The Cortical Basis of Motion Perception

PSYCH-508: Core Concepts in Perception Michael Beyeler 11/1/18

From Retinal Input to Perception



Hierarchical Processing of Vision



Visual Motion Pathway

- Primary visual cortex (V1)
 - tuned to simple attributes of shape, motion, color, texture, depth
- Middle temporal (MT) area
 - tuned to coherent local motion (retinal flow)
- Medial superior temporal (MST) area
 - tuned to global, complex motion
 - self-motion (MSTd), object motion (MSTv)
 - tuned to visual, vestibular cues
- Posterior parietal cortex (PPC)
 - polysensory: visual, vestibular, somatosensory, auditory cues
 - self-motion vs. object motion
 - spatial planning, path integration(?)
 - eye, arm, head movements



(Britten, Annu Rev Neurosci, 2008)

V1 simple cells





Hubel & Wiesel, 1968

Space-time oriented receptive fields



Space-time oriented receptive fields



Spatiotemporal-energy models

- Three steps:
 - 1. Linear filtering
 - 2. Motion energy
 - 3. Opponent energy
- 1. Linear filtering:
 - motion is an orientation in space-time
 - V1 simple cells: spacetime oriented receptive fields



Spatiotemporal-energy models

- 2. Motion energy
 - Constant grating motion gives oscillating output → not a useful indicator of motion
 - Measure the energy of the signal instead:

$$E_s = \int_{-\infty}^{\infty} |x(t)|^2 dt$$

 Shortcut: Square and sum quadrature pair to remove phase dependence from the output



Simoncelli & Heeger model



Simoncelli & Heeger model

- What if multiple motion components are present in the stimulus?
- Component-directionselective (CDS):
 - responds to both components in the plaid
- Pattern-directionselective (PDS) cell:
 - responds to global/pattern/perceived direction of motion



Aperture problem

- Problem: Localvelocity sample is different from the object velocity
- Goal: Disambiguate local-velocity samples and integrate them into an accurate estimate of the global (object) velocity



Aperture problem

- Problem: Localvelocity sample is different from the object velocity
- Goal: Disambiguate local-velocity samples and integrate them into an accurate estimate of the global (object) velocity



Aperture Problem

- Problem: Local-velocity samples are intrinsically ambiguous
- Solution: Infer global (object) velocity from many local-motion samples
- Vector average
- Feature tracking
- Intersection-of-constraints (IOC):
 - Each local velocity sample *constrains* the global object velocity
 - Find object velocity by integrating local samples
 - MT firing rates may represent the velocity of moving objects using IOC



(Bradley & Goyal, Annu Rev Neurosci, 2008)

Spatiotemporal energy models

- Toward Fourier transform:
 - a) horizontal slice reveals a sinusoidal modulation of intensity vs. x (ω_x)
 - b) same for y (ω_y)
 - c) looking at a fixed location as the grating moves (ω_t)
- The velocity (direction and speed) of the grating is completely characterized by a single 3–D frequency: (ω_x, ω_y, ω_t)



Born & Bradley, 2005

Spatiotemporal energy models

- Space-time oriented filters:
 - Space-time domain: Gabor function
 - Fourier domain: "fuzzy blob", the density of which decreases with distance from the center frequency



(Bradley & Goyal 2009)

Aperture problem in the Simoncelli & Heeger model

- The Simoncelli & Heeger (1998) model provides an IOC solution to the aperture problem
- Each 1D local-motion measurement (by a CDS cell) is consistent with a number of possible 2D velocities, all of which must fall on a line in velocity space
- A PDS cell in MT is created by summing the outputs of CDS cells with spectra centered on a common plane



Middle temporal (MT) area

- Rust et al. (2006):
 - Presented sequence of random pattern stimuli
 - Made from 12 gratings drifting in different directions
 - Fit to L-N cascade model
 - Explained full range of pattern motion selectivity found in MT



L-N cascade model



L-N cascade model



(Rust et al 2006)

3D visual response properties of MSTd emerge from an efficient, sparse population code

