Utrecht University

Shape Matching
Part 2: Feature-based

Remco Veltkamp
Center for Geometry, Imaging, and Virtual Environments (GIVE)
Utrecht University
www.cs.uu.nl/centers/give

Two step approach

- Step 1: feature extraction
  result: a vector of numerical values
- Step 2: compute distance between 2 vectors
  result: single distance value

Elementary descriptors

- Area A
- Perimeter l
- Compactness $c = l^2/(4\pi A)$
- Circularity, roundness $1/c$
- Centroid (center of mass)
- Major and minor axes λ_1, λ_2
- Eccentricity $||\lambda_1||/||\lambda_2||$
- Minimal bounding box area $A_m = h\, b$
- Rectangularity A/A_m

Moments

- Object uniquely defined by infinite sequence of moments $M_{p,q}$:

$$m_{p,q} = \int_{Obj} x^p y^q \, dx \, dy$$

- In terms of pixels [1,n]x[1,m] image *f(x,y)*:

$$m_{p,q} = \sum_{x=1}^{n} \sum_{y=1}^{m} f(x,y) \, x^p y^q$$

Moment invariants

- Translation invariant: central moments

$$\mu_{p,q} = \int_{Obj} \left(x - \frac{m_{1,0}}{m_{0,0}} \right)^p \left(y - \frac{m_{0,1}}{m_{0,0}} \right)^q dx\,dy$$

- Invariant under uniform scaling with factor α :

$$\eta_{p,q} = \frac{\mu_{p,q} / \alpha^{p+q+2}}{\mu_{0,0}^{(p+q+2)/2}}$$

Moment invariants

Rotation invariant:

$\varphi_1 = \mu_{20} + \mu_{02}$

$\varphi_2 = (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2$

$\varphi_3 = (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{03})^2$

$\varphi_4 = (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2$

$\varphi_5 = (\mu_{30} - 3\mu_{12}) + (\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] +$
$(3\mu_{21} - \mu_{03}) + (\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2]$

$\varphi_6 = (\mu_{20} - \mu 02)\ [(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] + 4\mu_{11}\ (\mu_{30} + \mu_{12})(\mu_{21} + \mu_{03})$

Reflection and rotation invariant:

$\varphi_7 = (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] +$
$(\mu_{30} - 3\mu_{12})(\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2]$

Modal Matching

- Take n samples along contour
- Matrix $D=(d_{ij})$ describes interaction between points i and j
- Determine n modes, or eigenshapes, the eigenvectors of D:

$$D\, e_i = \lambda e_i$$

Modal Matching

- n modes e_i of query, n modes e_j' of target
- Some dissimilarity measures between modes: $m(e_i,e_j')$
- For fixed i_0, determine value j_0 of j for which $m(e_{i0},e_j')$ is minimal
- If i for which $m(e_i,e_{j0}')$ is minimal is i_0, then point i of query and point j of target match

Fourier Descriptors

- FD of some shape signature:
  - Complex Coordinates: z(t)
  - Central Distance: r(t)
  - Chordlength: r*(t)
  - Curvature: K(t)
  - Cumulative Angles: φ(t)
  - Area function: A(t)

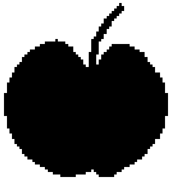
Complex Coordinates

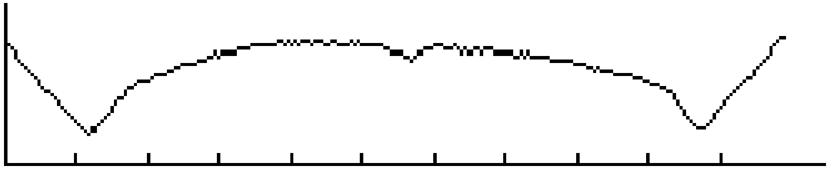
$$z(t) = [x(t) - x_c] + i[y(t) - y_c]$$

$$x_c = \frac{1}{N}\sum_{t=0}^{N-1} x(t), \quad y_c = \frac{1}{N}\sum_{t=0}^{N-1} y(t)$$

