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

Giacomo Boracchi[*] *and Alessandro Foi*[**]

[*] Dipartimento di Elettronica e Informazione, Politecnico di Milano
via Ponzio 34/5, 20133, Milano, Italy
web: http://home.dei.polimi.it/boracchi email: firstname.lastname@polimi.it

[**] Department of Signal Processing, Tampere University of Technology
P.O. Box 553, 33101, Tampere, Finland
web: http://www.cs.tut.fi/~foi email: firstname.lastname@tut.fi

ABSTRACT

We consider the problem of the joint denoising of a number of raw-data images from a digital imaging sensor. In particular, we exploit a recently proposed image modeling [8] that incorporates both the signal-dependent nature of noise and the clipping of the data due to under- or over-exposure of the sensor.

Our denoising approach is based on the V-BM3D algorithm [5], coupled with a set of homomorphic pre- and post-processing transformations derived for variance-stabilization, debiasing, and declipping [6]. The spatio-temporal nonlocality of V-BM3D frees us from the need of an explicit registration of the frames. It results in a practical algorithm directly applicable to raw-data processing, in particular for heavy-noise conditions such those encountered in low-light imaging or imaging at fast shutter speeds.

Experiments with synthetic images and with real raw-data from CCD sensor show the feasibility of the approach and provide an indicative measure of the advantage of multiframe versus single-frame processing.

1. INTRODUCTION

Pictures acquired by digital imaging sensors are always subject to noise. While the signal-to-noise ratio (SNR) can be improved by using a longer exposure time, this is often not feasible because scene motion (e.g., due to moving objects) or camera motion – also referred to as camera shake – during the acquisition would result in blur. The problem is particularly evident when acquiring images at low-light conditions.

A number of diverse solutions have been devised to cope with this kind of problems. These range from hardware solutions, such as optical stabilization based on real-time motion-adaptive sensor or lens actuation, to different acquisition paradigms. Particularly effective for compensating the impact of motion or hand-held camera shake blur is the approach based on pairs of differently exposed images [14, 13, 17, 18, 16, 15]. The key idea is to capture two images: one image taken with a short exposure-time, which ensures that the blur is negligible at the expense of heavy noise, and another image taken with a longer exposure, which reduces the noisiness but results in visible blur. Provided some registration, the noisy image is used in order to estimate the blur point-spread function (PSF), thus enabling a non-blind or semi-blind deconvolution of the blurred image. However, scene motion or camera shake very seldom can be faithfully described as a linear, shift-invariant blur; thus, heavy regularization is necessary to reduce artifacts [17].

An alternative strategy is based on the joint denoising of multiple images captured sequentially, thus making the problem conceptually equivalent to a video-denoising problem. In this paper, we follow this direction and consider the problem in the very specific setting of raw-data processing, through an observation model [8] that explicitly incorporates both the signal-dependent nature of noise and the clipping of the data due to under- or over-exposure of the sensor.

In our approach, we rely on the Video Block-Matching 3-D denoising algorithm (V-BM3D) [5], coupled with a set of homomorphic pre- and post-processing transformations derived for variance-stabilization, debiasing, and declipping [6]. The spatiotemporal nonlocality of V-BM3D frees us from the need of an explicit registration of the frames, while the homomorphic transformations enable an accurate estimation of the true image. Overall, it results in a practical algorithm directly applicable to multiframe raw-data processing, which simultaneously extends [5] and [6].

The rest of the paper is organized as follows: Section 2 introduces the observation model and the principal ideas of the V-BM3D filter. The proposed denoising algorithm is detailed in Section 3. Experiments with synthetic images and with real raw-data from CCD sensor are presented in Section 4, where we also compare against the approach based on blurred-noisy image pairs. We conclude the paper with few remarks about the impact of redundancy on the denoising performance.

2. PRELIMINARIES

2.1 Observation model

Let $\{\tilde{z}_i\}_{i=1}^{N}$, be a sequence set of N raw-data images. According to [8, 6], each image $\tilde{z}_i : X \rightarrow [0,1]$ can be modeled as

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

where

$$z_i(x) = y_i(x) + \sigma(y_i(x))\,\xi_i(x), \tag{2}$$

$y_i : X \rightarrow Y \subseteq \mathbb{R}$ is a deterministic unknown original image and $\sigma(y_i(x))\xi_i(x)$ is a zero-mean random error with signal-dependent standard-deviation $\sigma(y_i(x))$. Here, $\sigma : \mathbb{R} \rightarrow \mathbb{R}^+$ is a deterministic function while $\xi_i(x)$ is a random variable with unitary variance. For simplicity, the latter shall be approximated as a standard normal and all errors are assumed to be independent, thus treating ξ_i as i.i.d. with $\xi_i(\cdot) \sim \mathcal{N}(0,1)$. As discussed in [8], this is a suitable approximation when dealing with the noise in the raw data from CMOS and CCD digital imaging sensors. For these raw data, the typical form of the function σ is

$$\sigma^2(y_i(x)) = a y_i(x) + b, \tag{3}$$

with the constants $a \in \mathbb{R}^+$ and $b \in \mathbb{R}$ depending on the sensor's specific characteristics and on the particular acquisition settings (e.g., analog gain or ISO value, temperature, pedestal, etc.) [8]. These two constants are assumed fixed and invariant during the acquisition of the images and thus the same for all $i = 1, \ldots, N$.

This work was supported by the Academy of Finland (application no. 213462, Finnish Programme for Centres of Excellence in Research 2006-2011, and application no. 118312, Finland Distinguished Professor Programme 2007-2010) and by CIMO, the Finnish Centre for International Mobility (fellowship number TM-07-4952). Part of this work was carried out during the first author's visit at Tampere International Center for Signal Processing (TICSP) in July-October 2007.

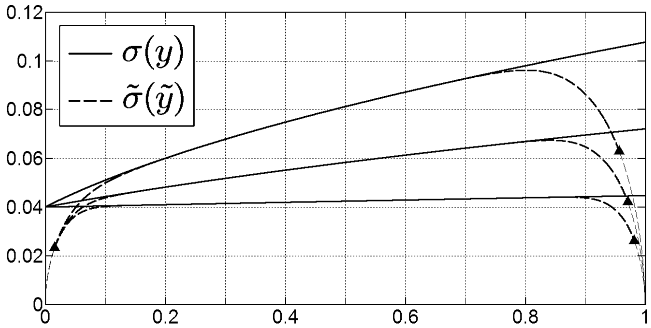


Figure 1: Some examples of the standard-deviation functions σ (solid lines) and $\tilde\sigma$ (dashed lines) from the models (2) and (5) for different combinations of the constants a and b of Equation (3): (left) $a = 0.02^2, 0.06^2, 0.10^2$, $b = 0.04^2$ and (right) $a = 0.4^2$, $b = 0.02^2, 0.06^2, 0.10^2$. The small black triangles indicate the points $(\tilde y, \tilde\sigma(\tilde y))$ which correspond to $y = 0$ and $y = 1$.

