PROXIMITY/DISTANCE MEASURES- PART 2

Dr. Umarani Jayaraman
Assistant Professor

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY,
DESIGN AND MANUFACTURING,
KANCHEEPURAM

Topic

- Normalized Euclidean Distance
- Quadratic Form Distance

L_p Norm

$$L_p(X,Y) = \left(\sum_{i=1}^{d}(|\, x_i - y_i \,|)^p\right)^{\frac{1}{p}}$$

- Where p=1,2,..., ∞ and d is the dimension
- Depending on the value of p, we get different distance measures.
  - L_2 : Euclidean
  - L_1 : Manhattan (city block distance)
  - L_∞ : Max (chess board distance)
  - $L_{-\infty}$: Min
- This is also called as *Minkowski Norm*

L_2 Norm/Euclidean Distance

- When $p = 2$, in L_p norm, we get the *Euclidean distance.*
- This is also called the L_2 norm.

$$D(X,Y) = \left(\sum_{i=1}^{d} | x_i - y_i |^2 \right)^{\frac{1}{2}}$$

Normalized Euclidean Distance

- When $p = 2$, in L_p norm, we get the *Euclidean distance*.
- This is also called the L_2 norm.

$$NED(X,Y) = \left(\sum_{i=1}^{d} | x_i' - y_i' |^2\right)^{\frac{1}{2}}$$

Each dimension is mean-centered and normalized

$$x_i' = (x_i - \mu_i)/\sigma_i$$

μ_i and σ_i are the mean and standard deviation of dimension i for all data, i.e., the i^{th} row of **D**

Metric? Yes

Quadratic form distance

$$d_Q(x, y) = \sqrt{(x-y)^T A (x-y)}$$

- Quadratic form distance is a cross bin distance
- It specifies cross-dependencies of the dimensions
- It allows comparison of histograms across different bin locations

Quadratic form distance

$$d_Q(x, y) = \sqrt{(x-y)^T A (x-y)}$$

- An example in case of 2-D vectors $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ $y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$

$$[(x_1 - y_1) \quad (x_2 - y_2)] \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} (x_1 - y_1) \\ (x_2 - y_2) \end{bmatrix} \longrightarrow \text{Scaler value}$$

Quadratic form distances

$$[(x_1-y_1) \quad (x_2-y_2)]\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}\begin{bmatrix} (x_1-y_1) \\ (x_2-y_2) \end{bmatrix}$$

$$= [(x_1-y_1) \quad (x_2-y_2)]\begin{bmatrix} A_{11}(x_1-y_1) + A_{12}((x_2-y_2) \\ A_{21}(x_1-y_1) + A_{22}(x_2-y_2) \end{bmatrix}$$

$$= (x_1-y_1)A_{11}(x_1-y_1) + (x_1-y_1)A_{12}(x_2-y_2) + (x_2-y_2)A_{21}(x_1-y_1) + (x_2-y_2)A_{22}(x_2-y_2)$$

$$= A_{11}(x_1-y_1)^2 + A_{12}(x_1-y_1)(x_2-y_2) + A_{21}(x_2-y_2)(x_1-y_1) + A_{22}(x_2-y_2)^2$$

Quadratic form distance

$$d_Q(x, y) = \sqrt{(x-y)^T A (x-y)}$$

- A is similarity matrix of size dxd
- A_{ij} denotes the similarity (or weight) of dimension i with dimension j
- Note: A is positive semi-definite (for distance to be ≥ 0)

