MULTIFRAME RAW-DATA DENOISING BASED ON BLOCK-MATCHING AND 3-D FILTERING FOR LOW-LIGHT IMAGING AND STABILIZATION

LNLA 2008 *- The 2008 International Workshop on*
Local and Non-Local Approximation in Image Processing,
23, 24 August, Lausanne, Switzerland

Giacomo BORACCHI
Politecnico di Milano

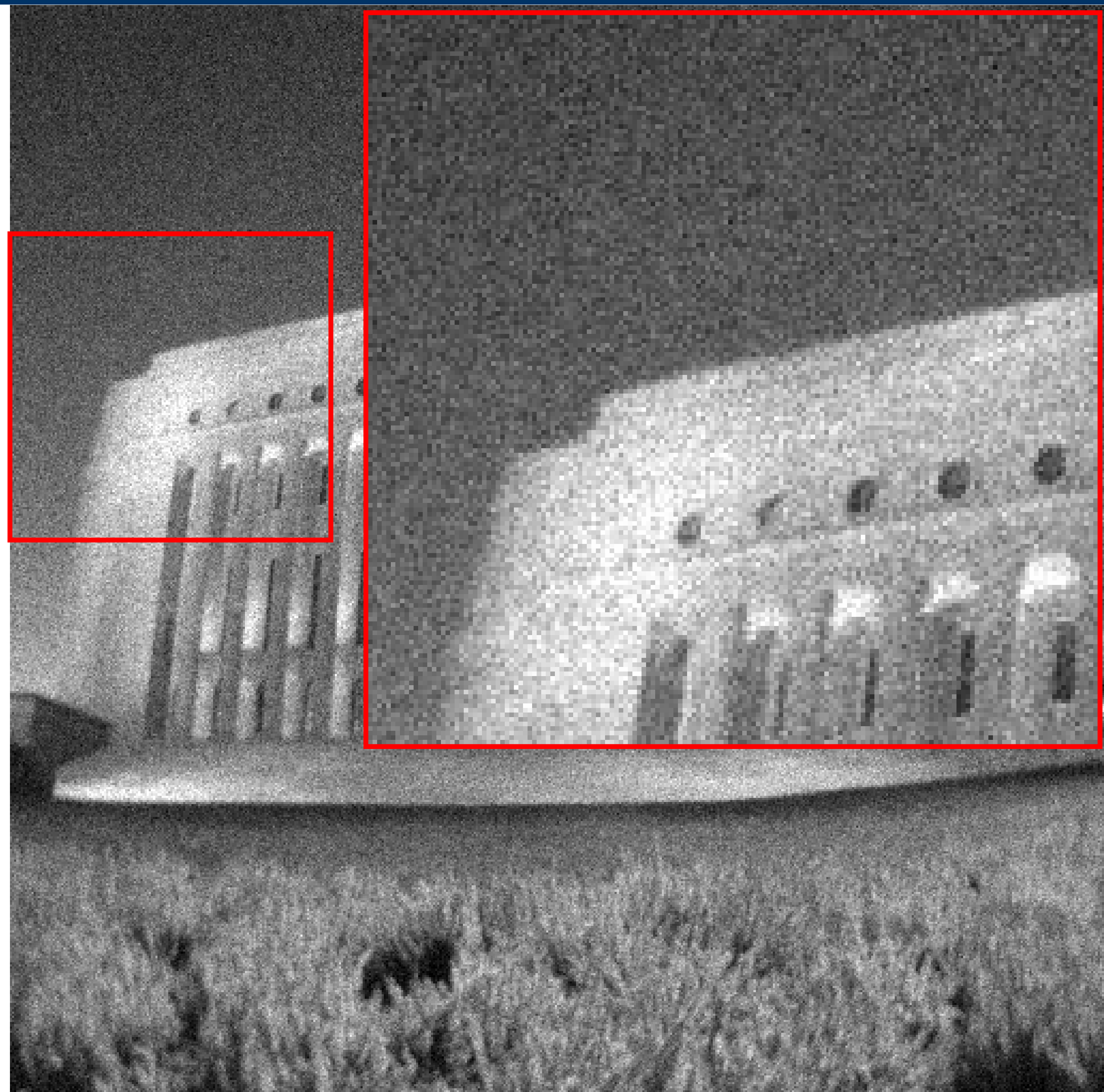
Alessandro FOI
Tampere University of Technology

Introduction

- Taking satisfactory pictures at low-light conditions is challenging.
- Pictures acquired with a *short* exposure-time have low SNR and are **very noisy** because of a high gain (ISO number).

Introduction

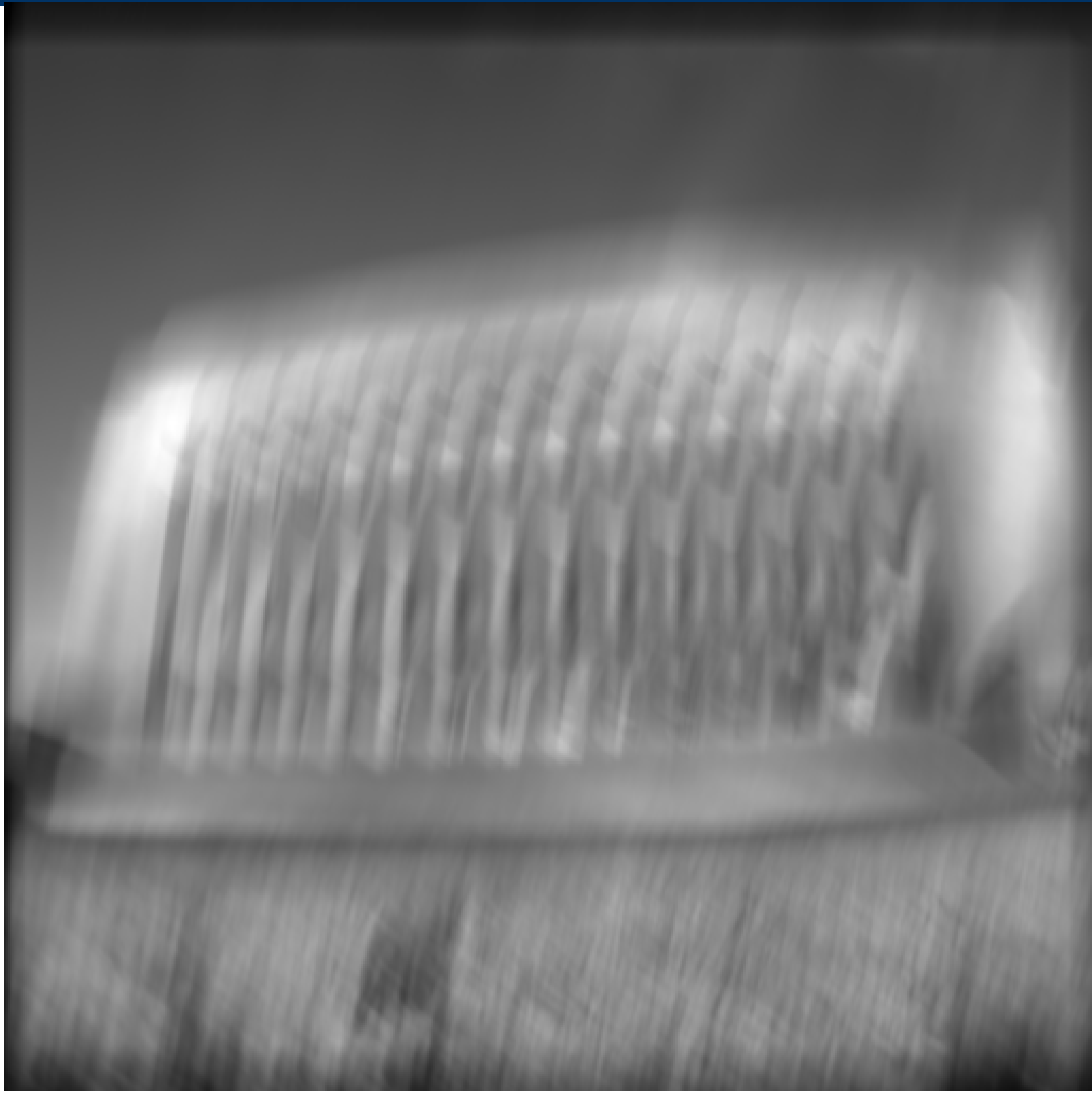
Introduction



Introduction

- Taking satisfactory pictures at low-light conditions is challenging.
- Pictures acquired with a *short* exposure-time have low SNR and are **very noisy** because of a high gain (ISO number).
- Typically the exposure time is increased in order to improve the SNR of the acquired image.
- But this also increases the risk of blur, because of movements occurring in the extended exposure.

Introduction



Solutions

- A variety of solutions:
  - Lenses Stabilization
  - Different Acquisition Strategies

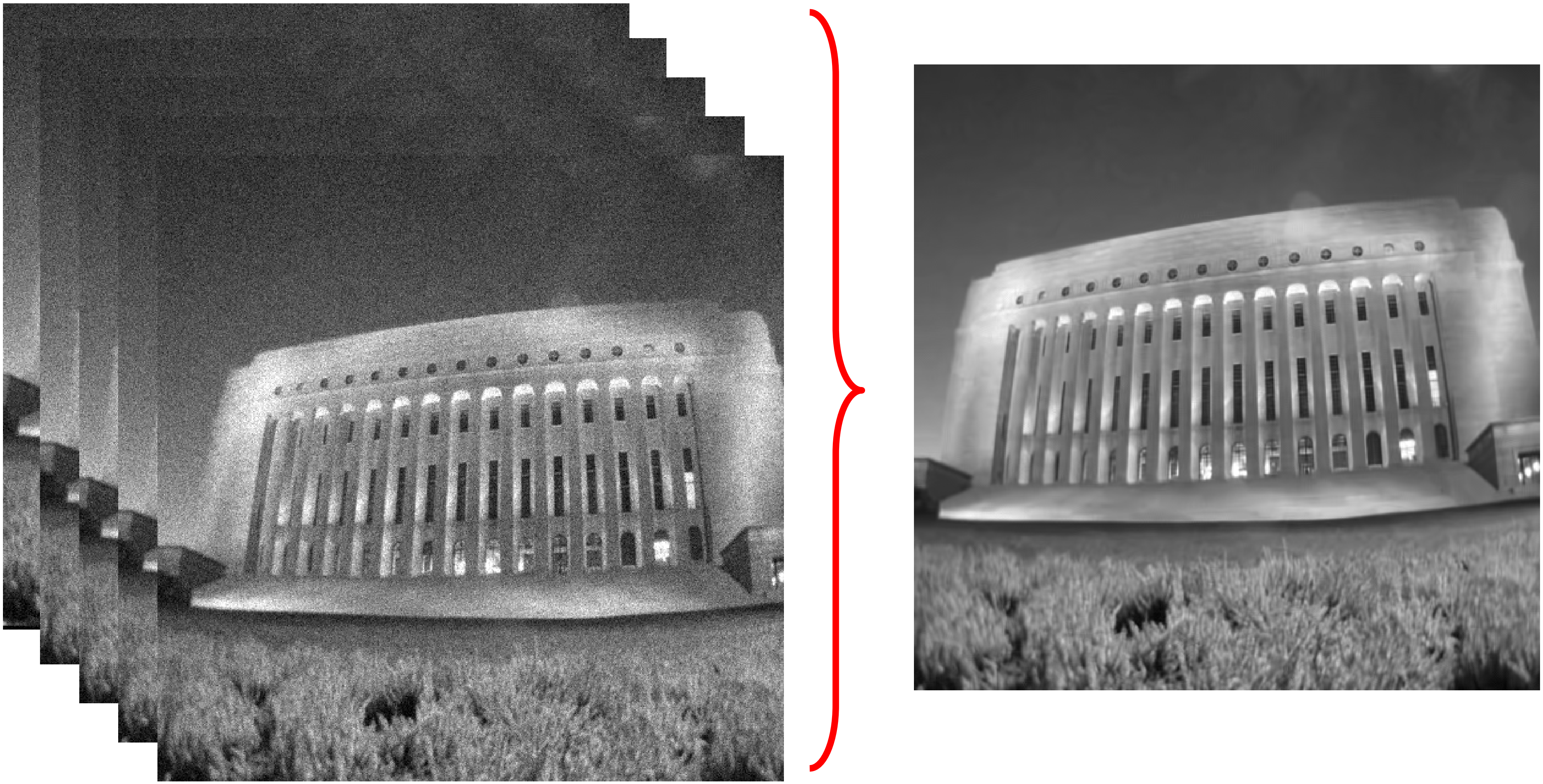
- In particular **[Tico06]** and **[Yuan07]** proposed two methods that use differently exposed images
  - one with a **long** exposure time (blurred but with negligible noise)
  - one with a **short** exposure time (noisy but with negligible blur)

- The noisy image is used to estimate the blur PSF allowing to restore the blurred image (deblurring)

[Tico06] Tico, M., "Estimation of motion blur point spread function from differently exposed image frames," Proc. 14th Eur. Signal Process. Conf., EUSIPCO 2006, Florence, Italy, September 2006
[Yuan07] Yuan, L., J. Sun, L. Quan, and H.-Y. Shum, "Image deblurring with blurred/noisy image pairs," ACM Trans. Graph., vol. 26, no. 3, July 2007

Alternative Solution

- Acquire a sequence of short exposure images (frames) and jointly denoise them, using a video denoising algorithm



Alternative Solution

- Acquire a sequence of short exposure images (frames) and jointly denoise them, using a video denoising algorithm

- Issues that has to be considered
  - **Movements** (camera viewpoint or scene objects) between frames
  - **Noise**
  - **Clipping**

  raw-data processing

Alternative Solution

- Acquire a sequence of short exposure images (frames) and jointly denoise them, using a video denoising algorithm

- Issues that has to be considered
  - **Movements** (camera viewpoint or scene objects) between frames
  - **Noise**
  - **Clipping**

  raw-data processing

- The proposed solution combines
  - An algorithm for estimating noise in clipped raw data
  - Homomorphic transformations
  - Video Denoising Algorithm (V-BM3D) for AWGN

The Observation Model

- The observation is a sequence of N raw-data frames $\{\tilde{z}_i\}_{i=1}^{N}$ modeled as the **noisy** and **clipped** images

$$\tilde{z}_i(x) = \max\{0, \min\{z_i(x), 1\}\}, \qquad x \in X \subset \mathbb{Z}^2,$$

where

$$z_i(x) = y_i(x) + \sigma(y_i(x))\,\xi_i(x),$$

$y_i : X \to Y \subseteq \mathbb{R}$ is an *original* frame

$\sigma(y_i(x))\,\xi_i(x)$ is an zero-mean random error

$\xi_i(\cdot) \sim \mathcal{N}(0,1)$

$\sigma : \mathbb{R} \to \mathbb{R}^{+}$

Raw Data

- For raw data

$$\sigma^2(y_i(x)) = a y_i(x) + b,$$

a and *b* depend on the sensor hardware characteristics and on the acquisition settings only.

