Mixed Gaussian-Impulse Noise Removal from Highly Corrupted Images via Adaptive Local and Nonlocal Statistical Priors

Nasser Eslahi, Hami Mahdavinataj, Ali Aghagolzadeh
Department of Electrical and Computer Engineering
Babol University of Technology, Babol, Iran
Email: {nasser.eslahi, mahdavinataj}@stu.nit.ac.ir, aghagol@nit.ac.ir

Abstract

The motivation of this paper is to introduce a novel framework for the restoration of images corrupted by mixed Gaussian-impulse noise. To this aim, first, an adaptive curvelet thresholding criterion is proposed which tries to adaptively remove the perturbations appeared during denoising process. Then, a new statistical regularization term, called joint adaptive statistical prior (JASP), is established which enforces both the local and nonlocal statistical consistencies, simultaneously, in a unified manner. Furthermore, a novel technique for mixed Gaussian plus impulse noise removal using JASP in a variational scheme is developed—we refer to it as De-JASP. To efficiently solve the above variational scheme, an efficient alternating minimization algorithm is developed based on split Bregman iterative framework. Extensive experimental results manifest the effectiveness of the proposed method comparing with the current state-of-the-art methods in mixed Gaussian-impulse noise removal.

Index Terms

Mixed Gaussian-impulse noise removal, Adaptive curvelet thresholding, nonlocal self-similarity.

I. INTRODUCTION

Image denoising as a fundamental problem in image restoration has been extensively studied in image processing and computer vision (see [1]-[12] for few various representative works). In this paper, we address the problem of image restoration from mixed Gaussian-impulse noise data. Suppose the clean image \( u \in \mathbb{R}^n \) is corrupted by both additive white Gaussian noise (AWGN) of \( \eta \in \mathbb{R}^n \) (with variance \( \sigma^2 \)) and an impulse noise. Also, let \( N_{imp} \) represent the process of image degradation with impulse noise. Accordingly, the overall degradation by mixed Gaussian-impulse noise can be modeled as the following:

\[
f = N_{imp}(u + \eta),
\]

where \( f \in \mathbb{R}^n \) is the observed image (noisy measurements). In the literature, there are two common types of impulse noise: salt-and-pepper (SP) noise and random-valued (RV) noise. Let denote \( u_{a,b} \) as the pixel value of image \( u \) at location \((a, b)\in I\), and \([d_{\text{min}}, d_{\text{max}}]\) be the dynamic range of \( u \). The operator \( N_{imp}(\cdot) \) is defined as follows:

\[
\tilde{f}_{a,b} = N_{imp}(u_{a,b}) = \begin{cases} 
  d_{a,b}, & \text{with probability } r/2, \\
  d'_{a,b}, & \text{with probability } r/2, \\
  u_{a,b}, & \text{with probability } 1 - r,
\end{cases}
\]

(2)
where \( d_{a,b} \) and \( d'_{a,b} \) are in the range of \([d_{\text{min}}, d_{\text{max}}]\) and independent from \( u_{a,b} \). When \( d_{a,b} = d_{\text{min}} \) and \( d'_{a,b} = d_{\text{max}} \), the model is SP noise; however, if \( d_{a,b} = d'_{a,b} \) is identically and uniformly distributed random numbers in \([d_{\text{min}}, d_{\text{max}}]\) with probability \( r \), the model is called RV noise. Median-type filters, e.g. adaptive median filter (AMF) [13] and adaptive center-weighted median filter (ACWMF) [14], are widely used in the literature to remove impulse noise.

Let \( N \) denote the set of locations of noisy pixel candidates corrupted by impulse noise, and accordingly, the set of pixels that are likely to be uncorrupted by impulse noise is defined as \( A = I / N \). Therefore, the degradation matrix \( \Phi \) can be defined as \( \Phi_{a,b} := \{1 : (a,b) \in A \text{ or } 0 : (a,b) \in N\} \). More precisely, the pixels probably being corrupted by impulse noise are assigned with 0 while the rest of pixel values are more likely to be corrupted by AWGN. Delving further into this issue reveals this fact that, \( \Phi \circ f \) follows an approximate Gaussian distribution, where \( \circ \) denotes the element-wise multiplication between two matrices. Also, due to the \textit{ill-posed} nature of (1), regularization-based methods are often used to convert (1) into a \textit{well-posed} problem by imposing \textit{a priori} information of the underlying signal [7], [9]-[12]. Hence, from the Bayesian inference, the following regularization-based framework for mixed Gaussian-impulse noise removal is written as:

\[
\min_u \frac{1}{2} \| \Phi \circ (f - u) \|_2^2 + \lambda J(u),
\]

