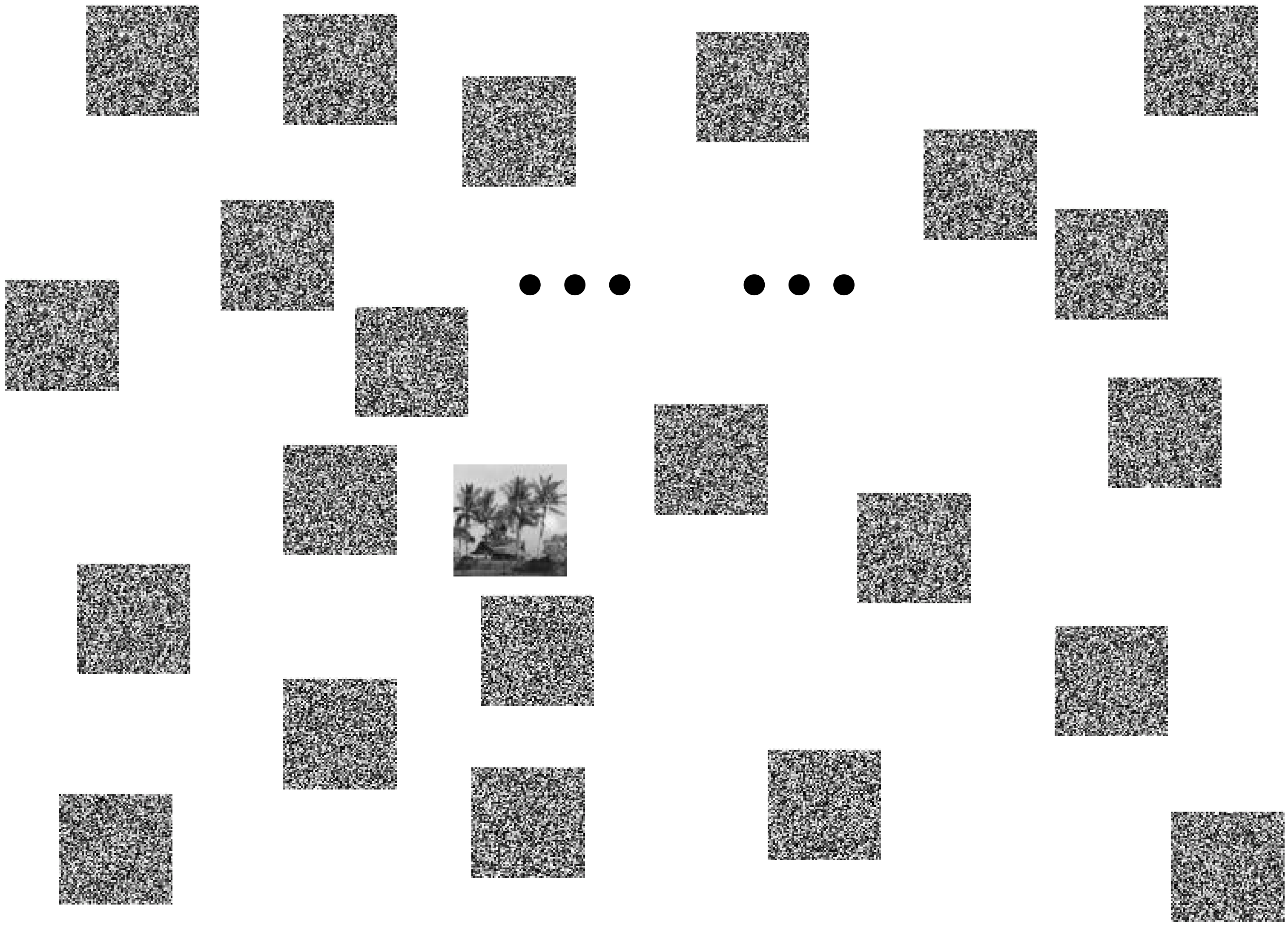
Modeling Images with Field of Gaussian Scale Mixtures

Siwei Lyu

Advisor: Eero P. Simoncelli

Center for Neural Science
Courant Institute of Mathematical Sciences
New York University

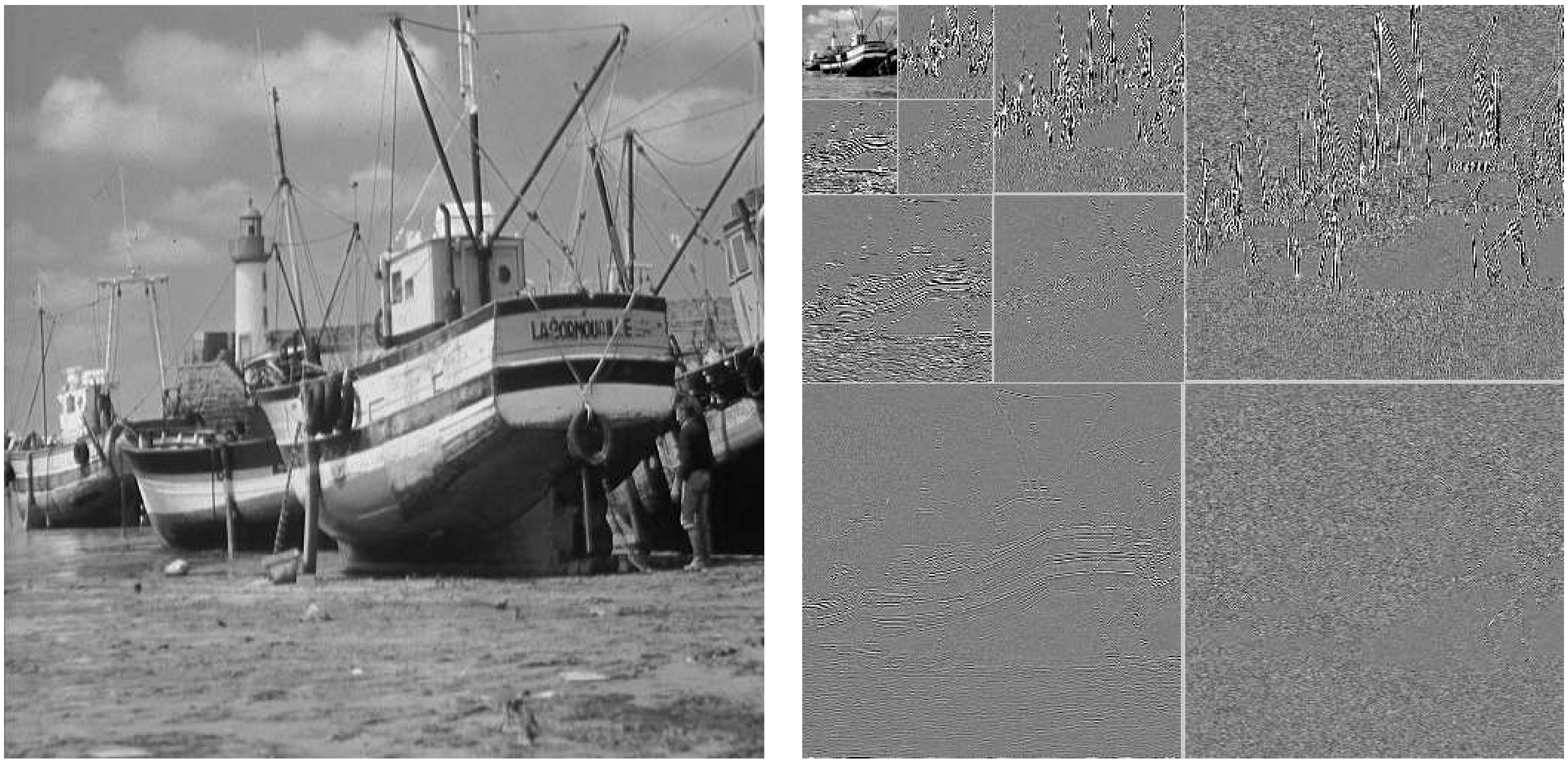
CIAR Summer School, University of Toronto, August 2007



