이학박사 학위논문

Group sparse representation for restoring images with non-Gaussian noise

(비가우시안 잡음 영상 복원을 위한 그룹 희소 표현)

2020년 2월

서울대학교 대학원

수리과학부

이상원

Group sparse representation for restoring images with non-Gaussian noise

(비가우시안 잡음 영상 복원을 위한 그룹 희소 표현)

지도교수 강명주

이 논문을 이학박사 학위논문으로 제출함

2019년 10월

서울대학교 대학원

수리과학부

이상원

이상원의 이학박사 학위논문을 인준함

2019년 12월

위 원 장 _____ (인)

부 위 원 장 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

Group sparse representation for restoring images with non-Gaussian noise

A dissertation
submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
to the faculty of the Graduate School of
Seoul National University

by

Sangwon Lee

Dissertation Director : Professor Myungjoo Kang

Department of Mathematical Science
Seoul National University

February 2020

Abstract

For the image restoration problem, recent variational approaches exploiting nonlocal information of an image have demonstrated significant improvements compared with traditional methods utilizing local features. Hence, we propose two variational models based on the sparse representation of image groups, to recover images with non-Gaussian noise. The proposed models are designed to restore image with Cauchy noise and speckle noise, respectively. To achieve efficient and stable performance, an alternating optimization scheme with a novel initialization technique is used. Experimental results suggest that the proposed methods outperform other methods in terms of both visual perception and numerical indexes.

Key words: Image restoration, Image denoising, Image deblurring, Speckle noise, Cauchy noise, Group sparse representation, Alternating direction method of multipliers
Student Number: 2014-21191

Contents

Chapter 1

Introduction

Image restoration is the most typical problem in imaging processing. It considers the recovery of an image from its inevitable degraded observation in real applications. Image restoration includes various problems: image denoising [9, 13], deblurring [10, 22, 46], inpainting [12, 47], compressive sensing [29, 65], and super-resolution [11, 63]. The degradation process consists of the application of a non-invertible operator, e.g. blurring or masking operator, and the contamination with noise which is usually assumed to follow the Gaussian distribution.

The image restoration is an ill-posed inverse problem, hence the prior assumption is needed to specify the solution. This prior assumption, which is called a regularizer, utilizes the knowledge about desirable image characteristics. The variational model reconstructs the original image as a solution of the minimization problem of some functional, which usually consists of a data fidelity term and a regularization term.

Classically, the regularization based on the local property of the image characteristic was widely used. The most popular one is the total variation (TV) regularization, which was originated from the Rudin-Osher-Fatemi

model [50]. It assumes that the gradient of the image has small ℓ_1 norm, which results in successful recovery of smooth regions and preservation of edges. However, it has a drawback of unwanted staircase artifacts and overly smooth details. The local patch based regularization also gained much attention. Elad and Aharon [25] proposed a denoising algorithm based on the assumption that local patches are sparsely represented by the learned dictionaries. Although it achieved remarkable results, the training process of dictionary is burdensome.

Methods exploiting the nonlocal properties of images to improve the performance of the local regularizer have been studied extensively. Buades et al. [8] proposed a nonlocal means algorithm which improved the local smoothing filters. Dabov et al. [16] used block matching and three-dimensional transform domain collaborative filtering for the image denoising problem. Dong et al. [23] exploited the nonlocal information of images to centralize the sparse coding coefficents in problems of image denoising, deblurring, and super-resolution.

Recently, Zhang et al. [66] improved the traditional patch sparse representation model to the group sparse representation (GSR) modeling of images. By using the local sparsity of patches and nonlocal self-similarity of repeated structures and patterns between patches, it achieved state-of-the-art results in image deblurring, inpainting, and compressive sensing. Additionally, the GSR model was utilized in various applications including image deblocking [67], low lighting image enhancement [55], and synthetic aperture radar (SAR) image despeckling [39].

As discussed, most works on the image restoration mainly focus on the problem of the Gaussian noise. However, various non-Gaussian noise distributions have been considered as well, including Cauchy distribution [43, 52], Gamma distribution [4, 48, 54], and Rician distribution [5, 28]. In fact, these non-Gaussian distributions can better represent the real noises in some applications. For example, the Cauchy distribution is known to better repre-

sent the heavy-tailed property of noise in radar and sonar applications or biomedical ultrasound images [35, 36, 51]. Also the speckle noise in SAR imagery, laser, and ultrasound can be modeled by using the Gamma distribution [2, 57, 62].

Herein, we focus on the image restoration problem of denoising and deblurring images in the presence of non-Gaussian noises, especially the Cauchy noise or the speckle noise. In these cases, the variational models designed for the Gaussian noise are not suitable, and should be adapted to handle different types of noises. Especially the data fidelity term, which is usually computed with the ℓ_2 norm, has to be changed to consider the different noise distributions. The derivation of appropriate data fidelity term can be made based on the Bayesian statistics, specifically the maximum a posteriori probability (MAP) estimation.

In addition, compared with the uniform amplitude of the Gaussian noise, the non-Gaussian noises have a rather irregular characteristic. The local regularization based methods have trouble in distinguishing the sharp noise from the details of the image. On the other hand, GSR exploits the local sparsity and the nonlocal self-similarity of the image patches, which is expected to effectively capture the non-uniform property of the non-Gaussian noises.

Hence, in this dissertation, we adopt GSR approaches and propose two models for restoring images containing Cauchy noise and speckle noise, respectively. An alternating minimization scheme is utilized and self-adaptive group dictionaries are learned for efficient optimization of the proposed models. To overcome the instability of the nonconvex model, an initialization technique is introduced. From the experimental results, we observe that the proposed models obtained improvements compared with existing methods.

The remainder of this dissertation is organized as follows. In Chapter 2, we provide introductions and literature reviews for the Cauchy noise and the speckle noise, and introduce the GSR model and the ADMM algorithm. In Chapter 3, we propose two models based on GSR for Cauchy noise and

speckle noise, respectively. Numerical experiments are provided to demonstrate the performance of the proposed models, and the relevant technique and analyses are discussed. In Chapter 4, we summarize this dissertation.

Chapter 2

Preliminaries

In this chapter, the required preliminaries for this dissertation is provided. We first introduce mathematical descriptions of Cauchy noise and speckle noise, and review the related works for restoring these kinds of noises. Subsequently, the GSR framework for image restoration is briefly introduced and the ADMM algorithm is presented, which are necessary to understand our proposed models.

2.1 Cauchy Noise

In this section, we introduce the Cauchy noise, which is an additive noise following the Cauchy distribution. The Cauchy noise has quite different feature compared with the additive Gaussian noise, which is explained with examples. Consequently, the adapted models are required to deal with this kind of noise accurately. We investigate the related works for the Cauchy noise in the image processing area.

2.1.1 Introduction

In the image restoration problem, the degradation of an image can be mathematically formulated as

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}, \tag{2.1}$$

where $\boldsymbol{x}$ is the unknown original image, $\boldsymbol{y}$ is the observed noisy image, $\boldsymbol{H}$ is the linear degradation operator, and $\boldsymbol{n}$ is the additive noise. Herein, we focus on the Cauchy noise, which is defined as the additive noise following the Cauchy distribution. The probability density function of the Cauchy distribution is given by [52]

$$p(x) = \frac{1}{\pi}\frac{\gamma}{\gamma^2 + (x-\delta)^2}, \tag{2.2}$$

where $\gamma > 0$ is the scale parameter acting as a variance; $\delta \in \mathbb{R}$ is the localization parameter that represents the median and may assumed as 0.

In Fig. 2.1, the probability density functions of Gaussian distribution and Cauchy distribution are plotted. The Gaussian distribution follows $\mathcal{N}(\mu, \sigma^2)$ with mean $\mu = 0$ and variance $\sigma^2 = \frac{\pi}{2}$, and the Cauchy distribution follows $\mathcal{C}(\gamma, \delta)$ with scale parameter $\gamma = 1$ and localization parameter $\delta = 0$. As shown, the peak values at the origin coincide but the Cauchy distribution has higher values on tails implying the impulsive characteristic.

In Fig. 2.2, the original cameraman image (Fig. 2.2a) is degraded by Gaussian noise (Fig. 2.2b) and Cauchy noise (Fig. 2.2c), respectively. Figs. 2.2d to 2.2f are the zoomed-in of the arm in the Figs. 2.2a to 2.2c, respectively. The parameters of the noises are adjusted so that the two noisy images have the same mean square error with the original image. Explicitly, the Gaussian noise has mean $\mu = 0$ and variance $\sigma^2 = 29.5^2$, and the Cauchy noise has scale parameter $\gamma = 5$ and localization parameter $\delta = 0$. As shown, the Gaussian noise has uniform intensity in both black and gray area (Fig. 2.2e).

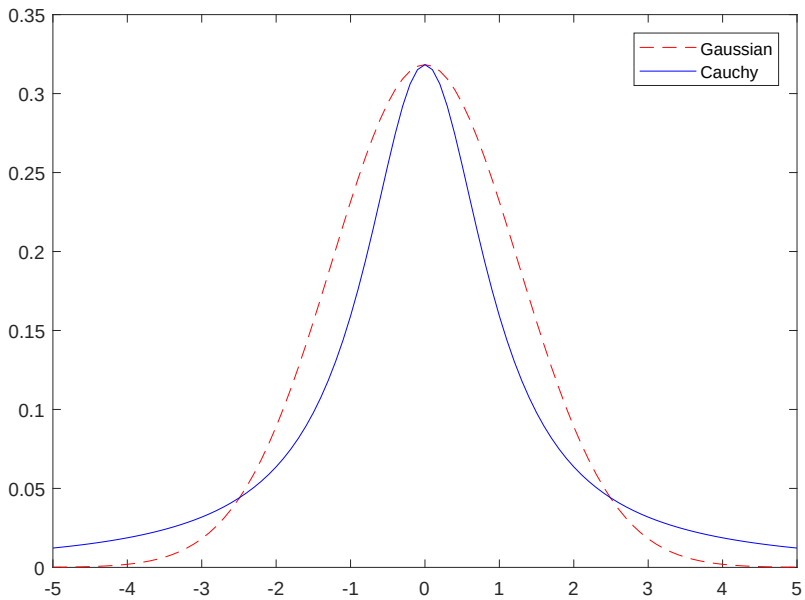


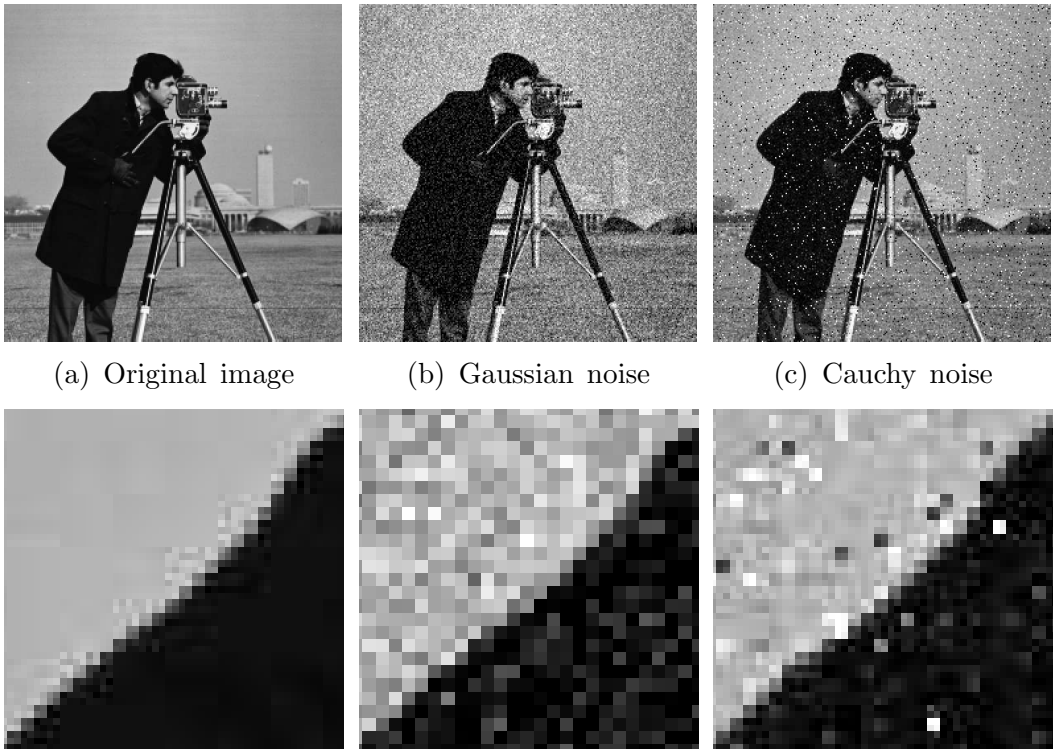
Figure 2.1: Comparison of the probability density functions of the Gaussian distribution $\mathcal{N}(\mu = 0, \sigma^2 = \frac{\pi}{2})$ and Cauchy distribution $\mathcal{C}(\gamma = 1, \delta = 0)$.

On the other hand, the Cauchy noise is rather impulsive so the overall noise intensity is low but it has very sharp noise, which can be seen in black points in gray region and white points in black region (Fig. 2.2f).

2.1.2 Literature Review

Many studies have been performed to mitigate Cauchy noise, especially in the wavelet domain. Achim and Kuruoglu [1] utilized bivariate isotropic Cauchy and Gaussian distributions in a complex wavelet domain for image denoising. Bhuiyan et al. [6] used a Cauchy probability density function as a prior for the wavelet coefficients of log-transformed speckle noise in synthetic aperture radar images. Loza et al. [41] described a multimodal image fusion algorithm based on the non-Gaussian modeling of wavelet coefficients.

Recently, variational models based on TV regularizer have garnered sig-



(a) Original image (b) Gaussian noise (c) Cauchy noise

(d) Zoom of (a) (e) Zoom of (b) (f) Zoom of (c)

Figure 2.2: Comparison of different noisy images. (a) Original image; (b) Gaussian noise ($\sigma^2 = 29.5^2$); (c) Cauchy noise ($\gamma = 5$); (d)-(f) Zoomed-in versions of (a)-(c), respectively.

nificant attention. The discrete TV of an image $\boldsymbol{u} \in \mathbb{R}^{n\times n}$ is defined by

$$\begin{aligned}\|\nabla \boldsymbol{u}\|_1 &= \sum_{1\le i,j\le n} \sqrt{(\nabla_x \boldsymbol{u})_{i,j}{}^2 + (\nabla_y \boldsymbol{u})_{i,j}{}^2} \\ &= \sum_{1\le i,j\le n} \sqrt{(\boldsymbol{u}_{i+1,j} - \boldsymbol{u}_{i,j})^2 + (\boldsymbol{u}_{i,j+1} - \boldsymbol{u}_{i,j})^2}. \end{aligned} \tag{2.3}$$

Sciacchitano et al. [52] first utilized the TV regularizer to restore images

corrupted by Cauchy noise,

$$\arg\min_{\boldsymbol{u}} \|\nabla \boldsymbol{u}\|_1 + \frac{\lambda}{2}\Big(\big\langle \log\big(\gamma^2 + (\boldsymbol{H}\boldsymbol{u} - \boldsymbol{y})^2\big), \mathbf{1}\big\rangle + \mu\|\boldsymbol{H}\boldsymbol{u} - \boldsymbol{u}_0\|_2^2\Big), \quad (2.4)$$

where $\langle \cdot, \cdot \rangle$ denotes the Frobenius inner product, $\mathbf{1} \in \mathbb{R}^{n\times n}$ is a matrix of ones, $\boldsymbol{u}_0 = med(\boldsymbol{y})$ is the median filtered version of $\boldsymbol{y}$, the second term is the data fidelity term for Cauchy noise, and the third term has been added to guarantee the convexity of (2.4). Mei et al. [43] applied the nonconvex alternating direction method of multipliers (ADMM) to directly solve the nonconvex TV model

$$\arg\min_{\boldsymbol{u}} \|\nabla \boldsymbol{u}\|_1 + \frac{\lambda}{2}\big\langle \log\big(\gamma^2 + (\boldsymbol{H}\boldsymbol{u} - \boldsymbol{y})^2\big), \mathbf{1}\big\rangle. \quad (2.5)$$

Although the TV regularizer can preserve sharp edges and restore smooth images, it suffers from staircase artifacts and oversmoothing. Improvements have been achieved by considering adaptive higher-order TV [64] or overlapping group sparsity of TV [21]. Laus et al. [37] suggested a nonlocal myriad filter for estimating the localization and the scale parameter of the Cauchy distribution, and proposed an unsupervised image denoising method.

2.2 Speckle Noise

In this section, we introduce the speckle noise, which is a multiplicative noise following the Gamma distribution. The speckle noise is fundamentally different from the common additive Gaussian noise or the additive Cauchy noise introduced in Section 2.1. It is a signal-dependent noise, which is caused by the multiplicative process in the noise generation. Hence, it is more difficult to handle this kind of noise and the adapted models are required. We also

provide the literature review for the speckle noise in the related area.

2.2.1 Introduction

Unlike the image degradation with additive noise introduced in (2.1), the multiplicative noise degradation can be formulated as

$$\boldsymbol{y} = \boldsymbol{x} * \boldsymbol{n}, \tag{2.6}$$

where $\boldsymbol{x}$ is the unknown original image, $\boldsymbol{y}$ is the observed noisy image, and $\boldsymbol{n}$ is the multiplicative noise. Here, we assume that the image is degraded by the noise only, not by the linear degradation operator. This type of noises occurs in many applications. For instance, Possion distribution is suitable for modeling the counting processes in medical images such as PET, SPECT, and fluorescent confocal microscopy imaging [49]. Also Gamma distribution can be used for modeling the speckle noise in SAR imagery, laser, and ultrasound [2, 57, 62].

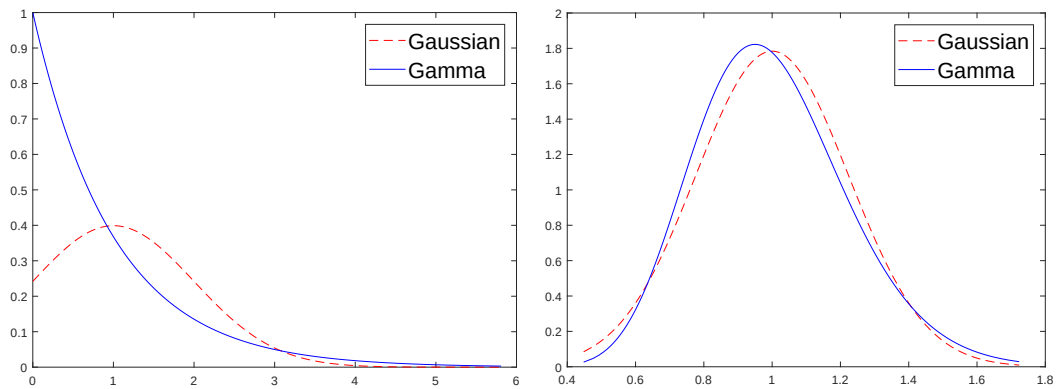
Herein, we focus on the speckle noise, which is the multiplicative noise following the Gamma distribution. The probability density function of the Gamma distribution is given by [24]

$$p(x; k, \theta) = \begin{cases} \dfrac{x^{k-1} e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)} & \text{for } x \geq 0, \\ 0 & \text{for } x < 0, \end{cases} \tag{2.7}$$

where Γ is the Gamma function satisfying $\Gamma(n) = (n-1)!$ for positive integers n, $k > 0$ is the shape parameter, and $\theta > 0$ is the scale parameter. By the central limit theorem, it is known that for a large k, the Gamma distribution approximates the Gaussian distribution with mean $\mu = k\theta$ and variance $\sigma^2 = k\theta^2$ [33].

Explictly, the speckle noise in SAR image with L looks can be modeled by the Gamma distribution with $k = L$ and $\theta = 1/L$, so that mean $\mu = k\theta = 1$ and variance $\sigma^2 = k\theta^2 = 1/L$ [4, 62]

$$p(x) = \begin{cases} \dfrac{L^L x^{L-1} e^{-Lx}}{\Gamma(L)} & \text{for } x \geq 0, \\ 0 & \text{for } x < 0. \end{cases} \tag{2.8}$$



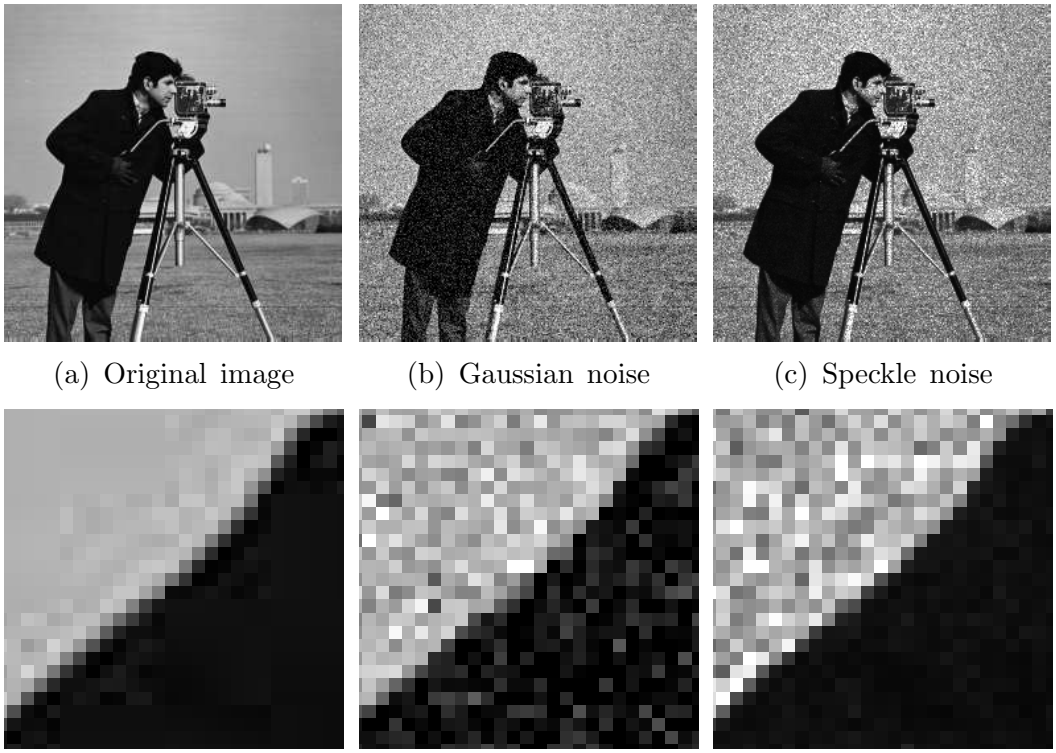
(a) $L = 1$; Gamma distribution $\Gamma(k = 1, \theta = 1)$, Gaussian distribution $\mathcal{N}(\mu = 1, \sigma^2 = 1)$.

(b) $L = 20$; Gamma distribution $\Gamma(k = 20, \theta = 1/20)$, Gaussian distribution $\mathcal{N}(\mu = 1, \sigma^2 = 1/20)$.

Figure 2.3: Comparison of the probability density functions of the Gamma distributions and the Gaussian distributions.

In Fig. 2.3, the probability density functions of the Gaussian distributions and the Gamma distributions are compared. In Figs. 2.3a and 2.3b, the Gamma distribution $\Gamma(k = L, \theta = 1/L)$ and the Gaussian distribution $\mathcal{N}(\mu = 1, \sigma^2 = 1/L)$ are plotted for $L = 1$ and $L = 20$, respectively. As shown, the Gamma distribution approximates the Gaussian distribution for high values of L, but it has completely different shape with skewness, especially in low values of L.