where the first term is a penalty that represents the closeness of the solution to the observed scene; the second term is a regularization term which represents a priori information of the original scene; and \( \lambda \) is a regularization parameter that balances the contribution of both terms. The term \( J(u) \) could be various choices, e.g., Tikhonov regularization [15], Geman and Reynolds’ half quadratic variational model [16], Rudin, Osher and Fatemi’s total variation model [17], Mumford-Shah model [18], and framelet based model [7]. These approaches do not need to find the damaged pixels and perform well in impulse noise removal. However, for the case of images corrupted by mixed Gaussian impulse noise, the Gaussian noise is not treated properly [12]. In recent works, the local smoothness and the nonlocal self-similarity properties (or image sparsity prior) exhibited in natural images are combined into the final cost functional of image restoration solution to achieve better performance, e.g., see [10]-[12].

The main contributions of this paper are listed as follows. First, we propose an \textit{adaptive curvelet thresholding} (ACT) criterion, in a statistical manner, for adaptively characterizing the perturbations appeared during noise removal process—exploiting local consistency. Second, we establish a new statistical regularization term, called \textit{joint adaptive statistical prior} (JASP) which enforces both the local and nonlocal statistical consistencies, simultaneously, in a unified manner. Third, we propose a novel technique for mixed Gaussian plus impulse noise removal using JASP in a variational scheme, we refer to it as De-JASP. To solve the above variational scheme, an efficient alternating minimization algorithm is developed based on split Bregman iterative framework [19].

The rest of this paper is organized as follows. Section II provides a brief background on a recently proposed modeling for nonlocal self-similarity. The proposed ACT is introduced in Section III. The proposed regularization term, along with its incorporation into the variational framework of mixed noise removal, and implementation details are introduced in Section IV. Numerical results are given in Section V and finally, Section VI concludes the paper.

\section{II. Background on Nonlocal Statistical Modeling (NLSM)}

Motivated by the success of BM3D [4] in image denoising, Zhang \textit{et al}. [11] proposed a model for nonlocal self-similarity prior information, which is employed efficiently in image restoration. The NLSM

\footnote{In this paper, similar to [5], [7]-[12] and for the sake of a fair comparison, \( \Phi \) is determined by AMF and ACWMF for SP and RV noise detection, respectively.}
Fig. 1: (a)-(b) The histograms of two successive orientations at one scale of the curvelet coefficients of the clean image “Fence” (which is shown in Fig. 2(a)). (c)-(d) The histograms of the same orientations at the same scale of the curvelet coefficients of the initiated image $u_{\text{init}}$ (which is shown in Fig. 2(c)), where, it is reconstructed by employing AMF on noisy image of Fig. 2(b). (e)-(f) PDFs of noise distribution (empirical perturbations) in natural log domain at the same orientations and scale (corresponding to the histograms of Figs. 1(a)-1(d)) along with their best fitted PDFs: the noise distribution (blue), the Gaussian fit (red), and the Laplacian fit (black).

explores the nonlocal self-similarity by means of the distribution of the transform coefficients, which are obtained by transforming the 3D array generated by stacking similar image patches. To elucidate on, First, the image $u$ (of size $\sqrt{n} \times \sqrt{n}$) is divided into $P$ overlapped patches of equal sizes, i.e., $u_{p,l'}$ (of size $\sqrt{B_p} \times \sqrt{B_p}$) at location $l'$, where $l' = 1, 2, \cdots, P$. Then, for each patch the best matched patches are found within a searching window (of size $\omega \times \omega$). Next, these similar patches are stacked into a 3D array, $G_{u_{p,l'}}$, which is called a group. Now, a 3D transform, $T_{3D}^3(\cdot)$, is conducted on the group to obtain its transform coefficients, followed by arranging them in a lexicographic order. The mathematical formulation of the NLSM for self-similarity in transform domain can be written as:

$$\Psi_{\text{NLSM}}(u) = \|\Theta_u\|_{\ell_1} = \sum_{l'=1}^{P} \|T_{3D}(G_{u_{p,l'}})\|_{\ell_1},$$

where $\Psi_{\text{NLSM}}(\cdot)$ corresponds to the NLSM operator. To put it simply and briefly, after obtaining $\Theta_u$, the new estimate of $u$ is achieved by $\hat{u} = \Omega_{\text{NLSM}}(\Theta_u)$, where $\Omega_{\text{NLSM}}(\cdot)$ corresponds the inverse operator of $\Psi_{\text{NLSM}}$.

III. THE PROPOSED ADAPTIVE CURVELET THRESHOLDING (ACT)

The curvelet transform [20] is a multi-scale directional transform which provides a new multi-resolution representation with several features that are superior to existing representation, i.e., wavelets and steerable pyramids, thereby gaining good performance in image denoising [1], [21]. For these reasons, here, an adaptive denoising/shrinkage procedure in curvelet domain via estimation of clean curvelet coefficients from available rough estimated data is proposed.

