Object description

What is an object descriptor?

- An object is a collection of pixels in an image
- A descriptor is a representation for some property of the group of pixels
  - generally a scalar

Need for a descriptor: quantification, recognition

Desired properties of a descriptor

1. **Completeness** – Two objects can have identical descriptors iff the objects are identical in shape
2. **Congruence** – Similar objects must have similar descriptors
3. **Invariance** – Descriptor must be invariant to scale, rotation, orientation, etc.
4. **Compactness** – Descriptor must be an efficient representation for the object.

   i.e capture the uniqueness with minimal information

Classes of object descriptors

- Shape boundary-based
  - Chain codes
  - Fourier descriptor

- Region-based
  - Basic (area, perimeter,compactness, dispersion)
  - Moments (First order, Centralised, Zernike)

Boundary descriptor 1

{Boundary, region}

- Boundary – A pixel which belongs to the object and has at least one background pixel as a neighbour
  - ➢ The entire set is found by contour following

- Region – A pixel in the <u>interior</u> of the object
  - ➢ Pixel which belongs to the object but not on the boundary

The above will change depending on connectivity rules

Boundary descriptor 2

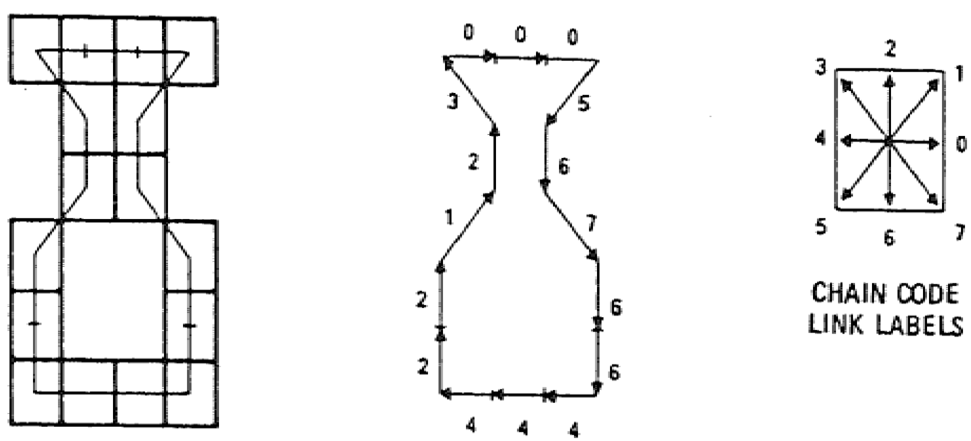
Chain Codes

- A representation for shape of an object

Version1: Store coordinates of boundary pixels

version 2: Store only their relative positions

Unlike version1 version2 preserves order and is more compact



Kuhl and Giardina CGIP 1983

Fourier descriptor

- Fourier theory applied to shapes
  - descriptor encodes the shape in terms of frequency

Approach:

- Treat the locus (trace) of the boundary points as a periodic function ➔ Expand using Fourier series

1-D Fourier transform- review

- N-length sequence $f[n]$ ⇔ DFT $F[k]$

$$F[k] = \frac{1}{N}\sum_{n=0}^{N-1} f[n] e^{-j\frac{2\pi kl}{N}}; \quad k = 0,1..N-1$$

$$f[n] = \sum_{k=0}^{N-1} F[k] e^{j\frac{2\pi kl}{N}} \quad n = 0,1..N-1$$

- f can be real or complex
- F is generally complex

Fourier descriptor – for shapes

Given a set of N boundary points for the shape, with coordinates (x,y])

1. define $b[l]= x[l] + j\, y[l]$
2. Find the DFT of $b[l]$ as

$$B[k] = \frac{1}{N}\sum_{l} b(l)\, e^{-j\frac{2\pi kl}{N}} = \frac{1}{N}\sum_{l} x(l)\, e^{-j\frac{2\pi kl}{N}} + j\,\frac{1}{N}\sum_{l} y(l)\, e^{-j\frac{2\pi kl}{N}}$$

