Lecture 2 Image Denoising

Yuyao Zhang PhD
zhangyy8@shanghaitech.edu.cn
SIST Building-3 420

Outline

- Non local mean
- BM3D
- Supervised Deep Learning Denoising

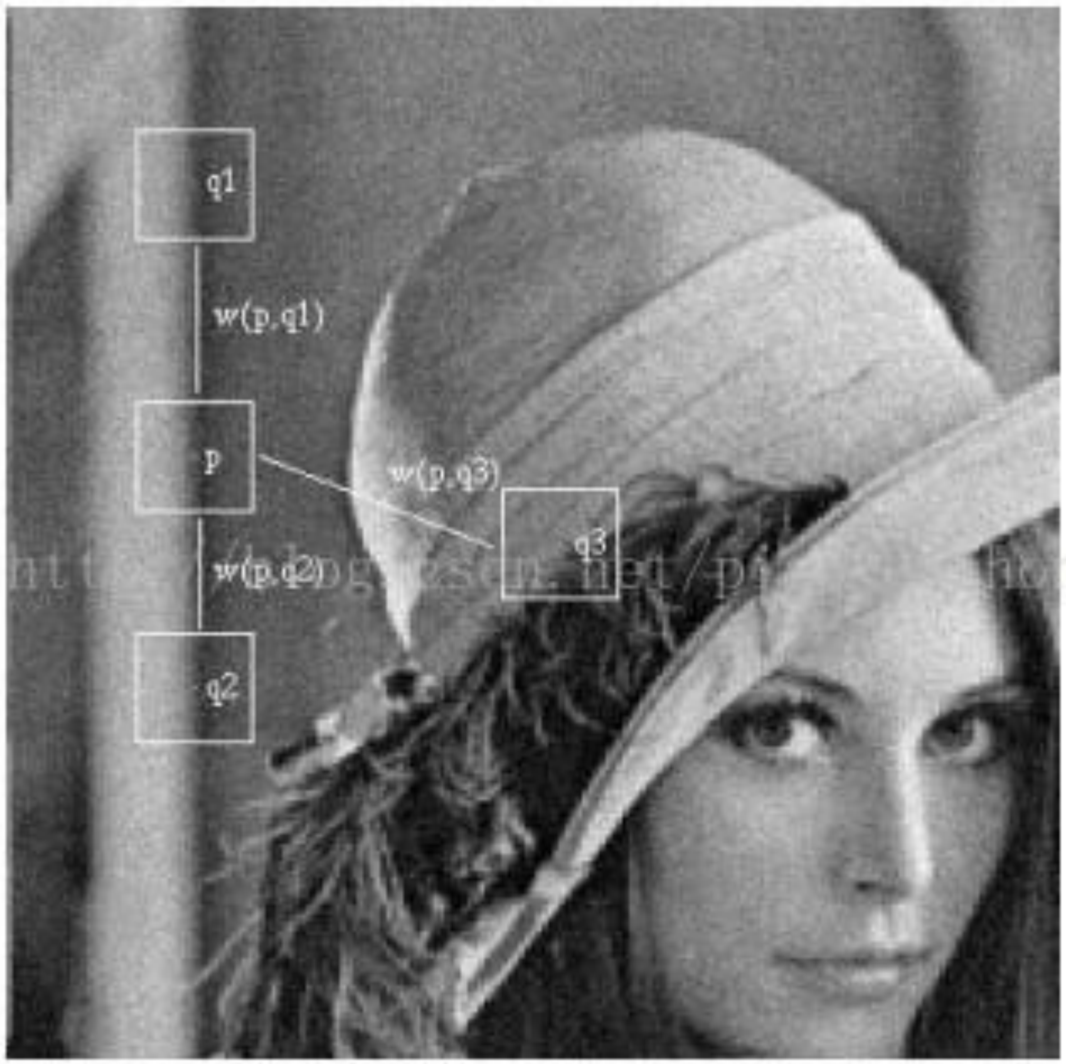
Resource: https://github.com/wenbihan/reproducible-image-denoising-state-of-the-art

Non-local Means

Reference:

[1] A. Buades, B. Coll, J.M. Morel “A non local algorithm for image denoising” IEEE Computer Vision and Pattern Recognition 2005, Vol 2, pp: 60-65, 2005.

Non-local Means



No need to stop at neighborhood. Instead search everywhere in the image.

Figure 1. Scheme of NL-means strategy. Similar pixel neighborhoods give a large weight, w(p,q1) and w(p,q2), while much different neighborhoods give a small weight w(p,q3).

Non-local Means

- Given a discrete noisy image $\boldsymbol{v} = \{\boldsymbol{v}(\boldsymbol{i}) | \boldsymbol{i} \in \boldsymbol{I}\}$, the estimated value $\boldsymbol{NL}[\boldsymbol{v}](\boldsymbol{i})$, for a pixel $\boldsymbol{i}$, is computed as a weighted average of all the pixels in the image,

$$\boldsymbol{NL}[\boldsymbol{v}](\boldsymbol{i}) = \sum_{\boldsymbol{j} \in \boldsymbol{I}} \boldsymbol{\omega}(\boldsymbol{i}, \boldsymbol{j}) \boldsymbol{v}(\boldsymbol{j})$$

where $\boldsymbol{\omega}(\boldsymbol{i}, \boldsymbol{j})$ denotes the weight of image that contributed to the denoised patch, which depends on the similarity between the pixels i and j, and satisfy the usual conditions $0 \le \boldsymbol{\omega}(\boldsymbol{i}, \boldsymbol{j}) \le 1$ and $\sum_{\boldsymbol{j} \in \boldsymbol{I}} \boldsymbol{\omega}(\boldsymbol{i}, \boldsymbol{j}) = \boldsymbol{1}$

