A NOTE ON MULTI-IMAGE DENOISING

Toni Buades,[1,4]*, Yifei Lou*[2]*, J.M. Morel*[3,1] *and Zhongwei Tang*[3]

[1] Universitat de les Illes Balears, Departament de Matematiques e Informatica, Spain
[2] Department of Mathematics, University of California Los Angeles, USA
[3] Centre de Mathématiques et de Leurs Applications, Ecole Normale Supérieure de Cachan, France
[4] MAP5, Université René Descartes, France.

ABSTRACT

Taking photographs under low light conditions with a hand-held camera is problematic. A long exposure time can cause motion blur due to the camera shaking and a short exposure time gives a noisy image. We consider the new technical possibility offered by cameras that take image bursts. Each image of the burst is sharp but noisy. In this preliminary investigation, we explore a strategy to efficiently denoise multi-images or video. The proposed algorithm is a complex image processing chain involving accurate registration, video equalization, noise estimation and the use of state-of-the-art denoising methods. Yet, we show that this complex chain may become risk free thanks to a key feature: the noise model can be estimated accurately from the image burst. Preliminary tests will be presented. On the technical side, the method can already be used to estimate a non parametric camera noise model from any image burst.

1. INTRODUCTION

It is a frustrating experience, even for professional photographers, to take pictures under bad lighting conditions with hand-held camera. If the camera is set to a long exposure time, the photograph gets motion blur. If it is taken with short exposure, the image is noisy. This dilemma can be solved by taking a burst of images, each with short-exposure time, as shown in Fig. 1. But then, as classical in video processing, an accurate registration technique is required to align the images. Denote by $u(x)$ the ideal non noisy image color at a pixel x. Such an image can be obtained from a still scene by a camera in a fixed position with a long exposure time. The observed value for a short exposure time τ is a random Poisson variable with mean $\tau u(x)$ and standard deviation proportional to $\sqrt{\tau u(x)}$. Thus, the SNR increases with the exposure time proportionally to $\sqrt{\tau}$. The core idea of the *burst denoising* method, which we sketch in these notes, is a slight extension of the same law. The only assumption is that the various values at a cross-registered pixel obtained by a burst are i.i.d.. Thus, averaging the registered images amounts to averaging several realizations of these random variables. An easy calculation shows that this increases the SNR by a factor proportional to $\sqrt{n}$, where n is the number of shots in the burst. (We call SNR of a given pixel the ratio of its temporal variance to its temporal mean). Fig. 1 summarizes the possibilities offered by an image burst. A long exposure image is exposed to motion blur. The short exposure image is noisy, but sharp. Finally, the image obtained by averaging the images of the burst after registration is both sharp and noiseless. In this real example the burst taken in a gallery had 16 images. The noise should therefore be divided 4.

The idea of combining multiple images to get a desired one is called image fusion. Most recent works on fusion use a pair of pictures taken with different camera parameters. Liu et al. [1] combine a blurred image with long-exposure time, and a noisy one with short-exposure time for the purpose of both denoising the second and deblurring the first. In another direction, Beltramio and Levine [2] improve the dynamic range of the final image by combining an underexposed snapshot with an overexposed one. Combining again two snapshots, one with and the other without flash, is investigated by Eisemann *et. al.* [3] and Fattal *et. al* [4]. Both papers report spectacular results. In contrast, we shall only consider classic image bursts, taken with the very same camera parameters. The number of images ranges from 9 to 36, thus promising a division of the noise by 3, 4 or 6. As is apparent in the above numbers, the denoising power of *burst denoising* is eventually hemmed by the low growth of the square root. On the other hand, dividing the noise by the mentioned factors and getting an artifact free image is in no way a negligible ambition. Indeed, even the best state of the art denoising methods can create slightly annoying artifacts, such as adhesion effects and shocks in NL-means [5] or patterns in the transform thresholding methods [6], [7]. Simple accumulation instead is the essence of photography. The first Nicephore Niepce photograph [8] was obtained after a several hours exposure. The only objection to long exposure is the variation of the scene. The more this variation can be compensated, the longer the exposure can be.

There is a strong argument in favor of denoising by simple averaging of the registered samples instead of block-matching strategies. If a fine non-periodic texture is present in an image, it is virtually indistinguishable from noise, and actually contains a flat spectrum part which has the same Fourier spectrum as the white noise. Such fine tex-

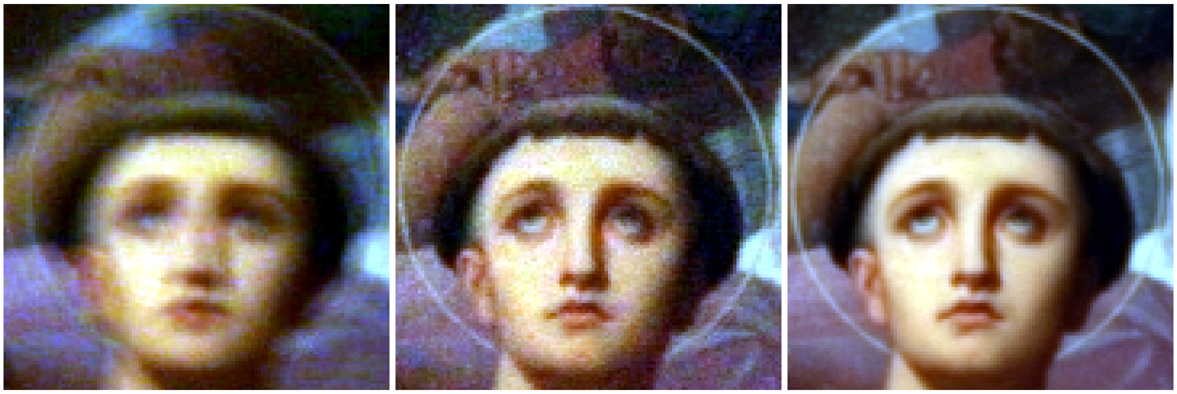

Figure 1. From left to right: one long-exposure image (time = 0.4 sec, ISO=100), one of 16 short-exposure images (time = 1/40 sec, ISO = 1600) and the average after registration. All images have been color-balanced to show the same contrast. The long exposure image is blurry due to camera motion. The middle short-exposure image is noisy, and the third one is some 4 times less noisy, being the result of averaging 16 short-exposure images.

tures can be distinguished from noise only if several samples of the same texture are present in other frames and can be accurately registered. Now, state of the art denoising methods are based on nonlocal block matching. In the case of a burst, the block matching would ideally find only one block in each image. But it doesn't. Precisely because of the noise, low contrasted textures are at risk of being mismatched across frames. The experimental section will show that this can cause a loss of resolution for such textures. A registration process more global than block matching, using strong features elsewhere in the image, should permit a safer denoising by accumulation.

