A Model of Visual Categorization Based on Structural Parameterization

Christoph Rasche

Justus-Liebig Universität Giessen
Germany

When we see a novel image…

Rosch et al 75, Oliva & Torralba 2001

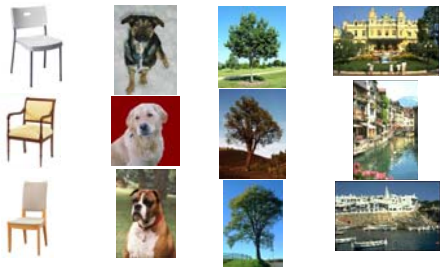
…we assign it to a basic-level category…

Chair Dog Tree House

…occurs within ca. *150ms*!
[for canonical views Palmer et al 81]

Structural Variability

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

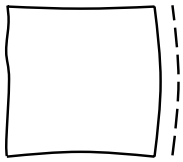


- Perona et al 101 categories: PCA, classifier, human-supervised
- Oliva & Torralba 8 urban/natural scenes [*super-ordinate*]: FT, classifier

→ Preprocessing not useful for analysis of 'parts' (or simulating a visual search by eye-movements)

Structural Description

Guzman 71, … , Marr 82, … , Biederman 87,…NN



integrating local orientations?
→ template matching

coarse scale?
→ loss of accuracy

→ local/global
→ distance distributions
obtaining parameters
→ buffer variability in a multi-dim. space

Contour Description

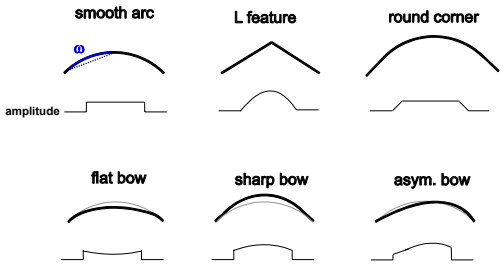
many animal silhouettes
undulating
curvature or inflexion
low edginess

natural scenes
wiggly
irregular
high edginess (zig-zag)

→ requires description that identifies those aspects:
elementary segment: **bow**

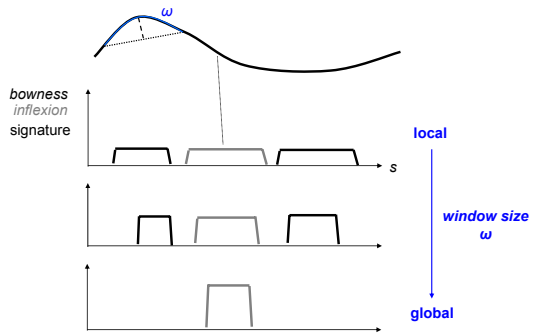
Bowness Signature

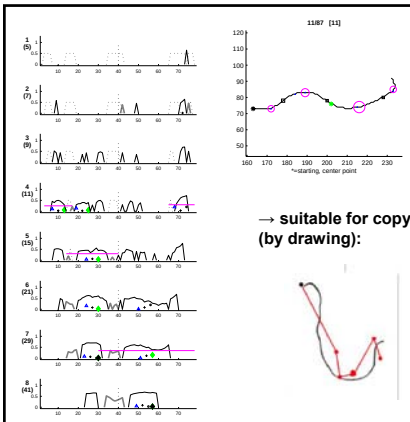
iteration with fixed window (chord/stick)



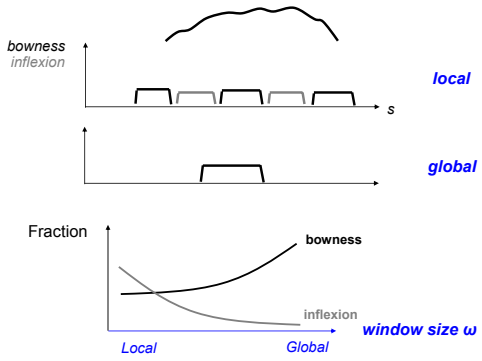
Advantage: signature (1D function) easier to analyze than 2D contour
→ curvature, circularity, edginess, symmetry,…