- The noisy clipped observation can be expressed as

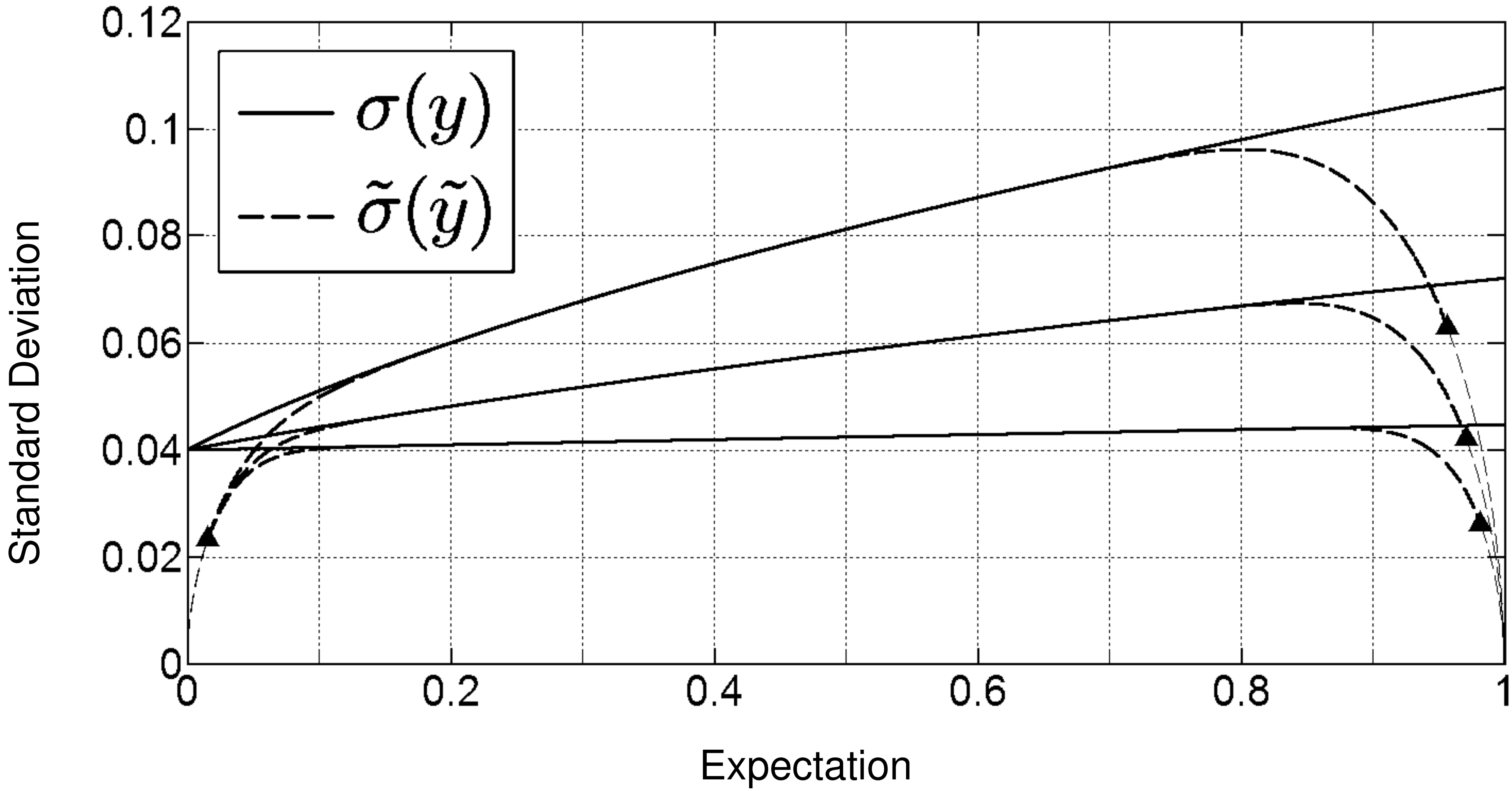
$$\tilde{z}_i(x) = \tilde{y}_i(x) + \tilde{\sigma}(\tilde{y}_i(x))\,\tilde{\xi}_i(x)\,, \qquad x \in X \subset \mathbb{Z}$$

where

$$\tilde{y}_i(x) = E\{\tilde{z}_i(x)\} \in [0,1]\,,$$
$$\tilde{\sigma}(\tilde{y}_i(x)) = \mathrm{std}\{\tilde{z}_i(x)\} \geq 0.$$

Expectation vs. Standard Deviation curves

- Each curve is determined by *a* and *b* only.

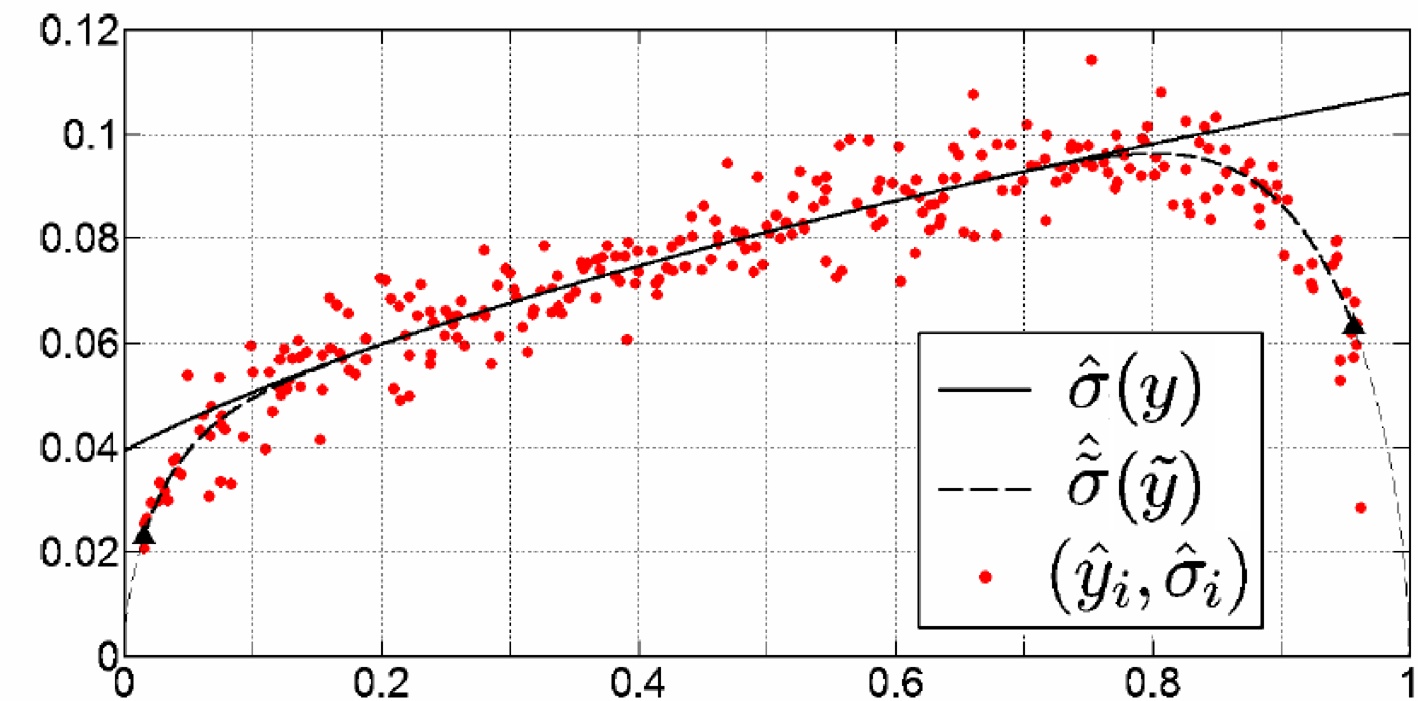


Algorithm Outline

- Noise Parameters Estimation
- Noise Variance Stabilization
- Video Denoising
- Debiasing and Inversion of Noise Variance Stabilizing Transformation
- Declipping

Noise Estimation

- The parameters *a* and *b* of the noise can estimated from a single noisy and clipped image using the algorithm presented in **[Foi08a]**
- This algorithm can be used on a single frame of the original sequence as *a* and *b* are constant.

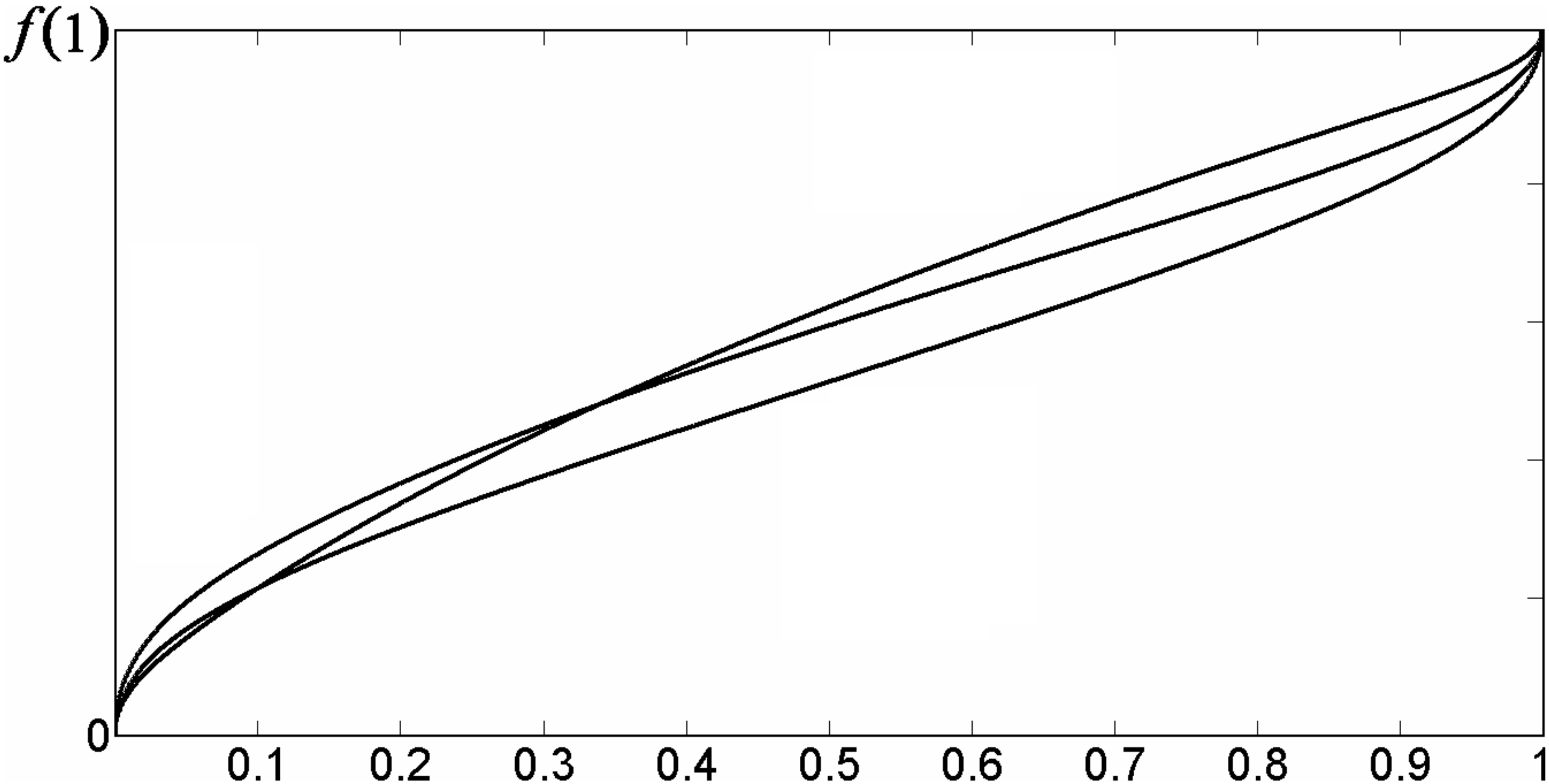


[Foi08a] Foi,A.,M.Trimeche,V.Katkovnik, and K.Egiazarian, “Practical Poissonian-Gaussian noise modeling and Þtting for single image raw-data”, to appear in *IEEE Trans. Image Process*.

Noise Variance Stabilization

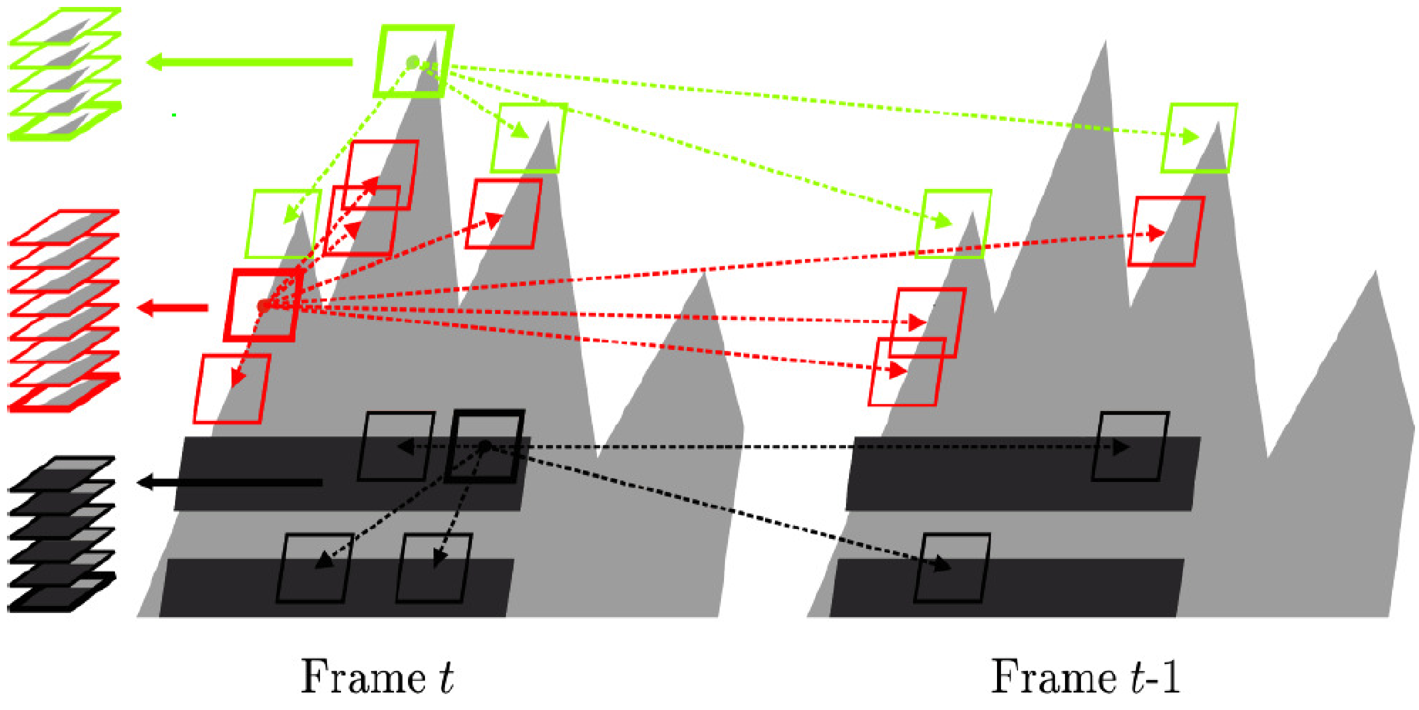
- Each frame is pixel-wise transformed in the following way

$$f(t) = \int_{t_0}^{t} \frac{c}{\tilde{\sigma}(s)} ds, \qquad t, t_0 \in [0, 1]$$