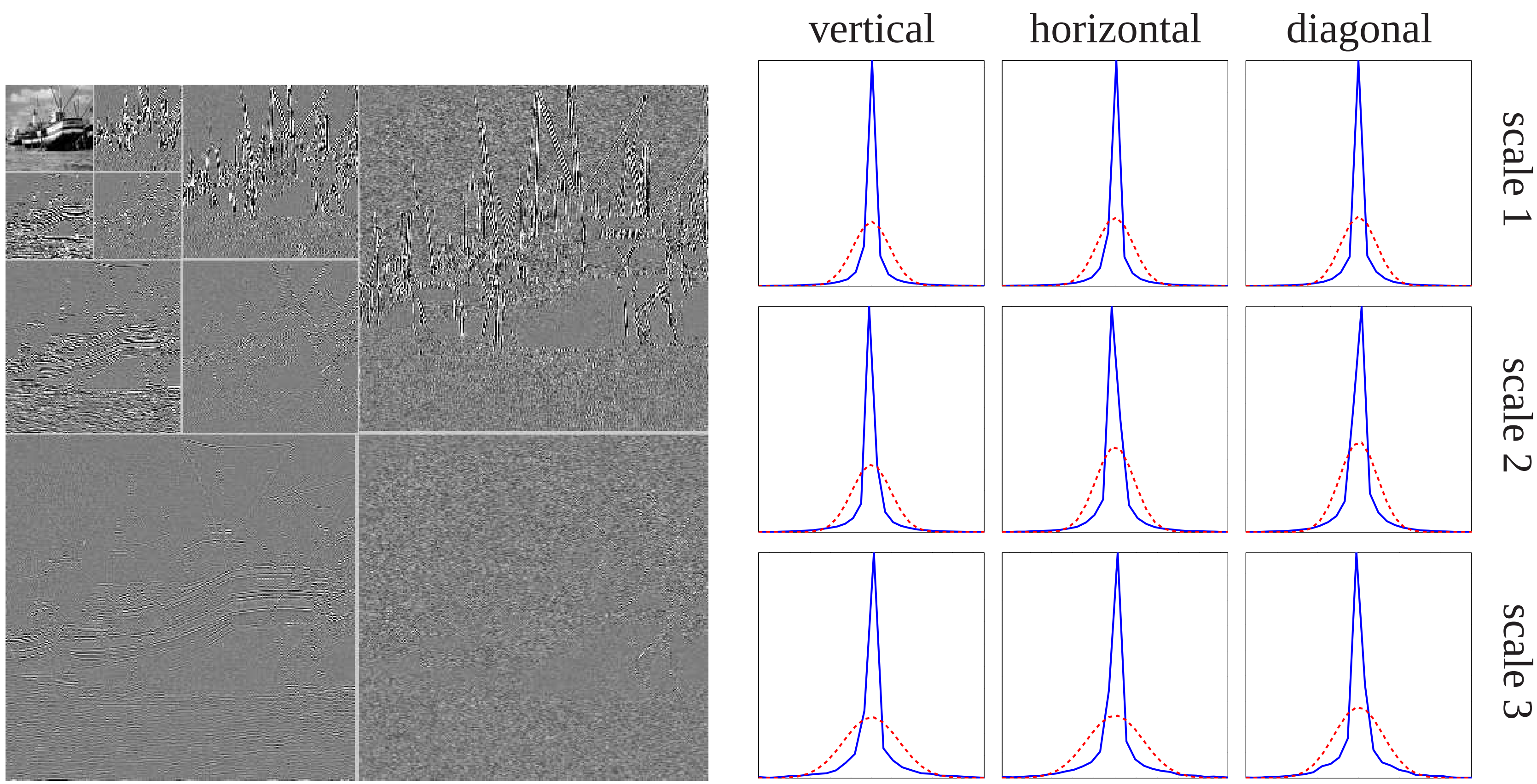
$\approx 10^{10000}$ 8-bit 65×65 images

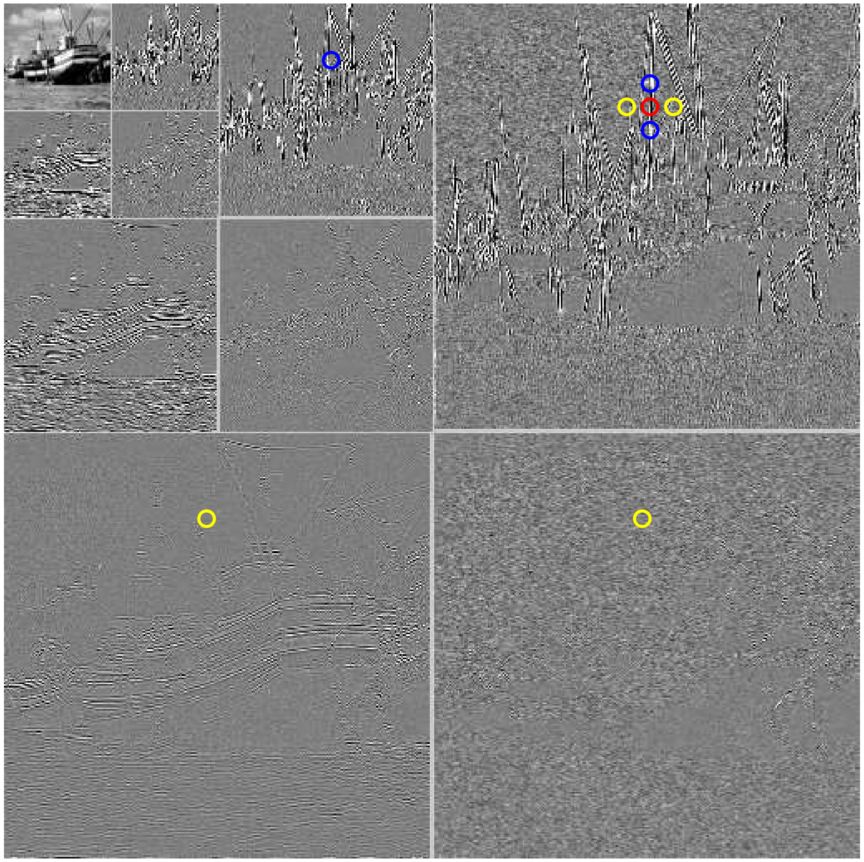
$\approx 10^{100}$ atoms in the universe

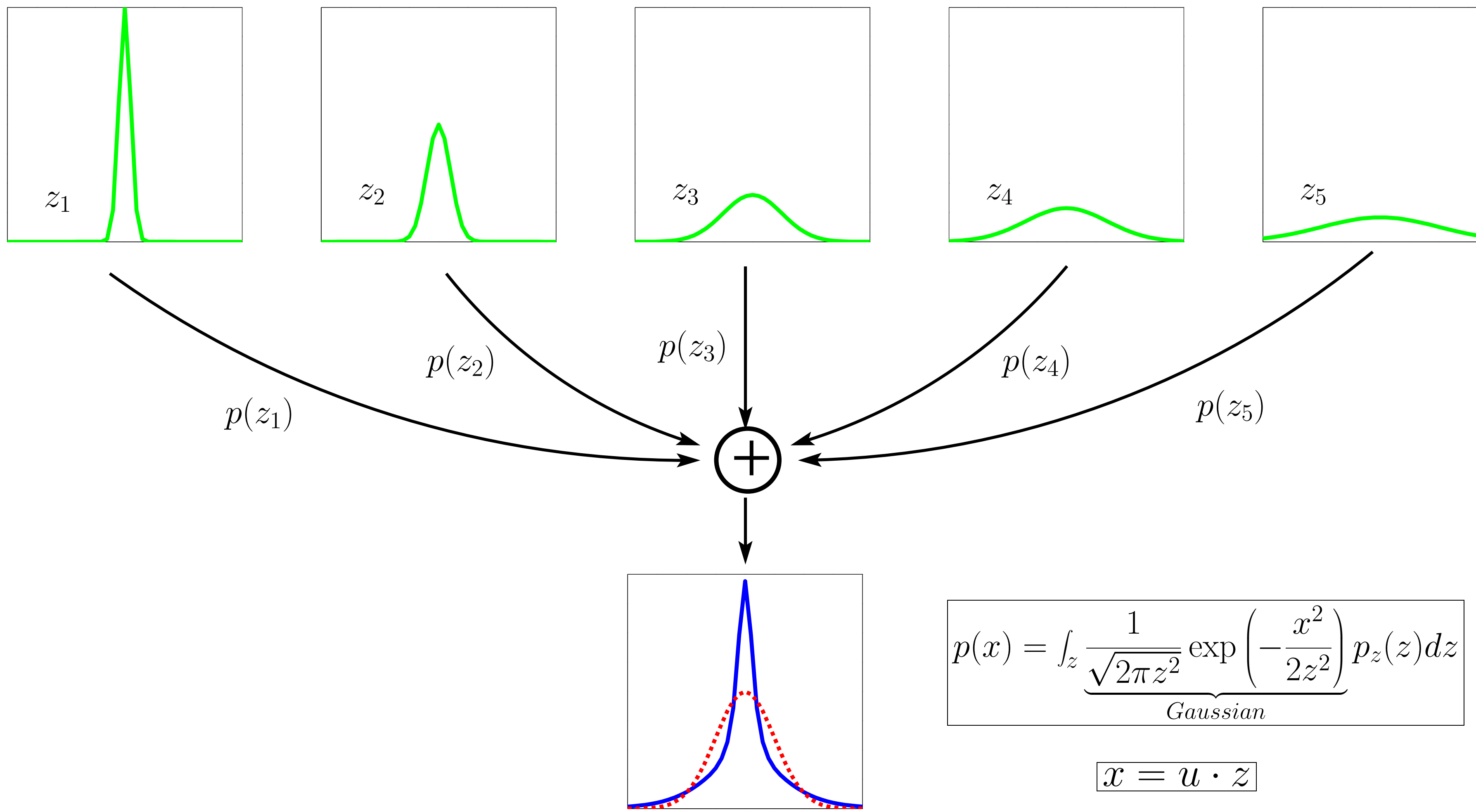
- compression (JPEG2000, 2002)
- denoising (Simoncelli & Adelson, 1996; Portilla et al., 2003; Roth & Black, 2005; Gehler & Welling, 2006)
- inpainting (Roth & Black, 2005)
- super-resolution (Freeman et al., 2000)
- texture synthesis(Zhu et al., 1998; Portilla & Simoncelli, 2000)
- segmentation (Figueiredo, 2005)
- forensics (Lyu & Farid, 2005; Lyu & Farid, 2006)



Multi-scale multi-orientation image decomposition (wavelet)