Central Distance

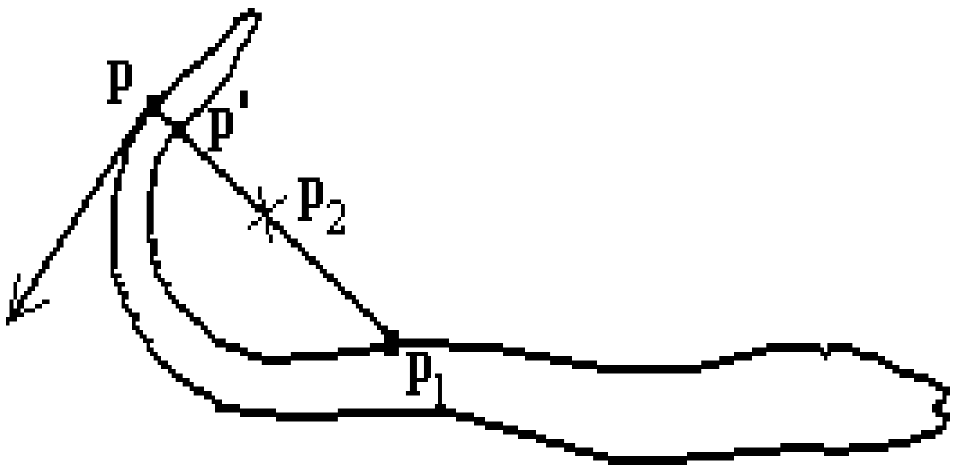
$$r(t) = ([x(t) - x_c]^2 + [y(t) - y_c]^2)^{1/2}$$





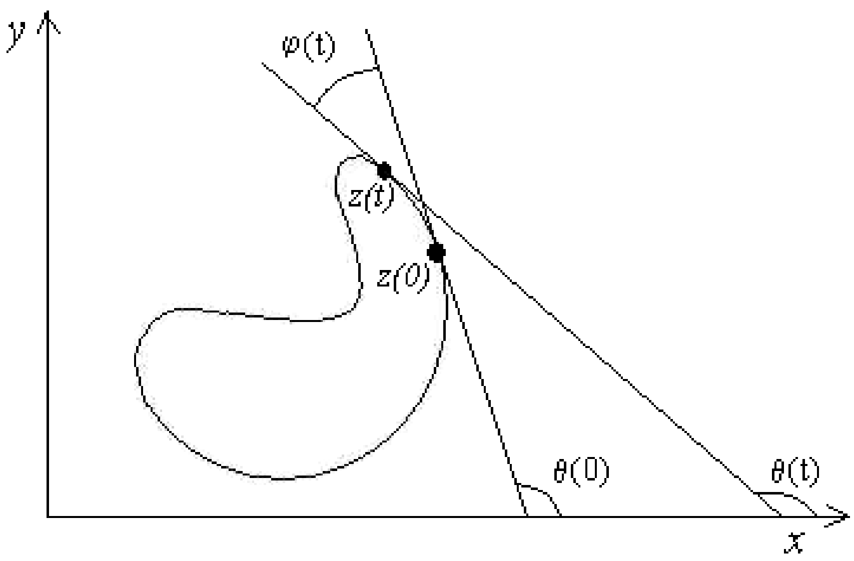
Chordlength

r*(t) = length of chord in object perpendicular to tangent at *p*, as a function of *p*



Cumulative Angular Function

- Also called turning angle function
- $\varphi(t) = [\theta(t) - \theta(0)]\mathrm{mod}(2\pi)$

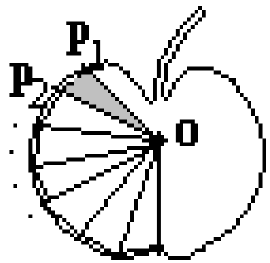


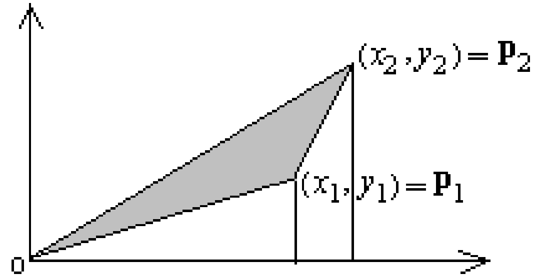
Curvature Function

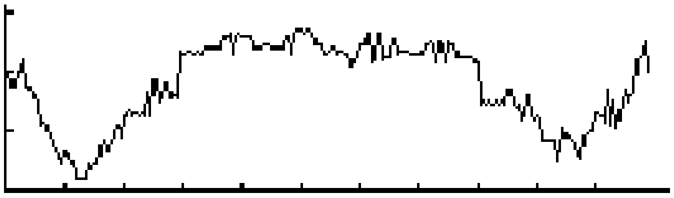
- $K(t) = \theta(t) - \theta(t\text{-}1)$
- $$\theta(t) = \arctan \frac{y(t) - y(t+w)}{x(t) - x(t+w)}$$
- w is jumping step in selecting next pixel

