

IMAGE RESTORATION UNDER CAUCHY NOISE WITH SPARSE REPRESENTATION PRIOR AND TOTAL GENERALIZED VARIATION*

Miyoun Jung

Department of Mathematics, Hankuk University of Foreign Studies, Yongin, Korea

Email: mjung@hufs.ac.kr

Myungjoo Kang

Department of Mathematical Sciences, Seoul National University, Seoul, Korea

Email: mkang@snu.ac.kr

Abstract

This article introduces a novel variational model for restoring images degraded by Cauchy noise and/or blurring. The model integrates a nonconvex data-fidelity term with two regularization terms, a sparse representation prior over dictionary learning and total generalized variation (TGV) regularization. The sparse representation prior exploiting patch information enables the preservation of fine features and textural patterns, while adequately denoising in homogeneous regions and contributing natural visual quality. TGV regularization further assists in effectively denoising in smooth regions while retaining edges. By adopting the penalty method and an alternating minimization approach, we present an efficient iterative algorithm to solve the proposed model. Numerical results establish the superiority of the proposed model over other existing models in regard to visual quality and certain image quality assessments.

Mathematics subject classification: 68U10, 65K10.

Key words: Image restoration, Cauchy noise, Sparse representation prior, Dictionary learning, Total generalized variation.

1. Introduction

Image restoration is a fundamental problem in image processing. It refers to estimating the original clean image from an observed image. This work focuses on image denoising and/or deblurring problem in the presence of Cauchy noise. The Cauchy noise is a type of impulsive noise that emerges in atmospheric and underwater acoustic noise, radar and sonar applications, wireless communication systems, biomedical images, and synthetic aperture radar images; see for instance [1–3].

Let $u \in \mathbb{R}^{m \times n}$ be a true 2-dimensional (2D) discrete grayscale image. Assume that the observed image $f \in \mathbb{R}^{m \times n}$ is given by

$$f = Au + \eta, \tag{1.1}$$

where $A : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \times n}$ represents either the identity operator or a blurring operator, defined as $Au = \kappa * u$ with κ as a blurring kernel and $*$ denoting the convolution, and η represents

* Received October 17, 2018 / Revised version received March 3, 2019 / Accepted July 19, 2019 /
Published online February 3, 2020 /

some Cauchy noise. That is, η is a random variable following a Cauchy distribution with the probability density function (PDF) [4, 5] as follows

$$P(\eta) = \frac{\gamma}{\pi(\gamma^2 + (\eta - \delta)^2)}, \quad (1.2)$$

where $\delta \in \mathbb{R}$ is the parameter specifying the location of the peak, and $\gamma > 0$ is the scale parameter that determines the level of noise. Here, we intend to retrieve a clean image u from an observed image f , which is an ill-posed inverse problem.

In recent years, several approaches to eliminating Cauchy noise have been suggested. Chang et al. [6] proposed a recursive restoration algorithm based on a Markov random field (MRF) model driven by Cauchy noise, and the proposed method was shown to provide better edge preservation than similar algorithms using Gaussian or Laplacian noise. In [7], the authors derived a new statistical model in the complex wavelet domain to remove Cauchy noise, from a bivariate *maximum a posteriori* (MAP) estimator. In addition, Loza et al. [8] proposed a statistical image fusion method based on a non-Gaussian distribution in the wavelet domain, which showed a noticeable improvement in terms of fusion quality and noise reduction. In [9], Wan et al. introduced a novel image segmentation method for noisy color images corrupted by Cauchy noise. The authors developed a noise reduction approach in the wavelet domain, relying on the bivariate Cauchy density, that was utilized for segmentation.

In addition to these MRF or wavelet-based denoising methods, a variational model was proposed in [10] to restore images degraded by Cauchy noise. This model involves a total variation (TV) regularization [11] and a nonconvex data-fidelity term derived from the PDF in (1.2). Specifically, assuming that η follows a *zero-centered* Cauchy law ($\delta = 0$), the authors proposed a TV model for restoring images corrupted by Cauchy noise as follows:

$$\min_u \frac{\lambda}{2} \langle \log(\gamma^2 + (Au - f)^2), \mathbf{1} \rangle + \text{TV}(u), \quad (1.3)$$

where $\lambda > 0$ is a tuning parameter, $\langle \cdot, \cdot \rangle$ represents the inner product, and TV denotes the discrete version of the isotropic TV norm: denoting u_s by the pixel value of an image u at location $s = (i, j)$ ($i = 1, \dots, m, j = 1, \dots, n$), $\text{TV}(u) = \|\nabla u\|_1 = \sum_s \sqrt{|(\nabla u)_s^1|^2 + |(\nabla u)_s^2|^2}$, where $\nabla u = [\partial_x u, \partial_y u]^T$ is a discrete gradient operator whose components $\partial_x u$ and $\partial_y u$ are the finite difference operators that estimate the partial derivatives of the image u along the x -axis and y -axis, respectively. The TV enabled the recovery of images with well-preserved structures and important edges. However, due to the deficiency of the nonconvexity, the same authors introduced a convex model by adding a quadratic penalty term, $\|u - \tilde{u}\|_2^2$, into the nonconvex model. This quadratic term contains a pre-denoised image, \tilde{u} , obtained by applying the median filter to the data f . However, the median filtering does not always bring adequate denoising results. Recently, Mei et al. [12] returned to the original nonconvex problem (1.3), and showed the effectiveness of the model combined with the alternating direction method of multipliers (ADMM) [13].

Despite its several benefits, TV regularization has a tendency to produce staircase artifacts in smoothly varying regions, as it favors solutions that are piecewise constant, and thus it smoothes textures and fine details in images. Thus, as one way of ameliorating staircasing effects, higher-order regularization based models were suggested in [14–18]. As an early work, a inf-convolution TV (ICTV) was proposed in [14], which takes the infimal convolution of TV and second-order TV. Moreover, Li et al. [17] proposed a denoising model, involving a convex

combination of TV and second-order TV as a regularizer. On the other hand, as a generalization of the ICTV, the total generalized variation (TGV) regularizer was proposed in [18, 19], and it was shown that TGV regularization restores fairly accurate pixels in homogeneous regions and wipes off the staircase effect.

Moreover, to better preserve textures and small features, wavelet-based methods [20–24] and nonlocal methods [25–28] have been proposed. Although wavelet-based methods better preserve textures, they may exhibit pseudo-Gibbs phenomena and bring artifacts into the recovered image [24, 29]. Moreover, while nonlocal-based approaches take advantage of the similarity of image patches, a low similarity or dissimilarity of the image patches limits their applicability. To cope with these limitations and better take into consideration the structure of the processed image, sparse representation methods via dictionary learning (DL) were proposed and have been widely adopted in image processing. This is due to the fact that natural images can be well approximated by a linear combination of only a few elements (atoms) of a dictionary. Aharon et al. [30] introduced an efficient DL algorithm, called the K-SVD algorithm, to attain sparse signal representations. Moreover, it was implemented in [31] to remove additive Gaussian noise. This denoising method preserved textural patterns and fine features well and has been generalized to handle various image processing problems including image sequence denoising [32–34], deblurring [35, 36], decomposition [37], and restoration problems under different non-Gaussian noise, e.g., multiplicative [38, 39], Poisson [40–42] or impulsive noise [43].

In this work, we introduce a new variational model for recovering images in the presence of Cauchy noise and/or blur. The model utilizes both a patch-based sparse representation prior via DL and a local derivative-based prior, TGV. The sparse representation prior enables efficient denoising in homogeneous regions with preserving textures and details, and TGV further helps removing noise in smooth regions while maintaining edges. In addition, we present an efficient iterative algorithm to solve the proposed model.

This paper is organized as follows: In Section 2, we review some properties of the Cauchy distribution, K-SVD based denoising method, and TGV regularization. In Section 3, we present our proposed model and an optimization algorithm to solve the proposed model. Section 4 exhibits the numerical results of the proposed model, with comparison to other existing methods. Lastly, we conclude our work in Section 5.

2. Preliminaries

2.1. Properties of Cauchy distribution

A continuous random variable with the PDF given in (1.2), where $\delta \in \mathbb{R}$ and $\gamma > 0$ are parameters, is said to have a Cauchy distribution. The Cauchy distribution is unimodal and symmetric about the point δ , which is its mode and median. The scale parameter γ specifies the half-width at half-maximum, and is also equal to half the interquartile range.

The Cauchy distribution looks similar to a Gaussian distribution with a bell-shaped curve. However, unlike Gaussian distributions, Cauchy distributions are heavy-tailed distributions. Denote the Cauchy distribution by $\mathcal{C}(\delta, \gamma)$ and let $\mathcal{N}(\mu, \sigma^2)$ be the Gaussian distribution with mean $\mu \in \mathbb{R}$ and standard deviation $\sigma > 0$. Figure 2.1 (a) exhibits the standard distributions $\mathcal{C}(0, 1)$ and $\mathcal{N}(0, 1)$, which shows that the Cauchy distribution has thicker tails than the Gaussian distribution. The tail heaviness of the Cauchy distribution is determined by the scale parameter γ . In Figure 2.1 (b), comparing PDFs corresponding to $\gamma = 5, 10, 15$, it can be seen that as the value of γ increases, the tails of the bells become heavier. Due to tail heaviness,

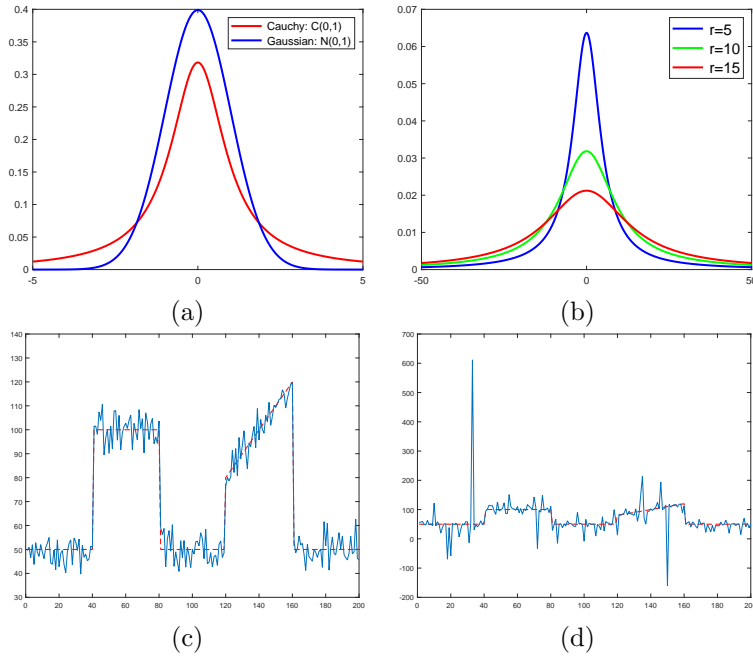


Fig. 2.1. Comparison of Cauchy noise with Gaussian noise. (a) PDFs of Cauchy distribution $\mathcal{C}(0,1)$ (red) and Gaussian distribution $\mathcal{N}(0,1)$ (blue), (b) PDFs of Cauchy distribution $\mathcal{C}(0,\gamma)$ with $\gamma = 5$ (blue), 10 (green), 15 (red), (c) noisy signal degraded by Gaussian noise with $\sigma = 5$, (d) noisy signal degraded by Cauchy noise with $\gamma = 5$.

rare events occur with large probability in the Cauchy distribution. Thus, the noise generated from the Cauchy distribution is more impulsive than the Gaussian one. Figure 2.1 (c) and (d) present noisy signals degraded by the Gaussian noise following $\mathcal{N}(0,5^2)$ and Cauchy noise following $\mathcal{C}(0,5)$ to the original clean signal (red curve), respectively. The vertical scale for the noisy signal corrupted by Gaussian noise is similar to that for the original signal, [30, 130], but it goes from -200 to 700 for the noisy signal distorted by the Cauchy noise.

Lastly, Cauchy distribution is also known to be the distribution of the ratio of two independent, normally-distributed Gaussian random variables. If X and Y are two independent Gaussian distributed random variables with mean 0 and variance 1, then the ratio X/Y follows the standard Cauchy distribution, i.e., $\mathcal{C}(0,1)$ [44, 45]. In practice, noisy images corrupted by Cauchy noise can be generated by applying this property.

