CrossMark

Removing Mixture of Gaussian and Impulse Noise by Patch-Based Weighted Means

Haijuan Hu[1,2,3] · Bing Li[4] · Quansheng Liu[3]

Abstract We first establish a law of large numbers and a convergence theorem in distribution to show the rate of convergence of the non-local means filter for removing Gaussian noise. Based on the convergence theorems, we propose a patch-based weighted means filter for removing an impulse noise and its mixture with a Gaussian noise by combining the essential idea of the trilateral filter and that of the non-local means filter. Experiments show that our filter is competitive compared to recently proposed methods. We also introduce the notion of degree of similarity to measure the impact of the similarity among patches on the non-local means filter for removing a Gaussian noise, as well as on our new filter for removing an impulse noise or a mixed noise. Using again the convergence theorem in distribution, together with the notion of degree of similarity, we obtain an estimation for the PSNR value of the denoised image by the non-local means filter or by the new proposed filter, which is close to the real PSNR value.

Keywords Gaussian noise · Impulse noise · Mixed noise · Trilateral filter · Non-local means filter · Convergence theorems · Degree of similarity · Estimation of PSNR

✉ Quansheng Liu
  Quansheng.Liu@univ-ubs.fr

  Haijuan Hu
  huhaijuan61@126.com

  Bing Li
  libcs@263.net

1 School of Mathematics and Statistics, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China

2 College of Information Science and Engineering, Northeastern University, Shenyang 110819, China

3 CNRS UMR 6205, LMBA, Université de Bretagne-Sud, Campus de Tohannic, 56000 Vannes, France

4 Department of Mathematics, Zhongshan Polytechnic, Zhongshan 528404, China

⚹ Springer

1 Introduction

Images are produced to record or display useful information. Due to the visibility of images and the rapid development of science and technology, images play an increasingly important role in our lives. However, because of imperfections in the imaging and capturing process, the recorded image invariably represents a degraded version of the original scene [3,5]. The undoing of these imperfections is crucial to many of the subsequent image processing tasks.

There exists a wide range of different degradations. A very important example is the existence of noise. Noise may be introduced by the medium through which the image is created and transmitted. In this paper, we concentrate on removing impulse noise and its mixture with Gaussian noise.

We present a numerical image by a $M \times N$ matrix $u = \{u(i) : i \in I\}$, where $I = \{0, 1, \ldots, M - 1\} \times \{0, 1, \ldots, N - 1\}$ is the image domain, and $u(i) \in \{0, 1, 2, \ldots, 255\}$ represents the gray value at the pixel i for 8-bit gray images. The additive Gaussian noise model is:

$$v(i) = u(i) + \eta(i),$$

where $u = \{u(i) : i \in I\}$ is the original image, $v = \{v(i) : i \in I\}$ is the noisy one, and η is the Gaussian noise: $\eta(i)$ are independent and identically distributed Gaussian random variables with mean 0 and standard deviation σ. In the sequel we always denote by u the original image, and v the noisy one. The random impulse noise model is:

$$v(i) = \begin{cases} \eta(i) & \text{with probability } p, \\ u(i) & \text{with probability } (1 - p), \end{cases}$$

where $0 < p < 1$ is the impulse probability (the proportion of the occurrence of the impulse noise), and $\eta(i)$ are independent random variables uniformly distributed on $[\min\{u(i) : i \in I\}, \max\{u(i) : i \in I\}]$, generally taken as [0,255] for 8-bit gray images. For the mixture of Gaussian noise and impulse noise model, an image is first added by Gaussian noise and then contaminated by impulse noise. The task of image denoising is to recover the unknown original image u as well as possible from the degraded one v.

Many denoising methods have been developed in the literature. To remove a Gaussian noise, there are approaches based on wavelets [7,10,15], approaches based on variational models [16,32], and weighted means approaches [4,23,33,34,41], etc. An important progress was marked by the proposition of the *non-local means filter* [4], abbreviated as *NL-means*; this filter estimates original images by weighted means along similar local patches. Since then, many researchers have combined the basic idea of NL-means with some other methods to remove noise, see for instance [11,18,20,27]. There are also many methods to remove an impulse noise, including median based filters [1,8,30], fuzzy filters [42], and variational based methods [6,14,29].

The above-mentioned methods can only be applied to remove one kind of noise (a Gaussian noise or an impulse noise), and can not be used to remove a mixture of a Gaussian noise and an impulse noise. To remove a mixed noise, a successful method is the *trilateral filter* proposed in [19], where an interesting statistic called ROAD (*Rank of Ordered Absolute Differences)* is introduced to detect the impulse noisy pixels; this filter combines the ROAD statistic with the *bilateral filter* [33,34] to remove the noise. The trilateral filter [19] is also effective for removing an impulse noise; a variant of the ROAD statistic, named ROLD (*Rank-Ordered Logarithmic Difference*), has been proposed in [14], where it is combined

with the edge-preserving variational method [6] for removing impulse noise. Other methods have also been developed recently to remove a mixed noise. The papers [12,13,21,24,38] use the patch-based idea of NL-means to remove an impulse noise and a mixed noise. In [21,24], generalizations of NL-means are proposed for removing an impulse noise and its mixture with a Gaussian noise, using the ROAD statistic [19]; the main idea therein is to define weights in terms of the ROAD statistic and the similarity of local patches, which are nearly zero for impulse noisy points. In [38], NL-means is adapted by estimating the similarity of patches with the reference image obtained in an impulse noise detection mechanism. The papers [12,13] also use a patch-based approach, where a robust distance is introduced (inspired by order statistics) to estimate the similarity between patches using the tail of the binomial distribution, and the maximum likelihood estimator (MLE) is used to estimate the original image. The methods in [40] and [37,43] use the ideas of BM3D [11] and K-SVD [18] (which are the state-of-the-art algorithms for Gaussian noise removal) respectively to remove a mixed noise; the algorithm proposed in [25] is based on a Bayesian classification of the input pixels, which is combined with the kernel regression framework.

The NL-means filter [4] explores in a nice way the similarity phenomenon existing very often in natural images. As stated above, many filters have been proposed based on the basic idea of NL-means. But the theoretic aspects have not been so much studied. A probabilistic explanation called *similarity principle* was given in [24]. In this paper we will improve this principle by proving a Marcinkiewicz law of large numbers and a convergence theorem in distribution, which describe the rate of convergence of NL-means.

Based on the convergence theorems, we will propose a new filter, called *Patch-based Weighted Means Filter (PWMF)* to improve the *Mixed Noise Filter (MNF)* introduced in [24]. Compared to MNF, the new filter simplifies the joint impulse factor in MNF, adds a spatial factor so that the filter extends entirely both the trilateral filter and the NL-means filter, and adjusts the choice of parameters. Experimental results show that our new filter is competitive for removing both the impulse noise and the mixed noise, compared to recently developed filters [13,19,37,38,40].

To well understand the impact of the similarity on the quality of NL-means, we will introduce the notion of *degree of similarity*. We will see that, in general, the larger the value of degree of similarity, the better the restoration result by NL-means. Furthermore, using the degree of similarity together with the convergence theorem in law, we will give a good estimation of the PSNR value of the denoised image (without knowing the original image). Our simulations show that the estimated PSNR value is quite close to the real one when the original image is known.

This paper is an extended and improved version of our conference paper [21]; it also develops and improves the earlier work [24].

The rest of this paper is organized as follows. In Sect. 2, we recall the non-local means filter [4], and present two convergence theorems (Theorems 1 and 2) to show the rate of convergence of NL-means. The proofs of the convergence theorems are postponed in an appendix by the end of the paper. In Sect. 3, we recall the trilateral filter [19] and introduce our new patch-based weighted means filter. Experiments are presented in Sect. 4 to compare the new filter with some recently proposed ones. In Sect. 5, the notion of degree of similarity is first introduced, and then applied to estimations of PSNR values, using the convergence theorem in law. Conclusions are made in Sect. 6. The paper is ended by an appendix in which we give the proofs of the convergence theorems, by establishing two more general convergence theorems for random weighted means of l-dependent random variables.

2 Convergence Theorems for Non-local Means

In this section, we present a Marcinkiewicz law of large numbers and a convergence theorem in distribution for NL-means [4] to remove Gaussian noise.

2.1 Non-local Means

Begin with some notation. For $i \in I$ and $d > 1$ an odd integer, let

$$\mathcal{N}_i(d) = \left\{ j \in I : \|j - i\|_\infty \leq \frac{d - 1}{2} \right\} \quad \text{and} \quad \mathcal{N}_i^0(d) = \mathcal{N}_i(d) \backslash \{i\},$$

with $\| \cdot \|_\infty$ denoting the sup norm:

$$\|j - i\|_\infty = \max(|j_1 - i_1|, |j_2 - i_2|) \quad \text{if} \quad i = (i_1, i_2) \text{ and } j = (j_1, j_2).$$

In other words, $\mathcal{N}_i(d)$ is the window with center i and size $d \times d$, and $\mathcal{N}_i^0(d)$ is the same window but with the center i deleted. We sometimes simply write $\mathcal{N}_i$ and $\mathcal{N}_i^0$ for $\mathcal{N}_i(d)$ and $\mathcal{N}_i^0(d)$, respectively. Denote by

$$v(\mathcal{N}_i) = \{v(k) : k \in \mathcal{N}_i\}$$

the vector composed of the gray values of v in the window $\mathcal{N}_i$ arranged lexicographically; it represents the local oriented image patch defined on the window $\mathcal{N}_i$.

For another odd integer D, let $\mathcal{N}_i(D)$ be the window with center i and size $D \times D$, defined in the same way as we did for $\mathcal{N}_i(d)$. *The denoised image by NL-means is given by*

$$\text{NLM}(v)(i) = \frac{\sum_{j \in \mathcal{N}_i(D)} w(i, j) v(j)}{\sum_{j \in \mathcal{N}_i(D)} w(i, j)}, \quad i \in I, \tag{1}$$

with

$$w(i, j) = e^{-\|v(\mathcal{N}_i) - v(\mathcal{N}_j)\|_a^2 / (2\sigma_r^2)}, \tag{2}$$

where $\sigma_r > 0$ is a control parameter,

$$\|v(\mathcal{N}_i) - v(\mathcal{N}_j)\|_a^2 = \frac{\sum_{k \in \mathcal{N}_i(d)} a(i, k) |v(k) - v(\mathcal{T}(k))|^2}{\sum_{k \in \mathcal{N}_i(d)} a(i, k)}, \tag{3}$$

$a(i, k) > 0$ being some fixed weights usually chosen to be a decreasing function of the Euclidean norm $\|i - k\|$ or the sup norm $\|i - k\|_\infty$, and $\mathcal{T} = \mathcal{T}_{ij}$ is the translation mapping i to j (thus mapping $\mathcal{N}_i$ onto $\mathcal{N}_j$):

$$\mathcal{T}(k) = k - i + j, \quad k \in \mathcal{N}_i.$$

We call $\mathcal{N}_i(D)$ search windows, and $\mathcal{N}_i = \mathcal{N}_i(d)$ local patches. Theoretically, the search window $\mathcal{N}_i(D)$ in (1) can be chosen as the whole image I; but in practice, it is better to choose $\mathcal{N}_i(D)$ with an appropriate number D not too large.

2.2 Convergence Theorems

We now present two convergence theorems for NL-means using probability theory. Following [24], first give a definition of similarity and recall the notion of l-dependence.

For simplicity, use the same notation $v(\mathcal{N}_i)$ to denote both the observed image patch centered at i and the corresponding random variable (in fact the observed image is just a

realization of the corresponding variable). Therefore the distribution of the observed image $v(\mathcal{N}_i)$ is just that of the corresponding random variable.