Michael Beyeler^{1,2}, Nikil Dutt¹, Jeffrey L. Krichmar¹

¹University of California, Irvine ²University of Washington



Retinal optic flow

 Apparent motion on the retina caused by relative movement between observer and environment



(Raudies & Neumann, CVIU, 2012)

Motion field model



• Project 3D world points $\vec{P} = (X, Y, Z)^T$ onto 2D image points $\vec{p} = (x, y)^T$

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \frac{1}{z} \begin{bmatrix} -f & 0 & x \\ 0 & -f & y \end{bmatrix} \begin{bmatrix} v_X \\ v_Y \\ v_Z \end{bmatrix} + \frac{1}{f} \begin{bmatrix} xy & -(f^2 + x^2) & fy \\ f^2 + y^2 & -xy & -fx \end{bmatrix} \begin{bmatrix} \omega_X \\ \omega_Y \\ \omega_Z \end{bmatrix}$$
translation
$$\text{rotation}$$

$$\text{(Longuet-Higgins \& Prazdny, Proc Royal Soc London, 1980)}$$

MSTd response properties

- Main visual input from MT
- Large receptive fields (up to 40° × 40°)
- encodes visual and vestibular signals
- responds to optic flow
 - responds to 3D observer translations (heading)
 - responds to 3D observer rotation
 - act like template matches
- encodes eye position / eye velocity



Heading template model of MSTd

- MSTd units receive input from a mosaic of MT local-motion detectors
- Detected heading corresponds to the position of the most active MSTd unit
- Many models have followed this idea
 - Perrone & Stone (1994,1998)
 - Lappe et al. (1996)
 - Beintema & van den Berg (2000)
 - Browning et al. (2009)
- Limitations:
 - Combinatorial explosion of required heading templates
 - Heading in MSTd is a population code
 - Cannot explain some of the less intuitive response properties







(Perrone & Stone, J Opt Soc Am A,1992)

Complex motion selectivity in MSTd

- MSTd responds to complex motion
 - No perfect heading templates
 - Neurons prefer an intricate mixture of flow components
 - Tuning varies in different subfields of the receptive field
- MSTd encodes vestibular signals
- MSTd codes in both retinal and head-centered coordinates



(Mineault et al., PNAS, 2012)

Complex motion selectivity in MSTd

- MSTd simultaneously encodes multiple perceptual variables
 - Distributed representation of heading, eye position, eye velocity
 - Most neurons participate in encoding multiple variables ("basis functions")
 - Nonlinear interactions



• How can these nonintuitive response properties be explained?

(Ben Hamed et al., *J Neurophys*, 2003; Mineault et al., *PNAS*, 2012)

Efficient coding of optic flow

- What if MSTd is trying to find an efficient encoding of visual input from MT?
 - such that any optic flow pattern / relevant perceptual variable can be decoded from a small number of MSTd neurons (accuracy)
 - such that any given stimulus activated only a small number of MSTd neurons (population sparsity)
 - such that any MSTd neuron responded to only a small number of stimuli (lifetime sparsity)

Basis function representations

- Form of population code
- Allows for flexible intermediate representations
- Most functions of interest can be computed using a linear combination of basis functions
- Examples of a basis set:
 - Coordinate axes in 3D space
 - Fourier transform: Any function = linear sum of a series of sin / cos
 - Radial basis functions (Gaussians)
- Many different basis sets possible...
 - Cartesian (x,y,z)? Spherical (r, θ , ϕ)?
 - Which one to choose?
 - What are the constraints?



