BROWN

Efficient Belief Propagation for MRFs in Low-Level Vision

Stefan Roth
Department of Computer Science
Brown University

Joint work with Xiangyang Lan, Dan Huttenlocher, and Michael Black

Image Inpainting

[Bertalmio et al., 2000]

old photograph

user-supplied mask

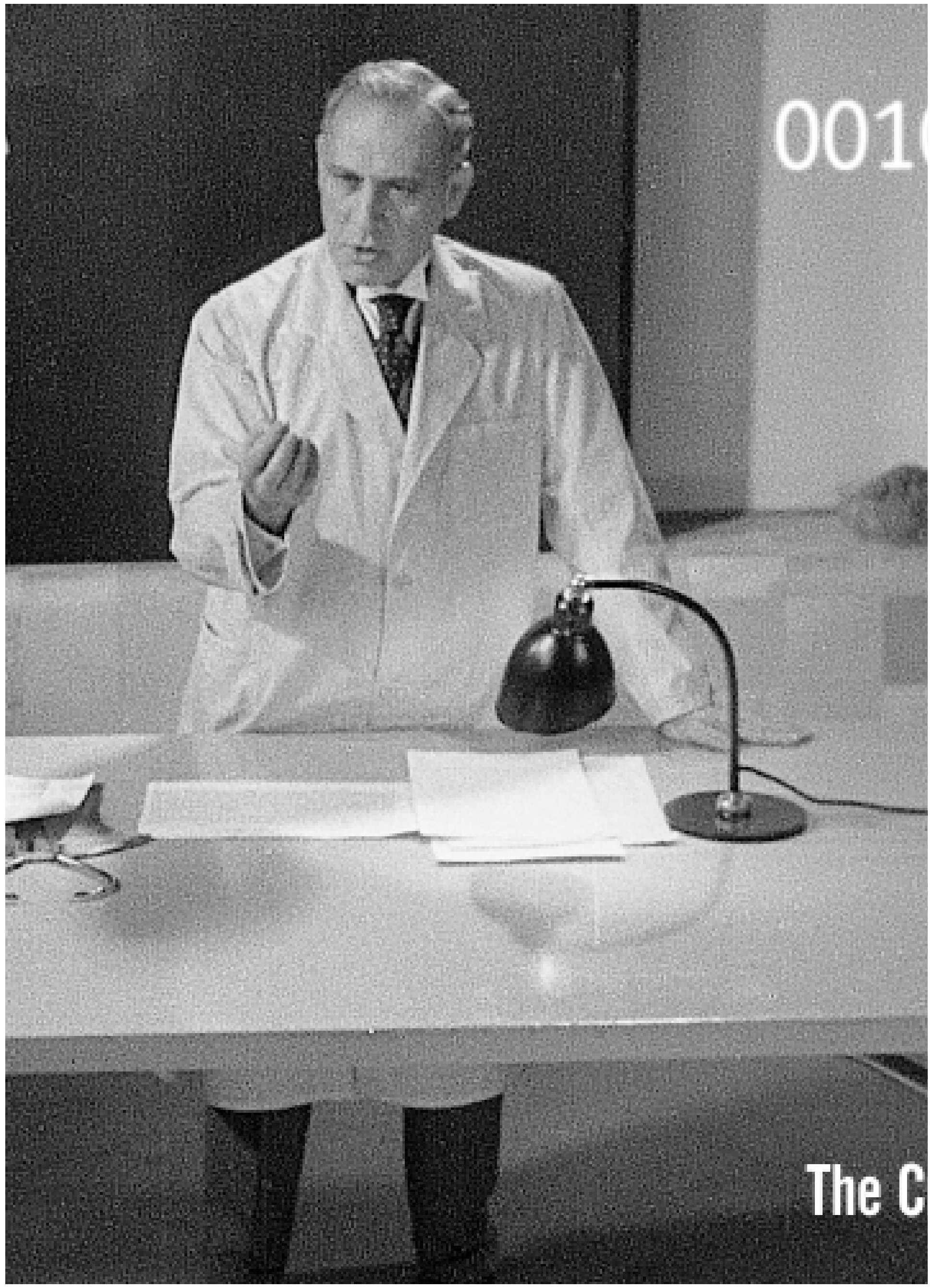
Image Inpainting

reconstructed photo

user-supplied mask

Image and Video Denoising

movie frame with "film grain"

