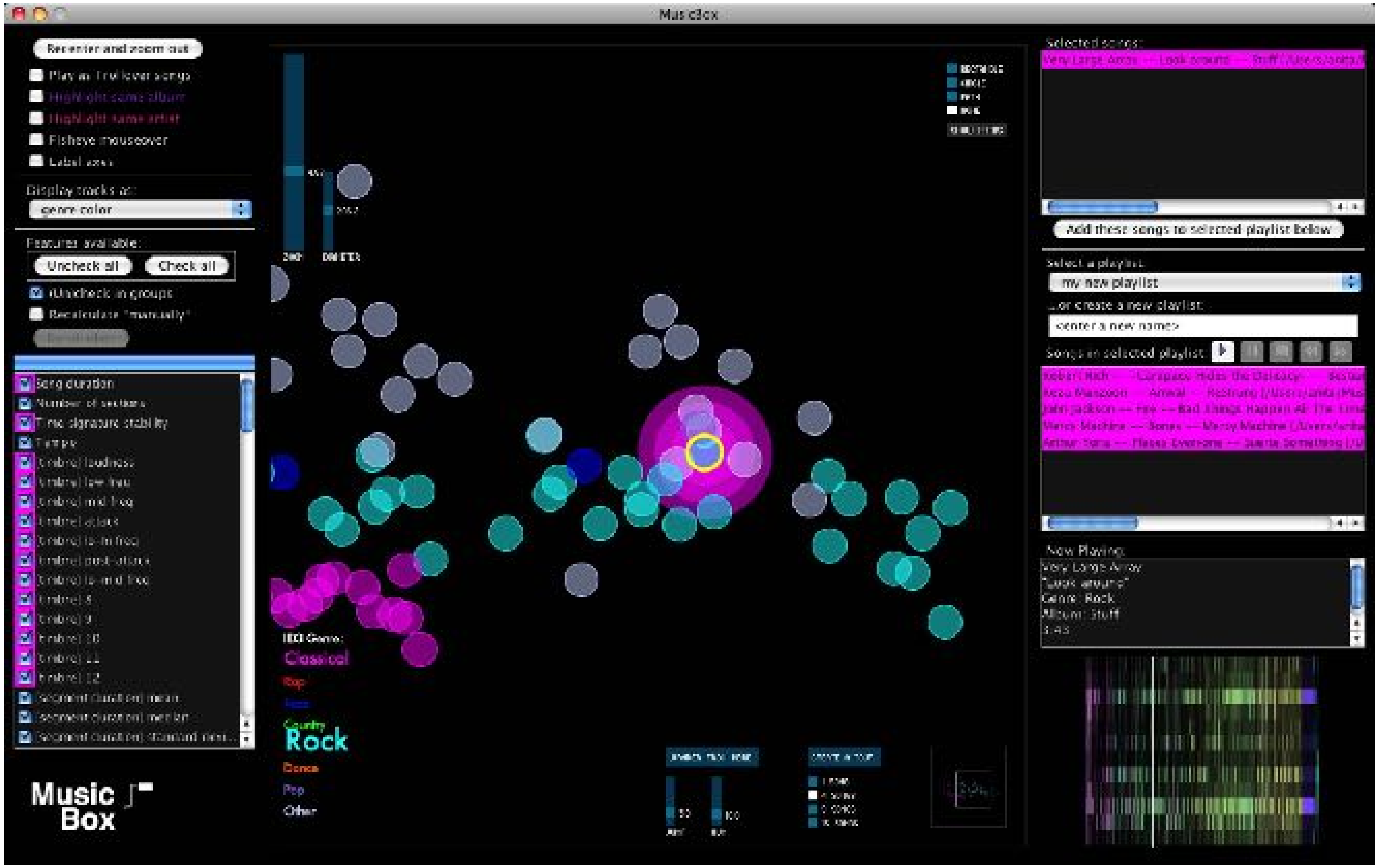


F. Sha

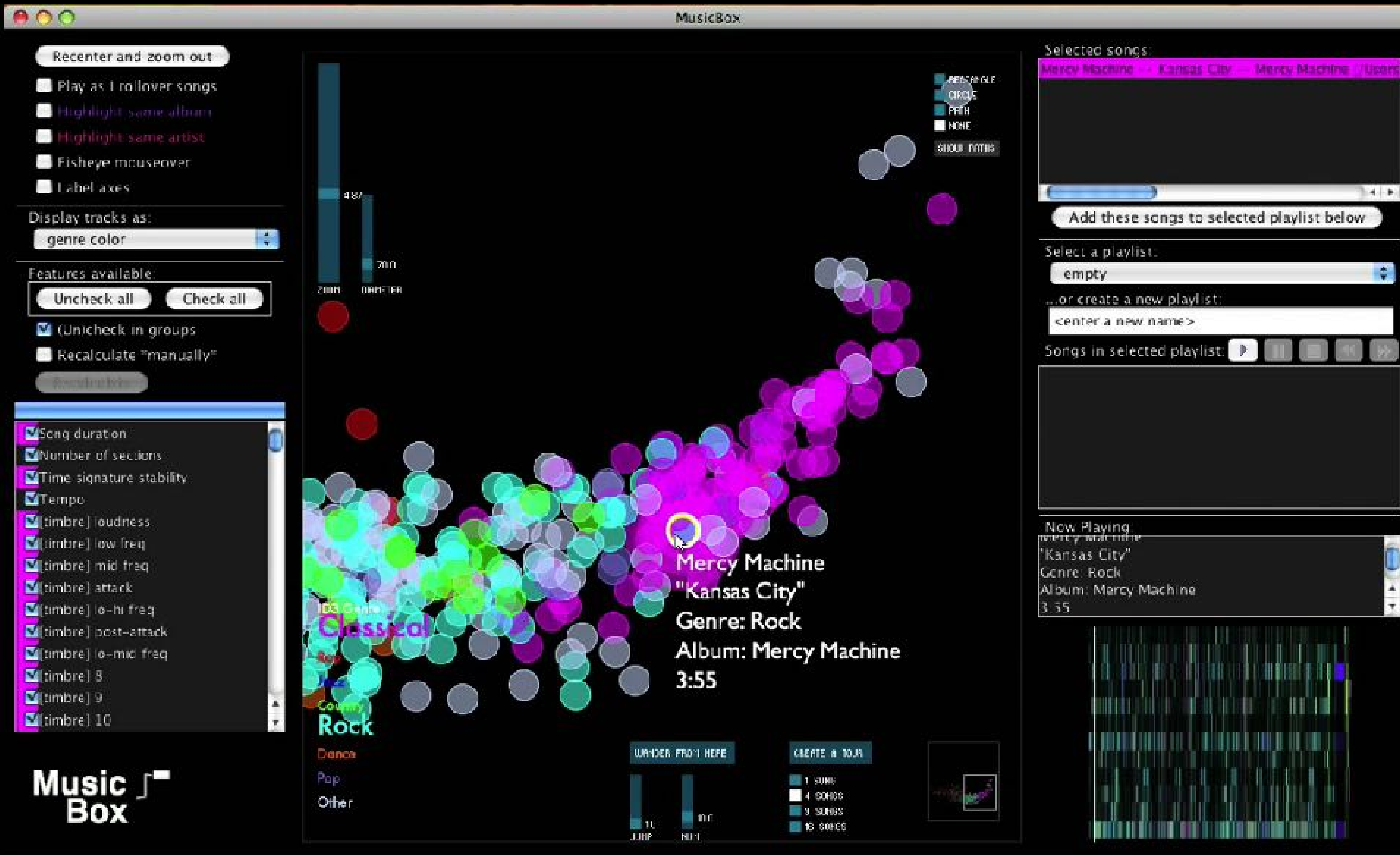
Principal Components Analysis (PCA)

Uses

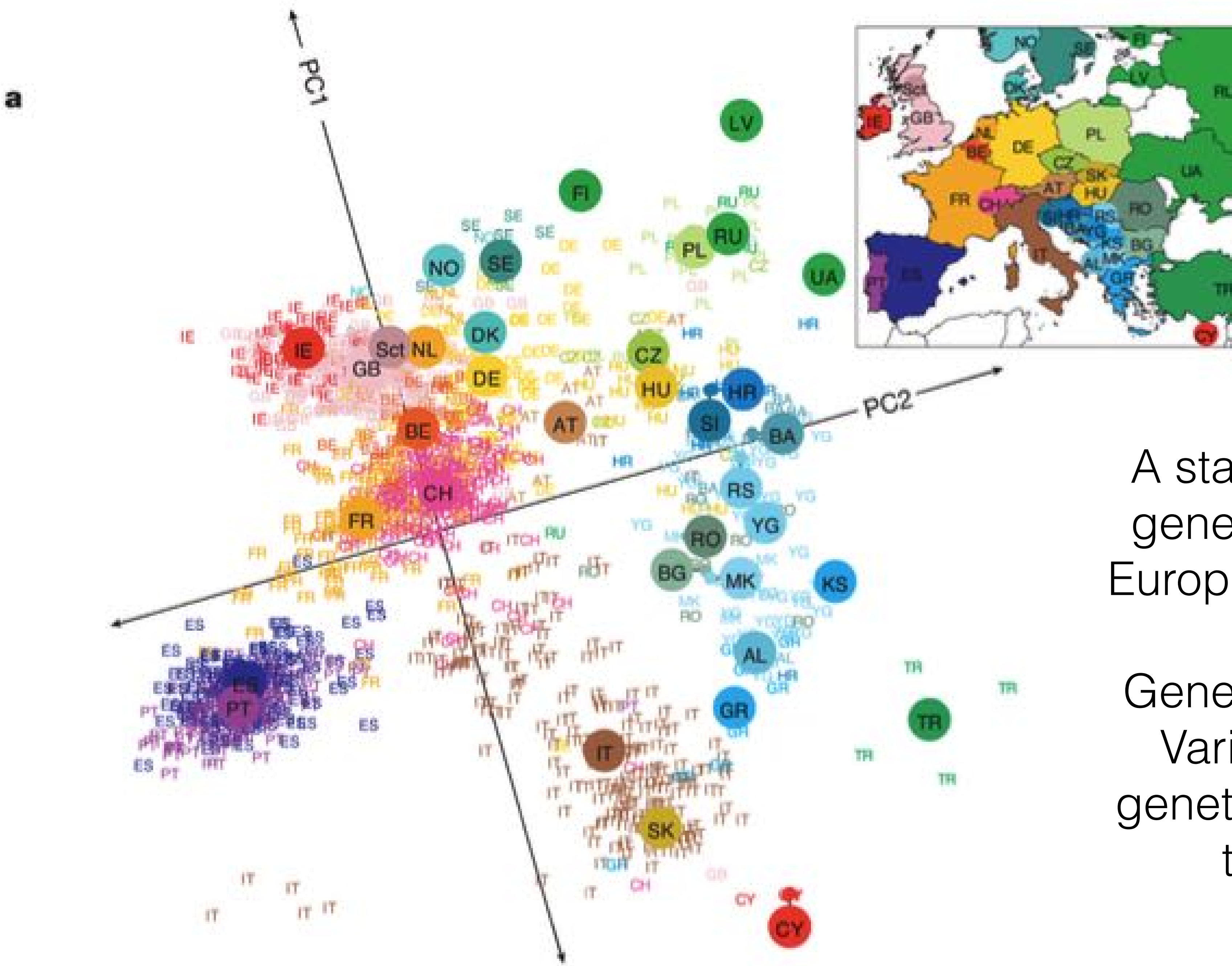
- Dimensionality reduction for supervised learning (e.g., Principle Component Regression)
- Compression
- Visualization



MusicBox, Anita Lillie, MIT


MusicBox
Recenter and zoom out
Play as I rollover songs
Highlight same album
Highlight same artist
Fisheye mouseover
Label axes
Display tracks as:
genre color
Features available:
Uncheck all
Check all
(Un)check in groups
Recalculate "manually"
Song duration
Number of sections
Time signature stability
Tempo
[timbre] loudness
[timbre] low freq
[timbre] mid freq
[timbre] attack
[timbre] lo-hi freq
[timbre] post-attack
[timbre] lo-mid freq
[timbre] 8
[timbre] 9
[timbre] 10
Music Box
Mercy Machine
"Kansas City"
Genre: Rock
Album: Mercy Machine
3:55
Classical
Rock
Dance
Pop
Other
WANDER FROM HERE
CREATE A TOUR
Selected songs:
Add these songs to selected playlist below
Select a playlist:
empty
...or create a new playlist:
<enter a new name>
Songs in selected playlist:
Now Playing:
"Kansas City"
Genre: Rock
Album: Mercy Machine
3:55

Population Ancestry

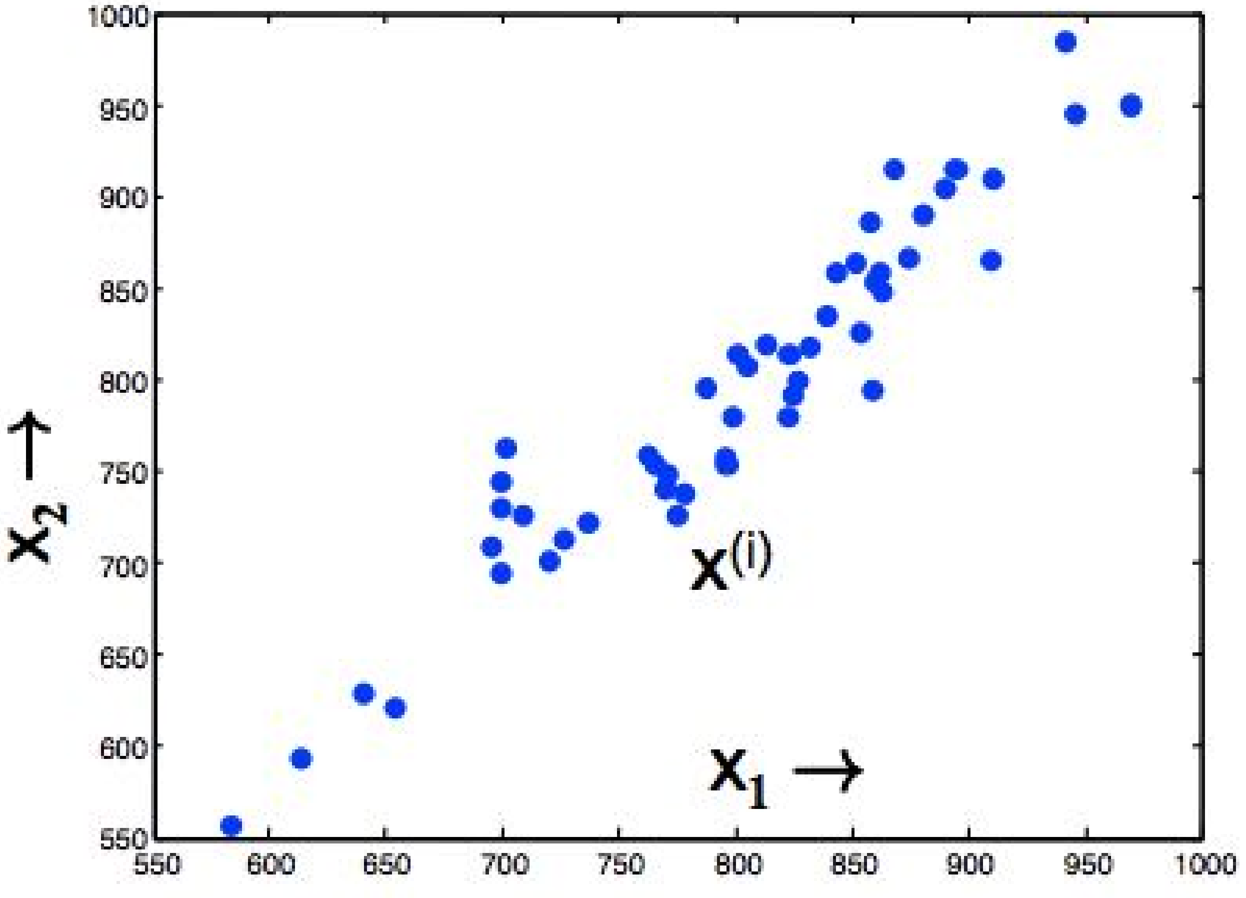


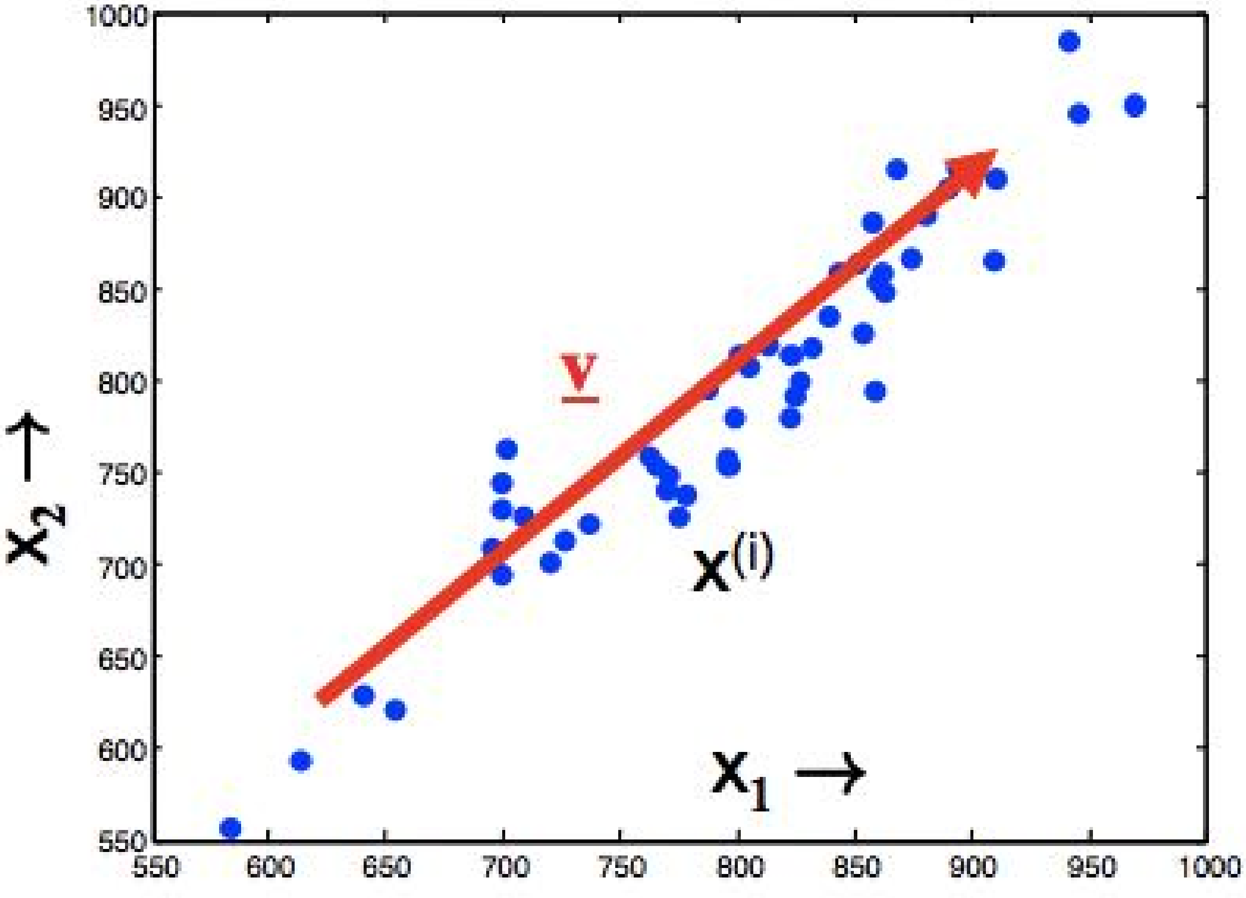
A statistical summary of genetic data from 1,387 Europeans based on PCA

Genes mirror geography
Variation at 100,000s genetic markers reduced to first two PCs

