Non-local medians filter for joint Gaussian and impulsive image denoising

Alexandre L. M. Levada
Computing Department, Federal University of São Carlos, São Carlos, SP, Brazil
alexandre.levada@ufscar.br

***Abstract*—Image denoising concerns with the development of filters to remove or attenuate random perturbations in the observed data, but at the same time, preserving most of edges and fine details in the scene. One problem with joint additive Gaussian and impulsive noise degradation is that they are spread over all frequencies of the signal. Hence, the most effective filters for this kind of noise are implemented in the spatial domain. In this paper, we proposed a Non-Local Medians filter that combine the medians of every patch of a search window using two distinct similarity measures: the Euclidean distance and the Kullback-Leibler divergence between Gaussian densities estimated from the patches. Computational experiments with 25 images corrupted by joint Gaussian and impulsive noises show that the proposed method is capable of producing, on average, significant higher PSNR and SSIM than the combination of the median filter and the Non-Local Means filter applied independently.**

I. Introduction

Image denoising is a crucial pre-processing stage in many pattern recognition and computer vision tasks. Among them, we can cite low-level image segmentation and tracking of objects in a video. Hence, the problem of estimating the true underlying signal from a degraded observed version is a topic of interest by many researchers. In summary, the main goal in denoising is to filter random perturbations in the image with a smoothing kernel, but keeping the maximum amount of details untouched [1], [2].

The typical statistical model assumed by most denoising filters is the AWGN (Additive White Gaussian Noise), which assumes that the observed pixel is the sum of two components, the underlying true value, and the noise term, generated from a Gaussian distribution with zero mean and variance σ_n^2. Moreover, the noise is uncorrelated with the own noise and the underlying signal [3]. Many filters have been proposed to deal with this kind of noise in the literature and the state-of-the-art in AWGN denoising is achieved by Non-Local Means [4] in the spatial domain and BM3D [5] in a sparse wavelet-based domain [6].

Another kind of degradation often presented in digital image processing is the impulsive noise, that are short duration "on/off" pulses caused by problems in acquisition, such as dead pixels in screens or digital cameras [7]. One of the best approaches for impulsive noise filtering are the non-linear rank-order filters, especially the median filter, which is more robust against the presence of outliers than the mean filter.

In this paper, we are interested in denoising images corrupted by joint Gaussian and impulsive noises. Recently, there has been a large interest in dealing with mixture of noises by combining different approaches. In the work of Hu et. al [8], a patch-based weighted means is proposed to remove a mixture of Gaussian and impulsive noises by combining ideas of the Trilateral filter and Non-Local Means. In the work of Awad [9], a method based on a cascade of stages is proposed to deal with mixture of Gaussian and impulsive noises. The idea consists in removing outliers in the first stage and the remaining of the noise is filtered in the subsequent stages. Adaptive total variation using L_1 norm regularization has been applied with success in effectively removing impulsive noise from images [10]. An optimization algorithm based on low-rank decomposition with tensor robust Principal Component Analysis has been proposed to recover the underlying signal corrupted with sparse noise/outliers using the definition of tensor logarithmic norm to avoid over-penalization of large eigenvalues [11].

Our proposed method to denoise joint Gaussian and impulsive noises is an attempt to combine the median and the Non-Local Means filters into a single method. In the first variant, we use as distance function the traditional Euclidean distance to measure the similarity between two patches. In the second variation, inspired by the work of Bindilatti and Mascarenhas to extend the Non-Local Means filter to Poisson noise [12], we employ information-theoretic divergences, namely, the KL-divergence or relative entropy, to measure the similarity between patches from the same search window.

It has been reported in the literature that some computerized tomography images can be corrupted with a mixture of Gaussian and impulsive noise during data acquisition process [13]. Hence, the usefulness of the proposed method can be relevant in medical image processing.

The main contributions of this work are twofold: 1) we propose Non-Local Medians, an extension to the traditional Non-Local Means filter to deal with impulsive noise by using the L_1 norm and the KL divergence as similarity measures; and 2) results obtained with 25 different images corrupted by a mixture of Gaussian and impulsive noises indicate that the proposed method is capable of producing quantitative results in terms of PSNR and SSIM that are, on average, superior than the results obtained by the successive application of the individual filters (first Median, then NLM).

The remaining of the paper is organized as follows: Section 2 describes the mathematical formulation of the proposed noise model, the median and the Non-Local Means filters, as

well as the Kullback-Leibler divergence and the computation of its symmetrized version in the univariate Gaussian case. Section 3 describes the proposed method in details: Non-Local Medians (using the Euclidean distance) and Non-Local Medians KL (using the KL divergence). Section 4 shows the experiments and the obtained results in terms of two quantitative metrics: the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index (SSIM). Finally, Section 5 presents the conclusions, final remarks and future directions.

II. Related work

In this section, we define the mathematical model of the denoising problem and briefly discuss the median and Non-Local Means filters.