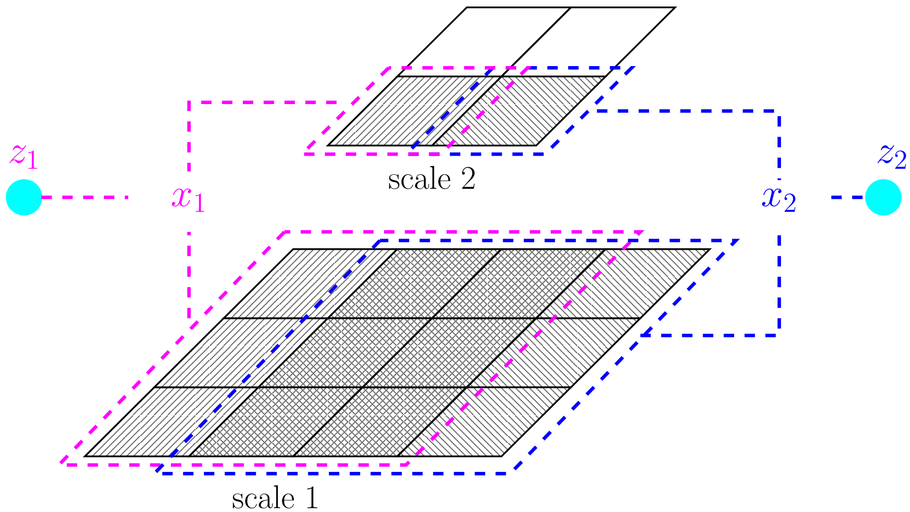

vertical
horizontal
diagonal
scale 1
scale 2
scale 3





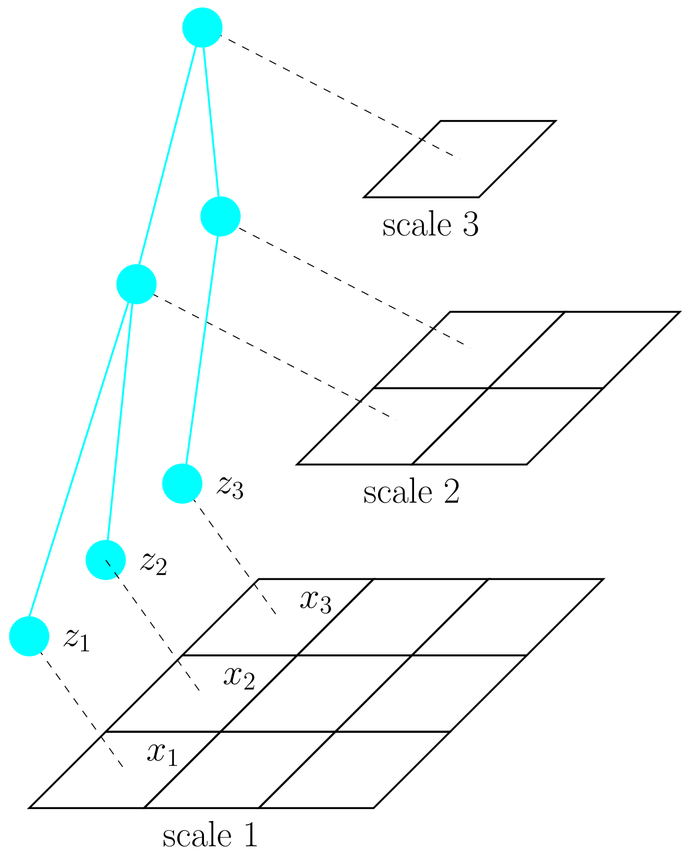
Gaussian scale mixtures (GSM)

Block GSM (Portilla et al., 2003)



overlapping blocks are independent GSM samples

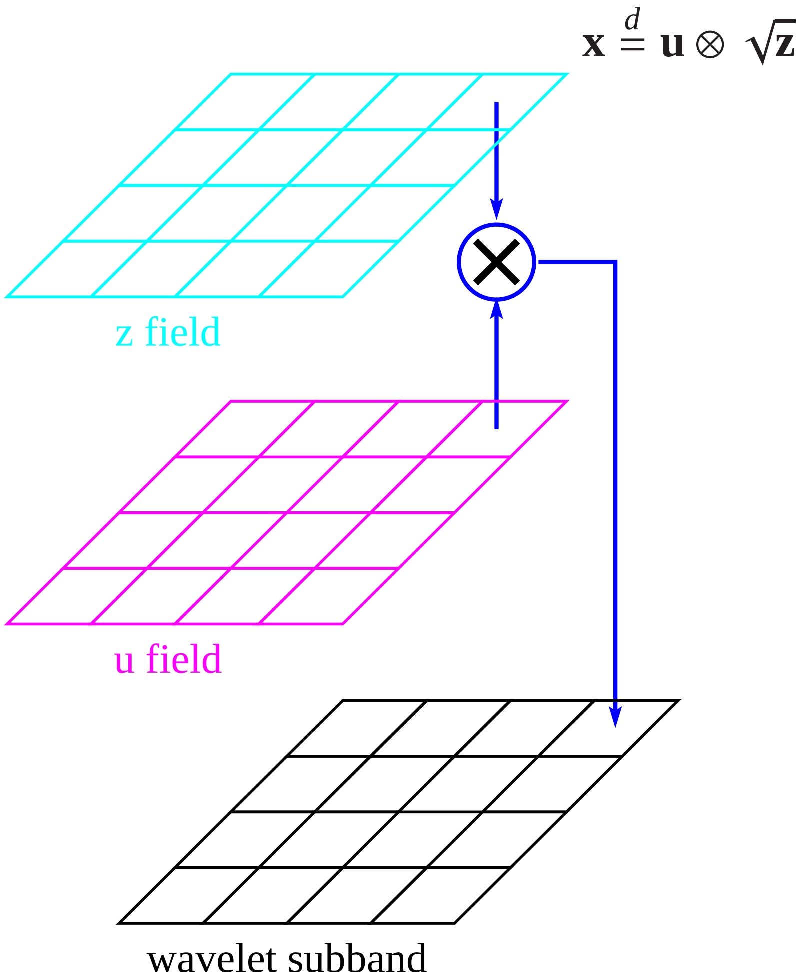
Multi-scale tree of GSM (Wainwright et al., 2001)



$$\mathbf{x} \stackrel{d}{=} \mathbf{u} \otimes \sqrt{\mathbf{z}}$$

$\mathbf{x}$	$\mathbf{u}$	$\mathbf{z}$	global consistency	no artificial boundary
block GSM	Gaussian vector	scalar random variable	✗	✓
GSM tree	independent Gaussian	multi-scale tree	✓	✗
field of GSM	Gaussian MRF	positive MRF	✓	✓

Field of Gaussian scale mixture (FoGSM)



- modeling wavelet subband
- spatial homogeneity
- $\mathbf{z}$: exponentiated Gauss MRF

