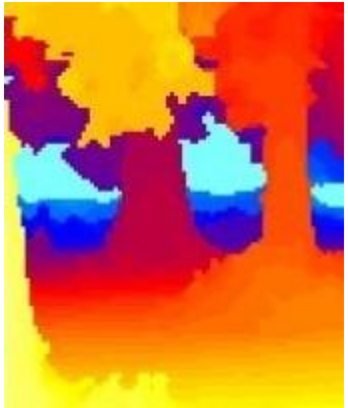
Learning Depth from Single Monocular Images

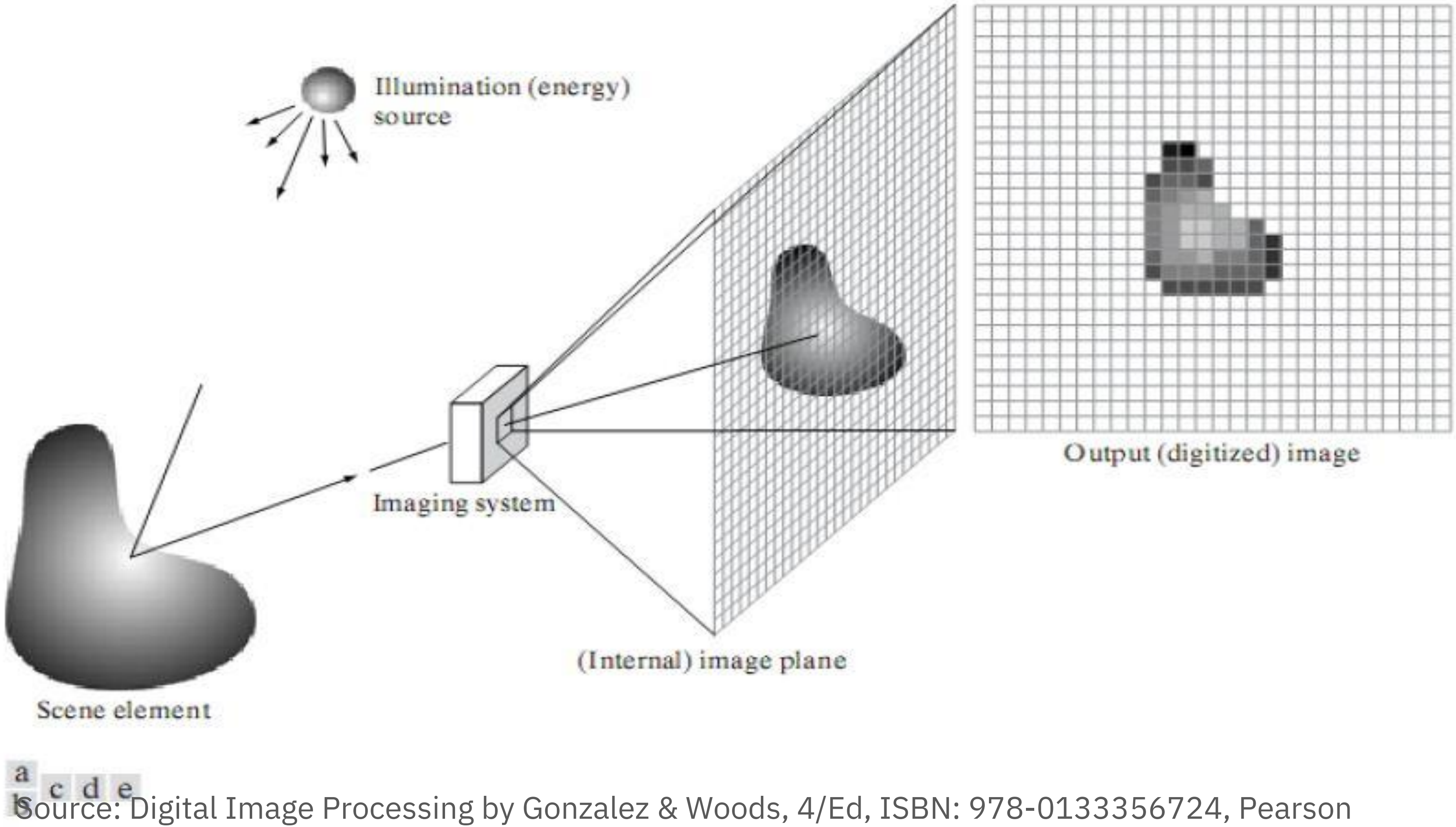
Saurabh Mathur

Ashutosh Saxena, Sung H. Chung, Andrew Y. Ng. In Neural Information Processing Systems 18, 2005.

