



**HAL**  
open science

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

Julie Delon, Agnès Desolneux

► **To cite this version:**

Julie Delon, Agnès Desolneux. A patch-based approach for removing impulse or mixed Gaussian-impulse noise. *SIAM Journal on Imaging Sciences*, 2013, 6 (2), pp. 1140-1174. 10.5201/ipol.2016.161 . hal-00718612v2

**HAL Id: hal-00718612**

**<https://hal.science/hal-00718612v2>**

Submitted on 20 Dec 2012

**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 IMPULSE OR MIXED GAUSSIAN-IMPULSE NOISE \*

JULIE DELON <sup>†</sup> AND AGNÈS DESOLNEUX <sup>‡</sup>

**Abstract.** In this paper, we address the problem of the restoration of images which have been affected by impulse noise or 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.** Impulse noise in images is generally due to data loss happening during the acquisition stage (for instance in the presence of faulty pixels), storage (due to defective memory locations) or image transmission. This loss usually occurs after, or possibly at the same time as the first stages of image acquisition. During these stages, other sources of noise affect images: mainly, the photon shot noise, which is Poisson distributed, and the readout noise, which is Gaussian distributed [1]. When no data is lost, variance stabilization techniques [24] can efficiently transform Poisson-Gaussian noise mixtures into an additive white Gaussian noise, which explains the huge literature solely dedicated to Gaussian denoising. In practice, a user may have to deal with an image that is damaged by a mixture of Poissonian, Gaussian and impulse noise. Removing this kind of mixture while preserving image details and textures is of great importance before most image analysis tasks (edge detection, segmentation, etc). In this paper, we study a simplified version of this noise model, and concentrate on the mixture of Gaussian and 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  $[0, M]$  is 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  $\{0, 1, \dots, M\}$ . 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 pure impulse noise, 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 [36, 30]. 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 [38], the adaptive center weighted median filter (ACWMF) [11] or the pixel-wise median absolute deviation [12]. Unfortunately, median-based methods tend to destroy details and textures

---

\*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.

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

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

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 [25], 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 authors of [8, 21], 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 the pixels affected by impulse noise) a Gaussian noise removal method. This is for instance the case of [41], 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 [7, 6, 28, 33] where a median type filter (ACWMF in general) is used to detect the impulse noise and then a variational framework (TV norm, or  $l_1$  norm of framelet coefficients for instance) is used to restore the whole image. Let us mention that some of these papers (namely [7, 6, 33]) tackle at the same time the problem of blur in images. The framework is thus a bit different, being at the same time more difficult (there is an additional deterioration of the image because of the blur), and an easier one (impulse noise is more efficiently detected on a blurry and therefore smoother image). Detecting the impulse noise is also the first step of [26] where an impulse noise estimator is used to modify an anisotropic diffusion model, or of [42] where impulse noise is first removed by an adaptive cascade of median filters before removing Gaussian noise using BM3D [13]. The method of [34] takes advantage of the full probabilistic mixture model: an external method is used to predict original pixel values, followed by an EM algorithm which permits to learn the full noise model, and finally some kernel regression to get the restored image. The authors of [32] use the ROAD detector of [25] 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) [5]. 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 [23, 2, 5, 19]. Since the introduction of the NL-means by Buades et al. [5] 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 [27, 15]. 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 [29, 39, 37, 22]. 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<sup>1</sup> in the paper. Experiments and comparisons with recent approaches are displayed in Section 6. For the sake of completeness, let us mention that a conference proceeding version of this work has appeared in [20].

## 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  $\Omega$ , 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

$$U = (1 - T).(u^0 + W) + T.V, \quad (2.1)$$

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  $V(x)$ ,  $x \in \Omega$  are i.i.d. random variables with a discrete uniform distribution on the range  $[0, M]$ ;
- the values  $W(x)$ ,  $x \in \Omega$  are independent centered Gaussian random variables with variance  $\sigma^2$ :  $W(x) \sim \mathcal{N}(0, \sigma^2)$ .

In other words, the value  $u^0(x)$  of a pixel  $x$  is first modified by adding a centered Gaussian noise and then, with probability  $p$ , replaced by a uniform value on  $[0, M]$ . Pure impulse noise can be considered as a particular case of the previous model, obtained when  $\sigma = 0$ .

**2.2. The fundamentals of patch-based denoising.** The idea of patch-based approaches is to take advantage of the redundancy of patches (also sometimes called the self-similarity) in images. The central hypothesis of these approaches is that for each patch  $\mathbf{P}_1$  in  $u$ , there exist other patches  $\mathbf{P}_2, \dots, \mathbf{P}_n$  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  $\hat{\mathbf{P}}$  of  $\mathbf{P}^0$ , computed from its different damaged versions  $\mathbf{P}_1, \dots, \mathbf{P}_n$ .

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 [5], 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, and these weights then balance the influence of the different patches in the final estimator  $\hat{\mathbf{P}}$ . Alternatively, the authors of [35] 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  $\hat{\mathbf{P}}$  keeps only patches whose distance to  $\mathbf{P}_1$  is below this threshold. Other valid solutions only retain the  $n$ -nearest neighbors of  $\mathbf{P}_1$  [3], or make use of clustering [4, 9] or of PCA [18] 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, 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  $U_1, \dots, U_n$   $n$  i.i.d. random vectors following a distribution  $\mathcal{P}(\theta)$  with  $\theta$  a vector parameter, find a

---

<sup>1</sup>PARIGI stands for *Patch based Approach for the Restoration of Images affected by Gaussian and Impulse noise*.

good estimator  $\hat{\theta}$  of  $\theta$ . When the noise is purely Gaussian,  $\mathcal{P}(\theta) = \mathcal{N}(\theta, \sigma^2)$ , with  $\sigma$  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 [5]. Recently, many authors have adopted the point of view of quadratic risk minimization as a way to optimize the parameters of the NL-means [29, 39, 22] or to propose further improvements [37]. 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 Strojceki [37], 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 [17]), and we will denote by  $\Omega_f = [-f, f] \times [-f, f]$  the domain of definition of the patches. A patch  $\mathbf{P}_x$  in an image  $u$  is then an image defined on  $\Omega_f$  by:  $\forall \delta \in \Omega_f, \mathbf{P}_x(\delta) = u(x + \delta)$ . Moreover we define the *support* of  $\mathbf{P}_x$  as the set  $x + \Omega_f$ . As a consequence, a pixel  $x$  belongs to the support of all the patches  $\mathbf{P}_{x+\delta}$ ,  $\delta \in \Omega_f$ . In order to restore the value  $u^0(x)$ , we can thus rely on all the values  $\widehat{\mathbf{P}_{x+\delta}}(-\delta)$ ,  $\delta \in \Omega_f$ . 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 [5]. Another natural choice, studied for instance in [37] is to take the mean

$$\frac{\sum_{\delta \in \Omega_f} \widehat{\mathbf{P}_{x+\delta}}(-\delta)}{(2f + 1)^2}. \quad (2.2)$$

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  $\widehat{\mathbf{P}_{x+\delta}}(-\delta)$ ,  $\delta \in \Omega_f$ , 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  $\delta \in \Omega_f$  such that  $\mathbf{P}_{x+\delta}$  and  $\mathbf{P}_{y+\delta}$  are similar.

**3. Robust distance between patches.** This section focuses on the search of 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 ( $\ell^2$ ) 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 [16], is to make use of the generalized maximum likelihood ratio

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

In the case of pure random-valued impulse noise, this ratio becomes

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

where  $n_{sim}$  is the number of pixels in  $\Omega_f$  that have exactly the same grey level value in  $\mathbf{P}$  and  $\mathbf{Q}$ , and  $|\mathbf{P}| = (2f + 1)^2$  is the size of the definition domain of  $\mathbf{P}$ . Obviously, this measure is too rigid to be used in practice. 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  $f_\theta$  the Gaussian distribution with mean  $\theta$  and variance  $\sigma^2$ ):

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

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  $x \in \Omega_f$ :  $num(x)$  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 (because its value is a sup that is achieved for  $\mathbf{T}(x) = (\mathbf{P}(x) + \mathbf{Q}(x))/2$ ) and  $num(x)$  approximately equals  $(\frac{p}{M+1} + \frac{1-p}{\sigma\sqrt{2\pi}})(\frac{p}{M+1} + \frac{1-p}{\sigma\sqrt{2\pi}}e^{-(\mathbf{P}(x)-\mathbf{Q}(x))^2/2\sigma^2})$  when  $(\mathbf{P}(x) - \mathbf{Q}(x))^2/\sigma^2$  is large (it is achieved for  $\mathbf{T}(x) \simeq \mathbf{P}(x)$  or  $\mathbf{T}(x) \simeq \mathbf{Q}(x)$ ). 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  $\sigma$  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 [14]. It is known that, for non-noisy patches, the  $\ell^2$  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  $\ell^2$  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  $|P - Q|_{(1)} \leq |P - Q|_{(2)} \leq \dots \leq |P - Q|_{(2f+1)^2}$  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

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

where  $w_1, \dots, w_n$  are positive weights. There is a wide choice for the weights  $w_k$ . For instance, if they are all equal, we find again the  $\ell^2$  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  $D_{\text{trimmed}}$  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<sup>2</sup>. 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  $k^{\text{th}}$  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).