Definition 1 (*Similarity*) Two patches $v(\mathcal{N}_i)$ and $v(\mathcal{N}_j)$ are called similar if they have the same probability distribution.

Definition 1 is a probabilistic interpretation of the similarity phenomenon. According to this definition, two observed patches $v(\mathcal{N}_i)$ and $v(\mathcal{N}_j)$ are similar if they are issued from the same probability distribution. In practice, we consider that two patches $v(\mathcal{N}_i)$ and $v(\mathcal{N}_j)$ are similar if their Euclidean distance is small enough, say $\|v(\mathcal{N}_i) - v(\mathcal{N}_i)\| < T$ for some threshold T. We will come back to this point later in Sect. 5 when we consider the notion of degree of similarity.

Note that Definition 1 is equivalent to say that the non-noisy patches $u(\mathcal{N}_i)$ and $u(\mathcal{N}_j)$ are equal. In fact, when $v(\mathcal{N}_i)$ and $v(\mathcal{N}_j)$ have the same distribution, they have the same expected value, so that $u(\mathcal{N}_i) = u(\mathcal{N}_j)$. The converse is also easy: when $u(\mathcal{N}_i) = u(\mathcal{N}_j)$, then $v(\mathcal{N}_i)$ and $v(\mathcal{N}_j)$ have the same distribution, as so do $\eta(\mathcal{N}_i)$ and $\eta(\mathcal{N}_j)$. The definition seems to be somehow restrictive, but this models the ideal situation, and coincides with our intuition that $u(\mathcal{N}_i)$ and $u(\mathcal{N}_j)$ are very close when the patches are similar.

Definition 2 (*l-dependence*) For an integer $l \geq 0$, a sequence of random variables $X_1, X_2, \ldots$ is called to be l-dependent if each subsequence $X_{k_1}, X_{k_2}, \ldots$ is independent whenever $|k_m - k_n| > l$ for all $m, n \geq 1$. (That is, random variables with distances strictly greater than l are independent of each other.)

Fix two odd integers $d > 1$ and $D > 1$. For $i \in I$, define

$$I_i = \{j \in \mathcal{N}_i(D): \quad v(\mathcal{N}_i) \text{ and } v(\mathcal{N}_j) \text{ are similar}\}. \tag{4}$$

For convenience, we write I_i in the form

$$I_i = \{j_1, j_2, \ldots j_n\} \quad \text{with } n = |I_i| \tag{5}$$

(throughout the paper for a set S we write $|S|$ for the cardinality of S). The elements $j \in I_i$ are ordered according to the increasing order of $\|j - i\|_\infty$, and for the j's with the same distance $\|j - i\|_\infty$, say $\|j - i\|_\infty = c$ for some constant c, they are ordered anticlockwise beginning from the upper left corner of the quadrilateral $\{x \in \mathbb{R}^2 : \|x - i\|_\infty = c\}$. Notice that $v(\mathcal{N}_i)$ and $v(\mathcal{N}_j)$ are independent if $\mathcal{N}_i \cap \mathcal{N}_j = \emptyset$. Since $\mathcal{N}_j \cap \mathcal{N}_j \neq \emptyset$ if and only if $\|j - i\|_\infty \leq d - 1$, there are precisely $(2d - 1)^2 - 1$ windows $\mathcal{N}_j$ with $j \neq i$ which intersect $\mathcal{N}_i$ (as the elements $j \neq i$ with $\|j - i\|_\infty \leq d - 1$ constitute a window centered at i of size $(2d - 1) \times (2d - 1)$, whose center is deleted). Therefore, we have:

Lemma 1 *For each $i \in I$, with the order of I_i that we defined above, the sequence of random vectors $\{v(\mathcal{N}_j) : j \in I_i\}$ is l-dependent for $l = (2d - 1)^2 - 1$.*

Actually, very often we can take l smaller, as we only consider similar patches. Notice that if we choose the lexicographical order for I_i, then we need to choose l much larger for the l-dependence. This is why we ordered $j \in I_i$ according to the increasing order of $\|j - i\|_\infty$.

As usual, for two sequences of real numbers a_n and b_n, we write

$$a_n = o(b_n) \text{ if } \lim_{n \to \infty} \frac{a_n}{b_n} = 0, \text{ and } a_n = O(b_n) \text{ if } \limsup_{n \to \infty} \frac{|a_n|}{|b_n|} < \infty.$$

Accordingly, when a_n and b_n are random, $a_n = o(b_n)$ (resp. $a_n = O(b_n)$) almost surely means that

$$\lim_{n \to \infty} \frac{a_n}{b_n} = 0 \quad \left(\text{resp.} \ \limsup_{n \to \infty} \frac{|a_n|}{|b_n|} < \infty \right) \quad \text{almost surely.}$$

The following theorem is a kind of Marcinkiewicz law of large numbers. It gives an estimation of the almost sure convergence rate of the estimator to the original image for the non-local means filter.

Theorem 1 *Let $i \in I$ and let I_i be the set of j such that the patches $v(\mathcal{N}_i)$ and $v(\mathcal{N}_j)$ are similar (in the sense of Definition 1). Set*

$$v^0(i) = \frac{\sum_{j \in I_i} w^0(i, j) v(j)}{\sum_{j \in I_i} w^0(i, j)}, \tag{6}$$

where

$$w^0(i, j) = e^{-\|v(\mathcal{N}_i^0) - v(\mathcal{N}_j^0)\|_a^2 / (2\sigma_r^2)}. \tag{7}$$

Then for any $\epsilon \in (0, \frac{1}{2}]$, as $|I_i| \to \infty$,

$$v^0(i) - u(i) = o\left(|I_i|^{-(\frac{1}{2} - \epsilon)} \right) \quad \text{almost surely}, \tag{8}$$

where $|I_i|$ denotes the cardinality of I_i.

Notice that when $\epsilon = \frac{1}{2}$, (8) means that

$$\lim_{|I_i| \to \infty} v^0(i) = u(i) \quad \text{almost surely} \tag{9}$$

which is the similarity principle in [24].

Recall that $\mathcal{N}_i^0 = \mathcal{N}_i \setminus \{i\}$. Theorem 1 shows that $v^0(i)$ is a good estimator of the original image $u(i)$ if the number of similar patches $|I_i|$ is sufficiently large. Here we use the weight $w^0(i, j)$ instead of $w(i, j)$, as $w^0(i, j)$ has the nice property that it is independent of $v(j)$ if $j \notin \mathcal{N}_i$. This property is used in the proof, and makes the estimator $v^0(i)$ to be nearly non-biased: in fact, if the family $\{v(j)\}_j$ is independent of the family $\{w^0(i, j)\}_j$ (e.g. this is the case when the similar patches are disjoint), then it is evident that $\mathbb{E}v^0(i) = u(i)$. We can consider that this non-biased property holds approximately as for each j there are few k such that $w^0(i, k)$ is dependent of $v(j)$. A closely related explanation about the biased estimation of NL-means can be found in [39].

Notice that when $v(\mathcal{N}_j)$ is not similar to $v(\mathcal{N}_i)$, the weight $w^0(i, j)$ is small and negligible. Therefore it is also reasonable to take all patches for the calculation. Indeed, taking just similar patches or all patches does not make much difference for the denoising results, as shown in Table 1 of [2]. However, selecting only similar patches can slightly improve the restoration result, and can also speed up the algorithm, as shown in [2,26] where a pre-classification of patches is proposed so that we only need calculate the weights for similar patches. On the other hand, as illustrated by [39], the difference between $w^0(i, j)$ and $w(i, j)$ is also small,

Table 1 Choice of search window sizes D for PWMF

	$\sigma = 0$	$\sigma = 10$	$\sigma = 20$	$\sigma = 30$
D	7	7	11	15

but $w^0(i)$ often gives a little better restoration result. Therefore, Theorem 1 not only gives a mathematical justification of the original non-local means filter (showing that $\mathrm{NLM}(v)(i)$ is a reasonable estimator of $u(i)$), but also suggests some improvements by taking just similar patches instead of all patches, and using the weights $w^0(i, j)$ instead of $w(i, j)$.

The next result is a convergence theorem in distribution. It states that $v^0(i) - u(i) \to 0$ with a rate as $1/\sqrt{|I_i|}$ in the sense of distribution.

Theorem 2 *Under the condition of Theorem 1, assume additionally that $\{v(\mathcal{N}_j) : j \in I_i\}$ is a stationary sequence of random vectors. Then as $|I_i| \to \infty$,*

$$\sqrt{|I_i|}\big(v^0(i) - u(i)\big) \xrightarrow{d} \mathcal{L},$$

where $\xrightarrow{d}$ means the convergence in distribution, and $\mathcal{L}$ is a mixture of centered Gaussian laws in the sense that it has a density of the form

$$f(t) = \int_{\mathbb{R}^{|\mathcal{N}_i^0|}} \frac{1}{\sqrt{2\pi}\,c_x} e^{-\frac{t^2}{2c_x^2}} \mu(dx), \tag{10}$$

μ being the law of $v(\mathcal{N}_i^0)$ and

$$c_x = \frac{1}{\mathbb{E}a_1}\sqrt{\mathbb{E}(a_1^2)\mathbb{E}(v_1 - \mathbb{E}v_1)^2 + 2\sum_{k=2}^{d^2} \mathbb{E}(a_1 a_k (v_1 - \mathbb{E}v_1)(v_k - \mathbb{E}v_k))}, \tag{11}$$

with

$$a_k = e^{-\|x - v(\mathcal{N}_{j_k}^0)\|^2/(2\sigma_r^2)}, \quad v_k = v(j_k). \tag{12}$$

Moreover, writing $m = |\mathcal{N}_i^0| = d^2 - 1$, $\nu = \mathbb{E}(v(\mathcal{N}_{j_k}^0))$ and

$$c(x) = \sigma^{m/2+1}\left(\frac{(\sigma^2 + \sigma_r^2)^2}{\sigma^2\sigma_r^2(2\sigma^2 + \sigma_r^2)}\right)^{m/4}\exp\left(\frac{\sigma^2\|x\|^2}{2(\sigma_r^2 + \sigma^2)(\sigma_r^2 + 2\sigma^2)}\right), \tag{13}$$

we have the approximations

$$c_x \approx c(x - \nu), \quad x \in \mathbb{R}^m, \tag{14}$$

and

$$f(t) \approx \tilde{f}(t) := \int_{\mathbb{R}^m} \frac{1}{\sqrt{2\pi}\,c(x)} e^{-\frac{t^2}{2c^2(x)}} \frac{1}{(\sqrt{2\pi}\sigma)^m} e^{-\frac{\|x\|^2}{2\sigma^2}}\,dx, \quad t \in \mathbb{R}. \tag{15}$$

To apply Theorem 2 we often need to calculate the probability of the form $\mathcal{L}(a, b) = \int_a^b f(t)dt$, where a, b are real numbers such that $a < b$. In practice, to this end we can replace the density function $f(t)$ by its approximation $\tilde{f}(t)$. But a direct calculation of $\tilde{f}(t)$ is not easy when m is large (as we have a multiple integral of order m whose numerical calculation is not easy). An efficient way for the calculation is to use the Monte-Carlo simulation as follows.