BM3D Denoising

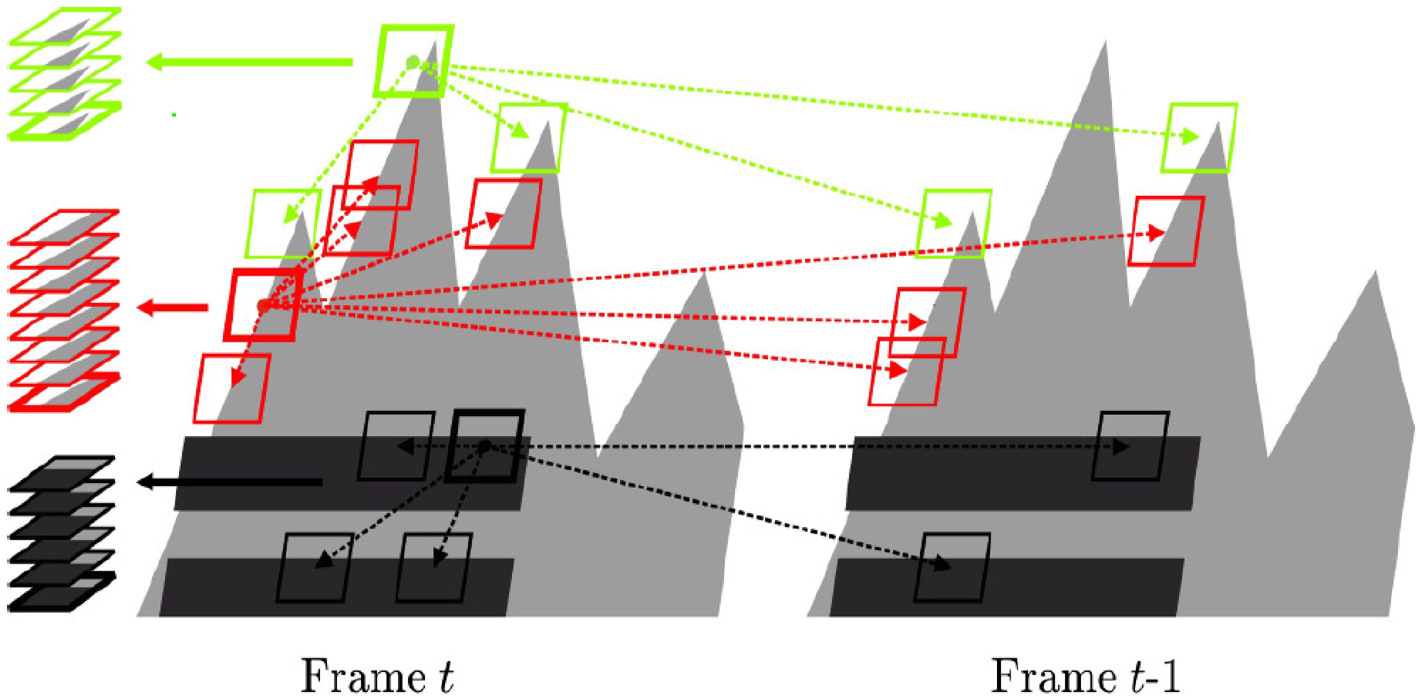
- Block Matching 3D (BM3D) **[Dabov07],** is a **nonlocal** method that filters the image in a block-wise manner as



[Dabov07] Dabov, K., A. Foi, and K. Egiazarian, "Video denoising by sparse 3D transform-domain collaborative filtering", *Proc. 15th Eur. Signal Process. Conf., EUSIPCO 2007*, Poznan, September 2007

BM3D Denoising

- Block Matching 3D (BM3D) **[Dabov07],** is a **nonlocal** method that filters the image in a block-wise manner as

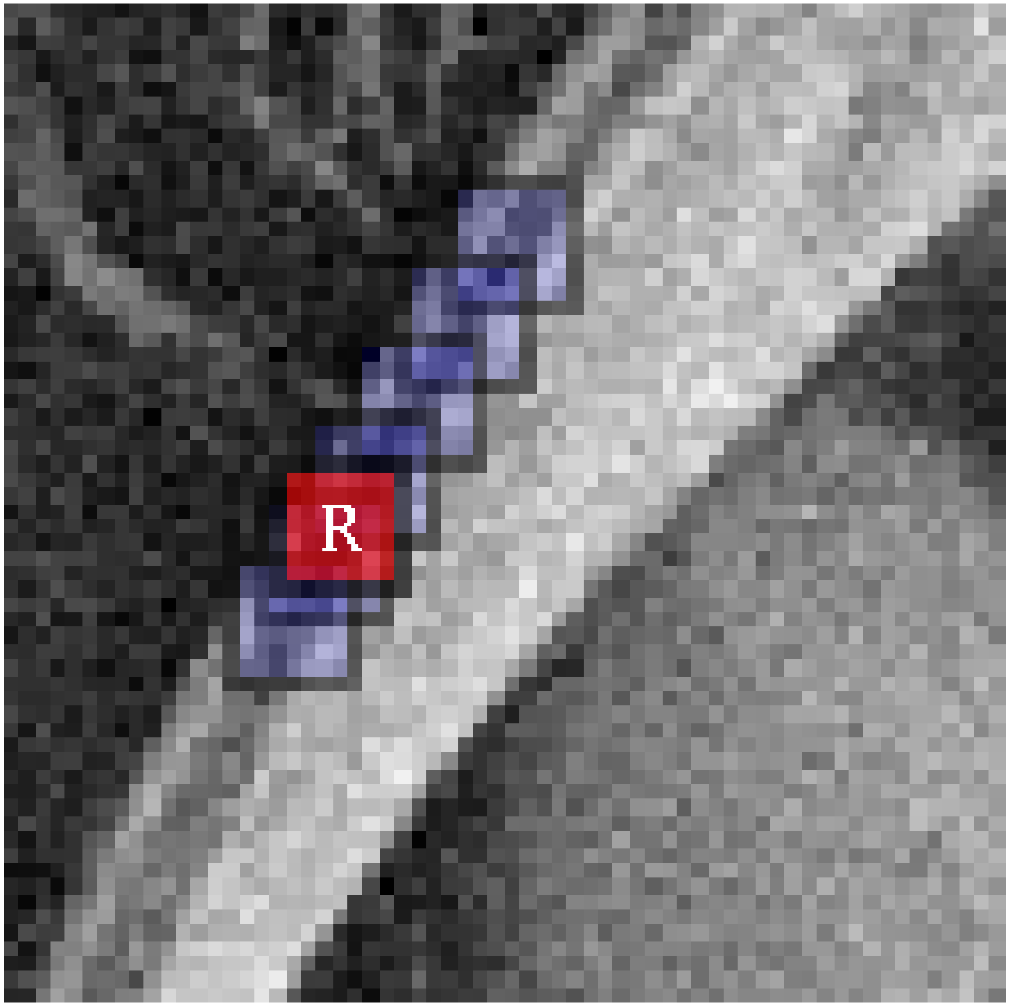


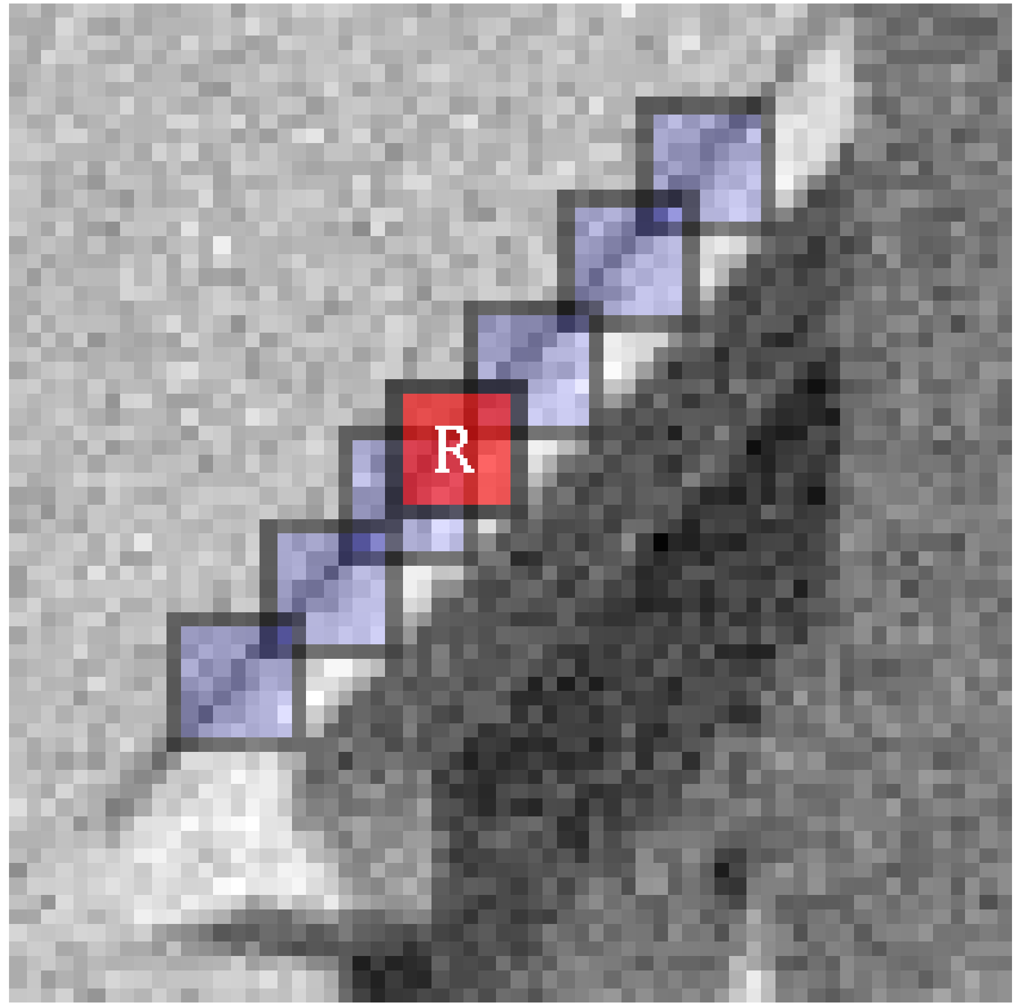
1. **Grouping**: search for similar blocks and stack them together in a 3D array

[Dabov07] Dabov, K., A. Foi, and K. Egiazarian, "Video denoising by sparse 3D transform-domain collaborative filtering", *Proc. 15th Eur. Signal Process. Conf., EUSIPCO 2007*, Poznan, September 2007

BM3D Denoising

- Example of Grouping





BM3D Denoising – Grouping

- Grouping is performed by matching

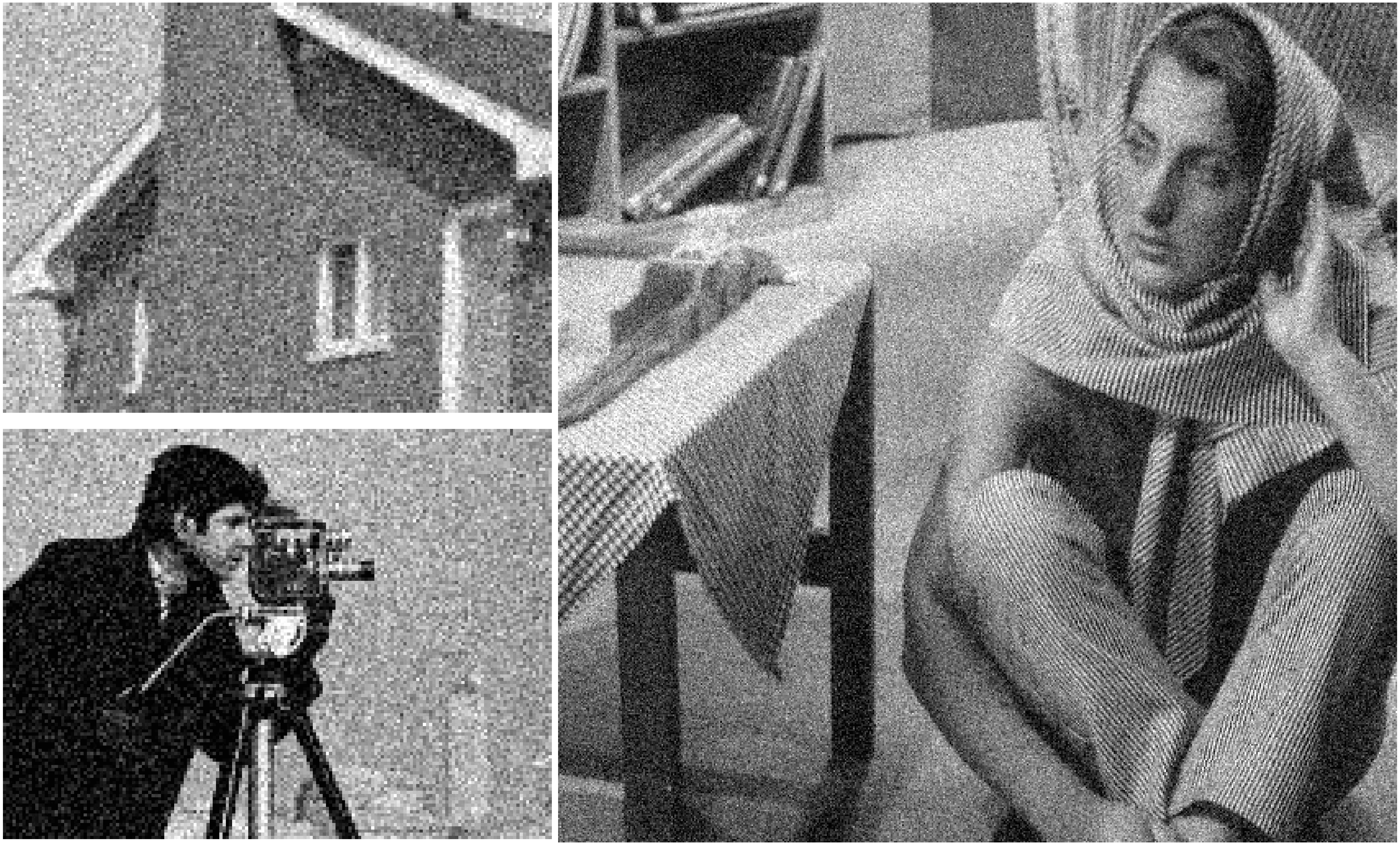
- We measure the distance between blocks and the reference block and we group those having minimum distance

$$d^{\text{noisy}}\left(Z_{x_R}, Z_x\right) = \frac{\left\|Z_{x_R} - Z_x\right\|_2^2}{\left(N_1^{\text{ht}}\right)^2}.$$

