Some Interesting Problems in Pattern Recognition and Image Processing

JEN-MEI CHANG

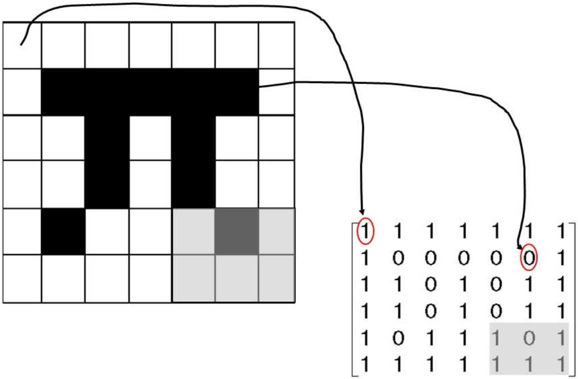
Department of Mathematics and Statistics
California State University, Long Beach
jchang9@csulb.edu

University of Southern California Women in Math

Outline

Background: Data Matrix

An r-by-c gray scale digital image corresponds to an r-by-c matrix where each entry enumerates one of the 256 possible gray levels of the corresponding pixel.



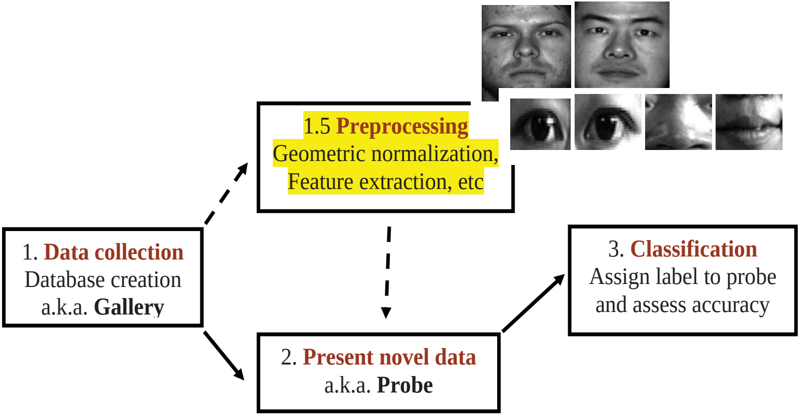
Background: Data Vector

Realize the data matrix by its columns and concatenate columns into a single column vector.

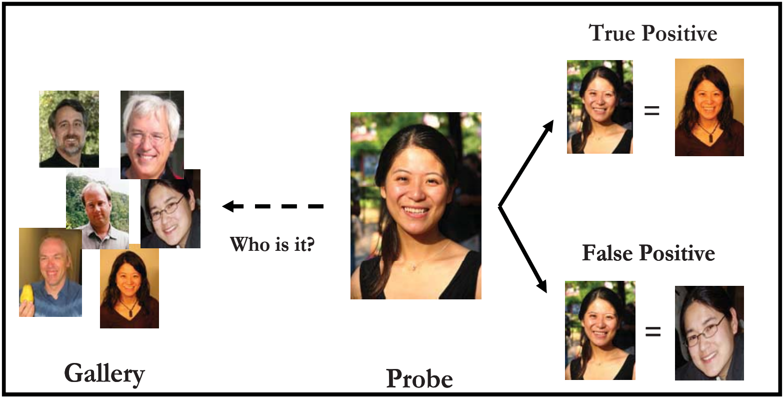
$$
\begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 1 & 0 & 1 & 1 \\
1 & 1 & 0 & 1 & 0 & 1 & 1 \\
1 & 0 & 1 & 1 & 1 & 0 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1
\end{bmatrix}
\longrightarrow
\begin{bmatrix}
1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ \hline 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ \hline \vdots \\ \hline 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1
\end{bmatrix}
$$