Remark 1 (Calculation of $\int_a^b f(t)dt$ by Monte-Carlo simulation) Let $-\infty < a < b < \infty$, and let X_i and U_i be independent random variables such that each X_i has the normal law $N(0, \sigma^2 Id_m)$ on $\mathbb{R}^m$ (with Id_m denoting the identity matrix of size $m \times m$), and each U_i has the uniform law on (a, b). Then for $a < b$ and k large enough,

$$\int_a^b f(t)dt \approx \int_a^b \tilde{f}(t)dt \approx (b-a)\sum_{i=1}^k \frac{1}{\sqrt{2\pi}\,c(X_i)} e^{-\frac{U_i^2}{2[c(X_i)]^2}}. \tag{16}$$

To see the approximation, it suffices to notice that, by the law of large numbers,

$$\frac{1}{b-a}\int_a^b \tilde{f}(t)dt = \lim_{k\to\infty}\sum_{i=1}^k \frac{1}{\sqrt{2\pi}c(X_i)}e^{-\frac{v_i^2}{2[c(X_i)]^2}} \quad \text{almost surely.} \tag{17}$$

By Theorems 1 and 2, the larger the value of $|I_i|$, the better the approximation of $v^0(i)$ to $u(i)$, that is to say, the larger the number of similar patches, the better the restored result. This will be confirmed in Sect. 5 where we shall introduce the notion of degree of similarity for images, showing that the larger the degree of similarity, the better the quality of restoration.

The proofs of the theorems will be given in Appendix.

3 Patch-Based Weighted Means Filter

In this section, we first introduce our new filter which combines the basic idea of NL-means [4] and that of the trilateral filter [19]. We then analyse the convergence of this new filter for removing mixed noise.

3.1 Trilateral Filter

The authors of [19] proposed a neighborhood filter called the trilateral filter as an extension of the bilateral filter [33,34] to remove random impulse noise and its mixture with Gaussian noise. Firstly, they introduced the statistic ROAD (Rank of Ordered Absolute Differences) to measure how like a point is an impulse noisy point defined by

$$\text{ROAD}(i) = r_1(i) + \cdots + r_m(i), \tag{18}$$

$r_k(i)$ being the k-th smallest term in the set $\{|u(i) - u(j)| : j \in \mathcal{N}_i(d)\backslash\{i\}\}$, d and m two constants taken as $d = 3, m = 4$ in [19]. If i is an impulse noisy point, then ROAD(i) is large; otherwise it is small. Therefore, the ROAD statistic serves to detect impulse noisy points.

Secondly, with the ROAD statistic, they defined the impulse factor $w_I(i)$ and the joint impulse factor $J_I(i, j)$:

$$w_I(i) = e^{-\frac{\text{ROAD}(i)^2}{2\sigma_I^2}}, \tag{19}$$

$$J_I(i, j) = e^{-\frac{1}{2\sigma_J^2}\left(\frac{\text{ROAD}(i)+\text{ROAD}(j)}{2}\right)^2}, \tag{20}$$

where σ_I and σ_J are control parameters [1]. If i is an impulse noisy point, then the value of $w_I(i)$ is close to 0; otherwise it is close to 1. Similarly, if either i or j is an impulse noisy point, then the value of $J_I(i, j)$ is close to 0; otherwise it is close to 1.

Finally, the restored image by the trilateral filter is

$$\text{TriF}(v)(i) = \frac{\sum_{j\in\mathcal{N}_i(D)} w(i, j)v(j)}{\sum_{j\in\mathcal{N}_i(D)} w(i, j)}, \tag{21}$$

where

$$w(i, j) = w_S(i, j)w_R(i, j)^{J_I(i,j)}w_I(j)^{1-J_I(i,j)},$$

---

[1] In fact, [19] defines the joint impulse factor as $J(i, j) = 1 - J_I(i, j)$. Following [24], we use $J_I(i, j)$ rather than $J(i, j)$, which seems more convenient.

with

$$w_S(i,j) = e^{-\frac{|i-j|^2}{2\sigma_S^2}}, \quad w_R(i,j) = e^{-\frac{(v(i)-v(j))^2}{2\sigma_R^2}}.$$

3.2 Patch-Based Weighted Means Filter

As in the non-local means filter [4], our filter estimates each point by the weighed means of its neighbors, and the weight for each neighbor is determined by the similarity of local patches centered at the estimated point and the neighbor. Due to the existence of impulse noise, some points are totally destroyed, so that noisy values are not related to original values at all. So we have to diminish the influence of impulse noisy points. Similarly to [21,24], we introduce the following weighted norm:

$$\|v(\mathcal{N}_i) - v(\mathcal{N}_j)\|_M^2 = \frac{\sum_{k \in \mathcal{N}_i^0} w_{S,M}(i,k) F\big(k, \mathcal{T}(k)\big) |v(k) - v\big(\mathcal{T}(k)\big)|^2}{\sum_{k \in \mathcal{N}_i^0} w_{S,M}(i,k) F\big(k, \mathcal{T}(k)\big)}, \qquad (22)$$

where

$$w_{S,M}(i,k) = e^{-\frac{|i-k|^2}{2\sigma_{S,M}^2}}, \quad F\big(k, \mathcal{T}(k)\big) = w_I(k) w_I\big(\mathcal{T}(k)\big). \qquad (23)$$

Recall that here $k = (k_1, k_2)$ represents a two-dimensional spatial location of a pixel, w_I is defined in (19), and $\mathcal{T}$ is the translation mapping i to j (and thus $\mathcal{N}_i(d)$ onto $\mathcal{N}_j(d)$). $F\big(k, \mathcal{T}(k)\big)$ is a joint impulse factor: if k or $\mathcal{T}(k)$ is an impulse noisy point, then $F\big(k, \mathcal{T}(k)\big)$ is close to 0, so that these points contribute little to the weighted norm; otherwise $F\big(k, \mathcal{T}(k)\big)$ is close to 1.

We now define our filter that we call *Patch-Based Weighted Means Filter* (PWMF). The restored image by PWMF is defined as

$$\text{PWMF}(v)(i) = \frac{\sum_{j \in \mathcal{N}_i(D)} w(i,j) v(j)}{\sum_{j \in \mathcal{N}_i(D)} w(i,j)}, \qquad (24)$$

where

$$w(i,j) = w_S(i,j) w_I(j) w_M(i,j),$$

with

$$w_S(i,j) = e^{-|i-j|^2/(2\sigma_S^2)}, \quad w_M(i,j) = e^{-\|v(\mathcal{N}_i)-v(\mathcal{N}_j)\|_M^2/(2\sigma_M^2)}, \qquad (25)$$

and $w_I(j)$ is defined in (19). In addition, we mention that, in this paper, we use the joint impulse factor $F\big(k, \mathcal{T}(k)\big) = w_I(k) w_I\big(\mathcal{T}(k)\big)$, which is different from the choice in [24] and [21], where $F\big(k, \mathcal{T}(k)\big) = J_I\big(k, \mathcal{T}(k)\big)$. In fact, we can see that with this new choice, we simplify the methods in [24] and [21] by eliminating a parameter and speeding up the implementation. Furthermore, we empirically find that the new choice leads to an improvement of the quality of restored images, especially for impulse noise.

3.3 Convergence Analysis

Our new filter can be regarded as an application of the mathematical justifications of the non-local means filter stated in Sect. 2 to the remained image obtained after filtering the impulse noisy points by the weighted norm (22), which can be considered to contain only Gaussian noise.

Let us explain this more precisely. For a fixed pixel i, let

$$P = \{j \in \mathcal{N}_i(D) : v(j) \text{ is an impulse noisy point}\} \tag{26}$$

be the set of impulse noisy points, and $P^c = \mathcal{N}_i(D) \backslash P$ be its complementary. We will show that

$$\text{PWMF}(v)(i) = \frac{\sum_{j \in \mathcal{N}_i(D)} w(i, j)v(j)}{\sum_{j \in \mathcal{N}_i(D)} w(i, j)} \approx \frac{\sum_{j \in P^c} w(i, j)v(j)}{\sum_{j \in P^c} w(i, j)}, \tag{27}$$

which demonstrates that the estimated value by this filter is close to the weighted average of gray value of the non-impulse noisy points. Since $w_I(j) \approx 0$ for impulse noisy point j, and $w_I(j) \approx 1$ for other points, we can think that PWMF is close to NL-means applied to images containing only Gaussian noisy points. By Theorem 1, the right-hand side of (27) is close to $u(i)$, which indicates the convergence $\text{PWMF}(v)(i)$ to $u(i)$.

To obtain (27), set

$$R_1 = \sum_{j \in P^c} w(i, j)v(j), \quad R_2 = \sum_{j \in P} w(i, j)v(j),$$

$$J_1 = \sum_{j \in P^c} w(i, j), \quad J_2 = \sum_{j \in P} w(i, j),$$

then it holds that

$$\text{PWMF}(v)(i) - \frac{\sum_{j \in P^c} w(i, j)v(j)}{\sum_{j \in P^c} w(i, j)} = \frac{R_1 + R_2}{J_1 + J_2} - \frac{R_1}{J_1}$$

$$= -\frac{R_1}{J_1} \frac{J_2}{J_1 + J_2} + \frac{R_2}{J_2} \frac{J_2}{J_1 + J_2}.$$

Since

$$0 \leq \frac{R_1}{J_1} \leq \max_j v(j) \leq 255 \quad \text{and} \quad 0 \leq \frac{R_2}{J_2} \leq \max_j v(j) \leq 255,$$

it follows that

$$\left| \text{PWMF}(v)(i) - \frac{R_1}{J_1} \right| \leq 255 \times \frac{J_2}{J_1 + J_2}. \tag{28}$$

We claim that J_2/J_1 is very small, and so is $J_2/(J_1 + J_2)$. In fact,

$$\frac{J_2}{J_1} = \frac{\sum_{j \in P} w(i, j)}{\sum_{j \in P^c} w(i, j)} = \frac{\sum_{j \in P} w_S(i, j)w_I(j)w_M(i, j)}{\sum_{j \in P^c} w_S(i, j)w_I(j)w_M(i, j)}.$$

For $j \in P$, we can approximately replace the ROAD statistic by its mean value $\bar{R}(P)$ over P; for $j \in P^c$, we do the same by the mean value $\bar{R}(P^c)$ over P^c. Therefore writing

$$W_I(P) = e^{-(\bar{R}(P))^2/(2\sigma_I^2)} \quad \text{and} \quad W_I(P^c) = e^{-(\bar{R}(P^c))^2/(2\sigma_I^2)},$$

we get

$$\frac{J_2}{J_1} \approx \frac{W_I(P)}{W_I(P^c)} \frac{\sum_{j \in P} w_S(i, j)w_M(i, j)}{\sum_{j \in P^c} w_S(i, j)w_M(i, j)}$$

$$\approx \frac{W_I(P)}{W_I(P^c)} \frac{p}{1 - p},$$

where the last step holds by Theorem 1 of [31]. Therefore, it follows that

$$\frac{J_2}{J_1 + J_2} \approx \frac{p W_I(P)}{p W_I(P) + (1 - p) W_I(P^c)}. \tag{29}$$

The value of right-hand side of (29) is generally small. For example, for impulse noise $p = 0.2$, by results in [19], $\bar{R}(P)$ is about 242, and $\bar{R}(P^c)$ is about 17. Therefore, with the parameter $\sigma_I = 50$ used in the experiments,

$$\frac{J_2}{J_1 + J_2} \approx \frac{p e^{-242^2/(2\sigma_I^2)}}{p e^{-242^2/(2\sigma_I^2)} + (1 - p) e^{-17^2/(2\sigma_I^2)}} = 2.2 \times 10^{-6}.$$

Thus, by (28), the value PWMF$(v)(i)$ is very close to $\frac{R_1}{J_1}$: we have approximately

$$| \text{PWMF}(v)(i) - \frac{R_1}{J_1} | \le 5.6 \times 10^{-4}. \tag{30}$$

4 Choices of Parameters and Comparison with Other Filters by Simulations

4.1 Choices of Parameters

Notice that PWMF reduces to NL-means when $\sigma_I = \sigma_S = \infty$. So for removing Gaussian noise, a reasonable choice is to take σ_I and σ_S large enough. Now present the choices of parameters for removing impulse noise and mixed noise, which are determined empirically and important for our filter. The noise level σ and p, are supposed to be known, otherwise, there are methods to estimate them in the literature, for example [22,28,40].