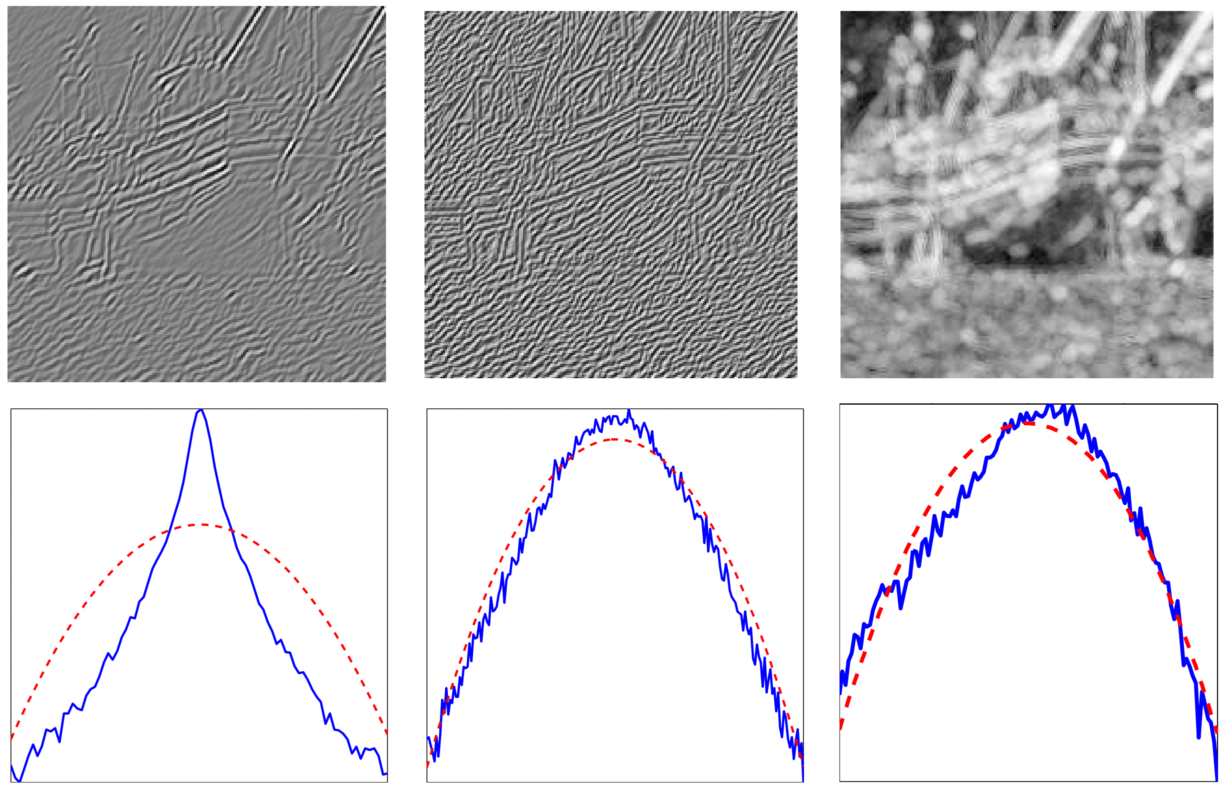
FoGSM: $\boxed{\mathbf{x} \stackrel{d}{=} \mathbf{u} \otimes \sqrt{\mathbf{z}}}$

- $\mathbf{u}$: zero mean homogeneous Gauss MRF
- $\mathbf{z}$: exponentiated homogeneous Gauss MRF
- $\mathbf{x}|\mathbf{z}$: *inhomogeneous* Gauss MRF
- $\mathbf{x} \oslash \sqrt{\mathbf{z}}$: homogeneous Gauss MRF
- marginal distribution is GSM
- generative model: efficient sampling

Fitting FoGSM model:

$$\boxed{(\hat{\mathbf{z}}, \hat{\theta}_u, \hat{\theta}_z) = \underset{\mathbf{z}, \theta_u, \theta_z}{\operatorname{argmax}} \log p(\mathbf{z}|\mathbf{x}; \theta_u, \theta_z)}$$

$$\begin{aligned}
&\cdots\ (\mathbf{z}, \theta_u, \theta_z) \\
\mathbf{z} &= \underset{\mathbf{z}}{\operatorname{argmax}} \log p(\mathbf{z}|\mathbf{x}; \theta_u, \theta_z) \\
\theta_u &= \underset{\theta_u}{\operatorname{argmax}} \log p(\mathbf{z}|\mathbf{x}; \theta_u, \theta_z) \\
\theta_z &= \underset{\theta_z}{\operatorname{argmax}} \log p(\mathbf{z}|\mathbf{x}; \theta_u, \theta_z) \\
&\cdots\ (\mathbf{z}, \theta_u, \theta_z)
\end{aligned}$$



