Modeling Human Visual Recognition

[Prof B Velichkovsky]

C Rasche

1st part: existing approaches
2nd part: toward contour/region based approach
Notes contain links to selected websites

When we see a novel image…

Rosch et al 75, Oliva & Torralba 2001

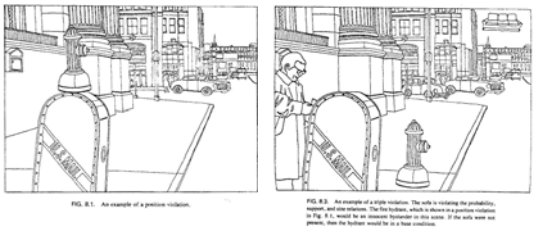
Categorization/Classification

Furniture	Animal	Plant	Building	Super-Ordinate
Chair	Dog	Tree	House	Basic-Level (5000)
Kitchen Chair	Dachs	Ahorn	Mansion	Sub-Ordinate

- occurs within 150ms (+- 50ms)!...that is before 1st saccade is made
- for canonical views
- How much do we see within this first glance? → 'Gist Perception'

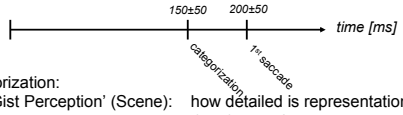
Scene Memory

Biederman



- helps in visual search (saccadic eye movements)
- we expect objects in certain positions with certain size. scene knowledge.

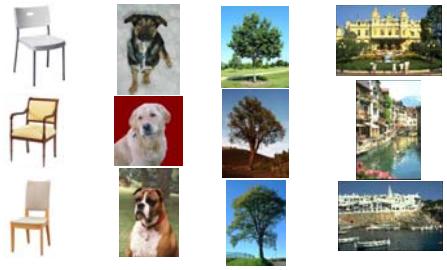
Visual Recognition



- Categorization:
  'Gist Perception' (Scene): how detailed is representation?
  how much is it *bottom-up* or *feedforward*?
  retrieves scene memory
- Visual Search (*top-down, feedback*)
  Saccadic eye-movements (3-4/s)
  Attentional shifts (as fast as 33ms per shift)
  Detailed understanding
  Learning and memorizing?
- When modeling, we encounter two problems
  structural variability, fragmented contour image

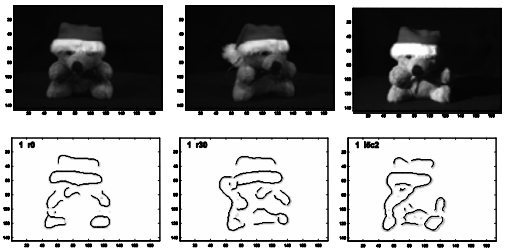
Structural Variability

Witkin & Tenenbaum 83 , Draper et al 96, Palmer 99, Rasche 05, Basri & Jacobs 97



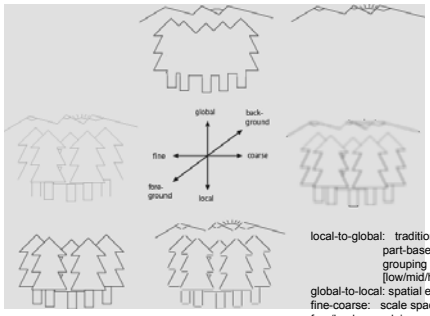
- variability in contours, object parts, configuration
- …that's why so far any theory or model has failed…
- …we're dealing with the most complex electrical device

Fragmented Contour Image



- Gray-scale image = intensity image ≈ landscape
- Contour extraction metaphor ≈ determining structure of landscape
- Image = noise source + signal

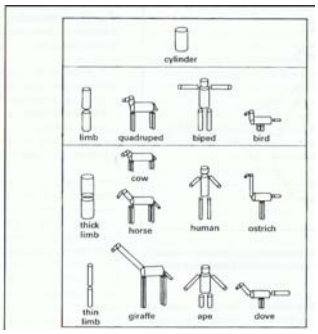
Evolvement Concepts



local-to-global: traditional neuroscientific view
part-based approaches (part-whole)
grouping
[low/mid/high level vision]
global-to-local: spatial envelope
fine-coarse: scale space, pyramid
fore/background: image segmentation

Marr [early 80's]

A catalogue of 3-D model descriptions at different levels of specificity. Reproduced from Marr and Nishihara (1978) with permission of The Royal Society.