Area Function

$$A(t) = \frac{1}{2} \left| x_1(t) y_2(t) - x_2(t) y_1(t) \right|$$







Fourier Descriptors

- **Fourier transform of the signature *s*(*t*)**

$$u_n = \frac{1}{N}\sum_{t=0}^{N-1} s(t)\exp\left(\frac{-j2\pi nt}{N}\right)$$

- u_n, $n = 0, 1, \ldots, N-1$, are called FD denoted as FD_n

- **Normalised FD**

$$\mathbf{f} = \left[\frac{|FD_1|}{|FD_0|}, \frac{|FD_2|}{|FD_0|}, \ldots, \frac{|FD_m|}{|FD_0|}\right]$$

Where *m*=*N*/2 for central distance, curvature and angular function
***m*=*N* for complex coordinates**

FD Convergence Speed

- Finite number of coefficients are used to approximate the signal. Partial Fourier sum of degree *n* of *u*(*t*) is given by

$$(S_n u)(t) = \sum_{|k|\leq n} \hat{u}(k) e^{jkt}$$

- For piecewise smooth function *u*(*t*), there exists one-to-one correspondence between *u*(*t*) and the limit of their Fourier series expansion

$$\lim_{n\to\infty}(S_n u)(t) = u(t)$$

- For shape retrieval application, number of coefficients to represent shape should not be large, therefore, the convergence speed of the Fourier series derived from the signature function is crucial

FD Convergence Speed

- How fast get the FD coefficients below a threshold?

Signature functions	Number of normalized FD coefficients > 0.1	Number of normalized coefficients > 0.01	
r(t)	15	120	
r*(t)	40	360	
A(t)	20	210	
z(t)	10	50	←fast
ψ(t)	40	280	
κ(t)	∞	∞	←slow

FD Matching

- Similarity between a query shape and a target shape in the database is

$$d = \left(\sum_{i=1}^{m} (f_i^q - f_i^t)^2\right)^{1/2}$$

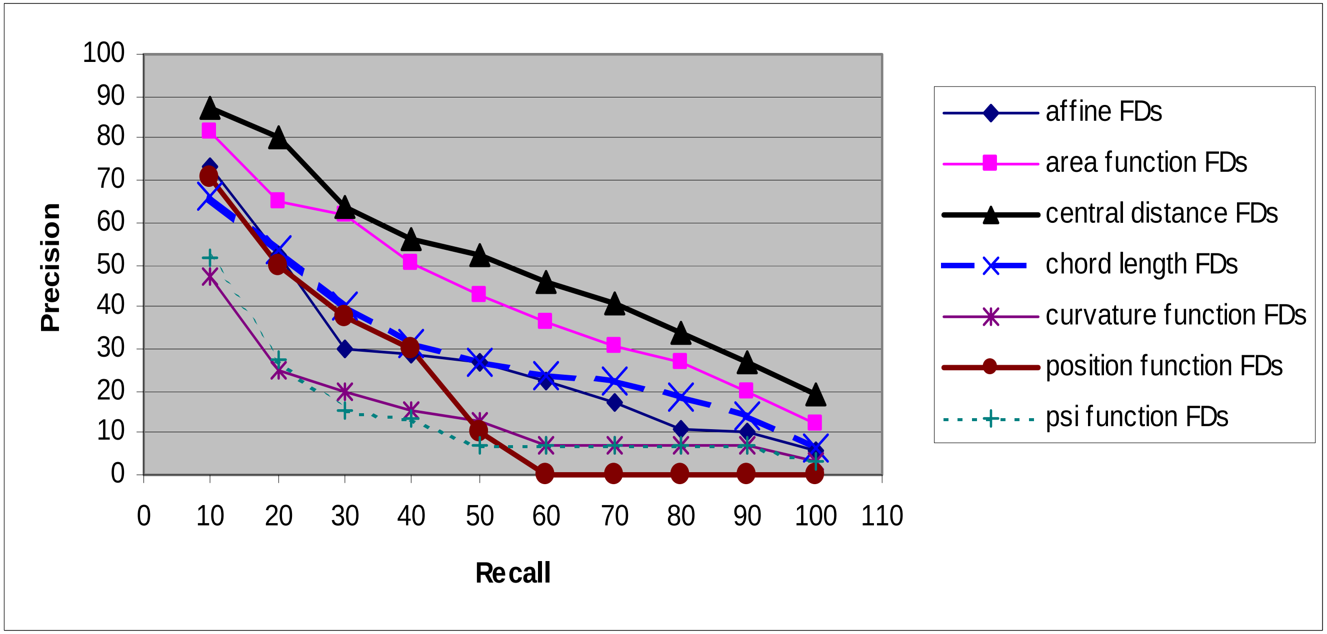
$$where\ \mathbf{f}_q = (f_q^1, f_q^2, \ldots, f_q^m)\ and\ \mathbf{f}_t = (f_t^1, f_t^2, \ldots, f_t^m)\ are\ the$$

