

The Kadir Operator Saliency, Scale and Image Description

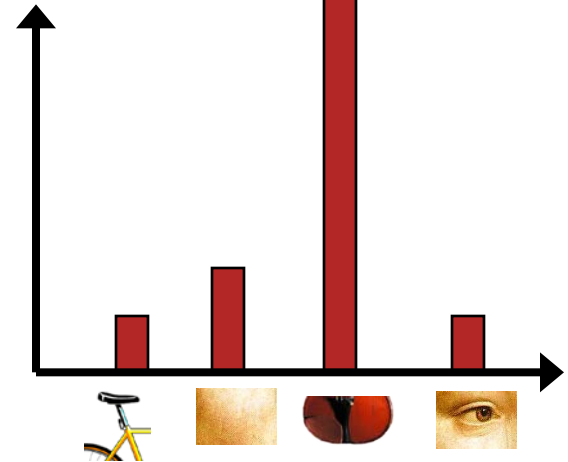
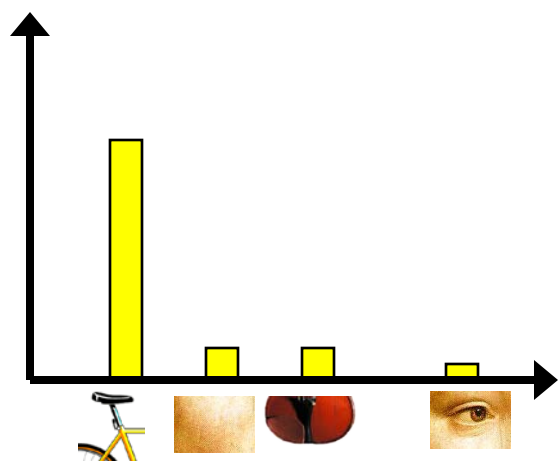
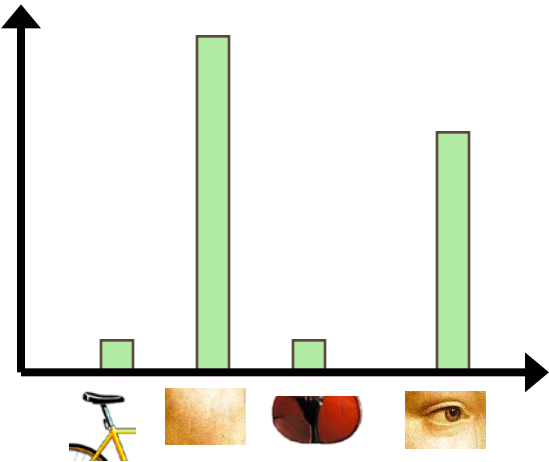
Timor Kadir and Michael Brady
University of Oxford

The issues...

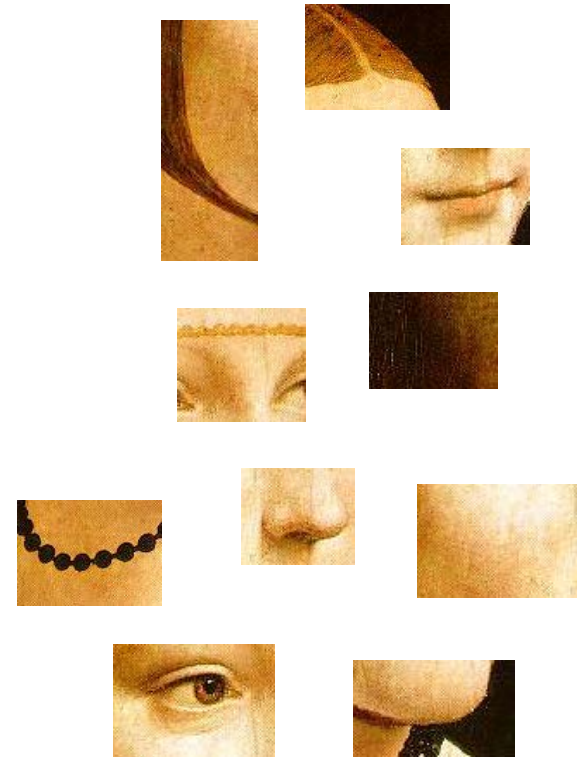
- salient – standing out from the rest, noticeable, conspicuous, prominent
- scale – find the best scale for a feature
- image description – create a descriptor for use in object recognition