<sup>2</sup> $\mathcal{B}(n, k, q) = \sum_{i=k}^n \binom{n}{i} q^i (1 - q)^{n-i}$ .

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  $D_{\text{weighted}}$ :

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

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

The choice of the distance  $D_{\text{weighted}}$  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  $\delta_{(b-a)^2}$  denotes the Dirac distribution at  $(b-a)^2$  and  $dF$  is a probability distribution on  $\mathbb{R}_+$  (that we don't need to compute). In this framework, we have the following theorem that shows that choosing  $w_k = \mathcal{B}((2f+1)^2, k, (1-p)^2)$  for the weighted distance in (3.4) is, in some sense, optimal to discriminate the two constant patches.

**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  $\delta_\alpha$  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  $p_\alpha := \mathbb{P}(Y \leq \alpha) \rightarrow 0$  when  $\alpha \rightarrow 0$  where  $Y$  is a random variable distributed following the law  $dF$ . Then when  $\alpha$  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  $dF$ .

Now, a direct consequence of this theorem is that 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 (this corresponds to  $\alpha = 0$ ) from two constant but non-equal patches ( $\alpha > 0$ ), is the one such that  $\sum_{k=1}^n w_k \mathcal{B}(n, k, q)$  is as large as possible. 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).

*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  $N_\alpha$  be the random variables that counts the number of  $X_i^{(\alpha)}, 1 \leq i \leq n$ , that are equal to  $\alpha$ . Then  $N_\alpha$  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

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

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

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

Finally, since  $0 \leq m_{k,l}^-(\alpha) \leq \alpha$ , and  $p_\alpha$  goes to 0 when  $\alpha$  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

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

which is the announced result.  $\square$

**3.3. Comparison of the different distances.** When an image is not affected by noise, the  $\ell^2$  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  $\ell^2$  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  $\ell^2$  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  $D$  between patches, and for  $n \geq 1$  integer, let  $x_1, \dots, x_n$  denote the centers of the patches that achieve the  $n$  smallest values of the distance  $D$  to the noisy patch  $\mathbf{P}$ :

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



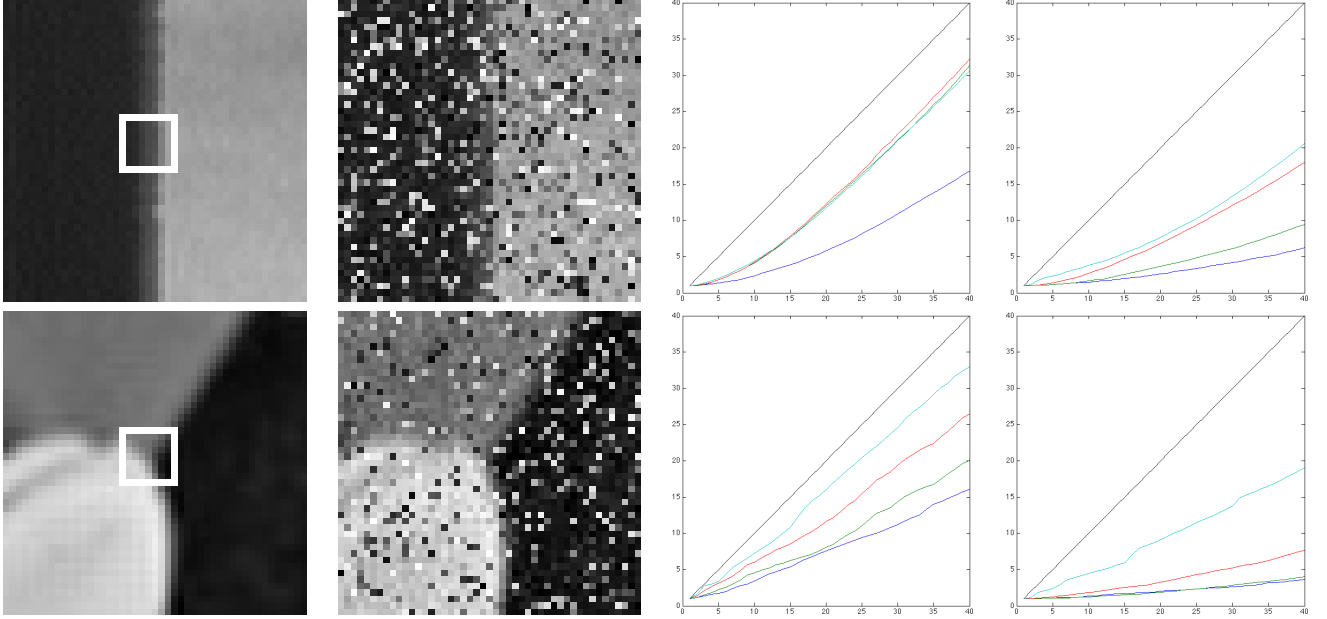


FIG. 3.1. Comparison of different distances on noisy patches. From top to bottom: an edge patch and a T-junction patch. From left to right: the original patch (marked by a white window of size  $7 \times 7$  pixels and its neighborhood  $V$ ), a noisy version of the image ( $p = 0.2$  and  $\sigma = 10$ ), and the mean number of selected patches for the different distances when  $\sigma = 10$ ,  $p = 0.2$  (third column) and  $p = 0.5$  (fourth column). The  $\ell^2$  distance corresponds to blue curves, the trimmed distance  $D_{\text{trimmed}}$  to green curves, the weighted distance  $D_{\text{weighted}}$  to red curves and the “impulse-controlled” (IC) distance used by [32] to indigo curves. See the text for some comments on these graphics.

Now, we count how many of these patches would have been also chosen among the  $n$ -nearest patches for the  $\ell^2$  distance in the original image. More precisely, if we denote by  $x_1^0, \dots, x_n^0$  the centers of the patches that achieve the  $n$  smallest  $\ell^2$  distances to the patch  $\mathbf{P}^0$  in  $u^0$ , we then compute

$$N_D(n) = \#\{1 \leq i \leq n \text{ such that } x_i \in \{x_1^0, \dots, x_n^0\}\}.$$

We make  $n$  vary from 1 to 40 and we plot the number  $N_D(n)$  of “well-selected” patches as a function of  $n$ . We ran this experiment for four different distances between patches: the  $\ell^2$  distance (blue curves), the trimmed distance  $D_{\text{trimmed}}$  (green curves), the weighted distance  $D_{\text{weighted}}$  defined in Equation (3.5) (red curves) and the “impulse-controlled” (IC) distance used by [32] (indigo curves). This last distance is also a weighted  $\ell^2$  distance but where the weights are now pixel-dependent and related to the noise detector ROAD [25] (see Section 5.1 for a more detailed description of ROAD). The results are shown on Figures 3.1 and 3.2, for different patches: an edge patch (its boundary is marked by the white window in the top left image), a T-junction patch, and three different texture patches; and for different noise parameters:  $\sigma = 10$  and  $p = 0.2$ ; or  $\sigma = 10$  and  $p = 0.5$ . The obtained curves are averaged from 200 noise samples for each set of parameters. The black diagonal (that recalls the value of  $n$ ) is the value that  $N_D(n)$  should try to reach in order to be a “good” distance between the noisy patches. As a consequence, the higher the curve  $n \mapsto N_D(n)$  is, the better the distance is.

From these experiments we can make the following remarks:

- The  $\ell^2$  distance (blue curves) between the noisy patches always achieve the worst performances.
- The trimmed distance (green curves) becomes unable to correctly select the patches when the impulse noise is too important (the parameter  $T$  used to trimmed the distance is too small - for