In the calculation of ROAD [cf.(18)], we choose 3×3 neighborhoods and $m = 4$. For impulse noise or mixed noise with $p = 0.4, 0.5$, to further improve the results, 5×5 neighborhoods and $m = 12$ are used to calculate ROAD.

The patch size $d = 9$ is used in all cases; the search window sizes D are shown in Table 1.

Now come to the choice of $\sigma_I, \sigma_M, \sigma_S, \sigma_{S,M}$ appearing in (19), (25) and (23). To apply our filter easily in practice, simple and uniform formulas in terms of p and σ are searched empirically. Assume that there is some simple relation between the desired parameters and the given parameters (the values of σ and p). We try some simple relations of the form $a\sigma + bp + c$, then determine a, b, c by experiments in different cases, and then verify their validity again by many experiments.

- To remove impulse noise, use $\sigma_M = 3 + 20p, \sigma_S = 0.6 + p$, and omit the factor $w_{S,M}$ (i.e. $\sigma_{S,M}$ can be taken a value large enough); $\sigma_I = 50$ for $p = 0.2, 0.3$, and $\sigma_I = 160$ for $p = 0.4, 0.5$.
- For mixed noise, choose $\sigma_I = 50 + 5\sigma/3, \sigma_M = 3 + 0.4\sigma + 20p, \sigma_{S,M} = 2$, and omit the factor w_S, while $\sigma = 10, 20, 30$ and $p = 0.2, 0.3$.
- For other values of σ or p, choose parameters by linear interpolation, linear extension or according to the adjacent values of σ or p.

It is not easy to find appropriate parameters for a filter. Different choices of parameters can have great influence to the restored images. See also [13]. Some similar research for NL-means can be found in [17,35,39].

Fig. 1 Original 512×512 images of Lena, Bridge, Peppers512, Boats. Since in the original Peppers images, there are black boundaries of width of one pixel in the left and top which can be considered as impulse noise, to make an impartial comparison, we compute PSNR for Peppers images after removing all the four boundaries, that is with images of size 510×510 for Peppers512 and 254×254 for Peppers256

Note that our choice of parameters is different from [24]: for the patch size, we use $d = 9$, while [24] uses $d = 3$ in most cases; for impulse noise with $p = 0.4, 0.5$, we use 5×5 neighborhoods for ROAD, while [24] always uses 3×3 neighborhoods.

4.2 Experiments and Comparisons

We use standard gray images to test the performance of our filter[2]. Original images are shown in Fig. 1.[3] As usual we use PSNR (Peak Signal-to-Noise Ratio)

$$\text{PSNR}(\bar{v}) = 10 \log_{10} \frac{255^2 |I|}{\sum_{i \in I} (\bar{v}(i) - u(i))^2} \text{dB}$$

[2] The code of our method and the images can be downloaded at
https://www.dropbox.com/s/oylg9to8n6029hh/to_j_sci_comput_paper_code.zip.

[3] The images Lena, Peppers256 and Boats are originally downloaded from
http://decsai.ugr.es/~javier/denoise/test_images/index.htm; the image Peppers512 is from http://perso.telecom-paristech.fr/~delon/Demos/Impulse and the image Bridge is from www.math.cuhk.edu.hk/~rchan/paper/dcx/.

to measure the quality of a restored image, where u is the original image, and $\bar{v}$ the restored one. For the simulations, the gray value of impulse noise is uniformly distributed on the interval [0,255]. We add Gaussian noise and then add impulse noise for the simulation of mixed noise. The same realizations of noisy images for comparisons of different methods are used when codes are available, that is, for TriF [19], ROLD-EPR [14], NLMixF [21] and MNF [24]. For other methods, the reported results in corresponding papers are listed.

The results for TriF are obtained by the program made by ourselves. To compare the performance of our filter with those of TriF fairly, we make our effort to obtain the best results as we can according to the suggestion of [19].

- Use $\sigma_I = 40$, $\sigma_J = 50$, $\sigma_S = 0.5$, and $\sigma_R = 2\sigma_{QGN}$, where σ_{QGN} is an estimator for the standard deviation of "quasi-Gaussian" noise defined in [19].
- For impulse noise, apply one iteration for $p = 0.2$, two iterations for $p = 0.3, 0.4$, and four iterations for $p = 0.5$. For mixed noise, apply TriF twice with different values of σ_S as suggested in [19]: with all impulse noise levels p, for $\sigma = 10$, first use $\sigma_S = 0.3$, then $\sigma_S = 1$; for $\sigma = 20$, first $\sigma_S = 0.3$, then $\sigma_S = 15$; for $\sigma = 30$, first $\sigma_S = 15$, then $\sigma_S = 15$.

For ROLD-EPR, the listed values are the best PSNR values along iterations with the code from the authors of [14].

Table 2 demonstrates the performances of PWMF for removing impulse noise by comparing with TriF [19], ROLD-EPR [14], PARIGI [13], and NLMixF [21]. For ease of comparison, in this and following tables, the best results and the results where the differences from the best ones are less than 0.1 dB are displayed in bold. We can observe that our filter PWMF attains the best performance in term of PSNR. Some visual comparisons are shown in Figs. 2 and 3. Carefully comparing these images, we observe that TriF loses some small textured details, while ROLD-EPR is not smooth enough. PARIGI and PWMF show better results.

Different papers consider different mixtures of Gaussian noise and impulse noise. We demonstrate the performance of PWMF for removing mixed noise in Tables 3, 4, 5, and 6 by comparing it with TriF [19], NLMixF [21], PARIGI [13] IPAMF+BM [40], Xiao [37], MNF [24], and Zhou [43]. All these comparisons illustrate good performance of our filter except for Barbara when comparing with PARIGI. Our method does not work very well as PARIGI for Barbara, because the ROAD statistics is a very local statistics and can not use the redundancy of this image very well to detect impulse noisy pixels, while PARIGI is particularly powerful for the restoration of textured regions. From Fig. 4, it can be seen that the results of our filter are visually better than TriF. From Fig. 5, we can see that when the standard deviation σ is high, our filter is smoother than PARIGI, while PARIGI seems to preserve more weak textured details, but it has evident artifacts throughout the whole image (see the electronic version of this paper at full resolution).

Finally, we compare the CPU time of TriF [19], NLMixF [21], and our method PWMF for removing mixed nose in seconds in the platform of MATLAB R2011a with unoptimized mex files. The computer is equipped with 2.13GHZ Intel (R) Core (TM) i3 CPU and 3.0 GB memory. The results are presented in Table 7, which demonstrate that PWMF is rather fast: much faster than NLMixF and even faster than TriF when the noise level is low, thanks to the simplified joint impulse factor $F\left(k, \mathcal{T}(k)\right)$ defined in (23). The results also show that our method is faster than Zhou [43].

Table 2 PSNR values (dB) to remove impulse noise for TriF [19], ROLD-EPR [14], PARIGI [13], NLMixF [21] and our filter PWMF

	$p = 0.2$	$p = 0.3$	$p = 0.4$	$p = 0.5$
Lena				
TriF	34.75	32.54	31.28	29.37
ROLD-EPR	34.87	32.08	30.81	29.51
PARIGI	35.45	–	31.75	–
NLMixF	35.69	33.13	31.69	29.87
PWMF	**35.90**	**33.45**	**31.98**	**30.17**
Bridge				
TriF	26.81	25.25	24.41	23.35
ROLD-EPR	27.60	25.58	24.42	23.45
PARIGI	27.68	–	**24.80**	–
NLMixF	27.77	25.54	24.45	23.33
PWMF	**28.10**	**26.11**	24.74	**23.64**
Peppers256				
TriF	30.55	28.81	27.67	25.98
ROLD-EPR	31.03	28.10	27.34	25.96
PARIGI	–	–	–	–
NLMixF	31.78	29.38	28.10	26.47
PWMF	**31.93**	**29.64**	**28.34**	**26.78**
Peppers512				
TriF	34.52	31.93	31.27	29.76
ROLD-EPR	34.46	32.31	31.12	30.03
PARIGI	34.75	–	31.63	–
NLMixF	34.77	**32.56**	31.73	30.23
PWMF	**35.08**	**32.59**	**31.95**	**30.38**
Boats				
TriF	30.22	28.55	27.52	26.10
ROLD-EPR	30.75	28.19	26.95	25.91
PARIGI	31.21	–	27.56	–
NLMixF	31.32	29.01	27.42	26.10
PWMF	**31.83**	**29.58**	**27.67**	**26.48**

5 Degree of Similarity and Estimated PSNR Values

In this section, we introduce the notion of degree of similarity (DS) of images corrupted by Gaussian noise and mixed noise. With this notion and Theorem 2, estimations of the PSNR values of denoised images are obtained. For simplicity, in this section, impulse noise is considered as particular mixed noise with $\sigma = 0$.

For Gaussian noise model, if two patches $v(\mathcal{N}_i)$ and $v(\mathcal{N}_j)$ have the same distribution and are independent of each other, i.e. $u(\mathcal{N}_i) = u(\mathcal{N}_j)$, then $\{v(k) - v(\mathcal{T}(k)) : k \in \mathcal{N}_i\}$ are independent variables with the same law $N(0, 2\sigma^2)$ (recall that $\mathcal{T}$ is the translation mapping of $\mathcal{N}_i$ onto $\mathcal{N}_j$). Therefore

$$\|v(\mathcal{N}_i) - v(\mathcal{N}_j)\|^2 = \sum_{k \in \mathcal{N}_i} |v(k) - v(\mathcal{T}(k))|^2 = 2\sigma^2 X,$$
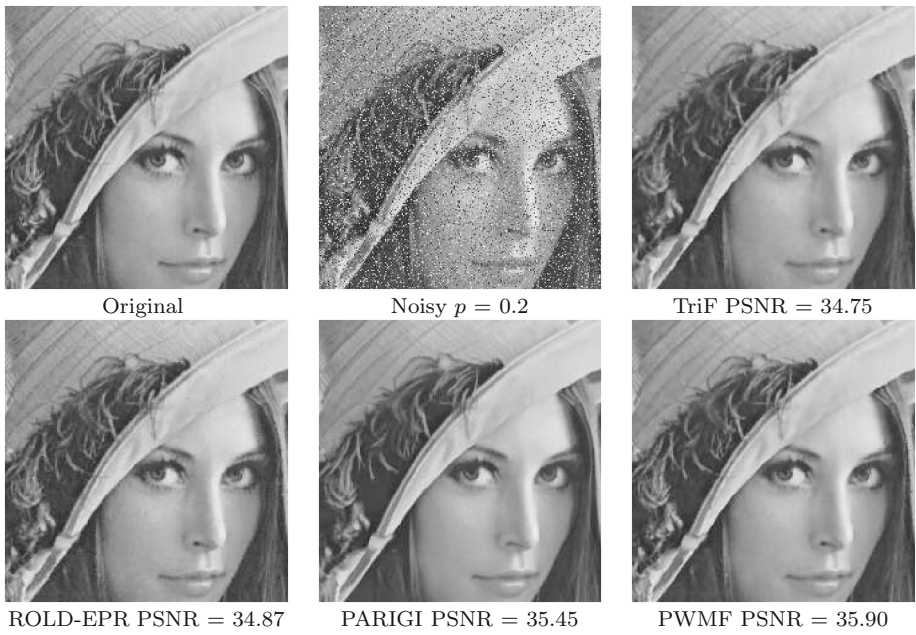
Original	Noisy $p = 0.2$	TriF PSNR $= 34.75$
ROLD-EPR PSNR $= 34.87$	PARIGI PSNR $= 35.45$	PWMF PSNR $= 35.90$