$$k, l = 0,1..N-1$$

$B[k]$ is the Fourier descriptor of the shape *(Elliptic form)*

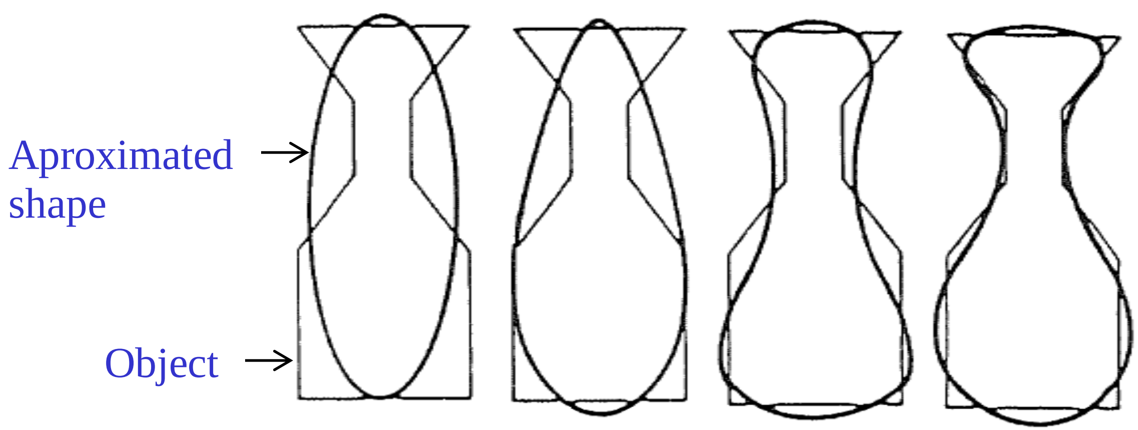
$l = 0$ gives the average value of the boundary points (position/centroid of the shape)

$l > 0$ encodes details

Reconstruction from M<N coeffs.