Recognition by Cylinders (local-to-global)

Problem: Can't be applied

Marr

(influenced by Gibson)

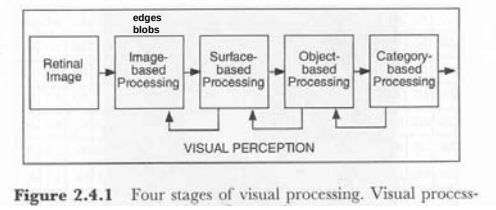


Figure 2.4.1 Four stages of visual processing. Visual process-

Early theories very 3D dominated – surface perception as a first step.

Biederman [late 80's]

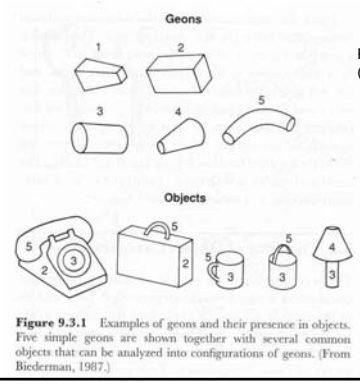


Figure 9.3.1 Examples of geons and their presence in objects. Five simple geons are shown together with several common objects that can be analyzed into configurations of geons. (From Biederman, 1987.)

Recognition by Parts (vertex detection)

Neural Network

- Layer 1: local orientations
- local-to-global
- (model not so popular with neuroscientists)

Problem: Still can't be applied

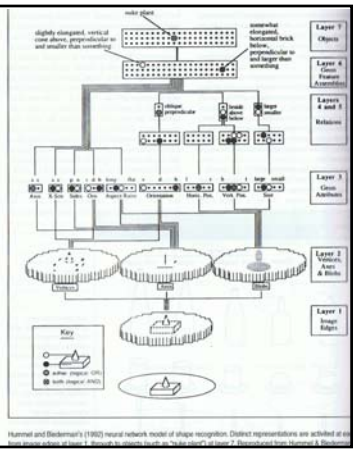
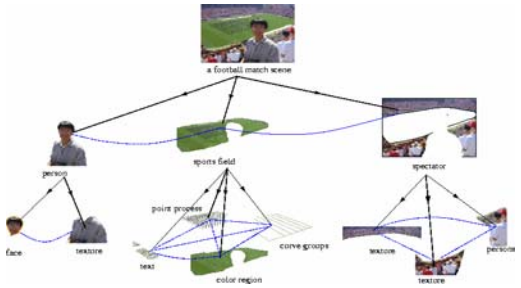


Image Segmentation

(toward scenes)



- foreground/background
- problem: can't be applied either for categorization
- easier when when category is known!

Scale Space

• Goal: to smear image to obtain more information
→ smoother but more accidental contour extraction

• repeated convolution (*) of original image with a Gaussian filter:
3rd dimension scale: $I(x,y,\sigma)$

• sometimes subsampled: pyramid

• does it solve the variability problem?
…only limitedly.

fine

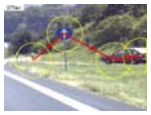
*

coarse

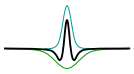
Saliency Map

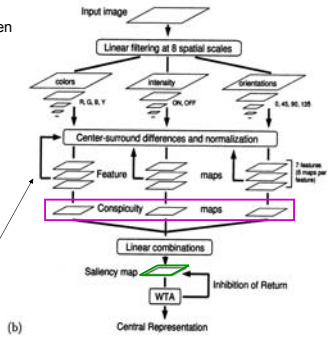
(Itti & Koch)

