A patch-based approach for removing mixed Gaussian-impulse noise

Julie Delon, Agnès Desolneux

▶ **To cite this version:**

Julie Delon, Agnès Desolneux. A patch-based approach for removing mixed Gaussian-impulse noise. 2012. hal-00718612v1

HAL Id: hal-00718612

https://hal.science/hal-00718612v1

Preprint submitted on 17 Jul 2012 (v1), last revised 20 Dec 2012 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A PATCH-BASED APPROACH FOR REMOVING MIXED GAUSSIAN-IMPULSE NOISE *

JULIE DELON † AND AGNÈS DESOLNEUX ‡

Abstract. In this paper, we address the problem of the restoration of images which have been affected by a mixture of Gaussian and impulse noise. We rely on a patch-based approach, which requires careful choices for both the distance between patches and for the statistical estimator of the original patch. Experiments are led in the case of pure impulse noise and in the case of a mixture. The method proves to be particularly powerful, especially for the restoration of textured regions, and compares favorably to recent restoration methods.

Key words. restoration, denoising, Gaussian noise, impulse noise, non local methods, image patch, maximum likelihood estimator, order statistics.

1. Introduction. Gaussian noise and impulse noise generally appear during the acquisition or the transmission of images. It is not always possible to separate them, and the user has thus to deal with an image that is damaged by a mixture of Gaussian and impulse noise (it means that the image is first affected by Gaussian noise, and then by impulse noise). Removing this kind of noise while preserving image details and textures is of great importance before most image analysis tasks (edge detection, segmentation, etc). The Gaussian noise is an additive noise, and it is very different in nature from the impulse noise. Two models of impulse noise are generally used in the literature. In the first one, called *salt-and-pepper noise*, each gray level is replaced with a given probability by 0 or M, where $[0, M]$ designs the range of the original image ($M = 255$, in general). In this paper, we will focus on the second model of impulse noise, called *random-valued impulse noise*, where each gray level value is replaced with probability p, called *noise ratio*, by a random value in the set $\{0, 1, \ldots, M\}$. Observe that detecting and removing random-valued impulse noise is much more difficult in practice than removing salt-and-pepper noise.

There is a broad literature on the pure impulse noise case, but the mixture of Gaussian and impulse noise is generally less studied, despite the fact that it is a more realistic noise model. The traditional approaches for impulse noise removal act locally and non linearly on images. Among them, let us mention the median and its extensions [29, 24]. These approaches modify all pixels indifferently, while impulse noise affects only a portion of the pixels. In order to avoid this shortcoming, the trend for nearly twenty years has been to propose different impulse noise detectors and to restrict the restoration to pixels detected as corrupted. For instance, this idea underlies the switching median filter [31], the adaptive center weighted median filter (ACWMF) [8] or the pixel-wise median absolute deviation [9]. Unfortunately, median-based methods tend to destroy details and textures in images when the noise ratio is large. A successful alternative consists in combining a well chosen impulse detector, generally relying on local order statistics of gray level differences, with a global or at least semi-local restoration approach. This is the case of [20], which combines the Rank Order Absolute Differences (ROAD) for detection with a trilateral filter for restoration. Moreover, this method is used in both cases: pure impulse noise removal or mixture of Gaussian and impulse noise removal. This detection/restoration scheme is also followed by the

*The authors would like to thank Vincent Duval, for his fruitful comments and suggestions, as well as Yiqiu Dong, who kindly provided us with her ROLD-EPR denoising code.

†LTCI, TÉLÉCOM PARIS (CNRS UMR 5141) 46 RUE BARRAULT, 75634 PARIS CEDEX 13 (JULIE.DELON 'AT' ENST.FR).

‡CMLA (CNRS UMR 8536), ENS CACHAN, 61 AVENUE DU PRÉSIDENT WILSON, 94235 CACHAN CEDEX (AGNES.DESOLNEUX 'AT' CMLA.ENS-CACHAN.FR).

authors of [5, 17], who rely on ACWMF, or ROLD (Rank Ordered Logarithmic Difference) for detection, before applying a variational approach to restore corrupted pixels.
Other approaches for removing a mixture of Gaussian and impulse noise generally start by estimating or detecting the impulse noise and then adapt (by reducing the influence of impulse affected pixels) a Gaussian noise removal method. This is for instance the case of [33], where outliers are first detected by a median type filter and then a K-SVD dictionary learning is performed on impulse-free pixels to finally solve a $l_1 - l_0$ minimization problem. This is also the case of [21] where an impulse noise estimator is used to modify an anisotropic diffusion model, or the case of [34] where impulse noise is first removed by an adaptive cascade of median filters and then, in a second stage, Gaussian noise is removed (using a BM3D filter). The method of [27] takes advantage of the full probabilistic mixture model: it first executes an external method to have a first prediction of original pixel values, then it uses EM to learn the full noise model, and finally it performs some kernel regression to get the restored image. The authors of [26] use the ROAD detector of [20] to obtain a modified ("impulse controlled") distance between image patches that they use to compute the weights in the denoising scheme of Non Local means (NL-means) [4].
The approach presented in this paper for removing mixed Gaussian-impulse noise will also be patch-based, which means that it will rely on the patch redundancy inside images. In the last fifteen years, a great deal of image processing techniques have been developed in order to take advantage of self-similarities of images [19, 1, 4, 15]. Since the introduction of the NL-means by Buades et al [4] to tackle Gaussian noise, the idea of relying on patch redundancy to reduce noise variance has proved to be particularly powerful. The NL-means have since been extended with success to other noise models [22, 11]. The idea of these restoration methods is both simple and nice, since it relies on the assumption that in a natural image, a given patch can be found almost identically in different places. The patch can then be restored in two steps : first, by finding all of its corresponding patches in the image, and second by estimating the real underlying patch behind these different damaged versions. The mathematical framework adapted to deal with this redundancy is the one of statistical estimation, as underlined in the recent contributions [23, 32, 30, 18]. The goal of this paper is to model properly the two steps of this estimation scheme in the case of images suffering from a mixture of impulse and Gaussian noise.

The paper is organized as follows. We start in Section 2 by giving some details about the important steps of patch-based approaches for denoising. In Section 3, we compare different measures of similarity between patches, designed to be robust to impulse and Gaussian noise. The estimation step is then tackled in Section 4, in which we also analyse the statistical properties of several possible estimators. In Section 5 we present in details our full denoising scheme. For practical reasons, this scheme is named PARIGI [1] in the paper. Experiments and comparisons with recent approaches are displayed in Section 6. A conference proceeding version of this work has appeared in [16].

2. Patch-based approaches for denoising.

2.1. Noise model. In the rest of the paper, the discrete damaged image is denoted by u and the original image is denoted by u^0. These images are defined on a discrete domain Ω, assumed to be a bounded rectangle of $\mathbb{Z}^2$. We consider in this paper that the damaged image u is a realization of a random image U which can be written

$$U = (1 - T).V + T.W, \tag{2.1}$$

[1] PARIGI stands for *Patch based Approach for the Restoration of Images affected by Gaussian and Impulse noise.*

where T, V and W are independent random images such that

- the values $T(x),\ x \in \Omega$ are i.i.d. random variables with a Bernoulli distribution of parameter p (in the sense that $\mathbb{P}[T(x) = 1] = p$); p is called the *noise ratio*;
- the values $W(x),\ x \in \Omega$ are i.i.d. random variables with a discrete uniform distribution on the range $[0, M]$;
- the values $V(x),\ x \in \Omega$ are independent random variables and $V(x) \sim \mathcal{N}(u^0(x), \sigma^2)$.

In other words, the value $u^0(x)$ of a pixel x is either replaced (with probability p) by a uniform value on $[0, M]$ or modified (with probability $1 - p$) by adding a centered Gaussian noise. Impulse noise can be considered as a particular case of the previous model, obtained when $\sigma = 0$ (in this case, the Gaussian distribution degenerates at each point x into a Dirac centered at $u^0(x)$).

2.2. The fundamentals of patch-based denoising. The idea of patch-based approaches is to take advantage of the redundancy of geometry in images. The central hypothesis of these approaches is that for each patch $\mathbf{P}_1$ in u, there exist other patches $\mathbf{P}_2, \ldots, \mathbf{P}_n$ in u such that all the $\mathbf{P}_i$ are realizations of the same random patch, obtained from an underlying non-noisy patch $\mathbf{P}^0$ of the original image u^0. The first step of the whole restoration process consists in finding, for each patch $\mathbf{P}_1$, this set of corresponding replicas. In a second step, a denoised version of $\mathbf{P}_1$ can be obtained by relying on any estimator $\widehat{\mathbf{P}}$ of $\mathbf{P}^0$, computed from its different damaged versions $\mathbf{P}_1, \ldots, \mathbf{P}_n$.

The first step generally boils down to the choice of a similarity measure between patches. This measure should constitute a good trade-off between discriminative power and robustness to noise, as presented in Section 3. In the classical version of NL-means [4], the similarity between patches relies on the euclidean distance, chosen in particular for its robustness to Gaussian noise. This distance is used to compute dissimilarity weights, which permit to balance the influence of the different patches in the final estimator $\widehat{\mathbf{P}}$. Alternatively, the authors of [28] propose to threshold the similarity measure in such a way that *"a patch and one of its replicas have a probability of 0.99 to be considered as similar"*. The estimator $\widehat{\mathbf{P}}$ keeps only patches whose distance to $\mathbf{P}_1$ is below this threshold. Other valid solutions only retain the n-nearest neighbors of $\mathbf{P}_1$ [2], or make use of clustering [3, 6] or of PCA [14] strategies in order to determine the set of replicas for $\mathbf{P}_1$. For the sake of simplicity, the denoising scheme PARIGI described in this paper relies on the n-nearest neighbors solution for this first step, the main difficulty to be tackled being the design of a robust similarity measure in presence of impulse noise.

The second step of the restoration amounts to a very classical estimation problem : given $U_1, \ldots, U_n$ n i.i.d. random vectors following a distribution $\mathcal{P}(\theta)$ with θ a vector parameter, find a good estimator $\hat{\theta}$ of θ. When the noise is purely Gaussian, $\mathcal{P}(\theta) = \mathcal{N}(\theta, \sigma^2)$, with σ known, the best estimator in terms of Mean Squared Error (or quadratic risk) is the mean, which explains the usual averaging formulation of the NL-means [4]. Recently, many authors have adopted the point of view of quadratic risk minimization as a way to optimize the parameters of the NL-means [23, 32, 18] or to propose further improvements [30]. In Section 4.2, we will also make use of the quadratic risk in order to compare the efficiency of different estimators in presence of impulse and Gaussian noise.

These two steps permit to restore any patch in the noisy image. Now, since a given pixel x belongs to several patches, multiple choices are possible to obtain a restored value $\hat{u}(x)$ at x. This last step, called *reprojection* in the work of Salmon and Strozecki [30], corresponds to the general framework of estimators aggregation in statistics. Let us denote by $\mathbf{P}_x$ a patch centered at x (in this paper, we assume for the sake of simplicity that patches are $(2f + 1) \times (2f + 1)$ squares, but more general shapes could be considered, as proposed in [13]). The pixel x belongs to the support of all the patches $\mathbf{P}_{x+\delta}$, $\delta \in [-f, f] \times [-f, f]$. In order to restore the value $u^0(x)$, we can thus rely

on all the values $\widehat{\mathbf{P}_{x+\delta}}(-\delta)$, $\delta \in [-f,f] \times [-f,f]$. A common choice for the estimator of $u^0(x)$ is to take only into account the value $\widehat{\mathbf{P}_x}(0)$ of the restored patch centered at x, as proposed in the original NL-means [4]. Another natural choice, studied for instance in [30] is to take the mean

$$\frac{\sum_{\delta \in [-f,f]^2} \widehat{\mathbf{P}_{x+\delta}}(-\delta)}{(2f+1)^2}. \tag{2.2}$$

The interest of this simple aggregation is that it divides the variance of the estimator of $u^0(x)$ by $(2f+1)^2$ if the different estimations $\widehat{\mathbf{P}_{x+\delta}}(-\delta)$, $\delta \in [-f,f]^2$, are i.i.d. This assumption is not true in practice since the patches are overlapping, but the quality of the results is nevertheless visibly improved by this procedure, at least in the case of Gaussian noise. In this paper, we propose to merge this aggregation step with the estimation of the denoised patches. More precisely, the final estimator $\widehat{u(x)}$ of $u^0(x)$, described in Section 5, will rely on all the values $u(y)$ for which there exists $\delta \in [-f,f]^2$ such that $\mathbf{P}_{x+\delta}$ and $\mathbf{P}_{y+\delta}$ are similar.