A. The noise model

Considering joint additive zero mean Gaussian noise and the impulsive noise, the mathematical model for the denoising problem is given by:

$$y_i = x_i + n_i \tag{1}$$

$$z_i = h(y_i) \tag{2}$$

where y_i is the corrupted pixel, x_i is the respective noise-free pixel, n_i is the additive Gaussian noise, $h(.)$ is a non-linear function that turns y_i to 0 or 255 with probability p. In the classical approach, we have $n_i \sim N(0, \sigma_n^2)$. Note that the noise is uncorrelated, that is, $E[n_i n_j] = \sigma_n^2 \delta_{i,j}$, where $\delta_{i,j} = 1$ if $i = j$ and $\delta_{i,j} = 0$ if $i \neq j$ and $E[.]$ denotes the expectation operator. Thus, according to equation (1), we have $E[y_i] = E[x_i] = \mu$ and $\sigma_z^2 = \sigma_x^2 + \sigma_n^2$.

A digital impulse, $\delta(n)$, is defined as a signal with duration of one sample that can be expressed as:

$$\delta(n) = \begin{cases} 1 & \text{if } n = 0 \\ 0 & \text{if } n \neq 0 \end{cases} \tag{3}$$

By computing the Fourier transform of the impulse signal, we have:

$$F\{\delta(n)\} = \sum_{n=-\infty}^{\infty} \delta(n) e^{-j2\pi fn} = 1 \tag{4}$$

which shows that the energy of an impulse is equally spread among all frequency components [7]. Usually the best filters to deal with this kind of noise work in the spatial domain.

B. The median filter

Given the model defined by equations (1) and (2), it is possible to recover y_i by inverting the non-linear function $h(.)$, that is, $y_i = h^{-1}(z_i)$. One way to achieve this goal is to use the median filter. The median filtering of a 2D image comprises the following steps: 1) Apply a sliding window of size $k \times k$, where k is chosen to be an odd integer; 2) At each pixel $p_{i,j}$, store the pixel values of the window in an array of size k^2; 3) Sort the resulting array; 4) Replace the central pixel $p_{i,j}$ of the image by the median (the element in the middle of the sorted array). Note that a sorting algorithm is required for finding the median of a set of values. There exist various sorting algorithm with complexity of $O(n\ log(n))$, as Quicksort, for example. In larger windows, the difference between different sorting algorithms is noticeable in the computational time. Median filter is suitable for impulsive noise since the median of a set is quite robust to the presence of outliers. Figure 1 illustrates the process.

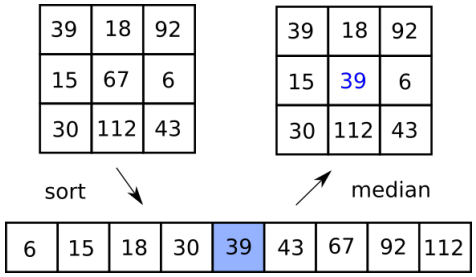

Fig. 1. Median filtering of a pixel using a kernel of size 3×3

C. The non-local means filter

The non-local means algorithm tries to take advantage of the high degree of redundancy in natural images, by scanning a vast portion of the image in search of similar pixels, using the concept of a similarity measure [4], as Figure 2 illustrates. Given a noisy image $x = \{x_i | i \in I\}$, the estimated value of the noise-free pixel, denoted by $NL[x](i)$, is computed as a weighted average of all the pixels in the image:

$$NL[x](i) = \sum_{j \in I} w(i,j) x_j \tag{5}$$

where the family of weights $\{w(i,j)\}_j$ depend on the similarity between pixels i and j, and satisfy the usual conditions $0 \leq w(i,j) \leq 1$ and $\sum_j w(i,j) = 1$. The similarity between two pixels i and j depends on the similarity of the intensity gray level vectors $x(\eta_i)$ and $x(\eta_j)$, where η_k denotes a patch, that is, a square neighborhood of fixed size and centered at a pixel k. The pixels with a similar gray level neighborhood to $x(\eta_i)$ have larger weights on average. These weights are defined as:

$$w(i,j) = \frac{1}{Z(i)} exp\left\{ -\frac{||x(\eta_i) - x(\eta_j)||_2^2}{h^2} \right\} \tag{6}$$

where $Z(i)$ is the normalizing constant given by:

$$Z(i) = \sum_j exp\left\{ -\frac{||x(\eta_i) - x(\eta_j)||_2^2}{h^2} \right\} \tag{7}$$

and h is a parameter that controls the decay of the exponential, acting as a degree of smoothing. Considering an input image of size n, the NLM algorithm has order of $O(n^2\Omega)$, where Ω denotes the patch size. To achieve a good trade-off between cost and performance, the non-local estimate of the noise-free pixel, given by the summation in equation (5), does not involve all the image pixels. Instead, it considers a search window of

size $s \times s$ around the current pixel i, reducing the complexity to $O(nsΩ)$. It is possible to extend the NLM filter by using different distance functions instead of the traditional Euclidean metric. Some works consider stochastic distances based on the parameters of statistical models estimated inside each patch to deal with other kind of noises, such as signal dependent noise (Poisson) [12] and multiplicative noise (speckle) [14].

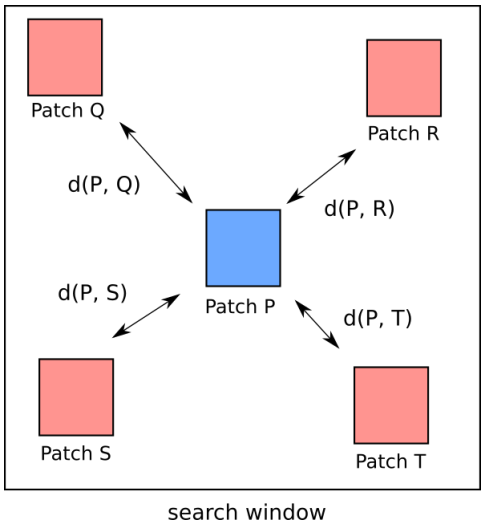

Fig. 2. Non-Local Means employs the concept of similarity measures between patches to compute the weights used in the mean value.

