Fast dejittering for digital video images using local non-smooth and non-convex functionals

Mila Nikolova

CMLA, ENS Cachan, CNRS, UniverSud
61 Av. President Wilson, F-94230 Cachan, FRANCE

http://www.cmla.ens-cachan.fr/∼nikolova
nikolova@cmla.ens-cachan.fr

Imaging with Modulated/Incomplete Data (SFB Workshop), July 2010, Graz Austria

Bayesian TV	Bake & Shake	Proposed≡Original
MAE=11.7, PSNR=22	MAE=7.4, PSNR=23	MAE=0, PSNR=∞

Jitter occurs in digital video when the synchronization signal is corrupted, in fax documents

Original image $\boldsymbol{f} \in \mathbb{R}^{r\times c}$, jittered image $\boldsymbol{g} \in \mathbb{R}^{r\times c}$ ($\boldsymbol{f_i}$ is a row of $\boldsymbol{f}$, $\boldsymbol{g_i}$ is a row of $\boldsymbol{g}$):

$$1 \leq j \leq c,\ \ 1 \leq i \leq r,\quad \boldsymbol{g_i}(\boldsymbol{j}) = \begin{cases} \boldsymbol{f}_i(\boldsymbol{j} + \boldsymbol{d}_i) & \text{if } \boldsymbol{1 \leq j + d_i \leq c} \\ \boldsymbol{0} & \textbf{otherwise} \end{cases} \qquad \boldsymbol{d_i} \in \mathbb{Z},\ \ |\boldsymbol{d_i}| \leq \boldsymbol{M}. \tag{1}$$

Intrinsic dejittering = restore the image frame from the jittered data (Kokaram 1997)

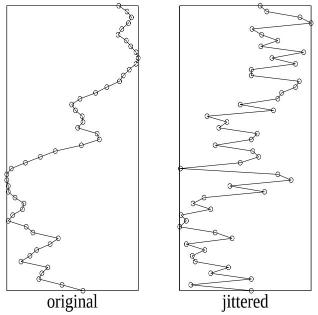
Main existing methods:

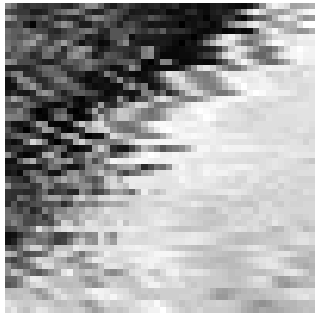
- A. Kokaram et al. (1997 and 1998) : 2D AR model + drift compensation
- J. Shen (2004) : Fully bayesian method with TV prior (B-TV)
- S.-H. Kang & J. Shen (2006) : Bake and Shake (based on a PDE model) (B&S)
- S.-H. Kang & J. Shen (2007) : Bayesian regularization of the vertical slicing moments of images

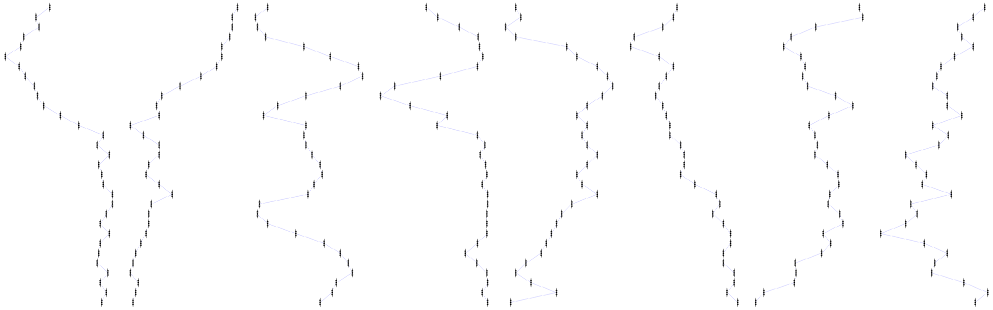
Our goal: For each jittered row $\boldsymbol{g_i}$ we wish to estimate its displacement $\widehat{\boldsymbol{d}}_{\boldsymbol{i}}$ based on the previously restored rows $\widehat{\boldsymbol{f}}_{\boldsymbol{i-1}}, \widehat{\boldsymbol{f}}_{\boldsymbol{i-2}}, \ldots$

$\Rightarrow$ We need a good prior on the gray-value of the columns of natural images

1. Choice of a criterion







Original (b) One column Jittered

(b) Gray value of the same column in original & jittered image

Gray-values of pieces of columns for various natural images

The gray-value of the columns of natural images can be seen as pieces of 2^{nd} or 3^{rd} order polynomials which is hard to claim for their jittered versions.