2.2. K-SVD based denoising method

The K-SVD [30] is a dictionary learning algorithm that finds a dictionary for sparse representations via a singular value decomposition (SVD).

The K-SVD based denoising method in [31] assumes that image patches admit a sparse representation. More precisely, each image patch of size $\sqrt{N} \times \sqrt{N}$, reformed as a single column vector, can be represented sparsely as a linear combination of atoms (columns) taken from a dictionary $D \in \mathbb{R}^{N \times K}$ ($N \ll K$). Assuming the observed noisy image $f = u + b$, where $b \in \mathbb{R}^{m \times n}$ is the zero-mean Gaussian noise, Elad et al. [31] proposed the following minimization

problem for Gaussian noise removal:

$$\min_{\{\alpha_s\}, D, u} \frac{\lambda}{2} \|f - u\|_2^2 + \sum_{s \in \mathcal{B}} \mu_s \|\alpha_s\|_0 + \sum_{s \in \mathcal{B}} \|D\alpha_s - R_s u\|_2^2, \quad (2.1)$$

where $\|\cdot\|_2^2 = \langle \cdot, \cdot \rangle$, \mathcal{B} denotes the index set where a patch exists in the image, $\alpha_s \in \mathbb{R}^K$ is a coefficient vector, and $R_s \in \mathbb{R}^{N \times mn}$ is a binary matrix that extracts a $\sqrt{N} \times \sqrt{N}$ patch from the image u at location s , so that $R_s u \in \mathbb{R}^N$. The l_0 term, $\|\alpha\|_0$, counts for the number of nonzero entries in the vector α . The values μ_s are patch-specific weights and are hidden parameters determined by the optimization procedure, and $\lambda > 0$ is a tuning parameter.

To realize the model (2.1), the authors proposed an iterative block-coordinate relaxation method that fixes all unknowns except from the one that is to be updated. The K-SVD algorithm iterates the following two stages to obtain the best dictionary D :

- *Sparse coding stage*: Use any pursuit algorithm to compute the representation vector α_s ($s \in \mathcal{B}$) per each patch $R_s u$.
- *Dictionary update stage*: For each column $\ell = 1, \dots, K$ in D , select the set of patches that use this atom, $\omega_\ell = \{s \mid \alpha_s(\ell) \neq 0\}$. For each $s \in \omega_\ell$, compute its residual, $e_s^\ell = R_s u - \sum_{j \neq \ell} d_j \alpha_s(j)$. Set $E_\ell = (e_s^\ell)_{s \in \omega_\ell}$. Update the dictionary column d_ℓ and the coefficient values $(\alpha_s(\ell))_{s \in \omega_\ell}$ by applying the SVD to E_ℓ .

Finally, the denoised image u is explicitly obtained as

$$u = \left(\lambda I + \sum_{s \in \mathcal{B}} R_s^T R_s \right)^{-1} \left(\lambda f + \sum_{s \in \mathcal{B}} R_s^T D \alpha_s \right).$$

2.3. Total generalized variation

The total generalized variation (TGV) [18, 19] was proposed as a penalty functional for modeling images with edges as well as smooth variations. It can be interpreted as a sparse penalization of optimal balancing from the first up to the k -th derivative. In particular, the discrete second-order TGV can be defined as

$$\text{TGV}_\rho^2(u) := \min_w \|\nabla u - w\|_1 + \rho \|\mathcal{E}(w)\|_1, \quad (2.2)$$

where $\|\cdot\|_1$ represents the sum of the ℓ^2 (Frobenius) norms of 2×1 vectors (2×2 matrices), ρ is a weighted parameter that controls the balance between the first- and second-order terms, and $\mathcal{E}(w)$ represents the symmetrized gradient operator that can be separately expressed as

$$\mathcal{E}(w) = \begin{bmatrix} \partial_x w_1 & \frac{1}{2}(\partial_y w_1 + \partial_x w_2) \\ \frac{1}{2}(\partial_y w_1 + \partial_x w_2) & \partial_y w_2 \end{bmatrix}. \quad (2.3)$$

From the formulation (2.2) of TGV_ρ^2 , it can be interpreted that TGV_ρ^2 can automatically find an appropriate balance between the first- and the second-order derivative of u with respect to ρ . As a consequence, it reduces the staircase effects generated by the bounded variation functional.

3. Description of the Proposed Model and Algorithm

In this section we present a new model for the restoration of images deteriorated by Cauchy noise and/or blurring. Specifically, we consider the estimation of a clean image u from the observed image f in (1.1) corrupted by Cauchy noise following $\mathcal{C}(0, \gamma)$.

A classical statistical approach is to find the MAP estimate of u by maximizing the conditional posterior probability $P(u|f)$, the probability that u occurs when f is observed. From the Bayes' Theorem, i.e.,

$$P(u|f) = \frac{P(f|u)P(u)}{P(f)},$$

and by taking negative logarithms, the MAP estimate of u can be retrieved by minimizing the functional $-\log(P(f|u)) - \log P(u)$, where the expression $-\log P(f|u)$ can be viewed as a data-fidelity term and the term $-\log P(u)$ is used to regularize a solution. Assuming that f and u are written as column vectors of $\mathbb{R}^{m \times n}$ indexed by $\mathcal{A} = \{1, 2, \dots, m\} \times \{1, 2, \dots, n\}$, the image is discretized and the pixels are mutually independent and identically distributed, thus $P(u) = \prod_{s \in \mathcal{A}} P(u_s)$. Besides, since $P(f_s|u_s) = \frac{\gamma}{\pi(\gamma^2 + ((Au)_s - f_s)^2)}$, one can have that

$$-\log P(f|u) = - \sum_{s \in \mathcal{A}} \log(P(f_s|u_s)) = \sum_{s \in \mathcal{A}} \log(\gamma^2 + ((Au)_s - f_s)^2) + \log \pi - \log \gamma. \quad (3.1)$$

Then we can integrate the data-fidelity term in (3.1) with the sparse representation prior via dictionary learning in (2.1). However, this patch-based prior may cause some blocking artifacts (for instance, see [46]), so we combine two priors: the patch-based sparse representation prior and the local derivative-based prior, TGV_ρ^2 . TGV regularization prefers to locally smooth images like TV, so it can diminish the artifacts generated by the patch-based prior. In addition, it produces visually desirable results with almost no staircase artifacts present in the restored images, providing a natural visual quality than TV. But, it may smooth textural patterns and details. The patch-based prior makes use of the structure similarity of image patterns, so it furnishes a recovery in textures, conserving a more geometrical structure. As a result, TGV regularization is effective in denoising smooth regions, whereas the patch-based prior is well suited to textures and delicate structures. Moreover, both priors are complementary by ameliorating mutual drawbacks.

Consequently, we propose the following model for restoring images degenerated by Cauchy noise and/or blurring:

$$\min_{\{\alpha_s\}, D, u} \frac{\lambda}{2} \langle \log(\gamma^2 + (Au - f)^2), \mathbf{1} \rangle + \sum_{s \in \mathcal{B}} \mu_s \|\alpha_s\|_0 + \sum_{s \in \mathcal{B}} \|D\alpha_s - R_s u\|_2^2 + \zeta \text{TGV}_\rho^2(u), \quad (3.2)$$

where $D \in \mathbb{R}^{N \times K}$ ($N \ll K$), $\alpha_s \in \mathbb{R}^K$, $R_s \in \mathbb{R}^{N \times mn}$ are defined in (2.1), $\mathcal{B} = \{1, 2, \dots, m - \sqrt{N} + 1\} \times \{1, 2, \dots, n - \sqrt{N} + 1\}$, and $\zeta > 0$ is a parameter that balances the sparse representation prior and TGV term. We note that the combination of the sparse representation prior and TV regularization has been introduced and utilized for image restoration problems under some non-Gaussian noise, such as multiplicative noise [39], Poisson noise [41], and impulsive noise [43]. Here we benefit from TGV instead of TV to further enhance the quality of restoration.

3.1. Algorithm for solving model (3.2)

The proposed minimization problem (3.2) is nonconvex due to the nonconvex data-fitting term, ℓ_0 term, and the product between the unknowns D and α_s . Furthermore, the TGV regularization term is nondifferentiable, so it is not trivial to solve the problem (3.2).

First, we employ the variable splitting scheme, which is a very simple procedure that consists of creating new variables to convert an unconstrained minimization problem to its equivalent constrained one. The motivation behind the variable splitting technique is that it may be easier to solve the constrained problem than its unconstrained counterpart, and this idea has been recently used in several image processing applications [47–50]. Here, we introduce auxiliary variables $p \in \mathbb{R}^{mn}$, $q \in (\mathbb{R}^{mn})^2$, $r \in (\mathbb{R}^{mn})^4$, and $z \in \mathbb{R}^{mn}$, and then the unconstrained problem (3.2) can be reformulated as

$$\begin{aligned} \min_{\{\alpha_s\}, D, u, w, p, q, r, z} & \frac{\lambda}{2} \langle \log(\gamma^2 + (z - f)^2), \mathbf{1} \rangle + \sum_{s \in \mathcal{B}} \mu_s \|\alpha_s\|_0 + \sum_{s \in \mathcal{B}} \|D\alpha_s - R_s p\|_2^2 + \zeta (\|q\|_1 + \rho \|r\|_1), \\ \text{subject to: } & z = Au, \quad p = u, \quad q = \nabla u - w, \quad r = \mathcal{E}(w). \end{aligned} \quad (3.3)$$

Then, to solve the constrained minimization problem (3.3), we employ the penalty method [51, 52], which replaces a constrained problem by a series of unconstrained problems, by adding penalty terms for constraints. Its solution ideally converges to the solution of the original constrained problem. Therefore, using the penalty method, the constrained problem (3.3) can be relaxed to the following unconstrained one

$$\begin{aligned} \min_{\{\alpha_s\}, D, u, w, p, q, r, z} & \frac{\lambda}{2} \langle \log(\gamma^2 + (z - f)^2), \mathbf{1} \rangle + \sum_{s \in \mathcal{B}} \mu_s \|\alpha_s\|_0 \\ & + \sum_{s \in \mathcal{B}} \|D\alpha_s - R_s p\|_2^2 + \zeta (\|q\|_1 + \rho \|r\|_1) + \frac{\beta}{2} \|p - u\|_2^2 \\ & + \frac{\tau_1}{2} \|q - \nabla u\|_2^2 + \frac{\tau_2}{2} \|r - \mathcal{E}(w)\|_2^2 + \frac{\xi}{2} \|z - Au\|_2^2, \end{aligned} \quad (3.4)$$

where β , τ_1 , τ_2 and ξ are positive parameters. The relation between (3.4) and (3.3) is straightforward: as the parameters β , τ_1 , τ_2 , ξ approaches $+\infty$, problem (3.4) turns back to problem (3.3).

To realize the problem (3.4), we adopt an alternating minimization algorithm (AMA). The AMA [53] basically minimizes a function of two variables, and its essential idea is to keep one variable fixed while minimizing the other variable and iterate this process. The AMA was extended to solve the minimization problem with more than three variables and its convergence was proven in [54] under certain restrictions. These methods have performed well in practice, although the objective function is non-convex. The AMA applied to (3.4) results in the following several subproblems:

$$(\{\alpha_s\}_{s \in \mathcal{B}}, D) \in \arg \min_{\{\alpha_s\}, D} \sum_{s \in \mathcal{B}} \mu_s \|\alpha_s\|_0 + \sum_{s \in \mathcal{B}} \|D\alpha_s - R_s p\|_2^2, \quad (3.5)$$

$$p \in \arg \min_p \sum_{s \in \mathcal{B}} \|D\alpha_s - R_s p\|_2^2 + \frac{\beta}{2} \|p - u\|_2^2, \quad (3.6)$$

$$q \in \arg \min_q \zeta \|q\|_1 + \tau_1 \|q - (\nabla u - w)\|_2^2, \quad (3.7)$$

$$r \in \arg \min_r \zeta \rho \|r\|_1 + \tau_2 \|r - \mathcal{E}(w)\|_2^2, \quad (3.8)$$

$$z \in \arg \min_z \frac{\lambda}{2} \langle \log(\gamma^2 + (z - f)^2), \mathbf{1} \rangle + \frac{\xi}{2} \|z - Au\|_2^2, \quad (3.9)$$

$$\begin{aligned} (u, w) \in \arg \min_{u, w} & \frac{\beta}{2} \|u - p\|_2^2 + \frac{\tau_1}{2} \|q - (\nabla u - w)\|_2^2 \\ & + \frac{\tau_2}{2} \|r - \mathcal{E}(w)\|_2^2 + \frac{\xi}{2} \|z - Au\|_2^2. \end{aligned} \quad (3.10)$$

All of which can be efficiently solved or have closed solutions, which will be given in the following subsections.