In Fig. 2.4, the original cameraman image (Fig. 2.4a) is degraded by the



(a) Original image (b) Gaussian noise (c) Speckle noise

(d) Zoom of (a) (e) Zoom of (b) (f) Zoom of (c)

Figure 2.4: Comparison of different noisy images. (a) Original image; (b) Additive Gaussian noise ($\sigma^2 = 29.8^2$); (c) Speckle noise ($L = 21$); (d)-(f) Zoomed-in versions of (a)-(c), respectively.

additive Gaussian noise (Fig. 2.4b) and the speckle noise (Fig. 2.4c), respectively. Figs. 2.4d to 2.4f are the zoomed-in of the arm in the Figs. 2.4a to 2.4c, respectively. The parameters of the noises are adjusted so that the two noisy images have the same mean square error with the original image. Explicitly, the Gaussian noise has mean $\mu = 0$ and variance $\sigma^2 = 29.8^2$, and the speckle noise has look $L = 21$. As shown, the Gaussian noise has uniform intensity in both black and gray areas (Fig. 2.4e). On the other hand, the speckle noise is signal-dependent so it has weak noise in black region and strong noise in gray region (Fig. 2.4f).

2.2.2 Literature Review

Many researches have been conducted to remove the speckle noise based on the TV regularization, which are well summarized in [44, 60] and briefly described in this subsection. Aubert and Aujol (AA) [4] first introduced the TV regularization model for removing speckle noise by using a MAP estimator

$$\arg\min_{\boldsymbol{u}>0} \|\nabla \boldsymbol{u}\|_1 + \lambda\Big\langle \log \boldsymbol{u} + \frac{\boldsymbol{y}}{\boldsymbol{u}}, \mathbf{1}\Big\rangle, \tag{2.9}$$

where $\langle \cdot, \cdot \rangle$ denotes the Frobenius inner product, $\mathbf{1} \in \mathbb{R}^{n\times n}$ is a matrix of ones, the first term is the discrete TV operator defined in (2.3), and the second term is the data fidelity term for speckle noise.

Since the data fidelity term of (2.9) is nonconvex, adapted models are suggested to obtain the convex optimization problems. Shi and Osher [54] utilized the logarithmic transformation $\boldsymbol{w} = \log \boldsymbol{u}$ and replaced the TV operator $\|\nabla \boldsymbol{u}\|_1$ in (2.9) with $\|\nabla \log \boldsymbol{u}\|_1 = \|\nabla \boldsymbol{w}\|_1$ to propose the following convex model

$$\arg\min_{\boldsymbol{w}} \|\nabla \boldsymbol{w}\|_1 + \lambda\big\langle \boldsymbol{w} + \boldsymbol{y}e^{-\boldsymbol{w}}, \mathbf{1}\big\rangle. \tag{2.10}$$

This model was further expanded to spatially adapted TV model [14], or the total generalized variation model [27].

Also Steidl and Teuber [56] proposed a convex model consisting of the I-divergence data fidelity term and the TV regularization term

$$\arg\min_{\boldsymbol{u}>0} \|\nabla \boldsymbol{u}\|_1 + \lambda\big\langle \boldsymbol{u} - \boldsymbol{y}\log \boldsymbol{u}, \mathbf{1}\big\rangle. \tag{2.11}$$

Although the I-divergence data fidelity term is known to be appropriate for the Poisson noise, which is derived by the MAP estimation, the authors showed that the two models (2.10) and (2.11) are equivalent.

Later, Dong and Zeng [24] introduced a quadratic penalty term into (2.9),

which turned out to be

$$\arg\min_{\boldsymbol{u}>0} \|\nabla \boldsymbol{u}\|_1 + \lambda\Big\langle \log \boldsymbol{u} + \frac{\boldsymbol{y}}{\boldsymbol{u}} + \alpha\left(\sqrt{\frac{\boldsymbol{y}}{\boldsymbol{u}}} - 1\right)^2, \mathbf{1}\Big\rangle, \tag{2.12}$$

where the model is proven to be strictly convex if the penalty parameter α satisfies $\alpha \geq 2\sqrt{6}/9$. This model was also expanded to total generalized variation model [53], or an exp-model [42].

Various approaches based on patch regularization are also proposed. Inspired by the dictionary learning [25] and the AA model [4], Huang et al. [34] proposed a variational model for multiplicative noise removal based on the combination of a TV regularization and a sparse representation in an adaptive dictionary of image patches. Parrilli et al. [48] adapted the block matching and three dimensional transform filtering [16] to SAR despeckling, and Cozzolino et al. [15] improved it to propose a fast nonlocal despeckling filter.

Recently, the nonlocal based approaches received much attention. Fang et al. [26] proposed a SAR image denoising method based on texture strength and weighted nuclear norm minimization. Liu et al. [38] adapted the nonlocally centralized sparse representation algorithm [23] to propose its SAR-oriented version. Guan et al. [32] suggested a SAR image despeckling method based on the nonlocal low-rank minimization model.

The GSR was also utilized for SAR image despeckling. Liu et al. [39] proposed an over-complete dictionary, which consists of the prespecified dictionaries and learned dictionary, to adapt GSR for SAR image despeckling. Liu et al. [40] included a mean filter in the modeling process of GSR based dictionary learning algorithm for SAR image despeckling. But these two models didn't consider the appropriate fidelity term for the speckle noise, which will be considered in our proposed algorithm.

2.3 GSR

In this section, the process of the group construction and the GSR modeling for image restoration is briefly introduced. For detailed explanations and discussions of GSR, we refer the reader to [66].

2.3.1 Group Construction

Suppose that the original image $\boldsymbol{x}$ of size $\sqrt{N} \times \sqrt{N}$ is divided into n overlapping patches $\boldsymbol{x}_k$ of size $\sqrt{P} \times \sqrt{P}$, $k = 1, 2, \ldots, n$, with stride s. The image and patches are represented as column vectors $\boldsymbol{x} \in \mathbb{R}^N$ and $\boldsymbol{x}_k \in \mathbb{R}^P$, respectively.

For each patch $\boldsymbol{x}_k$, the most similar c patches are searched in the training window of size $L \times L$. Subsequently, similar patches are stacked as columns to comprise a matrix $\boldsymbol{x}_{\boldsymbol{G}_k} = [\boldsymbol{x}_{\boldsymbol{G}_k,1}, \boldsymbol{x}_{\boldsymbol{G}_k,2}, \ldots, \boldsymbol{x}_{\boldsymbol{G}_k,c}] \in \mathbb{R}^{P \times c}$, which is called a group.

The extraction of a group from an image can be defined as a linear operator

$$\boldsymbol{x}_{\boldsymbol{G}_k} = \boldsymbol{R}_{\boldsymbol{G}_k}(\boldsymbol{x}), \tag{2.13}$$

and the transpose $\boldsymbol{R}_{\boldsymbol{G}_k}^T$ is an operator that returns the group to the original position in the image, with entries possibly overlapping and padded with zeros elsewhere.

The image $\boldsymbol{x}$ is reconstructed from the groups $\{\boldsymbol{x}_{\boldsymbol{G}_k}\}$ by averaging all the groups as

$$\boldsymbol{x} = \sum_{k=1}^{n} \boldsymbol{R}_{\boldsymbol{G}_k}^T(\boldsymbol{x}_{\boldsymbol{G}_k}) ./ \sum_{k=1}^{n} \boldsymbol{R}_{\boldsymbol{G}_k}^T(\mathbf{1}_{P \times c}), \tag{2.14}$$

where ./ is the entry-wise division of vectors; $\mathbf{1}_{P \times c} \in \mathbb{R}^{P \times c}$ is a matrix of ones.

2.3.2 GSR Modeling

The GSR model assumes that the groups can be sparsely represented by the atoms in the self-adaptive group dictionary $\boldsymbol{D}_{\boldsymbol{G}_k} = [\boldsymbol{d}_{\boldsymbol{G}_k,1}, \boldsymbol{d}_{\boldsymbol{G}_k,2}, \ldots, \boldsymbol{d}_{\boldsymbol{G}_k,m}]$, which is chosen during optimization. Explicitly, each group $\boldsymbol{x}_{\boldsymbol{G}_k}$ is approximated as a linear sum of the atoms $\boldsymbol{d}_{\boldsymbol{G}_k,i} \in \mathbb{R}^{P\times c}$, whose coefficients are given by the sparse codes $\boldsymbol{\alpha}_{\boldsymbol{G}_k} = [\boldsymbol{\alpha}_{\boldsymbol{G}_k,1}, \boldsymbol{\alpha}_{\boldsymbol{G}_k,2}, \ldots, \boldsymbol{\alpha}_{\boldsymbol{G}_k,m}] \in \mathbb{R}^m$:

$$\boldsymbol{x}_{\boldsymbol{G}_k} \approx \boldsymbol{D}_{\boldsymbol{G}_k}\boldsymbol{\alpha}_{\boldsymbol{G}_k} := \sum_{i=1}^{m} \boldsymbol{\alpha}_{\boldsymbol{G}_k,i}\boldsymbol{d}_{\boldsymbol{G}_k,i}. \tag{2.15}$$

The image $\boldsymbol{x}$ is reconstructed from the sparse codes $\{\boldsymbol{\alpha}_{\boldsymbol{G}_k}\}$ by averaging all approximations of the groups as

$$\boldsymbol{x} \approx \boldsymbol{D}_{\boldsymbol{G}} \circ \boldsymbol{\alpha}_{\boldsymbol{G}} := \sum_{k=1}^{n} \boldsymbol{R}_{\boldsymbol{G}_k}^T(\boldsymbol{D}_{\boldsymbol{G}_k}\boldsymbol{\alpha}_{\boldsymbol{G}_k})./\sum_{k=1}^{n} \boldsymbol{R}_{\boldsymbol{G}_k}^T(\mathbf{1}_{P\times c}), \tag{2.16}$$

where $\boldsymbol{D}_{\boldsymbol{G}}$ and $\boldsymbol{\alpha}_{\boldsymbol{G}}$ are concatenations of all $\boldsymbol{D}_{\boldsymbol{G}_k}$ and $\boldsymbol{\alpha}_{\boldsymbol{G}_k}$, respectively.

Using the ℓ_2 error as the data fidelity term and the sparsity of codes as the regularization term for the image restoration problem (2.1), the GSR model can be formulated as

$$\hat{\boldsymbol{\alpha}}_{\boldsymbol{G}} = \arg\min_{\boldsymbol{\alpha}_{\boldsymbol{G}}} \frac{1}{2}\|\boldsymbol{H}\boldsymbol{D}_{\boldsymbol{G}} \circ \boldsymbol{\alpha}_{\boldsymbol{G}} - \boldsymbol{y}\|_2^2 + \lambda\|\boldsymbol{\alpha}_{\boldsymbol{G}}\|_0, \tag{2.17}$$

where λ is the regularization parameter that balances the two terms. Subsequently, the original image can be reconstructed by $\hat{\boldsymbol{x}} = \boldsymbol{D}_{\boldsymbol{G}} \circ \hat{\boldsymbol{\alpha}}_{\boldsymbol{G}}$.

2.4 ADMM

In this section, the ADMM algorithm is presented. The ADMM algorithm [7] solves the following minimization problem:

$$\text{minimize } f(\boldsymbol{u}) + g(\boldsymbol{v}), \text{ subject to } \boldsymbol{A}\boldsymbol{u} + \boldsymbol{B}\boldsymbol{v} = \boldsymbol{c} \tag{2.18}$$

with variables $\boldsymbol{u} \in \mathbb{R}^N$ and $\boldsymbol{v} \in \mathbb{R}^M$, where $\boldsymbol{A} \in \mathbb{R}^{P\times N}$, $\boldsymbol{B} \in \mathbb{R}^{P\times M}$, and $\boldsymbol{c} \in \mathbb{R}^P$. The functions f, g are typically assumed as convex, but nonconvex functions can be considered as well [58]. The augmented Lagrangian for (2.18) is given by

$$L_\mu(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) = f(\boldsymbol{u}) + g(\boldsymbol{v}) + \boldsymbol{w}^T(\boldsymbol{A}\boldsymbol{u} + \boldsymbol{B}\boldsymbol{v} - \boldsymbol{c}) + \frac{\mu}{2}\|\boldsymbol{A}\boldsymbol{u} + \boldsymbol{B}\boldsymbol{v} - \boldsymbol{c}\|_2^2, \tag{2.19}$$

where $\boldsymbol{w}$ is the dual variable and $\mu > 0$ is a penalty parameter. For convenience, the scaled form of (2.19) can be written as

$$L_\mu(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{b}) = f(\boldsymbol{u}) + g(\boldsymbol{v}) + \frac{\mu}{2}\|\boldsymbol{A}\boldsymbol{u} + \boldsymbol{B}\boldsymbol{v} - \boldsymbol{c} + \boldsymbol{b}\|_2^2 - \frac{\mu}{2}\|\boldsymbol{b}\|_2^2, \tag{2.20}$$

where $\boldsymbol{b} = \boldsymbol{w}/\mu$ is the scaled dual variable. The ADMM for solving (2.18) can be expressed as follows:

Algorithm 1 ADMM for solving the minimization problem (2.18)

Input: $\mu > 0$

Initialization: $t = 0,\ \boldsymbol{u}^{(0)} = \mathbf{0},\ \boldsymbol{v}^{(0)} = \mathbf{0},\ \boldsymbol{b}^{(0)} = \mathbf{0}$.

repeat

$\quad \boldsymbol{u}^{(t+1)} = \arg\min_{\boldsymbol{u}} \left(f(\boldsymbol{u}) + \frac{\mu}{2} \|\boldsymbol{A}\boldsymbol{u} + \boldsymbol{B}\boldsymbol{v}^{(t)} - \boldsymbol{c} + \boldsymbol{b}^{(t)}\|_2^2 \right)$

$\quad \boldsymbol{v}^{(t+1)} = \arg\min_{\boldsymbol{v}} \left(g(\boldsymbol{v}) + \frac{\mu}{2} \|\boldsymbol{A}\boldsymbol{u}^{(t+1)} + \boldsymbol{B}\boldsymbol{v} - \boldsymbol{c} + \boldsymbol{b}^{(t)}\|_2^2 \right)$

$\quad \boldsymbol{b}^{(t+1)} = \boldsymbol{b}^{(t)} + \boldsymbol{A}\boldsymbol{u}^{(t+1)} + \boldsymbol{B}\boldsymbol{v}^{(t+1)} - \boldsymbol{c}$

$\quad t \leftarrow t + 1$

until stopping criterion is satisfied

Output: $\boldsymbol{u}$, $\boldsymbol{v}$

Chapter 3

Proposed Models

In this chapter, we propose two models for non-Gaussian noises based on GSR, which are the GSR model for Cauchy noise (GSRC) and the GSR model for speckle noise (GSRS). First, we will derive the variational model based on MAP estimation for each noise type. The appropriate patch distances are introduced, and both models are solved by the ADMM algorithm which was introduced in Section 2.4. In addition, the numerical experiments and related discussions are provided for both models.

3.1 Proposed Model 1: GSRC

In this section, we apply the GSR approach to restore images corrupted by Cauchy noise. Since the Cauchy noise is the additive noise with Cauchy distribution, the original GSR model, which is designed for Gaussian noise, should be adapted in some ways. We consider the adaptations, which include the modeling of GSRC via MAP estimation and the patch distance for Cauchy noise. The initialization technique and the analyses of the parameters and convergence are also discussed.

3.1.1 GSRC Modeling via MAP Estimator

The original GSR model (2.17) uses the ℓ_2 norm for the data fidelity term, which is known to be appropriate for Gaussian noise but not for other types of noise. In the TV model for Cauchy noise [43, 52], the data fidelity term appropriate for Cauchy noise is derived using MAP estimate. We follow these to derive the modeling of GSRC analogously.

Recall that we want to restore the original image $\boldsymbol{x}$ from the noisy observation $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}$, where $\boldsymbol{n}$ follows the Cauchy distribution. We consider $\boldsymbol{y}(i)$, $\boldsymbol{x}(i)$ as random variables for each $i = 1, 2, \ldots, N$. We want to find the MAP estimator $\hat{\boldsymbol{x}}$, which maximizes the posterior probability $P(\boldsymbol{x} \mid \boldsymbol{y})$. Using Bayes' theorem,

$$\begin{aligned}\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} P(\boldsymbol{x} \mid \boldsymbol{y}) &= \arg\max_{\boldsymbol{x}} \frac{P(\boldsymbol{y} \mid \boldsymbol{x})P(\boldsymbol{x})}{P(\boldsymbol{y})} \\ &= \arg\max_{\boldsymbol{x}} P(\boldsymbol{y} \mid \boldsymbol{x})P(\boldsymbol{x}). \end{aligned} \tag{3.1}$$

we can equivalently find the minimizer of the negative logarithm of (3.1) as the following:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} -\log P(\boldsymbol{y} \mid \boldsymbol{x}) - \log P(\boldsymbol{x}). \tag{3.2}$$

Since we are assuming that the values of the image are independent and identically distributed, we have $P(\boldsymbol{y} \mid \boldsymbol{x}) = \prod_{i=1}^{N} P(\boldsymbol{y}(i) \mid \boldsymbol{x})$. From the probability density function (2.2) of Cauchy noise, we know that

$$P(\boldsymbol{y}(i) \mid \boldsymbol{x}) = \frac{1}{\pi} \frac{\gamma}{\gamma^2 + (\boldsymbol{H}\boldsymbol{x}(i) - \boldsymbol{y}(i))^2}. \tag{3.3}$$

For the prior probability $P(\boldsymbol{x})$, we use the regularizer as

$$P(\boldsymbol{x}) = \exp(-2\lambda\Psi(\boldsymbol{x})), \tag{3.4}$$

where $\lambda > 0$ is a regularization parameter. Substituting these into (3.2) gives

$$\begin{aligned}\hat{\boldsymbol{x}} &= \arg\min_{\boldsymbol{x}} -\log \prod_{i=1}^{N} P(\boldsymbol{y}(i) \mid \boldsymbol{x}) + 2\lambda\Psi(\boldsymbol{x}) \\ &= \arg\min_{\boldsymbol{x}} -\sum_{i=1}^{N} \log\left(\frac{1}{\pi}\frac{\gamma}{\gamma^2 + (\boldsymbol{H}\boldsymbol{x}(i) - \boldsymbol{y}(i))^2}\right) + 2\lambda\Psi(\boldsymbol{x}) \\ &= \arg\min_{\boldsymbol{x}} \sum_{i=1}^{N} \log\left(\gamma^2 + (\boldsymbol{H}\boldsymbol{x}(i) - \boldsymbol{y}(i))^2\right) + 2\lambda\Psi(\boldsymbol{x}), \end{aligned} \tag{3.5}$$

where the irrelevant constants are not considered. The variational model for removing Cauchy noise can be written as

$$\arg\min_{\boldsymbol{x}} \frac{1}{2}\left\langle \log\left(\gamma^2 + (\boldsymbol{H}\boldsymbol{x} - \boldsymbol{y})^2\right), \mathbf{1}\right\rangle + \lambda\Psi(\boldsymbol{x}), \tag{3.6}$$

where $\langle \cdot, \cdot \rangle$ denotes the standard inner product and $\mathbf{1} \in \mathbb{R}^N$ is a vector of ones.

Now, in the framework of GSR, we are assuming that $\boldsymbol{x} \approx \boldsymbol{D_G} \circ \boldsymbol{\alpha_G}$ (see (2.16)) and the regularization term as $\Psi(\boldsymbol{\alpha_G}) = \|\boldsymbol{\alpha_G}\|_0$. Hence the GSRC model can be formulated as

$$\arg\min_{\boldsymbol{\alpha_G}} \frac{1}{2}\left\langle \log\left(\gamma^2 + (\boldsymbol{H}\boldsymbol{D_G} \circ \boldsymbol{\alpha_G} - \boldsymbol{y})^2\right), \mathbf{1}\right\rangle + \lambda\|\boldsymbol{\alpha_G}\|_0. \tag{3.7}$$

Although the general ℓ_0 minimization is an NP-hard problem, by selecting an appropriate self-adaptive group dictionary, the minimization problem (3.7) can be solved explicitly, which will be explained in Section 3.1.3.

3.1.2 Patch Distance for Cauchy Noise

In the process of group construction (Section 2.3.1), when searching for similar patches, the Euclidean distance can be affected considerably by the impulsive property of Cauchy noise; thus, similar patches cannot be found effectively. Hence, the appropriate patch distance for Cauchy noise should be used. The patch similarity under non-Gaussian noise has been presented in [19, 20]. Given two patches $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, the generalized likelihood ratio for Cauchy noise is given by

$$\mathcal{L}_G(\boldsymbol{x}_1, \boldsymbol{x}_2) = \prod_{i=1}^{P} \left[1 + \left(\frac{\boldsymbol{x}_1(i) - \boldsymbol{x}_2(i)}{2\gamma} \right)^2 \right]^{-2}, \tag{3.8}$$

and the patch distance for Cauchy noise can be defined as the negative logarithm of (3.8),

$$d(\boldsymbol{x}_1, \boldsymbol{x}_2) = 2 \sum_{i=1}^{P} \log \left[1 + \left(\frac{\boldsymbol{x}_1(i) - \boldsymbol{x}_2(i)}{2\gamma} \right)^2 \right]. \tag{3.9}$$

Therefore, when we collect similar patches to construct the group $\boldsymbol{x}_{\boldsymbol{G}_k}$, (3.9) is used as the distance between patches.

3.1.3 The ADMM Algorithm for Solving (3.7)

The optimization problem (3.7) is difficult to solve directly, since it is a nonconvex problem with complicated structure. For an efficient minimization, we adopt the ADMM algorithm in Section 2.4. Recently, the nonconvex ADMM was proven to generate a sequence of iterates that has a convergent subsequence to a stationary point of the augmented Lagrangian of (3.7) [58].

By introducing an auxiliary variable $\boldsymbol{u}$, the equivalent constrained formulation of (3.7) can be written as

$$\arg\min_{\boldsymbol{\alpha_G},\boldsymbol{u}} \frac{1}{2}\left\langle \log\left(\gamma^2 + (\boldsymbol{H}\boldsymbol{u}-\boldsymbol{y})^2\right), \boldsymbol{1}\right\rangle + \lambda\|\boldsymbol{\alpha_G}\|_0, \quad \text{s.t. } \boldsymbol{u} = \boldsymbol{D_G}\circ\boldsymbol{\alpha_G}. \tag{3.10}$$

Setting $f(\boldsymbol{\alpha_G}) = \lambda\|\boldsymbol{\alpha_G}\|_0$ and $g(\boldsymbol{u}) = \frac{1}{2}\left\langle \log\left(\gamma^2 + (\boldsymbol{H}\boldsymbol{u}-\boldsymbol{y})^2\right), \boldsymbol{1}\right\rangle$, the minimization of (3.10) is obtained from Algorithm 1, which involves iteratively solving the two subproblems of $\boldsymbol{\alpha_G}$ and $\boldsymbol{u}$, with the update of the dual variable $\boldsymbol{b}$ as follows:

$$\boldsymbol{\alpha}_{\boldsymbol{G}}^{(t+1)} = \arg\min_{\boldsymbol{\alpha_G}} \lambda\|\boldsymbol{\alpha_G}\|_0 + \frac{\mu}{2}\|\boldsymbol{u}^{(t)} - \boldsymbol{D_G}\circ\boldsymbol{\alpha_G} + \boldsymbol{b}^{(t)}\|_2^2 \tag{3.11}$$

$$\boldsymbol{u}^{(t+1)} = \arg\min_{\boldsymbol{u}} \frac{1}{2}\left\langle \log\left(\gamma^2 + (\boldsymbol{H}\boldsymbol{u}-\boldsymbol{y})^2\right), \boldsymbol{1}\right\rangle + \frac{\mu}{2}\|\boldsymbol{u} - \boldsymbol{D_G}\circ\boldsymbol{\alpha}_{\boldsymbol{G}}^{(t+1)} + \boldsymbol{b}^{(t)}\|_2^2 \tag{3.12}$$

$$\boldsymbol{b}^{(t+1)} = \boldsymbol{b}^{(t)} + \boldsymbol{u}^{(t+1)} - \boldsymbol{D_G}\circ\boldsymbol{\alpha}_{\boldsymbol{G}}^{(t+1)} \tag{3.13}$$

In the following, we provide the details to solve the two subproblems above: (3.11) and (3.12).

