Perception and the statistics of natural scenes

Yair Weiss
School of Computer Science and Engineering
The Hebrew University of Jerusalem
www.cs.huji.ac.il/ ~yweiss

Collaborators: Anat Levin, Assaf Zomet

Outline

- Perception and statistics.
- Perception of transparency.
- Some applications to computer vision.

Perception and statistics

The idea that the statistics of the sensory stimuli we receive from the environment are important for perception and cognition is not new, and surprisingly clear statements about it can be found before 1950 in the writings of Mach (1886), Pearson (1892), Helmholtz (1925), Craik (1943) and others.

(Barlow 2001)

Rennaisance of Bayesian approach to perception

Books, workshops, journal special issues.

> The past decade has seen an explosion of interest in the application of statistical techniques to modeling vision in both biological and artificial systems. $\cdots$ The mathematics of statistics, stochastic processes and Bayesian inference provide the natural framework for understanding these aspects of visual processing.

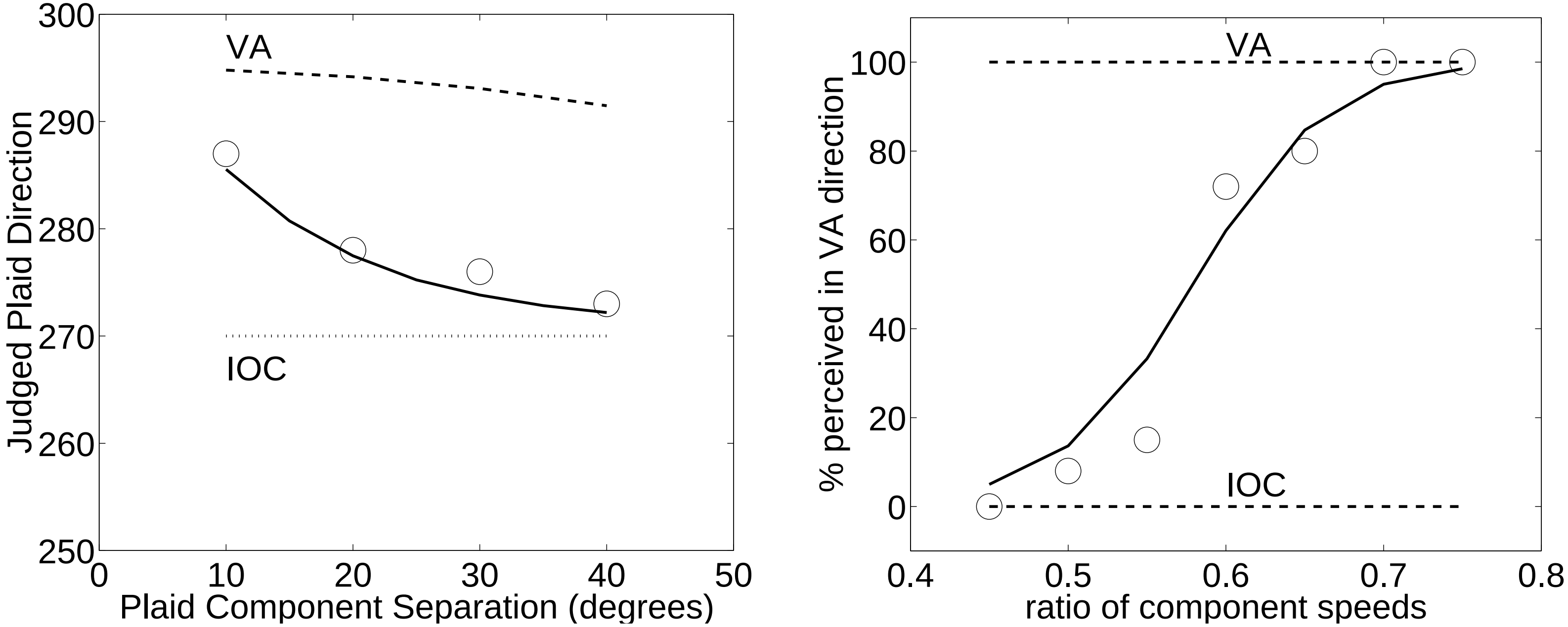
(CFP JOSA A special issue summer 2003)

"Bayesian" approach to perception

$$\mathsf{Pr}(percept|image) = \frac{\mathsf{Pr}(percept)\,\mathsf{Pr}(image|percept)}{\mathsf{Pr}(image)}$$

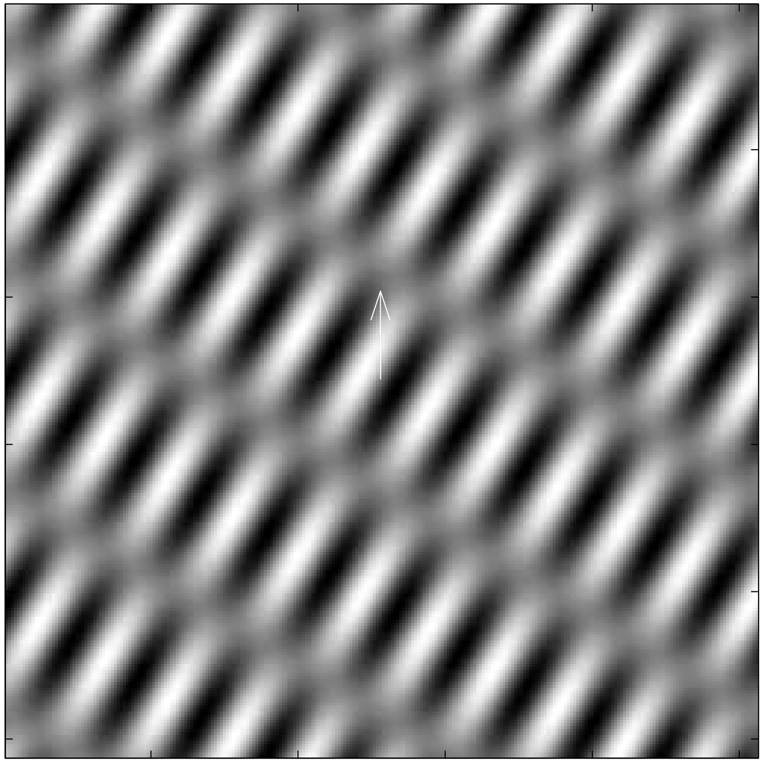
- $\mathsf{Pr}(percept)$ prior.
- $\mathsf{Pr}(image|percept)$ likelihood.