- Distance is measured in transform domain, performing a *preliminary* denoising

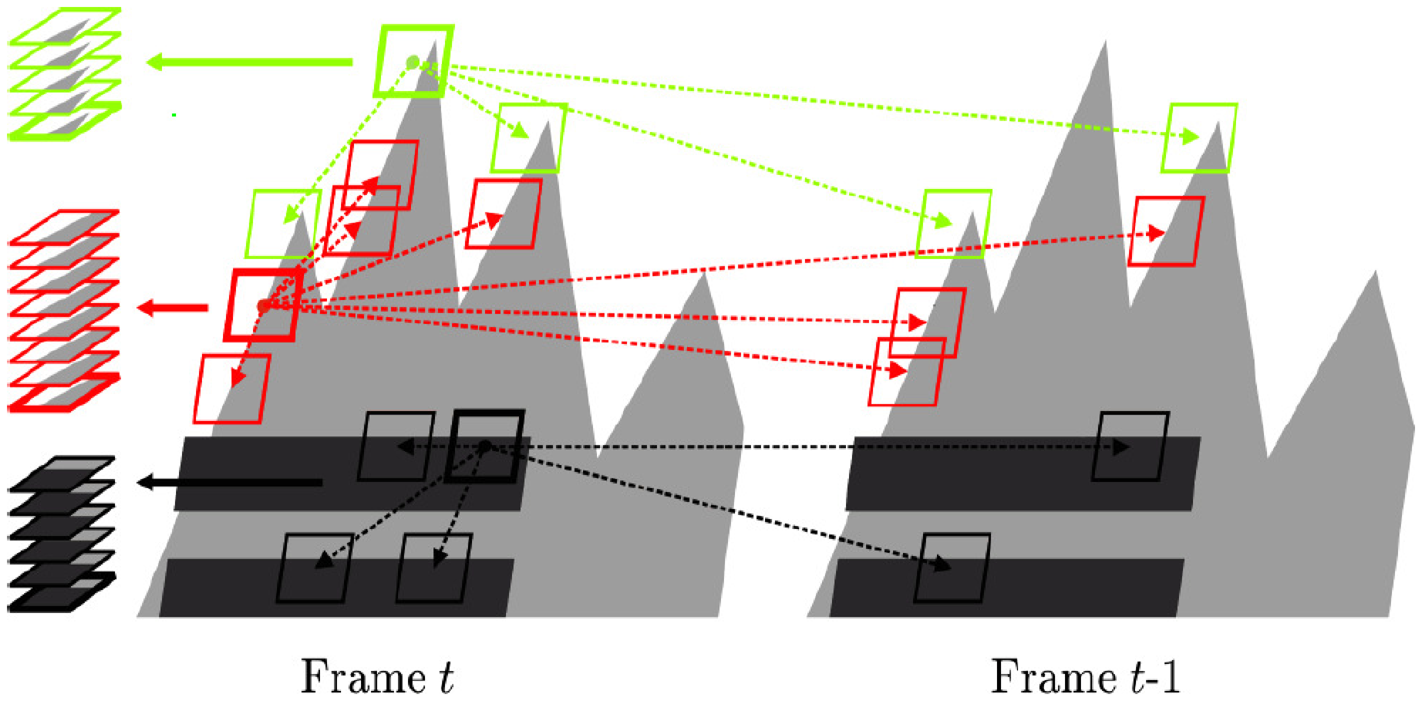
$$d\left(Z_{x_R}, Z_x\right) = \frac{\left\|\Upsilon'\left(\mathcal{T}_{2\text{D}}^{\text{ht}}\left(Z_{x_R}\right)\right) - \Upsilon'\left(\mathcal{T}_{2\text{D}}^{\text{ht}}\left(Z_x\right)\right)\right\|_2^2}{\left(N_1^{\text{ht}}\right)^2},$$

BM3D Denoising



BM3D Denoising

- Block Matching 3D (BM3D) **[Dabov07]**, is a **spatiotemporal nonlocal** method that filters the image in a block-wise manner as



2. **Collaborative Filtering**: filter the groups by 3D transform-domain shrinkage, obtaining **individual estimates for all grouped blocks**

[Dabov07] Dabov, K., A. Foi, and K. Egiazarian, “Video denoising by sparse 3D transform-domain collaborative filtering”, *Proc. 15th Eur. Signal Process. Conf., EUSIPCO 2007*, Poznan, September 2007

BM3D Denoising – Collaborative Filtering

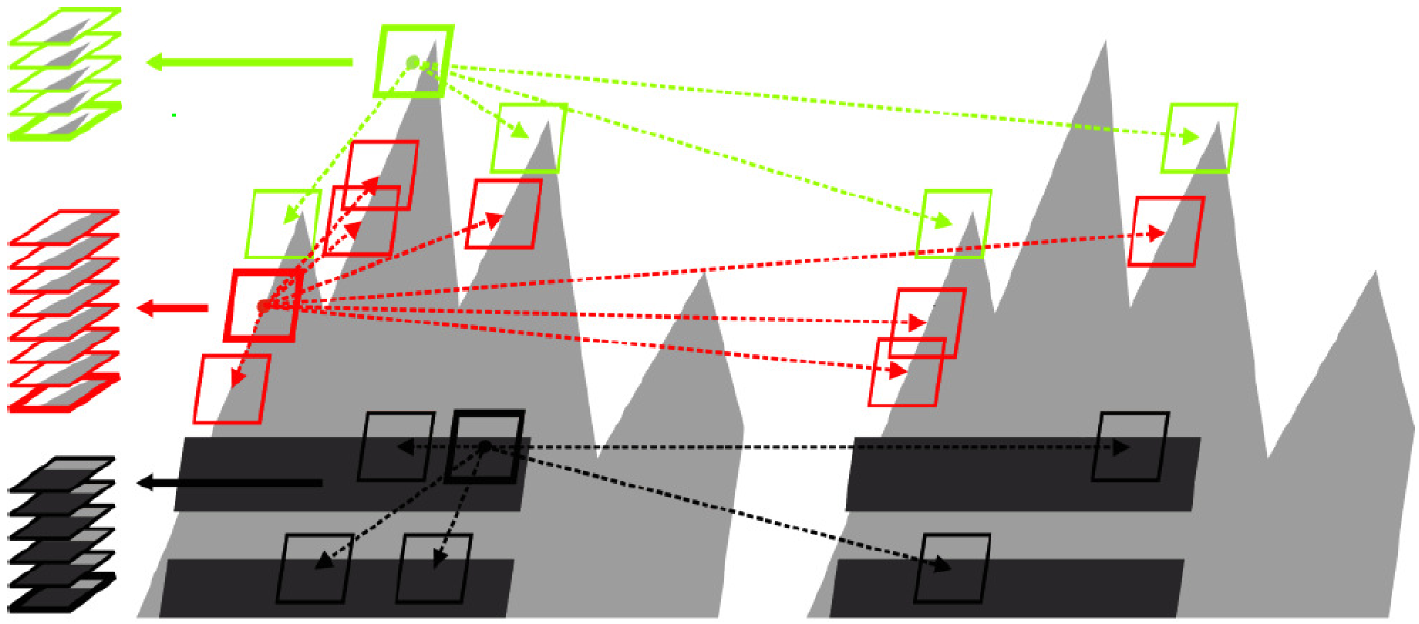
- The collaborative filtering is realized hard-thresholding in 3D transform domain
  - apply a 3D linear transform on each group.
  - Shrink the transform coefficients of each block to attenuate noise.
  - Invert the linear transform to produce estimates for each fragments in the group.

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

- We thus obtain an overcomplete representation of the image

BM3D Denoising

- Block Matching 3D (BM3D) **[Dabov07],** is a **nonlocal** method that filters the image in a block-wise manner as



3. **Aggregation**: restored frames are obtained by weighted averages of the filtered blocks when they are overlapping

[Dabov07] Dabov, K., A. Foi, and K. Egiazarian, "Video denoising by sparse 3D transform-domain collaborative filtering", *Proc. 15th Eur. Signal Process. Conf., EUSIPCO 2007*, Poznan, September 2007

BM3D – Aggregation

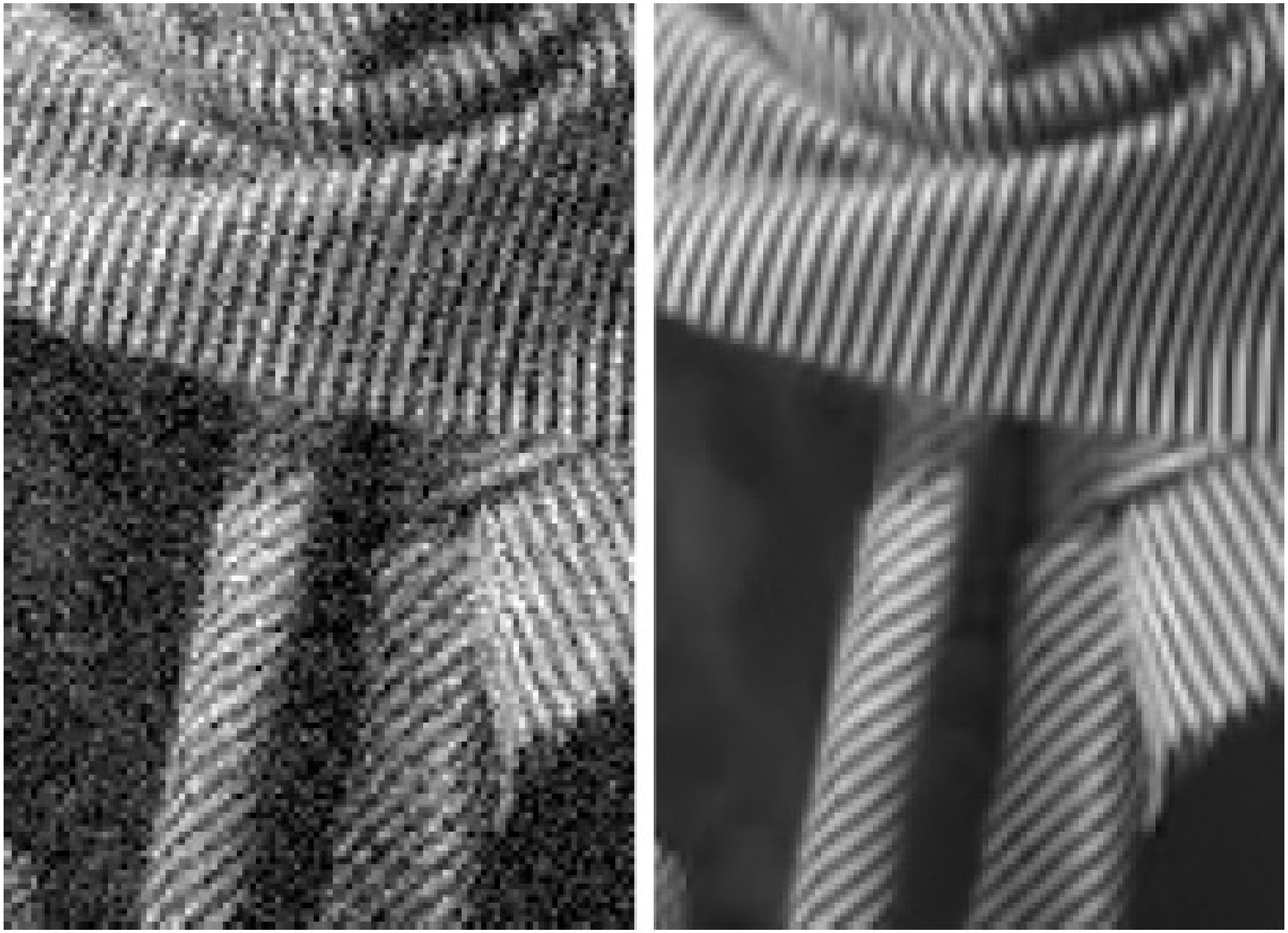
- The aggregation is performed with weighted averages of the pixels where there are overlapping blocks estimates.

$$\widehat{y}^{\text{basic}}(x) = \frac{\sum_{x_R \in X} \sum_{x_m \in S^{\text{ht}}_{x_R}} w^{\text{ht}}_{x_R} \widehat{Y}^{\text{ht},x_R}_{x_m}(x)}{\sum_{x_R \in X} \sum_{x_m \in S^{\text{ht}}_{x_R}} w^{\text{ht}}_{x_R} \chi_{x_m}(x)}, \forall x \in X,$$

- In such a way each grouped fragment collaborates for filtering the others and vice versa