$$feature\ vectors\ of\ the\ two\ shapes\ respectively$$

FD Performance

- Precision P is the ratio of the number of relevant retrieved shapes r to the total number of retrieved shapes n
- Recall R is the ration of the number of relevant retrieved shapes r to the total number m of relevant shapes in the whole database

$$P = \frac{r}{n} \qquad R = \frac{r}{m}$$



Local descriptors

To process the shape locally on salience: search for

local maximum {s(t)}

Supported by perceptual evidence that:

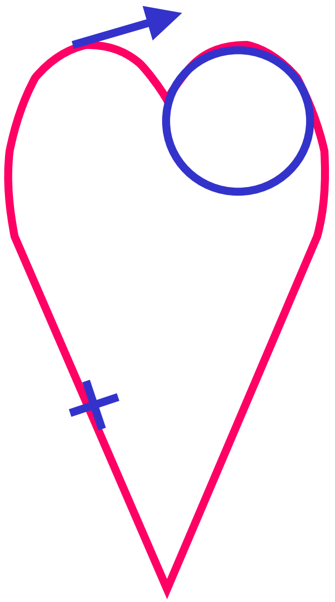
- humans focus on high curvature points
- humans keep the memory of figure by a few, salient points
- humans are very poor in quantitative measurement of overall shape

Local descriptors

Location: $\mathbf{x}(t)$

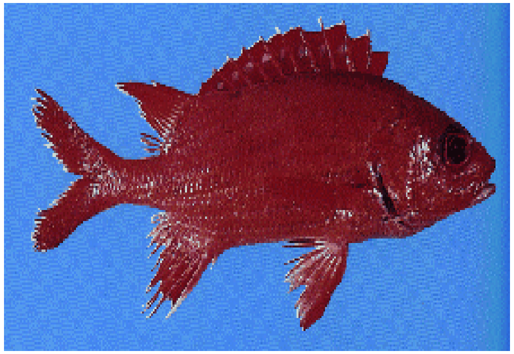
Tangent: $\phi(t) = \mathbf{x}'(t)$

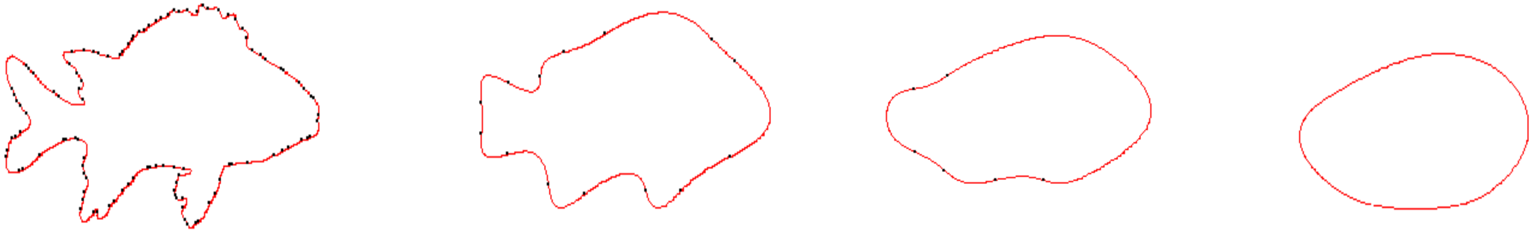
Curvature: $\kappa(t) = \phi'(t) = \mathbf{x}''(t)$



