Adaptive regularization of the NL-means: Application to image and video denoising

Camille Sutour, Charles-Alban Deledalle, Jean-François Aujol

Adaptive regularization of the NL-means: Application to image and video denoising

Camille Sutour, Charles-Alban Deledalle, and Jean-François Aujol,

***Abstract*—Image denoising is a central problem in image processing and it is often a necessary step prior to higher level analysis such as segmentation, reconstruction or super-resolution. The non-local means (NL-means) perform denoising by exploiting the natural redundancy of patterns inside an image; they perform a weighted average of pixels whose neighborhoods (patches) are close to each other. This reduces significantly the noise while preserving most of the image content. While it performs well on flat areas and textures, it suffers from two opposite drawbacks: it might over-smooth low-contrasted areas or leave a residual noise around edges and singular structures. Denoising can also be performed by total variation minimization – the ROF model – which leads to restore regular images, but it is prone to over-smooth textures, staircasing effects, and contrast losses. We introduce in this paper a variational approach that corrects the over-smoothing and reduces the residual noise of the NL-means by adaptively regularizing non-local methods with the total variation. The proposed regularized NL-means algorithm combines these methods and reduces both of their respective defaults by minimizing an adaptive total variation with a non-local data fidelity term. Besides, this model adapts to different noise statistics and a fast solution can be obtained in the general case of the exponential family. We develop this model for image denoising and we adapt it to video denoising with 3D patches.**

***Index Terms*—Non-local means, total variation regularization, image and video denoising, adaptive filtering**

I. Introduction

IMAGE denoising is a central problem in image processing. It is often a necessary step prior to higher level analysis such as segmentation, reconstruction or super-resolution. The goal of the denoising process is to recover a high quality image from a degraded version. Among the main denoising techniques, variational methods minimize an energy that constrains the solution to be regular. One of the most famous variational models used for image denoising is the ROF model [1] that minimizes the total variation (TV) of the image, hence pushing the image towards a piecewise constant solution. This method is quite adapted to denoising while preserving edges, but it presents three major drawbacks: the textures tend to be overly smoothed, the flat areas are approximated by a piecewise constant surface resulting in a *staircasing effect*, and the image suffers from losses of contrast. A possible solution to reduce these undesirable artifacts is to spatially balance the regularization as shown in [2].

Among state-of-the-art denoising methods, the non-local means (NL-means) [3] algorithm performs spatial filtering by using spatial redundancy occurring in an image. Instead of averaging pixels that are spatially close to each other, it compares patches extracted around each pixel to perform a weighted average of pixels whose surroundings are close. It has led to many non-local methods and improvements such as [4]–[8]. This technique shows good performance on smooth areas and repetitive textures for which the redundancy is high, but on singular structures the algorithm might fail to find enough similar patches and thus performs insufficient denoising. This is referred to as the *rare patch effect*, and has been studied for instance in [9] and [10]. On the other hand, the presence of noise in the compared patches can lead to false detections. It can result in averaging several pixel values that do not truly belong to the same underlying structure, creating an over-smoothing [6] sometimes referred to as the *patch jittering blur effect* [9].

Variational and non-local methods have been put together by defining "non-local" regularization terms [11], [12]. Such approaches enforce the regularity on neighborhood systems built from the similarity between surrounding patches. This allows to deal separately with smooth and textured areas. In [12], the authors perform a non-local regularization by defining a non-local gradient that enables them to smooth the flat areas while preserving fine structures. Unlike the total variation, this approach is free of the *staircasing effect* but it is still subject to the *rare patch effect*. Non-local regularization has been adapted to many inverse problems, for example in [13] for inpainting or compressive sensing or in [14] for deconvolution.

Louchet and Moisan [9] have also proposed to combine the NL-means with TV regularization in their TV-means algorithm. They adapt the total variation into a local filter, and they perform the NL-means based on the local TV regularization. They manage to reduce both the *staircasing effect* and the *rare patch effect* with an iterative scheme.

In the context of super-resolution, Protter *et al.* [15] and d'Angelo and Vandergheynst [16] suggest minimizing an energy with a non-local data-fidelity term, based on the weights of the NL-means, and a total-variation regularization (instead of a non-local regularization term as studied above). This sums up to minimizing the total-variation of the NL-means solution. The same idea was explored in [17] for deconvolution.

In this article, we follow a similar approach to combine TV with the NL-means in order to correct their respective defaults. We first reduce the *patch jittering blur effect* and next we apply a local regularization where the NL-means perform insufficient denoising, to reduce the *rare patch effect*. The prior *dejittering* step reduces bias at a cost of a larger variance. This ensures that the denoising performed by the NL-

C. Sutour is with IMB and LaBRI, Université de Bordeaux, Talence, France, e-mail: camille.sutour@math.u-bordeaux1.fr

J.-F. Aujol and C.-A. Deledalle are with IMB, CNRS – Université de Bordeaux, Talence, France, e-mail: {jean-francois.aujol, charles-alban.deledalle}@math.u-bordeaux1.fr

means is reliable up to residual noise fluctuations, meaning that not too many irrelevant candidates have been selected in the averaging process. Then we define a confidence index based on the level of residual noise after *dejittering*. This index locally ponders the TV regularization, hence creating an adaptive model as suggested in [2]. Contrary to [15] and [16], our model preserves the NL-means result where it is reliable without introducing over-smoothing, *staircasing* or contrast losses inherent to the non-adaptive TV minimization. On singular structures, the adaptive regularization corrects the *rare patch effect*. Besides, we propose a model that adapts to different noise models and we derive a simple resolution scheme in the general case of the exponential family, often encountered in imaging problems.

We have then extended this model to video denoising. Indeed, Buades *et al.* have shown in [18] that the NL-means can denoise image sequences without applying motion compensation. In fact, they even show that it is harmful to do so instead of using all possible candidates in all frames. Besides, in order to guarantee temporal coherence between frames, we use 3D patches, as proposed in [19]. Instead of comparing square patches computed in a single frame in a spatio-temporal neighborhood, we compute 3D patches that take into account temporal consistency in the computing of weights. This selects less (if better) candidates, hence enhancing the *rare patch effect*, that is then corrected thanks to the adaptive TV regularization, that we apply locally in the spatio-temporal domain. Hence, we have adapted our proposed model to video denoising by computing spatio-temporal NL-means combined with spatio-temporal TV regularization.

The organization of the paper is as follows: in Sections II and III we remind the reader of the principle of the ROF model and the NL-means. We give details as to how to solve these problems and how to adapt them to different noise statistics. In section IV, we detail our first contribution concerning the *dejittering* procedure used to reach a satisfying bias-variance trade-off. Section V presents the main contribution with the proposed model that performs an adaptive regularization of the NL-means. Section VI presents related approaches that combine the NL-means with variational methods in a denoising framework. In Section VII, we present a third contribution with the generalization of our R-NL model to other noise statistics, along with an implementation. Finally, our last contribution is presented in section VIII that extends the model to video denoising. Section IX presents results and compares them to state-of-the-art methods.

II. Variational methods

A. ROF model

The general problem in denoising is to recover the image $f \in \mathbb{R}^N$, N being the number of pixels in the image domain Ω, based on the noisy observation $g \in \mathbb{R}^N$. The usual model is the case of additive white Gaussian noise:

$$g = f + \epsilon \tag{1}$$

where f is the true (unknown) image and ϵ is a realization of Gaussian white noise of zero-mean and standard deviation σ. The variational methods consist in looking for an image that minimizes a given energy in order to fit the data while respecting some smoothness constraints. Among these methods, the ROF model [1] relies on the total variation (TV), hence forcing smoothness while preserving edges. The restored image u^{TV} is obtained by minimizing the following energy:

$$u^{\mathrm{TV}} = \underset{u \in \mathbb{R}^N}{\operatorname{argmin}} \lambda \|u - g\|^2 + \mathrm{TV}(u). \tag{2}$$

The term $\|u - g\|^2 = \sum_{i \in \Omega} (u_i - g_i)^2$ is a data fidelity term, $\mathrm{TV}(u) = \sum_{i \in \Omega} \|(\nabla u)_i\|$ is a regularization term and $\lambda > 0$ is the parameter that sets the compromise between data fidelity and smoothness. A lot of methods have been developed in order to solve such minimization problems, among which Chambolle's algorithm [20], the forward-backward algorithm [21] or Chambolle-Pock's algorithm [22].

B. Adaptation to other noise statistics

Formula (2) is well adapted to Gaussian noise since it can be seen from a Bayesian point of view as a maximum a posteriori with a data fidelity corresponding to the log-likelihood, with a TV *a priori* on the image. This model can thus be extended to other types of (uncorrelated) noise with an energy of the following form:

$$u^{\mathrm{TV}} = \underset{u \in \mathbb{R}^N}{\operatorname{argmin}} \; -\lambda \sum_{i \in \Omega} \log p(g_i | u_i) + \mathrm{TV}(u) \tag{3}$$

where $p(g_i|u_i)$ is the conditional likelihood of the true pixel value u_i given the observation of the noisy value g_i.

C. Limits and discussion

Minimizing the total variation forces the solution to be piece-wise regular, which is well adapted to denoising while preserving edges. However, a compromise has to be found between regularity on flat areas and preservation of textures, based on the choice of the parameter λ. If flat areas are to be properly denoised, λ needs to be small, so fine textures tend to be over-smoothed. On the other hand, preserving small structures requires a higher λ that will not allow to recover perfectly the flat areas, resulting in a *staircasing effect* and a loss of contrast. This trade-off makes the choice of λ difficult and strongly image-dependent [23], [24]. The authors of [2] show that this trade-off can be spatially reached by adapting locally the regularization parameter to the local variance in order to smooth flat areas while preserving textures.

III. Non-local filtering

A. The NL-means algorithm

One of the most recent popular denoising methods is the NL-means algorithm described by Buades *et al.* in [3]. It is based on the natural redundancy of the image structures, not just locally but in the whole image. Instead of averaging pixels that are spatially close to each other, the NL-means algorithm compares patches, ie small windows extracted around each pixel, in order to average pixels whose surroundings are similar. Weights are computed in order to reflect how much two noisy pixels are likely to represent the same true gray

level, then pixels are averaged according to these weights. For each pixel $i \in \Omega$, the solution of the NL-means is given by the following weighted average:

$$u_i^{\mathrm{NL}} = \sum_{j \in \Omega} w_{i,j}^{\mathrm{NL}} g_j \tag{4}$$

which is equivalent to the solution of the following minimization problem:

$$u^{\mathrm{NL}} = \underset{u \in \mathbb{R}^N}{\operatorname{argmin}} \sum_{i \in \Omega} \sum_{j \in \Omega} w_{i,j}^{\mathrm{NL}} (g_j - u_i)^2. \tag{5}$$

In both cases the weights $w_{i,j}^{\mathrm{NL}} \geq 0$ are computed in order to select the pixels j whose surrounding patches are similar to the one extracted around the pixel of interest i as:

$$w_{i,j}^{\mathrm{NL}} = \frac{1}{Z_i} \varphi \left[d(g(P_i), g(P_j)) \right] \tag{6}$$

where $Z_i = \sum_{j \in \Omega} w_{i,j}^{\mathrm{NL}}$ is a normalization factor that ensures that the weights sum to 1, φ is a kernel decay function and d a distance function that measures the similarity between the two patches P_i and P_j of size $|P|$ extracted around the pixels i and j. Note that in practice, the weighted average at pixel i is restricted to a large search window $\mathcal{N}_i$ centered on i such that $w_{i,j}^{\mathrm{NL}} = 0$ when $j \notin \mathcal{N}_i$. In [3], φ is a negative exponential and the distance function is an Euclidean norm convolved by a Gaussian kernel. In our implementation, we use the following kernel:

$$\varphi_h [d] = \exp \left(- \frac{|d - m_d^P|}{s_d^P \times h^2} \right) \tag{7}$$

where $h > 0$ is a filtering parameter and m_d^P and s_d^P are respectively the expectation and standard deviation of the dissimilarity d when applied between two patches of size $|P|$ following the same distribution. With this setting, the parameter h becomes less sensitive to the noise level, the choice of the dissimilarity criterion d as well as to the size of the patches. Hence, a same value of h can fairly be kept constant whatever the noise level, the patch size, or the dissimilarity criterion d.