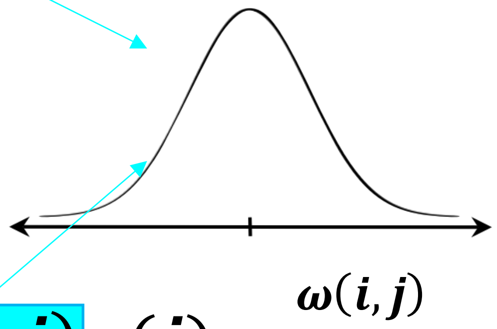
$$\boldsymbol{\omega}(\boldsymbol{i}, \boldsymbol{j}) = \frac{\boldsymbol{1}}{\boldsymbol{Z}(\boldsymbol{i})} \boldsymbol{e}^{-\frac{\left\| \boldsymbol{v}(\boldsymbol{\mathcal{N}}_{\boldsymbol{i}}) - \boldsymbol{v}(\boldsymbol{\mathcal{N}}_{\boldsymbol{j}}) \right\|_{2,a}^{2}}{\boldsymbol{h}^{2}}}; \boldsymbol{Z}(\boldsymbol{i}) = \sum_{\boldsymbol{j}} \boldsymbol{e}^{-\frac{\left\| \boldsymbol{v}(\boldsymbol{\mathcal{N}}_{\boldsymbol{i}}) - \boldsymbol{v}(\boldsymbol{\mathcal{N}}_{\boldsymbol{j}}) \right\|_{2,a}^{2}}{\boldsymbol{h}^{2}}}$$

Non-local Means vs Bilateral Filtering

- Non-local means Filtering

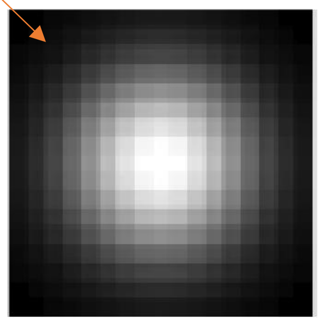
$$NL[v](i) = \sum_{j \in I} \omega(i,j)v(j)$$

Intensity range weighting: favor similar pixels (patches in case of non-local means)



- Bilateral Filtering

$$NL[v](i) = \sum_{j \in I} g(i,j)\omega(i,j)v(j)$$



Spatial weighting: favor nearby pixels

Everything put together

- Gaussian filtering

  Smooths everything nearby (even edges) Only depends on spatial distance Bilateral Filtering.

- Bilateral filtering

  Smooths 'close' pixels in space and intensity Depends on spatial and intensity distance.

- Non-local means Filtering

  Smooths similar patches no matter how far away Only depends on intensity distance.

Non-local Means

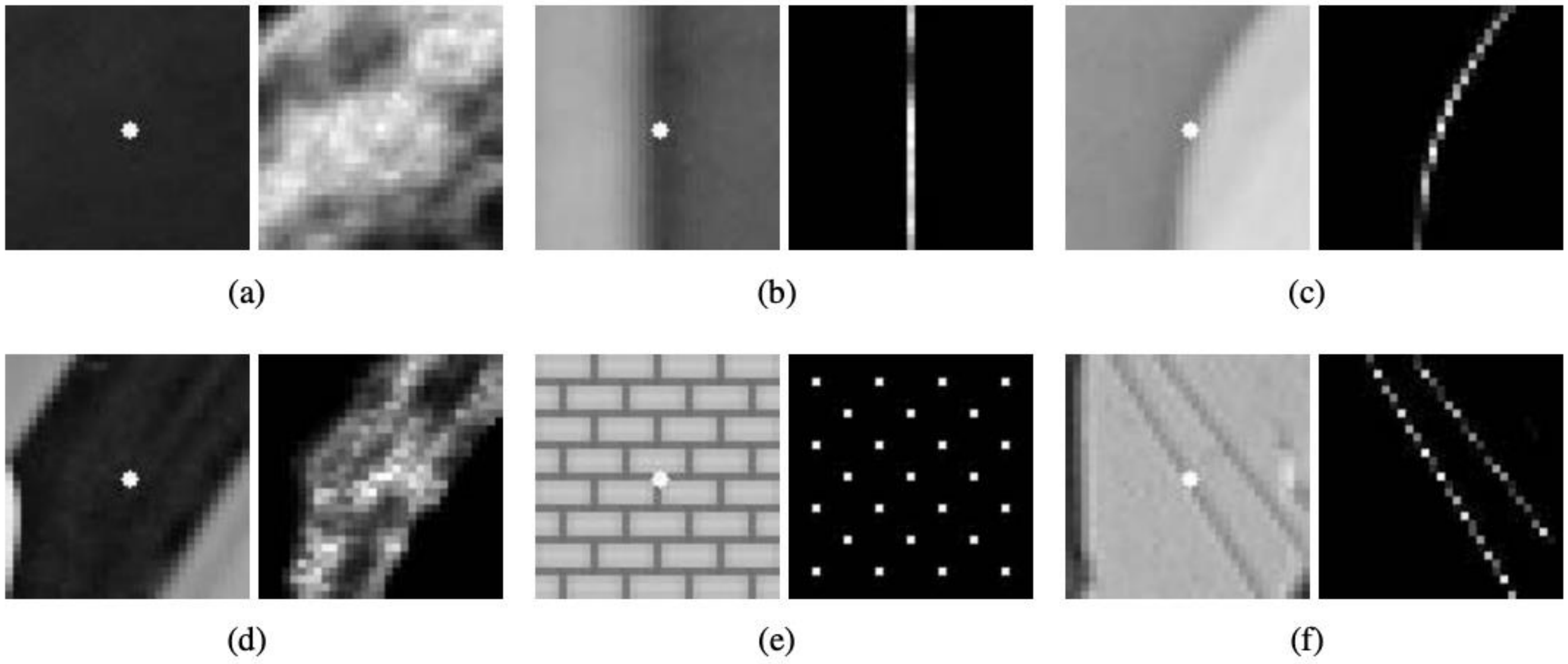


Figure 2. Display of the NL-means weight distribution used to estimate the central pixel of every image. The weights go from 1(white) to zero(black).