Curvature Scale Space

- Smooth curvature along contour





Curvature Scale Space

- Contour *C(s)=(x(s),y(s))*
- Convolution with Gaussian kernel of width σ:

$$x_\sigma(s) = x(s) * \phi(t) = \int x(s)\phi(t-s)\,dt$$

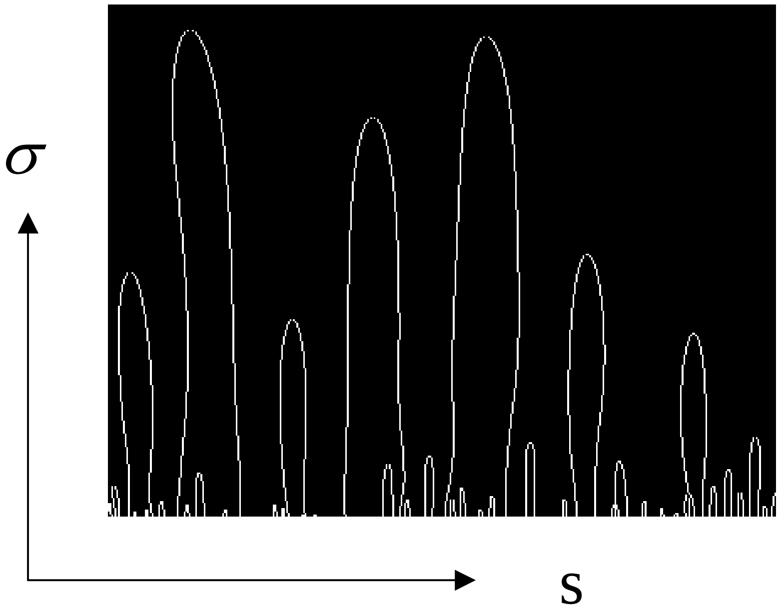
$$\phi_\sigma(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{t^2}{2\sigma^2}}$$

same for *y(s)*

Curvature Scale Space

- Increasing σ: positions of zero crossings move together, then annihilate
- Number of zero crossings decreases until contour is convex
- Matching: match points of annihilation in *(s, σ)*-plane:

http://www.ee.surrey.ac.uk/Research/VSSP/imagedb/demo.html



Step 2: Matching Feature Vectors

- Result of feature extraction: numerical values $x_1,\ldots,x_n$ assembled in a "feature vector" $x=(x_1,\ldots,x_n)$
- Example: 64 values for hue histogram, 8 for edge directions histogram, 16 for wavelet coefficients, 16 for Fourier coefficients of contour $\Rightarrow$ n=104
- n-dimensional feature space

L_p metrics

- Also called Minkowski distance
- $x=(x_1,\ldots,x_n)$, $y=(y_1,\ldots,y_n)$

$$L_p = \left(\sum_{i=1}^{n} | x_i - y_i |^p \right)^{1/p}$$

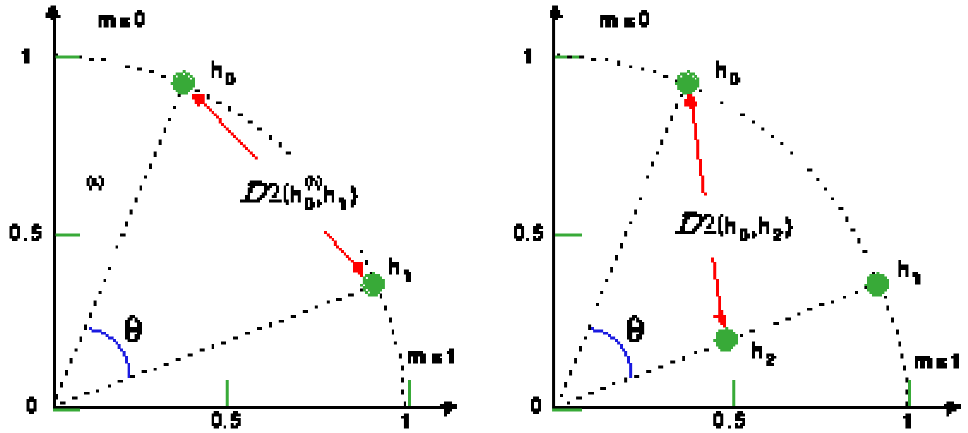
- Metric for $p \geq 1$ (otherwise no triangle inequality)
- L_1: taxicab, city block, Manhattan, rectilinear distance
- L_2: Euclidean distance
- L_∞: max, chess board distance, $\max_i |x_i - y_i|$

Cosine Distance

- $x=(x_1,\ldots,x_n),\ y=(y_1,\ldots,y_n)$

$$d(x,y) = 1 - \cos(\angle(x,y)) = 1 - \frac{x\,y}{|x|\,|y|}$$

- Metric
- Only angle relevant, not vector lengths



Quadratic Form Distance

$$d(x,y)=\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i-y_i|\,w_{ij}\,|x_j-y_j|=|x-y|^{\mathrm{T}}\,W\,|x-y|$$