Yet, this method rises serious technical objections. The main technical objection is: how to register globally the images of a burst? Fortunately, there are several situations where the series of snapshots are indeed related to each other by a homography, and we shall explore these situations first. The homography assumption is actually valid if one of the assumptions is satisfied:

1. the only motion of the camera is an arbitrary rotation around its optic center;
2. the photographed objects share the same plane in the 3D scene;
3. the whole scene is far away from the camera.

In those cases, image registration is equivalent to computing the underlying image homography. But this registration should be sub-pixel accurate. To this aim we will introduce a precise variant of SIFT [9] and a generalization of ORSA (Optimized Random Sampling Algorithm, [10]) to register all the images together.

Yet, in general, the images of 3D scene are **not** related by a homography, but by an epipolar geometry [11]. Even if the camera is well-calibrated, 3D point-to-point correspondence is impossible to obtain without knowing the depth of the 3D scene. Therefore, we should not expect that a simple homography will work everywhere in the image, but only on a significant part. On this part, we shall say that we have a dominant homography. [1]

To go further, we shall need several tools whose list follows. The main one is the accurate estimation of the noise model from a partial registration.

- **High accurate keypoint detection:** By canceling the subsampling in SIFT, a subpixel precision of the key point detection will be reached. As a result, the dominant homography will be computed accurately from the matching points.
- **Noise estimation:** At each pixel that is well-registered, the registered samples are i.i.d. samples of the same underlying Poisson model. As a result, a signal dependent noise model will be accurately estimated for each colour channel. This model simply is a curve of image intensity versus the standard deviation of the noise.
- **Color equalization** However, the noise estimation will require an extra step, the histogram equalization of all images. Indeed, the images taken with indoor lights often show fast variations of the contrast and brightness. It is only after this equalization that the empirical standard deviation of the samples becomes a measurement of the noise standard deviation.
- **Hybrid denoising scheme:** Averaging does not work at the mis-registered pixels, and block matching methods are at risk on the fine image structures. Thus they will be combined. The simple combination used here will be a convex combination of them, the weight function being based on the noise curve and on the observed standard deviation of the values for the accumulation at a certain pixel. If this standard deviation is compatible with the noise model,

[1] Needless to be said, an accurate camera calibration correcting the optical distortion could also play a role, particularly on the image borders.

the denoised value will be the mean of the samples. Otherwise, the standard deviation test will imply that the registration at this point is inaccurate, and a conservative denoising will be preferred at the pixel. (More prudently, the denoised value will be a weighted average of both denoised values, the weights being steered by the test.)

References and preliminaries on the used techniques are given in Sec. 2. The tentative algorithm is described in Sec. 3 including how to register all the images, estimate the noise and combine two denoising schemes. Experiments on various kinds of real data sets are examined in Sec. 4.

2. PRELIMINARIES, ANTERIOR WORKS

2.1. Image matching

To find key points in images and match them is a fundamental step for many computer vision and image processing applications. One of the most robust is the Scale Invariant Feature Transform (SIFT) [9]. There are other attempts to match key points in a more invariant fashion [12, 13, 14, 15, 16, 17]. Applications of image matching include scene parsing[18], object/image retrieval [19] and motion estimation [20]. The image stitching [21, 22] generates a panorama from several images of the same landscape. The underlying technical problems are basically the same as for the burst denoising problem. In particular, the registration accuracy is a key issue in image stitching. In [21], bundle adjustment is used to minimize the homography projection error. This technique requires a knowledge of the camera internal parameters for initialization. Because wrong matches occur in the SIFT method used here for the registration, an accurate estimate of the dominant homography will require the elimination of outliers. The standard method to eliminate outliers is RANSAC (RANdom SAmple Consensus) [23]. However, it is efficient only when outliers are a small portion of the whole matching set. There are other variants of RANSAC to improve the performance of outlier elimination and the estimation of fundamental matrix, such as [24, 25, 26, 27]. We chose the method based on a contrario model proposed by Moisan and Stival [10]. It has zero parameter and is effective even the matching set contains up to 90% of outliers.

2.2. Noise Estimation

Most computer vision algorithms should adjust their parameters to the image noise level. Surprisingly, there are few papers dealing with the noise estimation problem and most of them only estimate a signal-independent noise. The standard procedure is the following: (1) compute the mean and standard deviation for each $N \times N$ block in the image (N is small, e.g. $N = 3$ or $N = 5$); (2) classify the standard deviations according to their mean, and (3) take the median value of all standard deviations for each mean. Instead of computing the variance of patches, Olsen [28] and posteriorly Rank *et. al.* [29] consider the patches of the image derivative, since it is more robust to the noise. As a variant, Donoho *et. al.* [6] proposed to estimate the noise standard deviation as the median of absolute values of wavelet coefficients at the finest scale. All the algorithms mentioned above usually give a reasonable estimation of the standard deviation when the noise is uniform. Yet, when applying these algorithms to estimate signal dependent noise, the results are poor. An exception is the work of C. Liu *et. al.* [30], which estimates the upper bound on the noise level from a single image. However, the real CCD camera noise is not simply additive, neither is it uniform over the gray levels. For obvious compression requirements, our experiments will treat JPEG bursts that have undergone an unknown contrast change (gamma-correction). As we shall see, the resulting estimated curve model is strongly image dependent and cannot be estimated by a parametric method.