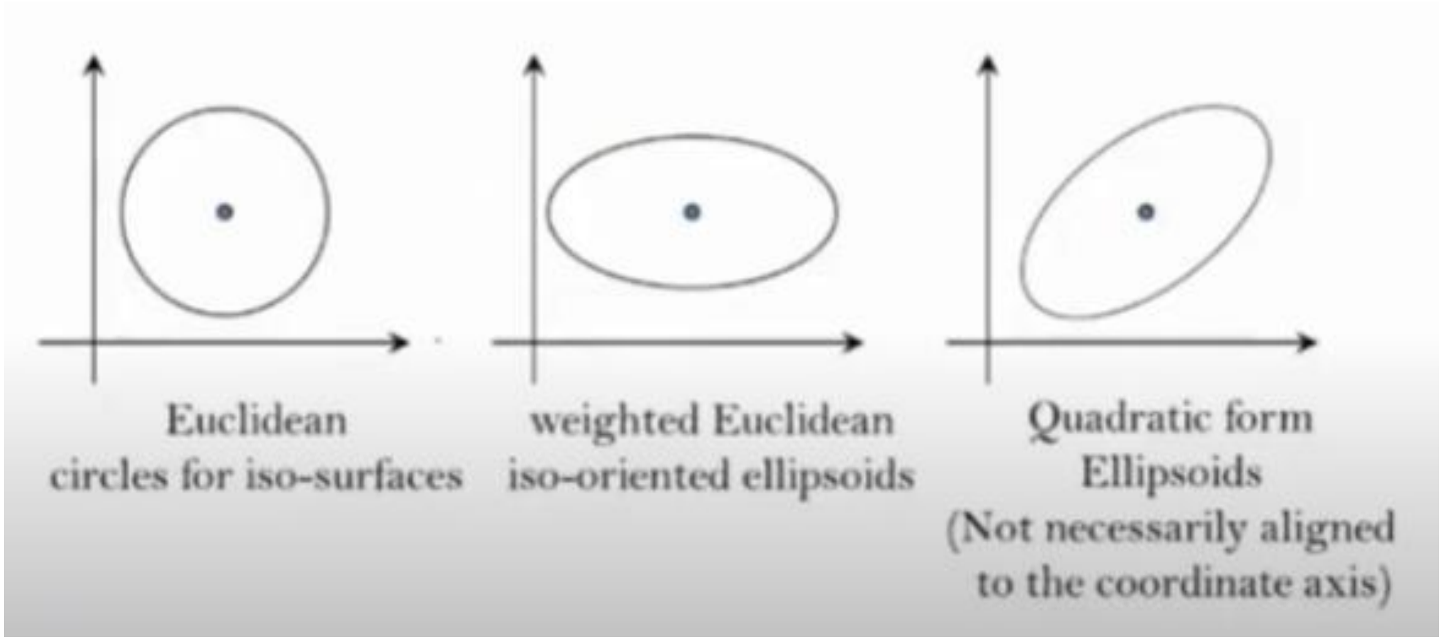
Quadratic form distance

$$d_Q(x,y) = \sqrt{(x-y)^T A (x-y)}$$

- For example $A_{ij}=1-c_{ij}/c_{max}$ for color histograms
- c_{ij} is bin-to-bin distance and c_{max} the maximum distance
- Note
  - If A is an identity matrix, then Euclidean
  - If A is a diagonal matric, then weighted Euclidean
  - **Is it a Metric? Yes , if A is positive definite**