Local/Global Space

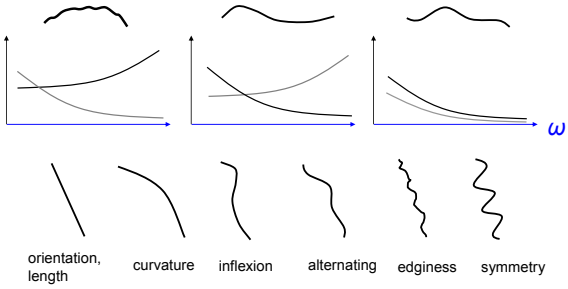




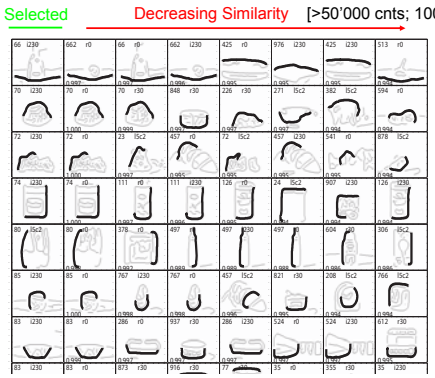
Toward Abstraction of Global Geometry



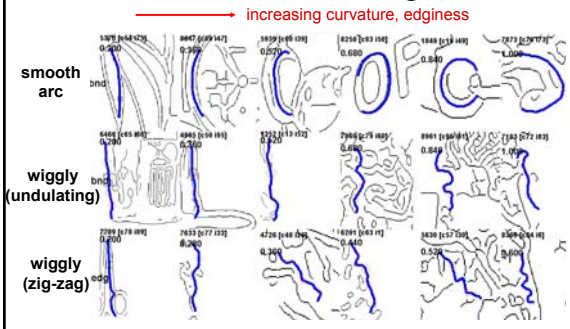
Parameterization



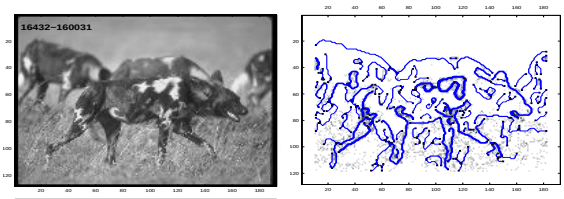
c*(o, l, a, t, x, b, e, s)*



Parameter Changes

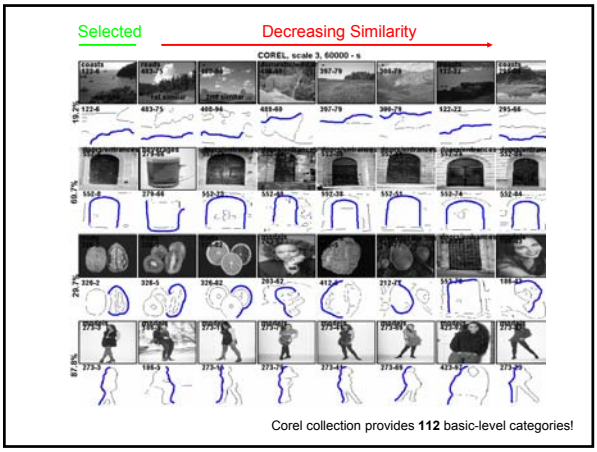


For Gray-Scale Scenes: Adding Appearance Parameters



$\boldsymbol{c}(\ o, l, c, t, a, x, e\ \ r, c_m, c_s, f_m, f_s\ )$

geometry appearance



Corel collection provides **112** basic-level categories!

