Optimal Denoising of Natural Images and their Multiscale Geometry and Density

Boaz Nadler

Department of Computer Science and Applied Mathematics
Weizmann Institute of Science, Israel.

Joint work with Anat Levin (WIS), Fredo Durand and Bill Freeman (MIT).

June 2013

[Levin and Nadler, CVPR 2011]
[Levin, Nadler, Durand, Freeman, ECCV 2012]

Classical Statistical Estimation vs. Image Restoration

Classical Statistical Inference:

$\{x_i\}_{i=1}^n$ are n i.i.d. random samples (observations) from some unknown density $p(x)$

(in parametric setting, $p(x) = p_\theta(x)$).

Task:
Given $\{x_i\}$ wish to estimate $p(x)$ (or θ).

Classical Statistical Estimation vs. Image Restoration

Classical Statistical Inference:

$\{x_i\}_{i=1}^n$ are n i.i.d. random samples (observations) from some unknown density $p(x)$

(in parametric setting, $p(x) = p_\theta(x)$).

Task:
Given $\{x_i\}$ wish to estimate $p(x)$ (or θ).

Key result: Consistency

$$\textit{As } n \to \infty,\ \hat{p}(x) \to p(x),\ \textit{and } \hat{\theta} \to \theta$$

Restoration Tasks and Image Priors

Many image restoration tasks: denoising, super-resolution, deblurring, etc, are *different*:

Restoration Tasks and Image Priors

Many image restoration tasks: denoising, super-resolution, deblurring, etc, are *different*:

Have only one (or even less) noisy observation per each unknown.

Restoration Tasks and Image Priors

Many image restoration tasks: denoising, super-resolution, deblurring, etc, are *different*:

Have only one (or even less) noisy observation per each unknown.

Image Restoration Tasks are *ill-posed* problems.

Restoration Tasks and Image Priors

Many image restoration tasks: denoising, super-resolution, deblurring, etc, are *different*:

Have only one (or even less) noisy observation per each unknown.

Image Restoration Tasks are *ill-posed* problems.

Given a degraded image y, there are multiple valid explanations x.

Restoration Tasks and Image Priors

Many image restoration tasks: denoising, super-resolution, deblurring, etc, are *different*:

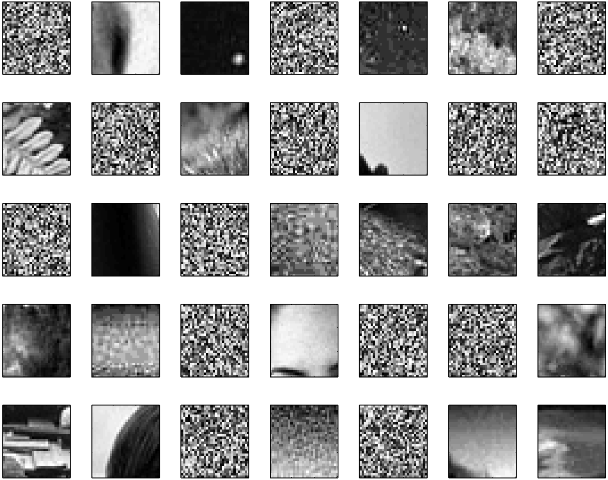
Have only one (or even less) noisy observation per each unknown.

Image Restoration Tasks are *ill-posed* problems.

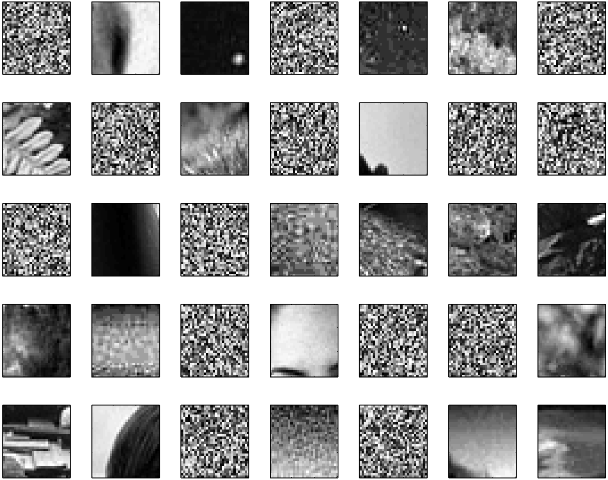
Given a degraded image y, there are multiple valid explanations x.