<u>Genes mirror geography within Europe, J. Novembre et al., Nature 2008</u>

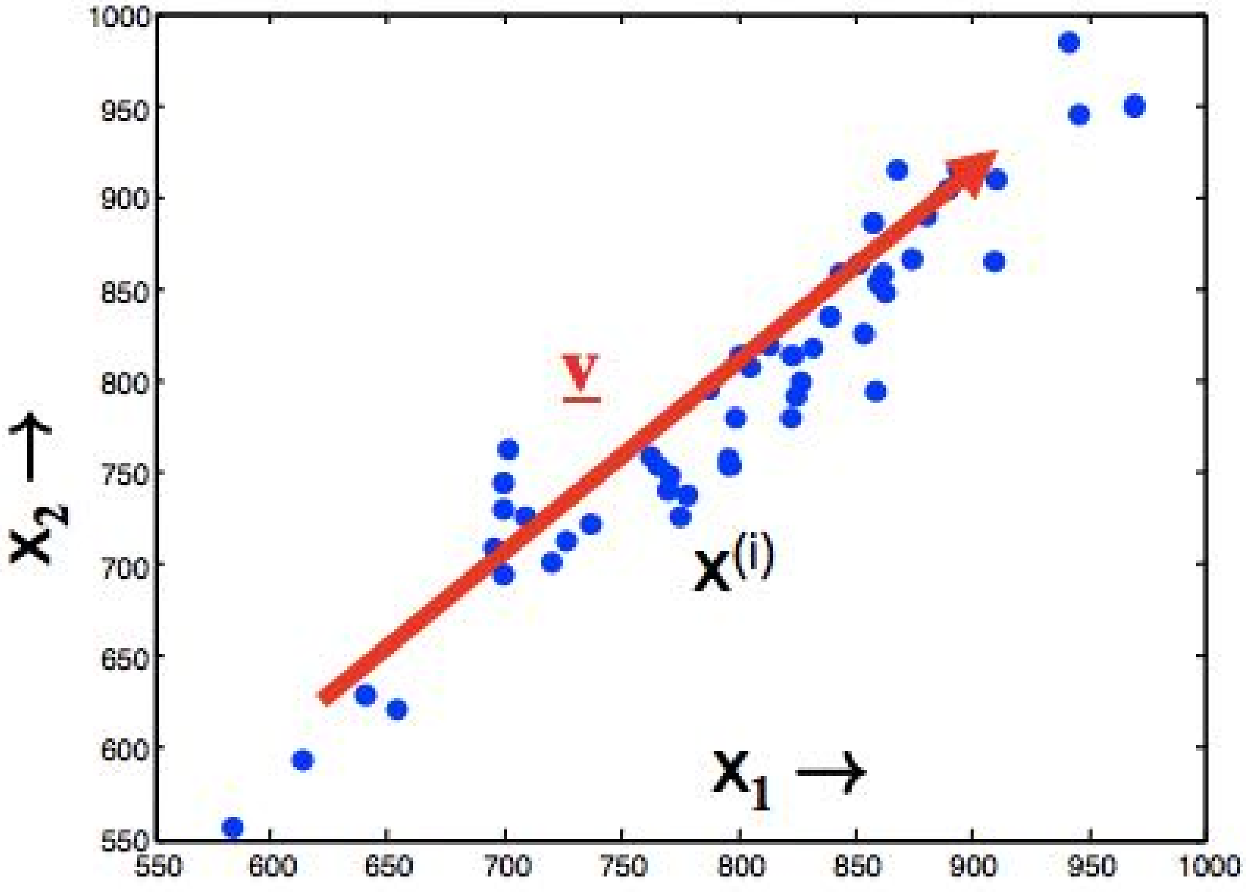
PCA: Main Idea





A. Ihler

PCA: Main Idea

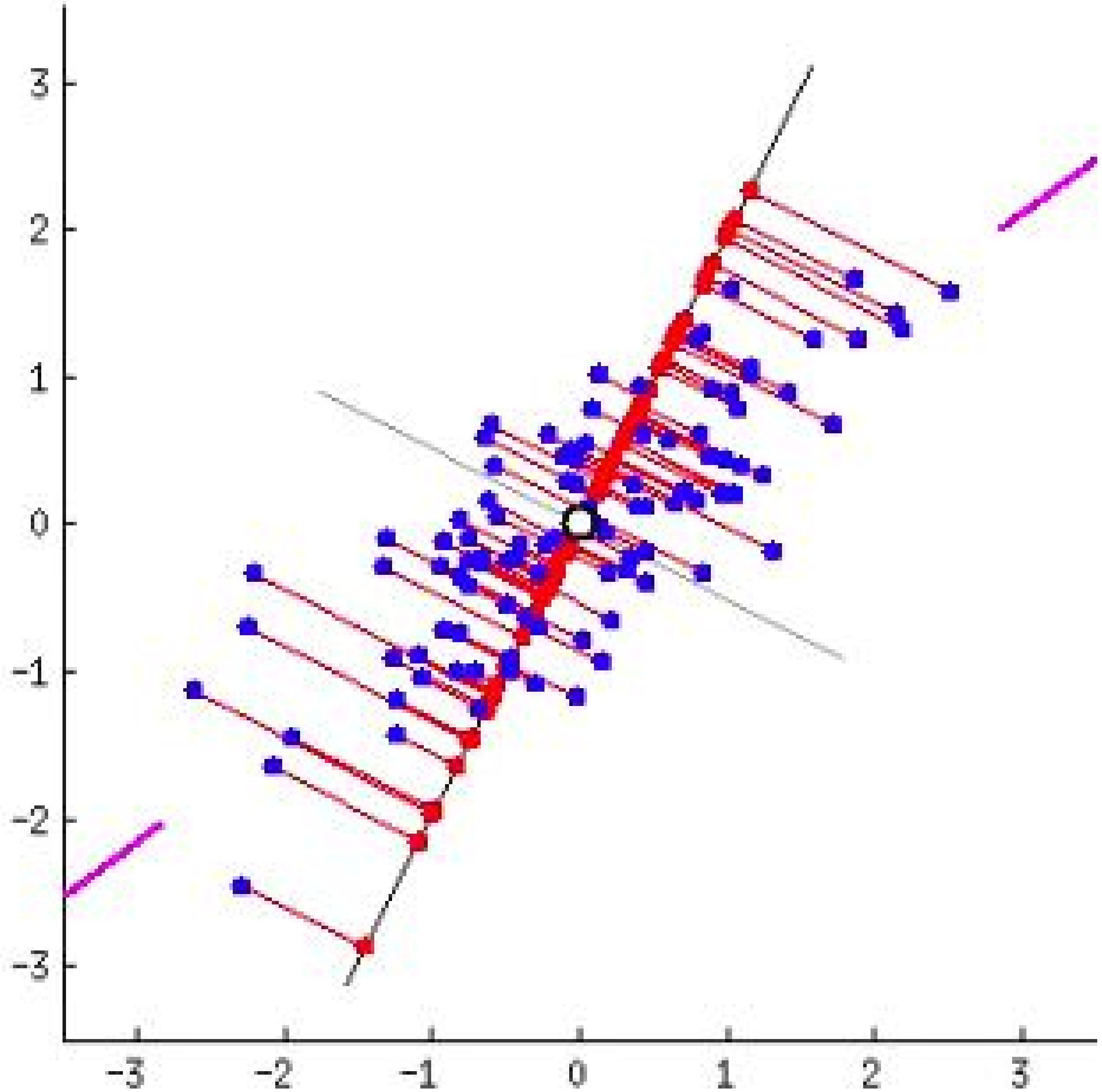


Identify a new coordinate axis **v** through the data that is the (least-squares) line of best fit through the plotted data

Equivalent: **v** is the direction of maximum variance (max. the spread along **v**)

A. Ihler

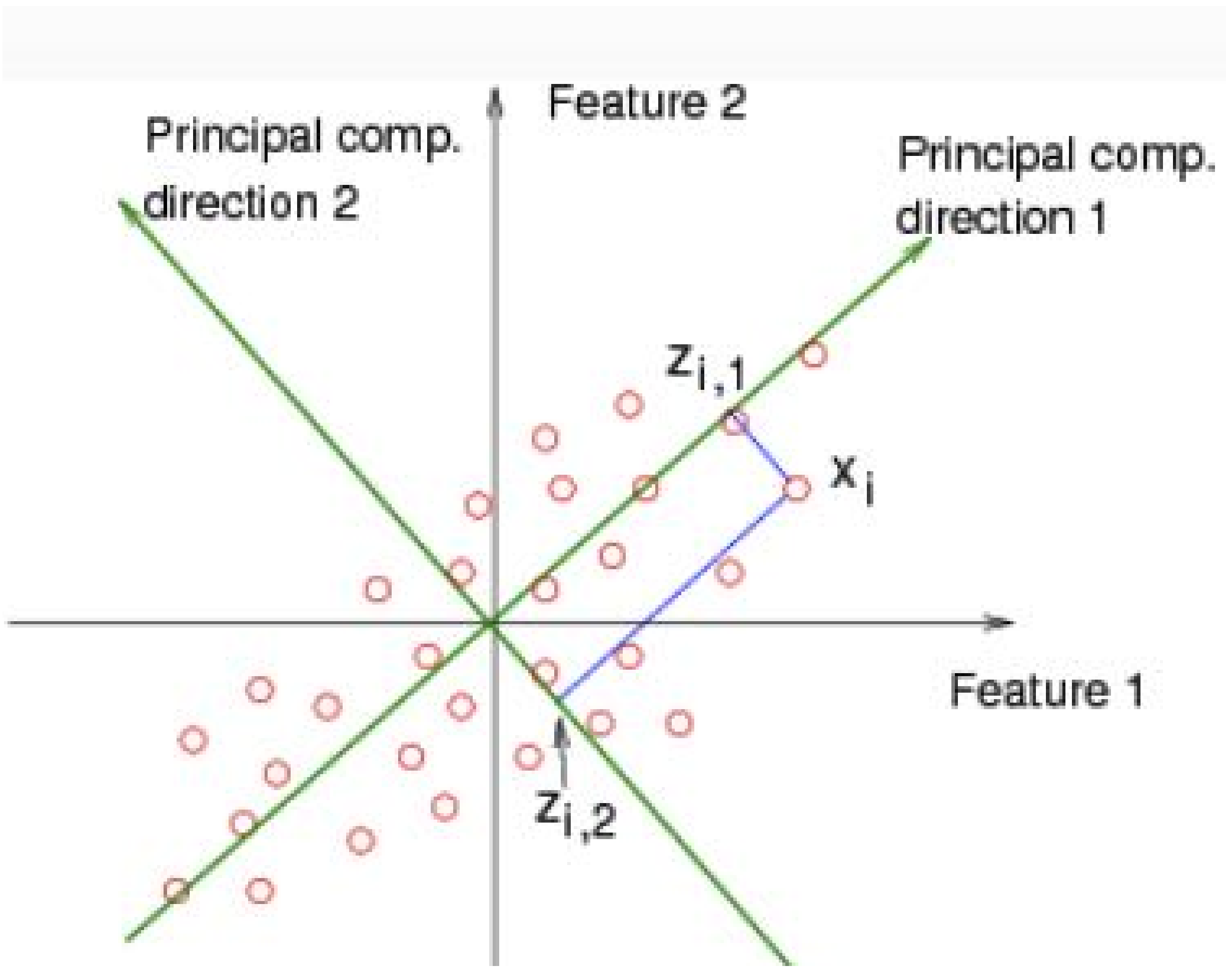
PCA: Main Idea



To maximize the variance, look for the projection with the smallest average (mean-squared) distance between the original vectors and their projections onto the new axis.

http://stats.stackexchange.com/questions/2691/making-sense-of-principal-component-analysis-eigenvectors-eigenvalues

PCA: Main Idea



New axis is called the *first principal component* of the data

Additional PC vectors are orthogonal to each other

Basic PCA Algorithm

- Subtract mean from data (center X)
- (Typically) scale each dimension by its variance
    - Helps to pay less attention to magnitude of dimensions
- Compute covariance matrix S $\mathbf{S} = \frac{1}{N}\mathbf{X}^\intercal\mathbf{X}$
- Compute k largest eigenvectors of S
- These eigenvectors are the k principal components

http://www.dsc.ufcg.edu.br/~hmg/disciplinas/posgraduacao/rn-copin-2014.3/material/SignalProcPCA.pdf

Examples

Eating in the UK (a 17D example)

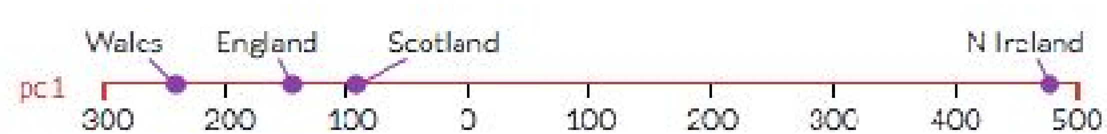
Original example from Mark Richardson's class notes Principal Component Analysis

What if our data have way more than 3-dimensions? Like, **17** dimensions?! In the table is the average consumption of 17 types of food in grams per person per week for every country in the UK.

The table shows some interesting variations across different food types, but overall differences aren't so notable. Let's see if PCA can eliminate dimensions to emphasize how countries differ.

	England	N Ireland	Scotland	Wales
Alcoholic drinks	375	135	458	475
Beverages	57	47	53	73
Carcase meat	245	267	242	227
Cereals	1472	1494	1462	1582
Cheese	105	66	103	103
Confectionery	54	41	62	64
Fats and oils	193	209	184	235
Fish	147	93	122	160
Fresh fruit	1102	674	957	1137
Fresh potatoes	720	1033	566	874
Fresh Veg	253	143	171	265
Other meat	685	586	750	803
Other Veg	488	355	418	570
Processed potatoes	198	187	220	203
Processed Veg	360	334	337	365
Soft drinks	1374	1506	1572	1256
Sugars	156	139	147	175

Here's the plot of the data along the first principal component. Already we can see something is different about Northern Ireland.



http://setosa.io/ev/principal-component-analysis/

Example in R

```
> # Load data
> data(iris)
> head(iris, 3)
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
2          4.9         3.0          1.4         0.2  setosa
3          4.7         3.2          1.3         0.2  setosa
> # log transform
> log.ir <- log(iris[, 1:4])
> ir.species <- iris[, 5]
>
> # apply PCA - scale. = TRUE is highly
> # advisable, but default is FALSE.
> ir.pca <- prcomp(log.ir,
+                  center = TRUE,
+                  scale. = TRUE)
```

Recommended for this data

Subtract mean and scale to unit variance

https://tgmstat.wordpress.com/2013/11/28/computing-and-visualizing-pca-in-r/

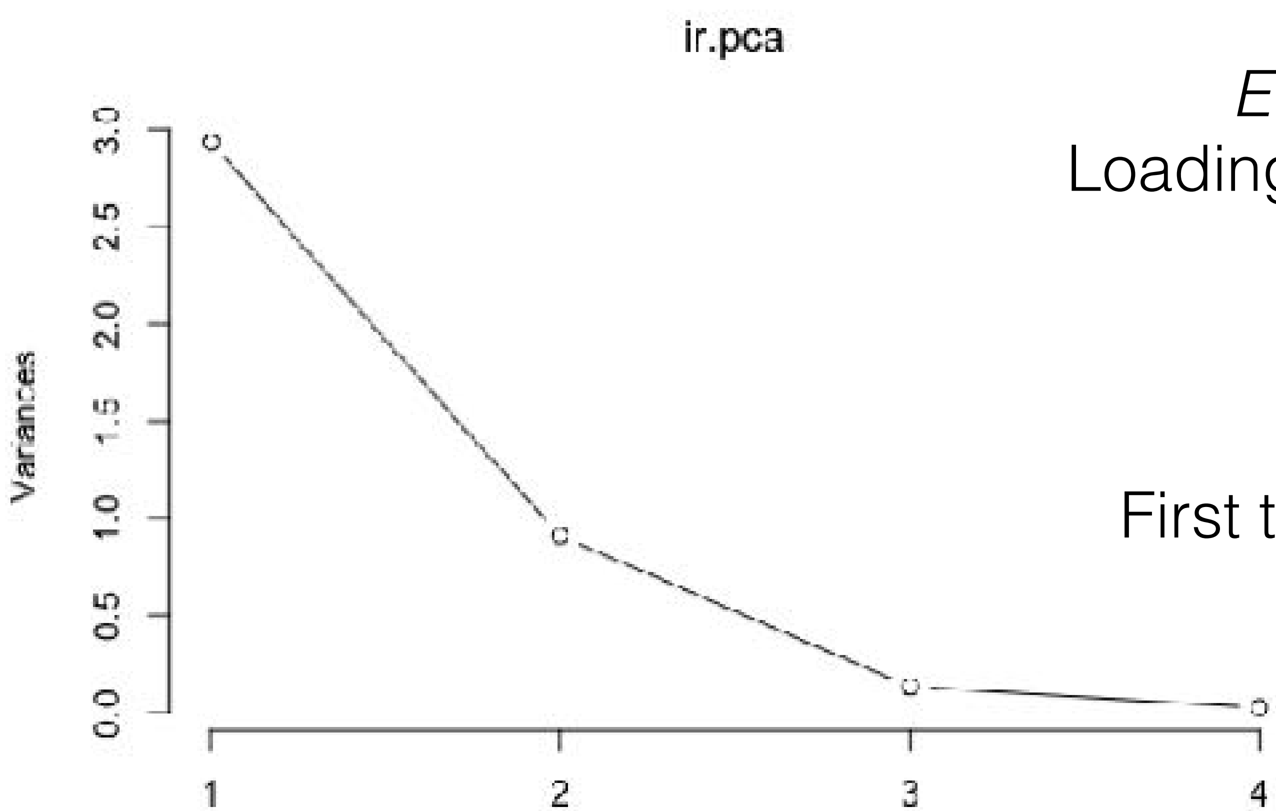
Example in R

```
> # print method
> print(ir.pca)
Standard deviations:
[1] 1.7124583 0.9523797 0.3647029 0.1656840

Rotation:
                    PC1         PC2        PC3         PC4
Sepal.Length  0.5038236 -0.45499872  0.7088547  0.19147575
Sepal.Width  -0.3023682 -0.88914419 -0.3311628 -0.09125405
Petal.Length  0.5767881 -0.03378802 -0.2192793 -0.78618732
Petal.Width   0.5674952 -0.03545628 -0.5829003  0.58044745
> # plot method
> plot(ir.pca, type = "l")
```

Rotation of feature to PC space

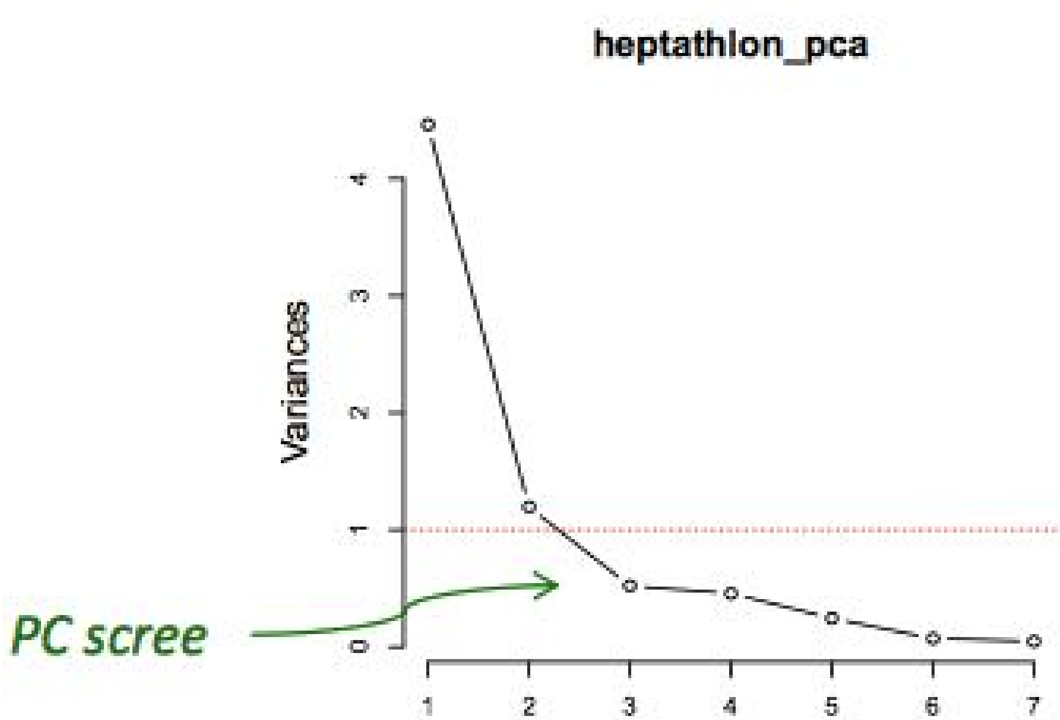
PC4 *loading vector*
Entries are known as *loadings*
Eigenvector is unit vector in same direction
Loading vector = eigenvector * sqrt(eigenvalues)



Plot of variances, aka *screeplot*
First two PCs explain most variance in the data

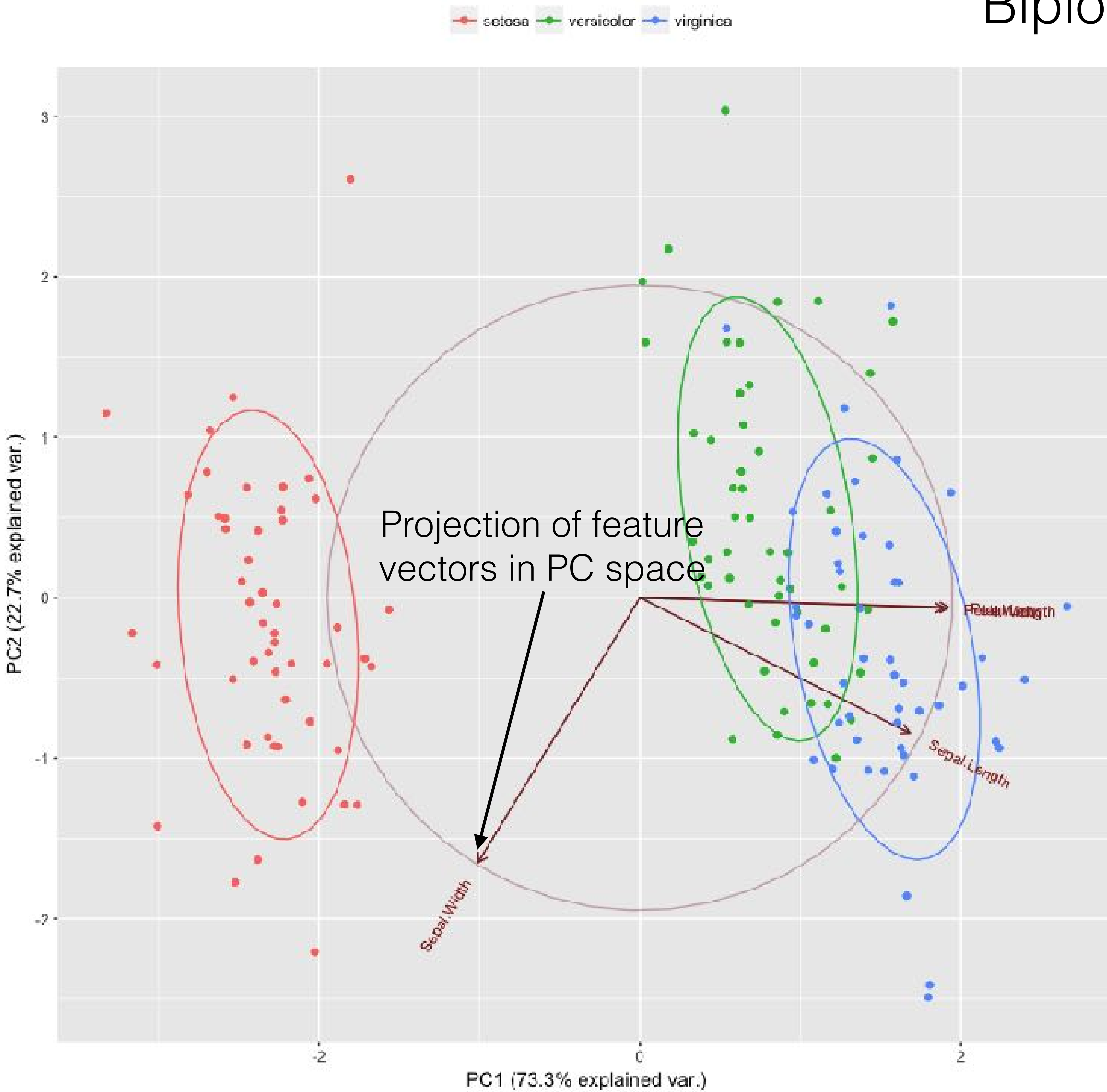
Scree

- *Scree* is a collection of broken rock fragments at the base of crags, mountain cliffs, volcanoes or valley shoulders that has accumulated through periodic rockfall from adjacent cliff faces. [wikipedia]