Quantitative predictions

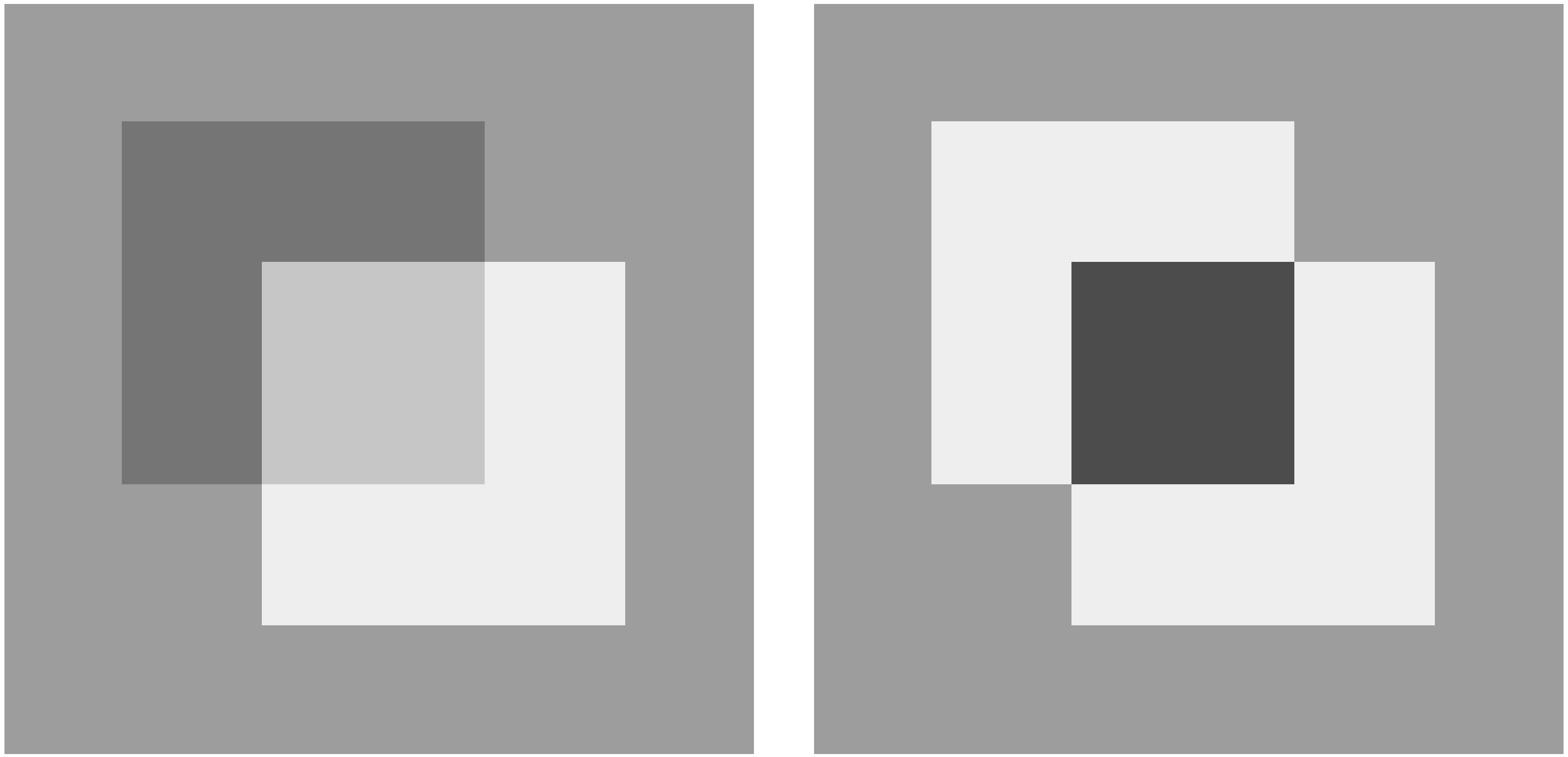


(Weiss et al. 2002)

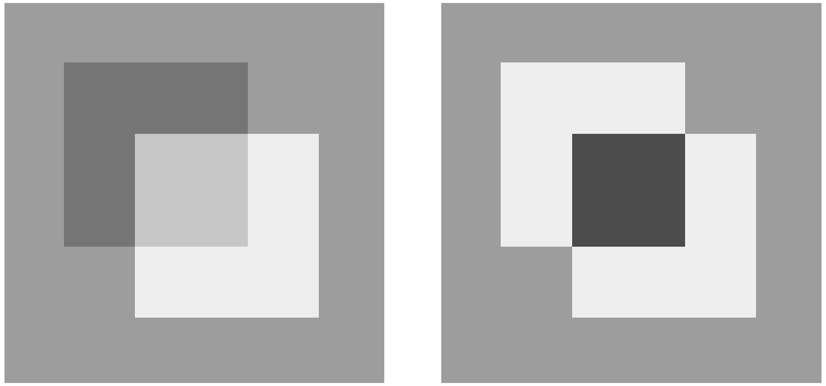
Where do priors come from?



Transparency

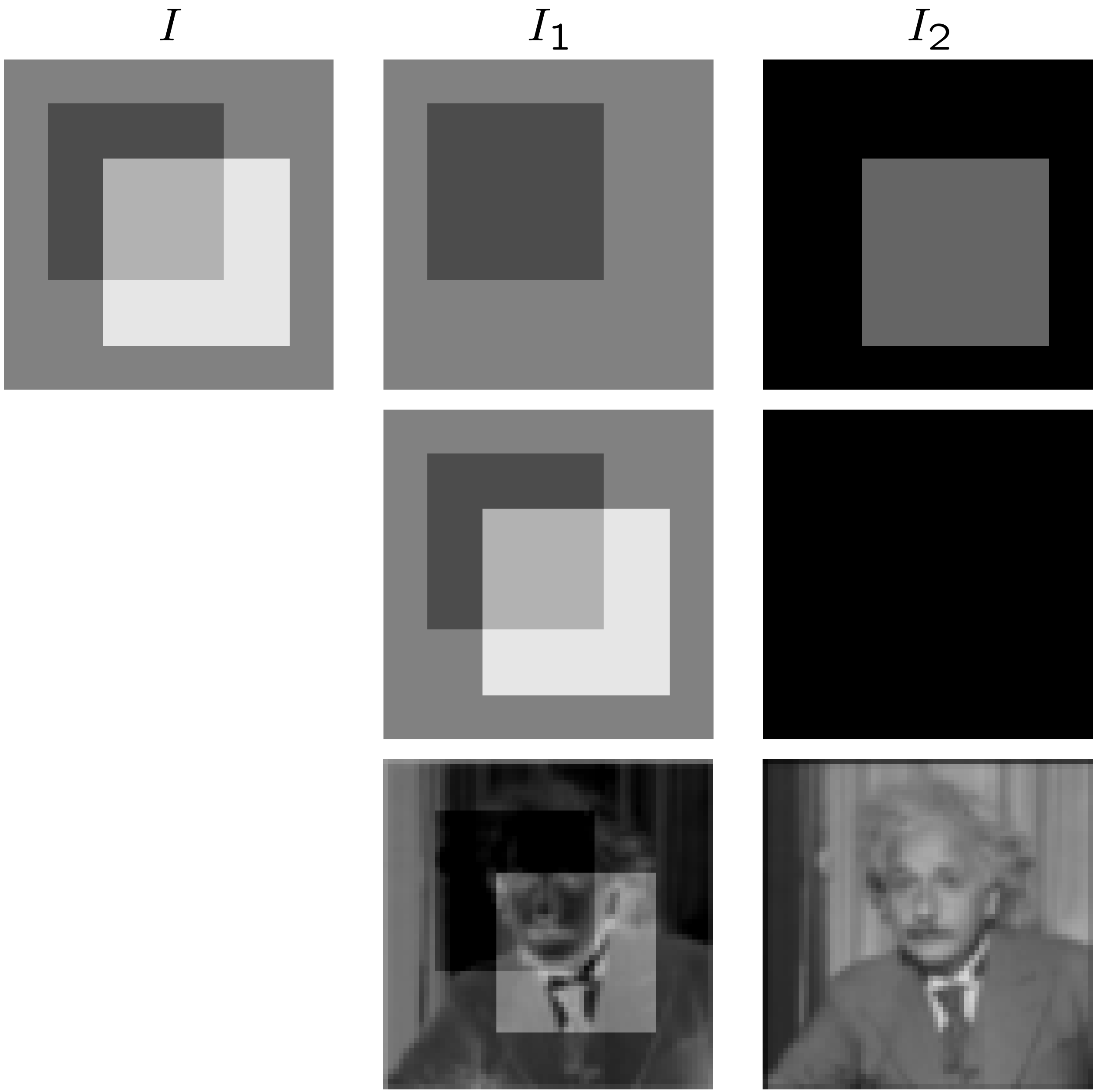


Puzzle of Transparency

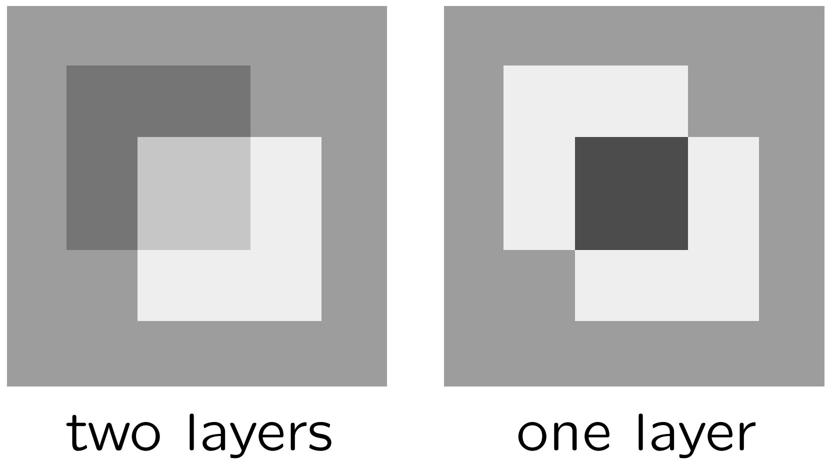


$$I(x,y) = I_1(x,y) + I_2(x,y)$$

- why not "simpler" one layer solution?
- which two layer out of infinite possibilities ?