For stability, the lowest degree differences that fit real-world images is better

$\Rightarrow$ Consider differences of the form $\left|g_i(j-d_i) - 2\widehat{f}_{i-1}(j) + \widehat{f}_{i-2}(j)\right|$

Each row of g has at one of its extremes at most N (overestimate$\geq M$) null pixels

$\Rightarrow$ The N columns at the extremities of g are globally meaningless.

$\Rightarrow$ They must be excluded from our criterion.

$\Rightarrow$ The other columns $N+1, \cdots, c-N$ bear sound information on the true image.

$\Rightarrow$ A criterion that uses only columns $N+1, \cdots, c-N$

Estimated row displacement:

$$\widehat{d}_i = \arg\min \left\{ \mathcal{J}(d_i) \; : \; d_i \in \{-N, \cdots, 0, \cdots, N\} \right\}$$

$$\mathcal{J}(d_i) = \frac{1}{n-m+1} \sum_{j=m}^{n} \left| g_i(j-d_i) - 2\widehat{f}_{i-1}(j) + \widehat{f}_{i-2}(j) \right|^{\alpha}, \quad \alpha \in \{0.5, 1\}$$

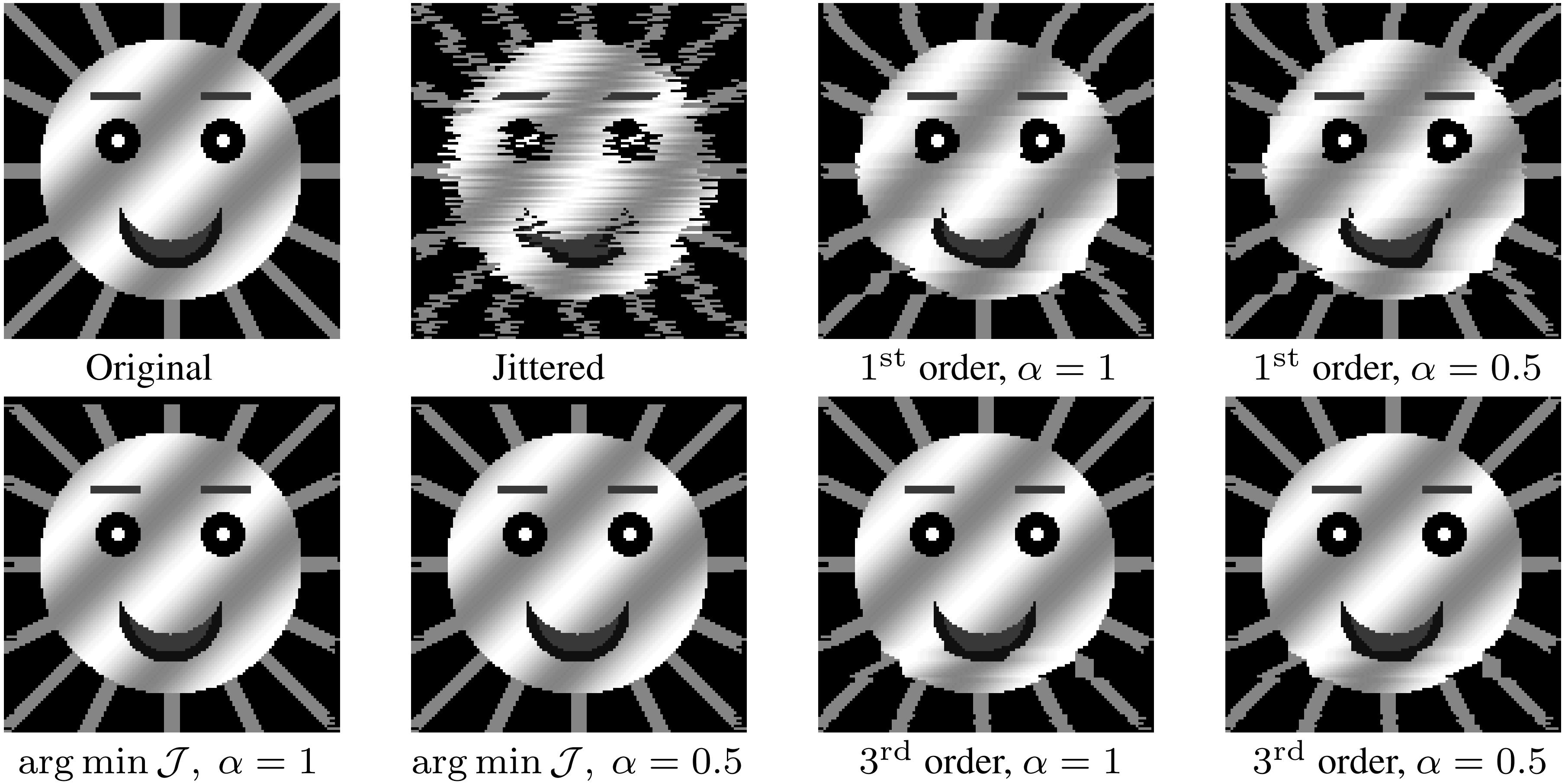
$$\begin{aligned} m &= N + 1 + \max\{d_i, \widehat{d}_{i-1}, \widehat{d}_{i-2}\} \\ n &= c - N + \min\{d_i, \widehat{d}_{i-1}, \widehat{d}_{i-2}\} \end{aligned}$$