IMAGE $\longrightarrow$ MATRIX $\longrightarrow$ VECTOR

Classification/Recognition Problem



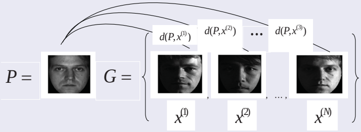
Face Recognition Problem



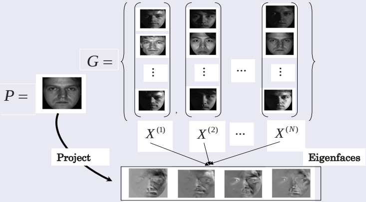
Architectures

Historically

- single-to-single

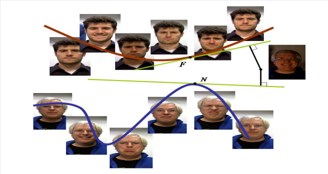


- single-to-many

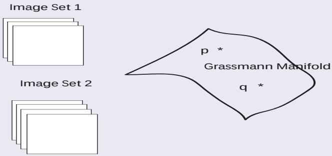


Currently

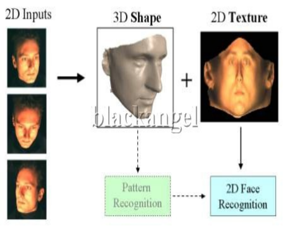
- subspace-to-subspace



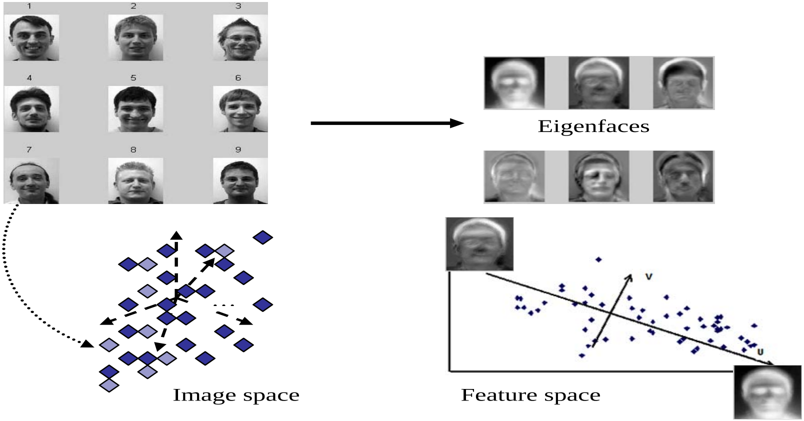
- many-to-many



Commercial Applications

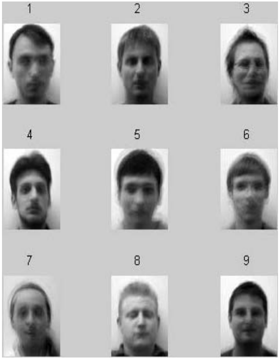


A Possible Mathematical Approach

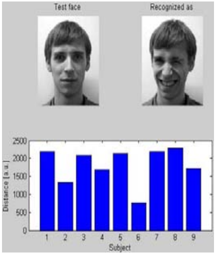


(Adapted from Vladimir Bondarenko at University of Konstanz, ST; http://www.inf.uni-konstanz.de/cgip/lehre/na_08/Lab2/5_FaceRecognition/html/myFaceRecognition.html

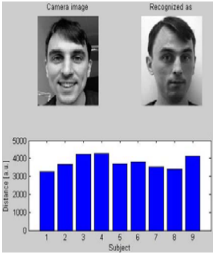
A Possible Mathematical Approach



Database in feature space

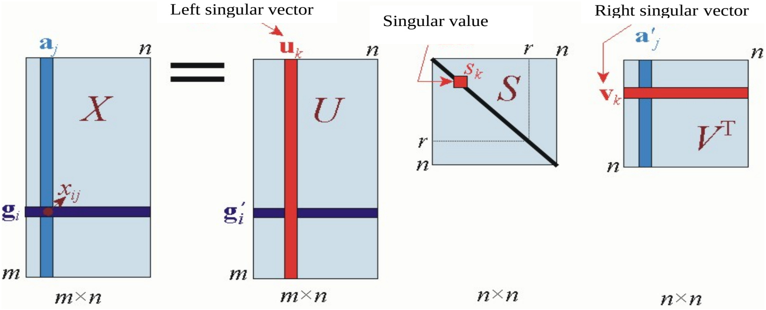


Classification in feature space



(Adapted from Vladimir Bondarenko at University of Konstanz, ST; http://www.inf.uni-konstanz.de/cgip/lehre/na_08/Lab2/5_FaceRecognition/html/myFaceRecognition.html

$\longrightarrow X = USV^T$



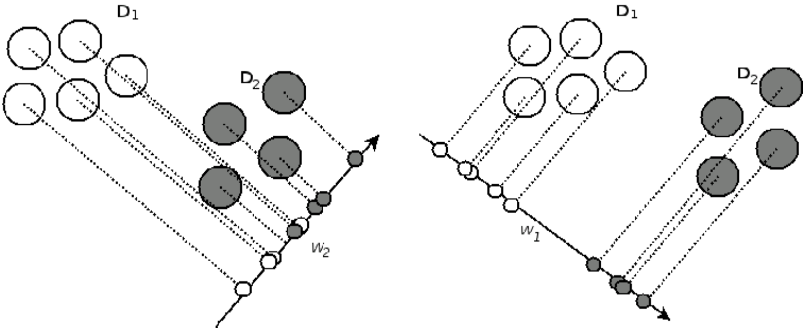
- $V = [v_1, \ldots, v_r, v_{r+1}, \ldots, v_n]$ is orthogonal with v_i's eigenvectors of $X^T X$.
- $S = \text{diag}(s_1, \ldots, s_r, 0, \ldots, 0)$ is diagonal with s_i's square root of eigenvalues of $X^T X$.
- $U = S^{-1} XV$ is orthogonal with u_i's eigenvectors of XX^T.