Without loss of generality, in (1) we are considering data given on the range [0, 1], where the extreme values 0 and 1 correspond to the minimum and maximum pixel values for the considered raw-data format. Values below or above these bounds are replaced by the bounds themselves: this clipping corresponds to the behavior of digital imaging sensors in the case of under- or over-exposure. While the probability density function (p.d.f.) of the unobserved (virtual) non-clipped noisy data $z_i(x)$ is simply $\frac{1}{\sigma(y_i(x))}\phi\left(\frac{\zeta - y_i(x)}{\sigma(y_i(x))}\right)$, because of the clipping (1), the observed $\tilde z_i(x)$ is distributed according to a special doubly censored Gaussian distribution [2] having a generalized p.d.f. $\wp_{\tilde z_i(x)}$ of the form

$$\wp_{\tilde z_i(x)}(\zeta) = \frac{1}{\sigma(y_i(x))}\phi\left(\frac{\zeta - y_i(x)}{\sigma(y_i(x))}\right)\chi_{[0,1]} + \\ + \Phi\left(\frac{-y_i(x)}{\sigma(y_i(x))}\right)\delta_0(\zeta) + \left(1 - \Phi\left(\frac{1 - y_i(x)}{\sigma(y_i(x))}\right)\right)\delta_0(1 - \zeta). \quad (4)$$

Here, $\chi_{[0,1]}$ denotes the characteristic function of the interval [0, 1], δ_0 is the Dirac delta impulse at 0, and ϕ and Φ are the p.d.f. and cumulative distribution function (c.d.f.) of the standard normal $\mathcal{N}(0, 1)$, respectively. The last two addends in (4) correspond to the probabilities of clipping from below and from above (under- or over-exposure).

The expectation and the standard deviation of $\tilde z_i(x)$ are denoted as

$$\begin{aligned} \tilde y_i(x) &= E\{\tilde z_i(x)\} \in [0, 1], \\ \tilde\sigma(\tilde y_i(x)) &= \operatorname{std}\{\tilde z_i(x)\} \geq 0. \end{aligned}$$

These equations define a function $\tilde\sigma : [0, 1] \to \mathbb{R}^+$ that maps the expectation of $\tilde z(x)$ to its standard-deviation, leading us to the counterpart of the signal-dependent noise model (5) for $\tilde z$:

$$\tilde z_i(x) = \tilde y_i(x) + \tilde\sigma(\tilde y_i(x))\tilde\xi_i(x), \quad x \in X \subset \mathbb{Z}^2. \quad (5)$$

Here, $\tilde\xi_i(x)$ is another (non Gaussian) random variable with zero mean and unitary variance, $E\{\tilde\xi_i(x)\} = 0$, $\operatorname{var}\{\tilde\xi_i(x)\} = 1$. As opposed to $\xi_i(x)$, $\tilde\xi_i(x)$ is not identically distributed: different distributions are found for different i and different x, as can be easily derived from (4) and (5). The functions σ and $\tilde\sigma$ are illustrated in Figure 1. We refer the reader to [8] for further details on the above model, including the derivation of the direct and inverse functional relations between y_i, $\tilde y_i$, σ, and $\tilde\sigma$.

Our goal is to recover the unknown y_i, $i = 1, \ldots, N$, of (2), given the raw-data image sequence $\{\tilde z_i\}_{i=1}^N$ (1), possibly exploiting the redundancy due to portions of image content shared by different frames.

2.2 V-BM3D

In order to exploit the similarities between the image content across multiple frames without the requirement of an explicit motion estimation or registration, we use the V-BM3D denoising filter [5]. A difference between this filter and conventional block-based denoisers lies in the fact that V-BM3D takes advantage not only from the similarity between different frames but also of the self-similarity found within individual frames. It is a spatiotemporal *nonlocal* method. A detailed description of the V-BM3D denoising filter can be found in [5]. In brief, given a noisy sequence, the filter works as follows.

- *Blockwise estimates*. Each image in the sequence is processed in sliding-block manner. For each block the filter performs:
  - *Grouping*. Searching within all images in the sequence, find blocks that are similar to the currently processed one, and then stack them together in a 3D array (group).
  - *Collaborative filtering*. Apply a 3-D transform to the formed group, attenuate the noise by shrinkage (hard-thresholding or empirical Wiener filtering) of the transform coefficients, invert the 3D transform to produce estimates of all grouped blocks, and return these estimates of the blocks to their original place.
- *Aggregation*. Compute the estimates of the output images by weighted averaging all of the obtained blockwise estimates that are overlapping.

Due to the similarity between the grouped blocks, the transform can achieve a highly sparse representation of the true signal so that the noise or small distortions can be well separated by shrinkage. In this way, the collaborative filtering reveals even the finest details shared by grouped fragments and at the same time it preserves the essential unique features of each individual fragment.

The V-BM3D is an extension to image sequences of the Block-Matching 3D filtering (BM3D) image denoising algorithm [4]. The two algorithms coincide when the sequence is composed of a unique image.

Both BM3D and V-BM3D are developed and implemented for observations degraded by additive white Gaussian noise (AWGN). There is a big deal of difference between such ideal observations (where there is no clipping and where the noise has constant signal-independent variance) and the clipped observations with signal-dependant noise described in the previous section. Direct application of a denoising algorithm for AWGN to the raw-data is highly ineffective, leading to visible oversmoothing and undersmoothing of various parts of the image [due to the non-constant variance $\tilde\sigma(\tilde y_i)$ in (5)] and to an essentially biased estimate (because of clipping, the random differences between the observed $\tilde z_i$ and the desired y_i have a non-zero mean).

3. ALGORITHM

In [6], we recently proposed a complete pre- and post-processing framework to enable efficient and effective raw-data image filtering using standard denoising algorithms for AWGN. The developed procedure is based on a set of homomorphic transformations specifically designed for the particular noise model at hand. These transformations are pixelwise operations and are thus applicable to images as well as to videos without any modification.

The overall denoising algorithm, using notation for the multi-frame case, can be then summarized as follows:

1. Estimate the noise parameters a and b of the noise (3).

2.a. Calculate a variance stabilizing transformation $f : [0, 1] \to \mathbb{R}$, such that $\operatorname{std}\{f(\tilde z_i(x))\} \simeq c$, where $c > 0$ is a fixed constant which does not depend on $y_i(x)$.

2.b. Apply f to $\tilde z_i(x)$, $\forall x \in X$ and $i = 1, \ldots, N$, and thus obtain a sequence $\{f(\tilde z_i)\}_{i=1}^N$ with approximately constant variance equal to c^2.