Non-local Means: Result

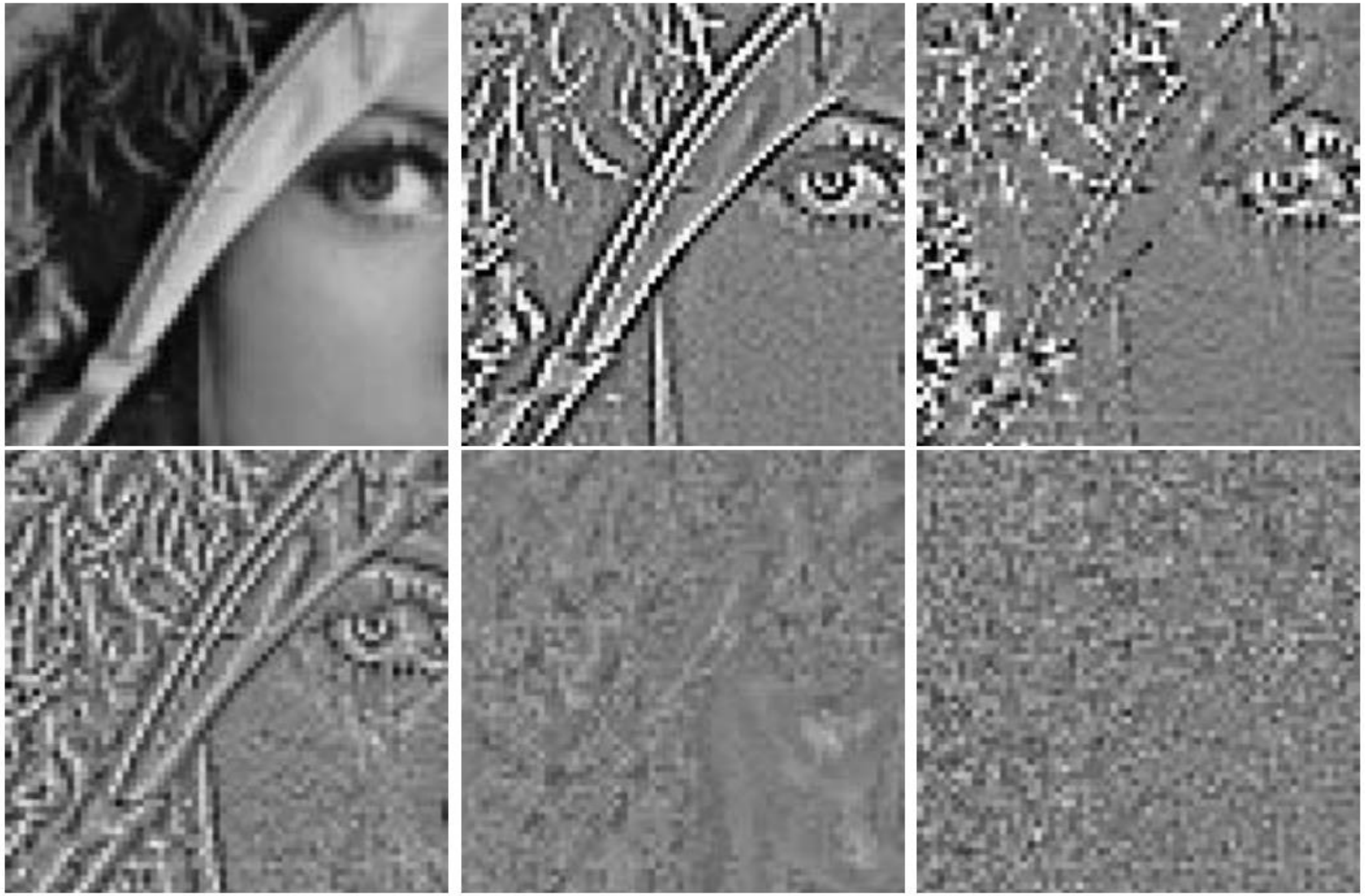
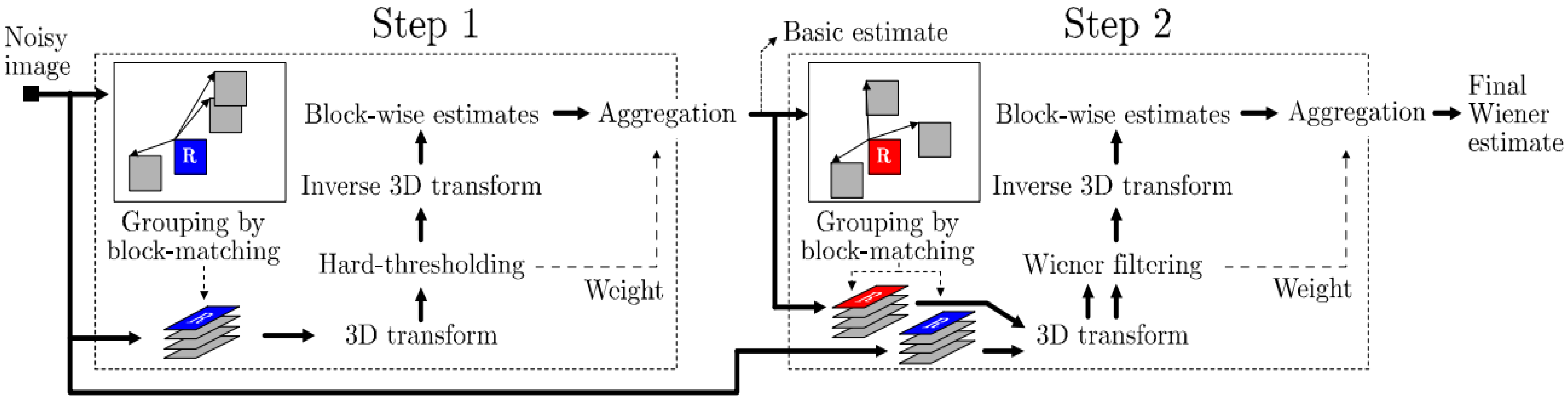