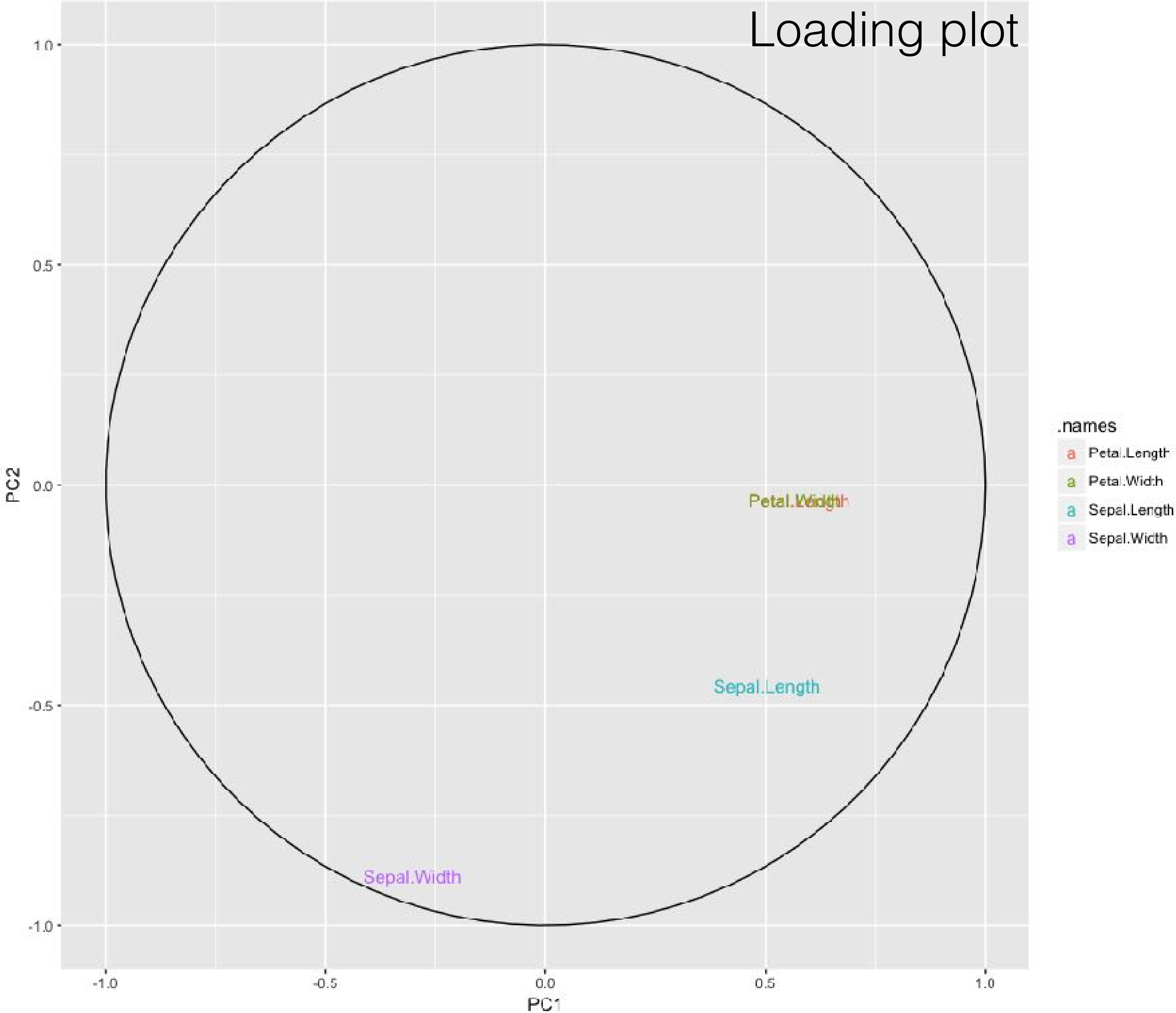
K. B. Hanscombe

Biplot

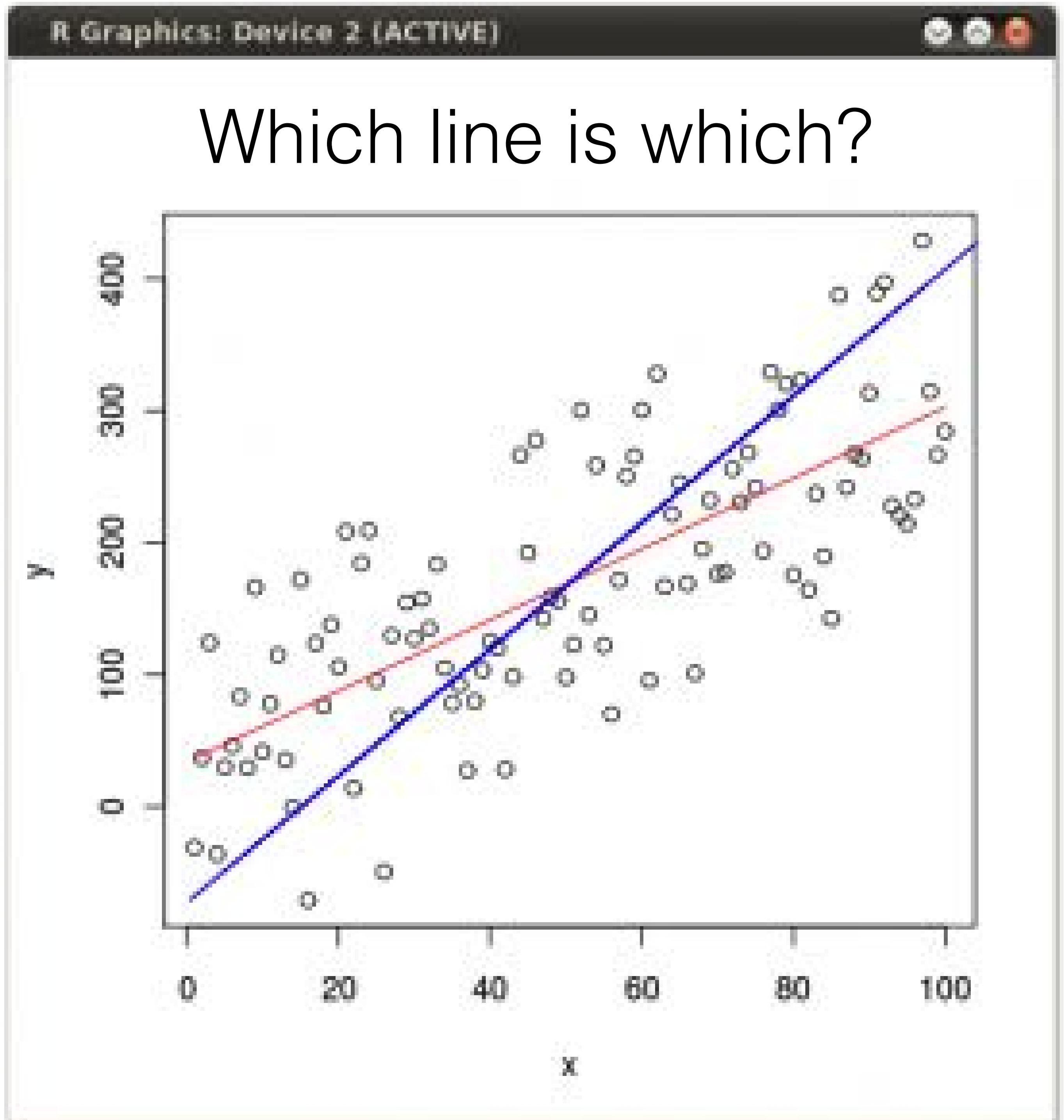


setosa
versicolor
virginica
Projection of feature vectors in PC space
Sepal.Length
Sepal.Width
PC1 (73.3% explained var.)
PC2 (22.7% explained var.)
-2
0
2
-2
-1
0
1
2
3

Loading plot



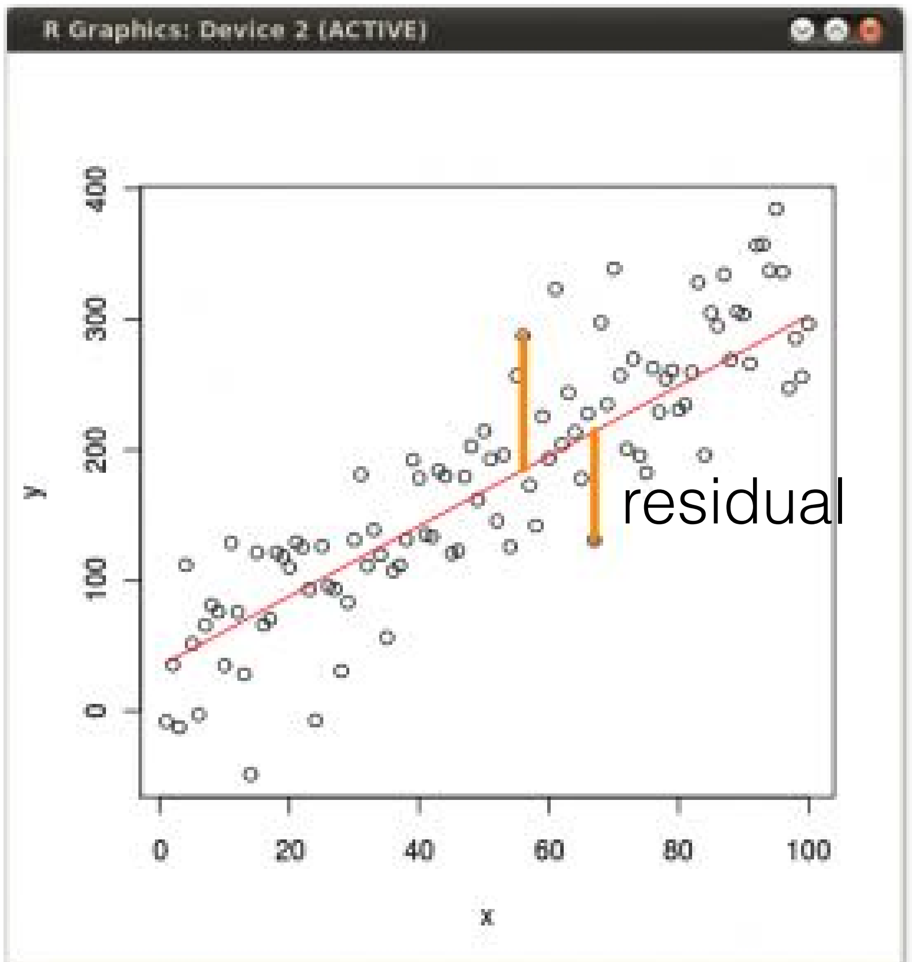
Linear Regression vs. PCA



http://www.cerebralmastication.com/2010/09/principal-component-analysis-pca-vs-ordinary-least-squares-ols-a-visual-explination/

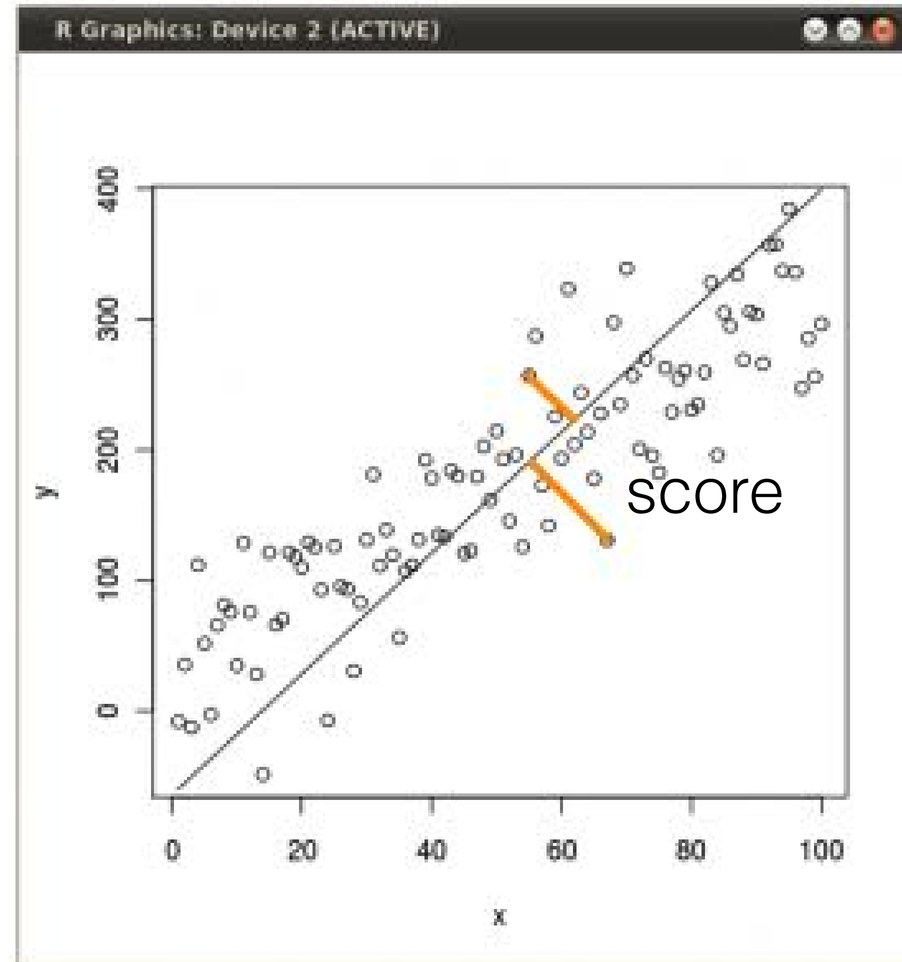
Linear Regression vs. PCA

Linear Regression



Projection along dimensions of X

PCA



Projection along dimensions of PCs

http://www.cerebralmastication.com/2010/09/principal-component-analysis-pca-vs-ordinary-least-squares-ols-a-visual-explination/

PCA in Vision

PCA for Face Images

T. Yang

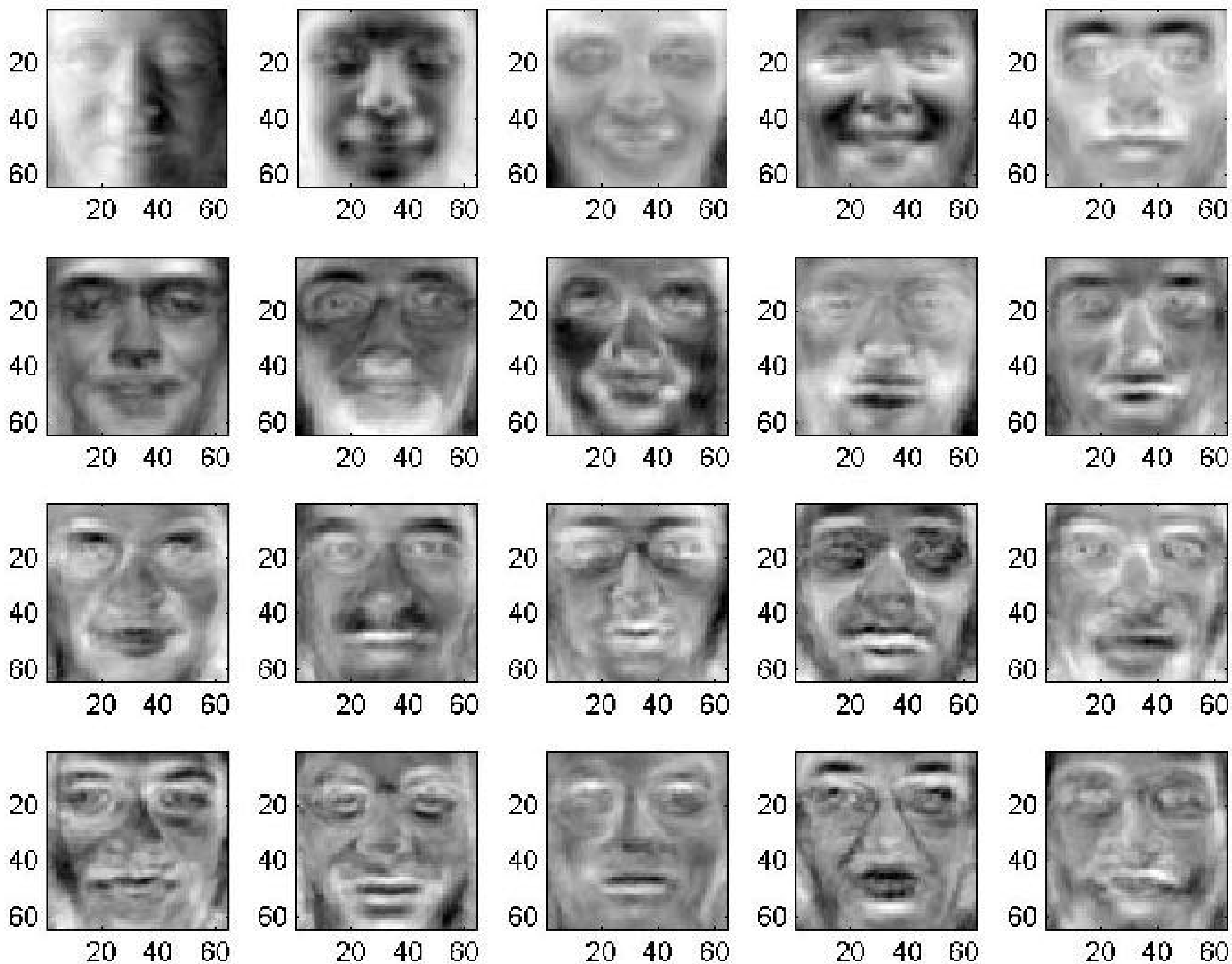
Suppose we have 136 face images, each with 64x64 pixels. What is the dimensionality of this data?

PCA for Face Images

- 64x64 images of faces = 4096 dimensional data

A. Ihler

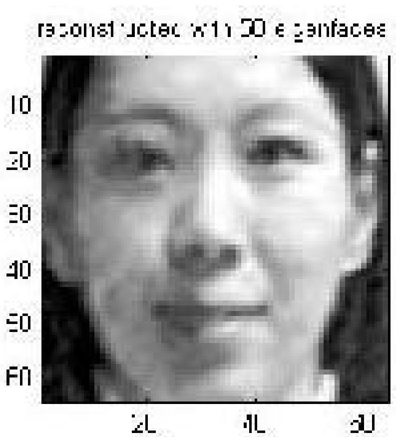
“Eigenfaces”



A. Ihler

“Eigenfaces”

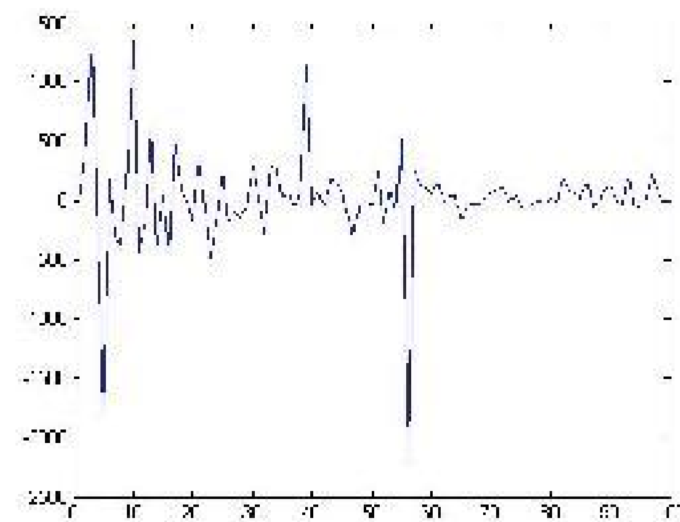
We can reconstruct each face as a linear combination of “basis” PC vectors, or eigenfaces [M. Turk and A. Pentland (1991)]