Early Vision Motivation

- pre-attentive stage: features pop out
- attentive stage: relationships between features and grouping

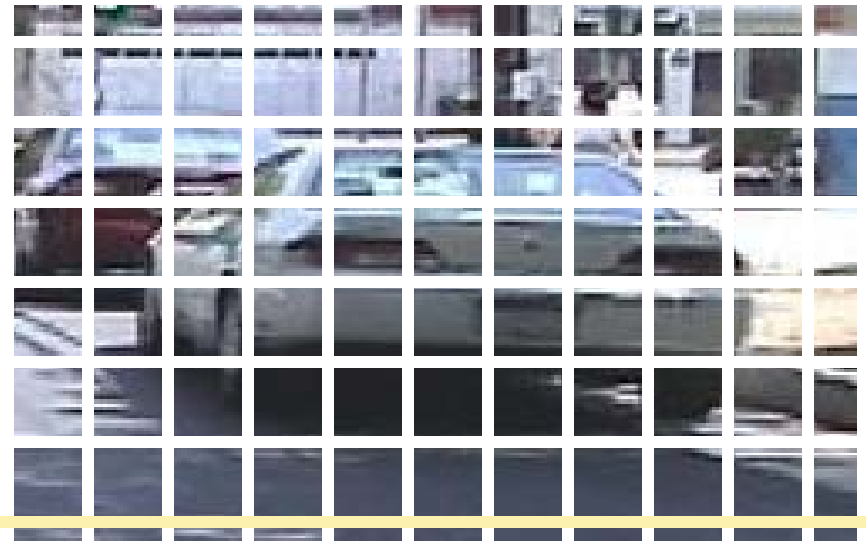


Detection of Salient Features for an Object Class



How do we do this?

1. fixed size windows
(simple approach)
2. Harris detector,
Lowe detector, etc.
3. Kadir's approach