3. Robust distance between patches. This section focuses on the quest for similar patches in an image affected by a mixture of impulse and Gaussian noise. As underlined before, the success of any patch-based denoising procedure depends greatly on the ability to find the replicas of a patch in the noisy image u. More precisely, for a given patch $\mathbf{P}$ in u, we aim at discovering all patches $\mathbf{Q}$ such that the unknown original patches $\mathbf{P}^0$ and $\mathbf{Q}^0$ in u^0 are equal or at least similar. In presence of impulse noise, the Euclidean (L^2) distance between $\mathbf{P}$ and $\mathbf{Q}$ contains outliers and cannot be trusted. In the following, we compare several alternative measures designed to be robust to this kind of noise.

3.1. Generalized likelihood ratio. A first possibility, suggested by the recent work [12], is to make use of the generalized maximum likelihood ratio

$$GLR(\mathbf{P},\mathbf{Q}) = \frac{\sup_{\mathbf{T}} \mathbb{P}[\mathbf{P},\mathbf{Q}|\mathbf{P}^0=\mathbf{Q}^0=\mathbf{T}]}{\sup_{\mathbf{T}} \mathbb{P}[\mathbf{P}|\mathbf{P}^0=\mathbf{T}] \sup_{\mathbf{T}} \mathbb{P}[\mathbf{Q}|\mathbf{Q}^0=\mathbf{T}]}. \tag{3.1}$$

In the case of pure impulse noise, this ratio becomes

$$GLR(\mathbf{P},\mathbf{Q}) = \left(1 + \frac{(M+1)(1-p)}{p}\right)^{2(n_{sim}-|\mathbf{P}|)}, \tag{3.2}$$

where n_{sim} is the number of pixels that have the same value in $\mathbf{P}$ and $\mathbf{Q}$, and $|\mathbf{P}| = (2f+1)^2$ is the size of $\mathbf{P}$. Obviously, this measure is too rigid in practice to be used even if the noise is purely impulsive. Indeed, two patches can be very similar and at the same time be such that $n_{sim} = 0$. In the mixed Gaussian-impulse case, the ratio becomes (with f_θ the Gaussian distribution with mean θ and variance σ^2):

$$GLR(\mathbf{P},\mathbf{Q}) = \frac{\sup_{\mathbf{T}} \prod_{x\in\mathbf{P}} \left(\frac{p}{M+1} + (1-p) f_{\mathbf{T}(x)}(\mathbf{P}(x))\right)\left(\frac{p}{M+1} + (1-p) f_{\mathbf{T}(x)}(\mathbf{Q}(x))\right)}{\prod_{x\in\mathbf{P}} \left(\frac{p}{M+1} + (1-p) f_{\mathbf{P}(x)}(\mathbf{P}(x))\right)\left(\frac{p}{M+1} + (1-p) f_{\mathbf{Q}(x)}(\mathbf{Q}(x))\right)}. \tag{3.3}$$

The denominator of this formula can be simplified into the constant $(\frac{p}{M+1} + \frac{1-p}{\sigma\sqrt{2\pi}})^{2|\mathbf{P}|}$, independent of $\mathbf{P}$ and $\mathbf{Q}$. The numerator is more complex. It consists of a product of different terms $num(x)$

when x spans $\mathbf{P}$: $num(x)$ equals $(\frac{p}{M+1} + \frac{1-p}{\sigma\sqrt{2\pi}}e^{-(\mathbf{P}(x)-\mathbf{Q}(x))^2/8\sigma^2})^2$ when $(\mathbf{P}(x) - \mathbf{Q}(x))^2/\sigma^2$ is small and $(\frac{p}{M+1} + \frac{1-p}{\sigma\sqrt{2\pi}})(\frac{p}{M+1} + \frac{1-p}{\sigma\sqrt{2\pi}}e^{-(\mathbf{P}(x)-\mathbf{Q}(x))^2/2\sigma^2})$ when $(\mathbf{P}(x) - \mathbf{Q}(x))^2/\sigma^2$ is large. As a consequence, the log of $GLR(\mathbf{P}, \mathbf{Q})$ is a sum over x of a function of $(\mathbf{P}(x) - \mathbf{Q}(x))^2$, the function depending on whether $(\mathbf{P}(x) - \mathbf{Q}(x))^2/2\sigma^2$ is small or not. The main difficulty when using this generalized likelihood ratio (GLR) is that we do not know σ in practice, but its main advantage is that it automatically adapts to large or small values of $(\mathbf{P}(x) - \mathbf{Q}(x))^2$. It is thus related to robust similarity measures, that we explain in the following section.

3.2. Weighted euclidean distances. A second possibility is to rely on a robust similarity measure, inspired by order statistics [10]. It is known that, for non-noisy patches, the L^2 distance is a good distance to measure the similarity between the patches. But when impulse noise is present, many large values of $(P-Q)_i^2, 1 \leq i \leq n$, are observed that are due to the noise and the L^2 distance then contains many outliers. Our distance is built to be similar to the euclidean distance but robust to large values created by the noise. Let $|P - Q|_{(1)} \leq |P - Q|_{(2)} \leq \cdots \leq |P - Q|_{(2f+1)^2}$ be the values obtained by ordering the $(2f + 1)^2$ values of the differences $|\mathbf{P}(y) - \mathbf{Q}(y)|$. We propose to rely on a distance of the form

$$D(\mathbf{P}, \mathbf{Q}) = \sum_{k=1}^{(2f+1)^2} w_k |P - Q|_{(k)}^2, \tag{3.4}$$

where $w_1, \ldots, w_n$ are positive weights. There is a wide choice for the weights w_k. For instance, if they are all equal, we find again the L^2 distance. If we set $w_k = 1$ for k below a number T and w_k equal to 0 above, we obtain a trimmed sum of order statistics : the distance relies only on the smallest distances between gray levels in the support of the patch. We denote this distance D_{trimmed} and we choose for T the value $T = \sup\{k; \mathcal{B}((2f + 1)^2, k, (1 - p)^2) > 0.99\}$, where $\mathcal{B}$ denotes the tail of the binomial distribution [2]. This choice comes from the following property: If $\mathbf{P}$ and $\mathbf{Q}$ are two independent random patches such that $\mathbf{P}^0 = \mathbf{Q}^0$, the probability that the k^{th} difference $|P - Q|_{(k)}$ stems from two untouched pixels is $\mathcal{B}((2f + 1)^2, k, (1 - p)^2)$ (with the approximation that the smallest distances correspond to untouched pixels).
In cases of low Signal to Noise Ratio, this trimmed distance becomes difficult to use, since the number T of pixels taken into account rapidly tends toward zero. In order to avoid this problem, we propose to keep all the weight w_k positive and equal to the probability that the value $|P - Q|_{(k)}$ stems from untouched pixels. We denote this distance D_{weighted}:

$$D_{\text{weighted}}(\mathbf{P}, \mathbf{Q}) = \sum_{k=1}^{(2f+1)^2} \mathcal{B}((2f + 1)^2, k, (1 - p)^2) |P - Q|_{(k)}^2. \tag{3.5}$$

Notice that when there is no impulse noise ($p = 0$), then the two distances D_{weighted} and D_{trimmed} are both the usual L^2 distance between patches.

The choice of the distance D_{weighted} is reinforced by modeling the problem the following simple way: assume the two patches $\mathbf{P}^0$ and $\mathbf{Q}^0$ are both constant patches, and denote by a and b their respective constant gray level value. Then, when the two patches are damaged independently by some impulse noise of intensity p, the random variables $(\mathbf{P} - \mathbf{Q})_i^2, 1 \leq i \leq n$, are i.i.d. following a distribution of the form $(1 - p)^2\delta_{(b-a)^2} + (1 - (1 - p)^2)dF$, where dF is a probability distribution on $\mathbb{R}_+$ (that we don't need to compute). In this framework, we have the following theorem.

[2] $\mathcal{B}(n, k, q) = \sum_{i=k}^{n} \binom{n}{i} q^i (1 - q)^{n-i}$.

THEOREM 3.1. *Let dF be a probability distribution on $\mathbb{R}_+$ and let $X_1^{(\alpha)}, \dots, X_n^{(\alpha)}$ be n i.i.d. random variables, distributed according to the mixture distribution $q\delta_\alpha + (1-q)dF$, where $q \in [0,1]$, and δ_α denotes the Dirac distribution at $\alpha \in [0, +\infty)$. Let $\mathbf{w} = (w_1, \dots, w_n)$ be a vector of weights in $[0, +\infty)$, and let*

$$D_{\mathbf{w}}(\alpha) = \mathbb{E}\left(\sum_{k=1}^{n} w_k X_{(k)}^{(\alpha)}\right),$$

where $X_{(k)}^{(\alpha)}$ denotes the k-th order statistic of $X_1^{(\alpha)}, \dots, X_n^{(\alpha)}$. Assume that the probability distribution dF satisfies $p_\alpha := \mathbb{P}(Y \leq \alpha) \to 0$ when $\alpha \to 0$ where Y is a random variable distributed following the law dF. Then when α is small, we have

$$D_{\mathbf{w}}(\alpha) = D_{\mathbf{w}}(0) + \alpha \sum_{k=1}^{n} w_k \mathcal{B}(n,k,q) + o(\alpha),$$

and notice that this first-order Taylor expansion is independent of the probability distribution dF.

Proof. Let us first introduce some notations. We will denote $m_{k,l}^+(\alpha)$ and $m_{k,l}^-(\alpha)$ the following expected values:

$$m_{k,l}^+(\alpha) = \mathbb{E}(Y_{(k)} \,|\, Y_1 \geq \alpha, \dots, Y_l \geq \alpha) \text{ and } m_{k,l}^-(\alpha) = \mathbb{E}(Y_{(k)} \,|\, Y_1 \leq \alpha, \dots, Y_l \leq \alpha),$$

where $Y_1, \dots, Y_l$ are i.i.d. random variables following the law dF. Let also N_α be the random variables that counts the number of $X_i^{(\alpha)}$, $1 \leq i \leq n$, that are equal to α. Then N_α follows the binomial distribution of parameters n and q, and we will denote, for $0 \leq k \leq n$, $b_k = \mathbb{P}(N_\alpha = k) = b(n,k,q) = \binom{n}{k} q^k (1-q)^{n-k}$. Now for $\alpha \geq 0$ we can write

$$\begin{aligned}
D_{\mathbf{w}}(\alpha) &= \sum_{k_0=0}^{n} \sum_{k=1}^{n} w_k \mathbb{E}\left(X_{(k)}^{(\alpha)} \,|\, N_\alpha = k_0\right) \mathbb{P}(N_\alpha = k_0) \\
&= \sum_{k_0=0}^{n} b_{k_0} \sum_{l=0}^{n-k_0} b(n-k_0, l, p_\alpha) \left(\sum_{k=1}^{l} w_k m_{k,l}^-(\alpha) + (w_{l+1} + \ldots + w_{l+k_0})\alpha \right. \\
&\quad \left. + \sum_{k=l+k_0+1}^{n} w_k m_{k-l-k_0, n-l-k_0}^+(\alpha) \right).
\end{aligned}$$

But notice that we also have, for any $\alpha \geq 0$,

$$D_{\mathbf{w}}(0) = \sum_{k_0=0}^{n} b_{k_0} \sum_{l=0}^{n-k_0} b(n-k_0, l, p_\alpha) \left(\sum_{k=1}^{l} w_{k+k_0} m_{k,l}^-(\alpha) + \sum_{k=l+k_0+1}^{n} w_k m_{k-l-k_0, n-l-k_0}^+(\alpha) \right).$$

Consequently, we get

$$D_{\mathbf{w}}(\alpha) = D_{\mathbf{w}}(0) + \sum_{k_0=0}^{n} b_{k_0} \sum_{l=0}^{n-k_0} b(n-k_0, l, p_\alpha) \left(\sum_{k=1}^{l} (w_k - w_{k+k_0}) m_{k,l}^-(\alpha) + (w_{l+1} + \ldots + w_{l+k_0})\alpha \right).$$

Finally, since $0 \leq m_{k,l}^{-}(\alpha) \leq \alpha$, and p_α goes to 0 when α goes to 0, we have that the first-order Taylor expansion is given by the term $l = 0$ in the sum above. More precisely, we get

$$\begin{aligned} D_{\mathbf{w}}(\alpha) &= D_{\mathbf{w}}(0) + \sum_{k_0=0}^{n} b_{k_0}(w_1 + \ldots + w_{k_0})\alpha + o(\alpha) \\ &= D_{\mathbf{w}}(0) + \alpha \sum_{k=1}^{n} w_k \mathcal{B}(n,k,q) + o(\alpha), \end{aligned}$$

which is the announced result. □

Now, among all the distances such that $\sum w_k^2$ is fixed, the one that is the most able to discriminate, on the average, two constant equal patches from two constant but non-equal patches is the one such that $\sum_{k=1}^{n} w_k \mathcal{B}(n,k,q)$ is as large as possible, and this is achieved when w_k is proportional to $\mathcal{B}(n,k,q)$ where $q = (1-p)^2$. We recognize the weighted distance defined in (3.5).