=

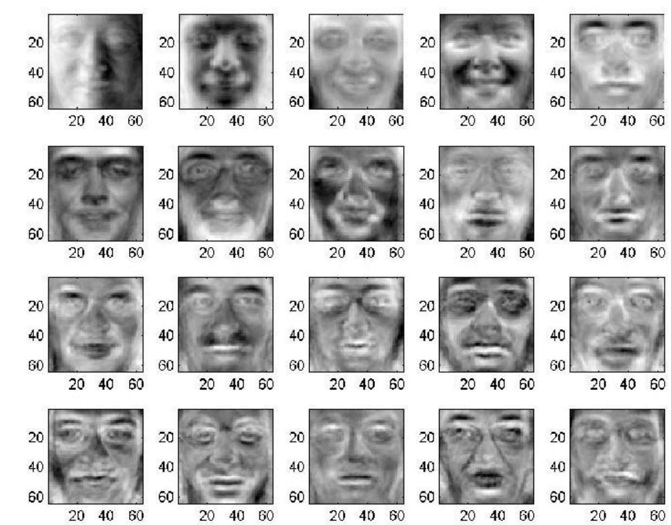
Average face

$+\sum$



Eigenvalues

x

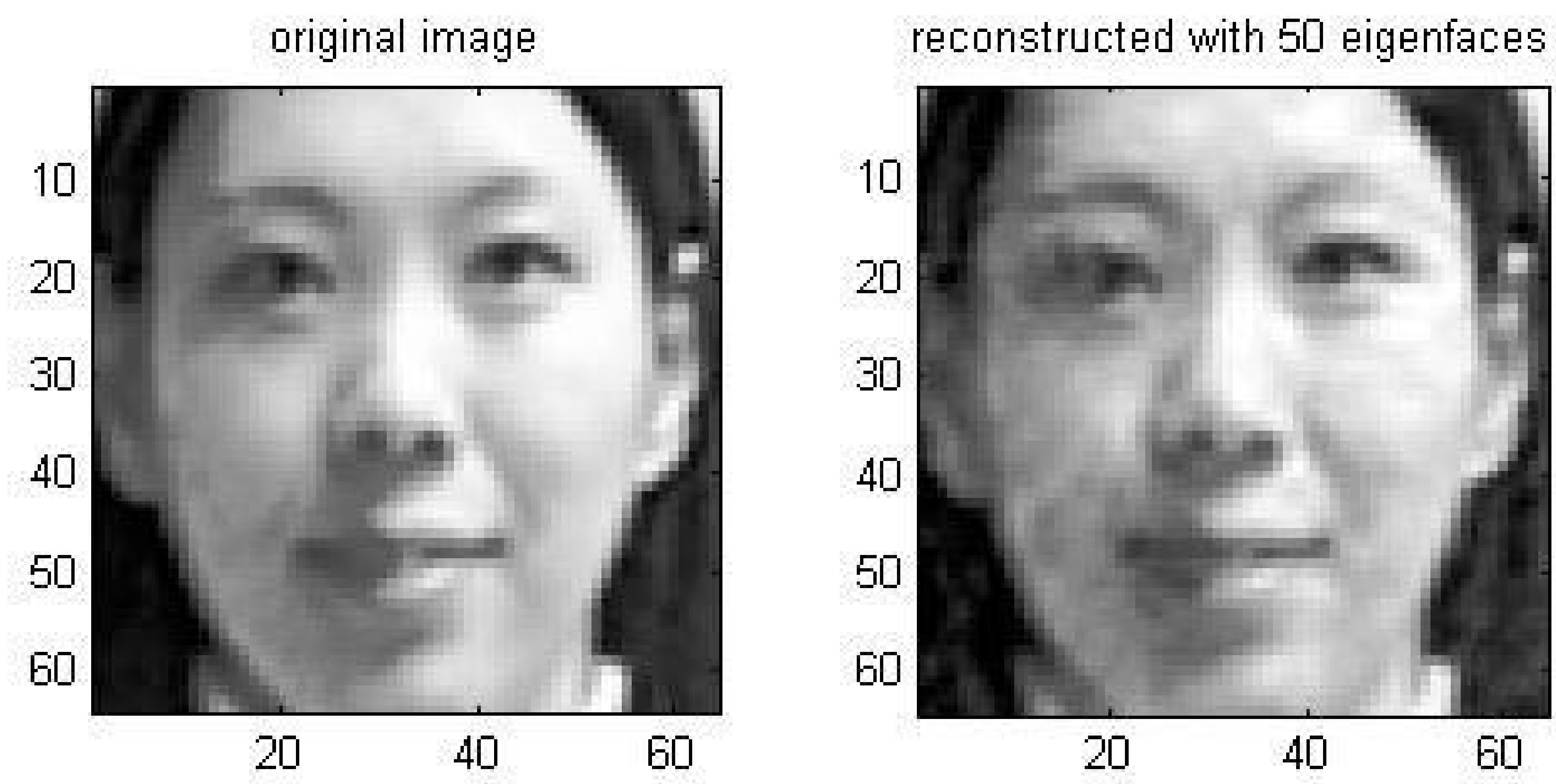


Eigenvectors (“Eigenfaces”)

A. Ihler

Reconstruction

90% variance is captured by the first 50 eigenvectors



PCA for Handwritten Digits

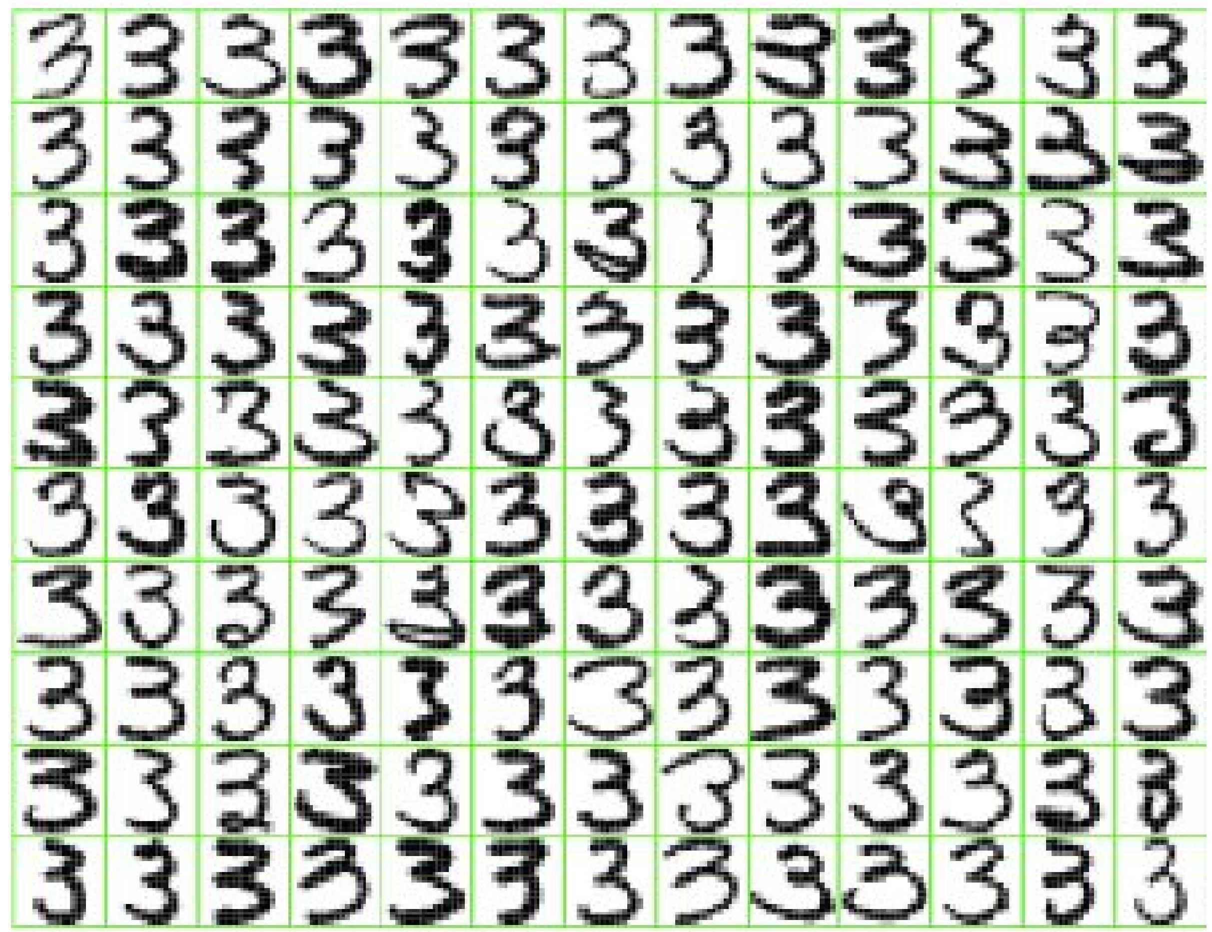
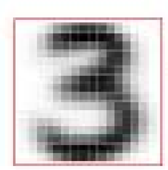
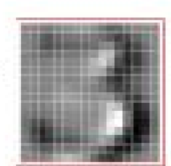
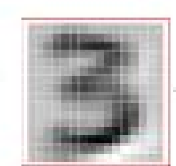


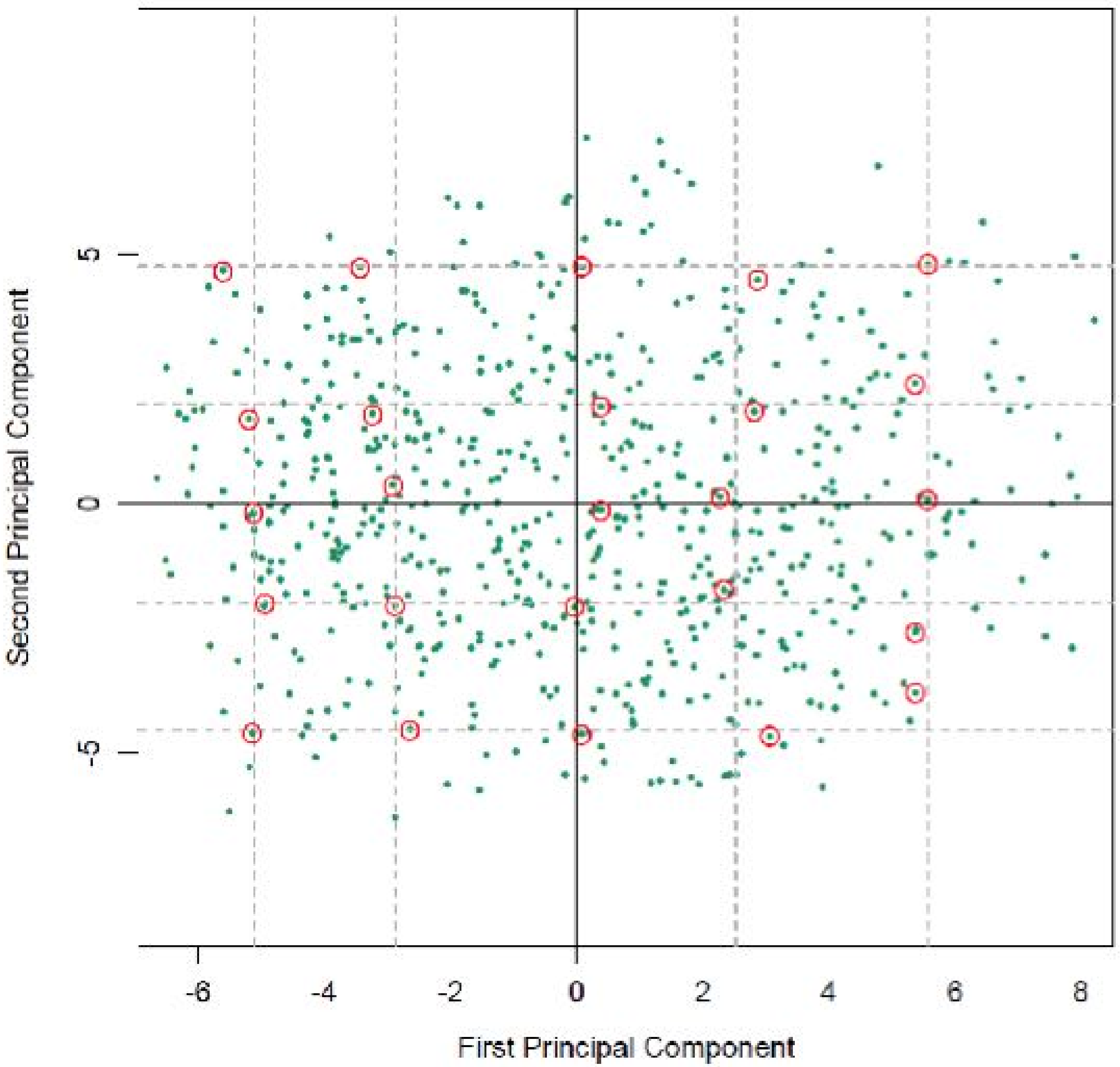
FIGURE 14.22. *A sample of 130 handwritten 3's shows a variety of writing styles.*

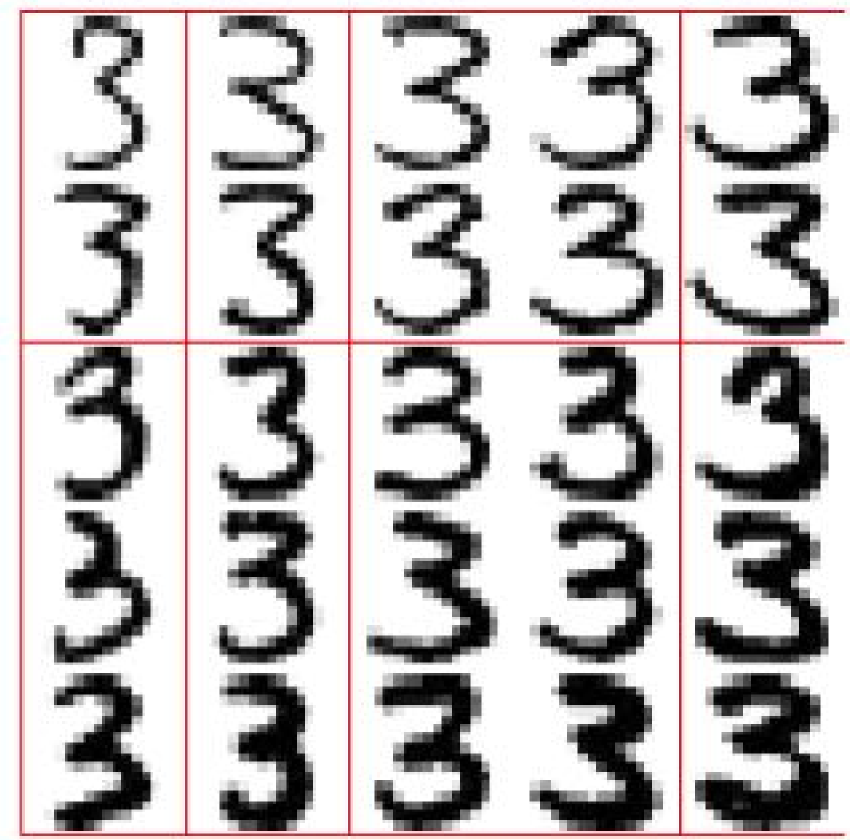
$$\hat{f}(\lambda) = \bar{x} + \lambda_1 v_1 + \lambda_2 v_2$$

$=$  $+ \lambda_1 \cdot$  $+ \lambda_2 \cdot$  .

Hastie et al., ELSL

PCA for Handwritten Digits



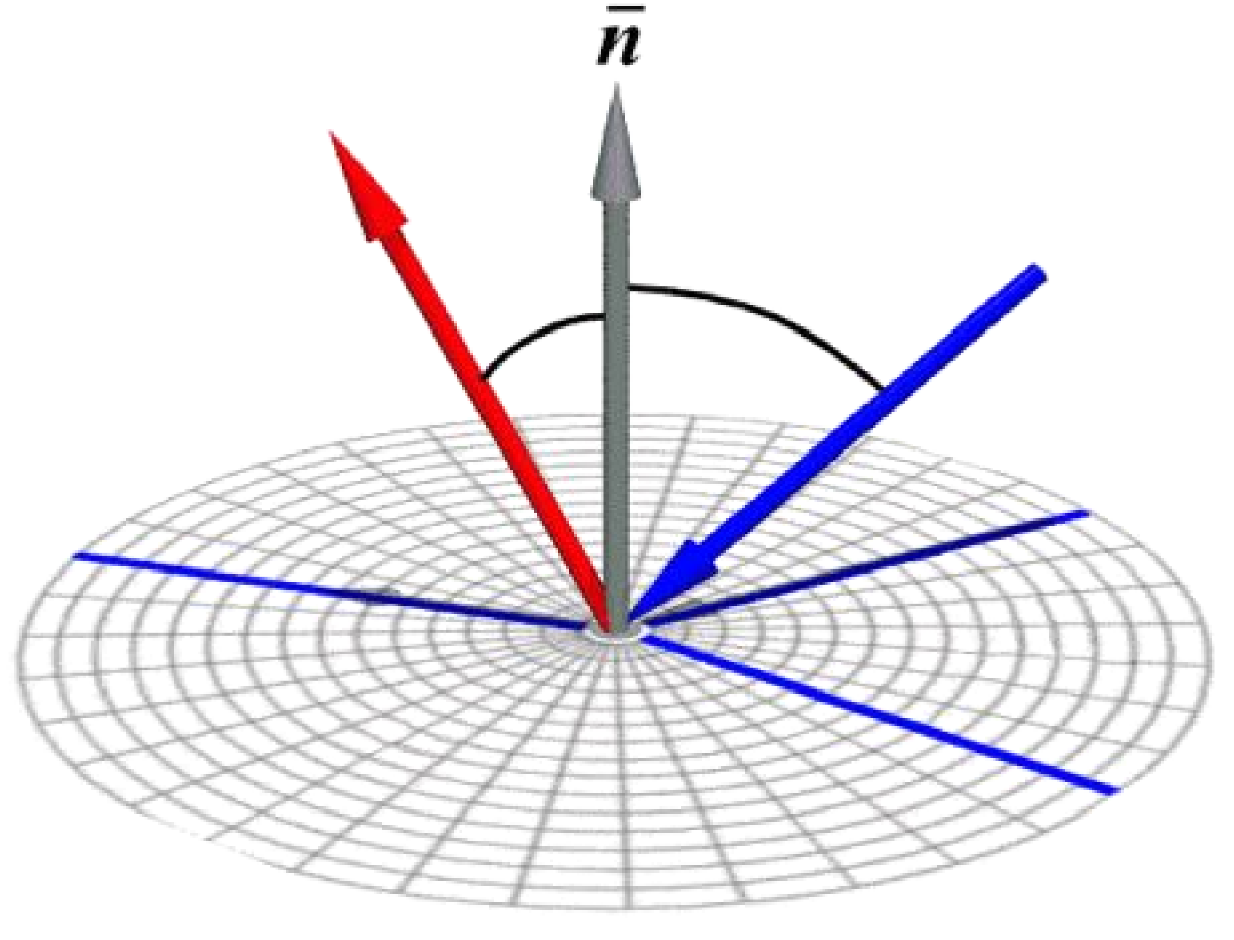


Hastie et al., ELSL

PCA in Graphics

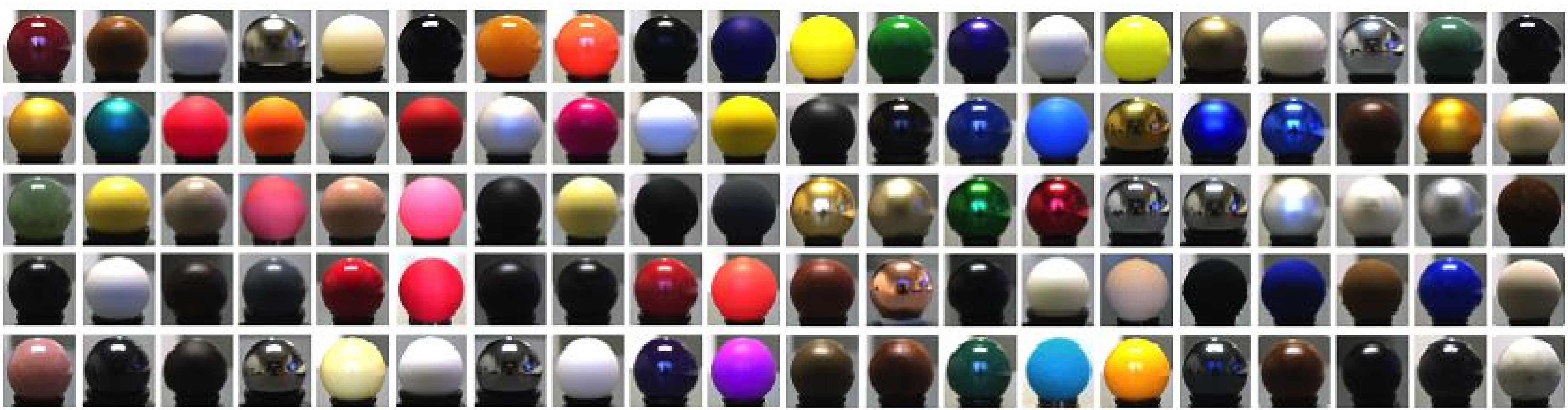
Data-Driven BRDFs

- Bi-Directional Reflectance Distribution Functions



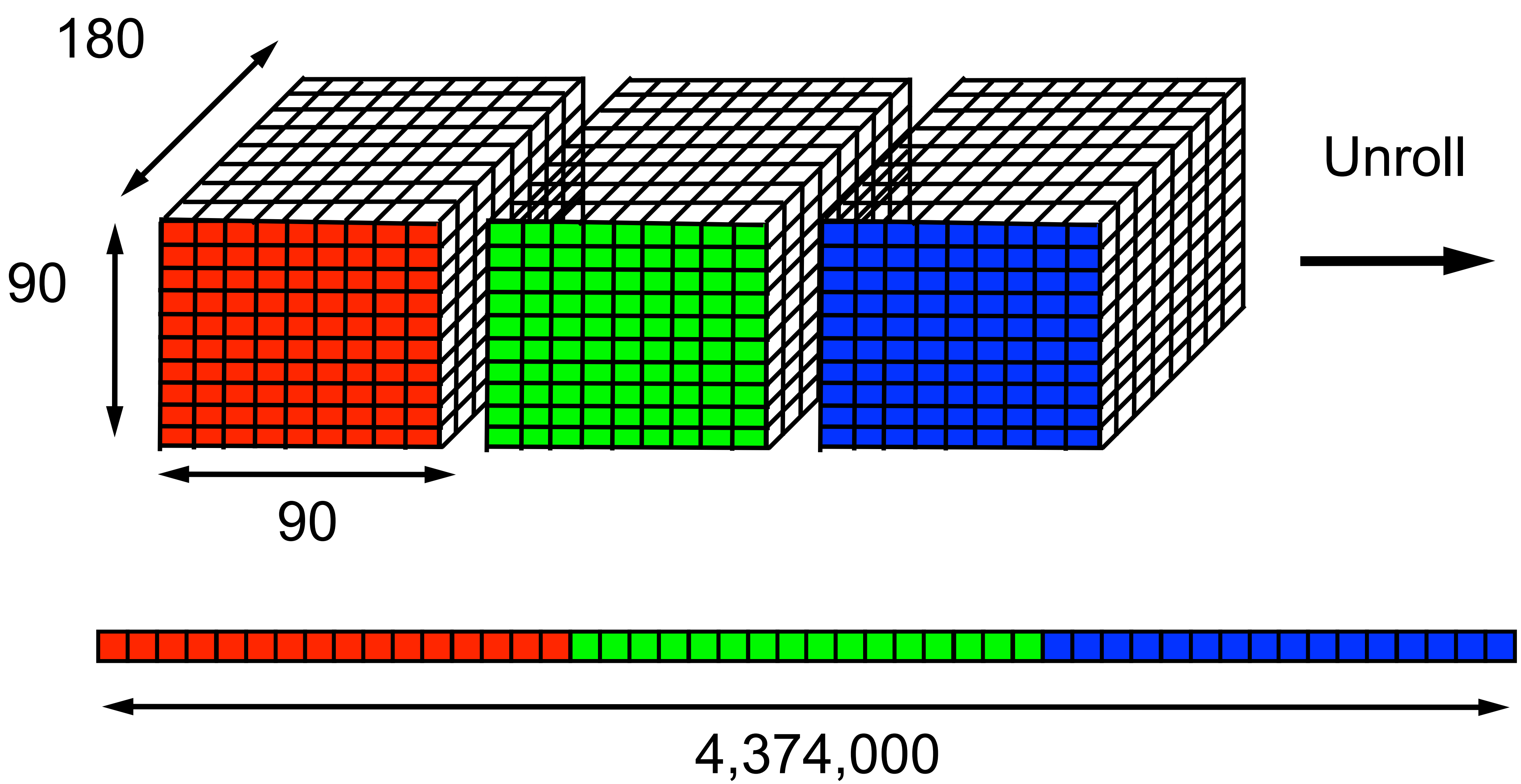
Data-Driven BRDFs

- Measure light reflected off a sphere
- 20-80 million measurements (6000 images) per material