Constants $m,\ n$ guarantee that all indexes within the sum belong to $\{N+1, \cdots, c-N\}$

Normalization $(n-m+1)^{-1}$ is needed because for each shift d_i, the sum contains a different number of terms while the solution $\widehat{d}$ is the minimizer over a set of $\mathcal{J}(d_i)$

Restored row:

$$\widehat{f}_i(j) = g_i(j - \widehat{d}_i) \quad \text{if } 1 \le j \le c \text{ and } 1 \le j - \widehat{d}_i \le c$$



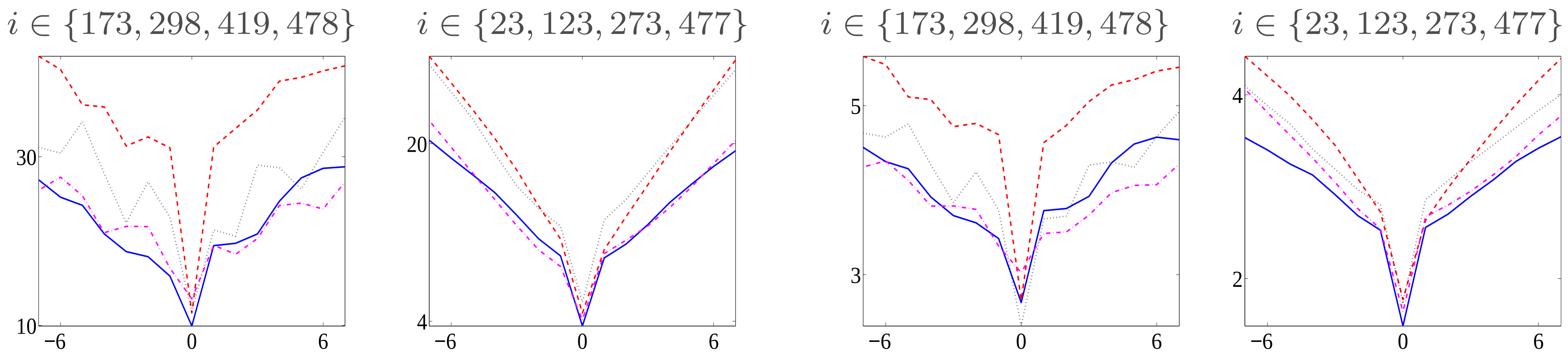
1^{st} order looks for constant vertical pieces. 3^{rd} order is too “loose” to discriminate between a column of a natural image and its slightly wrong displacements.

