Natural scene statistics and visual inference

Tai Sing Lee
Computer Science Department &
Center for the Neural Basis of Cognition
Carnegie Mellon University

April 26, 2007

Mathematical Bioscience Institute, Ohio State University.

Outline of the talk

1. Conjectures on the role of feedback in cortex: hierarchical Bayesian inference.
2. Neural evidence of feedback (shape computation, priors and surprise).
3. Natural 3D scene statistics -- correlation between 3D structures and 2D images.
4. Neural evidence for the encoding of natural scene statistical priors.
5. Neural dynamics of inference under the influence of priors.

Jason Samonds  Matt Smith  Brian Potetz  Ryan Kelly
What is the functional role of feedback?

The integration of priors, context and memory in visual inference.
Conjectures:

• The visual cortex is a hierarchical graph for Bayesian inference.
• Neurons encode probability distribution of hypotheses or beliefs.
• Communication between neurons implement belief propagation.
• Neurons in the hierarchy can encode priors and higher order factors, constraining the rendering of images in the high-resolution buffer (V1 and LGN).

Lee and Mumford (2003) JOSA.

Perception of the 3D world

Shading, texture, perspective, occlusion …
Complexity of inference: multiple causes are involved in explaining a real image

3D Shape
- gross 3D geometry
- occlusion
- 3D texture

Illumination
- diffusion
- distance from scene
- intensity & color

Surface Material Properties
- surface color
- reflectance (shiny vs matte)

Camera Model

Model of Image Formation

2D Image

An example on 3D inference

Inference of hidden 2.5D sketch and rendering of images according to different illumination

Input image

Inferred depth map

Potetz (2007, CVPR)
Belief propagation in factor graph

- **Smoothness Nodes**: Encode prior on and smoothness nodes
- **Lambertian Constraint**: $p = \frac{\partial z}{\partial y}$
- **Integrability Constraint**: $q = \frac{\partial z}{\partial y}$

$\begin{align*}
q &= \frac{\partial z}{\partial y} \\
p &= \frac{\partial z}{\partial x}
\end{align*}$

$p_{x+1, y} - p_{x, y}$

Does the early visual areas exhibit sensitivity to higher order computation (as a result of feedback), and sensitivity to priors useful for Bayesian inference?
Perceptual grouping that is influenced by 3D perception

Ramachandran (1988), Science.

Strong priors of the interaction between lighting direction and shapes
Evidence of Shape from Shading computation

Weaker contrast
Strong pop-out

Stronger contrast
Weak pop-out

Sensitivity of V1 and V2 neurons to shape from shading
pop-out revealed in long-latency response

Smith, Kelly and Lee (submitted)

Effect very late.

(Smith MA, Bair W & Movshon JA (2006))
Perceptual asymmetry in LB and LA pop-out

LA pop-out  
LB pop-out

Pop-out slightly faster

Behavioral study (2 alternative force choice)

Fixation (150 ms)  
Stimulus (300 ms)  
Choice (300 ms)  
Saccade

Smith, Kelly and Lee (submitted)
Monkey behavior shows LB (concave) indeed pop-out "better".

Greater neural pop-out responses for LB (concave) oddball.
Findings

1. Early visual areas are sensitive to shape from shading -- a higher order inference beyond orientation and other basic tunings.
   - Pop-out response for 3D shapes but not for 2D patterns

2. Interaction between V2 and V1 -- feedback.
   - First observed in V2, and then in V1 only after monkeys were exposed to the task demands.
   - Latency distinct from (much later than) cross-orientation inhibition and surround suppression.
Findings

3. Asymmetry in convex and concave pop-out -- neural pop-out responses and behaviors

Why?

Conjecture:
A reflection of the statistical prior in the natural scene.
Natural scene statistics and **visual inference**

**James J. Gibson:** Ask not what’s inside your head, but what your head’s inside of!

Statistical priors on the relationship between 3D structures and 2D images in natural scenes

Riegl LMS-Z360  
Potetz and Lee (2003) JOSA
The most likely image associated with a convex shape is indeed a LA image patch.

\[
P[\text{convex}] = 0.69 \\
P[\text{convex} | \text{LA}] = 0.87 \\
P[\text{convex} | \text{LB}] = 0.43
\]

Findings:

- In natural scenes, convexity is more strongly associated with LA pattern, and appear to be more often than concavity (LB) patterns.

- Asymmetry in convex and concave pop-out -- neural pop-out responses and behaviors reflect the priors and surprise.

Consistent with Treisman’s general theory:

Unusual object pops out more readily among common objects than vice versa, or the idea that ‘surprise’ carries more information and attracts more attention.
A surprising finding: Da Vinci correlation

- Correlation between the center image pixel and the center range pixel across patches is negative.
- Brighter pixels tend to be near.

Surprise!

Correlation between intensity at center pixel (13,13) and distance at (x,y) -- the other pixels in the patch

Da Vinci Correlation

Among bodies equal in size and distance, that which shines the more brightly seems to the eye nearer. - Leonardo da Vinci

The first empirical demonstration of the source of this correlation based on natural image statistics.
Where does the effect come from?