B. Adaptation to other noise statistics

A possible extension of the NL-means to other (uncorrelated) noise statistics consists in replacing the weighted average by the weighted maximum likelihood estimate [25], [26]:

$$u^{\mathrm{NL}} = \underset{u \in \mathbb{R}^N}{\operatorname{argmin}} \; - \sum_{i \in \Omega} \sum_{j \in \Omega} w_{i,j}^{\mathrm{NL}} \log p(g_j | u_i) \tag{8}$$

and the distance between two noisy patches g_1 and g_2 by the following likelihood ratio [27]:

$$d(g_1, g_2) = - \log \frac{\sup_u p(g_1 | u_1 = u) p(g_2 | u_2 = u)}{\sup_u p(g_1 | u_1 = u) \sup_u p(g_2 | u_2 = u)}. \tag{9}$$

where u_1 and u_2 refers to the underlying noise-free patches. Other extensions could have been considered, for instance the extension of [28] applied to signal-dependent noise in [29], [30]. Our chosen extension is known to apply directly to various noise models (including gamma noise and Poisson noise) as demonstrated in [26], [27], [31] and it retrieves exactly the classical NL-means in the case of Gaussian noise. In comparison, [28] differs from the classical NL-means since it leads to a two-step filtering scheme and its applicability to some specific noise distribution might require further modifications: for instance an additional "a priori mean" step had to be introduced in [30].

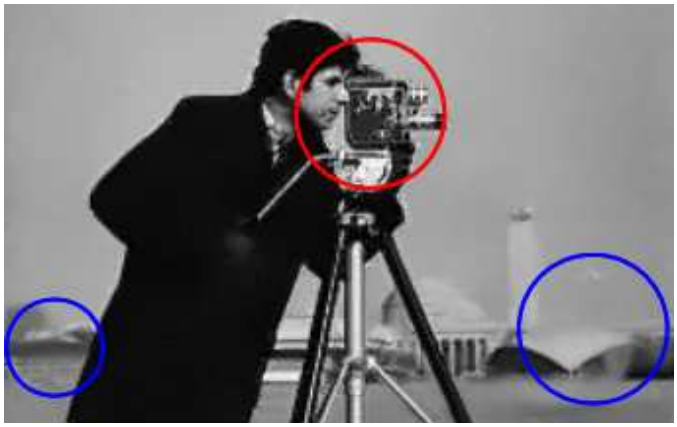


Fig. 1. Illustration of the defaults of the NL-means : the *rare patch effect* (red circle) can be observed around the head and the camera, while the patch *jittering effect* (blue circles) can be observed on the background.

C. Limits and discussion

Non-local methods achieve overall good performances but they suffer from two opposite drawbacks. The algorithm might select too many irrelevant candidates, resulting in the *patch jittering effect*: structures in these areas are overly smoothed due to the combination of candidates with different underlying values. On the contrary, around singular structures or edges, it can be difficult to find enough similar patches so the pixels are not properly denoised, resulting in a residual noise called the *rare patch effect*. These two problems oppose each other; they are controlled by the filtering parameter h, the search window size and the patch size that act as a bias-variance trade-off, as interpreted in [32]. Figure 1 illustrates these defaults: around the face of the cameraman and on the camera, the *rare patch effect* can be observed: similar patches are hard to find so residual noise remains. On the fence and the buildings on the background, over-smoothing results in a loss of resolution and produces some artifacts. The trade-off between the two problems makes again the choice of these parameters delicate and image dependent [7], [32].

IV. Dejittering of the NL-means (NLDJ)

One way to deal with the *jittering effect* has been addressed by Kervrann and Boulanger in [6]. The authors use adaptive search windows whose size is automatically adjusted to the local content of the image in order to reduce the number of potentially wrong candidates. The window size is locally selected based on a bias-variance trade-off principle. An approximation of the residual variance in the estimated image

u^{NL} is given at pixel i by

$$(\hat{\sigma}_i^{\text{residual}})^2 = \Big[\sum_{j\in\Omega}(w_{i,j}^{\mathrm{NL}})^2\Big](\sigma_i^{\text{noise}})^2. \tag{10}$$

where $(\sigma_i^{\text{noise}})^2$ is the noise variance, assumed to be constant in the non-local neighborhood of pixel i, and $w_{i,j}$ are assumed to be constant with respect to g_i (or their dependency can be neglected given the reasonably large patch size used in practice). The quantity $(\hat{\sigma}_i^{\text{residual}})^2$ plays an important role as an indicator of the total amount of noise that has been removed at pixel i. However, the residual variance cannot solely be used to assess the quality of the denoising. Indeed, the *jittering effect* arises from large noise reduction with mixes up of populations (i.e., bias). In [6], bias is detected iteratively by expanding the search window size, building up confidence interval based on (10), and selecting the largest window size that provides an estimate u_i^{NL} included in all confidence intervals obtained with smaller windows. This model deals efficiently with the *jittering effect*, but is difficult to extend to cases where noise is non-Gaussian.

We propose a unified framework to tackle the *jittering effect*, by limiting the amount of denoising performed by the NL-means when it is judged to be biased. We follow the idea of [33], [34] proposed initially for local adaptive filtering and extended to non-local adaptive filtering in [31].

We assume that in the non-local neighborhood of pixel i, the observations g_j are all realizations of the random variable $g_i = f_i + \epsilon_i$ where f_i and ϵ_i are two independent random variables. The quantity f_i models signal fluctuations and ϵ_i noise fluctuations. We assume that the signal f_i has a mean u_i^{NL} and a standard deviation σ_i^{signal} and ϵ_i has a zero mean with a known standard deviation σ_i^{noise}. A significant value σ_i^{signal} is an indicator of jittering; it assesses that observations in the non-local neighborhood of i belong to different populations.

In the scope of the Local Linear Minimum Mean Square Estimator (LLMMSE) strategy [33], [34], we locally perform a convex combination between the non-local estimation u^{NL} and the noisy data g, according to the following formula:

$$u_i^{\mathrm{NLDJ}} = (1-\alpha_i)u_i^{\mathrm{NL}} + \alpha_i g_i \tag{11}$$

where α_i is a confidence index defined by:

$$\alpha_i = \frac{(\sigma_i^{\text{signal}})^2}{(\sigma_i^{\text{signal}})^2 + (\sigma_i^{\text{noise}})^2} \approx \frac{|(\hat{\sigma}_i^{\mathrm{NL}})^2 - (\sigma_i^{\text{noise}})^2|}{|(\hat{\sigma}_i^{\mathrm{NL}})^2 - (\sigma_i^{\text{noise}})^2| + (\sigma_i^{\text{noise}})^2}. \tag{12}$$

Note that in the case of Gaussian noise, the noise variance is assumed to be known and constant in the image; it corresponds to σ^2 given for the model in (1). For signal-dependent noise, the non-local variance can be estimated as a function of u_i^{NL} as will be detailed in section VII. The approximation by the last term follows from the non-local independence assumption between f_i and ϵ_i, hence $\mathrm{Var}[g_i] = (\sigma_i^{\text{signal}})^2 + (\sigma_i^{\text{noise}})^2$. The signal variance $Var[g_i]$ can be estimated directly from the data g_j located in the non-local neighborhood using the following formula:

$$(\hat{\sigma}_i^{\mathrm{NL}})^2 = \sum_{j\in\Omega} w_{i,j}^{\mathrm{NL}} g_j^2 - \Big(\sum_{j\in\Omega} w_{i,j}^{\mathrm{NL}} g_j\Big)^2. \tag{13}$$

Pixels in the same non-local neighborhood should belong to the same population, so the estimated standard deviation $\hat{\sigma}_i^{\mathrm{NL}}$ should be close to the one of the noise σ^{noise}. Hence, if the estimated standard deviation $\hat{\sigma}_i^{\mathrm{NL}}$ is close to the expected one σ_i^{noise}, the index α_i is close to 0. According to (11), the non-local estimation u_i^{NL} is kept unchanged. On the other hand, if the non-local variance is found to be far from the expected one, then α is closer to 1. This results in re-injecting some noise back, to balance the bias-variance trade-off.

Note that the solution u^{NLDJ} can be rewritten as the following weighted sum

$$\begin{aligned} u_i^{\mathrm{NLDJ}} &= \sum_{j\in\Omega} w_{i,j}^{\mathrm{NLDJ}} g_j \\ \text{where}\quad w_{i,j}^{\mathrm{NLDJ}} &= (1-\alpha_i)w_{i,j}^{\mathrm{NL}} + \alpha_i \delta_{i,j} \end{aligned} \tag{14}$$

where $\delta_{i,j} = 1$ if $i = j$, 0 otherwise. Hence, as for u^{NL} and equation (10), the residual variance of the dejittered solution u^{NLDJ} can be approached at pixel i by:

$$(\hat{\sigma}_i^{\text{residual}})^2 = \Big[\sum_{j\in\Omega}(w_{i,j}^{\mathrm{NLDJ}})^2\Big](\sigma_i^{\text{noise}})^2. \tag{15}$$

The quantity $\hat{\sigma}_i^{\text{residual}}$ gives an indicator on the level of residual variance at pixel i obtained after *dejittering*, i.e. after a bias-variance trade-off has been achieved. Hence, unlike (10), this indicator can assess the quality of the denoising: it accounts for both bias and variance since bias in u^{NL} leads to residual variance in u^{NLDJ}.

This indicator is at the heart of the proposed regularization of the NL-means as we will see in the next section. In the following, since the *dejittering* step is performed in the NL-means algorithm, we will generally refer to the (possibly dejittered) solution of the NL-means as $u_i^{\mathrm{NL}} = \sum_{j\in\Omega} w_{i,j} g_j$, where $w_{i,j}$ are the weights in (14).

V. Regularized NL-means (R-NL)

The proposed model combines both the NL-means and the TV minimization in order to reduce the defaults observed in each method, in particular the *jittering effect* and the *rare patch effect* originating from the NL-means and the *staircasing effect* and the loss of contrast due to the TV minimization. We perform a TV minimization with a non-local data fidelity term as follows:

$$u^{\text{R-NL}} = \underset{u\in\mathbb{R}^N}{\operatorname{argmin}} \sum_{i\in\Omega}\lambda_i \sum_{j\in\Omega} w_{i,j}(g_j - u_i)^2 + \mathrm{TV}(u). \tag{16}$$

where $\lambda_i > 0$ are spatially varying regularization parameters. With non-local weights $w_{i,j} = \delta_{i,j}$, we recover the solution of the ROF problem (2). With non-local weights defined as in Section III and $\lambda_i = +\infty$, we can retrieve from (16), the solutions of the NL-means presented in (4) and (14).

Problem (16) is also equivalent to the following:

$$\underset{u\in\mathbb{R}^N}{\operatorname{argmin}} \sum_{i\in\Omega}\lambda_i \left(u_i - u_i^{\mathrm{NL}}\right)^2 + \mathrm{TV}(u). \tag{17}$$

This equivalence can be obtained by developing the data fidelity terms in equations (16) and (17) using the definition u^{NL} in (4) or (14), then keeping only the terms that take part

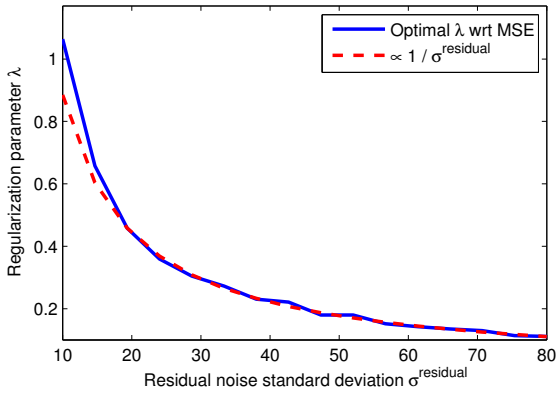


Fig. 2. Evolution of the optimal regularization parameter λ (blue line) with respect to the minimum square error (MSE) as a function of the residual noise standard deviation σ^{residual}. It is shown to be proportional to the inverse residual noise standard deviation $(\sigma^{\text{residual}})^{-1}$ (red dotted line).

in the minimization process. Hence $u^{\text{R-NL}}$ can be interpreted as a regular solution fitting u^{NL} where the λ_i are high.