2.3. Image/Video denoising algorithms that will be involved

Image denoising methods are based on various models of the original noise-free image, which permit to separate it from noise. One of the assumptions is the sparsity in an basis, orthogonal or over-complete. Sparsity is widely used in the many applications of image processing, such as denoising [31], color denoising, inpainting [32] and super-resolution [33, 34]. Non-Local means [5] assumes an image self-similarity and restores an unknown pixel using other similar pixels. The similarity is considered in terms of a patch centered at each pixel, not just the intensity of the pixel itself. In order to denoise a pixel, it is better to average the nearby pixels with similar structures (patches). This idea was extended to movie denoising [35, 36, 37]. The denoising algorithm by Dabov *et. al.* [38] combines self-similarity block matching, and threshold in the transform domain. The sparse representation is enhanced in transform domain by grouping similar 2D image patches into a 3D block. The weighted averaging of all the block-wise estimates are aggregated for the final output. Extensions to other applications were discussed by the same group of the authors, such as color denoising [7], grayscale video denoising [39], image sharpening [40] and restoration [41]. So far BM3D represents the state of the art for stand alone denoising. G. Boracchi and A. Foi [42] extend BM3D or V-BM3D to signal-dependent noise. They assume a parametric noise model, in which the parameters can be estimated using [43]. Then BM3D is applied on the images after a variance-stabilizing transformation to make noise homogeneous and post-processing follows.

The present paper can be understood as an extension and explanation of the multiple image denoising attempt by Zhang *et. al.* [44]. These authors propose a global registration of an image burst before applying a block matching multiimage strategy to the registered images. They remark that their denoising performance stalls when the number of frames grows and write that this difficulty should be overcome. Yet, their observed denoising performance curves grow like the square root of the number of frames, which indicates that their algorithm relies on accumulation. Thus, this performance is in fact optimal. The only non-synthetic experiments are made by these authors on a flat static real scene, actually a white board. The method proposed here is definitely an extension: It uses a hybrid scheme which chooses the best of accumulation or block denoising, depending on the reliability of the match. Without the accurate nonparametric noise estimation, this strategy would be unreliable.

3. THE MAIN TOOLS OF THE BURST DENOISING CHAIN

In this section, we discuss how to register all the images into one in the image sequence, which is taken as template. The average of the registered images gives a desired denoising result, but this only works at well-registered pixels. As for the pixels that are not well-registered, classic state of the art denoising (NLM, BM3D) will be tested. The decision maker, i.e. whether to use averaging or a denoising algorithm, will be based on the noise model, which will be estimated from the samples of each well-registered pixel along time. Thus having an accurate noise model obtained from the burst itself in crucial in the strategy. In summary, burst denoising is a relatively complex chain that:

- registers the images of the burst by subpixel accurate SIFT and estimation of the best dominant homography;
- equalizes the histograms of the registered images to remove lighting effects;
- estimates accurately from the many samples offered after registration the noise for each channel and at each level;
- thanks to this estimation, proceeds to denoising by averaging at all pixels where the correct registration is confirmed, and applies a state of the art denoising elsewhere.

Short preliminary discussion: is that safe? In spite of its complexity the chain is safe. Indeed, the dominant registration yields many samples permitting robust estimation of the noise. The averaging is applied only at pixels where the observed standard deviation after registration is close to the one predicted by the noise model. Thus, there is no risk whatsoever with averaging. At the other pixels, standard state of the art video denoising is applied. Block matching is only made safer by the previous registration and equalization. The experimental section will confirm the safety of the method by showing that the final result always is better than classic video denoising alone.

3.1. Registration of an Image Sequence

We shall use SIFT as the tool for the key point detection. A sub-pixel precision for denoising is required but, unfortunately, the precision of the SIFT points decreases through the octaves. Indeed, SIFT simulates the scale space by sub-sampling the images by factor two through each octave. Thus the sub-pixel key point detection, which is sub-pixel accurate in the first octave, can be several pixels inaccurate in the last octaves. To maintain a constant precision through the octaves, the SIFT sub-sampling between the octaves was simply canceled.

This cancelation of the sub-sampling entails two adjustments of SIFT. The first one is to adjust the Laplacian threshold, an important parameter in the SIFT method removing key points due to noise. Canceling the sub-sampling between octaves is equivalent to up-sampling the images by a power of two. Thus the Laplacian of the pixel on the twice up-sampled image is four times smaller than the corresponding one on the original image, because

$$\triangle\left(u(\frac{x}{2},\frac{y}{2})\right)=\frac{1}{4}\triangle u(x,y) \tag{1}$$

where $u(x,y)$ is the image and $\triangle$ is the Laplace operator.

The second adjustment after the cancelation of the SIFT sub-sampling is the construction of the descriptors. In our case, the blur is increasing through octaves, and so is the size of the domaing associated with each descriptor. To keep the scale invariance, the domain of each descriptor in the n-th octave is therefore sub-sampled by a 2^{n-1} ratio.

In summary, the precision of SIFT key points is improved by canceling the sub-sampling through octaves. The SIFT descriptor construction and the Laplacian threshold are adapted to keep them as in the original SIFT. As will be proved in simulations, the accurate SIFT retains a rather constant precision through octaves.

3.2. Reliable dominant homography estimation

An adaptation to multi-images of the Moisan-Stival ORSA algorithm [10] will be used. We adapt their notations here. Assume the set of match pairs is

$$S=\left(\mathbf{x}_i=(x_i,y_i),\mathbf{x}'_i=(x'_i,y'_i)\right)_{i=1\dots n},$$

We are interested in the homography matrix $\mathbf{H}$, that is best compatible with these matches (and not in the fundamental matrix itself [45, 11]). Also, we want to keep a safe subset of inliers T in S, with size k ($4<k\leq n$). Following [10] define the rigidity of T associated with $\mathbf{H}$ by

$$\alpha_{\mathbf{H}}(T)=\frac{\pi}{A'}\Big(\max_{(\mathbf{x},\mathbf{x}')\in T}\mathrm{dist}(\mathbf{x}',\mathbf{H}\mathbf{x})\Big)^2, \tag{2}$$

where A' the area of the second image domain. The rigidity is in fact a geometric probability. It is obtained by dividing the area of a disk with radius the maximal $\mathbf{H}$-projection error for T, by the image area A'. Following the *a contrario* method, if the rigidity is too small to be explained by randomness, the deduction is that there are only "inliers" in T. It is difficult to compute the probability $P(\inf_{\mathbf{H}}\alpha_{\mathbf{H}}(T)<t)$ to select the best subset T, even if we assume all the points are uniformly distributed in images. Instead of computing this probability directly, Moisan and

Stival [10] use a Bonferroni-like estimate, namely the expected number of false alarms (**NFA**), also referred to as the meaningfulness:

$$\epsilon(\alpha, n, k) := (n-4) \cdot {}^{n}_{k}\mathbf{C} \cdot {}^{k}_{4}\mathbf{C} \cdot \alpha^{(k-4)}. \tag{3}$$

This number incorporates the size of the matching set, the size of the subset and the rigidity. This algorithm has zero parameter and does not require any assumptions on the camera motion or the estimation of noise variance.