We assume that the initially estimated image $u_{\text{init}}$ achieved by the median filter (AMF or ACWMF) is perturbed and model that perturbation by additive noise, with no prejudging on being AWGN or any specific species of noise. To elucidate on, this modeling of perturbations by additive noise helps us to model the noise probability density function (PDF) effectively and being able to solve the ensuing optimization problem efficiently. Suppose, the initiated image coefficients in curvelet domain ($\hat{\Theta} = \Psi_{u_{\text{init}}}$) can be

\[3\] In matrix notation, we can write $\hat{\Theta} = \Psi u$, and $u = \Psi^\dagger \hat{\Theta}$, where $\hat{\Theta}$ is the curvelet coefficient of function $u$. $\Psi$ is a forward curvelet transform and $\Psi^\dagger = (\Psi^T \Psi)^{-1} \Psi^T$ is its Moore-Penrose pseudo-inverse transform (corresponding to the minimal dual synthesis frame). Due to the Parseval tight frame property of curvelet transform (i.e., $\Psi^T \Psi = I$), we have $\Psi^\dagger = \Psi^T$. 


written as:
\[ \hat{\vartheta}_{j,l,k} = \vartheta_{j,l,k} + n_{j,l,k}, \]
where \( \hat{\vartheta} \) and \( \vartheta \) are the initiated image curvelet coefficient (available rough coefficient which is assumed to be noisy) and the noise-free (or clean) curvelet coefficients, respectively, and \( n \) is unknown additive noise (or perturbation) vector. Also, the three indices of \( j,l,k \) are curvelet triple index of scale, rotation and position, respectively. The problem of image denoising in curvelet domain can be expressed as estimation of clean coefficients from noisy data with Bayesian estimation techniques. Either minimum mean-squared error (MMSE) or maximum a posteriori (MAP) estimator is used for solving this problem, the solution requires a prior knowledge about the distribution of curvelet coefficients. More specifically, these methods are optimized w.r.t. the marginal statistics of the coefficients within each sub-band, by imposing a prior distribution on the clean transform coefficients \[6\].

Figs. 1(a)-1(b) show the histograms of two successive orientations at one scale of the curvelet coefficients of the clean image “Fence” \((256 \times 256)\) —see Fig. 2(a). Obviously, it can be observed that the distributions are characterized by a very sharp peak at zero amplitude and extended tails to both sides of the peak \((\text{leptokurtic distribution})\). This leptokurtic property is observed in all histograms of each scale and orientation of all test images. In this paper, Laplacian distribution is chosen to model the marginal distributions of curvelet coefficients of the clean image, i.e., \( \Pr_\vartheta(\vartheta_{j,l,k}) = (1/\sqrt{2\pi}\sigma_{j,l,k}) \cdot \exp \left(-\sqrt{2}|\vartheta_{j,l,k}|/\sigma_{j,l,k}\right) \), where \( \sigma_{j,l,k} \) is the standard deviation of clean coefficients. This choice is despite the fact that marginal distributions of curvelet coefficients have significantly heavier tails than a Laplacian distribution; tend to go zero slower than a Gaussian distribution; and can be well modelled by a hyper-Laplacian distribution. But, actually, this choice makes a trade-off between modeling the coefficients statistics accurately and being able to solve the ensuing optimization problem efficiently.

The histograms of the same two successive orientations at the same scale of the coefficients of the initially estimated image \( u_{\text{init}} \) —see Fig. 2 (c)—are shown in Figs. 1(c)-1(d). More specifically, these histograms are obtained using the curvelet coefficients of the initiated image \( u_{\text{init}} \), which is achieved by the initial median filtering step. As can be seen, here, the leptokurtic behavior and heavy tails do not exist in the shown distributions. More accurately, these two properties do not exist in all distributions of each scale and orientation of all initiated images achieved by this step.

Considering (5), the noise distribution can be obtained using the curvelet coefficients of the clean image and the initiated image. Figs. 1(e)-1(f) show the PDFs of noise distribution at the same two successive orientations of the same scale corresponding to those of Figs. 1(a)-1(d). It can be observed that the empirical distributions of noise coefficients \( n_{j,l,k} \) can be well characterized by Gaussian distributions, while the Laplacian distributions have much larger fitting errors. This observation motivates us to model the PDF of \( n_{j,l,k} \) by a standard Gaussian distribution: \( \Pr_n(n_{j,l,k}) = (1/\sqrt{2\pi}\sigma_{n,j,l}) \cdot \exp \left(-n_{j,l,k}^2/2\sigma_{n,j,l}^2\right) \), where \( \sigma_{n,j,l}^2 \) is the noise variance in curvelet domain, which can be estimated using robust median estimator \[22\] and Monte-Carlo simulations \[20\].