As discussed in the previous section, an indicator of the quality of the denoising in u_i^{NL} is given by $\sigma_i^{\text{residual}}$ defined in eq. (15) (the lower, the better). Hence, the parameter λ_i can be chosen as a (non negative) decreasing function of $\sigma_i^{\text{residual}}$. The choice of this regularization function is critical to ensure good restoration without over-smoothing the edges and textures. A simple dimensional analysis of (2) indicates that λ should be chosen as inversely proportional to σ^{residual}. This relation is also suggested in [35] and in other regularizations involving ℓ_1 norms, see for instance [36]. To validate this point, we perform the following experiment. From an image u, we generate several noisy versions with increasing variance. In our context, these versions aim at modeling a solution u^{NL} in (17) with increasing levels of (spatially constant) residual noise $\hat{\sigma}^{\text{residual}}$. We next compute solutions of (17) for different λ parameters. For each noise level, we select the parameter minimizing the square error (known in this simulation). We perform a Monte-Carlo scheme in order to obtain mean values of the optimal λ for each noise level. Figure 2 gives the evolution of the mean optimal λ value as a function of the residual noise level $\hat{\sigma}^{\text{residual}}$. This experiment suggests again that the value of λ should be chosen inversely proportional to the standard deviation $\hat{\sigma}^{\text{residual}}$. More precisely, we find that λ_i should be chosen proportional to our approximation of the amount of noise reduction, i.e.:

$$\lambda_i = \gamma \left(\frac{\hat{\sigma}_i^{\text{residual}}}{\sigma_i^{\text{noise}}} \right)^{-1} = \gamma \left(\sum_{j \in \Omega} w_{i,j}^2 \right)^{-1/2}. \qquad (18)$$

where $\gamma > 0$ is a fixed parameter that sets the strength of the adaptive regularization. The relation with the non-local weights arises from equation (15). While λ in (2) controls the regularity globally on the whole image, here the parameter γ is influential only locally where the NL-means are unable to reduce significantly the noise.

With the proposed model, the NL-means and the TV regularization complete each other: on areas where the redundancy is important (homogeneous areas for example), the NL-means select many candidates so the residual variance is low. In the energy to minimize, the data fidelity term is then prominent over the regularization term, so the solution is close to the NL-means. This provides good smoothing and prevents the *staircasing effect* observed on smooth areas when treated with TV minimization. Around singular structures and edges where the redundancy is low, the NL-means select fewer candidates so the residual variance is high. This is also the case if wrong candidates are selected since the *dejittering* procedure reintroduces noises, hence leading to a higher residual variance. The regularization term becomes prominent over the data fidelity term, so it costs less to minimize the total variation of the image. The solution tends to a TV solution, preserving edges while reducing the *rare patch effect*.

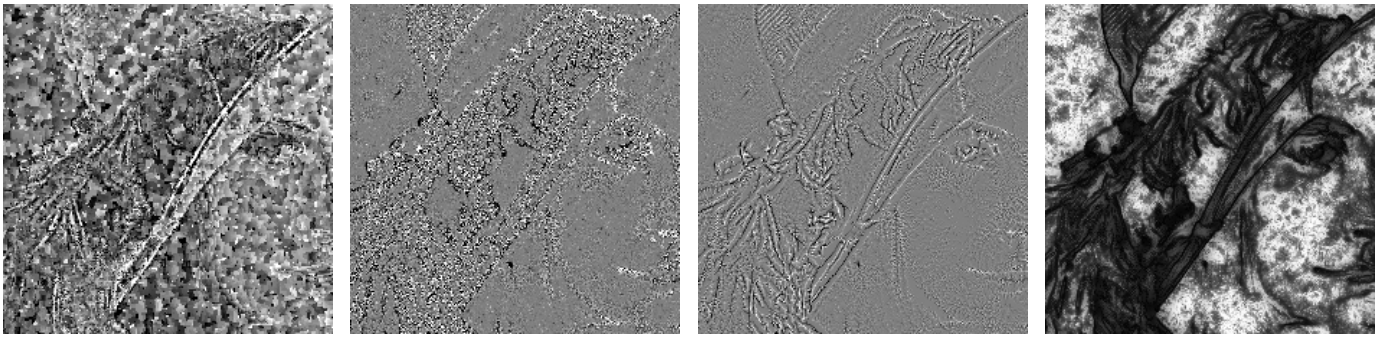
While the regularization parameter of the TV model and the parameters of the NL-means have been shown to be highly image-dependent, here the optimal parameters are not strongly influenced by the image content. Indeed, the model intrinsically adapts to the image content, hence the parameters do not require additional tuning: equivalent denoising strength can be achieved with a fixed set of parameters for all images as done in section IX.

The method is intuitive since it is based on the strengths and weaknesses of both the NL-means and the TV minimization. In Section VII, we will propose a simple implementation, in a more general framework, derived directly from mixing an NL-means algorithm with a TV minimization solver.

Figure 3 gives an illustration of solutions obtained with TV minimization, with the NL-means, before and after the *dejittering* procedure, and with the proposed R-NL algorithm. Figure 4 highlights the effect of the *dejittering* step and the adaptive regularization compared to the NL-means and TV. Figure 4-(a) displays the difference between R-NL and TV. Two defaults inherent to TV and corrected with R-NL are illustrated on this picture: the *staircasing effect* as well as the losses of contrast. Figure 4-(b) shows the difference between the NL-means before and after the *dejittering* step. We can see that some noise has been re-introduced on the hat and in the feathers which were over-smoothed by NL-means. Figure 4-(c) displays the difference between the (dejittered) NL-means and the regularized R-NL solution. Here, the difference is much more localized around the edges where the TV has been selectively applied. Figure 4-(d) shows a map of the confidence index λ_i that weights the data-fidelity term in the minimization process. Accordingly with the previous figure, the confidence index is high on flat areas, where no further regularization is applied to the already satisfying solution provided by the NL-means. On edges however, the NL-means did not find enough candidates so the residual noise is important, resulting in a smaller confidence. This is also the case on small structures such as the side of the hat where the *dejittering* has re-introduced some noise, hence increased the residual noise, to correct the fact that some irrelevant candidates might have been selected. This figure illustrates how the combination of the NL-means with TV minimization can effectively correct their respective defaults.

(a) TV (b) NL-means (c) Dejittered NL-means (d) R-NL

Fig. 3. Denoising of Gaussian noise with standard deviation $\sigma = 20$. The parameters that have been used for all the experiments are detailed in section IX-A. We notice the *staircasing effect* on the face of Lena and on the background on the TV denoised image (a). The NL-means result (b) suffers from over-smoothing (on the hat and the feathers for example) due to the *jittering effect*, that is corrected on the dejittered NL-means (c). The proposed method R-NL (d) combines these methods to provide efficient denoising free of the *staircasing effect*, the *jittering effect* and the *rare patch effect*.



(a) Diff. between R-NL and TV (b) Diff. NL-means and NLDJ (c) Diff. R-NL and NLDJ (d) Map of the weights

Fig. 4. (a) Difference between R-NL and TV. We note two major differences inherent to the defaults of the TV minimization: the flat areas are much smoother on R-NL since it does not suffer from the *staircasing effect*, and the change of intensity in the difference map reveals a loss of contrast on the TV denoised image. (b) Difference between the standard NL-means and the associated dejittered solution. The difference is stronger on the hair and the hat where some noise has been re-injected to balance the over-smoothing. (c) Difference between NLDJ and R-NL. Here, the difference is very local since the adaptive TV regularization performs solely where some residual noise remains. (d) Map of the confidence index λ_i. As expected, the confidence in the NL-means is smaller around edges where the candidates are hard to find and on small structures where the *jittering effect* has been corrected.

VI. Related approaches: combined non-local variational models

The idea to combine the NL-means with variational methods or to interpret the NL-means from a variational point of view has been studied in different approaches [37], [38]. Besides, since the NL-means provide interesting results in denoising, several authors have adapted the non-local methods to other problems such as deconvolution, inpainting or super-resolution [17], [39], [40]. This has often been achieved through a minimization framework that relies on non-local properties.

One of the most famous hybrid methods is the non-local TV (NL-TV) proposed by Gilboa and Osher in [12]. Based on the work on graph Laplacian of Zhou and Scholkopf [41] and Bougleux et al. [42], as well as the definition of non-local regularization terms of Kindermann et al. [11], they define a non-local gradient as follows:

$$(\nabla_w u)_{i,j} = (u_i - u_j)\sqrt{w_{i,j}} \tag{19}$$

where $w_{i,j}$ is the weight that measures the similarity between pixels i and j. This leads to the definition of a non-local framework, including the non-local ROF model:

$$u^{\text{NLTV}} = \underset{u}{\operatorname{argmin}} \|u - g\|^2 + \lambda \sum_{i\in\Omega} \|(\nabla_w u)_i\| \tag{20}$$

$$\text{with} \quad \sum_{i\in\Omega} \|(\nabla_w u)_i\| = \sum_{i\in\Omega} \sqrt{\sum_{j\in\Omega} (u_i - u_j)^2 w_{i,j}}.$$

This model has been introduced to deal separately with textures and smooth areas. It has been adapted to deblurring, inpainting or compressive sensing [13], [14]. Gilboa and Osher have also adapted this non-local regularization to non-local diffusion [43], which offers good denoising results. Figure 5-(d) shows the denoising in the case of Gaussian noise with NL-TV. We can see that small structures such as the cables and the writings on the boat are well-preserved while the *staircasing effect* is reduced on flat areas, thanks to the adaptive non-local regularization. If the approach can seem similar to the R-NL model, the philosophy is in fact quite opposite. NL-TV performs a non-local regularization in order to preserve textures and to reduce the *staircasing effect* that emanates from the TV minimization. It is possible to add in the implementation a step that reduces the *rare patch effect*, but the sole NL-TV model does not address this problem. On