$\boldsymbol{\alpha_G}$-subproblem

Because the $\boldsymbol{\alpha_G}$-subproblem (3.11) is not changed from the original GSR model, we refer to [66] to explain the solution and construction of the self-adaptive group dictionary. The $\boldsymbol{\alpha_G}$-subproblem can be written as

$$\arg\min_{\boldsymbol{\alpha_G}} \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{r}\|_2^2 + \frac{\lambda}{\mu}\|\boldsymbol{\alpha_G}\|_0, \tag{3.14}$$

where $\boldsymbol{x} = \boldsymbol{D_G} \circ \boldsymbol{\alpha_G}$, and $\boldsymbol{r} = \boldsymbol{u}^{(t)} + \boldsymbol{b}^{(t)}$ is a noisy observation of $\boldsymbol{x}$. In the following proposition, it is proven that the error term of the image can be well approximated by the error term of the groups with a reasonable assumption.

Proposition 3.1. *Let $\boldsymbol{x}, \boldsymbol{r} \in \mathbb{R}^N$, $\boldsymbol{x_{G_k}}, \boldsymbol{r_{G_k}} \in \mathbb{R}^{P \times c}$ and assume that each entry of $\boldsymbol{x} - \boldsymbol{r}$ is independent and identically distributed with mean 0 and variance σ^2. Then, for any $\epsilon > 0$,*

$$\lim_{\substack{N \to \infty \\ K \to \infty}} \Pr \left\{ \left| \frac{1}{N} \|\boldsymbol{x} - \boldsymbol{r}\|_2^2 - \frac{1}{K} \sum_{k=1}^{n} \|\boldsymbol{x_{G_k}} - \boldsymbol{r_{G_k}}\|_F^2 \right| < \epsilon \right\} = 1, \tag{3.15}$$

where $K = P \times c \times n$.

Proof. It follows from the law of large numbers, see [66]. □

Substituting the approximation of (3.15) into (3.14) and using $\|\boldsymbol{\alpha_G}\|_0 = \sum_{k=1}^{n} \|\boldsymbol{\alpha_{G_k}}\|_0$, which is obvious because $\boldsymbol{\alpha_G}$ is the concatenation of all $\boldsymbol{\alpha_{G_k}}$'s, (3.14) is equivalent to

$$\arg\min_{\boldsymbol{\alpha_G}} \sum_{k=1}^{n} \left(\frac{1}{2} \|\boldsymbol{x_{G_k}} - \boldsymbol{r_{G_k}}\|_F^2 + \tau \|\boldsymbol{\alpha_{G_k}}\|_0 \right), \tag{3.16}$$

where $\tau = \frac{\lambda K}{\mu N}$. The minimization of (3.16) can be achieved by solving n subproblems of $\boldsymbol{\alpha_{G_k}}$; however, it is a ℓ_0 minimization problem, which is NP-hard in general. Nevertheless, by selecting a self-adaptive dictionary in the following manner, (3.16) can be solved simply.

Let the singular value decomposition of $\boldsymbol{r_{G_k}} \in \mathbb{R}^{P \times c}$ be given by

$$\boldsymbol{r_{G_k}} = \boldsymbol{U_{G_k}} \boldsymbol{\Sigma_{G_k}} \boldsymbol{V_{G_k}}^T = \sum_{i=1}^{m} \boldsymbol{\gamma}_{\boldsymbol{r_{G_k}},i} (\boldsymbol{u_{G_k},}_i \boldsymbol{v_{G_k},}_i{}^T), \tag{3.17}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{G}_k} = \text{diag}(\boldsymbol{\gamma}_{\boldsymbol{r}_{\boldsymbol{G}_k}})$, $\boldsymbol{\gamma}_{\boldsymbol{r}_{\boldsymbol{G}_k}} = [\boldsymbol{\gamma}_{\boldsymbol{r}_{\boldsymbol{G}_k},1}, \boldsymbol{\gamma}_{\boldsymbol{r}_{\boldsymbol{G}_k},2}, \ldots, \boldsymbol{\gamma}_{\boldsymbol{r}_{\boldsymbol{G}_k},m}] \in \mathbb{R}^m$ with $m = \min(P, c)$ is the singular value vector of $\boldsymbol{r}_{\boldsymbol{G}_k}$ and $\boldsymbol{u}_{\boldsymbol{G}_k,i}, \boldsymbol{v}_{\boldsymbol{G}_k,i}$ are the columns of $\boldsymbol{U}_{\boldsymbol{G}_k}, \boldsymbol{V}_{\boldsymbol{G}_k}$, respectively. The atoms for group $\boldsymbol{x}_{\boldsymbol{G}_k}$ are defined by

$$\boldsymbol{d}_{\boldsymbol{G}_k,i} = \boldsymbol{u}_{\boldsymbol{G}_k,i}\boldsymbol{v}_{\boldsymbol{G}_k,i}{}^T, \quad i = 1, 2, \ldots, m, \tag{3.18}$$

and the self-adaptive dictionary for $\boldsymbol{x}_{\boldsymbol{G}_k}$ is defined by

$$\boldsymbol{D}_{\boldsymbol{G}_k} = [\boldsymbol{d}_{\boldsymbol{G}_k,1}, \boldsymbol{d}_{\boldsymbol{G}_k,2}, \ldots, \boldsymbol{d}_{\boldsymbol{G}_k,m}]. \tag{3.19}$$

The $\boldsymbol{\alpha}_{\boldsymbol{G}_k}$-subproblem is to obtain $\boldsymbol{x}_{\boldsymbol{G}_k} = \boldsymbol{D}_{\boldsymbol{G}_k}\boldsymbol{\alpha}_{\boldsymbol{G}_k}$, given that $\boldsymbol{r}_{\boldsymbol{G}_k} = \boldsymbol{D}_{\boldsymbol{G}_k}\boldsymbol{\gamma}_{\boldsymbol{r}_{\boldsymbol{G}_k}}$. By the unitary property of $\boldsymbol{U}_{\boldsymbol{G}_k}$ and $\boldsymbol{V}_{\boldsymbol{G}_k}$, the n subproblems for the $\boldsymbol{\alpha}_{\boldsymbol{G}_k}$ of (3.16) are equivalent to

$$\arg\min_{\boldsymbol{\alpha}_{\boldsymbol{G}_k}} \frac{1}{2}\|\boldsymbol{\alpha}_{\boldsymbol{G}_k} - \boldsymbol{\gamma}_{\boldsymbol{r}_{\boldsymbol{G}_k}}\|_F^2 + \tau\|\boldsymbol{\alpha}_{\boldsymbol{G}_k}\|_0. \tag{3.20}$$

(3.20) can be minimized with the entry-wise hard thresholding operator [3],

$$\hat{\boldsymbol{\alpha}}_{\boldsymbol{G}_k} = hard(\boldsymbol{\gamma}_{\boldsymbol{r}_{\boldsymbol{G}_k}}, \sqrt{2\tau}) = \boldsymbol{\gamma}_{\boldsymbol{r}_{\boldsymbol{G}_k}} * \mathbf{1}\left(|\boldsymbol{\gamma}_{\boldsymbol{r}_{\boldsymbol{G}_k}}| - \sqrt{2\tau}\right), \tag{3.21}$$

where $*$ is the entry-wise product of vectors and $\mathbf{1}(\cdot)$ is the indicator function $\mathbf{1}(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \boldsymbol{x}_i > 0 \\ 0 & \text{if } \boldsymbol{x}_i \leq 0 \end{cases}$.

$\boldsymbol{u}$-subproblem

To minimize (3.12), we first derive the optimality condition as follows.

Proposition 3.2. *The optimality condition of* (3.12) *is*

$$\boldsymbol{H}^T \frac{\boldsymbol{H}\boldsymbol{u} - \boldsymbol{y}}{\gamma^2 + (\boldsymbol{H}\boldsymbol{u} - \boldsymbol{y})^2} + \mu(\boldsymbol{u} - \boldsymbol{D}_{\boldsymbol{G}} \circ \boldsymbol{\alpha}_{\boldsymbol{G}}^{(t+1)} + \boldsymbol{b}^{(t)}) = 0, \tag{3.22}$$

where the division implies the entry-wise division.

Proof. Let $\boldsymbol{u} = (u_1, u_2, \ldots, u_N) \in \mathbb{R}^N$, $\boldsymbol{y} = (y_1, y_2, \ldots, y_N) \in \mathbb{R}^N$, $\boldsymbol{D}_{\boldsymbol{G}} \circ \boldsymbol{\alpha}_{\boldsymbol{G}}^{(t+1)} - \boldsymbol{b}^{(t)} = (v_1, v_2, \ldots, v_N) \in \mathbb{R}^N$ and $\boldsymbol{H} = (h_{ij}) \in \mathbb{R}^{N \times N}$. Then, the objective function in (3.12) can be written as

$$F(u_1, u_2, \ldots, u_N) := \frac{1}{2} \sum_{i=1}^{N} \log \left(\gamma^2 + \left(\sum_{j=1}^{N} h_{ij} u_j - y_i \right)^2 \right) + \frac{\mu}{2} \sum_{i=1}^{N} (u_i - v_i)^2. \tag{3.23}$$

By taking the partial derivative of F with respect to u_k,

$$\begin{aligned} \frac{\partial}{\partial u_k} F(u_1, u_2, \ldots, u_N) &= \sum_{i=1}^{N} \frac{h_{ik} \left(\sum_{j=1}^{N} h_{ij} u_j - y_i \right)}{\gamma^2 + \left(\sum_{j=1}^{N} h_{ij} u_j - y_i \right)^2} + \mu(u_k - v_k) \\ &= \sum_{i=1}^{N} h_{ik} \left[\frac{\boldsymbol{H}\boldsymbol{u} - \boldsymbol{y}}{\gamma^2 + (\boldsymbol{H}\boldsymbol{u} - \boldsymbol{y})^2} \right]_i + \mu(u_k - v_k) \end{aligned} \tag{3.24}$$

Therefore, the gradient of F is represented as (3.22) in the vector form. □

The solution of (3.22) is dependent on the degradation operator $\boldsymbol{H}$. In denoising, $\boldsymbol{H}$ is the identity operator and (3.22) becomes a cubic equation by multiplying with the denominator, which can be written as

$$a(\boldsymbol{u} - \boldsymbol{y})^3 + b(\boldsymbol{u} - \boldsymbol{y})^2 + c(\boldsymbol{u} - \boldsymbol{y}) + d = 0, \tag{3.25}$$

where $a = \mu$, $b = \mu(\boldsymbol{y} - \boldsymbol{D_G} \circ \boldsymbol{\alpha}_{\boldsymbol{G}}^{(t+1)} + \boldsymbol{b}^{(t)})$, $c = 1 + \mu\gamma^2$, $d = \mu\gamma^2(\boldsymbol{y} - \boldsymbol{D_G} \circ \boldsymbol{\alpha}_{\boldsymbol{G}}^{(t+1)} + \boldsymbol{b}^{(t)})$. The cubic equation (3.25) can be solved explicitly using the cubic root formula, which can be expressed in terms of real functions as the following [68]:

Proposition 3.3. *For the cubic equation*

$$ax^3 + bx^2 + cx + d = 0, \text{ where } a \neq 0, \tag{3.26}$$

define $p = \frac{3ac-b^2}{9a^2}$, $q = \frac{2b^3-9abc+27a^2d}{27a^3}$ *and the discriminant* $\Delta = 4p^3 + q^2$. *If* $\Delta > 0$, *the cubic equation* (3.26) *has only one real root, which is given by*

$$x = \sqrt[3]{\frac{-q + \sqrt{q^2 + 4p^3}}{2}} + \sqrt[3]{\frac{-q - \sqrt{q^2 + 4p^3}}{2}} - \frac{b}{3a}. \tag{3.27}$$

Otherwise, if $\Delta \leq 0$, *then* (3.26) *has three real roots (possibly equal), which are given by*

$$x = 2\sqrt{-p}\cos\frac{\cos^{-1}\frac{-q}{2\sqrt{-p^3}} + 2k\pi}{3} - \frac{b}{3a}, \quad k = 0, 1, 2. \tag{3.28}$$

If the equation (3.25) has three real roots, the correct solution is selected to yield the minimal value of the objective function (3.12).

In deblurring and denoising, $\boldsymbol{H}$ is not the identity operator and the optimality condition (3.22) cannot be solved explicitly. Here, the gradient descent algorithm is utilized as

$$\boldsymbol{u} \leftarrow \boldsymbol{u} - \eta\left[\boldsymbol{H}^T\frac{\boldsymbol{Hu} - \boldsymbol{y}}{\gamma^2 + (\boldsymbol{Hu} - \boldsymbol{y})^2} + \mu(\boldsymbol{u} - \boldsymbol{D_G} \circ \boldsymbol{\alpha}_{\boldsymbol{G}}^{(t+1)} + \boldsymbol{b}^{(t)})\right]. \tag{3.29}$$

By assuming the periodic boundary condition, the blurring operator $\boldsymbol{H}$ and its transpose $\boldsymbol{H}^T$ are block circulant with circulant blocks, which can be efficiently computed using the two-dimensional (2-D) fast Fourier transform as follows [17, 45]:

$$\boldsymbol{u} \leftarrow \boldsymbol{u} - \eta \Bigg[\mathcal{F}^{-1} \left(\overline{\mathcal{F}(\boldsymbol{h})} * \mathcal{F} \left(\frac{\mathcal{F}^{-1}(\mathcal{F}(\boldsymbol{h}) * \mathcal{F}(\boldsymbol{u})) - \boldsymbol{y}}{\gamma^2 + (\mathcal{F}^{-1}(\mathcal{F}(\boldsymbol{h}) * \mathcal{F}(\boldsymbol{u})) - \boldsymbol{y})^2} \right) \right) + \mu(\boldsymbol{u} - \boldsymbol{D}_{\boldsymbol{G}} \circ \boldsymbol{\alpha}_{\boldsymbol{G}}^{(t+1)} + \boldsymbol{b}^{(t)}) \Bigg], \tag{3.30}$$

where the variables are viewed as matrices, $*$ is the entry-wise product of the matrices, $\mathcal{F}$ is the 2-D discrete Fourier transform, and $\boldsymbol{h}$ is the padded blur kernel of $\boldsymbol{H}$.

In Algorithm 2, the full description of the proposed algorithm is provided.

3.1.4 Numerical Experiments

Various numerical experiments were designed to investigate the performance of the proposed algorithm, including image denoising and deblurring. The test images are presented in Fig. 3.1: 12 grayscale images of size 256×256, whose values are in the range of $[0, 255]$. All experiments were performed using MATLAB R2017b and Windows 10 on a PC with Intel(R) Core(TM) i7-4790 CPU @ 3.60 GHz, 16.0 GB RAM.

In the following experiments, Cauchy noise with probability density function (2.2), scale parameter γ, and localization parameter $\delta = 0$ is generated using [37]

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n} = \boldsymbol{H}\boldsymbol{x} + \gamma \frac{\boldsymbol{\eta}_1}{\boldsymbol{\eta}_2}, \tag{3.31}$$

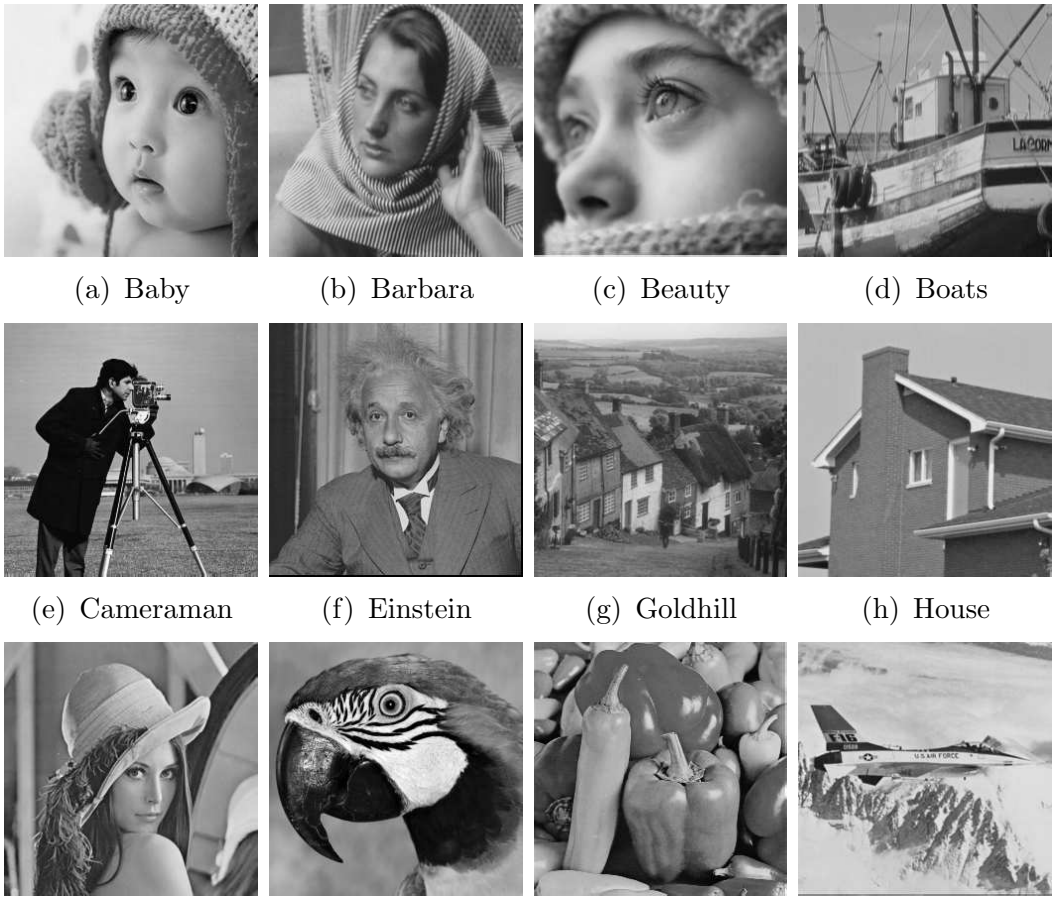
where $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ are two independent Gaussian random variables with mean 0 and variance 1. We assume that $\boldsymbol{y}$ is restricted to $[0, 255]$ after the noise

Algorithm 2 GSR algorithm for Cauchy noise (GSRC)

Input: the noisy image $\boldsymbol{y}$, the degradation operator $\boldsymbol{H}$, parameters P, s, L, c, T, λ, μ.

Initialization: $t = 0$, $\boldsymbol{u}^{(0)} = \boldsymbol{y}$, $\boldsymbol{b}^{(0)} = \boldsymbol{0}$.
repeat
  Set $\boldsymbol{r} = \boldsymbol{u}^{(t)} + \boldsymbol{b}^{(t)}$, $\tau = \lambda K/\mu N$.
  Construct the group $\boldsymbol{r}_{\boldsymbol{G}_k}$ by obtaining similar patches using the distance (3.9).
  for each group $\boldsymbol{r}_{\boldsymbol{G}_k}$ **do**
    Construct the dictionary $\boldsymbol{D}_{\boldsymbol{G}_k}$ from (3.17) and (3.18).
    Compute $\hat{\boldsymbol{\alpha}}_{\boldsymbol{G}_k}$ by (3.21).
  end for
  Update $\boldsymbol{D}_{\boldsymbol{G}}^{(t+1)}$ by concatenating all $\boldsymbol{D}_{\boldsymbol{G}_k}$.
  Update $\boldsymbol{\alpha}_{\boldsymbol{G}}^{(t+1)}$ by concatenating all $\hat{\boldsymbol{\alpha}}_{\boldsymbol{G}_k}$.
  Compute $\boldsymbol{D}_{\boldsymbol{G}}^{(t+1)} \circ \boldsymbol{\alpha}_{\boldsymbol{G}}^{(t+1)}$ by (2.16).
  if $\boldsymbol{H}$ is the identity operator **then**
    Update $\boldsymbol{u}^{(t+1)}$ by solving (3.25), using (3.27) or (3.28).
  else if $\boldsymbol{H}$ is a blurring operator **then**
    Update $\boldsymbol{u}^{(t+1)}$ by (3.30).
  end if
  Update $\boldsymbol{b}^{(t+1)}$ by (3.13).
  $t \leftarrow t + 1$.
until $t < T$

Output: the restored image $\boldsymbol{u}^{(T)}$.



(a) Baby (b) Barbara (c) Beauty (d) Boats

(e) Cameraman (f) Einstein (g) Goldhill (h) House

(i) Lena (j) Parrot (k) Peppers (l) Plane

Figure 3.1: Original test images.

is generated; hence, $\boldsymbol{y} = \max(0, \min(\boldsymbol{y}, 255))$. For comparison, the degraded image $\boldsymbol{y}$ is restored by three different methods: our algorithm (GSRC), the median filter (MED) [30], and the total variation with overlapping group sparsity model (OGSTV). OGSTV [21] is a recently developed method that improves TV models using the sparsity and group sparsity of the gradient.

The quality of the recovered images from different algorithms are examined by two widely used criteria: peak signal-to-noise ratio (PSNR) and

structural similarity index (SSIM) [59], which are defined as

$$\mathrm{PSNR}(\boldsymbol{u}, \tilde{\boldsymbol{u}}) = 10 \log_{10} \frac{255^2}{\|\boldsymbol{u} - \tilde{\boldsymbol{u}}\|_F^2 / N},$$
$$\mathrm{SSIM}(\boldsymbol{u}, \tilde{\boldsymbol{u}}) = \frac{(2\mu_{\boldsymbol{u}}\mu_{\tilde{\boldsymbol{u}}} + c_1)(2\sigma_{\boldsymbol{u}\tilde{\boldsymbol{u}}} + c_2)}{(\mu_{\boldsymbol{u}}^2 + \mu_{\tilde{\boldsymbol{u}}}^2 + c_1)(\sigma_{\boldsymbol{u}}^2 + \sigma_{\tilde{\boldsymbol{u}}}^2 + c_2)},$$

where $\boldsymbol{u}$ is the original image and $\tilde{\boldsymbol{u}}$ is the recovered image; N is the image size; $\mu_{\boldsymbol{u}}$ and $\mu_{\tilde{\boldsymbol{u}}}$ are their respective averages, $\sigma_{\boldsymbol{u}}$ and $\sigma_{\tilde{\boldsymbol{u}}}$ are their respective standard deviations; $\sigma_{\boldsymbol{u}\tilde{\boldsymbol{u}}}$ is the covariance of $\boldsymbol{u}$ and $\tilde{\boldsymbol{u}}$; c_1, c_2 are constants.

The parameters are as follows: The image size N is 256×256, and the patch size P is set to 6×6. The stride between overlapping patches s is set to 2; thus, the number of overlapping patches n is 128×128. The training window size $L \times L$ is set as 20×20, and the number of similar patches in a group c is selected as 60. The maximum iteration number T of the algorithm is set to 10 for denoising and 30 for deblurring and denoising combined. The regularization parameter λ and the penalty parameter μ are adjusted to yield the best results in different simulations, which are presented in the following sections. For a fair comparison, the parameters of all algorithms are fixed for different image types. A detailed discussion of the parameter selection is provided in Section 3.1.5.

Image Denoising

The pure denoising problem is simulated, where $\boldsymbol{H}$ is the identity operator. The test images are corrupted by Cauchy noise with two different noise levels, $\gamma = 5$ and $\gamma = 10$. The parameters λ and μ are set as follows: For $\gamma = 5$, $\lambda = 0.13, \mu = 0.008$ are used; for $\gamma = 10$, $\lambda = 0.09, \mu = 0.004$ are used. However, to produce fine results, different λ values are used at the first iteration: $\lambda^{(0)} = 2$ for both noise levels. This initialization technique is discussed in detail in Section 3.1.5.