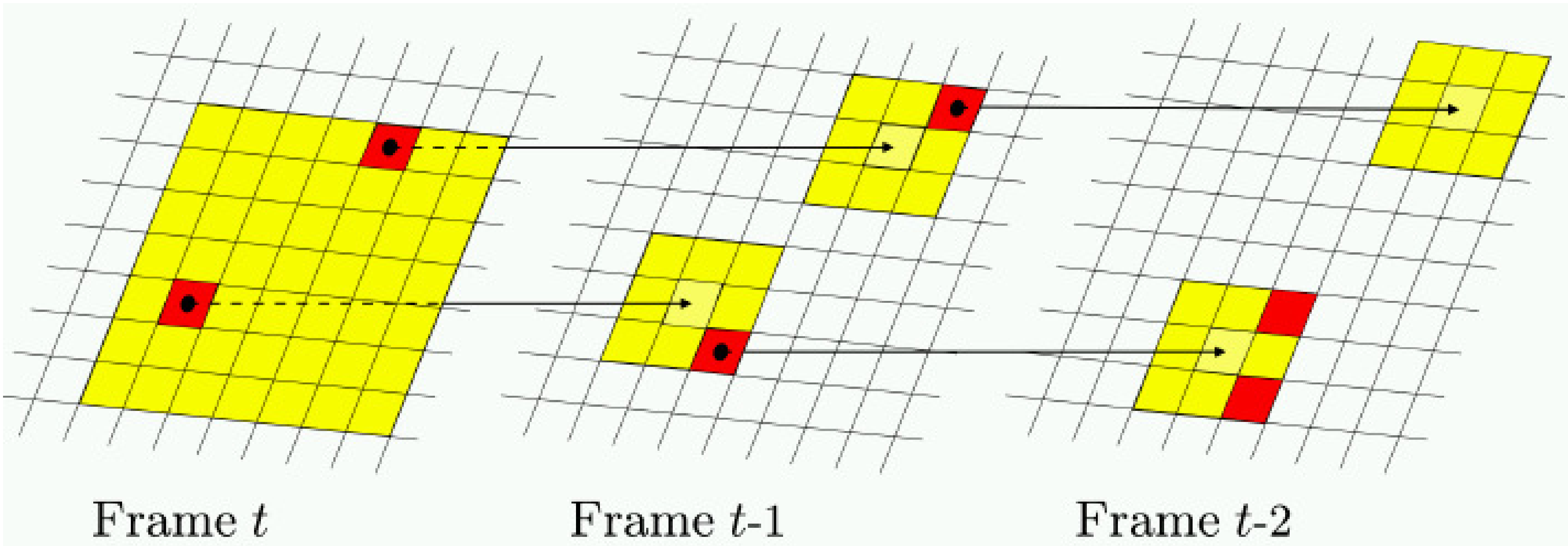
BM3D – Aggregation

BM3D – Aggregation



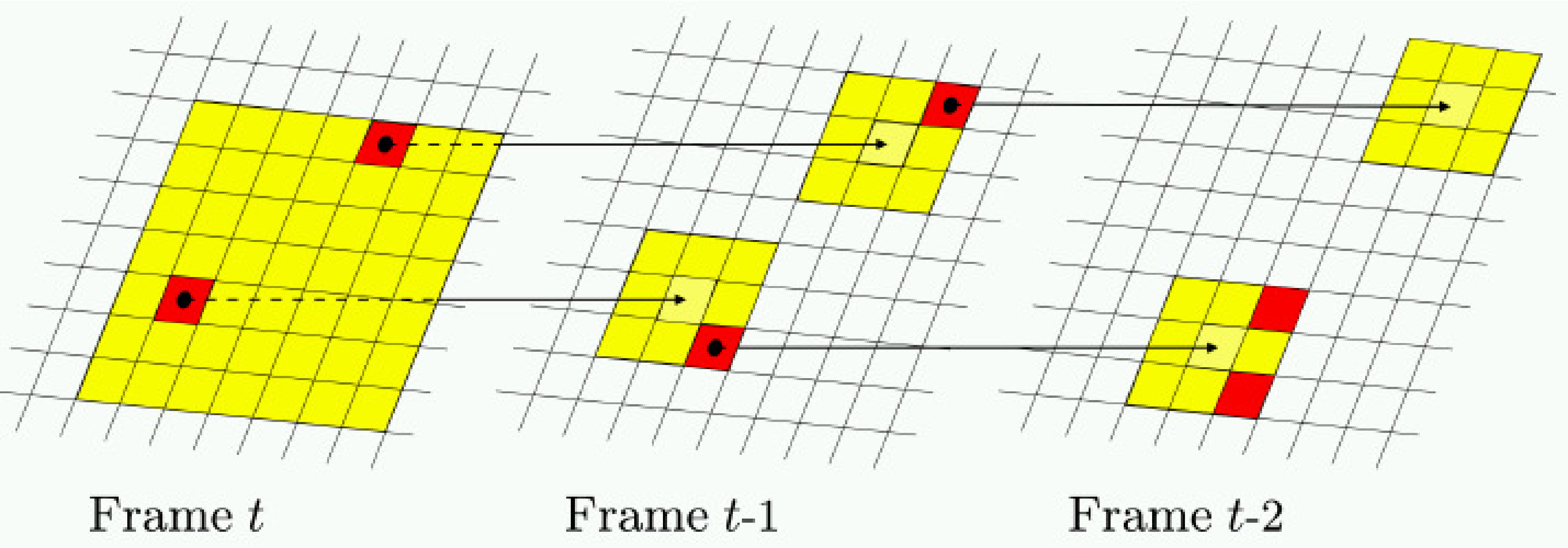
V-BM3D: Video BM3D

- The non local search spans in a both the frame and time dimensions of each block



V-BM3D: Video BM3D

- The non local search spans in a both the frame and time dimensions of each block



- Grouping: instead of using full image or fixed-size neighborood search, we use data-adaptive neighborhoods, using the *predictive search*.

Variance Stabilizing Transform Inversion

- Since *f* is nonlinear there is estimation bias:

$$\mathbf{D}(f(\tilde{z}_i(x))) \approx E\{f(\tilde{z}_i(x))\} \neq f(E\{\tilde{z}_i(x)\})$$

being **D** the V-BM3D denoising operator

- Debiasing **[Foi08b]**

$$h^{-1}(\mathbf{D}(f(\tilde{z}_i(x)))) \approx f(E\{\tilde{z}_i(x)\})$$

- and then inversion

$$f^{-1}\big(h^{-1}(\mathbf{D}(f(\tilde{z}_i(x))))\big) \approx E\{\tilde{z}_i(x)\}$$

[Foi08b] Foi, A., "Practical denoising of clipped or overexposed noisy images", Proc. 16th European Signal Process. Conf., EUSIPCO 2008, Lausanne, Switzerland, August 2008

Declipping

- We obtain an estimate of clipped data

$$E\{\tilde{z}_i\} = \tilde{y}_i \neq E\{z_i\}$$

- To obtain an estimate the **original** signal we need to invert the bias due to clipping with the transform **[Foi08b]**

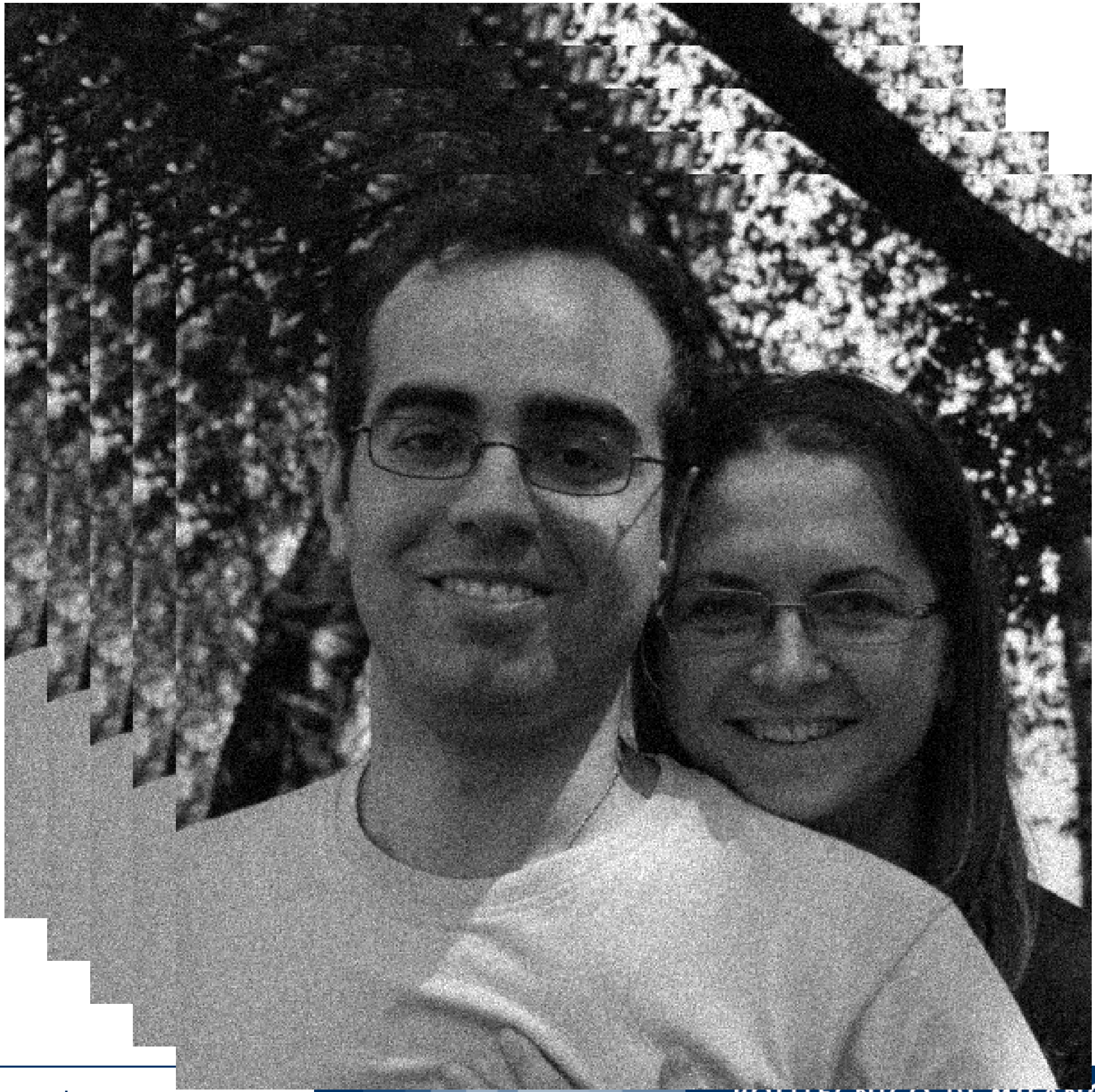
$$\mathcal{C} : E\{\tilde{z}_i\} \longmapsto E\{z_i\}$$

note that $\mathcal{C} : [0,1] \longrightarrow Y$, where *Y is* the range of the original image.

- Thus the range of the restored image is increased w.r.t. the observation range

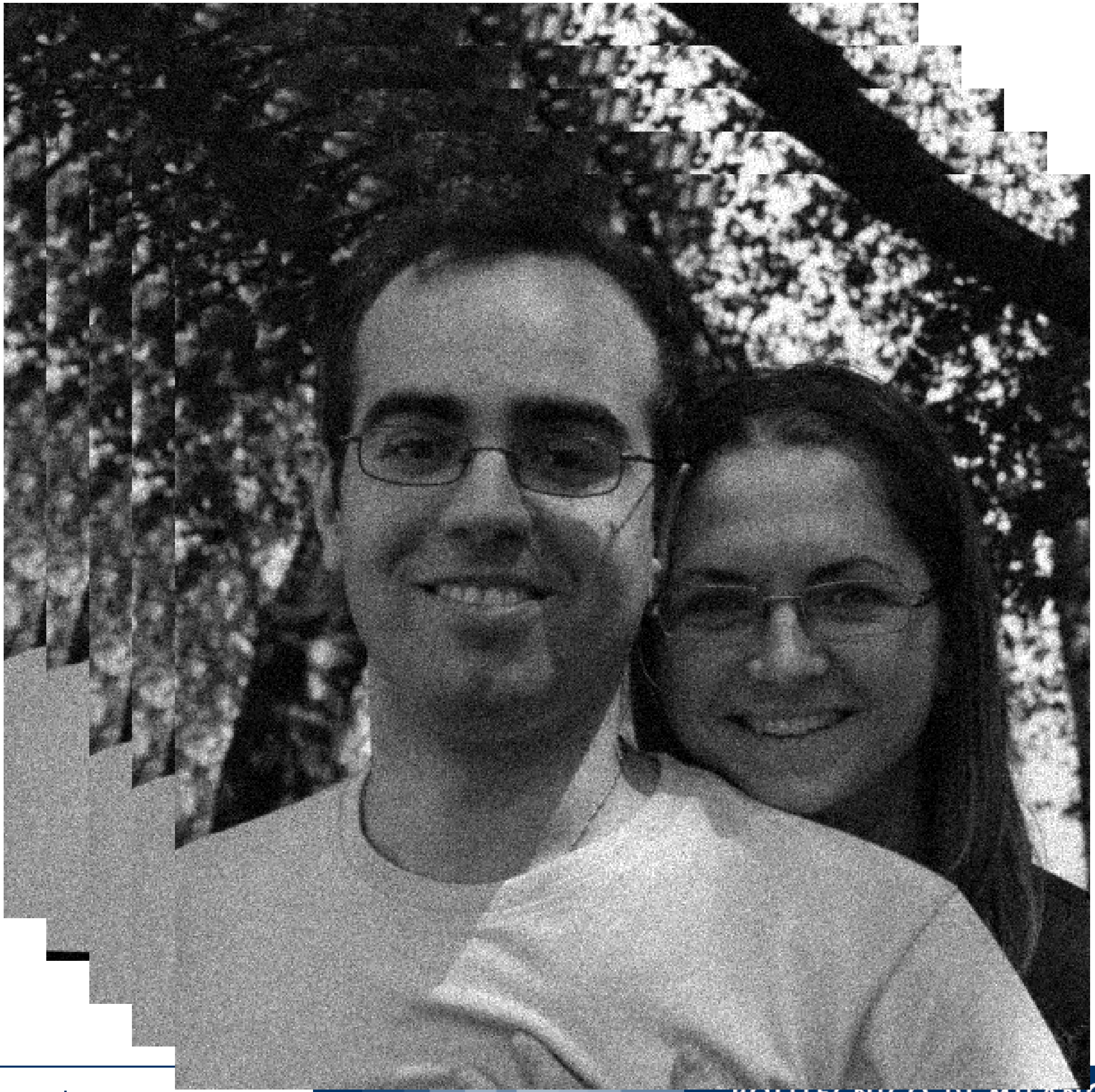
Synthetic Experiments- *Luca & Tania* Sequences