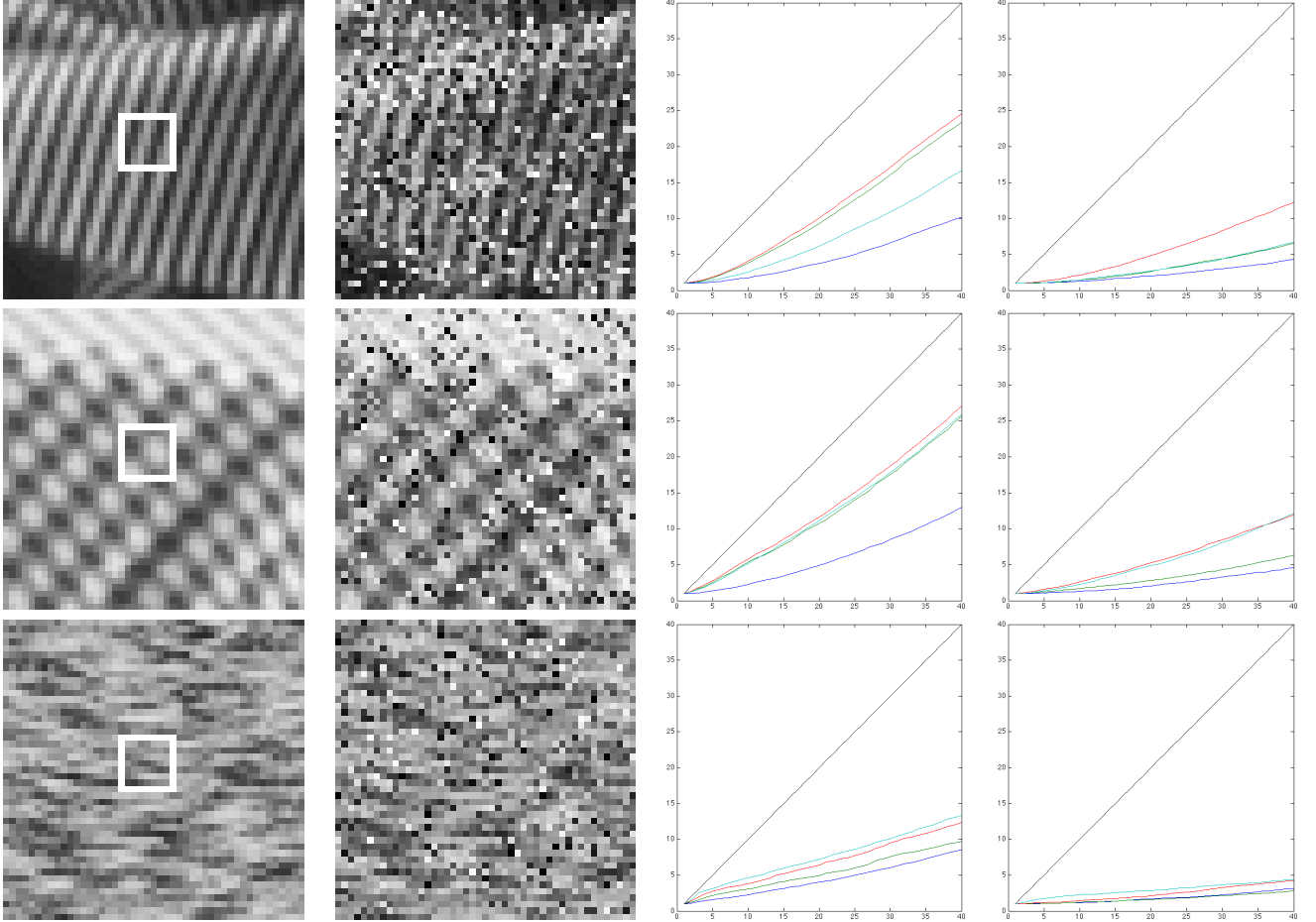


FIG. 3.2. Comparison of different distances on noisy texture patches. From top to bottom: three different texture patches (the scarf of Barbara, the tablecloth in Barbara, the fur of the baboon). From left to right: the original patch (marked by a white window of size  $7 \times 7$  pixels and its neighborhood  $V$ ), a noisy version of the image ( $p = 0.2$  and  $\sigma = 10$ ), and the mean number of selected patches for the different distances when  $\sigma = 10$ ,  $p = 0.2$  (third column) and  $p = 0.5$  (fourth column). The  $\ell^2$  distance corresponds to blue curves, the trimmed distance  $D_{\text{trimmed}}$  to green curves, the weighted distance  $D_{\text{weighted}}$  to red curves and the “impulse-controlled” (IC) distance used by [32] to indigo curves. See the text for some comments on these graphics.

instance its value is only  $T = 6$  when  $p = 0.5$  for patches of size  $7 \times 7$  pixels).

- The IC distance (indigo curves) and the weighted distance (red curves) have similar performances on the edge patch and on the second and third texture patches.
- On the T-junction patch, the IC distance performs better, while on the first texture patch (a part of the scarf of Barbara), the weighted distance seems to be a more relevant choice.
- On the third texture, all distances have bad performances (and even very bad ones when  $p = 0.5$ ). This can be explained by the fact that this texture looks like “noise”, and in that case, the nearest patches selected by the distance are quite randomly chosen.

Our conclusion of these experiments is the following: since the only case where the IC distance performs clearly better than the weighted distance is on the T-junction patch and since such a patch is rather usually a “rare” patch in images, we decide in all the following to use the weighted distance between noisy patches. Indeed, the IC distance of [32] is rather empirically defined, and relies on several parameters whose influence is not easy to interpret in practice. Conversely, the

weighted distance has a simple definition and only relies on the value of  $p$ . We will see in Section 5.1 how the noise ratio  $p$  can be efficiently estimated.

**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  $\mathbf{P}_1, \dots, \mathbf{P}_n$ . 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  $\delta \in \Omega_f$  be a fixed point in the domain of definition of the patches (for instance  $\delta = 0$  corresponds to the center of the patches) and let us denote by  $X_1, \dots, X_n$  the random variables corresponding to the realizations  $\mathbf{P}_1(\delta), \dots, \mathbf{P}_n(\delta)$ . The  $X_i$  are then real-valued random variables that are i.i.d. and follow the mixture distribution

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

where  $\mathcal{U}_{[0, M]}$  denotes the uniform law on the discrete set  $[0, \dots, M]$  and  $\mathcal{N}(\mu, \sigma^2)$  denotes the Gaussian distribution of mean  $\mu$  and variance  $\sigma^2$ . We wish to estimate  $\mu$  (and possibly  $\sigma$ ). 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  $\mu$ .

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

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

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

In the following, we study the relevance of different estimators of  $\mu$ .

**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  $\mu$  for additive Gaussian noise. For pure impulse noise, the mean of  $\overline{X}_n$  equals

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

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

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

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  $\sqrt{\text{Var}[\tilde{X}_n]} \simeq \frac{108}{\sqrt{n}}$ , 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  $\mu$ . Assume, for the sake of simplicity, that  $n$  is odd: the median of  $X_1, \dots, X_n$  is  $X_{(\frac{n+1}{2})}$ , where  $X_{(k)}$  is the  $k^{\text{th}}$  order statistic (or  $k^{\text{th}}$  smallest value) of  $X_1, \dots, X_n$ . The distribution function of the median can thus be easily computed as:

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

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

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

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:

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

Figure 4.1 shows the values of  $\sqrt{\mathcal{R}[X_{(\frac{n+1}{2})}]}$  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  $\mu$  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  $p$ .

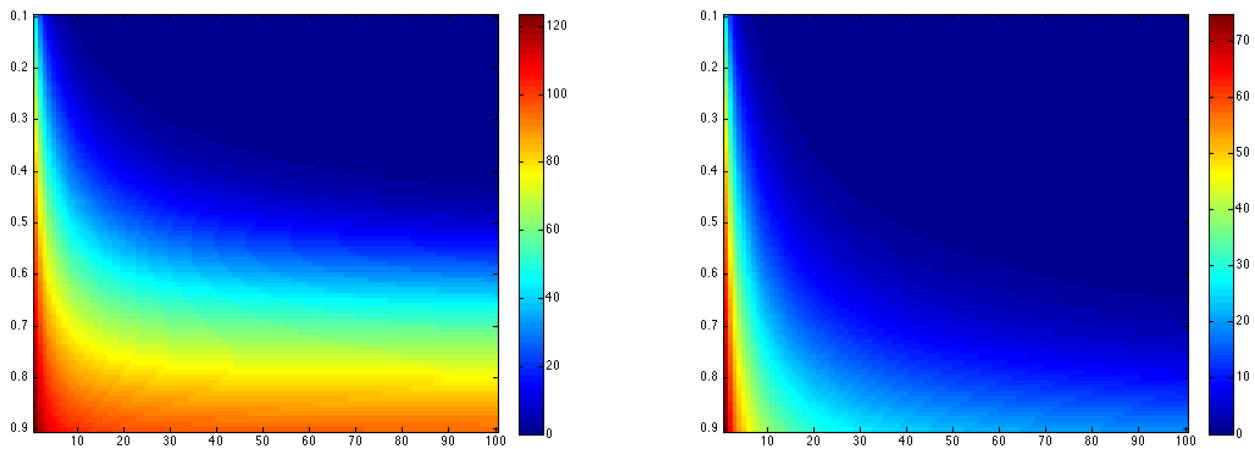


FIG. 4.1. Square root of the risk (or MSE)  $\sqrt{\mathcal{R}[X_{(\frac{n+1}{2})}]}$  as a function of  $p$  (between 0.1 and 0.8 on the vertical axis) and  $n$  (between 1 and 100 on the horizontal axis), with  $M = 255$ ,  $\mu = 20$  on the left and  $\mu = 100$  on the right.