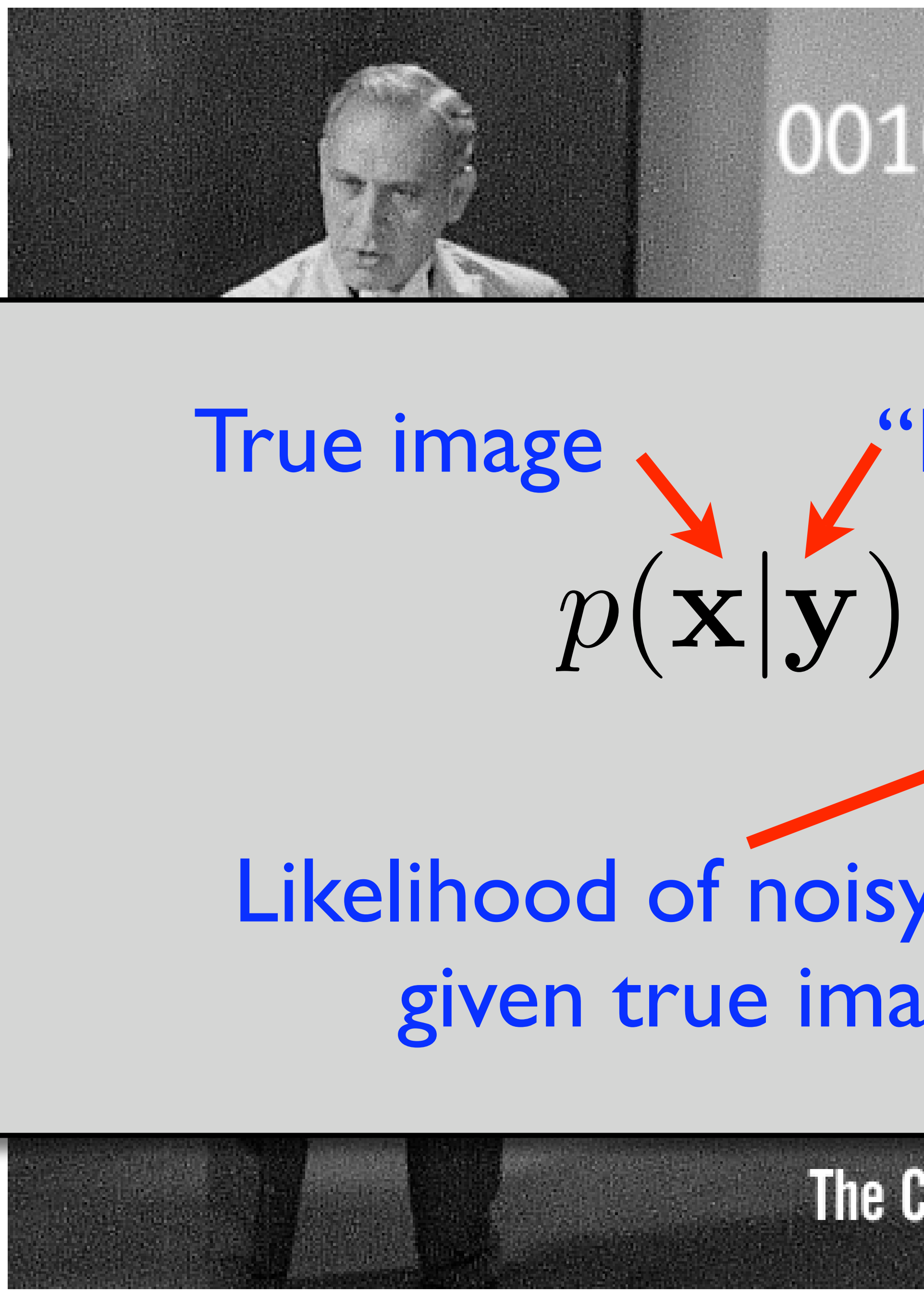


denoised frame

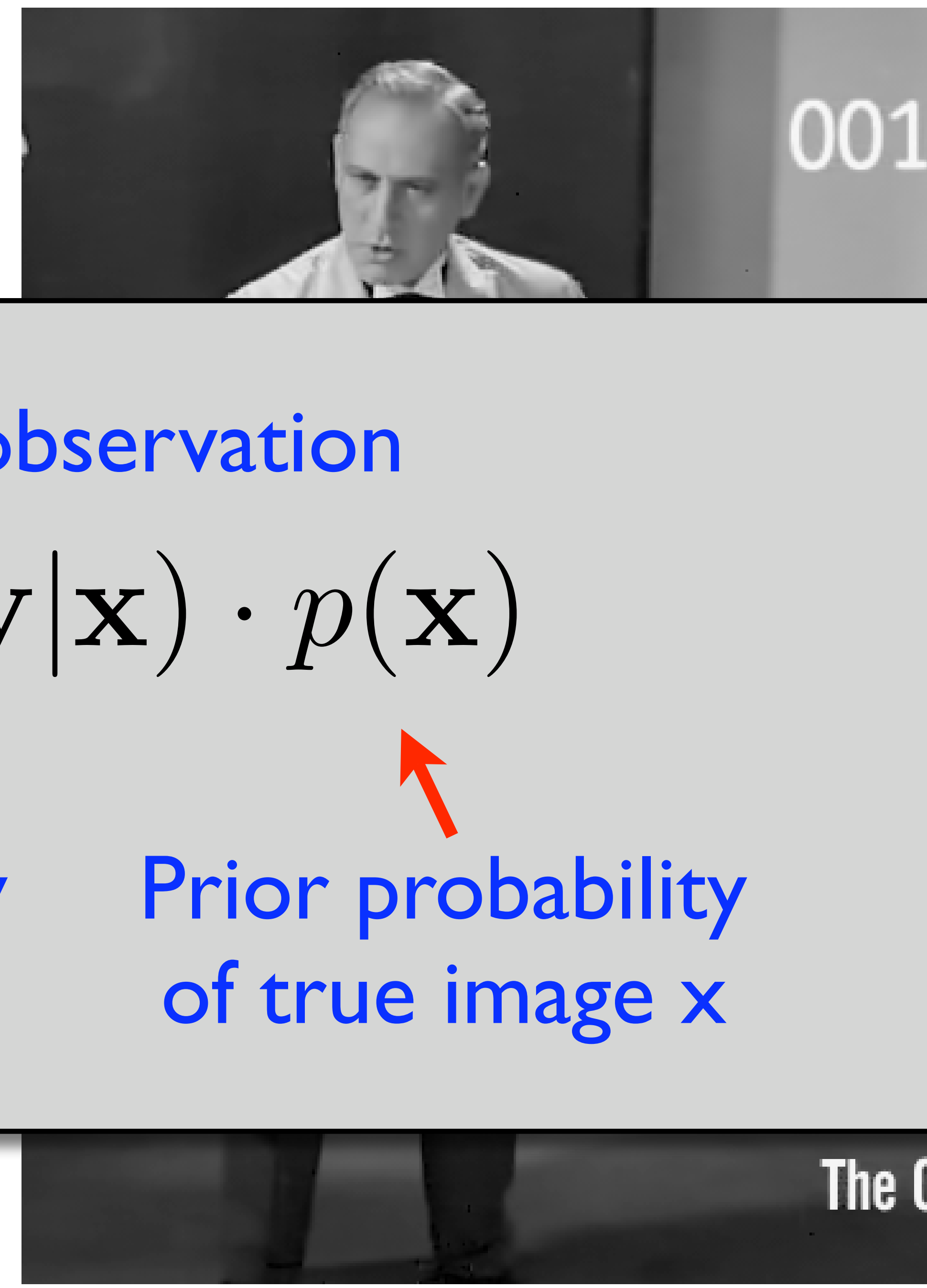
Thanks to Kevin Manbeck and Jay Cassidy (MTI)

Image and Video Denoising

movie frame with "film grain"