Bankruptcy prediction is the art of predicting bankruptcy and various measures of financial distress of public firms. It is a vast area of finance and accounting research. The importance of the area is due in part to the relevance for creditors and investors in evaluating the likelihood that a firm may go bankrupt[1].

- Form a feature vector for each firm where each entry in the feature vector is a numerical value for a certain characteristic, e.g., number of customers and annual profit.
- This is a two-class (bankrupt or non-bankrupt) classification problem.

[1] adapted from Wikipedia

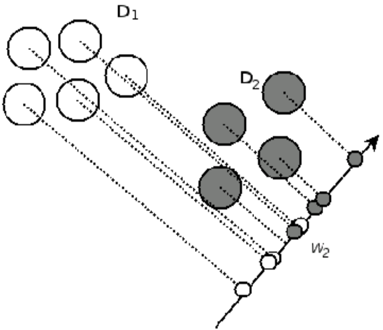
A Possible Mathematical Approach



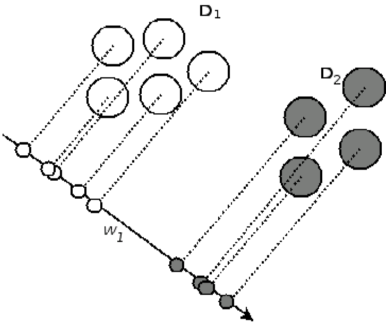
Bad projection Good projection

Question: What are the characteristics of a **good** projection?

A Possible Mathematical Approach



Bad projection

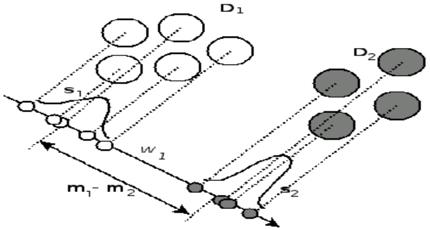


Good projection

Question: What are the characteristics of a **good** projection?

Two-Class LDA

$$m_1 = \frac{1}{n_1} \sum_{x \in D_1} w^T x, \quad m_2 = \frac{1}{n_2} \sum_{y \in D_2} w^T y$$



Look for a projection w that

- maximizes (inter-class) distance in the projected space,
- and minimizes the (intra-class) distances in the projected space.

Two-Class LDA

Namely, we desire a w^* such that

$$w^* = \arg\max_w \frac{(m_1 - m_2)^2}{S_1 + S_2},$$

where $S_1 = \sum_{x \in D_1} (w^T x - m_1)^2$ and $S_2 = \sum_{y \in D_2} (w^T y - m_2)^2$.

Alternatively, (with scatter matrices)

$$w^* = \arg\max_w \frac{w^T S_B w}{w^T S_W w}, \quad (1)$$

with $S_W = \sum_{i=1}^{2} \sum_{x \in D_i} (x - \mathbf{m}_i)(x - \mathbf{m_i})^T$,

$S_B = (\mathbf{m}_2 - \mathbf{m}_1)(\mathbf{m}_2 - \mathbf{m}_1)^T$.

Two-Class LDA

Namely, we desire a w^* such that

$$w^* = \arg\max_w \frac{(m_1 - m_2)^2}{S_1 + S_2},$$

where $S_1 = \sum_{x \in D_1} (w^T x - m_1)^2$ and $S_2 = \sum_{y \in D_2} (w^T y - m_2)^2$.

Alternatively, (with scatter matrices)

$$w^* = \arg\max_w \frac{w^T S_B w}{w^T S_W w}, \tag{1}$$

with $S_W = \sum_{i=1}^{2} \sum_{x \in D_i} (x - \mathbf{m}_i)(x - \mathbf{m_i})^T,$

$S_B = (\mathbf{m}_2 - \mathbf{m}_1)(\mathbf{m}_2 - \mathbf{m}_1)^T.$

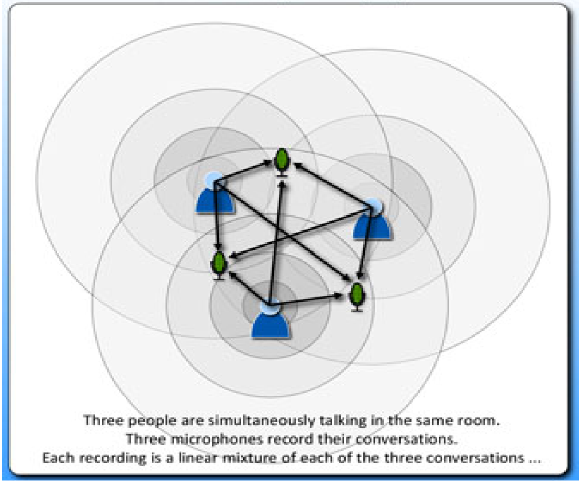
LDA

- The criterion in Equation (1) is commonly known as the generalized Rayleigh quotient, whose solution can be found via the generalized eigenvalue problem