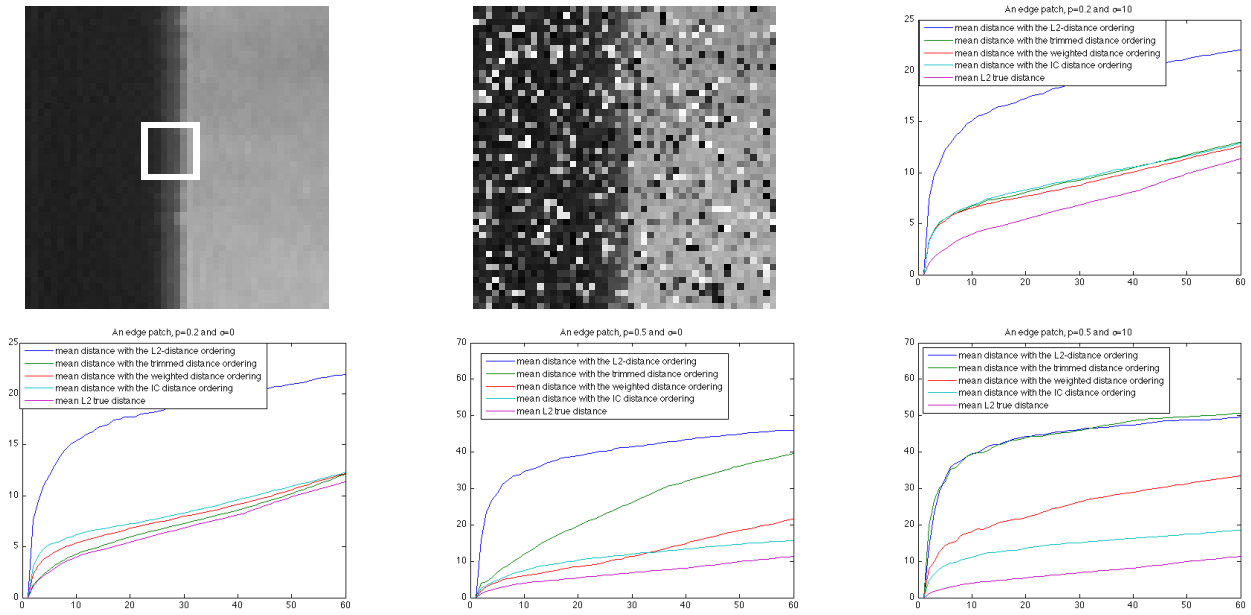


FIG. 3.1. *Comparison of different distances between an edge patch and similar patches. On the first raw, we show the original patch (white window) of size 7×7 pixels and its neighborhood V; a noisy version of the image ($p = 0.2$ and $\sigma = 10$); and the mean distances obtained in that case. On the second raw, we show the curves obtain for other noise parameters: respectively $p = 0.2$ and $\sigma = 0$, $p = 0.5$ and $\sigma = 0$, $p = 0.5$ and $\sigma = 10$. See the text for some comments on these graphics.*

3.3. Comparison of the different distances. When an image is not affected by noise, the L^2 distance is a good way to measure the similarity between patches. A nice way to compare different robust distances in presence of noise is to measure their ability to find similar patches at the same locations as those found by the L^2 distance in the non-noisy image. A robust distance should be able to preserve as much as possible the ordering of the patches provided by the L^2 distance in the non-noisy image. The comparison procedure works as follows. Given an original patch $\mathbf{P}^0$ in an original image u^0, and a neighborhood V (that is a square window of half-size t centered at the center of the patch), we will denote by $\mathbf{P}_x^0$ the patch of u^0 centered at $x \in V$. Let the previous patches in the noisy image u be respectively denoted by $\mathbf{P}$ and $\mathbf{P}_x$. For a distance

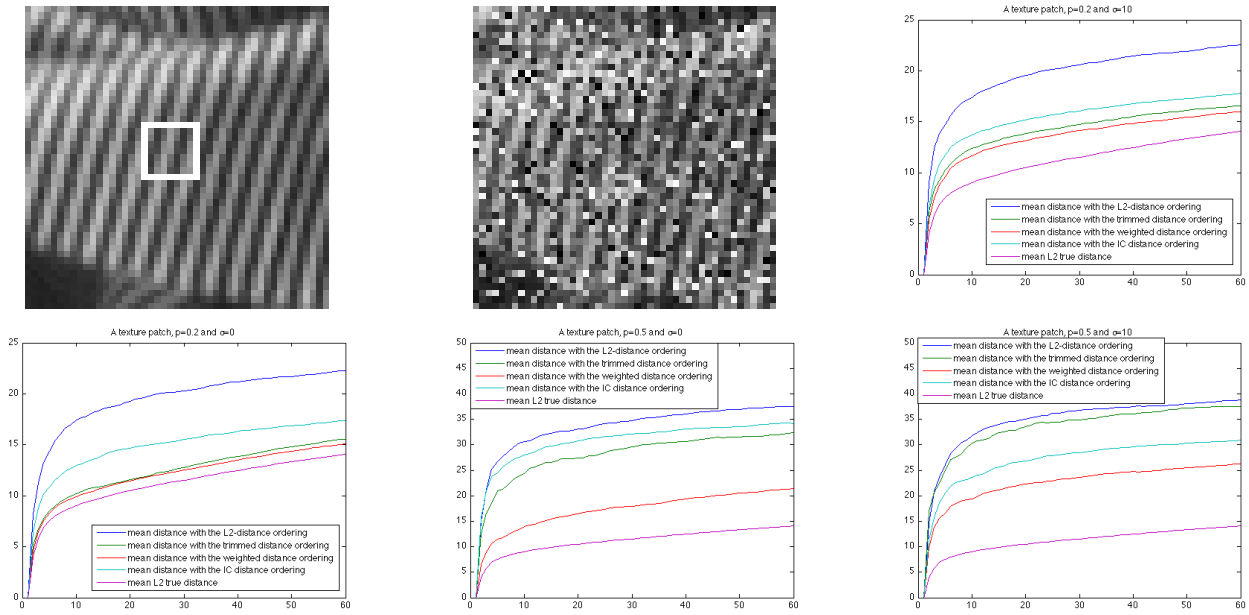


FIG. 3.2. *Comparison of different distances between a texture patch and similar patches. On the first raw, we show the original patch (white window) of size* 7×7 *pixels and its neighborhood* V*; a noisy version of the image (*$p = 0.2$ *and* $\sigma = 10$*); and the mean distances obtained in that case. On the second raw, we show the curves obtain for other noise parameters: respectively* $p = 0.2$ *and* $\sigma = 0$*,* $p = 0.5$ *and* $\sigma = 0$*,* $p = 0.5$ *and* $\sigma = 10$*. See the text for some comments on these graphics.*

D between patches, and for $n \geq 1$ integer, let $x_1, \ldots, x_n$ denote the centers of the patches that achieve the n smallest values of the distance D to the noisy patch $\mathbf{P}$:

$$0 = D(\mathbf{P}, \mathbf{P}_{x_1}) \leq D(\mathbf{P}, \mathbf{P}_{x_2}) \leq \ldots D(\mathbf{P}, \mathbf{P}_{x_n}).$$

Now, consider the mean of the L^2-distances of these patches in the original image: $\frac{1}{n}(\| \mathbf{P}^0 - \mathbf{P}^0_{x_1} \|_2 + \ldots + \| \mathbf{P}^0 - \mathbf{P}^0_{x_n} \|_2)$. We want this mean distance to be as close as possible to the mean of the n smallest L^2-distances in the original image (called the "mean L^2 true distance" in Figure 3.2), given by $\frac{1}{n}(\| \mathbf{P}^0 - \mathbf{P}^0_{x^0_1} \|_2 + \ldots + \| \mathbf{P}^0 - \mathbf{P}^0_{x^0_n} \|_2)$, where $x^0_1, \ldots, x^0_n$ are the centers of the patches that achieve the n smallest L^2 distances to the patch $\mathbf{P}^0$ in u^0. We make n vary from 1 to 60 and we plot the mean distances as a function of n.

We made this for four different distances between patches: the L^2 distance, the trimmed distance D_{trimmed}, the weighted distance D_{weighted} (defined in Equation (3.5)) and the "impulse-controlled" (IC) distance used by [26]. This last distance is also a weighted L^2 distance but where the weights are now pixel-dependent and related to the noise detector ROAD [20] (see Section 5.1 for a more detailed description of ROAD). The results are shown on Figures 3.1 and 3.2 for two different patches: an edge patch (its boundary is marked by the white window in Figure 3.1), and a texture patch (also marked by a white window, in Figure 3.2); and for different noise parameters: $p = 0.2$ or $p = 0.5$ and $\sigma = 0$ or $\sigma = 10$. The obtained curves are averaged from 50 noise samples for each set of parameters. The "mean L^2 true distance" is the purple lowest curve and the "best" distance between noisy patches should be the one that is as close as possible to this purple curve.

From these experiments we can draw the following conclusions: the L^2 distance between noisy patches always achieve the worst performances and the trimmed distance becomes unable to find

the right ordering of patches when the impulse noise is too important (the parameter T used to trimmed the distance is too small - for instance its value is only $T = 6$ when $p = 0.5$ for patches of size 7×7 pixels). On the edge patch, the IC distance and the weighted distance provide similar results, except for $p = 0.5$ and $\sigma = 10$ where the IC distance is more efficient. On the texture patch, the situation is different, the weighted distance always yields the best result and the IC distance sometimes fails completely (for instance in the case of pure impulse noise: $p = 0.5$ and $\sigma = 0$). The conclusion of this experiment goes in the direction of Theorem 3.1. In the rest of the paper, we choose to use the distance D_{weighted} between noisy patches for its ability to deal properly with different patch geometries.

4. Choice of the estimator $\widehat{\mathbf{P}}$. In this section, we aim at defining a good estimator $\widehat{\mathbf{P}}$ of the underlying patch $\mathbf{P}^0$ behind different damaged versions $\mathbf{P}_1$,... $\mathbf{P}_n$. For the sake of simplicity, we assume that these patches are independent realizations of the same random patch, following the noise model (2.1).

Let x be a pixel in the support of $\mathbf{P}^0$ and let us denote by $X_1, \ldots, X_n$ the random variables corresponding to the realizations $\mathbf{P}_1(x), \ldots, \mathbf{P}_n(x)$. The X_i are i.i.d. and follow the mixture distribution

$$X_i \sim (1-p).\mathcal{N}(\mu, \sigma^2) + p.\mathcal{U}_{[0,M]}, \tag{4.1}$$

where $\mathcal{U}_{[0,M]}$ denotes the uniform law on the discrete set $[0, \ldots, M]$ and $\mathcal{N}(\mu, \sigma^2)$ denotes the Gaussian distribution of mean μ and variance σ^2. We wish to estimate μ (and possibly σ). Let us start with the simpler problem of pure impulse noise : in this case, the Gaussian part of the mixture degenerates into a Dirac distribution at μ.