3.1.1. Solving $(\{\alpha_s\}, D)$ -subproblem

We solve the subproblem (3.5) for $\{\alpha_s\}$ and D while p fixed. First, with fixed D and p , we solve the minimization with respect to α_s for each $s \in \mathcal{B}$, which is given by

$$\min_{\alpha_s} \mu_s \|\alpha_s\|_0 + \|D\alpha_s - R_s p\|_2^2. \quad (3.11)$$

This subproblem can be efficiently solved by using the orthogonal matching pursuit (OMP) algorithm [55]. The OMP is an iterative greedy algorithm that chooses at each step the column of D which is the most correlated with the current residuals. Solving (3.11) using OMP is easy, gathering one atom at a time, and stopping when the error $\|D\alpha_s - R_s p\|_2^2$ goes below ϵ^2 . In such a way, the selection of μ_s is implicitly handled. A main benefit of OMP is its simplicity and fast implementation.

To find the dictionary D , we apply the K-SVD algorithm to the problem (3.5). However, if we update the dictionary whenever p is updated, the computational cost becomes high. Thus we only update the dictionary D in the outer loop to reduce the computing time. Moreover, the dictionary learning procedure significantly affects on the quality of the recovered image. We use the processed image u to learn the dictionary D , instead of p . Indeed, u and p become almost the same after all since β is set to approach ∞ . Therefore, updating D in the outer loop is reasonable since the dictionary remains almost the same when u reaches convergence. In practice, we repeat the *sparse coding* and *dictionary update* stages J times in the K-SVD algorithm.

3.1.2. Solving p, q, r, z -subproblems

The subproblem for p in (3.6) is a least squares problem that results in a closed form solution as

$$p = \left(\beta I + 2 \sum_{s \in \mathcal{B}} R_s^T R_s \right)^{-1} \left(\beta u + 2 \sum_{s \in \mathcal{B}} R_s^T D \alpha_s \right). \quad (3.12)$$

The expression (3.12) shows that averaging of the denoised patches can be done with some relaxation acquired by averaging with the current denoised image u . The matrix $R_s^T R_s$ has the value 1 only in the diagonal entries corresponding to the indices of the image patch at location s , otherwise it has zero values. Thus the matrix $\beta I + 2 \sum_{s \in \mathcal{B}} R_s^T R_s$ is diagonal, so the computation of (3.12) can be also done on a pixel-by-pixel basis.

The minimization problems (3.7) and (3.8) for q and r , respectively, are the l_1 -regularized least squares problems. These can be explicitly solved by an elementwise soft-thresholding as

$$q = \mathit{shrink}_2 \left(\nabla u - w, \frac{\zeta}{\tau_1} \right), \quad r = \mathit{shrink}_F \left(\mathcal{E}(w), \frac{\zeta \rho}{\tau_2} \right), \quad (3.13)$$

where the shrinkage operator $\mathit{shrink}_t(\cdot, \cdot)$ is defined as

$$\mathit{shrink}_t(\nu, a)_s = \begin{cases} \frac{\nu_s}{\|\nu_s\|_t} \cdot \max(\|\nu_s\|_t - a, 0), & \text{if } \|\nu_s\|_t \neq 0, \\ 0, & \text{if } \|\nu_s\|_t = 0, \end{cases} \quad (3.14)$$

with $\|\cdot\|_F$ representing the Frobenius norm of a matrix.

It can be easily proven that the minimization problem (3.9) is strictly convex if $8\gamma \geq \lambda/\xi$, and we can attain a minimizer of it by solving its Euler-Lagrange equation as

$$G(z) = \lambda \frac{(z-f)}{\gamma^2 + (z-f)^2} + \xi(z-Au) = 0. \quad (3.15)$$

To solve the normal equation (3.15), the Newton's method is employed. That is, z can be obtained by iterating the following equation for $t = 1, 2, \dots, N_z$:

$$z^{t+1} = z^t - \frac{G(z^t)}{G'(z^t)}, \quad (3.16)$$

which converges after a few iterations in practice.

Algorithm 3.1. Solving the proposed model (3.2)

1: **Input:** choose the parameters $\lambda, \zeta, \rho, \beta, \tau_1, \tau_2, \xi > 0$, the patch size, the number of columns of D , the iteration number for the inner loop (N_I), and the growth rates $r_\beta, r_{\tau_1}, r_{\tau_2}, r_\xi$.

2: **Initialization:** Set $u = \max(\min(f, 255), 0)$, $w = \mathbf{0}$, $D = DCT$, $p = u$, $z = Au$.

3: **repeat**

4: **for** $\ell = 0, 1, 2, \dots, N_I$ **do**

5: Update $\alpha_s^{\ell+1}$ ($s \in \mathcal{B}$) by solving (3.11) using OMP;

6: $p^{\ell+1} = \left(\beta I + 2 \sum_{s \in \mathcal{B}} R_s^T R_s \right)^{-1} \left(\beta u^\ell + 2 \sum_{s \in \mathcal{B}} R_s^T D \alpha_s^{\ell+1} \right)$;

7: $q^{\ell+1} = \mathit{shrink}_2 \left(\nabla u^\ell - w^\ell, \frac{\zeta}{\tau_1} \right)$;

8: $r^{\ell+1} = \mathit{shrink}_F \left(\mathcal{E}(w^\ell), \frac{\zeta \rho}{\tau_2} \right)$;

9: $z^{\ell+1}$ by solving (3.15) using Newton's method;

10: $u^{\ell+1}, w^{\ell+1}$ by solving (3.17a)–(3.17b) using FFT;

11: **end for**

12: Update D by using the K-SVD;

13: $\beta = \beta \cdot r_\beta$;

14: $\tau_1 = \tau_1 \cdot r_{\tau_1}$;

15: $\tau_2 = \tau_2 \cdot r_{\tau_2}$;

16: $\xi = \xi \cdot r_\xi$;

17: **until** s stopping condition is satisfied.

18: **Output:** restored image u .

3.1.3. Solving (u, w) -subproblem

The minimization problem (3.10) is also a least squares problem, which brings the normal equations as

$$(\beta + \tau_1 \nabla^T \nabla + \xi A^T A) u - \tau_1 \nabla^T w = \beta p + \tau_1 \nabla^T q + \xi A^T z, \quad (3.17a)$$

$$-\tau_1 \nabla u + \tau_1 w + \tau_2 \mathcal{E}^T \mathcal{E}(w) = -\tau_1 q + \tau_2 \mathcal{E}^T(r), \quad (3.17b)$$

where ∇^T and \mathcal{E}^T are the adjoint operators of ∇ and \mathcal{E} , respectively. Indeed, $\nabla^T = -\text{div}$, where div is a discrete divergence: $\text{div } v = \partial_x v_1 + \partial_y v_2$ for $v = [v_1, v_2]^T$. And $\mathcal{E}^T = -\text{div}^2$, where

$$\text{div}^2 W = \begin{bmatrix} \partial_x w_{11} + \partial_y w_{12} \\ \partial_x w_{21} + \partial_y w_{22} \end{bmatrix} \quad \text{for } W = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}.$$

Following the ideas in [56], we can obtain u and w by applying the 2D fast Fourier transform (FFT) to (3.17a)–(3.17b). Specifically, assuming that ∇_1 and ∇_2 are the circulant matrices corresponding to the forward finite difference operators with periodic boundary conditions along the x -axis and y -axis, respectively, and using the fact that the circulant matrices can be diagonalized under the Fourier transform, we can attain the closed-form solutions to (3.17a)–(3.17b). First, we rearrange the formula (3.17a)–(3.17b) as the following linear system:

$$\begin{bmatrix} d_1 & d_4^T & d_5^T \\ d_4 & d_2 & d_6^T \\ d_5 & d_6 & d_3 \end{bmatrix} \begin{bmatrix} u \\ w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix}, \quad (3.18)$$

where $w = [w_1, w_2]^T$,

$$\begin{aligned} d_1 &= \beta + \tau_1(\nabla_1^T \nabla_1 + \nabla_2^T \nabla_2) + \xi A^T A, & d_2 &= \tau_1 + \tau_2 \left(\nabla_1^T \nabla_1 + \frac{1}{2} \nabla_2^T \nabla_2 \right), \\ d_3 &= \tau_1 + \tau_2 \left(\frac{1}{2} \nabla_1^T \nabla_1 + \nabla_2^T \nabla_2 \right), & d_4 &= -\tau_1 \nabla_1, & d_5 &= -\tau_1 \nabla_2, & d_6 &= \frac{\tau_2}{2} \nabla_1^T \nabla_2, \\ C_1 &= \beta p + \tau_1 \nabla^T q + \xi A^T z, & [C_2, C_3]^T &= -\tau_1 q + \tau_2 \mathcal{E}^T(r). \end{aligned}$$

Then denoting F by a matrix representing the 2D discrete Fourier transform, we multiply a preconditioner matrix from the left to the linear system (3.18) such that the coefficient matrix is blockwise diagonal:

$$\begin{bmatrix} F & 0 & 0 \\ 0 & F & 0 \\ 0 & 0 & F \end{bmatrix} \begin{bmatrix} d_1 & d_4^T & d_5^T \\ d_4 & d_2 & d_6^T \\ d_5 & d_6 & d_3 \end{bmatrix} \begin{bmatrix} F & 0 & 0 \\ 0 & F & 0 \\ 0 & 0 & F \end{bmatrix}^* \begin{bmatrix} Fu \\ Fw_1 \\ Fw_2 \end{bmatrix} = \begin{bmatrix} F & 0 & 0 \\ 0 & F & 0 \\ 0 & 0 & F \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix}, \quad (3.19)$$

where F^* is the conjugate of F . By letting $\tilde{d}_j = \text{diag}(Fd_j F^*)$ and $\tilde{d}_j^T = \text{diag}(Fd_j^T F^*) = \text{conj}(\text{diag}(Fd_j F^*))$, we have

$$\begin{cases} \tilde{d}_1 \cdot (Fu) + \tilde{d}_4^T \cdot (Fw_1) + \tilde{d}_5^T \cdot (Fw_2) = FC_1, \\ \tilde{d}_4 \cdot (Fu) + \tilde{d}_2 \cdot (Fw_1) + \tilde{d}_6^T \cdot (Fw_2) = FC_2, \\ \tilde{d}_5 \cdot (Fu) + \tilde{d}_6 \cdot (Fw_1) + \tilde{d}_3 \cdot (Fw_2) = FC_3, \end{cases} \quad (3.20)$$

where \cdot is the componentwise multiplication.

Analogously to the scalar case, Fu , Fw_1 , and Fw_2 can be achieved by applying Cramer's



Fig. 3.1. Original clean images. Top to bottom (left to right): Barbara, Boat, Building (321×481), Face, Cameraman, Lena, Walkbridge, Lake, Mandrill, Parrot, Peppers, Policemen (321×481), Castle (321×481).