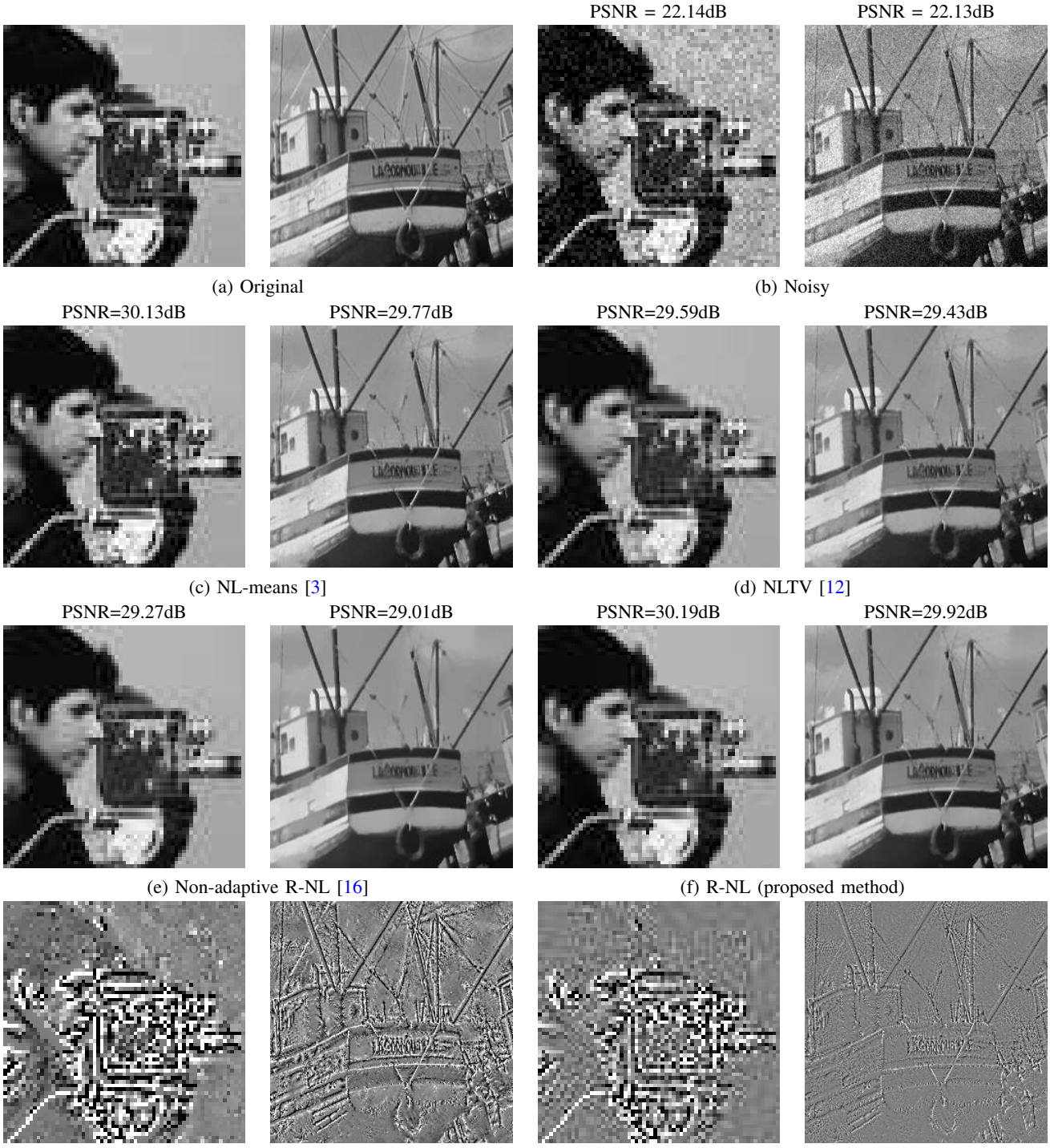


Fig. 5. Denoising of Gaussian noise, standard deviation $\sigma = 20$. From top to bottom, left to right: (a) Original and (b) noisy images of the Cameraman and the boat, processed with (c) NL-means, (d) NL-TV, (e) non-adaptive R-NL and (f) R-NL. (g) Difference between the solution of the NL-means and the result of (e) and (h) Difference between the solution of the NL-means and the result of (f). We observe on these images the ability of each method to combine variational and non-local methods in order to denoise efficiently while reducing the *staircasing effect* and/or the *rare patch effect*. The parameters are given in section IX-A. See text for more details.

the contrary, our approach performs a local TV regularization thanks to a non-local data-fidelity term, the main goal being to correct the *rare patch effect* and the *jittering effect* that come from the NL-means. A positive side-effect is that since the TV regularization is applied very locally and selectively, it does not suffer from the *staircasing effect* either.

Louchet and Moisan [9] also combine the NL-means with TV minimization in order to reduce both the *rare patch effect* and the *staircasing effect*. First they adapt the gradient to create a local filter and they show that their local TV denoising prevents the *staircasing effect*. Then they reduce the *rare patch effect* occurring in the NL-means thanks to

local TV minimization: if the NL-means do not select enough candidates to provide sufficient denoising, they apply local TV regularization to the patches to obtain a sufficient number of candidates. This amounts to applying adaptive TV regularization prior to the NL-means in order to ensure that it will find enough candidates in the averaging process. Our idea in the R-NL algorithm actually works the other way, since we first compute the solution of the NL-means and apply adaptive TV regularization afterward on the areas that did not have enough good candidates, based on the confidence index obtained from the NL-means.

NL-means have also been associated to TV minimization in a super-resolution context. Protter et al. [15] and d'Angelo and Vandergheynst [16] use a non-local data fidelity term combined to TV minimization in order to obtain a high-quality image from a low-resolution image sequence. Their formulation shows some similarities to our proposed model, but the philosophy behind it remains quite different. Figure 5-(e) illustrates this algorithm in a denoising context. This amounts to applying global TV regularization on the solution of the NL-means. We can see that the *rare patch effect* can be reduced compared to the solution of the NL-means. However, since the regularization is not adaptive, reducing the *rare patch effect* requires performing an excessive smoothing resulting in the loss of small structures and a contrast diminution. Compared to our model, the main difference resides in the fact that our data-fidelity is weighted by our confidence index, which allows to perform a locally adaptive TV minimization.

Figure 5 shows the results obtained from the different methods described in this section that combine the NL-means with TV minimization. We can observe that the NL-TV algorithm of Gilboa and Osher does effectively preserve fine structures such as the cables of the boat and it reduces the *staircasing effect*, but the flat areas like the sky around the Cameraman's head are not as smoothed as in the NL-means for example. The method based on the super-resolution algorithm of [16] reduces the *rare patch effect* of the NL-means thanks to the TV-regularization, but the compromise is hard to find: for example, the cables of the boat tend to be over-smoothed if the residual noise on the writings is to be efficiently removed. Besides, the difference map between this method and the NL-means on Fig. 5-(g) reveals that the solution obtained with non-adaptive regularization suffers from a loss of contrast, due to the TV minimization. R-NL on Fig. 5-(f) offers a simple way to deal with the compromise between efficient denoising and preservation of textures, thanks to the information provided by the estimated residual noise that acts as a measure of confidence in the denoising performed by the NL-means. We can observe on Fig. 5-(h) the difference between R-NL and the NL-means. The adaptive TV regularization is locally applied only on edges and singular structures, based on the confidence in the NL-means provided by our residual noise variance estimation. This prevents over-smoothing, staircasing and loss of contrast.

VII. R-NL for the exponential family

Both TV and NL-means are robust to different kind of noises thanks to the possible adaptations described in sections II and III. The proposed model can then be extended to other types of (uncorrelated) noise with a weighted data fidelity of the form $-\sum\sum w_{i,j}\log p(g_j|u_i)$, following the idea in [25], and with weights adapted as presented in Section III. This extended model can be solved efficiently in the general case of the exponential family that includes additive white Gaussian noise, Poisson noise and some multiplicative noises that are frequently encountered in image processing problems such as medical imaging, astronomy, remote sensing applications, etc. A probability law belongs to the exponential family [44] if it can be written under the following form:

$$p(T(g)|u) = c(g)\exp(\eta(u)T(g) - A(u)) \tag{21}$$

where c, T, η and A are known functions. The extended model is then the following:

$$u^{\text{R-NL}} = \underset{u\in\mathbb{R}^N}{\operatorname{argmin}} \sum_{i\in\Omega}\lambda_i \sum_{j\in\Omega} w_{i,j}[A(u_i) - \eta(u_i)T(g_i)] + \text{TV}(u) \tag{22}$$

where $\lambda_i = \gamma\big(\sum_{j\in\Omega} w_{i,j}^2\big)^{-1/2}$. As in the Gaussian case, it can also be reformulated with a weighted NL-means based fidelity term:

$$u^{\text{R-NL}} = \underset{u\in\mathbb{R}^N}{\operatorname{argmin}} \; -\sum_{i\in\Omega}\lambda_i \log p\left(u_i^{\text{NL}}|u_i\right) + \text{TV}(u) \tag{23}$$

where $\mathcal{L}_w(u_i) = -\lambda_i \log p\left(u_i^{\text{NL}}|u_i\right)$ is the weighted log-likelihood, $u_i^{\text{NL}} = \sum_{j\in\Omega} w_{i,j}T(g_j)$ and λ_i can be calculated with a quick implementation of the NL-means. We refer the interested reader to [45] for a more complete description of a fast way to compute the weights. Then the minimization step is achieved thanks to standard minimization algorithms, depending on the type of noise involved. A general implementation of the R-NL algorithm is given in Algorithm 1. More details regarding the minimization step will be given in the following subsections, according to the type of noise involved.

A. Gaussian case

Additive white Gaussian corruptions are said to be homoscedastic or signal-independent. The amplitude of the noise fluctuations is then constant in the image with a variance of σ^2. Hence, as already mentioned, the expected non-local variance involved in the *dejittering* step is chosen as: $(\sigma_i^{\text{noise}})^2 = \sigma^2$. In this case, solving (23) is equivalent to solving

$$\underset{u\in\mathbb{R}^N}{\operatorname{argmin}} \sum_{i\in\Omega}\lambda_i \frac{\left(u_i - u_i^{\text{NL}}\right)^2}{2\sigma^2} + \text{TV}(u). \tag{24}$$

which is up to a re-parametrization of the proposed model given in Section V, eq. (16). In practice, by taking into account the multiplicative constant that intervene in the negative log-likelihood, we can choose the same value of γ (that defines the scaling of λ) for every noise model, as done in section IX.

A lot of methods exist to solve this sort of problem. We have chosen to use Chambolle-Pock's algorithm [22], whose details are given in Algorithm 2. This method deals with minimization problems of the following form:

$$\min_u F(Ku) + G(u) \tag{25}$$

Algorithm 1 R-NL

Require: g: noisy input image,
h: filtering parameter,
$|P|$: patch size,
$\mathcal{N}$: size and shape of the search neighborhood
γ: regularization parameter

for $i \in \Omega$ **do**

NL-means step

Compute $w_{i,j} \leftarrow \varphi_h\big[d(g(P_i), g(P_j))\big],\ \forall j \in \mathcal{N}_i$
Normalize $w_{i,j} \leftarrow w_{i,j} / \sum_j w_{i,j},\ \forall j \in \mathcal{N}_i$
Compute $u_i^{\mathrm{NL}} \leftarrow \sum_j w_{i,j} T(g_j)$
Compute $(\hat{\sigma}_i^{\mathrm{NL}})^2 \leftarrow \sum_j w_{i,j} T(g_j)^2 - (u_i^{\mathrm{NL}})^2$
Compute $(\sigma_i^{\mathrm{noise}})^2$ according to the type of noise

Dejittering step

Compute $\alpha_i \leftarrow \frac{|(\hat{\sigma}_i^{\mathrm{NL}})^2 - (\sigma_i^{\mathrm{noise}})^2|}{|(\hat{\sigma}_i^{\mathrm{NL}})^2 - (\sigma_i^{\mathrm{noise}})^2| + (\sigma_i^{\mathrm{noise}})^2}$
Update $u_i^{\mathrm{NL}} \leftarrow (1-\alpha_i) u_i^{\mathrm{NL}} + \alpha_i g_i$
Update $w_{i,j} \leftarrow (1-\alpha_i) w_{i,j} + \alpha_i \delta_{i,j}$
Compute $\lambda_i \leftarrow \gamma \left(\sum_j w_{i,j}^2 \right)^{-1/2}$

end for

Minimization step

$$u^{\text{R-NL}} = \underset{u}{\operatorname{argmin}} \sum_{i \in \Omega} \lambda_i [A(u_i) - \eta(u_i) u_i^{\mathrm{NL}}] + \mathrm{TV}(u)$$

return $u^{\text{R-NL}}$

Algorithm 2 Chambolle-Pock algorithm [22]

Initialization: choose τ, $\sigma > 0$
$u_0 = u^{\mathrm{NL}}$ the result of the NL-means step
$y_0 = \nabla u_0$, $\overline{u}_0 = u_0$
Iterations ($k \geq 0$):

$$\begin{cases} y_{k+1} &= (I + \sigma \partial F^*)^{-1} (y_k + \sigma \nabla \overline{u}_k) \\ u_{k+1} &= (I + \tau \partial \mathcal{L}_w)^{-1} (u_k + \tau \nabla^* y_{k+1}) \\ \overline{u}_{k+1} &= 2u_{k+1} - u_k \end{cases}$$

that can be re-written in a primal-dual form:

$$\min_u \max_y \langle Ku, y \rangle - F^*(y) + G(u) \tag{26}$$

In this case, the function G is the weighted fidelity term, ie the weighted log-likelihood $\mathcal{L}_w(u_i)$, and the function $F(Ku)$ is the total variation $\|\nabla u\|_1$, ie $F = \|.\|_1$ and $K = \nabla$.