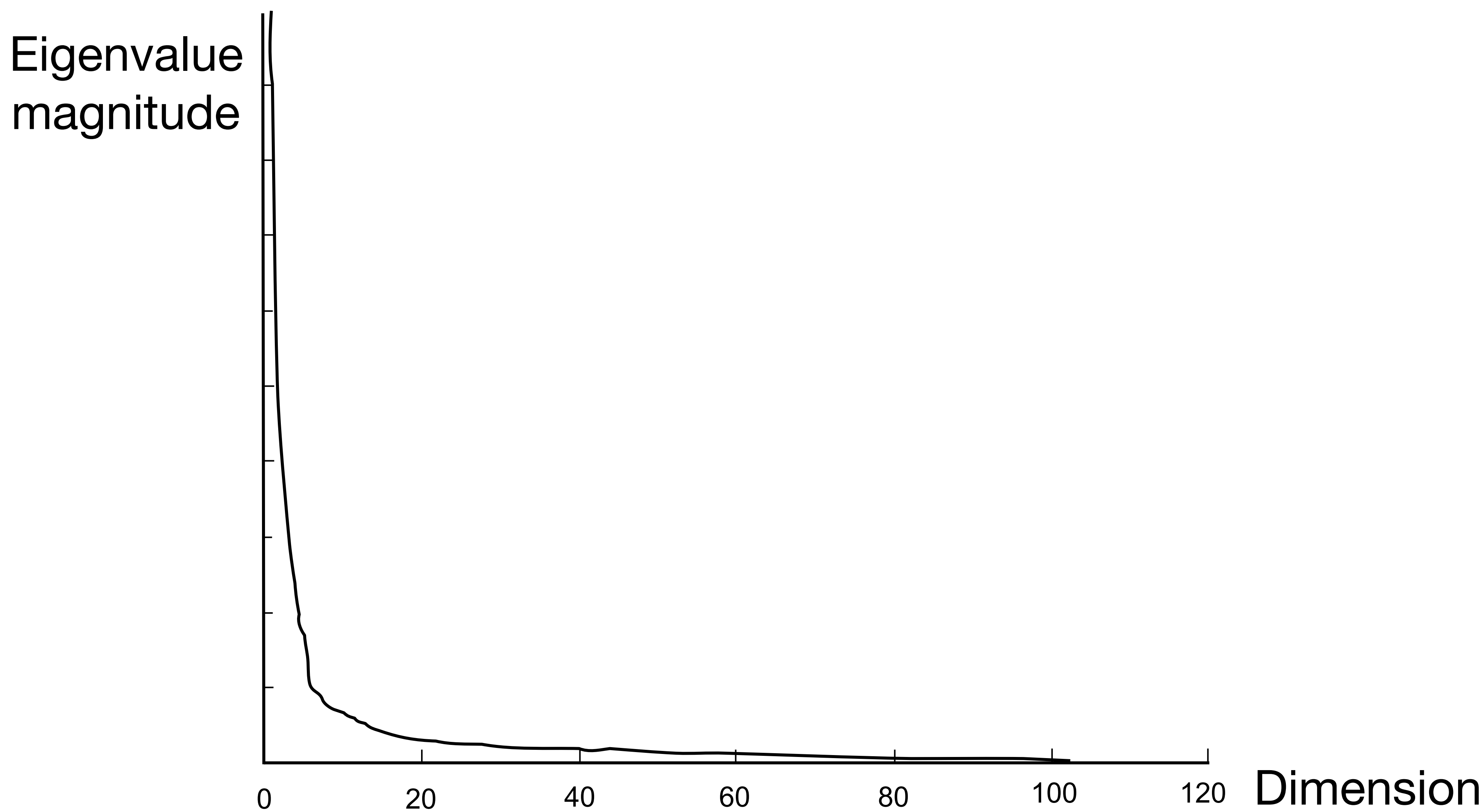


Data-Driven BRDFs

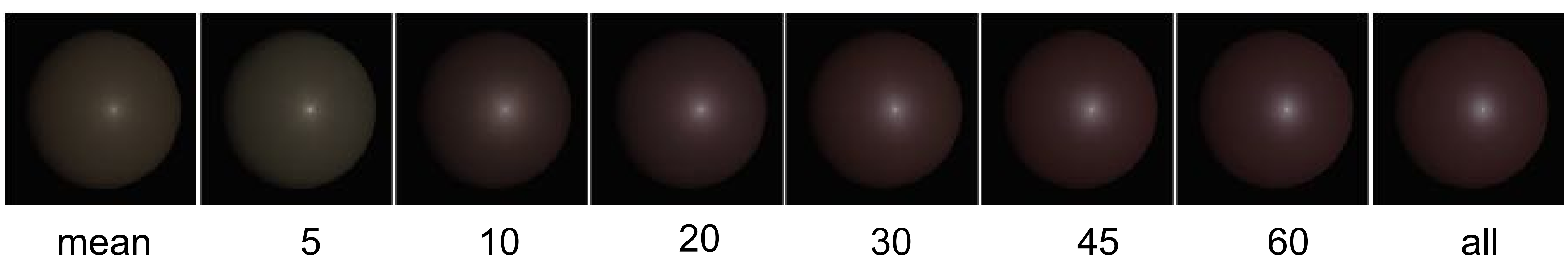
- Each tabulated BRDF is a vector in 90x90x180x3 =4,374,000 dimensional space



PCA



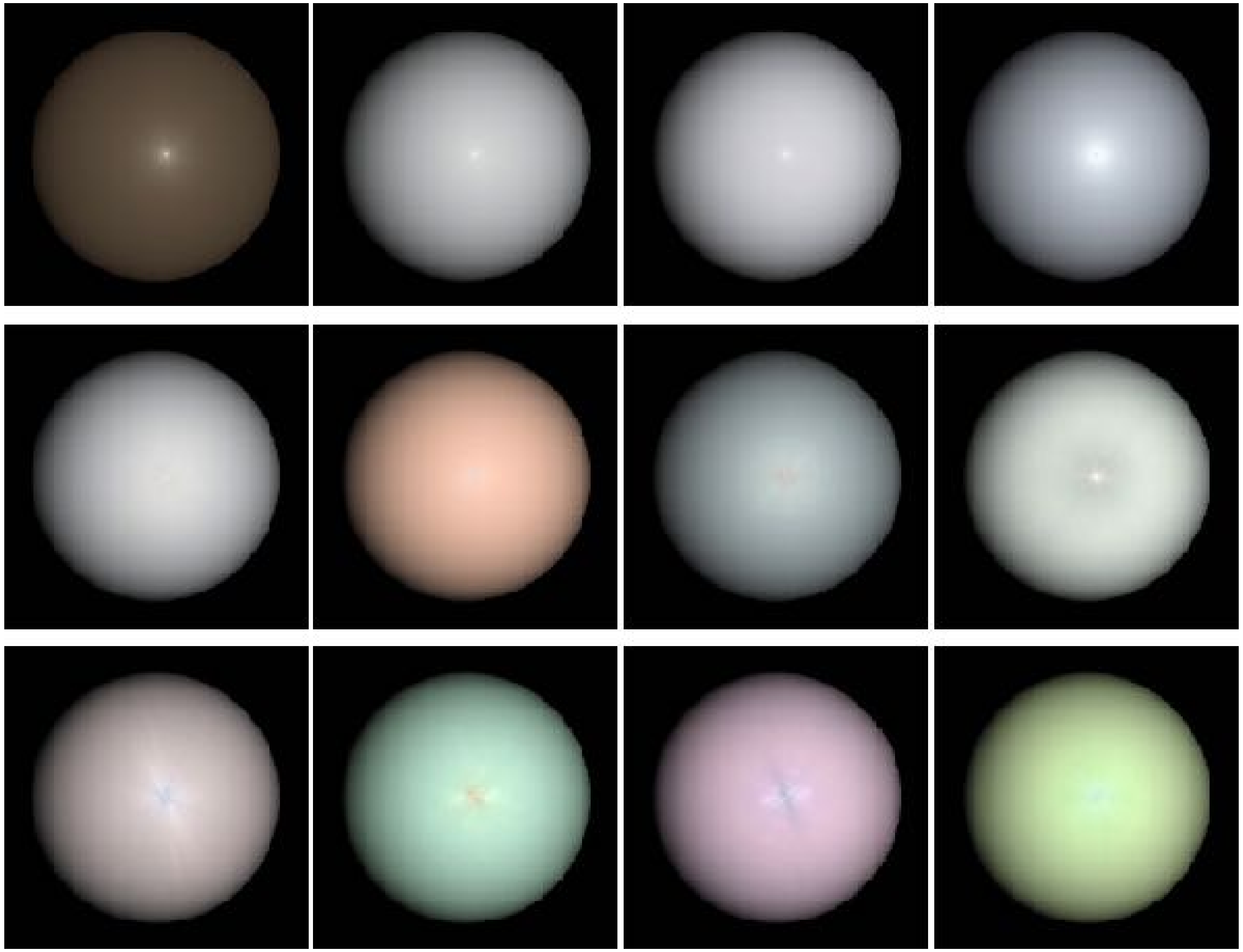
Eigenvalue magnitude
0
20
40
60
80
100
120
Dimension



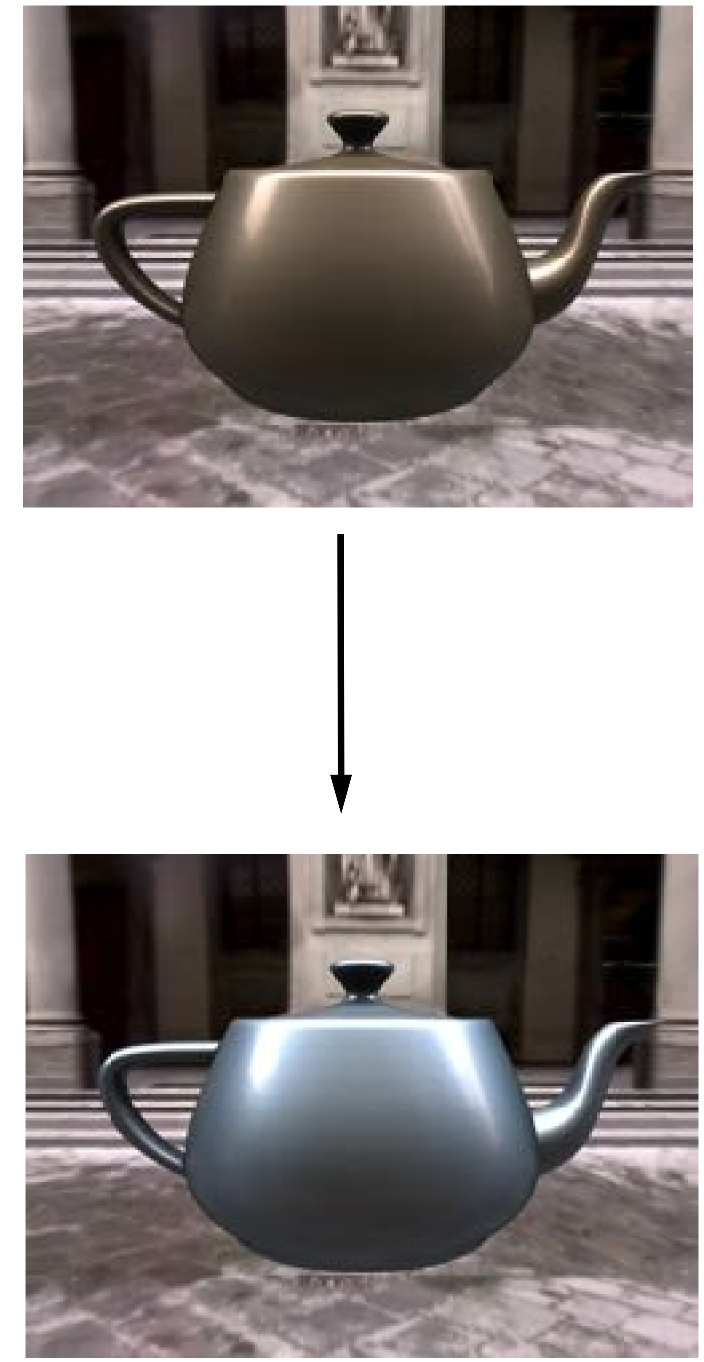
mean
5
10
20
30
45
60
all

PCA

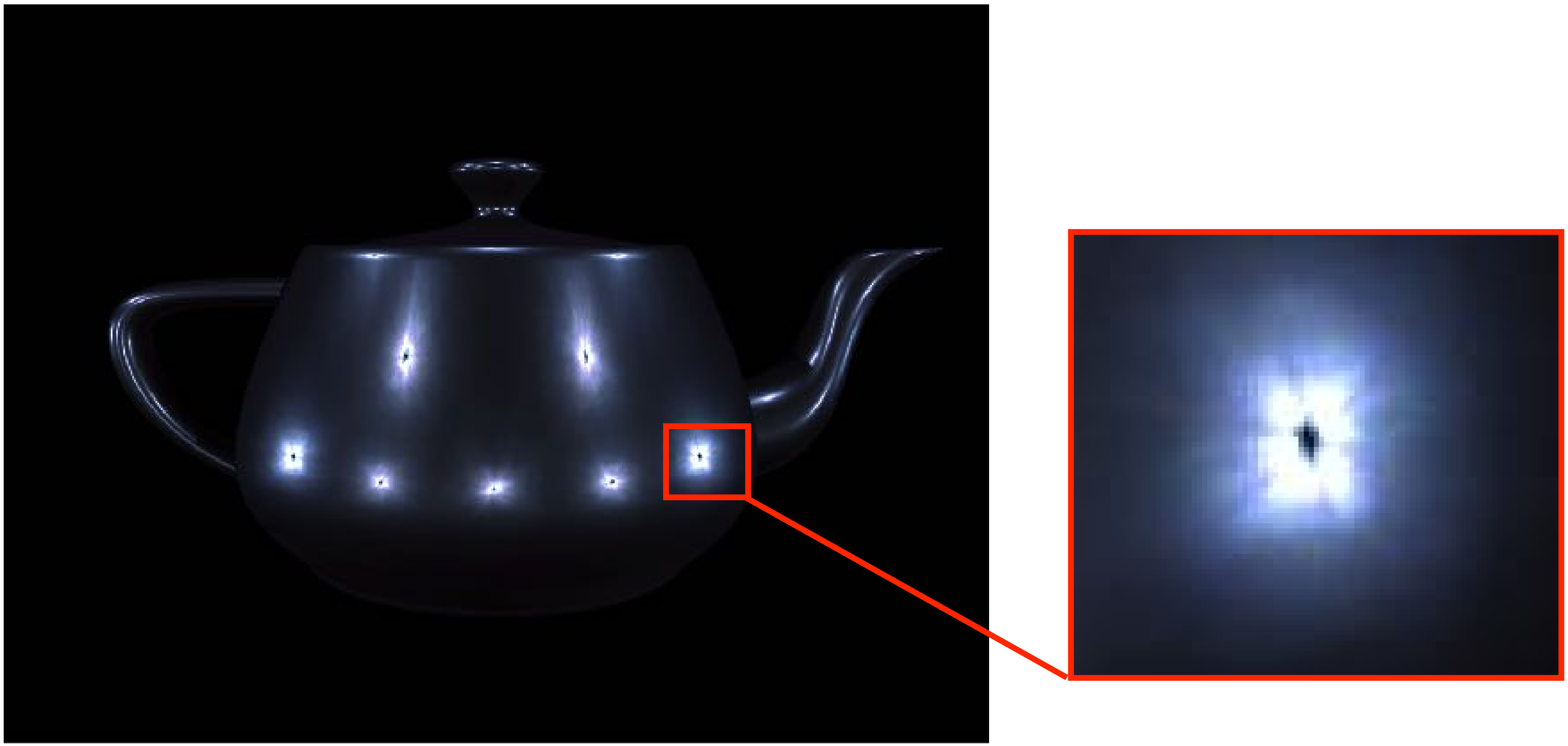
- First 11 PCA components



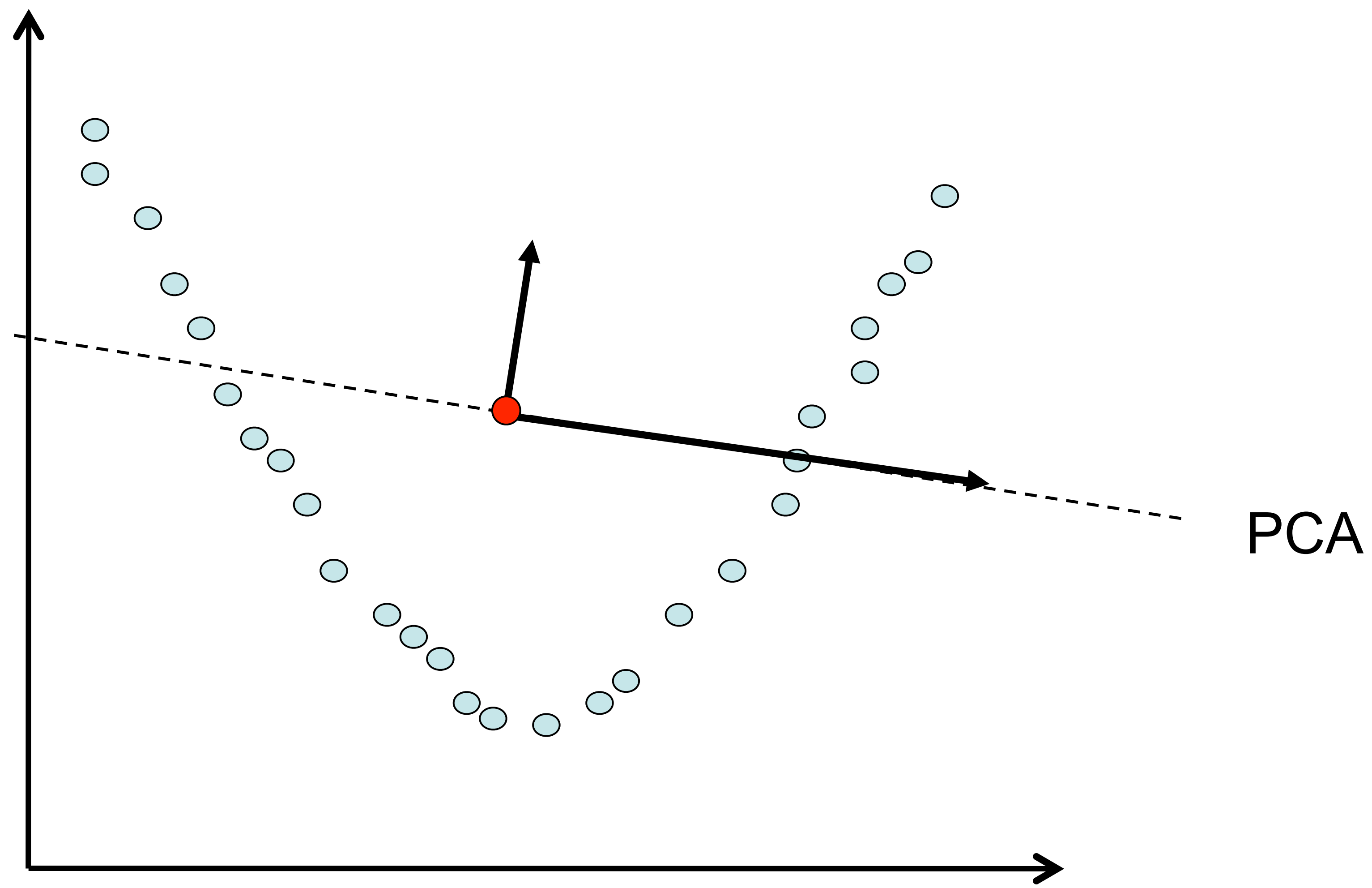
PCA Interpolation



Then, one day...