denoised frame



True image

"Noisy" observation

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x}) \cdot p(\mathbf{x})$$

Likelihood of noisy image y given true image x

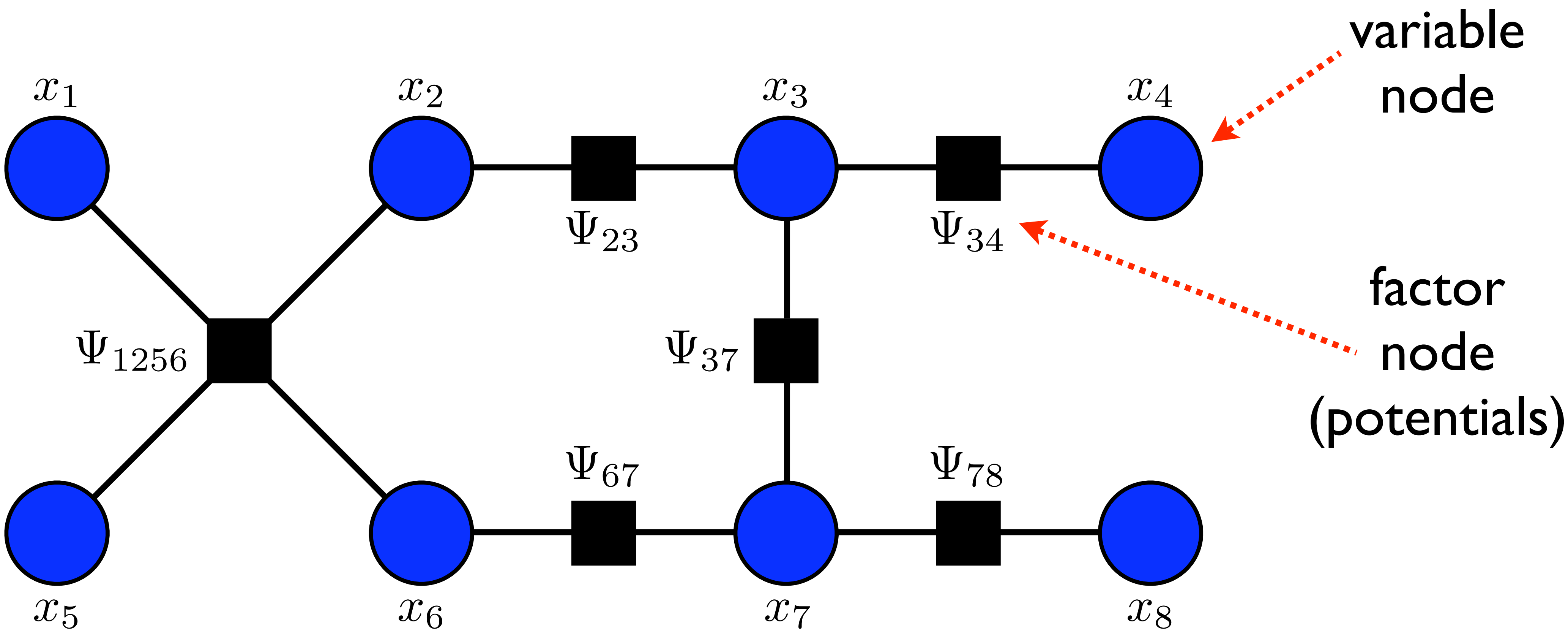
Prior probability of true image x

Thanks to Kevin Manbeck and Jay Cassidy (MTI)

Low-level Vision & MRFs

- Other applications of interest:
  - Stereo
  - Optical flow
  - Super-resolution
- Both likelihood and prior commonly formulated as Markov random field (MRF).
- Consider two model types here:
  - Classical pairwise MRFs
  - More expressive high-order MRFs.

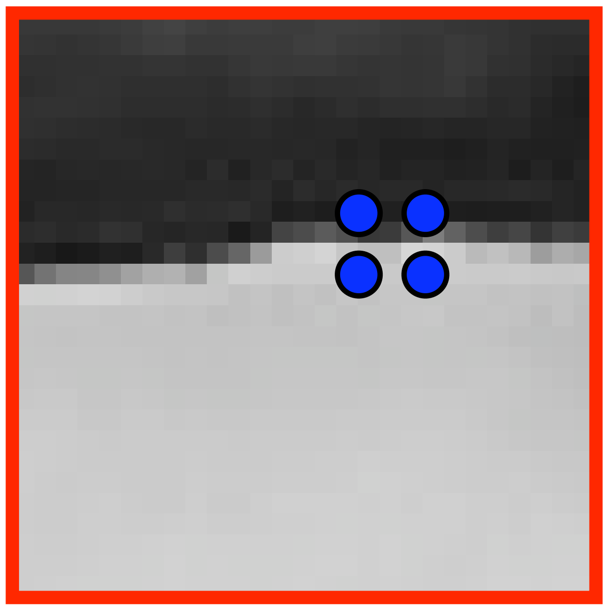
Review: Factor Graphs



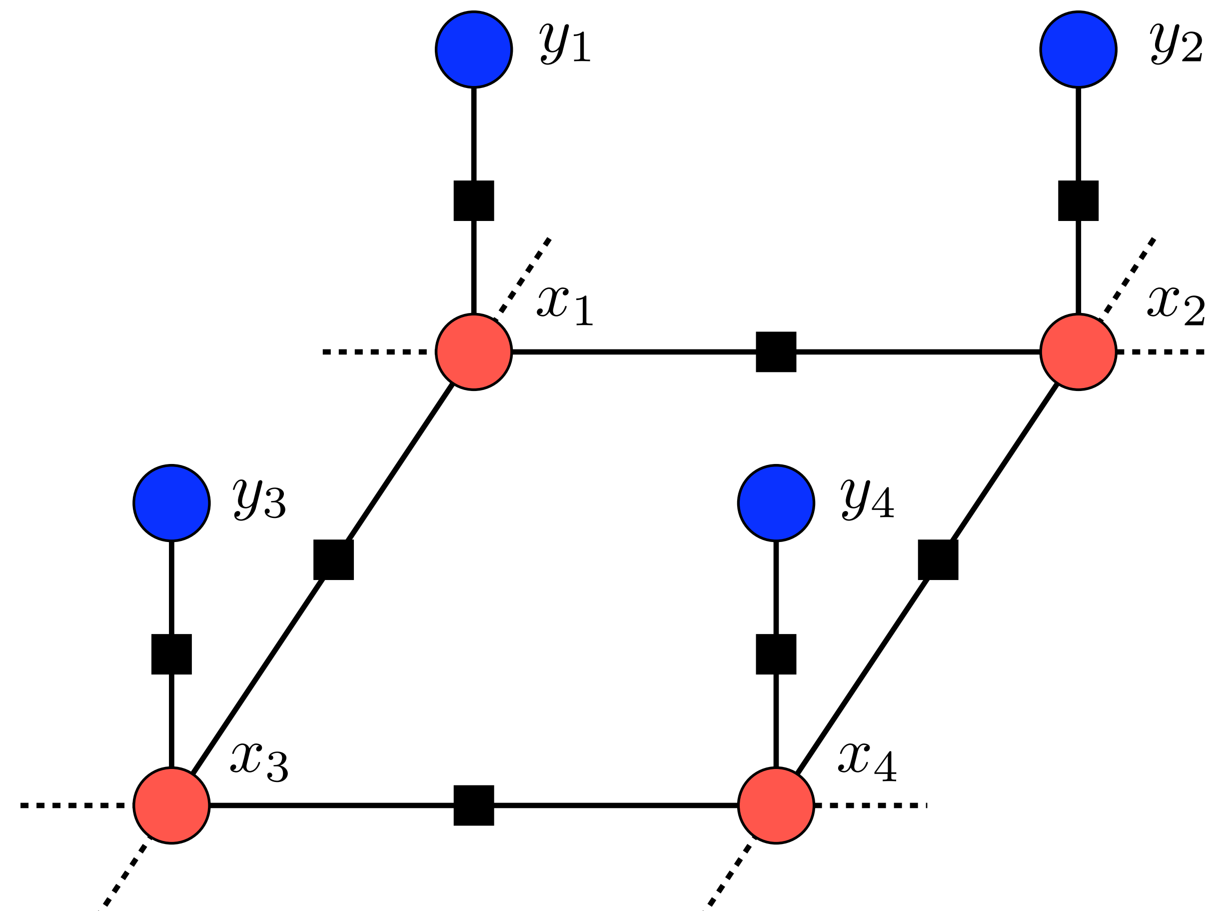
$$p(x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8) = \frac{1}{Z} \cdot \Psi_{1256}(x_1,x_2,x_5,x_6) \cdot \Psi_{23}(x_2,x_3) \cdot \Psi_{37}(x_3,x_7) \cdot \Psi_{67}(x_6,x_7) \cdot \Psi_{34}(x_3,x_4) \cdot \Psi_{78}(x_7,x_8)$$

Pairwise Markov Random Fields (MRFs)

e.g., [Geman & Geman, 1984]



input image



$$p(\mathbf{x}|\mathbf{y}) = \frac{1}{Z} \cdot \underbrace{\prod_i \Psi_L(x_i, y_i)}_{\text{likelihood}} \cdot \underbrace{\prod_{\substack{\text{neighbors} \\ x_i, x_j}} \Psi_P(x_i, x_j)}_{\text{prior}}$$

recovered image

Pairwise Potentials

$$p(\mathbf{x}|\mathbf{y}) = \frac{1}{Z} \cdot \underbrace{\prod_i \Psi_L(x_i, y_i)}_{\text{likelihood}} \cdot \underbrace{\prod_{\substack{\text{neighbors} \\ x_i, x_j}} \Psi_P(x_i, x_j)}_{\text{prior}}$$