Note that in Algorithm 2, ∇^* refers to the adjoint of the gradient, so $\nabla^* = -\mathrm{div}$, and the terms $(I + \sigma \partial F^*)^{-1}$ and $(I + \tau \partial \mathcal{L}_w)^{-1}$ are the proximal operators associated to the functions F^* and $\mathcal{L}_w$. It is easy to check as in [22] that the proximal operator of F^* is a soft thresholding and we have:

$$y = (I + \sigma \partial F^*)^{-1} (\widetilde{y}) \Leftrightarrow y_i = \frac{\widetilde{y}_i}{\max(1, |\widetilde{y}_i|)}. \tag{27}$$

We can also show that [22]:

$$u = (I + \tau \partial \mathcal{L}_w)^{-1} (\widetilde{u}) \Leftrightarrow u_i = \frac{\widetilde{u}_i + \tau/\sigma^2 \lambda_i u_i^{\mathrm{NL}}}{1 + 2\tau/\sigma^2 \lambda_i}. \tag{28}$$

B. Poisson case

Poisson noise is also of particular interest since it occurs often in medical imaging, astronomy or night vision where the number of photons is limited. The negative log-likelihood of $u > 0$ for an observed intensity g is given by $\mathcal{L}(g|u) = \frac{u}{Q} - \frac{g}{Q} \log \frac{u}{Q} + \log \frac{g}{Q}!$ where g/Q is a non-negative integer. Poisson-corrupted data is signal-dependent with a variance proportional to the expectation. Hence, the expected non-local variance involved in the *dejittering* step is chosen as $(\sigma_i^{\mathrm{noise}})^2 = Q u_i^{\mathrm{NL}}$. The solution of the NL-means and the adaptive regularization parameters λ_i can then be computed accordingly and the variational problem becomes:

$$u^{\text{R-NL}} = \underset{u \geq 0}{\operatorname{argmin}} \sum_{i \in \Omega} \lambda_i \left[\frac{u_i}{Q} - \frac{u_i^{\mathrm{NL}}}{Q} \log \left(\frac{u_i}{Q} \right) \right] + \mathrm{TV}(u). \tag{29}$$

This functional is strictly convex, so the uniqueness of the solution is guaranteed. However, the data fidelity term is not differentiable. One solution would be to regularize the fidelity term (ie the logarithm). Instead, we use the primal-dual algorithm adapted to the Poisson case. Its general form is the one presented in Algorithm 2, and Anthoine et al. have determined explicitly the proximal operators required for Chambolle-Pock's algorithm in [46]:

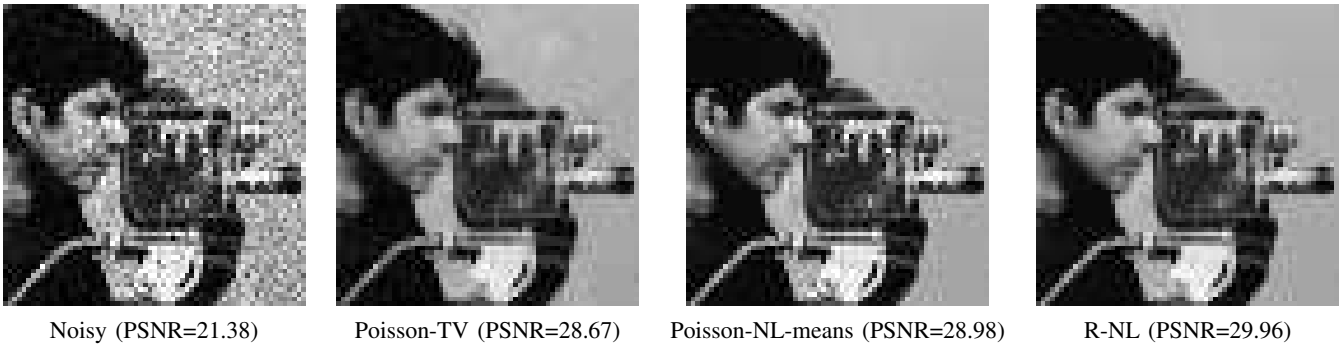
$$u = (I + \tau \partial \mathcal{L}_w)^{-1} (\widetilde{u}) \Leftrightarrow \tag{30}$$
$$u_i = \begin{cases} \frac{1}{2} \left(\widetilde{u}_i - \tau\lambda_i + \sqrt{(\widetilde{u}_i - \tau\lambda_i)^2 + 4\tau\lambda_i u_i^{\mathrm{NL}}} \right) & \text{if } u_i^{\mathrm{NL}} > 0, \\ \max(\widetilde{u}_i - \tau\lambda_i, 0) & \text{otherwise} \end{cases}$$

C. Gamma case

We also focus on gamma multiplicative noise which models *speckle* fluctuations encountered for example in synthetic aperture radar imagery or ultrasound imagery. The negative log-likelihood of $u > 0$ for an observed intensity $g > 0$ is given by $\mathcal{L}(g|u) = \frac{Lg}{u} + L \log u + \log \Gamma(L) - L \log L - (L-1) \log g$, where L is the "number of looks" that sets the level of the noise. As for Poisson-corrupted data, gamma corruptions are signal-dependent but with a standard deviation proportional to the expectation. Hence, the expected non-local variance involved in the *dejittering* step is chosen as $(\sigma_i^{\mathrm{noise}})^2 = (u_i^{\mathrm{NL}})^2 / L$. The solution of the NL-means and the adaptive regularization parameters λ_i can then be computed accordingly and the variational problem becomes:

$$u^{\text{R-NL}} = \underset{u > 0}{\operatorname{argmin}} \sum_{i \in \Omega} \lambda_i \left[L \log(u_i) + L \frac{u_i^{\mathrm{NL}}}{u_i} \right] + \mathrm{TV}(u). \tag{31}$$

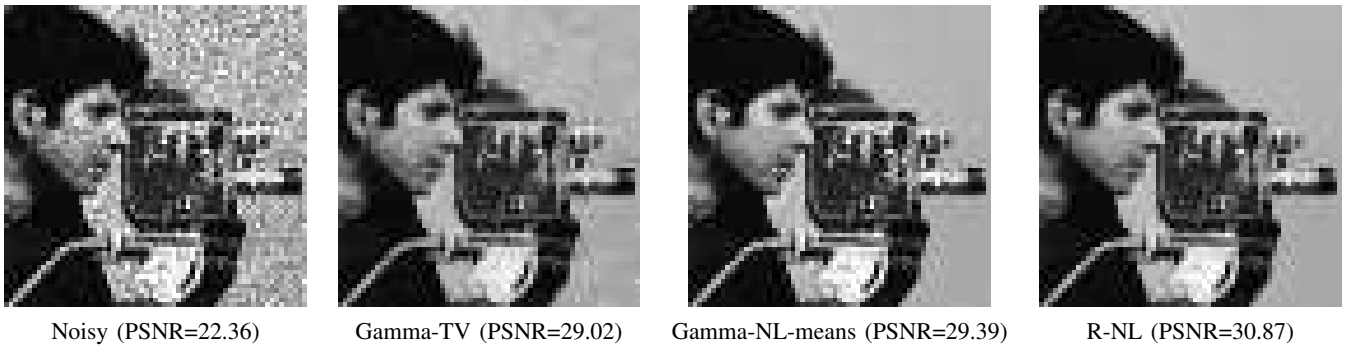
This functional is not convex, so there is no guarantee as to the existence of a unique minimizer. However, we can show that a minimization algorithm will converge towards a stationary point [47]. Since we initialize the algorithm to the solution of the NL-means, it is reasonable to believe that the solution will be close to the global one. However, for strong gamma noise (with a low number of looks L, for example $L = 4$ in table I), the sole TV minimization (with arbitrary or null initialization) will not necessarily converge to a satisfying result. Besides,



Noisy (PSNR=21.38) Poisson-TV (PSNR=28.67) Poisson-NL-means (PSNR=28.98) R-NL (PSNR=29.96)

Noisy (PSNR=20.97) Poisson-TV (PSNR=28.66) Poisson-NL-means (PSNR=28.79) R-NL (PSNR=29.43)

Fig. 6. Denoising of Poisson noise, $Q = 4$. Small structures such as the writing and the cables of the boat are better preserved with R-NL, while denoising is better achieved on the fine structures than with NL-means.



Noisy (PSNR=22.36) Gamma-TV (PSNR=29.02) Gamma-NL-means (PSNR=29.39) R-NL (PSNR=30.87)

Noisy (PSNR=22.12) Gamma-TV (PSNR=28.50) Gamma-NL-means (PSNR=29.35) R-NL (PSNR=30.22)

Fig. 7. Denoising of Gamma noise, $L = 12$. We observe on the NL-means some residual noise around the face of the cameraman or the writings on the boat, corrected by TV regularization on the R-NL results. The smoother areas are correctly treated by the NL-means, so we do not observe the *staircasing effect* associated to TV denoising.

we cannot use Chambolle Pock's algorithm since the proximal operator derived from the data-fidelity term is not easy to compute. However, the data fidelity term is differentiable so we can use the forward-backward algorithm [48]. Its purpose is to minimize the following problem:

$$\min_u F(u) + G(u), \tag{32}$$

where G needs to be differentiable with ∇G $1/\beta$-Lipschitz and F simple, meaning that its proximal operator is easy to compute. In the present case, G is the data-fidelity term $\mathcal{L}_w$, and F is the total variation. The algorithm is described in Algorithm 3. Some accelerations such as the FISTA algorithm [49] can also be used, or a generalized version GFB [50]

Algorithm 3 FB algorithm [21]

Initialization: choose $u_0 = u^{\mathrm{NL}}$ and $\kappa > 0$
Iterations ($k \geq 0$):
$u_{k+1} = (I + \kappa \partial F)^{-1} (u_k - \eta \nabla \mathcal{L}_w(u_k))$

that allows to deal with n functions. In this algorithm, we cannot compute directly the proximal operator of the total variation $(I + \kappa \partial F)^{-1}$, so we have to proceed to an inner loop that will calculate iteratively this proximal operator. The calculation of this proximal operator is achieved through a forward-backward algorithm or its fast version FISTA [51].

Figures 3, 5, 6 and 7 compare the R-NL algorithm with the TV minimization and the NL-means on images in $[0, 255]$. In the Gaussian case, Gaussian noise with standard deviation $\sigma = 20$ has been added. In the Poisson and gamma case, Poisson and gamma noise have been applied respectively to the original images with a noise level set in order to achieve a PSNR of 22dB. The parameters that were used for all the experiments are detailed in section IX-A.

VIII. Video denoising using the R-NL algorithm

The NL-means have been adapted in [52] to video denoising, so we would like to adapt our algorithm R-NL to video denoising as well, in order to bring the same advantages. The temporal NL-means achieve spatio-temporal filtering without prior motion-compensation. Indeed, motion estimation is a difficult task that may be nearly impossible to solve in constant areas where the aperture problem is too high. Buades *et al.* have shown that motion compensation is in fact counter-productive in this case: in image sequence denoising, the NL-means use a 3D neighborhood whose third dimension corresponds to the temporal frames. Adding motion compensation reduces the number of eligible candidates, when in the contrary in NL-means the more the merrier.

However, Liu et al. advocate in [53] that motion compensation is in fact essential, even in a non-local context. They integrate optical flow estimation in the NL-means framework in order to perform more efficient denoising while guaranteeing better temporal stability. In both NL-means and R-NL, if a small number of frames is used to compute the solution (which is often the case in order to lower the computational costs), no temporal regularity is guaranteed. Indeed, motion compensation is based on the Lambertian assumption that stipulates that a pixel has the same gray level during its whole trajectory, forcing a temporal regularity. In the case of the NL-means, such a hypothesis is not verified, so a pixel value may change from one frame to another. This results in a *glittering effect* when looking at the video, even though it is not perceptible when looking at only one frame or at the PSNR. Such a phenomenon is not harmful when the whole image is in motion, but when only part of the image in the video is in motion while the rest is static, the *glittering effect* appears like an undesirable consequence of the estimation variability.