III. The Proposed Method

In this section, we describe the proposed method in details, describing how each variation works. In summary, the idea behind Non-Local Medians is to deal with the impulsive noise by computing a weighted median instead of a weighted mean.

A. Non-Local Medians

In the first variation of the proposed method, namely, Non-Local Medians, the main difference in comparison with NLM consists in modifying equation (5) to:

$$NL[x](i) = \sum_{j \in I} w(i,j) m_j \tag{8}$$

where m_j denotes the median value of the patch $x(\eta_j)$. Note that, by replacing x_j by m_j, we eventually suppress impulsive noise within that particular patch. As we are trying to estimate the median in a non-local way, instead of using the Euclidean L_2 norm, we employ the L_1 norm as similarity measure in Non-Local Medians, since it has been shown that in statistics, L_2 regression estimates the mean while L_1 regression estimates the median.

B. Non-Local Medians KL

The idea here is to replace the L_1 norm by an information-theoretic distance function: the KL-divergence or relative entropy.

1) Kullback-Leibler divergence: One of the most challenging tasks in multivariate data analysis is quantifying meaningful similarity measures between data points in an unsupervised manner [15]. Finding alternative distance functions can bring benefits to data clustering and filtering, especially in scenarios in which the standard Euclidean distance becomes an unreasonable choice [16]. Information-theoretic measures have been successfully applied in statistics to quantify a degree of similarity between random variables [17]. In this context, the concepts of entropy and relative entropy can be used as a solid mathematical background for unsupervised metric learning [18]. We begin by introducing the entropy of a random variable x as the expected value of the self-information:

$$H(p) = -\int p(x)[log\ p(x)]dx = -E\left[log\ p(x)\right] \tag{9}$$

where $p(x)$ is the probability density function (pdf) of x. Assuming x is normally distributed as $N(\mu, \sigma^2)$, its entropy is given by:

$$\begin{aligned} H(p) &= \frac{1}{2} log\ (2\pi\sigma^2) + \frac{1}{2\sigma^2} E[(x-\mu)^2] \\ &= \frac{1}{2}\left(1 + log\ (2\pi\sigma^2)\right) \end{aligned} \tag{10}$$

In a similar way, we can define the cross-entropy between two probability density functions as:

$$H(p,q) = -\int p(x)[log\ q(x)]dx \tag{11}$$

The Kullback-Leibler divergence, or simply relative entropy, is the difference between the cross-entropy of $p(x)$ and $q(x)$ and the entropy of $p(x)$ [19], that is:

$$\begin{aligned} D_{KL}(p,q) &= H(p,q) - H(p) \\ &= -\int p(x)[log\ q(x)]dx + \int p(x)[log\ p(x)]dx \\ &= \int p(x) log\left(\frac{p(x)}{q(x)}\right) dx = E_p\left[log\left(\frac{p(x)}{q(x)}\right)\right] \end{aligned} \tag{12}$$

It should be mentioned that the relative entropy is always non-negative, that is, $D_{KL}(p,q) \geq 0$, being equal to zero if, and only if, $p(x) = q(x)$. Let $p(x)$ and $q(x)$ be univariate Gaussian densities, $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$. Then, the KL-divergence between them is given by:

$$\begin{aligned} D_{KL}(p,q) = log\left(\frac{\sigma_2}{\sigma_1}\right) &+ \frac{1}{2\sigma_2^2} E_p[(x-\mu_2)^2] \\ &- \frac{1}{2\sigma_1^2} E_p[(x-\mu_1)^2] \end{aligned} \tag{13}$$

It is straightforward to note that:

$$E_p[(x-\mu_1)^2] = \sigma_1^2 \tag{14}$$

$$E_p[(x-\mu_2)^2] = E[x^2] - 2E[x]\mu_2 + \mu_2^2 \tag{15}$$

$$E[x^2] = Var[x] + E^2[x] = \sigma_1^2 + \mu_1^2 \tag{16}$$

which finally leads to:

$$D_{KL}(p,q) = log\left(\frac{\sigma_2}{\sigma_1}\right) + \frac{1}{2\sigma_2^2}(\sigma_1^2 + \mu_1^2 - 2\mu_1\mu_2 + \mu_2^2) - \frac{1}{2}$$
$$= log\left(\frac{\sigma_2}{\sigma_1}\right) + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2} \quad (17)$$

Note that $D_{KL}(p,q) \neq D_{KL}(q,p)$, that is, the relative entropy is not symmetric. The symmetrized KL-divergence between $p(x)$ and $q(x)$ is:

$$D_{KL}^{sym}(p,q) = \frac{1}{2}[D_{KL}(p,q) + D_{KL}(q,p)] \quad (18)$$
$$= \frac{1}{4}\left[\frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{\sigma_2^2} + \frac{\sigma_2^2 + (\mu_1 - \mu_2)^2}{\sigma_1^2} - 2\right]$$
$$= \frac{1}{4\sigma_1^2\sigma_2^2}\left[\left(\sigma_1^2 - \sigma_2^2\right)^2 + \left(\mu_1 - \mu_2\right)^2\left(\sigma_1^2 + \sigma_2^2\right)\right]$$

