# 11. OBJECT RECOGNITION

In computer vision, we may wish to recognise either the classes of features and objects (e.g. a cat as opposed to a dog), or their individual instances (e.g. Felix as opposed to Tabby).

Recognition typically involves finding a *match* between an image instance and a some kind of *model*. There have been a great variety of models proposed for both natural and computer systems. Broadly speaking, models can be either "hardwired" or acquired by learning.

This Unit will follow chronologically the development of approaches to recognition, from first attempts (which are still used!) to the more recent systems. For the approaches which are well described in books and scientific journals, only the very basic details are given in the handout. A list of references at the back of this Unit points to relevant publications.

## Template matching

- matching is carried out in image space
- very limited method, but can be fast in specific applications

*Notes*

## Feature analysis

- Hubel & Wiesel's "feature detectors"
- Pandemonium system
- Selfridge (1959)
- Lindsay & Norman (1972)

*Notes*

## Structural descriptions

- Features
- Relationships
  - 2D
  - 3D
- From $2\frac{1}{2}$ D to 3D descriptions

*Notes*

## Volumetric representation & modular organisation

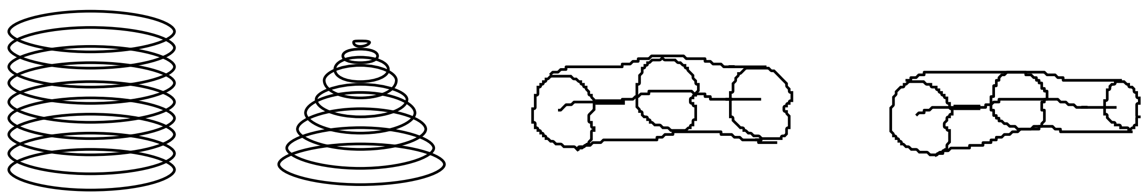
Generalised cones were first introduced by Marr & Nishihara (1978) as a representational framework for vision as a part of their thesis that in order to understand vision it is necessary to understand the underlying information processing tasks.

**Approach**
Representing the image
- Find zero-crossings on multiple scales (LOG)
- Create the Raw Primal Sketch
  edges, blobs, bars and terminators
- Derive Full Primal Sketch by grouping by similarity and proximity

Derive surface information from the Primal Sketch (21/2 D sketch)
- Stereo matching
- For each point on the surface provide its normal vector and distance from the observer

Represent shapes for recognition
- Generalised cones

Examples of generalised cones



Superquadratics (Pentland (1986)) extend the generalised cones approach to more generic shapes.

## Recognition by components (RBC)

*Biederman I. (1987) Recognition by components: a theory of human image understanding. Psychological Review 94, 115-147*

**Geons**

*Geons* are "geometric icons" - a set of 3-dimensional visual primitives characterised by:
- viewpoint invariance
- non-accidental properties
- canonical views

**Recognition By Components (RBC) theory**

Biederman proposed the following approach to object recognition:
- segment an image of a complex object into geons
  - extract all the line segments
    - breakpoints are set at matched concavities
    - the segmented regions are convex or simply concave
  - detect components of geons, using viewpoint invariant properties (VIPs), e.g.
    - edges straight or curved
    - pairs of edges parallel or not
    - vertices formed at co-terminations
  - group the components into geons
    - the VIP coded features activate the closest fitting geon
- recognise the object by matching the detected geons to the model object geons
  - model objects are represented by geons and their relationships
  - image objects are represented by geons and their relationships

The RBC theory claims to have solved the following fundamental problems in object recognition:
- viewpoint invariance (w.r.t. translation, size and viewpoint)
- grouping (organisation) of image elements into appropriate parts
- a basis of determining invariant object centred relations
- a basis for computing the similarity and equivalence of object images

Biederman also postulated how his computational theory can be implemented in a natural system (phase-locking and fast enabling links) and how the enabling links may develop through individual and evolutionary "learning experience".

