Scale Saliency

Timor Kadir, Michael Brady

Pat Tressel
13-Apr-2005

The issues…

- Typical features…
  - Geometry: *gradients, filters, basis projection*
  - Morphology: *corners, blobs*
- …are not good for everything
  - Each specific to limited classes of objects
  - Poor recognition, poor scale detection
  - Could throw in many features, but slow

Instead, want features that are…

- Independent of specific object types
- Not fooled by (planar) warping
  - affine transformations
  - scaling
- Insensitive to intensity fluctuation
- Helps detect appropriate scale
- Usable with many underlying features
  - color, texture, gradient
  - optical flow

What to do?

- For general-purpose features...
  - Join the stampede – appeal to info theory
  - Define:
    salience = surprise = unpredictability = entropy
  - Doesn’t depend on a metric
  - Histogram low-level features around each point
  - Any low-level features will do:
    - intensity, color, texture, gradient
    - optical flow

What to do?

- To handle scale...
  - Histogram over simple region around point
  - Region size controlled by scale parameter
  - New cross-scale salience factor: how much histograms differ across scales
  - Search over scale for highest salience
- To handle planar transformations...
  - Use elliptical regions
  - Also search over orientation & eccentricity

Inference with the new input

- Goal is system identification – predict firing rate given a new input
  - Input is stimulus *and* last AP interval
- Given an input:
  - Compute the probability of membership in both classes
  - Use Bayes rule to get probability of spike:

$$p(spike \mid x) = \frac{p(x \mid spike)\, p(spike)}{\sum_{class} p(x \mid class)\, p(class)}$$

Finding salient points

- Define (raw, discrete) "scale saliency":

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

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

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

$\mathbf{x} = \text{point}$

$\mathbf{s} = (\text{s}, \text{r}, \theta) = (\text{scale, eccentricity, orientation})$

$D = \text{low-level feature domain}$

$p_{\mathbf{s},\mathbf{x}}(d) = \text{histogram of values of } D \text{ in region } \mathbf{s},\mathbf{x}$

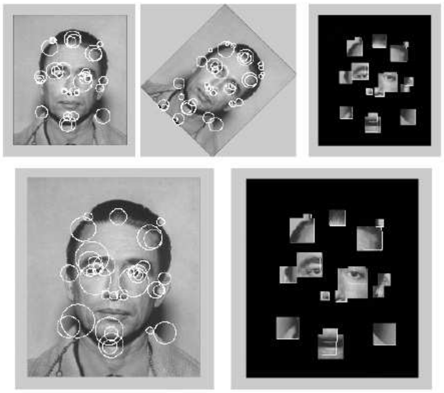
Finding salient points

- For each point & region shape, find maxima over scale
  - If monotonic, then none
- Over all points, keep most salient regions
  - E.g. top 10%, threshold

$$\mathbf{S} = \left\{ \begin{aligned} \mathbf{s},\mathbf{x} : &H_D((s-1,r,\theta),\mathbf{x}) < H_D((s,r,\theta),\mathbf{x}), \\ &H_D((s+1,r,\theta),\mathbf{x}) < H_D((s,r,\theta),\mathbf{x}), \\ &Y_D(\mathbf{s},\mathbf{x}) \text{ meets some cutoff criterion} \end{aligned} \right\}$$

Finding salient points - example

- Circular regions



Finding salient points - example

- Ellipses

Finding the salient points

- What underlying feature to use?
  - Feature is as random as possible at s.p.
  - So, no use for "describing" the points there
  - Elsewhere, single feature value is acceptable local match
- Want few salient points
  - Choose generally non-salient features
  - Use composite of these as underlying feature

Finding the salient points

- These *only* provide locations
  - Feature *D* is as random as possible there
  - No use for further "describing" the points
- They propose:
  - Repeat process with different feature
  - "At each level", use "more powerful" features
  - Yields "hierarchy of salient points"
- Combine nearby s.p.
- Annotate s.p. with other features

Using the salient points

- Tracking
  - Hand-select & crop each object in one frame
  - Find set of s.p. for each
  - Annotate with small image patches
- Segmentation
  - S.p. *opposite* of good region representatives
  - Fixup:
    - Pick points far from any s.p.
    - Grow regions starting there
    - Clusters of s.p. wall off regions

Benefits

- Not tied to specific object features
- S.p. sable across resizing
- Selects relevant scale

Issues

- “Salient” != object interest point
  - Noise is “salient”
  - Jumble of tiny objects is “salient”
  - Occluded object is “salient” at boundary
  - So not necessarily even object point
- “Salient” != salient
  - Periodic tiling (cougar spots) gets dense s.p.
  - But, it’s wallpaper, camouflage
  - Should be considered uniform

Issues

- Image resizing vs. zoom
  - Don’t want new s.p. during zoom
  - “Top n %” over smaller region adds points
  - Fixup:
    - Apply % at outset & get equivalent threshold
    - Stick with that threshold (at least through zoom)
    - Stable over resize with fixed % implies stable over zoom with threshold
- Not insensitive to variable illumination
  - Changes local statistics
  - Brighter yields higher salience

Issues

- Invariant under local pixel scrambling
  - Any arrangement within **s**,**x** region is same
- Two problems when using ellipses
  - Sensitive to noise
  - Slow – they're doing exhaustive search
  - Fixup: Try standard optimization

Meta-issues

- Much effort spent tying salience and...
  - Attentive / pre-attentive dichotomy
  - Operation of human visual system
  - Dropped *entirely* for summary paper
- Attentive / pre-attentive paradigm claims
  - Salience is main goal of low-level h.v.s.
  - Low-level h.v.s. features can't be orientation or scale sensitive
  - Can't depend on context

Meta-issues

- Couldn't be more wrong if they tried
- From neurobiology...
  - Main function of low-level h.v.s.:
    - Dimension reduction
    - "Fast", "cheap"
    - Appropriate for human tasks
  - Low-level h.v.s. features are *all* orientation, scale sensitive
  - Center / surround
  - Bar detectors
    - At various angles
    - Various speeds of bar movement

Meta-issues

- From neurobiology...
  - Yes, it's "context" dependent – it adapts
  - Values of features depend on local conditions
    - Aperture changes
    - Subconscious head motion to target important locations

Meta-meta-issues

- Why the disconnect?
  - Examine the “evidence”
  - Who cites whom?
- Postulate
  - There are distinct populations of researchers
    - Computer vision
    - Psychology
    - Machine learning
    - Neurobiology
    - Neurocomputation

Meta-meta-issues

- Postulate
  - Graph of relationships is sparse
    - Computer vision folks pay attention to psychology
    - Neurocomputation folks pay attention to neurobiology and machine learning
    - Psych folks aware of computer vision folks
  - Is change coming?
    - Neurobio folks have discovered what psych and comp vision folks are up to

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