Interpretation of $\mathcal{J}$

• $\mathcal{J}$ finds a $\widehat{\boldsymbol{d}}_i$ such that $\widehat{\boldsymbol{f}}_i(\boldsymbol{j}+\widehat{\boldsymbol{d}}_i) \approx \mathbf{2}\widehat{\boldsymbol{f}}_{i-1}(\boldsymbol{j}) - \widehat{\boldsymbol{f}}_{i-2}(\boldsymbol{j})$ for a maximum number of $\boldsymbol{j}$ at current row i. Constraint is stronger for $\alpha < 1$

• The contribution to $\mathcal{J}$ of $\left|\widehat{\boldsymbol{f}}_i(\boldsymbol{j}+\widehat{\boldsymbol{d}}_i) - \mathbf{2}\widehat{\boldsymbol{f}}_{i-1}(\boldsymbol{j}) + \widehat{\boldsymbol{f}}_{i-2}(\boldsymbol{j})\right|^{\boldsymbol{\alpha}} \gg \mathbf{0}$ (a breakpoint) decreases as far as $\boldsymbol{\alpha} \leq \mathbf{1}$ decreases—a vertical edge at $\widehat{\boldsymbol{f}}_i(\boldsymbol{j}+\widehat{\boldsymbol{d}}_i)$ can be recovered.

$\Rightarrow$ $\alpha \in (0,1)$. For stability $|.|^{\alpha}$—increasing enough: we consider $\boldsymbol{\alpha} \in \left[\frac{\mathbf{1}}{\mathbf{2}}, \mathbf{1}\right]$



$\alpha = 1$ (Barbara) $\alpha = 1$ (Peppers) $\alpha = \frac{1}{2}$ (Barbara) $\alpha = \frac{1}{2}$ (Peppers)

X-axis: $d_i \in \{-M, \ldots, M\}$ for $M = 7$.

Y-axis: $\mathcal{J}(d_i) = \dfrac{1}{n-m+1} \displaystyle\sum_{j=m}^{n} \left|f_i(j+\boldsymbol{d_i}) - 2f_{i-1}(j) + f_{i-2}(j)\right|$.

$\boldsymbol{f}$ is the original image and the true displacement of row i is naturally $\widehat{d}_i = 0$.

2. Specific error measures

- The percentage of displaced rows in $\widehat{f}$ w.r.t. the original: $e_0(\widehat{d}, d) \stackrel{\text{def}}{=} \frac{100}{r} \|d - \widehat{d}\|_0 \%$
- The maximum horizontal error: $e_\infty(\widehat{d}, d) \stackrel{\text{def}}{=} \frac{100}{c} \|d - \widehat{d}\|_\infty \%$
- The changes in $d - \widehat{d}$: $e_0^{\Delta}(\widehat{d}, d) \stackrel{\text{def}}{=} \frac{100}{r-1} \#\{(\widehat{d}_i - d_i) - (\widehat{d}_{i+1} - d_{i+1}) \neq 0,\ 1 \leq i \leq r-1\} \%$

The proposed method leads to essentially piecewise constant $d - \widehat{d}$ whose largest part $= 0$

Example: Consider that $d - \widehat{d}$ is composed of L pieces

$$\{1, \cdots, i_{2-1}\},\ \{i_2, \cdots, i_{3-1}\},\ \cdots,\ \{i_L, \cdots, r\}$$

where $d_i = \widehat{d}_i$, $i_2 \leq i \leq i_{3-1}$ (the largest constant piece). Then

$$\begin{aligned} e_0(\widehat{d}, d) &= i_{2-1} + r - i_3 + 1 \quad \text{(can be high for any } L \geq 2\text{)} \\ e_0^{\Delta}(\widehat{d}, d) &= L \times \frac{100}{r-1} \% \end{aligned}$$

By the latter, L pairs of consecutive rows are misplaced relatively to each other.
Let the maximal displacement between two consecutive rows is K pixels. Then

$$e_\infty(\widehat{d}, d) = K \times \frac{100}{c} \%$$

For a 512×512 image, i.e. $r = c = 512$, errors like $L = 4$ and $K = 2$ remain visually indistinguishable from the original. Their measures read $e_0^{\Delta} = 0.78\%$ and $e_\infty = 0.39\%$.

Remark: If both e_∞ and e_0^Δ are small (e.g. $e_\infty \leq \mathbf{0.4\%}$ and $e_0^\Delta \leq \mathbf{0.8\%}$), we are guaranteed that dejittering is nearly perfect, independently of any other error measure.

E.g. for a 512×512 image—no more than 4 rows have a horizontal error up to 2 pixels. Such an error is invisible to the naked eye.

If one of these measures is higher, nothing can be claimed on the quality of $\widehat{\boldsymbol{f}}$.