Kadir's Approach

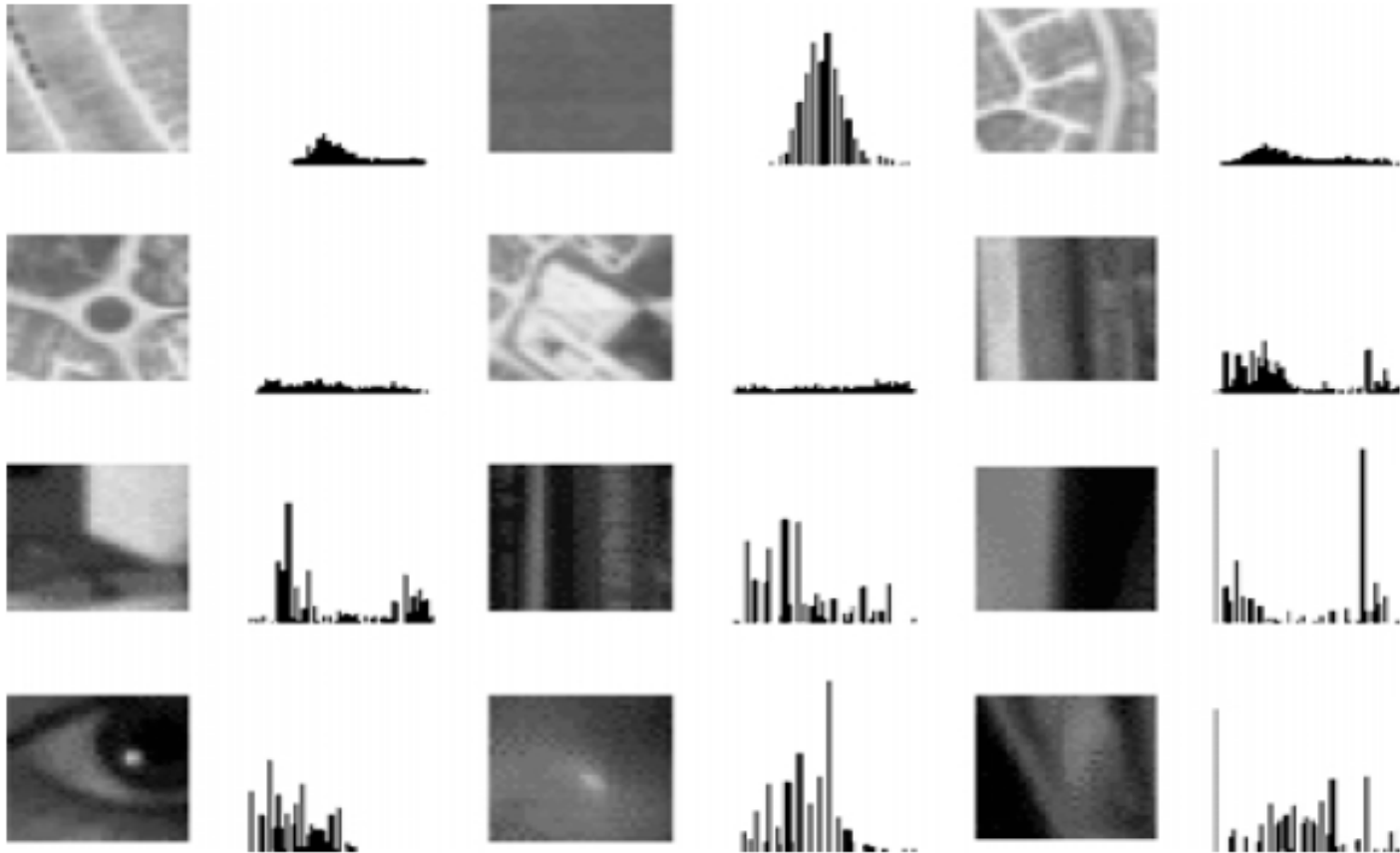
- Scale is intimately related to the problem of determining **saliency** and extracting relevant descriptions.
- Saliency is related to the local image complexity, ie. **Shannon entropy**.
- entropy definition $H = -\sum_{\substack{i \text{ in set} \\ \text{of interest}}} P_i \log_2 P_i$

Specifically

- x is a point on the image
- R_x is its local neighborhood
- D is a descriptor and has values $\{d_1, \dots, d_r\}$.
- $P_{D,R_X}(d_i)$ is the probability of descriptor D taking the value d_i in the local region R_x . (The normalized histogram of the gray tones in a region estimates this probability distribution.)

$$H_{D,R_X} = - \sum_i P_{D,R_X}(d_i) \log_2 P_{D,R_X}(d_i)$$

Local Histograms of Intensity



Neighborhoods with structure have flatter distributions which converts to higher entropy.

Problems Kadir wanted to solve

1. Scale should not be a global, preselected parameter
2. Highly textured regions can score high on entropy, but not be useful
3. The algorithm should not be sensitive to small changes in the image or noise.

Kadir's Methodology

- use a scale-space approach
- features will exist over multiple scales
 - Berghoml (1986) regarded features (edges) that existed over multiple scales as best.
- Kadir took the opposite approach.
 - He considers these too self-similar.
 - Instead he looks for **peaks in (weighted) entropy over the scales.**

The Algorithm

1. For each pixel location x
 - a. For each scale s between s_{min} and s_{max}
 - i. Measure the local descriptor values within a window of scale s
 - ii. Estimate the local PDF (use a histogram)
 - b. Select scales (set S) for which the entropy is peaked (S may be empty)
 - c. Weight the entropy values in S by the sum of absolute difference of the PDFs of the local descriptor around S .