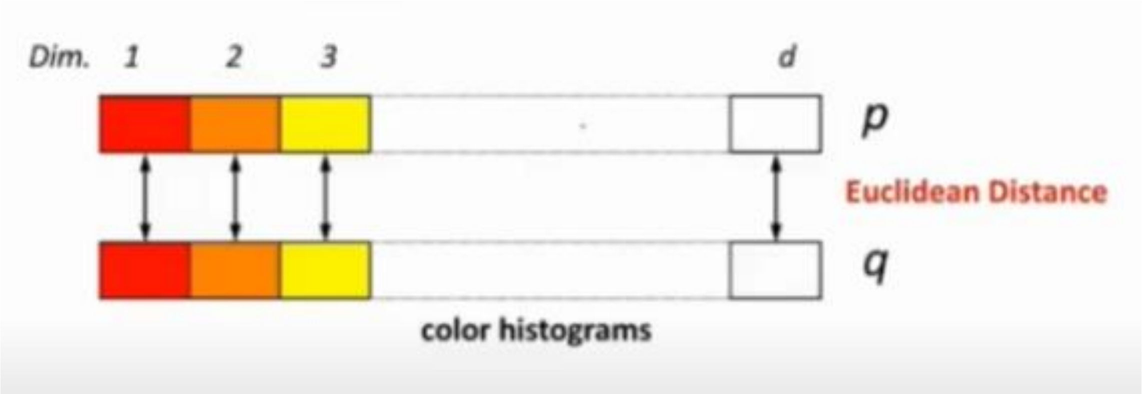
Quadratic form distance

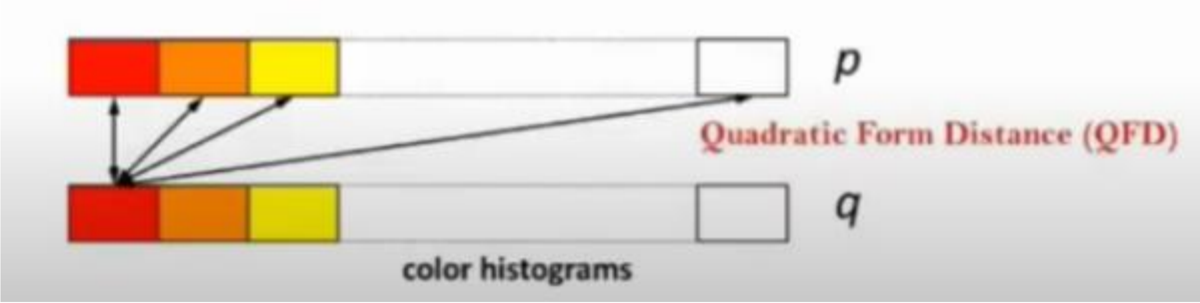
- QFD represents correlation between dimensions