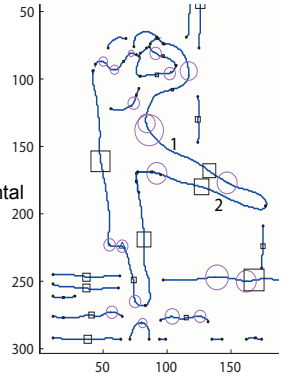
Contour Partitioning

[contextual issue]

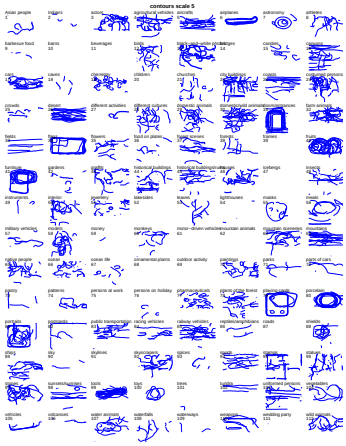
1 'hard' rule only: U-turns

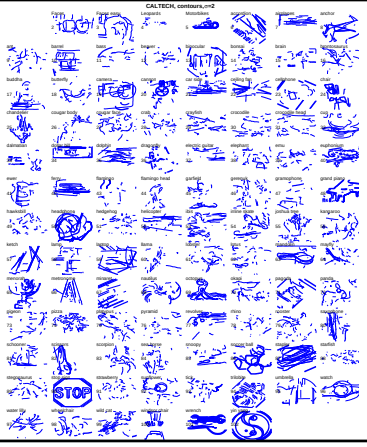
- category specific <> accidental
- local segments useful for grouping/other features

Extracting:
- large bows
- large straight segments



Category-Specific Contours



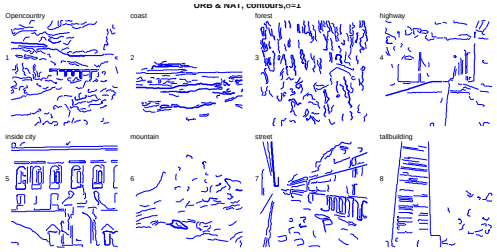


CALTECH 101

Li et al 06

Urban & Natural

Oliva, Torralba 01



Can Explain Parallel Contour Pop-Out

Treisman & Gormican 88

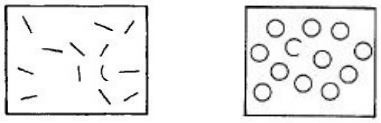


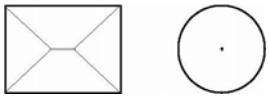
Fig 5 Fig 11

Traditional: saliency implemented by templates
Alternative: a systematic decomposition/parameterization
→ saliency = *var(**c**)*

Criticism: many pop-outs possible…

And now to Structural Relations…

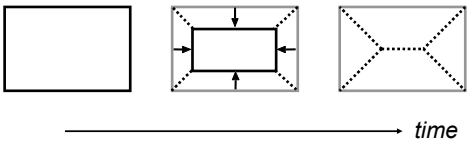
…using Blum's Symmetric-Axis Transform



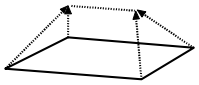
- global-local processing Kovacs & Julesz 94, Lee et al 98
- biological motion perception Kovacs
- saccadic landing positions Richards & Kaufman, Kowler

Grass-Fire

Blum 73

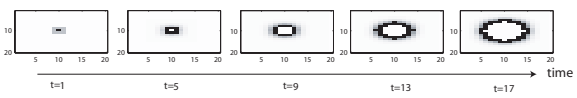


- segments very informative
- previous implementations
  [closed shapes, iterations]
  Feldman & Singh, 2006



Wave Propagation

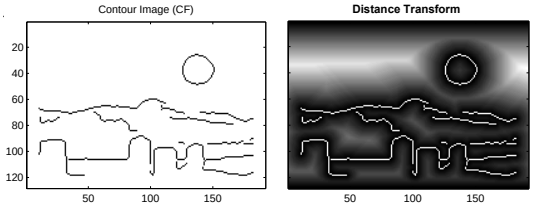
Traveling wave in a 2D map of I&F neurons