Consider the marginal distribution of the clean curvelet coefficients. Since in image denoising problems, we do not have access to the clean image, the variance of the clean image curvelet coefficients \( \sigma_{j,l,k}^2 \) can be estimated by maximum likelihood (ML) estimator, through averaging over the neighboring initiated image coefficients (available \( \hat{\vartheta}_{j,l,k} \)) within a squared window \[23\].

In order to estimate \( \vartheta_{j,l,k} \) from the available observation \( \hat{\vartheta}_{j,l,k} \), the MAP estimator is employed:
\[ \vartheta_{j,l,k} = \arg \max_{\vartheta_{j,l,k}} \left\{ \Pr_{\vartheta|\hat{\vartheta}}(\vartheta_{j,l,k}|\hat{\vartheta}_{j,l,k}) \right\}. \]
By considering (5) and using Bayes’ rule, we get:

\[
\theta_{j,l,k} = \arg \max_{\hat{\theta}_{j,l,k}} \left\{ \Pr_{\hat{\theta}_{j,l,k}} (\hat{\theta}_{j,l,k} | \theta_{j,l,k}) \cdot \Pr_{\theta}(\theta_{j,l,k}) \right\}
\]

\[
= \arg \max_{\hat{\theta}_{j,l,k}} \left\{ \Pr_n (\hat{\theta}_{j,l,k} - \hat{\theta}_{j,l,k}) \cdot \Pr_{\theta}(\theta_{j,l,k}) \right\}.
\]

Therefore, (7) allows us to write this estimation in terms of the PDF of the noise coefficient (Prₙ), and the PDF of the clean image coefficient (Prₜ). By substituting Prₙ and Prₜ into (7) and doing some manipulations, the estimation of \( \theta_{j,l,k} \) is achieved as:

\[
\theta_{j,l,k} = \text{sign}(\hat{\theta}_{j,l,k}) \max\left(0, |\hat{\theta}_{j,l,k}| - \frac{\sqrt{2} \sigma_{jn,k}^2}{\sigma_{j,l,k}} \right),
\]

where, in fact, (8) is the classical soft shrinkage function [24] defined as \( S_{\rho}(y) := \text{sgn}(y) \cdot \max(0, |y| - \rho) \).

Not only can the proposed ACT applied for the initiated image achieved by the initial median filtering step, but also it can be employed in further iterations—by using the last approximated image achieved in previous iteration as the noisy image for the current iteration—which leads to a simple and effective mixed Gaussian-impulse noise removal scheme, e.g., see Fig. 2(d).

IV. THE PROPOSED IMAGE DE_NOISING VIA JASP (DE-JASP)

The proposed joint adaptive statistical prior (JASP) is defined by integrating both the local information prior (which depicts the local smoothness and geometric regularity of image structures achieved by discrete curvelet transform (DCuT) [20]) and the nonlocal self-similarity prior (corresponding to the NLSM in 3D transform domain achieved by the method introduced in Section II). In a mathematical expression, the proposed JASP is expressed as:

\[
\mathcal{J}_{\text{JASP}}(u) = \Psi_{\text{DCuT}}(u) + \tau \Psi_{\text{NLSM}}(u)
\]

\[
= \|\Psi u\|_{\ell_1} + \tau \|\Theta u\|_{\ell_1},
\]

where the first and the second terms represent the image local smoothness prior and nonlocal self-similarity prior, respectively. Also, \( \tau \) is a regularization parameter which controls the trade-off between two competing (statistical) terms.
By substituting the proposed JASP (9) for the regularization term $J(u)$ in the regularization-based framework of (3), and by introducing two auxiliary variables $\vartheta$ and $w$, the proposed optimization problem for image recovery is expressed as follow:

$$
\min_{u, \vartheta, w} \left\{ \frac{1}{2} \| \Phi \circ (f - u) \|^2_{\ell_2} + \lambda \| \vartheta \|_{\ell_1} + \tau' \| \Theta_w \|_{\ell_1} \right\}
$$

s.t. $\vartheta = \Psi u$, $u = w$, $w = \Omega_{NLSM}(\Theta_w)$.