Metric if $w_{ij}=w_{ji}$ and $w_{ii}=1$

Earth Mover's Distance

- $A=\{(x_i,w_i)\}, \sum w_i=W, B=\{(y_j,u_j)\}, \sum u_j=U$
- f_{ij} flow from x_i to y_j over d_{ij}
- $f_{ij}\geq 0$
- $\sum_j f_{ij} \leq w_i$
- $\sum_i f_{ij} \leq u_i$
- $\sum_i \sum_j f_{ij} = \min(W,U)$

$$EMD(A,B) = \frac{\min_F \sum_{ij} f_{ij} d_{ij}}{\min(W,U)}$$

Properties EMD

- Invariant under rigid motion
- Respects scaling
- Metric if d metric, and $W=U$
- If $W \neq U$:
  - No positivity, surplus not taken into account
  - No triangle inequality

Proportional Transportation Dist

- $A=\{(x_i,w_i)\}$, $\sum w_i=W$, $B=\{(y_j,u_j)\}$, $\sum u_j=U$
- f_{ij} flow from x_i to y_j over d_{ij}
- $f_{ij} \geq 0$
- $\sum_j f_{ij} = w_i$
- $\sum_i f_{ij} \leq u_i W/U$
- $\sum_i \sum_j f_{ij} = W$

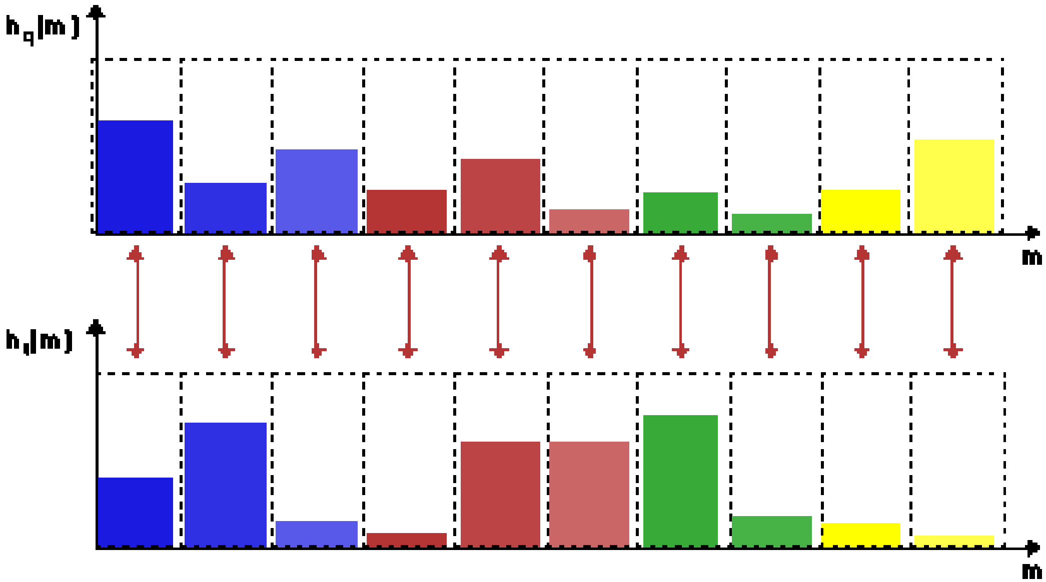
$$PTD(A,B) = \frac{\min_F \sum_{ij} f_{ij} d_{ij}}{W}$$

Properties PTD

- Invariant under rigid motion
- Respects scaling
- PTD is pseudo-metric:
  - Triangle inequality holds
  - No positivity
    - *but only when same relative weights*
    - *surplus taken into account*