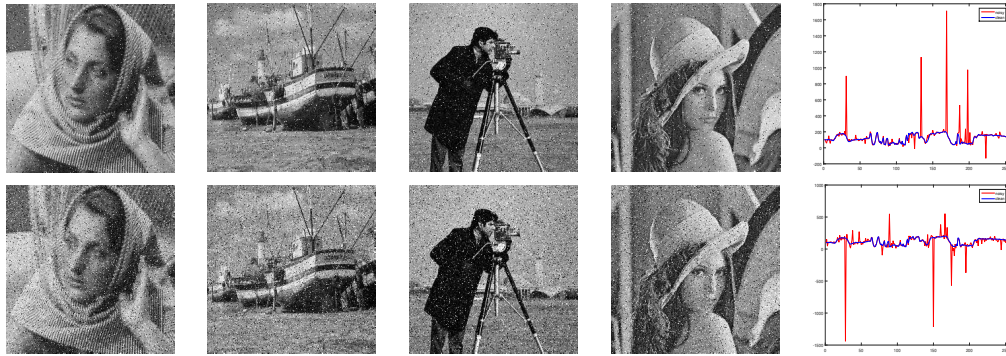


Fig. 3.2. Noisy images and signals corrupted by Cauchy noise with noise level γ (Top row: $\gamma = 5$, Bottom row: $\gamma = 10$). 1st-4th columns: noisy images, 5th column: horizontal cross-sectional lines of original Lena image (blue line) and noisy Lena image (red line).

rule. Therefore u and w have the following closed-forms:

$$\begin{cases} u = F^* \left(\begin{array}{c|c} FC_1 & \widetilde{d}_4^T & \widetilde{d}_5^T \\ FC_2 & \widetilde{d}_2 & \widetilde{d}_6^T \\ FC_3 & \widetilde{d}_6 & \widetilde{d}_3 \end{array} \right) ./\text{denom}, \\ w_1 = F^* \left(\begin{array}{c|c} \widetilde{d}_1 & FC_1 & \widetilde{d}_5^T \\ \widetilde{d}_4 & FC_2 & \widetilde{d}_6^T \\ \widetilde{d}_5 & FC_3 & \widetilde{d}_3 \end{array} \right) ./\text{denom}, \\ w_2 = F^* \left(\begin{array}{c|c} \widetilde{d}_1 & \widetilde{d}_4^T & FC_1 \\ \widetilde{d}_4 & \widetilde{d}_2 & FC_2 \\ \widetilde{d}_5 & \widetilde{d}_6 & FC_3 \end{array} \right) ./\text{denom} \end{cases}, \quad \text{denom} = \begin{vmatrix} \widetilde{d}_1 & \widetilde{d}_4^T & \widetilde{d}_5^T \\ \widetilde{d}_4 & \widetilde{d}_2 & \widetilde{d}_6^T \\ \widetilde{d}_5 & \widetilde{d}_6 & \widetilde{d}_3 \end{vmatrix}_*, \quad (3.21)$$

where $./$ is the componentwise division, and $|\cdot|_*$ is defined to be the determinant of the matrix $(a_{ij})_{1 \leq j \leq 3}$.

Lastly, the parameters β , τ_1 , τ_2 , ξ should be set as very large values to enforce the solution of (3.4) to reach that of (3.3), which in turn is identical to (3.2). But, if too large values are initially set for the parameters, numerical stability problems can arise (see [57], Chapter 17). Thus, from the idea of the FTVD method [47], we start with small values for the parameters and gradually enlarge them during the iterative process, which imposes convergence to the solution

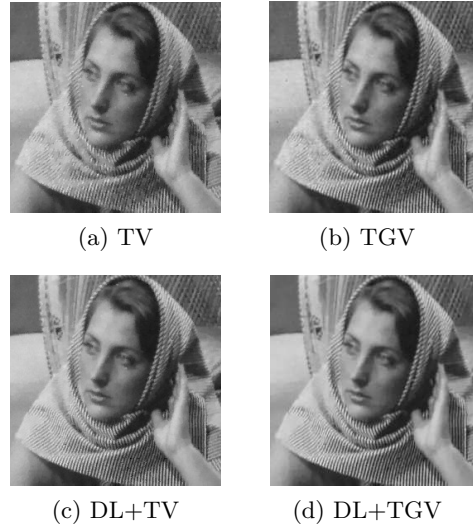


Fig. 3.3. Comparison of denoising results when $\gamma = 5$ with different regularization terms (TV, TGV, DL+TV, DL+TGTV) and the same data-fitting term in (3.2). PSNR: (a) 27.007, (b) 29.022, (c) 30.396, (d) 31.110. Parameter λ for our model: 450.

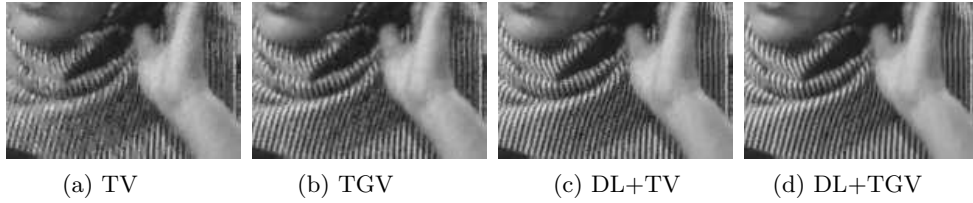


Fig. 3.4. Zoomed images of the restored images in Figure 3.3.

of the model (3.3). Hence, the whole algorithm for solving the model (3.2) is summarized in Algorithm 3.1.

4. Numerical Results

This section presents the experimental results for the proposed model (3.2) and compare them to the nonconvex TV model [10], given in (1.3). For denoising, we also present the denoising results of the median filtering since the Cauchy noise has an impulsive character. We test with 13 natural images, provided in Figure 3.1, and the range of the intensity values in clean images is assumed to be $[0, 255]$. The size of images is 256×256 except some. The observed images are generated by $f = Au + \eta = Au + \gamma \frac{\eta_1}{\eta_2}$, where η_1 and η_2 follow the standard Gaussian distribution, $\mathcal{N}(0, 1)$, independently, and $\gamma = 5, 10, \text{ or } 15$. The operator A is defined as the identity operator for denoising and a blurring operator for deblurring and denoising. When A is a blurring operator, we consider two types of blurring kernels: a Gaussian blurring kernel of size 7×7 with standard deviation 1, and a uniform blurring kernel of size 5×5 .

To quantify the quality of restored images, we compute the Peak-Signal-to-Noise-Ratio

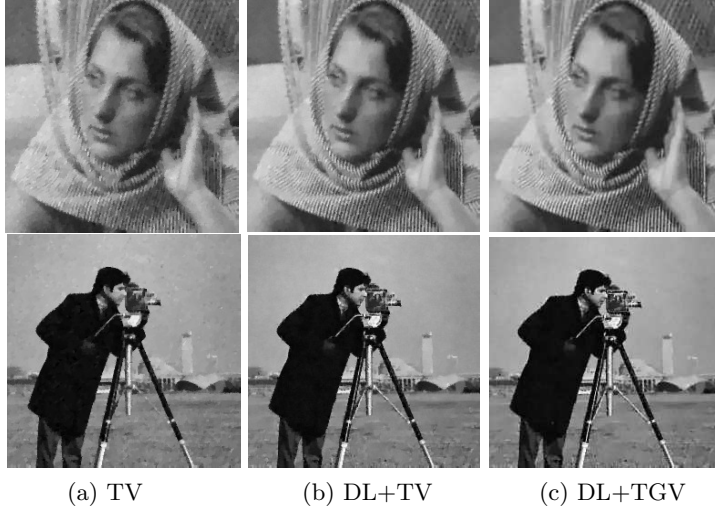


Fig. 3.5. Comparison of denoising results when $\gamma = 10$ with different regularization terms (TV, TGV, DL+TV, DL+TGV) and the same data-fitting term in (3.2). PSNR: (top) (a) 25.410, (b) 27.436, (c) 28.006; (bottom) (a) 26.283, (b) 27.916, (c) 28.298. Parameter λ for our model: 750 (top), 850 (bottom).

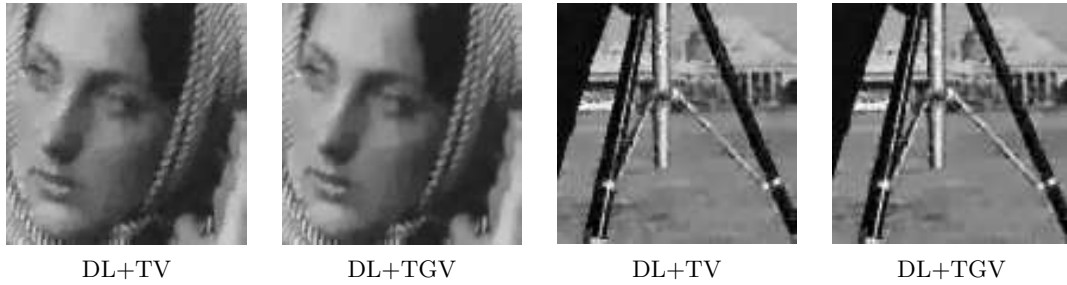


Fig. 3.6. Zoomed images of the restored images in Figure 3.5.

(PSNR) value, which is given by

$$\text{PSNR}(u, \bar{u}) = 10 \log_{10} \left(\frac{255^2 mn}{\|u - \bar{u}\|_2^2} \right),$$

where mn is the size of the image, and \bar{u} represents the original image. The structure similarity (SSIM) index [58] and quality index based on local variance (QILV) [59] are also measured. The SSIM index and QILV have a great (intentional) similarity, but the SSIM index is the mean of the local statistics of the images while QILV deals with global statistics of the local variances of the images.

The stopping criterion for all methods is given by

$$\frac{\|u^{iter+1} - u^{iter}\|_2}{\|u^{iter+1}\|_2} < tol, \quad \text{or} \quad iter > M, \quad (4.1)$$

where tol is a given tolerance, and M is a maximum iteration number. For all models, the value for tol is set as $tol = 5 \times 10^{-5}$. The maximum iteration numbers for the TV model is fixed as $M = 200$. For our model, we use the stopping criterion (4.1) for both the inner and outer loops, and we set $M = 60$ for the inner loop. The K-SVD algorithm for learning the dictionary

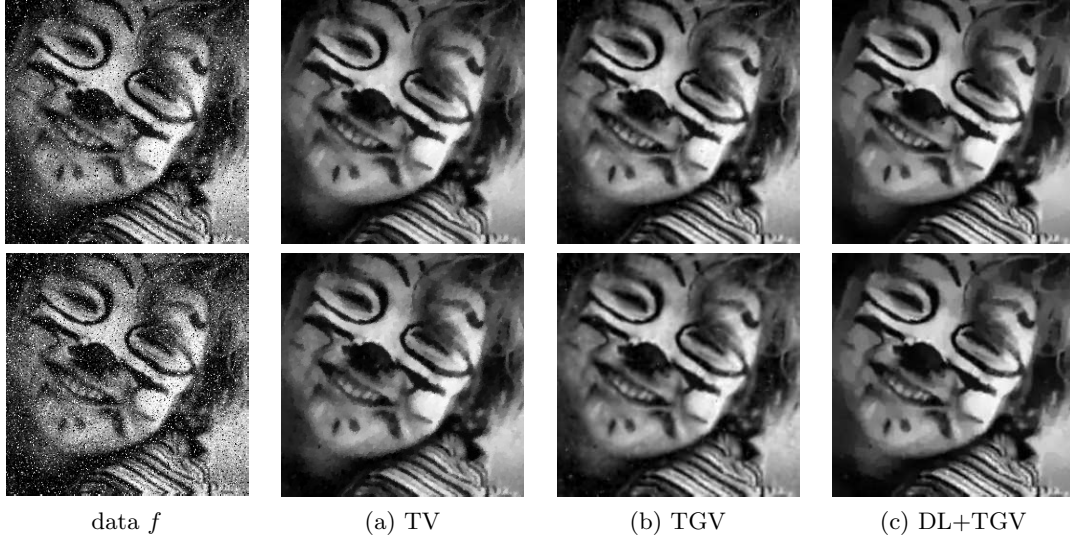


Fig. 3.7. Comparison of denoising results when $\gamma = 10$ (top), $\gamma = 15$ (bottom) with different regularization terms (TV, TGV, DL+TGV) and the same data-fitting term in (3.2). PSNR: (top) (a) 29.245, (b) 31.036, (c) 31.678; (bottom) (a) 27.659, (b) 28.600, (c) 29.799. Parameter λ for our model: 700 (top), 900 (bottom).