In burst denoising the ideal way would be to partition the image domain into different regions, each of which shares the same homography, to compute an homography on each of them and finally to register each image to the reference one by inverting the homography for each region. But, if we apply separately ORSA between the template and each other image, it is not guaranteed that the same region with a dominant homography will be chosen for each pair. A natural solution is to find a region and a homography common to all pairs of images. Therefore, ORSA is adapted by defining a "joint meaningfulness" as indicated in Algorithm 1.

Algorithm 1 multiple ORSA

Input The set S_0 of the common SIFT points in the template and the corresponding matching points in the $j-$th image, denoted as S_j.
Set $\epsilon = +\infty$
while the number of trials does not exceed N **do**
  Pick up 4 random points from S_0
  for (each $j > 0$) **do**
    Compute the homography using these 4 points and the corresponding ones in S_j
    Find the most meaningful subset of S with respect to S_j under this homography, save the meaningfulness parameter as ϵ_j
  end for
  Compute the joint meaningfulness $\epsilon_{joint} = \sum \epsilon_j$
  If $\epsilon_{joint} < \epsilon$, then $\epsilon = \epsilon_{joint}$, and save the meaningful subset for each pair of images as T_j and the 4 points, $P4$.
end while
Return ϵ_{joint}, T_j and $P4$.

It is impossible to try all 4 points combinations. Instead, an optimized random sampling algorithm (ORSA)is used as suggested in [10]. The algorithm stops once $\epsilon_{joint} < 1.0$. Then it is iterated for a small number of trials, typically $N/10$.

3.3. Video Equalization

There still is an extra step before noise estimation: the burst equalization! The images taken under indoor lights usually consist of fast variations of the contrast and brightness. We want to make them consistent through all the images, so that the standard deviation along time is indeed due to the noise, not to the changes of lights. This is done by a joint histogram equalization of all images. The best exponent of joint equalization is the Midway method [46, 47] which is summarized in a simple and elegant formula. Let $v : \Omega \to [0, 1]$ be an image and h its intensity histogram. The cumulative histogram of v is

$$H(x) = \int_0^x h(s)ds.$$

Starting with a series of images $v_j,\ j = 1, \cdots, N$ with cumulative histogram $H_j(x)$, the Midway cumulative histogram H is defined as a compromise of all H_j by $H = \left(\frac{1}{N}(\sum_j H_j^{-1})\right)^{-1}$. Once H is computed, each image v_j is replaced by $\phi_j(v_j) = H^{-1}(H_j(v_j))$. The necessity of histogram equalization to get a reliable noise model is illustrated in Fig. 4.

3.4. Signal-dependent Noise estimation

Here is the crucial step of the chain. The sequence of registered images is used to estimate the signal-dependent noise curve. If one pixel is well-registered, its values along the time give samples permitting to estimate the noise model. Therefore the standard deviations are classified according to their mean. The main question is to have an estimate robust to the wrongly registered pixels. The histogram of the mean of each pixel along time is constructed, with $n = 100$ uniform bins. Inside each bin, the median value of the standard deviations of all pixels is computed. This yields a curve of mean versus standard deviation. The median is robust to outliers by itself, but several precautions can be taken to make the estimate still more reliable. First, all edge points in all images, on which the interpolation error is stronger, are simply ruled out. This is done by a simple Canny edge detector. Second, the pixels whose standard deviation is too large are also considered outliers. The threshold is set to be the double of the peak value in the histogram of all standard deviations. Finally, bins that contain less than 100 items are simply not retained. The noise curves for the three color channels are estimated separately, but show a striking coincidence up to a multiplicative factor.

3.5. Hybrid Denoising Scheme

The noise estimate is crucial to meet a safe decision about which kind of denoising can be applied at each pixel. Suppose we have two denoising results: the one from averaging u_{A} and the other from NLM or BM3D, u_{BM3D}, the hybrid scheme will return

$$u_{\text{H}} = \alpha \cdot u_{\text{BM3D}} + (1-\alpha) \cdot u_{\text{A}}$$

For each registered pixel x in the template image, the average $u(x)$ of its samples after registration is looked up in the noise model. The noise curve gives the expected standard deviation $\sigma(x) := \sigma(u(x))$. At the same registered pixel the empirical standard deviation $\hat{\sigma}(x)$ of the samples is also computed. If this pixel is correctly registered, $\hat{\sigma}$ should close to σ, in which case a small value should be given to the weight α. A simple choice for α uses the sigmoid function:

$$\alpha(x) := \frac{1}{1 + \exp(c - \hat{\sigma}(x)/\sigma(x))}.$$

To avoid any impulse noise created by a local conflict between estimates, the weight function $\alpha(x)$ is slightly smoothed out by a 3×3 spatial average. Algorithm 2 summarizes the steps of the proposed multi-image denoising.

Algorithm 2 multi-image denoising

Input: ImageSequence $V = \{V_0, \cdots, V_N\}$
compute SIFT points on V_0, saved as CommonSIFTpts
for $(j > 0)$ **do**
  compute SIFT points on V_j
  save the matching points in V_j and V_0 as currSIFTpts
  update CommonSIFTpts = CommonSIFTpts $\cap$ currSIFTpts
end for
Apply multiple ORSA (Algo. 1) on the set of CommonSIFTpts to get the most meaningful 4 points P_4
for $(j > 0)$ **do**
  compute homography between V_j and V_0 using P_4
  $V\text{reg}_j$ = register V_j back to V_0 by this homography
end for
Video equalization
Noise estimation
Hybrid denoising scheme combining the average and block mathing denoising applied on $V\text{reg}$

4. EXPERIMENTS

4.1. Accurate SIFT

A check was made on the accuracy gain of the accurate SIFT described in Sec. 3.1. We applied SIFT and accurate SIFT on two images respectively. One of the images was generated from the other image by a simple rotation+translation, as shown in Fig. 2. The key points in both images were matched by using Lowe's classic matching method. After eliminating the outliers by ORSA the homograghy from one image to the other image was estimated. This homography allows us to project the key points of one image on the other image, and to estimate the average error. Table 1 shows the average error estimated in each octave in the scale space underlying SIFT. The experiment confirms that the precision for the classical SIFT decreases when the octave index increases. For accurate SIFT, the precision remains stable through octaves.