$$S_B w = \lambda S_W w.$$

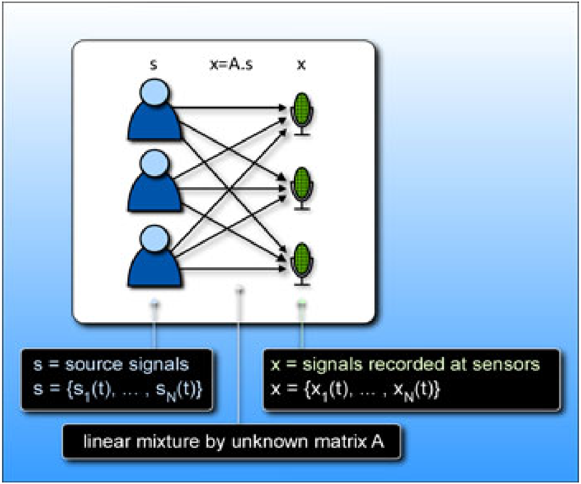
- Once we have the good projection w, we can predict whether a given bank will go bankruptcy by projecting it onto the real line using w.
- Multi-class LDA follows similarly.

Cocktail Party Problem



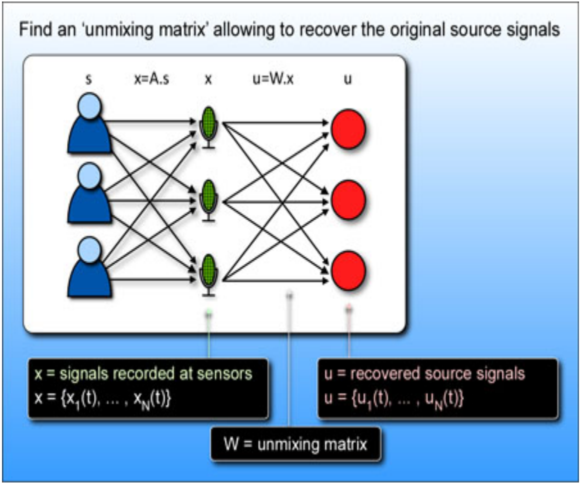
(adapted from André Mouraux)

Cocktail Party Problem



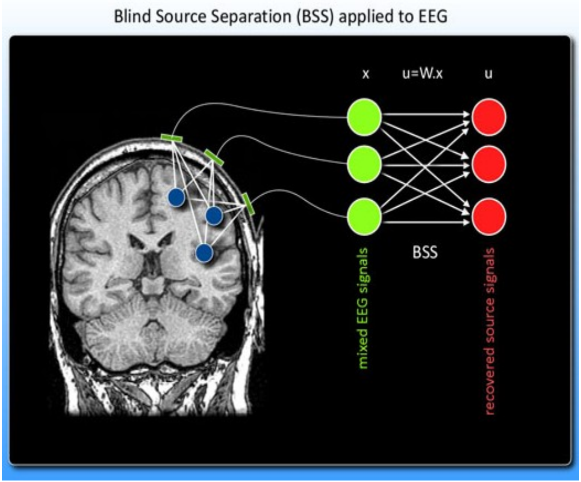
(adapted from André Mouraux)

Cocktail Party Problem



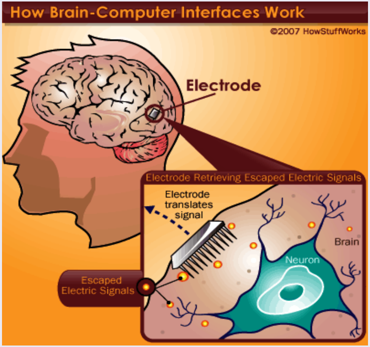
(adapted from André Mouraux)

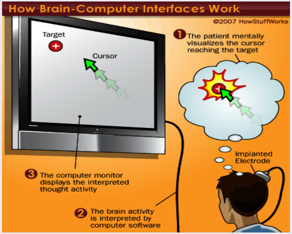
A Similar Problem: EEG

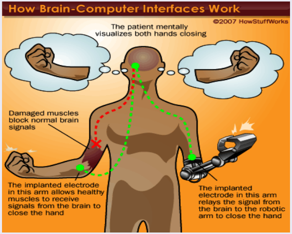


(adapted from André Mouraux)

Commercial Applications







(http://computer.howstuffworks.com/brain-computer-interface1.htm)

A Possible Mathematical Approach

- Decompose observed data into its *noise* and *signal* components:
$$\mathbf{x}^{(\mu)} = \mathbf{s}^{(\mu)} + \mathbf{n}^{(\mu)},$$
or, in terms of data matrices,
$$X = S + N. \quad (S = \text{ signal }, N = \text{ noise })$$
- The optimal first basis vector, ϕ, is taken as a superposition of the data, i.e.,
$$\phi = \psi_1 \mathbf{x}^{(1)} + \cdots + \psi_P \mathbf{x}^{(P)} = X\psi.$$
- May decompose ϕ into signal and noise components
$$\phi = \phi_{\mathbf{n}} + \phi_{\mathbf{s}},$$
where $\phi_{\mathbf{s}} = S\psi$ and $\phi_{\mathbf{n}} = N\psi$.