**4.1.3. Maximum likelihood.** Now, let us study the statistical properties of the maximum likelihood estimator (MLE) of  $\mu$ . 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

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

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  $\mu$  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 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  $\mu$  can be written

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

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  $\mu$  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  $\mu$ . Among the  $n - K$  remaining samples with values different from  $\mu$ , 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

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

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

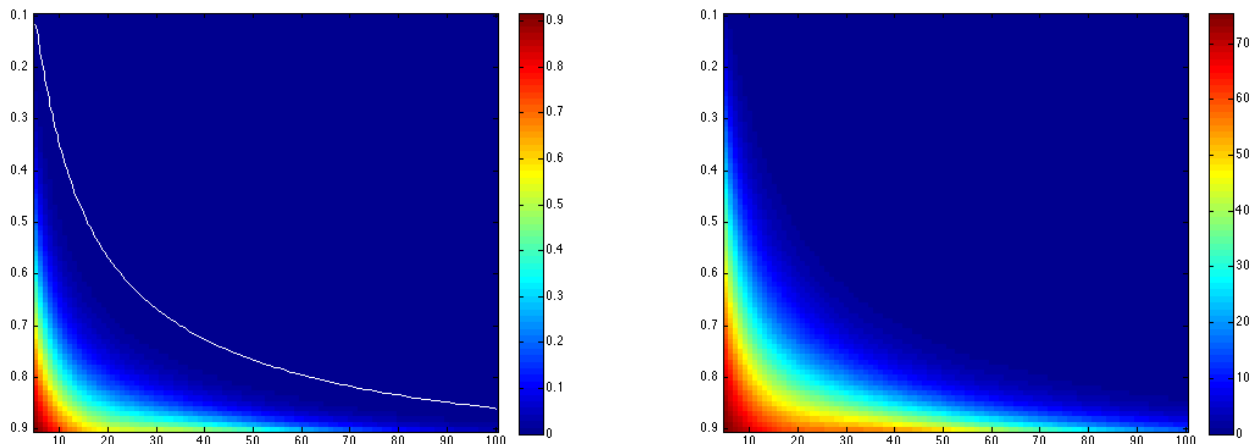


FIG. 4.2. *Left: parameter  $q_{n,p}$  of the distribution (4.9) in function of  $p$  (vertical axis) and  $n$  (horizontal axis). The white curve corresponds to the level line  $q_{n,p} = 10^{-3}$ . This line yields the number  $n$  of samples necessary to keep the square root of the risk always below 5 gray levels (whatever the value of  $\mu$ ). 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  $\mu = 100$ .*

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  $\mu$ ), 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  $\mu$ , 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  $\mu$  and the variance  $\sigma^2$  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  $\sigma$  (that is pixel-dependent).

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

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

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  $\mu$  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  $(\mu, \sigma)$ , 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  $\mu'$  and variance  $\sigma'^2$ . 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  $\sigma$  and for  $\mu = 100$  (as observed in the pure impulse case, this MSE does not depend too much on the value of  $\mu$ ). 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  $\sigma$  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  $\sigma$  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  $\sigma$ , this number of trusted patches can be computed empirically and used to estimate  $\mu$  in a reliable way.

Now, recall that  $X_1, \dots, X_n$  correspond in practice to realizations  $\mathbf{P}_1(\delta), \dots, \mathbf{P}_n(\delta)$  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  $\sigma$  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 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  $\sigma$  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,

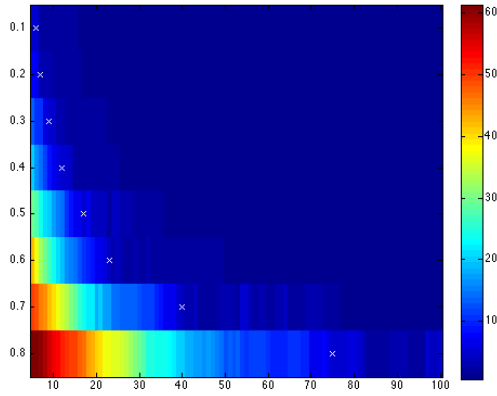
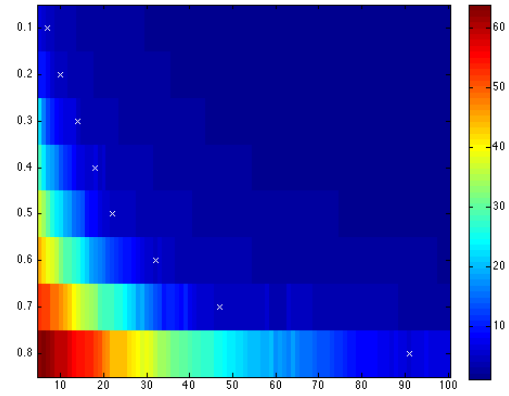
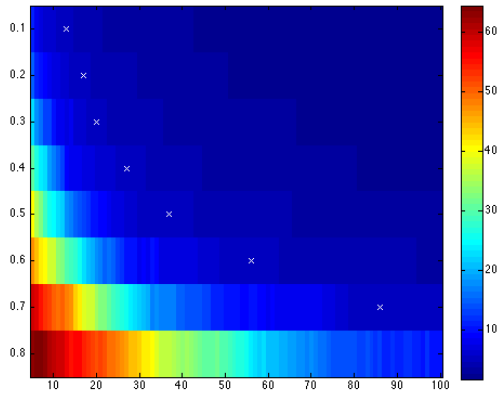
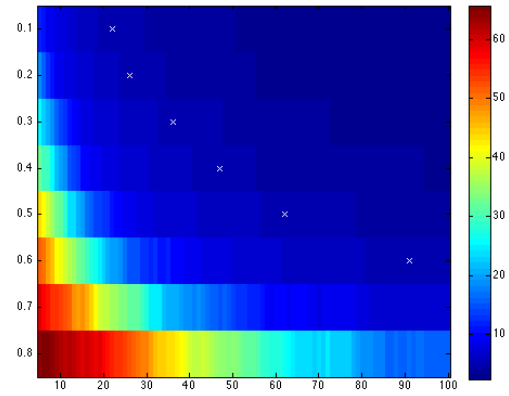
(a)  $\sigma = 5$ (b)  $\sigma = 10$ (c)  $\sigma = 15$ (d)  $\sigma = 20$ 

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.

these values are given in Tables 4.1 and 4.2. Let us notice at this point that these tabulated values of  $n$  are used once for all in our experiments without actually fixing  $\sigma$  in the algorithm: the standard deviation for each set of patches is estimated by the maximum likelihood step.

$p$	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.

**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



$p$	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.

$p$  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 [25], ACWMF [11], ROLD [21] or the pixel-wise MAD [12], 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 [25], 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  $3 \times 3$  patch around  $x$ . These differences are ordered. The value  $\text{ROAD}(x)$  is obtained by computing the sum of the 4 smallest differences. This value measures how close  $u(x)$  is from its neighbors. When  $\text{ROAD}(x)$  is above a given threshold  $\tau$ , set as 70 in our experiments,  $x$  is considered as noisy.

The detector ACWMF [11] works as follows. For a given pixel  $x$ , a weighted median of order  $k$  is defined as

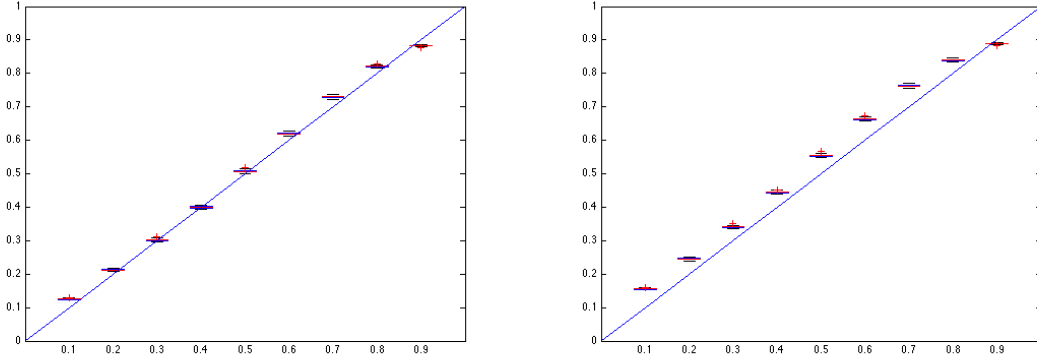
$$\text{med}^k(x) = \text{median} \left( \left\{ u(y), y \in V_x \right\} \cup \underbrace{\left\{ u(x), \dots, u(x) \right\}}_{2k \text{ times}} \right), \quad (5.1)$$

where  $V_x$  is a square window 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  $d_k = |\text{med}^k(x) - u(x)|$  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 [11] recommend to use  $3 \times 3$  patches and to compute the four thresholds  $t_k$ ,  $k = 1, \dots, 4$  as

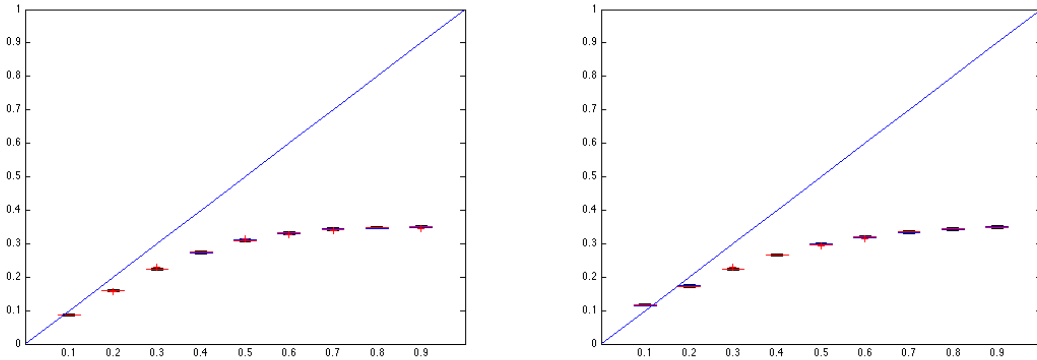
$$t_k = s. \text{median} \left( \{|u(y) - \text{med}^0(x)|, y \in V_x\} \right) + \delta_k, \quad (5.2)$$