In summary, Non-Local Medians KL works as follows:

1) For each patch p_j in the search window, we compute the median m_j and the standard deviation σ_j;
2) Discard all samples that do not belong to the 95% confidence interval around the median, that is, we select to compose our novel patch $\hat{p}_j$ the pixels that satisfy:
$$m_j - 1.96\sigma_j < x_j < m_j + 1.96\sigma_j \quad (19)$$
3) Estimate the local mean and the local variance of the novel patch $\hat{p}_j$;
4) Compute the KL-divergence between the central patch $\hat{p}_i$ and $\hat{p}_j$ using equation (18);
5) Compute the weight $w(i,j)$ as:
$$w(i,j) = \frac{1}{Z(i)} exp\left\{-\frac{D_{KL}^{sym}(p,q)}{h^2}\right\} \quad (20)$$
6) Estimate the noise-free pixel x_i as:
$$NL[x](i) = \sum_{j \in I} w(i,j) m_j \quad (21)$$

IV. Experiments and results

To test and evaluate the performance of the proposed Non-Local Medians approach for image denoising, we conducted a set of computational experiments with 25 different gray level images obtained from the USC-SIPI database, shown in Figure 3. Basically, we compared the performance of both variations of the proposed method, namely, Non-Local Medians (NLMED) and Non-Local Medians KL (NLMEDKL) with the Median filter alone (MED) and the successive application of the original Median (first) and NLM (second) filters (MED+NLM). To compare the different methods, we selected two quantitative measures: PSNR [20] and SSIM [21]. Tables I and II shows the obtained PSNR's and SSIM's for each one of the 25 images. All images were degraded by additive Gaussian noise with variance $\sigma_n^2 = 10$ and salt and pepper noise in 5% randomly selected image pixels. Figure 4 shows some noisy images used in the experiments.

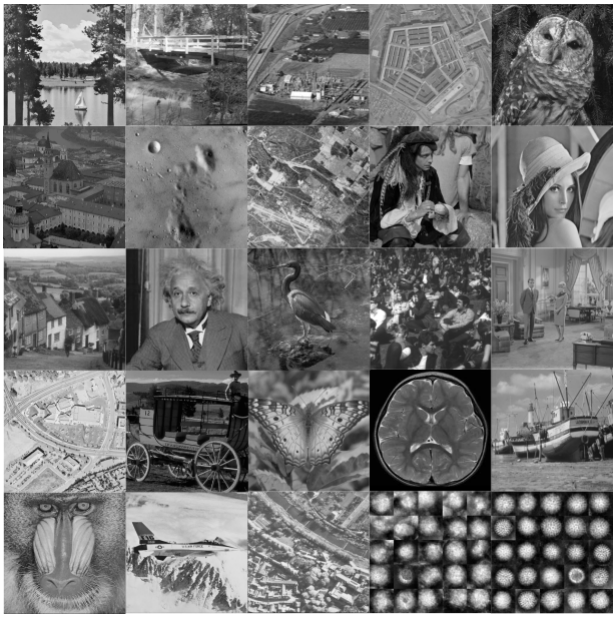

Fig. 3. Selected gray level images for the computational experiments.

It is worth noticing that, according to the PSNR metric, the proposed method (NLMED or NLMEDKL) produced the best results in 24 out of 25 images, that is, in 96% of the cases. In terms of SSIM, the proposed method (NLMED or NLMEDKL) produced the best results in 22 out of 25 images, which represents 88% of the cases. To test if the results obtained by the proposed method are statistically superior, we performed a non-parametric Friedman test [22]. In terms of PSNR, there are strong evidences against the null hypothesis that all groups are identical ($p = 1.11 \times 10^{-16}$) for a significance level $\alpha = 0.01$. To check which groups are significantly different, we performed a Nemenyi post-hoc test [23]. According to the test, for a significance level $\alpha = 0.01$, there are strong evidences that the proposed method NLMEDKL produced significantly higher PSNR's than the Median filter MED ($p = 1.43 \times 10^{-6}$) and MED+NLM ($p = 4.32 \times 10^{-8}$). Similarly, the proposed method NLMED produced significantly higher PSNR's than the Median filter MED ($p = 1.44 \times 10^{-6}$) and MED+NLM ($p = 4.31 \times 10^{-8}$). There are no evidences that the variations of the proposed method NLMED and NLMEDKL are different in terms of PSNR for these images ($p = 0.511$), with the same for MED and MED+NLM ($p = 0.999$).