Figure 4. Method noise experience on a natural image. Displaying of the image difference. From left to right and from top to bottom: original image, Gauss filtering, anisotropic filtering, Total variation minimization, Neighborhood filtering and NL-means algorithm. The visual experiments corroborate the formulas of section 2.

Block-matching and 3D Filtering（BM3D）

reference :
[1] Dabov K, Foi A, Katkovnik V, et al. Image denoising by sparse 3-D transform-domain collaborative filtering[J]. IEEE Transactions on image processing, 2007, 16(8): 2080-2095.

BM3D: Overview



BM3D: Block matching

- Grouping and Collaborative Hard thresholding

$$d\left(Z_{x_R}, Z_x\right) = \frac{\left\| \Upsilon'\left(\mathcal{T}_{2\mathrm{D}}^{\mathrm{ht}}\left(Z_{x_R}\right)\right) - \Upsilon'\left(\mathcal{T}_{2\mathrm{D}}^{\mathrm{ht}}(Z_x)\right)\right\|_2^2}{\left(N_1^{\mathrm{ht}}\right)^2} \qquad (4)$$

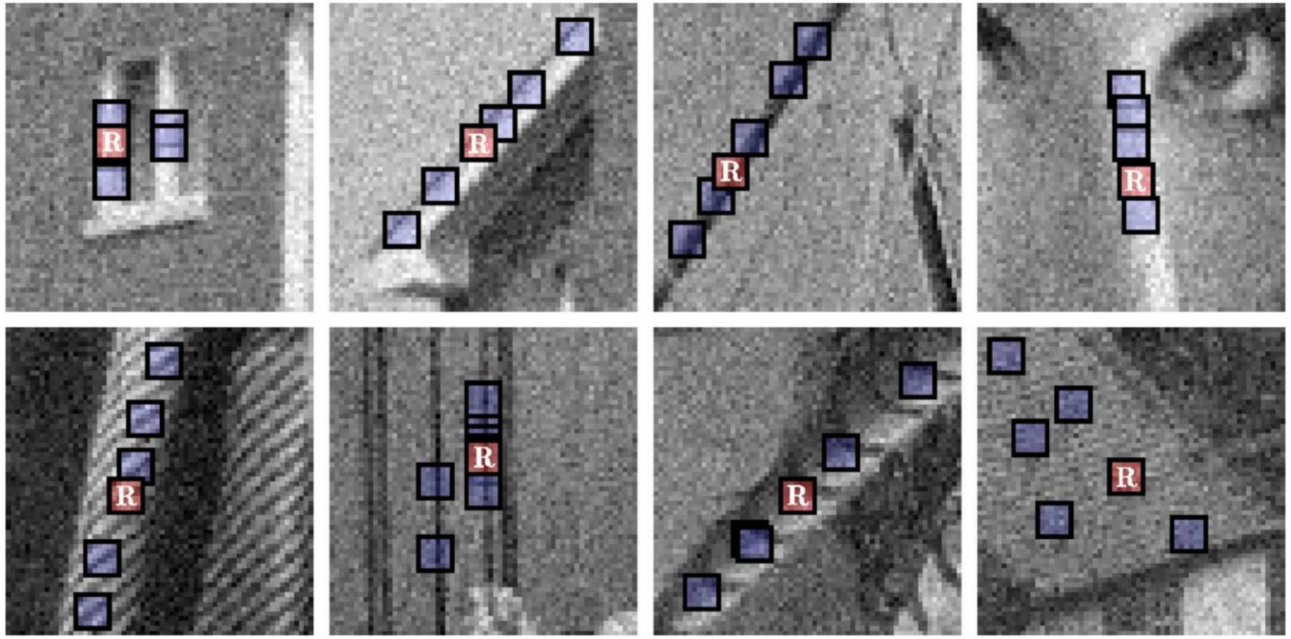
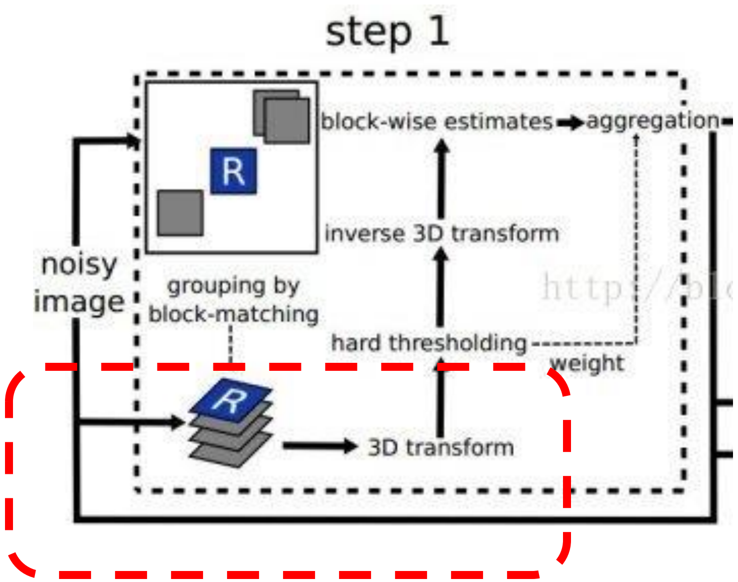


Fig. 1. Illustration of grouping blocks from noisy natural images corrupted by white Gaussian noise with standard deviation 15 and zero mean. Each fragment shows a reference block marked with "R" and a few of the blocks matched to it.