3. Filter the sequence $\{f(\tilde z_i)\}_{i=1}^N$ using V-BM3D video denoising algorithm for AWGN (homoskedastic filtering). The denoised output of the algorithm is a sequence denoted as

$$\left\{\mathbf{D}_{\text{ho}}(f(\tilde z_i))\right\}_{i=1}^N \triangleq \mathbf{VBM3D}\left(\{f(\tilde z_i)\}_{i=1}^N\right).$$

4.a. Calculate the estimation bias due to the nonlinearity of f as the function $h : \left[0, f(1)\right] \to \left[0, f(1)\right]$ implicitly defined by

$$f(E\{\tilde{z}_i\}) \xmapsto{h} E\{f(\tilde{z}_i)\} = h(f(E\{\tilde{z}_i\}))\,. \tag{6}$$

4.b. Apply f inverse to the debiased $h^{-1}\big(\mathbf{D}_{\text{ho}}(f(\tilde{z}_i))\big)$, obtaining an estimate of $\{\tilde{y}_i\}_{i=1}^{N}$.

5. Compensate the bias due to clipping by applying the transformation $\mathcal{C} : \tilde{y}_i \longmapsto y_i$ [8, 6] to the debiased $f^{-1}\big(h^{-1}\big(\mathbf{D}_{\text{ho}}(f(\tilde{z}_i))\big)\big)$.

Thus, the estimate $\left\{\hat{y}_i\right\}_{i=1}^{N}$ of $\{y_i\}_{i=1}^{N}$can be expressed as

$$\hat{y}_i = \mathcal{C}\left(f^{-1}\big(h^{-1}\big(\mathbf{D}_{\text{ho}}(f(\tilde{z}_i))\big)\big)\right). \tag{7}$$

Let us comment and give additional details on these various steps.

3.1 Noise estimation

In [8], is presented an algorithm for automatic estimation of the parameters a and b of the clipped signal-dependent noise model (3) from a single noisy image. We use this algorithm as the very first step when processing sequences of raw-data images. Remember that, as stated in Section 2.1, these two parameters (and thus the standard-deviation function σ) are independent of the particular frame index i, as they are influenced only by acquisition parameters, which are assumed to be fixed while the different images are captured. Thus, the parameters a and b need to be estimated only once, and the estimation can be carried out on any of the frames $\tilde{z}_i$ (e.g., $\tilde{z}_1$) or on a mosaic composed by tiling many frames, one next to the other. The latter solution can improve the noise estimation especially for small images, where there could be not enough pixels within a single image for accurately estimating the noise parameters. For the experiments presented in this paper we always estimate the noise parameters from the first image $\tilde{z}_1$ alone.

3.2 Variance stabilization

The variance-stabilizing transformation $f : [0,1] \to \mathbb{R}$ used in this work is the standard indefinite integral

$$f(t) = \int_{t_0}^{t} \frac{c}{\tilde{\sigma}(s)}\,ds, \qquad t, t_0 \in [0,1]\,. \tag{8}$$

Because of its simplicity, this classical transformation appears frequently in many works on mathematical statistics (e.g., [3] and references therein) and signal processing (e.g., [12], [1], [9], [11]). One of the main theoretical results proved in [6] is the fact that the indefinite integral (8) is actually bounded for the function $\tilde{\sigma}$ corresponding to the raw-data model (5). Of course, the resulting f is always nonlinear, because $\tilde{\sigma}$ cannot be constant.

3.3 Denoising

As discussed in [8] and [6], transform-domain algorithms where the basis functions have supports with sufficient number of non-zero samples (e.g., 4×4 pixel blocks), are naturally suited for filtering variance-stabilized clipped observations. This is because the distribution of the transform coefficients turns out to be essentially a Gaussian with fixed standard-deviation equal to c. The fact that the distribution approaches a Gaussian is a direct consequence of the central-limit theorem and was illustrated in [8], for Daubechies wavelets, and in [6], for the block-DCT transforms. The fixed standard-deviation is due to the variance-stabilization.

For the particular case of BM3D and V-BM3D algorithms (and of many other block-based nonlocal algorithms) it is worth to briefly discuss about the ℓ^2-norm of blockwise differences of the noisy data, which is used to estimate the block similarity for the matching. The variance and the mean of this similarity estimate are actually dependent not only on the variance, but also on the actual distributions of the noisy data (the Gaussian case is studied in [4]). Nevertheless, in these algorithms, the thresholds used for the matching are typically selected from deterministic speculations about the suitable value of the blockwise difference, mainly ignoring the statistical characteristics of the noisy components [4]. Moreover, during the second stage of the algorithms (with collaborative empirical Wiener filtering), the matching is not performed any longer on the noisy data, but on the image/sequence estimate obtained from the first stage (with collaborative hard-thresholding), which can be practically considered as a noise-free image. Finally, contrary to nonlocal algorithms based on averaging, the collaborative filtering is quite robust against possible errors in the matching, because shrinkage can preserve large inter-block dissimilarities [4],[10]. Thus, the non-Gaussianity of the data does not constitute an impairment to a satisfactory block-matching.

Let us emphasize the concrete meaning of collaborative filtering. First, each filtered frame $\mathbf{D}_{\text{ho}}(f(\tilde{z}_i))$, $i \in \{1,\ldots,N\}$, is obtained exploiting mutually similar blocks; these blocks can be taken from all frames of the noisy sequence $\{f(\tilde{z}_i)\}_{i=1}^{N}$ and more than one block can be taken from each frame. Second and more important, while producing an estimate for a block in a particular frame, we are also producing individual estimates for all mutually similar blocks (including blocks from other frames) which are used in estimation.

The output of the denoising filter is an estimate of the (conditional) expectation of the input:

$$\mathbf{D}_{\text{ho}}(f(\tilde{z}_i(x))) \approx E\{f(\tilde{z}_i(x))\}, \quad \forall x \in X,\ i = 1,\ldots,N.$$

3.4 Inversion of the variance-stabilizing transformation

Because of the nonlinearity of f we have that

$$E\{f(\tilde{z}_i(x))\} \neq f(E\{\tilde{z}_i(x)\}) \tag{9}$$

Therefore, f cannot be inverted right after denoising. The nonlinear function h (6) takes care of compensating the discrepancy (9) between the expectation of the transformed data and the transformed expectation of the data. Thus, we have that

$$h^{-1}\big(\mathbf{D}_{\text{ho}}(f(\tilde{z}_i(x)))\big) \approx f(E\{\tilde{z}_i(x)\}), \tag{10}$$