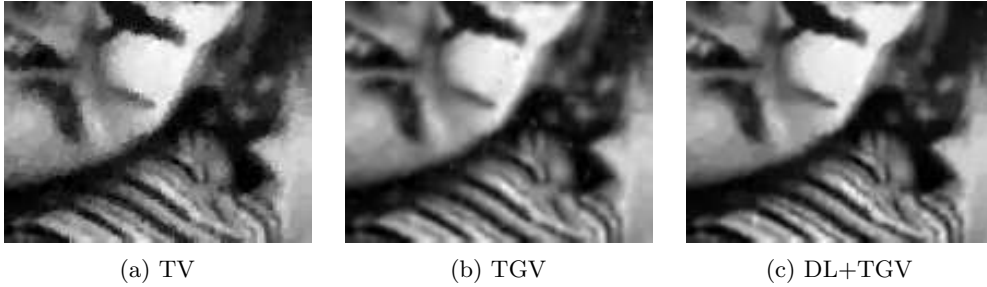


Fig. 3.8. Zoomed images of the restored images when $\gamma = 15$ in Figure 3.7.

D is terminated when the iteration number reaches $J = 20$. In addition, we fix the iteration number of Newton's method for z -subproblem as $N_z = 5$.

In all experiments, we set the parameters for our model as follows. The size of the image patch is fixed as 4×4 , and the size of the dictionary D is set to be 16×256 . The initial penalty parameters are set as $(\beta, \tau_1, \tau_2, \xi) = (1, 0.2, 0.2, 10)$, and the growth rates of the penalty parameters are fixed as $(r_\beta, r_{\tau_1}, r_{\tau_2}, r_\xi) = (2, 1.5, 1.5, 2)$. The remaining parameter λ , ζ and ρ are the regularization parameters, which mainly affect the quality of the restored images. The regularization parameters (ζ, ρ) are fixed as $(30, 0.2)$ for denoising, $(1, 0.2)$ (Gaussian blur) and $(1, 0.1)$ (uniform blur) for deblurring-denoising. Then the parameter λ is adjusted, depending on the noise level, to achieve the best restored images. For instance, for the denoising examples, $\lambda \in \{450, 500, 550, 600\}$ if $\gamma = 5$, $\lambda \in \{700, 750, 800, 850\}$ if $\gamma = 10$, and $\lambda \in \{900, 950, 1000, 1050, 1100\}$ if $\gamma = 15$. When executing the OMP for $\{\alpha_s\}$ and D , atoms are accumulated until the average error passes the threshold ϵ , chosen empirically to be $\epsilon = 1.15 \cdot \sigma$ with a fixed $\sigma = 3$. Finally, from the experiments in [12], we choose the initial condition for u to be $u = \max(\min(f, 255), 0)$ for all models.

The regularization parameter λ in the TV model was empirically tuned to attain the best

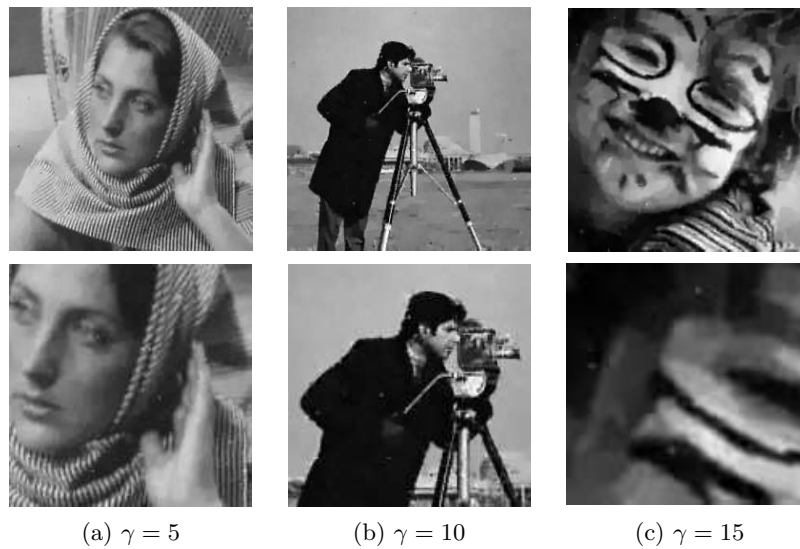


Fig. 3.9. Denoising results of the L^1 data-fitting term ($\lambda\|f - Au\|_1$) with the DL+TGV regularization (Top row). Bottom row: zoomed images of the restored images in the top row. PSNR (top): (a) 30.417, (b) 27.742, (c) 28.874.



Fig. 3.10. Denoised images when noise level $\gamma = 5$, using (a) Median filtering, (b) TV model [10], (c) our model (3.2). PSNR (from (a) to (c)): (top) 25.168/26.894/28.730, (middle) 26.272/27.721/30.026, (bottom) 28.387/29.434/31.043. Parameter λ for our model: (top to bottom) 600, 550, 550.



Fig. 3.11. Zoomed images of the restored images (top: TV, bottom: Proposed) in Figure 3.10.

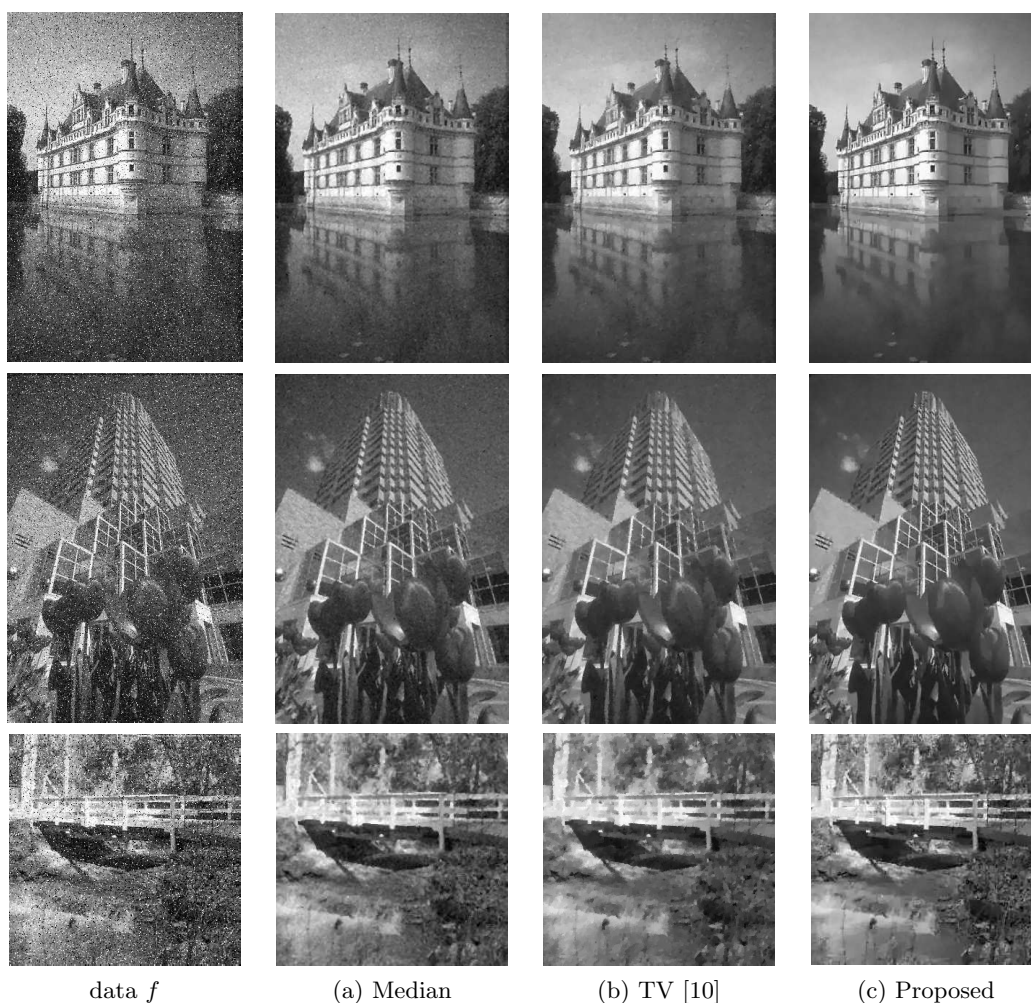


Fig. 3.12. Denoised images when noise level $\gamma = 10$, using (a) Median filtering, (b) TV model [10], (c) our model (3.2). PSNR (from (a) to (c)): (top) 25.272/26.787/28.702, (middle) 25.267/25.569/28.020, (bottom) 22.330/23.241/24.235. Parameter λ for our model: (top to bottom) 750, 800, 850.

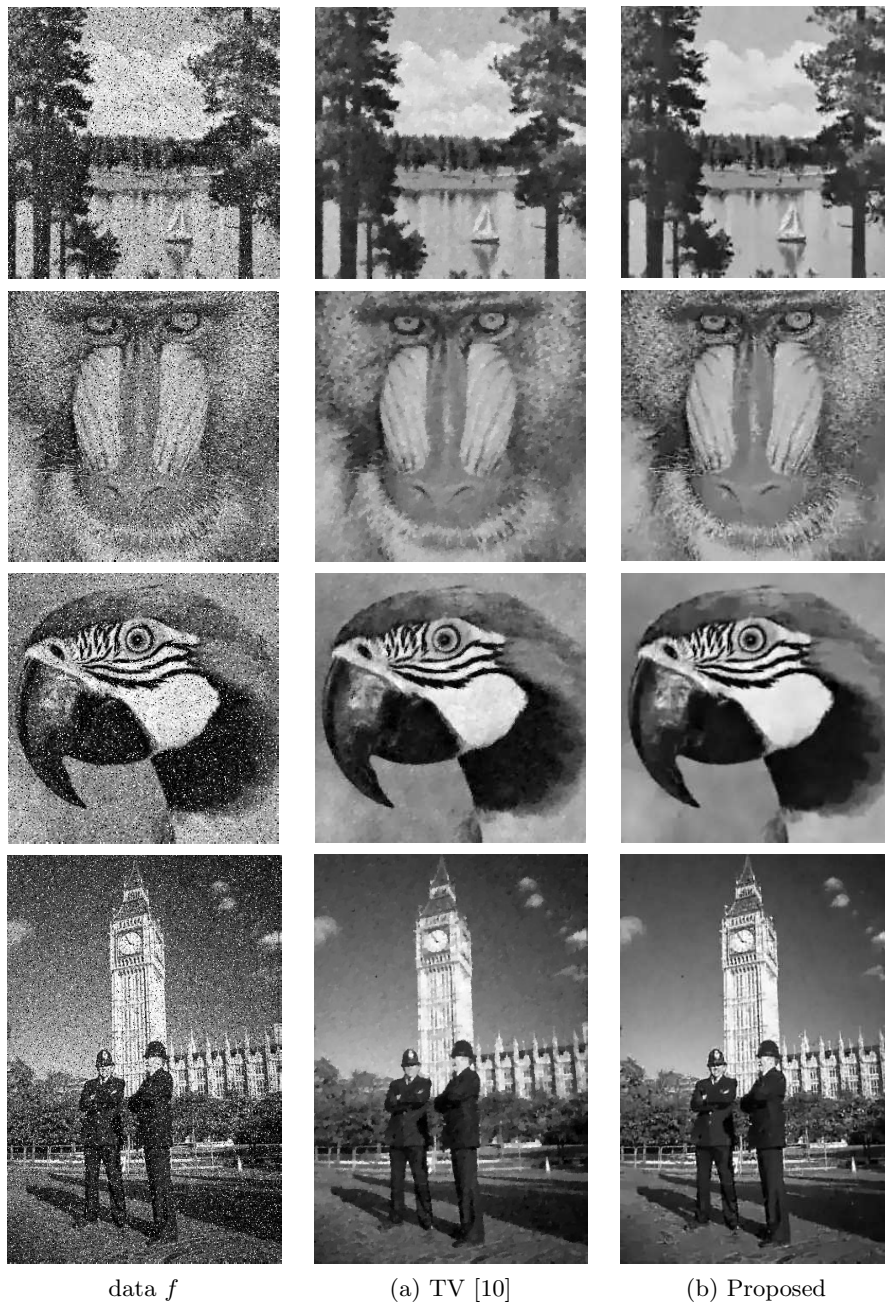


Fig. 3.13. Denoised images when noise level $\gamma = 15$, using (a) TV model [10], (b) our model (3.2). PSNR (1st row to 4th row): (a) 24.522/20.956/26.066/22.506, (b) 25.316/21.644/26.858/24.331. Parameter λ for our model: (top to bottom) 1000, 1100, 900, 1000.

quality of the restored images. The TV model (1.3) was solved by the nonconvex ADMM [13], as in [12], and the convex ADMM [60] was adopted to solve the subproblem for u .