Natural image priors are crucial to rule out unlikely solutions.

On the Strength of Natural Image Priors



On the Strength of Natural Image Priors



Natural $k \times k$ patches are *extremely sparse* in the set of $256^{k \times k}$ patches

Optimal Image Restoration

Practical / Computational issues limit most current image priors to local patches.

Optimal Image Restoration

Practical / Computational issues limit most current image priors to local patches.

Hence, the restoration results are *suboptimal*.

Towards Optimal Image Restoration

Fundamental Theoretical Questions:

Towards Optimal Image Restoration

Fundamental Theoretical Questions:

- How accurate are currently used priors ?

Towards Optimal Image Restoration

Fundamental Theoretical Questions:

- How accurate are currently used priors ?

- How much gain is possible by increasing patch size ?

Towards Optimal Image Restoration

Fundamental Theoretical Questions:

- How accurate are currently used priors ?

- How much gain is possible by increasing patch size ?

- Computational issues aside, what is the optimal possible restoration ? Does it have a zero error ?

Towards Optimal Image Restoration

Fundamental Theoretical Questions:

- How accurate are currently used priors ?

- How much gain is possible by increasing patch size ?

- Computational issues aside, what is the optimal possible restoration ? Does it have a zero error ?

- How do these questions relate to natural image statistics ?

Towards Optimal Image Restoration

Fundamental Theoretical Questions:

- How accurate are currently used priors ?

- How much gain is possible by increasing patch size ?

- Computational issues aside, what is the optimal possible restoration ? Does it have a zero error ?

- How do these questions relate to natural image statistics ?

This talk will try to address some of these issues for a specific task: *denoising*

Outline

Denoising of Natural Images:

- Ill posed problem - solved using priors
- Optimal prior = (unknown) density of natural images

Outline

Denoising of Natural Images:

- Ill posed problem - solved using priors
- Optimal prior = (unknown) density of natural images

Questions:

Outline

Denoising of Natural Images:

- Ill posed problem - solved using priors
- Optimal prior = (unknown) density of natural images

Questions:

i) What is the optimal achievable denoising ? How is it related to the multiscale geometry and density of natural images ?

Outline

Denoising of Natural Images:

- Ill posed problem - solved using priors
- Optimal prior = (unknown) density of natural images

Questions:

i) What is the optimal achievable denoising ? How is it related to the multiscale geometry and density of natural images ?

ii) How far from optimal are current algorithms with approximate priors ?

Outline

Denoising of Natural Images:

- Ill posed problem - solved using priors
- Optimal prior = (unknown) density of natural images

Questions:

i) What is the optimal achievable denoising ? How is it related to the multiscale geometry and density of natural images ?

ii) How far from optimal are current algorithms with approximate priors ?

iii) How can we further improve current algorithms ?

Image Denoising - The Final Frontier

[Is denoising dead ? / Chatterjee & Milanfar]

[Is denoising dead ? / Chatterjee & Milanfar]

Should you do your Ph.D. in image denoising ?

[Is denoising dead ? / Chatterjee & Milanfar]

Should you do your Ph.D. in image denoising ?

Answer at end of talk...

Natural Image Denoising

Problem Setting:

x : unknown noise free natural image

$y = x + n$: observed image corrupted by noise.

In this talk, n - additive iid Gaussian $N(0, \sigma^2)$.

Natural Image Denoising

Problem Setting:

x : unknown noise free natural image

$y = x + n$: observed image corrupted by noise.

In this talk, n - additive iid Gaussian $N(0, \sigma^2)$.

Task:

Given y, construct clean version $\hat{y}$ (estimate of x).

Natural Image Denoising

Problem Setting:

x : unknown noise free natural image

$y = x + n$: observed image corrupted by noise.

In this talk, n - additive iid Gaussian $N(0, \sigma^2)$.

Task:

Given y, construct clean version $\hat{y}$ (estimate of x).

Performance Measure:

$$\text{MSE} = \mathbb{E}[\|x - \hat{y}\|^2]$$

Natural Image Denoising

Problem Setting:

x : unknown noise free natural image

$y = x + n$: observed image corrupted by noise.

In this talk, n - additive iid Gaussian $N(0, \sigma^2)$.

Task:

Given y, construct clean version $\hat{y}$ (estimate of x).