Passive propagation [sub-threshold]
Active propagation [above-threshold]
(Similar to dendritic active propagation)

Temporal evolvement for point: amphitheater

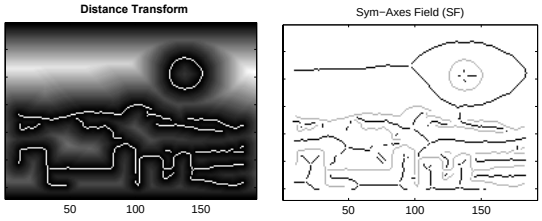
Temporal Evolvement → Landscape



- characteristics:
  Rivers ≈ Contours
  Ridges ≈ Sym-axes

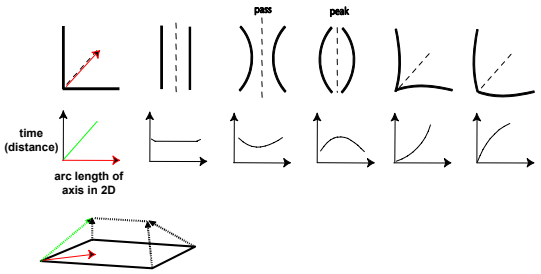
- works for fragmented contour images

Selecting Sym-Axes with High-Pass Filter [e.g. DOG]



- biologically plausible: retina or V1 (single propagation + DOG)
  → *translation independent*
- next: segregation at intersections

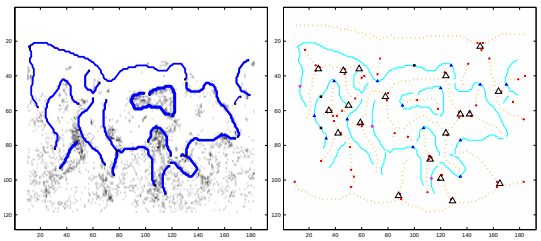
Freq. Occurring SAX Signatures



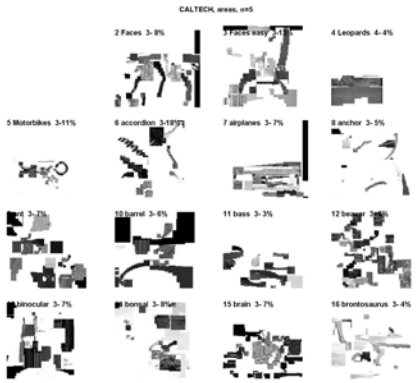
→ parameterization: $o, l, s_1, s_2, s_{fx}, p_{fx}, b, e,$

Area (Relation) Vector

$$\boldsymbol{a}(\underbrace{o, l, s_1, s_2, s_{fx}, p_{fx}, b, e,}_{\text{geometry}} \underbrace{c_m, c_s, f_m, f_s}_{\text{appearance}})$$

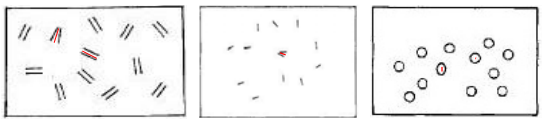


Sym-Ax Vector



Et Voilá…

Treisman & Gormican 88

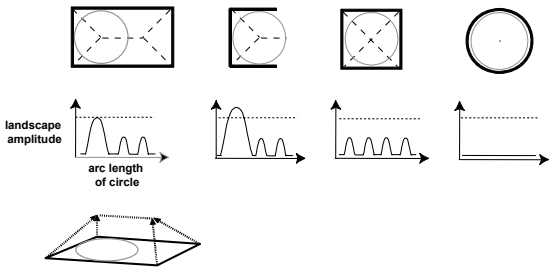


• Summary: flexible, dynamic decomposition (instead of a fixed architecture)