- *Fixed* Sequence



Synthetic Experiments- *Luca & Tania* Sequences

- *Shaked* Sequence

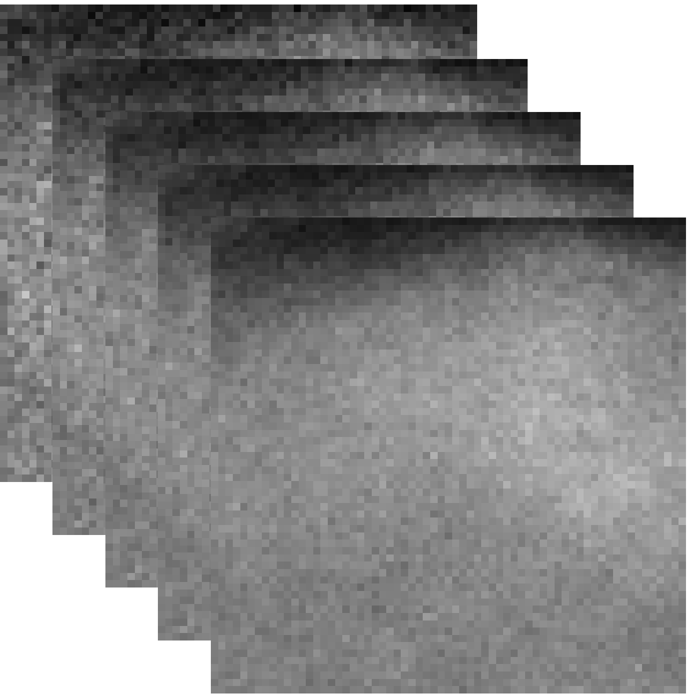


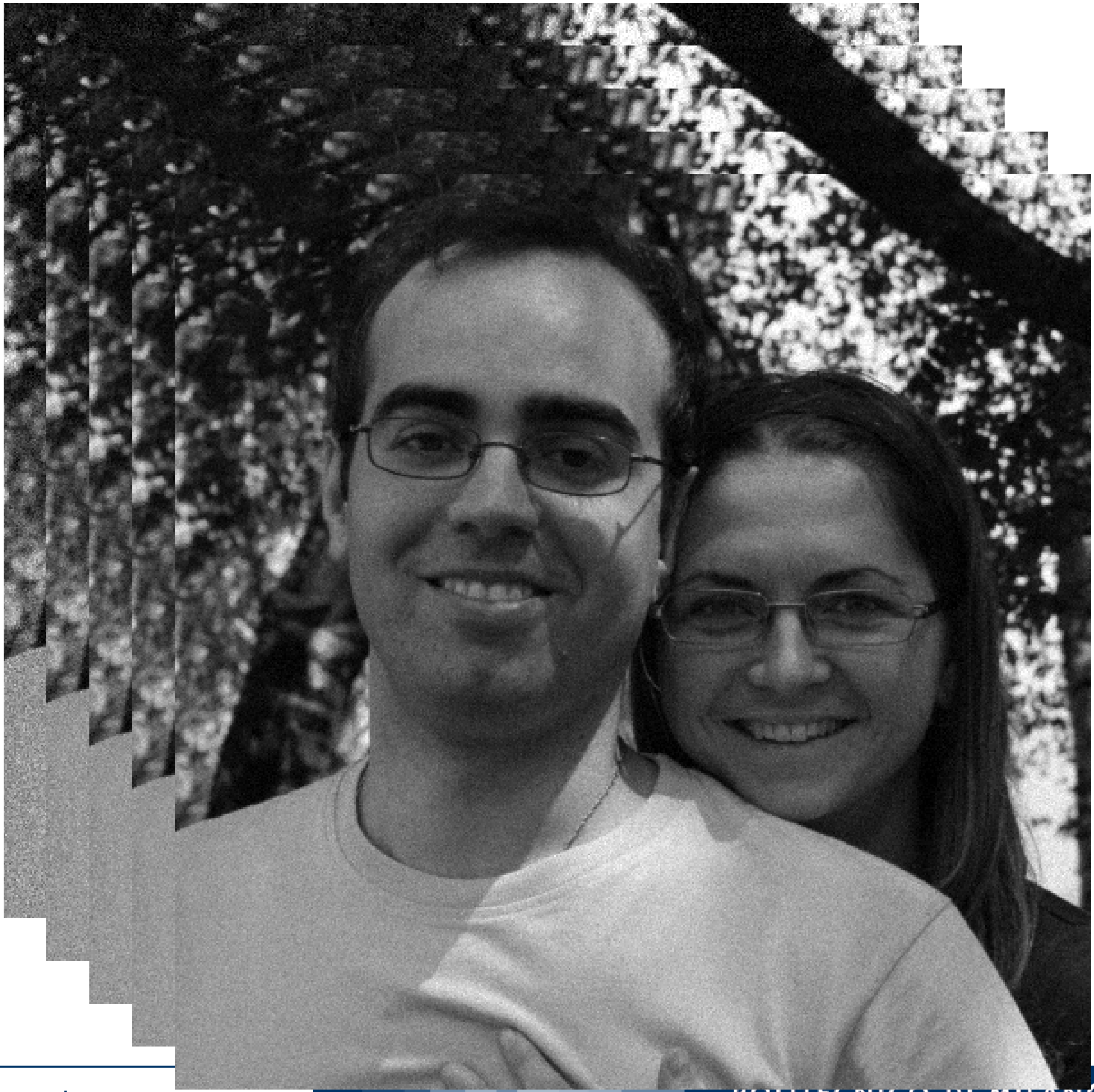
Synthetic Experiments- *Luca & Tania* Sequences

- *Mixed* Sequence

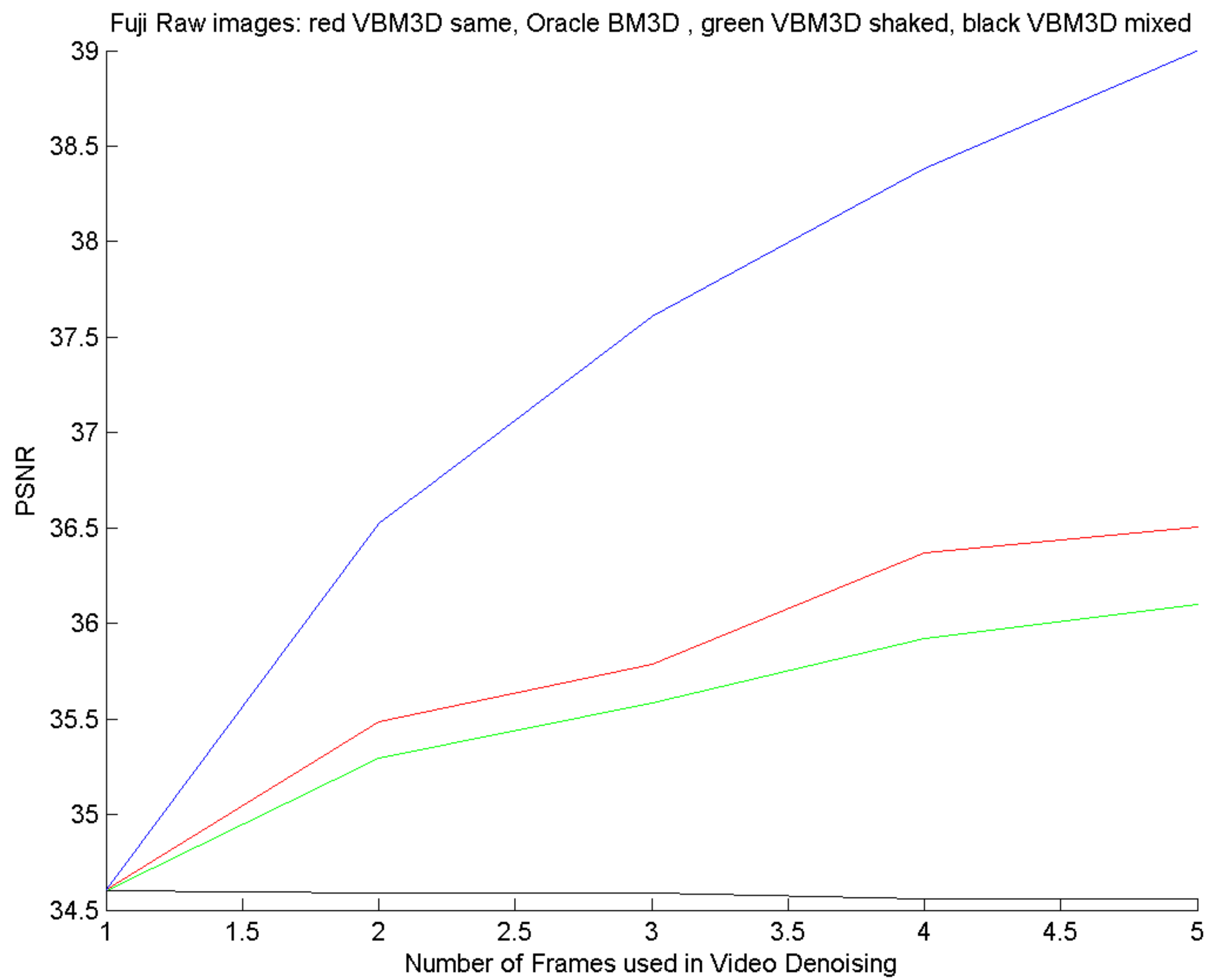
Synthetic Experiments- *Luca & Tania* Sequences