Performance Measure:

$$\text{MSE} = \mathbb{E}[\|x - \hat{y}\|^2]$$

Challenge:

Ill posed problem: $\#\text{unknowns} = \#\text{observations}$

Denoising Algorithms

Priors: All denoising algorithms employ implicitly or explicitly some prior on the unknown (noise-free) image.

Many different priors and denoising algorithms.

1980's - Gabor filters, anisotropic diffusion

1990's - wavelet based methods, total variation

2000's - sparse representations,

Denoising Algorithms

Priors: All denoising algorithms employ implicitly or explicitly some prior on the unknown (noise-free) image.

Many different priors and denoising algorithms.

1980's - Gabor filters, anisotropic diffusion

1990's - wavelet based methods, total variation

2000's - sparse representations,

05' Non-Local Means Algorithm

The NL-Means Algorithm

L. Yaroslavsky already suggested similar ideas in 1985.

[Buades & al 2005]
[Awate & Whittaker 2006]
[Barash & Comaniciu 2004]

$$\hat{y}(x_i) = \frac{1}{D(x_i)} \sum_{pixels\ j} K_\varepsilon(\mathbf{y}(x_i), \mathbf{y}(x_j)) y(x_j)$$

where

$$D(x_i) = \sum_{pixels\ j} K_\varepsilon(\mathbf{y}(x_i), \mathbf{y}(x_j))$$

The NL-Means Algorithm

L. Yaroslavsky already suggested similar ideas in 1985.

[Buades & al 2005]
[Awate & Whittaker 2006]
[Barash & Comaniciu 2004]

$$\hat{y}(x_i) = \frac{1}{D(x_i)} \sum_{pixels\ j} K_\varepsilon(\mathbf{y}(x_i), \mathbf{y}(x_j)) y(x_j)$$

where

$$D(x_i) = \sum_{pixels\ j} K_\varepsilon(\mathbf{y}(x_i), \mathbf{y}(x_j))$$

In matrix form

$$\hat{\mathbf{y}} = \mathbf{D}^{-1}\mathbf{W}\mathbf{y}$$

One interpretation: Random walk on image patches.

[Singer, Shkolnisky, N. 08']

Example: Input Image

Example: Original noise-free image

Example: Result of BM3D

Example: Result of KSVD

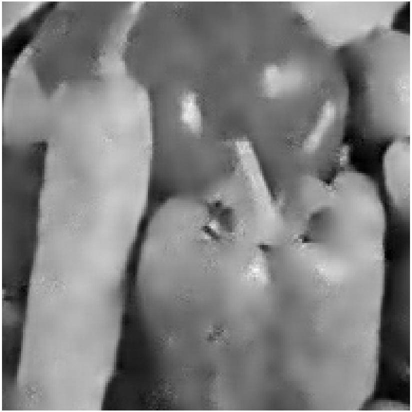


Image Denoising, Smoothness, Sparsity, Priors

Key Questions About *Natural* Image Denoising:

Image Denoising, Smoothness, Sparsity, Priors

Key Questions About *Natural* Image Denoising:

- *Optimal* Denoising in this setup ? Relation to geometry and density of natural image patches ?

Key Questions About *Natural* Image Denoising:

- *Optimal* Denoising in this setup ? Relation to geometry and density of natural image patches ?

- Are we there yet ? How much better can we improve on BM3D or other denoising algorithms ?

Statistical Framework

First consider denoising algorithms that estimate the central pixel x_c from a $k \times k$ window around it, dimension $d = k^2$.

In other words, think of x, y as both $k \times k$ patches.

Goal: Given $y = x + n$, estimate central pixel of x as best as possible.

Quality Measure: Mean Squared Error (MSE)

$$MSE = \mathbb{E}[(\hat{y}_c - x_c)^2],$$

$$PSNR = -10 \log_{10} MSE$$

Optimal Natural Image Denoising

The Bayesian Minimum Mean Squared Error Estimator

$$\begin{aligned}\mu(y) &= \mathbb{E}[x_c \mid y] = \int p(x_c \mid y) x_c dx \\ &= \frac{\int p(y|x) x_c dx}{\int p_d(x) p(y|x) dx}\end{aligned}$$

where $p_d(x)$ - density of $d = k \times k$ patches of natural images.