MNF/BBS

- The basis vector ϕ is said to have maximum noise fraction (MNF) if the ratio

$$D(\phi) = \frac{\phi_{\mathbf{n}}^T \phi_{\mathbf{n}}}{\phi^T \phi}$$

  is a maximum.

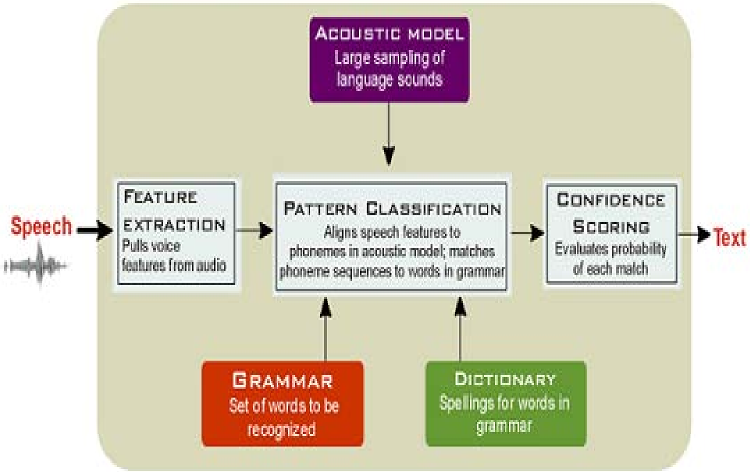
- A steepest descent method yields the *symmetric definite generalized eigenproblem*

$$N^T N \psi = \mu^2 X^T X \psi.$$

  This problem may be solved without actually forming the product matrices $N^T N$ and $X^T X$, using the generalized SVD (gsvd).

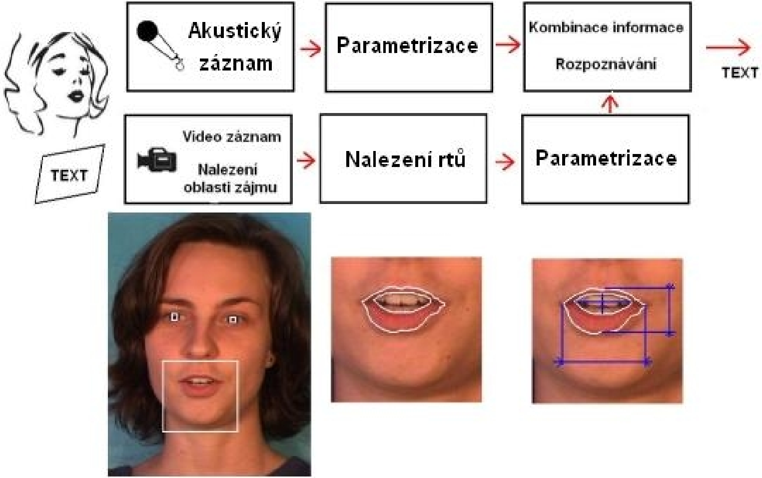
- Note that the same orthonormal basis vector ϕ optimizes the signal-to-noise ratio. And this technique is called **Blind Source Separation (BSS).**

Audio



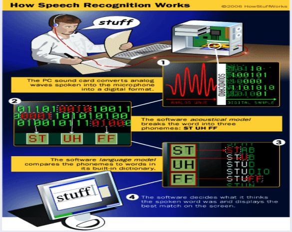
(adapted from AT&T Lab Inc. - http://www.research.att.com/viewProject.cfm?prjID=49)

Audio-Visual



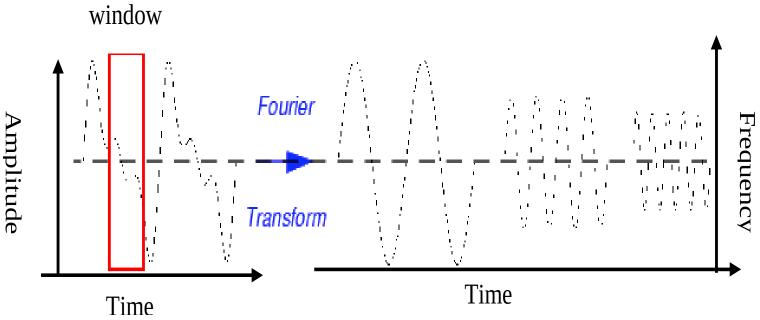
(adapted from Project MUSSLAP - http://musslap.zcu.cz/img/audiovizualni-zpracovani-reci/schema.jpg)