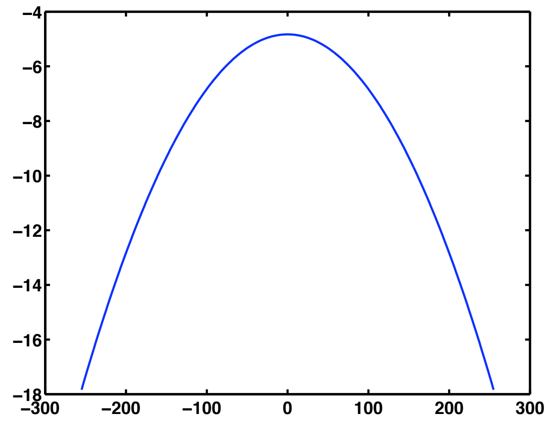
- What are the potential functions in low-level vision applications?
- Likelihood:
  - Application specific
  - Often a simple Gaussian, e.g.:

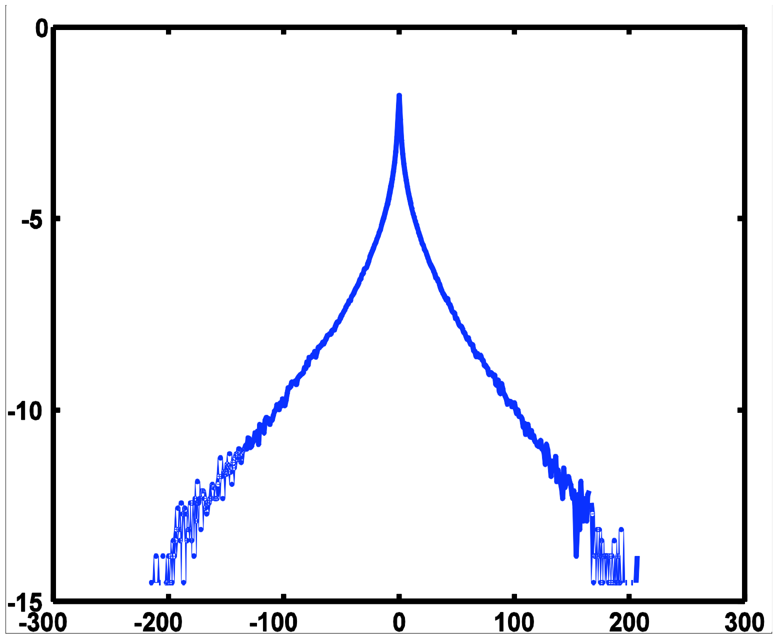
$$\Psi_L(x_i, y_i) \propto e^{-\frac{1}{2\sigma_L^2}(x_i - y_i)^2}$$

Pairwise Potentials (II)

- Prior:
  - Gaussian potentials:

$$\Psi_P(x_i, x_j) = e^{-\frac{1}{2\sigma^2}(x_i - x_j)^2}$$



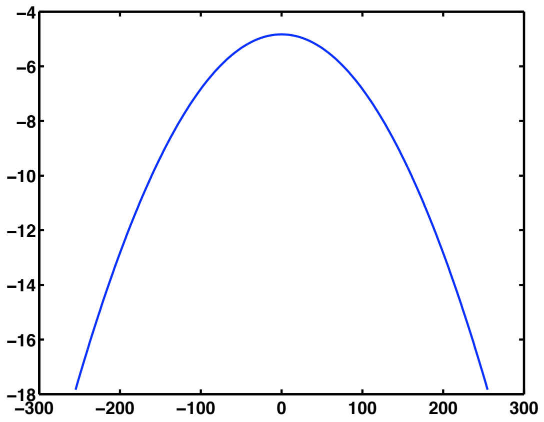


Log-histogram of the image gradient
[Ruderman, 1997], [Huang & Mumford, 1999]

Pairwise Potentials (II)

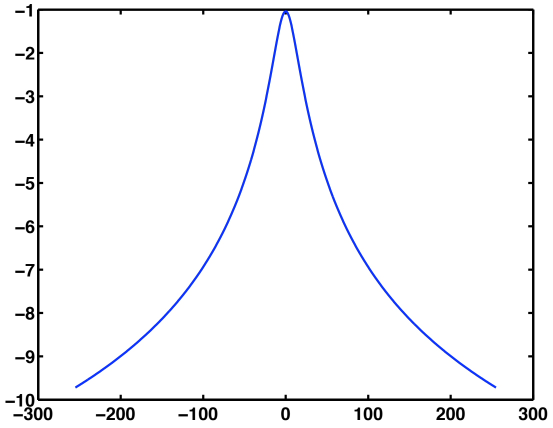
- Prior:
  - Gaussian potentials:

$$\Psi_P(x_i, x_j) = e^{-\frac{1}{2\sigma^2}(x_i - x_j)^2}$$



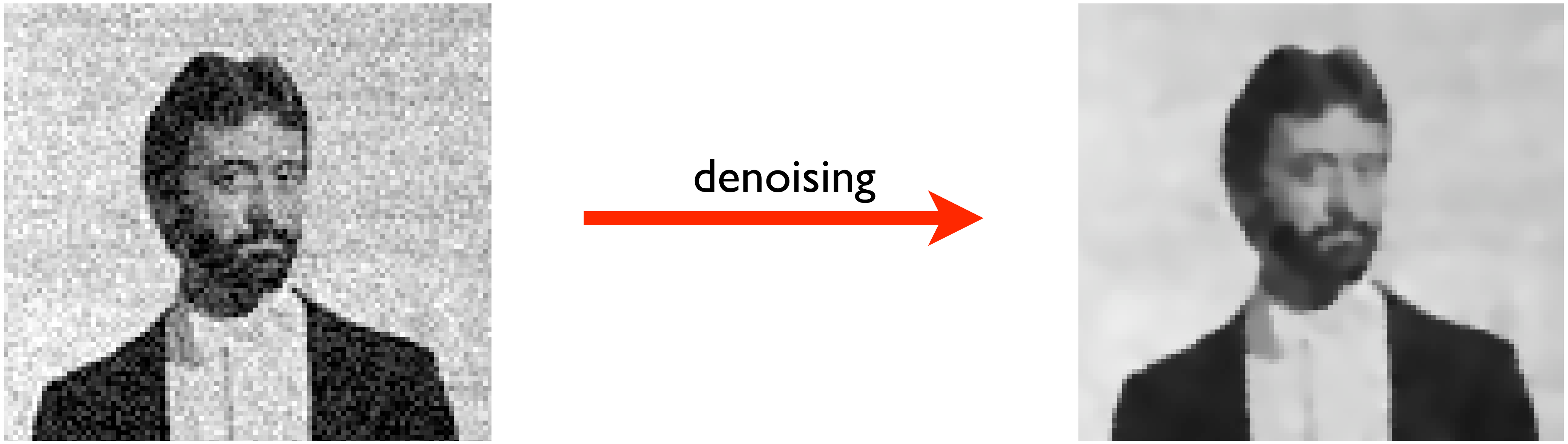
  - "Robust" potentials (e.g., t-distribution):

$$\Psi_P(x_i, x_j) = \left(1 + \frac{1}{2\sigma^2}(x_i - x_j)^2\right)^{-\alpha}$$