4.1. Pure impulse noise. In this pure impulse noise case, the mean and variance of the X_i's can be written

$$\mathbb{E}[X_i] = (1-p).\mu + p.\frac{M}{2} \quad \text{and} \tag{4.2}$$

$$\text{Var}[X_i] = \mu.p(1-p)(\mu - M) + p\frac{2M^2 + M}{6} - p^2\left(\frac{M}{2}\right)^2. \tag{4.3}$$

In the following, we study the relevance of different estimators of μ.

4.1.1. Mean. The mean $\overline{X_n} = \frac{1}{n}\sum_{i=1}^n X_i$ is known to be the maximum likelihood estimator (MLE) of μ for additive Gaussian noise. For pure impulse noise, the mean of $\overline{X_n}$ equals

$$\mathbb{E}[\overline{X_n}] = \mathbb{E}[X_i] = \mu + p\left(\frac{M}{2} - \mu\right), \tag{4.4}$$

We can thus derive a first unbiased estimator of μ, as $\widetilde{X}_n = \frac{1}{1-p}\left(\overline{X_n} - p\frac{M}{2}\right)$. The quadratic risk (*i.e* the mean squared error or MSE) of this unbiased estimator equals

$$\mathcal{R}[\widetilde{X}_n] = \mathbb{E}[|\widetilde{X}_n - \mu|^2] = \text{Var}[\widetilde{X}_n] = \left(\frac{1}{1-p}\right)^2 \frac{\text{Var}[X_i]}{n}. \tag{4.5}$$

As a consequence, this risk makes the estimator useless for practical purposes. Indeed, for example, when $M = 255$, $p = 0.5$ and $\mu = 100$, then $\sqrt{\text{Var}[\widetilde{X}_n]} \simeq \frac{108}{\sqrt{n}}$, which means that we need more than 468 samples to make the square root of the risk go below 5 gray levels.

4.1.2. Median. The spatial median filter is one of the most famous filters used to remove impulse noise. Let us study its properties as an estimator of μ. Assume, for the sake of simplicity, that n is odd: the median of $X_1, \dots X_n$ is $X_{\left(\frac{n+1}{2}\right)}$, where $X_{(k)}$ is the k^{th} order statistic (or k^{th} smallest value) of $X_1, \dots X_n$. The distribution function of the median can thus be easily computed as:

$$\mathbb{P}[X_{\left(\frac{n+1}{2}\right)} \leq k] = \mathcal{B}\left(n, \frac{n+1}{2}, F(k)\right) := \sum_{j \geq \frac{n+1}{2}}^{n} \binom{n}{j} F(k)^j (1 - F(k))^{n-j}, \tag{4.6}$$

where $F(k) := \mathbb{P}[X_i \leq k] = p\frac{k+1}{M+1} + (1-p)\mathbf{1}_{k \geq \mu}$ is the distribution function of the X_i's. After some computations, it follows that

$$\mathbb{E}[X_{\left(\frac{n+1}{2}\right)}] = M - \sum_{k=0}^{M-1} \mathcal{B}\left(n, \frac{n+1}{2}, F(k)\right). \tag{4.7}$$

Numerically, this bias is negligible as long as $p \leq 0.4$. Nonetheless, it increases dangerously when p becomes larger than 0.5. The variance of the estimator can be computed in the same way. At the end, the quadratic risk of the estimator equals:

$$\mathcal{R}[X_{\left(\frac{n+1}{2}\right)}] = (M-\mu)^2 - \sum_{k=0}^{M-1} (2(k-\mu)+1)\mathcal{B}\left(n, \frac{n+1}{2}, F(k)\right). \tag{4.8}$$

Figure 4.1 shows the values of $\sqrt{\mathcal{R}[X_{\left(\frac{n+1}{2}\right)}]}$ for different values of p and n, with $M = 255$ and $\mu = 20$ or $\mu = 100$. For $p = 0.5$ and $\mu = 100$, $n = 18$ samples are necessary to make the square root of the MSE decrease below 5, which is quite reasonable. However, for $\mu = 20$, this number of necessary samples becomes larger than 100, as can be observed on the left part of Figure 4.1. This huge difference in behavior is due to the bias of the median estimator which attains very large values as soon as $p \geq 0.5$ and for gray levels μ far from the middle of $[0, M]$. As a conclusion, the median estimator is clearly more interesting than the one built upon the mean, but in practice, its quadratic risk is not controlled, except for small values of p.

4.1.3. Maximum likelihood. Now, let us study the statistical properties of the maximum likelihood estimator (MLE) of μ. The MLE corresponds to the most represented value among the samples: if we compute the empirical histogram h of $\{X_1, \dots, X_n\}$ on $[0, M]$, the MLE is the place where the histogram attains its maximum. Indeed,

$$\begin{aligned}\widehat{X_n} &= \arg\max_{\mu'} \log \mathbb{P}[X_1, \dots, X_n | \mu'] = \arg\max_{\mu'} \sum_{i=1}^{n} \log \mathbb{P}[X_i | \mu'] \\ &= \arg\max_{\mu'} \sum_{i=1}^{n} \log\left((1-p)\delta_{X_i = \mu'} + \frac{p}{M+1}\right) = \arg\max(h).\end{aligned}$$

If the histogram has several maxima, one of them is chosen randomly, with equal probabilities. Consequently, if we denote by $\widehat{X_n}$ the random variable corresponding to this MLE, the law of $\widehat{X_n}$ is given by

$$\widehat{X_n} \sim (1 - q_{n,p}).\delta_\mu + q_{n,p}.\mathcal{U}_{[0,M]}, \tag{4.9}$$

where $q_{n,p} \in [0,1]$ is a probability that depends on p, n (and M). The probability for the estimator to estimate correctly μ is then $\mathbb{P}[\widehat{X_n} = \mu] = 1 - q_{n,p} + \frac{q_{n,p}}{M+1}$. The mean and the variance of $\widehat{X_n}$ are

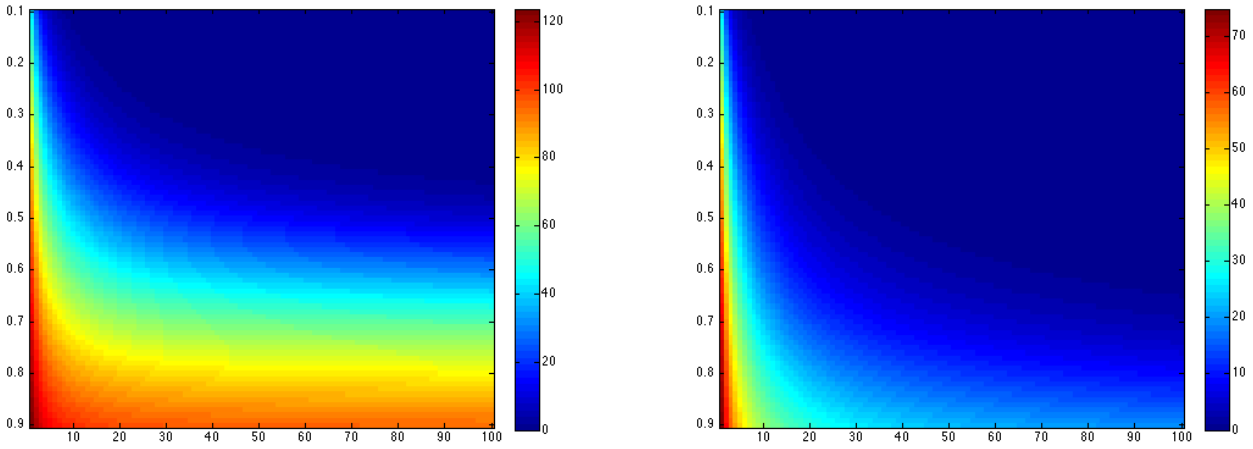


FIG. 4.1. *Square root of the risk (or MSE)* $\sqrt{\mathcal{R}[X_{(\frac{n+1}{2})}]}$ *as a function of* p *(between* 0.1 *and* 0.8 *on the vertical axis) and* n *(between* 1 *and* 100 *on the horizontal axis), with* $M = 255$, $\mu = 20$ *on the left and* $\mu = 100$ *on the right.*

easily obtained by replacing p by $q_{n,p}$ in Equations (4.3) and (4.2). It follows that the quadratic risk of $\widehat{X_n}$ as an estimator of μ can be written

$$\mathcal{R}[\widehat{X_n}] = \mathbb{E}[|\widehat{X_n} - \mu|^2] = (Bias[\widehat{X_n}])^2 + \mathrm{Var}(\widehat{X_n}) = q_{n,p}\left(\mu(\mu - M) + \frac{2M^2 + M}{6}\right). \tag{4.10}$$

and is proportional to $q_{n,p}$. In order to keep $\sqrt{\mathcal{R}[\widehat{X_n}]}$ smaller than 5 gray levels for $M = 255$ and any value of μ in $[0, M]$, the probability $q_{n,p}$ must be smaller than 10^{-3}. Now, it can be shown that the value of $q_{n,p}$ is quickly decreasing when n increases. Indeed, let $K := \#\{i,\ X_i = \mu\}$ be the random variable corresponding to the number of samples equal to μ. Among the $n - K$ remaining samples with values different from μ, let C_K be the number of K-uples with equal values. Observe that the event $\widehat{X_n} = \mu$ is realized when $C_K = 0$. Of course, $C_K = 0$ whenever $K \geq \frac{n+1}{2}$. As a consequence, if we assume for sake of simplicity that n is odd, we get

$$\begin{aligned} 1 - q_{n,p} + \frac{q_{n,p}}{M+1} = \mathbb{P}[\widehat{X_n} = \mu] &\geq \mathbb{P}[K \geq \frac{n+1}{2}] + \mathbb{P}[2 \leq K \leq \frac{n-1}{2} \text{ and } C_K = 0] \\ &= \sum_{k=\frac{n+1}{2}}^{n} \binom{n}{k} \widehat{p}^{n-k}(1-\widehat{p})^k + \sum_{k=2}^{\frac{n-1}{2}} \binom{n}{k} \widehat{p}^{n-k}(1-\widehat{p})^k(1-\alpha_k), \end{aligned}$$

where $\widehat{p} = \mathbb{P}[X_1 \neq \mu] = p - \frac{p}{(M+1)}$ and where $\alpha_k := \mathbb{P}[C_k \geq 1] \leq \mathbb{E}[C_k] = \binom{n-k}{k}\frac{1}{M^{k-1}}$. It follows that

$$\begin{aligned} 1 - q_{n,p} + \frac{q_{n,p}}{M+1} = \mathbb{P}[\widehat{X_n} = \mu] &\geq \mathcal{B}(n, \frac{n+1}{2}, 1-\widehat{p}) + \sum_{k=2}^{\frac{n-1}{2}} \binom{n}{k} \widehat{p}^{n-k}(1-\widehat{p})^k \left(1 - \binom{n-k}{k}\frac{1}{M^{k-1}}\right) \\ &= 1 - \widehat{p}^n - n(1-\widehat{p})\widehat{p}^{n-1} - \sum_{k=2}^{\frac{n-1}{2}} \binom{n}{k}\binom{n-k}{k}\frac{1}{M^{k-1}}(1-\widehat{p})^k\widehat{p}^{n-k}. \end{aligned}$$

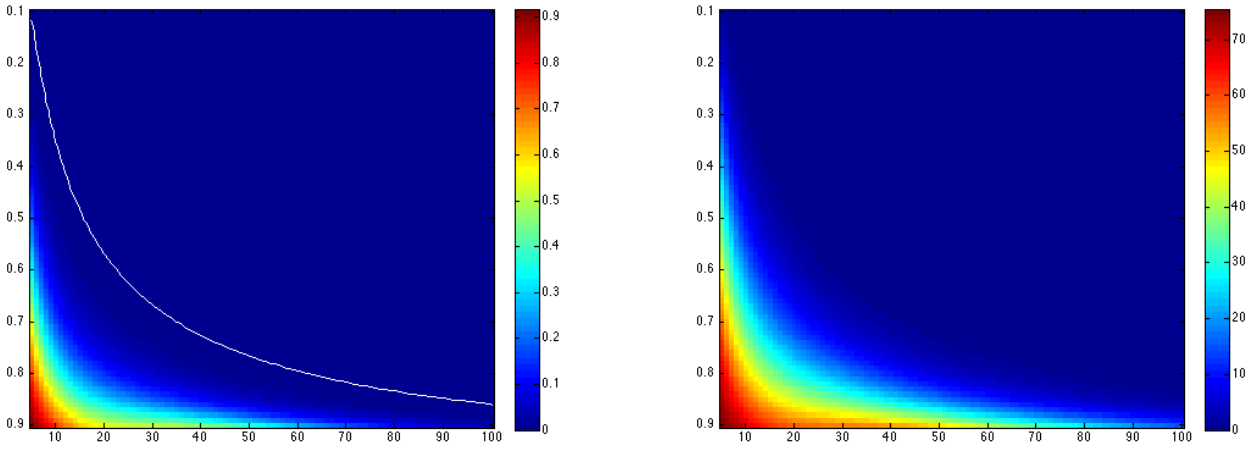


FIG. 4.2. *Left: parameter $q_{n,p}$ of the distribution* (4.9) *in function of p (vertical axis) and n (horizontal axis). The white curve corresponds to the level line $q_{n,p} = 10^{-3}$. This line yields the number n of samples necessary to keep the square root of the risk always below 5 gray levels (whatever the value of μ). Right: Square root of the risk $\sqrt{\mathbb{E}[|\widehat{X_n} - \mu|^2]}$ as a function of p (vertical axis) and n (horizontal axis), for $\mu = 100$.*