Commercial Applications



A Possible Mathematical Approach

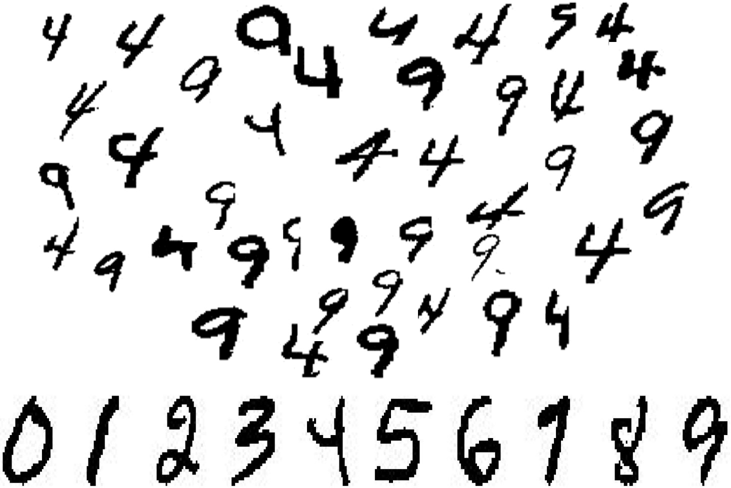
- Continuous $F(\omega) = \int_{-\infty}^{\infty} f(t) e^{i\omega t}\, dt$
- Discrete $f(\omega) = \sum_{k \in \mathbb{Z}} c_k e^{ik\omega}$



Discrete Fourier Transform (DFT)

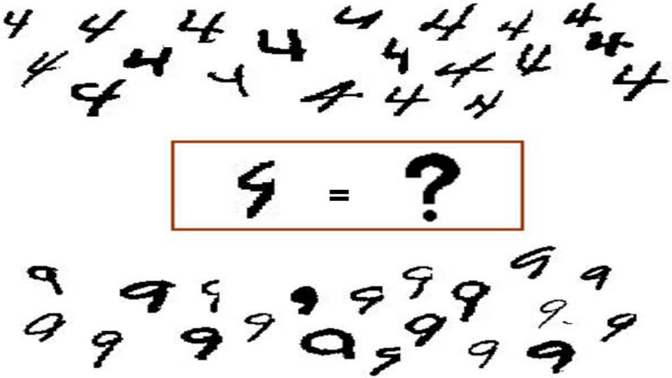
- Fourier analysis is applied to speech waveform in order to discover what frequencies are present at any given moment in the speech signal with time on the horizontal axis and frequency on the vertical.
- The speech recognizer has a database of several thousand such graphs (called a codebook) that identify different types of sounds the human voice can make.
- The sound is “identified” by matching it to its closest entry in the codebook.

Handwritten Digits



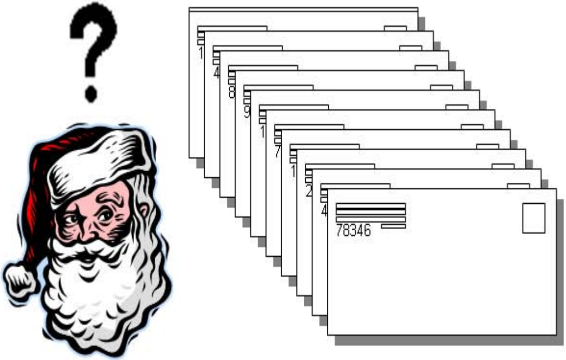
Handwritten Digit Classification

How do we tell whether a new digit is a 4 or a 9?



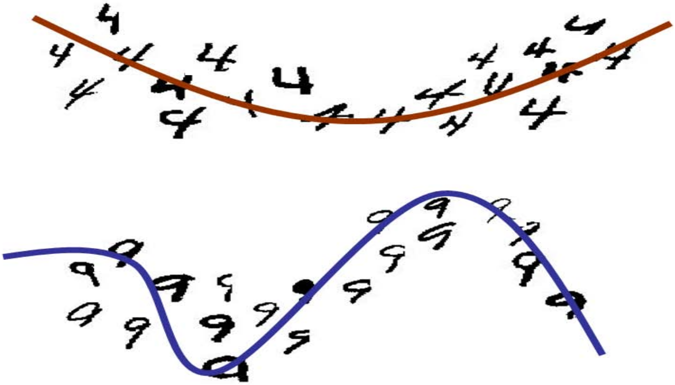
Commercial Applications

Santa thought to himself, "only if these mails can go to the right place according to their zip code".