Fig. 2 Comparison of the performances of TriF [19], ROLD-EPR [14], PARIGI [13] and our filter PWMF for removing impulse noise with $p = 0.2$ for Lena

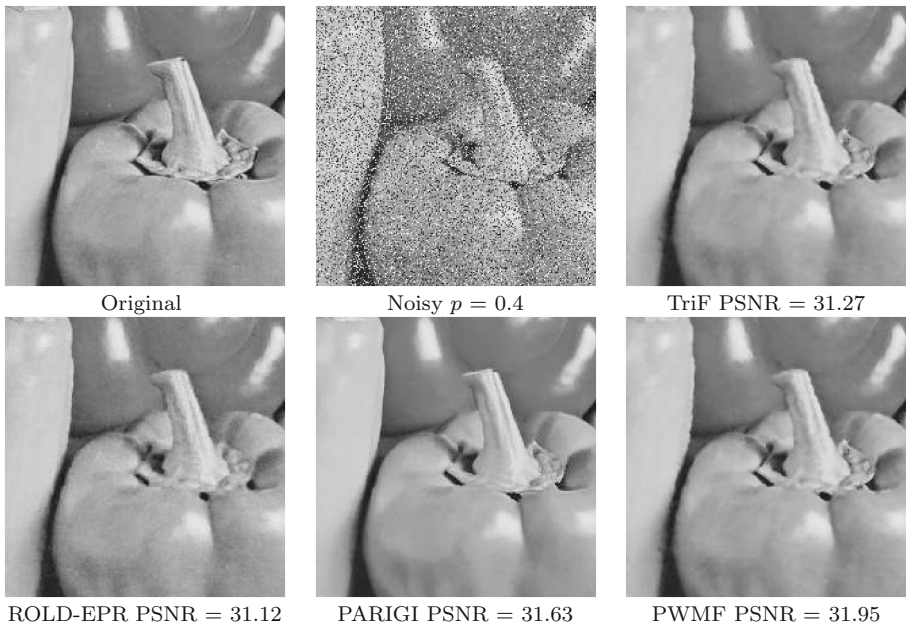|		
Original	Noisy $p = 0.4$	TriF PSNR $= 31.27$
ROLD-EPR PSNR $= 31.12$	PARIGI PSNR $= 31.63$	PWMF PSNR $= 31.95$

Fig. 3 Comparison of the performances of TriF [19], ROLD-EPR [14], PARIGI [13] and our filter PWMF for removing impulse noise with $p = 0.4$ for Peppers512

Table 3 PSNR values (dB) to remove mixed noise for TriF [19], NLMixF [21] and our filter PWMF

p	$\sigma = 10$			$\sigma = 20$			$\sigma = 30$		
	TriF	NLMixF	PWMF	TriF	NLMixF	PWMF	TriF	NLMixF	PWMF
Lena									
0.2	31.70	**32.85**	**32.93**	28.75	**30.51**	**30.47**	26.54	**28.70**	**28.67**
0.3	30.77	**31.28**	**31.30**	28.01	**29.41**	**29.38**	25.82	**27.68**	**27.65**
Bridge									
0.2	25.28	26.14	**26.35**	23.84	**24.48**	**24.53**	22.50	**23.33**	**23.35**
0.3	24.66	24.73	**25.00**	23.34	23.59	**23.70**	22.08	**22.66**	**22.72**
Peppers256									
0.2	29.09	**30.57**	**30.65**	26.82	**28.56**	**28.53**	24.70	**26.73**	**26.71**
0.3	27.89	**28.68**	**28.73**	26.01	**27.16**	**27.13**	23.97	**25.55**	**25.57**
Peppers512									
0.2	31.84	**32.64**	**32.70**	29.01	**30.74**	**30.69**	26.82	**29.04**	**29.07**
0.3	30.92	**31.06**	30.93	28.27	**29.55**	29.39	26.01	**27.92**	**27.86**
Boats									
0.2	28.37	29.75	**29.91**	26.40	**27.71**	**27.74**	24.58	**26.23**	**26.23**
0.3	27.63	28.15	**28.38**	25.84	**26.66**	**26.75**	24.09	**25.44**	**25.48**

For each PSNR value, the corresponding level of Gaussian noise and impulse noise are shown in the top and left of the table

Table 4 PSNR values (dB) to remove mixed noise for PARIGI [13] and our filter PWMF

	Lena	Barbara	Cameraman	Boat
$p = 0.1\, \sigma = 5$				
PARIGI	34.72	**31.55**	34.98	31.41
PWMF	**35.80**	30.86	**35.80**	**32.60**
$p = 0.3\, \sigma = 15$				
PARIGI	29.22	**27.33**	28.59	26.57
PWMF	**30.25**	25.58	**30.27**	**27.45**

Table 5 PSNR values (dB) for mixed noise removal with (Xiao) [37], (IPAMF+BM) [40], (Zhou) [43] and our filter PWMF

Lena $\sigma = 10$	$p = 0.1$	$p = 0.2$	$p = 0.3$
Xiao	32.75	31.66	30.42
IPAMF+BM	33.61	32.12	30.69
Zhou	**34.25**	32.68	**31.21**
PWMF	34.10	**32.93**	**31.30**

where X is the sum of squares of independent random variables with normal law $N(0, 1)$, and has the law χ^2 of density function f_κ with $\kappa = d^2$, where

$$f_\kappa(t) = \begin{cases} \dfrac{x^{\kappa/2-1}e^{-t/2}}{2^{\kappa/2}\Gamma(\kappa/2)} & t \geq 0, \\ 0 & \text{else.} \end{cases}$$

Table 6 PSNR values (dB) for mixed noise removal with MNF [24] and our filter PWMF

Lena	$\sigma = 10, p = 0.2$	$\sigma = 20, p = 0.2$	$\sigma = 30, p = 0.2$
MNF	31.63	29.33	28.40
PWMF	**32.93**	**30.47**	**28.67**

Noisy $\sigma = 30, p = 0.2$ TriF PSNR $= 26.54$ PWMF PSNR $= 28.67$

Noisy $\sigma = 20, p = 0.2$ TriF PSNR $= 26.40$ PWMF PSNR $= 27.74$
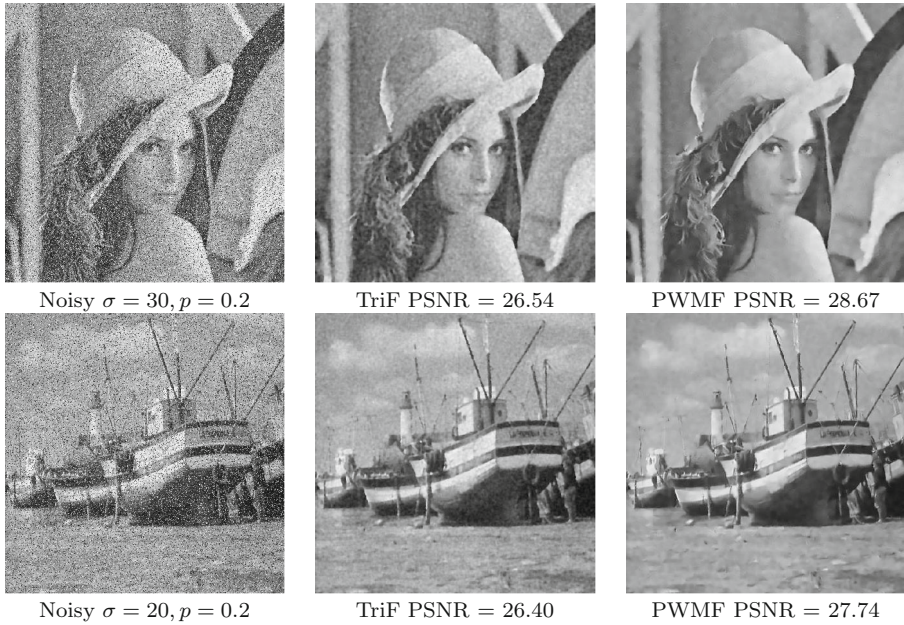
Fig. 4 Comparison of the performances of TriF [19] and our filter PWMF for removing mixed noise

Let $\alpha \in (0, 1)$ represent the risk probability, chosen to be small enough. And let $T_\alpha > 0$ be determined by

$$\int_0^{T_\alpha} f_\kappa(t)dt = 1 - \alpha, \quad \text{with } \kappa = d^2,$$

so that

$$\mathbb{P}(\|v(\mathcal{N}_i) - v(\mathcal{N}_j)\| \leq T_\alpha) = 1 - \alpha.$$

When $\|v(\mathcal{N}_i) - v(\mathcal{N}_j)\| \leq T_\alpha$, we consider that $v(\mathcal{N}_i)$ and $v(\mathcal{N}_j)$ are similar with confidence level $1 - \alpha$. This leads us to the following definition of the degree of similarity.

Definition 3 *(Degree of similarity of images corrupted by Gaussian noise)* Let $\alpha \in (0, 1)$ and let $T_\alpha > 0$ be defined as above. For $i \in I$, let

$$DS_i = \frac{\#\{j \in \mathcal{N}_i(D) : \|v(\mathcal{N}_i) - v(\mathcal{N}_j)\| \leq T_\alpha\}}{D^2}$$

be the proportion of patches $v(\mathcal{N}_j)$ similar to $v(\mathcal{N}_i)$ in the search window $\mathcal{N}_i(D)$, and let
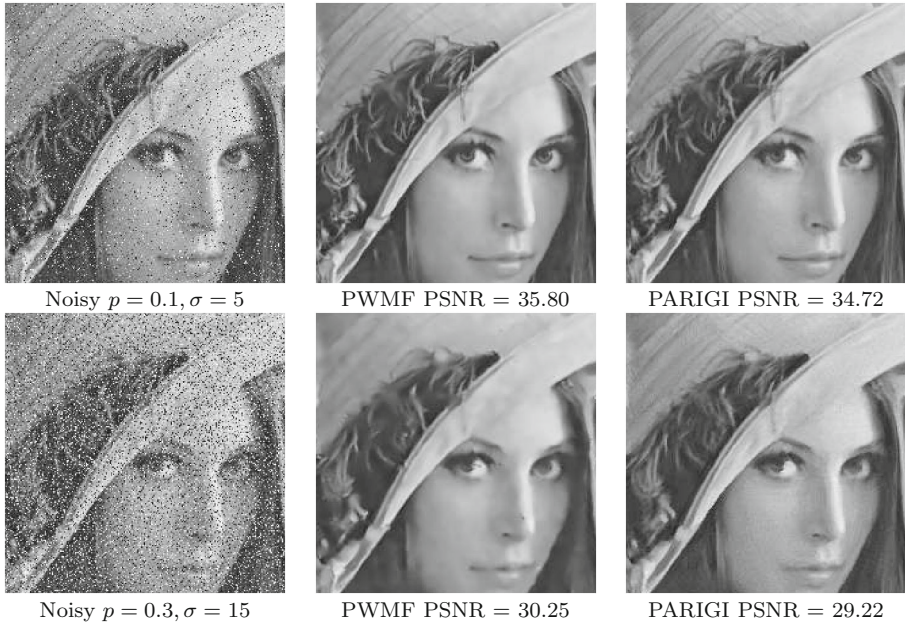
$$DS = \frac{\sum_{i \in I} DS_i}{MN}$$

Noisy $p = 0.1, \sigma = 5$ PWMF PSNR $= 35.80$ PARIGI PSNR $= 34.72$

Noisy $p = 0.3, \sigma = 15$ PWMF PSNR $= 30.25$ PARIGI PSNR $= 29.22$

Fig. 5 Comparison of the performances of our filter PWMF and PARIGI [13] for removing mixed noise with Lena

Table 7 Time(s) for TriF [19], NLMixF [21], and PWMF

Image	Noise levels	TriF	NLMixF	PWMF
Lena	$\sigma = 10, p = 0.2$	6.9	60.6	5.4
Lena	$\sigma = 20, p = 0.3$	7.9	170.0	16.7

be their mean over the whole image. We call *DS* the degree of similarity of the image v with confidence level $1 - \alpha$.