The Bayesian Minimum Mean Squared Error Estimator

$$\mu(y) = \mathbb{E}[x_c \,|\, y] = \int p(x_c \,|\, y) x_c dx = \frac{\int p(y|x) x_c dx}{\int p_d(x) p(y|x) dx}$$

where $p_d(x)$ - density of $d = k \times k$ patches of natural images.

NL-means

$$\hat{y}(x_i) = \frac{\sum_j K(\mathbf{y}(x_j), \mathbf{y}(x_i)) y_c(x_j)}{\sum_j K(\mathbf{y}(x_i), \mathbf{y}(x_j))}$$

Optimal Natural Image Denoising

The Bayesian Minimum Mean Squared Error Estimator

$$\begin{aligned} \mu(y) &= \mathbb{E}[x_c \,|\, y] = \int p(x_c \,|\, y) x_c dx \\ &= \frac{\int p(y|x) x_c dx}{\int p_d(x) p(y|x) dx} \end{aligned}$$

where $p_d(x)$ - density of $d = k \times k$ patches of natural images.

NL-means

$$\hat{y}(x_i) = \frac{\sum_j K(\mathbf{y}(x_j), \mathbf{y}(x_i)) y_c(x_j)}{\sum_j K(\mathbf{y}(x_i), \mathbf{y}(x_j))}$$

can be viewed as non-parametric approximation of the MMSE estimator using noisy image patches.

Upper and Lower Bounds on MMSE

If $p_d(x)$ were known - in principle could compute MMSE and see how far we are from optimality (at fixed window size d).

Upper and Lower Bounds on MMSE

If $p_d(x)$ were known - in principle could compute MMSE and see how far we are from optimality (at fixed window size d).

Challenge: Derive lower bound on $\mathbf{MMSE}_d$ w/out knowing $p_d(x)$?

Upper and Lower Bounds on MMSE

If $p_d(x)$ were known - in principle could compute MMSE and see how far we are from optimality (at fixed window size d).

Challenge: Derive lower bound on $\mathbf{MMSE}_d$ w/out knowing $p_d(x)$?

Solution: A trick ...

A brute force / elegant non-parametric approach

Don't know $p_d(x)$, but can sample from it.

A brute force / elegant non-parametric approach

Don't know $p_d(x)$, but can sample from it.

Approximate two identical yet *different* representations of MSE.

A brute force / elegant non-parametric approach

Don't know $p_d(x)$, but can sample from it.

Approximate two identical yet *different* representations of MSE.

ERROR representation

$$MSE = \int p(x) \int p(y \,|\, x)(\hat{\mu}(y) - x_c)^2 dydx$$

A brute force / elegant non-parametric approach

Don't know $p_d(x)$, but can sample from it.

Approximate two identical yet *different* representations of MSE.

ERROR representation

$$MSE = \int p(x) \int p(y \,|\, x)(\hat{\mu}(y) - x_c)^2 dydx$$

VARIANCE representation

$$MSE = \int p_\sigma(y) \int p(x \,|\, y)(\hat{\mu}(y) - x_c)^2 dxdy$$

A brute force / elegant non-parametric approach

Don't know $p_d(x)$, but can sample from it.

A brute force / elegant non-parametric approach

Don't know $p_d(x)$, but can sample from it.

Consider a huge set of $N = 10^{10}$ natural image patches.
Approximate MMSE_d non-parametrically.

$$\hat{\mu}(y) = \frac{\frac{1}{N}\sum_i p(y|x_i)x_{i,c}}{\frac{1}{N}\sum_i p(y|x_i)}, \tag{1}$$

where for Gaussian noise

$$p(y|x) = \frac{1}{(2\pi\sigma^2)^{d/2}} e^{-\frac{\|x-y\|^2}{2\sigma^2}} \tag{2}$$

and $d = k^2$.