Depth: What is the z-value at each pixel ?

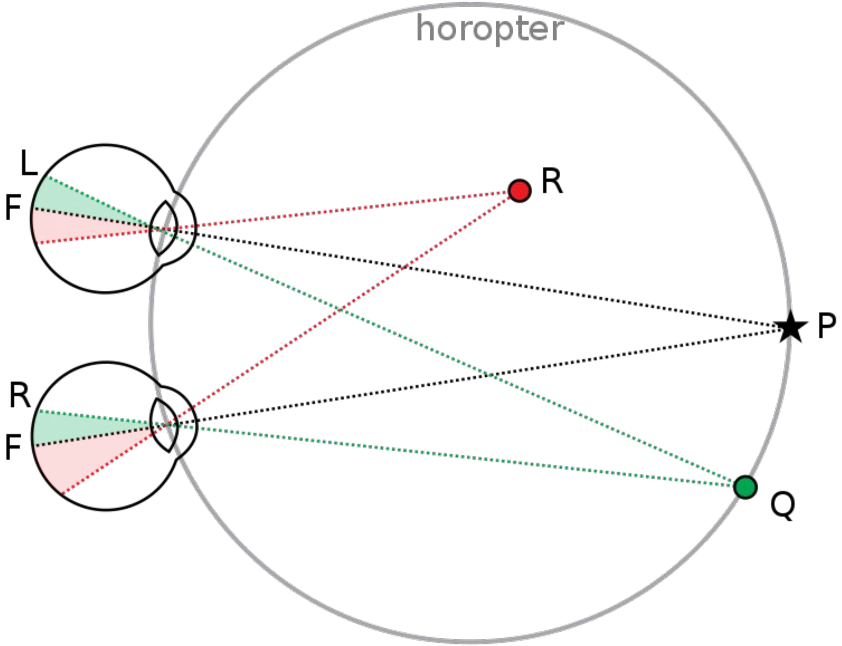


Source: http://www.cs.cornell.edu/~asaxena/learningdepth/depthmaps.jpg



Source: Digital Image Processing by Gonzalez & Woods, 4/Ed, ISBN: 978-0133356724, Pearson

Binocular/Stereo vision



Source: https://en.wikipedia.org/wiki/Binocular_vision

Why estimate depth from monocular images ?

- Small embedded systems can't carry a lot of depth sensors.
- Sensors have range limitations.
- Monocular images are easier to acquire.

What does depth estimation require ?

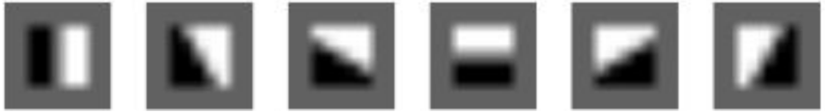
- Prior knowledge about the scene
- Local features: Texture variation, gradients, occlusion
- Global context