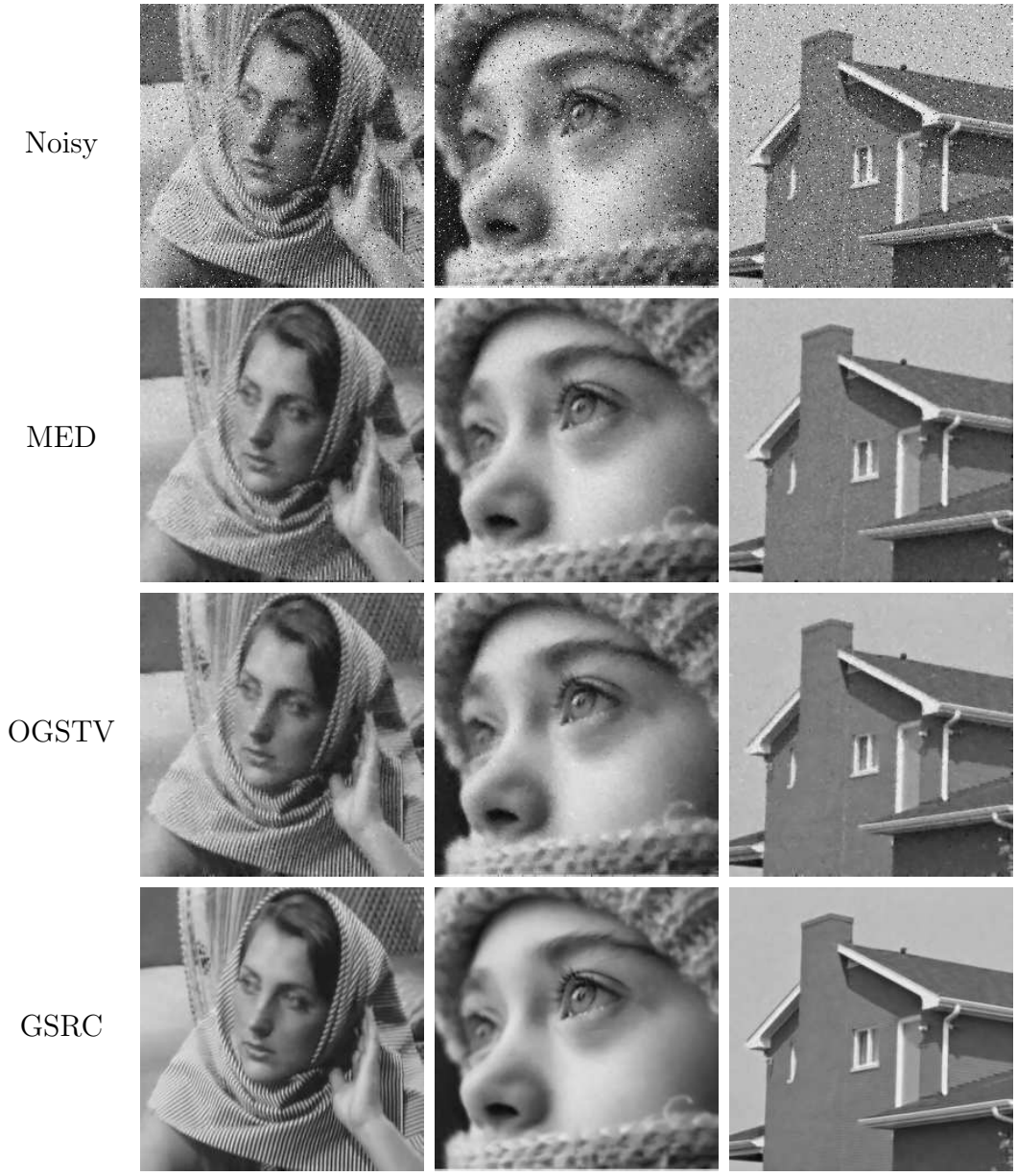


Figure 3.2: Comparison of different methods for restoring images degraded by Cauchy noise ($\gamma = 5$).

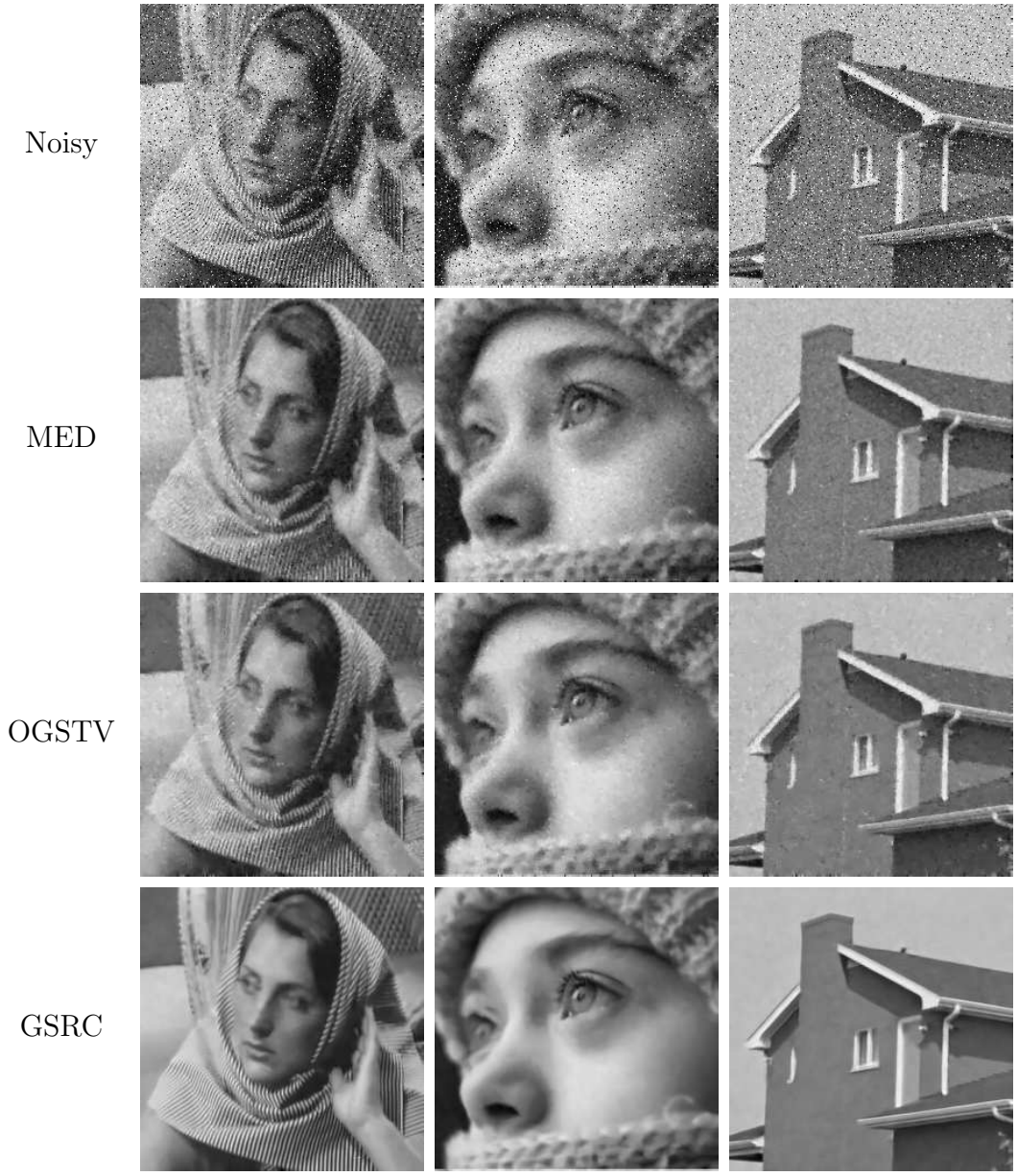


Figure 3.3: Comparison of different methods for restoring images degraded by Cauchy noise ($\gamma = 10$).

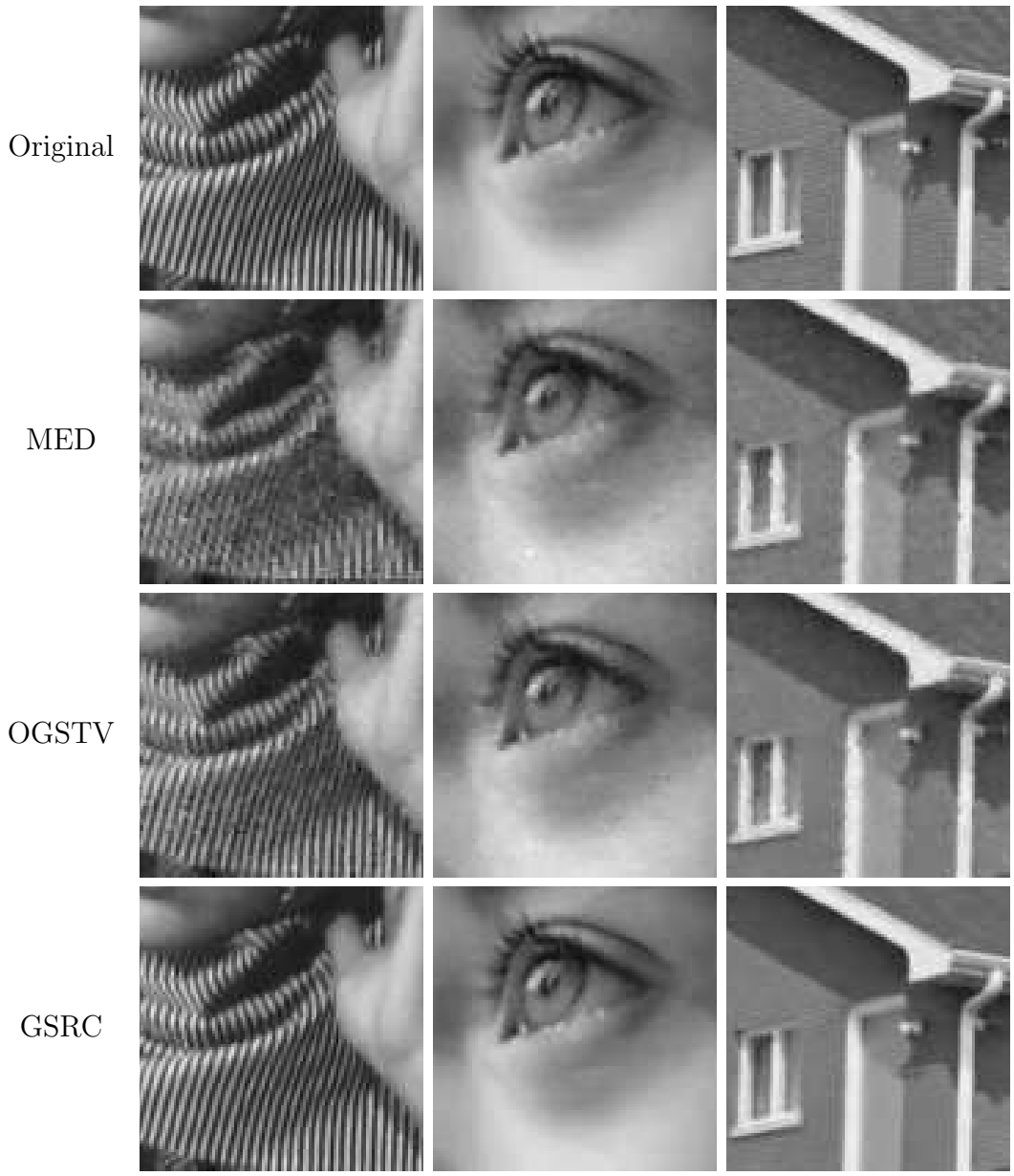


Figure 3.4: The zoomed-in version of the restored images in Fig. 3.2.

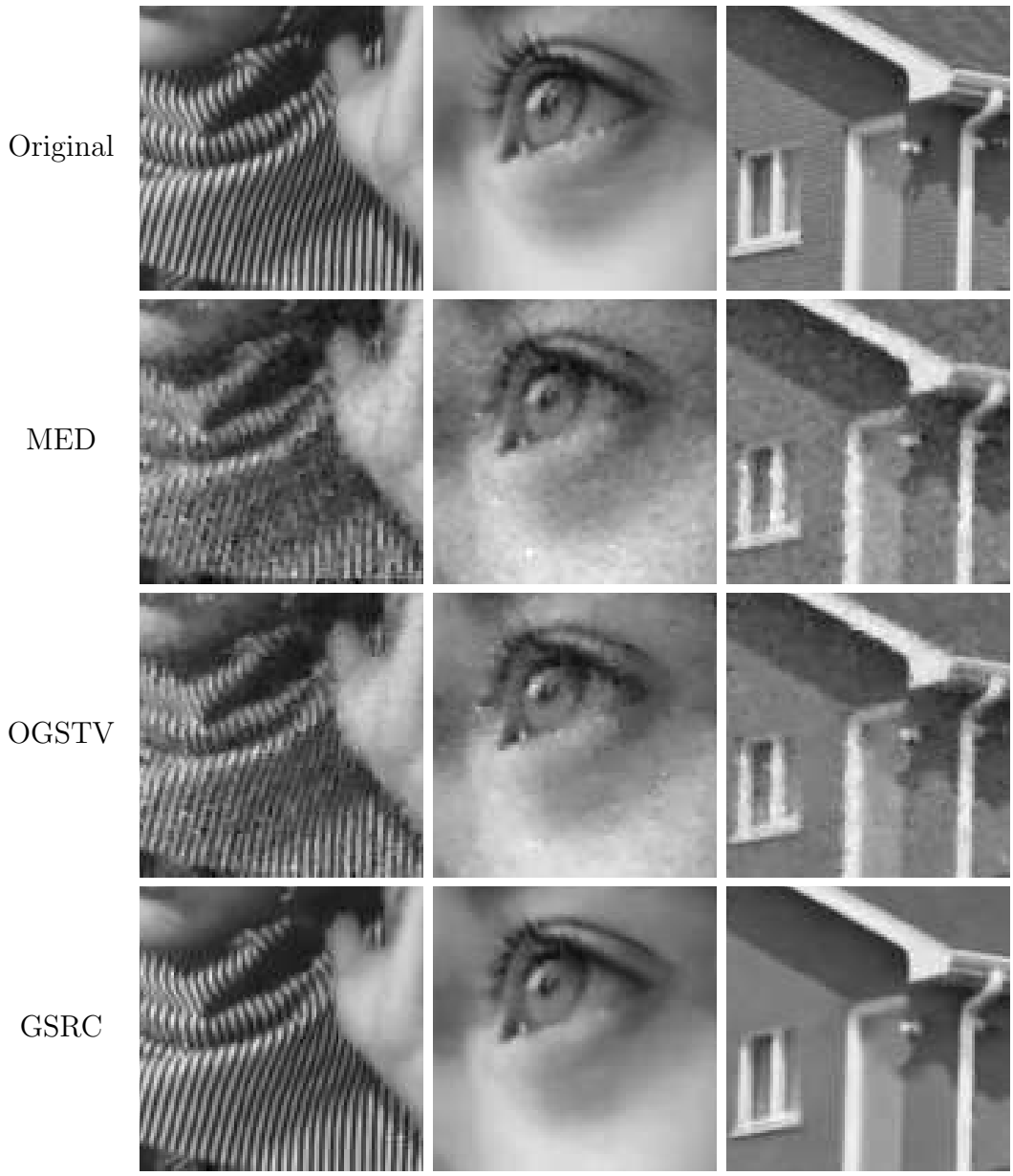


Figure 3.5: The zoomed-in version of the restored images in Fig. 3.3.

Table 3.1: The PSNR and SSIM values of the images degraded by Cauchy noise ($\gamma = 5$) and the restored images by different methods. The best values are marked in bold.

Image	PSNR				SSIM			
	Noisy	MED	OGSTV	GSRC	Noisy	MED	OGSTV	GSRC
Baby	19.27	31.44	32.56	**34.77**	0.3093	0.8762	0.9069	**0.9454**
Barbara	19.23	25.86	28.92	**34.46**	0.4380	0.7906	0.8657	**0.9517**
Beauty	19.20	29.87	33.01	**38.84**	0.2932	0.8958	0.9122	**0.9663**
Boats	19.24	29.18	30.92	**33.93**	0.4046	0.8340	0.8805	**0.9318**
Cameraman	19.15	26.14	29.07	**30.56**	0.3542	0.7842	0.8710	**0.8910**
Einstein	19.21	30.48	31.43	**32.71**	0.3358	0.8001	0.8371	**0.8576**
Goldhill	19.27	29.32	30.78	**32.04**	0.3944	0.8132	0.8512	**0.8899**
House	19.30	31.01	32.83	**35.84**	0.2911	0.8030	0.8597	**0.9079**
Lena	19.20	29.62	30.99	**33.63**	0.3624	0.8555	0.8855	**0.9338**
Parrot	19.20	27.11	29.73	**30.69**	0.3873	0.8243	0.8807	**0.9045**
Peppers	19.21	29.44	30.77	**32.23**	0.3867	0.8514	0.8838	**0.9113**
Plane	19.31	28.65	30.65	**31.81**	0.3844	0.8624	0.8985	**0.9253**
Average	19.23	29.01	30.97	**33.46**	0.3618	0.8326	0.8777	**0.9180**

In Figs. 3.2 and 3.3, we show the noisy images and the restored images using different methods at the two noise levels, $\gamma = 5$ and $\gamma = 10$. As shown, the median filter reduces the intensive variation of the Cauchy noise well but does not remove low-intensity noise. In contrast, OGSTV captures the overall noise better but yields staircase artifacts. Compared with these methods, the proposed GSRC efficiently removes Cauchy noise and restores a smooth region, while preserving the fine details without any artifacts.

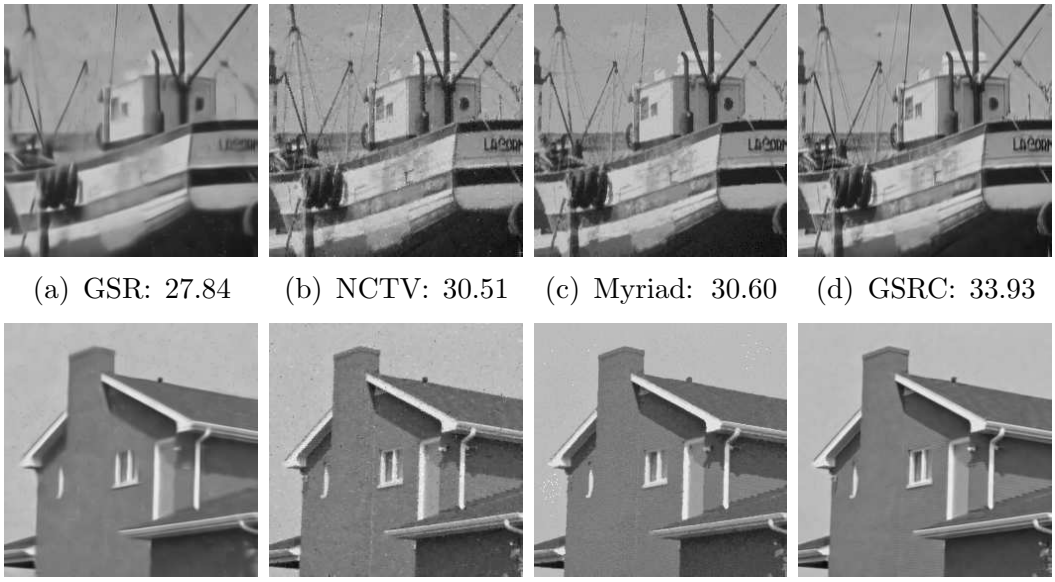
The visual difference can be highlighted in zoomed-in images, which are provided in Figs. 3.4 and 3.5. While other methods cannot clearly restore the texture pattern of Barbara, the smooth regions of Beauty, and the edges of

Table 3.2: The PSNR and SSIM values of the images degraded by Cauchy noise ($\gamma = 10$) and the restored images by different methods. The best values are marked in bold.

Image	PSNR				SSIM			
	Noisy	MED	OGSTV	GSRC	Noisy	MED	OGSTV	GSRC
Baby	16.40	28.52	29.82	**32.42**	0.1904	0.7434	0.8411	**0.9197**
Barbara	16.35	24.61	26.42	**31.84**	0.2980	0.6990	0.7763	**0.9208**
Beauty	16.33	27.75	30.34	**35.66**	0.1703	0.7750	0.8432	**0.9490**
Boats	16.36	27.00	28.24	**31.36**	0.2719	0.7336	0.8041	**0.8924**
Cameraman	16.24	24.88	27.13	**28.78**	0.2421	0.6606	0.7952	**0.8434**
Einstein	16.31	28.09	28.96	**31.01**	0.2107	0.6971	0.7616	**0.8120**
Goldhill	16.35	27.20	28.28	**29.80**	0.2487	0.7183	0.7698	**0.8194**
House	16.40	28.22	30.03	**33.93**	0.1847	0.6728	0.7830	**0.8747**
Lena	16.31	27.27	28.32	**31.40**	0.2399	0.7410	0.8073	**0.9041**
Parrot	16.31	25.43	27.37	**28.79**	0.2700	0.7084	0.8096	**0.8684**
Peppers	16.30	27.13	28.09	**30.49**	0.2616	0.7496	0.8143	**0.8831**
Plane	16.43	26.62	28.02	**29.76**	0.2662	0.7431	0.8313	**0.8940**
Average	16.34	26.89	28.42	**31.27**	0.2379	0.7202	0.8031	**0.8817**

House, it is obvious that GSRC yields outstanding results. For a quantitative comparison, we list the PSNR and SSIM values of the noisy images and restored images in Tables 3.1 and 3.2. Our method always yields the highest PSNR and SSIM values, which represent the superior quality of the restored images. Particularly, compared with OGSTV, our method yields 2.49 dB higher PSNR values on average for the low noise case ($\gamma = 5$) and 2.85 dB higher for the heavy noise case ($\gamma = 10$).

To further compare with other methods, in Fig. 3.6, we provide the results of removing Cauchy noise with $\gamma = 5$ using the original GSR method [66], the nonconvex TV model (NCTV) [43], and the nonlocal myriad filter



(a) GSR: 27.84 (b) NCTV: 30.51 (c) Myriad: 30.60 (d) GSRC: 33.93

(e) GSR: 31.42 (f) NCTV: 31.95 (g) Myriad: 32.30 (h) GSRC: 35.84

Figure 3.6: Comparison of the restored images by different methods for removing Cauchy noise ($\gamma = 5$). The PSNR values are shown below each image. From left to right: GSR, NCTV, Myriad, GSRC.

[37]. It is clear that the original GSR model produces oversmoothed images, thus justifying the adaptation of GSR for the Cauchy noise. NCTV and the myriad filter suffer from noise or artifacts. It is obvious that GSRC performs the best in both visual and numerical comparisons.

Image Deblurring and Denoising

The deblurring and denoising problems were simulated simultaneously, where $\boldsymbol{H}$ is given by a blurring operator. The test images were first blurred with either a Gaussian blur kernel of size 9 and standard deviation 1, or a motion blur kernel of length 8 and angle 30°; subsequently, Cauchy noise with $\gamma = 5$ was added. The parameters λ and μ were set as follows: For the Gaussian blur, $\lambda = 0.05, \mu = 0.003$ were used; for the motion blur, $\lambda = 0.05, \mu = 0.001$

were used. In both cases, the initial different values of $\lambda^{(0)}$ were not used, and λ was constant for all iterations. The gradient descent algorithm (3.30) was executed for 300 iterations, with a step size $\eta = 80$.

Table 3.3: The PSNR and SSIM values of the images degraded by a Gaussian blur (size 9 and standard deviation 1) and Cauchy noise ($\gamma = 5$) and the restored images by different methods. The best values are marked in bold.