The same analysis is performed in terms of SSIM. There are strong evidences against the null hypothesis that all groups are identical ($p = 2.99 \times 10^{-14}$) for a significance level $\alpha = 0.01$. To check which groups are significantly different, we performed a Nemenyi post-hoc test. According to the test, for a significance level $\alpha = 0.01$, there are strong evidences that the proposed method NLMEDKL produced significantly higher SSIM's than the Median filter MED ($p = 8.24 \times 10^{-7}$), MED+NLM ($p = 2.62 \times 10^{-7}$). Similarly, the proposed method NLMED produced significantly higher SSIM's than

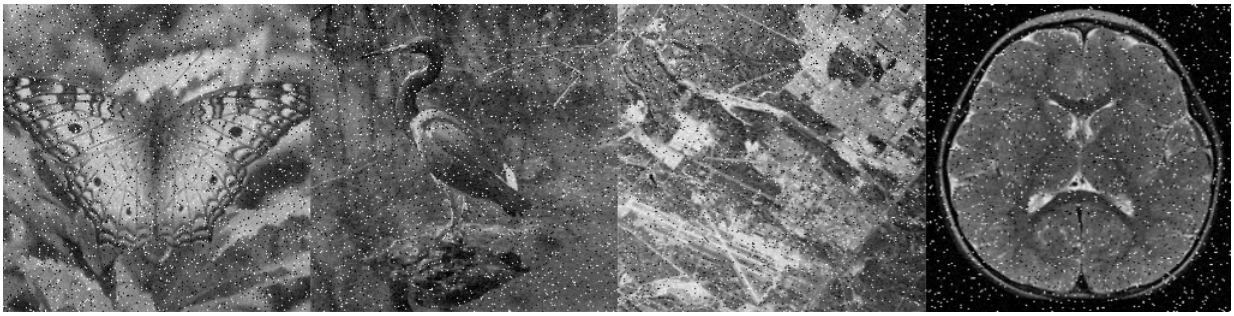

Fig. 4. Noisy images corrupted with joint Gaussian and impulsive noises.

TABLE I
PSNR VALUES OBTAINED AFTER PERFORMING DENOISING WITH MEDIAN FILTER (MED), MEDIAN AND NLM (MED+NLM) AND THE PROPOSED METHOD VARIATIONS NON-LOCAL MEDIANS (NLMED) AND NON-LOCAL MEDIANS WITH KL-DIVERGENCE (NLMEDKL).

	MED	**MED+NLM**	**NLMED**	**NLMEDKL**
Aerial	26.243	26.001	26.438	**26.458**
Baboon	25.109	24.921	25.188	**25.211**
Brain	28.556	28.699	**28.943**	28.579
Boat	28.975	28.912	**29.244**	29.160
Crowd	28.746	28.612	**28.860**	28.795
Couple	30.342	30.300	30.580	**30.611**
Goldhill	30.383	30.318	30.634	**30.761**
Lena	30.829	31.303	**31.336**	31.198
Man	28.798	28.683	**29.135**	29.084
Miramar	26.414	25.885	26.452	**26.566**
Moon	29.618	29.846	29.976	**30.016**
Duck	28.619	28.541	28.753	**28.833**
Plant	27.724	27.630	**27.752**	27.745
River	27.598	27.224	27.651	**27.767**
Pentagon	29.856	29.414	30.006	**30.119**
Airplane	27.208	27.301	**29.409**	27.350
Sail	26.138	26.040	26.349	**26.384**
City	24.225	24.012	**24.280**	24.270
Einstein	32.101	32.478	32.642	**32.688**
Owl	28.299	27.736	28.373	**28.506**
Butterfly	29.842	29.91	30.144	**30.081**
OldCity	26.898	27.005	**27.080**	27.052
VirusA	28.298	27.896	28.403	**28.491**
VirusB	28.943	28.766	29.206	**29.316**
Chariots	23.692	23.715	**23.769**	23.697
Average	28.138	28.046	28.344	**28.350**
Median	28.556	28.541	**28.753**	28.579
Minimum	23.692	23.715	**23.769**	23.697
Maximum	32.101	32.478	32.642	**32.688**

TABLE II
SSIM VALUES OBTAINED AFTER PERFORMING DENOISING WITH MEDIAN FILTER (MED), MEDIAN AND NLM (MED+NLM) AND THE PROPOSED METHOD VARIATIONS NON-LOCAL MEDIANS (NLMED) AND NON-LOCAL MEDIANS WITH KL-DIVERGENCE (NLMEDKL).

	MED	**MED+NLM**	**NLMED**	**NLMEDKL**
Aerial	0.8648	0.8487	0.8692	**0.8704**
Baboon	0.7046	0.6583	0.7077	**0.7104**
Brain	0.8269	0.8568	**0.8587**	0.8461
Boat	0.8281	0.8332	**0.8428**	0.8425
Crowd	0.8733	0.8795	**0.8870**	0.8857
Couple	0.8239	0.8253	0.8368	**0.8374**
Goldhill	0.8409	0.8375	0.8560	**0.8579**
Lena	0.8566	**0.9006**	0.8907	0.8854
Man	0.8471	0.8405	0.8561	**0.8580**
Miramar	0.8185	0.7672	0.8211	**0.8250**
Moon	0.6782	0.6740	0.6904	**0.6954**
Duck	0.7749	0.7597	0.7805	**0.7874**
Plant	0.8297	0.8048	0.8330	**0.8365**
River	0.8221	0.7791	0.8235	**0.8277**
Pentagon	0.8368	0.8012	0.8404	**0.8473**
Airplane	0.8229	**0.8714**	0.8640	0.8537
Sail	0.8206	0.8405	**0.8443**	0.8406
City	0.7567	0.7243	0.7605	**0.7625**
Einstein	0.8553	0.8752	0.8741	**0.8827**
Owl	0.8451	0.8029	0.8463	**0.8523**
Butterfly	0.8783	0.8896	**0.8925**	0.8900
OldCity	0.6894	0.6809	**0.7015**	0.6996
VirusA	0.8954	0.8777	0.8963	**0.8975**
VirusB	0.9135	0.9025	0.9170	**0.9187**
Chariots	0.744	0.741	**0.7469**	0.7438
Average	0.8179	0.8109	0.8295	**0.8302**
Median	0.8281	0.8332	0.8443	**0.8461**
Minimum	0.6782	0.6583	0.6904	**0.6954**
Maximum	0.9135	0.9025	0.9170	**0.9187**

