Hierarchical inference of disparity

REDWOOD CENTER for Theoretical Neuroscience

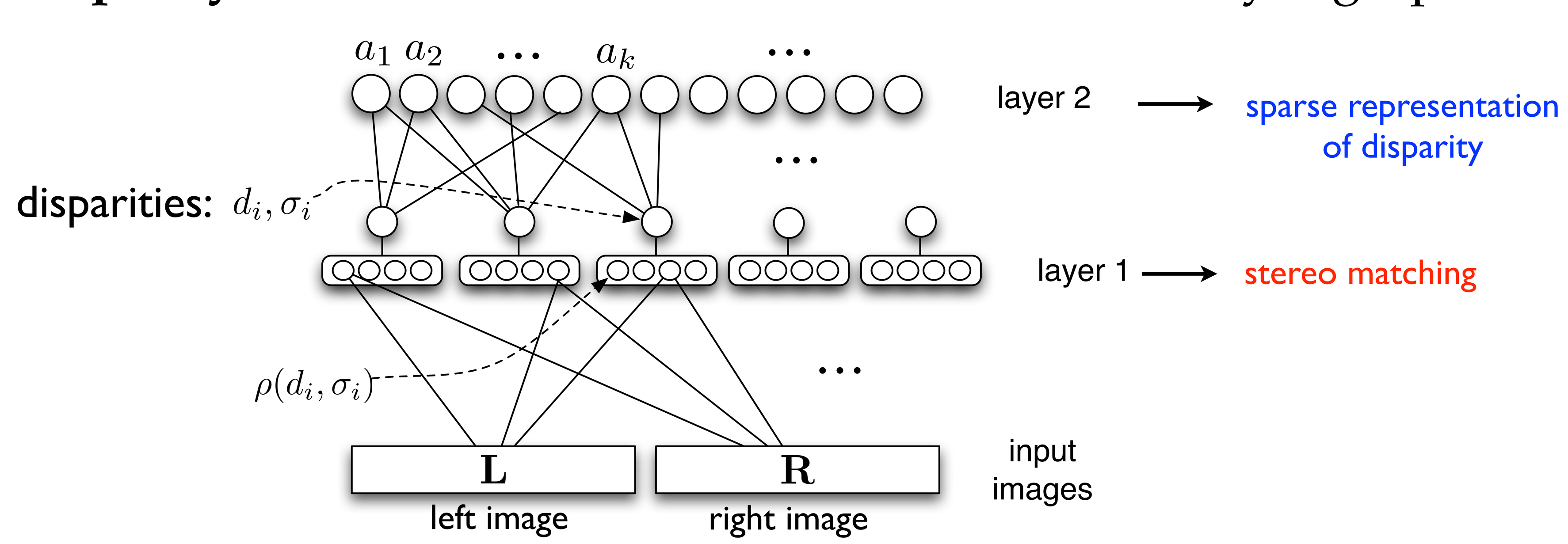
Ivana Tošić and Bruno A. Olshausen, *University of California, Berkeley*

Motivations

- The stereo correspondence problem is hard to solve due to ambiguous matches between similar image features
- Disparity selective cells in V2 respond less to false matches than cells in V1 [1]:
  - hierarchical processing might have a role in resolving ambiguities
  - existing models do not exploit this hierarchy [2,3]
- We propose **a two-layer graphical model for disparity inference**
  - upper layer nodes act as priors that disambiguate false from correct matches
  - model parameters learned from natural disparities

Model

- **Disparity estimation:** iterative inference on a two-layer graph



- **Disparity sparse generative model with uncertainty**
  - disparity estimation/measurement is erroneous: uncertainty in the model

$$\boxed{\mathbf{d} = \Psi \mathbf{a} + \boldsymbol{\epsilon} + \boldsymbol{\eta}}$$

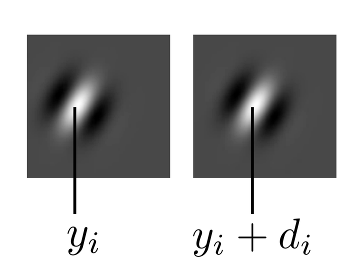
sparse representation; non-stationary uncertainty $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \Sigma)$, $\Sigma = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)$; approximation error $\boldsymbol{\eta} \sim \mathcal{N}(0, \sigma_0^2)$

- **Stereo matching likelihood**

$$\boxed{P(\mathbf{L}, \mathbf{R} | \mathbf{d}, \Sigma) \propto e^{-\sum_i \rho(d_i, \sigma_i)}}$$