BM3D: Block matching

$$\widehat{\mathbf{Y}}^{\mathrm{ht}}_{S^{\mathrm{ht}}_{x_R}} = \mathcal{T}^{\mathrm{ht}^{-1}}_{3\mathrm{D}} \left(\Upsilon \left(\mathcal{T}^{\mathrm{ht}}_{3\mathrm{D}} \left(\mathbf{Z}_{S^{\mathrm{ht}}_{x_R}} \right) \right) \right) \qquad (6)$$

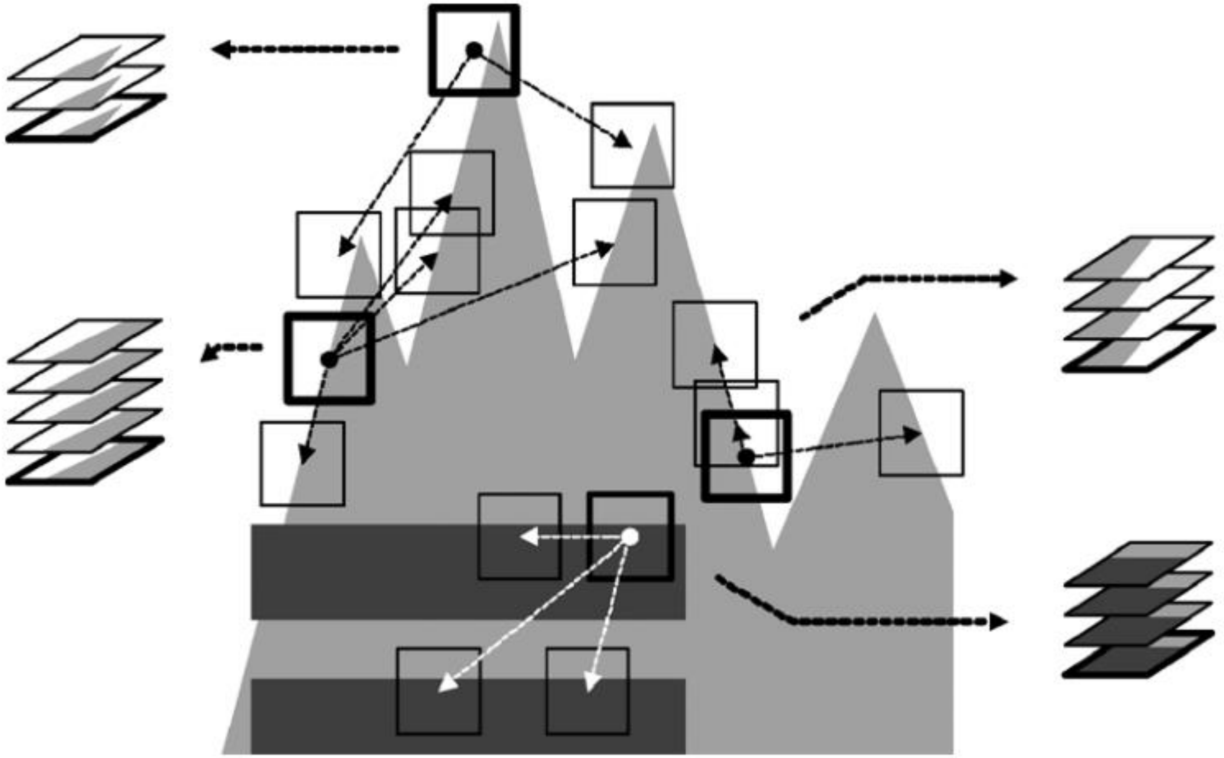
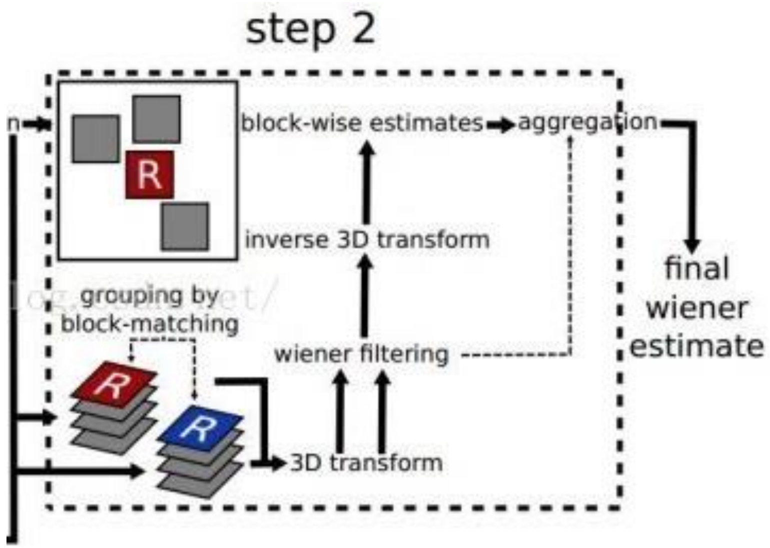
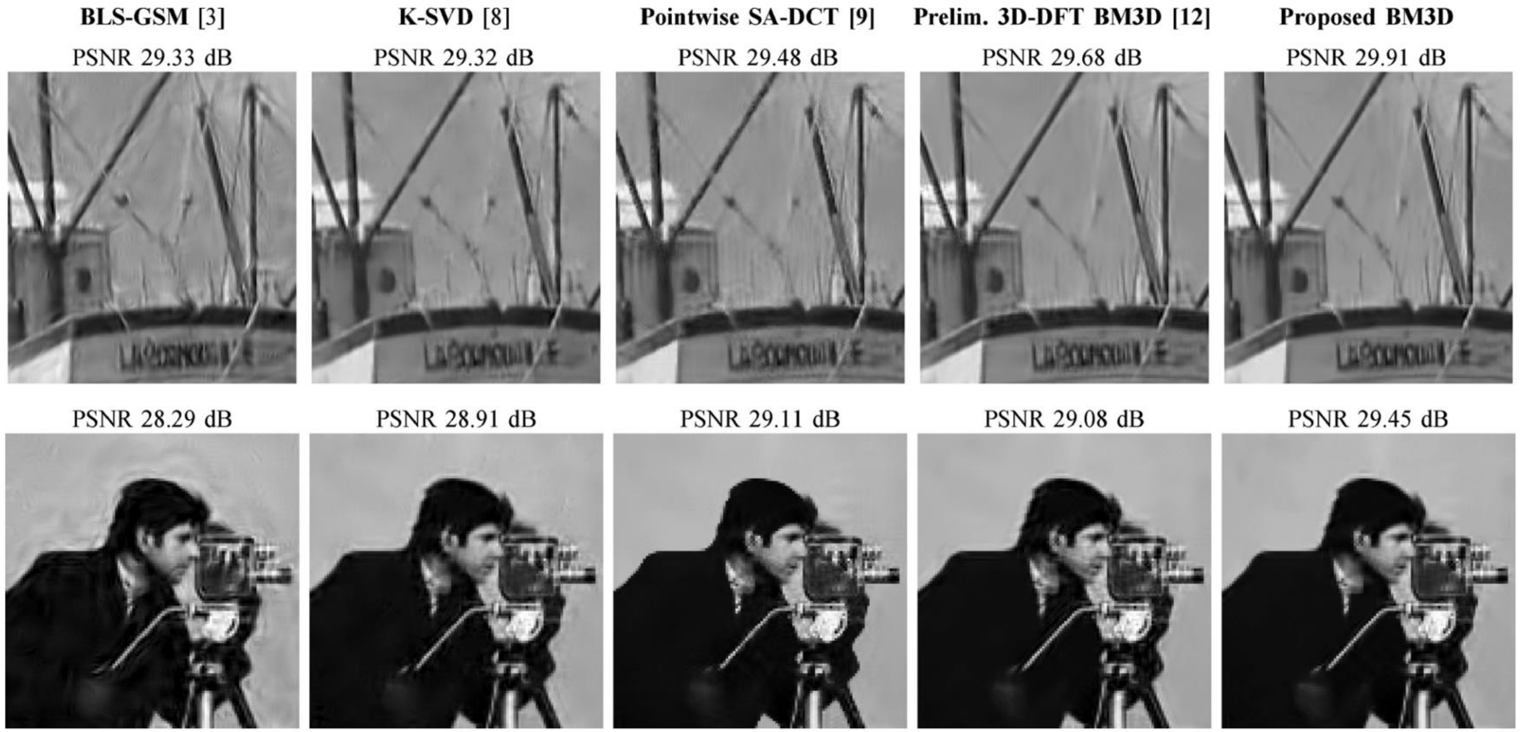