I
I_1
I_2

Transparency in human vision



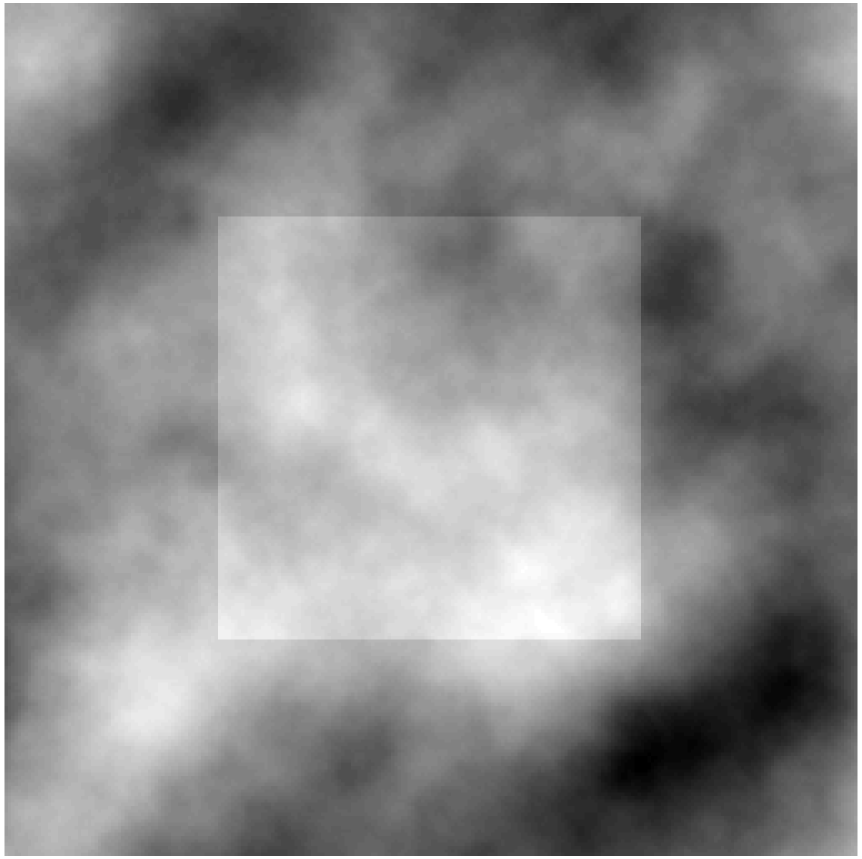
- Metelli's conditions (Metelli 74)

- T-junctions, X-junctions, doubly reversing junctions (Adelson and Anandan 90, Anderson 99)

Not obvious how to apply "junction catalogs" to real images.

Transparency in the “real world”

Junction catalogs ?



(Bart Anderson)

Our approach

$$I(x, y) = I_1(x, y) + I_2(x, y)$$

Probability distribution $\text{Pr}(I_1), \text{Pr}(I_2)$ Percept is most probable decomposition.

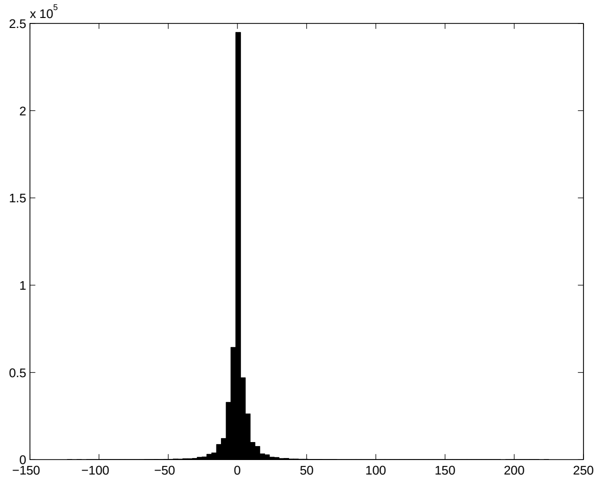
Can we make this work?

Assumption: visual system summarizes statistics of images via statistics of simple features.

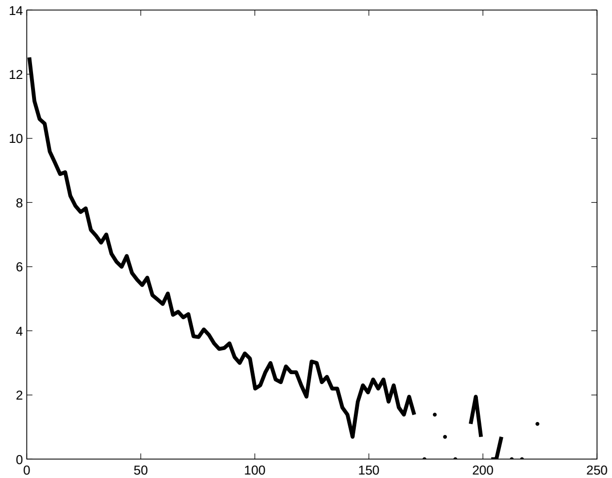
(e.g. Zhu and Mumford 97, Della Pietra Della Pietra and Lafferty 96, Heeger Bergen 95).