Finding salient points

- the math for saliency discretized

$$Y_D(\mathbf{s}, \mathbf{x}) = H_D(\mathbf{s}, \mathbf{x}) W_D(\mathbf{s}, \mathbf{x})$$

$$H_D(\mathbf{s}, \mathbf{x}) = - \sum_{d \in D} p_{\mathbf{s}, \mathbf{x}}(d) \log_2 p_{\mathbf{s}, \mathbf{x}}(d)$$

$$W_D(\mathbf{s}, \mathbf{x}) = \frac{s^2}{2s - 1} \sum_{d \in D} |p_{\mathbf{s}, \mathbf{x}}(d) - p_{s-1, \mathbf{x}}(d)|$$

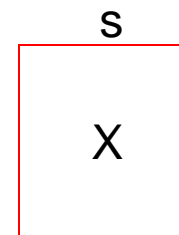
\mathbf{x} = point

$\mathbf{s} = (s, r, \theta) = (\text{scale}, \text{[redacted]})$

D = low - level feature domain (gray tones)

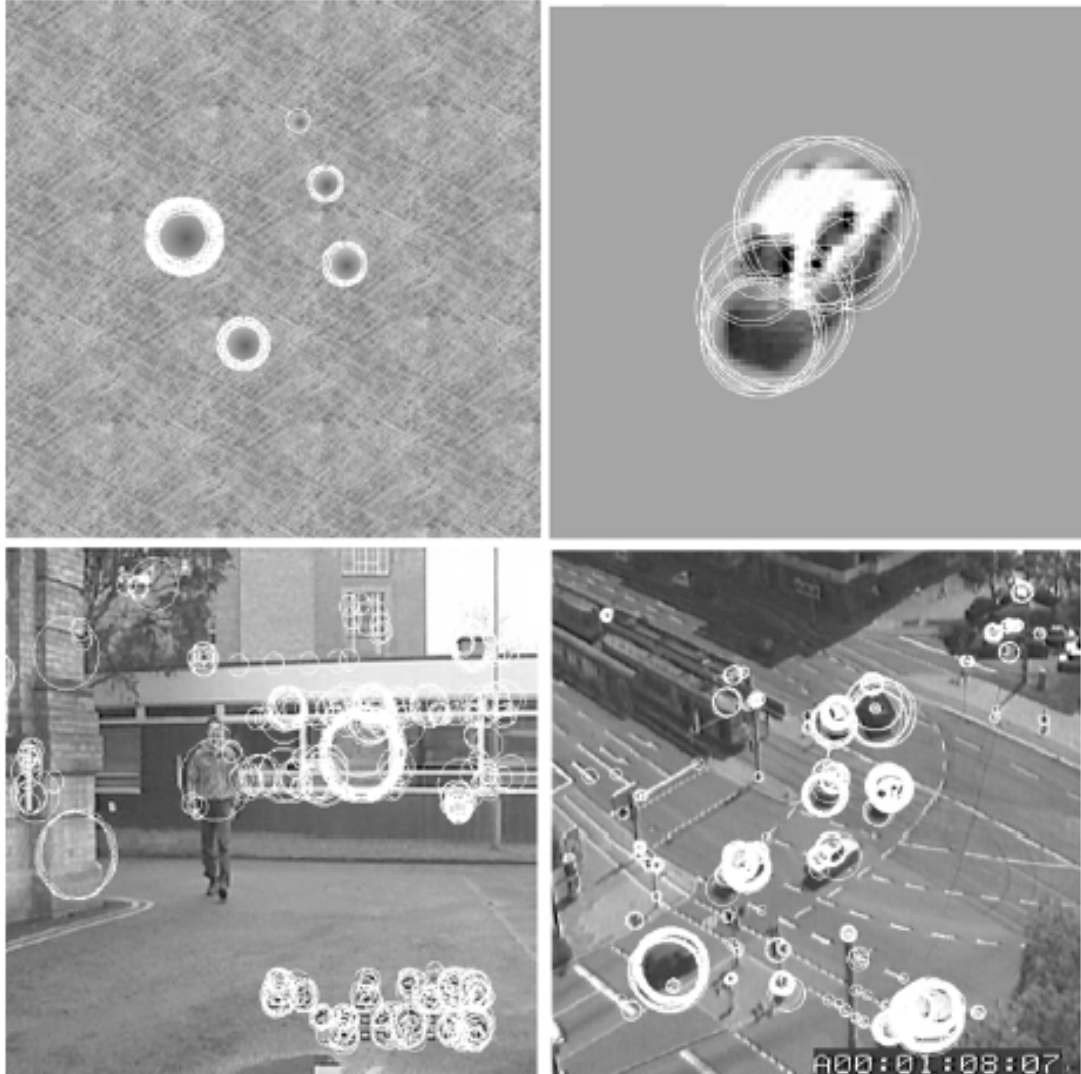
$p_{\mathbf{s}, \mathbf{x}}(d)$ = probability of descriptor D taking value d in the region centered at \mathbf{x} with scale s