subband $\mathbf{x}$ estimated $\mathbf{u}$ estimated $\log \mathbf{z}$

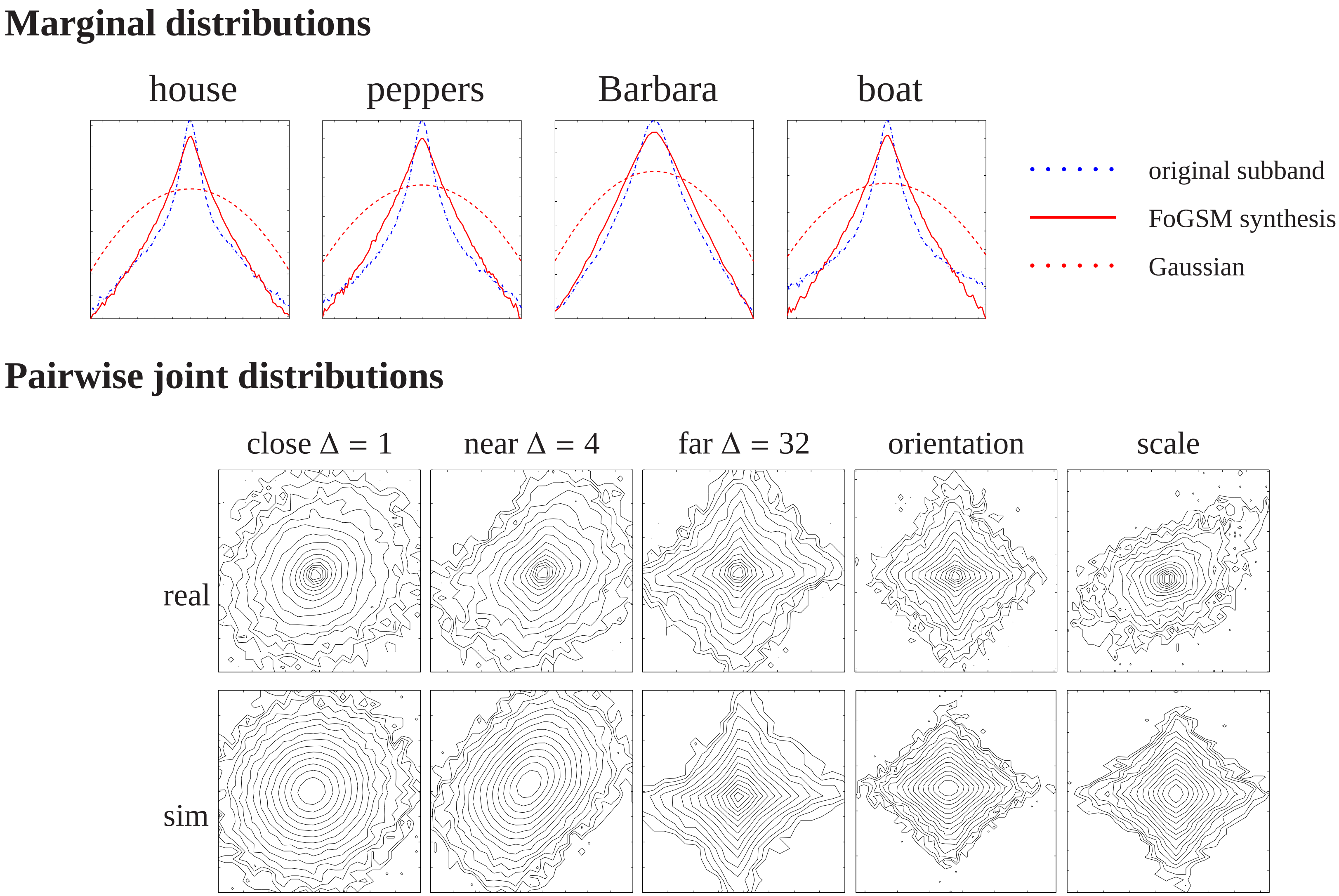
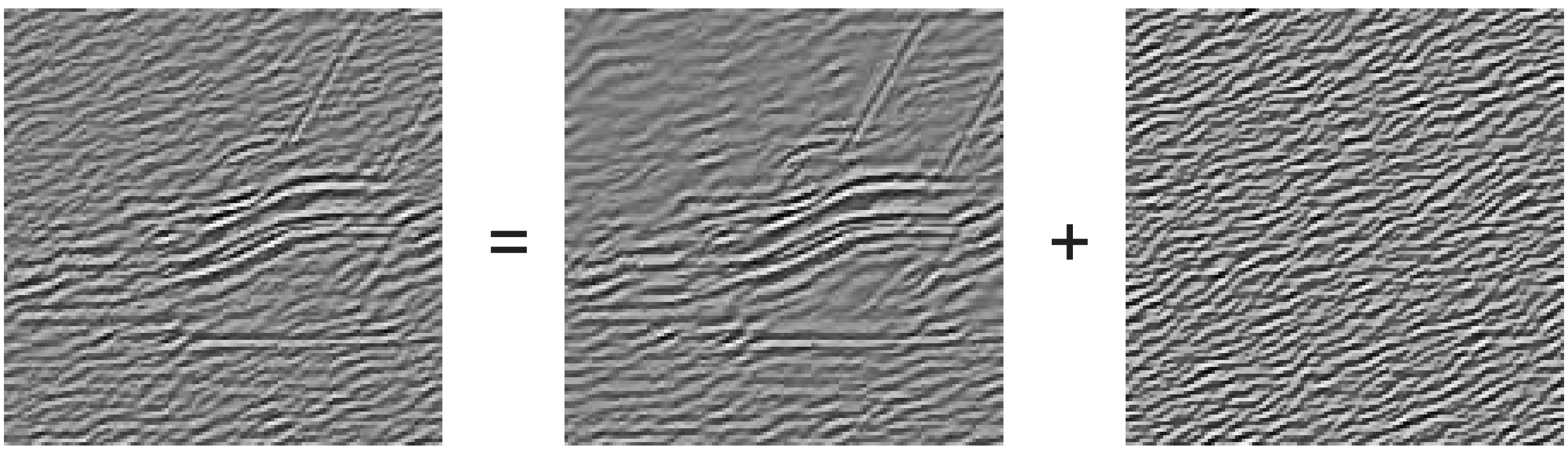

Marginal distributions
house
peppers
Barbara
boat
original subband
FoGSM synthesis
Gaussian
Pairwise joint distributions
close Δ = 1
near Δ = 4
far Δ = 32
orientation
scale
real
sim

Image denoising

In wavelet domain



y **x** **w**

- assume $\mathbf{x} \sim$ FoGSM
- estimate $\mathbf{x}$ from $\mathbf{y}$
  - *Maximum a posteriori*: $\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmax}}\, p(\mathbf{x}|\mathbf{y}) = \underset{\mathbf{x}}{\operatorname{argmax}} \int_{\mathbf{z}} p(\mathbf{x}, \mathbf{z}|\mathbf{y}) d\mathbf{z}$
  - *Bayes least squares*: $\hat{\mathbf{x}} = E(\mathbf{x}|\mathbf{y}) = \int_{\mathbf{x},\mathbf{z}} \mathbf{x} p(\mathbf{x}, \mathbf{z}|\mathbf{y}) d\mathbf{x} d\mathbf{z}$
  - *Maximum joint mode*: $(\hat{\mathbf{x}}, \hat{\mathbf{z}}) = \underset{\mathbf{x},\mathbf{z}}{\operatorname{argmax}} \log p(\mathbf{x}, \mathbf{z}|\mathbf{y})$
- combine with parameter estimation:

$$\boxed{(\hat{\mathbf{x}}, \hat{\mathbf{z}}, \hat{\theta}_u, \hat{\theta}_z) = \underset{\mathbf{x},\mathbf{z},\theta_u,\theta_z}{\operatorname{argmax}} \log p(\mathbf{x}, \mathbf{z}|\mathbf{y}; \theta_u, \theta_z)}$$