**Notes:**

## Model based recognition

### Bottom - up techniques

- start from a 2D image
- derive descriptions
- match descriptions with the model

The techniques discussed thus far all fall in this category.

### Top-down techniques

- start from a model (3D based)
- look for instances of a model in an image

Examples of using this approach to object recognition are systems developed by Lowe (1987), described below; and Sullivan et al (1989).

### Perceptual Organisation and Visual Recognition (SCERPO system)

*David G Lowe, "Perceptual Organisation and Visual Recognition", Kluwer, Boston, 1987.*

Three-dimensional object recognition form single two-dimensional images.
Uses principles of perceptual organisation but not explicit depth reconstruction.

**Approach**
Process of perceptual organisation is used to form groupings and structures in the image that are likely to be invariant over a wide range of viewpoints:

- proximity
- parallelism
- colinearity

Probabilistic ranking method is used to reduce the size of the search space during model-based matching.

- measuring the probability that an instance of a particular grouping has arisen by accident low probability

Process of spatial correspondence brings the projections of three-dimensional models into direct correspondence with the image.

- matching consists of individually comparing each of the perceptual groupings in the image against each of the structures of the object model which is likely to give rise to that form of grouping

**Notes:**

## Point distribution model (PDM) - a statistical model of shape

*Cootes, Taylor & Cooper (1992)*

The key idea is that the views of an object, or a class of objects, are statistically similar. A statistical model (using PCA - principal component analysis) of a shape of a class of objects can be built from shape examples.

The steps of the algorithm:

- Label matching points in a training set of object shapes - $\{x_i\} i = 1,...,N$
- Align the training set
- Compute the "mean shape"
  $\underline{x} = 1/N \sum x_i$
- Calculate a covariance matrix
  $S = 1/N \sum dx_i \, dx_i^T$ , $dx_i = x_i - \underline{x}$
- Extract eigenvectors of S, $p_i$
  Eigenvectors corresponding to the largest eigenvalues describe the most significant modes of variation in shape
- Any shape in the model can be approximated by
  $x = \underline{x} + Pb$
  where: $P = (p_1, ..., p_t)$ - eigenvectors
  $b = (b_1, ..., b_t)^T$ - weights
- New examples of shapes can be generated by varying $b_i$ in:
  $b_i = P^T (x - \underline{x})$, $-3\sqrt{\lambda_i} \le b_i \le 3\sqrt{\lambda_i}$, where $\lambda_i$ is the i-th eigenvalue of S

Point distribution model

- Captures typical shape
- Allows variability
- Can be computed directly from a training set

**Notes:**

## Further reading and exploration


### Feature analysis

Lindsay PH, Norman, DA (1972) *Human Information Processing*, AP.

Selfridge, OG (1959) Pandemonium: a paradigm for learning. In: *The Mechanisation of Thought Processes*, HMSO, London.

### Structural descriptions

Winston, PH (1975) Learning structural descriptions from examples. In: Winston PH (Ed) *The Psychology of Computer Vision*, McGraw-Hill.

### Volumetric representation

Marr D, Nishihara HK (1978) Representation and recognition of the spatial organization of three-dimensional shapes. *Proc Royal Soc London*, Series B, 200, 269-294.

Pentland A (1986) Local shading analysis. In: Pentland A (Ed) *From Pixels to Predicates*, Ablex.

### Recognition by components

Biederman I (1987) Recognition by components: a theory of human image understanding. *Psychological Review* 94, 115-145.

### Recognition based on 3D models

Lowe, DG (1987) Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence* 31, 355-395.

Sullivan (1991) Relational model construction and 3D object recognition from single 2D monochromatic image. *Proc BMVC 1991*, 240-248.

### View-based recognition

Breuel, TM (1993) View-based recognition. *IDIAP Technical Report* 93-09.

Cootes TF, Taylor CJ, Cooper DH, Graham J (1992) Training models of shape from sets of examples. *Proc BMVC* 1992, 9-18.

Korn MR, Dyer CR (1987) 3D multiview object representations for model-based object recognition. *Pattern Recognition* 20, 91-103.