E. g., if the image is planar (or constant) on a horizontal slice, large errors e_∞ and e_0^Δ remain invisible.

Bayesian TV (B-TV) MAE=13.36, PSNR=20.82

Bake & Shake (B&S) MAE=12.5, PSNR=20.27

Proposed $\mathcal{J}$, $\alpha \in \{0.5, 1\}$ MAE=0.16, PSNR=42.87 e_∞=0.39%, e_0^Δ=0.25%

3. Main algorithm

1. $\boldsymbol{g} = \left[\, \boldsymbol{g}^L \vdots \overline{\boldsymbol{g}} \vdots \boldsymbol{g}^R \,\right]$ where $\boldsymbol{g}^L \in \mathbb{R}^{r\times N}$, $\overline{\boldsymbol{g}} \in \mathbb{R}^{r\times(c-2N)}$ and $\boldsymbol{g}^R \in \mathbb{R}^{r\times N}$
2. for each i, optimal displacement $\widehat{d}_i$ is computed based on $\overline{\boldsymbol{g}}_i$ by minimizing $\mathcal{J}$
3. then $\widetilde{\boldsymbol{f}}_i(j+N) = \boldsymbol{g}_i(j-\widehat{d}_i)$, $1 \le j \le c$ Note: $\widetilde{\boldsymbol{f}}_i \in \mathbb{R}^{1\times(c+2N)}$
4. at step r, $\widetilde{\boldsymbol{f}} \in \mathbb{R}^{r\times(c+2N)}$ and $\widehat{\boldsymbol{f}}$ is an $r \times c$ submatrix of $\widetilde{\boldsymbol{f}}$.

Computation times: For a 512×512 image and $N = 7$, the solution is got 0.62 second for $\alpha = 1$ and in 1 second for $\alpha = 0.5$. (Matlab 7.2, PC with a Pentium 4 CPU 2.8GHz and 1GB RAM, Windows XP Professional service pack 2)

4. Large-scale experiment

1000 experiments with Lena, Peppers, Barbara, Boat for 2 different types of independent jitter.

- The means of e_0^Δ and e_∞ are small.
- $\widehat{d} - d$ is either null or piecewise constant with values near 0.
- The mean and the variance of MAE are small.
- In almost all cases, $\alpha = 0.5$ is better than $\alpha = 1$.

Jitter

Jittered image (512×512)