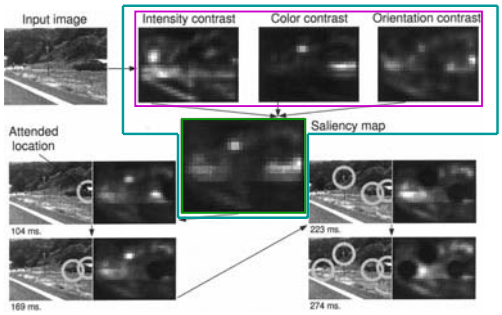
Goal: to mimic a bottom-up driven visual search of attentional and saccadic shifts



Band-Pass Filter: DOG
(difference of Gaussians)



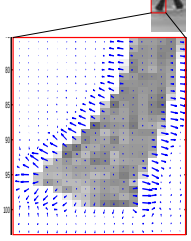


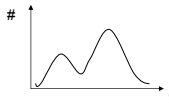


• popular in psychology/neuroscience due to attempt to connect to visual search studies and neurophysiology
• limitedly used in computer science due to time-intensive DOG filtering

(Object Detection)

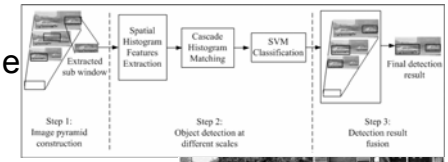
- Object search
  biologically a bit implausible
1. '**appearance**'-based:
   e.g. intensity gradients
   histogramming (circle, square)
   multi-dimensional vectors
2. **contour**-based:





- Region of interest (ROI)
- Interest point detection

Example



• views very similar

(Image Retrieval)

- Google uses text annotation
- Content-based search engines:
  primarily relying on color
  secondarily relying on edge information (histograms)
- JPEG descriptors

Caltech 101/256

(Perona)

- Sub-ordinate object categorization
- Purely computationally motivated

- Each pixel = 1 dimension (high-dimensional vector)

1. generating 'sparse code' with PCA
2. classifier
3. weakly supervised learning by human

- Orientation histograms

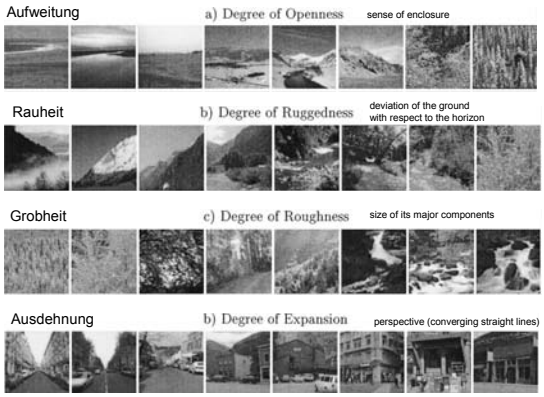
- Drawback: can not interpret parts of structure
- (can not be assigned to any of the evolvement axes)

Spatial Envelope

Oliva & Torralba

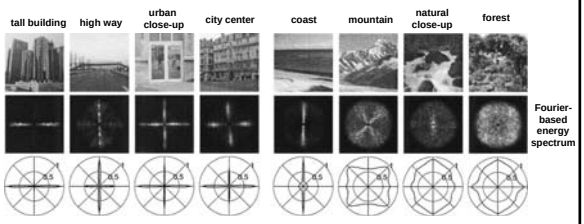
- Super-ordinate category system
- Using psychophysical experiments:
  scenes are described by 5 perceptual dimensions:
  naturalness, openness, roughness, expansion, ruggedness

- 'holistic' representation (global-to-local)