4.2. Multi-image Registration

Video data provided by the company DxO Labs capture a series of images of a rotating pattern with a fixed pedestal. We show three images from the sequence and the ones after registration in Fig. 3. In this easy case the dominant homography is a rotation of the main circular pattern, which contains more SIFT points than the pedestal region.

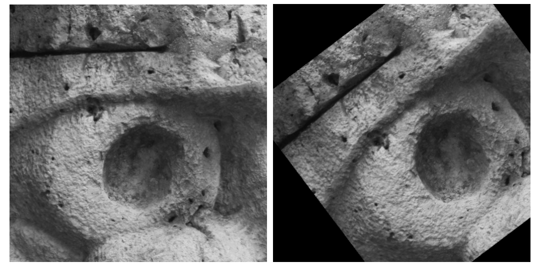

Figure 2. Two images used to test the accurate SIFT. The right image is generated from the left one by a translation+rotation.

average error		
	classical SIFT	improved SIFT
octave -1	0.036	0.036
octave 0	0.064	0.032
octave 1	0.263	0.033
octave 2	no keypoints	0.040

Table 1. The average error in each octave for Lowe's classical SIFT and for accurate SIFT. The precision decreases for Lowe's classical SIFT, while accurate SIFT remains stable through octaves. This is essentially obtained by removing the sub-sampling step in the SIFT method.

4.3. Video Equalization

Fig. 4 shows the efficiency of video histogram equalization. The images were taken under ceiling lights with changing illumination.

4.4. Noise Estimation

In the real scenario, the noise is inherent to the image, each pixel being modeled as a Poisson process. This model is valid except in the very dark regions where thermal and electronic noise dominate, and in the bright regions because the sensor gain is anyway nonlinear. The original image was simulated as a Poisson noise whose mean was a good quality image, after geometric homographies simulating the camera shaking. The noise estimation algorithm is demonstrated on three examples: Barbara, Couple and Hill. As shown in Fig. 5, the standard deviation (Y-axis) of the noise curves follows nicely the square root of the intensity (X-axis). The noise curves of the real datasets are given in Fig. 6.

4.5. Multi-image Denoising

For the experiments on synthetic data, the quantitative measurement of the denoising performance will be measured by the root-mean-square errors (RMSE) of different denoising methods in Tab. 2. The accumulation is based on 16 images, thus yielding a theoretical noise reduction by 4. A 3.5 noise reduction is experimentally attained in the images. In all cases, the difference between the theoretical factor 4 and the observed one is probably due to the

Figure 3. Multi-image registration. Top: three frames from an image sequence with a rotating pattern and a fixed pedestal. Bottom: the corresponding ones after registration. The dominant homography we find is on the plane of the rotating pattern, since it contains more SIFT points than the pedestal region. As a result we observe the rotating pedestal and its background after registration. The images are a courtesy of DxO Labs, Boulogne.

Figure 4. Video Equalization. Top: three frames from an image sequence with different illuminations. Bottom: after registration and equalization.

Figure 5. Noise curve. From top to bottom: the original image, one of the simulated images by moving the image and adding Poisson noise, and the noise curve from our algorithm using 16 images. The standard deviation of the noise (Y-axis) fits to the square root of the intensity (X-axis).

Figure 6. Noise curves of the real data sets. Left: one of the images in the sequence; right: the noise curves of the three color channels.

Table 2. RMSE for different methods

	Barbara	Couple	Hill
noisy	11.30	11.22	10.27
NLM	10.83	5.43	6.73
BM3D	4.33	3.39	3.90
AR	**3.55**	**3.03**	**2.73**
GT	**2.85**	**2.89**	**2.63**

Table 3. RMSE on synthetic data with 16 images. AR and GT stand for "average after registration" and "ground-truth" in the sense of registration back by the ground-truth motion. In principle GT divides the RMSE by 4, while AR is very close but higher than GT due to misregistration and interpolation errors. In all cases video BM3D gets close to the ratio 4 limit, but is overcome by AR.

fact that the simulated images are seriously aliased, which caused interpolation errors after registration.

The denoising results are now given for several real data sets, each of which consists of 16 JPEG images bursts. For a better illustration, the comparison shows the intermediate steps: the simple average, Non-Local Means on the registered images, BM3D on the registered images, and the result of the hybrid scheme. These results are shown on several well-chosen zoomed-in regions.

Since the proposed algorithm only finds a dominant homography, which is the rotating pattern in Fig. 3, the simple average fails to denoise the region of the fixed pedestals. It also fails to remove some dust that was incidentally stick to the camera objective, as zoomed-in and shown in Fig. 7. On the other hand, fine texture details are dramatically lost by Non-Local Means, which instead gives good denoising on contrasted regions such as the pedestals. The hybrid scheme with NLM, combining both averaging and NLM captures the virtue of each method. As expected the result is still better with a hybrid scheme using BM3D: Indeed BM3D is the best denoising algorithm and is actually quite close in performance to the direct averaging, as has been shown in Table 2.

The images in Fig. 4 captured a 3D scene with a single-lens reflex (SLR) camera, Canon EOS 30D. The scene consists of 2 books, a newspaper and a moving mouse. We enlarge three illustrative parts in Fig. 8, in which the structure lines on the book and letters in the newspaper are smoothed out by non-local means. In contrast, the letters turn out to be readable when averaging. As for the moving mouse, the average fails completely, while block-matching succeeds, since it uses the similarity patches in the template image itself.

Finally we show a burst of images of a painting. This is a good direct application for our algorithm, since the images of the painting are in principle related by homography if the painting is flat and the camera distortion-free. As a result, the average is always favored by the hybrid scheme. The details are compared in Fig. 9, where the dynamics of the patches are equalized for a fair comparison.

Acknowledgment

Many thanks for data and fruitful discussions to Frédéric Cao and Frédéric Guichard, DxO Lab and to Guillermo Sapiro. Research partially financed by the Centre National d'Etudes Spatiales and the Office of Naval research under grant N00014-97-1-0839.

5. REFERENCES

[1] Lu Yuan, Jian Sun, Long Quan, and Heung-Yeung SHum, "Image deblurring with blurred/noisy image pairs," in *SIGGRAPH*, 2007.

[2] M. Bertalmio and S. Levine, "Fusion of bracketing pictures," Tech. Rep., 2009.