(10)

In order to solve the above minimization problem, an alternating split Bregman iterative algorithm [19] is invoked. We finally obtain the following schemes:

$$(u^{k+1}, \vartheta^{k+1}, w^{k+1}) \leftarrow \arg \min_{u, \vartheta, w} \left\{ \frac{1}{2} \| \Phi \circ (f - u) \|^2_{\ell_2} + \lambda \| \vartheta \|_{\ell_1} + \tau' \| \Theta_w \|_{\ell_1} + \frac{\mu_1}{2} \| \vartheta - \Psi u - b^k \|^2_{\ell_2} + \frac{\mu_2}{2} \| u - w - c^k \|^2_{\ell_2} \right\}$$

(11a)

$$b^{k+1} \leftarrow b^k - \vartheta^{k+1} + \Psi u^{k+1}$$

(11b)

$$c^{k+1} \leftarrow c^k - u^{k+1} + w^{k+1}.$$  

(11c)

Here, $\mu_1$ and $\mu_2$ are fixed value parameters for improving the numerical stability of the algorithm.

Given $\vartheta^k$ and $w^k$, the $u$ sub-problem of (11a) consists of minimizing a strictly convex quadratic function that can be solved easily, i.e., $u^{k+1} = \left( \Phi^T \Phi + \mu I \right)^{-1} \circ D$, where $\mu = \mu_1 + \mu_2$ and $D = \left( \mu_1 \Psi^T (\vartheta^k - b^k) + \mu_2 (w^k + c^k) + \Phi^T \Phi \circ f \right)$. In order to avoid computing the matrix inversion in the first term, the matrix inversion lemma of the Sherman-Morrison formula is applied to that term. Thus, the $u$ sub-problem can be obtained as:

$$u^{k+1} = \frac{1}{\mu} \left( I - \frac{\Phi^T \Phi}{\mu + \Phi \Phi^T} \right) \circ D.$$  

(12)

By given $u$ in hand (according to (12)) and considering $\hat{\vartheta} = \Psi u + b$ (for simplicity, the superscript $k$ is dropped without confusion), the $\vartheta$ sub-problem of (11a) becomes:

$$\min_{\vartheta} \frac{1}{2} \| \vartheta - \hat{\vartheta} \|^2_{\ell_2} + \lambda \| \vartheta \|_{\ell_1}.$$  

(13)

By these transformations, we regard $\vartheta'$ as some type of the noisy unknown coefficient $\vartheta$. More precisely, the sub-problem of (13) can be interpreted as the denoising in curvelet coefficient domain. Considering the fact that the unknown variable $\vartheta$ is component-wise separable in curvelet domain, each of its component $\vartheta_{j,l,k}$ can be obtained by a component-wise shrinkage procedure:

$$\vartheta_{j,l,k} = \arg \min_{\vartheta_{j,l,k}} \left\{ \frac{1}{2} | \vartheta_{j,l,k} - \vartheta_{j,l,k} |^2 + \frac{\lambda_{j,l,k}}{\mu_1} | \vartheta_{j,l,k} | \right\},$$  

(14)

where its solution is the same as the one introduced in (8) with regarding $\frac{\lambda_{j,l,k}}{\mu_1} = \frac{\sqrt{2} \sigma_{j,l,k}}{\sigma_{j,l,k}}$. To elucidate on, the $\vartheta$ sub-problem of (13) can be interpreted as the proposed ACT introduced in Section III. Thus, the parameter $\lambda$ is self-adaptive.

Given $u$, analogously, the $w$ sub-problem of (11a) can be written as:

$$\min_{w} \frac{1}{2} \| w - z \|^2_{\ell_2} + \frac{\tau'}{\mu_2} \| \Theta_w \|_{\ell_1},$$  

(15)

where $z = u - c$ (for simplicity the superscript $k$ is omitted without confusion). Owing to the complicated definition of $\Theta_w$, it seems difficult to solve (15) directly. In order to solve (15) amendable, according to
TABLE I: THE PROPOSED DE-JASP FRAMEWORK

| Input: | $f$, $B_p$, $\omega$, $\sigma'$, $\mu_1$, $\mu_2$, $Tol$, $k_{max}'$: maximum iteration number |
| Output: | $u^*$: Recovered image; |
| Initialization: | Set $k = 0$; $(b^0, c^0) = (0, 0)$ |
| $u_{init} = u^0$: = median-filtering ($f$); compute $\Phi$ |
| While a stop criterion is not satisfied do |
| 1. | $u^{k+1} = \bar{u} = u^k$, $\hat{\omega}^{k+1} = \Psi u^{k+1} + b^k$, $z^{k+1} = u^{k+1} - c^k$ |
| 2. | solve $\omega$ sub-problem using (14) and concatenating all coefficients |
| 3. | solve $w$ sub-problem using (19) |
| 4. | solve $u$ sub-problem using (12) |
| 5. | compute $s^{k+1} = \text{SSIM}(u^{k+1}, \bar{u})$ |
| 6. | compute $\text{diff} = \text{abs}(s^{k+1} - s^k)$ |
| 7. | update $b^k$, $c^k$ using (11b) and (11c), respectively |
| 8. | $k \leftarrow k + 1$ |
| end While |
| Stopping criterion: | $k = k_{max}'$ or $\text{diff} \leq Tol$. |