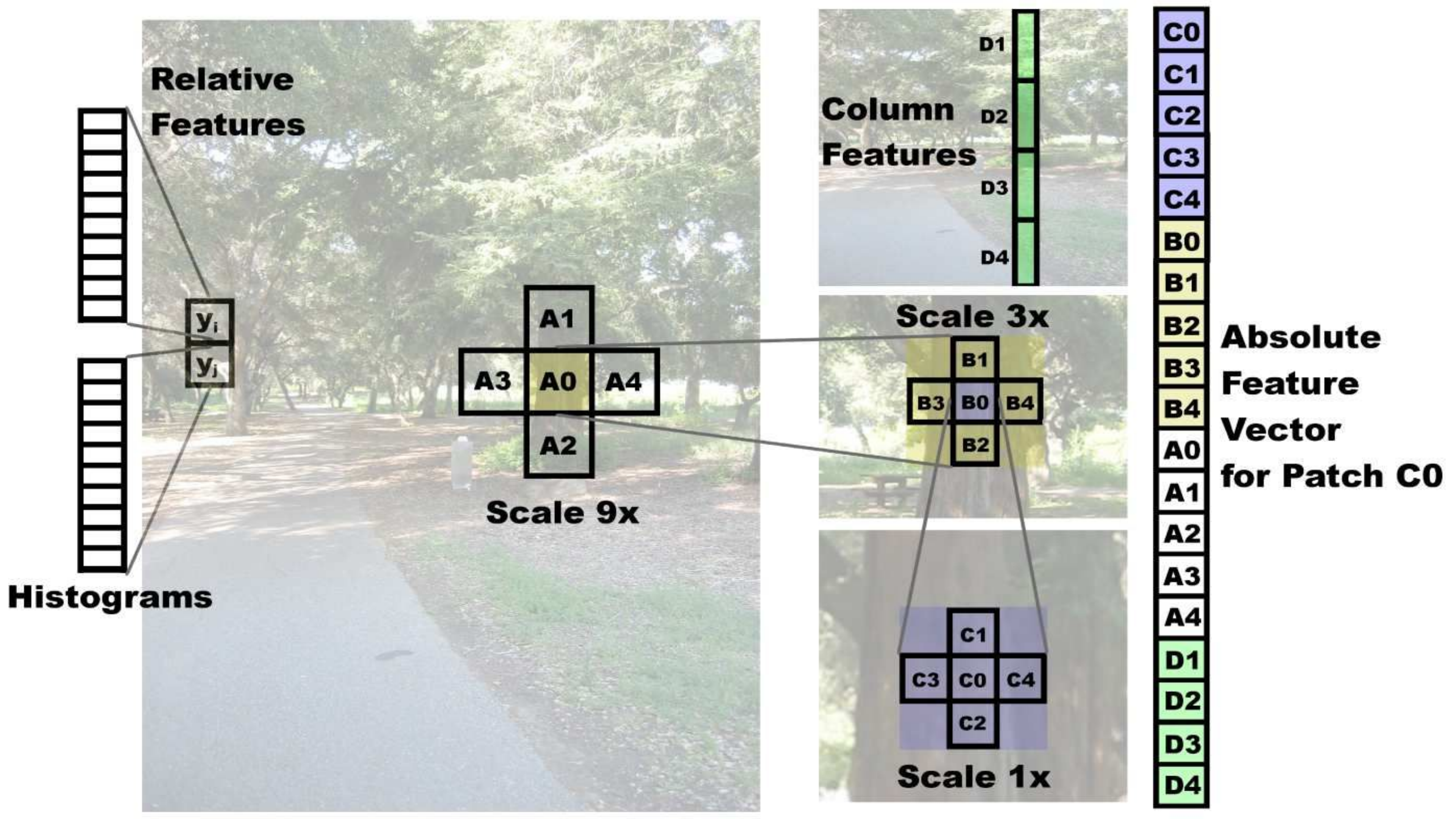
Nine Law's texture masks

Law's texture masks: $v_i^T v_j$

- v_1 = [1 4 6 4 1], Averaging (Gaussian)
- v_2 = [-1 -2 0 2 1], Edge (Gradient)
- v_3 = [-1 0 2 0 -1], Spot (Laplacian)

Six Edge detectors spaced at 30°

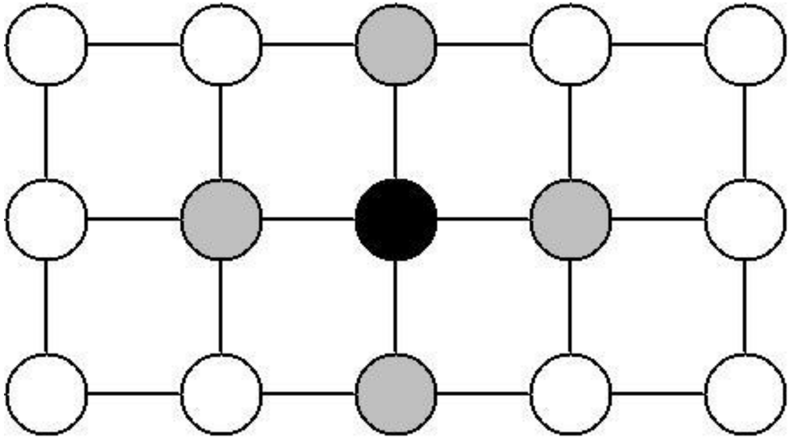



Relative Features
y_i
y_j
Histograms
A1
A3
A0
A4
A2
Scale 9x
Column Features
D1
D2
D3
D4
Scale 3x
B1
B3
B0
B4
B2
C1
C3
C0
C4
C2
Scale 1x
C0
C1
C2
C3
C4
B0
B1
B2
B3
B4
A0
A1
A2
A3
A4
D1
D2
D3
D4
Absolute Feature Vector for Patch C0

Approach

1. Collect image data and ground truth depth
2. Extract features from images
3. Train a supervised learning model on features
4. ??
5. Profit !?