[3] E. Eisemann and F. Durand, "Flash photography enhancement via intrinsic relighting," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 673–678, 2004.

[4] Raanan Fattal, Maneesh Agrawala, and Szymon Rusinkiewicz, "Multiscale shape and detail enhancement from multi-light image collections," in *ACM SIGGRAPH*, 2007, p. 51.

[5] A. Buades, B. Coll, and J-M. Morel, "On image denoising methods," *SIAM Multiscale Modeling and Simulation*, vol. 4, no. 2, pp. 490–530, 2005.

[6] David Donoho and Iain M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, pp. 1200–1224, 1995.

[7] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Color image denoising via sparse 3d collaborative filtering with grouping constraint in luminance-chrominance space," in *Proc. IEEE Int. Conf. Image Process., ICIP*, 2007.

[8] C. Chevalier, G. Roman, and J.N. Niepce, *Guide du photographe.*, C. Chevalier, 1854.

[9] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[10] L. Moisan and B. Stival, "A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix," *International Journal of Computer Vision*, vol. 57, no. 3, pp. 201–218, 2004.

[11] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, ISBN: 0521540518, second edition, 2004.

[12] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector," *ECCV*, vol. 1, pp. 128–142, 2002.

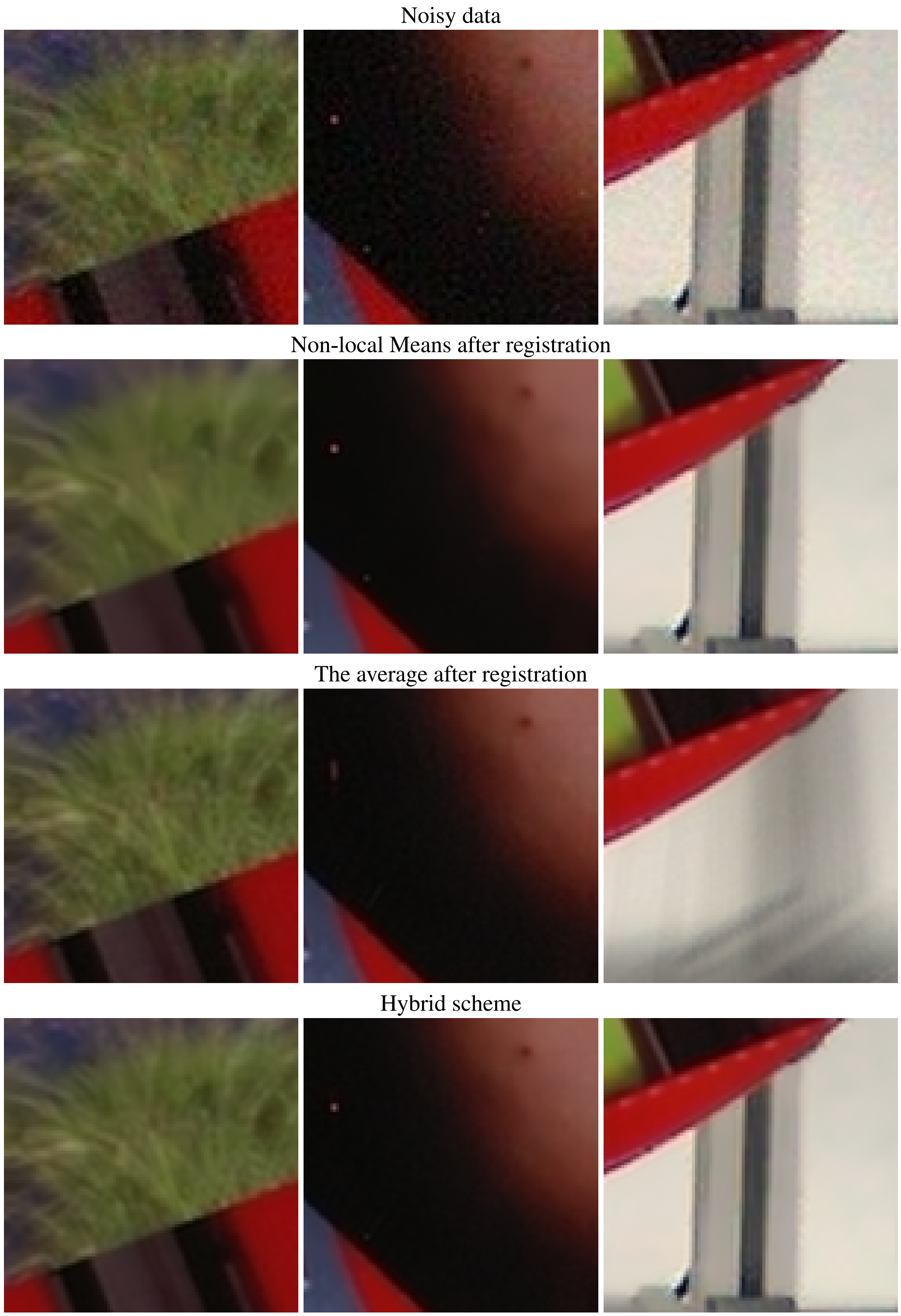

Figure 7. From top to bottom: the noisy data, NLM on registered images, the average and the hybrid scheme. Fine details are lost with NLM, for instance the grass. Due to mis-registration, the simple average fails to denoise the region of the pedestals. In the middle example, it does not remove some dust stick to the camera objective. The hybrid scheme works everywhere and gives roughly the same result with NLM and BM3D.