[to clarify: not supporting FIT]

Toward Shape Abstraction

RF at point of intersection



→ parameterization: max_{amp} (openness-closeness)

Categorization Results

[vector matching; 10 training images]

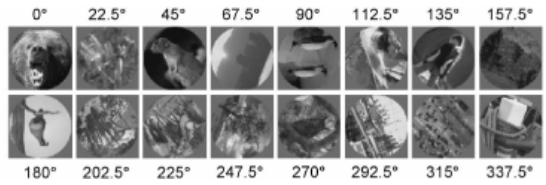
COREL	CALTECH	URB&NAT	
18%	19%	42%	(human-unsupervised)

However, my approach:

- one decomposition **for all inputs**
  (texture, shape, object, scene)
- can understand **parts** of the structure

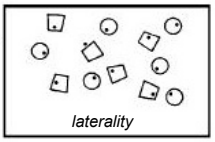
Image-Rotation Invariance

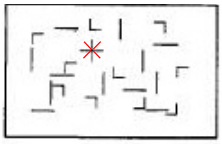
Guyonneau & Kirchner & Thorpe 06



- Orientation knock-out is robust

Future?: More Grouping





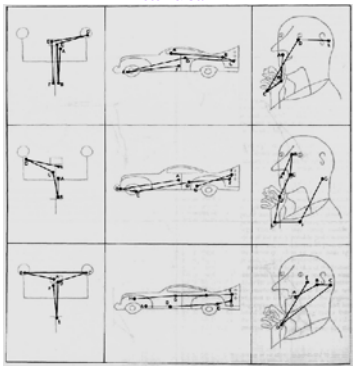
- It's in the temporal landscape
  (context analysis of sym-axis signatures)

→ structural relations (Gestaltist's idea: self-collapsing)
→ probabilistic combinations of descriptors
→ more appearance dimensions (gradient, 3rd moment, texture)

Predicting Fixation Locations?

Noton & Stark 71

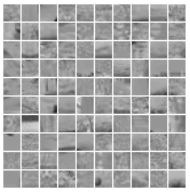
can explain all pop-out



In Gray-Scale Scenes?

Kienzle et al 06

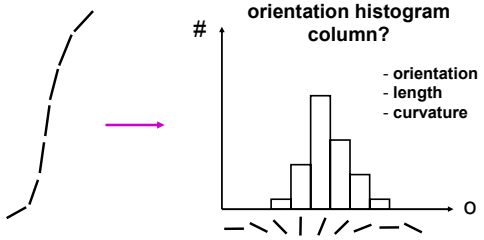
worst



best

- …intersections are part of it:
  decomposition/grouping delivers them.

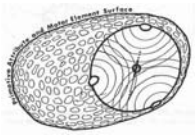
Neural Network?



- V1: long-range horizontal connections
- translation independent

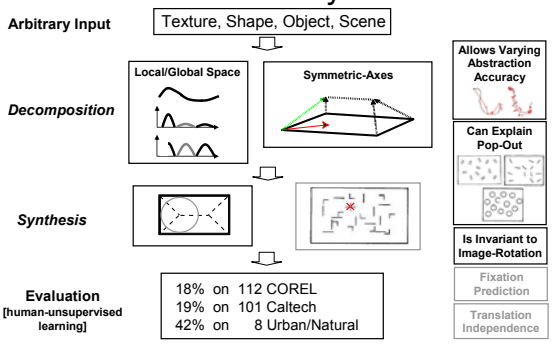
Translation Invariance

Thorpe



Blum 67, Deutsch 62, McCulloch 65

Summary



Acknowledgments

Lab support: *Karl Gegenfurtner*

COREL Categorization: *Nadine Hartig*

Gaze Com

Gaze-Based Communication Project
European Commission within the Information Society Technologies contract no. IST-C-033816