Main Algorithm, $\boldsymbol{\alpha = 0.5}$

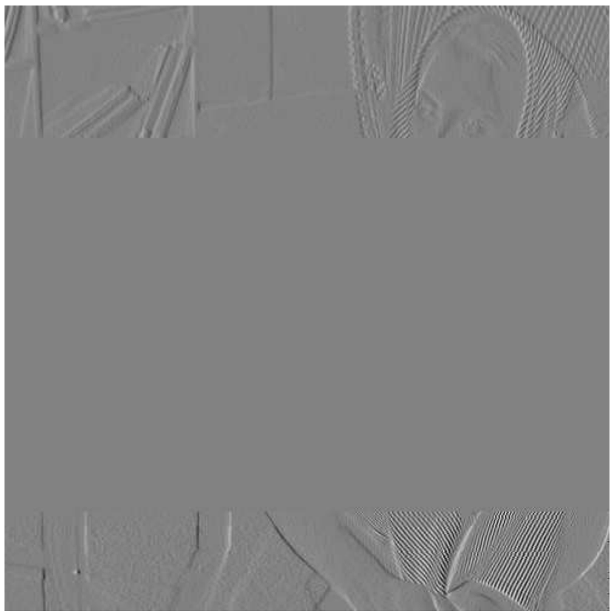
MAE$=$ 4.16

PSNR=25.53

e_1=0.52

e_∞=0.39%

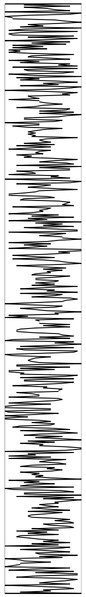
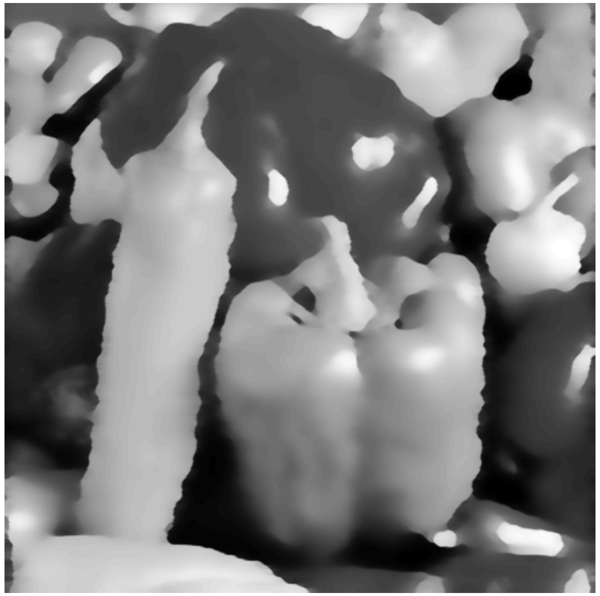
e_0^Δ=0.59%



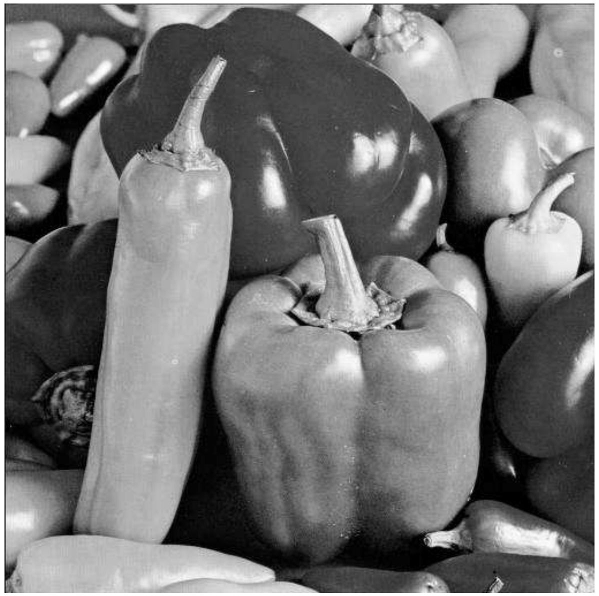
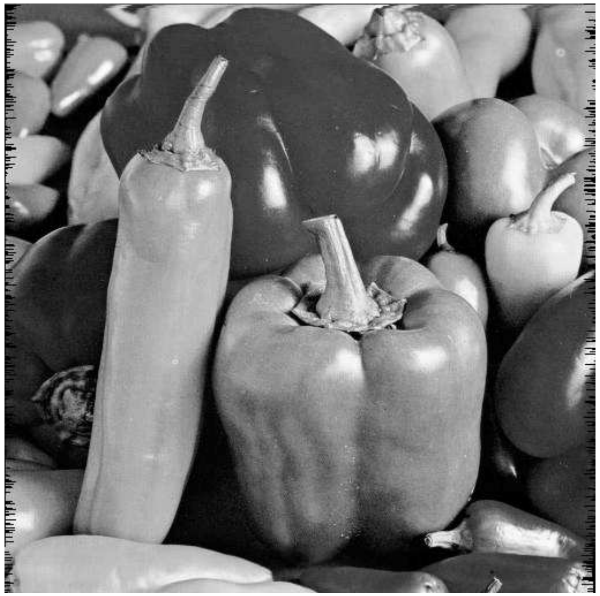
Error $\widehat{f} - f$ (error crosses the nose)

Zoom dejittered

Zoom original

Jitter   Jittered Image M=10   B-TV, MAE=15.52, PSNR=20.1   B&S, MAE=14.12, PSNR=22.39

Original   Our: α=0.5, MAE=1.35, PSNR=31.51   Our: Error $f - \widehat{f}$

5. Color images

The jitter is the same for all color channels (R,G,B)

In step 2 of Main Algorithm (sec. 3) $\overline{g}_i$ is replaced by $|\overline{g}_i^R| + |\overline{g}_i^G| + |\overline{g}_i^B|$ (gray value)

The row shift $\widehat{d}$ is estimated for the gray value image. Then $\widehat{d}$ is inserted in the color image