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  $512 \times 512$  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 [11], 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  $\bar{p}$  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  $\hat{P}$ . 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  $\Omega$ , we seek the  $n$  nearest neighbors  $\mathbf{P}_y$  of  $\mathbf{P}_x$  for the distance  $D_{\text{weighted}}$  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)

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

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  $\widehat{u}$  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

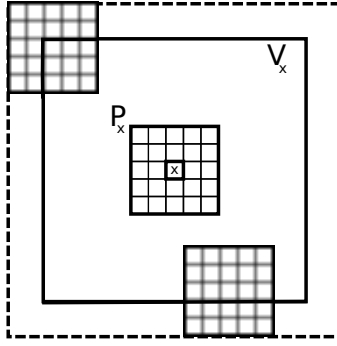


FIG. 5.2. Neighborhood  $\mathcal{V}_x$  and possible corresponding patches.

deviation  $\hat{\sigma}(x)$  at each point  $x$  for defining a map of noisy pixels:

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

We then compute the mixture of  $\hat{u}$  and  $u$ :  $u_2(x) = \hat{u}(x) \cdot \mathbf{1}_{x \in \mathcal{M}} + u(x) \cdot (1 - \mathbf{1}_{x \in \mathcal{M}})$ , and eventually apply the previous denoising steps to  $u_2$  in order to obtain  $\hat{u}_2$ . 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 6.7. 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 [37]. Notice that a point  $x$  belongs to all patches  $\mathbf{P}_{x+\delta}$ ,  $\delta \in \Omega_f$ . 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

$$\{u(y - \delta); \delta \in \Omega_f \text{ and } y \in \mathcal{V}_{x+\delta}^n\}.$$

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 [25, 21] on pure impulse noise and with [25, 41] on the mixture of impulse and Gaussian noise. Comparison results are provided both under the form of PSNR tables and of visual experiments. In the case of pure impulse noise, we also show MAE (Maximum Absolute Error) and SSIM (Structural SIMilarity index) tables. 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. Tables 6.1, 6.2 and 6.3 give the PSNR, MAE and SSIM obtained with the three methods ([25, 21] and our approach) on the  $512 \times 512$  classical images *Lena*, *Bridge*, *Baboon*, *Barbara*, *Cameraman*, *Boat*, *Peppers* and *Goldhill*. Let us mention at this point that most of these test images are available in more than one version on the internet, the main differences concerning their

---

**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  $n = \text{NumberNearestNeighbors}(p)$ ; // 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

        |  $D_{\text{weighted}}$  in  $u_i$ ;

**for** each  $\delta \in \Omega_f$  **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

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

    where

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

**end**

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

---

size, contrast, compression level and the way they are cropped<sup>3</sup>. In consequence, it is sometimes difficult to compare performance results provided by different papers, which may have used different image versions: on two different  $512 \times 512$  versions of *Lena*, we observed PSNR differences of more than 3dB for the same algorithm. In this paper, we always run the different algorithms on the same images and noise samples (when the code is available), instead of using the performance results provided in the literature. We also provide the different images used in the paper on our website <http://perso.telecom-paristech.fr/~delon/Demos/Impulse/>.

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  $u^0$ . An alternative to this quality measure is given by the

---

<sup>3</sup>See for instance <http://www.ece.rice.edu/~wakin/images/> for a discussion on the *standard* gray level version of *Lena*. This *standard* version is the one used in all our experiments.

Structural Similarity, an index (SSIM) measuring the similarity between images [40], designed to be more consistent with human perception than the usual measures PSNR or MSE. The code used here to compute SSIM is the one kindly provided by the authors of [40] on <http://www.cns.nyu.edu/~lcv/ssim/>, with the default parameters. For the sake of completeness, we also provide the Mean Absolute Error for all results:

$$\text{MAE}(u^0, v) = \frac{\sum_{x \in \Omega} |u^0(x) - v(x)|}{|\Omega|},$$

The results of Tables 6.1, 6.2 and 6.3 are all obtained on the same noise samples, using the codes kindly provided by the authors of the papers, both for the trilateral filter <sup>4</sup> and for ROLD-EPR. The results for ROAD-trilateral [25] are obtained with the parameters set as recommended by the authors, *i.e.*  $\sigma_s = 1$ ,  $\sigma_I = 40$ ,  $\sigma_J = 30$  and 4 iterations of the algorithm. We observed that the parameter  $\sigma_R$  of the trilateral filter, which controls its radiometric influence, has a great impact on the performances of the code and that its optimal value strongly depends on the noise level. For this reason, we keep for each experiment the parameter  $\sigma_R$  which yields the best PSNR (the same value of  $\sigma_R$  is used for the MAE and SSIM tables). For ROLD-EPR, we iterate the algorithm 7 times and keep all parameters as proposed in [21]. In the PSNR table, we also put into brackets the best PSNR of the algorithm along the iterations.

Table 6.1 shows quite similar performances between PARIGI and ROLD-EPR on most images (the SSIM results are particularly close), with a slight advantage of the patch-based approach for small noise ratios, and the opposite for large noise ratios. This proves that a patch-based approach is well founded for random-valued impulse noise removal. Much larger differences in PSNR, MAE and SSIM 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 regular textures in images. Let us insist on the fact that the parameters of our approach (patch size, research neighborhood size, number of iterations, ROAD threshold, number of trusted patches in function of  $p$ ) are fixed for all these experiments, as explained in Section 5.2. Better results can be obtained by optimizing these values for each image.

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 close to 0.1: for instance, for the scheme PARIGI on the image *Lena* and  $p = 0.4$ , the empirical standard deviation is 0.096. This means that if we observe, between two methods, PSNR differences smaller than 0.2, these differences can be considered as meaningless in practice.

Figures 6.1, 6.3, 6.2 and 6.4 provide a visual comparison of the different approaches on the different images when  $p = 20\%$  and  $p = 40\%$  (images should be seen at full resolution on the electronic version of the paper). When the impulse noise ratio is low, all methods are efficient and results are visually quite similar in smooth regions. Now, fine geometrical structures (for instance the ropes on the *Boat* image) are clearly better handled by our patch-based scheme. Conversely, ROLD-EPR seems to preserve better the image grain than other methods.

When the noise ratio increases, one can observe that the result of PARIGI is clearly smoother than the ones of ROLD-EPR [21] and ROAD-Trilateral [25], 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.

---

<sup>4</sup>The authors of [25] provide their code on their webpage [www.ssc.wisc.edu/~thuegeri](http://www.ssc.wisc.edu/~thuegeri).

		ROAD-Trilateral [25]	ROLD-EPR [21]	PARIGI
Lena	$p = 20\%$	33.83	33.52 (33.52)	35.45
	$p = 40\%$	30.38	30.84 (30.92)	31.75
	$p = 60\%$	24.80	28.10 (28.10)	27.38
Bridge	$p = 20\%$	27.23	26.74 (27.40)	27.68
	$p = 40\%$	24.32	24.37 (24.87)	24.80
	$p = 60\%$	21.15	22.60 (22.71)	22.03
Baboon	$p = 20\%$	23.85	23.32 (24.56)	24.17
	$p = 40\%$	21.96	21.62 (22.04)	22.02
	$p = 60\%$	19.87	20.59 (20.59)	20.13
Barbara	$p = 20\%$	24.98	24.80 (26.38)	33.91
	$p = 40\%$	22.97	23.45 (23.61)	29.92
	$p = 60\%$	20.87	22.49 (22.55)	24.93
Cameraman	$p = 20\%$	31.82	31.79 (32.47)	35.1
	$p = 40\%$	28.88	27.95 (29.38)	29.73
	$p = 60\%$	23.29	25.91 (26.36)	24.36
Boat	$p = 20\%$	30.04	29.31 (29.68)	31.21
	$p = 40\%$	26.94	26.41 (26.97)	27.56
	$p = 60\%$	23.14	24.68 (24.83)	23.68
Peppers	$p = 20\%$	33.59	33.62 (33.63)	34.75
	$p = 40\%$	30.70	30.97 (31.23)	31.63
	$p = 60\%$	24.72	28.57 (28.61)	27.58
Goldhill	$p = 20\%$	32.07	32.13 (32.25)	32.74
	$p = 40\%$	28.73	29.44 (29.63)	30.06
	$p = 60\%$	24.30	27.25 (27.26)	26.65

TABLE 6.1

PSNR results of different restoration filters for the  $512 \times 512$  images Lena, Bridge, Baboon, Barbara, Cameraman, Boat, Peppers and Goldhill for pure impulse noise. For the results of the trilateral filter, the parameter  $\sigma_R$  is optimized between 10 and 50 for each experiment. As for ROLD-EPR, we show into brackets the best possible PSNR obtained along the iterations of the algorithm. Other parameters are set as explained in the text.