Application of QFD

- Comparison of color histograms
  - Considers similarity between colors i and j

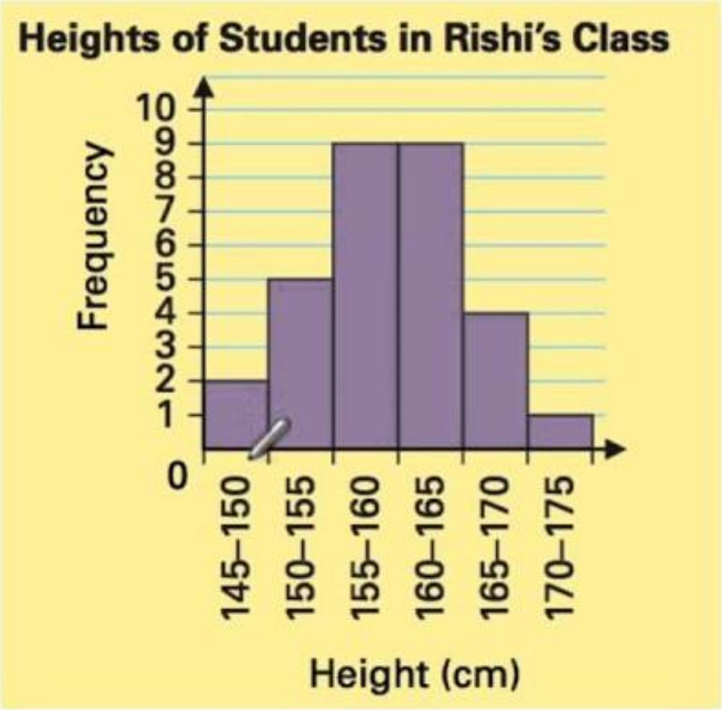




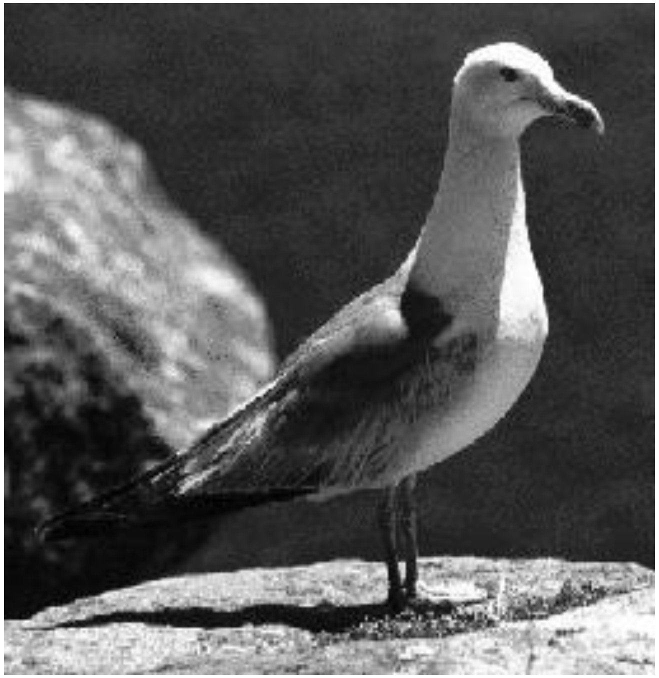
Application of QFD

- Example of histogram

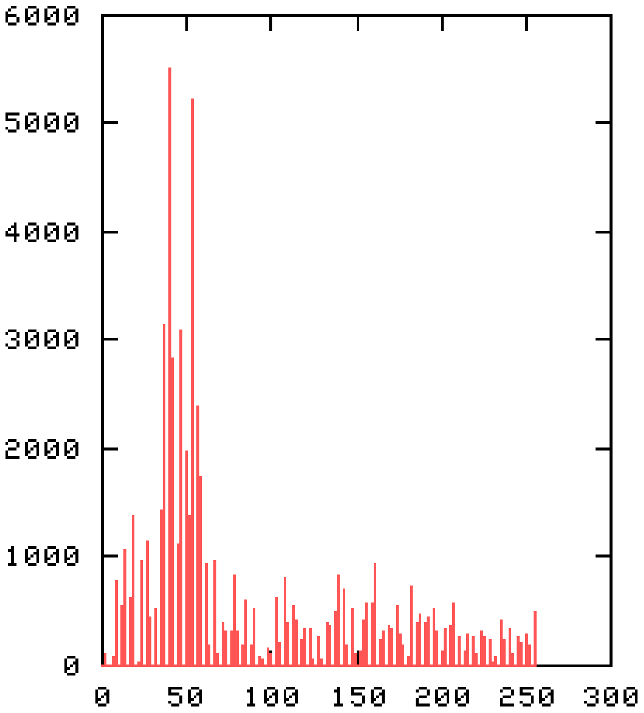
Height (cm)	Frequency
145–150	2
150–155	5
155–160	9
160–165	9
165–170	4
170–175	1



Application of QFD



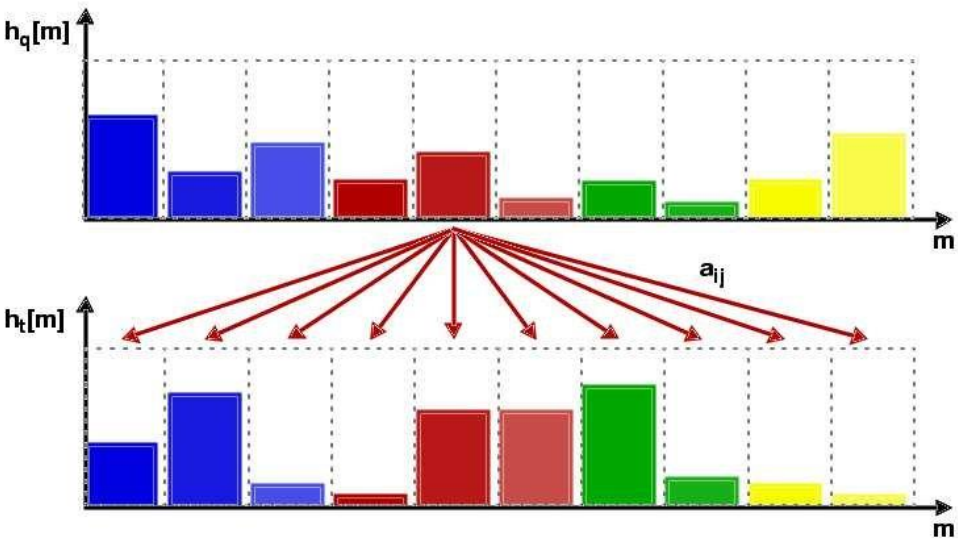
Image



Histogram of the given image

Comparison of color histograms

- How to find the similarity weights?
  - Distance between the bins
- Bin Distance = difference of indices



Comparison of color histograms

- *Let* $d(i,j)$ represents distance of bin i and j
- Similarity matrix can be computed as

$$A_{ij} = e^{-\sigma . d(i,j)}$$

where the parameter σ controls the global shape of the similarity matrix

Comparison of color histograms

Test Image

Database Image 1

Database Image 2

Mahalanobis distance

- We know the quadratic form distance

$$d_Q(x, y) = \sqrt{(x - y)^T A (x - y)}$$

- Replace A in quadratic form distance by inverse of covariance matrix Σ to get **Mahalanobis distance**

$$d_M(x, y) = \sqrt{(x - y)^T \Sigma^{-1} (x - y)}$$

Quadratic form distance

$$d_Q(x, y) = \sqrt{(x - y)^T A (x - y)}$$

- For example $A_{ij}=1-c_{ij}/c_{max}$ for color histograms
- c_{ij} is bin-to-bin distance and c_{max} the maximum distance
- Note
  - If A is an identity matrix, then Euclidean
  - If A is a diagonal matric, then weighted Euclidean
  - If A is an inverse of covariance matrix, then Mahalanobis distance
  - **Is it a Metric? Yes , if A is positive definite**

Summary

- Metric distance measure
- Quadratic form distance
- Application of QFD

THANK YOU