  - non-convex energy
- Not covered here: Parameter estimation

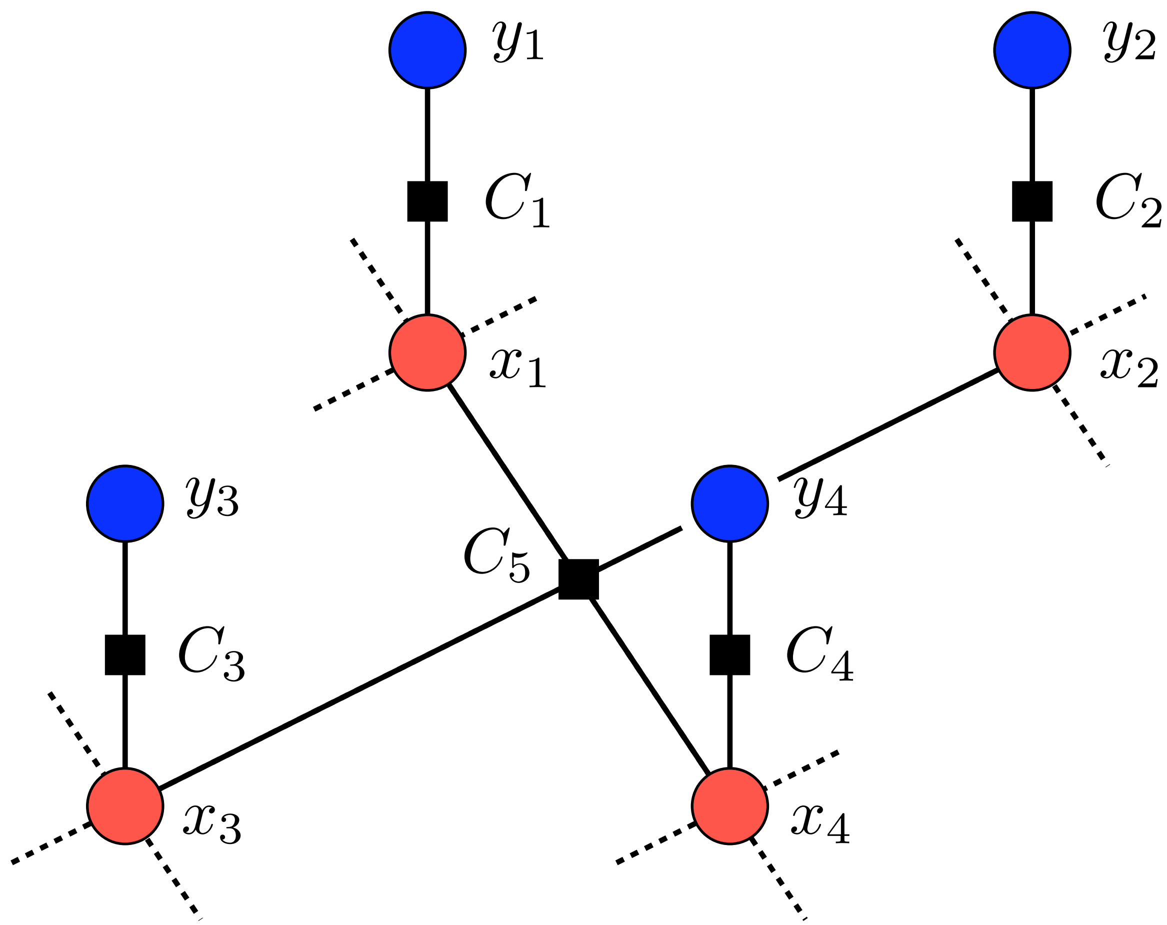
Beyond pairwise MRFs



- Pairwise MRFs do not capture the rich spatial structure of natural images:
  - Interactions are too local.
- How do we resolve that?
  - Turn to richer, high-order models for the prior.
  - E.g. Fields of Experts [Roth & Black, 2005].

High-order MRF Models

- Likelihood model stays the same.
- Simplest case: Prior has 2x2 factors (cliques).
  - larger factors possible (e.g. 3x3, or 5x5)



$$\mathbf{x}_{C_1} = (x_1, y_1)$$
$$\vdots$$
$$\mathbf{x}_{C_4} = (x_4, y_4)$$
$$\mathbf{x}_{C_5} = (x_1, x_2, x_3, x_4)$$

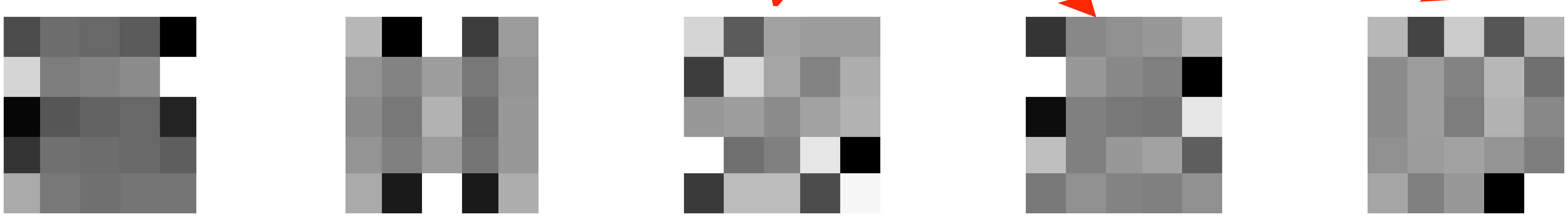
Fields of Experts

[Roth & Black, 2005]

- Model high-order factor using Product of Experts [Hinton, 1999].
- Formalization:

Expert distribution

$$\Psi_P(\mathbf{x}_C) = \prod_{k=1}^{K} \phi(\mathbf{J}_k^{\mathrm{T}} \mathbf{x}_C; \alpha_k)$$



Example filters

Review: Probabilistic Inference

- Our goals are:
  - to compute marginals of the posterior,
  - or to compute an assignment that maximizes the posterior (MAP).
- Loopy belief propagation is very popular for approximate inference [Weiss, 1997]:
  - Sum-product BP for (approximately) computing marginals.
  - Max-product BP for (approximately) computing MAP assignments.