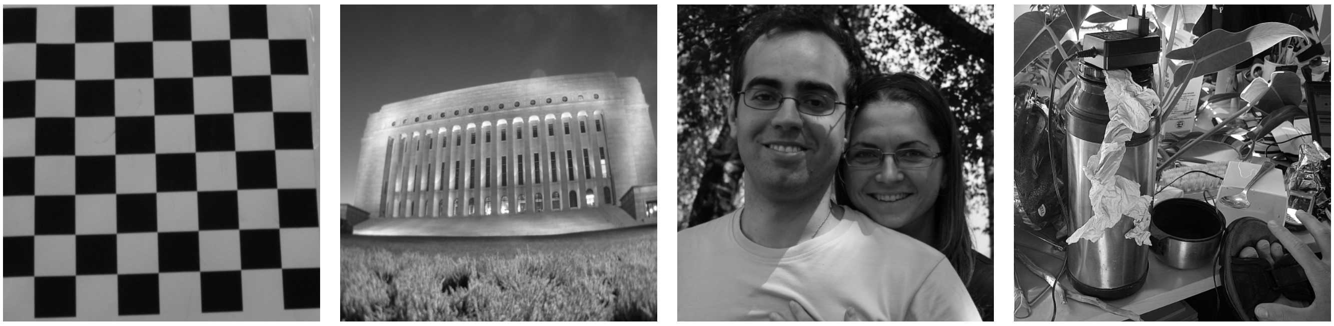
and hence that

$$f^{-1}\big(h^{-1}\big(\mathbf{D}_{\text{ho}}(f(\tilde{z}_i(x)))\big)\big) \approx E\{\tilde{z}_i(x)\}. \tag{11}$$

Let us observe that, in most image processing algorithms exploiting variance-stabilization in the integral form (8), the role of h is neglected and the two terms in (9) are mistakenly assumed as equal (see e.g., [12], [1], [9], [11]). However, the compensation (10) turns out to be crucial, particularly when dealing with asymmetric distributions such as (4).

3.5 Declipping

After (11), we have a sequence of estimates of $E\{\tilde{z}_i\} = \tilde{y}_i$, $i = 1,\ldots,N$. However, as declared in Section 2.1, our goal is to estimate the non-clipped sequence $\{y_i\}_{i=1}^{N}$ (2). Thus, we wish to invert the bias due to clipping by applying on the left-hand side of (11) a transformation $\mathcal{C} : \tilde{y}_i \longmapsto y_i$, as in (7). The analytical expression of the transformation $\mathcal{C}$ was derived in [8] and the effectiveness of its use for debiasing denoised estimates of clipped images was shown in [7] and [6]. While the range of $\tilde{y}_i$ is at most the interval [0, 1], the range Y of y_i can be much wider. This typically results in flattened out portions of the image due to saturation of the imaging device (overexposure). The essential impact of the transformation (7) is indeed the potential increase of the range of the denoised estimates, up to reaching the whole Y. We refer to [6] for the analysis of the increase of the image range. Roughly speaking, in the case of image sequences, a result of this analysis could be stated as follows: provided that the original sequence $\{y_i\}_{i=1}^{N}$, is sufficiently smooth (in either space or time), the larger is the noise variance σ, the wider can be the range of the reconstructed sequence $\left\{\hat{y}_i\right\}_{i=1}^{N}$.



Checkerboard *Eduskuntatalo* *Luca & Tania* *Mess*

Figure 2: The first frames y_1 of the four sequences used for the experiments with synthetic noise.

Figure 3: The five noisy frames $\tilde{z}_1, \tilde{z}_2, \tilde{z}_3, \tilde{z}_4$, and $\tilde{z}_5$ of the "shaked" sequence *Mess*. Noise parameters: $a = 1/200$, $b = (10/255)^2$.

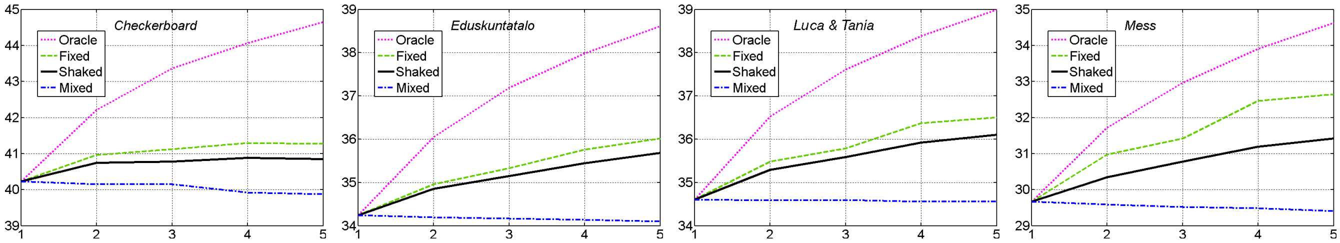


Figure 4: PSNR versus the number M of frames used in V-BM3D. Noise parameters: $a = 1/200$, $b = (10/255)^2$.

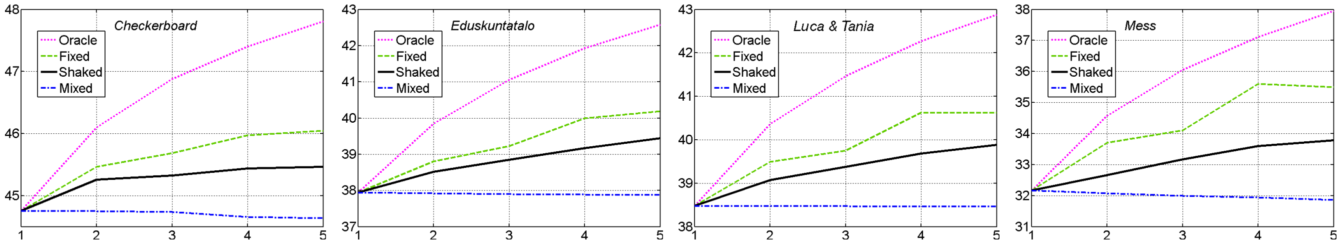


Figure 5: PSNR versus the number M of frames used in V-BM3D. Noise parameters: $a = 1/800$, $b = (5/255)^2$.

4. EXPERIMENTS

The proposed denoising procedure has been tested on sequences of images corrupted by synthetic noise, as well as on real raw-data image sequences acquired by a digital camera.

In all the experiments, the maximum number of blocks in a group was set to 16 for the hard-thresholding step, and to 32 for the Wiener filtering step. In both steps, the 3-D transform used for the collaborative filtering is the separable composition of the 2-D 8×8 DCT and 1-D Haar wavelet transforms.