To reduce this effect, we use 3D patches instead of 2D patches. Indeed, in the original version on the NL-means, Buades *et al.* compute patches in the current image (in 2D) and compare them to 2D patches in a 3D spatio-temporal search zone. This does not force regularity from one frame to another, resulting in the *glittering effect*. Instead, we compute 3D patches that compare spatio-temporal neighborhoods in order to force temporal consistency, as suggested in an inpainting context in [19]. In the NL-means-3D alone, the use of three-dimensional patches favors the *rare patch effect* since candidates are harder to find.

The *rare patch effect* can then be corrected using the adaptive TV regularization. The natural way to do it would be to apply spatial TV regularization, based on the adaptive regularization parameter λ we define. This provides an effective reduction of the *rare patch effect*, but reduces the temporal stability because of the fact that the TV minimization is applied on each frame independently, so the regularization is not consistent through time. To remedy this, we introduce a spatio-temporal TV regularization. This approach does seem counter-intuitive in a general framework, since the assumption of a piece-wise constant model in the temporal dimension does not hold (although it does not hold in the spatial domain either, but is used nonetheless). Video denoising using only spatio-temporal TV regularization suffers from the same defaults we might encounter on image denoising using TV regularization, but in the three dimensions: the *staircasing effect* appears also in the temporal domain, creating a blur and a loss of contrast around moving structures. In our framework however, the TV regularization is applied very locally only where the NL-means are judged to be too weak, so the *staircasing effect* is also avoided here. This adaptive temporal TV regularization forces temporal consistency when TV regularization is applied, but does not hurt the content of the video elsewhere. The implementation is quite straight-froward compared to spatial TV regularization; it only needs the computation of a 3-dimensional gradient and divergence operator.

IX. Results and discussion

A. Parameters

The algorithms described above have been carefully adapted to limit the influence of the statistic and the level of the noise on the different parameters. We have used three levels of noise: low, medium and high, that correspond in the Gaussian case to standard deviation $\sigma = 20, 30, 40$ respectively. We have then adapted the parameters Q and L for Poisson and gamma distributions to achieve the same initial PSNRs of about 22dB for low noise level, 18dB for medium noise level and 16dB for high noise level.

The kernel defined in (7) for the computation of the non-local weights is not very sensitive to the noise statistic, so we have set the filtering parameter h to 1 for all types and levels of noise. We have used standard 7×7 patches and 21×21 search windows. Besides, the constants were carefully kept in the log-likelihoods used for the TV minimization in order to obtain a regularization scaling parameter γ of the same scale for each type of noise. It has been set to 66 for low noise level and 100 for medium and high noise levels, for images quantified in 8 bits.

TABLE II
MEAN PSNRS ON DENOISED SEQUENCES WITH NL-MEANS-2D, R-NL-2D AND V-BM3D (2D PATCHES), AND NL-MEANS-3D R-NL-3D AND BM4D (3D PATCHES).

	Target	Tennis	Bicycle
NL-means 2D [52]	30.84	30.07	*32.24*
R-NL-2D	31.27	30.16	32.29
V-BM3D [59]	30.21	29.79	32.92
NL-means 3D	32.47	30.52	31.72
R-NL-3D	*33.35*	*30.60*	32.16
BM4D [60]	**34.53**	**31.06**	**33.37**

For video denoising, we have used spatio-temporal search windows of size $7\times7\times9$ and spatio-temporal patches of size $7\times7\times5$, with a regularization parameter (for low noise level) $\gamma = 50$.

B. Image denoising

We present here some numerical results to compare our R-NL algorithm to other approaches that rely on non-local and/or variational methods. We can see in Table I that these methods do provide an increase in PSNR and SSIM [54] compared to the NL-means or the TV minimization, confirming the visual observations from Figures 3-7. In the case of Poisson or gamma noise, we did not include the NL-TV algorithm since it applies only directly to Gaussian noise. Table I illustrates also the benefit of our adaptive regularization compared to the non-adaptive model of d'Angelo et al. [16]. Extensions of their model to Poisson ($\mathcal{P}$) and gamma ($\mathcal{G}$) noise have also been included under the name of *NA/R-NL* for non-adaptive R-NL. We have furthermore added in this table the results obtained with BM3D [55] in the case of Gaussian noise, and with BM3D applied after variance stabilization using Anscombe transform for Poisson noise [56] and logarithm transform for gamma noise [56]. Since the main goal of this article is to offer an improvement to the NL-means through an interpretation of the variance reduction as a measure of confidence, we do not claim to offer a better PSNR than BM3D, but we do offer an intuitive interpretation along with a simple and general implementation.

C. Video restoration

We also study here the behavior of our proposed method to video denoising, on three image sequences: *target*, *tennis* and *bicycle*. The original, noisy and denoised videos are available for download[1]. Table II displays the mean PSNR of the denoised videos using the standard NL-means adapted to video denoising (NL-means 2D), the R-NL algorithm adapted to video denoising (R-NL 2D), the NL-means algorithm with 3-dimensional patches (NL-means 3D), the R-NL algorithm with 3-dimensional patches (R-NL 3D) and the state-of-the-art video denoising algorithms V-BM3D [59] and BM4D [60]. We can see that our results with the proposed R-NL method with 3 dimensional patches are quite competitive.

[1] http://image.math.u-bordeaux1.fr/RNL

TABLE III
TEMPORAL STANDARD DEVIATION ON DENOISED SEQUENCES WITH NL-MEANS AND R-NL USING 2D PATCHES, V-BM3D, NL-MEANS AND R-NL USING 3D PATCHES, AND BM4D. THANKS TO THE USE OF 3D PATCHES, R-NL-3D AND BM4D PROVIDE THE BEST TEMPORAL STABILITY, AND VISUAL COMFORT.

	Target	Tennis (1-24)	Tennis (90-148)	Bicycle
NL-means 2D [52]	5.45	4.27	4.36	1.64
R-NL-2D	4.47	3.34	3.57	1.25
V-BM3D [59]	5.25	3.66	4.60	1.76
NL-means 3D	5.20	4.04	4.15	1.37
R-NL-3D	3.85	2.82	**3.09**	**0.90**
BM4D [60]	**3.67**	**2.78**	3.14	0.94

We can also show that the R-NL-3D algorithm enforces temporal consistency thanks to the use of three-dimensional patches, while limiting the *rare patch effect* thanks to the adaptive TV regularization. The superiority of spatio-temporal patches over spatial patches is also demonstrated by comparing BM4D with V-BM3D. We measure temporal variance on different image sequences in order to illustrate the Lambertian assumption: on areas that do not move during a part of the sequence, the pixel value should be unchanged from one frame to another, so the temporal standard deviation should be close to zero. Based on the ground truth of the original sequences, we can select areas that do not vary with time and then calculate the standard deviation obtained on the different denoised versions. Table III displays the standard deviation computed on such constant areas on the denoised versions of the same three image sequences.

Based on Table III, we can see that temporal stability is best guaranteed with R-NL-3D and BM4D, thanks to the use of 3-dimensional patches. The difference between R-NL-3D (resp. R-NL-2D) and NL-means-3D (resp. NL-means-2D) shows that the adaptive spatio-temporal TV regularization helps reducing the *glittering effect* by enforcing spatio-temporal consistency.

D. Discussion

The approach we have presented is based on a two-step improvement of non-local methods. It consists in correcting the *jittering effect* that occurs if candidates issued from a different underlying value are averaged together (this happens for example on the grass of the cameraman, see figure 8), and in reducing the *rare patch effect*. Since these drawbacks are linked to the difficulty to compute the non-local weights on noisy patches and to find relevant candidates, they are inherent to non-local methods. Our method uses the local estimation of the noise variance reduction as an indicator to correct first the *jittering effect* then the *rare patch effect*. We can then extend the model we developed for the NL-means to other non-local methods based on the computation of non-local weights, for example BM3D [55], or the improved version of the non-local means SAFIR [6] and SAIF [8]. We estimate the residual variance of the non-local solution, that we use to perform the *dejittering* step then the regularization step. Figure 8 displays the result of denoising of Gaussian noise using BM3D, SAFIR and SAIF, along with the associated regularized versions using the *dejittering* and adaptive regularization steps. We can see

TABLE I
PSNR/SSIM VALUES OF DENOISED IMAGES USING DIFFERENT METHODS FOR IMAGES CORRUPTED BY ADDITIVE WHITE GAUSSIAN NOISE, POISSON NOISE AND GAMMA NOISE.