Denoising + parameter estimation

$$
\begin{aligned}
& \cdots \quad (\mathbf{x}, \mathbf{z}, \theta_u, \theta_z) \\
\mathbf{x} &= \underset{\mathbf{x}}{\operatorname{argmax}} \log p(\mathbf{x}, \mathbf{z} | \mathbf{y}; \theta_u, \theta_z) \\
\mathbf{z} &= \underset{\mathbf{z}}{\operatorname{argmax}} \log p(\mathbf{x}, \mathbf{z} | \mathbf{y}; \theta_u, \theta_z) \\
\theta_u &= \underset{\theta_u}{\operatorname{argmax}} \log p(\mathbf{x}, \mathbf{z} | \mathbf{y}; \theta_u, \theta_z) \\
\theta_z &= \underset{\theta_z}{\operatorname{argmax}} \log p(\mathbf{x}, \mathbf{z} | \mathbf{y}; \theta_u, \theta_z) \\
& \cdots \quad (\mathbf{x}, \mathbf{z}, \theta_u, \theta_z)
\end{aligned}
$$

- standard image set as used in (Portilla et al., 2003)
- steerable pyramid (Simoncelli & Freeman, 1995) of 8 orientations
- 5 level decomposition for image of 512×512
- 4 level decomposition for image of 256×256
- different noise levels
- initialization with block GSM denoised subband (Portilla et al., 2003)
- heuristic choice of model structure
- Running time: On average 4.5 hours to denoise a 512×512 image on PowerPC G5 (2.3 Ghz processor, 16 GB RAM) and unoptimized MATLAB (version R14) code.

σ/PSNR	Barbara	barco	boat	fingerprint	goldhill
1/48.13	**48.75** (48.40)	**49.88** (49.25)	**48.59** (48.43)	**48.67** (48.46)	**48.33** (48.33)
5/34.15	**38.65** (37.78)	**38.98** (38.39)	**37.39** (36.99)	**37.28** (36.69)	**37.16** (36.91)
10/28.13	**35.01** (34.01)	**35.05** (34.42)	**34.12** (33.58)	**33.28** (32.45)	**33.78** (33.38)
15/24.61	**32.85** (31.83)	**32.92** (32.27)	**32.31** (31.68)	**31.07** (30.15)	**31.99** (31.51)
25/20.17	**30.10** (29.07)	**30.44** (29.73)	**30.03** (29.34)	**28.45** (27.44)	**29.91** (29.37)
30/18.59	**29.12** (28.11)	**29.61** (28.88)	**29.22** (28.52)	**27.56** (26.54)	**29.22** (28.67)
50/14.15	**26.40** (25.45)	**27.36** (26.63)	**27.01** (26.35)	**25.11** (24.13)	**27.38** (26.82)
75/10.63	**24.29** (23.61)	**25.64** (24.96)	**25.33** (24.78)	**23.16** (22.39)	**25.93** (25.46)
100/8.13	**23.01** (22.61)	**24.44** (23.84)	**24.20** (23.79)	**21.78** (21.21)	**24.88** (24.53)
σ/PSNR	Flintstones	house	Lena	peppers	baboon
1/48.13	**49.79** (48.26)	**48.57** (48.87)	**47.92** (48.47)	**49.03** (48.42)	**50.08** (48.18)
5/34.15	**36.43** (35.65)	**38.98** (38.62)	**38.66** (38.48)	**37.91** (37.30)	**36.61** (35.06)
10/28.13	**32.47** (31.78)	**35.63** (35.27)	**35.94** (35.60)	**34.38** (33.73)	**31.69** (30.42)
15/24.61	**30.63** (29.86)	**33.89** (33.54)	**34.28** (33.91)	**32.34** (31.70)	**29.19** (28.04)
25/20.17	**28.29** (27.48)	**31.64** (31.32)	**32.11** (31.70)	**29.78** (29.18)	**26.34** (25.33)
30/18.59	**27.42** (26.60)	**30.82** (30.50)	**31.32** (30.89)	**28.89** (28.30)	**25.41** (24.45)
50/14.15	**24.82** (24.02)	**28.51** (28.23)	**29.12** (28.62)	**26.43** (25.93)	**23.03** (22.28)
75/10.63	**22.72** (21.94)	**26.69** (26.50)	**27.37** (26.92)	**24.53** (24.11)	**21.47** (20.99)
100/8.13	**21.24** (20.49)	**25.33** (25.31)	**26.12** (25.77)	**23.17** (22.80)	**20.58** (20.32)

PSNR: $20\log_{10}(255/\sigma_e)$, where σ_e is the standard deviation of the error.

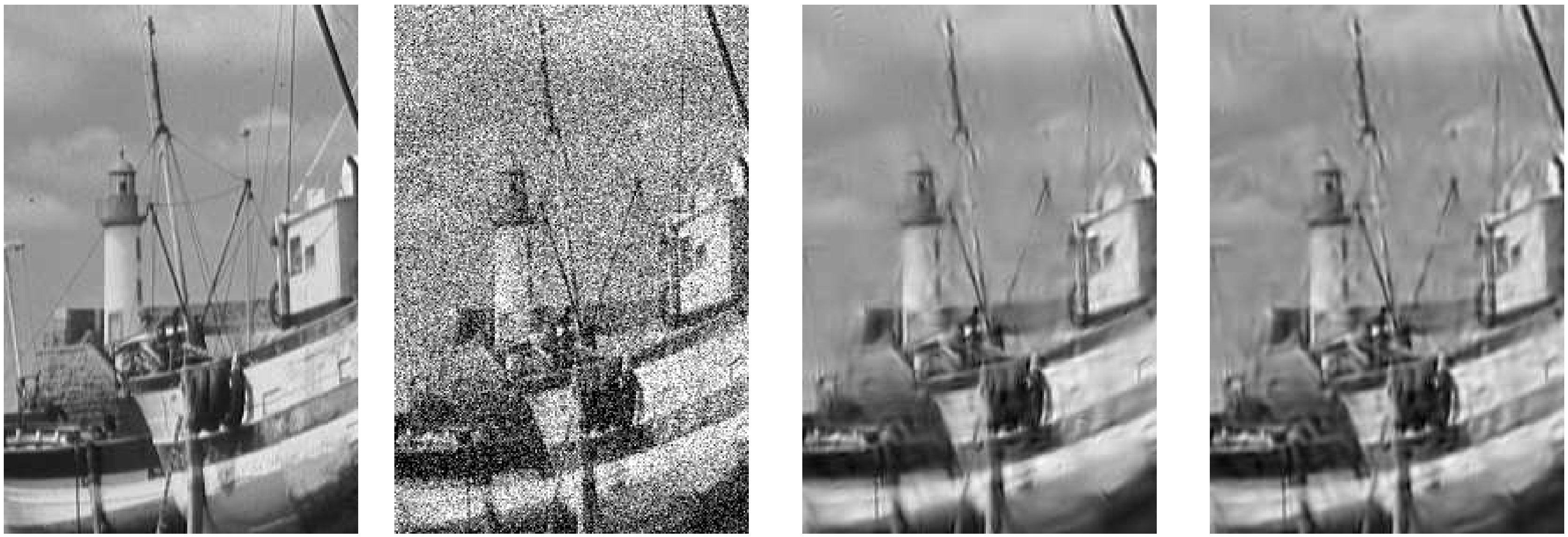
PSNR: **FoGSM** vs. (block GSM) (Portilla et al., 2003)

original image

noisy image (σ = 50) (PSNR = 14.15dB)

block GSM (Portilla et al., 2003) (PSNR = 26.34dB)