- **where:** $\rho(d_i, \sigma_i) = \frac{1}{\sigma_i^2} \sum_\theta [\langle \mathbf{L}, \phi_\theta(x_i, y_i) \rangle - \langle \mathbf{R}, \phi_\theta(x_i, y_i + d_i) \rangle]^2$

$\phi_\theta(x_i, y_i)$: Gabor atom at position (x_i, y_i), with orientation θ



- **Inference of disparity:** MAP formulation

$$P(\mathbf{d}, \Sigma, \mathbf{a} | \mathbf{L}, \mathbf{R}) \propto \underbrace{P(\mathbf{L}, \mathbf{R} | \mathbf{d}, \Sigma, \mathbf{a})}_{\text{stereo matching likelihood}} \underbrace{P(\mathbf{d} | \Sigma, \mathbf{a}) P(\Sigma) P(\mathbf{a})}_{\text{disparity prior}}$$

Inference of disparity

- Inference on the proposed graph with the following priors:
  - Jeffreys prior for uncertainty: $P(\Sigma) = \prod_{i=1}^N 1/|\sigma_i|$
  - Laplace prior for coefficients: $P(\mathbf{a}) \propto e^{-\lambda \sum_k |a_k|}$

- **Iterative algorithm**

1. Initialize disparity d_i at each pixel
2. Infer sparse coefficients $\mathbf{a}$ and σ_i's with non-stationary sparse coding [4]

$$\min_{\mathbf{a}, \{\sigma_i\}} \sum_{i=1}^N \left[\log \sigma_i^2 + \frac{(d_i - \hat{d}_i)^2}{2\sigma_i^2} \right] + \lambda \|\mathbf{a}\|_1, \quad \text{where } \hat{\mathbf{d}} = \Psi \mathbf{a}$$

3. Inference of disparity using gradient descent

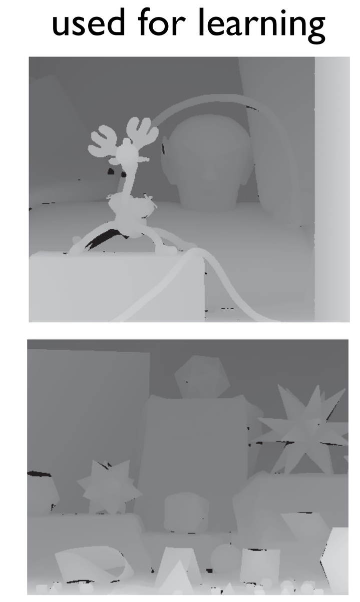
$$\min_{\mathbf{d}} \sum_{i=1}^N \left[\frac{(d_i - \hat{d}_i)^2}{2\sigma_i^2} + \rho(d_i, \sigma_i) \right], \quad \text{where } \hat{\mathbf{d}} = \Psi \mathbf{a}$$

4. Back to 2 or end if convergence

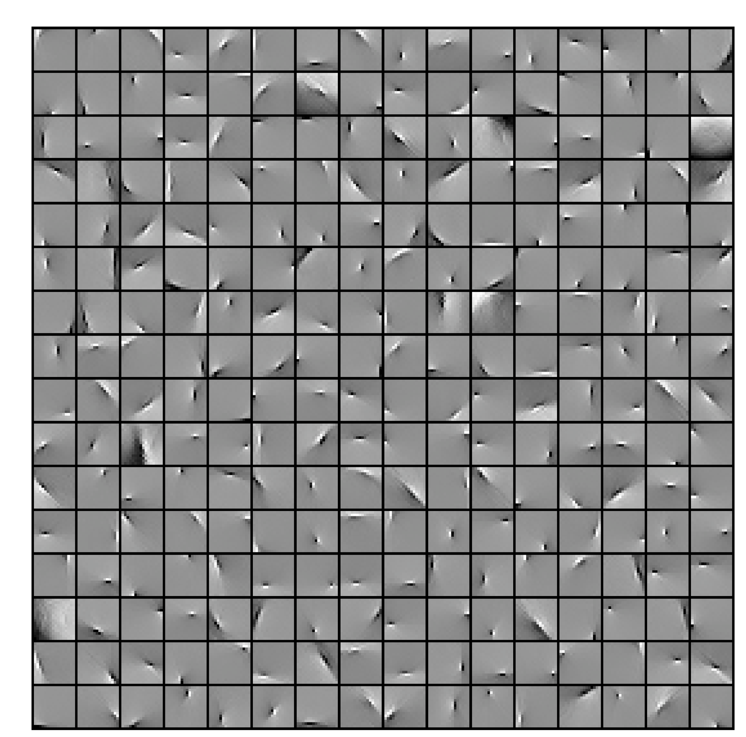
Learning disparity representation

- For disparity inference, we learn a dictionary Ψ
- Learning: non-stationary sparse coding [4]
  - on the disparity maps from the Middlebury database
  - unwhitened data
  - patches 16x16 pixels