- *Oracle* Sequence



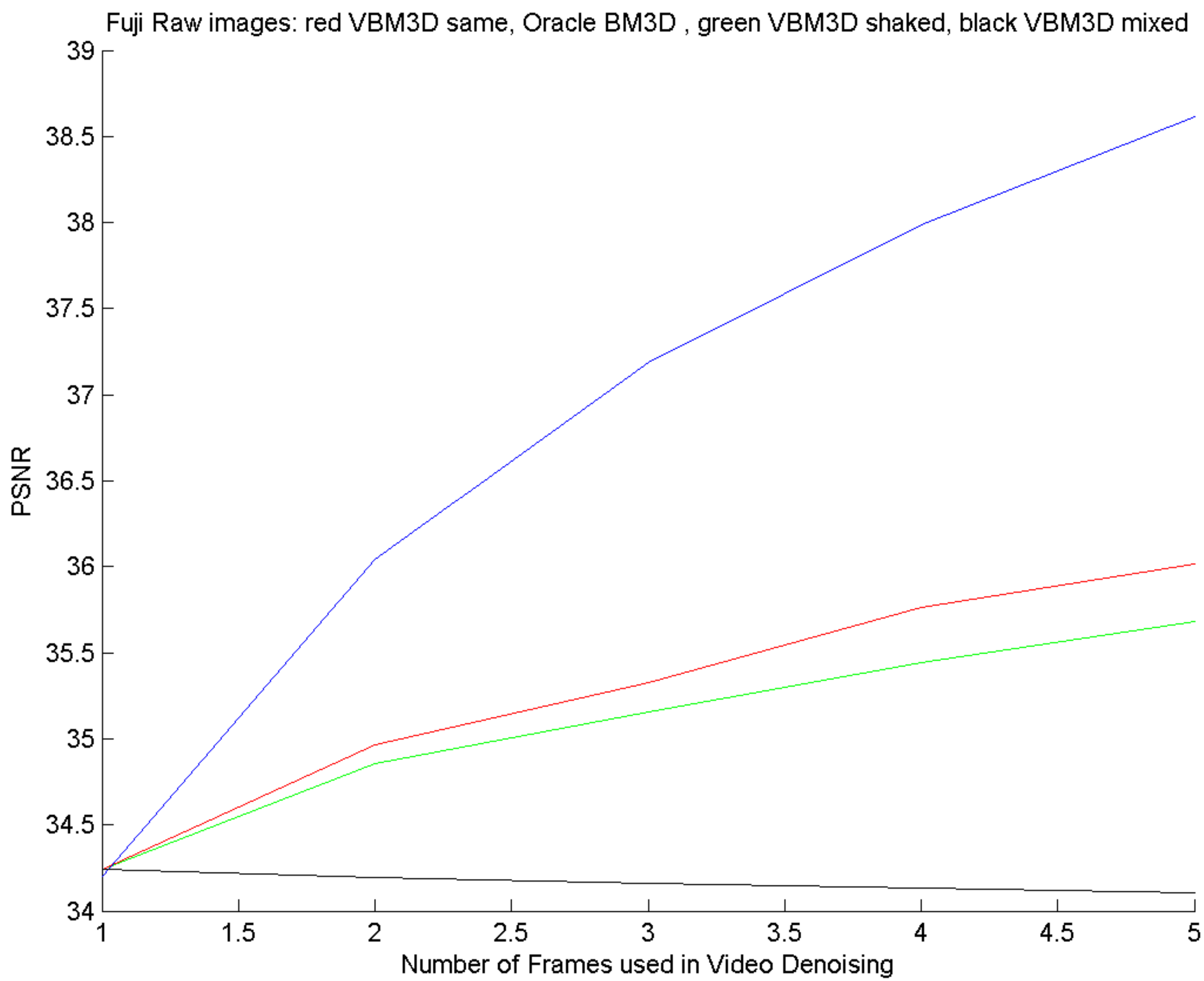


Restoration RMSE - *Luca & Tania* Sequences



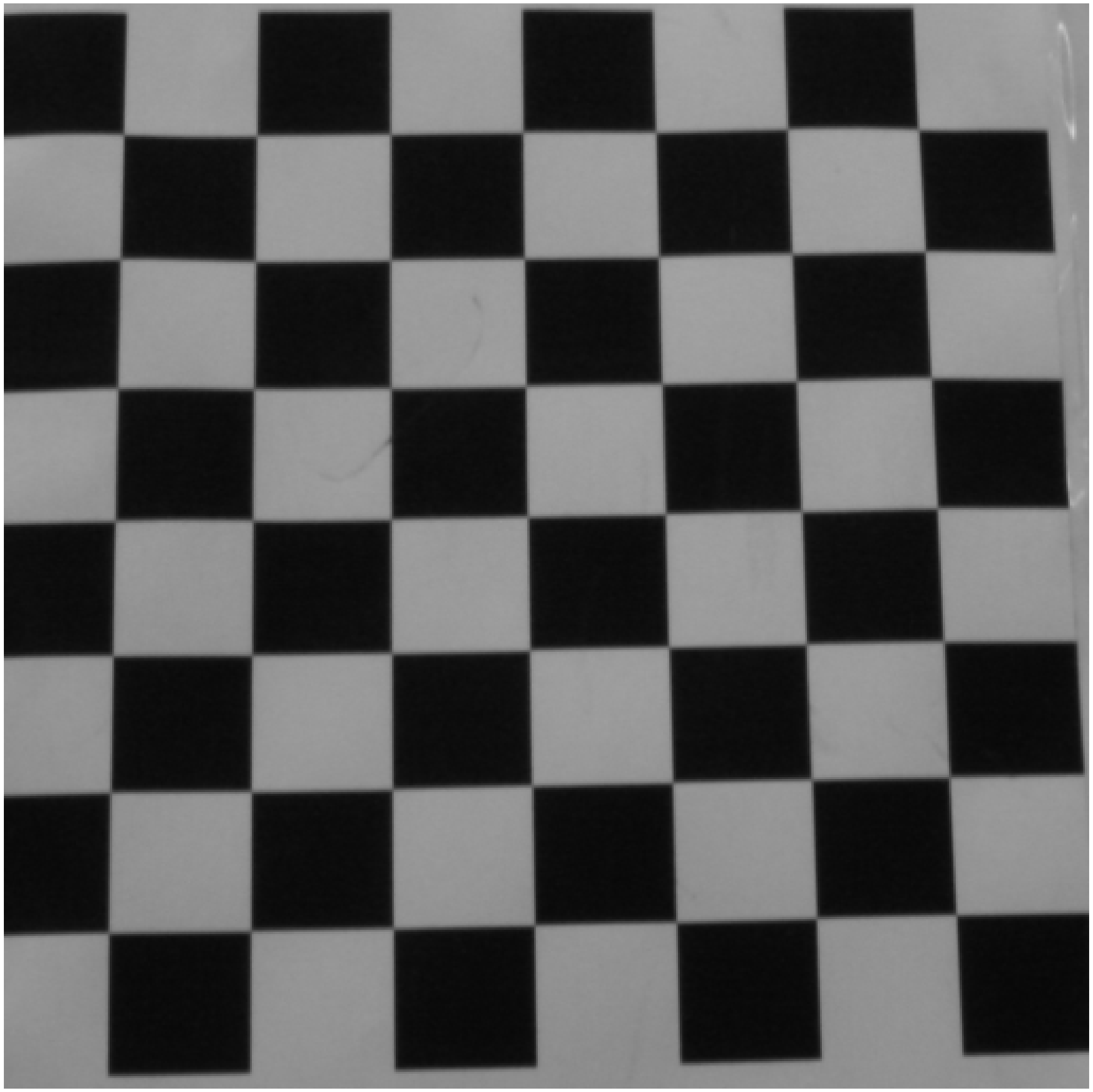
Etalo Sequences

Restoration RMSE – *Etalo* Sequences

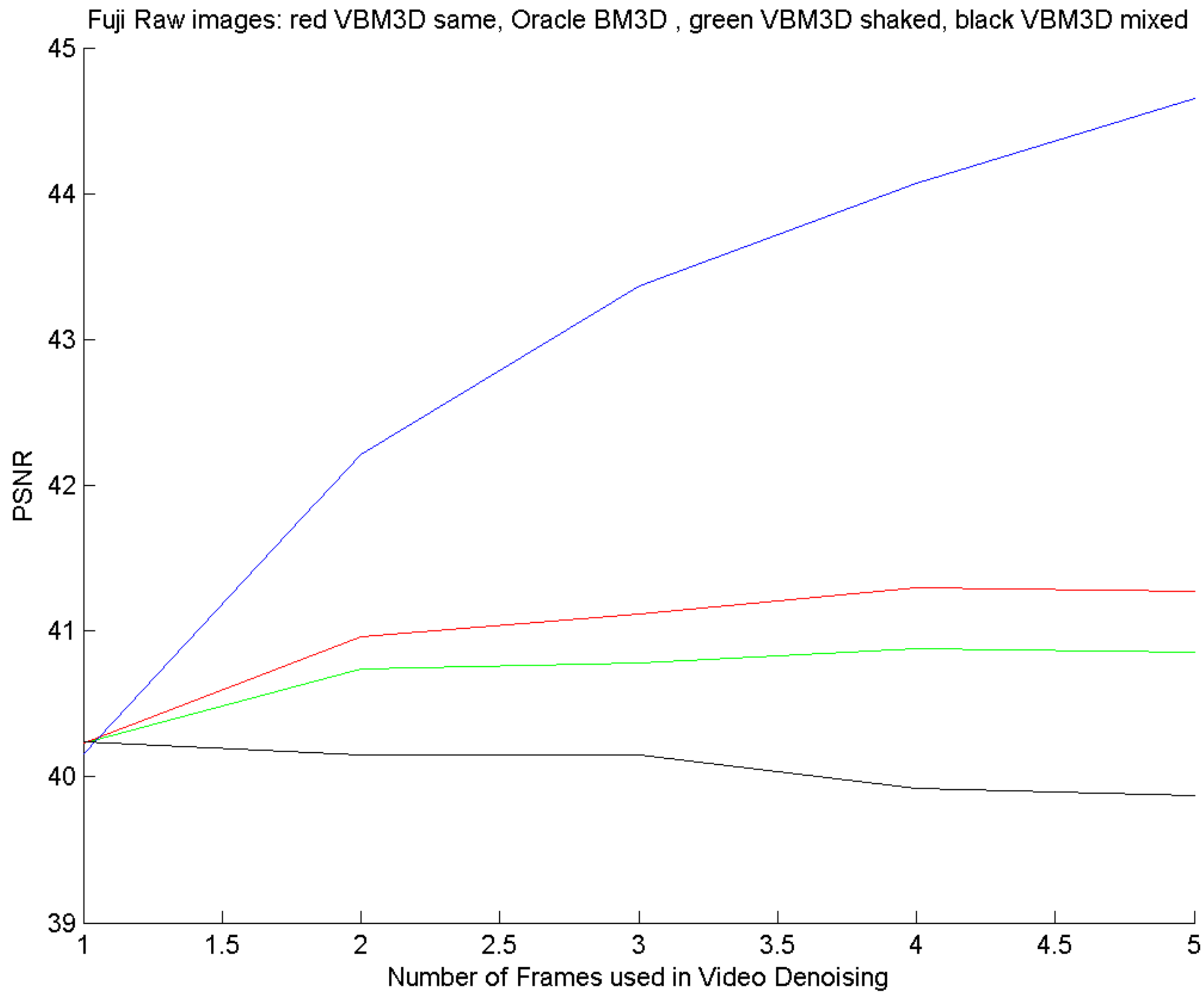


Checkerboard Sequences

Using a “more redundant”
Image as the original image



Restoration RMSE - *Checkerboard* Sequences



Experiments on camera raw data

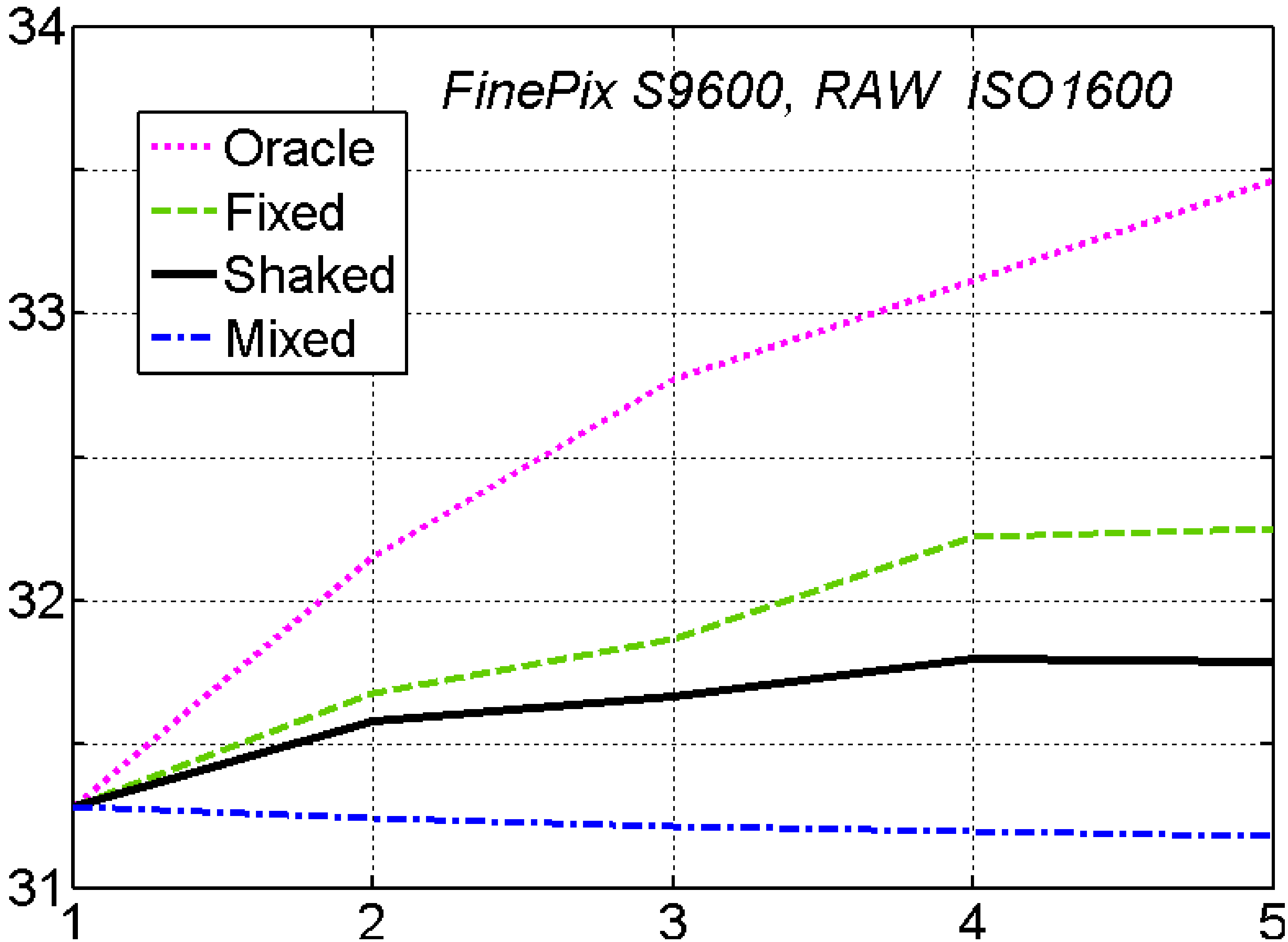
- We performed the following experiment on 3 sequences of raw data
  - “fixed”   : a sequence of short exposure images acquired with the camera a tripod
  - “shaked”: a sequence of short exposure images acquired with an hand held camera
  - “mixed” : a sequence of images of depicting completely different scenes

camera raw data

camera raw data

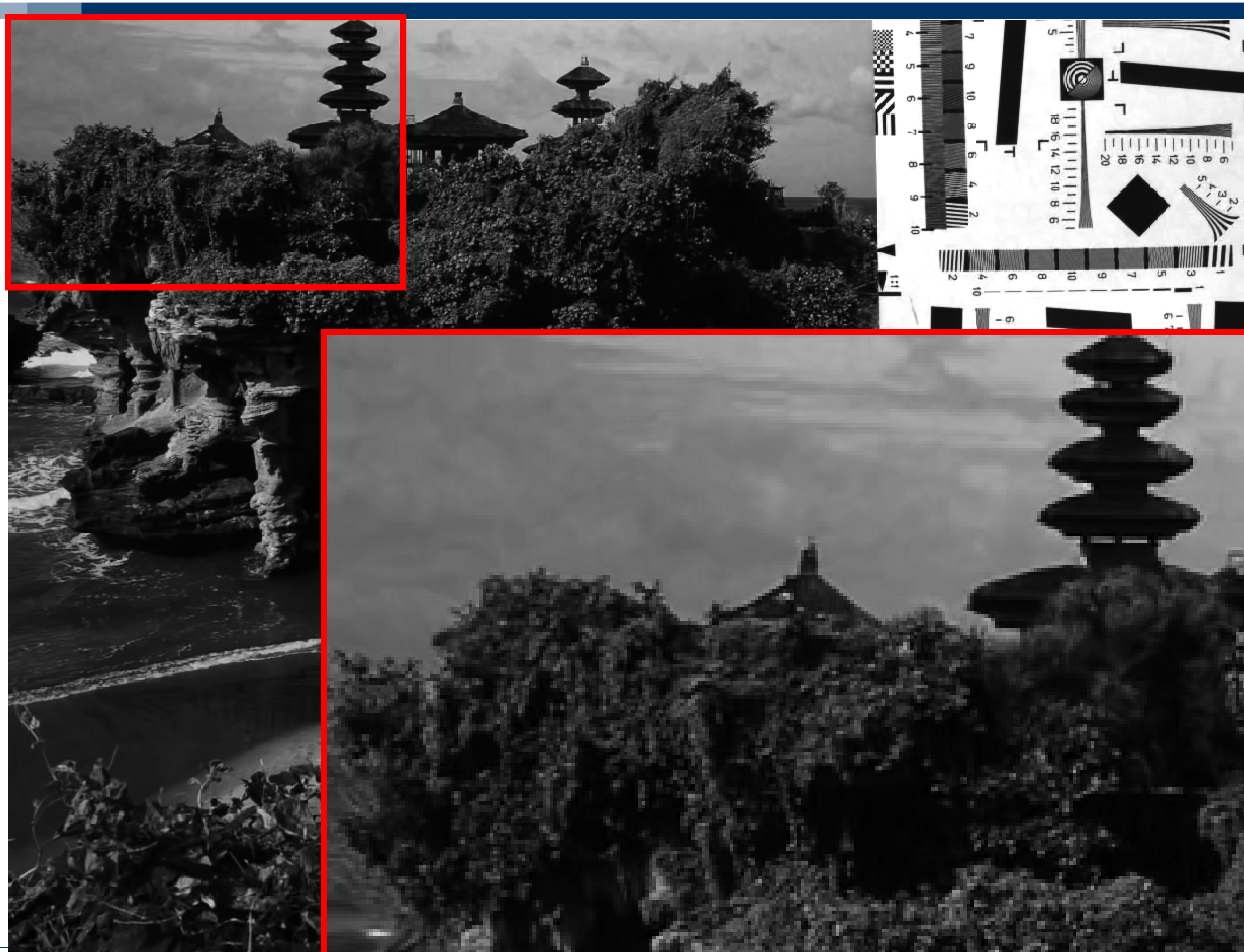
PSNR of restored images

- The behavior is consistent with synthetic experiments



restored using 5 frames[raw_shake5.tif]

restored using 5 frames[raw_shake5.tif]



Denoising vs Deblurring

- We acquired with an hand held camera the following triplet of images of a dim scene

  1. a long exposure image (ISO 100)
  2. a short exposure image (ISO 1600)
  3. a short exposure image (ISO 1600)

- We asked both *Tico et al.* and *Yuan et al*. to restore with their method the image pair 1,2

- While we restore with our method the pair 2,3

Denoising vs Deblurring

- Long exposure, camera shaked image

Denoising vs Deblurring

- One of the short exposure, noisy image

Denoising vs Deblurring

- A detail from restored with *Tico et al.* algorithm
- Visible artifacts due to mismatches between assumed blur model (invariant PSF, linearity) and real blur.

Denoising vs Deblurring

A detail from image restored with our algorithm

- There are less artifacts.
- Modeling is accurate.
- Denoising is less ill-posed than deblurring.