Statistics of derivative filters

image

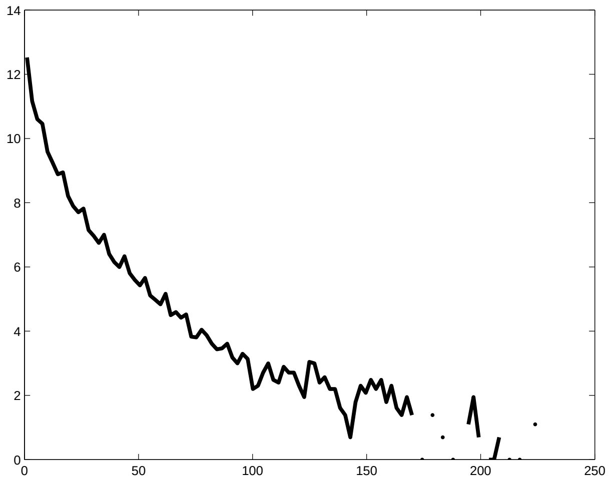


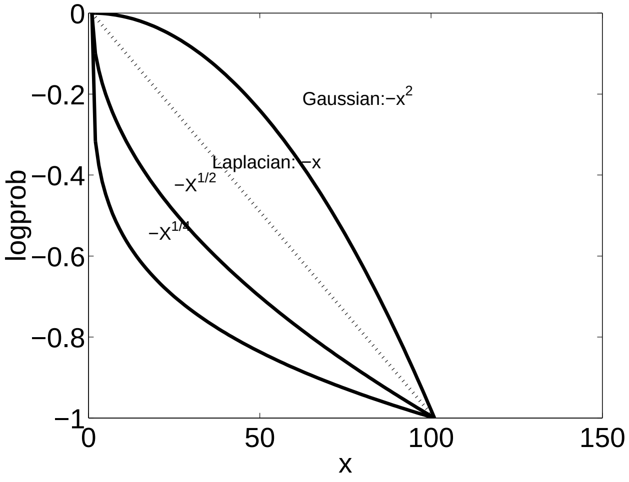
dx hist



dx log hist

Sparsity of derivative filters



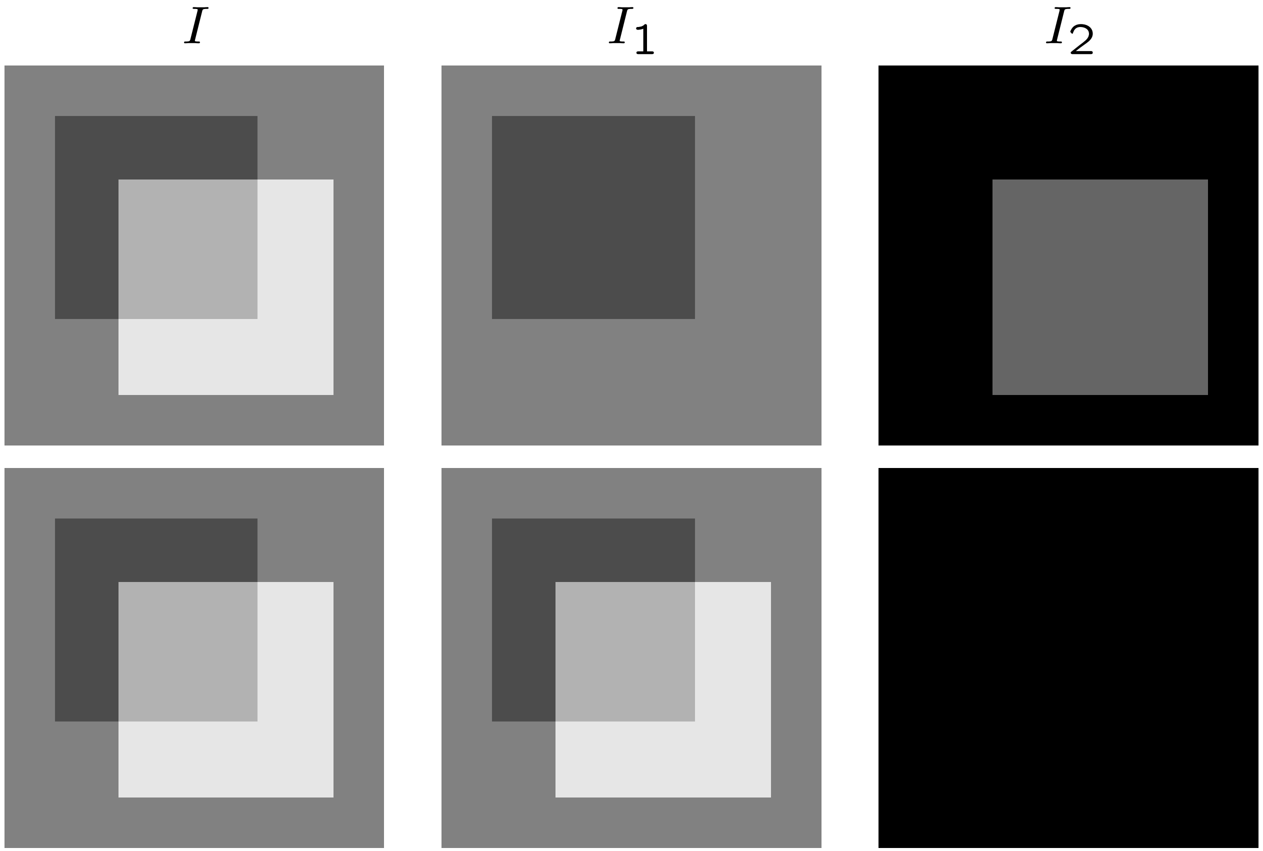


Generalized Gaussian distribution (Mallat 89, Simoncelli 95)

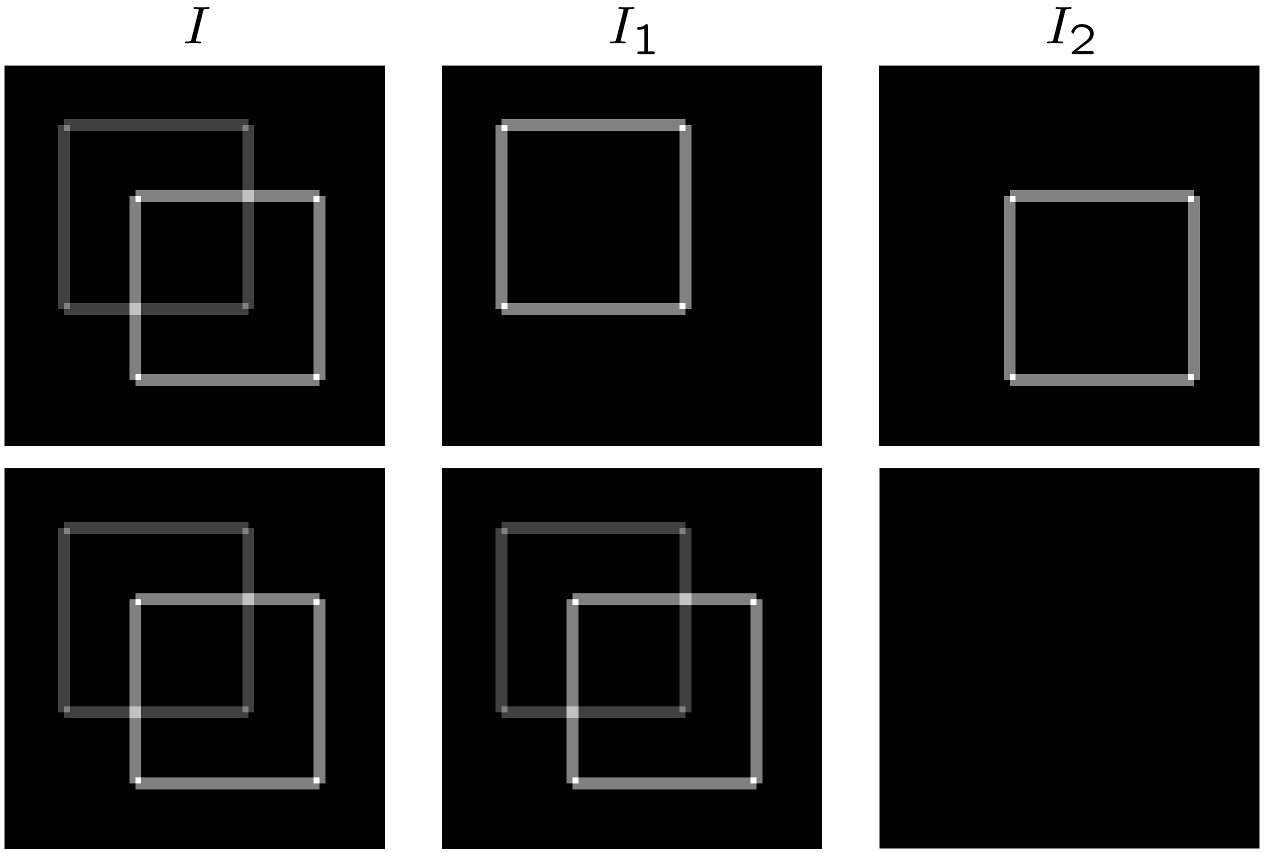
$$\mathrm{Pr}(x) \propto e^{-x^{\alpha}/s}$$

with $\alpha < 1$.

Is sparsity enough?



Is sparsity enough?