Image	PSNR				SSIM			
	Blurred	MED	OGSTV	GSRC	Blurred	MED	OGSTV	GSRC
Baby	18.97	28.96	30.87	**33.07**	0.2678	0.8308	0.8591	**0.9213**
Barbara	18.39	24.18	25.01	**29.33**	0.2988	0.6842	0.7217	**0.8648**
Beauty	19.04	30.23	31.47	**36.62**	0.2714	0.8829	0.8846	**0.9536**
Boats	18.81	27.06	28.88	**31.02**	0.3256	0.7720	0.8161	**0.8819**
Cameraman	18.30	24.47	26.52	**28.00**	0.2588	0.7356	0.7880	**0.8317**
Einstein	18.75	27.46	29.11	**31.53**	0.2511	0.7447	0.7748	**0.8290**
Goldhill	18.87	27.56	28.84	**30.19**	0.3016	0.7383	0.7768	**0.8322**
House	19.03	29.00	30.96	**33.66**	0.2409	0.7757	0.8028	**0.8586**
Lena	18.77	27.25	28.88	**30.95**	0.2982	0.7970	0.8298	**0.8980**
Parrot	18.38	24.47	27.19	**28.64**	0.3104	0.7777	0.8214	**0.8694**
Peppers	18.40	25.28	26.78	**29.59**	0.3056	0.7870	0.8157	**0.8769**
Plane	18.76	26.11	28.18	**29.96**	0.3112	0.7977	0.8365	**0.8949**
Average	18.71	26.84	28.56	**31.05**	0.2868	0.7770	0.8106	**0.8760**

In Figs. 3.7 and 3.8, we present the degraded images and restored images from different methods for the cases of Gaussian blur and motion blur, respectively. We observed that the median filter smoothes out the details and noise while not deblurring the images, which remain fuzzy especially in the motion blur case. OGSTV exhibits better deblurring performance but produces noise and artifacts. By contrast, our method shows exceptional performance in both deblurring and denoising, while maintaining the fine details

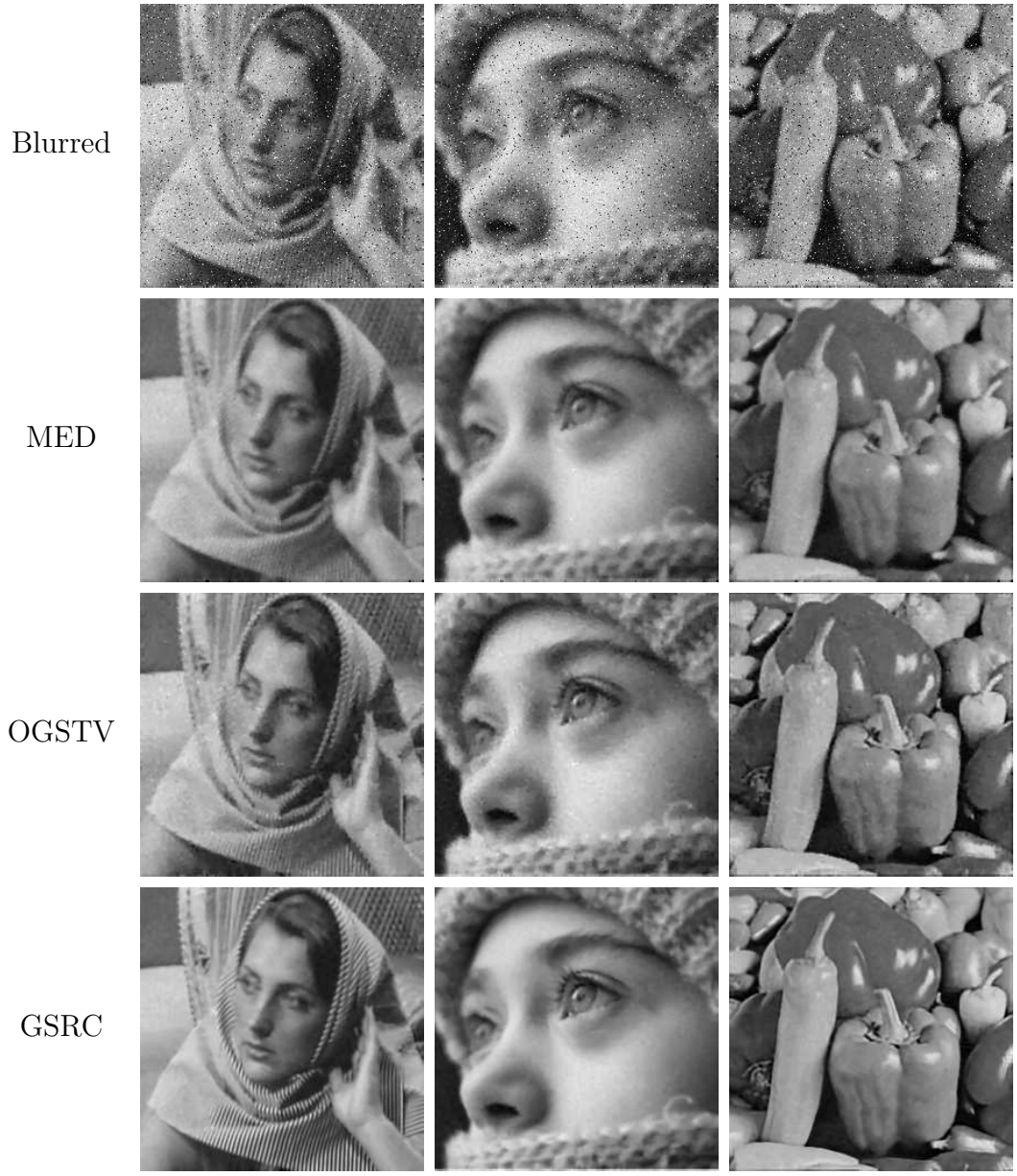


Figure 3.7: Comparison of different methods for restoring images degraded by a Gaussian blur (size 9 and $\sigma = 1$) and Cauchy noise ($\gamma = 5$).

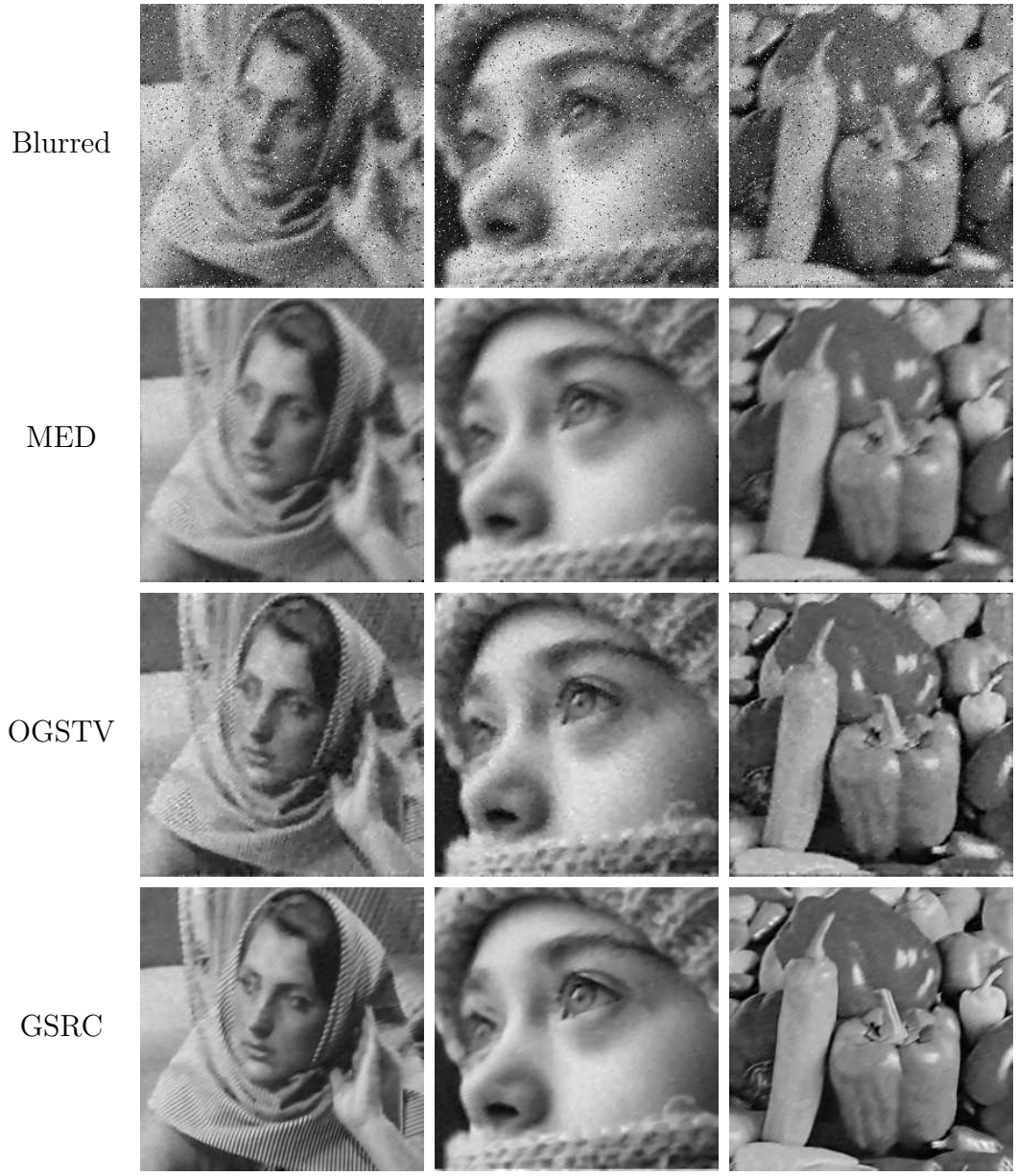


Figure 3.8: Comparison of different methods for restoring images degraded by a motion blur (length 8 and angle 30°) and Cauchy noise ($\gamma = 5$).

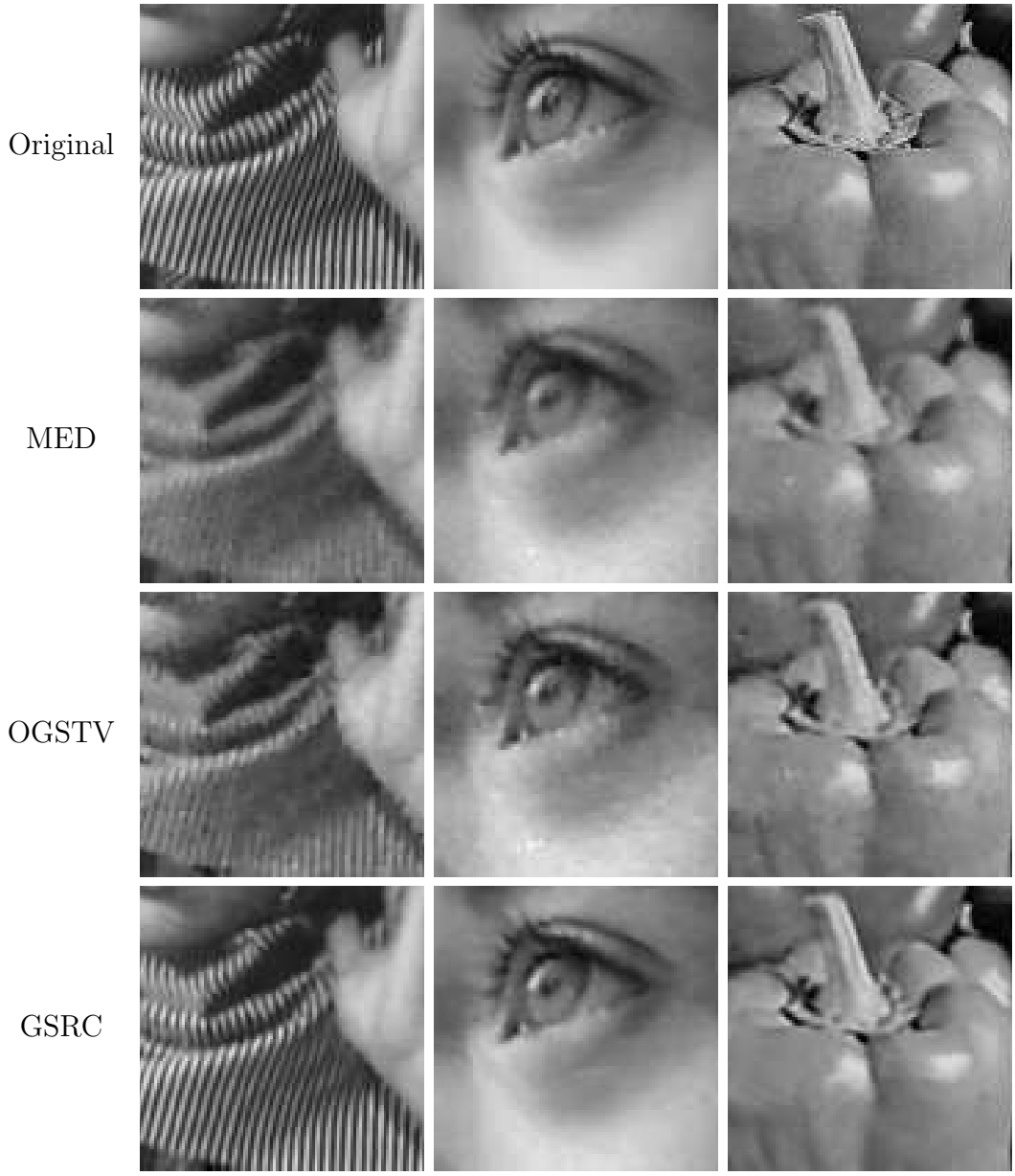


Figure 3.9: The zoomed-in version of the restored images in Fig. 3.7.

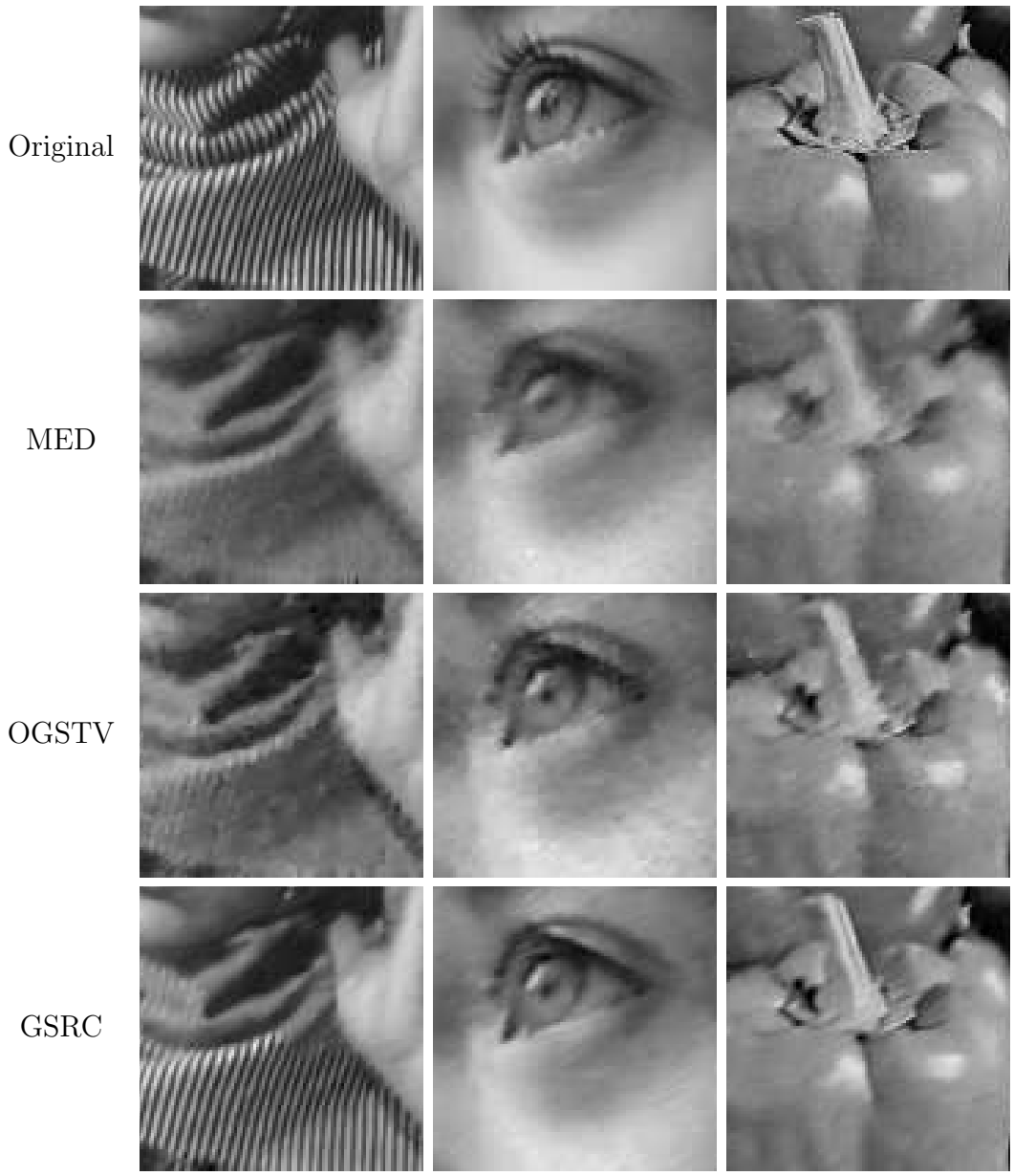


Figure 3.10: The zoomed-in version of the restored images in Fig. 3.8.

Table 3.4: The PSNR and SSIM values of the images degraded by a motion blur (length 8 and angle 30°) and Cauchy noise ($\gamma = 5$) and the restored images by different methods. The best values are marked in bold.

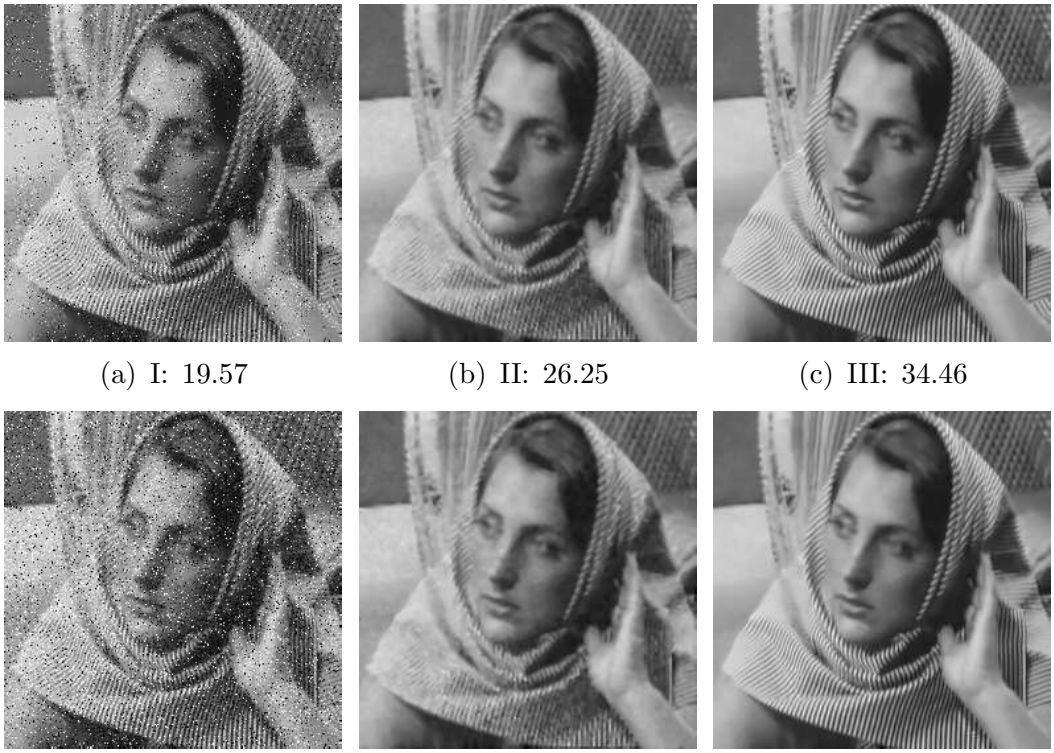
Image	PSNR				SSIM			
	Blurred	MED	OGSTV	GSRC	Blurred	MED	OGSTV	GSRC
Baby	18.48	25.86	28.44	**30.91**	0.2287	0.7590	0.7982	**0.8618**
Barbara	17.72	22.92	23.96	**27.19**	0.2154	0.5936	0.6624	**0.7982**
Beauty	18.87	28.77	29.84	**33.85**	0.2520	0.8473	0.8417	**0.9090**
Boats	18.23	24.51	26.54	**28.65**	0.2599	0.6694	0.7361	**0.8069**
Cameraman	17.48	22.17	24.49	**26.85**	0.1934	0.6480	0.7300	**0.7734**
Einstein	18.22	24.98	26.89	**29.60**	0.1837	0.6267	0.6900	**0.7534**
Goldhill	18.45	25.47	27.15	**28.59**	0.2402	0.6333	0.7071	**0.7720**
House	18.62	26.34	29.10	**31.87**	0.2074	0.7169	0.7597	**0.8121**
Lena	18.21	24.62	27.01	**29.12**	0.2449	0.7083	0.7690	**0.8409**
Parrot	17.56	22.09	25.18	**27.29**	0.2583	0.6993	0.7728	**0.8188**
Peppers	17.80	23.19	25.13	**28.64**	0.2549	0.7045	0.7580	**0.8321**
Plane	18.08	23.69	25.84	**27.76**	0.2449	0.7074	0.7671	**0.8252**
Average	18.14	24.55	26.63	**29.19**	0.2320	0.6928	0.7493	**0.8170**

of the images.

From the zoomed-in results in Figs. 3.9 and 3.10, the differences with other methods can be clarified. Compared with the blurry results from MED and the noisy details from OGSTV, our method can restore the stripes of Barbara, the eye of Beauty, and the stem of Peppers more clearly without artifacts. As shown in Tables 3.3 and 3.4, GSRC always obtains the highest PSNR and SSIM values. Specifically, GSRC outperforms OGSTV by 2.49 dB on average in the Gaussian blur case and by 2.56 dB in the motion blur case.

3.1.5 Discussion

Initialization Technique



(a) I: 19.57 (b) II: 26.25 (c) III: 34.46

(d) I: 16.51 (e) II: 25.49 (f) III: 31.84

Figure 3.11: Results of GSRC from different initial conditions for the images degraded by Cauchy noise (first row: $\gamma = 5$; second row: $\gamma = 10$). The PSNR values are shown below each image. From left to right: (I) $\boldsymbol{u}^{(0)} = \boldsymbol{y}$, λ fixed; (II) $\boldsymbol{u}^{(0)} = med(\boldsymbol{y})$, λ fixed; (III) $\boldsymbol{u}^{(0)} = \boldsymbol{y}$, $\lambda^{(0)} = 2$.