examples of disparity maps used for learning



Learned dictionary for disparity

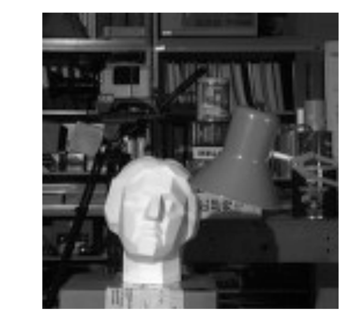


- Learned dictionary:
  - mostly edges (consistent with disparity RF's in V2 [5])
  - some slanted surfaces

Results

- Tsukuba and Venus datasets (not in the training set)
- Initialization: integer disparity values that give minimal stereo matching term

Tsukuba experiment: Gabor parameters: envelope 1, wavelength 1, eight orientations



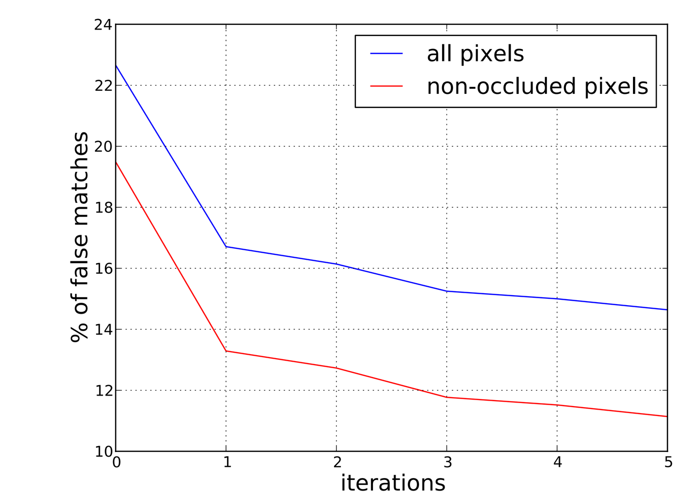
left image (128x128)

ground truth

layer 1 only (22.8% false matches)

layers 1 and 2 (14.6% false matches)

percentage of false matches, accuracy 1 pixel

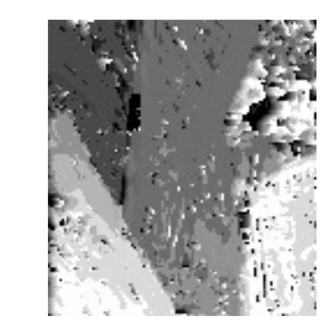


The percentage of false matches decreases with iterations (up to 35%-42% less false matches)

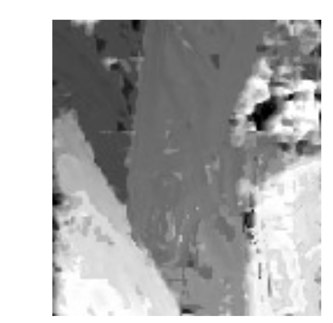
Venus experiment: Gabor parameters: envelope 2, wavelength 8, eight orientations

left image (128x128)

ground truth



layer 1 only (22.9% false matches)



layers 1 and 2 (18.4% false matches)

Conclusions

- Contributions:
  - a new hierarchical model for disparity inference based on natural disparity statistics
  - learned representation of disparity consistent with electrophysiology
- Future work:
  - competition between disparity detectors in the first layer
  - learning from human data: stereo views dependent on fixation
  - learning joint statistics of disparity and natural images

- References:

[1] J.S. Bakin, K. Nakayama and C.D. Gilbert, "Visual Responses in Monkey Areas V1 and V2 to Three-Dimensional Surface Configurations", *Journal of Neuroscience*, vol. 20(21), 2000.
[2] D. Marr and T. Poggio, "Cooperative computation of stereo disparity", *Science*, vol. 194(4262), 1976.
[3] J. Read and B. Cumming, "Sensors for impossible stimuli may solve the stereo correspondence problem", *Nature Neuroscience*, vol. 10(10), 2007.
[4] I. Tosic, B.A. Olshausen and B.J. Culpepper, "Learning representations of depth", *Journal on Selected Topics in Signal Processing,* submitted, 2010.
[5] R. von der Heydt, H. Zhou and H. S. Friedman, "Representation of stereoscopic edges in monkey visual cortex", *Vision Research*, vol. 40(15), 2000.

FNSNF Swiss National Science Foundation

This work has been supported by the Swiss National Science foundation under fellowships PBELP2-127847 and PA00P2_134159.