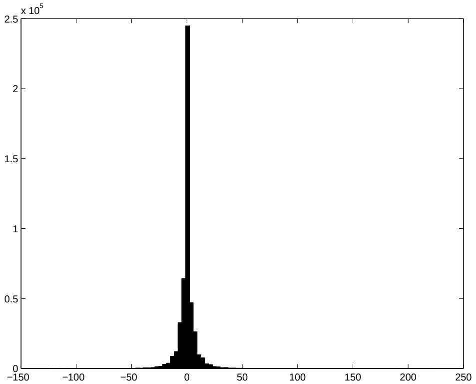


Cornerness operator

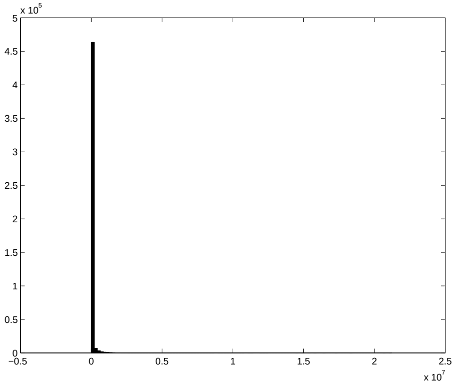
$$c(x_0, y_0) = det(\sum w(x,y) \begin{pmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{pmatrix})$$

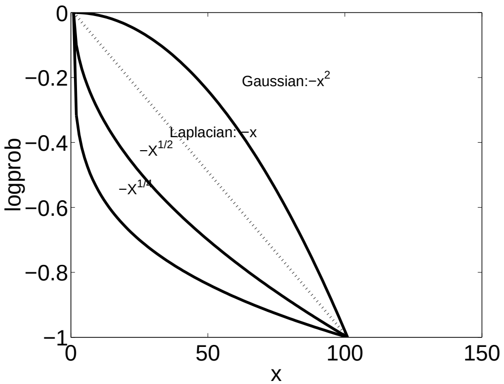
Corner Histograms

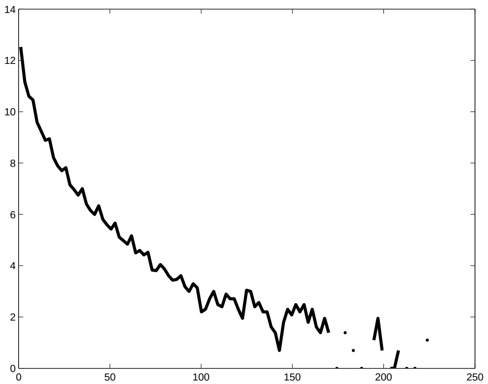
deriv filter

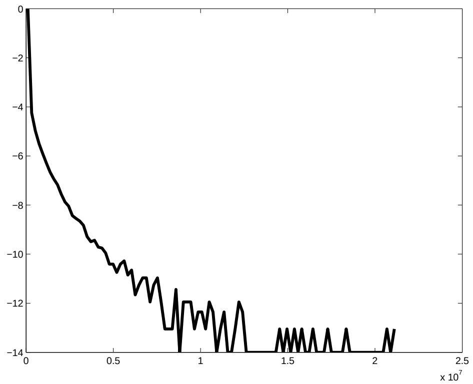


corner operator



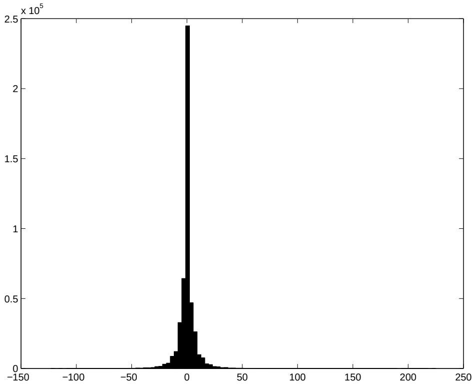


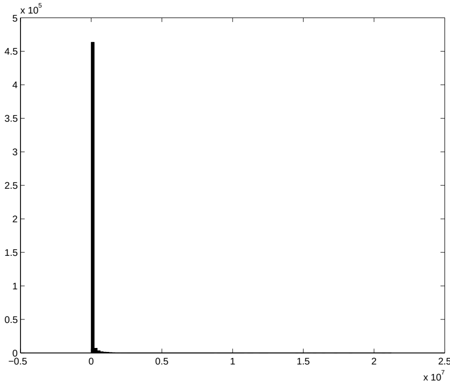


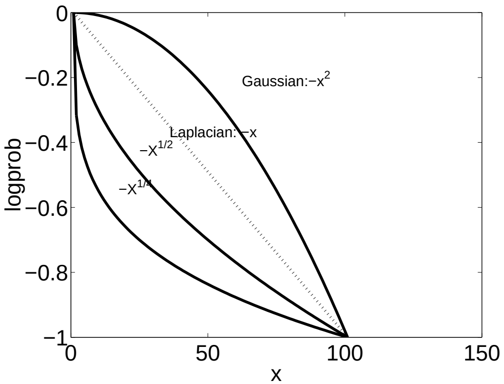


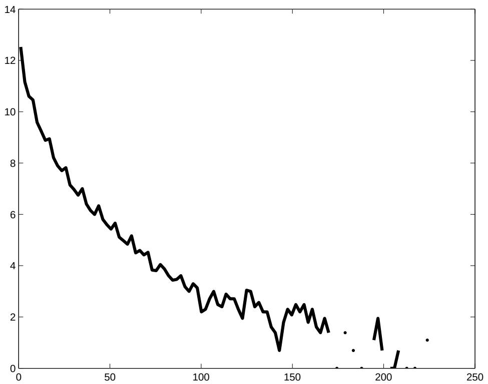
Fitting $P(x) \propto e^{-x^{\alpha}}$

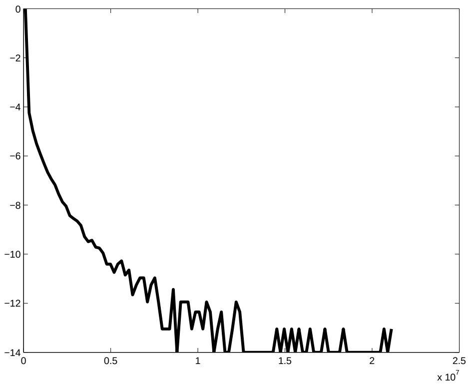
deriv filter | corner operator











typical exponents: $\alpha = 0.7$ $\beta = 0.2$

The model

$$I(x,y) = I_1(x,y) + I_2(x,y)$$

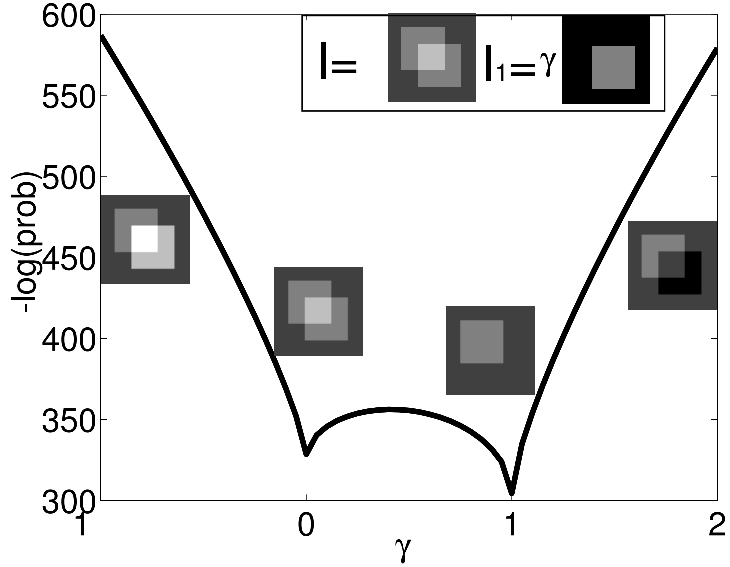
$\mathsf{Pr}(I_1), \mathsf{Pr}(I_2)$ defined via gradients:

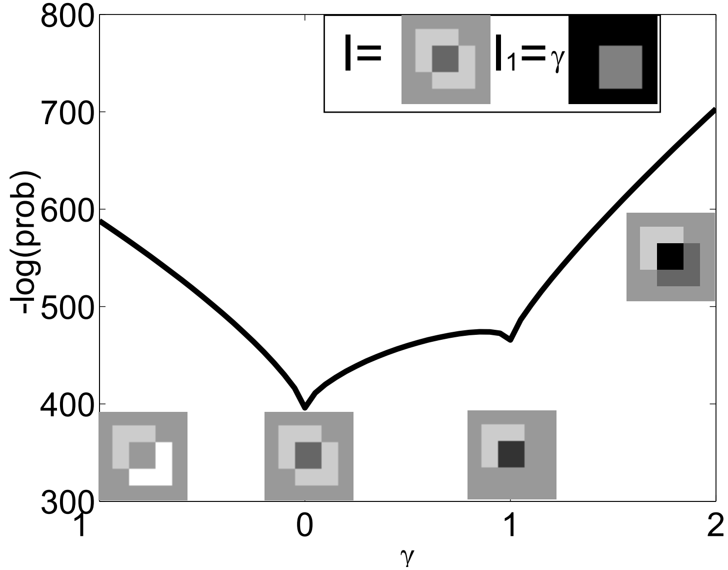
$$\begin{aligned}\log \mathsf{Pr}(I_1) &= \log \mathsf{Pr}(I_2) \\ &= -\log Z - \sum_{x,y} |\nabla I(x,y)|^{\alpha} + \eta c(x,y)^{\beta}\end{aligned}$$

with $\alpha = 0.7, \beta = 0.25, \eta > 1$.

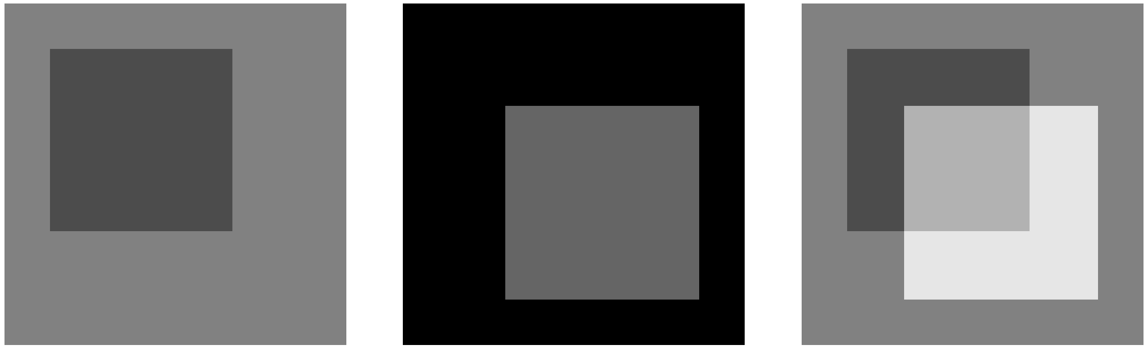
Does this predict transparency?

$$\log P = -\log Z - \sum_{x,y} |\nabla I(x,y)|^{\alpha} + \eta c(x,y)^{\beta}$$





Intuition:



- Adding typical images will only increase number of corners.

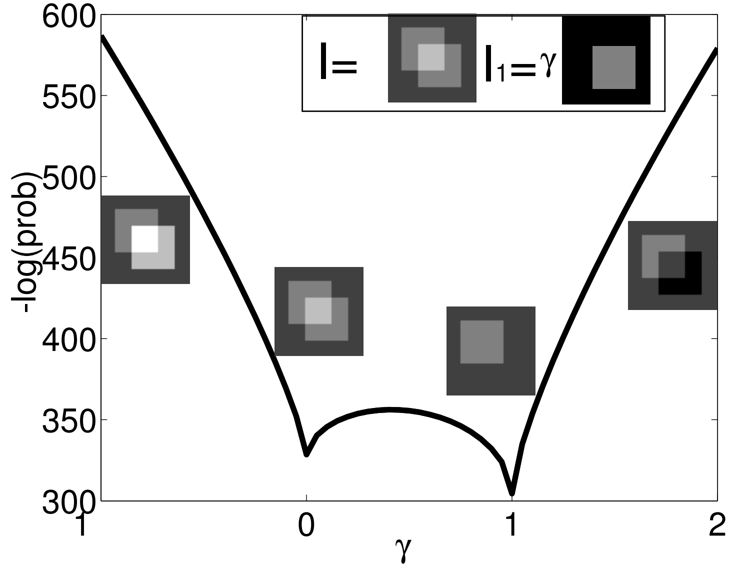
- Not true for white noise

How important are the statistics ?

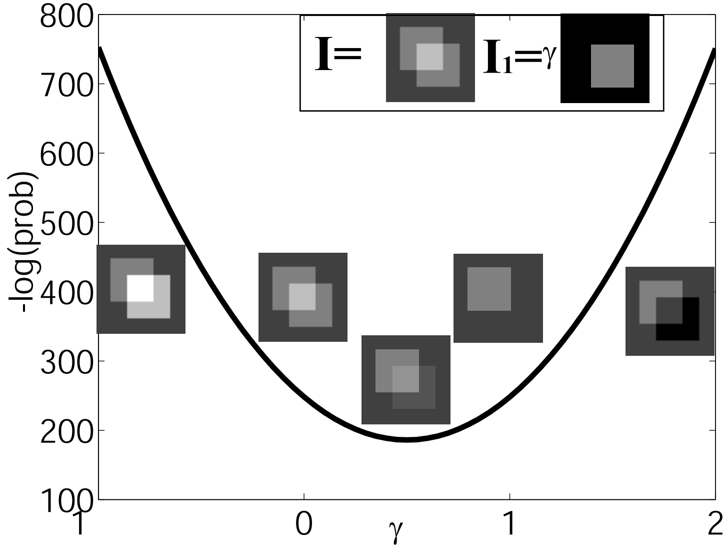
Is it important that the statistics are non Gaussian? Would any cost that penalized high gradients and corners work?

The importance of being non Gaussian

$$\log P = -\log Z - \sum_{x,y} |\nabla I(x,y)|^{\alpha} + \eta c(x,y)^{\beta}$$



$\alpha = 0.7, \beta = 0.25$



$\alpha = 2, \beta = 2$

The "scalar transparency" problem"

$$a + b = 1$$

with $a > 0$,$b > 0$.

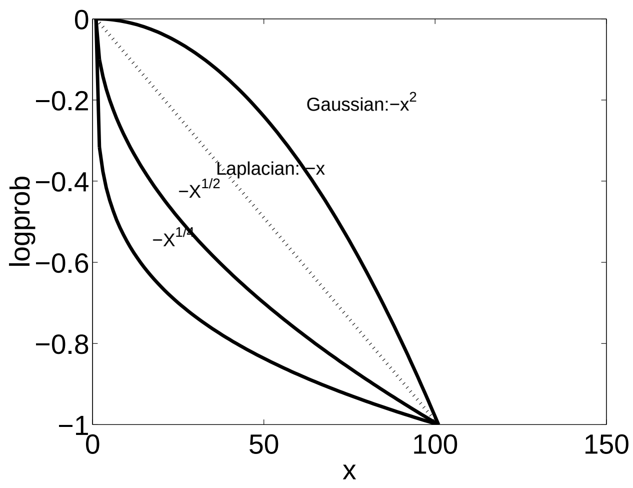
Consider a prior over positive scalars $\Pr(x) \propto e^{-x^\alpha}$, what is the MAP solution ?