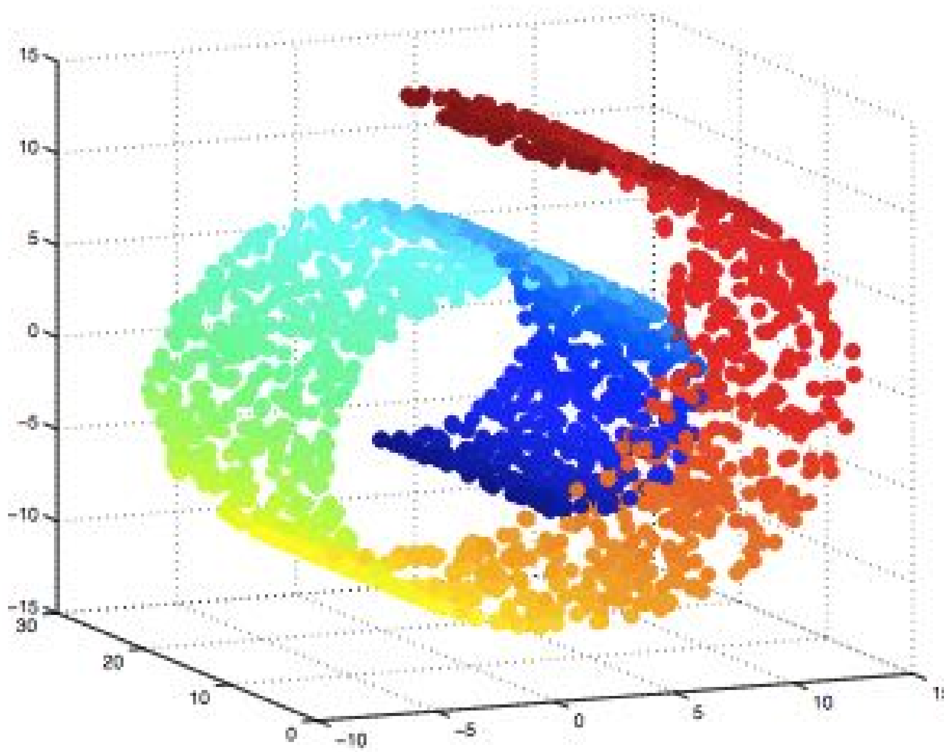


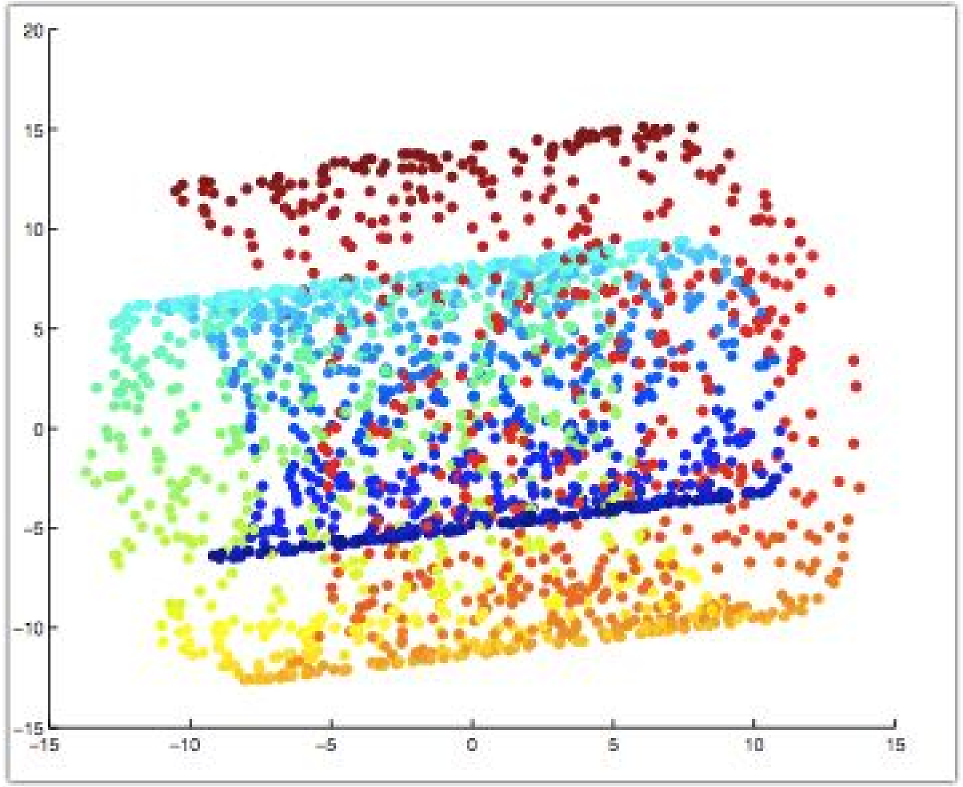
Why do linear models fail?



Why do linear models fail?

- Classic “Swiss Roll” example



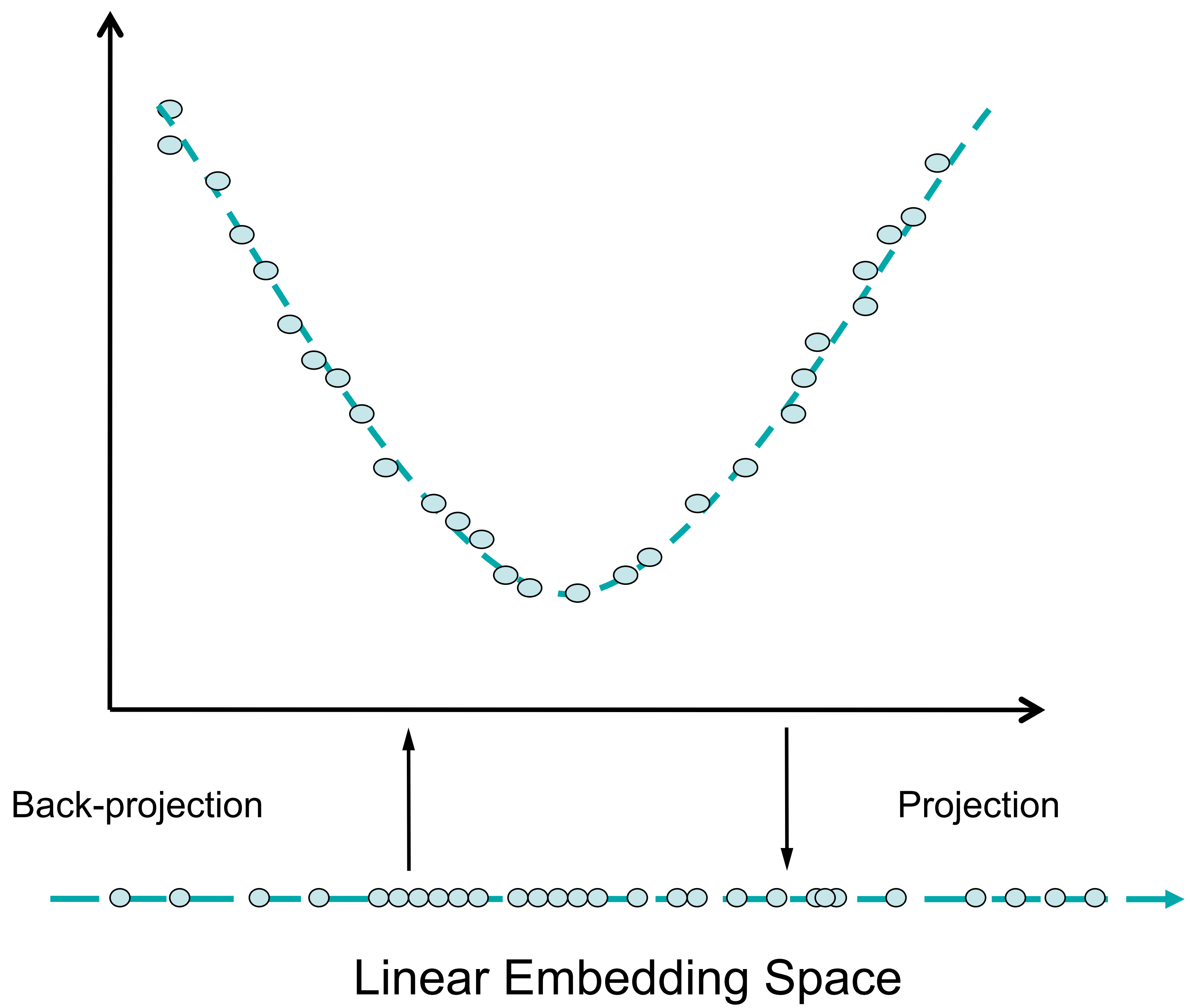


PCA

F. Sha

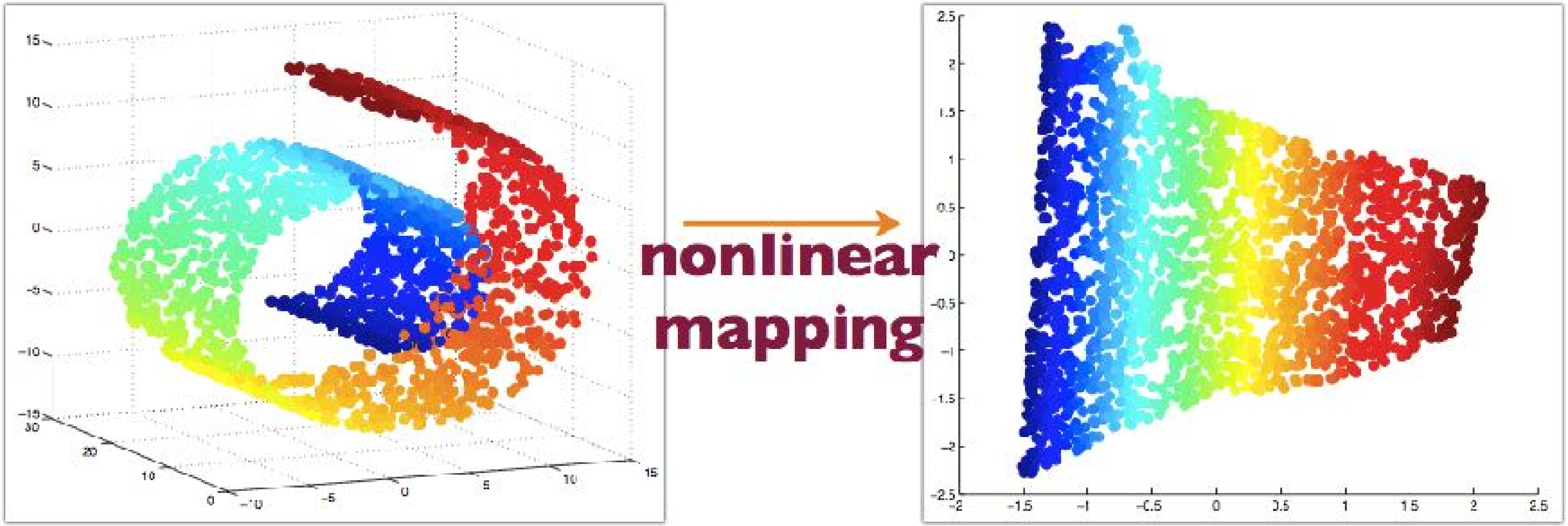
Nonlinear Methods

Non-Linear Manifold Methods



Non-Linear Manifold Methods

- Intuition: Distortion in local areas, but faithful in the global structure



F. Sha

Non-Linear BRDF Model

- 15-dimensional space (instead of 45 PCs)
- More robust - allows extrapolations

Linear Model Extrapolation

Non-linear Model Extrapolation

Dimensionality Reduction

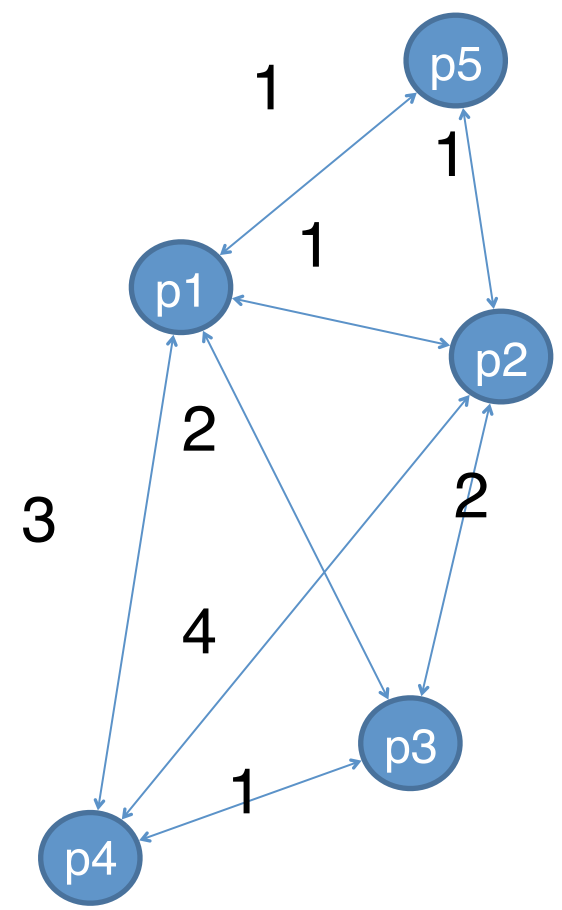
- Linear methods:
  - Principal Component Analysis (PCA) – Hotelling [33]
  - Singular Value Decomposition (SVD) – Eckart/Young [36]
  - Multidimensional Scaling (MDS) – Young [38]
- Nonlinear methods:
  - IsoMap – Tenenbaum [00]
  - Locally Linear Embeddings (LLE) – Roweis [00]

Multidimensional Scaling (MDS)

Multi-Dimensional Scaling

Find a set of points whose pairwise distances match a given distance matrix

	p1	p2	p3	p4	p5
p1	0	1	2	3	1
p2	1	0	2	4	1
p3	2	2	0	1	3
p4	3	4	1	0	1
p5	1	1	3	1	0



Classical MDS

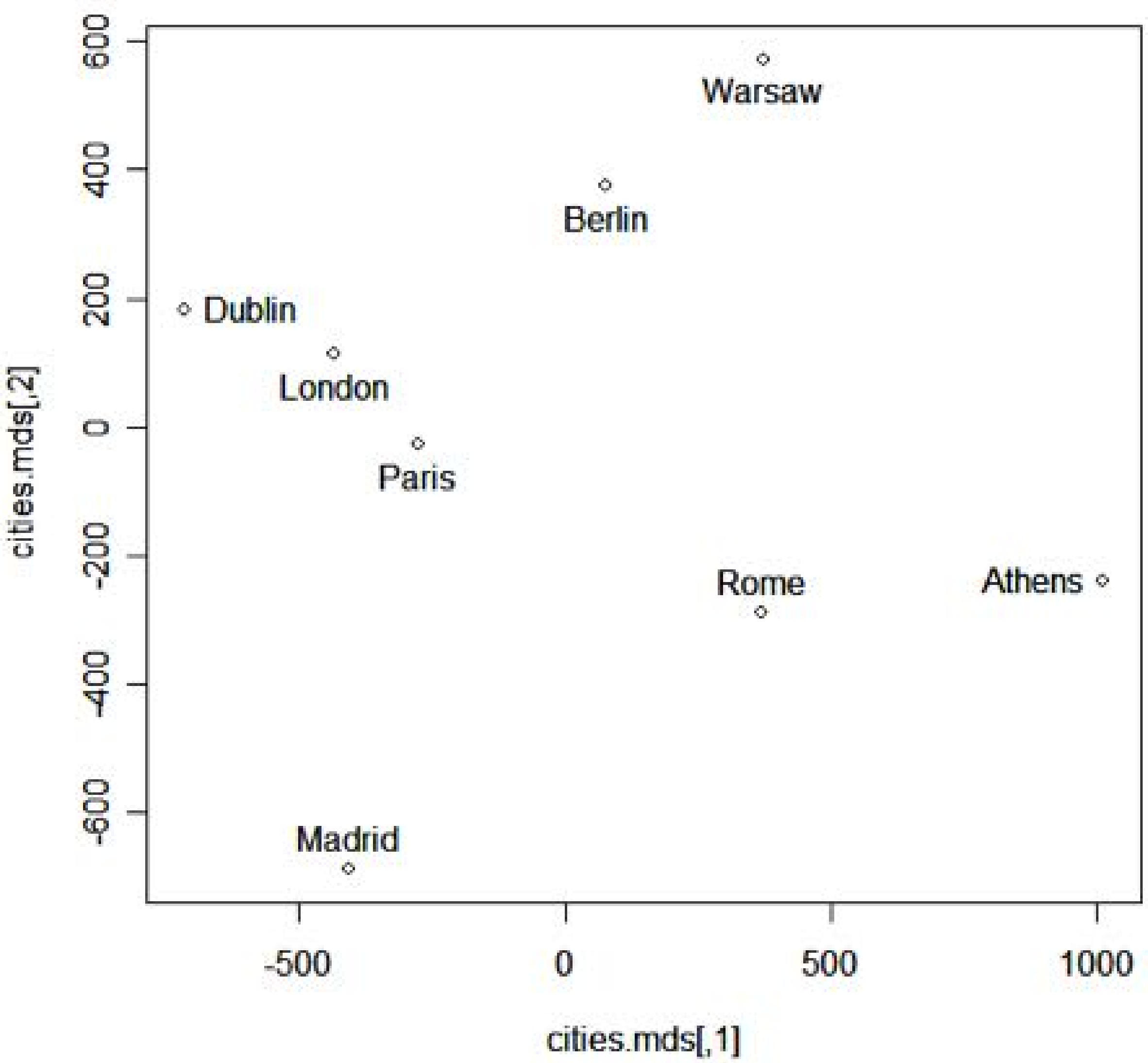
- Given n x n matrix of pairwise distances between data points
- Compute n x k matrix X with coordinates of distances with some linear algebra magic
- Perform PCA on this matrix X

European Cities Data

- Distances between European cities

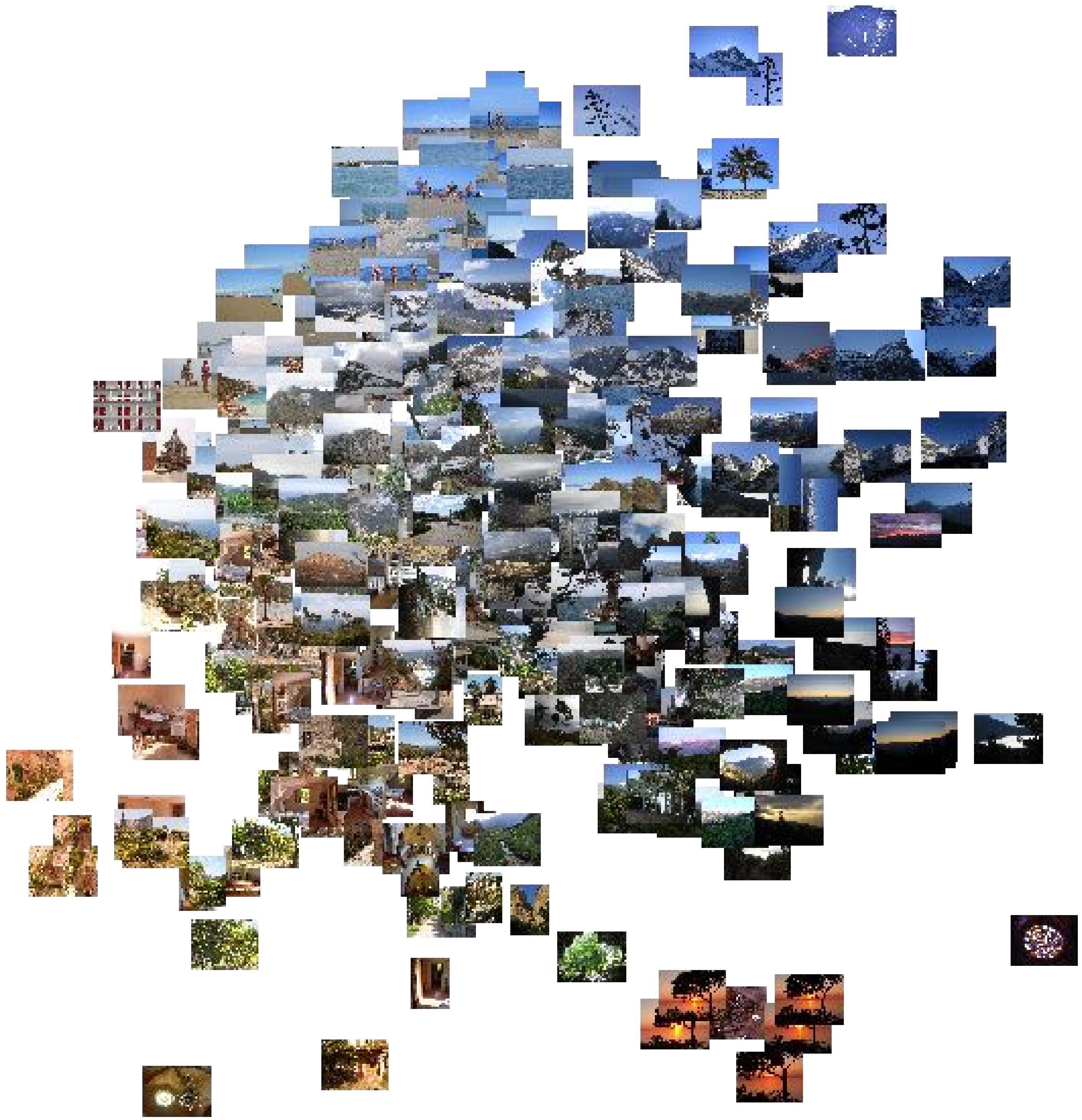
	Athens	Berlin	Dublin	London	Madrid	Paris	Rome	Warsaw
Athens	0	1119	1777	1486	1475	1303	646	1013
Berlin	1119	0	817	577	1159	545	736	327
Dublin	1777	817	0	291	906	489	1182	1135
London	1486	577	291	0	783	213	897	904
Madrid	1475	1159	906	783	0	652	856	1483
Paris	1303	545	489	213	652	0	694	859
Rome	646	736	1182	897	856	694	0	839
Warsaw	1013	327	1135	904	1483	859	839	0