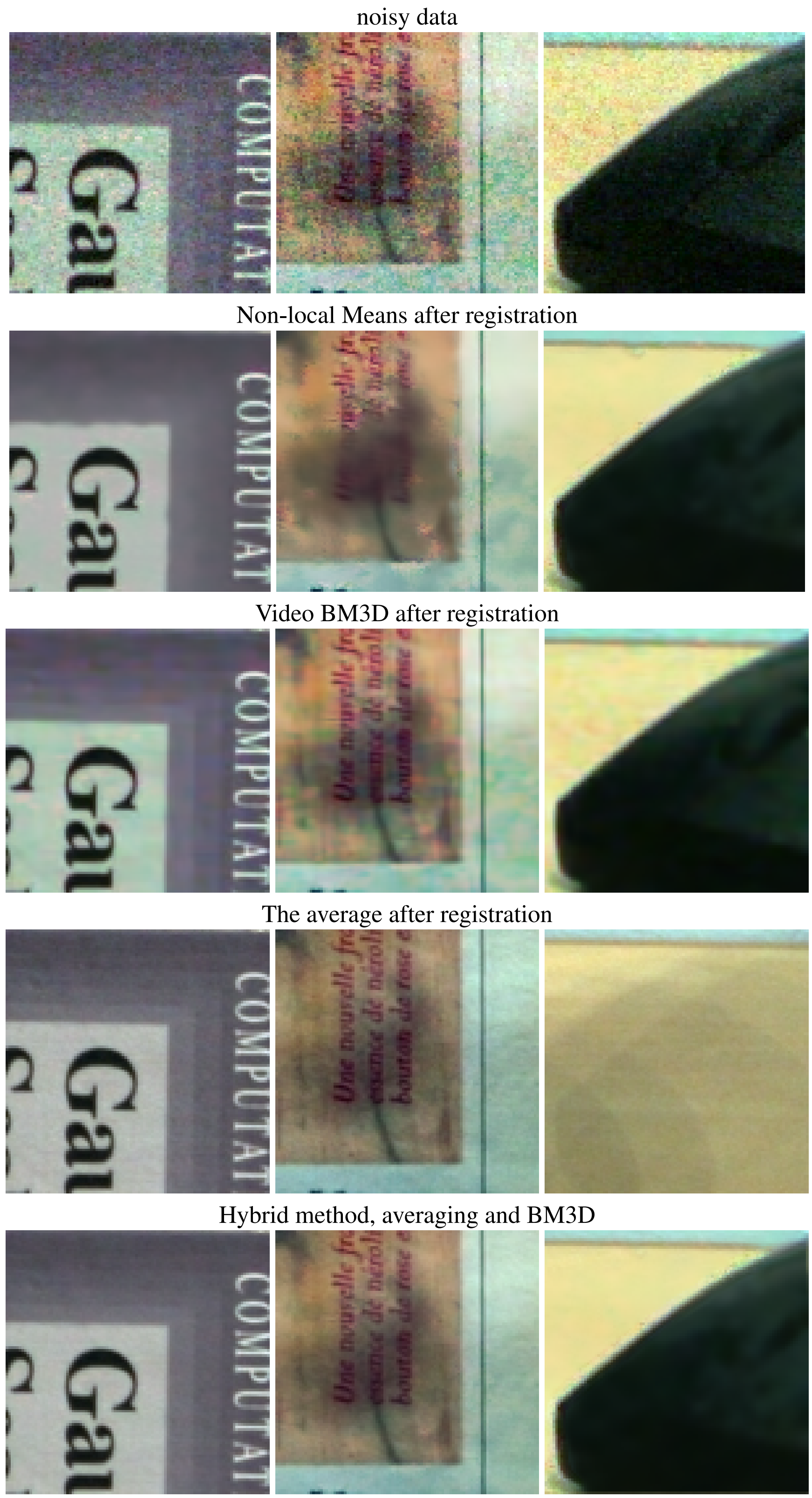

Figure 8. From top to bottom: the noisy data, NLM on registered images, video BM3D on registered data, the average and the hybrid scheme using BM3D. The average fails completely with the moving mouse on the right example, while block matching succeeds since it uses the similarity patches in the template image itself.

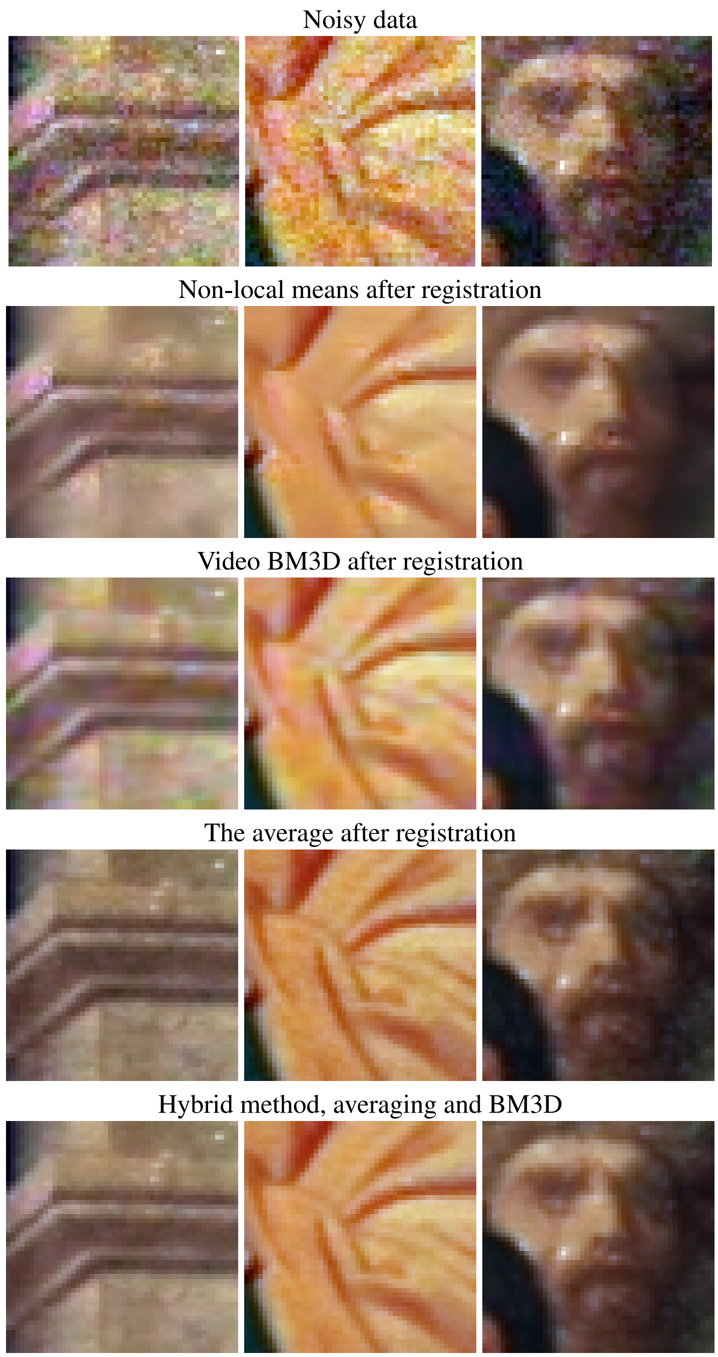

Figure 9. From top to bottom: the noisy data, NLM on registered images, video BM3D on the registered images, finally the average and the hybrid scheme with BM3D. The last two results are almost identical, which indicates that the registration has been detected correct almost everywhere.

epipolar geometry," *Artificial Intelligence*, vol. 78, no. 1-2, pp. 87–119, 1995.