4.1. Image Denoising results

First, in Figures 3.3-3.8, we compare the denoising results when using different regularization terms with the same data-fidelity term given in (3.2). Specifically, we compare the denoising

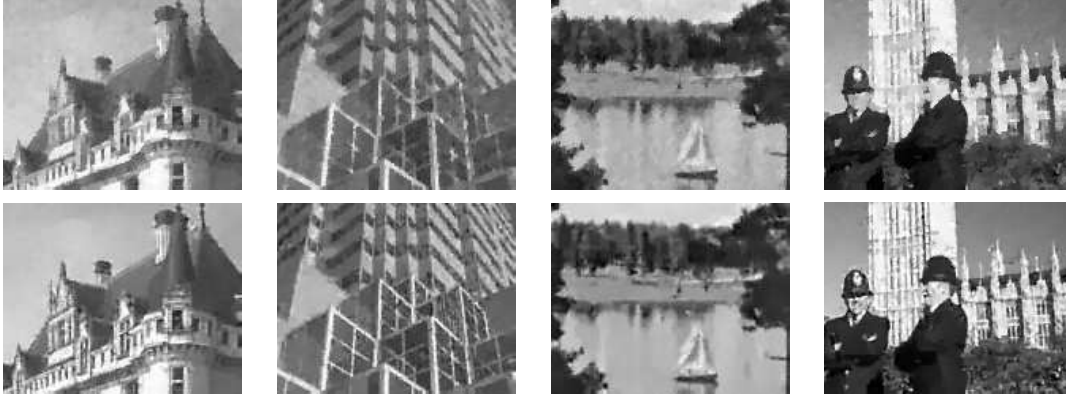


Fig. 3.14. Zoomed images of the restored images (top: TV, bottom: Proposed) in Figures 3.12 and 3.13.

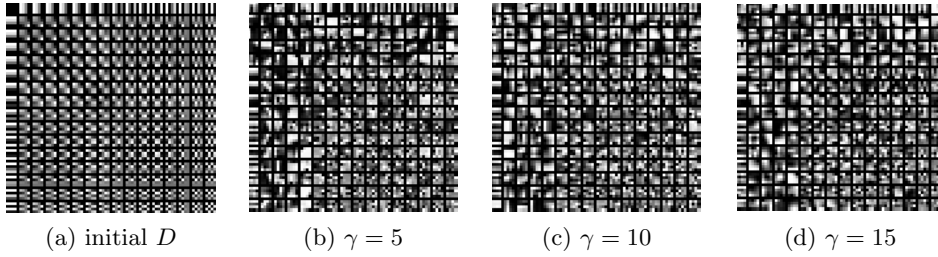


Fig. 3.15. Learned dictionaries with patch size 4×4 . (a) overcomplete DCT dictionary, (b)-(d) learned dictionaries of (b) Cameraman with $\gamma = 5$, (c) Castle with $\gamma = 10$, (d) Lake with $\gamma = 15$.

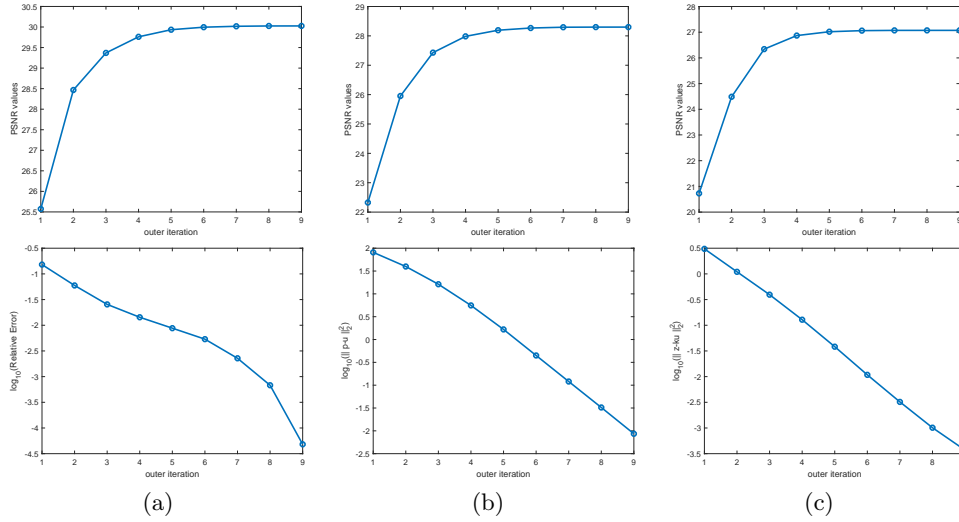


Fig. 3.16. Plots of energy values via the outer iteration. Top: (a)-(c) PSNR values when noise level $\gamma = 5, 10, 15$, respectively. Bottom: (a) $\log_{10}(\|u^{iter+1} - u^{iter}\|_2 / \|u^{iter+1}\|_2)$, (b) $\log_{10}(\|p - u\|_2^2 / (mn))$, (c) $\log_{10}(\|z - Ku\|_2^2 / (mn))$, when noise level $\gamma = 5$.

results of our DL+TGV regularization with those of TV, TGV, or DL+TV, where DL+TV indicates the integrated regularization of the sparse representation prior via DL and TV. Figure 3.3 presents the denoised images of all four regularization terms when the noise level $\gamma = 5$. It

Table 4.1: Denoising results when noise level $\gamma = 5$.

Model	(a) Median filter			(b) TV model [10]			(d) Proposed model		
Image	PSNR	SSIM	QILV	PSNR	SSIM	QILV	PSNR	SSIM	QILV
barbara	26.045	0.7970	0.5272	27.007	0.8401	0.6950	31.110	0.9077	0.9478
boat	25.168	0.7103	0.7389	26.894	0.8000	0.8750	28.730	0.8403	0.9775
building	26.550	0.8190	0.7600	27.914	0.8677	0.8925	30.100	0.8992	0.9788
cameraman	26.272	0.7970	0.8434	27.721	0.8481	0.9161	30.026	0.8716	0.9855
castle	26.361	0.8007	0.7117	27.423	0.8582	0.7252	30.974	0.8956	0.9798
face	31.646	0.9172	0.9810	31.907	0.9249	0.9883	34.894	0.9455	0.9950
lake	25.193	0.7749	0.8803	26.281	0.8202	0.9149	28.271	0.8628	0.9833
lena	28.387	0.8106	0.9020	29.434	0.8509	0.9442	31.043	0.8813	0.9839
mandrill	20.571	0.5047	0.3666	22.051	0.6641	0.5531	23.955	0.7660	0.8803
parrot	27.405	0.8329	0.9487	28.687	0.8700	0.9720	30.366	0.8888	0.9945
peppers	30.288	0.8619	0.9532	31.022	0.8905	0.9629	32.555	0.9107	0.9851
policemen	22.802	0.7504	0.6526	24.137	0.8170	0.7668	27.862	0.8906	0.9773
walkbridge	23.011	0.6389	0.7730	24.304	0.7376	0.8521	25.875	0.8247	0.9705

Table 4.2: Denoising results when noise level $\gamma = 10$.

Model	(a) Median filter			(b) TV model [10]			(d) Proposed model		
Image	PSNR	SSIM	QILV	PSNR	SSIM	QILV	PSNR	SSIM	QILV
barbara	24.837	0.7049	0.5407	25.410	0.7598	0.4820	28.006	0.8375	0.8479
boat	24.295	0.6368	0.7644	25.588	0.7277	0.7808	26.628	0.7503	0.8803
building	25.267	0.7105	0.7796	25.569	0.8012	0.6352	28.020	0.8559	0.9238
cameraman	25.198	0.6737	0.8438	26.283	0.8044	0.8197	28.298	0.8441	0.9603
castle	25.272	0.6685	0.7474	26.556	0.8126	0.6241	28.702	0.8546	0.8862
face	28.766	0.8259	0.9779	29.245	0.8705	0.9804	31.678	0.9058	0.9838
lake	24.182	0.6980	0.8801	25.473	0.7772	0.9006	26.486	0.8115	0.9477
lena	26.719	0.7134	0.9044	27.809	0.7991	0.8707	28.889	0.8321	0.9353
mandrill	20.167	0.4645	0.4145	21.259	0.5638	0.3158	22.182	0.6388	0.6242
parrot	25.819	0.7226	0.9449	27.067	0.8252	0.9423	28.200	0.8486	0.9738
peppers	27.920	0.7536	0.9491	29.191	0.8425	0.9454	30.245	0.8729	0.9597
policemen	22.194	0.6396	0.6746	23.070	0.7573	0.6063	25.625	0.8382	0.9101
walkbridge	22.330	0.5975	0.7954	23.241	0.6523	0.7249	24.235	0.7298	0.9051

can be seen that TGV improves TV by recovering more textures and outcoming no staircase artifacts. But, the inclusion of the sparse representation prior much enhances the denoising results, by better retaining textures while sufficiently denoising homogeneous regions. This shows the benefit of the patch-based prior. On the other hand, comparing DL+TGV with DL+TV, the denoised images look similar, but DL+TGV supplies cleaner textures, as shown in the zoomed images in Figure 3.4. Moreover, it diminishes the staircase artifacts that still appear in the denoised image of DL+TV, leading to the higher PSNR value. These can be more clearly observed as the noise level increases. In Figures 3.5 and 3.6, comparing the denoised images of DL+TV and DL+TGV when the noise level $\gamma = 10$, we can see that DL+TGV preserves more textures and details, such as the textural regions of Barbara and some parts of the tripod of Cameraman. Besides, DL+TGV provides more natural visual quality by ameliorating the staircase artifacts, as seen in the face of Barbara, and also gives higher PSNR values. Lastly, in Figures 3.7 and 3.8, we compare the denoising results of TV, TGV, DL+TGV, when the noise level is high, $\gamma = 10$ or 15. It can be apparently seen that TGV reduces the staircase

Table 4.3: Denoising results when noise level $\gamma = 15$.

Model	(a) Median filter			(b) TV model [10]			(d) Proposed model		
Image	PSNR	SSIM	QILV	PSNR	SSIM	QILV	PSNR	SSIM	QILV
barbara	22.011	0.5534	0.2791	24.826	0.7127	0.5200	26.409	0.7924	0.8302
boat	22.791	0.5488	0.3505	24.845	0.6817	0.7270	25.527	0.7061	0.7991
building	22.979	0.6371	0.3102	25.039	0.7565	0.6292	26.280	0.8029	0.7506
cameraman	23.147	0.6587	0.5587	25.710	0.7711	0.7919	27.071	0.8169	0.9125
castle	23.592	0.6539	0.3442	26.015	0.7820	0.5600	27.431	0.8256	0.7987
face	26.439	0.7346	0.9622	27.659	0.8321	0.9535	29.799	0.8761	0.9668
lake	22.256	0.6195	0.5756	24.522	0.7338	0.8419	25.316	0.7723	0.8939
lena	25.444	0.6896	0.6374	26.822	0.7624	0.7922	27.650	0.7959	0.8609
mandrill	19.797	0.3535	0.1133	20.956	0.5128	0.2312	21.644	0.5952	0.5738
parrot	23.482	0.7091	0.7048	26.066	0.7783	0.9289	26.858	0.8175	0.9395
peppers	26.634	0.7646	0.7565	27.955	0.8150	0.9023	28.887	0.8439	0.9239
policemen	20.097	0.5546	0.2689	22.506	0.7199	0.5142	24.331	0.7982	0.8278
walkbridge	21.131	0.4586	0.4142	22.930	0.6349	0.7342	23.218	0.6519	0.8083

artifacts emerged in the restored images of TV, and DL+TGV further improves TGV by better maintaining textures while sufficiently eliminating noise in the smooth regions. As a result, all these examples show the effectiveness of DL+TGV regularization comparing to the other prior terms. Particularly, even though the restored images of DL+TGV and DL+TV look visually similar, DL+TGV better keeps textures and details and furnishes more natural looking images, yielding higher PSNR values.