Markov Random Field

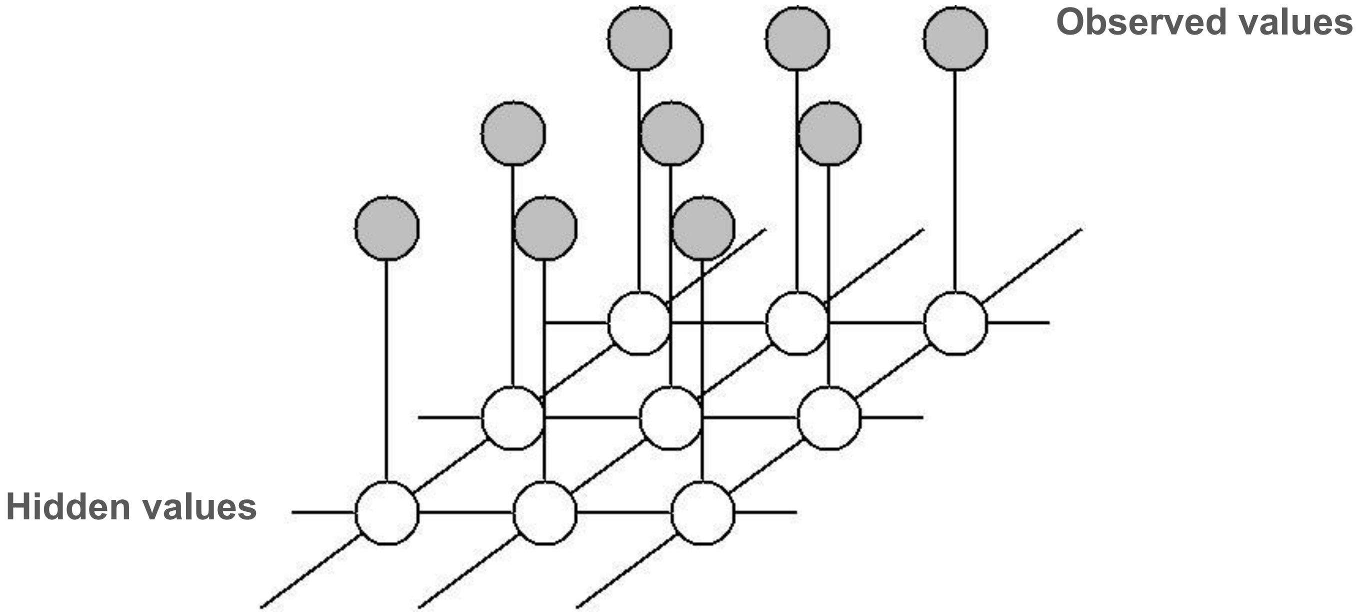


Source: http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/AV0809/ORCHARD/

Markov Random Field

- Undirected graph G = (V, E)
- Each node is independent given its 4-neighborhood (Markov Blanket)
- Joint P(x) = (1/Z) * exp(-E(x))
- P(X) ~ Gaussian, P(X) ~ Laplacian etc.
- Loops !

Hidden Markov Random Field



Source: http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/AV0809/ORCHARD/

Hidden Markov Random Field

- Ising Prior
- Discriminatively trained
- Exact posterior is intractable
- Approximate MAP inference by linear programming.

Gaussian MRF

$$P(d|X;\theta,\sigma) = \frac{1}{Z}\exp\left(-\sum_{i=1}^{M}\frac{(d_i(1) - x_i^T\theta_r)^2}{2\sigma_{1r}^2} - \sum_{s=1}^{3}\sum_{i=1}^{M}\sum_{j\in N_s(i)}\frac{(d_i(s) - d_j(s))^2}{2\sigma_{2rs}^2}\right)$$