For the mixed noise model, if the original patches satisfy $u(\mathcal{N}_i) = u(\mathcal{N}_j)$, then, letting

$$\|v(\mathcal{N}_i) - v(\mathcal{N}_j)\|_m^2 = \sum_{k \in \mathcal{N}_i} F\big(k, \mathcal{T}(k)\big) |v(k) - v\big(\mathcal{T}(k)\big)|^2, \tag{31}$$

we have

$$\frac{1}{2\sigma^2} \|v(\mathcal{N}_i) - v(\mathcal{N}_j)\|_m^2 \approx \frac{1}{2\sigma^2} \sum_{k \in P^c, \mathcal{T}(k) \in P^c} |v(k) - v\big(\mathcal{T}(k)\big)|^2,$$

where P^c represents the set of non-impulse noisy points contained in $\mathcal{N}_i$, and the right-hand side has the law χ^2 with density function $f_\kappa(t)$ with $\kappa \approx (1-p)^2 d^2$.

Similar to Definition 3, let $T'_\alpha > 0$ be determined by

$$\int_0^{T'_\alpha} f_\kappa(t) dt = 1 - \alpha, \quad \text{with } \kappa = (1-p)^2 d^2.$$

Definition 4 *(Degree of similarity of images corrupted by mixed noise)* Let $\alpha \in (0, 1)$ and let $T'_\alpha > 0$ be defined as above. For $i \in I$, let

$$DS_i = \frac{(1-p)|\{j \in \mathcal{N}_i(D) : \|v(\mathcal{N}_i) - v(\mathcal{N}_j)\|_m \le T'_\alpha\}|}{D^2}, \tag{32}$$

and

$$DS = \frac{\sum_{i \in I} DS_i}{MN}.$$

We call DS the degree of similarity of the image v with confidence level $1 - \alpha$.

Note that, on average, the proportion of significant pixels in a search window $\mathcal{N}_i(D)$ is $(1 - p)$, which explains the factor $(1 - p)$ in the right-hand side of (32).

For each pixel $i \in I$, denote by $\bar{v}(i)$ the gray value of the restored image at pixel i by NL-means for Gaussian noise or by PWMF for mixed noise. Then by Theorem 2, it holds that

$$\mathbb{P}(|\bar{v}(i) - u(i)| < A_\beta/\sqrt{n}) \approx \int_{-A_\beta}^{A_\beta} \tilde{f}(t)dt,$$

where n represents the number of similar patches in the search window $\mathcal{N}_i(D)$.

Definition 5 *(Estimated PSNR value)* For $\beta \in (0, 1)$, let $A_\beta > 0$ be such that

$$\int_{-A_\beta}^{A_\beta} \tilde{f}(t)dt = 1 - \beta, \tag{33}$$

where $\tilde{f}(t)$ is defined by (15). Define

$$\mathrm{PSNR}_e = 20 \log_{10}(255\sqrt{n}/A_\beta), \tag{34}$$

where $n = D^2 \cdot DS$ is the estimated number of similar patches by means of the degree of similarity DS. We call PSNR_e the estimated PSNR value with confidence level $1 - \beta$.

The value of β is chosen to be small enough, representing the risk probability. For mixed noise, we can replace σ_r by σ_M for the calculation of $c(x)$ in $\tilde{f}(t)$. For given β, we can use the Monte-Carlo simulation to obtain the approximate value of A_β (see Remark 1 in Sect. 2.2).

We compute the DS values and estimated PSNR values for different images corrupted by Gaussian noise with $\sigma = 10, 20, 30$, and mixed noise with $\sigma = 20$, $p = 0.2$ or 0.3 shown in Tables 8 and 9. Note that the DS value depends on the choice of α, and the estimated PSNR value depends on the choice of β and α. To have a good approximation to the true PSNR

	σ		Lena	Peppers512	Peppers256	Boats	Bridge
Table 8 The degree of similarity(DS), the estimated PSNR values (PSNR_e) and true PSNR values (PSNR) with NL-means for images corrupted by Gaussian noise		DS	0.53	0.49	0.37	0.33	0.11
	10	PSNR_e	34.99	34.81	33.56	33.05	28.00
		PSNR	34.92	34.51	33.76	32.96	30.51
		DS	0.39	0.38	0.27	0.27	0.09
	20	PSNR_e	31.32	31.13	29.65	29.65	25.01
		PSNR	31.59	31.68	30.08	29.45	26.50
		DS	0.30	0.29	0.21	0.22	0.09
	30	PSNR_e	29.07	28.94	27.53	27.80	23.82
		PSNR	29.43	29.76	27.75	27.41	24.56

Table 9 The degree of similarity(DS), the estimated PSNR values (PSNR$_e$) and true PSNR values (PSNR) with PWMF for images corrupted by mixed noise

$\sigma = 20$		Peppers512	Lena	Peppers256	Boats	Bridge
	DS	0.45	0.43	0.33	0.35	0.19
$p = 0.2$	PSNR$_e$	30.17	30.00	28.95	29.16	26.35
	PSNR	30.69	30.47	28.53	27.74	24.53
	DS	0.25	0.24	0.18	0.19	0.11
$p = 0.3$	PSNR$_e$	28.20	27.90	26.65	27.04	24.56
	PSNR	29.39	29.38	27.13	26.75	23.70

value, we take $\beta = 0.03$; $\alpha = 0.03, 0.3, 0.5$ for Gaussian noise $\sigma = 10, 20, 30$ respectively, and $\alpha = 0.04$ for mixed noise. It can be seen that, generally, the larger the DS value, the larger the estimated PSNR value and the true PSNR value, and the estimated PSNR value is close to the true PSNR value.

6 Conclusions

Two convergence theorems, one for the almost sure convergence and the other for the convergence in law, are established to show the rate of convergence of NL-means [4]. Based on the convergence theorems, a new filter called patch-based weighted means filter (PWMF) is proposed to remove mixed noise, leading to an extension of NL-means. The choice of parameters has been carefully discussed. Simulation results show that the new proposed filter is competitive compared to recently developed known algorithms. The notion of degree of similarity is introduced to describe the influence of the proportion of similar patches in the application of NL-means or PWMF. With the notion of degree of similarity and the convergence theorem in law, we obtain a good estimation for PSNR values of denoised images by NL-means or PWMF.

Appendix

In this appendix we will prove Theorems 1 and 2.

Convergence Theorems for Random Weighted Means

We first show a Marcinkiewicz law of large numbers (Theorem 3) and a convergence theorem in distribution for random weighted means (Theorem 4) for l-dependent random variables, which we will use to prove Theorems 1 and 2.

Theorem 3 *Let* $\{(a_k, v_k)\}$ *be a sequence of l-dependent identically distributed random variables, with* $\mathbb{E}|a_1|^p < \infty$ *and* $\mathbb{E}|a_1 v_1|^p < \infty$ *for some* $p \in [1, 2)$, *and* $\mathbb{E}a_1 \neq 0$. *Then*

$$\frac{\sum_{k=1}^{n} a_k v_k}{\sum_{k=1}^{n} a_k} - \frac{\mathbb{E} a_1 v_1}{\mathbb{E} a_1} = o(n^{-(1-1/p)}) \quad \text{almost surely.}$$

We need the following lemma to prove it.

Lemma 2 [24] *If $\{X_n\}$ are l-dependent and identically distributed random variables with $\mathbb{E} X_1 = 0$ and $\mathbb{E} |X_1|^p < \infty$ for some $p \in [1, 2)$, then*

$$\lim_{n \to \infty} \frac{X_1 + \cdots + X_n}{n^{1/p}} = 0 \quad \text{almost surely.}$$

This lemma is a direct consequence of Marcinkiewicz law of large numbers for independent random variables (see e.g. [9], p. 118), since for all $k \in \{1, \ldots, l+1\}$, $\{X_{i(l+1)+k} : i \geq 0\}$ is a sequence of i.i.d. random variables, and for each positive integer n, we have

$$X_1 + \cdots + X_n = \sum_{k=1}^{l+1} \sum_{i=0}^{m-1} X_{i(l+1)+k} + \sum_{1 \leq k \leq k_0} X_{m(l+1)+k},$$

where m, k_0 are positive integers determined by $n = m(l+1) + k_0, 0 \leq k_0 \leq l$.

Proof of Theorem 3 Notice that

$$n \left(\frac{\sum_{k=1}^{n} a_k v_k}{\sum_{k=1}^{n} a_k} - \frac{\mathbb{E} a_1 v_1}{\mathbb{E} a_1} \right) = \frac{n}{(\sum_{k=1}^{n} a_k)} \frac{1}{\mathbb{E} a_1} \sum_{k=1}^{n} a_k z_k,$$

where

$$z_k = v_k \mathbb{E} a_1 - \mathbb{E} a_1 v_1.$$

Since $\mathbb{E}|a_1| \leq (\mathbb{E}|a_1|^p)^{1/p} < \infty$, by Lemma 2 with $p = 1$, we have

$$\lim_{n \to \infty} \frac{\sum_{k=1}^{n} a_k}{n} = \mathbb{E} a_1 \quad \text{almost surely.}$$

Since $\mathbb{E} a_1 z_1 = 0$, and $\mathbb{E}|a_1 z_1|^p < \infty$, again by Lemma 2, we get

$$\lim_{n \to \infty} \frac{\sum_{k=1}^{n} a_k z_k}{n^{1/p}} = 0 \quad \text{almost surely.}$$

Thus the conclusion follows. □

Theorem 4 *Let $\{(a_k, v_k)\}$ be a stationary sequence of l-dependent and identically distributed random variables with $\mathbb{E} a_1 \neq 0$, $\mathbb{E} a_1^2 < \infty$, and $\mathbb{E}(a_1 v_1)^2 < \infty$. Then*

$$\sqrt{n} \left(\frac{\sum_{k=1}^{n} a_k v_k}{\sum_{k=1}^{n} a_k} - \frac{\mathbb{E} a_1 v_1}{\mathbb{E} a_1} \right) \xrightarrow{d} N(0, c^2),$$

that is,

$$\lim_{n \to \infty} \mathbb{P} \left\{ \frac{\sqrt{n}}{c} \left(\frac{\sum_{k=1}^{n} a_k v_k}{\sum_{k=1}^{n} a_k} - \frac{\mathbb{E} a_1 v_1}{\mathbb{E} a_1} \right) \leq z \right\} = \Phi(z), \quad z \in \mathbb{R},$$

where $\Phi(z)$ is the cumulative distribution function of the standard normal distribution,

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt,$$

and

$$c = \frac{1}{(\mathbb{E}a_1)^2}\sqrt{\mathbb{E}(a_1 z_1)^2 + 2\lambda} \tag{35}$$

with $\lambda = \sum_{k=1}^{l} \mathbb{E}a_1 z_1 a_{1+k} z_{1+k}$, $z_k = v_k \mathbb{E}a_1 - \mathbb{E}a_1 v_1$.