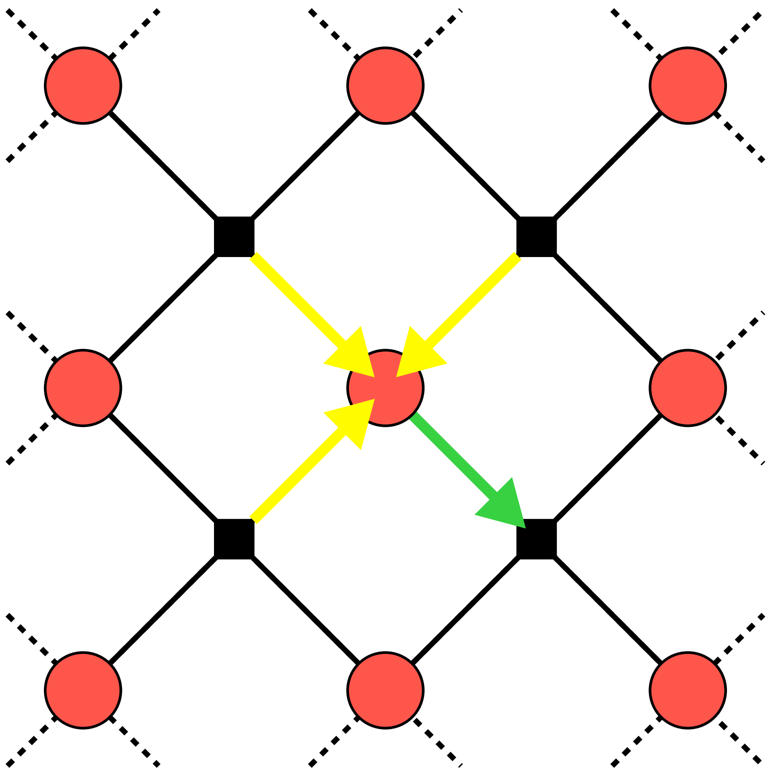
Loopy BP on Factor Graphs

- Equivalent to standard loopy BP on pairwise graphs, but more general.
- Two types of messages:
  - From variable node i to factor node C: $n_{i \to C}(x_i)$
  - From factor node C to variable node i: $m_{C \to i}(x_i)$
- Belief for variable node i:

$$b(x_i) \propto \prod_{C \in \mathcal{N}(i)} m_{C \to i}(x_i)$$

Variable Node to Factor Node

$$n_{i \to C}(x_i) \propto \prod_{C' \in \mathcal{N}(i) \setminus C} m_{C' \to i}(x_i)$$



- Very easy to compute for discrete variables.
- Applies to sum-product and max-product.

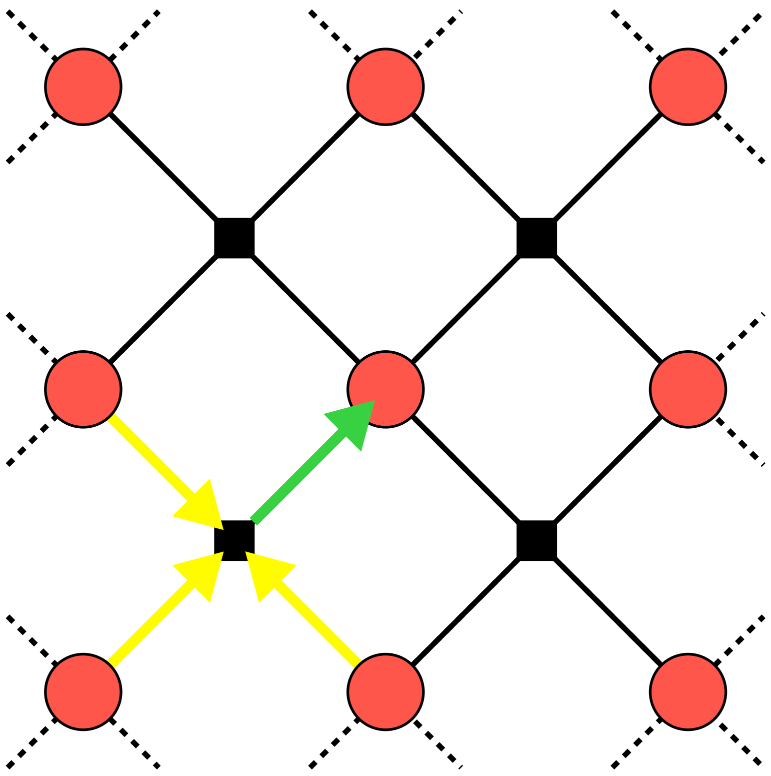
Factor Node to Variable Node

$$m_{C \to i}(x_i) \propto \sum_{\mathbf{x}_C \setminus x_i} \Psi_C(\mathbf{x}_C) \prod_{i' \in \mathcal{N}(C) \setminus i} n_{i' \to C}(x_i')$$

sum-product BP

$$m_{C \to i}(x_i) \propto \max_{\mathbf{x}_C \setminus x_i} \Psi_C(\mathbf{x}_C) \prod_{i' \in \mathcal{N}(C) \setminus i} n_{i' \to C}(x_i')$$

max-product BP



- Often expensive to compute: Have to sum or max over a potentially huge space.

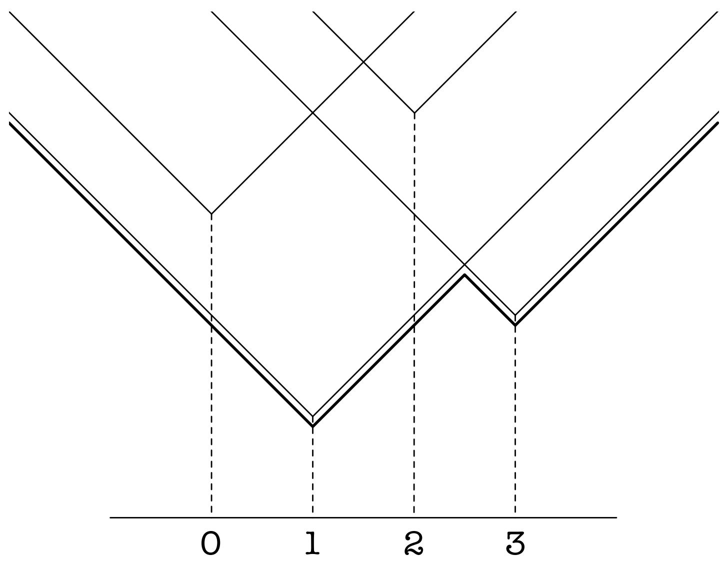
Computational Burden

- Per message cost (N - number of discrete bins, often as many as 256)
  - Pairwise model: $\mathcal{O}(N^2)$
  - mxm factors: $\mathcal{O}(N^{m^2})$
- What can we do to make this tractable?
  - Pairwise model: Apply distance transform [Felzenszwalb & Huttenlocher, 2004].
  - 2x2 factors: Restrict the number of bins.

Distance Transform

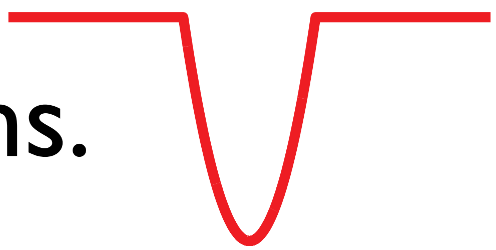
[Felzenszwalb & Huttenlocher, 2004]