**6.2. Gaussian and impulse noise mixture.** We now 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 [25, 41]. To the best of our knowledge, the paper of Xiao et al. [41] on Gaussian-impulse noise can be considered as the state of the art on the subject, and [25] is one of the first paper proposing to handle such a noise mixture. To be completely fair, we do not include the results of ROLD-EPR in the comparison, mainly because it was not developed to handle such a noise mixture and does not work well as soon as the Gaussian standard deviation increases. Nevertheless, we think that a similar variational approach could be adapted to this form of degradation.

Table 6.4 shows the PSNR results of the three methods for different combinations of Gaussian and impulse noise. The results for the  $l_1 - l_0$  minimization approach of Xiao *et al.* are taken from Tables 8 and 9 in [41], and thus come from different noise samples. The results of the trilateral filter are computed on the same noise samples as our approach. Again, for the trilateral filter, we choose for each mixture experiment the parameter  $\sigma_R$  between 10 and 50 which yields the best PSNR, the optimal  $\sigma_R$  begin strongly dependent on the noise ratio. Again, the PSNR standard deviation

		ROAD-Trilateral [25]	ROLD-EPR [21]	PARIGI
Lena	$p = 20\%$	2.60	1.45	1.71
	$p = 40\%$	4.10	2.80	2.75
	$p = 60\%$	7.33	4.87	4.62
Bridge	$p = 20\%$	5.90	4.36	4.95
	$p = 40\%$	9.31	7.95	7.81
	$p = 60\%$	13.70	11.20	11.88
Baboon	$p = 20\%$	8.83	7.43	7.48
	$p = 40\%$	11.81	11.70	11.39
	$p = 60\%$	16.75	14.76	15.49
Barbara	$p = 20\%$	6.37	5.36	1.97
	$p = 40\%$	9.21	7.98	3.52
	$p = 60\%$	12.87	10.16	6.88
Cameraman	$p = 20\%$	2.13	1.41	1.26
	$p = 40\%$	3.52	2.91	2.51
	$p = 60\%$	7.41	4.92	4.97
Boat	$p = 20\%$	3.93	2.61	2.82
	$p = 40\%$	5.92	5.02	4.59
	$p = 60\%$	9.51	7.38	7.74
Peppers	$p = 20\%$	2.88	1.50	1.96
	$p = 40\%$	4.02	2.85	2.92
	$p = 60\%$	7.24	4.63	4.53
Goldhill	$p = 20\%$	3.51	1.92	2.67
	$p = 40\%$	5.32	3.82	4.03
	$p = 60\%$	8.76	6.02	6.32

TABLE 6.2

MAE results of different restoration filters for the  $512 \times 512$  images *Lena*, *Bridge*, *Baboon*, *Barbara*, *Cameraman*, *Boat*, *Peppers* and *Goldhill* for pure impulse noise. For the results of the trilateral filter, the parameter  $\sigma_R$  is optimized between 10 and 50 for each experiment. Other parameters are set as explained in the text.

for one experiment is generally around 0.1, which means that PSNR differences of this order of magnitude should not be considered as meaningful. The performances of the different schemes are quite similar, with the notable exception of *Barbara*. As is the case for the pure impulse noise experiments, the performance differences on this image can be explained by the presence of several regular textures, better preserved by our patch-based approach. These results are also illustrated on Figures 6.5 and 6.6.

**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 other images. The visual experiments 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 6.7. In this example, we compare several denoising procedures on a synthetic  $256 \times 256$  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 (because of the image regularity, the result is even better by choosing larger patches). The variational approach of [21] and the trilateral

		ROAD-Trilateral [25]	ROLD-EPR [21]	PARIGI
Lena	$p = 20\%$	0.93	0.94	0.94
	$p = 40\%$	0.87	0.89	0.91
	$p = 60\%$	0.70	0.80	0.83
Bridge	$p = 20\%$	0.85	0.86	0.88
	$p = 40\%$	0.71	0.74	0.76
	$p = 60\%$	0.55	0.61	0.55
Baboon	$p = 20\%$	0.77	0.78	0.78
	$p = 40\%$	0.65	0.62	0.63
	$p = 60\%$	0.47	0.50	0.46
Barbara	$p = 20\%$	0.84	0.84	0.96
	$p = 40\%$	0.72	0.75	0.92
	$p = 60\%$	0.57	0.66	0.82
Cameraman	$p = 20\%$	0.96	0.95	0.97
	$p = 40\%$	0.91	0.92	0.93
	$p = 60\%$	0.70	0.84	0.86
Boat	$p = 20\%$	0.88	0.90	0.90
	$p = 40\%$	0.79	0.82	0.83
	$p = 60\%$	0.63	0.72	0.70
Peppers	$p = 20\%$	0.90	0.93	0.92
	$p = 40\%$	0.85	0.88	0.88
	$p = 60\%$	0.67	0.80	0.82
Goldhill	$p = 20\%$	0.89	0.92	0.90
	$p = 40\%$	0.79	0.84	0.84
	$p = 60\%$	0.63	0.74	0.73

TABLE 6.3

*SSIM results of different restoration filters for the  $512 \times 512$  images Lena, Bridge, Baboon, Barbara, Cameraman, Boat, Peppers and Goldhill for pure impulse noise.*

filter [25] 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 [21] 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



		ROAD-Trilateral [25]	Xiao [41]	PARIGI
Lena	$p = 10\% \sigma = 5$	34.49	34.98	34.72
	$p = 30\% \sigma = 5$	31.12	32.04	32.57
	$p = 10\% \sigma = 15$	30.20	30.85	30.31
	$p = 30\% \sigma = 15$	28.48	29.11	29.22
Bridge	$p = 10\% \sigma = 5$	28.05	–	26.96
	$p = 30\% \sigma = 5$	25.57	–	25.45
	$p = 10\% \sigma = 15$	25.59	–	25.34
	$p = 30\% \sigma = 15$	24.14	–	23.38
Baboon	$p = 10\% \sigma = 5$	24.62	–	24.81
	$p = 30\% \sigma = 5$	22.74	–	23.05
	$p = 10\% \sigma = 15$	23.12	–	23.63
	$p = 30\% \sigma = 15$	21.84	–	21.81
Barbara	$p = 10\% \sigma = 5$	25.62	30.48	31.55
	$p = 30\% \sigma = 5$	23.75	25.92	29.28
	$p = 10\% \sigma = 15$	23.91	27.31	28.8
	$p = 30\% \sigma = 15$	22.85	24.55	27.33
Cameraman	$p = 10\% \sigma = 5$	32.25	–	34.98
	$p = 30\% \sigma = 5$	30.18	–	31.40
	$p = 10\% \sigma = 15$	29.60	–	30.33
	$p = 30\% \sigma = 15$	28.22	–	28.59
Boat	$p = 10\% \sigma = 5$	30.51	–	31.41
	$p = 30\% \sigma = 5$	28.16	–	28.81
	$p = 10\% \sigma = 15$	27.48	–	28.21
	$p = 30\% \sigma = 15$	26.16	–	26.57
Peppers	$p = 10\% \sigma = 5$	34.57	–	33.90
	$p = 30\% \sigma = 5$	31.65	–	32.38
	$p = 10\% \sigma = 15$	30.50	–	30.28
	$p = 30\% \sigma = 15$	28.96	–	29.39
Goldhill	$p = 10\% \sigma = 5$	32.52	–	32.60
	$p = 30\% \sigma = 5$	29.65	–	30.64
	$p = 10\% \sigma = 15$	28.70	–	29.08
	$p = 30\% \sigma = 15$	27.45	–	27.99

TABLE 6.4

PSNR results of different restoration filters for the  $512 \times 512$  images Lena, Bridge, Baboon, Barbara, Cameraman, Boat, Peppers and Goldhill, and different mixtures of Gaussian and impulse noise. For the trilateral filter, we take the best PSNR obtained for all values of  $\sigma_R$  between 10 and 50.

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 [10] and [31] in the case of Gaussian noise.