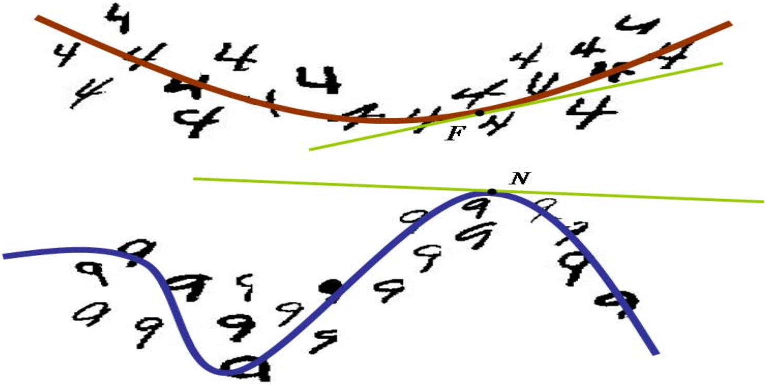
A Possible Mathematical Approach

Imagine a high-D surface (red curve) where all 4's live on and a high-D surface (blue curve) where all 9's live on.

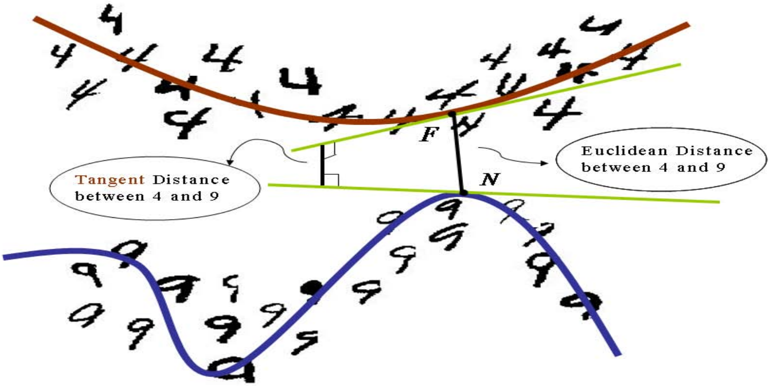


Manifold Learning

Create a Tangent Space of the 4's at F and create a Tangent Space of the 9's at N.

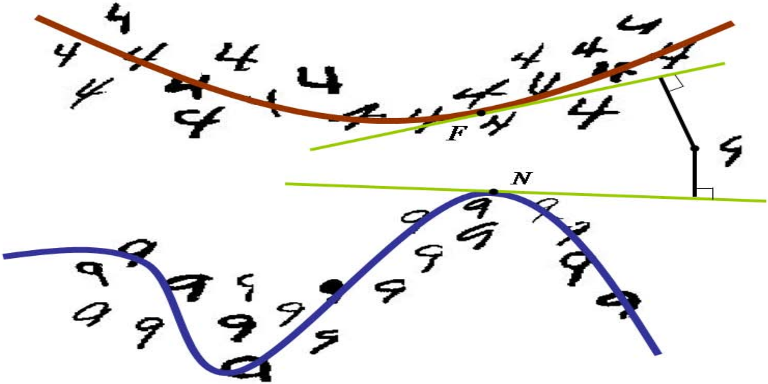


Which Distance?



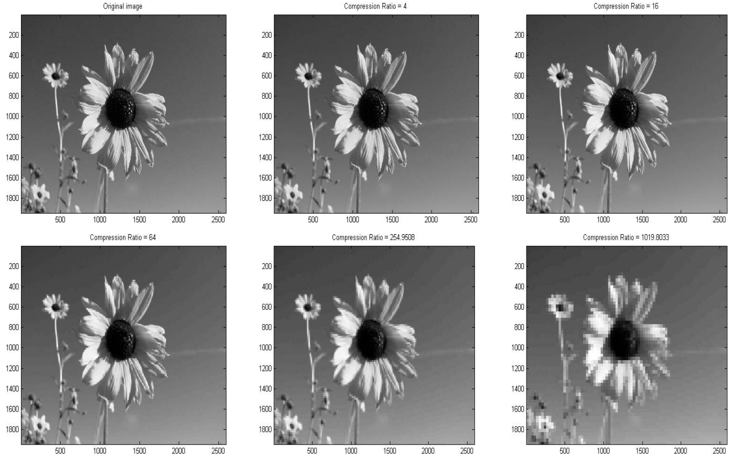
Classification

So, is it a 4 or a 9?



High-Resolution Image

The objective of image compression is to reduce redundancy of the image data in order to be able to store or transmit data in an efficient form.