Fig. 2. Simple example of grouping in an artificial image, where for each reference block (with thick borders) there exist perfectly similar ones.

BM3D: Grouping and Collaborative Wiener



$$S_{x_R}^{\text{wie}} = \left\{ x \in X : \frac{\left\| \widehat{Y}_{x_R}^{\text{basic}} - \widehat{Y}_{x}^{\text{basic}} \right\|_2^2}{\left(N_1^{\text{wie}}\right)^2} < \tau_{\text{match}}^{\text{wie}} \right\}. \qquad (7)$$



Wiener Filtering

- Expected value of mean square error

$$e^2 = E\left\{\left(f - \hat{f}\right)^2\right\}$$

- The estimate of f in frequency domain

$$\hat{F}(u,v) = \left[\frac{1}{H(u,v)} \frac{H^*(u,v)}{|H(u,v)|^2 + K}\right] G(u,v)$$

上海科技大学 ShanghaiTech University

Total Variation Denoising

[1] L. Rudin and S. Osher, “Total variation based image restoration with free local constraints,” in Proc. 1st IEEE Int. Conf. Image Processing, vol. 1, 1994, pp. 31–35
[2] L. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” Phys. D, vol. 60, pp. 259–268, 1992.

Total Variation

- The intuitive idea of ***Total Variation*** is based on the premise that signals exhibiting excessive and potentially spurious details possess high total variation (the integral of the image gradient magnitude).
- For a digital signal x_n, we can, for example, define the total variation as

$$V(x) = \sum_n |x_{n+1} - x_n|$$

- To recover the clean signal x_n from y_n is defined by total-variation denoising problem amounts to minimizing the following discrete functional over the signal :

$$E(x,y) + \lambda V(x) = \frac{1}{n}\sum_n (x_n - y_n)^2 + \lambda * \sum_n |x_{n+1} - x_n|$$

- In the original approach, this function is derived using **Euler–Lagrange equation**. Since it is a convex function, techniques from **convex optimization** can be used to minimize it and find the solution.