$\alpha = 1$

MAE=0.2

PSNR=40.34

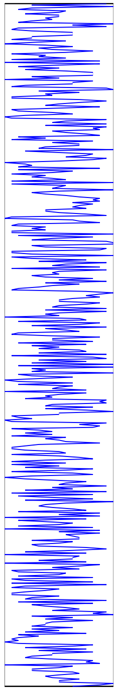
α=0.5

e_∞=0.24%

e_0^Δ=0.37%

Original

Jittered, M=20 α=0.5, e_∞=0.35%, e_0^Δ=0.28 Original 707×579

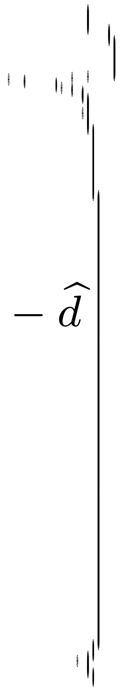


Original

original

restored

$d - \widehat{d}$



$\alpha = 1$

MAE=1.45

PSNR=33.82

errors mainly
in flat strips

6. Noisy jittered images

For weak noise ($\geq 20 - 30$ dB) just use Main Algorithm

20dB SNR+Jitter M=6

Noisy, Dejittered, $\alpha = 1$

Denoising: hard shrinkage

Original (256×256)

Strong noise

Main idea

For $i = 1, \cdots, r$

- Denoise row $\overline{\boldsymbol{g}}_{\boldsymbol{i}}$ using fast shrinkage estimator to ensure the 2nd order assumption
- Estimate $\widehat{d}_i$ using Main Algorithm

Insert $\widehat{d}$ into jittered noisy data to get a noisy dejittered image

Use a denoising method to restore this noisy dejittered image

Changes in Main Algorithm:

- changes in step 2
  - if RGB image—transform $\overline{g}_i$ into gray-value (as in sec. 5)
  - replace $\overline{g}_i$ by $\gamma_i = W^* y_i^T$ where
    for $W : \mathbb{R}^{1\times c} \to \mathbb{R}^{1\times c}$ — 1D wavelet transform and W^* its inverse,

$$y_i = W\overline{g}_i \in \mathbb{R}^{1\times c}$$

and for a (small) $T > 0$, $y_i^T(j) = \begin{cases} 0 & \text{if } |y_i(j)| \leq T \\ y_i(j) & \text{otherwise} \end{cases} \quad 1 \leq j \leq c.$

  - replace $\mathcal{J}$ by

$$\widetilde{\mathcal{J}}(d_i) = \frac{1}{n-m+1} \sum_{j=m}^{n} \varphi\Big(\big|\gamma_i(j) - 2\gamma_{i-1}(j) + \gamma_{i-2}(j)\big| + \beta\big|\gamma_i(j) - \gamma_{i-1}(j)\big|\Big)$$

where φ is edge preserving and $\beta \geq 0$

- Step 3 is the same
- Add step 5: classical denoising of $\widehat{f}$ (e.g. shrinkage estimation)

Gaussian noise + binomial jitter Bayesian TV Our method (+curvelets) Original (256×256)

10 dB SNR + Jitter, M=8

Bayesian TV

Bake & Shake

Our dejitter: $\varphi(t) = |t|^{0.5}, \beta = 0$

Denoising: curvelets shrinkage

Original

Jitter 10dB SNR+Sin-jitter, M =6 Dejitter: $\varphi(t)=|t|^{0.5}, \beta = 0$ Denoise: curvelets shrinkage

Jitter 10dB SNR+Sin jitter, M=6 Dejitter: $\varphi(t)=|t|^{0.5}, \beta = 0$ Denoise: curvelets shrinkage

7. Conclusions and perspectives

- Very fast and simple dejittering method yielding remarkable results
- A better exploration for the parameters in presence of noise is necessary
- The case of general impairments + jitter is unexplored
- Go further to restore full jittered sequences (on-going)

Full paper: One-iteration dejittering of digital video images, Journal of Visual Communication and Image Representation, Vol. 20, 2009, pp. 254-274

see also: http://www.cmla.ens-cachan.fr/∼nikolova/, Journal papers

Acknowledgements

- the French *Agence Nationale de la Recherche (ANR)*, under grant FREEDOM (ANR07-JCJC-0048-01), ”Films, REstauration Et DOnnées Manquantes”.
- *Louis LABORELLI* (Institut Nat. de l'Audiovisuel, France) for discussion on practical questions.
- Dr. *Jackie SHEN* (Manager at Barclays Capital, Wall Street) who gave me his Matlab codes.
- Dr. *Sung-Ha KANG*, Ass. Prof., Georgia Institute of Technology, Atlanta, who realized comparative experiments for me!