(Pouget & Snyder, *Nat Neurosci*, 2000)

Nonnegative matrix factorization (NMF)

- Dimensionality reduction technique
 - Similar to PCA, ICA
 - Unsupervised statistical learning
 - Reconstruct input data matrix from product of two much smaller matrices
- Use NMF to find basis set
- V: Observed data
- W: Basis vectors
- H: Hidden coefficients
- Linear decomposition: V ≈ WH
- NMF minimizes the following cost function:

$$c(\mathbf{W}, \mathbf{H}) = \|\mathbf{V} - \mathbf{W}\mathbf{H}\|^2$$

$$\forall ij: W_{ij} \ge 0, H_{ij} \ge 0$$



- F: features
- S: stimuli (observations)
- B: basis vectors

(Paatero & Tapper, *Environmetrics*, 1994; Lee & Seung, *Nature*, 1999)

NMF applied to face images



Sparse decomposition model of MSTd

- Input stimuli: All naturally occurring optic flow fields
- MT-like units: local motion sensors (direction / speed)
- MSTd-like units: small basis set of complex motion templates



(Beyeler et al., J Neurosci, 2016)

NMF applied to optic flow

NMF



November 2, 2018

Using NMF to model neuronal response properties





Takahashi et al., JNeurosci, 2007



(Beyeler et al., J Neurosci, 2016)

3D translation / rotation selectivity

- All 3D directions are represented
- Bias towards lateral headings
 - Only few neurons respond to straight-ahead headings
 - Only few neurons respond to roll rotation



(Beyeler et al., *J Neurosci*, 2016)

Decoding heading from the population response

- Decode (x_{FOE}, y_{FOE}) of arbitrary expansive flow fields from a population of *N* MSTd-like model units
 - linear regression, cross-validated
 - randomly sampled 10,000 flow fields
 - azimuth between
 [45°, 135°], elevation
 between [-45°, +45°]
 - learn a set of $N \times 2$ linear decoding weights



Sparse population code for heading



Encoding of multiple perceptual variables

Translation	FOE (x, y)	Eye velocity (ω_x, ω_z)
Ben Hamed et al. (2003)	$(3.62^{\circ} \pm 6.78^{\circ}, 3.87^{\circ} \pm 4.96^{\circ})$	(1.39° ± 3.69°, 1.38° ± 3.02°)
Sparse decomposition model	$(5.75^{\circ} \pm 5.62^{\circ}, 6.02^{\circ} \pm 5.51^{\circ})$	$(0.82^{\circ} \pm 0.89^{\circ}, 0.92^{\circ} \pm 0.99^{\circ})$

- A strong decoding weight suggests a neuron contributes strongly to the decoding
 - most model units involved in encoding both FOE and P
 - same is true for neurons in MSTd
- These neurons are generalists, not specialists

FOE = focus of expansion P = pursuit (eye movement velocity)



(Beyeler et al., J Neurosci, 2016)

Heading discrimination

- Heading discrimination is best around straight-ahead headings
- MSTd is causally linked to heading perception
 - behavioral performance due to neurons preferring straight-ahead headings with sharp tuning curves? (Duffy & Wurtz, 1995)
- But most MSTd neurons prefer lateral headings!



(Gu et al., Neuron, 2010)

Population code underlying heading discrimination

- Calculate tuning curves:
 - Preferred direction: Heading where neuronal activity is maximal (peak of tuning curve)
 - Tuning width: Spread of tuning curve at half-maximum
 - Peak discriminability: Heading where derivative of tuning curve is maximal



(Beyeler et al., *J Neurosci*, 2016)

Population code underlying heading discrimination



subset of MSTd neurons preferring forward headings with sharp tuning curve

(Beyeler et al., *J Neurosci*, 2016)

Conclusion

- Sparse decomposition model of MSTd
 - Offers a biologically plausible account of a wide range of visual response properties ranging from single-unit selectivity to population statistics
 - Suggests that most visual response properties are a by-product of MSTd neurons performing dimensionality reduction on their inputs
 - Provides a further step towards a scientific understanding of nonintuitive MSTd response properties

Sparseness

Sparseness metric s for a signal r with N sample points (Vinje & Gallant, 2000):

$$s = \left(1 - \frac{1}{N} \frac{(\sum_i r_i)^2}{\sum_i r_i^2}\right) / \left(1 - \frac{1}{N}\right)$$