Upper and Lower Bound on MMSE

Given a set of M noisy pairs $\{(\tilde{x}_j, y_j)\}_{j=1}^M$ and another independent set of N clean patches $\{x_i\}_{i=1}^N$, both randomly sampled from natural images, we compute

$$\mathrm{MMSE}^U = \frac{1}{M}\sum_j (\hat{\mu}(y_j) - \tilde{x}_{j,c})^2 \quad (3)$$

$$\mathrm{MMSE}^L = \frac{1}{M}\sum_j \hat{\mathcal{V}}(y_j) \quad (4)$$

where $\hat{\mathcal{V}}(y_j)$ is the approximated variance:

$$\hat{\mathcal{V}}(y_j) = \frac{\frac{1}{N}\sum_i p(y_j|x_i)(\hat{\mu}(y_j) - x_{i,c})^2}{\frac{1}{N}\sum_i p(y_j|x_i)} \quad (5)$$

Upper and Lower Bound on MMSE

Given a set of M noisy pairs $\{(\tilde{x}_j, y_j)\}_{j=1}^{M}$ and another independent set of N clean patches $\{x_i\}_{i=1}^{N}$, both randomly sampled from natural images, we compute

$$\mathrm{MMSE}^U = \frac{1}{M}\sum_j (\hat{\mu}(y_j) - \tilde{x}_{j,c})^2 \quad (3)$$

$$\mathrm{MMSE}^L = \frac{1}{M}\sum_j \hat{\mathcal{V}}(y_j) \quad (4)$$

where $\hat{\mathcal{V}}(y_j)$ is the approximated variance:

$$\hat{\mathcal{V}}(y_j) = \frac{\frac{1}{N}\sum_i p(y_j|x_i)(\hat{\mu}(y_j) - x_{i,c})^2}{\frac{1}{N}\sum_i p(y_j|x_i)} \quad (5)$$

Key Difference:
MMSE^U knows true noise-free patch $\tilde{x}$,
MMSE^L does not know $\tilde{x}$,

The Theoretical Part

MMSE^U is the MSE of yet another (not very fast) denoising algorithm. It is an *upper bound* on the optimal MSE.

The Theoretical Part

MMSE^U is the MSE of yet another (not very fast) denoising algorithm. It is an *upper bound* on the optimal MSE.

Claim: In expectation, MMSE^L is a *lower bound* on the MSE.

The Theoretical Part

MMSE^U is the MSE of yet another (not very fast) denoising algorithm. It is an *upper bound* on the optimal MSE.

Claim: In expectation, MMSE^L is a *lower bound* on the MSE.

If $N \gg 1$ is sufficiently large so that $\text{MMSE}^U \approx \text{MMSE}^L$ - then we have an accurate estimate of the MMSE !

Claim

$$\mathbb{E}_N[\hat{\mathcal{V}}(y)] = \mathcal{V}(y) + C(y)\mathcal{B}(y) + o\left(\frac{1}{N}\right), \qquad (6)$$

with

$$\begin{aligned} \mathcal{B}(y) &= \mathbb{E}[x_c^2|y] - 3\mathbb{E}[x_c|y]^2 - 2\mathbb{E}_{\sigma^*}[x_c^2|y] \\ &\quad +4\mathbb{E}[x_c|y]\mathbb{E}_{\sigma^*}[x_c|y] \qquad (7) \\ C(y) &= \frac{1}{N}\frac{p_{\sigma^*}(y)}{(4\pi\sigma^2)^{d/2}p(y)^2} \qquad (8) \end{aligned}$$

where p_{σ^*} and $\mathbb{E}_{\sigma^*}[\cdot]$ denote probability and expectation of random variables whose noise standard deviation is reduced from σ to $\sigma^* = \sigma/\sqrt{2}$.

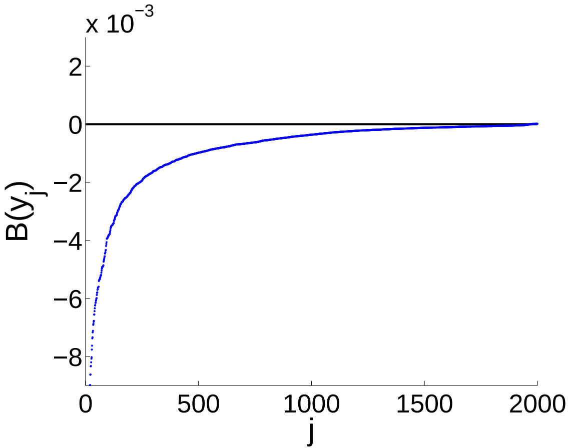
Theory Part II

Next, we consider a local Gaussian approximation for $p(x)$ around $x = y$, e.g., a Laplace approximation consisting of a second order Taylor expansion of $\ln p(x)$.
For a Gaussian distribution analytic computation of bias