上海科技大学 ShanghaiTech University

Total Variation

- We now consider 2D signals x, such as images. The total-variation norm proposed by the 1992 article is

$$V(x) = \sum_{i,j} \sqrt{\left|x_{i+1,j} - x_{i,j}\right|^2 + \left|x_{i,j+1} - x_{i,j}\right|^2}$$

- It is isotropic and not differentiable. A variation that is sometimes used, since it may sometimes be easier to minimize, is an anisotropic version

$$V_{aniso}(x) = \sum_{i,j} \left|x_{i+1,j} - x_{i,j}\right| + \left|x_{i,j+1} - x_{i,j}\right|$$

- The standard total-variation denoising problem is still of the form

$$\min_x [E(x, y) + \lambda V(x)]$$

- where $E(x, y)$ is the 2D L2 norm. In contrast to the 1D case, solving this denoising is non-trivial. The authors in [1] solved this by s the primal dual method [2].

[1] L. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” Phys. D, vol. 60, pp. 259–268, 1992.
[2] Chambolle, A. (2004). "An algorithm for total variation minimization and applications". Journal of Mathematical Imaging and Vision. **20** 89–97.

Marcov Random Fields

[1] Boykov, Y. and Kolmogorov, V. (2004) An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Trans PAMI 26(9):1124-1137.
[2] Boykov, Y., Veksler, O., and Zabih, R. (2001) Fast approximate energy minimization via graph cuts. IEEE Trans. PAMI, 23(11):1222-1239.

Dictionary-learning Denoising

reference :
[1] C. Taswell, "The what, how, and why of wavelet shrinkage denoising," in Computing in Science & Engineering, vol. 2, no. 3, pp. 12-19, May-June 2000, doi: 10.1109/5992.841791.
[2] Meyer Scetbon; Michael Elad; Peyman Milanfar, et al. Deep K-SVD Denoising. IEEE Transactions on image processing, 2021, 16(8): 5944-5955.
[3] M. Aharon, M. Elad, and A. Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans. on Signal Processing, 54(11):4311, 2006

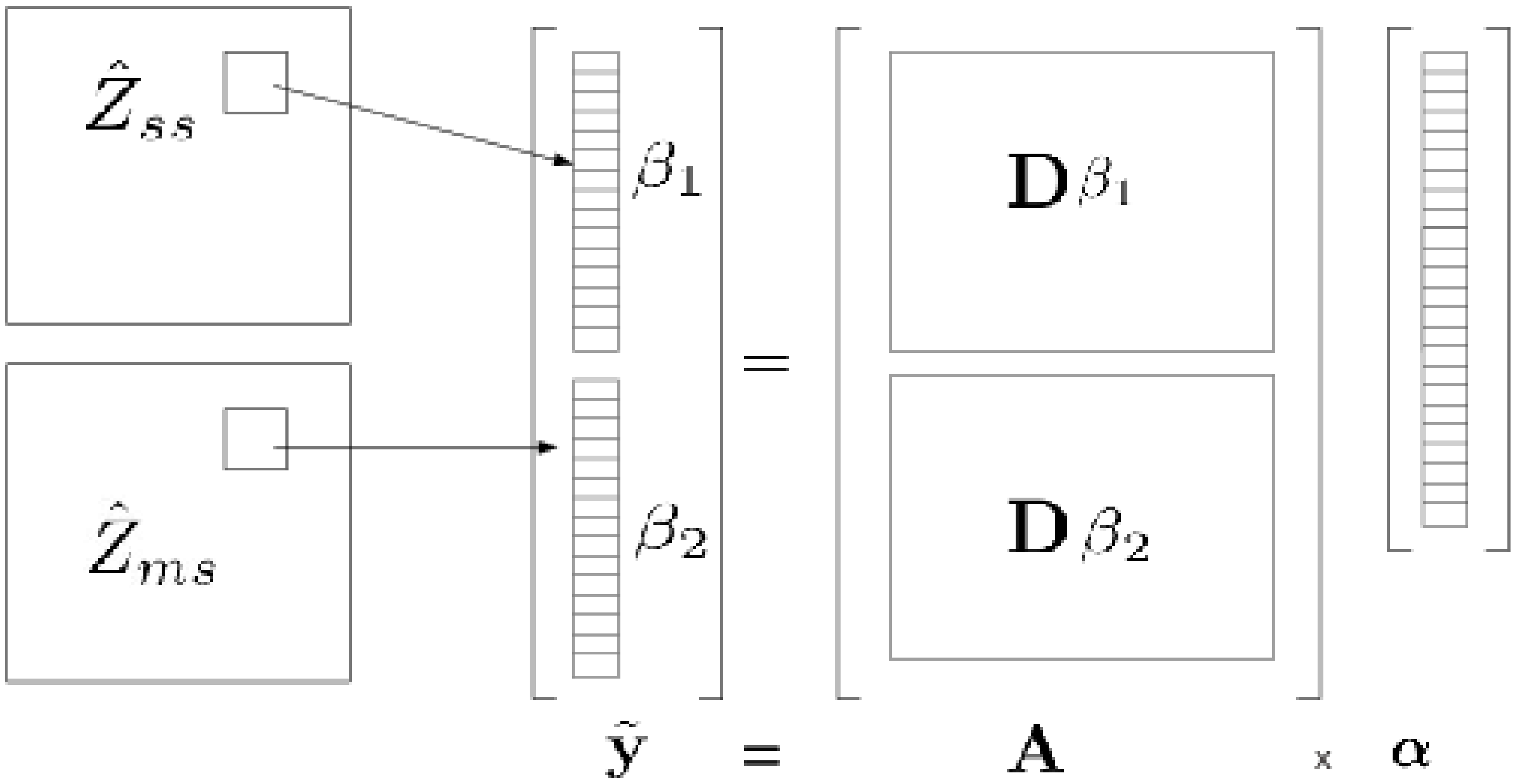
上海科技大学
ShanghaiTech University

Overview

- The basic idea behind wavelet denoising, or any linear transform thresholding, is that the transform leads to a sparse representation for many real-world signals and images.
- What this means is that the wavelet transform concentrates signal and image features in a few large-magnitude wavelet coefficients.
- Wavelet coefficients which are small in value are typically noise and you can "shrink" those coefficients or remove them without affecting the signal or image quality.
- After you threshold the coefficients, you reconstruct the data using the inverse transform