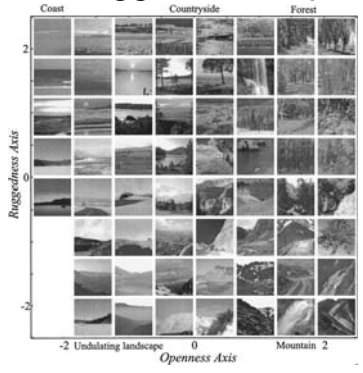
Energy Spectra

for 8 super-ordinate categories:



- PCA, classifier
- No contours, no image segmentation, no grouping due to speed
- system has potential for basic-level categorization

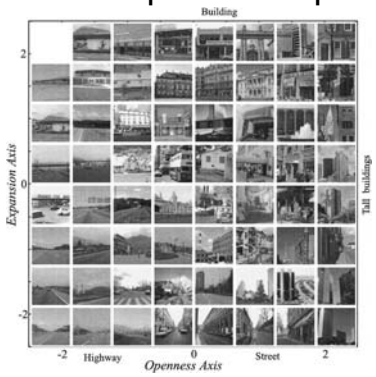
Ruggedness-Openness



Ruggedness: deviation of the ground with respect to the horizon

Openness: sense of enclosure

Expansion-Openness



Expansion: perspective (converging straight lines)

Criticism

- Previous systems (Caltech 101, Spatial Envelope) are successful because categories are relatively distinct and manually selected by humans.
- However:

  1) preprocessing (PCA, Fourier) can not be used to interpret parts of the image/structure

  2) can not deal well with overlapping content (perception subjectivity)

  3) Other images?

→ Contour, Regions (NN: template matching)

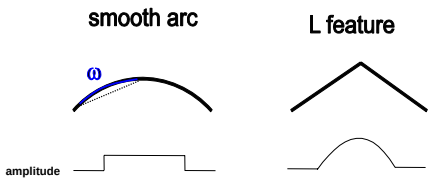
Contour Approach

Where should this shape be segregated?

→ At points of highest curvature. How to locate them?
No algorithm yet.
Scale space (coarse/fine)? Unfortunately not.
→ Local/global space: creating signatures Fischler, Bolles, 1983

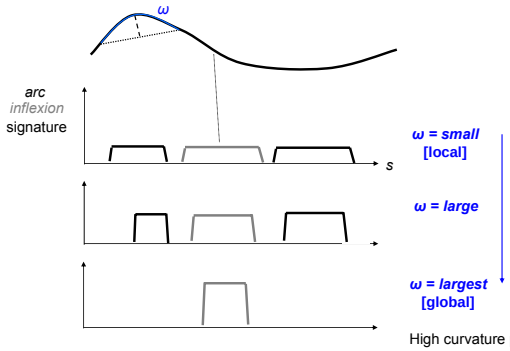
Signature

For One Window Size

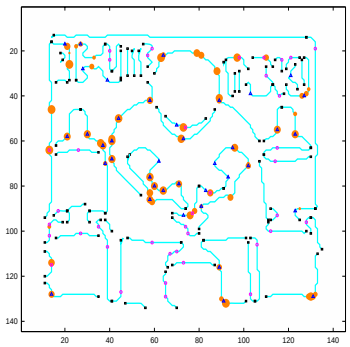


Iteration with fixed contour segment (window)
taking amplitude → signature
Advantage: signature (1D function) easier to analyze than 2D contour (can determine high-curvature points)

Local/Global Space



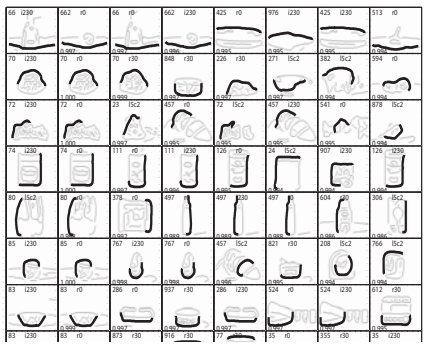
High curvature points
Geometric parameters