Because the data fidelity term of the proposed GSRC (3.7) is nonconvex, the solution may depend on the initial guess $\boldsymbol{u}^{(0)}$ of the image. To investigate this, three experiments are designed to denoise the Barbara image corrupted by Cauchy noise with $\gamma = 5$ and $\gamma = 10$. The proposed GSRC is applied with

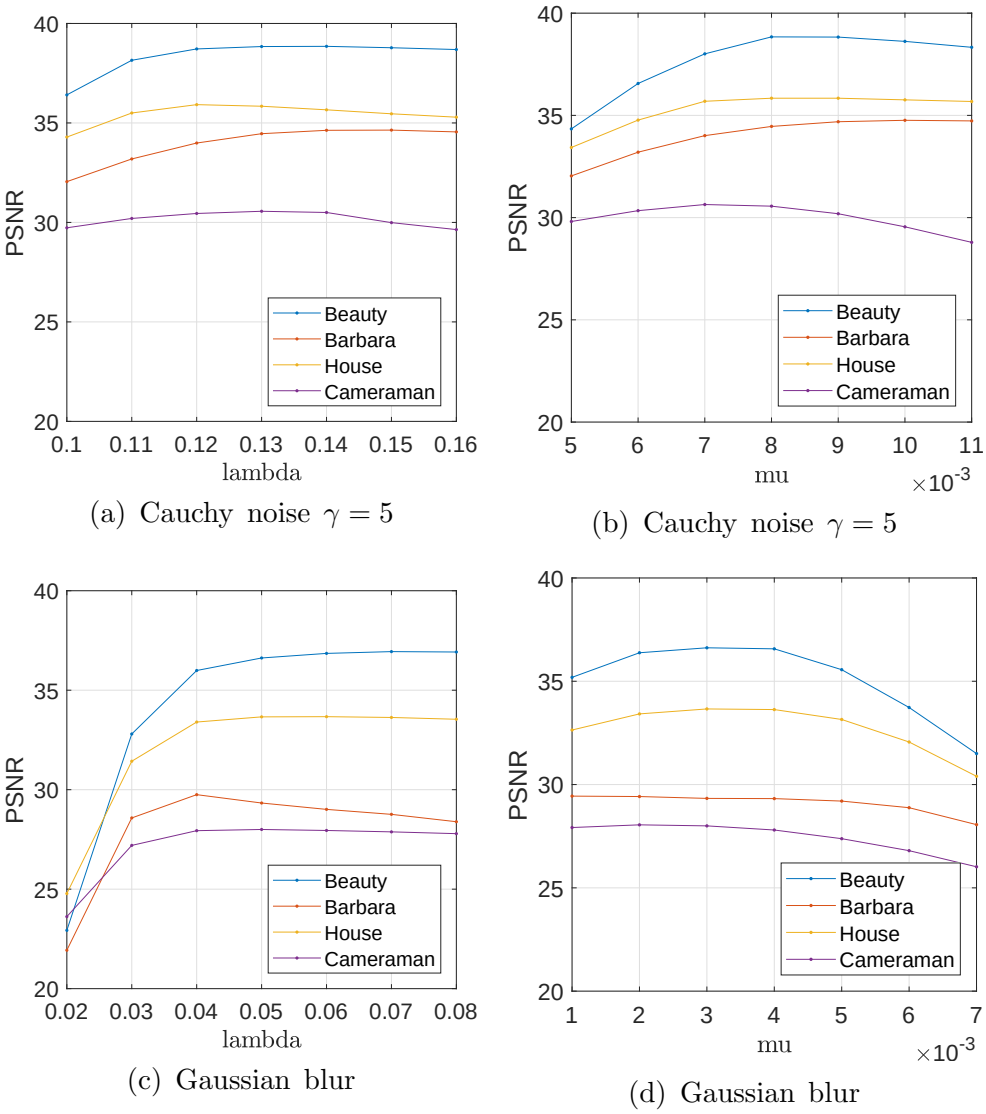
different initial conditions as follows. (I): $\boldsymbol{u}^{(0)}$ is the noisy image $\boldsymbol{y}$ and λ is fixed. (II): $\boldsymbol{u}^{(0)}$ is the median filtered noisy image $med(\boldsymbol{y})$ and λ is fixed. (III): $\boldsymbol{u}^{(0)}$ is the noisy image $\boldsymbol{y}$ and λ is fixed except at $\lambda^{(0)} = 2$. In all cases, the fixed value of the regularization parameter λ and other parameter values are the same as those in Section 3.1.4.

The results are presented in Fig. 3.11. As shown, in (I), denoising barely occurred while in (II), the fine features are not restored clearly. To obtain better results from the noisy image in (I), λ should be increased, but then the denoising result will be too smoothed to capture the details. In other words, λ should be sufficiently low to avoid oversmoothing and high enough to smooth out noise from the initial image. To balance this, consider the $\boldsymbol{\alpha_G}$-subproblem (3.11) at the first iteration of the algorithm. By enforcing the sparsity regularization intensively with a high value of $\lambda^{(0)}$ at the first iteration, the initial noise is effectively reduced. Subsequently, the small value of λ is used to remove noise while maintaining the details. Compared with (I) or (II), the initialization technique of selecting a high $\lambda^{(0)}$ yields by far the best results in (III).

Analysis of Parameters

Plenty of parameters can be adjusted in the GSRC algorithm, as shown in Section 3.1.4. Therefore, the tuning of the optimal parameters while considering all variations of the parameters is difficult. Nonetheless, we performed extensive tests to tune the parameters; for example, we used the patch size $P = [6 \times 6, 7 \times 7, 8 \times 8]$, the stride between patches $s = [2, 3, 4]$, the training window size $L \times L$ for $L = [20, 30, 40]$, and the number of patches in a group $c = [40, 50, 60, 70, 80]$.

The test results of these parameters are not discussed herein. Nonetheless, we provide the results for two crucial parameters: the regularization parameter λ and the penalty parameter μ. In Fig. 3.12, the PSNR values of



(a) Cauchy noise $\gamma = 5$

(b) Cauchy noise $\gamma = 5$

(c) Gaussian blur

(d) Gaussian blur

Figure 3.12: Plots of the PSNR values versus the parameter values of GSRC for 4 test images in the two experimental simulations. (a) and (b): PSNR values versus λ and μ, respectively, for noisy images with $\gamma = 5$; (c) and (d): PSNR values versus λ and μ, respectively, for Gaussian blurred and noisy images with $\gamma = 5$.

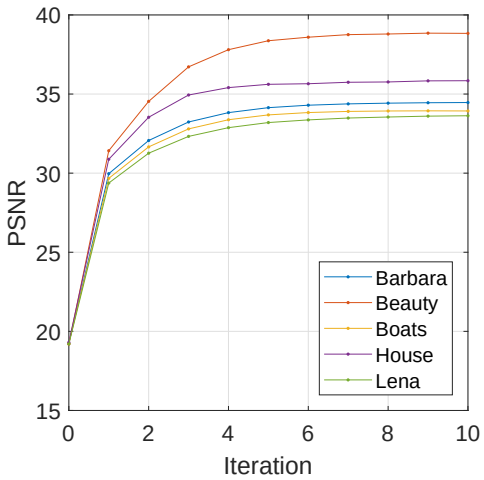
the results from varying λ and μ are plotted, in the experiments of restoring noisy images with $\gamma = 5$ and Gaussian blurred and noisy images with $\gamma = 5$. Because of the time consumption, we tuned the parameters based on the experiments on the image Cameraman; however, Fig. 3.12 indicates that λ and μ exhibit consistent behaviors, implying the stable performance of the proposed algorithm. In fact, we noticed that the fine tuning for each image could yield better results.

Analysis of Convergence

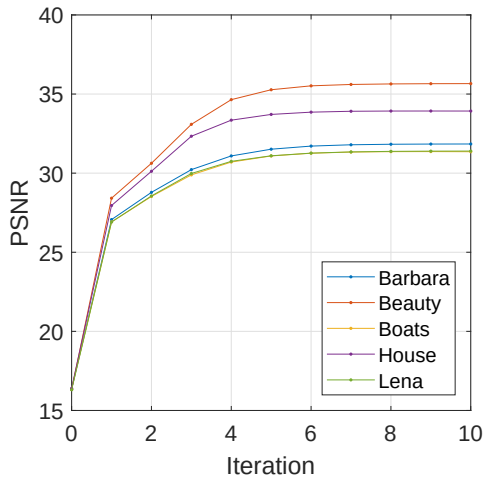
As shown in Section 3.1.5, the nonconvex property of GSRC causes it to depend on the initial image, and it does not have a global optimizer. Hence, a theoretical proof of the convergence of the algorithm to a global minimum cannot be obtained. Nonetheless, we present the numerical analysis of the convergence. In Fig. 3.13, the PSNR values versus the iteration number are shown for different types of images in all experimental simulations of Section 3.1.4. As shown, the PSNR values increase monotonically and converge asymptotically, which support the numerical convergence of the algorithm.

3.2 Proposed Model 2: GSRS

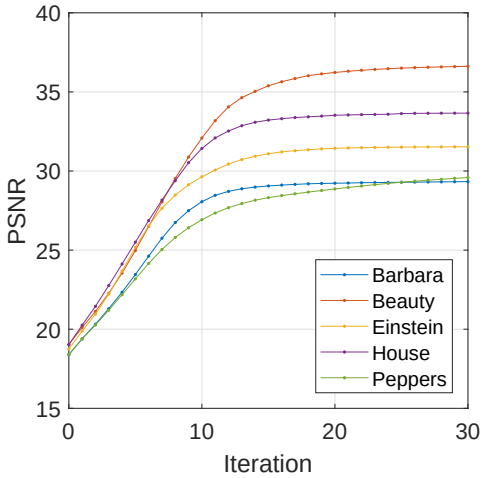
In this section, we utilize GSR framework to restore images corrupted by speckle noise. Since the speckle noise is the multiplicative noise with Gamma distribution, the original GSR model should be adapted in a proper manner. We propose the modeling of GSRS via MAP estimation and introduce the patch distance for speckle noise. The relevant techniques and the analyses are also discussed.



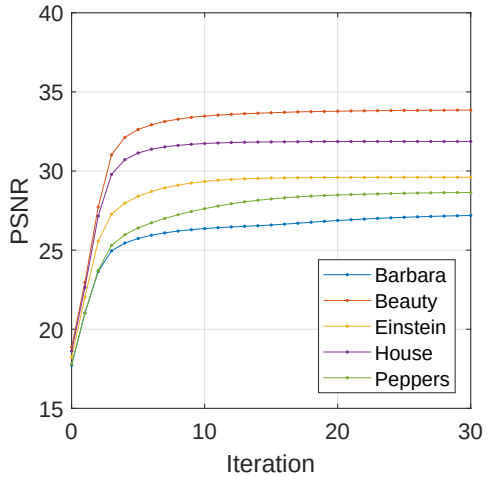
(a) Cauchy noise $\gamma = 5$



(b) Cauchy noise $\gamma = 10$



(c) Gaussian blur



(d) motion blur

Figure 3.13: Plots of the PSNR values versus the iteration number of GSRC for 5 test images in experiments for restoring images degraded by (a) Cauchy noise with $\gamma = 5$; (b) Cauchy noise with $\gamma = 10$; (c) Gaussian blur and Cauchy noise with $\gamma = 5$; (d) motion blur and Cauchy noise with $\gamma = 5$.

3.2.1 GSRS Modeling via MAP Estimator

The ℓ_2 data fidelity term in (2.17) and the Cauchy data fidelity term in (3.7) are not suitable for the speckle noise. In the AA model [4], the data fidelity term appropriate for speckle noise is derived using MAP estimation, and the convex data fidelity term are introduced in [54, 56]. Following these, we derive the modeling of GSRS.

Recall the multiplicative noise formulation $\boldsymbol{y} = \boldsymbol{x} * \boldsymbol{n}$, where we want to restore the original image $\boldsymbol{x}$ from the noisy image $\boldsymbol{y}$, and $\boldsymbol{n}$ follows the Gamma distribution. We consider $\boldsymbol{y}(i)$, $\boldsymbol{x}(i)$ as random variables for each $i = 1, 2, \ldots, N$. The MAP estimation is to find $\hat{\boldsymbol{x}}$, which maximizes the posterior probability $P(\boldsymbol{x} \mid \boldsymbol{y})$. Using Bayes' theorem,

$$\begin{aligned}\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} P(\boldsymbol{x} \mid \boldsymbol{y}) &= \arg\max_{\boldsymbol{x}} \frac{P(\boldsymbol{y} \mid \boldsymbol{x})P(\boldsymbol{x})}{P(\boldsymbol{y})} \\ &= \arg\max_{\boldsymbol{x}} P(\boldsymbol{y} \mid \boldsymbol{x})P(\boldsymbol{x}). \end{aligned} \tag{3.32}$$

It is equivalent to find the minimizer of the negative logarithm of (3.1) as the following:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} -\log P(\boldsymbol{y} \mid \boldsymbol{x}) - \log P(\boldsymbol{x}). \tag{3.33}$$

From the assumption that the image values are independent and identically distributed, we have $P(\boldsymbol{y} \mid \boldsymbol{x}) = \prod_{i=1}^{N} P(\boldsymbol{y}(i) \mid \boldsymbol{x}(i))$. Considering the probability density function (2.8) of speckle noise and the fact that $P_{y|x}(y \mid x) = P_n(\frac{y}{x})\frac{1}{x}$ (see [4, 31]),

$$P(\boldsymbol{y}(i) \mid \boldsymbol{x}(i)) = \frac{L^L \boldsymbol{y}(i)^{L-1}}{\Gamma(L)\boldsymbol{x}(i)^L} e^{-L\frac{\boldsymbol{y}(i)}{\boldsymbol{x}(i)}}. \tag{3.34}$$

We use the regularizer for the prior probability $P(\boldsymbol{x})$ as

$$P(\boldsymbol{x}) = \exp(-L\lambda\Psi(\boldsymbol{x})), \tag{3.35}$$

where $\lambda > 0$ is a regularization parameter. Substituting these into (3.33), we have

$$\begin{aligned}
\hat{\boldsymbol{x}} &= \arg\min_{\boldsymbol{x}} -\log \prod_{i=1}^{N} P(\boldsymbol{y}(i) \mid \boldsymbol{x}(i)) + L\lambda\Psi(\boldsymbol{x}) \\
&= \arg\min_{\boldsymbol{x}} -\sum_{i=1}^{N} \log \left(\frac{L^L \boldsymbol{y}(i)^{L-1}}{\Gamma(L)\boldsymbol{x}(i)^L} e^{-L\frac{\boldsymbol{y}(i)}{\boldsymbol{x}(i)}} \right) + L\lambda\Psi(\boldsymbol{x}) \\
&= \arg\min_{\boldsymbol{x}} \sum_{i=1}^{N} \log \left(\boldsymbol{x}(i)^L e^{L\frac{\boldsymbol{y}(i)}{\boldsymbol{x}(i)}} \right) + L\lambda\Psi(\boldsymbol{x}) \\
&= \arg\min_{\boldsymbol{x}} \sum_{i=1}^{N} L \left(\log \boldsymbol{x}(i) + \frac{\boldsymbol{y}(i)}{\boldsymbol{x}(i)} \right) + L\lambda\Psi(\boldsymbol{x})
\end{aligned} \tag{3.36}$$

where we omit the unnecessary constants. The variational model for speckle noise reduction can be written as

$$\arg\min_{\boldsymbol{x}} \left\langle \log \boldsymbol{x} + \frac{\boldsymbol{y}}{\boldsymbol{x}}, \mathbf{1} \right\rangle + \lambda\Psi(\boldsymbol{x}), \tag{3.37}$$

where $\langle \cdot, \cdot \rangle$ denotes the standard inner product and $\mathbf{1} \in \mathbb{R}^N$ is a vector of ones. This is the derivation of the AA model if $\Psi(\boldsymbol{x}) = \|\nabla \boldsymbol{x}\|_1$ is used.

To derive the I-divergence fidelity term in [56], we use

$$\Psi(\boldsymbol{x}) = \int \frac{\Phi'(\boldsymbol{x})}{\boldsymbol{x}} = \frac{\Phi(\boldsymbol{x})}{\boldsymbol{x}} + \int \frac{\Phi(\boldsymbol{x})}{\boldsymbol{x}^2}. \tag{3.38}$$

Then we have the optimality condition of (3.37) as

$$\frac{\mathbf{1}}{\boldsymbol{x}} - \frac{\boldsymbol{y}}{\boldsymbol{x}^2} + \lambda \frac{\Phi'(\boldsymbol{x})}{\boldsymbol{x}} = 0, \tag{3.39}$$

which is equivalent to (for $\boldsymbol{x} > 0$)

$$\mathbf{1} - \frac{\boldsymbol{y}}{\boldsymbol{x}} + \lambda \Phi'(\boldsymbol{x}) = 0. \tag{3.40}$$

Corresponding to this optimality condition is the I-divergence model [56]

$$\arg\min_{\boldsymbol{x}} \langle \boldsymbol{x} - \boldsymbol{y} \log \boldsymbol{x}, \mathbf{1} \rangle + \lambda \Phi(\boldsymbol{x}). \tag{3.41}$$

From the assumption of GSR, we consider $\boldsymbol{x} \approx \boldsymbol{D_G} \circ \boldsymbol{\alpha_G}$ (see (2.16)) and the regularization term as $\Phi(\boldsymbol{\alpha_G}) = \|\boldsymbol{\alpha_G}\|_0$. Hence the modeling of GSRS can be formulated as

$$\arg\min_{\boldsymbol{\alpha_G}} \langle \boldsymbol{D_G} \circ \boldsymbol{\alpha_G} - \boldsymbol{y} \log \left(\boldsymbol{D_G} \circ \boldsymbol{\alpha_G} \right), \mathbf{1} \rangle + \lambda \|\boldsymbol{\alpha_G}\|_0. \tag{3.42}$$

3.2.2 Patch Distance for Speckle Noise

Since the image patches have different noise density distribution, the appropriate patch distance for speckle noise should be considered in the process of group construction (Section 2.3.1). By using the suitable patch distance, similar patches can be found well and the overall performance of algorithm can be improved. The patch similarity for the speckle noise has been presented in [18, 48] as follows.

Given two patches $\boldsymbol{y}_1, \boldsymbol{y}_2$ of initial noisy image, the patch distance for

speckle noise in the initial step is defined by

$$d_0(\boldsymbol{y}_1, \boldsymbol{y}_2) = (2L-1)\sum_{i=1}^{P} \log\left[\frac{\sqrt{\boldsymbol{y}_1}(i)}{\sqrt{\boldsymbol{y}_2}(i)} + \frac{\sqrt{\boldsymbol{y}_2}(i)}{\sqrt{\boldsymbol{y}_1}(i)}\right]. \tag{3.43}$$

After the initial step of the algorithm, we have an estimate $\boldsymbol{x}$ of original image. The patch distance for two patches $\boldsymbol{x}_1, \boldsymbol{x}_2$ in the following step is defined by

$$d(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sum_{i=1}^{P}\left[(2L-1)\log\left(\frac{\sqrt{\boldsymbol{y}_1}(i)}{\sqrt{\boldsymbol{y}_2}(i)} + \frac{\sqrt{\boldsymbol{y}_2}(i)}{\sqrt{\boldsymbol{y}_1}(i)}\right) + \gamma L\frac{|\boldsymbol{x}_1(i) - \boldsymbol{x}_2(i)|^2}{\boldsymbol{x}_1(i)\boldsymbol{x}_2(i)}\right]. \tag{3.44}$$

Hence, when we find similar patches to construct the group $\boldsymbol{x}_{\boldsymbol{G}_k}$, (3.43) is used at the first iteration and then (3.44) is used.

3.2.3 The ADMM Algorithm for Solving (3.42)

Although the data fidelity term of (3.42) is convex, the ℓ_0 regularization term is nonconvex and the optimization problem is hard to solve. To find the minimizer efficiently, we again adopt the ADMM algorithm which was introduced in Section 2.4. As mentioned, the nonconvex ADMM was proven to have subsequential convergence to a stationary point [58].

By introducing an auxiliary variable $\boldsymbol{u}$, we have the equivalent formulation of (3.42) as the constrained version

$$\arg\min_{\boldsymbol{\alpha}_{\boldsymbol{G}}, \boldsymbol{u}} \langle \boldsymbol{u} - \boldsymbol{y}\log\boldsymbol{u}, \boldsymbol{1}\rangle + \lambda\|\boldsymbol{\alpha}_{\boldsymbol{G}}\|_0, \quad \text{s.t. } \boldsymbol{u} = \boldsymbol{D}_{\boldsymbol{G}} \circ \boldsymbol{\alpha}_{\boldsymbol{G}}. \tag{3.45}$$

We apply the ADMM algorithm (see Algorithm 1) for an efficient optimization. By setting $f(\boldsymbol{\alpha}_{\boldsymbol{G}}) = \lambda\|\boldsymbol{\alpha}_{\boldsymbol{G}}\|_0$ and $g(\boldsymbol{u}) = \langle \boldsymbol{u} - \boldsymbol{y}\log\boldsymbol{u}, \boldsymbol{1}\rangle$, (3.45) can be minimized by iteratively solving the two subproblems of $\boldsymbol{\alpha}_{\boldsymbol{G}}$ and $\boldsymbol{u}$, with

the update of the dual variable $\boldsymbol{b}$ as follows:

$$\boldsymbol{\alpha}_{\boldsymbol{G}}^{(t+1)} = \arg\min_{\boldsymbol{\alpha}_{\boldsymbol{G}}} \lambda\|\boldsymbol{\alpha}_{\boldsymbol{G}}\|_0 + \frac{\mu}{2}\|\boldsymbol{u}^{(t)} - \boldsymbol{D}_{\boldsymbol{G}} \circ \boldsymbol{\alpha}_{\boldsymbol{G}} + \boldsymbol{b}^{(t)}\|_2^2 \tag{3.46}$$

$$\boldsymbol{u}^{(t+1)} = \arg\min_{\boldsymbol{u}} \langle \boldsymbol{u} - \boldsymbol{y}\log\boldsymbol{u}, \boldsymbol{1}\rangle + \frac{\mu}{2}\|\boldsymbol{u} - \boldsymbol{D}_{\boldsymbol{G}} \circ \boldsymbol{\alpha}_{\boldsymbol{G}}^{(t+1)} + \boldsymbol{b}^{(t)}\|_2^2 \tag{3.47}$$

$$\boldsymbol{b}^{(t+1)} = \boldsymbol{b}^{(t)} + \boldsymbol{u}^{(t+1)} - \boldsymbol{D}_{\boldsymbol{G}} \circ \boldsymbol{\alpha}_{\boldsymbol{G}}^{(t+1)} \tag{3.48}$$

In the following, we explain the details to solve the $\boldsymbol{\alpha}_{\boldsymbol{G}}$-subproblem (3.46) and the $\boldsymbol{u}$-subproblem (3.47).

$\boldsymbol{\alpha}_{\boldsymbol{G}}$-subproblem

Because the $\boldsymbol{\alpha}_{\boldsymbol{G}}$-subproblem (3.46) is the same as in GSRC and the original GSR model, we briefly explain and refer to [66]. The $\boldsymbol{\alpha}_{\boldsymbol{G}}$-subproblem can be summarized as

$$\arg\min_{\boldsymbol{\alpha}_{\boldsymbol{G}}} \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{r}\|_2^2 + \frac{\lambda}{\mu}\|\boldsymbol{\alpha}_{\boldsymbol{G}}\|_0, \tag{3.49}$$

where $\boldsymbol{x} = \boldsymbol{D}_{\boldsymbol{G}} \circ \boldsymbol{\alpha}_{\boldsymbol{G}}$, and $\boldsymbol{r} = \boldsymbol{u}^{(t)} + \boldsymbol{b}^{(t)}$ is a noisy version of $\boldsymbol{x}$.

Substituting the approximation of (3.15) into (3.49) yields the equivalent formulation

$$\arg\min_{\boldsymbol{\alpha}_{\boldsymbol{G}}} \sum_{k=1}^{n} \left(\frac{1}{2}\|\boldsymbol{x}_{\boldsymbol{G}_k} - \boldsymbol{r}_{\boldsymbol{G}_k}\|_F^2 + \tau\|\boldsymbol{\alpha}_{\boldsymbol{G}_k}\|_0 \right), \tag{3.50}$$

where $\tau = \frac{\lambda K}{\mu N}$. Applying the singular value decomposition to $\boldsymbol{r}_{\boldsymbol{G}_k} \in \mathbb{R}^{P\times c}$ gives

$$\boldsymbol{r}_{\boldsymbol{G}_k} = \boldsymbol{U}_{\boldsymbol{G}_k}\boldsymbol{\Sigma}_{\boldsymbol{G}_k}\boldsymbol{V}_{\boldsymbol{G}_k}^T = \sum_{i=1}^{m} \boldsymbol{\gamma}_{\boldsymbol{r}_{\boldsymbol{G}_k},i}(\boldsymbol{u}_{\boldsymbol{G}_k,i}\boldsymbol{v}_{\boldsymbol{G}_k,i}{}^T), \tag{3.51}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{G}_k} = \mathrm{diag}(\boldsymbol{\gamma}_{\boldsymbol{r}_{\boldsymbol{G}_k}})$, $\boldsymbol{\gamma}_{\boldsymbol{r}_{\boldsymbol{G}_k}} = [\boldsymbol{\gamma}_{\boldsymbol{r}_{\boldsymbol{G}_k},1}, \boldsymbol{\gamma}_{\boldsymbol{r}_{\boldsymbol{G}_k},2}, \ldots, \boldsymbol{\gamma}_{\boldsymbol{r}_{\boldsymbol{G}_k},m}] \in \mathbb{R}^m$ is the singular value vector and $\boldsymbol{u}_{\boldsymbol{G}_k,i}, \boldsymbol{v}_{\boldsymbol{G}_k,i}$ are the columns of $\boldsymbol{U}_{\boldsymbol{G}_k}, \boldsymbol{V}_{\boldsymbol{G}_k}$, respectively.