## REFERENCES

- [1] C. Aguerrebere, J. Delon, Y. Gousseau, and P. Musé. Study of the digital camera acquisition process and statistical modeling of the sensor raw data. <http://hal.archives-ouvertes.fr/hal-00733538>.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] J. Cai, R. Chan, and M. Nikolova. Two-phase approach for deblurring images corrupted by impulse plus Gaussian noise. *Inverse Problems and Imaging*, 2(2):187–204, 2008.
- [7] J. Cai, R. Chan, and M. Nikolova. Fast two-phase image deblurring under impulse noise. *Journal of Mathematical Imaging and Vision*, 36(1):46–53, 2010.
- [8] 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.
- [9] P. Chatterjee and P. Milanfar. Clustering-based denoising with locally learned dictionaries. *Image Processing, IEEE Transactions on*, 18(7):1438–1451, 2009.
- [10] P. Chatterjee and P. Milanfar. Is denoising dead? *IEEE Transactions on Image Processing*, 19(4):895–911, 2010.
- [11] T. Chen and H. Wu. Adaptive impulse detection using center-weighted median filters. *Signal Processing Letters, IEEE*, 8(1):1–3, 2001.
- [12] V. Crnojevic, V. Senk, and Z. Trpovski. Advanced impulse detection based on pixel-wise MAD. *Signal Processing Letters, IEEE*, 11(7):589–592, 2004.
- [13] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image restoration by sparse 3d transform-domain collaborative filtering. In *SPIE Electronic Imaging '08*, volume 6812-07, 2008.
- [14] H. David and H. Nagaraja. *Order statistics*. Wiley series in probability and mathematical statistics. Probability and mathematical statistics. John Wiley, 2003.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] C.-A. Deledalle, J. Salmon, and A. S. Dalalyan. Image denoising with patch based pca: local versus global. In *BMVC*, 2011.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian. Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. *Image Processing, IEEE Transactions on*, 17(10):1737–1754, 2008.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] Y.-M. Huang, M. Ng, and Y.-W. Wen. Fast Image Restoration Methods for Impulse and Gaussian Noises Removal. *IEEE Signal Processing Letters*, 16(6):457–460, 2009.
- [29] 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.
- [30] 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.
- [31] A. Levin and B. Nadler. Natural image denoising: Optimality and inherent bounds. In *CVPR*, pages 2833–2840, 2011.
- [32] 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.
- [33] Y.-R. Li, L. Shen, D.-Q. Dai, and B. W. Suter. Framelet Algorithms for De-Blurring Images Corrupted by Impulse Plus Gaussian Noise. *IEEE Transactions on Image Processing*, 20(7):1822–1837, 2011.
- [34] E. López-Rubio. Restoration of images corrupted by Gaussian and uniform impulsive noise. *Pattern Recognition*, 43(5):1835–1846, 2010.
- [35] C. Louchet and L. Moisan. Total variation as a local filter. *SIAM Journal on Imaging Sciences*, 4(2):651–694, 2011.
- [36] W. K. Pratt. Median filtering. Technical report, Image Proc. Inst., Univ. Southern California, 1975.
- [37] J. Salmon and Y. Strozecski. Patch rejections for Non Local methods. *Signal Processing*, 92(2):447–489, 2012.
- [38] T. Sun and Y. Neuvo. Detail-preserving median based filters in image processing. *Pattern Recognition Letters*, 15(4):341–347, 1994.
- [39] D. van de Ville and M. Kocher. SURE-Based Non-Local Means. *IEEE Signal Processing Letters*, 16:973–976, Nov. 2009.
- [40] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: From error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, 2004.
- [41] 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.
- [42] 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.

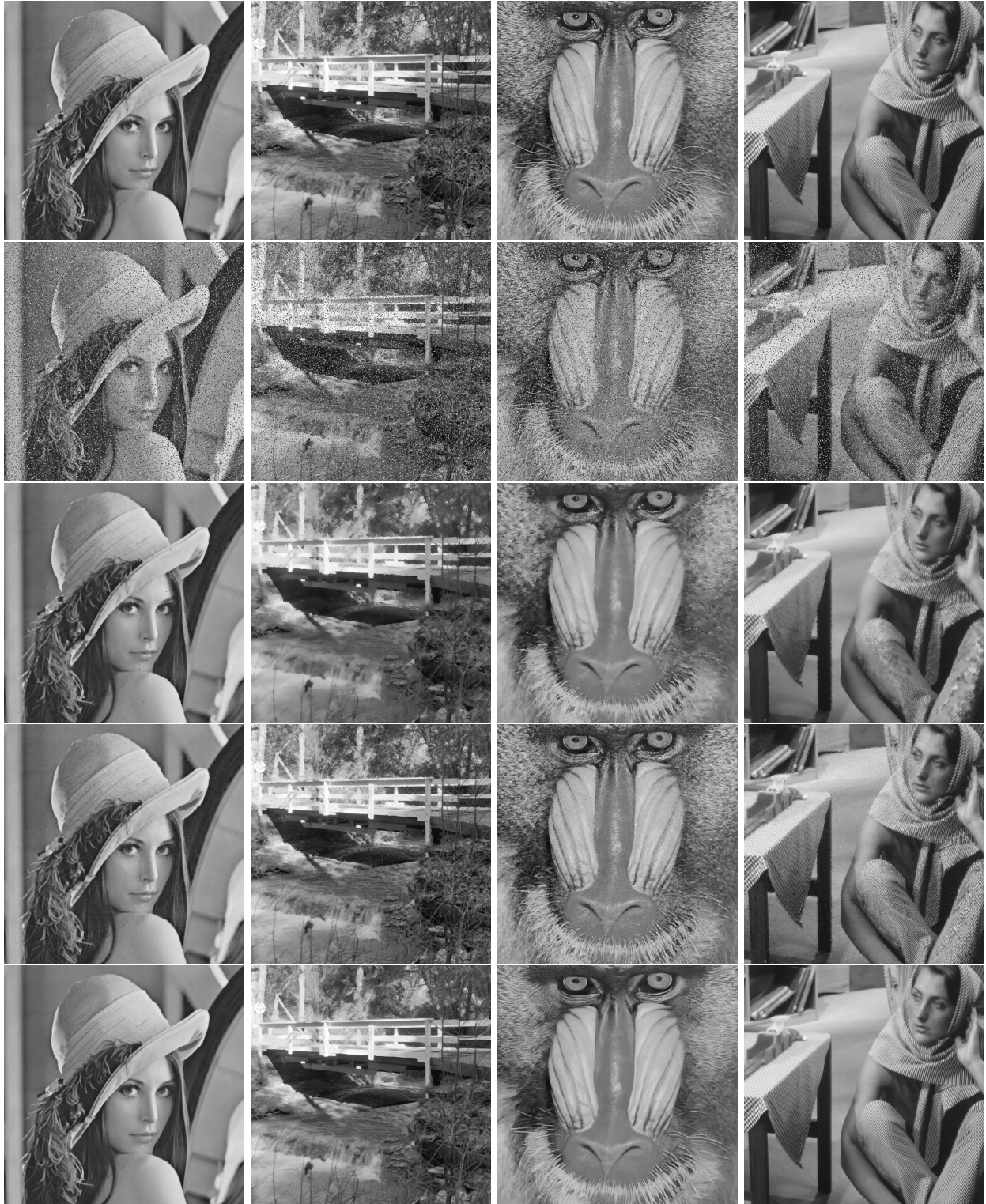


FIG. 6.1. Comparative results on *Lena*, *Bridge*, *Baboon* and *Barbara* with  $p = 20\%$  of random-valued impulse noise. For each column, from top to bottom: original image, noisy image, ROAD+trilateral [25], ROLD+EPR [21], PARIGI. Images should be seen at full resolution on the electronic version of the paper.