For reasonable values of p and n ($p \leq 0.8$ and $n \leq 50$), we checked empirically that the inequality in the above equation is in fact almost an equality, and that the last sum can be approximated by its first term ($k = 2$), which is broadly dominant. This allows us to estimate the parameter $q_{n,p}$ of the distribution (4.9) very easily in function of p. The left part of Figure 4.2 shows the values of $q_{n,p}$ as a function of p and n, when $M = 255$. The white line on the figure corresponds to the level line $q_{n,p} = 10^{-3}$. This line yields the number n of samples necessary to keep the square root of the risk always below 5 gray levels (whatever the value of μ), as a function of p. This number of samples remains quite reasonable in comparison to previous estimators. Moreover, the risk is controlled for all values of μ, unlike the median estimator. The risk of $\widehat{X_n}$ as a function of p and n is illustrated on the right part of Figure 4.2, for $\mu = 100$.

4.2. Mixture case. The previous study tends to prove that in the case of pure impulse noise, the Maximum Likelihood Estimator is preferable to the median estimator for denoising.

In the full mixture case, we would like to estimate both the mean μ and the variance σ^2 of the Gaussian part. Observe that in practice, this full mixture case is a realistic model for the estimation problem even in presence of pure impulse noise. Indeed, similar patches in an image are never exactly equal and it is a sound hypothesis to assume that their inner variability can be modeled with a standard deviation σ (that is pixel-dependent).

The expectation of the X_i's is unchanged (see Equation (4.2)) and their variance becomes

$$\mathrm{Var}[X_i] = \mu.p(1-p)(\mu - M) + p\frac{2M^2 + M}{6} - p^2\left(\frac{M}{2}\right)^2 + (1-p)\sigma^2. \tag{4.11}$$

The only difference with the pure impulse case consists in the additional term $+(1-p)\sigma^2$.

4.2.1. Mean and Median. The properties of the mean and median as estimators of μ are quite similar to the pure impulse case. The variance of the unbiased estimator $\widetilde{X}_n$ is still obtained from Equation (4.5). The law of the median is easily computed by replacing the distribution function

F in Equation (4.6) by $F(k) = p\frac{k+1}{M+1} + (1-p)G_{\mu,\sigma}(k)$, where $G_{\mu,\sigma}$ is the distribution function of the $\mathcal{N}(\mu, \sigma^2)$ distribution. Formulas (4.7) and (4.8) for the bias and risk of this estimator are generalized in the same way. Again, the risk of this estimator is uncontrolled as soon as p increases.

4.2.2. Maximum likelihood. In the mixture case, the Maximum Likelihood Estimator (MLE) of (μ, σ), that we will denote by $(\widehat{X_n}, \widehat{\sigma})$, is defined as

$$\begin{aligned}(\widehat{X_n}, \widehat{\sigma}) &= \arg\max_{\mu',\sigma'} \log \mathbb{P}[X_1, \dots, X_n | \mu', \sigma'] \\ &= \arg\max_{\mu',\sigma'} \sum_{i=1}^{n} \log\left(\frac{p}{M+1} + (1-p) g_{\mu',\sigma'}(X_i)\right),\end{aligned}$$

where $g_{\mu',\sigma'}$ is the Gaussian probability density of mean μ' and variance σ'^2. This can be rewritten by using the empirical distribution h of the values $\{X_1, \dots, X_n\}$ on $[0, M]$, which yields

$$(\widehat{X_n}, \widehat{\sigma}) = \arg\max_{\mu',\sigma'} h * f_{\sigma'}(\mu')$$

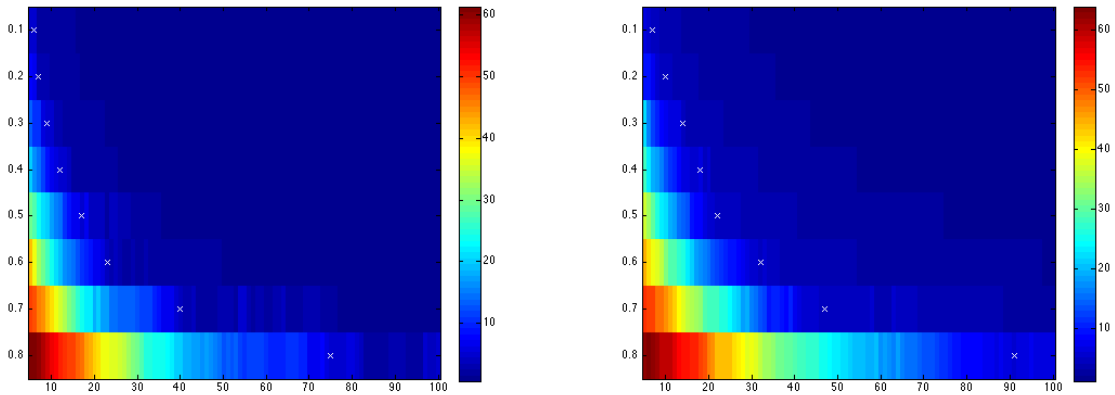
where $f_{\sigma'} : m \mapsto \log\left(\frac{p}{M+1} + \frac{1-p}{\sigma'\sqrt{2\pi}} e^{-\frac{m^2}{\sigma'^2}}\right)$. Observe that the MLE of pure impulse noise can be recovered from this formula by taking $\sigma = 0$. In the same way, the MLE in the case of pure Gaussian noise is obtained by taking $p = 0$ in the previous formula.

The study of the bias and variance of this maximum likelihood estimator is far more complex than in the pure impulse case. In particular, there is no obvious close formula for the distribution of $\widehat{X_n}$. In order to evaluate the quality of this estimator, we replace this study by an empirical estimation of the quadratic risk $\mathcal{R}[\widehat{X_n}] = \mathbb{E}[(\widehat{X_n} - \mu)^2]$ as a function of p and n for different values of σ and for $\mu = 100$ (as observed in the pure impulse case, this MSE does not depend too much on the value of μ). Figure 4.3 shows the empirical values of $\mathcal{R}[\widehat{X_n}]$ for $p \in \{0.1, \dots, 0.8\}$ and $n \in \{5, \dots, 100\}$. The four subfigures correspond respectively to $\sigma = 5, 10, 15$ and 20. For each σ and for each value of p, we draw a white cross at the minimum number n that should be chosen in order to ensure that the square root of the risk or MSE is smaller than 5. This number obviously increases with σ and p. It remains nevertheless reasonable for $p \leq 0.5$ and $\sigma \leq 15$.

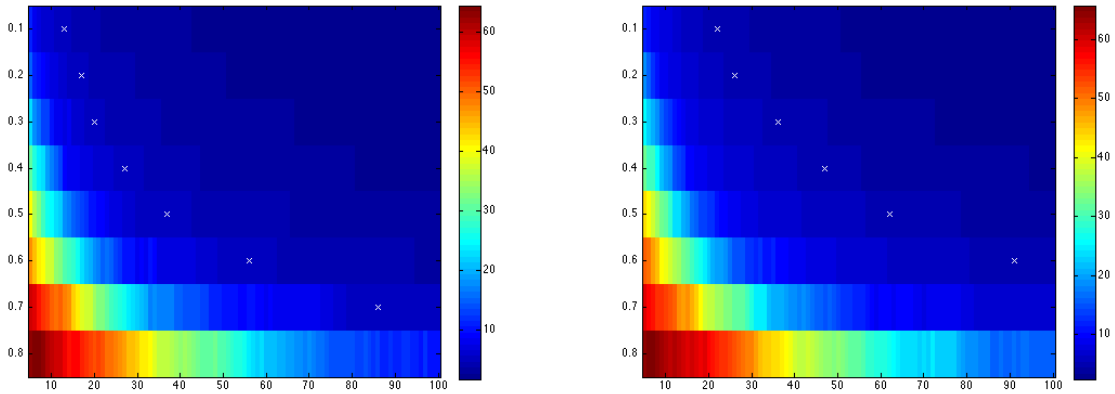
Choosing n in practice. The idea behind the previous empirical computation is to evaluate the appropriate number n of patches necessary to compute $\widehat{X_n}$ with a controlled risk. Assuming that we have an accurate estimation of p and σ, this number of trusted patches can be computed empirically and used to estimate μ in a reliable way.

Now, recall that $X_1, \dots, X_n$ correspond in practice to realizations $\mathbf{P}_1(x), \dots, \mathbf{P}_n(x)$ for a set of similar patches $\mathbf{P}_1, \dots, \mathbf{P}_n$. In natural images, even the most similar patches are seldom equal, and present some differences. This situation has two consequences. First, the value of σ in the model (4.1) should account both for the Gaussian part of the noise and for the inner variability between patches. Even in the case of pure impulse noise, choosing $\sigma > 0$ in order to estimate the right number of trusted patches n might improve the results. Second, the number of trusted patches n should always result from a compromise between the control of the theoretical risk and the redundancy of the image to be denoised. Indeed, the inner variability between patches considered as similar might increase with n, depending on the image regularity.

In practice, we observed that choosing $\sigma = 10$ to estimate n in the case of images suffering from pure impulse noise usually yields a good compromise between image redundancy, patch variability and risk control. Once p is estimated, the values of n used in all our experiments on pure impulse



(a) $\sigma = 5$ (b) $\sigma = 10$



(c) $\sigma = 15$ (d) $\sigma = 20$

FIG. 4.3. *Square root of the risk* $\sqrt{\mathcal{R}[\widehat{X}_n]}$ *for the MLE in the mixture case, as a function of* p *(vertical axis) and* n *(horizontal axis), for* $\mu = 100$. *White crosses correspond to the minimum values of* n *ensuring that the square root of the risk is smaller than* 5.

noise (see Section 6) correspond to the white crosses of Figure 4.3 (b). In the case of mixture noise, the situation is more complex, since the values of σ are unknown in practice. In order to reduce the number of parameters, we resolved to choose the values of n obtained for $\sigma = 15$ in all our experiments involving a mixed noise (white crosses of Figure 4.3 (c)). For the sake of completeness, these values are given in Tables 4.1 and 4.2.

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
n	8	10	14	18	22	34	47	91

TABLE 4.1

Number n *of patches used in the maximum-likelihood estimation for each value of* p *in the case of pure impulse noise.*

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
n	13	17	20	27	37	56	86	101

TABLE 4.2

Number n of patches used in the maximum-likelihood estimation for each value of p in the case of a mixture noise.

5. Denoising scheme. We are now in a position to fully describe PARIGI, a denoising scheme built upon the maximum likelihood estimator presented in Section 4.2.2 and the weighted distance between patches introduced in Section 3. This denoising scheme requires to estimate the noise ratio p beforehand. We describe in Section 5.1 different approaches devoted to estimate p globally on the damaged image. Section 5.2 then details the successive steps of our patch-based denoising scheme.

5.1. Estimation of p. Different methods are possible to estimate p globally on the damaged image. In this paper, we propose to rely on impulse noise detectors.

Many impulse noise detectors are proposed in the literature, such as the detectors ROAD [20], ACWMF [8], ROLD [17] or the pixel-wise MAD [9], to cite only a few. The goal of these schemes is to yield a map of noisy (impulse affected) pixels in an image damaged by impulse noise, or by a mixture of impulse and Gaussian noise. An estimation of p can naturally be derived from this estimation by computing the ratio between the number of noisy pixels and the image size. For the sake of simplicity, we restrict ourselves in this paper to the detectors ROAD and ACWMF, which are quite complementary.