$$b[l] = \sum_{k=-M/2}^{M/2} F[k] e^{j\frac{2\pi kl}{N}} \qquad l = 0,1..N-1$$

Reconstructed shape with *M =1* to *4*

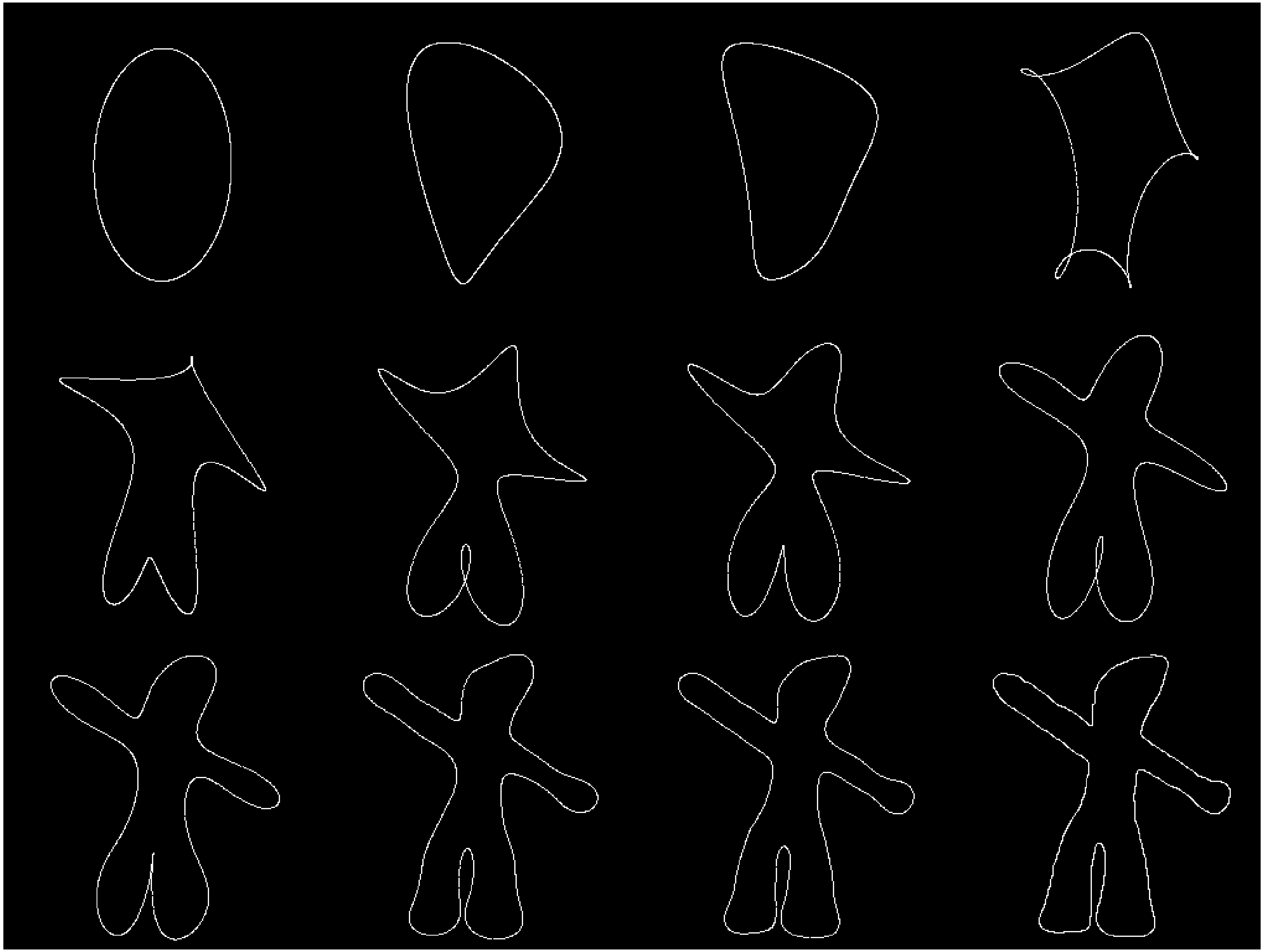


Kuhl and Giardina CGIP 1983

M=1 to 4

M=5 to 7

N=1257



M=10 M=20 M=30 M=1257

From http://fourier.eng.hmc.edu/e161/lectures/fd/node1.html

Variant of Fourier Descriptor

$B[k]$ is **not** invariant to scale, translation, rotation and starting point shift

Ex. Translation by $b_0 \rightarrow B[k] + b_0\delta[k]$

Why? (check by deriving it mathematically)

To achieve these, one can

- drop the phase and the $l = 0$ term
- do a normalisation

Normalised magnitude form of FD

$$C[k] = \frac{|B[k]|}{\max\{|B[k]|\}} \quad k = 1, (N-1), 2, (N-2)\ldots$$

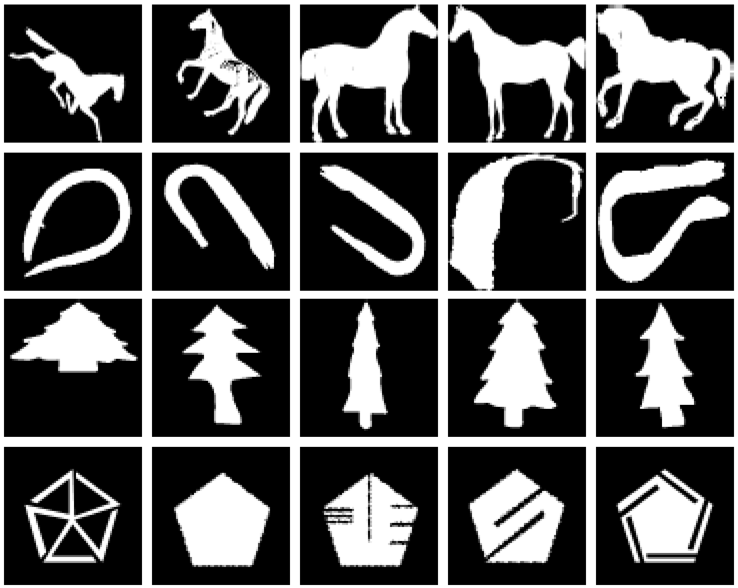
Fourier descriptors..contd.