Result of MDS



T. Yang

Color Images

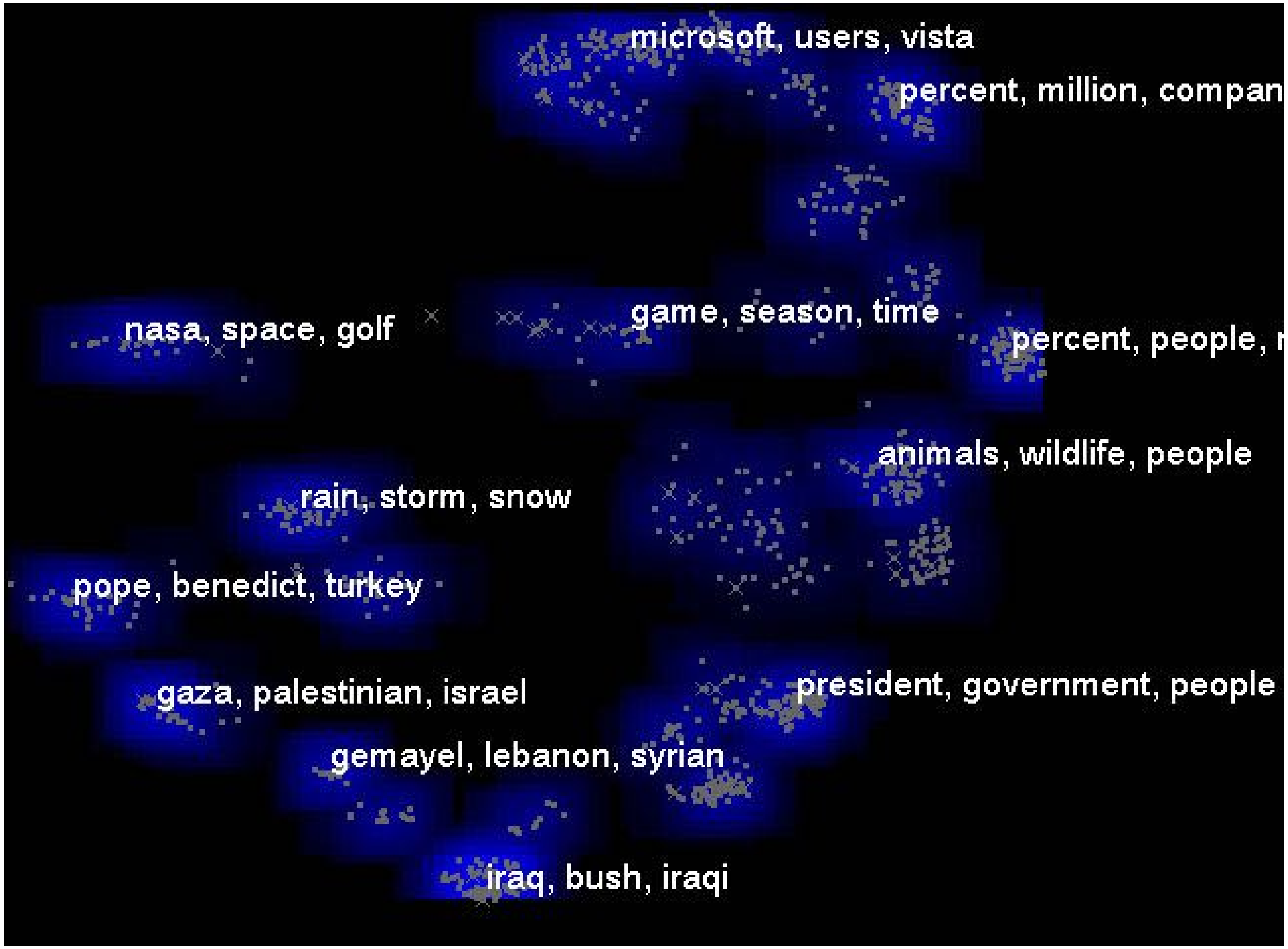


N. Bonneel

Facebook Friends

N. Bonneel

Documents

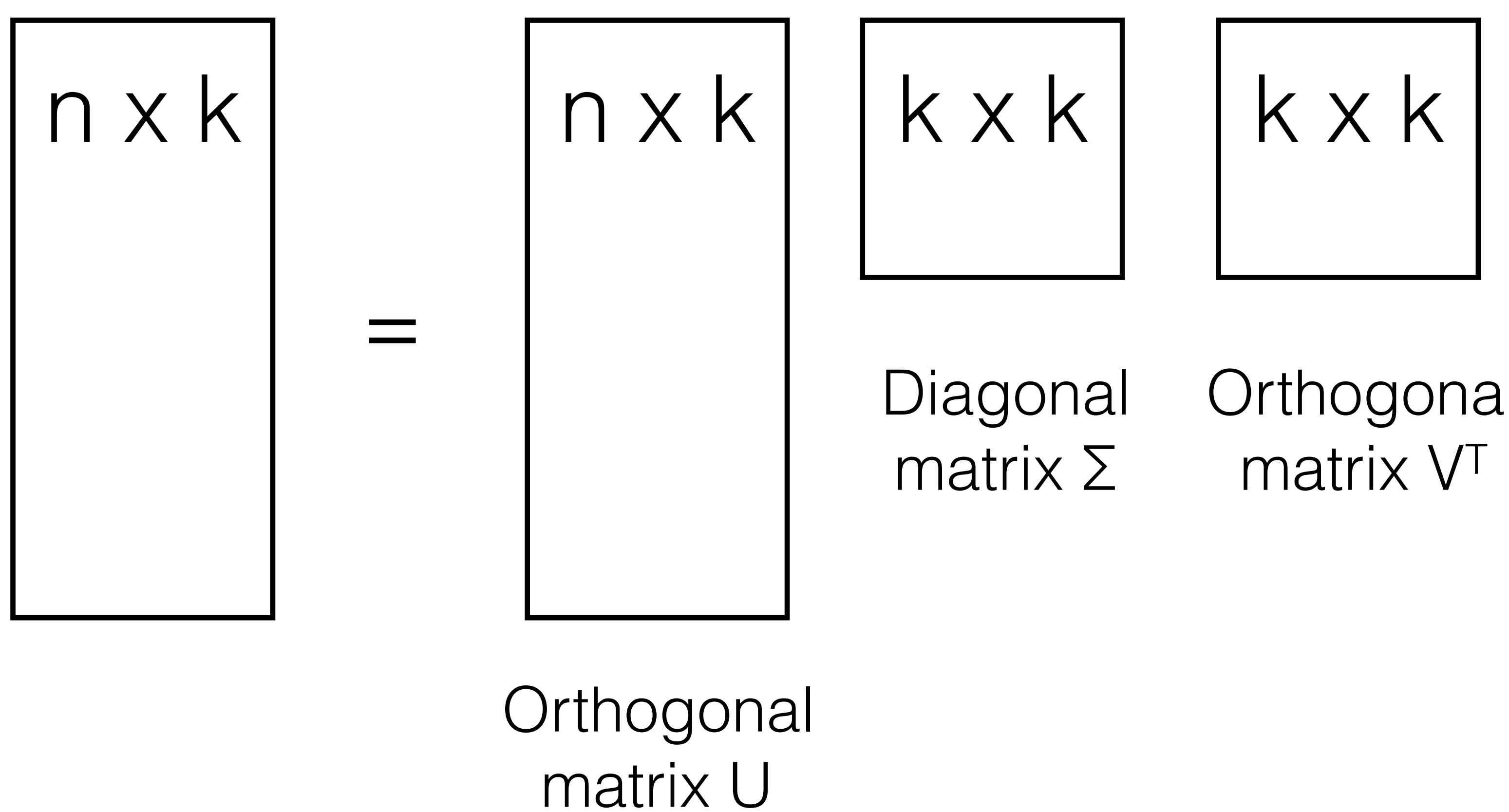


INSPIRE, PNNL

Singular Value Decomposition (SVD)

Singular Value Decomposition (SVD)

$$\mathbf{X} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{T}}$$



Based on slide from A. Ihler

PCA vs. SVD

$$\mathbf{XX}^{\top} = \mathbf{WDW}^{\top}$$

Sample covariance
(symmetric, diagonalizable)

$$\mathbf{X} = \mathbf{U\Sigma V}^{\top}$$

SVD

$$\mathbf{XX}^{\top} = (\mathbf{U\Sigma V}^{\top})(\mathbf{U\Sigma V}^{\top})^{\top}$$
$$\mathbf{XX}^{\top} = (\mathbf{U\Sigma V}^{\top})(\mathbf{V\Sigma U}^{\top})$$

Some linear algebra magic…

$$\mathbf{XX}^{\top} = \mathbf{U\Sigma}^{2}\mathbf{U}^{\top}$$

Eigendecomposition of XX^T
(up to scale factor 1/N)

SVD Properties

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$$

- Non-zero (diagonal) values of Σ are the singular values of X (i.e., square roots of eigenvalues of S)
- Columns of U are the principal components (i.e., eigenvectors of S)
- Using the SVD to perform PCA makes more sense numerically since $XX^\top$ can be unstable

Face Modeling & Animation

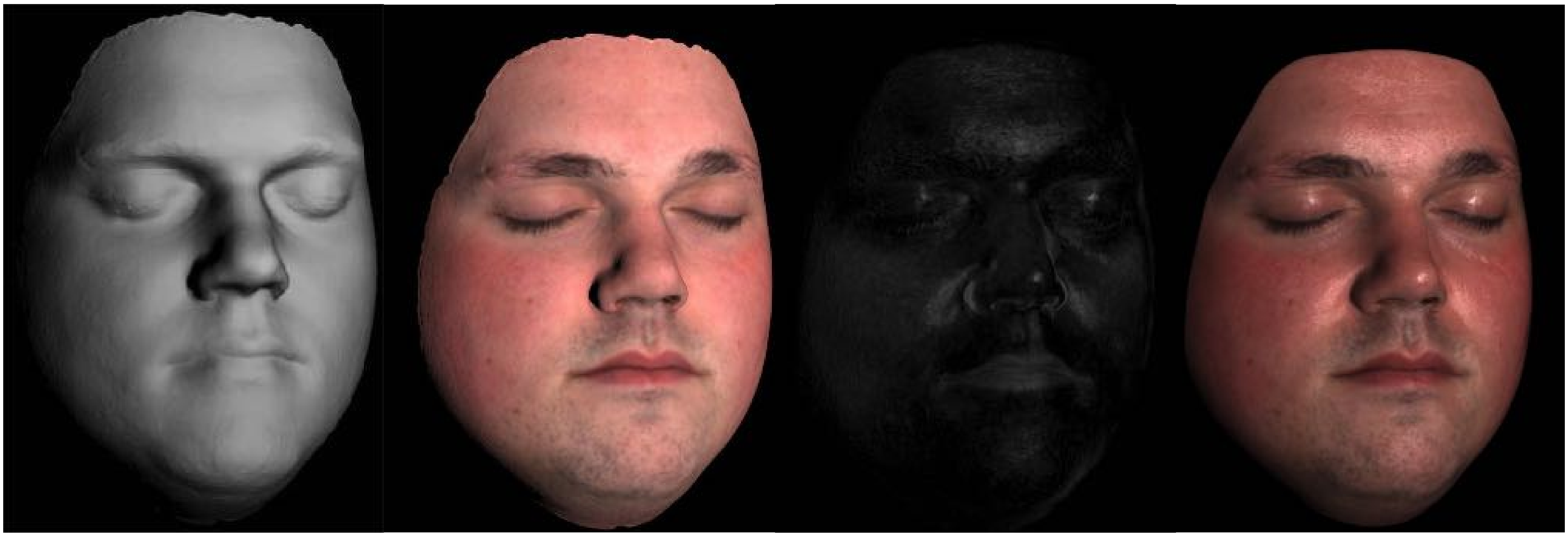
[Vlasic et al., SIGGRAPH 2005]

[Bickel et al., SIGGRAPH 2007]

[Dale et al., SIGASIA 2011]

The MERL Face Scanning Dome

Face Model



3D Geometry + Diffuse Scattering + Specular Reflection = Final Result

http://vcg.seas.harvard.edu/publications/analysis-human-faces-using-measurement-based-skin-reflectance-model

Which one is real?

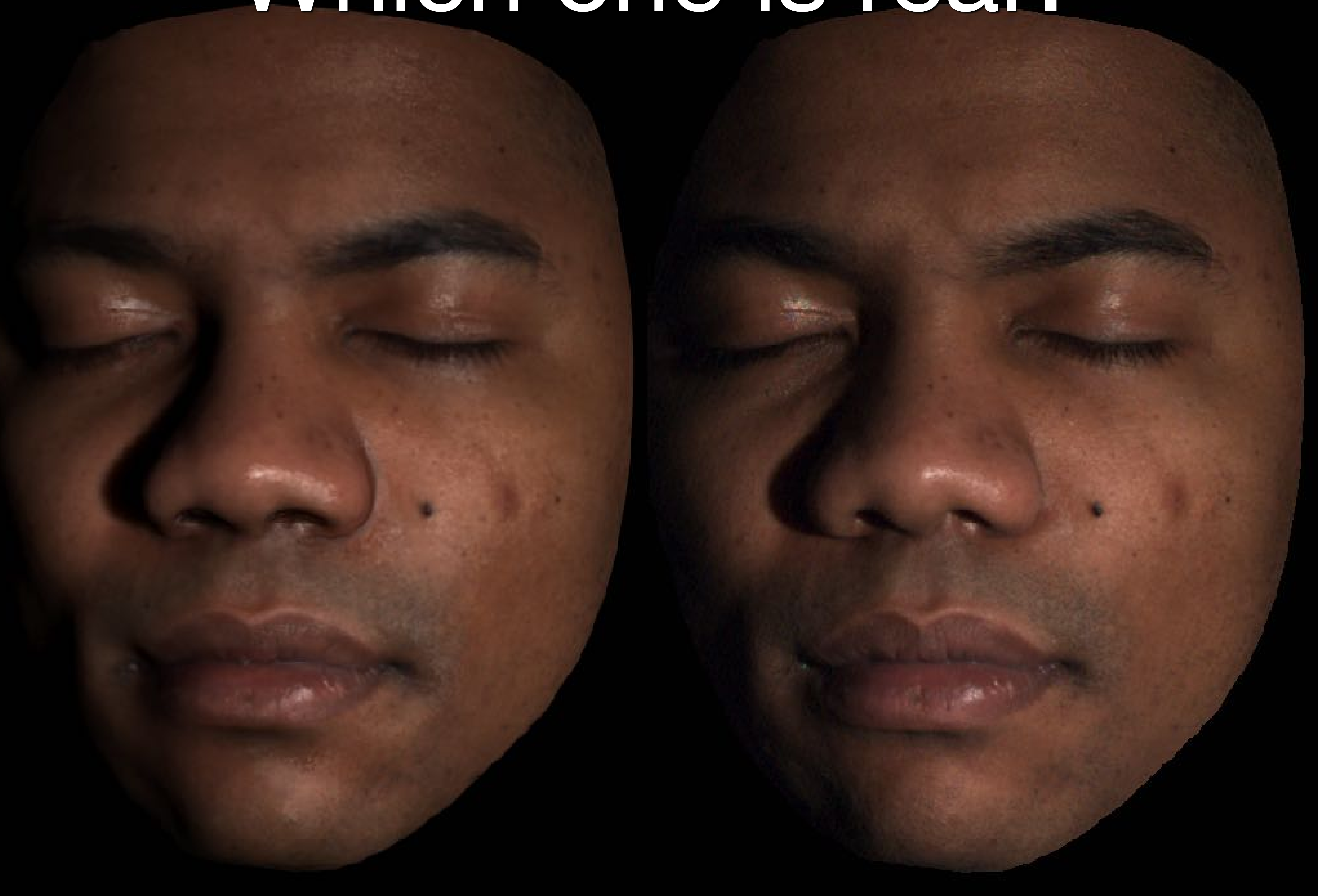


Photo vs. Model

Original Model

Appearance Change

The MERL Face Database

- Scanned ~500 subjects
- Classification by
  - Skin type, gender, age, …
  - Facial region

- Total data: ~12 terabytes
- Analysis of variations in model parameters

Analysis of Human Faces using a Measurement-Based Skin Reflectance Model

Tim Weyrich [*] Wojciech Matusik [†] Hanspeter Pfister [†] Bernd Bickel [*] Craig Donner [‡] Chien Tu [†]
Janet McAndless [†] Jinho Lee [†] Addy Ngan [§] Henrik Wann Jensen [‡] Markus Gross [*]

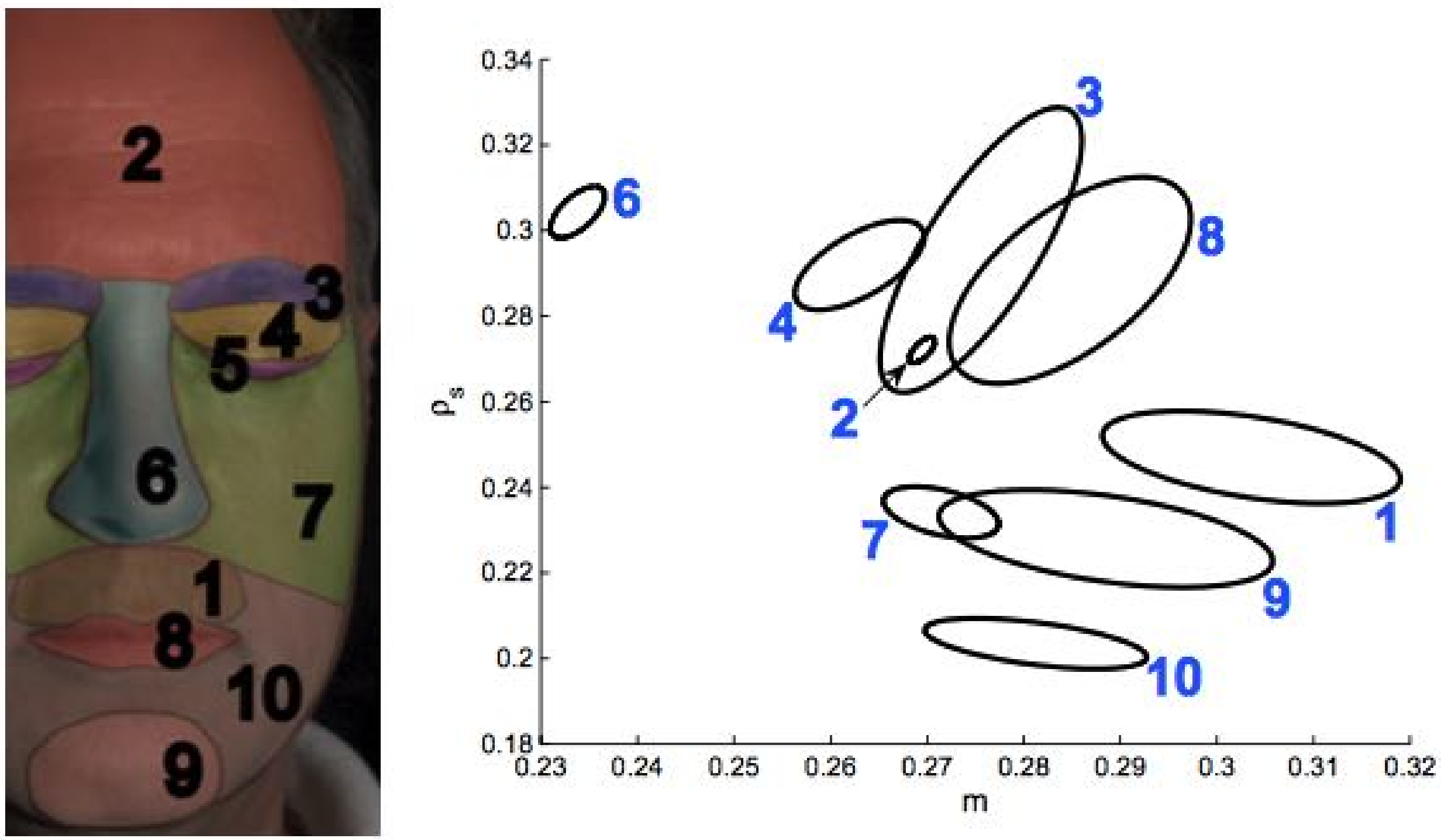
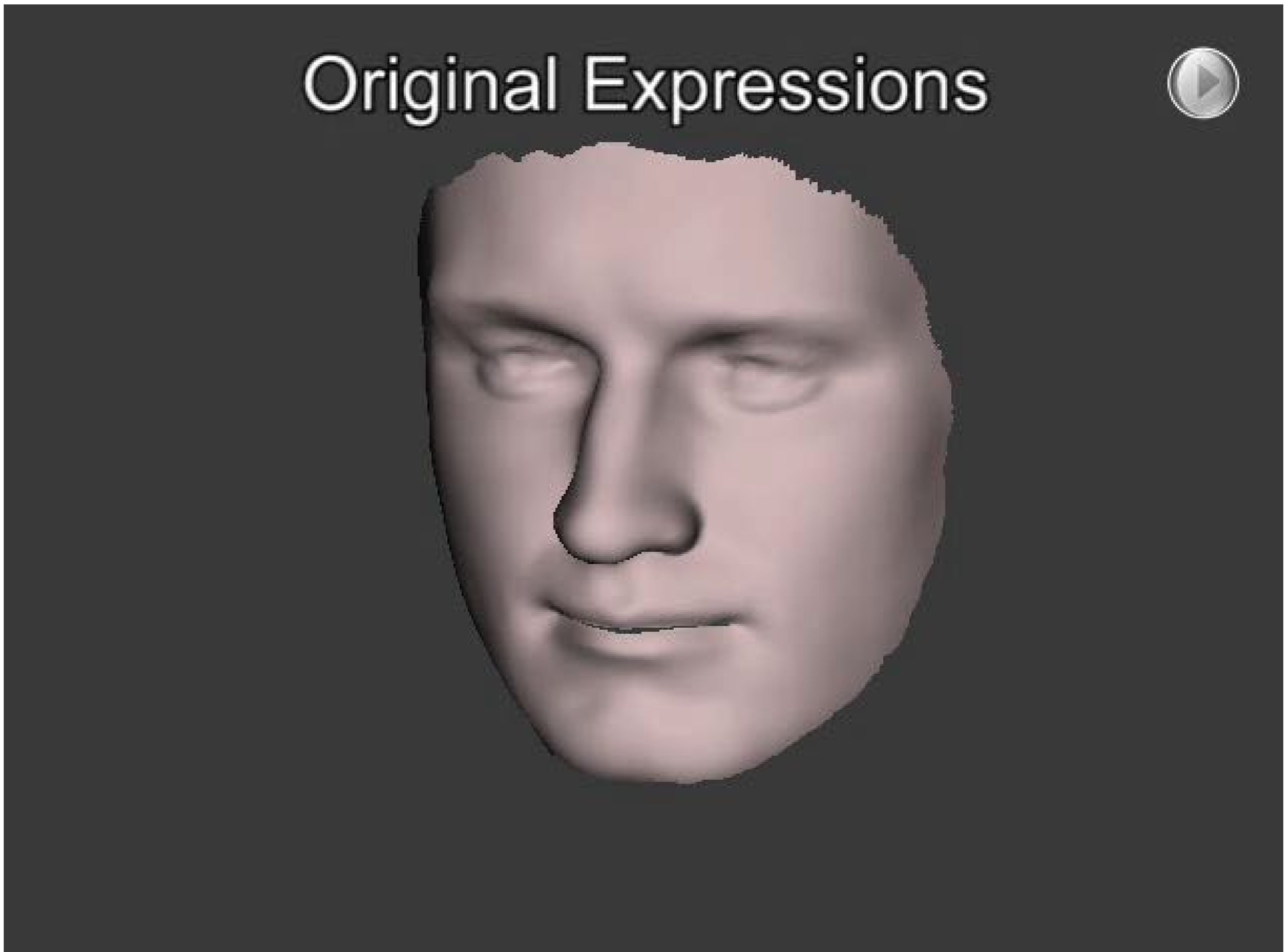