The detector ROAD (for "Rank Ordered Absolute Differences"), proposed in [20], can be described as follows: for each pixel x, the absolute differences between $u(x)$ and $u(y)$ are computed for all $y \neq x$ in a centered 3×3 patch around x. These differences are ordered. The value $\mathrm{ROAD}(x)$ is obtained by computing the sum of the 4 smallest differences. This value measures how close $u(x)$ is from its neighbors. When $\mathrm{ROAD}(x)$ is above a given threshold τ, set as 70 in our experiments, x is considered as noisy.

The detector ACWMF [8] works as follows. For a given pixel x, a weighted median of order k is defined as

$$\mathrm{med}^k(x) = \mathrm{median}\left(\{u(y), y \in \mathbf{P}_x\} \cup \underbrace{\{u(x), \ldots, u(x)\}}_{2k \text{ times}}\right), \tag{5.1}$$

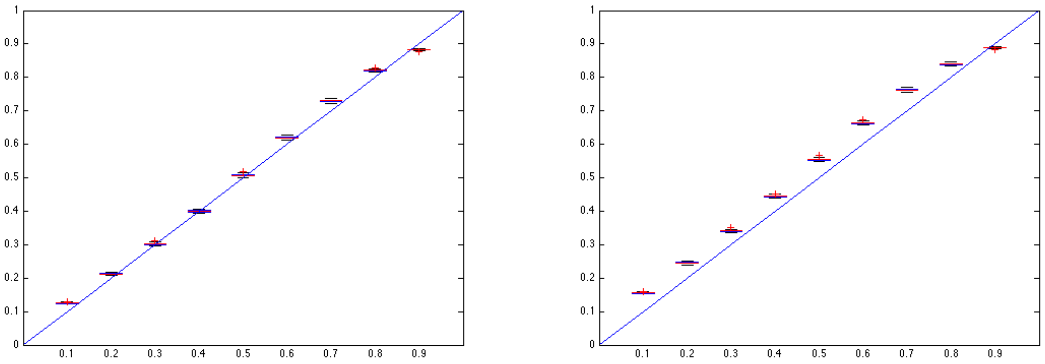
where $\mathbf{P}_x$ is a squared patch centered at x. This weighted median boils down to the usual median filter when $k = 0$ and to the identity when k is large enough. Now, let $d_k = |\,\mathrm{med}^k(x) - u(x)|$ and let t_k be a decreasing sequence of well chosen thresholds. The pixel x is presumed to be noisy if there exists one k such that d_k is above the threshold t_k. In practice, the authors of [8] recommend to use 3×3 patches and to compute the four thresholds t_k, $k = 1, \ldots, 4$ as

$$t_k = s.\,\mathrm{median}\left(\{|u(y) - \mathrm{med}^0(x)|, y \in \mathbf{P}_x\}\right) + \delta_k, \tag{5.2}$$

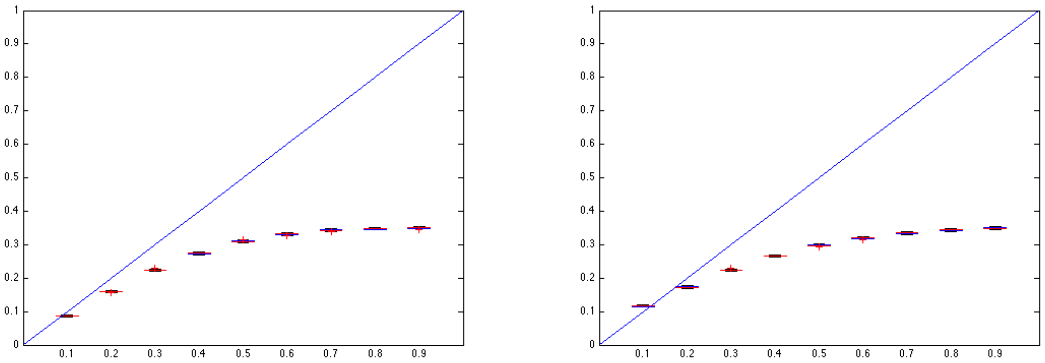
with $[\delta_0, \delta_1, \delta_2, \delta_3] = [40, 25, 10, 5]$ and $0 \leq s \leq 0.6$.

Figures 5.1 (a,b) show the quality of both estimators for different values of p on the 512×512 image *Lena* in the case of pure impulse noise (on the left) and in the case of a mixture of impulse noise and Gaussian noise with $\sigma = 10$ (on the right). For each value of p between 0.1 and 0.9 with a step of 0.1, we show the boxplot of the estimation of p for 100 different noise samples. The ROAD-based estimation of p, shown on Figure 5.1 (a), is quite accurate in both cases and fast to compute. As already underlined in [8], the detector ACWMF is efficient when the noise ratio is not too high, typically $p \leq 0.25$. Above this value, the number of pixels considered as touched by

noise is highly underestimated and the estimation of p is not usable in practice. As a consequence, we decided in all our experiments to use the estimation of p based on the detector ROAD.



(a) ROAD estimation of p when $\sigma = 0$ (on the left) and $\sigma = 10$ (on the right).



(b) ACWMF estimation of p when $\sigma = 0$ (on the left) and $\sigma = 10$ (on the right)

FIG. 5.1. *Left: statistical results of the estimation of the impulse noise parameter p on the Lena image when there is no Gaussian noise. The results $\overline{p}$ are shown as boxplot graphics, where the horizontal axis represents the tested values of p: from 0.1 to 0.9. Boxes are the statistics obtained from 100 samples for each value of p. On the right, same experiment made on the Lena image in the case of a mixed noise: impulse and Gaussian with $\sigma = 10$.*

5.2. Implementation details. The estimation of p is the first step of PARIGI. The second step consists in estimating n, the number of nearest neighbors taken into account in the space of patches to compute the estimator $\widehat{P}$. Assuming that the estimation of p is accurate, this number of trusted patches is computed empirically as explained in Section 4.2.2.

The algorithm continues as follows. For each point x in Ω, we seek the n nearest neighbors $\mathbf{P}_y$ of $\mathbf{P}_x$ for the distance D_{weighted} introduced in Section 3. We restrict this investigation to the points y spanning a $(2t+1) \times (2t+1)$ square $\mathcal{V}_x$ centered at x (see Figure 5.2)

$$\mathcal{V}_x = \{x + \delta; \quad \delta \in [-t, t] \times [-t, t]\}. \tag{5.3}$$

This permits to define $\mathcal{V}_x^n$, the subset of $\mathcal{V}_x$ corresponding to the n nearest neighbors of $\mathbf{P}_x$. The MLE $(\widehat{u(x)}, \widehat{\sigma(x)})$ is then computed at each x from the n-tuple $(u(y), y \in \mathcal{V}_x^n)$. The image $\widehat{u}$

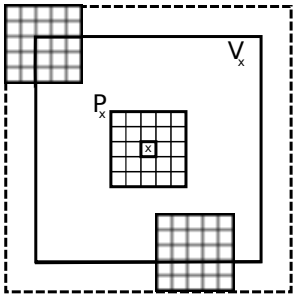


FIG. 5.2. *Neighborhood $\mathcal{V}_x$ and possible corresponding patches.*

constitutes a first denoised version of u. In practice, this restored image is sometimes a little bit too smooth. At the same time, some impulse pixels can remain. In order to recover the grain of the original image and to eliminate these last impulses, we take into account the estimated standard deviation $\widehat{\sigma}(x)$ at each point x for defining a map of noisy pixels:

$$\mathcal{M} = \{x \in \Omega;\ |\widehat{u}(x) - u(x)| > \widehat{\sigma}(x)\}.$$

We then compute the mixture of $\widehat{u}$ and u: $u_2(x) = \widehat{u}(x).\mathbf{1}_{x\in\mathcal{M}} + u(x).(1 - \mathbf{1}_{x\in\mathcal{M}})$, and eventually apply the previous denoising steps to u_2 in order to obtain $\widehat{u_2}$. These steps can be repeated a few times to improve the final result. The interest of these iterations is shown on the last line of Figure 6.5. In the rest of the experiments, we repeat this step twice and the output of the algorithm is given by u_3. In all our experiments, the half-size of the research neighborhood $\mathcal{V}_x$ is set to $t = 7$. The half-size of the patches is set to $f = 3$ for pure impulse noise and to $f = 8$ in the case of a mixture of Gaussian and impulse noise.

A refined version of the algorithm can be obtained by following the approach introduced in [30]. Notice that a point x belongs to all patches $\mathbf{P}_{x+\delta}$, $\delta \in [-f, f] \times [-f, f]$. The idea of the refinement is to take into account in the estimation all the information from these patches. In this version, the MLE is computed at each x from the set of values

$$\left\{u(y - \delta);\ \delta \in [-f, f] \times [-f, f] \text{ and } y \in \mathcal{V}^n_{x+\delta}\right\}.$$

This refined version is the one used in the experimental section. The whole refined algorithm is described in Algorithm 1.

6. Experiments and discussion. This section is devoted to the experimental analysis of the denoising scheme introduced in the previous section. We confront PARIGI with the recent state of the art approaches [20, 17] on pure impulse noise and with [20, 33] on the mixture of impulse and Gaussian noise. Comparison results are provided both under the form of PSNR tables and of visual experiments. The last part of the section is devoted to a short discussion on the link between image regularity models and denoising approaches.

6.1. Pure impulse noise. In this paragraph, we present some experiments on pure impulse noise. Table 6.1 describes the PSNR obtained with the three methods ([20, 17] and our approach)

Algorithm 1: Denoising algorithm PARIGI

input : Image u, half patch size f, half size of the research zone t, number of iterations N_{it} ($N_{it} = 2$ by default)

output: Denoised image $u_{N_{it}+1}$

Initialize $u_1 = u$;

for i from 1 to N_{it} **do**

  Estimate p on u_i; `// See Section 5.1`

  Compute $n = NumberNearestNeighbors(p)$; `// see Tables 4.1 and 4.2 in Section 4.2.2.`

  for each pixel x **do**

    $\mathcal{L}_x = \emptyset$;

  end

  for each pixel x **do**

    Find $\mathcal{V}_x^n$, the subset of $\mathcal{V}_x$ corresponding to the n nearest neighbors of $\mathbf{P}_x$ for D_{weighted} in u_i;

    for each $\delta \in [-f, f] \times [-f, f]$ **do**

      $\mathcal{L}_{x+\delta} = \mathcal{L}_{x+\delta} \cup \{u_i(y + \delta), y \in \mathcal{V}_x^n\}$;

    end

  end

  for each pixel x **do**

    Obtain $(\widehat{u(x)}, \widehat{\sigma(x)})$ as the MLE from the set $\mathcal{L}_x$; `// see Section 4.2.2.`

  end

  Compute

$$u_{i+1} = \widehat{u}.\mathbf{1}_{\mathcal{M}} + u.(1 - \mathbf{1}_{\mathcal{M}})$$

  where

$$\mathcal{M} = \{x \in \Omega;\ |\widehat{u(x)} - u(x)| > \widehat{\sigma(x)}\}.$$

end

Return $u_{N_{it}+1}$

on the 512×512 classical images *Lena*, *Bridge*, *Baboon* and *Barbara* [3]. Let us recall that the PSNR is a common way to measure the quality of a restored image v in comparison to the undamaged one u^0. It is given by the formula

$$\text{PSNR}(u^0, v) = 10 \log_{10} \frac{255^2 |\Omega|}{\sum_{x \in \Omega} (u^0(x) - v(x))^2},$$

where $|\Omega|$ is the size of the support of u^0. For *Lena*, *Bridge* and *Baboon*, the PSNR results of [20, 17] are taken from Table 1 in [17]. Our result on these images are thus obtained with the same noise ratios, but different noise samples. The results on *Barbara* are all obtained on the same noise samples, using the codes kindly provided by the authors, both for the trilateral filter [4] and for ROLD-EPR. Table 6.1 shows very similar performances between PARIGI and ROLD-EPR, at least for the first three images. This proves that a patch-based approach is well founded for

[3] The first three images are those available on the website `www.math.cuhk.edu.hk/~rchan/paper/dcx`. The Barbara image is available for instance at `http://www.irisa.fr/vista/Themes/Demos/Debruitage/images/`.