(Potetz and Lee 2003 JOSA, Potetz and Lee 2006 NIPS)

Correlational structures between range and optical images are revealed in the cross power spectra(ZI)

Each radial frequency:

- Imaginary part $B_i(\theta)$
- Real part $B_R(\theta)$

$\text{Real}(ZI) = \frac{B_R(\theta)}{f^3}$

$\text{Imaginary}(ZI) = \frac{B_i(\theta)}{f^3}$

Log-Log polar plot of the real part of the cross-spectra.

Man-made structures

![Man-made structures image]

![Graph of $B_I(\theta)$ and $B_R(\theta)$]

Natural structures

![Natural structures image]

![Graph of $B_I(\theta)$ and $B_R(\theta)$]
Correlation is strong in rural scenes

Overall Rural scenes Urban scenes

Man-made structures: shading dominates

$B_{I}(\theta)$ shading

$B_{R}(\theta)$ shadows
Natural structures: strong shadow effect

The effects of shadows and concavity -- interaction of illumination and non-smooth 3D surfaces

(Potetz and Lee 2003 JOSA, Potetz and Lee 2006 NIPS)
Findings

1. Correlation between shading images and 3D shapes
   - Explanation for immediate perception of 3D shape based on shading (LA, LB).
   - Prevalence of convexity might explain asymmetry in pop-out perception and neural responses.
   - Da Vinci correlation strong in natural (rural scenes) and likely arise from shadows.
   - Power cross-spectrum decomposed into shading and shadow components.

An application to illustrate the importance of this correlation:

Inferring high-resolution 3D based on low-res 3D data (with the help of high-resolution optical image)
Results: useful for inference


Does the visual cortex encode this statistical regularity in the natural scenes?

Prediction:
Neurons prefer nearer surface should prefer brighter surface.
Measuring stereo and luminance tunings

Disparity tunings in V1 neurons

Gian Poggio (1981)
Correlation between depth and luminance tunings

(Potetz, Samonds and Lee 2006, Soc of Neuroscience abstract)
Conclusion

- Neurons prefer near disparity do prefer brighter surface at a population level.

- This correlation is consistent with a potential strategy of neural encoding of the “Da Vinci correlation” prior.

Image Range Map

(dark-light/near-far)
Correlation between the center pixel and the rest of the pixels in the luminance patch across all patches

Correlation between the center pixel and the rest of the pixels in the range data patch across all patches

Findings

1. Smoothness constraint -- the priors on spatial variation in luminance and depth.
2. Surface more smooth than luminance patterns.

Correlation in surface luminance

Correlation in surface depth
Smoothness prior is present in almost all computer vision algorithms

Marr and Poggio’s (1976) algorithm for solving the stereo correspondence problem:

1. *Compatibility*: Black dots can match only black dots.
2. *Uniqueness*: Almost always, a black dot from one image can match no more than one black dot from the other image.
3. *Continuity*: The disparity of the matches varies smoothly almost everywhere over the image.

Measuring interaction of neurons in response to 3D images -- dynamic random dot stereograms

Samonds, Potetz and Lee, NIPS (2007)
Quantifying Interactions

\[ C_{xy}(t_1, t_2) = \frac{\langle x(t_1)y(t_2) \rangle - \langle x(t_1) \rangle \langle y(t_2) \rangle}{\sqrt{\left( \langle x(t_1)x(t_1) \rangle - \langle x(t_1) \rangle \langle x(t_1) \rangle \right) \left( \langle y(t_2)y(t_2) \rangle - \langle y(t_2) \rangle \langle y(t_2) \rangle \right)}} \]
Temporal Dynamics of Synchrony

Population Temporal Dynamics of Correlation
Neurons with similar disparity tunings are positively coupled across space

\[ n = 115 \text{ pairs (separate electrodes)} \]

\[ n = 30 \text{ pairs (26%)} \]

\[ r = 0.46 \quad p = 0.01 \]

Neurons at the same location are also coupled, positively when similarly tuned and negatively when dissimilarly tuned.

\[ n = 17 \text{ pairs (same electrode)} \]

\[ n = 15 \text{ pairs (88%)} \]

Significant correlation ±10 ms
(same electrode)
Cooperation and competition between neurons

Findings

1. Positive “functional connectivity” among V1 neurons of similar disparity tunings across space -- consistent with the smoothness constraint or prior.

2. Negative “functional connectivity” among V1 neurons at the same spatial location -- consistent with the uniqueness constraint.

3. Rapid competition and long latency cooperation generates -- consistent with local competition and long-range cooperation.
Summary

Statistical study of natural scenes reveal a number of statistical regularities:

- Power-law smoothness priors
- Da Vinci correlation
- Correlation between 3D shapes and luminance patterns
- Prevalence of convexity

A variety of neurophysiological evidence demonstrating that the visual system is sensitive to these statistical regularities, and likely encoded in the recurrent connections.

Challenges: Integrating these and other shape priors for 3D inference using graphical model, and understanding the mechanisms and representation for such computation.

Acknowledgment

Research supported by National Science Foundation and National Institute of Health.

Jason Samonds  Matt Smith  Brian Potetz  Ryan Kelly