Claim

For a Gaussian distribution, $\mathcal{B}(y) \leq 0$ *for all* y*.*

Numerical Example



Experiment Design

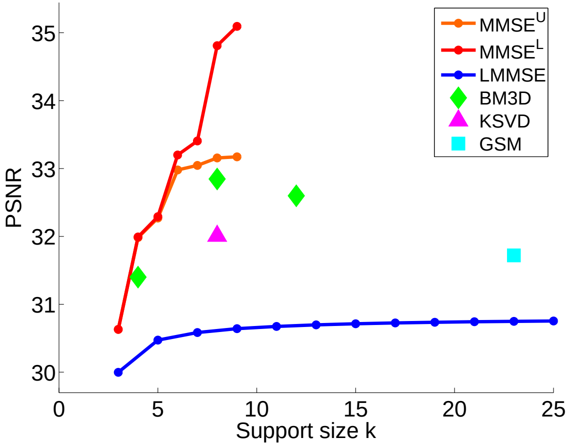
Set of $N = 10^{10}$ natural image patches out of 20K images from LabelMe dataset.

Independent set of $M = 2000$ patches from same dataset $\tilde{x}$.

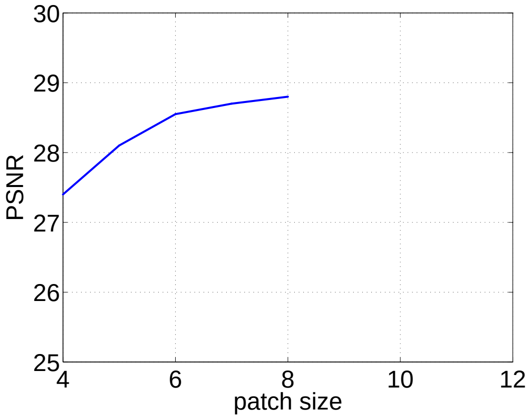
Add noise at different levels, $\sigma = 18, 55, 170$, and with different window sizes, $k = 3, 4, \ldots, 20$.

Compared MMSE^L, MMSE^U as well as MSE of various state-of-the-art denoising algorithms.

Experiments: $\sigma = 18$



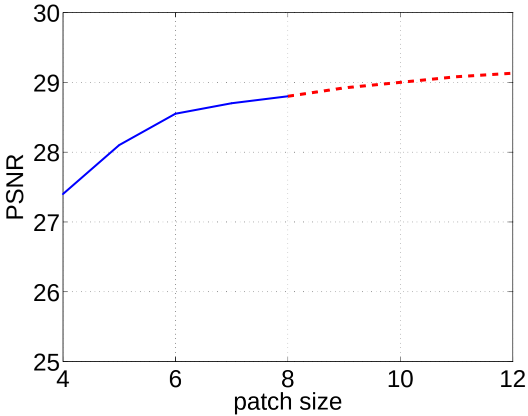
MMSE_d vs window size



For small patches / large noise non-parametric approach can accurately estimate MMSE.

[Levin and N. CVPR 2011]

MMSE_d vs window size



Extrapolation: What happens as $d \to \infty$?

Towards Optimal Image Denoising

Results of Levin and Nadler [CVPR 2011] imply that with small patches current algorithms are close to optimal.

Towards Optimal Image Denoising

Results of Levin and Nadler [CVPR 2011] imply that with small patches current algorithms are close to optimal.

Open Questions:

Towards Optimal Image Denoising

Results of Levin and Nadler [CVPR 2011] imply that with small patches current algorithms are close to optimal.

Open Questions:

-**How much gain is possible by increasing patch size ?**

Towards Optimal Image Denoising

Results of Levin and Nadler [CVPR 2011] imply that with small patches current algorithms are close to optimal.

Open Questions:

-**How much gain is possible by increasing patch size ?**

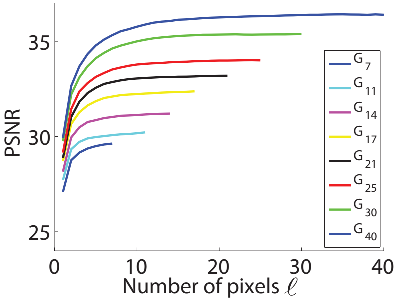
- Computational issues aside, what is the optimal possible restoration ? Does it have a zero error ?