[4] The authors of [20] provide their code on their webpage `www.ssc.wisc.edu/~thuegeri`.

random-valued impulse noise removal. Much larger differences in PSNR can be observed on the image *Barbara*. As we will see in the following, these remarkable differences in performance can be mostly explained by the way the three denoising approaches are handling geometrical textures in images.

In order to evaluate the precision of the PSNR results, for each value of p and each method, we compute the PSNR for 10 different noise realizations. In practice, the standard deviation of the PSNR for all methods is always very close to 0.1: for instance, for the scheme PARIGI on the image Lena and $p = 0.4$, the empirical average PSNR is 34.19 and the corresponding standard deviation is 0.096. This means that if we observe, between two methods, PSNR differences smaller than $\sqrt{2} \times 0.1 \simeq 0.14$, these differences can be considered as meaningless.

Figure 6.1 provides a visual comparison of the different approaches on the four images when $p = 20\%$. In this simple case, all methods are efficient and results are visually quite similar, although the result of our scheme might be slightly smoother. Figures 6.2 and 6.3 correspond to a more complex denoising situation, with a noise ratio of 40% on *Lena* and 50% on *Barbara*. On *Lena*, the result of PARIGI is smoother than the ones of ROLD-EPR [17] and ROAD-Trilateral [20], in which some clues of impulse noise remain. On *Barbara*, while all schemes yield reasonable results on constant regions, our approach is the only one to handle properly the regular stripes of the clothes, while the trilateral filter and ROLD-EPR replace them by mottled textures.

	ROAD-Trilateral [20]	ROLD-EPR [17]	PARIGI
Lena 512×512, $p = 20\%$	36.70	37.45	**38.33**
Lena 512×512, $p = 40\%$	31.12	32.76	**34.18**
Lena 512×512, $p = 60\%$	26.08	29.03	**29.96**
Bridge 512×512, $p = 20\%$	27.60	**27.86**	27.68
Bridge 512×512, $p = 40\%$	24.01	24.79	24.80
Bridge 512×512, $p = 60\%$	20.84	**22.59**	22.03
Baboon 512×512, $p = 20\%$	24.18	**24.49**	24.17
Baboon 512×512, $p = 40\%$	21.60	21.92	22.02
Baboon 512×512, $p = 60\%$	19.52	**20.38**	20.13
Barbara 512×512, $p = 20\%$	24.17	26.16	**33.91**
Barbara 512×512, $p = 40\%$	23.11	23.72	**29.92**
Barbara 512×512, $p = 60\%$	21.71	22.65	**24.93**

TABLE 6.1

PSNR results of different restoration filters for the 512×512 *images Lena, Bridge, Baboon and Barbara for pure impulse noise.*

6.2. Gaussian and impulse noise mixture. In this paragraph, we investigate the performance of PARIGI for the denoising of images suffering from a mixture of Gaussian and impulse noise, by confronting it with the recent approaches [20, 33]. Table 6.2 shows the PSNR results of the three methods for different combinations of Gaussian and impulse noise, on *Lena* and *Barbara*. The results for the $l_1 - l_0$ minimization approach of Xiao *et al.* are taken from Tables 8 and 9 in [33]. The results of the trilateral filter are computed on the same noise samples as our approach. We observed that the parameter σ_R of the trilateral filter, which controls its radiometric influence, has a great impact on the performances of the code. For this reason, we choose for each mixture experiment the parameter σ_R which yields the best PSNR. Other parameters are chosen as proposed by default in the code provided by the authors. For both images and for each mixture of

Gaussian and impulse noise, our scheme yields the highest PSNR. Observe that the ROLD-EPR approach has not been developed for such mixed noises, but still provides very reasonable results in this case. Again, the performance improvements observed on *Barbara* with the scheme PARIGI are higher than those observed on *Lena*, especially when p and σ become larger. These results are also illustrated on Figure 6.4.

	ROAD-Trilateral [20]	Xiao [33]	PARIGI
Lena 512×512, $p = 10\%$ $\sigma = 5$,	33.23	34.98	**36.03**
Lena 512×512, $p = 30\%$ $\sigma = 5$	29.90	32.04	**33.88**
Lena 512×512, $p = 10\%$ $\sigma = 10$,	31.45	32.75	**33.40**
Lena 512×512, $p = 30\%$ $\sigma = 10$	29.28	30.42	**31.97**
Lena 512×512, $p = 10\%$ $\sigma = 15$	29.96	30.85	**31.37**
Lena 512×512, $p = 30\%$ $\sigma = 15$	28.37	29.11	**30.18**
Barbara 512×512, $p = 10\%$ $\sigma = 5$,	27.66	30.48	**31.55**
Barbara 512×512, $p = 30\%$ $\sigma = 5$	24.44	25.92	**29.28**
Barbara 512×512, $p = 10\%$ $\sigma = 10$,	25.74	28.42	**30.12**
Barbara 512×512, $p = 30\%$ $\sigma = 10$	23.94	25.34	**28.19**
Barbara 512×512, $p = 10\%$ $\sigma = 15$	24.55	27.31	**28.8**
Barbara 512×512, $p = 30\%$ $\sigma = 15$	23.39	24.55	**27.33**

TABLE 6.2

PSNR results of different restoration filters for the 512×512 *images Lena and Barbara, and different mixtures of Gaussian and impulse noise.*

6.3. Discussion. It is remarkable that the patch-based method presented in this paper provides such good performances on *Barbara*, when compared to state of the art approaches. Such a difference cannot be observed on *Bridge*, *Baboon* or *Lena*. The visual experiments of Figures 6.3 and 6.4 suggest that this difference is mostly due to the way the different methods are handling the regular stripes on *Barbara*'s clothes. This is confirmed by the experiments shown on Figure 6.5. In this example, we compare several denoising procedures on a synthetic 256×256 image composed of perfect vertical stripes and suffering from pure impulse noise, with $p = 50\%$. Thanks to the huge patch redundancy of this particular image, PARIGI is able to recover the vertical stripes almost perfectly after a few iterations, providing a result very close the original image. The variational approach of [17] and the trilateral filter [20] do not fully take advantage of this redundancy, and this explains their quite poor results on this kind of image. On the contrary, we can observe on Figure 6.1 that the edge preserving regularization of [17] seems to preserve the fur of the mandrill image slightly better than our patch-based approach. An interesting direction of research, which is beyond the scope of this paper, would consist in identifying the image classes (or, at least, the texture classes) optimally restored by each kind of denoising approach.

7. Conclusion. In this paper, it is shown that a patch-based approach can be an efficient tool to remove mixtures of Gaussian and impulse noises. This result was known (and has been widely studied) in the particular case of Gaussian noise, but its extension to impulse degradations necessitates a careful choice of both the similarity measure between patches and the statistical estimator of the original patches. It is shown on several experiments that this patch-based approach permits to attain state of the art denoising performances on classical images. An important performance gain is demonstrated on geometrically regular textures. This work opens several perspectives. First, as explained in the previous section, we intend to explore the connections between image or texture

regularity models and the denoising performances of different kinds approaches (local filters, variational methods, patch-based approaches). In particular, it would be of great interest to determine which image class is optimal for each denoising method. Another possible extension of this work includes the collaboration of different denoising approaches, depending on the image local regularity and redundancy. Finally, we intend to study the theoretical bounds of impulse noise removal, as pioneered by [7] and [25] in the case of Gaussian noise.

REFERENCES

[1] S. Awate and R. Whitaker. Unsupervised, information-theoretic, adaptive image filtering for image restoration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(3):364–376, 2006.

[2] T. Brox and D. Cremers. Iterated nonlocal means for texture restoration. In *Proceedings of the 1st international conference on Scale space and variational methods in computer vision*, pages 13–24. Springer-Verlag, 2007.

[3] T. Brox, O. Kleinschmidt, and D. Cremers. Efficient nonlocal means for denoising of textural patterns. *Image Processing, IEEE Transactions on*, 17(7):1083–1092, 2008.

[4] A. Buades, B. Coll, and J. Morel. A review of image denoising algorithms, with a new one. *Multiscale Modeling and Simulation*, 4(2):490–530, 2006.

[5] R. Chan, C. Hu, and M. Nikolova. An iterative procedure for removing random-valued impulse noise. *Signal Processing Letters, IEEE*, 11(12):921–924, 2004.

[6] P. Chatterjee and P. Milanfar. Clustering-based denoising with locally learned dictionaries. *Image Processing, IEEE Transactions on*, 18(7):1438–1451, 2009.

[7] P. Chatterjee and P. Milanfar. Is denoising dead? *IEEE Transactions on Image Processing*, 19(4):895–911, 2010.

[8] T. Chen and H. Wu. Adaptive impulse detection using center-weighted median filters. *Signal Processing Letters, IEEE*, 8(1):1–3, 2001.

[9] V. Crnojevic, V. Senk, and Z. Trpovski. Advanced impulse detection based on pixel-wise MAD. *Signal Processing Letters, IEEE*, 11(7):589–592, 2004.

[10] H. David and H. Nagaraja. *Order statistics*. Wiley series in probability and mathematical statistics. Probability and mathematical statistics. John Wiley, 2003.

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

[12] C.-A. Deledalle, L. Denis, and F. Tupin. How to compare noisy patches? patch similarity beyond gaussian noise. *International Journal of Computer Vision*, 99:86–102, 2012.

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

[14] C.-A. Deledalle, J. Salmon, and A. S. Dalalyan. Image denoising with patch based pca: local versus global. In *BMVC*, 2011.

[15] J. Delon and A. Desolneux. Stabilization of flicker-like effects in image sequences through local contrast correction. *SIAM Journal on Imaging Sciences*, 3(4):703–734, 2010.

[16] J. Delon and A. Desolneux. A patch-based approach for random-valued impulse noise removal. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2012.

[17] Y. Dong, R. Chan, and S. Xu. A detection statistic for random-valued impulse noise. *IEEE Transactions on Image Processing*, 16(4):1112–1120, 2007.

[18] V. Duval, J. Aujol, and Y. Gousseau. On the parameter choice for the Non-Local Means. *SIAM Journal on Imaging Sciences*, 4(2):760–788, 2011.

[19] A. Efros and T. Leung. Texture synthesis by non-parametric sampling. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 2, pages 1033–1038. Ieee, 1999.

[20] R. Garnett, T. Huegerich, C. Chui, and W. He. A universal noise removal algorithm with an impulse detector. *IEEE Transactions on Image Processing*, 14(11):1747–1754, 2005.

[21] O. Ghita and P. F. Whelan. A new GVF-based image enhancement formulation for use in the presence of mixed noise. *Pattern Recognition*, 43(8):2646–2658, 2010.

[22] B. Goossens, Q. Luong, A. Pizurica, and W. Philips. An improved non-local denoising algorithm. In *Local and Non-Local Approximation in Image Processing, International Workshop, Proceedings*, page 143, 2008.

[23] C. Kervrann and J. Boulanger. Local adaptivity to variable smoothness for exemplar-based image regularization and representation. *International Journal of Computer Vision*, 79(1):45–69, 2008.

[24] S. J. Ko and Y. H. Lee. Center weighted median filters and their applications to image enhancement. *IEEE Transactions on Circuits and Systems*, 38(9):984–993, 1991.

[25] A. Levin and B. Nadler. Natural image denoising: Optimality and inherent bounds. In *CVPR*, pages 2833–2840, 2011.

[26] B. Li, Q. Liu, J. Xu, and X. Luo. A new method for removing mixed noises. *SCIENCE CHINA Information Sciences*, 54(1):51–59, 2011.

[27] E. López-Rubio. Restoration of images corrupted by Gaussian and uniform impulsive noise. *Pattern Recognition*, 43(5):1835–1846, 2010.

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

[29] W. K. Pratt. Median filtering. Technical report, Image Proc. Inst., Univ. Southern California, 1975.

[30] J. Salmon and Y. Strozecki. Patch reprojections for Non Local methods. *Signal Processing*, 92(2):447–489, 2012.

[31] T. Sun and Y. Neuvo. Detail-preserving median based filters in image processing. *Pattern Recognition Letters*, 15(4):341–347, 1994.