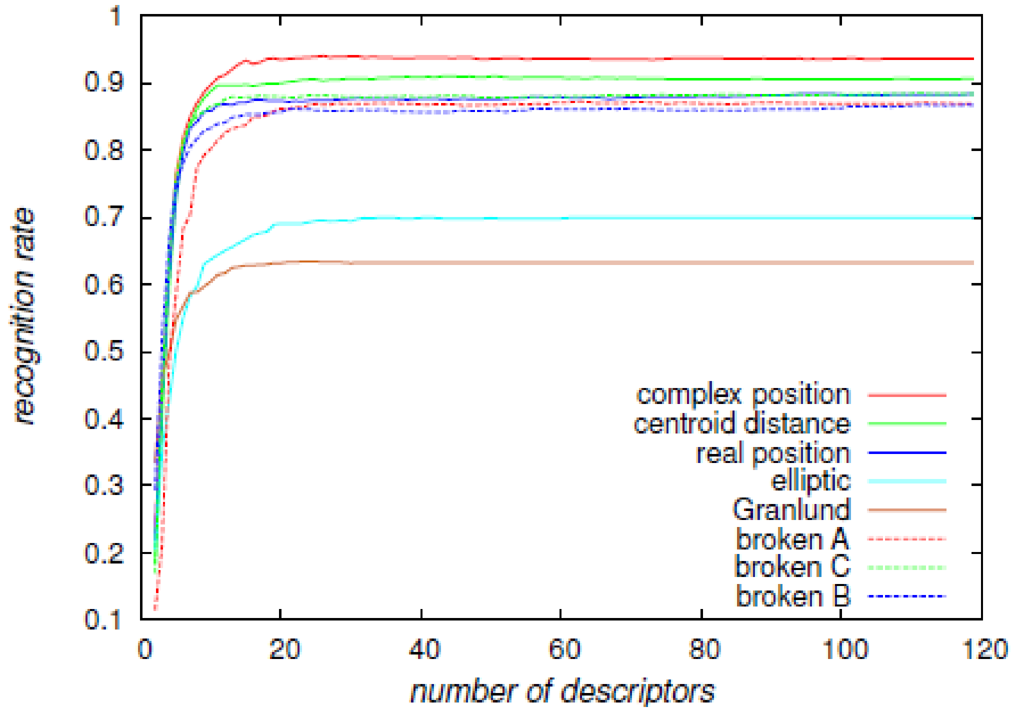
- Very popular for representing shape

- Many variants exist
  - All aimed at invariance+ improving discriminating power

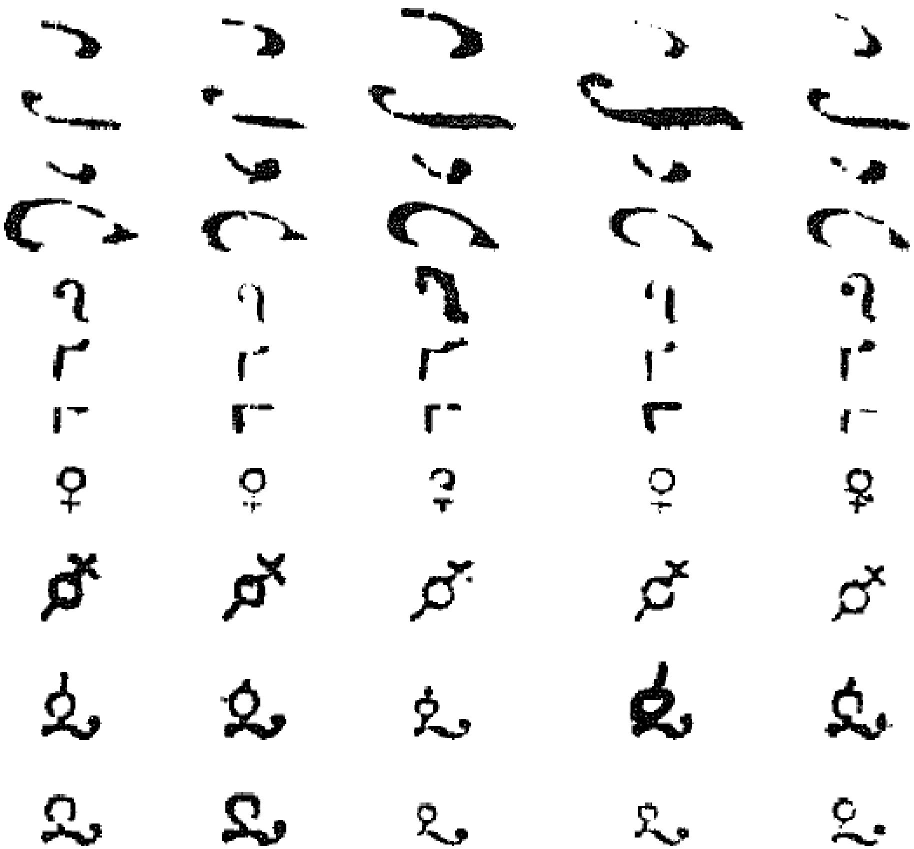
- Dalitz *et al.* (Eurasip J of Sig proc. 2013) propose one aimed at handling 'broken' shapes
  - Uses the radial distance $r[k]$ from centroid and a distance transform to encode shape