[11], a reasonable assumption is used which leads to obtain a closed-form solution of (15). By assuming that all elements of $w - z$ (such that $w, z \in \mathbb{R}^n$) are independent and identically distributed (i.i.d.) with zero-mean and variance $\sigma'^2$, and by invoking the law of large numbers in probability theory, for any $\epsilon' > 0$, it leads to:

$$\lim_{n \to \infty} \text{Pr}\left\{ \frac{1}{n} \| w - z \|_2^2 \sigma'^2 \sigma'^2 \leq \frac{\epsilon'}{2} \right\} = 1.$$  \hspace{1cm} (16)

Due to the orthogonal property of 3D transform $T_{3D}$, and denoting the error vector in 3D transform domain by $\Theta_e = \Theta_w - \Theta_z$ (such that $\Theta_w, \Theta_z \in \mathbb{R}^{K_{\Theta}}$, where $K_{\Theta} = B_p * c * P$), it is concluded that all elements of $\Theta_e$ are i.i.d. with zero-mean and variance $\sigma'^2$. Thanks to the law of large numbers, by doing the same manipulations with (16) to $\Theta_e$, for any $\epsilon' > 0$, it yields:

$$\lim_{K_{\Theta} \to \infty} \text{Pr}\left\{ \frac{1}{K_{\Theta}} \| \Theta_w - \Theta_z \|_2^2 \sigma'^2 \sigma'^2 \leq \frac{\epsilon'}{2} \right\} = 1.$$  \hspace{1cm} (17)

Therefore, according to (16) and (17) the following relationship is described (almost surely): $\frac{1}{n} \| w - z \|_2^2 = \frac{1}{K_{\Theta}} \| \Theta_w - \Theta_z \|_2^2$, where incorporating it into (15) leads to:

$$\arg \min_{\Theta} \frac{1}{2n} \| \Theta_w - \Theta_z \|_2^2 + \frac{K_{\Theta} \sigma'}{\mu_2 n} \| \Theta_w \|_{\ell_1}.$$  \hspace{1cm} (18)

Since the unknown variable $\Theta_w$ is component-wise separable in (18), each of its component $\Theta_w(j'), j' = 1, \cdots, K_{\Theta}$ can be independently obtained by a component-wise (soft) shrinkage procedure. Thus, the closed-form solution of (15) is written as:

$$w = \Omega_{\text{NLSM}}(\Theta_w) = \Omega_{\text{NLSM}}(S_{K_{\Theta} \sigma'}(\Theta_w)).$$  \hspace{1cm} (19)

The proposed De-JASP algorithm is delineated in Table I.

---

3It is worth emphasizing that the above assumption does not need to be Gaussian, Laplacian or generalized Gaussian process.
TABLE II: PSNR/SSIM COMPARISONS FOR GAUSSIAN PLUS SALT-AND-PEPPER NOISE REMOVAL.

| Test Image | r_gen | σ   | Noisy Image | Init. Image | IFASDA [7] | JSM [11] | De-JASP |
|------------|------|-----|-------------|-------------|------------|---------|---------|
| Barbara    | 0.2  | 20  | 12.10/0.147 | 22.48/0.529 | 24.92/0.756 | 27.73/0.795 | 28.06/0.835 |
|            | 0.3  | 30  | 11.69/0.133 | 19.57/0.385 | 20.46/0.514 | 22.76/0.537 | 24.83/0.676 |
| Boat       | 0.2  | 20  | 11.44/0.134 | 23.08/0.510 | 27.81/0.771 | 28.40/0.784 | 29.22/0.833 |
|            | 0.3  | 30  | 11.72/0.122 | 19.93/0.373 | 22.74/0.506 | 23.22/0.520 | 24.83/0.676 |
| Fence      | 0.2  | 20  | 11.77/0.122 | 21.58/0.560 | 24.99/0.782 | 26.60/0.791 | 26.79/0.812 |
|            | 0.3  | 30  | 11.39/0.197 | 19.13/0.443 | 21.03/0.577 | 22.35/0.583 | 24.15/0.693 |
| Parrot     | 0.2  | 20  | 11.89/0.095 | 23.15/0.407 | 27.78/0.759 | 28.83/0.765 | 29.65/0.837 |
|            | 0.3  | 30  | 11.50/0.087 | 19.97/0.277 | 22.65/0.424 | 23.51/0.438 | 26.21/0.640 |

V. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of the proposed method—De-JASP. The performances of our experiments are evaluated on 4 gray-scale test images of Barbara, Boat, Fence, and Parrot (of size 256 × 256). To evaluate our simulation results, we use two applicable quality assessors, the peak signal-to-noise ratio (PSNR) in dB and the structural similarity (SSIM) [25]. All the experimental results reported in this paper are averaged over 10 independent trials. Also, the second generation of DCuT via wrapping [20] is employed in our implementations. To demonstrate the effectiveness of the proposed method in image denoising, we have compared it with two competitive mixed noise removal techniques: (i) iterative framelet-based approximation/sparsity deblurring algorithm IFASDA [7], and (ii) image denoising via joint statistical modeling (JSM) [11]. It is worth emphasizing that these two recently proposed denoising methods are among the state-of-the-art techniques in mixed-noise removal which achieve the best performance so far. Note that, to make a fair comparison among the competing methods, we have carefully tuned their parameters to achieve the best performance. Also, for the sake of fair comparison, the same test conditions are used in all experiments; i.e., the same noisy measurements are applied for each method. Due to the space limitations, only parts of the experimental results are shown in this paper. Interested readers may contact the corresponding author for all the images and the source code.

Table II lists the PSNR and SSIM values of the recovered noisy test images obtained by the proposed De-JASP compared to those obtained by JSM and IFASDA in different noise level hypotheses (best results are emphasized in bold face). From Table II, it can be inferred that the performance of the proposed De-JASP is superior to those of compared algorithms in terms of both employed objective quality assessors. The average PSNR (SSIM) gain of De-JASP over JSM and IFASDA, in mixed AWGN+RV noise removal, can be up to 2.31 dB (0.154) and 3.55 dB (0.167), respectively. For visual comparison, some recovered images by the competing methods, in different noise levels, are shown in Figs. 4-7. It can be observed that De-JASP outperforms the other competing methods and reproducing clearer images. Better visual comparison can be made by zooming the images on the screen.
Fig. 4: Visual quality comparison of mixed Gaussian-impulse noise removal on image “Barbara”. (a) noisy image corrupted by AWGN+SP noise with $\sigma = 20$ and $r_{sp} = 20\%$. Denoised results by (b) IFASDA [7]; (c) JSM [11]; and (d) the proposed De-JASP. For numerical comparison, see Table II.

Fig. 5: Visual quality comparison of mixed Gaussian-impulse noise removal on image “Boat”. (a) noisy image corrupted by AWGN+SP noise with $\sigma = 10$ and $r_{sp} = 60\%$. Denoised results by (b) IFASDA [7] (PSNR=27.58 dB, SSIM=0.769); (c) JSM [11] (PSNR=28.06 dB, 0.793); and (d) the proposed De-JASP (PSNR=28.37 dB, SSIM=0.814).

Fig. 6: Visual quality comparison of mixed Gaussian-impulse noise removal on image “Parrot”. (a) noisy image corrupted by AWGN+RV noise with $\sigma = 20$ and $r_{rv} = 20\%$. Denoised results by (b) IFASDA [7] (PSNR=26.34 dB, SSIM=0.726); (c) JSM [11] (PSNR=26.96 dB, 0.789); and (d) the proposed De-JASP (PSNR=27.42 dB, SSIM=0.816)
VI. Conclusion

In this paper, first, an adaptive curvelet thresholding (ACT) criterion is introduced to adaptively characterize and abate the perturbations which are appeared during denoising process. Then, a novel joint adaptive statistical prior (JASP) and a new strategy for mixed Gaussian-impulse noise removal via JASP, called De-JASP, is proposed. Extensive experimental results clearly confirm that De-JASP significantly outperforms the current state-of-the-art mixed Gaussian-impulse noise removal techniques.

For future work, we would like to propose an effective nonlocal self-similarity modeling, and pursue this direction to extend our proposed statistical prior to other image restoration applications such as image deblurring and inpainting.

ACKNOWLEDGMENT

The authors would like to express their gratitude to Prof. J. Fadili (CNRS-ENSICAEN-Université de Caen) and Dr. J. Zhang (Peking University) for many fruitful discussions. They also would like to thank the authors of [7] and [11] for sharing the source code of their papers used in Section V.