- saliency
- entropy
- weight based on difference between scales



= normalized histogram count for the bin representing gray tone d .

Picking salient points and their scales



Getting rid of texture

- One goal was to **not** select highly textured regions such as grass or bushes, which are not the type of objects the Oxford group wanted to recognize
- **Such regions are highly salient with just entropy**, because they contain a lot of gray tones in roughly equal proportions
- But they are **similar at different scales** and thus the weights make them go away



Salient Regions

- Instead of just selecting the most salient points (based on weighted entropy), select **salient regions** (more robust).
- Regions are like volumes in scale space.
- Kadir used **clustering** to group selected points into regions.
- We found the clustering was a **critical** step.

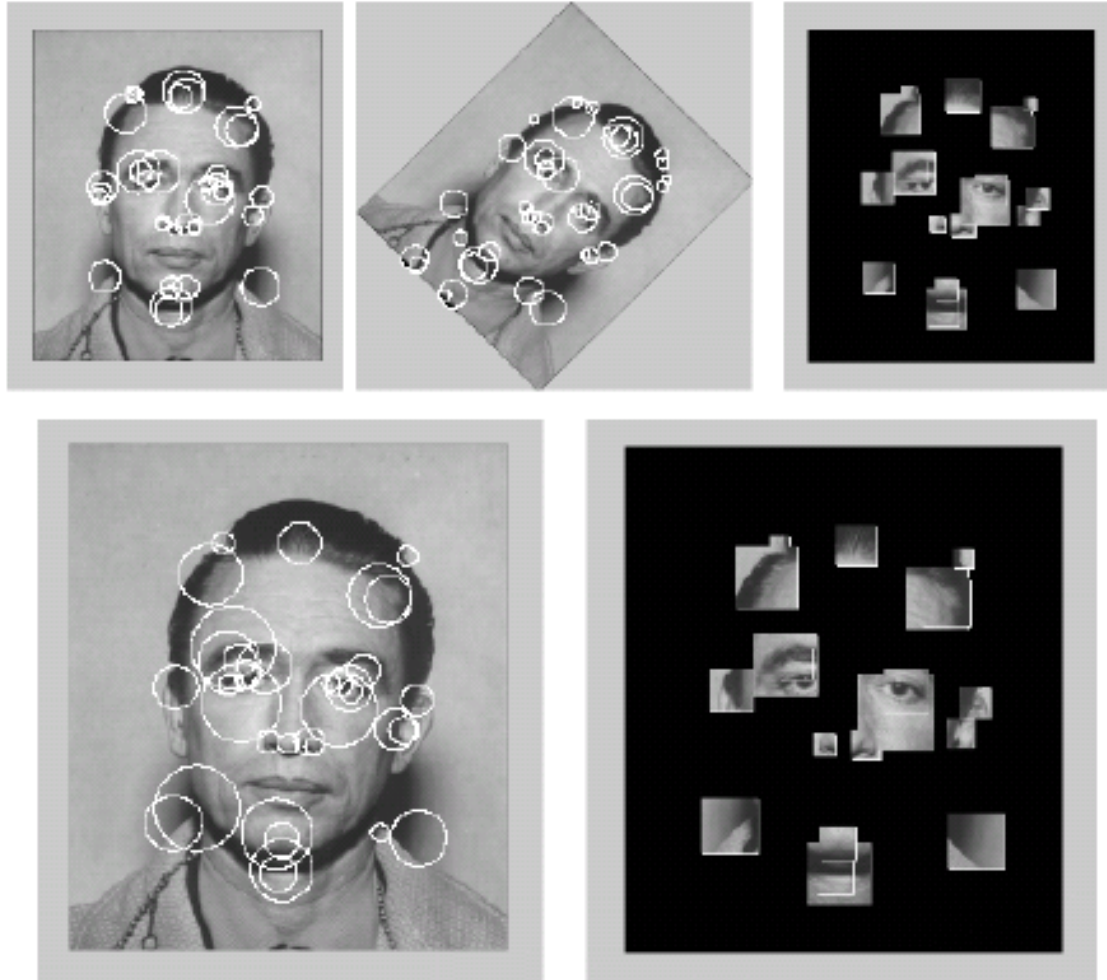
More examples



Robustness Claims

- **scale invariant** (chooses its scale)
- **rotation invariant** (uses circular regions and histograms)
- **somewhat illumination invariant** (why?)
- **not affine invariant** (able to handle small changes in viewpoint)