We use sequences of five frames, where, progressively, only the subsequence composed of the first $M = 1, \ldots, 5$ frames is processed by the algorithm. The following four scenarios are considered:

- "fixed": all noise-free images $y_1 = y_i$, $i = 2, \ldots, M$, are identical. This corresponds to the ideal scenario where the camera is fixed and the scene is static.
- "shaked": all frames portray the same scene, but the images y_i, $i = 1, \ldots, M$, are slightly shifted, rotated, or enlarged, with respect to each other. This corresponds to images taken with a hand-held camera, where the shake causes small changes in the camera position.
- "mixed": each frame in the sequence $\{y_i\}_{i=1}^M$ portrays a completely different scene. This is an extreme case where the noise free images have nothing in common and thus the denoising procedure cannot exploit redundancy between frames.
- "oracle": the sequence $\{y_i\}_{i=1}^M$ is as in "fixed". Instead of applying V-BM3D on the sequence $\{f(\tilde{z}_i)\}_{i=1}^M$, we apply BM3D on the average $\frac{1}{M}\sum_{i=1}^M f(\tilde{z}_i)$ assuming standard-deviation $c/\sqrt{M}$. In this way, the fact that the underlying images are identical is assumed as known and it is exploited in the best

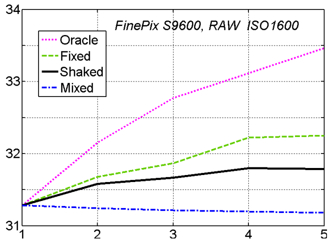


Figure 8: PSNR versus the number M of frames used in V-BM3D for the experiment with raw data from digital camera.

possible way. In a sense, this constitutes an upper bound to the performance achievable in all other cases.

We compare the denoising performance under these four scenarios, with the first frame $\tilde{z}_1$ (and y_1) being the same for the various sequences corresponding to a particular scene and noise. The results (figures and PSNR) are given for the estimate $\hat{y}_1$ of the first frame only.

4.1 Synthetic noise

The noise-free images y_i are generated by downsampling high-quality high-resolution images to the size 512×512. The frames of the "shaked" sequence are obtained by first applying some image transformations (including minor translations, rotations, scale modifications, cropping) to the high-resolution image, with the aim to replicate the effect of a handheld camera movement, and then downsampling. From these y_i, we simulate raw-data as the clipped noisy observations $\tilde{z}_i$ (5) with the noise term composed of two mutually independent parts, a *Poissonian* signal-dependent component η_{p} and a *Gaussian* signal-independent component η_{g}: $\sigma(y_i(x))\xi_i(x) = \eta_{\mathrm{p}}(y_i(x)) + \eta_{\mathrm{g}}(x)$ [8]. In particular, $\left(y_i(x) + \eta_{\mathrm{p}}(y_i(x))\right)\chi \sim \mathcal{P}(\chi y_i(x))$, where $\chi > 0$ and $\mathcal{P}$ denotes the Poisson distribution, and $\eta_{\mathrm{g}}(x) \sim \mathcal{N}(0, b)$. Thus, $\sigma^2(y_i(x)) = a y_i(x) + b$, with $a = \chi^{-1}$. As discussed in Section 2.1, the algorithm treats $\xi_i(x)$ as a standard normal variate. In these experiments we use $a = 1/200$, $b = (10/255)^2$ and (halving the standard-deviation σ) $a = 1/800$, $b = (5/255)^2$. Four scenes are used: *Checkerboard*, *Eduskuntatalo*, *Luca & Tania*, and *Mess*. The corresponding frames y_1 are shown in Figure 2. As an illustration of the kind of frame displacements existing in the "shaked" sequences, in Figure 3 we show all the five frames $\tilde{z}_i$, $i = 1, \ldots, 5$, of the "shaked" sequence *Mess* with $a = 1/200$, $b = (10/255)^2$.

The plots of Figures 4 and 5 give the PSNR of the restoration of the first frame in each sequence as function of the number M of frames involved in the denoising. The plots show that the restoration always improves significantly, provided that there is some degree of similarity between the frames in the sequence. In particular, one can observe that the gain of "fixed" over "shaked" is more pronounced for more complex scenes (such as *Mess*). A visual comparison for an enlarged fragment of *Mess* is shown in Figure 6. It can be observed that the central fragments of the second and third rows are visibly smoother than the respective right fragments of the same rows, for which the V-BM3D can recover additional details of the original image by exploiting the redundancy characterizing the "fixed" and "shaked" sequences. We also note the marginal decrease in PSNR for the "mixed" case: this follows from grouping together blocks from images of different scenes, which thus do not bring useful information about y_i. The results obtained by the "oracle" are much better than those corresponding to the other three scenarios.

4.2 Raw data from digital camera

The proposed multiframe denoising procedure has been tested also on raw data from a Fujifilm FinePix S9600 digital camera. The four scenarios have been realized by taking various shots of a printed poster as follows. With the remotely-controlled camera fixed on a tripod, we first acquired a few long-exposure images at ISO 100 and the five of short-exposure images at ISO 1600 for the "fixed" sequence. The long-exposure images have been averaged together to generate a ground-truth[1] y_1 to be used as reference for computing the PSNR of the estimates $\hat{y}_1$. After detaching the tripod, four more images were captured at ISO 1600, with the camera held in the hands. These four images, together with the first frame $\tilde{z}_1$ of the "fixed" sequence constitute the "shaked" sequence. Figure 7 shows the first noisy frame and two V-BM3D estimates, for M=1 and M=5. The PSNR curves for the four scenarios are given in Figure 8, showing a behavior consistent with the results obtained with synthetic noise.