FoGSM (PSNR = 27.02dB)



original image

noisy image (σ = 50) (PSNR = 14.15dB)

block GSM (Portilla et al., 2003) (PSNR = 26.34dB)

FoGSM (PSNR = 27.02dB)

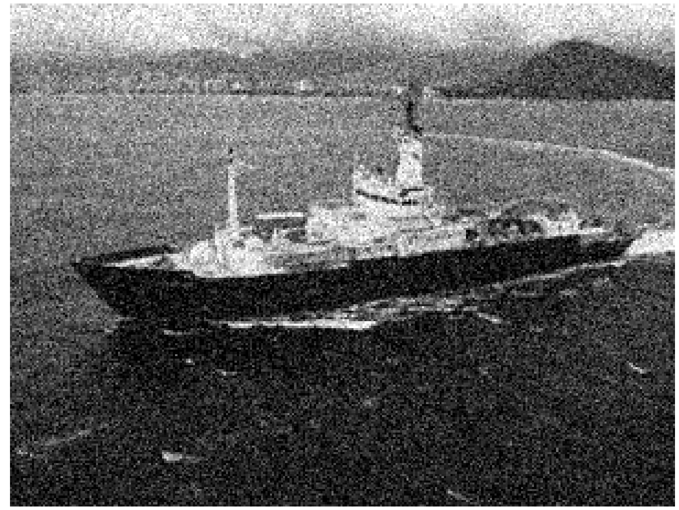
original image

noisy image (σ = 100) (PSNR = 8.13dB)

block GSM (Portilla et al., 2003) (PSNR = 22.61dB)

FoGSM (PSNR = 23.01dB)

original image

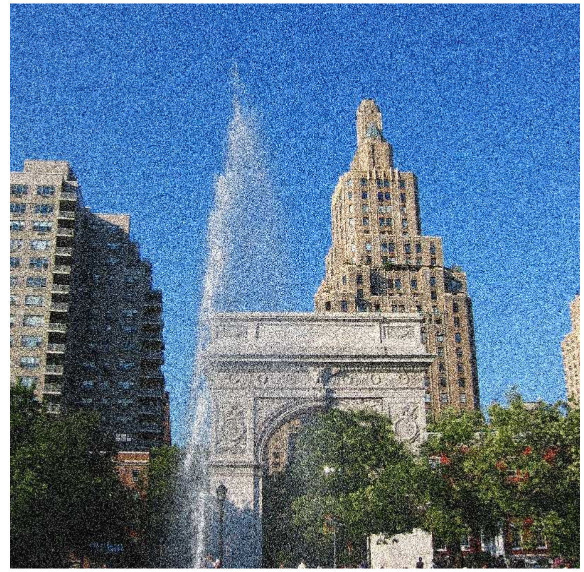


noisy image (σ = 25) (PSNR = 20.17dB)

local GSM (Portilla et al., 2003) (PSNR = 29.73dB)

FoGSM (PSNR = 30.44dB)

original image



noisy image (σ = 100) (PSNR = 8.13dB)

local GSM (Portilla et al., 2003) (PSNR = 22.61dB)

FoGSM (PSNR = 23.01dB)

Conclusion

- photographic images have statistical regularities in wavelet domain
- local GSM description + global MRF structure can capture such regularities
- construction with homogeneous Gauss MRF makes computation feasible
- applied to denoising, FoGSM achieved substantial improvement in performance

Future work

- general Markov neighborhoods across scales and orientations
- other modelings for the $\mathbf{z}$ field
- modeling of other image characteristics, e.g., local phases and orientations
- combining with neuronal spiking model

References

Figueiredo, M. (2005). Bayesian image segmentation using wavelet-based priors. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Diego.

Freeman, W. T., Pasztor, E. C., & Carmichael, O. T. (2000). Learning low-level vision. *International Journal of Computer Vision*, *40*, 25–47.

Gehler, P., & Welling, M. (2006). Products of “edge-perts”. In Y. Weiss, B. Schölkopf and J. Platt (Eds.), *Advances in neural information processing systems (nips)*, 419–426. Cambridge, MA: MIT Press.

JPEG2000 (2002). *JPEG2000: Image compression fundamentals, standards and practice*. Kluwer Academic.

Lyu, S., & Farid, H. (2005). How realistic is photorealistic? *IEEE Transactions on Signal Processing*, *53*, 845–850.

Lyu, S., & Farid, H. (2006). Steganalysis using higher-order image statistics. *IEEE Transactions on Information Forensics and Security*, *1*, 111–119.

Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *Int’l Journal of Computer Vision*, *40*, 49–71.

Portilla, J., Strela, V., Wainwright, M. J., & Simoncelli, E. P. (2003). Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Transactions on Image Processing*, *12*, 1338–1351.

Roth, S., & Black, M. (2005). Fields of experts: A framework for learning image priors. *IEEE Conference on Computer Vision and Patten Recognition (CVPR)* (pp. 860–867).

Simoncelli, E. P., & Adelson, E. H. (1996). Noise removal via Bayesian wavelet coring. *Proc 3rd IEEE Int’l Conf on Image Proc* (pp. 379–382). Lausanne: IEEE Sig Proc Society.

Simoncelli, E. P., & Freeman, W. T. (1995). The steerable pyramid: A flexible architecture for multi-scale derivative computation. *IEEE Int'l. Conf. on Image Proc.* (pp. 444–447).

Wainwright, M. J., Simoncelli, E. P., & Willsky, A. S. (2001). Random cascades on wavelet trees and their use in modeling and analyzing natural imagery. *Applied and Computational Harmonic Analysis*, *11*, 89–123.

Zhu, S. C., Wu, Y., & Mumford, D. (1998). Filters, random fields and maximum entropy (FRAME): Towards a unified theory for texture modeling. *Int'l. J. Comp. Vis.*, *27*, 107–126.