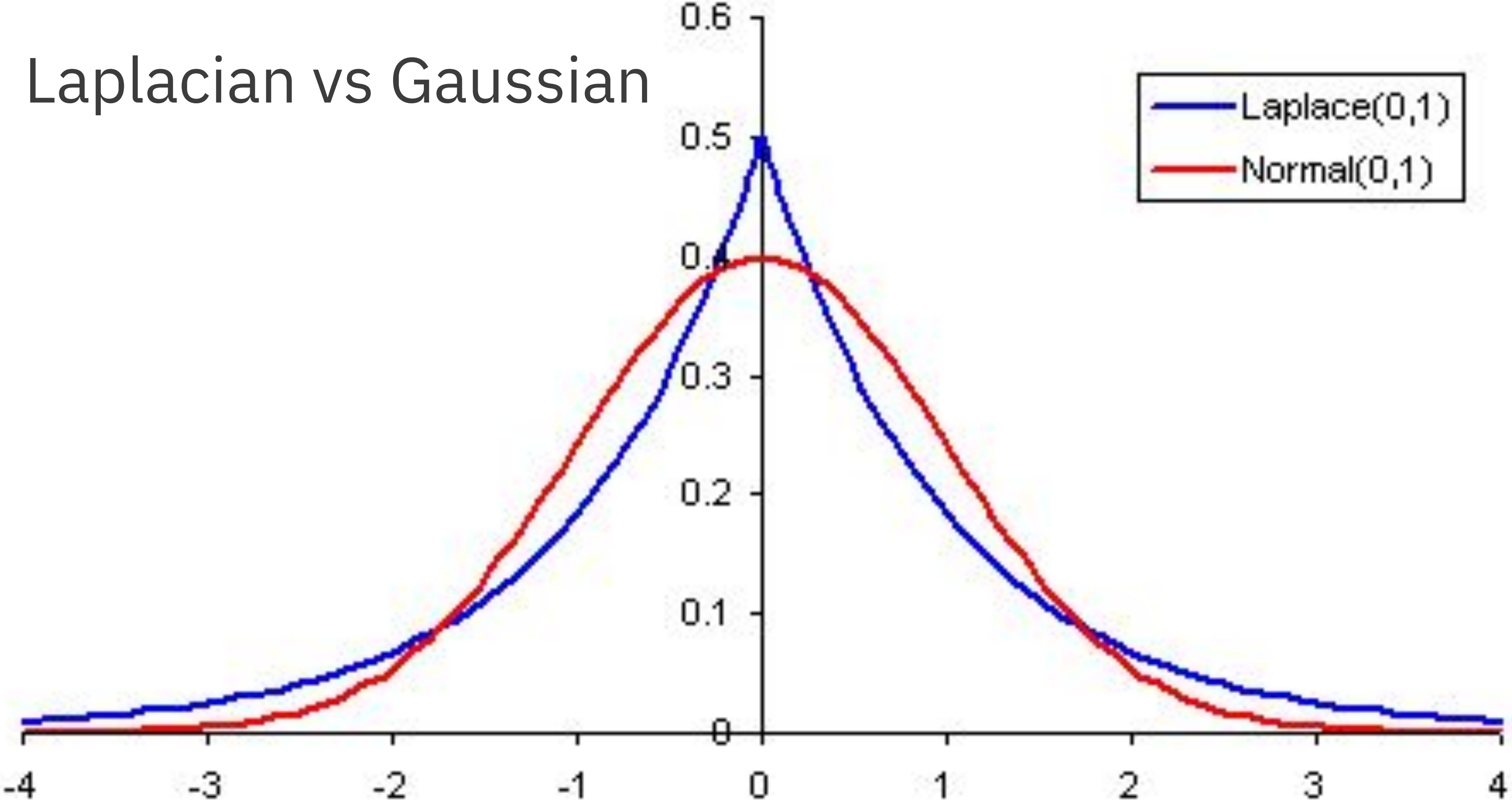
Laplacian MRF

$$P(d|X;\theta,\lambda) = \frac{1}{Z}\exp\left(-\sum_{i=1}^{M}\frac{|d_i(1) - x_i^T\theta_r|}{\lambda_{1r}} - \sum_{s=1}^{3}\sum_{i=1}^{M}\sum_{j\in N_s(i)}\frac{|d_i(s) - d_j(s)|}{\lambda_{2rs}}\right)$$

Laplacian vs Gaussian

- Relative depth histogram appears Laplacian.
- Heavier tails - more robust to outliers
- Sharper edges/transitions

Laplacian vs Gaussian



Results

Ground Truth | Gaussian | Laplacian