4.3 Multiframe denoising vs. restoration from blurred-noisy image pairs

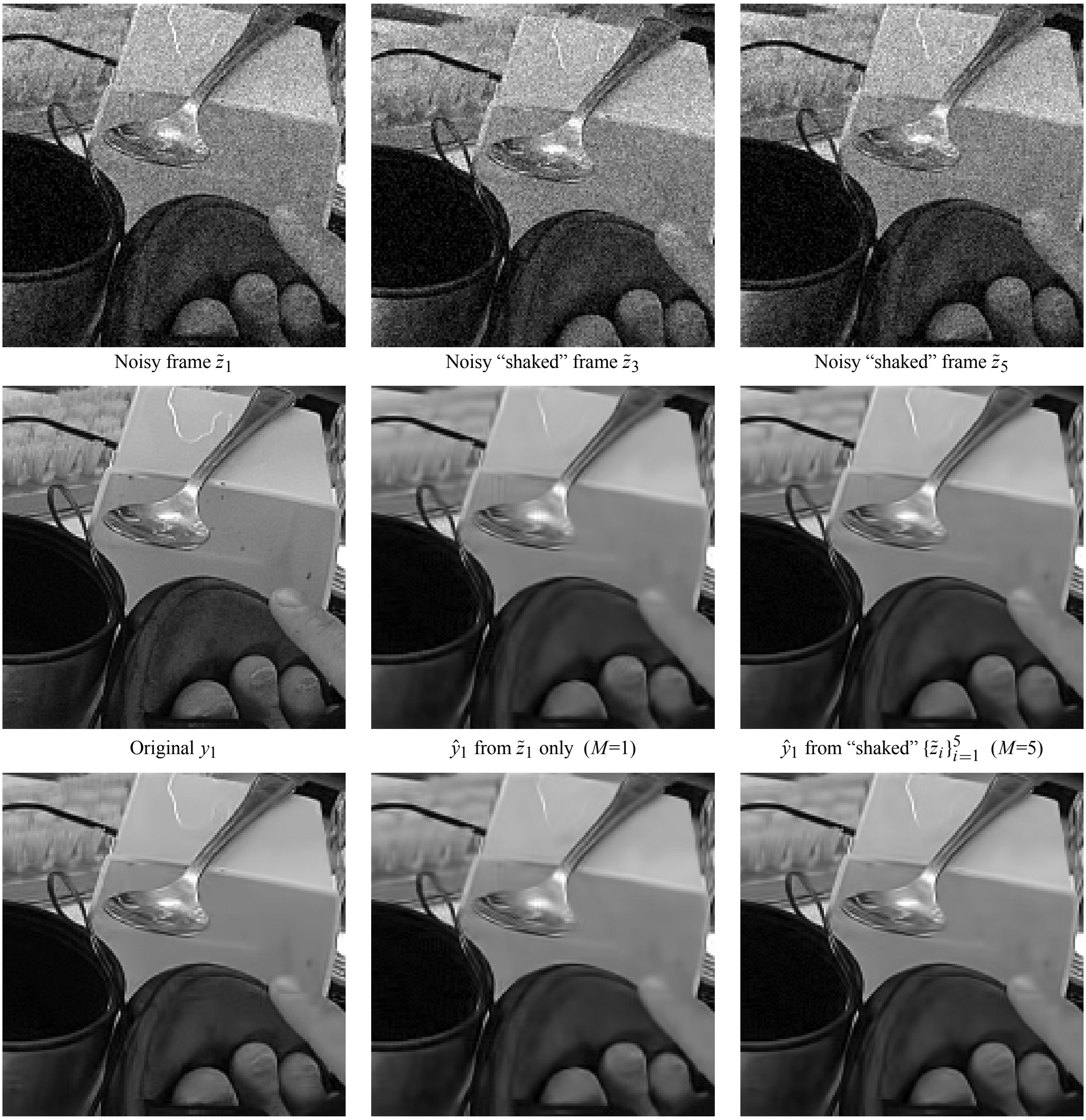
As an additional element of our experimental analysis, we wish to compare our multiframe denoising against the image restoration approach based on differently exposed blurred-noisy image pairs. In particular, in Figure 9, we provide a visual comparison against the algorithm by Tico et al. [14],[13] for pairs of raw-data images acquired with a Canon EOS 350D camera. A pair of short-exposure images is used as the input of our multiframe algorithm. Enlarged fragments of the estimates are shown in Figure 10. Even though the V-BM3D estimate appears much sharper and without artifacts, there are some details, such as the text "kirjasto" in one of the fragments, that cannot be recovered out of the short-exposure images, because far too low SNR in the input data. Despite severe artifacts, these details are indeed visible in the estimate obtained from the blurred-noisy pair. We wish to note that the used implementation of the algorithm [14],[13] was preliminary and did not fully exploit the considered noise model.

5. DISCUSSION AND CONCLUSIONS

The experiments on synthetic noise, as well as those on raw-data from digital camera, show that the denoising performance increases considerably when, in V-BM3D, the search for similar blocks spans different frames of the same scene. Such improvement is particularly interesting in the case of the "shaked" sequences, where the V-BM3D manages to exploit the scene redundancy also when the frames are acquired with small variations in the camera position. This makes the acquisition of multiple noisy frames and their joint denoising, an effective and practical solution for obtaining a single enhanced image in low-light conditions. The experiment presented in Section 4.3 shows that a direct comparison, between the restoration performance of multiframe denoising and that of blurred-noisy image pairs methods, is not straightforward. In fact, while effective deblurring-based methods still produce heavy artifacts and works restrictively to shift-invariant blur, they are potentially able to restore finer details of the scene that are not recoverable from the noisy images alone. It should be noticed that the proposed multiframe denoising algorithm is able to process indifferently both images acquired during camera shake and images of scenes with objects in relative motion. If these were to be captured with a prolonged exposure, they would lead to blur PSFs that may be assumed as shift-invariant for the former but definitely not for the latter, resulting in blurred images that are much more difficult to handle and to restore.

The marginal decrease in the PSNR for the "mixed" case follows from grouping together blocks from images of different scenes. This raises an interesting point of discussion. Loosely speaking, in the algorithm we consider blocks to be *similar* provided that the ℓ^2-norm of the blockwise differences (measured on the noisy sequence or on an intermediate denoised sequence) falls

[1] Note that this "ground truth" is not really as such, partly because of the fixed-pattern noise component, which cannot be suppressed by averaging, and partly because of the non-exactly linear response of the sensor. Thus, the computed PSNR values provide only a very rough indication of the actual goodness of the estimate.



Noisy frame $\tilde{z}_1$ | Noisy "shaked" frame $\tilde{z}_3$ | Noisy "shaked" frame $\tilde{z}_5$

Original y_1 | $\hat{y}_1$ from $\tilde{z}_1$ only (M=1) | $\hat{y}_1$ from "shaked" $\{\tilde{z}_i\}_{i=1}^5$ (M=5)

Oracle $\hat{y}_1$ from "fixed" $\sum_{i=1}^5 \tilde{z}_i/5$ | $\hat{y}_1$ from "mixed" $\{\tilde{z}_i\}_{i=1}^5$ | $\hat{y}_1$ from "fixed" $\{\tilde{z}_i\}_{i=1}^5$ (M=5)

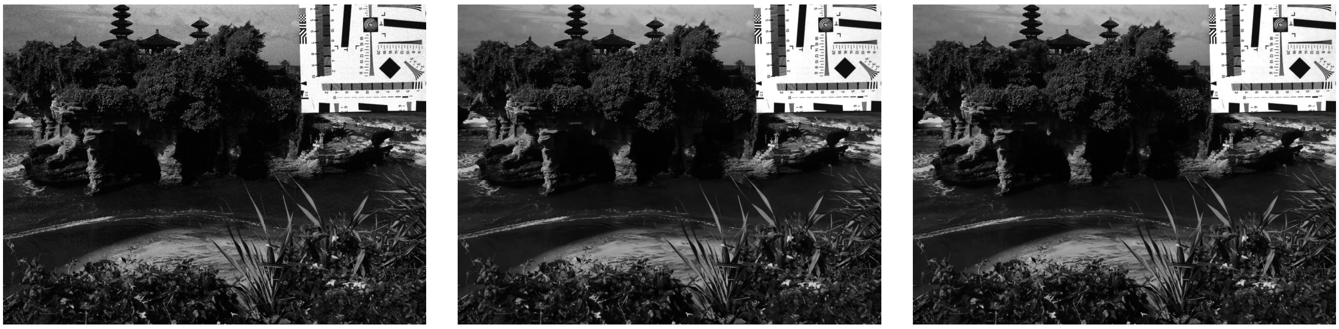
Figure 6: Enlarged fragments corresponding to the *Mess* scene. Noise parameters: $a = 1/200$, $b = (10/255)^2$.