The “scalar transparency” problem”

$$a + b = 1$$

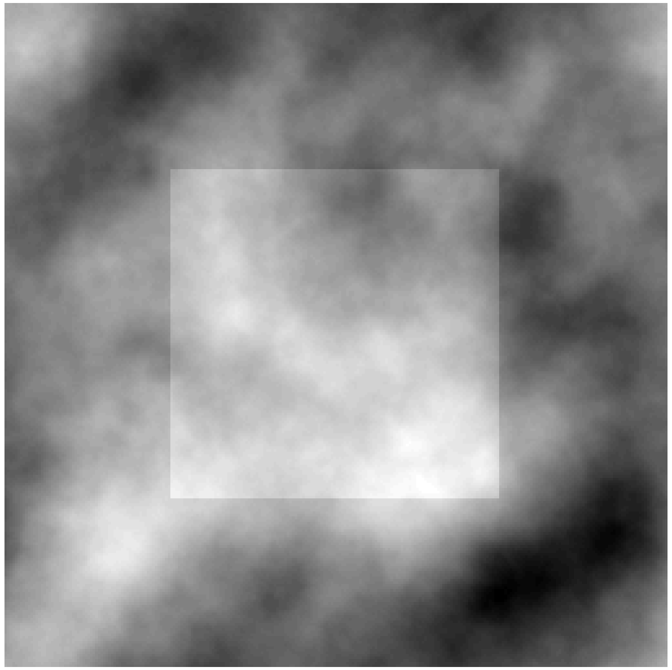
with $a > 0$,$b > 0$.

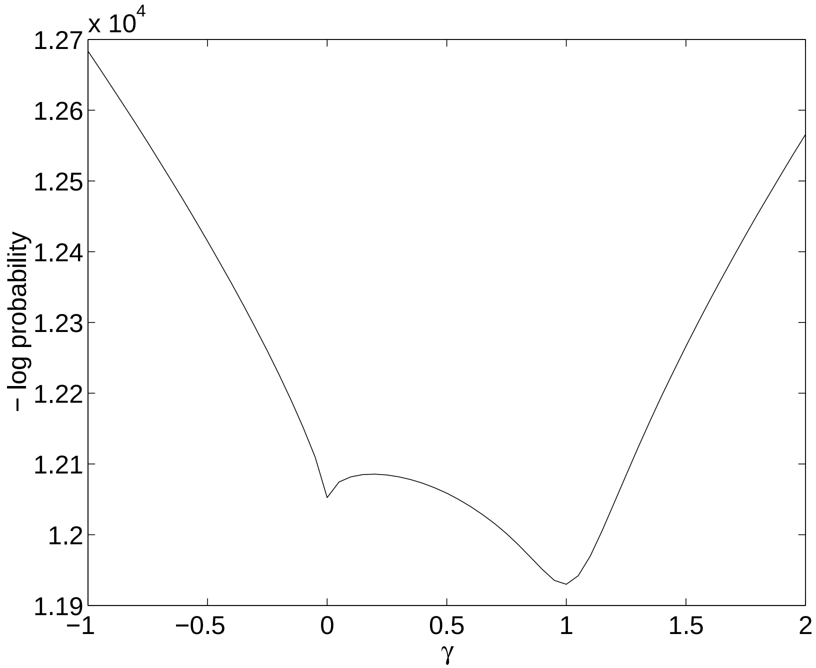
Observation: The solution is obtained with $a = 1, b = 0$ or $a = 0$,$b = 1$ if and only if $\log P(x)$ is below diagonal.



Does this predict transparency?

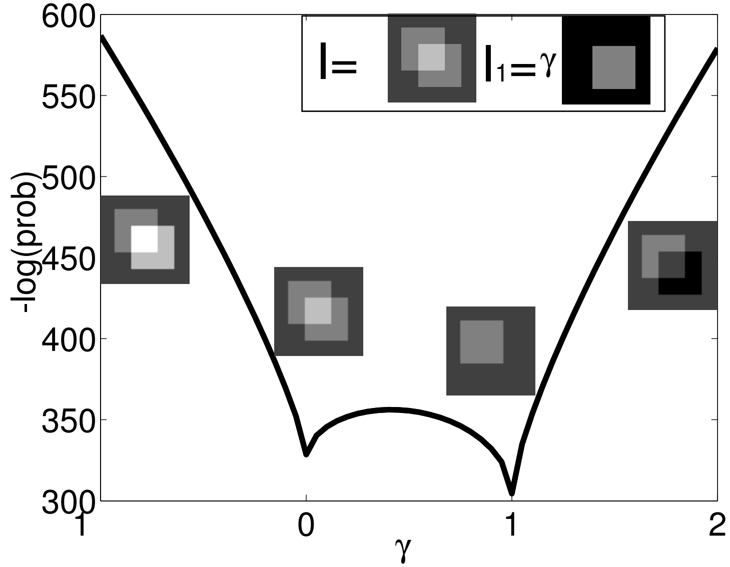
$$\log P = -\log Z - \sum_{x,y} |\nabla I|^{\alpha} + \eta c(x,y)^{\beta}$$

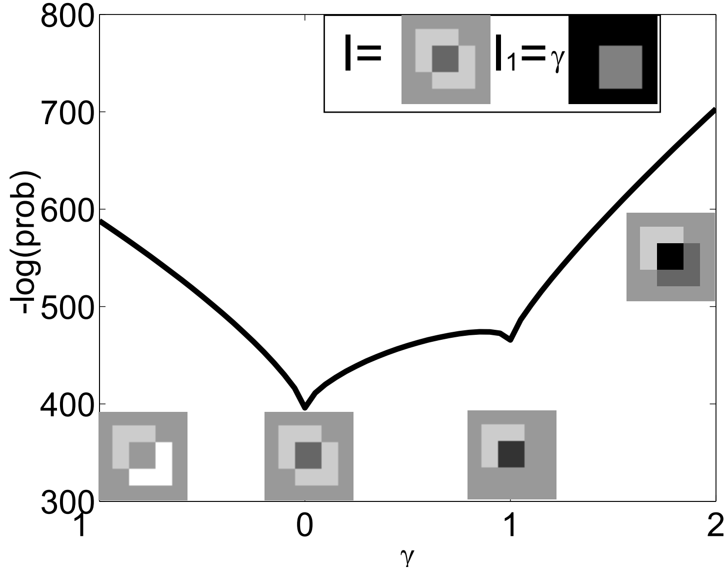




Can we perform global optimization?

$$\log P = -\log Z - \sum_{x,y} |\nabla I|^{\alpha} + \eta c(x,y)^{\beta}$$





Conversion to discrete MRF

$g_i = (g_{ix}, g_{iy})$: discretization of I_1 gradient at location i. f_i is discretization of I_2 gradient $f_i = (I_x, I_y) - g_i$.

$$
\begin{aligned}
\Pr(g) &= \frac{1}{Z} \prod_i \Psi_i(g_i) \prod_{<ijkl>} \Psi_{ijkl}(g_i, g_j, g_k, g_l) \\
\Psi_i(g_i) &= e^{-|g_i|^\alpha - |f_i|^\alpha} \\
\Psi_{ij}(g_i, g_j, g_k, g_l) &= e^{-det(g_i g_i^T + g_j g_j^T + g_k g_k^T + g_l g_l^T)^\beta}
\end{aligned}
$$

$$
e^{-det(f_i f_i^T + f_j f_j^T + f_k f_k^T + f_l f_l^T)^\beta}
$$

$\Psi_{ijkl}(g_i, g_j, g_k, g_l)$ also enforce integrability

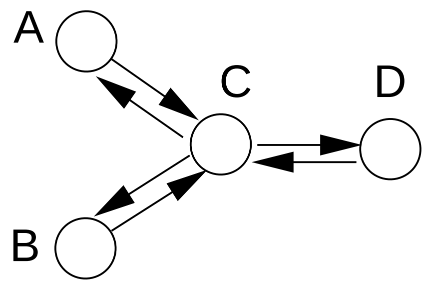
Optimizing discrete MRF

$g_i = (g_{ix}, g_{iy})$: discretization of I_1 gradient at location i. f_i is discretization of I_2 gradient $f_i = (I_x, I_y) - g_i$.

$$\Pr(g) = \frac{1}{Z} \prod_i \Psi_i(g_i) \prod_{<ijkl>} \Psi_{ijkl}(g_i, g_j, g_k, g_l)$$

$|g|^N$ possible assignments. Solution: use max-product belief propagation.