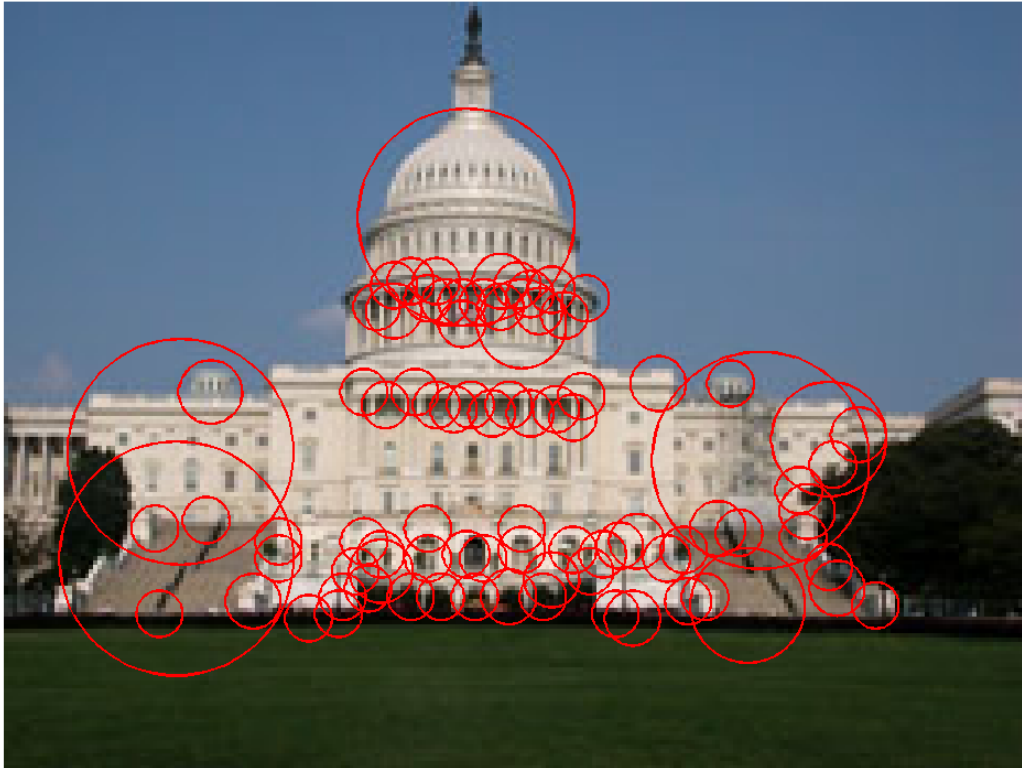
More Examples



Temple



Capitol



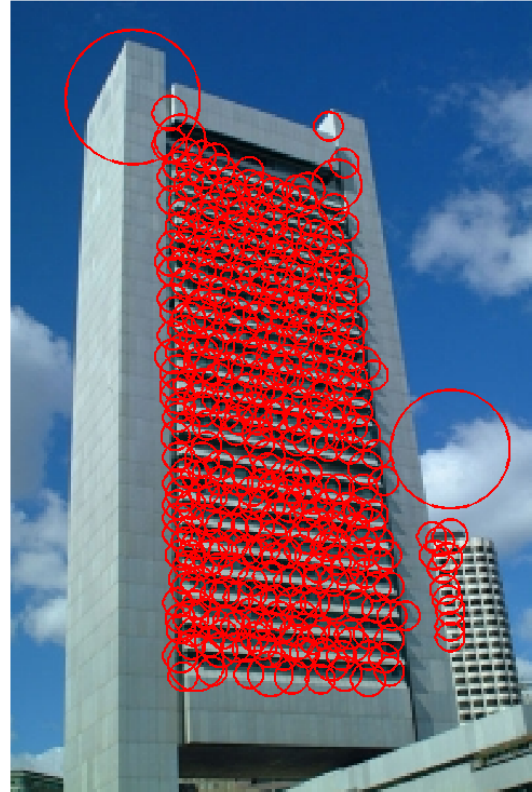
Houses and Boats



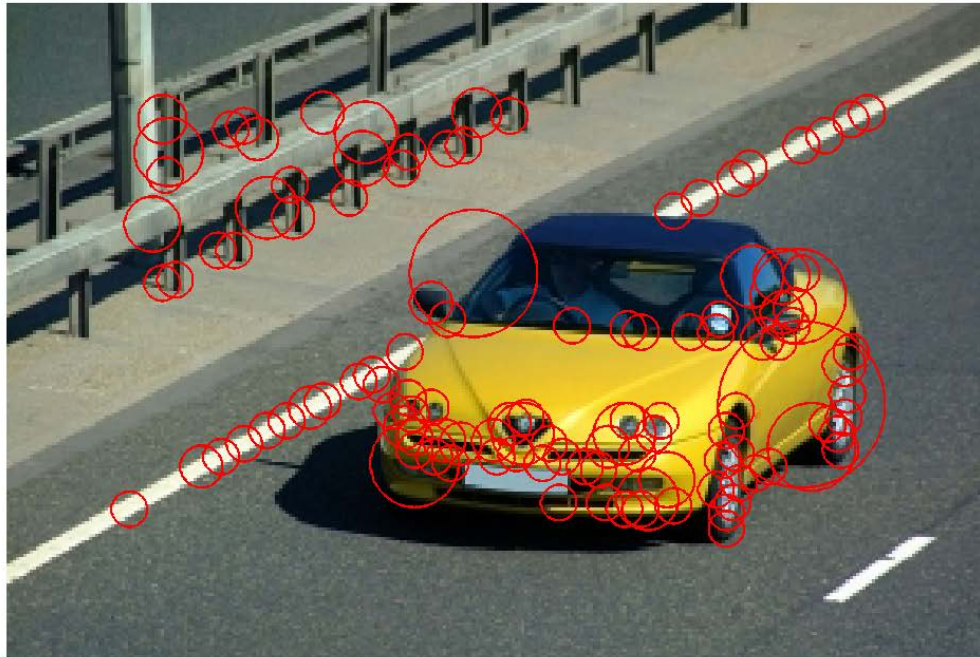
Houses and Boats



Sky Scraper



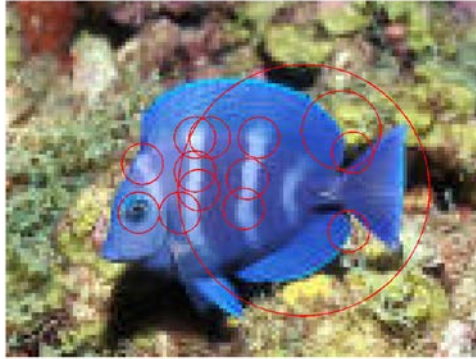
Car



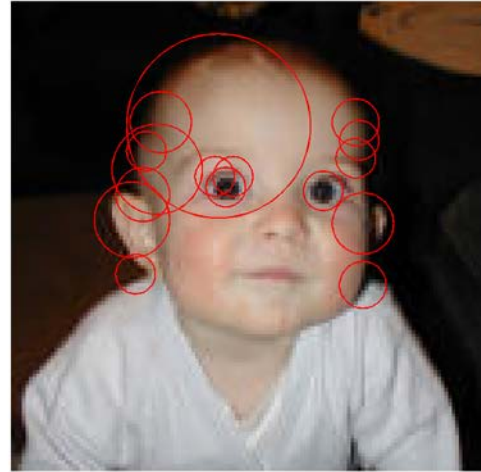
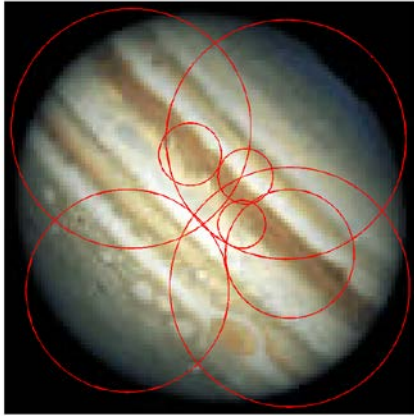
Trucks



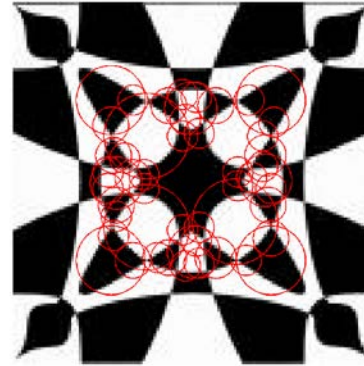
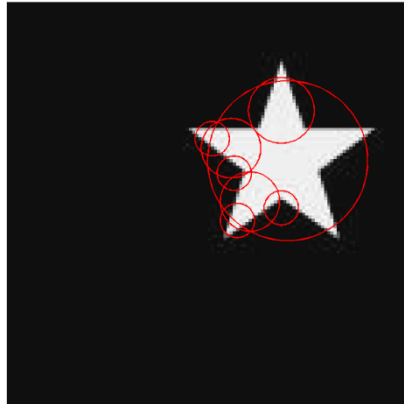
Fish



Other



Symmetry and More



Benefits

- General feature: not tied to any specific object
- Can be used to detect rather complex objects that are not all one color
- Location invariant, rotation invariant
- Selects relevant scale, so scale invariant
- What else is good?
- Anything bad?

References

Kadir, Brady; Saliency, scale and image description; *IJCV* 45(2), 83-105, 2001

Kadir, Brady; Scale saliency: a novel approach to salient feature and scale selection

Treisman; Visual coding of features and objects: Some evidence from behavioral studies; *Advances in the Modularity of Vision Selections*; NAS Press, 1990

Wolfe, Treisman, Horowitz; What shall we do with the preattentive processing stage: Use it or lose it? (poster); 3rd Annual Mtg Vis Sci Soc

References

Dayan, Abbott; Theoretical Neuroscience; MIT Press, 2001

Lamme; Separate neural definitions of visual consciousness and visual attention; Neural Networks 17, 861-872, 2004

Di Lollo, Kawahara, Zuvic, Visser; The preattentive emperor has no clothes: A dynamic redressing; *J Experimental Psych, General* 130(3), 479-492

Hochstein, Ahissar; View from the top: Hierarchies and reverse hierarchies in the visual system; Neuron 36, 791-804, 2002