	House (256)		Peppers (256)		Cameraman (256)		Boat (512)		Lena (512)		Barbara (512)		Man (512)	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Gaussian noise, $\sigma = 20$														
Noisy image	22.12	0.33	22.11	0.47	22.14	0.39	22.13	0.73	22.11	0.67	22.12	0.78	22.10	0.79
TV [1]	30.48	0.68	29.23	0.78	28.57	0.70	28.94	0.87	30.22	0.87	26.66	0.87	28.42	0.89
NL-TV [12]	27.04	0.73	29.99	0.81	29.59	0.79	29.43	0.87	29.16	0.89	27.92	0.90	28.05	0.88
NL-Means [3]	32.23	0.77	29.87	0.82	29.01	0.78	29.30	0.85	31.53	0.89	*30.09*	0.92	28.48	0.87
NLDJ [31]	32.31	0.77	30.45	0.83	30.13	0.81	29.77	*0.88*	31.73	*0.91*	29.98	*0.93*	29.10	*0.90*
NA/R-NL [16]	32.16	0.77	30.11	0.82	29.27	0.78	29.01	0.84	31.33	0.89	28.11	0.89	28.38	0.85
R-NL	*32.69*	*0.79*	*30.78*	*0.84*	*30.19*	*0.82*	*29.92*	*0.88*	*32.04*	*0.91*	29.76	*0.93*	*29.26*	0.89
BM3D [55]	**33.70**	**0.81**	**31.32**	**0.86**	**30.49**	**0.83**	**30.85**	**0.91**	**33.03**	**0.92**	**31.74**	**0.95**	**29.81**	**0.92**
Gaussian noise, $\sigma = 30$														
Noisy image	18.60	0.24	18.59	0.35	18.60	0.30	18.57	0.61	18.57	0.55	18.59	0.66	18.59	0.68
TV [1]	26.04	0.71	27.19	0.73	27.43	0.64	27.25	0.80	27.46	0.82	25.92	0.80	26.33	0.80
NL-TV [12]	25.15	0.61	25.97	0.68	26.26	0.63	26.04	0.81	26.15	0.83	25.68	0.85	25.75	0.82
NL-Means [3]	30.11	0.72	27.90	0.75	27.36	0.72	27.38	0.78	29.52	0.84	*27.80*	0.87	26.69	0.80
NLDJ [31]	30.06	0.71	28.11	0.76	27.96	0.72	27.74	*0.82*	29.62	*0.86*	27.61	*0.88*	27.10	*0.84*
NA/R-NL [16]	30.17	0.74	27.82	0.76	27.23	0.73	27.14	0.77	29.32	0.84	25.86	0.83	26.50	0.78
R-NL	*30.70*	*0.75*	*28.53*	*0.78*	*28.12*	*0.75*	*27.98*	*0.82*	*30.10*	*0.86*	27.50	*0.88*	*27.36*	0.83
BM3D [55]	**32.22**	**0.77**	**29.31**	**0.82**	**28.66**	**0.78**	**29.04**	**0.87**	**31.25**	**0.89**	**29.76**	**0.92**	**28.00**	**0.87**
Gaussian noise, $\sigma = 40$														
Noisy image	16.10	0.18	16.10	0.27	16.12	0.24	16.09	0.51	16.11	0.46	16.09	0.57	16.10	0.59
TV [1]	24.71	0.60	23.08	0.67	22.18	0.57	23.82	0.73	24.53	0.76	21.96	0.73	22.75	0.73
NL-TV [12]	24.87	0.50	25.45	0.61	25.81	0.54	25.99	0.76	25.99	0.76	24.45	0.80	25.13	*0.78*
NL-Means [3]	28.25	0.66	26.32	0.70	26.08	0.67	26.03	0.72	28.10	0.80	*26.07*	0.82	25.38	0.73
NLDJ [31]	28.28	0.64	26.43	0.71	26.54	0.67	26.34	*0.76*	28.22	0.82	25.92	*0.84*	25.74	*0.78*
NA/R-NL [16]	27.89	*0.71*	25.09	0.70	24.71	0.67	25.16	0.68	27.05	0.78	23.99	0.76	24.60	0.69
R-NL	*29.05*	*0.71*	*26.85*	*0.75*	*26.70*	*0.72*	*26.61*	*0.76*	*28.77*	*0.83*	25.72	0.83	*26.04*	*0.78*
BM3D [55]	**30.65**	**0.74**	**27.77**	**0.78**	**27.27**	**0.73**	**27.67**	**0.82**	**29.86**	**0.86**	**27.96**	**0.89**	**26.72**	**0.83**
Poisson noise, $Q = 4$														
Noisy image	20.70	0.31	21.23	0.45	21.38	0.84	20.97	0.70	21.21	0.64	21.40	0.76	22.59	0.84
$\mathcal{P}$-TV [46]	30.31	0.72	29.25	0.78	28.67	0.73	28.66	0.86	30.35	0.87	25.99	0.86	28.91	0.92
$\mathcal{P}$-NL-Means [27]	31.62	0.75	29.48	0.80	28.98	0.77	28.79	0.83	31.36	0.88	*29.55*	0.91	28.74	0.89
$\mathcal{P}$-NLDJ [31]	31.68	0.75	29.91	0.81	29.91	0.80	29.28	*0.87*	31.58	*0.90*	29.44	*0.92*	29.49	*0.92*
NA/R-NL [16]	31.43	0.76	29.44	0.80	29.00	0.78	28.40	0.81	31.22	0.88	27.19	0.88	28.82	0.88
R-NL	*32.19*	*0.78*	*30.22*	*0.83*	*29.96*	*0.81*	*29.43*	0.86	*31.95*	*0.90*	29.14	*0.92*	*29.67*	0.91
Ans. + BM3D [56]	**33.30**	**0.80**	**30.89**	**0.85**	**30.35**	**0.82**	**30.42**	**0.90**	**32.88**	**0.92**	**31.23**	**0.94**	**30.17**	**0.93**
Poisson noise, $Q = 8$														
Noisy image	17.71	0.23	18.22	0.35	18.36	0.33	17.96	0.59	18.18	0.53	18.39	0.67	19.63	0.76
$\mathcal{P}$-TV [46]	28.74	0.69	27.49	0.74	26.89	0.67	27.12	0.81	28.82	0.83	24.43	0.81	27.23	*0.88*
$\mathcal{P}$-NL-Means [27]	28.69	0.65	27.37	0.71	27.17	0.68	26.66	0.72	29.04	0.81	26.81	0.83	26.83	0.79
$\mathcal{P}$-NLDJ [31]	29.57	0.69	28.20	0.76	28.18	0.73	27.67	*0.81*	29.82	0.86	*27.47*	*0.88*	27.87	0.87
NA/R-NL [16]	29.67	*0.74*	27.82	0.76	27.56	0.75	27.03	0.75	29.71	0.85	25.71	0.83	27.30	0.83
R-NL	*30.18*	*0.74*	*28.59*	*0.79*	*28.35*	*0.77*	*27.88*	*0.81*	*30.33*	*0.87*	27.29	*0.88*	*28.08*	0.87
Ans. + BM3D [56]	**31.78**	**0.76**	**29.23**	**0.81**	**28.80**	**0.78**	**28.88**	**0.86**	**31.36**	**0.89**	**29.57**	**0.92**	**28.56**	**0.90**
Poisson noise, $Q = 12$														
Noisy image	15.91	0.19	16.43	0.30	16.54	0.29	16.23	0.53	16.42	0.48	16.65	0.61	17.86	0.70
$\mathcal{P}$-TV [46]	27.07	0.59	26.10	0.67	25.67	0.59	25.95	0.77	27.18	0.77	23.60	0.76	25.96	0.84
$\mathcal{P}$-NL-Means [27]	27.76	0.62	26.71	0.70	26.36	0.67	26.15	0.71	28.54	0.80	26.26	0.82	26.47	0.82
$\mathcal{P}$-NLDJ [31]	28.29	0.63	27.20	0.71	27.04	0.67	26.76	*0.78*	28.82	0.83	*26.46*	*0.85*	27.03	*0.85*
NA/R-NL [16]	28.53	*0.72*	26.84	0.74	26.49	0.72	26.07	0.71	28.81	0.83	24.44	0.79	26.49	0.80
R-NL	*29.12*	0.71	*27.75*	*0.76*	*27.23*	*0.73*	*27.04*	*0.78*	*29.49*	*0.85*	26.21	*0.85*	*27.31*	*0.85*
Ans. + BM3D [56]	**30.75**	**0.74**	**28.29**	**0.78**	**27.73**	**0.74**	**27.93**	**0.83**	**30.38**	**0.86**	**28.46**	**0.89**	**27.73**	**0.87**
Gamma noise, $L = 12$														
Noisy image	21.68	0.37	22.33	0.52	22.36	0.51	22.12	0.74	22.44	0.70	22.68	0.82	24.39	0.88
$\mathcal{G}$-TV [47]	29.36	0.65	29.13	0.76	29.02	0.71	28.50	0.86	29.84	0.84	27.18	0.89	29.65	0.93
$\mathcal{G}$-NL-Means [3]	32.24	0.77	29.72	0.82	29.39	0.80	29.35	0.85	32.22	0.91	30.22	0.93	29.79	0.92
$\mathcal{G}$-NLDJ [31]	32.59	0.77	30.69	0.83	30.68	0.83	30.01	*0.89*	32.58	*0.92*	*30.42*	*0.94*	30.82	*0.94*
NA/R-NL [16]	32.53	0.78	30.65	0.83	30.10	0.82	29.40	0.84	32.67	0.91	28.59	0.91	30.53	0.92
R-NL	*33.09*	*0.80*	*31.12*	*0.85*	*30.87*	**0.84**	*30.22*	0.88	*33.05*	*0.92*	30.33	*0.94*	*31.03*	*0.94*
log-BM3D [57], [58]	**34.01**	**0.82**	**31.70**	**0.87**	**30.99**	**0.84**	**31.10**	**0.91**	**33.81**	**0.93**	**32.01**	**0.95**	**31.45**	**0.95**
Gamma noise, $L = 4$														
Noisy image	16.94	0.24	17.60	0.37	17.66	0.40	17.40	0.59	17.75	0.55	17.95	0.68	19.65	0.78
$\mathcal{G}$-NL-Means [3]	29.20	0.69	27.19	0.74	26.81	0.71	26.96	0.75	29.62	0.84	*27.20*	0.86	27.21	0.85
$\mathcal{G}$-NLDJ [31]	29.20	0.66	27.56	0.74	27.65	0.72	27.48	*0.81*	29.78	0.86	27.15	*0.88*	28.11	*0.89*
NA/R-NL [16]	30.03	*0.74*	27.78	0.77	27.49	0.75	27.21	0.76	30.17	*0.87*	25.86	0.84	28.39	0.88
R-NL	*30.15*	*0.74*	*28.09*	*0.78*	*27.99*	*0.76*	*27.86*	*0.81*	*30.46*	*0.87*	27.18	*0.88*	*28.49*	*0.89*
log-BM3D [57], [58]	**31.41**	**0.76**	**28.79**	**0.80**	**28.50**	**0.78**	**28.60**	**0.85**	**31.15**	**0.88**	**29.21**	**0.91**	**28.81**	**0.91**

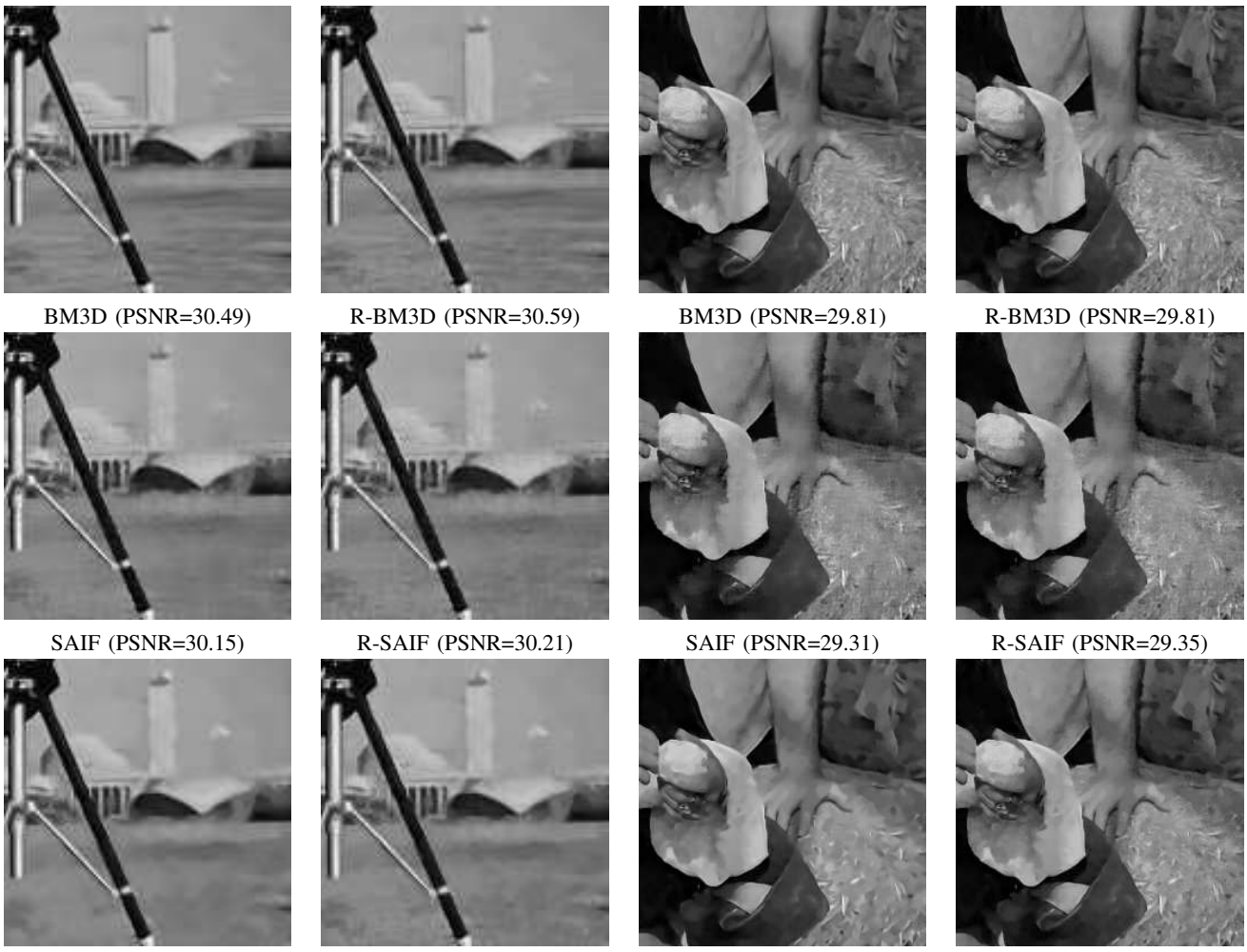


Fig. 8. Denoising of Gaussian noise (standard deviation σ=20) using BM3D, SAFIR and SAIF and associated regularized results R-BM3D, R-SAFIR and R-SAIF using the proposed *dejittering* and adaptive regularization. The adaptive regularization corrects some artifacts when necessary, for example on the grass and the small tower from "cameraman" or the straw from "man". However, thanks to the adaptivity of the model, it does not degrade the result when the targeted defaults do not occur, for example on the man with BM3D that does not suffer from either the *jittering effect* or the *rare patch effect*.

that the over-smoothing of fine structures such as the grass on "cameraman" and the straw on "man" is corrected, and that some artifacts (around the tower) are reduced. Besides, thanks to the adaptivity of our approach, when no *jittering* or *rare patch effect* has occurred, the noise reduction is estimated to be safe so no further correction is applied.