Low-Resolution Image

Compression with SVD

If we know the correct rank of A, e.g., by inspecting the singular values, then we can **remove the noise and compress the data** by approximating A by a matrix of the correct rank. One way to do this is to truncate the singular value expansion:

Theorem

(Approximation Theorem) If

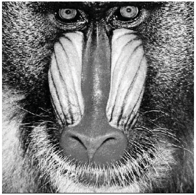
$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T \quad (1 \leq k \leq r)$$

then

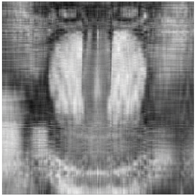
$$A_k = \min_{\text{rank}(B)=k} \|A - B\|_2 \quad \textit{and} \quad A_k = \min_{\text{rank}(B)=k} \|A - B\|_F .$$

Compression with SVD

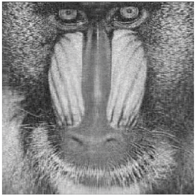
The error term of rank k approximation is given by the $(k+1)^{\text{th}}$ singular value σ_{k+1}.



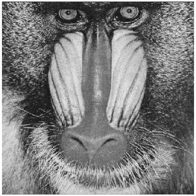
(a) full rank (rank 480)



(b) rank 10, rel. err. = 0.0551



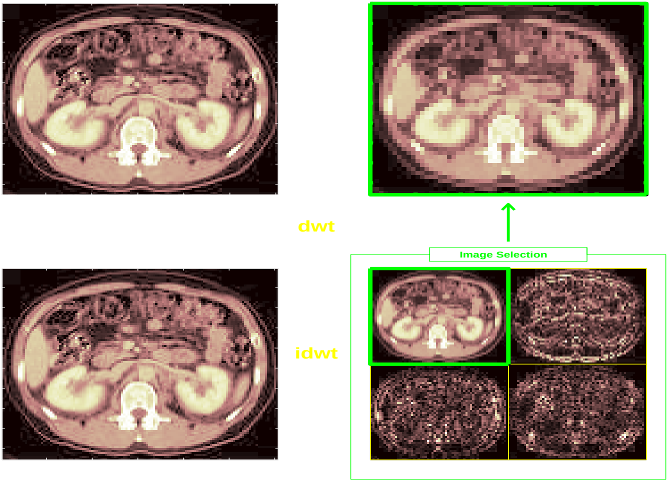
(c) rank 50, rel. err. = 0.0305



(d) rank 170, rel. err. = 0.0126

Compression with DWT

$$X(b,a) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t)\Psi^{*}\left(\frac{t-b}{a}\right)\,dt$$



Convolution as Filters

$$f(x,y) \star s(x,y) = \sum_{m=-a}^{a} \sum_{n=-b}^{b} s(m,n) f(x-m, y-n)$$

s

0	0	0	0	0
0	0	0	0	0
0	0	1	0	0
0	0	0	0	0
0	0	0	0	0

f

1	2	3
4	5	6
7	8	9

zero-padded s

0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

rotated f

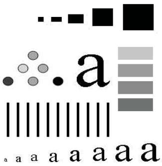
9	8	7
6	5	4
3	2	1

cropped result

0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	1	2	3	0	0
0	0	4	5	6	0	0
0	0	7	8	9	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0

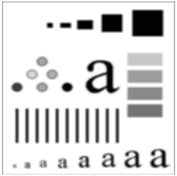
Smoothing with Low-pass Filters

Filtering with $k \times k$ low-pass filters $\frac{1}{k^2}\begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}$.



(a) original

(b) 3×3



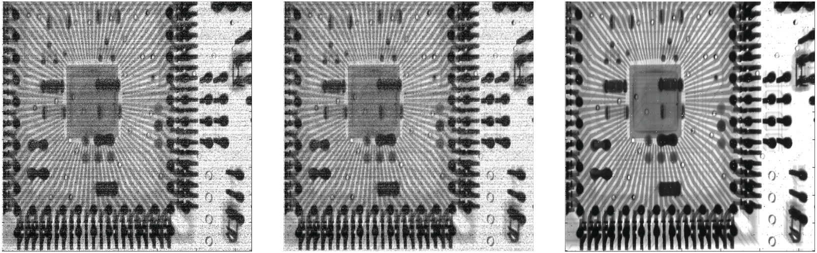
(c) 5×5

(d) 9×9

(e) 15×15

(f) 35×35

Smoothing with Median Filter



(a) (b) (c)

Figure: (a) X-ray image of circuit board corrupted by salt-and-pepper noise. (b) Noise reduction with a 3×3 averaging filter. (c) Noise reduction of a 3×3 median filter.

Sharpening with High-pass Filters

- The simplest *isotropic* filter (direction independent) derivative operator is the discrete Laplacian of two variables:

$$\nabla^2 f(x,y) = f(x+1,y) + f(x-1,y) + f(x,y+1) + f(x,y-1) - 4f(x,y).$$

- This equation can be implemented using the filter mask

$$\begin{bmatrix} & (x,y-1) & \\ (x-1,y) & (x,y) & (x+1,y) \\ & (x,y+1) & \end{bmatrix} \longrightarrow \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$