On the other hand, in Figure 3.9, we present the denoising results when using the L^1 data-fidelity term, $\lambda\|f - Ku\|_1$, with our DL+TGV regularization term, since the L^1 data-fitting term is well-known for removing impulse noise. The best denoised images are chosen to have the highest PSNR values. It can be seen that the denoised images using the L^1 fidelity term retain noise in all cases, unlike our results as given in Figures 3.3 (Barbara), 3.5 (Cameraman), 3.7 (Face). This also leads to higher PSNR values for our model than the model (3.2) using the L^1 data-fitting term. These examples demonstrate the effectiveness of our data-fitting term for eliminating Cauchy noise.

In the remaining numerical part, we compare our model with DL+TGV only to the existing TV model (1.3) since the models using TGV or DL+TV have not been proposed yet. In Figures 3.10-3.14, we present the denoising results of our model when the noise level $\gamma = 5, 10, \text{ or } 15$, respectively. We compare the results to those of the median filtering (MF) and TV model [10]. The size of the median filter is chosen as 3×3 when $\gamma = 5, 10$, and 5×5 when $\gamma = 15$.

Figure 3.2 displays noisy images corrupted by Cauchy noise with different noise levels $\gamma = 5$ (top row) and $\gamma = 10$ (bottom row). The red and blue curves, plotted in the last two figures, depict horizontal cross-sectional lines in the original Lena image and its noisy versions, respectively. It can be seen that the range for the original signal is $[0, 255]$, but those for the noisy signals are $[-200, 1800]$ (top row) and $[-1500, 600]$ (bottom row). These also illustrate the impulsive characteristics of Cauchy noise.

In Figures 3.10, we show the denoising results for noisy images given in Figure 3.2, when the noise level $\gamma = 5$. First, we can see that MF provides adequate denoising results, but it smooths edges and details even with a small size of the median filter. The TV model yields sharper edges than the MF, while adequately removing noise. However, some fine features are simultaneously lost, e.g., some parts of the tripod in the Cameraman image. However, comparing the results



Fig. 4.1. Deblurred and denoised images when data f degraded by a Gaussian blur and Cauchy noise with $\gamma = 5$. (a) TV model [10], (b) Our model. PSNR (top to bottom row): (a) 24.924/26.662/27.075, (b) 27.239/27.511/27.903. Parameter λ for our model: (top to bottom) 1500, 1400, 1350.

of these models to those of our model, our model significantly outperforms the others by well maintaining textures, details and edges. This can be clearly observed in the ropes and iron pillars in Boat, the face and camera areas in Cameraman, and the hair parts in Lena. Moreover, our model generates much cleaner homogeneous regions, such as background areas. Overall, our model produces decent denoising results with much cleaner and clearer images than the other models. Therefore, these example demonstrate the effectiveness of the combination of the sparse representation prior using patches and TGV prior for eliminating Cauchy noise.

In Figure 3.12, we show denoising results when the noise level $\gamma = 10$. Similarly, we can observe that the outcomes of our model show the best quality of denoised images, with conserved fine features and edges as well as adequately denoised smooth regions. As the noise level increases, the denoised images by the MF become noisier or smoother than those of TV model. However, TV still does not capture some parts of tricky structures such as continuous thin lines and textures, as seen in the zoomed images in 3.14. In contrast, our model not only



Fig. 4.2. Deblurred and denoised images when data f degraded by a uniform blur and Cauchy noise with $\gamma = 5$. (a) TV model [10], (b) Our model. PSNR (top to bottom row): (a) 24.485/28.159/25.037, (b) 25.403/28.858/25.575. Parameter λ for our model: (top to bottom) 1950, 1800, 1900.

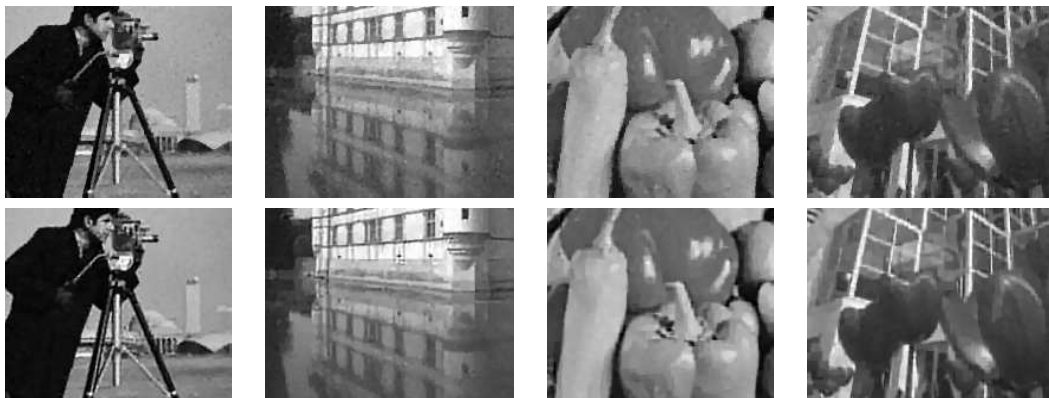


Fig. 4.3. Zoomed images of the restored images (top: TV, bottom: Proposed) in Figures 4.1 and 4.2.

Table 4.4: Deblurring-denoising results with different blur kernels when noise level $\gamma = 5$.

Blur	Gaussian blur with $\sigma = 1$		Uniform blur of size 5×5	
	(a) TV model [10]	(b) Proposed model	(a) TV model [10]	(b) Proposed model
Image	PSNR/SSIM/QILV	PSNR/SSIM/QILV	PSNR/SSIM/QILV	PSNR/SSIM/QILV
barbara	24.924/0.7325/0.3811	27.239/0.8036/0.8219	24.485/0.6735/0.4523	25.403/0.7210/0.6639
boat	25.637/0.7252/0.7890	26.021/0.7452/0.8446	24.493/0.6584/0.6565	24.998/0.6819/0.7251
building	26.489/0.8078/0.8322	27.401/0.8518/0.9003	25.037/0.7552/0.6846	25.575/0.7874/0.7580
cameraman	26.662/0.8047/0.8786	27.511/0.8418/0.9369	25.367/0.7723/0.7790	26.128/0.8057/0.8537
castle	27.075/0.8059/0.7767	27.903/0.8527/0.8793	26.021/0.7748/0.6632	26.774/0.8194/0.8259
face	30.954/0.9049/0.9894	33.031/0.9264/0.9932	28.935/0.87104/0.9682	30.344/0.8942/0.9755
lake	25.601/0.7837/0.9208	26.052/0.8080/0.9455	24.237/0.7298/0.8557	24.595/0.7500/0.8688
lena	28.375/0.8123/0.9148	28.776/0.8341/0.9338	26.988/0.7661/0.8721	27.417/0.7891/0.8494
mandrill	21.362/0.5530/0.3268	21.521/0.5718/0.4183	20.750/0.4656/0.2083	20.758/0.4568/0.1915
parrot	27.409/0.8318/0.9633	27.952/0.8604/0.9816	25.535/0.7962/0.9083	25.952/0.8226/0.9175
peppers	30.055/0.8608/0.9681	30.446/0.8797/0.9726	28.159/0.8235/0.9413	28.858/0.8500/0.9347
policemen	23.590/0.7630/0.7379	24.433/0.8182/0.8791	22.318/0.6990/0.5682	22.897/0.7513/0.7179
walkbridge	23.540/0.6685/0.8133	23.618/0.6803/0.8459	22.661/0.5825/0.7267	22.814/0.5821/0.7289

keeps such fine structures, but furnishes much cleaner homogeneous regions. These also show the superior performance of our model over the other models.

Figure 3.13 presents the denoising results of our model when the noise level $\gamma = 15$, and compares them to those of the TV model. First we can see in all examples that our model sufficiently denoises the homogeneous regions and also alleviates the staircasing artifacts that emerge in the restored images of the TV model, leading to much cleaner and more natural looking images. Moreover, our model tends to preserve delicate features and textural parts, as seen in the land areas in Lake, hairs in Mandrill and tower areas in Policemen, while the TV model over-smoothes them. This can be more clearly seen in the zoomed images in Figure 3.14. Hence, our model achieves satisfactory denoising results even when the noise level is high.

Figure 3.15 shows learned dictionaries with patch size 4×4 when the noise level $\gamma = 5, 10, 15$, respectively. We can obviously see that the learned dictionaries in all noise cases contain more relevant image patches to the original image than the initial DCT atoms. Furthermore, in the top row of Figure 3.16, the plots of the PSNR values via the outer iteration number are presented. This implies that in all noise cases, the PSNR values converge as the iteration progresses, and 9 outer iterations are sufficient to reach the stopping criterion. Indeed, in all experiments, the outer iterations are stopped after $M = 9$ (for denoising) or 10 (for deblurring-denoising). Figure 3.16 also shows in the bottom row the descent of the relative error of u and the penalty terms, $\|p - u\|_2^2$ and $\|z - Au\|_2$, as the outer iteration increases. This illustrates the convergence behavior of our iterative algorithm although it is not theoretically proven.

Lastly, in Tables 4.1-4.3, we report the PSNR, SSIM and QILV values of the restored images. The proposed model supplies the highest PSNR, SSIM and QILV values in all cases. Overall, our model provides the best denoising results with respect to these image quality measurements, and these results also establish the outstanding performance of our model over filtering-based or local derivative-based regularization models.

Despite the outstanding performance, our model takes 2-3 minutes, whereas the TV model takes around 5 seconds in MATLAB R2018a running on a 64-bit Windows 10 PC with an Intel i7-7700K CPU @4.20 GHz and 64 GB RAM. Reducing the computational time is a remaining issue for the DL-based approach.

4.2. Image Deblurring-Denoising results

In Figures 4.1 and 4.2, we present the deblurring-denoising results of our model with comparisons to the TV model only, since the MF is not appropriate for deblurring. The data in Figures 4.1 and 4.2 are blurred by the Gaussian blur and uniform blur, respectively, and are then polluted by Cauchy noise with noise level $\gamma = 5$.

First, we can observe from the Barbara images in both figures that our model preserves much texture in contrast to the local derivative-based model, which is due to the use of the patch-based prior that considers image structural patterns. It can be also seen from all examples that the sparse representation prior assists in reducing the staircasing artifacts emerged in the restored images of the TV model. This can be recognized in the smoothly varying areas, such as the face in Barbara and the sky and lake in Castle. On the other hand, unlike the denoising results, some blocking artifacts remain. Furthermore, our model conserves edges as well as the TV model, and also generates the cleaner restored images with sufficiently denoised smooth regions. These can be observed in the zoomed images in Figure 4.3. Thus, these examples illustrate that our model is also capable to efficiently restore images degraded by both blurring and Cauchy noise.

In Table 4.4, the PSNR, SSIM, and QILV values for our model and TV model are reported. In all cases our model brings the highest PSNR values, but the SSIM or QILV values are occasionally slighter smaller than those of the TV model, especially when the blurring is strong. Nonetheless, these measurements also demonstrate the effectiveness of our model for deblurring images in the presence of Cauchy noise, as compared to the TV model.