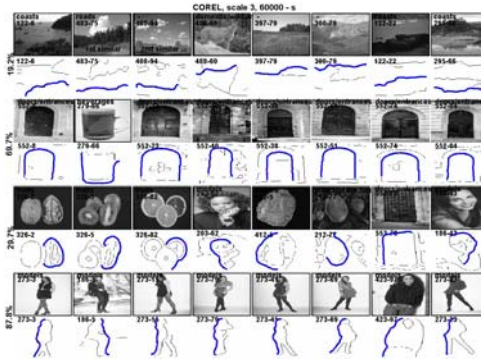
Proper segregation:
1) context dependent
2) matter of interpretation

→ how perfect must segregation be for categorization?

Selected Decreasing Similarity [>50'000 cnts; 1000 objs]



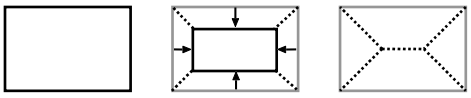
→ just a vector!



Symmetric-Axis Transform

Blum 73

(Gestaltists: self-collapsing structure)

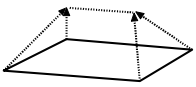


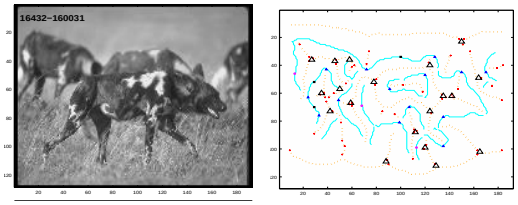
time

grass-fire (wave propagation)

- segments very informative
- previous implementations
  [closed shapes, iterations]

Feldman & Singh, 2006





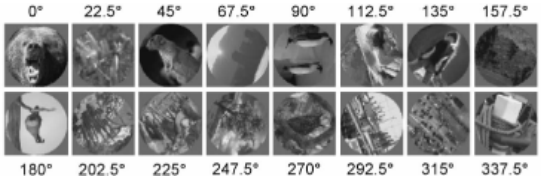
Works for fragmented contour images!

The sym-axes outline:

1) regions: useful to obtain texture information
2) geometry: useful to represent exact relations between contours

Image-Rotation Invariance

Guyonneau & Kirchner & Thorpe 06

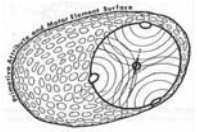


Images are still categorized within ca. 150ms!

Translation Invariance

Thorpe

attentional shifts? [33ms]



30 - 70 ms
50 - 90 ms

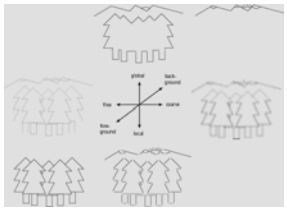
Blum 67, Deutsch 62, McCulloch 65

Neuromorphic Engineering

Mead 89

- Goal: to emulate the nervous system using analog electronic circuits – real time & low power.
- Fundamental principle: transistor run in analog domain (exponential I-V relationship)
- Noise: requires robust and simple models
- A large system (network) is distributed across several chips: communication by spikes (very much like in the brain) Deiss et al 99
- In short: there exists exciting technology, but no effort to develop models suitable for it

Evolvement Concepts



-- What is categorization?
Which one is the typical one/most used one?
-- Describe the principal recognition steps when a scene is seen (first ca. 250ms.)
-- Which evolvement concepts do you know?
-- What types of systems follow a local-to-global evolvement?
-- What are the difficulties of categorization?
-- What is one way to simulate a bottom-up driven visual search?
-- Which categorization systems do you know?
- On what kind of images do they work?
- Which system claims to correspond to a global-to-local evolvement?
-- How does a typical NN start analyzing its input?
-- What kind of method to you know that transforms a region or to express the relation of contours?
-- Is color a useful cue for categorization?