the Median filter MED ($p = 1.93 \times 10^{-5}$) and MED+NLM ($p = 7.07 \times 10^{-6}$). There are no evidences that the variations of the proposed method NLMED and NLMEDKL are different in terms of PSNR for these images ($p = 0.510$), with the same for MED and MED+NLM ($p = 0.826$).

To illustrate the difference between the methods, Figure 5 shows visual results for Miramar, Butterfly, Brain, Airplane and Old City images. It is possible to notice that there is a significant level of residual noise in the Median filtered images. On the other hand, when we successively apply the Median and the NLM filters, the images exhibit oversmoothing, in the sense the many fine details in the images are not preserved. The proposed Non-Local Medians variations provide a better tradeoff between noise reduction and edge preservation. Figure 6 shows additional qualitative results for the images Goldhill, Duck, Plant, Moon and River. Again, note how the proposed Non-Local Medians filter achieves a good balance between smoothing and conservation of high frequency content.

In future works, we intend to compare the proposed method against other state-of-the-art approaches for joint Gaussian and impulsive noise removal. Among these methods, we can cite a robust non-local median filter [24] and a PCA based filter [25]. However, we need to implement them from scratch, since the source codes are not public available.

V. Conclusion

Denoising of images corrupted by joint Gaussian and impulsive noises is a challenging task, since state-of-the-art methods based on domain transformations and sparse representations (BM3D) are not effective. Methods in the spatial domain are often better choices to deal with impulsive noise. Given the above, in this paper, we presented a Non-Local Medians filter that combine the features of rank-order and non-local strategies.

Our method can be considered as an extension of NLM to deal with impulsive noise. The proposed Non-Local Medians variations unify two distinct approaches that are very good at dealing with Gaussian and impulsive noise separately: the Non-Local Means filter and the Median filter. Computational experiments with several real images jointly corrupted by Gaussian and impulsive noises showed that the proposed method can produce, on average, significantly better results in terms of PSNR and SSIM than the successive application of the original Median and NLM filters.

Future works may include the use of other information-theoretic divergences and other family of entropies as similarity measures, such as the Hellinger, Bhattacharyya, Cauchy-Schwarz and Total Variation divergences, as well as, Renyi and Sharma-Mittal entropies. Dimensionality reduction based metric learning algorithms can be applied to learn a more compact and meaningful representation for the patches within a search window. We intend to apply PCA, Sparse PCA, Robust PCA and Parametric PCA before the computation of the Euclidean distances as a way to learn better similarity measures. Furthermore, Gaussian-Markov random field models can be used to capture the spatial dependence structure of the patches, as a way to induce an additional smoothing constraint to the denoising problem.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

References

[1] P. Chatterjee and P. Milanfar, "Is denoising dead?" *IEEE Transactions on Image Processing*, vol. 19, no. 4, pp. 895–911, 2010.

[2] C. Jin and N. Luan, "An image denoising iterative approach based on total variation and weighting function," *Multimedia Tools and Applications*, vol. 79, p. 20947–20971, 2020.

[3] C. Boncelet, "Chapter 7 - image noise models," in *The Essential Guide to Image Processing*, A. Bovik, Ed. Boston: Academic Press, 2009, pp. 143–167.

[4] A. Buades, B. Coll, and J. . Morel, "A non-local algorithm for image denoising," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2, 2005, pp. 60–65.

[5] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[6] M. Shi, F. Zhang, S. Wang, C. Zhang, and X. Li, "Detail preserving image denoising with patch-based structure similarity via sparse representation and svd," *Computer Vision and Image Understanding*, vol. 206, p. 103173, 2021.

[7] S. Vaseghi, "Impulsive noise," in *Advanced Signal Processing and Noise Reduction*. Vieweg+Teubner Verlag, 1996.

[8] H. Hu, B. Li, and Q. Liu, "Removing mixture of gaussian and impulse noise by patch-based weighted means," *Journal of Scientific Computing*, vol. 67, no. 1, pp. 103–129, 2016.

[9] A. Awad, "Denoising images corrupted with impulse, gaussian, or a mixture of impulse and gaussian noise," *Engineering Science and Technology, an International Journal*, vol. 22, no. 3, pp. 746–753, 2019.

[10] D. N. H. Thanh, L. T. Thanh, N. N. Hien, and S. Prasath, "Adaptive total variation l1 regularization for salt and pepper image denoising," *Optik*, vol. 208, p. 163677, 2020.

[11] L. Chen, X. Jiang, X. Liu, and Z. Zhou, "Robust low-rank tensor recovery via nonconvex singular value minimization," *IEEE Transactions on Image Processing*, vol. 29, pp. 9044–9059, 2020.