below a certain threshold; it turns out that the actual blockwise similarity is likely to be much higher when the matched blocks are from the same scene, at least for what concerns finer details that are difficult to recover. This is a delicate aspect that in this work we touch only superficially but that definitely deserves some future attention. Firstly, it exposes the importance and strength of the self-similarity prior within a natural image. Second, it suggests the idea that even very complex natural images contain blocks only from few, narrow submanifolds of the space of patches.

Finally, it should not surprise that the results obtained by the "oracle" are far better than those corresponding to the other three scenarios, including the "fixed" one. Since the input sequence for "oracle" and the "fixed" is exactly the same, the current gap between their respective results leaves plenty of room for potential improvement of the algorithm, and particularly of its block-matching (grouping) step.

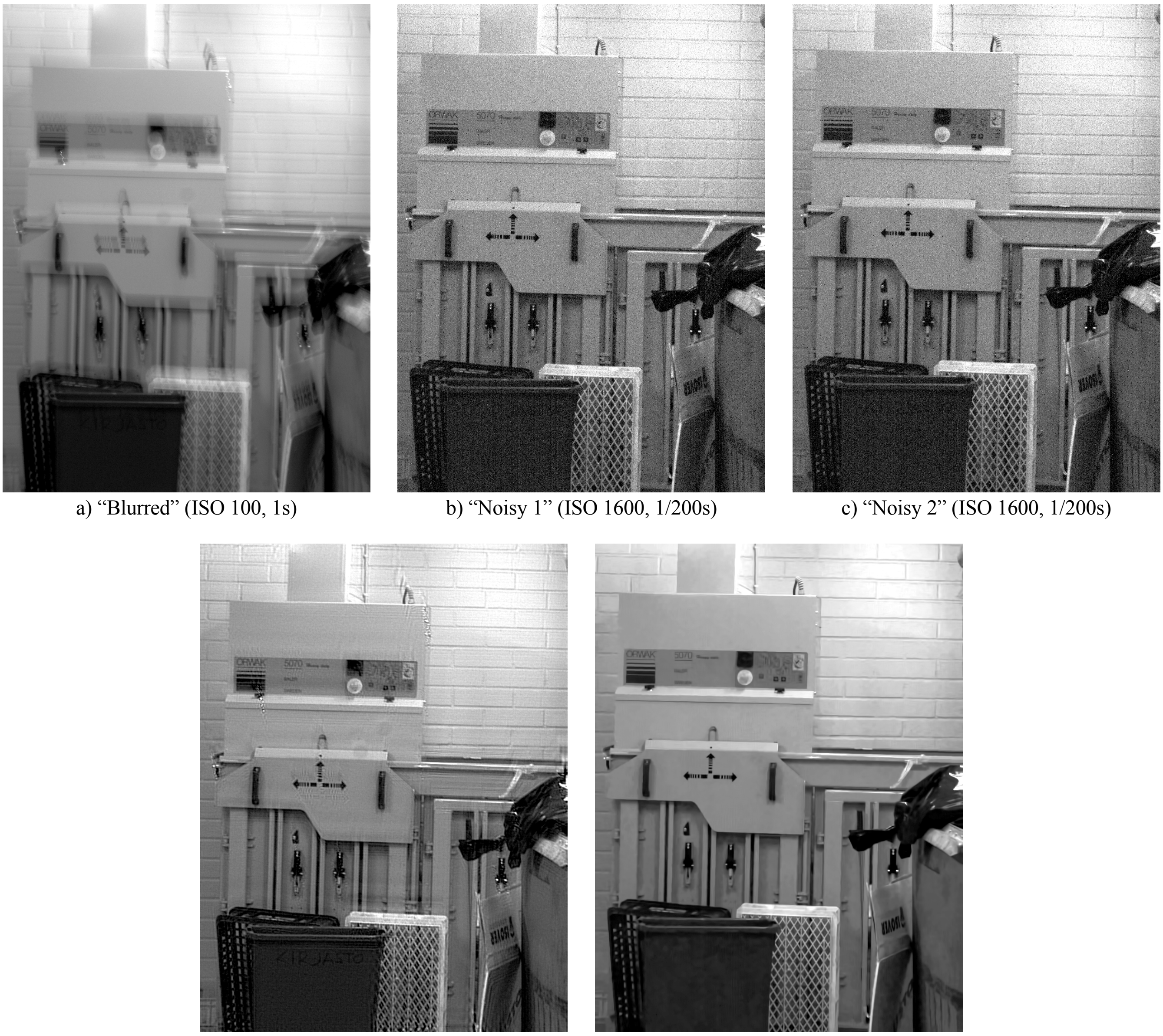
6. ACKNOWLEDGMENT

We wish to thank Marius Tico, for providing us the estimate from the differently exposed images, and Kostadin Dabov, for his help with the V-BM3D libraries.



Raw data frame $\tilde{z}_1$ $\hat{y}_1$ from $\tilde{z}_1$ only $\hat{y}_1$ from “shaked” $\{\tilde{z}_i\}_{i=1}^{5}$

Figure 7: Denoising of the “shaked” raw-data sequence: the first raw-data frame $\tilde{z}_1$ of the sequence and the corresponding V-BM3D estimates $\hat{y}_1$ for $M = 1$ and for $M = 5$.



a) “Blurred” (ISO 100, 1s) b) “Noisy 1” (ISO 1600, 1/200s) c) “Noisy 2” (ISO 1600, 1/200s)

d) “Blurred&noisy”-pair estimate [14],[13] e) V-BM3D “noisy&noisy”-pair estimate

Figure 9: Restoration from image pairs. a) image acquired with low ISO and long exposure, practically noise free but with severe blur due to camera shake; b) image taken with short exposure time and high ISO, significantly noisy, but without blur; c) same as b) but after a small displacement of the camera; d) restored image obtained by the algorithm [14],[13] using the “blurred&noisy” image pair a)+b); e) V-BM3D estimate obtained using the “noisy&noisy” image pair b)+c).