Patch Complexity vs. PSNR gain

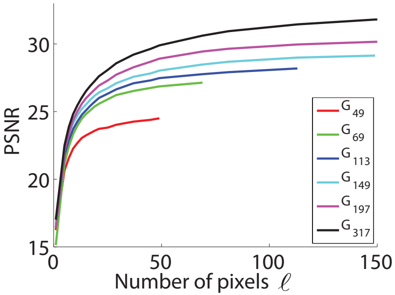
Empirical Results: Assign each pixel to group G_d where d is largest window size with reliable estimation.

Patch Complexity vs. PSNR gain

Empirical Results: Assign each pixel to group G_d where d is largest window size with reliable estimation.
PSNR vs. number of pixels in window, for different groups G_d



1D signals, $\sigma = 10$



2D signals, $\sigma = 20$

Patch Complexity vs. PSNR gain

Higher curves correspond to *smooth regions*, which flatten at larger patch dimensions, require smaller external database.

Lower curves correspond to *textured regions*, which not only run out of samples sooner, but also their curves flatten earlier, and require huge external database.

Patch Complexity vs. PSNR gain

Key conclusion:

Law of Marginal Return

Patch Complexity vs. PSNR gain

Key conclusion:

Law of Marginal Return

When increasing window size,

Smooth regions: large gain, no need to increase sample size

Patch Complexity vs. PSNR gain

Key conclusion:

Law of Marginal Return

When increasing window size,

Smooth regions: large gain, no need to increase sample size

Texture / edges: small gain, large increase in sample size

Adaptive Non-parametric Denoising

For each pixel x_i, find largest window w_d for which $\hat{\mu}_d$ is still reliable.

Adaptive Non-parametric Denoising

For each pixel x_i, find largest window w_d for which $\hat{\mu}_d$ is still reliable.

σ	20	35	50	75	100
Optimal Fixed	32.4	30.1	28.7	27.2	26.0
Adaptive	33.0	**30.5**	**29.0**	**27.5**	**26.4**
BM3D	**33.2**	30.3	28.6	26.9	25.6

Adaptive Non-parametric Denoising

For each pixel x_i, find largest window w_d for which $\hat{\mu}_d$ is still reliable.

σ	20	35	50	75	100
Optimal Fixed	32.4	30.1	28.7	27.2	26.0
Adaptive	33.0	**30.5**	**29.0**	**27.5**	**26.4**
BM3D	**33.2**	30.3	28.6	26.9	25.6

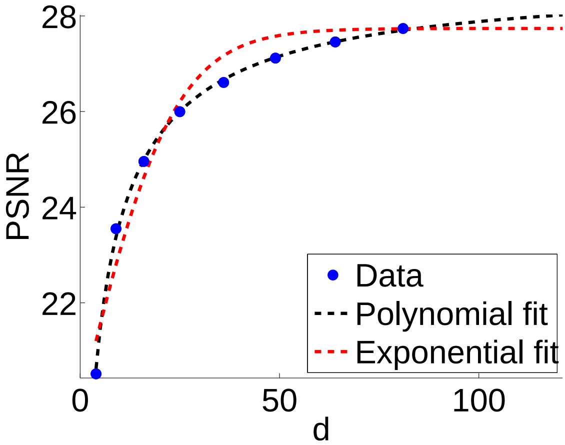
Similar idea suggested for adaptive Non-Local Means [Kervrann and Boulanger. 2006].

Challenge: Develop Practical Algorithm Based on this Adaptivity Principle

PSNR vs. window size

What is convergence rate of MMSE_d to MMSE_∞ ?

What is convergence rate of MMSE_d to MMSE_∞ ?



PSNR vs. window size

Empirically:

$$\text{MMSE}_d = e + \frac{c}{d^\alpha} \quad \text{with } \alpha \approx 1$$

PSNR vs. window size

Empirically:

$$\text{MMSE}_d = e + \frac{c}{d^\alpha} \quad \text{with } \alpha \approx 1$$

Can this be predicted theoretically ?

Dead Leaves Model

Simple Image Formation Model: *Dead Leaves* [Matheron 68']

Dead Leaves Model

Simple Image Formation Model: *Dead Leaves* [Matheron 68']

Image = random collection of finite size piece-wise constant regions

Region intensity = random variable with uniform distribution.

Multiscale Geometry of Natural Images

Scale Invariance: many natural image statistics are scale invariant: Down-sampling natural images does not change gradient distribution, segment sizes, etc.