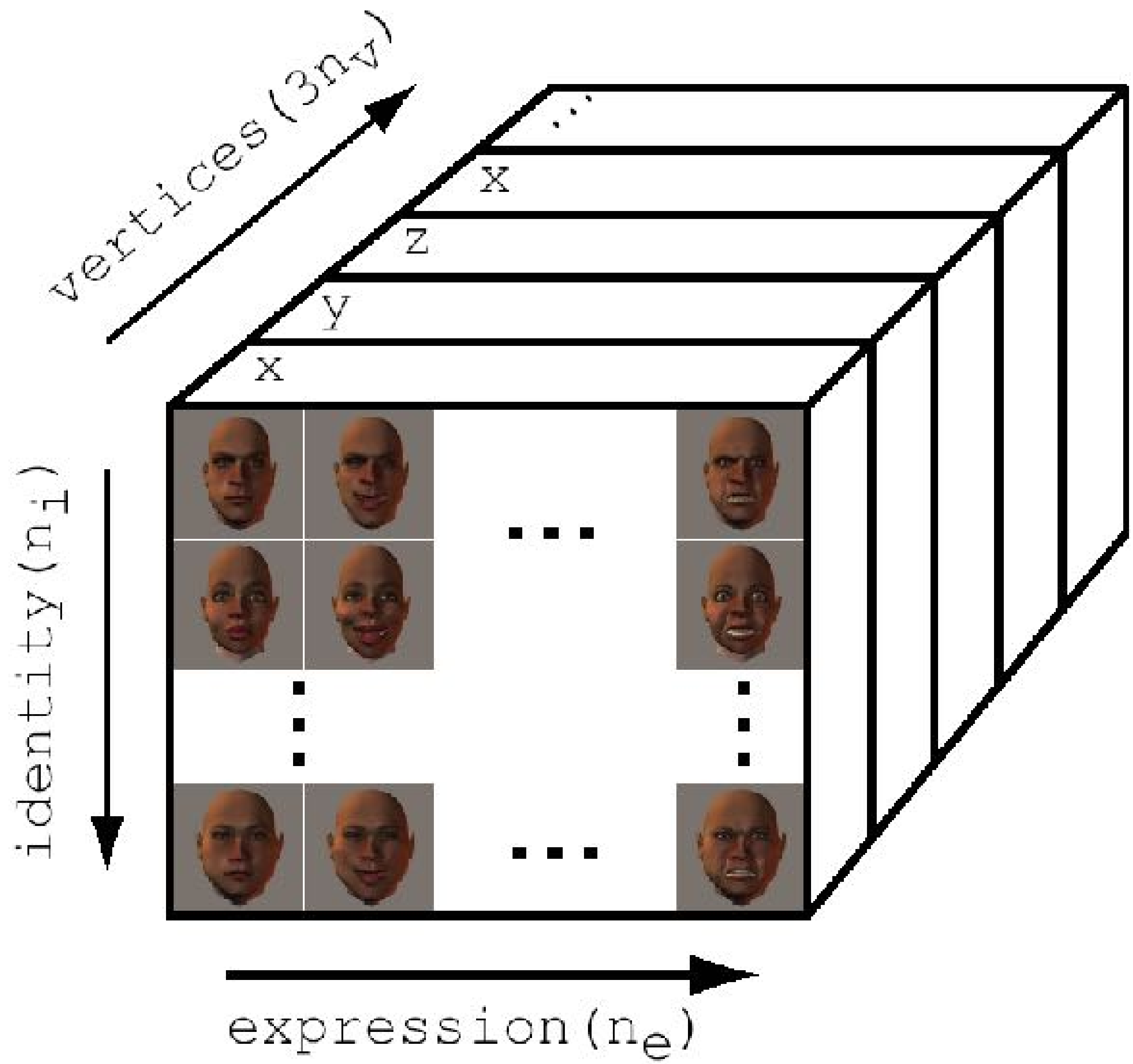
Figure 15: *Left: 10 face regions. Right: Variation of Torrance-Sparrow parameters per face region averaged over all subjects. The center of the ellipse indicates the mean. The axis of each ellipse shows the directions of most variation based on a PCA.*

Data Acquisition

16 identities x 5 visemes x 5 expressions = 400 scans



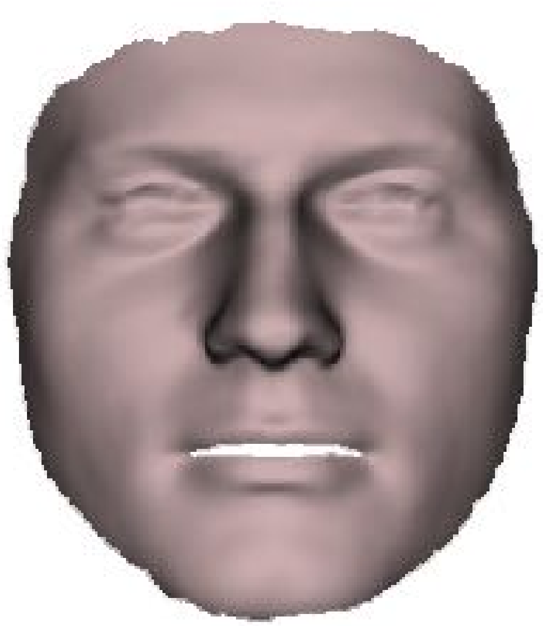
Data Tensor

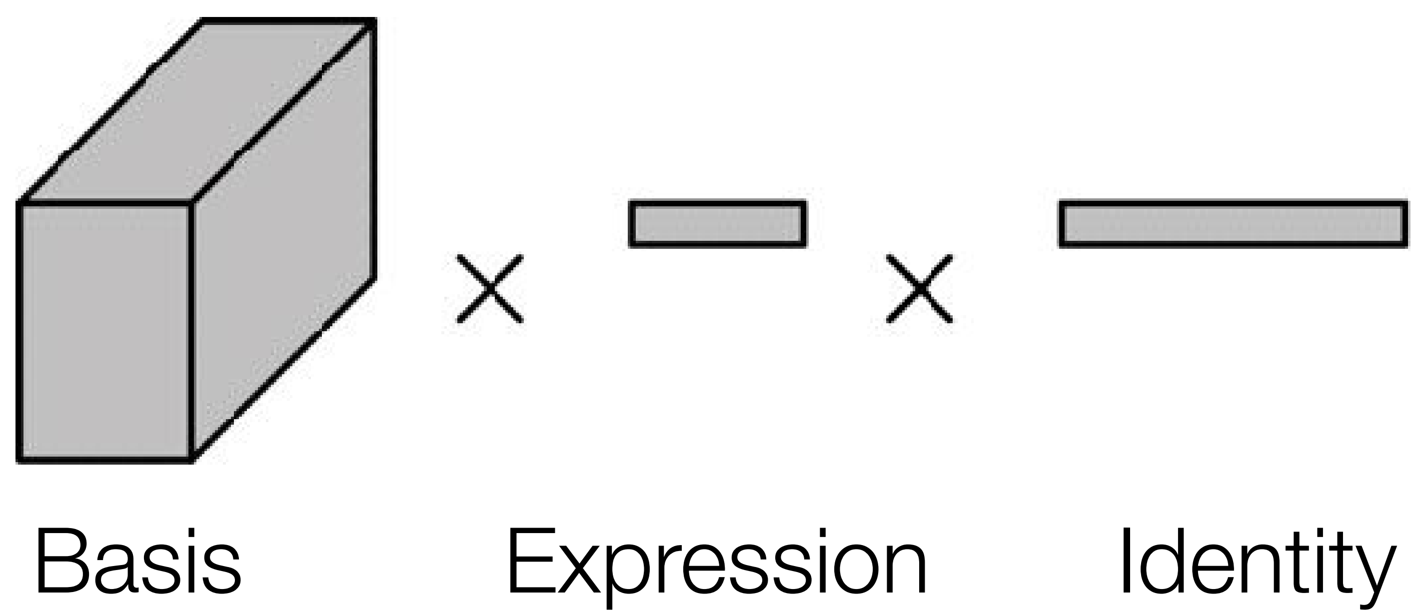


Multilinear Face Model

- Generalization of SVD to tensors
- Weights for identity and expression determine the synthesized face

$$\mathscr{T} = \mathscr{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3 \cdots \times_N \mathbf{U}_N$$

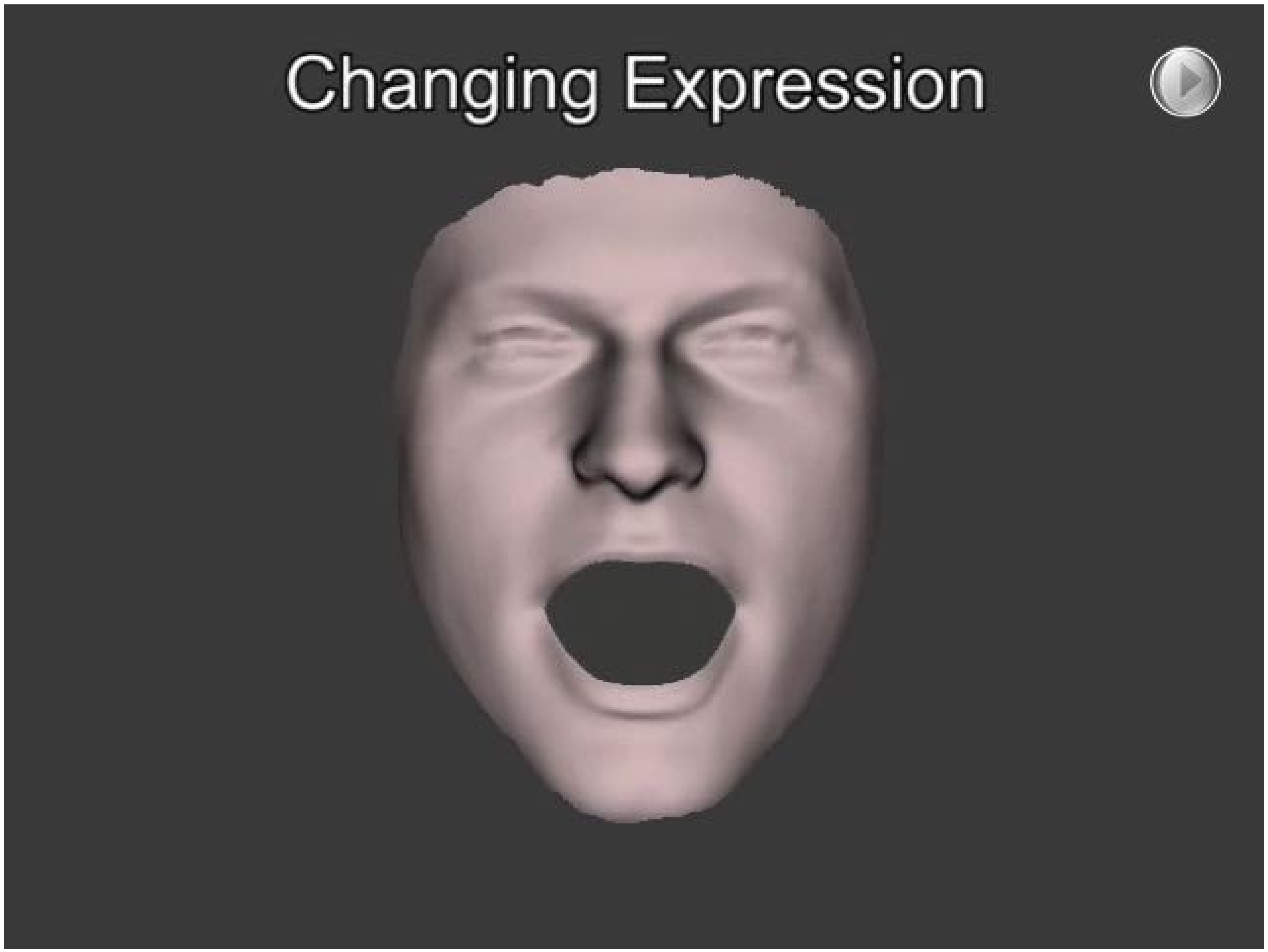




http://vcg.seas.harvard.edu/publications/face-transfer-multilinear-models

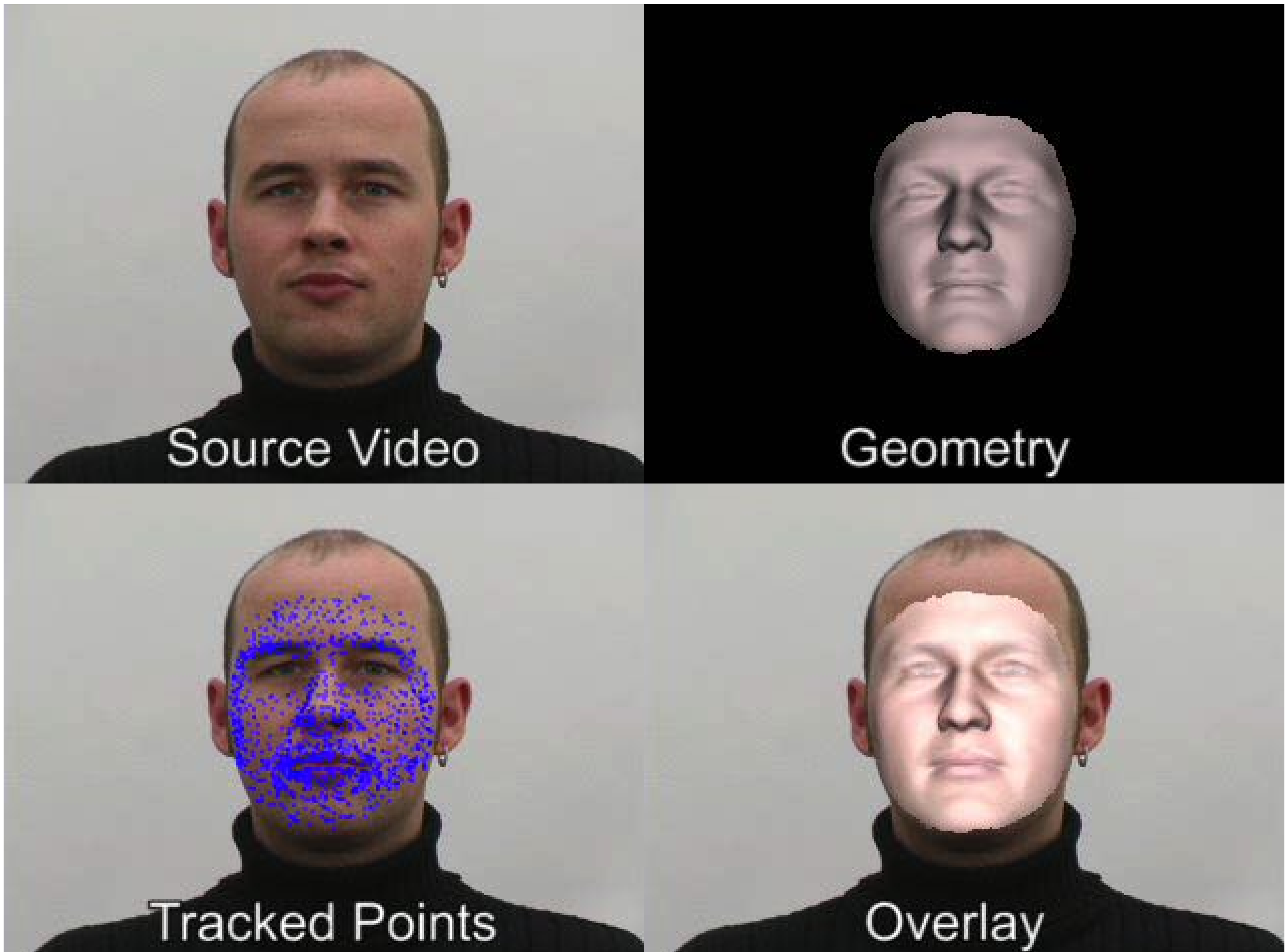
Example

Smoothly changing expressions and identity (interpolation)



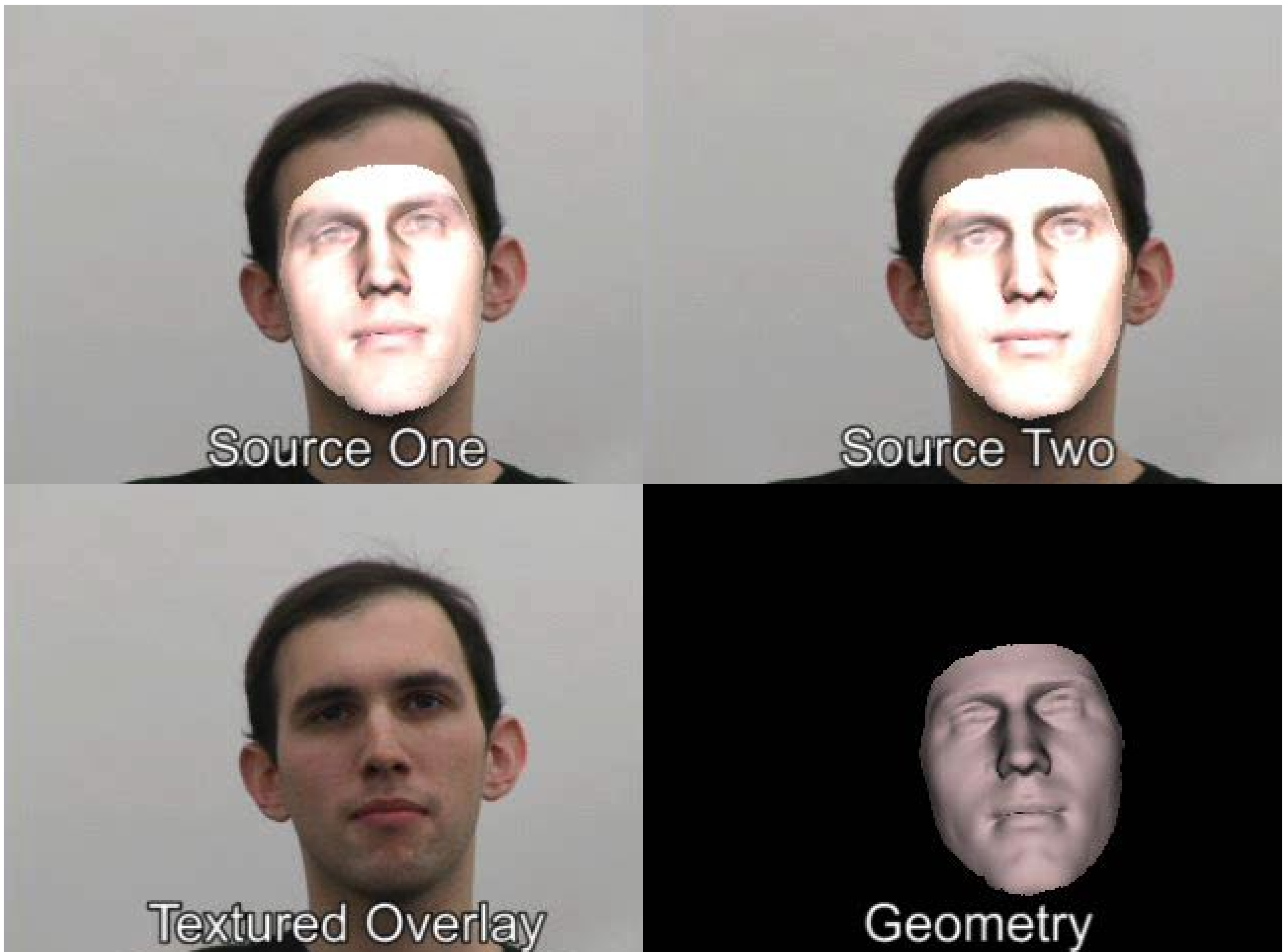
Matching to Video

Sequence tracked with optical flow



Video Rewrite

Blending a new performance into a source video



Video Puppetry

Blending a new performance with different actors

Video Face Replacement

http://vcg.seas.harvard.edu/publications/video-face-replacement

Video Face Replacement

http://vcg.seas.harvard.edu/publications/video-face-replacement

Video Face Replacement

http://vcg.seas.harvard.edu/publications/video-face-replacement

References

- "Principal Components Analysis" by Brian Junker and Cosma Shalizi, CMU
- "Principal Component Analysis" by Mark Richardson
- "Biomedical Data Science" by Rafael Irizarry and Michael Love