- Sparseness: $s \in [0,1]$
 - s = 0: minimal sparseness (dense code)
 - s = 1: maximal sparseness (local code)

- Population sparseness:
 - How many MSTd-like model units were activated by any given stimulus
 - r_i : response of *i*-th neuroto a particular stimulus
 - *N*: number of model units
- Lifetime sparseness:
 - How many stimuli any given MSTd-like model unit responded to
 - r_i : response of *i*-th stimulus
 - *N*: number of stimuli

Population Fisher Information

 Provides an upper limit on the precision with which an unbiased estimator can discriminate small variations in a variable (x) around a reference value (x_{ref}):

$$I_{\rm F}(x_{\rm ref}) = \sum_{i=1}^{N} \frac{\left(R_i'(x_{\rm ref})\right)^2}{\left(\sigma_i(x_{\rm ref})\right)^2}$$

- *N*: number of neurons in the population
- R'_i(x_{ref}): derivative of tuning curve for the *i*-th neuron at x_{ref}
- $\sigma_i(x_{ref})$: variance of *i*-th neuron's response at x_{ref}

(Pouget et al., Neural Comput, 1998; Seung & Sompolinsky, PNAS, 1993)

Population Code for Heading Discrimination

- Discrimination best around straight-ahead headings
- Traditional view: Performance due to neurons tuned to straight– ahead headings
- Gu et al. (2010): Performance due to neurons tuned to lateral headings
 - Broad tuning curves
 - Slope is max near straight-ahead max derivative
 max activity



- Consider two input stimuli with similar headings (e.g., 0°, 4°)
- Neurons with preferred direction (peak response) near 0° will contribute little to discrimination
 - Small heading deviations do not significantly alter response amplitude => not informative
 - Derivative ≈ 0
- Neurons whose tuning curve slope is steepest near 0° will contribute a lot to discrimination
 - Near steepest slope, small heading deviations cause large changes in response amplitude
 - Derivative is max
 - => "Peak discriminability"

Nonnegative sparse coding (NSC)

- Combination of dimensionality reduction and sparse coding
 - represents observed data V with a small number of dictionary elements in W
 - such that all dictionary elements are nonnegative
 - such that elements are sparsely activated (H)

$$\begin{split} & \min_{\mathbf{W},\mathbf{H}} \frac{1}{2} \|\mathbf{V} - \mathbf{W}\mathbf{H}\|^2 + \lambda \sum_{ij} f(H_{ij}) \\ & \text{nonnegative matrix} \\ & \text{factorization (NMF)} \\ & \text{V: Observed data} \quad & \text{W: Basis vectors} \quad & \text{H: Hidden coefficients} \end{split}$$

- can explain response properties in V1, V2
- shown to be equivalent to spike-timing dependent plasticity (STDP) and homeostatic synaptic scaling

(Hoyer, NNSP, 2002; Hoyer, JMLR, 2004; Carlson et al., J/CNN, 2013)

Nonnegative sparse coding (NSC)

- Sparse and (potentially) parts-based representations have been found throughout the brain
- Could NSC be a general principles to which neuronal computations adhere?

		Parts-	Experimental	Modeled	Computational
\mathbf{Area}	Sparse	based	evidence	by NSC	$\operatorname{support}$
Retina	Х	Х	[27]	Х	[27]
Early visual cortex	Х	Х	[17, 19, 28, 29]	Х	[17, 19, 30, 31]
Ventral visual stream	Х	Х	[20]	Х	[14, 32]
Dorsal visual stream	Х	Х	[33–35]	Х	[22]
Auditory cortex	Х	?	[36]	?	?
Olfactory cortex	Х	?	[37]	?	[38]
Retrosplenial cortex	Х	Х	[26]	Х	present paper
Motor cortex	Х	?	[39, 40]	?	[41]
Basal ganglia	Х	?	[3, 42]	Х	[3], advanced RDDR
Barrel cortex	Х	?	[43]	?	?

Table 1: Nonnegative sparse coding in the brain

(Beyeler et al., bioRxiv, 2017