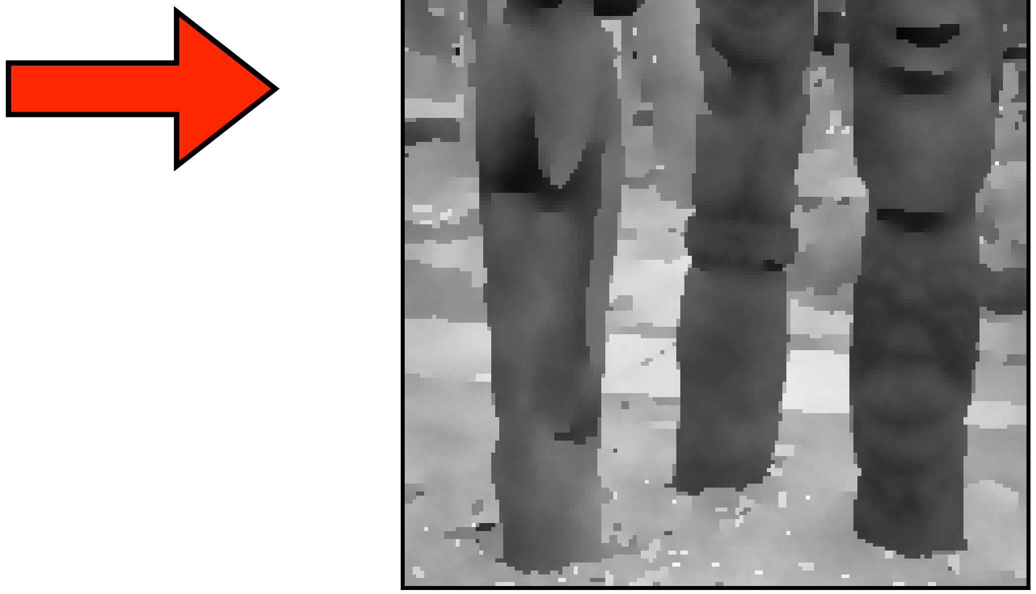
- Max-product (actually min-sum) with pairwise models.
- Speed up message computation using distance transform techniques:
  - Convex, symmetric potentials $\Psi(x_1, x_2) = \Psi'(|x_1 - x_2|)$
  - Can compute the lower envelope of potentials in linear time.
  - Allows us to compute message in $\mathcal{O}(N)$ instead of $\mathcal{O}(N^2)$.



Distance Transform (II)

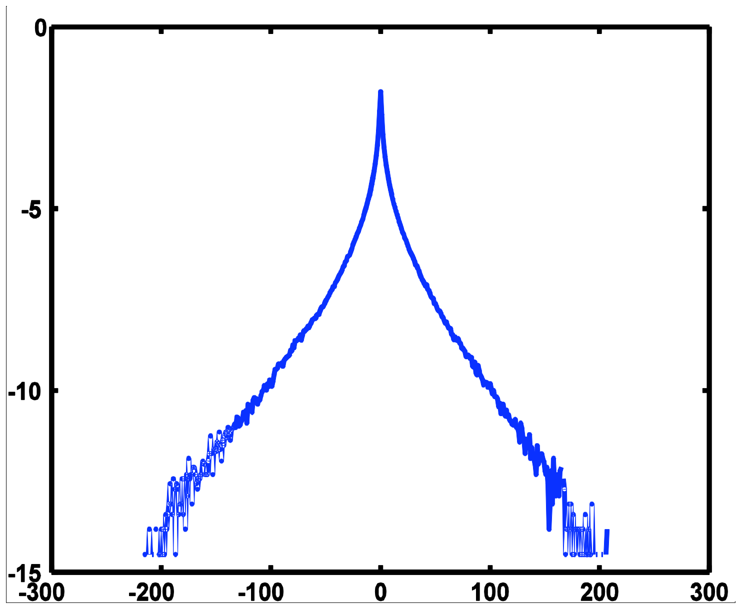
- Can be extended to combinations of convex potentials, e.g. truncated Gaussians.
- Very fast, but slightly disappointing results:

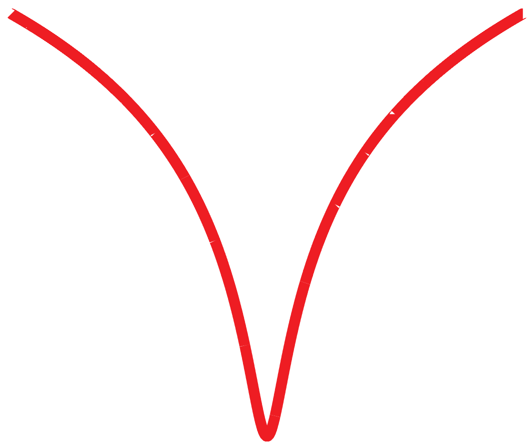




Non-convex Potentials

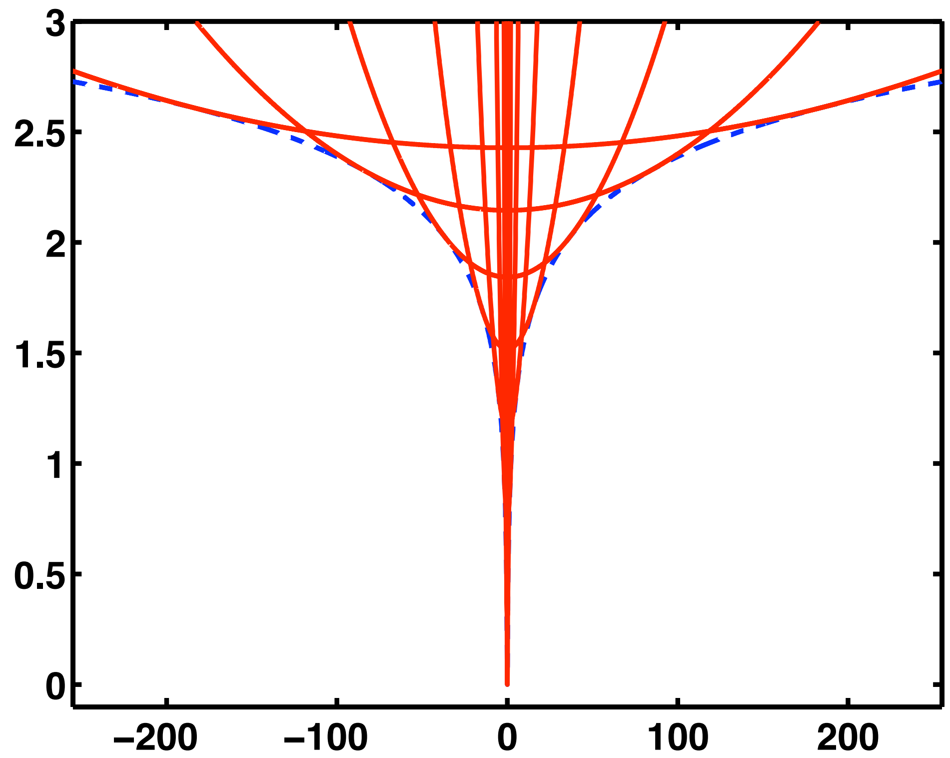
- What could be the problem?
  - Gaussian or truncated Gaussian potentials do not match the statistics of natural images well.
- We could use non-convex potentials, e.g. a t-distribution.
  - But: Distance transform doesn't apply to non-convex potentials!