X. Conclusion

In this paper, we have presented a new approach that combines the NL-means with TV regularization. We use a weighted non-local data-fidelity term, whose magnitude is driven by the estimation of the amount of denoising performed by the NL-means. The variance reduction offers a measure of confidence in the denoising performed by the NL-means, and the TV regularization is automatically adapted according to this value. This leads to a flexible algorithm that locally uses both the redundant and smooth properties of the images while offering both a reduction of the *jittering* and the *rare patch* effects associated to the NL-means and of the *staircasing effect* that occur in the TV minimization.

The proposed model has an intuitive interpretation that can be generalized to different noise statistics. Moreover, we have shown that a solution can be achieved quite simply for types of noises belonging to the exponential family.

We also propose an extension to video denoising that does not need motion estimation. The method offers good performances and guarantees temporal stability thanks to the use of three-dimensional patches.

While we have developed our regularized NL-means based on TV regularization, our model naturally extends to other regularization terms adapted to different image priors. Investigating such regularizations is the topic of future work. Beyond non-local denoising, we could also extend this model to other types of algorithms based on weighted averages, hence extending the applicability of this model to numerous other potential problems.

Acknowledgment

C. Sutour would like to thank the DGA and the Aquitaine region for funding her PhD. J.-F. Aujol acknowledges the

support of the Institut Universitaire de France. This study has been carried out with financial support from the French State, managed by the French National Research Agency (ANR) in the frame of the "Investments for the future" Programme IdEx Bordeaux - CPU (ANR-10-IDEX-03-02). The authors would also like to thank Guy Gilboa for the algorithms of the non-local TV he put at our disposition, and for the advice he gave us. We would also like to thank Cécile Louchet for the discussions we had regarding the TV-means algorithm.

REFERENCES

[1] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60(1):259-268, 1992.

[2] G. Gilboa, N. Sochen, and Y. Y. Zeevi, "Variational denoising of partly textured images by spatially varying constraints," *Image Processing, IEEE Transactions on*, vol. 15, no. 8, pp. 2281–2289, 2006.

[3] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling and Simulation*, vol. 4(2), pp. 490–530, Sep 2005.

[4] M. Mahmoudi and G. Sapiro, "Fast image and video denoising via non-local means of similar neighborhoods," *IEEE Signal Processing Letters*, vol. 12(12), 2005.

[5] C. Kervrann and J. Boulanger, "Optimal spatial adaptation for patch-based image denoising," *IEEE Trans. Image Processing*, vol. 15(10):2866-2878, 2006.

[6] ——, "Local adaptivity to variable smoothness for exemplar-based image regularization and representation," *International Journal of Computer Vision*, vol. 79, no. 1, pp. 45–69, 2008.

[7] M. V. de Ville and D. Kocher, "Non-local means with dimensionality reduction and SURE-based parameter selection," *IEEE Transactions on image processing*, vol. 20(9):2683-2690, 2011.

[8] H. Talebi, X. Xhu, and P. Milanfar, "How to SAIF-ly boost denoising performance," *IEEE Transactions on image processing*, vol. 22(4):1470-1485, 2013.

[9] C. Louchet and L. Moisan, "Total variation as a local filter," *SIAM Journal on Imaging Sciences*, vol. 4(2):651-694, 2011.

[10] C.-A. Deledalle, V. Duval, and J. Salmon, "Non-local methods with shape-adaptive patches (nlm-sap)," *Journal of Mathematical Imaging and Vision*, pp. 1–18, 2011.

[11] S. Kindermann, S. Osher, and P. Jones, "Deblurring and denoising of images by nonlocal functionals," *SIAM Journal Multiscale Model. Simul.*, vol. 4(4):1091-1115, 2005.

[12] G. Gilboa and S. Osher, "Nonlocal operators with applications to image processing," *Multiscale Modeling and Simulation*, vol. 7(3):1005-1028, 2008.

[13] G. Peyré, S. Bougleux, and L. D. Cohen, "Non-local regularization of inverse problems," *Inverse Problems and Imaging*, vol. 5, no. 2, pp. 511–530, 2011.

[14] X. Zhang, M. Burger, X. Bresson, and S. Osher, "Bregmanized non-local regularization for deconvolution and sparse reconstruction," *SIAM Journal Imaging Sciences*, vol. 3(3), pp. 253–276, 2010.

[15] M. Protter, M. Elad, H. Takeda, and P. Milanfar, "Generalizing the non-local means to super-resolution reconstruction," *IEEE Transactions on image processing*, vol. 18(1):36-51, 2009.

[16] E. d'Angelo and P. Vandergheynst, "Fully non-local super-resolution via spectral hashing," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1137–1140, 2011.

[17] M. Mignotte, "A non-local regularization strategy for image deconvolution," *Pattern recognition letters*, vol. 29, pp. 2206–2212, 2008.

[18] A. Buades, B. Coll, and J.-M. Morel, "Denoising image sequences does not require motion compensation," *Proc. IEEE Conf. Advanced Video and Signal based Surveillance*, pp. 70–74, 2005.

[19] Y. Wexler, E. Shechtman, and M. Irani, "Space-time completion of video," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 29(3):463-476, March 2007.

[20] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical Imaging and Vision*, vol. 20:89-97, 2004.

[21] P. Combettes and J. Pesquet, "Proximal splitting methods in signal processing," January 2010.

[22] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40(1):120-145, 2011.

[23] P. C. Hansen and D. P. O'Leary, "The use of the l-curve in the regularization of discrete ill-posed problems," *SIAM Journal on Scientific Computing*, vol. 14, no. 6, pp. 1487–1503, 1993.

[24] S. Ramani, T. Blu, and M. Unser, "Monte-carlo sure: A black-box optimization of regularization parameters for general denoising algorithms," *Image Processing, IEEE Transactions on*, vol. 17, no. 9, pp. 1540–1554, 2008.

[25] J. Polzehl and V. Spokoiny, "Propagation-separation approach for local likelihood estimation," *Probability Theory and Related Fields*, vol. 135, no. 3, pp. 335–362, 2006.

[26] C.-A. Deledalle, L. Denis, and F. Tupin, "Iterative weighted maximum likelihood denoising with probabilistic patch-based weights," *IEEE Trans. on Image Processing*, vol. 18(12):2661-2672, 2009.

[27] ——, "How to compare noisy patches ? Patch similarity beyond Gaussian noise," *International Journal of Computer Vision*, vol. 99(1):86-102, 2012.

[28] C. Kervrann, J. Boulanger, and P. Coupé, "Bayesian non-local means filter, image redundancy and adaptive dictionaries for noise removal," in *Scale Space and Variational Methods in Computer Vision*. Springer, 2007, pp. 520–532.

[29] P. Coupé, P. Hellier, C. Kervrann, and C. Barillot, "Bayesian non local means-based speckle filtering," in *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*. IEEE, 2008, pp. 1291–1294.

[30] H. Zhong, Y. Li, and L. Jiao, "Sar image despeckling using bayesian nonlocal means filter with sigma preselection," *Geoscience and Remote Sensing Letters, IEEE*, vol. 8, no. 4, pp. 809–813, 2011.

[31] C.-A. Deledalle, L. Denis, F. Tupin, A. Reigber, and M. Jäger, "NL-SAR: a unified non-local framework for resolution-preserving (pol)(in) SAR denoising," *Preprint*, 2013.

[32] V. Duval, J.-F. Aujol, and Y. Gousseau, "A biais-variance approach for the non-local means," *SIAM Journal on Imaging Sciences*, vol. 4(2):760-788, 2011.

[33] J.-S. Lee, "Refined filtering of image noise using local statistics," *Computer graphics and image processing*, vol. 15, no. 4, pp. 380–389, 1981.

[34] D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel, "Adaptive noise smoothing filter for images with signal-dependent noise," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, no. 2, pp. 165–177, 1985.

[35] A. Chambolle and P. Lions, "Image recovery via total variation minmization and related problems," *Numer. Math.*, vol. 76:167-188, 1997.

[36] D. L. Donoho and J. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[37] S. P. Awate and R. T. Whitaker, "Unsupervised, information-theoretic, adaptive image filtering for image restoration," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 3, pp. 364–376, 2006.

[38] T. Brox and D. Cremers, "Iterated nonlocal means for texture restoration," in *Scale Space and Variationnal Methods in Computer Vision*, vol. 4485. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2007, pp. 13–24.

[39] A. Buades, B. Coll, and J.-M. Morel, "Image enhancement by non-local reverse heat equation," *Technical report 22, CMLA, ENS-Cachan, Cachan, France*, 2006.

[40] P. Arias, V. Caselles, and G. Facciolo, "Analysis of a variational framework for exemplar-based image inpainting," *Multiscale Modeling & Simulation*, vol. 10, no. 2, pp. 473–514, 2012.

[41] D. Zhou and B. Scholkopf, "A regularization framework for learning from graph data," *Workshop on Statistical Relational Learning and its Connections to other fields*, 2004.

[42] S. Bougleux, A. Elmoataz, and M. Melkemi, "Discrete regularization on weighted graphs for image and mesh filtering," in *1st International Conference on Scale Space and Variationnal Methods in Computer Vision (SSVM)*, vol. 4485. Lecture Notes in Computer Science, 2007, pp. 128–139.

[43] G. Gilboa and S. Osher, "Nonlocal linear image regularization and supervised segmentation," *SIAM Multiscale Modeling and Simulation*, vol. 6(2):595-630, 2007.

[44] M. Collins, S. Dasgupta, and R. E. Schapire, "A generalization of principal components analysis to the exponential family," *NIPS*, pp. 617–624, 2002.

[45] J. Darbon, A.Cunha, T. Chan, S. Osher, and G. Jensen, "Fast non-local filtering applied to electron cryomicroscopy," *IEEE Int. Symposium on Biomedical Imaging (ISBI)*, pp. 1331–1334, 2008.

[46] S. Anthoine, J.-F. Aujol, Y. Boursier, and C. Mélot, "Some proximal methods for poisson intensity CBCT and PET," *Inverse Problems and Imaging*, vol. 6(4):565-598, 2012.
[47] G. Aubert and J.-F. Aujol, "A variational approcah to removing multiplicative noise," *SIAM Journal on Applied Mathematics*, vol. 68(4):925-946, 2008.
[48] P. Combettes and V. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling and Simulation*, vol. 4(4):1168-1200, 2005.
[49] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. on Imag. Sci.*, vol. 2(1), 2009.
[50] H. Raguet, J. Fadili, and G. Peyré, "Generalized forward-backward splitting," *SIAM Journal of Imaging Sciences*, 2012.
[51] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE TIP*, vol. 18(11), 2009.
[52] A. Buades, B. Coll, and J.-M. Morel, "Nonlocal image and movie denoising," *International Journal of Computer Vision*, vol. 76(2), pp. 123–139, 2008.
[53] C. Liu and W. Freeman, "A high-quality video denoising algorithm based on reliable motion estimation," *European Conference on Computer Vision (ECCV)*, pp. 706–719, 2010.
[54] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004.
[55] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," *IEEE Trans. on Image Processing*, vol. 16(8):2080-2095, 2007.
[56] M. Mäkitalo and A. Foi, "Poisson-gaussian denoising using the exact unbiased inverse of the generalized anscombe transformation," *ICASSP*, pp. 1081–1084, 2012.
[57] H. Xie, L. Pierce, and F. Ulaby, "Statistical properties of logarithmically transformed speckle," *IEEE Trans. Geosci. Remote Sens.*, vol. 40(3):721-727, 2002.
[58] M. Makitalo, A. Foi, D. Fevralev, and V. Lukin, "Denoising of single-look SAR images based on variance stabilization and nonlocal filters," in *Mathematical Methods in Electromagnetic Theory (MMET), 2010 International Conference on*. IEEE, 2010, pp. 1–4.
[59] K. Dabov, A. Foi, and K. Egiazarian, "Video denoising by sparse 3D transform-domain collaborative filtering," in *Proc. 15th European Signal Processing Conference*. EUSIPCO 2007, Poznan, Poland, 2007.
[60] M. Maggioni, G. Boracchi, A. Foi, and K. Egiazarian, "Video denoising, deblocking and enhancement through separable 4-D nonlocal spatiotemporal transforms," *IEEE Transactions on image processing*, vol. 21(9):3952-3966, 2012.