Belief Propagation:



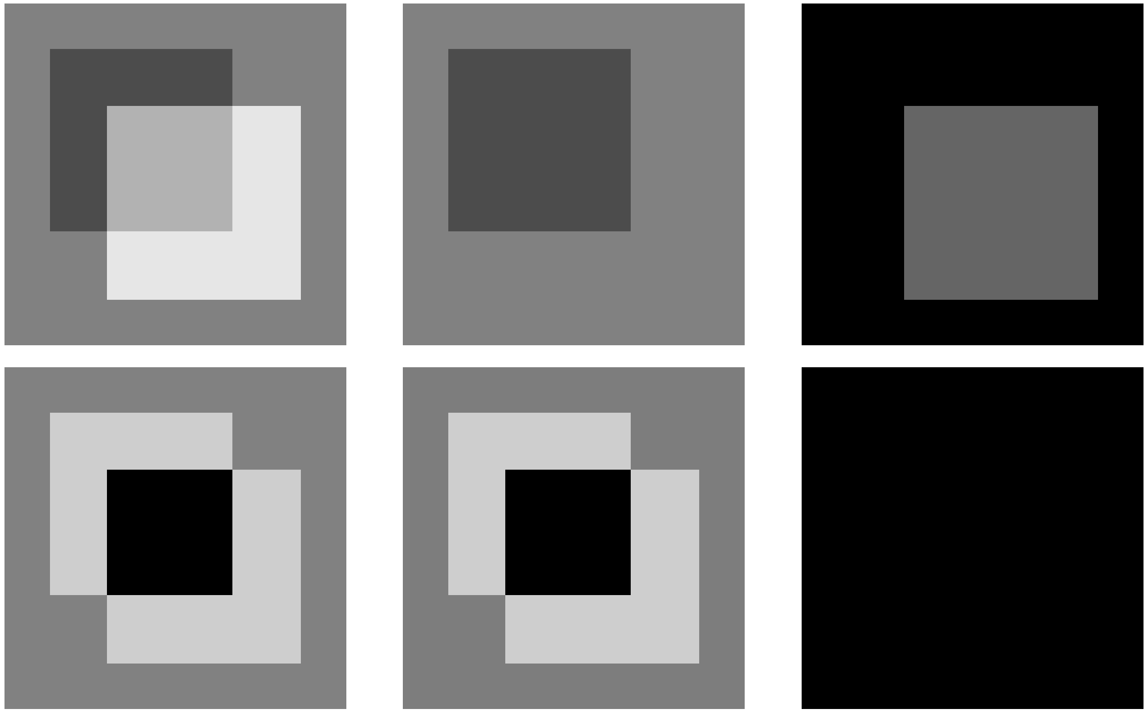
- A parallel message-passing algorithm.

- Every node sends a probability density to its neighbors.

Results

$$\log P = -\log Z - \sum_{x,y} |\nabla I|^{\alpha} + \eta c(x,y)^{\beta}$$

Input I Output I_1 Output I_2



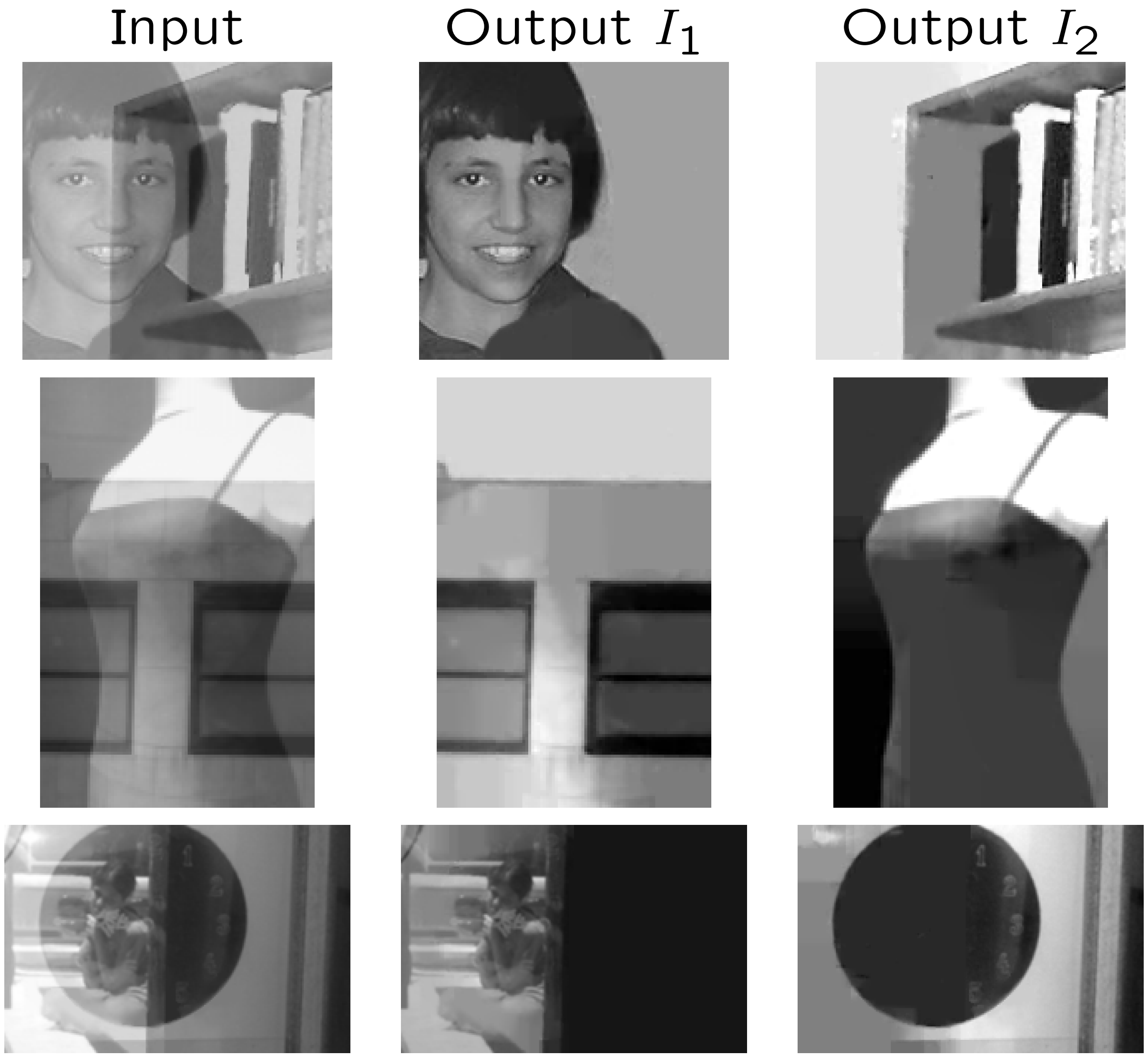
Caveats

Significant manual tweaking needed to get BP to converge on "real" images:

- BP convergence depends on update order.

- discretization artifacts

BP Patches Results



Input
Output I_1
Output I_2
1
2
3
4
5

EM results

Future work

- $\log P = -\log Z - \sum_{x,y} |\nabla I|^{\alpha} + \eta c(x,y)^{\beta}$ Learn features and weights automatically. (Zhu and Mumford 97, Della Pietra Della Pietra and Lafferty 96).

- Extend to more realistic imaging situations.

- Use similar prior to model amodal completion.

Conclusions

- natural scene statistics predict perception of transparency

- Algorithm can decompose a *single image* into the sum of two images.