Histogram Matching

- Histogram seen as feature vector
  e.g. $d(H_1,H_2)$ is Euclidean distance

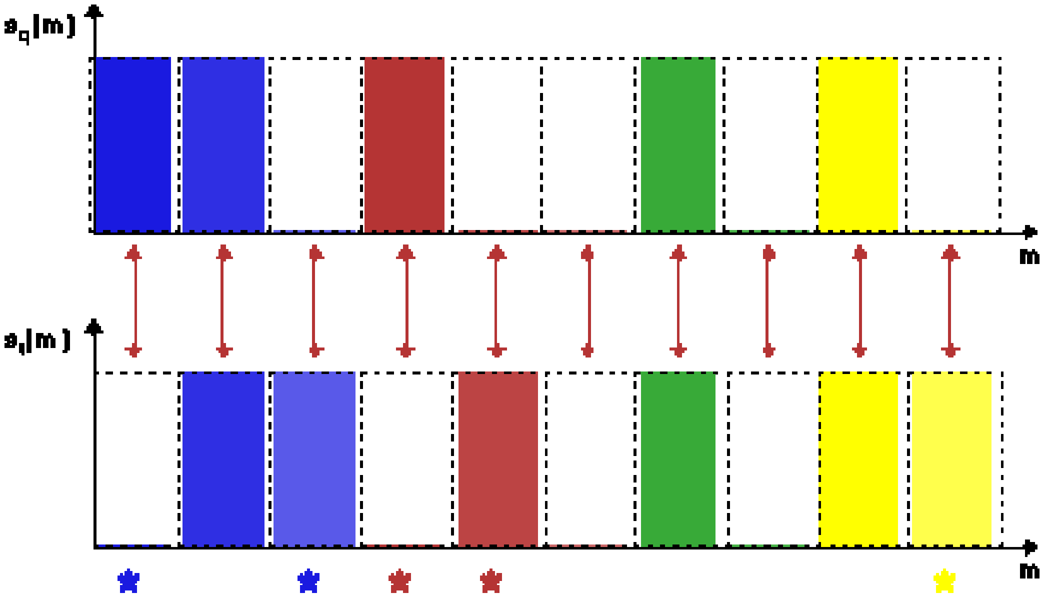


Hamming Distance

- Binary vectors/histograms

$$d(H_1, H_2) = \sum_{i=1}^{n} | H_1[i] - H_2[i] | = \sum_{i=1}^{n} (H_1 \ XOR \ H_2)$$

- Normalization: divide by *n*



Histogram Intersection

- To check occurrence of object in region
- Typically $\sum H_{obj}[i] < \sum H_{reg}[i]$
- Non-metric form (not symmetric):

$$d(H_{obj}, H_{reg}) = 1 - \frac{\sum_{i=1}^{n} \min(H_{obj}[i], H_{reg}[i])}{\sum_{i=1}^{n} H_{obj}[i]}$$

Histogram Intersection

- Metric form:

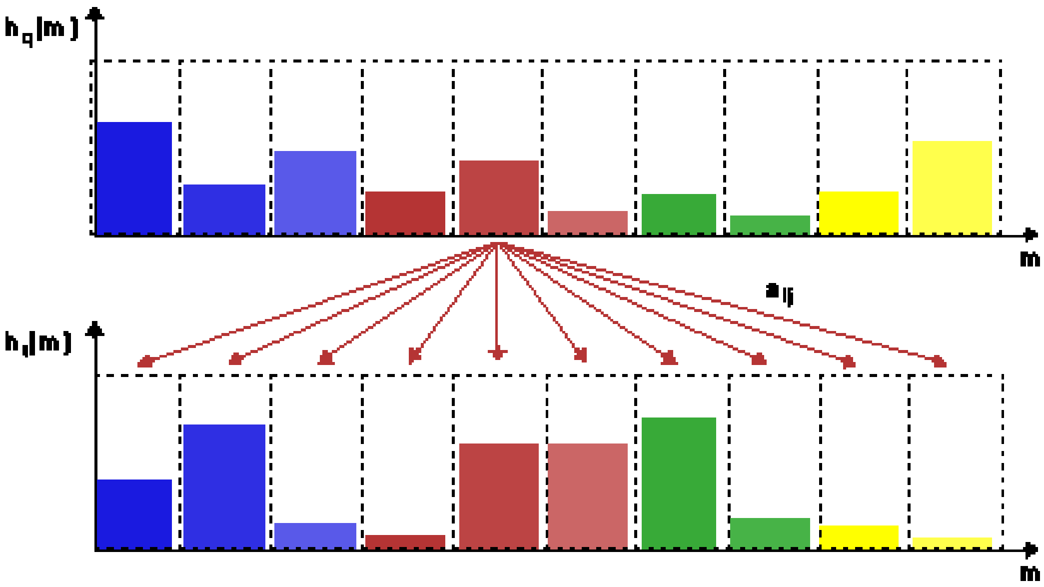
$$d(H_{obj}, H_{reg}) = 1 - \frac{\sum_{i=1}^{n} \min(H_{obj}[i], H_{reg}[i])}{\min\left(\sum_{i=1}^{n} H_{obj}[i], \sum_{i=1}^{n} H_{reg}[i]\right)}$$

- If $\sum H_{obj}[i] = \sum H_{reg}[i]$
  $d(H_{obj},H_{reg})=L_1(H_{obj},H_{reg})$