[Ruderman 97', Field 78, etc.]

Scale Invariance: many natural image statistics are scale invariant: Down-sampling natural images does not change gradient distribution, segment sizes, etc.

[Ruderman 97', Field 78, etc.]

Claim

Under scale invariance

$$\Pr[\textit{pixel belongs to region of size s pixels}\ ] \propto 1/s$$

[Alvarez, Gousseau, Morel, 99']

Our Key Contribution:

Claim

Dead leaves model with scale invariance (and edge oracle) implies strictly positive $MMSE_\infty$ and power law convergence

$$MMSE_d = e + \frac{c}{d^\alpha}$$

with $\alpha = 1$

Our Key Contribution:

Claim

Dead leaves model with scale invariance (and edge oracle) implies strictly positive $MMSE_\infty$ *and power law convergence*

$$MMSE_d = e + \frac{c}{d^\alpha}$$

with $\alpha = 1$

Note that

$$e = \mathrm{MMSE}_\infty$$

so we can extrapolate empirical curve and estimate PSNR_∞.

Extrapolated PSNR$_\infty$ vs. Current Algorithms

σ	35	50	75	100
Extrapolated bound	**30.6**	**28.8**	**27.3**	**26.3**
KSVD	28.7	26.9	25.0	23.7
BM3D	30.0	28.1	26.3	25.0
EPLL	29.8	28.1	26.3	25.1

Extrapolated $PSNR_\infty$ vs. Current Algorithms

σ	35	50	75	100
Extrapolated bound	**30.6**	**28.8**	**27.3**	**26.3**
KSVD	28.7	26.9	25.0	23.7
BM3D	30.0	28.1	26.3	25.0
EPLL	29.8	28.1	26.3	25.1

Still some modest room for improvement

Summary

- Understanding the relation between natural image statistics and restoration taks is a fundamental problem.

Summary

- Understanding the relation between natural image statistics and restoration taks is a fundamental problem.

- Statistical framework to study optimal natural image denoising.

Summary

- Understanding the relation between natural image statistics and restoration taks is a fundamental problem.

- Statistical framework to study optimal natural image denoising.

- Non-parametric methods - law of diminishing return.

Summary

- Understanding the relation between natural image statistics and restoration taks is a fundamental problem.

- Statistical framework to study optimal natural image denoising.

- Non-parametric methods - law of diminishing return.

- most further gain in smooth regions $\rightarrow$ *adaptive* approach.

Summary

- Understanding the relation between natural image statistics and restoration taks is a fundamental problem.

- Statistical framework to study optimal natural image denoising.

- Non-parametric methods - law of diminishing return.

- most further gain in smooth regions $\rightarrow$ *adaptive* approach.

- Natural image statistics pose inherent limits on denoising. Empirically, power law convergence vs. window size, supported by Dead Leaves + Scale Invariance model.

Summary

- Understanding the relation between natural image statistics and restoration taks is a fundamental problem.

- Statistical framework to study optimal natural image denoising.

- Non-parametric methods - law of diminishing return.

- most further gain in smooth regions $\rightarrow$ *adaptive* approach.

- Natural image statistics pose inherent limits on denoising. Empirically, power law convergence vs. window size, supported by Dead Leaves + Scale Invariance model.

- Still room for improvement (not by non-parametric method).

Summary

- Understanding the relation between natural image statistics and restoration taks is a fundamental problem.

- Statistical framework to study optimal natural image denoising.

- Non-parametric methods - law of diminishing return.

- most further gain in smooth regions $\rightarrow$ *adaptive* approach.

- Natural image statistics pose inherent limits on denoising. Empirically, power law convergence vs. window size, supported by Dead Leaves + Scale Invariance model.

- Still room for improvement (not by non-parametric method).

- Implications to other low level vision tasks - deconvolution, super-resolution.

Summary

- Understanding the relation between natural image statistics and restoration taks is a fundamental problem.

- Statistical framework to study optimal natural image denoising.

- Non-parametric methods - law of diminishing return.

- most further gain in smooth regions $\rightarrow$ *adaptive* approach.

- Natural image statistics pose inherent limits on denoising. Empirically, power law convergence vs. window size, supported by Dead Leaves + Scale Invariance model.

- Still room for improvement (not by non-parametric method).

- Implications to other low level vision tasks - deconvolution, super-resolution.

THE END / THANK YOU !