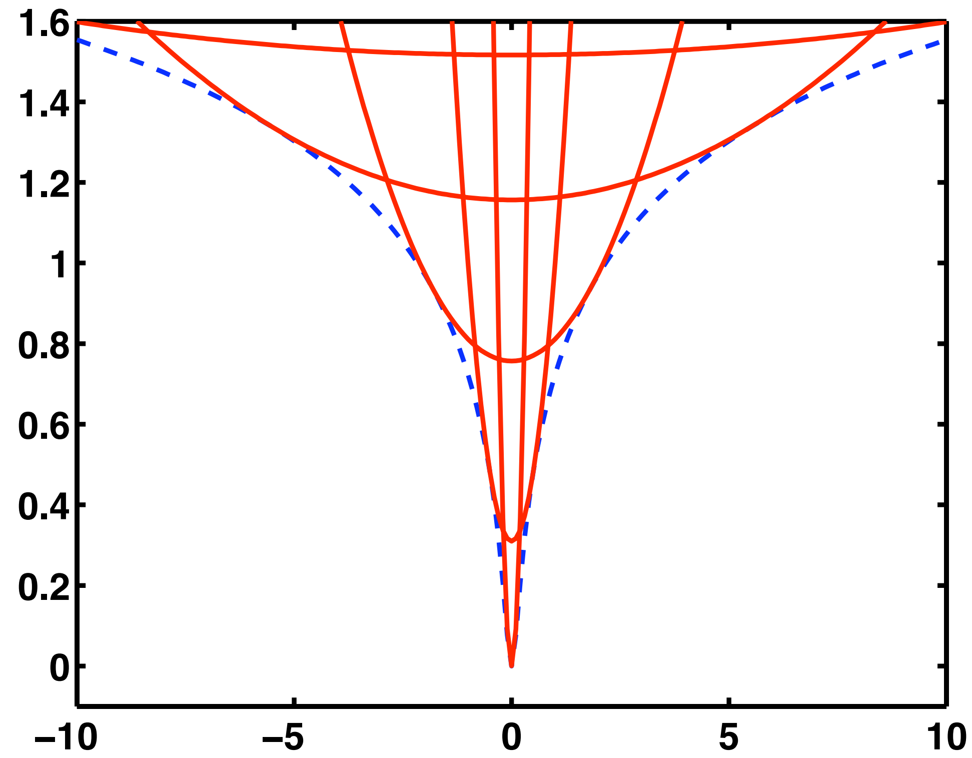




Key Idea

- Approximate non-convex potentials as the lower envelope of several convex potentials:





- Closed form expression for t-distributions:

$$-\log \Psi(x_i - x_j; \alpha, \sigma) = \inf_z \frac{(x_i - x_j)^2}{2\sigma^2} z + z - \alpha + \alpha \log \alpha z$$

Details

- Closed form for a number of "robust" potentials [Black & Rangarajan, 1996].
- Fit a given number of quadratics to potential by minimizing KL-divergence.
- Computational burden of message computation (q - number of quadratics):

$$\mathcal{O}(q \cdot N)$$

BROWN

Denoising Results

Noisy image

Denoised with t-distrib. potential

Approximate potentials (8 quadratics)

BROWN

High-order Models

- Decent performance with non-convex pairwise potentials.
- But: High-order potentials promise to be more powerful.
- Can we do unmodified BP on the factor graph even for 2x2 factors?
  - No, each message requires 2^{32} computations.

Key Idea

- For most pixels, we don't really need to represent the entire $[0..255]$ range.
- Limit computations to smaller range $[a..b]$
  - Determine a and b individually per pixel.
  - Denoising: Use neighborhood of pixels + noise scale.
  - Other applications: First approximate with pairwise model.
- Optional: Discretize $[a..b]$ coarsely.

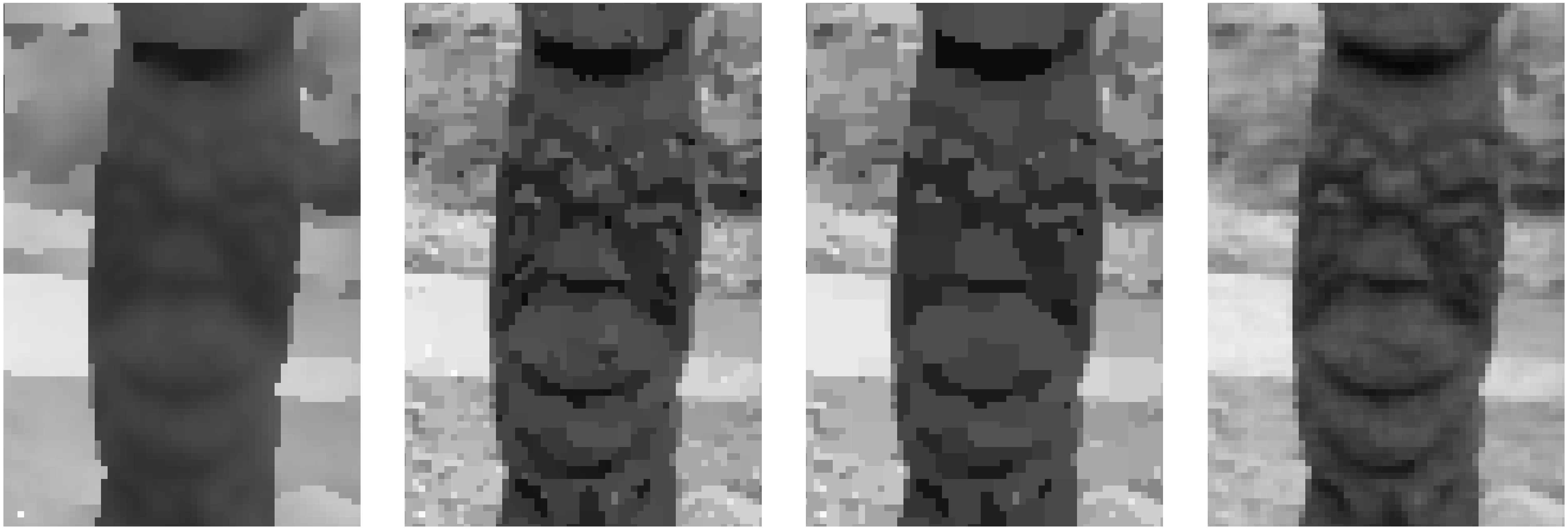
Results

Noisy image

Denoised with t-distrib. potential

Denoised with 2x2 FoE

Comparison



truncated Gaussian | Student-t | Student-t approximation | 2x2 FoE

Evaluation on 10 different images: Significant PSNR improvements (FoE over Student-t over truncated Gaussian)

Running Time

- Pairwise graph (256x256 image):
  - Standard algorithm: ~3 min / iteration
  - Distance transform with truncated Gaussian: ~5 sec / iteration
  - Distance transform with approximated non-convex potential: ~30 sec / iteration
- High-order graph (256x256 image):
  - Restricted value range: ~16 min / iteration

BROWN

Summary

- MRFs are a popular model for image processing, optical flow estimation, stereo etc.
- Loopy belief propagation for approximate inference has enjoyed enormous popularity.
- LBP has a large computational complexity, especially for high-order models.
    - Not always practical.

BROWN

Summary (II)

- Pairwise models:
  - Distance transform speed-up for convex potentials.
  - Approximate non-convex potentials as lower envelope of several convex ones.
- High-order models:
  - Standard algorithm impractical.
  - Restrict the value range individually for every pixel.

References

- X. Lan, S. Roth, D. Huttenlocher, and M. J. Black: *Efficient Belief Propagation with Learned Higher-Order Markov Random Fields*. ECCV 2006.
- Recent related work:
  - C. Pal, C. Sutton, and A. McCallum: *Sparse Forward-Backward using Minimum Divergence Beams for Fast Training of Conditional Random Fields*. ICASSP 2006.