MPEG-7 CE-1 dataset

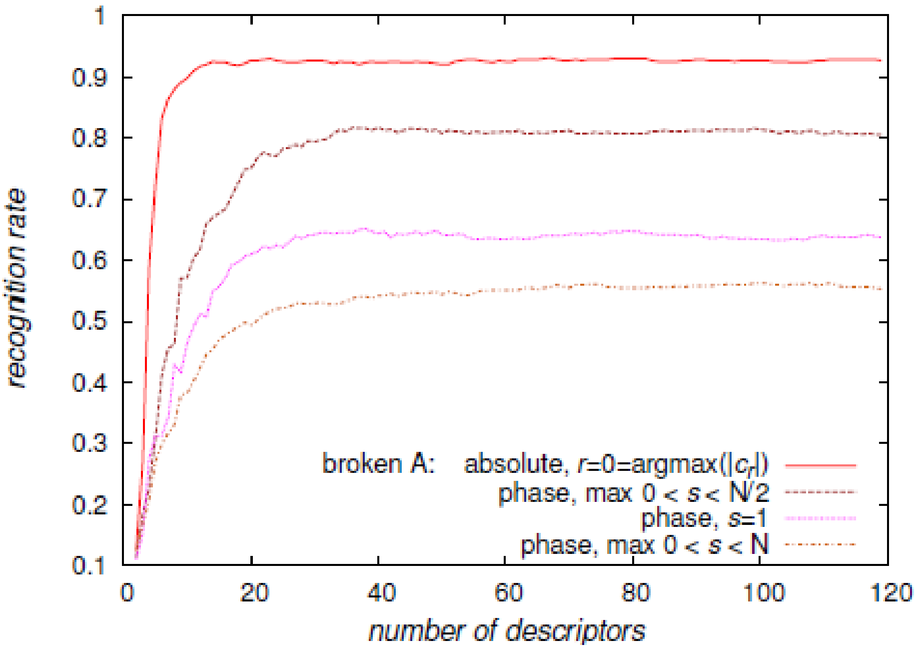




Neumes (broken shapes) dataset



Old music notation



Recognition vs number of descriptors

Region descriptors

Basic ones:

- *Area* A – total number of object pixels

- *Perimeter* P – total number of boundary pixels

- *Compactness* – $A/(P^2/4\pi)$
  - denominator is the area of a circle with perimeter P; max for a circular object

- *Dispersion* – major chord length/area
  - simpler defn: max radius/min radius

Moments (statistical)

Describes the layout of the object pixels

- **Moments** $$m_{pq} = \sum_x \sum_y x^p y^q I[x,y]$$
  - m_{00} = area under the shape I
  - *Centroid*: $[\bar{x}, \bar{y}] = [\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}}]$

- **Centralised moments** $$\mu_{pq} = \sum_x \sum_y (x-\bar{x})^p (y-\bar{y})^q I[x,y]$$
  - to attain invariance to translation
  - measure of skewness in x-direction: $\frac{\mu_{30}}{\sqrt{(m_{20})^3}}$
  - measure of skewness in y-direction $\frac{\mu_{03}}{\sqrt{(m_{02})^3}}$

Moment invariants

- Central moments are combined to create 7 moment invariants to attain invariance to
  - Translation
  - Rotation
  - Reflection

Orthogonal moments

- **Zernike polynomials**
  - set of orthogonal polynomials defined on a disc
- Start with orthogonal basis functions (in polar coords):

$$V_{nm}(\rho,\theta) = R_{n,m} e^{jm\theta};$$

$$n > 0; \quad |m| \quad \leq n; \quad n - |m| \quad is \quad even$$

$$R_{n,m}(\rho) = \sum_{s=0}^{(n-|m|)/2} (-1)^s . \frac{(n-s)!}{s!(\frac{n+|m|}{2}-s)!(\frac{n-|m|}{2}-s)!} \rho^{n-2s}$$

Zernike moments

- Find the projection of the image onto the polynomials defined onto a disc i.e. $(x^2 + y^2) = 1$

$$A_{nm} = \frac{n+1}{\pi} \sum_x \sum_y I(x, y) V^*{}_{nm}(\rho, \theta)$$

- If we use $|A_{nm}]$ the moments are rotationally invariant

- Have been used in retrieval, quantify mass volume, classify malignant/benign masses based on shape,