[12] A. A. Bindilatti and N. D. A. Mascarenhas, "A nonlocal poisson denoising algorithm based on stochastic distances," *IEEE Signal Processing Letters*, vol. 20, no. 11, pp. 1010–1013, 2013.

[13] J. J. Madhura and D. R. R. Babu, "An effective hybrid filter for the removal of gaussian-impulsive noise in computed tomography images," in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2017, pp. 1815–1820.

[14] Y. Tounsi, M. Kumar, A. Nassim, F. Mendoza-Santoyo, and O. Matoba, "Speckle denoising by variant nonlocal means methods," *Applied Optics*, vol. 58, no. 26, pp. 7110–7120, 2019.

[15] A. S. Shirkhorshidi, S. Aghabozorgi, and T. Y. Wah, "A comparison study on similarity and dissimilarity measures in clustering continuous data," *PLOS ONE*, vol. 10, no. 12, pp. 1–20, 12 2015.

[16] C. C. Aggarwal, A. Hinneburg, and D. A. Keim, "On the surprising behavior of distance metrics in high dimensional space," in *Database Theory — ICDT 2001*, J. Van den Bussche and V. Vianu, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 420–434.

[17] F. Pérez-Cruz, "Estimation of information theoretic measures for continuous random variables," in *Advances in Neural Information Processing Systems 21*, D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, Eds. Curran Associates, Inc., 2009, pp. 1257–1264.

[18] J. A. Costa and A. O. Hero, "Geodesic entropic graphs for dimension and entropy estimation in manifold learning," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2210–2221, 2004.

[19] S. Kullback and R. A. Leibler, "On information and sufficiency," *Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[20] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of psnr in image/video quality assessment," *Electronics Letters*, vol. 44, pp. 800–801(1), 2008.

[21] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[22] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *Journal of the American Statistical Association*, vol. 32, no. 200, pp. 675–701, 1937.

[23] M. Hollander, D. A. Wolfe, and E. Chicken, *Nonparametric Statistical Methods*, 3rd ed. Wiley, 2015.

[24] J. Matsuoka, T. Koga, N. Suetake, and E. Uchino, "Robust non-local median filter," *Optical Review*, vol. 24, no. 2, pp. 87–96, 1 2017.

[25] A. Awad, "Denoising images corrupted with impulse, gaussian, or a mixture of impulse and gaussian noise," *Engineering Science and Technology, an International Journal*, vol. 22, no. 3, pp. 746–753, 2019.

