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

- 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)


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**

\[\text{raw-data processing}\]
**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**

- 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 modeled as the *noisy* and *clipped* images

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

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

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

$$\sigma: \mathbb{R} \to \mathbb{R}^+$$
For raw data

\[ \sigma^2(y_i(x)) = ay_i(x) + b, \]

\(a\) and \(b\) depend on the sensor hardware characteristics and on the acquisition settings only.

The noisy clipped observation can be expressed as

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

where

\[
\begin{align*}
\tilde{y}_i(x) &= E\{\tilde{z}_i(x)\} \in [0, 1], \\
\tilde{\sigma}(\tilde{y}_i(x)) &= \text{std}\{\tilde{z}_i(x)\} \geq 0.
\end{align*}
\]
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
The parameters $a$ and $b$ of the noise can be 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.

Each frame is pixel-wise transformed in the following way

\[ f(t) = \int_{t_0}^{t} \frac{c}{\tilde{\sigma}(s)} ds, \quad t, t_0 \in [0, 1] \]
BM3D Denoising

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

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

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

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

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

\[
d (Z_{x_R}, Z_x) = \frac{\| \mathcal{Y}' (T_{2D}^{ht} (Z_{x_R})) - \mathcal{Y}' (T_{2D}^{ht} (Z_x)) \|_2^2}{(N_1^{ht})^2},
\]
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

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.

\[
\hat{Y}_{S^{ht}_{xR}} = T^{-1}_{3D} \left( T^{ht}_{3D} \left( Z_{S^{ht}_{xR}} \right) \right),
\]

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

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

\[
\hat{y}_{\text{basic}}(x) = \frac{\sum_{x_R \in X} \sum_{x_m \in S_{x_R}^{|x|}} w_{x_R}^{ht} \hat{Y}_{x_m}^{ht,x_R}(x)}{\sum_{x_R \in X} \sum_{x_m \in S_{x_R}^{|x|}} w_{x_R}^{ht} X_{x_m}(x)}, \forall x \in X,
\]

In such a way each grouped fragment collaborates for filtering the others and vice versa.
BM3D – Aggregation
BM3D – Aggregation
The non local search spans in both the frame and time dimensions of each block.
V-BM3D: Video BM3D

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

- Grouping: instead of using full image or fixed-size neighborhood search, we use data-adaptive neighborhoods, using the predictive search.
Variance Stabilizing Transform Inversion

- Since $f$ is nonlinear there is estimation bias:
  \[ D(f(\tilde{z}_i(x))) \approx E\{f(\tilde{z}_i(x))\} \neq f(E\{\tilde{z}_i(x)\}) \]

  being $D$ the V-BM3D denoising operator

- Debiasing [Foi08b]
  \[ h^{-1}(D(f(\tilde{z}_i(x)))) \approx f(E\{\tilde{z}_i(x)\}) \]

- and then inversion
  \[ f^{-1}(h^{-1}(D(f(\tilde{z}_i(x)))) \approx E\{\tilde{z}_i(x)\} \]

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]

\[ C : E\{\tilde{z}_i\} \longmapsto E\{z_i\} \]

note that \( C : [0, 1] \longrightarrow \mathcal{Y} \) where \( \mathcal{Y} \) 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

Fuji Raw images: red VBM3D same, Oracle BM3D, green VBM3D shaked, black VBM3D mixed

PSNR vs Number of Frames used in Video Denoising
Etalo Sequences
Fuji Raw images: red VBM3D same, Oracle BM3D, green VBM3D shaked, black VBM3D mixed

PSNR

Number of Frames used in Video Denoising

1 1.5 2 2.5 3 3.5 4 4.5 5

34 34.5 35 35.5 36 36.5 37 37.5 38 38.5 39
*Checkerboard Sequences*

Using a “more redundant” Image as the original image
Restoration RMSE - *Checkerboard Sequences*

 Fuji Raw images: red VBM3D same, Oracle BM3D, green VBM3D shaked, black VBM3D mixed
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 on 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
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
- 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 – 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 increase 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