5. Conclusion

In this paper, we proposed an image restoration model in the presence of Cauchy noise and/or blurring. The model is constituted from a nonconvex data-fidelity term derived from the statistics of the Cauchy distribution and two prior terms that are reciprocally complementary: a patch-based sparse representation prior over learned dictionaries and TGV regularization. Specifically, the sparse representation prior using image patches helped sufficiently denoise in smooth regions while preserving textural patterns and fine details. TGV regularization further facilitated effective denoising in smooth regions while maintaining edges. Besides, TGV alleviated some blocking artifacts driven from the patch-based prior. To solve the proposed model, we utilized a variable splitting scheme, the penalty method and an alternating minimization strategy, which resulted in an efficient iterative algorithm. The K-SVD algorithm was also employed to learn dictionaries from the processed images. The experimental results showed that the proposed model outcome more distinguished restoration results than other existing models, when it comes to visual aspect and image quality measurements. However, despite the numerical convergence of the proposed algorithm, its theoretical convergence analysis is a remaining issue. Moreover, the blind Cauchy denoising and/or deblurring with the unknown noise level γ is another challenging problem, which will be investigated in a future work.

Acknowledgments. Miyoun Jung was supported by Hankuk University of Foreign Studies Research Fund and the NRF (2017R1A2B1005363). Myungjoo Kang was supported by the NRF (2015R1A15A1009350, 2017R1A2A1A17069644).

References

- [1] E. Kuruoglu, W. Fitzgerald and P. Rayner, Near optimal detection of signals in impulsive noise modeled with asymmetric alpha-stable distribution, *IEEE Commun. Lett.*, **2**:10 (1998), 282–284.
- [2] M. Idan and J. Speyer, Cauchy estimation for linear scalar systems, *IEEE Trans. Automatic Control*, **55** (2010), 1329–1342.
- [3] Y. Peng, J. Chen, X. Xu and F. Pu, Sar images statistical modeling and classification based on the mixture of alpha-stable distributions, *Remote Sens.*, **5**:5 (2013), 2145–2163.
- [4] W. Feller, An Introduction to Probability Theory and Its Applications, volume II, New York: John Wiley & Sons Inc., 2 ed. edition, 1971.
- [5] N. Johnson, S. Kotz and N. Balakrishnan, Continuous Univariate Distributions, New York: Wiley, 1994.
- [6] Y. Chang, S. Kadaba, P. Doerschuk and S. Gelfand, Image restoration using recursive markov random field models driven by cauchy distributed noise, *IEEE Signal Process. Lett.*, **8**:3 (2001), 65–66.
- [7] A. Achim and E. Kuruoğlu, Image denoising using bivariate k-stable distributions in the complex wavelet domain, *IEEE Signal Process. Lett.*, **12**:1 (2005), 17–20.
- [8] A. Loza, D. Bull, N. Canagarajah and A. Achim, Non-gaussian model-based fusion of noisy images in the wavelet domain, *Comput. Vis. Image Underst.*, **114**:1 (2010), 54–65.
- [9] T. Wan, N. Canagarajah and A. Achim, Segmentation of noisy colour images using cauchy distribution in the complex wavelet domain, *IET Image Processing*, **5**:2 (2011), 159–170.
- [10] F. Sciacchitano, Y. Dong and T. Zeng, Variational approach for restoring blurred images with cauchy noise, *SIAM J. Imag. Sci.*, **8**:3 (2015), 1894–1922.
- [11] L. Rudin, S. Osher and E. Fatemi, Nonlinear total variation based noise removal algorithm, *Phys. D*, **60**:1-4 (1992), 259–268.
- [12] J. Mei, Y. Dong, T. Hunag and W. Yin, Cauchy noise removal by nonconvex admm with convergence guarantees, *J. Sci. Comput.*, **74**:2 (2018), 743–766.
- [13] Y. Wang, W. Yin and J. Zeng, Global convergence of admm in nonconvex nonsmooth optimization, *J. Sci. Comput.*, **78**:1 (2019), 29–63.
- [14] A. Chambolle and P. Lions, Image recovery via total variation minimization and related problems, *Numerische Mathematik*, **76**:2 (1997), 167–188.
- [15] T. Chan, A. Marquina and P. Mulet, High-order total variation-based image restoration, *SIAM J. Sci. Comput.*, **22**:2 (2000), 503–516.
- [16] M. Lysaker, A. Lundervold and X. Tai, Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time, *IEEE Trans. image process.*, **12**:12 (2003), 1579–1590.
- [17] F. Li, C. Shen, J. Fan and C. Shen, Image restoration combining a total variational filter and a fourth-order filter, *J. Vis. Commun. Image Represent.*, **18**:4 (2007), 322–330.
- [18] K. Bredies, K. Kunisch and T. Pock, Total generalized variation, *SIAM J. Imag. Sci.*, **3**:3 (2010), 492–526.
- [19] K. Bredies, K. Kunisch and T. Pock, Inverse problems with second-order total generalized variation constraints, Proceedings of SampTA 2011 - 9th International Conference on Sampling Theory and Applications, Singapore, 2011, IEEE.
- [20] D. Donoho and I. Johnstone, Ideal spatial adaptation by wavelet shrinkage, *Biometrika*, **81**:3 (1994), 425–455.
- [21] I. Daubechies, M. Defriese and C. Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Commun. Pure Appl. Math.*, **57**:11 (2004), 1413–1457.
- [22] B. Dong, H. Ji, J. Li, Z. Shen and Y. Xu, Wavelet frame based blind image inpainting, *Appl. Comput. Harmon. Anal.*, **32**:2 (2012), 268–279.
- [23] M. Figueiredo, J. Bioucas-Dias and R. Nowak, Majorization-minimization algorithms for wavelet-

- based image restoration, *IEEE Trans. Image Process.*, **16**:12 (2007), 2980–2991.
- [24] T. Zeng, X. Li and M. Ng, Alternating minimization method for total variation based wavelet shrinkage model, *Commun. Comput. Phys.*, **8**:5 (2010), 976–994.
- [25] A. Buades, B. Coll and J. Morel, A review of image denoising algorithms with a new one, *Multiscale Model. Simul.*, **4**:2 (2005), 490–530.
- [26] G. Gilboa and S. Osher, Nonlocal operators with applications to image processing, *SIAM J. Multiscale Model. Simul.*, **7**:3 (2008), 1005–1028.
- [27] F. Dong, H. Zhang and D.X. Kong, Nonlocal total variation models for multiplicative noise removal using split bregman iteration, *Math. Comput. Model.*, **5**:3-4 (2012), 939–954.
- [28] H. Hu, B. Li and Q. Liu, Removing mixture of gaussian and impulse noise by patch-based weighted means, *J. Sci. Comput.*, **67**:1 (2016), 103–129.
- [29] T.-F.Chan and H.-M.Zhou, Adaptive eno-wavelet transforms for discontinuous functions, 12th International Conference on Domain Decomposition Methods, pages 93–100, Chiba, Japan, 2001.
- [30] M. Aharon, M. Elad and A. Bruckstein, K-svd: An algorithm for designing of overcomplete dictionaries for sparse representation, *IEEE Trans. Signal Process.*, **54**:11 (2006), 4311–4322.
- [31] M. Elad and M. Aharon, Image denoising via sparse and redundant representations over learned dictionaries, *IEEE Trans. Image Process.*, **15**:12 (2006), 3736–3745.
- [32] M. Protter and M. Elad, Image sequence denoising via sparse and redundant representations, *IEEE Trans. Image Process.*, **18**:1 (2009), 27–35.
- [33] Y.Q. Zhao and J. Yang, Hyperspectral image denoising via sparse representation and low-rank constraint, *IEEE Trans. Geosci. Remote Sens.*, **53**:1 (2015), 296–308.
- [34] J. Yang, Y.Q. Zhao, J.W. Chan and et al., Coupled sparse denoising and unmixing with low rank constraint for hyper-spectral image, *IEEE Trans. Geosci. Remote Sens.*, **54**:3 (2016), 1818–1833.
- [35] Y. Lou, A. Bertozzi and S. Soatto, Direct sparse deblurring, *J. Math. Imag. Vis.*, **39**:1 (2011), 1–12.
- [36] H. Zhang, J. Yang, Y. Zhang and T. Huang, Sparse representation based blind image deblurring, Proc. IEEE Int. Conf. Multimedia Expo, pages 1–6, Barcelona, Spain, 2011, IEEE.
- [37] M. Kaleem, A. Guergachi and S. Krishnan, Empirical mode decomposition based sparse dictionary learning with application to signal classification, Digital Signal Processing and Signal Processing Education Meeting, pages 18–23, Napa, CA, USA, 2013, IEEE.
- [38] Y. Huang, L. Moisan, M. Ng and T. Zeng, Multiplicative noise removal via a learned dictionary, *IEEE Trans. Image Process.*, **21**:11 (2012), 4534–4543.
- [39] S. Li, G. Wang and X. Zhao, Multiplicative noise removal via adaptive learned dictionaries and tv regularization, *Digital Signal Processing*, **50** (2016), 218–228.
- [40] Y. Xiao and T. Zeng, Poisson noise removal via learned dictionary, 17th IEEE International Conference on Image Processing (ICIP), pages 1177–1180, Hong Kong, China, 2010, IEEE.
- [41] L. Ma, L. Moisan, J. Yu and T. Zeng, A dictionary learning approach for poisson image deblurring, *IEEE Trans. Med. Imag.*, **32**:7 (2013), 1277–89.
- [42] Y. Shi, H. Fang and G. Wang, Poissonian image deconvolution via sparse and redundant representations and framelet regularization, *Math. Probl. Eng.*, **2014** (2014).
- [43] L. Ma, J. Yu and T. Zeng, Sparse representation prior and total variation based image deblurring under impulse noise, *SIAM J. Imag. Sci.*, **6**:4 (2013), 2258–2284.
- [44] J. Nolan, Numerical calculation of stable densities and distribution functions, *Communications in statistics. Stochastic models*, **13**:4 (1997), 759–774.
- [45] N. Balakrishnan and V. Nevzorov, A primer on statistical distributions, New York: John Wiley & Sons, 2003.
- [46] W. Dong, L. Zhang, G. Shi and X. Wu, Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization, *IEEE Trans. Image Process.*, **20**:7 (2011), 1838–1857.
- [47] Y. Wang, J. Yang, W. Yin and Y. Zhang, A new alternating minimization algorithm for total

- variation image reconstruction, *SIAM J. Imag. Sci.*, **1**:3 (2008), 248–272.
- [48] Y. Huang, M. Ng and Y. Wen, A new total variation method for multiplicative noise removal, *SIAM J. Imag. Sci.*, **2**:1 (2009), 20–40.
- [49] T. Goldstein and S. Osher, The split bregman method for l^1 -regularized problems, *SIAM J. Imag. Sci.*, **2**:2 (2009), 323–343.
- [50] J. Bioucas-Dias and M. Figueiredo, Multiplicative noise removal using variable splitting and constrained optimization, *IEEE Trans. Image Process.*, **19**:7 (2010), 1720–1730.
- [51] R. Courant, Variational methods for the solution of problems of equilibrium and vibrations, *Bull. Amer. Math. Soc.*, **49** (1943), 1–23.
- [52] R. Fletcher, *Mathematical Programming The State of the Art: Penalty functions*, Springer, 1983.
- [53] I. Csiszár and G. Tusná, Information geometry and alternating minimization procedures, *Statistics and Decisions*, **1** (1984), 205–237.
- [54] Y. Xug and W. Yin, A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion, *SIAM J. Imag. Sci.*, **6**:3 (2013), 1758–1789.
- [55] Y. Pati, R. Rezaifar and P. Krishnaprasad, Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition, *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, pages 40–44, Pacific Grove, CA, USA, 1993, IEEE.
- [56] W. Guo, J. Qin and W. Yin, A new detail-preserving regularization scheme, *SIAM J. Imag. Sci.*, **7**:2 (2014), 1309–1334.
- [57] J. Nocedal and S. Wright, *Numerical Optimization*, Springer, 2006.
- [58] Z. Wang, A. Bovik, H. Sheikh and E. Simoncelli, Image quality assessment: From error visibility to structural similarity, *IEEE Trans. Image Process.*, **13**:4 (2004), 600–612.
- [59] S. Aja-Ferández, R. Estépar, C. Alberola-López and C. Westin, Image quality assessment based on local variance, *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, pages 4815–4818, New York, NY, USA, 2006, IEEE.
- [60] W. Deng and W. Yin, On the global and linear convergence of the generalized alternating direction method of multipliers, *J. Sci. Comput.*, **66** (2016), 889–916.