REFERENCES

[1] J. L. Starck, E. J. Candès, and D. L. Donoho, “The curvelet transform for image denoising,” IEEE Trans. Image. Process., vol. 11, no. 6, pp. 670-684 Jun. 2002.
[2] A. Buades, B. Coll, and J. M. Morel, “A review of image denoising algorithms, with a new one,” SIAM Multiscale Model. Sim., vol. 4, no. 2, pp. 490-530, 2005.
[3] M. Elad and M. Aharon, Image denoising via sparse and redundant representations over learned dictionaries,” IEEE Trans. Image Process., vol. 15, no. 12, pp. 3736-3745, Dec. 2006.
[4] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3D transform-domain collaborative filtering,” IEEE Trans. Image Process., vol. 16, no. 8, pp. 2080-2095, Aug. 2007.
[5] J. F. Cai, R. H. Chan, and M. Nikolova, “Two-phase approach for deblurring images corrupted by impulse plus Gaussian noise,” Inverse Probl. Imag., vol. 2, no. 2, pp. 187-204, 2008.
[6] H. Rabbani and S. Gazor, “Image denoising employing local mixture models in sparse domains,” IET Image Process., vol. 4, no. 5, pp. 413-428, Oct. 2010.
[7] Y. R. Li, L. Shen, D. Q. Dai, and B. W. Suter, “Framelet algorithms for de-blurring images corrupted by impulse plus Gaussian noise,” IEEE Trans. Image Process., vol. 20, no. 7, pp. 1822-1837, Jul. 2011.
[8] J. Liu, X. C. Tai, H. Huang, and Z. Huan, “A weighted dictionary learning model for denoising images corrupted by mixed noise,” IEEE Trans. Image Process., vol. 22, no. 3, pp. 1108-1120, Mar. 2013.
[9] M. Yan, “Restoration of images corrupted by impulse noise and mixed Gaussian impulse noise using blind inpainting,” SIAM J. Imaging Sci. vol. 6, no. 3, pp. 1227-1245, Jul. 2013.
[10] J. Jiang, L. Zhang, and J. Yang, “Mixed noise removal by weighted encoding with sparse nonlocal regularization,” IEEE Trans. Image Process., vol. 23, no. 6, pp. 2651-2662, Jun. 2014.
[11] J. Zhang, D. Zhao, R. Xiong, S. Ma, and W. Gao, “Image restoration using joint statistical modeling in space-transform domain,” IEEE Trans. Circuits Syst. video Technol., vol. 24, no. 6, pp. 915-928, Jun. 2014.
[12] J. Jiang, J. Yang, Y. Cui, W. K. Wong, and Z. Lai, “Sparse nonlocal priors based two-phase approach for mixed noise removal,” Signal Process., vol. 116, pp. 101-111, Nov. 2015.
[13] H. Hwang and R. Haddad, “Adaptive median filters: new algorithms and results,” IEEE Trans. Image Process, vol. 4, no. 4, pp. 499-502, Apr. 1995.
[14] S. J. Ko and Y. H. Lee, “Center weighted median filters and their applications to image enhancement,” IEEE Trans. Circuits Syst. video Technol., vol. 38, no. 9, pp. 984-993, Sep. 1991.
[15] A. N. Tikhonov and V. Y. Arsenin, Solutions of ill-posed problems, Washington, D.C., USA: V. H. Winston & Sons, 1977.
[16] D. Geman and G. Reynolds, “Constrained restoration and the recovery of discontinuities,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 14, no. 3, pp. 367-383, Mar. 1992.
[17] L. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” Phys. D., vol. 60, pp. 259-268, Nov. 1992.
[18] D. Mumford and J. Shah, “Optimal approximation by piecewise smooth functions and associated variational problems,” Comm. Pure Appl. Math., vol. 42, pp. 577-685, Jul. 1989.
[19] T. Goldstein and S. Osher, “The split Bregman method for $\ell_1$ regularized problems,” SIAM J. Imag. Sci., vol. 2, no. 2, pp. 323-343, Apr. 2009.
[20] E. Candès, L. Demanet, D. Donoho, and L. Ying, “Fast discrete curvelet transforms,” Multiscale Model. Simul., vol. 5, no. 3, pp. 861-899, Jan. 2006.
[21] B. Zhang, J. Fadili, and J. Starck, “Wavelets, ridgelets, and curvelets for Poisson noise removal,” IEEE Trans. Image Process., vol. 17, no. 7, pp. 1093-1108, Jul. 2008.
[22] D. L. Donoho and I. M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage,” Biometrika, vol. 81, pp. 425-455, Aug. 1994.
[23] M. K. Mihăilesc˘u, I. Kozintsev, K. Ramchandran, and P. Moulin, “Low-complexity image denoising based on statistical modeling of wavelet coefficients,” IEEE Signal Process. Lett., vol. 6, no. 12, Dec. 1999.
[24] D. L. Donoho, “De-noising by soft-thresholding,” IEEE Trans. Inf. Theory., vol. 41, no. 3, pp. 613-627, May. 1995.
[25] Z. Wong, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” IEEE Trans. Image Process., vol. 13, no. 4, pp. 600-612, Apr. 2004.