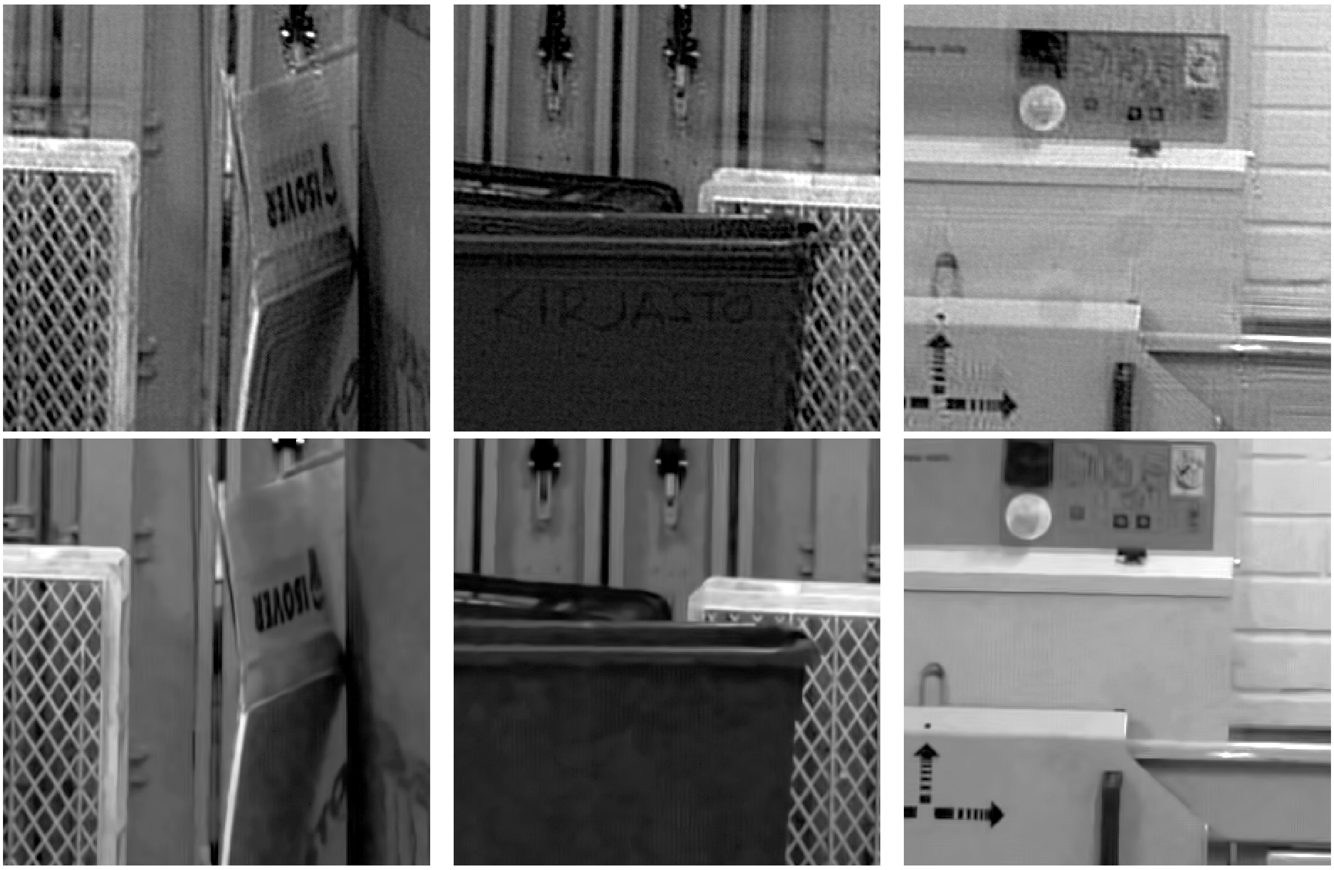


Figure 10: Enlarged fragments from the estimates d) (top) and e) (bottom) of Figure 9.

REFERENCES

[1] Arsenault, H.H., and M. Denis, "Integral expression for transforming signal-dependent noise into signal-independent noise," *Opt. Lett.* vol. 6, no. 5, pp. 210-212, May 1981.

[2] Cohen, A., *Truncated and censored samples*, CRC Press, 1991.

[3] Curtiss, J.H., "On transformations used in the analysis of variance", *The Annals of Mathematical Statistics*, vol. 14, no. 2, pp. 107-122, June 1943.

[4] Dabov, K., A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering", *IEEE Trans. Image Process.*, vol. 16, no. 8, Aug. 2007.

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

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

[7] Foi, A., *Pointwise Shape-Adaptive DCT Image Filtering and Signal-Dependent Noise Estimation*, D.Sc.Tech. Thesis, Institute of Signal Processing, Tampere University of Technology, Publication 710, December 2007.

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

[9] Kasturi, R., J.F. Walkup, and T.F. Krile, "Image restoration by transformation of signal-dependent noise to signal-independent noise," *Applied Optics*, vol. 22, no. 22, pp. 3537-3542, November 1983.

[10] Katkovnik, K., A. Foi, K. Egiazarian, and J. Astola, "Nonparametric regression in imaging: from local kernel to multiple-model nonlocal collaborative filtering," in *Proc. 2008 Int. Workshop on Local and Non-Local Approximation in Image Processing, LNLA 2008*, Lausanne, Switzerland, August 2008 (this volume).

[11] Hirakawa, K., and T.W. Parks, "Image denoising using total least squares", *IEEE Trans. Image Process.*, vol. 15, no. 9, pp. 2730-2742, Sept. 2006.

[12] Prucnal, P.R., and B.E.A. Saleh, "Transformation of image-signal-dependent noise into image signal-independent noise", *Optics Letters*, vol. 6, no. 7, July 1981.

[13] Tico, M., and M. Trimeche, "Motion blur identification based on differently exposed images," *Proc. 2006 IEEE Int. Conf. Image Process., ICIP 2006*, Atlanta, GA, USA, pp. 2021-2024, Oct. 2006.

[14] 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.

[15] Tico, M., and M. Vehvilainen, "Image stabilization based on fusing the visual information in differently exposed images," in *Proc. Int. Conf. Image Process., ICIP 2007*, San Antonio, TX, USA, September 2007.

[16] Tico, M., and M. Vehvilainen, "Motion deblurring based on fusing differently exposed images," in *Proc. SPIE Digital Photography III*, vol. 6502, 65020V, San Jose, CA, USA, 2007.

[17] 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.

[18] Yuan, L., J. Sun, L. Quan, and H.-Y. Shum, "Blurred/no-blurred image alignment using kernel sparseness prior," *Proc. Int. Conf. Computer Vision (ICCV)*, Rio de Janeiro, Brazil, October 2007.