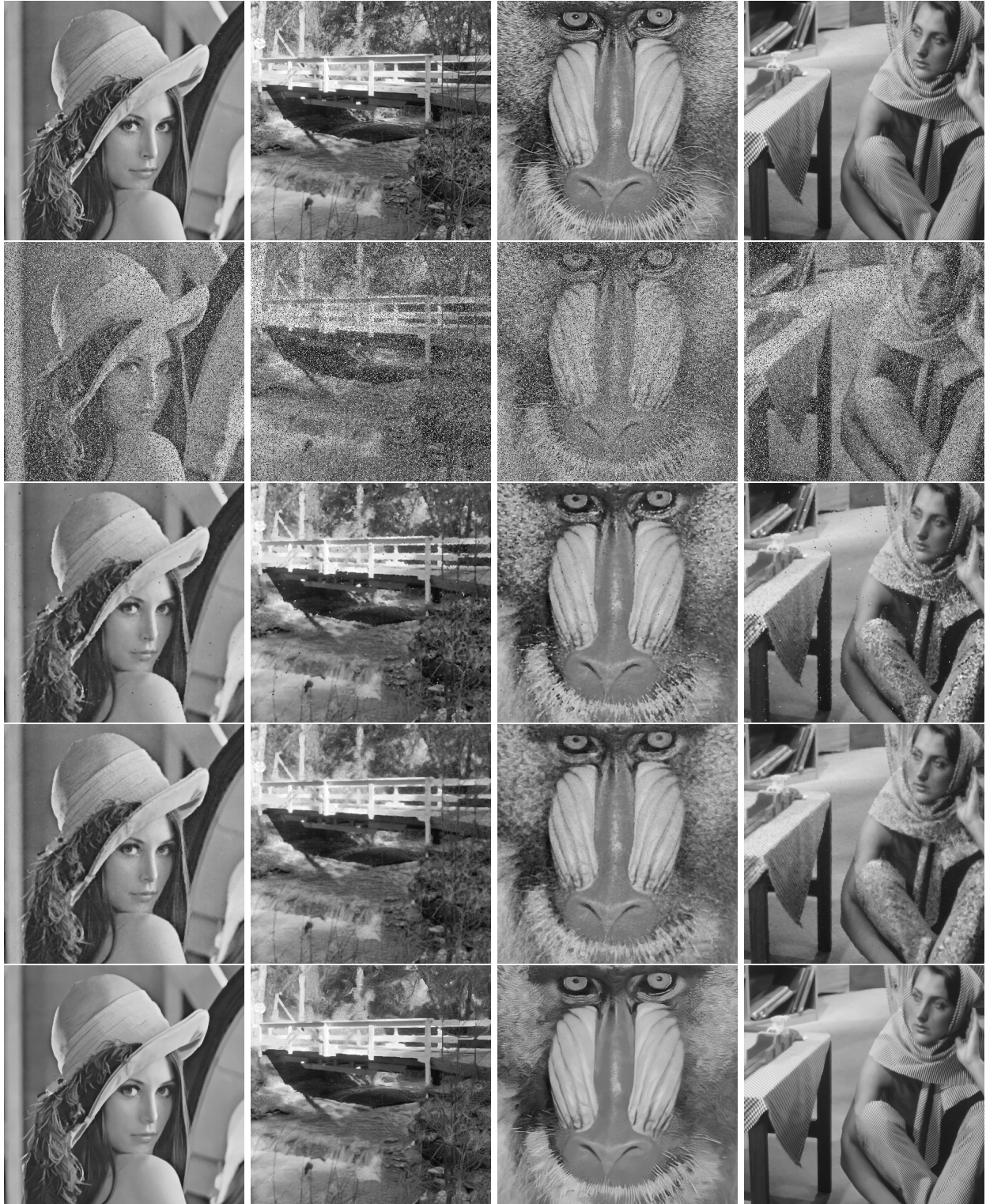


FIG. 6.2. Comparative results on *Lena*, *Bridge*, *Baboon* and *Barbara* with  $p = 40\%$  of random-valued impulse noise. For each column, from top to bottom: original image, noisy image, ROAD+trilateral [25], ROLD+EPR [21], PARIGI. Images should be seen at full resolution on the electronic version of the paper.



FIG. 6.3. Comparative results on *Cameraman*, *Boat*, *Peppers* and *Goldhill* with  $p = 20\%$  of random-valued impulse noise. For each column, from top to bottom: original image, noisy image, ROAD+trilateral [25], ROLD+EPR [21], PARIGI. Images should be seen at full resolution on the electronic version of the paper.



FIG. 6.4. Comparative results on *Cameraman*, *Boat*, *Peppers* and *Goldhill* with  $p = 40\%$  of random-valued impulse noise. For each column, from top to bottom: original image, noisy image, ROAD+trilateral [25], ROLD+EPR [21], PARIGI. Images should be seen at full resolution on the electronic version of the paper.



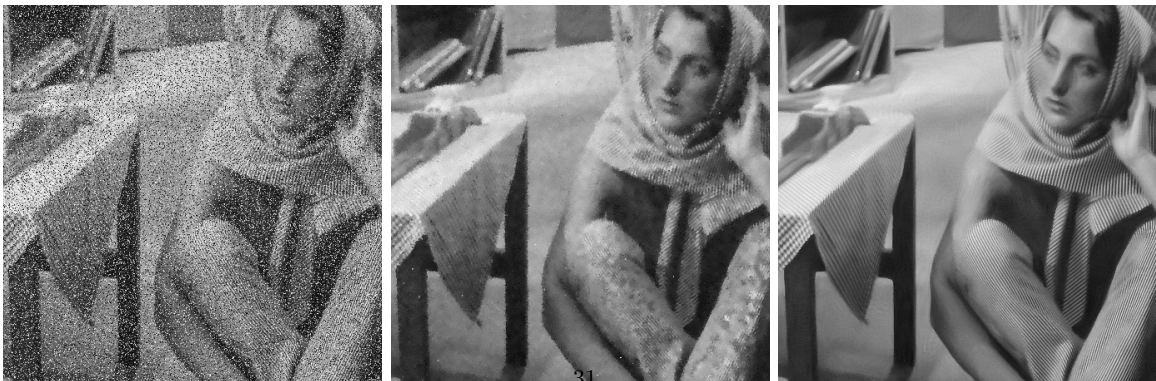
(a)  $p = 0.1, \sigma = 5$



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



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



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

FIG. 6.5. Comparative results on Lena and Barbara with  $p \in \{10\%, 30\%\}$  of random-valued impulse noise and  $\sigma \in \{5, 15\}$ . For each subfigure, from left to right: noisy image, result of the trilateral filter [25], PARIGI.





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



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

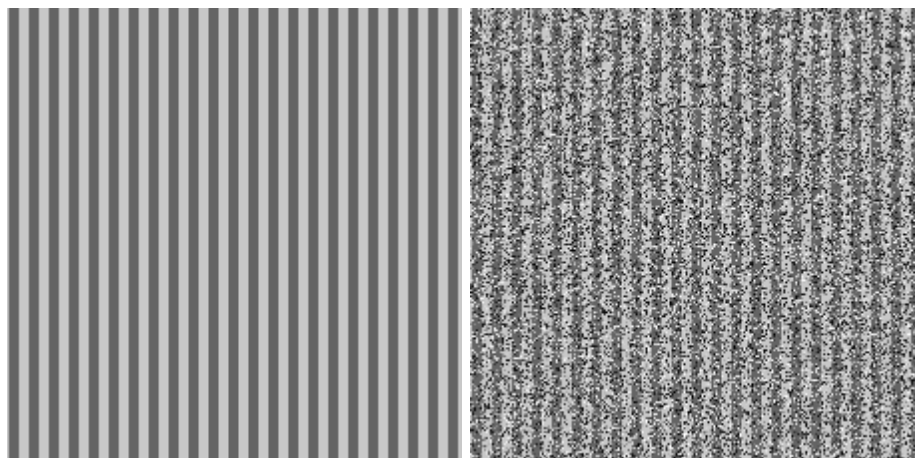


(c)  $p = 0.1, \sigma = 5$



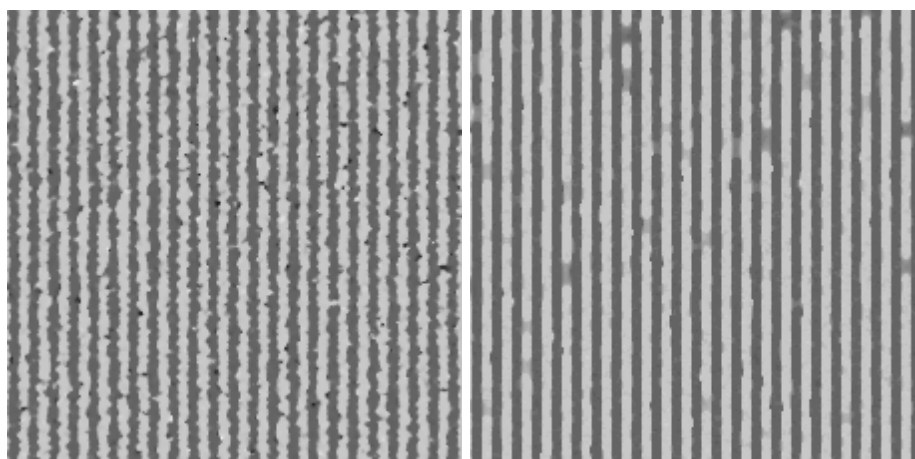
(d)  $p = 0.3, \sigma = 15$

FIG. 6.6. Comparative results on Cameraman and Boat with  $p \in \{10\%, 30\%\}$  of random-valued impulse noise and  $\sigma \in \{5, 15\}$ . For each subfigure, from left to right: noisy image, result of the trilateral filter [25], PARIGI.



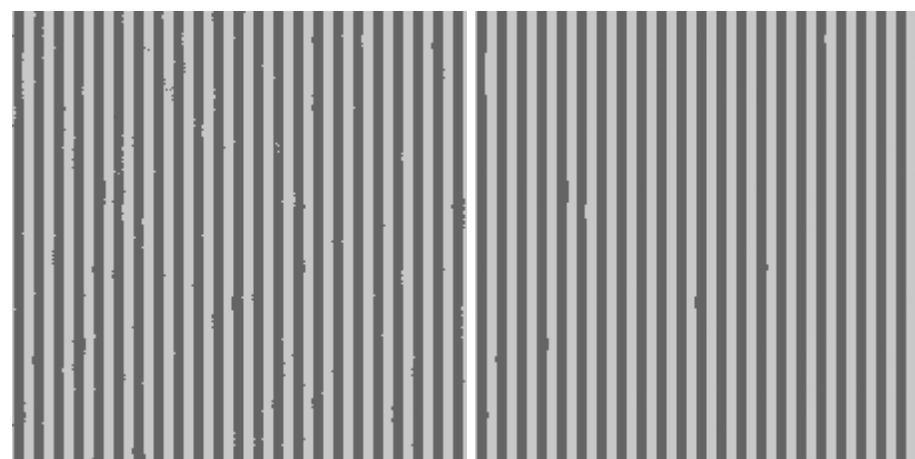
(a) Original image

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



(c) Trilateral filter [25], PSNR = 19.77

(d) ROLD-EPR [21], PSNR = 26.63



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

FIG. 6.7. Comparison of different denoising approaches on the image *Stripes*, with  $p = 50\%$ .