Dictionary Selection

- What D to use?
- A fixed overcomplete set of basis:
  - Steerable wavelet
  - Contourlet
  - DCT Basis
- Data Adaptive Dictionary – learn from data


$\hat{Z}_{ss}$
$\hat{Z}_{ms}$
β_1
β_2
=
$\mathbf{D}\beta_1$
$\mathbf{D}\beta_2$
$\tilde{\mathbf{y}}$ = $\mathbf{A}$ x $\boldsymbol{\alpha}$

Denoising Framework

- A cost function for： $\boldsymbol{Z = Y + n}$

$$\boldsymbol{arg\,min_Y \|Y - Z\|_2^2 + \lambda\|\cdot\|_p}$$

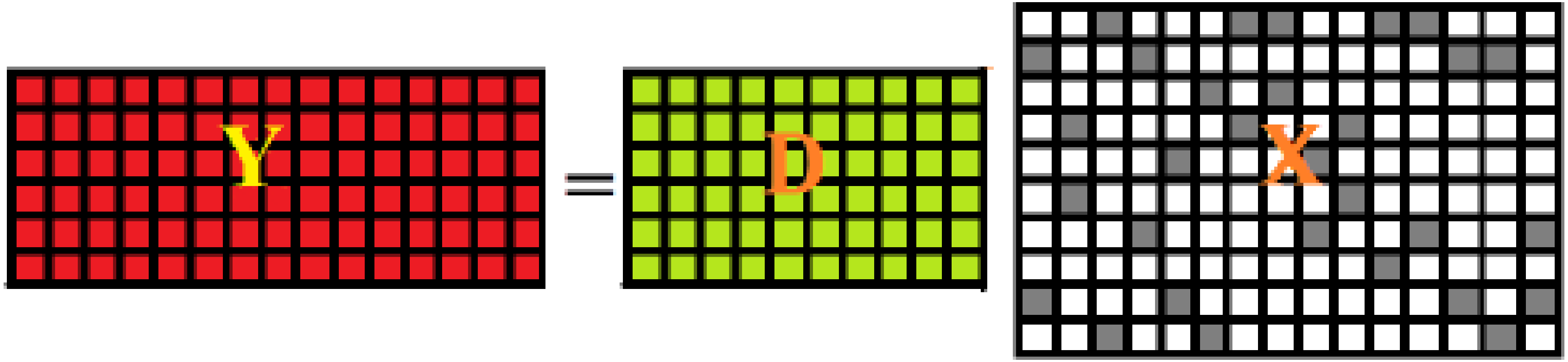
Prior Term

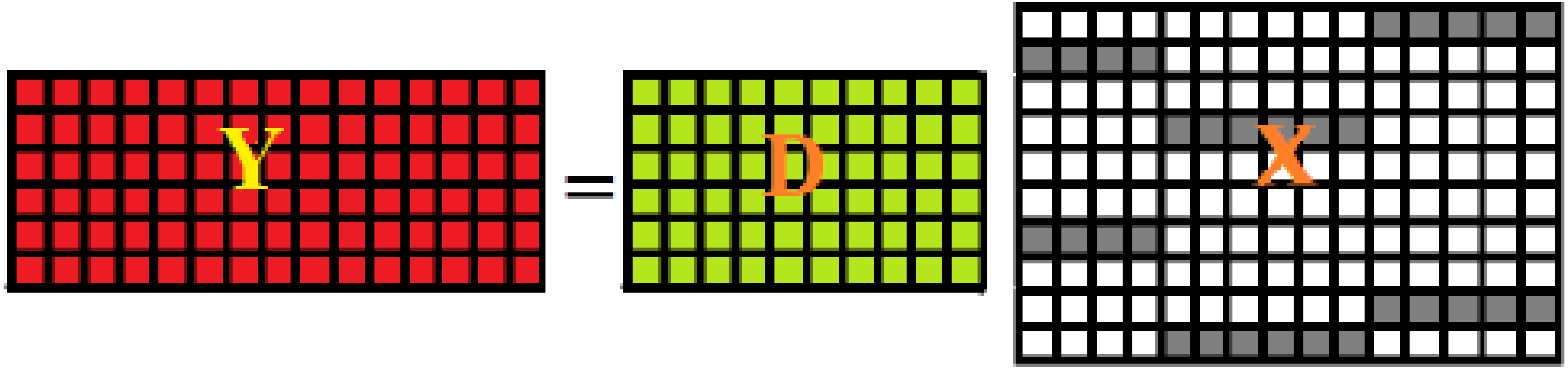
- Solve for

$$\boldsymbol{\hat{X}, \hat{Y} = arg\,min_{X,Y} \|Y - Z\|_2^2 + \lambda\|DX - Y\|_F^2 + \sum_i \mu_i \|x_i\|_0}$$

Dictionary Learning

Sparse Representation


Y
=
D
X


Y
=
D
X

Take home massage

- Image denoising is to recover signals hidden in a noisy background. Since noise is a statistical fluctuation governed by quantum mechanics, denoising is generally achieved by an mean/averaging operation.
- The key idea behind early denoising methods is to avoid smoothing on image edges.
- https://piazza.com/shanghaitech.edu.cn/spring2025/cs270b