Learning higher-order *structures* in natural images

Presented by
Santosh Kumar Divvala
Motivation for higher-order analysis

- Dependence amongst low-level features (i.e., correlated behavior of their magnitudes)
  - Low-level features => linear filters (ICA coefficients, texture, contours)

Figure 2.6: Magnitude dependence among pairs of coefficients.  

- Conditional histogram; each vertical slice shows the distribution of \( s_j \) conditional on values of \( s_i \) (shown on the ordinate, intensity rescaled independently for each slice). The “bow-tie” shape (originally described for wavelets in Simoncelli, 1997) indicates that variance of \( s_j \) increases with larger values of \(|s_i|\).  
- Probability density of \( s_j \) when \(|s_i| < .2 \) (red line) and \(|s_i| > 1.6 \) (blue line).  
- Joint distributions of coefficients are not consistent with factorial density functions (from Zetzsche and Röhrbein, 2001).
Motivation for higher-order analysis

• Low-level features vary within similar image regions

1. Efficient Coding
2. Good Discriminative Ability
Proposal

• Rather than encoding preferred features \((s)\), describe distributions \((v)\) over their inputs \((x)\)

\[
    x = As \quad \log(\sigma_s^2) = B v
\]

• “Variance coefficients \((v)\) provide an estimate of the probability density from which a sample ‘\(x\)’ was generated”

• “This density acts a statistical description of the context of the data sample”
ICA basis ($A_j$)

Variance of ICA coefficients ($\sigma_j$)

Figure 2.7: The variances of ICA coefficients change from context to context. Image patches were sampled from three regions in a natural image (top), and histograms for 7 ICA basis function coefficient computed (bottom).
Approach

• Parameters estimated in a MLE fashion using training data \((s)\)

\[
p(A, B|x) \propto \int_s p(x|A, s) \left( \int_v p(s|B, v)p(v)dv \right) p(A)p(B)ds
\]

\[
L = \log p(x|A, B)
\]

\[
\propto -\log |\det A| + \sum_i \left( -\frac{|B_{ij}|}{2} - \frac{\sqrt{2}s_i}{dB_{ij}/2} \right) - \sum |v_j|
\]

• Gradient descent

\[
v_{new} = v_{old} + \eta_v \frac{dL}{dv} \quad \text{where} \quad \frac{dL}{dv} = f(B_{old}, v_{old}, s)
\]

\[
B_{new} = B_{old} + \eta_B \frac{dL}{dB} \quad \text{where} \quad \frac{dL}{dB} = f(B_{old}, v_{new}, s)
\]
Result

- Captures the patterns of variability that underlie textures and other image elements
- Represents more abstract and invariant characteristics of image that are unobserved but significant
- Groups data points and finds statistically similar data
- Variance coefficients can be +ve or –ve, the components encode both indicated pattern and its converse

*Winner: Maximally active ‘v’ unit for that pixel*
Generalized Model

- Model the covariances distributions rather than only variances

Fig: Local image distributions in natural scenes show different correlational patterns. a. A natural scene with four distinct regions b. Each column shows the joint output of a pair of linear feature detectors (filters) from 20×20 image patches sampled from the different scene regions (rows 1-4). Both edges (row 1) and textures (rows 2-4) have high variability. Different visual features yield different distributions, but all of them overlap (row 5) and cannot be used to distinguish between the regions.
Result using generalized model

Result - 1
Variance Model

Result - 2
Covariance Model
Example winner maps on Images

- [http://www.cs.cmu.edu/~santosh/work/betterFeats/winnerMap.pdf](http://www.cs.cmu.edu/~santosh/work/betterFeats/winnerMap.pdf)
Discussion

• Does the theory hold for general set of features?
  – Are the low-level filters ‘linear’

**Location and Shape**
L1. Location: normalized x and y, mean
L2. Location: normalized x and y, 10th and 90th pctl
L3. Location: normalized y wrt estimated horizon, 10th, 90th pctl
L4. Location: whether segment is above, below, or straddles estimated horizon
L5. Shape: number of superpixels in segment
L6. Shape: normalized area in image

**Color**
C1. RGB values: mean
C2. HSV values: C1 in HSV space
C3. Hue: histogram (5 bins)
C4. Saturation: histogram (3 bins)

**Texture**
T1. LM filters: mean absolute response (15 filters)
T2. LM filters: histogram of maximum responses (15 bins)

**Perspective**
P1. Long Lines: (number of line pixels)/sqrt(area)
P2. Long Lines: percent of nearly parallel pairs of lines
P3. Line Intersections: histogram over 8 orientations, entropy
P4. Line Intersections: percent right of image center
P5. Line Intersections: percent above image center
P6. Line Intersections: percent far from image center at 8 orientations
P7. Line Intersections: percent very far from image center at 8 orientations
P8. Vanishing Points: (num line pixels with vertical VP membership)/sqrt(area)
P9. Vanishing Points: (num line pixels with horizontal VP membership)/sqrt(area)
P10. Vanishing Points: percent of total line pixels with vertical VP membership
P11. Vanishing Points: x-pos of horizontal VP—segment center (0 if none)
P12. Vanishing Points: y-pos of highest/lowest vertical VP wrt segment center
P13. Vanishing Points: segment bounds wrt horizontal VP
P14. Gradient: x, y center of mass of gradient magnitude wrt segment center