Histogram Matching

- Cross bin matching:

$$d(H_1, H_2) = \left(\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \, | H_1[i] - H_2[j] |^p \right)^{1/p}$$



2D Histogram Matching

- Instead of (2*n* terms)

$$\sum_{i=1}^{n} \ldots H_1[i] \ldots H_2[i] \ldots$$

do (n^2 terms)

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \ldots H_1[i, j] \ldots H_2[i, j] \ldots$$

Earth Mover's Distance

On histograms

- $H_1=\{(i,H_1[i])\}, \sum H_1[i]=W,$
  $H_2=\{(i,H_2[i])\}, \sum H_2[i]=U$
- f_{ij} flow from $H_1[i]$ to $H_2[j]$ over $d_{ij}=|i-j|$
- $f_{ij} \geq 0$
- $\sum_j f_{ij} \leq H_1[i]$
- $\sum_i f_{ij} \leq H_2[i]$
- $\sum_i \sum_j f_{ij} = \min(W,U)$

$$EMD(A,B) = \frac{\min_F \sum_{ij} f_{ij} d_{ij}}{\min(W,U)}$$

Kullback-Leibler

- Consider histograms as distibrutions:
  $\Sigma\, H_1[i] = \Sigma\, H_2[i] = 1,\ H_1[i], H_2[i] \geq 0$

$$d(H_1, H_2) = \sum_{i=1}^{n} \left(H_1[i] \log \frac{H_1[i]}{H_2[i]} \right)$$

- No metric: not symmetric

Divergence

- Symmetrized Kullback-Leibler

$$d(H_1, H_2) = d_{KL}(H_1, H_2) + d_{KL}(H_2, H_1)$$
$$= \sum_{i=1}^{n} \left((H_1[i] - H_2[i]) \log \frac{H_1[i]}{H_2[i]} \right)$$

Mahanalobis Distance

- Quadratic form with inverse covariance matrix: d(x,y)= (x-y) $\sum^{-1}$(x-y)
- Of N feature vectors $\mathbf{x}^{\mathbf{1}},\ldots,\mathbf{x}^{\mathbf{N}}$, each of length *n*, compute

$$\mu_i = \frac{1}{N}\sum_{j=1}^{N}\left(\mathbf{x}^{j}\right)_i, i = 1,\ldots,n$$

$$\sigma_{ij} = E((x_i - \mu_i)(x_j - \mu_j)) = \sum_{k,l=1}^{n}\frac{(x^k{}_i - \mu_i)(x^l{}_j - \mu_j)}{n^2}$$

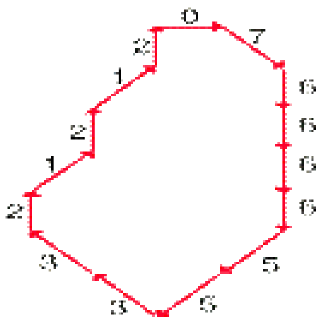
- $\sum = (\sigma_{ij})$

Edit Distance

- Feature interpreted as string of characters
- Edit distance operations
  - Insertion, where an extra character is inserted into the string
  - Deletion, where a character has been removed from the string
  - Transposition, in which two characters are reversed in their sequence
  - Substitution, which is an insertion followed by a deletion

Edit Distance

- Strings with a small edit distance are likely to be similar
- Edit distance is number of edit distance operations from one string to another
- Example: chaincodes
  12345678
  123845677 have distance 3



Edit Distance

- Use dynamic programming
- Given two strings $x = x_1x_2 .. x_n$ and $y = y_1y_2 ..y_m$
- Edit distance $f(i, j)$ is computed as best match of two substrings $x_1x_2 .. x_i$ and $y_1y_2 ..y_j$ where
  - $f(0,0) = 0$
  - $f(i, j) = min[f(i\text{-}1, j) + 1),$
    $f(i, j\text{-}1) +1, f(i\text{-}1, j\text{-}1) + d(x_i,y_j)]$

Linear weighting

- Combine k different feature distances d_i, e.g. color, texture and shape distances
- Linear weighting: (weighted average) $$d = \sum_{i=1}^{k} w_i d_i$$

- Affine weighting: $$\sum_{i=1}^{k} w_i = 1$$
- Convex weighting: $$\sum_{i=1}^{k} w_i = 1, w_i \geq 0$$

Non-linear weighting

- α-trimmed mean: weight only α percent highest of the *k* values
- Rank-based merging: sort values in decreasing order d_1', …, d_k'

$$d = \frac{1}{k}\sum_{j=1}^{k}\frac{1}{j}\sum_{i=1}^{j} d_k'$$