[32] D. van de Ville and M. Kocher. SURE-Based Non-Local Means. *IEEE Signal Processing Letters*, 16:973–976, Nov. 2009.

[33] Y. Xiao, T. Zeng, J. Yu, and M. K. Ng. Restoration of images corrupted by mixed Gaussian-impulse noise via l_1-l_0 minimization. *Pattern Recognition*, 44(8):1708–1720, 2011.

[34] J. X. Yang and H. R. Wu. Mixed Gaussian and uniform impulse noise analysis using robust estimation for digital images. In *Proceedings of the 16th international conference on Digital Signal Processing*, DSP'09, pages 468–472, Piscataway, NJ, USA, 2009. IEEE Press.

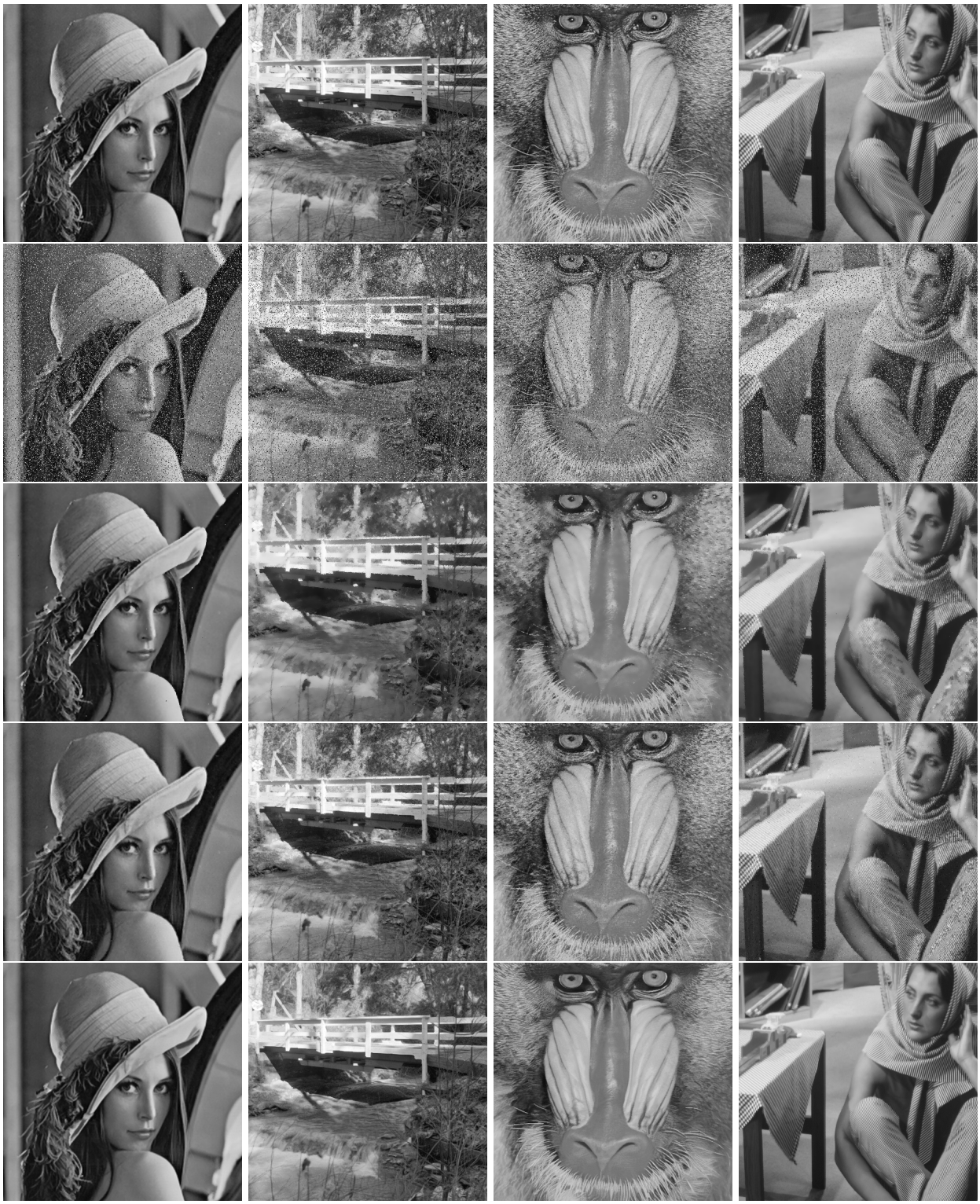


FIG. 6.1. *Comparative results on Lena,Bridge,Baboon and Barbara with $p = 20\%$ of uniform impulse noise. For each column, from top to bottom: original image, noisy image, ROAD+trilateral [20], ROLD+EPR [17], PARIGI. Images should be seen at full resolution on the electronic version of the paper.*

(a) Noisy, $p = 40\%$ (b) Trilateral filter [20]

(c) ROLD+EPR [17] (d) PARIGI

FIG. 6.2. *Comparative results on Lena in the case of pure impulse noise,* $p = 40\%$.

(a) Noisy, $p = 50\%$ (b) Trilateral filter [20]

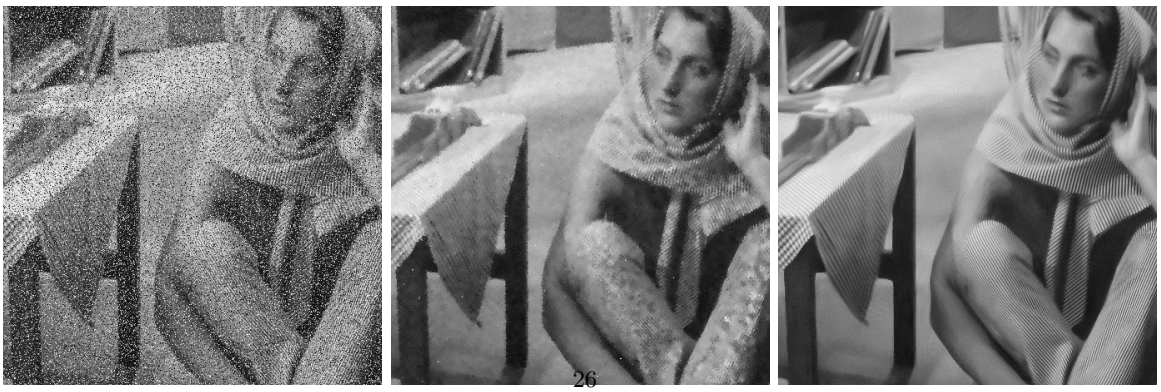
(c) ROLD+EPR [17] (d) PARIGI

FIG. 6.3. *Comparative results on Barbara in the case of pure impulse noise,* $p = 50\%$.

(a) $p = 0.1$, $\sigma = 5$

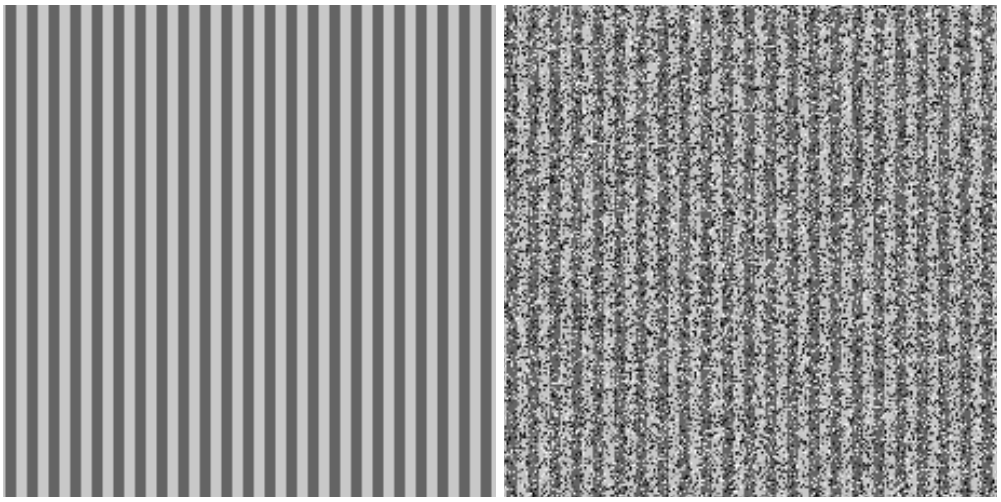
(b) $p = 0.3$, $\sigma = 15$

(c) $p = 0.1$, $\sigma = 5$

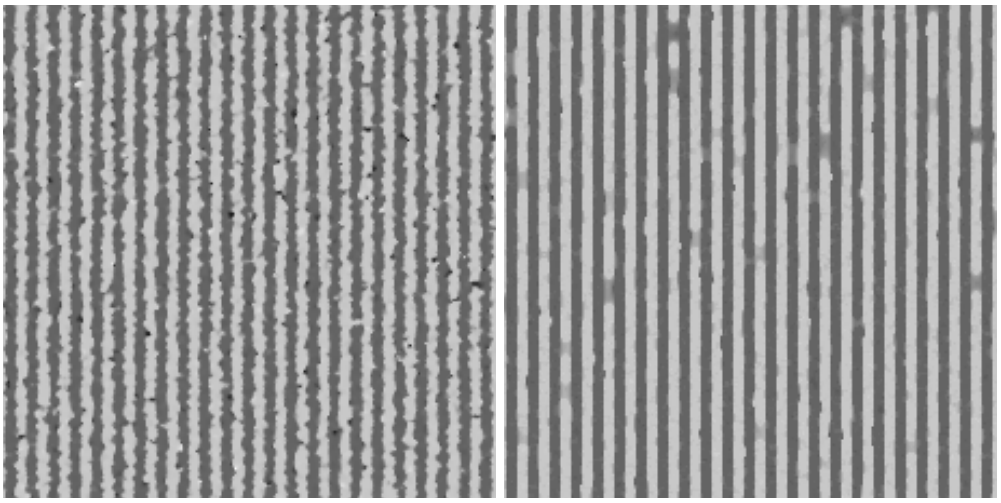


(d) $p = 0.3$, $\sigma = 15$

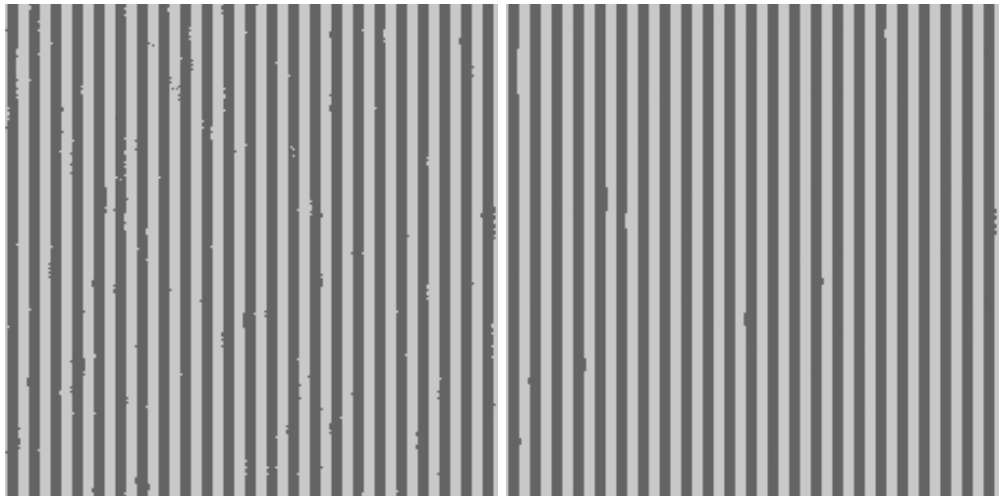
FIG. 6.4. *Comparative results on Lena and Barbara with* $p \in \{10\%, 30\%\}$ *of uniform impulse noise and* $\sigma \in \{5, 15\}$. *For each subfigure, from left to right: noisy image, result of the trilateral filter [20], PARIGI.*



(a) Original image (b) Noisy image, impulse noise, $p = 50\%$



(c) Trilateral filter [20], PSNR = 19.77 (d) ROLD-EPR [17], PSNR = 26.63



(e) PARIGI. On the left, $N_{it} = 1$, PSNR = 30.43. On the right, $N_{it} = 4$, PSNR = 36.74

FIG. 6.5. *Comparison of different denoising approaches on the image Stripes, with* $p = 50\%$.