Define the self-adaptive dictionary by

$$\boldsymbol{D}_{\boldsymbol{G}_k} = [\boldsymbol{d}_{\boldsymbol{G}_k,1}, \boldsymbol{d}_{\boldsymbol{G}_k,2}, \ldots, \boldsymbol{d}_{\boldsymbol{G}_k,m}], \tag{3.52}$$

where the atoms are defined by

$$\boldsymbol{d}_{\boldsymbol{G}_k,i} = \boldsymbol{u}_{\boldsymbol{G}_k,i}\boldsymbol{v}_{\boldsymbol{G}_k,i}{}^T, \quad i = 1, 2, \ldots, m. \tag{3.53}$$

The $\boldsymbol{\alpha}_{\boldsymbol{G}_k}$-subproblem is to find $\boldsymbol{x}_{\boldsymbol{G}_k} = \boldsymbol{D}_{\boldsymbol{G}_k}\boldsymbol{\alpha}_{\boldsymbol{G}_k}$, given that $\boldsymbol{r}_{\boldsymbol{G}_k} = \boldsymbol{D}_{\boldsymbol{G}_k}\boldsymbol{\gamma}_{\boldsymbol{r}_{\boldsymbol{G}_k}}$. Since $\boldsymbol{U}_{\boldsymbol{G}_k}$ and $\boldsymbol{V}_{\boldsymbol{G}_k}$ are unitary, the n subproblems for the $\boldsymbol{\alpha}_{\boldsymbol{G}_k}$ of (3.50) can be solved by

$$\arg\min_{\boldsymbol{\alpha}_{\boldsymbol{G}_k}} \frac{1}{2}\|\boldsymbol{\alpha}_{\boldsymbol{G}_k} - \boldsymbol{\gamma}_{\boldsymbol{r}_{\boldsymbol{G}_k}}\|_F^2 + \tau\|\boldsymbol{\alpha}_{\boldsymbol{G}_k}\|_0. \tag{3.54}$$

(3.54) is obtained by the entry-wise hard thresholding operator [3],

$$\hat{\boldsymbol{\alpha}}_{\boldsymbol{G}_k} = hard(\boldsymbol{\gamma}_{\boldsymbol{r}_{\boldsymbol{G}_k}}, \sqrt{2\tau}) = \boldsymbol{\gamma}_{\boldsymbol{r}_{\boldsymbol{G}_k}} * \mathbf{1}\left(|\boldsymbol{\gamma}_{\boldsymbol{r}_{\boldsymbol{G}_k}}| - \sqrt{2\tau}\right), \tag{3.55}$$

where $\mathbf{1}(\cdot)$ is the indicator function $\mathbf{1}(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \boldsymbol{x}_i > 0 \\ 0 & \text{if } \boldsymbol{x}_i \leq 0 \end{cases}$.

$\boldsymbol{u}$-subproblem

The optimality condition of (3.47) is given by

$$\mathbf{1} - \frac{\boldsymbol{y}}{\boldsymbol{u}} + \mu(\boldsymbol{u} - \boldsymbol{D}_{\boldsymbol{G}} \circ \boldsymbol{\alpha}_{\boldsymbol{G}}^{(t+1)} + \boldsymbol{b}^{(t)}) = 0, \tag{3.56}$$

where the division implies the entry-wise division. By multiplying with the denominator, (3.56) becomes a quadratic equation:

$$\mu \boldsymbol{u}^2 + (\mathbf{1} - \mu \boldsymbol{D_G} \circ \boldsymbol{\alpha}_{\boldsymbol{G}}^{(t+1)} + \mu \boldsymbol{b}^{(t)})\boldsymbol{u} - \boldsymbol{y} = 0. \tag{3.57}$$

It can be solved explicitly using the quadratic formula,

$$\boldsymbol{u} = \frac{1}{2\mu}\left(-(\mathbf{1} - \mu \boldsymbol{D_G} \circ \boldsymbol{\alpha}_{\boldsymbol{G}}^{(t+1)} + \mu \boldsymbol{b}^{(t)}) + \sqrt{(\mathbf{1} - \mu \boldsymbol{D_G} \circ \boldsymbol{\alpha}_{\boldsymbol{G}}^{(t+1)} + \mu \boldsymbol{b}^{(t)})^2 + 4\mu \boldsymbol{y}}\right). \tag{3.58}$$

Note that $\boldsymbol{u} > 0$ is satisfied.

In Algorithm 3, the proposed algorithm is summarized.

3.2.4 Numerical Experiments

Various numerical experiments were simulated to demonstrate the performance of the proposed algorithm for image despeckling. Test images are 8 grayscale images (6 natural images and 2 SAR images) of size 256×256, whose values are in the range of $[0, 255]$. The test images are presented in Fig. 3.14. All experiments were performed by using MATLAB R2017b and Windows 10 on a PC with Intel(R) Core(TM) i7-4790 CPU @ 3.60 GHz, 16.0 GB RAM.

In the following experiments, the speckle noise with L looks which has probability density function (2.8) is generated by

$$\boldsymbol{y} = \boldsymbol{x} * \boldsymbol{n}. \tag{3.59}$$

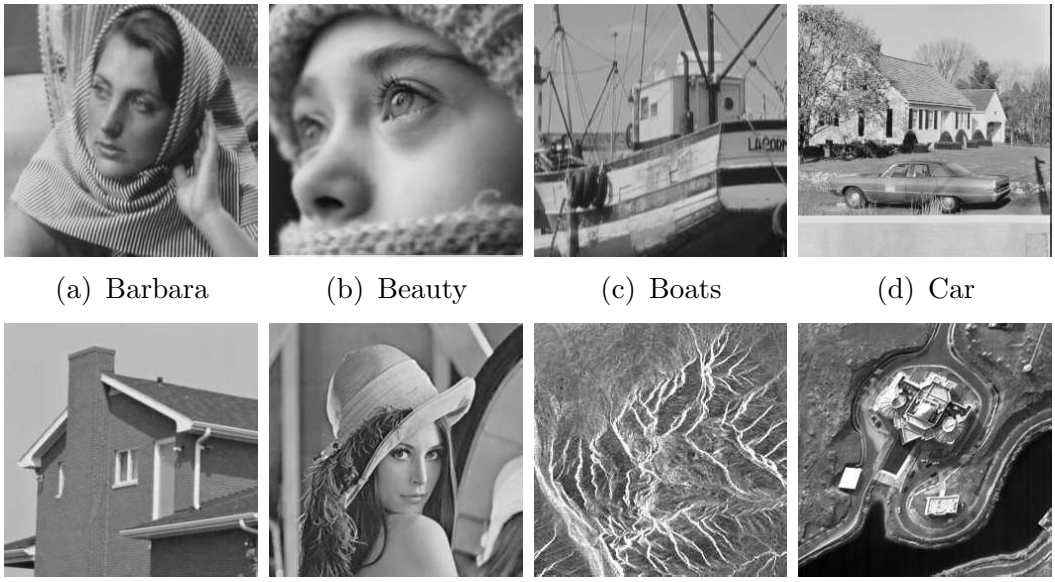
We simulated experiments for 4 different look numbers, $L = 1, 2, 4, 8$. For comparison, the speckled image $\boldsymbol{y}$ is restored by three methods: our algorithm (GSRS), the I-divergence model (IDIV) [56], and the block matching

Algorithm 3 GSR algorithm for speckle noise (GSRS)

Input: the noisy image $\boldsymbol{y}$, parameters P, s, L, c, T, γ, λ, μ.

Initialization: $t = 0$, $\boldsymbol{u}^{(0)} = \boldsymbol{y}$, $\boldsymbol{b}^{(0)} = \mathbf{0}$.
repeat
  Set $\boldsymbol{r} = \boldsymbol{u}^{(t)} + \boldsymbol{b}^{(t)}$, $\tau = \lambda K/\mu N$.
  Construct the group $\boldsymbol{r}_{\boldsymbol{G}_k}$ by obtaining similar patches using the distance (3.43) or (3.44).
  for each group $\boldsymbol{r}_{\boldsymbol{G}_k}$ **do**
    Construct the dictionary $\boldsymbol{D}_{\boldsymbol{G}_k}$ from (3.51) and (3.53).
    Compute $\hat{\boldsymbol{\alpha}}_{\boldsymbol{G}_k}$ by (3.55).
  end for
  Update $\boldsymbol{D}_{\boldsymbol{G}}^{(t+1)}$ by concatenating all $\boldsymbol{D}_{\boldsymbol{G}_k}$.
  Update $\boldsymbol{\alpha}_{\boldsymbol{G}}^{(t+1)}$ by concatenating all $\hat{\boldsymbol{\alpha}}_{\boldsymbol{G}_k}$.
  Compute $\boldsymbol{D}_{\boldsymbol{G}}^{(t+1)} \circ \boldsymbol{\alpha}_{\boldsymbol{G}}^{(t+1)}$ by (2.16).
  Update $\boldsymbol{u}^{(t+1)}$ by solving (3.58).
  Update $\boldsymbol{b}^{(t+1)}$ by (3.48).
  $t \leftarrow t + 1$.
until $t < T$

Output: the restored image $\boldsymbol{u}^{(T)}$.



(a) Barbara (b) Beauty (c) Boats (d) Car

(e) House (f) Lena (g) SAR1 (h) SAR2

Figure 3.14: Original test images.

and three dimensional algorithm for SAR denoising (BM3DS) [48]. The I-divergence model was implemented with PLAD algorithm [61], and the code for BM3DS was provided by the authors.

To measure the quality of the denoised images from different algorithms, we use peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [59], which are defined as

$$\mathrm{PSNR}(\boldsymbol{u}, \tilde{\boldsymbol{u}}) = 10 \log_{10} \frac{255^2}{\|\boldsymbol{u} - \tilde{\boldsymbol{u}}\|_F^2 / N},$$
$$\mathrm{SSIM}(\boldsymbol{u}, \tilde{\boldsymbol{u}}) = \frac{(2\mu_{\boldsymbol{u}}\mu_{\tilde{\boldsymbol{u}}} + c_1)(2\sigma_{\boldsymbol{u}\tilde{\boldsymbol{u}}} + c_2)}{(\mu_{\boldsymbol{u}}^2 + \mu_{\tilde{\boldsymbol{u}}}^2 + c_1)(\sigma_{\boldsymbol{u}}^2 + \sigma_{\tilde{\boldsymbol{u}}}^2 + c_2)},$$

where $\boldsymbol{u}$ is the original image and $\tilde{\boldsymbol{u}}$ is the restored image.

The parameters are as follows: The image size N is 256×256, and the

patch size P is 12×12. The stride between patches s is 4, and the number of total patches n is 64×64. The training window size $L \times L$ is 20×20, and the number of similar patches in a group c is set to 60. The maximum iteration number T is set to 10. The distance weight parameter γ in (3.44) is set to 3. The penalty parameter μ is set to 0.01. The regularization parameter λ is set $39/L$ for L look images, however, to produce fine results, different λ values are used at the first iteration: $\lambda^{(0)} = 10\lambda = 390/L$. Also we use a shifting technique [60], which shift values of the speckled image by 5 while processing. The initialization technique and the parameter selection are discussed in detail in Section 3.2.5.

Table 3.5: The PSNR and SSIM values of the images degraded by speckle noise ($L = 1$) and the restored images by different methods. The best values are marked in bold.

Image	PSNR				SSIM			
	Noisy	IDIV	BM3DS	GSRS	Noisy	IDIV	BM3DS	GSRS
Barbara	5.56	20.45	21.85	**22.33**	0.0654	0.4509	0.5441	**0.5953**
Beauty	5.42	22.53	23.55	**23.87**	0.0338	0.6092	0.6374	**0.6996**
Boats	5.36	20.80	22.13	**22.17**	0.0614	0.5079	0.5606	**0.5681**
Car	3.77	19.20	20.14	**20.21**	0.0663	0.4163	0.4452	**0.4757**
House	4.98	22.16	23.55	**23.84**	0.0329	0.5923	0.6028	**0.6303**
Lena	5.74	21.19	22.43	**22.69**	0.0597	0.5560	0.5870	**0.6235**
SAR1	5.42	17.12	17.87	**17.93**	0.0921	0.2855	**0.3558**	0.3503
SAR2	7.56	19.26	**20.24**	20.23	0.1350	0.4673	**0.5259**	0.5168
Average	5.48	20.34	21.47	**21.66**	0.0683	0.4857	0.5323	**0.5575**

In Figs. 3.15 to 3.18, we show the speckled images and the restored images using different models for 4 look numbers, $L = 1, 2, 4, 8$ respectively. As shown, the I-divergence model fails to remove speckle noise effectively, and has remaining noise which are represented as white dots in the restored

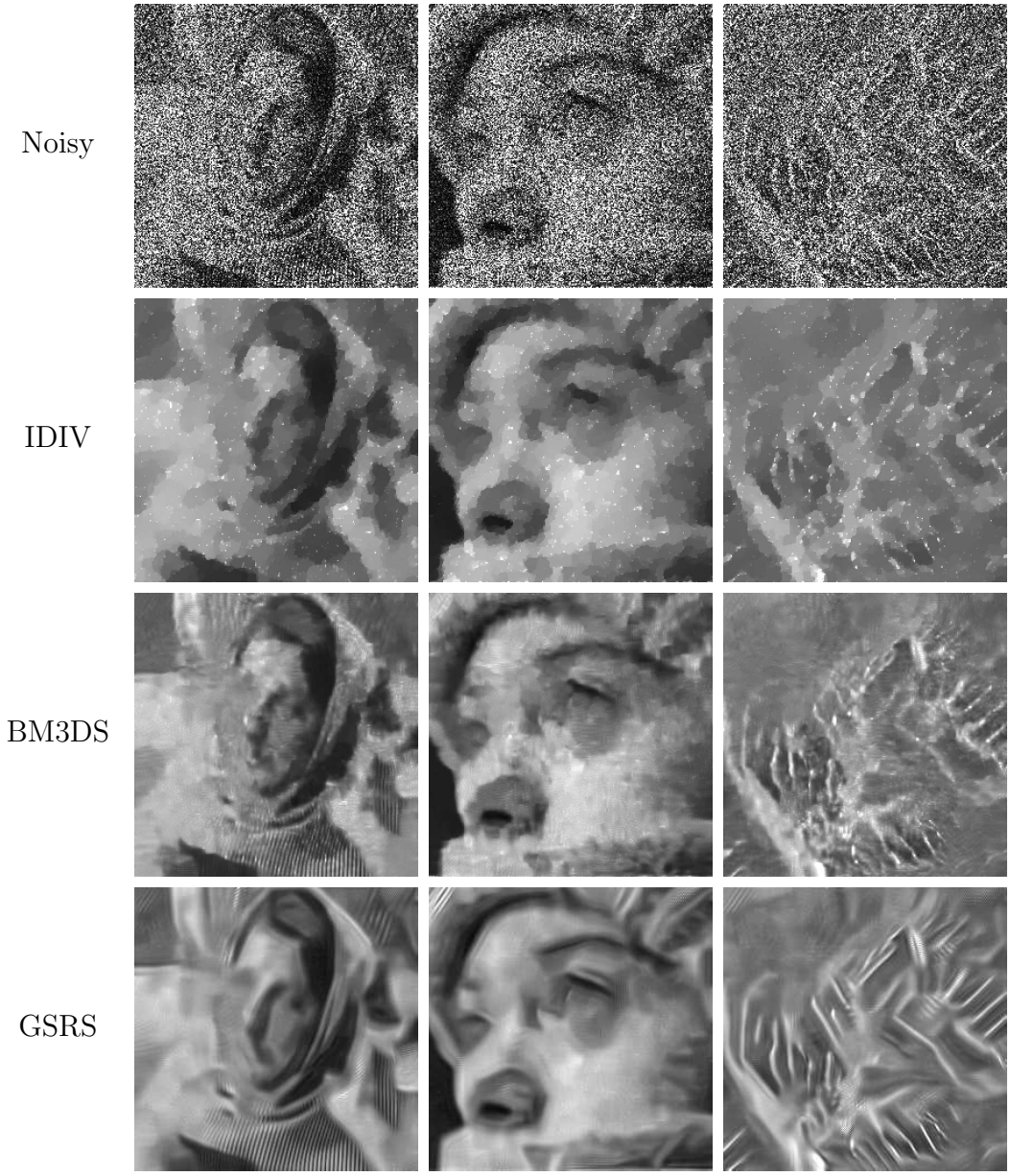


Figure 3.15: Comparison of different methods for restoring images degraded by speckle noise ($L = 1$).

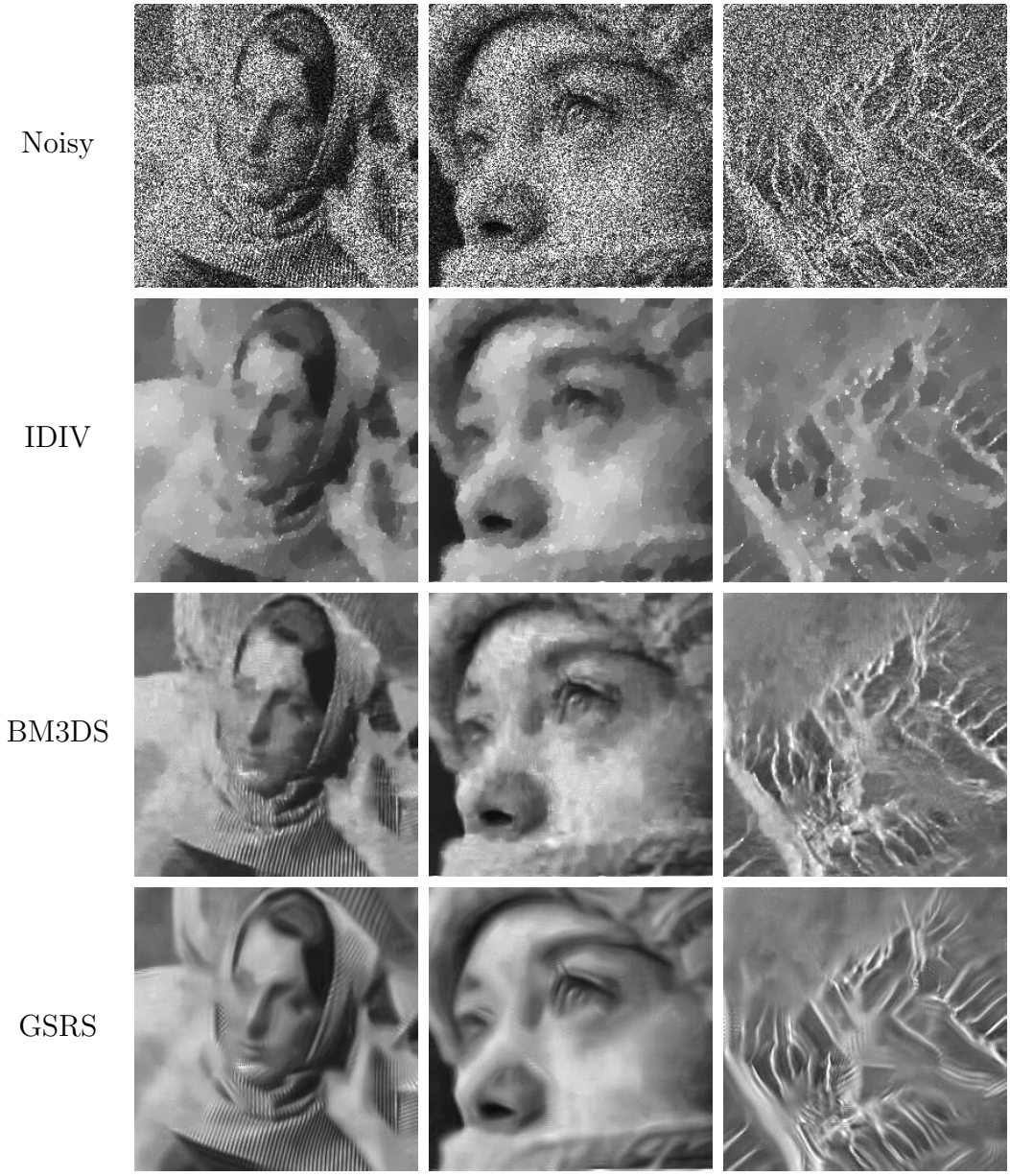


Figure 3.16: Comparison of different methods for restoring images degraded by speckle noise ($L = 2$).

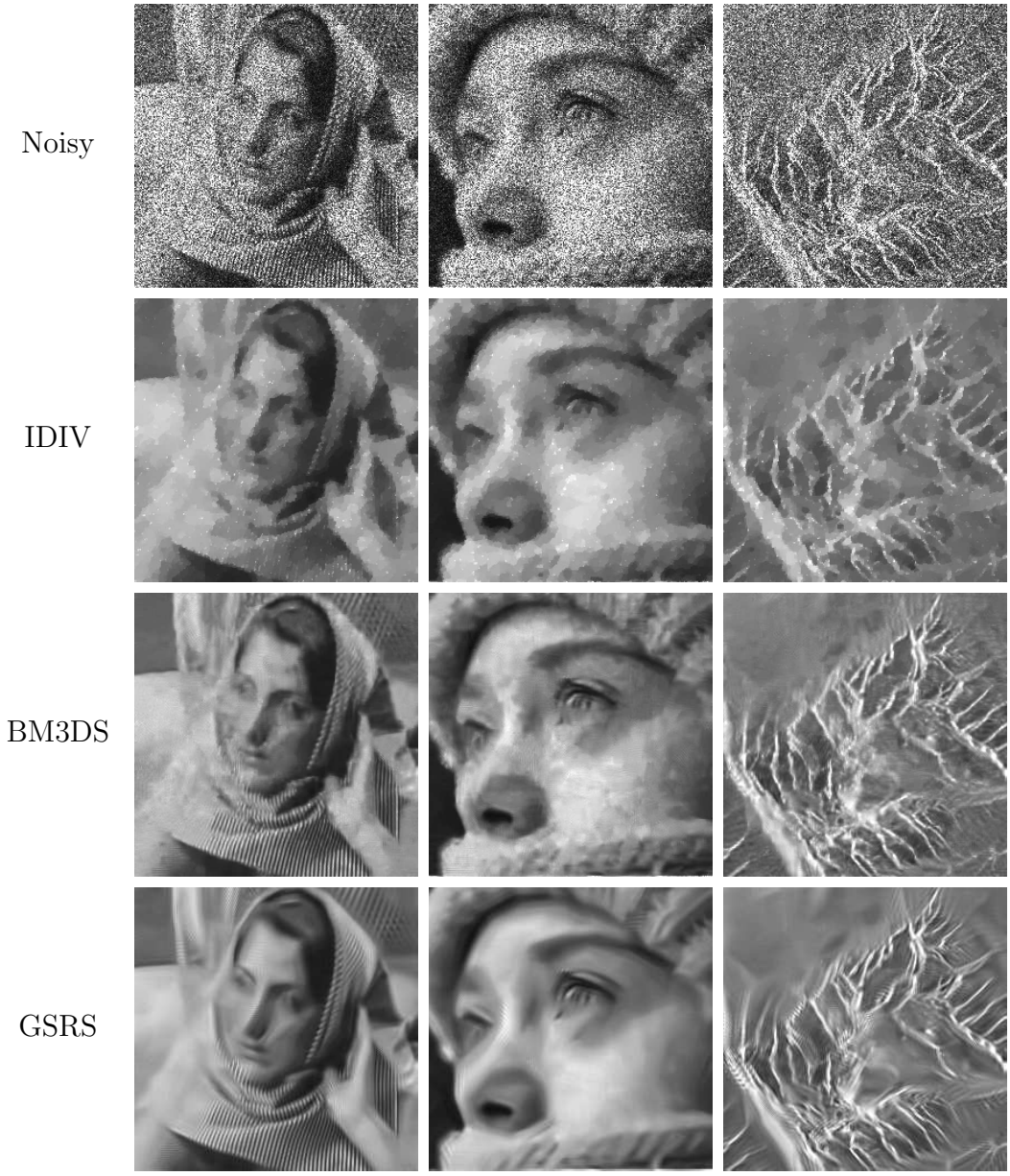


Figure 3.17: Comparison of different methods for restoring images degraded by speckle noise ($L = 4$).