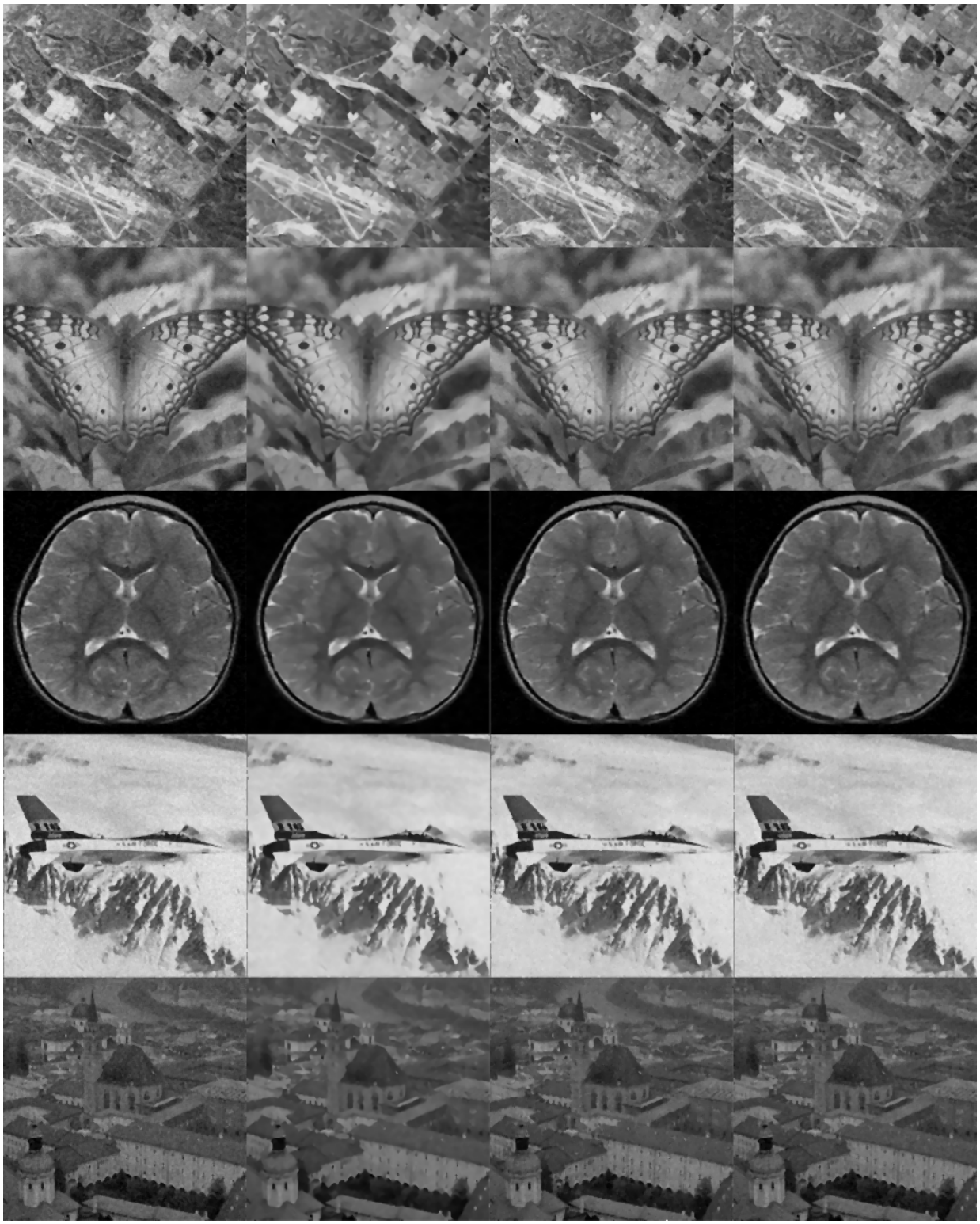

Fig. 5. Visual results for Miramar, Butterfly, Brain, Airplane and Old City images. From left to right, we have: Median, Median+NLM, Non-Local Medians and Non-Local Medians KL.

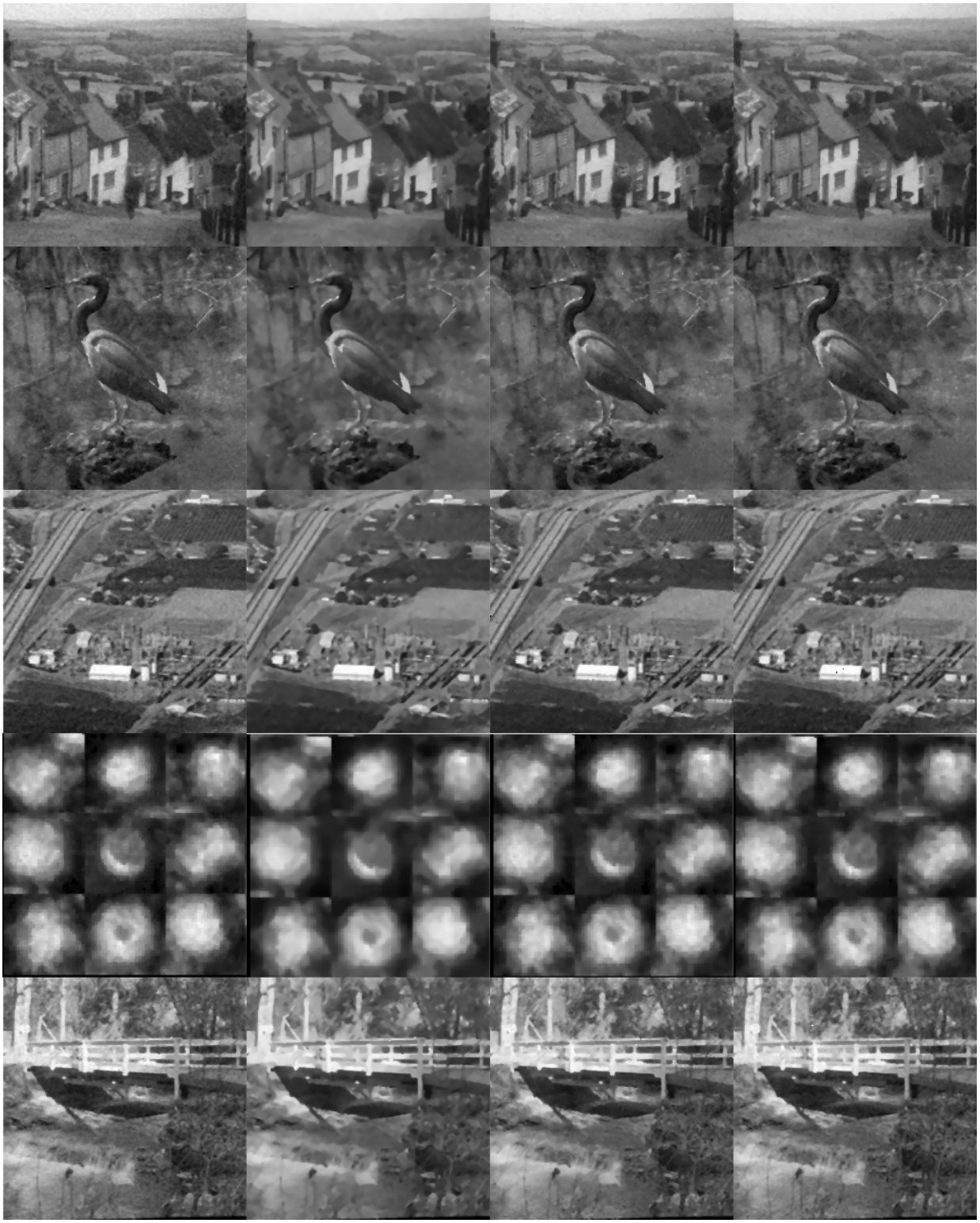

Fig. 6. Visual results for Goldhill, Duck, Plant, Moon and River images. From left to right, we have: Median, Median+NLM, Non-Local Medians and Non-Local Medians KL.