We need the following lemma to prove the theorem.

Lemma 3 [36] *Let* $\{X_n\}$ *be a stationary sequence of* l*-dependent and identically distributed random variables with* $\mathbb{E}X_1 = 0$ *and* $\mathbb{E}X_1^2 < \infty$. *Set* $S_n = X_1 + \cdots + X_n (n \geq 1)$,

$$c_1 = \mathbb{E}X_1^2 + 2\sum_{k=1}^{l} \mathbb{E}X_1 X_{1+k}, \quad \text{and} \quad c_2 = 2\sum_{k=1}^{l} k\mathbb{E}X_1 X_{1+k}.$$

Then $var(S_n) = c_1 n - c_2$ *for* $n > l$, *and as* $n \to \infty$,

$$\frac{S_n}{\sqrt{c_1 n}} \xrightarrow{d} N(0, 1).$$

Proof of Theorem 4. As in the proof of Theorem 3, we have

$$\sqrt{n}\left(\frac{\sum_{k=1}^{n} a_k v_k}{\sum_{k=1}^{n} a_k} - \frac{\mathbb{E}a_1 v_1}{\mathbb{E}a_1}\right) = \frac{n}{(\sum_{k=1}^{n} a_k)} \frac{1}{\mathbb{E}a_1} \frac{\sum_{k=1}^{n} a_k z_k}{\sqrt{n}},$$

where

$$z_k = v_k \mathbb{E}a_1 - \mathbb{E}a_1 v_1.$$

Notice that the l-dependence of $\{(a_k, v_k)\}$ and the stationarity imply those of $\{(a_k, z_k)\}$. Therefore by Lemma 3, we get

$$\frac{\sum_{k=1}^{n} a_k z_k}{\sqrt{n}c_0} \to N(0, 1),$$

where

$$c_0 = \sqrt{\mathbb{E}(a_1 z_1)^2 + 2\lambda}, \quad \text{with } \lambda = \sum_{k=1}^{l} \mathbb{E}a_1 z_1 a_{1+k} z_{1+k}.$$

Since $\mathbb{E}|a_1| \leq (\mathbb{E}|a_1|^p)^{1/p}$, by Lemma 2 with $p = 1$, we obtain

$$\lim_{n \to \infty} \frac{\sum_{k=1}^{n} a_k}{n} = \mathbb{E}a_1 \quad \text{almost surely.}$$

Thus the conclusion follows with $c = c_0/(\mathbb{E}a_1)^2$. $\qquad\square$

Proofs of Theorems 1 and 2

We now come to the proofs of Theorems 1 and 2, using Theorems 3 and 4.

For Theorem 1, we need to prove that for any $\epsilon \in (0, \frac{1}{2}]$, as $n \to \infty$,

$$\frac{\sum_{k=1}^{n} w^0(i, j_k) v(j_k)}{\sum_{k=1}^{n} w^0(i, j_k)} - u(i) = o(n^{-(\frac{1}{2}-\epsilon)}) \quad \text{almost surely,}$$

where

$$w^0(i, j_k) = e^{-\|v(\mathcal{N}_i^0) - v(\mathcal{N}_{j_k}^0)\|^2/(2\sigma_r^2)}.$$

We will apply Theorem 3 to prove this. Note that the sequence $\{w^0(i, j_k), v(j_k)\}$ $(k = 1, 2, \ldots, n)$ is usually not l-dependent, since the central random variable $v(\mathcal{N}_i^0)$ is contained in all the terms. To make use of Theorem 3, we first take a fixed vector to replace the central random variable.

Proof of Theorem 1 Fix $x \in \mathbb{R}^{|\mathcal{N}_i^0|}$. Let

$$a_k = w^0(x, j_k) = e^{-\|x - v(\mathcal{N}_{j_k}^0)\|^2/(2\sigma_r^2)}.$$

Then a_k and $v(j_k)$ are independent since $j_k \notin \mathcal{N}_{j_k}^0$, so that

$$\frac{\mathbb{E}a_k v(j_k)}{\mathbb{E}a_k} = \mathbb{E}v(j_k) = u(j_k) = u(i).$$

By Lemma 1, the sequence $\{v(\mathcal{N}_{j_k})\}$ is l-dependent for $l = (2d - 1)^2 - 1$; thus the sequence $\{(a_k, v(j_k))\}$ is also l-dependent. Since $v = u + \eta$, with the range of u being bounded and η being Gaussian, we have $\mathbb{E}|v(j_k)|^p < \infty$ for $p \in [1, 2)$. (In fact, it holds for all $p \geq 1$.) Hence $\mathbb{E}|a_k v(j_k)|^p < \infty$, as $a_k \leq 1$.

Applying Theorem 3, we have, for any fixed $x = v(\mathcal{N}_i^0) \in \mathbb{R}^{|\mathcal{N}_i^0|}$ and any positive integer k_0,

$$\frac{\sum_{k=k_0}^n w^0(x, j_k)v(j_k)}{\sum_{k=k_0}^n w^0(x, j_k)} - u(i) = o(n^{-(1-1/p)}) \quad \text{almost surely.} \tag{36}$$

Let $k_0 > l$, so that $v(\mathcal{N}_i^0)$ is independent of $v(\mathcal{N}_{j_k}^0)$ for all $k \geq k_0$. By Fubini's theorem, we can replace $w^0(x, j_k)$ in (36) by

$$w^0(i, j_k) = e^{-\|v(\mathcal{N}_i^0) - v(\mathcal{N}_{j_k}^0)\|^2/(2\sigma_r^2)}.$$

That is,

$$\frac{\sum_{k=k_0}^n w^0(i, j_k)v(j_k)}{\sum_{k=k_0}^n w^0(i, j_k)} - u(i) = o(n^{-(1-1/p)}) \quad \text{almost surely.} \tag{37}$$

To prove the theorem, we need to estimate the difference between the left-hand sides of (8) and (37). Let

$$A_0 = \sum_{k=1}^{k_0-1} w^0(i, j_k)v(j_k), \quad A_n = \sum_{k=k_0}^n w^0(i, j_k)v(j_k),$$

$$B_0 = \sum_{k=1}^{k_0-1} w^0(i, j_k), \quad B_n = \sum_{k=k_0}^n w^0(i, j_k).$$

Then as before, fixing $x \in \mathbb{R}^{|\mathcal{N}_i^0|}$, applying Theorem 3 with $p = 1$ and Fubini's theorem, and replacing x by $v(\mathcal{N}_i^0)$, we obtain

$$\lim_{n \to \infty} \frac{A_n}{n} = \mathbb{E}w^0(i, j_k)v(j_k), \quad \text{and} \quad \lim_{n \to \infty} \frac{B_n}{n} = \mathbb{E}w^0(i, j_k).$$

Using this and the fact that

$$\frac{A_0 + A_n}{B_0 + B_n} - \frac{A_n}{B_n} = \frac{A_0 B_n - A_n B_0}{B_n (B_0 + B_n)},$$

we see that

$$\left| \frac{\sum_{k=k_0}^n w^0(i, j_k)v(j_k)}{\sum_{k=k_0}^n w^0(i, j_k)} - \frac{\sum_{k=1}^n w^0(i, j_k)v(j_k)}{\sum_{k=1}^n w^0(i, j_k)} \right| = O\left(\frac{1}{n}\right) \quad \text{almost surely.} \qquad (38)$$

Therefore, (37) implies that

$$\frac{\sum_{k=1}^n w^0(i, j_k)v(j_k)}{\sum_{k=1}^n w^0(i, j_k)} - u(i) = o(n^{-(1-1/p)}) \quad \text{almost surely.} \qquad (39)$$

As (39) holds for any $p \in [1, 2)$, we see that (8) holds for all $\epsilon \in (0, \frac{1}{2}]$. $\qquad\square$

We will prove Theorem 2, which demonstrates that as $n \to \infty$,

$$\sqrt{n}\left(\frac{\sum_{k=1}^n w^0(i, j_k)v(j_k)}{\sum_{k=1}^n w^0(i, j_k)} - u(i) \right) \xrightarrow{d} \mathcal{L},$$

where

$$w^0(i, j_k) = e^{-\|v(\mathcal{N}_i^0) - v(\mathcal{N}_{j_k}^0)\|^2/(2\sigma_r^2)},$$

and $\mathcal{L}$ is a mixture of centered Gaussian laws in the sense that it has a density of the form (15).

Proof of Theorem 2 The procedure of the proof is similar to that of the proof of Theorem 1. Fix $x \in \mathbb{R}^{|\mathcal{N}_i^0|}$, and set

$$a_k = w^0(x, j_k) = e^{-\|x - v(\mathcal{N}_{j_k}^0)\|^2/(2\sigma_r^2)}, \quad k = 1, 2, \ldots, n.$$

Then a_k and $v(j_k)$ are independent, $\mathbb{E}a_k v(j_k)/\mathbb{E}a_k = \mathbb{E}v(j_k) = u(i)$, and $\{(a_k, v(j_k))\}$ is a sequence of l-dependent and identically distributed random vectors with $l = (2d - 1)^2 - 1$, and $\mathbb{E}|a_k v(j_k)|^2 \le \mathbb{E}|v(j_k)|^2 < \infty$. Hence applying Theorem 4, we get, for any fixed x and any positive integer k_0,

$$Z_n(x) := \sqrt{n}\left(\frac{\sum_{k=k_0}^n w^0(x, j_k)v(j_k)}{\sum_{k=k_0}^n w^0(x, j_k)} - u(i) \right) \xrightarrow{d} N(0, c_x^2),$$

where $c_x > 0$ will be calculated by the end of the proof. This means that for any $t \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{P}(Z_n(x) \le t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}c_x} e^{-\frac{z^2}{2c_x^2}} dz.$$

Let $k_0 > l$ be the positive integer such that $v(\mathcal{N}_i^0)$ is independent of $v(\mathcal{N}_{j_k}^0)$ for all $k \ge k_0$. Then by Fubini's theorem and Lebesgue's dominated convergence theorem, we have

$$\lim_{n\to\infty} \mathbb{P}\left(Z_n\big(v(\mathcal{N}_i^0)\big) \le t\right) = \int_{-\infty}^t f(z)dz,$$

where

$$f(z) = \int_{\mathbb{R}^{|\mathcal{N}_i^0|}} \frac{1}{\sqrt{2\pi}c_x} e^{-\frac{z^2}{2c_x^2}} \mu(dx),$$

with μ being the law of $v(\mathcal{N}_i^0)$. In other words,

$$\sqrt{n}\left(\frac{\sum_{k=k_0}^n w^0(i, j_k)v(j_k)}{\sum_{k=k_0}^n w^0(i, j_k)} - u(i) \right) \xrightarrow{d} \mathcal{L},$$

where $\mathcal{L}$ is the law with density f. This together with (38) prove the equation (15) of Theorem 2.