[13] K. Mikolajczyk and C. Schmid, "Scale and affine invariant interest point detectors," *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63–68, 2004.

[14] Chum O. Urban M. Matas J. and Pajdla T., "Robust wide baseline stereo from maximally stable extremal regions," *Proc. of British Machine Vision Conference*, pp. 384–396, 2002.

[15] F. Sur F. Cao P. Musé and Y. Gousseau, "Unsupervised thresholds for shape matchings," *Image Processing, 2003. Proceedings. 2003 International Conference on*, vol. 2, 2003.

[16] Pablo Musé, Frédéric Sur, Frédéric Cao, Yann Gousseau, and Jean-Michel Morel, "An a contraio decision method for shape element recognition," *International Journal of Computer Vision*, vol. 69, no. 3, pp. 295–315, 2006.

[17] J.M. Morel and G. Yu, "Asift: A new framework for fully affine invariant image comparison," *SIAM Journal on Imaging Science*, vol. 2, no. 2, 2009.

[18] J. Yuen C. Liu and A. Torralba, "Nonparametric scene parsing: label transfer via dense scene alignment," *CVPR*, 2009.

[19] J. Sivic and A. Zisserman, "Video google: a text retrieval approach to object matching in videos," *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on computer vision*, pp. 1470–1477 vol.2, 2003.

[20] A. Torralba J. Sivic C. Liu, J. Yuen and W. T. Freeman, "Sift flow: dense correspondence across different scenes," *ECCV*, 2008.

[21] Matthew Brown and David G. Lowe, "Automatic panoramic image stitching using invariant features," *International Journal of Computer Vision*, pp. 59–73, 2007.

[22] R. Szeliski M. Brown and S. Winder, "Multi-image matching using multi-scale oriented patches," *International Conference on Computer Vision and Pattern Recognition*, pp. 510–517, 2005.

[23] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *CACM*, vol. 24, no. 3, pp. 381–395, 1981.

[24] Chi-Keung Tang, Gerard G. Medioni, and Mi-Suen Lee, "N-dimensional tensor voting and application to epipolar geometry estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 8, pp. 829–844, 2001.

[25] R. Deriche O. D. Faugeras Z. Zhang and Q.T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown

[26] P. Torr and A. Zisserman, "MLESAC: A new robust estimator with application to estimating image geometry," *Computer Vision and Image Understanding*, vol. 78, pp. 138–156, 2000.

[27] D. Nistér, "Preemptive RANSAC for live structure and motion estimation," *Machine Vision and Applications*, vol. 16, no. 5, pp. 321–329, 2005.

[28] S. I. Olsen, "Estimation of noise in images: an evaluation," *CVGIP: Graph. Models Image Process.*, vol. 55, no. 4, pp. 319–323, 1993.

[29] M. Lendl K. Rank and R. Unbehauen, "Estimation of image noise variance.," in *Vision, Image and Signal Processing*, 1999, vol. 146, pp. 80–84.

[30] Ce Liu, William T. Freeman, Richard Szeliski, and Sing Bing Kang, "Noise estimation from a single image," *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, vol. 1, pp. 901–908, 2006.

[31] Michael Elad and Michal Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transations on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[32] J. Mairal, M. Elad, and G. Sapiro, "Learning multiscale sparse representations for image and video restoration," *IEEE Transactions on Image Processing*, vol. 17, no. 1, pp. 53–69, 2008.

[33] Jianchao Yang, John Wright, Thomas Huang, and Yi Ma, "Image super-resolution as sparse representation of raw image patches," in *Intl. Conf. on Comp. Vis. and Patt. Recog.*, 2008.

[34] Dmitry Datsenko and Michael Elad, "Example-based single document image super-resolution: a global map approach with outlier rejection," in *Multidim System Signal Processing*, 2007, number 18, pp. 103–121.

[35] A. Buades, B. Coll, and J.M Morel, "Nonlocal image and movie denoising," *International Journal of Computer Vision*, vol. 76, no. 2, pp. 123–139, 2008.

[36] Marius Tico, "Multiframe image denoising and stabilization," in *The 15th European Signal Processing Conference (EUSIPCO),*, 2008.

[37] Marius Tico, ""'adaptive block-based approach to image stabilization," in *Proc. of IEEE International Conference on Image Processing (ICIP)*.

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

[39] K. Dabov, A. Foi, and K. Egiazarian, "Video denoising by sparse 3d transform-domain collaborative filtering," in *Proc. European Signal Process. Conf., EUSIPCO*, 2007.

[40] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Joint image sharpening and denoising by 3d transform-domain collaborative filtering," in *Int. TICSP Workshop Spectral Meth. Multirate Signal Process., SMMSP*, 2007.

[41] K. Dabov, A. Foi, and K. Egiazarian, "Image restoration by sparse 3d transform-domain collaborative filtering," in *Proc. SPIE Electronic Imaging*, 2008.

[42] G. Boracchi and A. Foi, "Multiframe raw-data denoising based on block-matching and 3-d filtering for low-light imaging and stabilization," in *Proc. Int. Workshop on Local and Non-Local Approx. in Image Process., LNLA*, 2008.

[43] Alessandro Foi, Mejdi Trimeche, Vladimir Katkovnik, and Karen Egiazarian, "Practical poissonian-gaussian noise modeling and fitting for single-image raw-data," *IEEE TRANSACTIONS ON IMAGE PROCESSING*, , no. 10, pp. 1737–1754.

[44] Li Zhang, Sundeep Vaddadi, Hailin Jin, and Shree Nayar, "Multiple view image denoising," in *In IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2009.

[45] R. I. Hartley, "In defense of the eight-point algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 6, pp. 580–593, 1997.

[46] Julie Delon, "Midway image equalization," *Journal of Mathematical Imaging and Vision*, vol. 21, no. 2, pp. 119–134, 2004.

[47] J. Delon, "Movie and video scale-time equalization application to flicker reduction," *Image Processing, IEEE Transactions on*, vol. 15, no. 1, pp. 241–248, Jan. 2006.