Denoising vs Deblurring

- A detail from image restored with *Tico et al.* algorithm

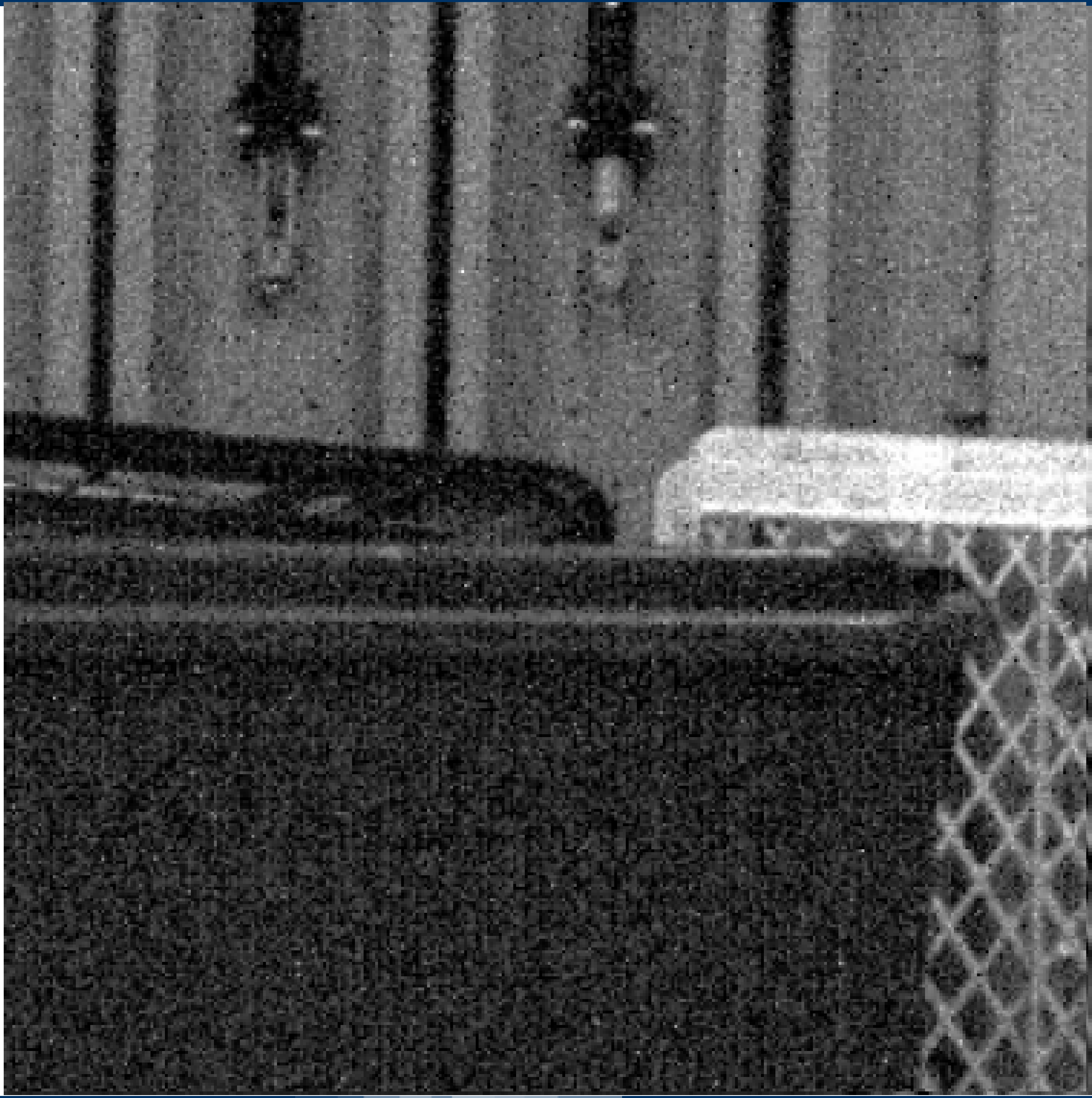
Denoising vs Deblurring

- A detail from image restored with our algorithm

- Not all details can be recovered by denoising because SNR is too low.

Denoising vs Deblurring

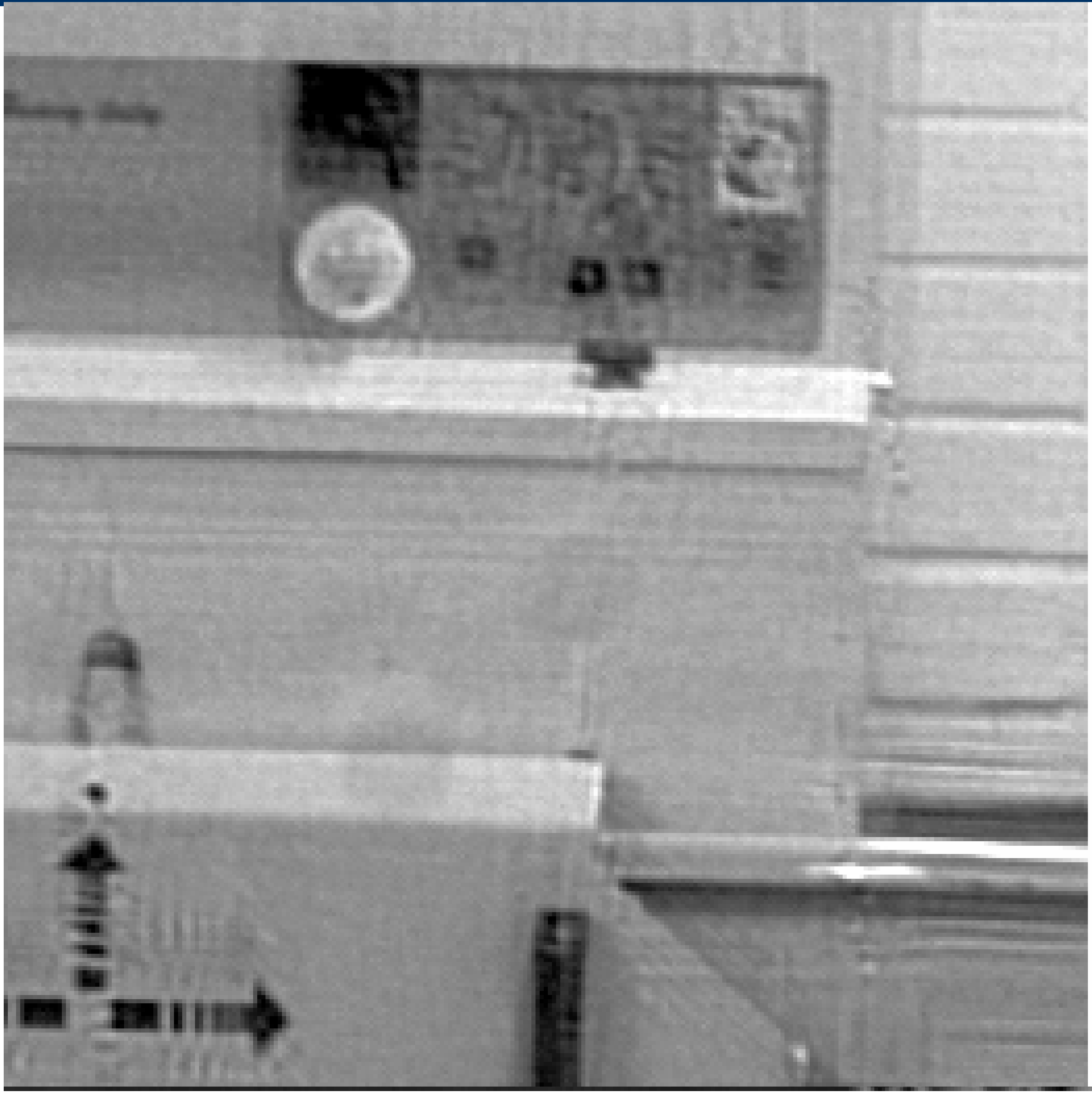
- A detail from the noisy image



Denoising vs Deblurring

- A detail from image restored with our algorithm

- Not all details can be recovered by denoising because SNR is too low.



Denoising vs Deblurring

- A detail from image restored with our algorithm

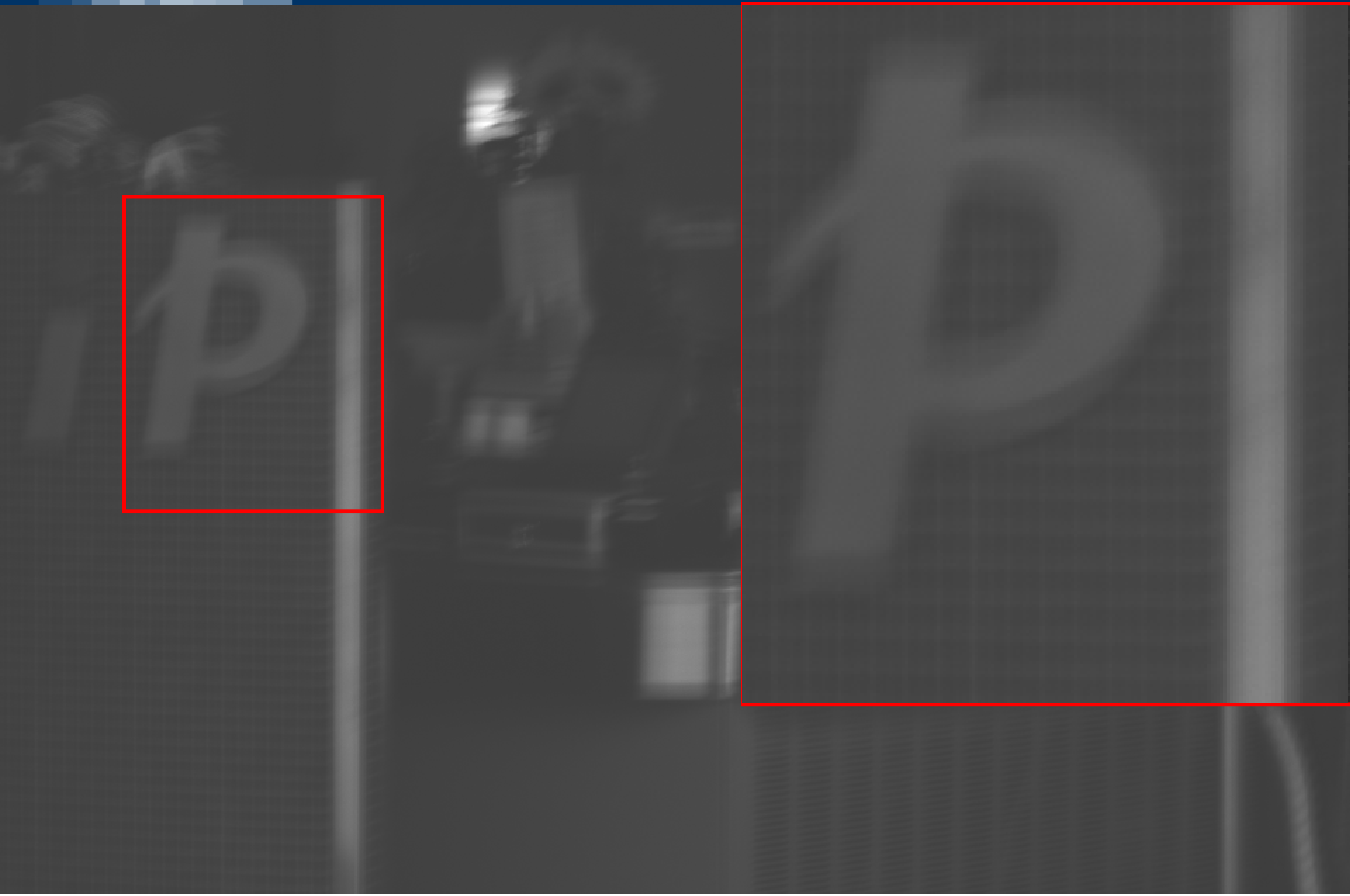
- Not all details can be recovered by denoising because SNR is too low.

Another Case

Another Case

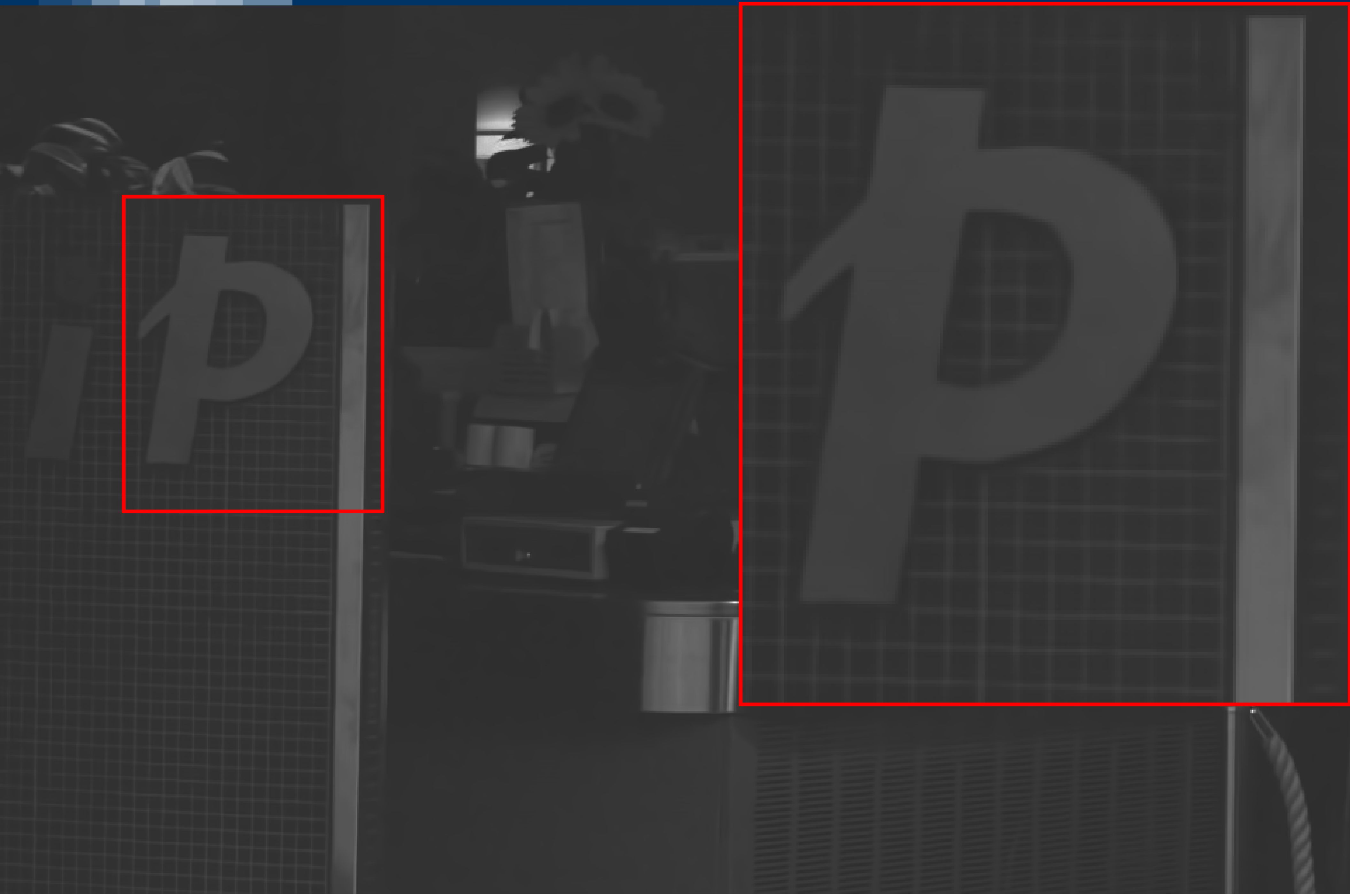
Another Case

Another Case



Another Case – Denoising-based approach

Another Case – Denoising-based approach



Another Case – Deblurring-based approach

Another Case – Deblurring-based approach

Concluding Remarks

- In "*shaked* sequences" the denoising performances always increases with the number of frames
- The gap between the "oracle" performances and the other leaves plenty of rooms for improvements.
- The proposed algorithm works indifferently in case of camera shake and object motion.

Thanks