We now turn to the calculation of c_x. Let $v_k = v(j_k)$ and $z_k = v_k \mathbb{E} a_1 - \mathbb{E}(a_1 v_1)$. Because of the independence of a_1 and v_1, we get $z_k = (v_k - \mathbb{E} v_1) \mathbb{E} a_1$. Then, it follows that

$$\mathbb{E}(a_1 z_1)^2 = \mathbb{E}(a_1^2 (v_1 - \mathbb{E} v_1)^2) \mathbb{E}^2 a_1 = \mathbb{E}(a_1^2) \mathbb{E}^2(a_1) \mathbb{E}(v_1 - \mathbb{E} v_1)^2,$$

and

$$\begin{aligned} \mathbb{E}(a_1 z_1 a_k z_k) &= \mathbb{E}(a_1 a_k (v_1 - \mathbb{E} v_1)(v_k - \mathbb{E} v_1) \mathbb{E}^2 a_1) \\ &= \mathbb{E}^2(a_1) \mathbb{E}(a_1 a_k (v_1 - \mathbb{E} v_1)(v_k - \mathbb{E} v_k)). \end{aligned}$$

Note that if (a_1, a_k) is independent of (v_1, v_k), it holds that

$$\mathbb{E}(a_1 z_1 a_k z_k) = \mathbb{E}^2(a_1) \mathbb{E}(a_1 a_k) \mathbb{E}(v_1 - \mathbb{E} v_1) \mathbb{E}(v_k - \mathbb{E} v_k) = 0, \qquad (40)$$

by the independence of v_1 and v_k. If v_1 is not contained in $v(\mathcal{N}_{j_k}^0)$, then (a_1, a_k) is independent of (v_1, v_k). Notice that according to the order of I_i defined in Sect. 2.2, when $k > d^2$, v_1 is not contained in $v(\mathcal{N}_{j_k}^0)$, so that (40) holds; therefore by Theorem 4,

$$\begin{aligned} c_x &= \sqrt{\mathbb{E}(a_1 z_1)^2 + 2\lambda}/(\mathbb{E} a_1)^2 \\ &= \frac{1}{\mathbb{E} a_1} \sqrt{\mathbb{E}(a_1^2) \mathbb{E}(v_1 - \mathbb{E} v_1)^2 + 2 \sum_{k=2}^{d^2} \mathbb{E}(a_1 a_k (v_1 - \mathbb{E} v_1)(v_k - \mathbb{E} v_k))}. \end{aligned}$$

We finally give an approximation of c_x. Recall that $a_k = e^{-\|x - v(\mathcal{N}_{j_k}^0)\|^2/(2\sigma_r^2)}$, and $v_k = v(j_k)$. Let $\mathcal{T}(j) = j - j_1 + j_k$ be the translation mapping j_1 to j_k (thus mapping $\mathcal{N}_{j_1}^0$ onto $\mathcal{N}_{j_k}^0$).

If $v(j_1)$ is not contained in $v(\mathcal{N}_{j_k}^0)$, we have already seen that $\mathbb{E}(a_1 z_1 a_k z_k) = 0$. If $v(j_1)$ is contained in $v(\mathcal{N}_{j_k}^0)$, to make (a_1, a_k) independent of (v_1, v_k), we can remove $v(j_1)$ from $v(\mathcal{N}_{j_k}^0)$ and the corresponding term $v(\mathcal{T}^{-1}(j_1))$ from $v(\mathcal{N}_{j_1}^0)$; remove $v(j_k)$ from $v(\mathcal{N}_{j_1}^0)$ and the corresponding term $v(\mathcal{T}(j_k))$ from $v(\mathcal{N}_{j_k}^0)$. The obtained values of a_1, a_k are very close to the initial values of a_1, a_k respectively. Hence, we can consider that $\mathbb{E}(a_1 z_1 a_k z_k) \approx 0$. Therefore

$$c_x \approx \frac{\sqrt{\mathbb{E}(a_1 z_1)^2}}{\mathbb{E}^2 a_1} = \frac{\sqrt{\mathbb{E} a_1^2}}{\mathbb{E} a_1} \sigma = \sigma \frac{\sqrt{\int_{R^m} e^{-\|x - v\|^2/\sigma_r^2} \mu(dv)}}{\int_{R^m} e^{-\|x - v\|^2/2\sigma_r^2} \mu(dv)},$$

where $v = v(\mathcal{N}_{j_k}^0)$ and $m = |\mathcal{N}_{j_k}^0|$. Recall that $\mu(dv)$ is the law of $v(\mathcal{N}_i^0)$, so it is also the law of $v(\mathcal{N}_{j_k}^0)$. Let

$$\nu = \mathbb{E}(v(\mathcal{N}_i^0)),$$

then $v \sim N(\nu, \sigma^2 I d_m)$, where $I d_m$ denotes the identity matrix of size $m \times m$. Since

$$\int_{R^m} e^{-\|x - v\|^2/a} \mu(dv) = \frac{1}{(\sqrt{2/a + 1/\sigma^2} \sigma)^m} \exp(-\|x - \nu\|^2/(a + 2\sigma^2)),$$

we get

$$c_x = \sigma^{m/2+1} \left(\frac{(\sigma^2 + \sigma_r^2)^2}{\sigma^2 \sigma_r^2 (2\sigma^2 + \sigma_r^2)} \right)^{m/4} \exp \left(\frac{\sigma^2}{2(\sigma_r^2 + \sigma^2)(\sigma_r^2 + 2\sigma^2)} \|x - \nu\|^2 \right).$$

This ends the proof of Theorem 2. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \square$

References

1. Akkoul, S., Ledee, R., Leconge, R., Harba, R.: A new adaptive switching median filter. IEEE Signal Process. Lett. **17**(6), 587–590 (2010)
2. Bilcu, R.C., Vehvilainen, M.: Fast nonlocal means for image denoising. In: Proceedings of SPIE, Digital Photography III 6502, 65020R (2007)
3. Bovik, A.: Handbook of Image and Video Processing. Academic Press, London (2005)
4. Buades, A., Coll, B., Morel, J.: A review of image denoising algorithms, with a new one. Multiscale Model. Simul. **4**(2), 490–530 (2005)
5. Castleman, K.: Digital Image Processing. Prentice Hall Press, Upper Saddle River (1996)
6. Chan, R., Hu, C., Nikolova, M.: An iterative procedure for removing random-valued impulse noise. IEEE Signal Process. Lett. **11**(12), 921–924 (2004)
7. Chang, S.G., Yu, B., Vetterli, M.: Adaptive wavelet thresholding for image denoising and compression. IEEE Trans. Image Process. **9**, 1532–1546 (2000)
8. Chen, T., Wu, H.: Adaptive impulse detection using center-weighted median filters. IEEE Signal Process. Lett. **8**(1), 1–3 (2001)
9. Chow, Y., Teicher, H.: Probability Theory: Independence, Interchangeability, Martingales. Springer, Berlin (2003)
10. Coifman, R., Donoho, D.: Translation-invariant denoising. Wavel. Stat. **103**, 125–150 (1995)
11. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Trans. Image Process. **16**(8), 2080–2095 (2007)
12. Delon, J., Desolneux, A.: A patch-based approach for random-valued impulse noise removal. In: 2012 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp. 1093–1096. IEEE (2012)
13. Delon, J., Desolneux, A.: A patch-based approach for removing impulse or mixed gaussian-impulse noise. SIAM J. Imaging Sci. **6**(2), 1140–1174 (2013)
14. Dong, Y., Chan, R., Xu, S.: A detection statistic for random-valued impulse noise. IEEE Trans. Image Process. **16**(4), 1112–1120 (2007)
15. Donoho, D., Johnstone, J.: Ideal spatial adaptation by wavelet shrinkage. Biometrika **81**(3), 425–455 (1994)
16. Durand, S., Froment, J.: Reconstruction of wavelet coefficients using total variation minimization. SIAM J. Sci. Comput. **24**(5), 1754–1767 (2003)
17. Duval, V., Aujol, J., Gousseau, Y.: A bias-variance approach for the nonlocal means. SIAM J. Imaging Sci. **4**, 760–788 (2011)
18. Elad, M., Aharon, M.: Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans. Image Process. **15**(12), 3736–3745 (2006)
19. Garnett, R., Huegerich, T., Chui, C., He, W.: A universal noise removal algorithm with an impulse detector. IEEE Trans. Image Process. **14**(11), 1747–1754 (2005)
20. Hu, H., Froment, J.: Nonlocal total variation for image denoising. In: Symposium on Photonics and Optoelectronics (SOPO), pp. 1–4. IEEE (2012)
21. Hu, H., Li, B., Liu, Q.: Non-local filter for removing a mixture of gaussian and impulse noises. In: Csurka, G., Braz, J. (eds.) VISAPP, vol. 1, pp. 145–150. SciTePress (2012)
22. Johnstone, I.M., Silverman, B.W.: Wavelet threshold estimators for data with correlated noise. J. R. Stat. Soc. B **B 59**(2), 319–351 (1997)
23. Kervrann, C., Boulanger, J.: Local adaptivity to variable smoothness for exemplar-based image regularization and representation. Int. J. Comput. Vis. **79**(1), 45–69 (2008)
24. Li, B., Liu, Q., Xu, J., Luo, X.: A new method for removing mixed noises. Sci. China Inf. Sci. **54**, 51–59 (2011)
25. López-Rubio, E.: Restoration of images corrupted by gaussian and uniform impulsive noise. Pattern Recognit. **43**(5), 1835–1846 (2010)
26. Mahmoudi, M., Sapiro, J.: Fast image and video denoising via nonlocal means of similar neighborhoods. IEEE Signal. Process. Lett. **12**(12), 839–842 (2005)
27. Mairal, J., Sapiro, G., Elad, M.: Learning multiscale sparse representations for image and video restoration. Multiscale Model. Simul. **7**, 214–241 (2008)
28. Muresan, D., Parks, T.: Adaptive principal components and image denoising. In: IEEE International Conference on image processing (2003)
29. Nikolova, M.: A variational approach to remove outliers and impulse noise. J. Math. Imaging Vis. **20**(1), 99–120 (2004)
30. Pratt, W.: Median filtering. Semiannual Report, Image Processing Institute, University of Southern California, Los Angeles, Tech. Rep pp. 116–123 (1975)

31. Pruitt, W.: Summability of independent random variables. J. Math. Mech. **15**, 769–776 (1966)
32. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Phys. D Nonlinear Phenom. **60**(1–4), 259–268 (1992)
33. Smith, S., Brady, J.: SUSAN—a new approach to low level image processing. Int. J. Comput. Vis. **23**(1), 45–78 (1997)
34. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: Proceedings of the international confernce on computer vision, pp. 839–846. IEEE (1998)
35. Van De Ville, D., Kocher, M.: Sure-based non-local means. IEEE Signal Process. Lett. **16**(11), 973–976 (2009)
36. Wen, Z.Y.: Sur quelques théorèmes de convergence du processus de naissance avec interaction des voisins. Bull. Soc. Math. Fr. **114**, 403–429 (1986)
37. Xiao, Y., Zeng, T., Yu, J., Ng, M.: Restoration of images corrupted by mixed gaussian-impulse noise via l sub (1-l) sub (0) minimization. Pattern Recognit. **44**(8), 1708–1720 (2011)
38. Xiong, B., Yin, Z.: A universal denoising framework with a new impulse detector and nonlocal means. IEEE Trans. Image Process. **21**(4), 1663–1675 (2012)
39. Xu, H., Xu, J., Wu, F.: On the biased estimation of nonlocal means filter. In: International conference on multimedia computing and systems/international conference on multimedia and expo, pp. 1149–1152 (2008)
40. Yang, J.X., Wu, H.R.: Mixed gaussian and uniform impulse noise analysis using robust estimation for digital images. In: Proceedings of the 16th international conference on digital signal processing, pp. 468–472 (2009)
41. Yaroslavsky, L.P.: Digital Picture Processing. Springer, Secaucus (1985)
42. Yuksel, M.: A hybrid neuro-fuzzy filter for edge preserving restoration of images corrupted by impulse noise. IEEE Trans. Image Process. **15**(4), 928–936 (2006)
43. Zhou, Y., Ye, Z., Xiao, Y.: A restoration algorithm for images contaminated by mixed gaussian plus random-valued impulse noise. J. Vis. Commun. Image Represent. **24**(3), 283–294 (2013)