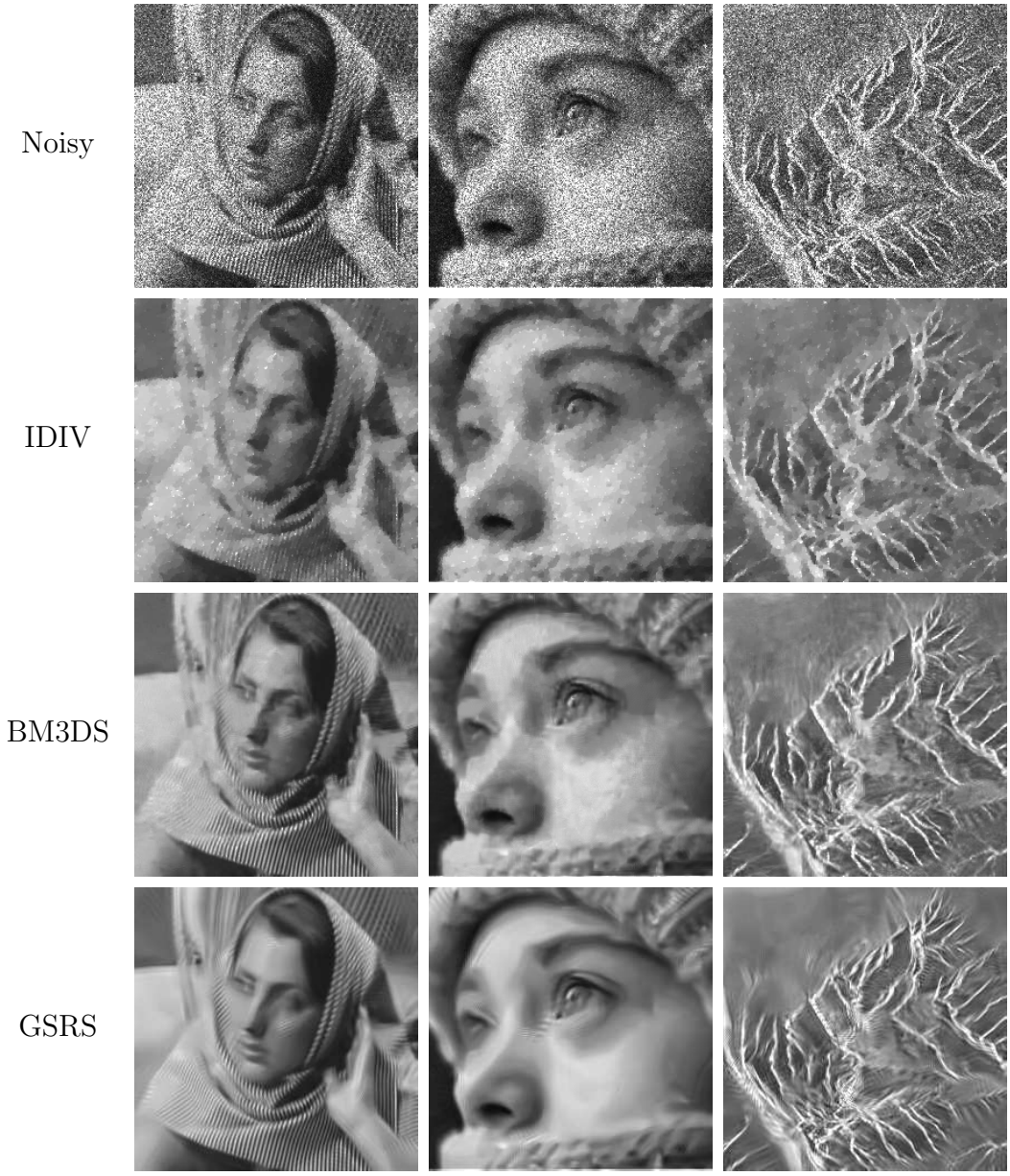


Figure 3.18: Comparison of different methods for restoring images degraded by speckle noise ($L = 8$).

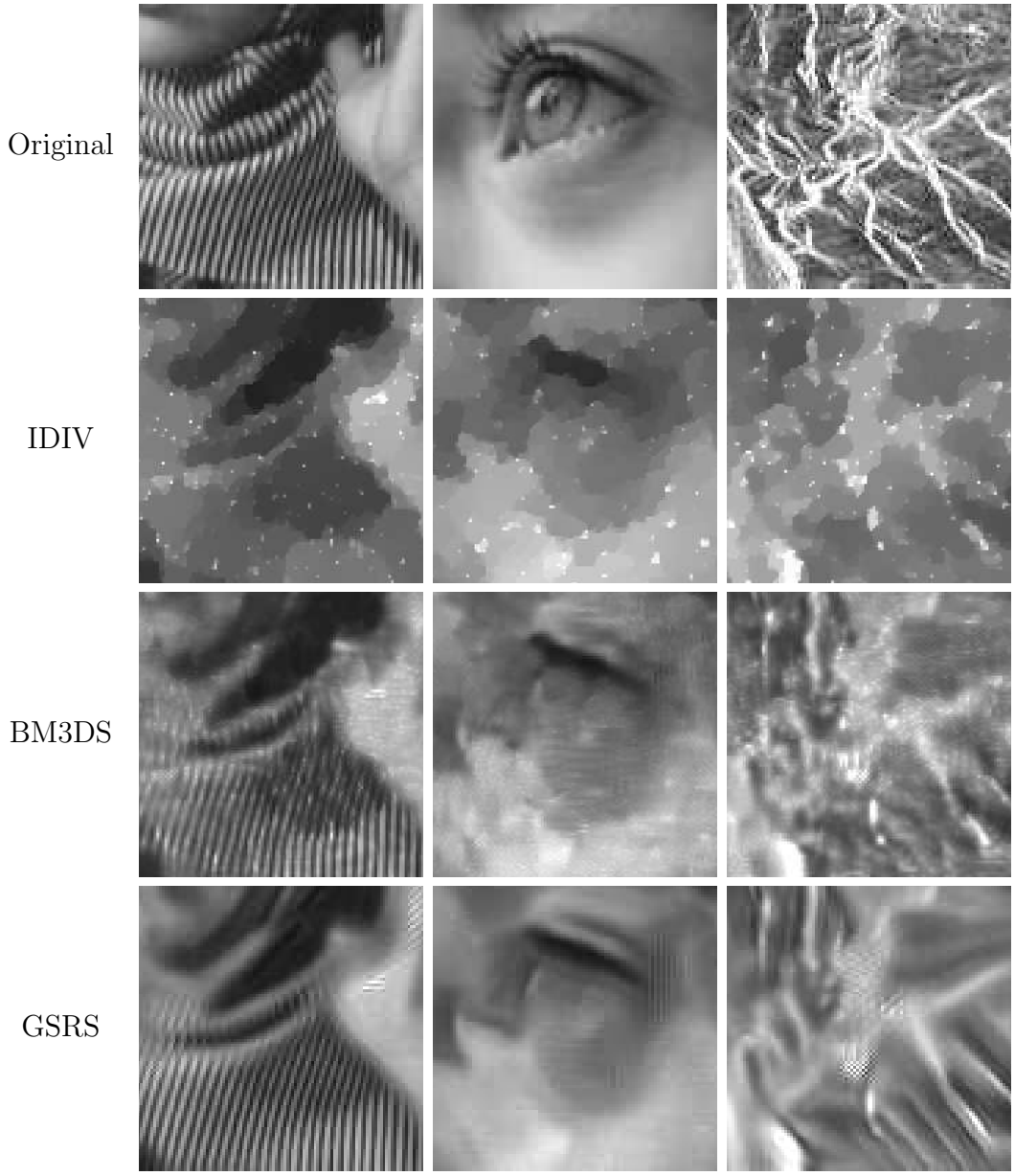


Figure 3.19: The zoomed-in version of the restored images in Fig. 3.15.

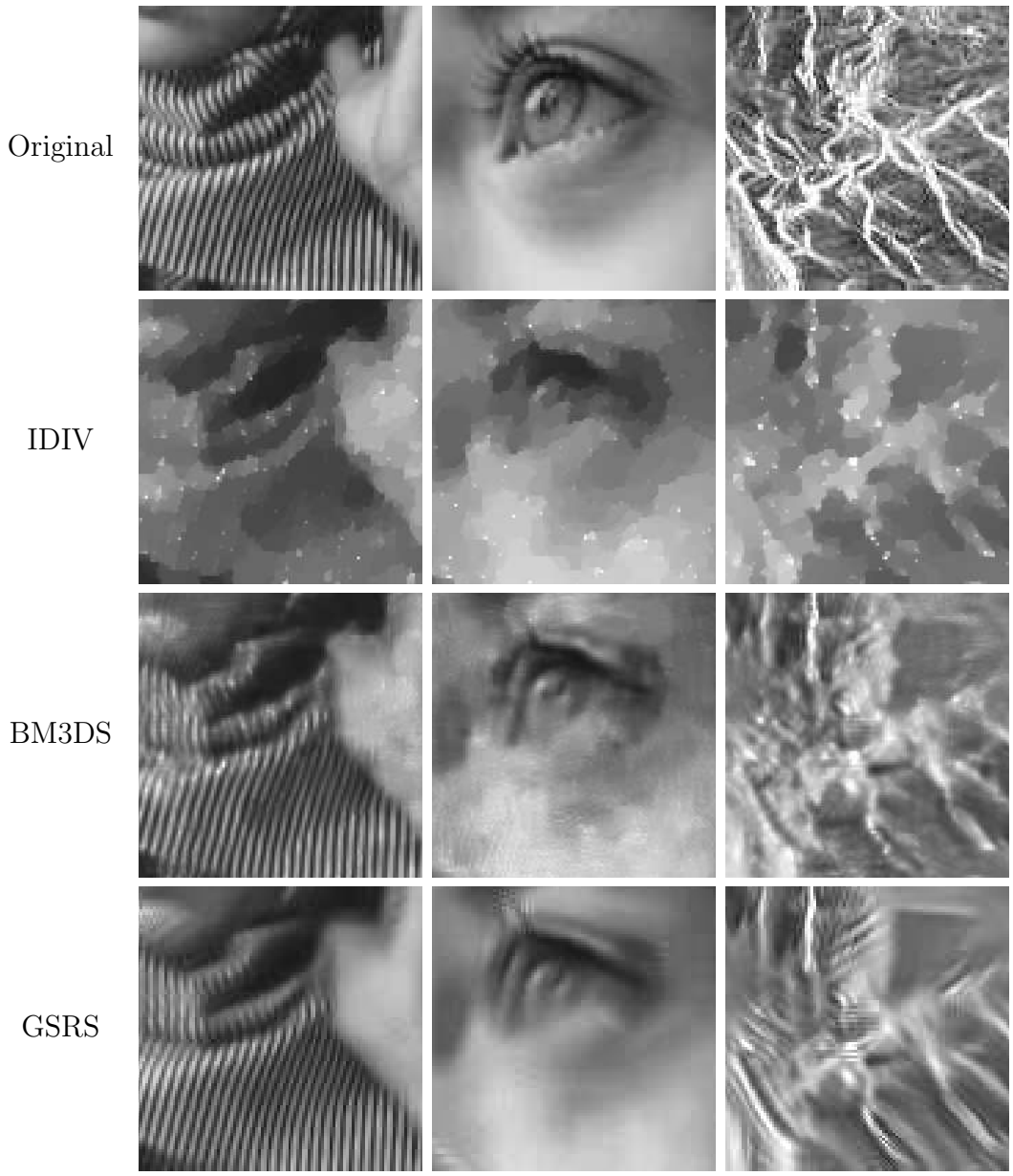


Figure 3.20: The zoomed-in version of the restored images in Fig. 3.16.

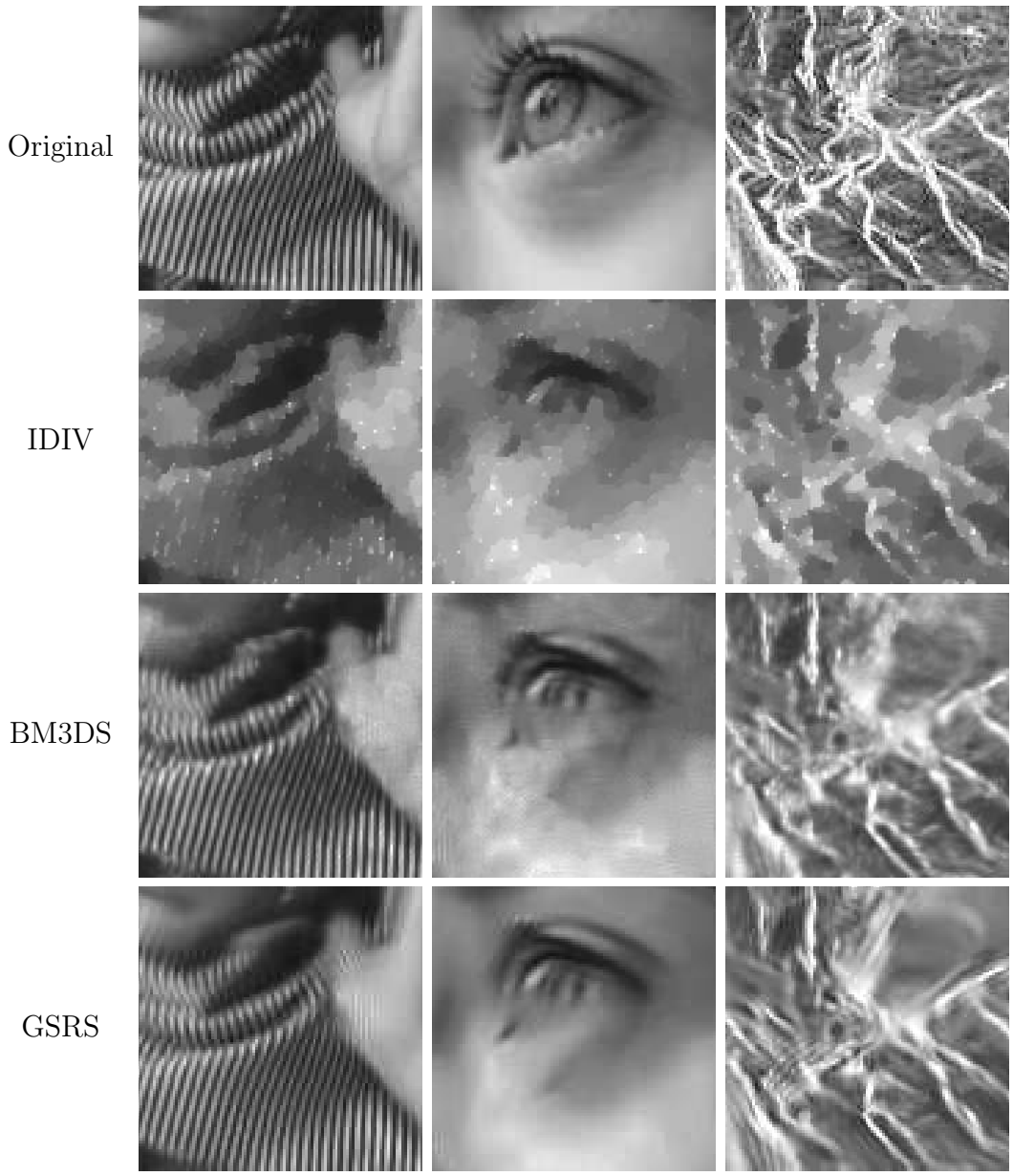


Figure 3.21: The zoomed-in version of the restored images in Fig. 3.17.

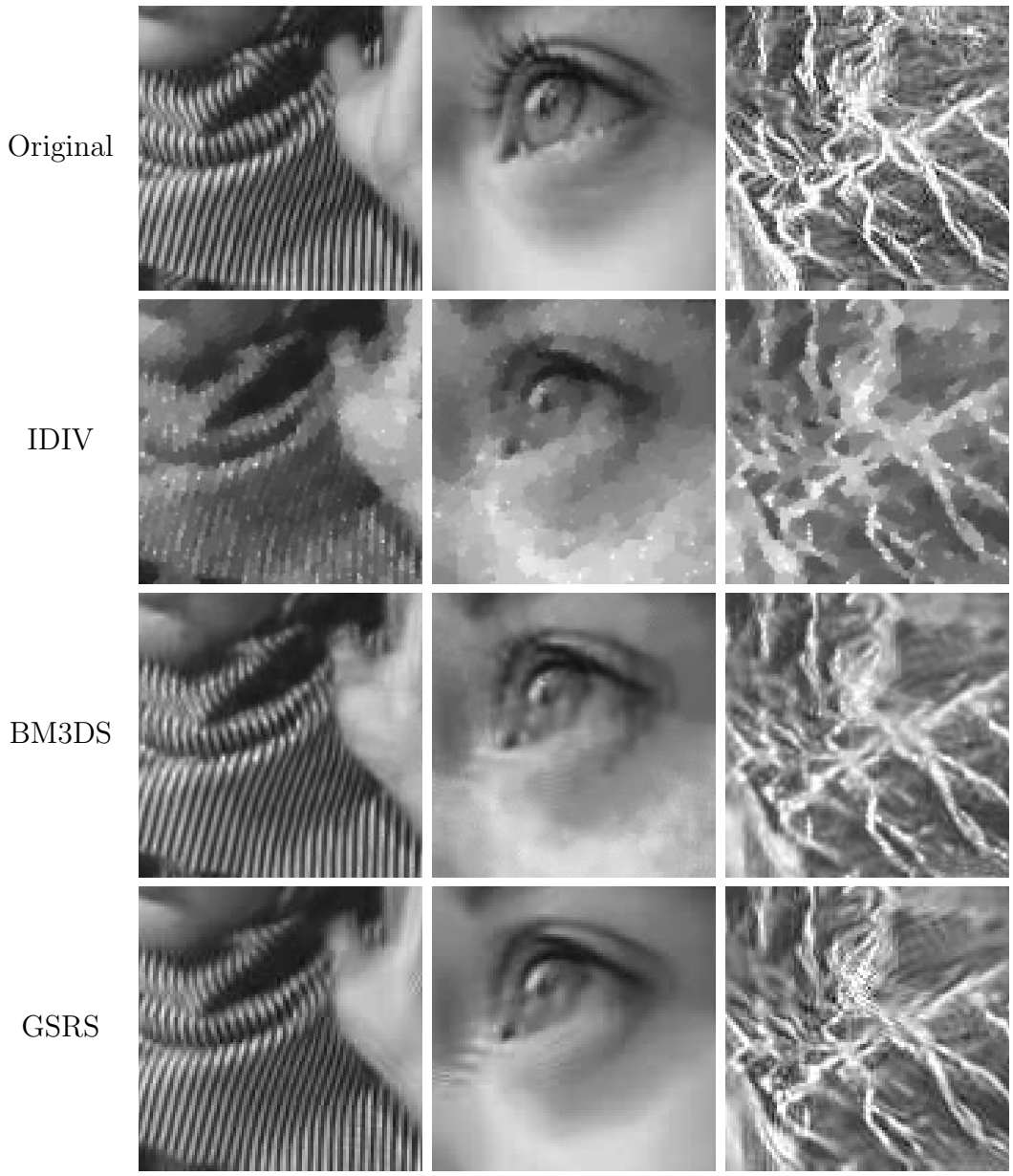


Figure 3.22: The zoomed-in version of the restored images in Fig. 3.18.

Table 3.6: The PSNR and SSIM values of the images degraded by speckle noise ($L = 2$) and the restored images by different methods. The best values are marked in bold.

Image	PSNR				SSIM			
	Noisy	IDIV	BM3DS	GSRS	Noisy	IDIV	BM3DS	GSRS
Barbara	8.45	21.68	23.61	**24.23**	0.1141	0.5130	0.6309	**0.6774**
Beauty	8.37	24.62	25.61	**26.22**	0.0601	0.6994	0.7236	**0.7850**
Boats	8.34	22.46	**23.93**	23.90	0.1052	0.5910	0.6365	**0.6469**
Car	6.73	20.42	**21.49**	21.45	0.1080	0.4960	0.5304	**0.5485**
House	7.92	24.20	25.59	**26.06**	0.0569	0.6909	0.6846	**0.7216**
Lena	8.65	23.01	24.28	**24.54**	0.1015	0.6491	0.6794	**0.7038**
SAR1	8.44	17.85	18.99	**19.01**	0.1579	0.3245	**0.4399**	0.4214
SAR2	10.44	20.23	**21.43**	21.25	0.2165	0.5146	**0.6008**	0.5904
Average	8.42	21.81	23.12	**23.33**	0.1150	0.5598	0.6158	**0.6369**

images. Although the BM3DS captures the speckle noise better, it does not recover the smooth region successfully. Compared with these methods, the proposed GSRS effectively reduces the speckle noise and restore smooth region well.

The visual comparison can be highlighted in zoomed-in versions, which are provided in Figs. 3.19 to 3.22. Compared with other methods, GSRS clearly restore the texture pattern of Barbara and smooth skin of Beauty. Also it recovers the branch parts of the SAR1 image as well as the BM3DS. To compare results quantitatively, we present the PSNR and SSIM values of the speckled images and restored images in Tables 3.5 to 3.8. Although our method doesn't yield the highest PSNR and SSIM values for all cases, it produces comparable results with BM3DS and slightly better values on average.

Table 3.7: The PSNR and SSIM values of the images degraded by speckle noise ($L = 4$) and the restored images by different methods. The best values are marked in bold.

Image	PSNR				SSIM			
	Noisy	IDIV	BM3DS	GSRS	Noisy	IDIV	BM3DS	GSRS
Barbara	11.48	22.50	25.11	**25.59**	0.1842	0.5570	0.7038	**0.7304**
Beauty	11.35	26.16	27.61	**28.19**	0.1004	0.7328	0.7858	**0.8325**
Boats	11.30	23.90	**25.53**	25.47	0.1698	0.6469	0.7072	**0.7153**
Car	9.75	21.88	**22.98**	22.62	0.1698	0.5569	0.6049	**0.6141**
House	10.92	25.41	27.64	**27.80**	0.0948	0.7095	0.7473	**0.7716**
Lena	11.74	24.48	26.19	**26.27**	0.1597	0.6973	0.7587	**0.7715**
SAR1	11.43	19.06	**20.42**	20.33	0.2580	0.4286	**0.5471**	0.5289
SAR2	13.43	21.52	**22.92**	22.77	0.3249	0.5913	**0.6861**	0.6733
Average	11.43	23.11	24.80	**24.88**	0.1827	0.6150	0.6926	**0.7047**

3.2.5 Discussion

Initialization Technique

Because the regularization term of the proposed GSRS (3.42) is nonconvex, the solution may depend on the initial condition. To show this, experiments are designed to denoise the Car image degraded by speckle noise with $L = 1$, $L = 4$, and $L = 8$. The GSRS algorithm is applied with two different initial conditions: (I): λ is fixed. (II): λ is fixed except at $\lambda^{(0)} = 10\lambda$. In all cases, the parameter values are the same as given in Section 3.2.4.

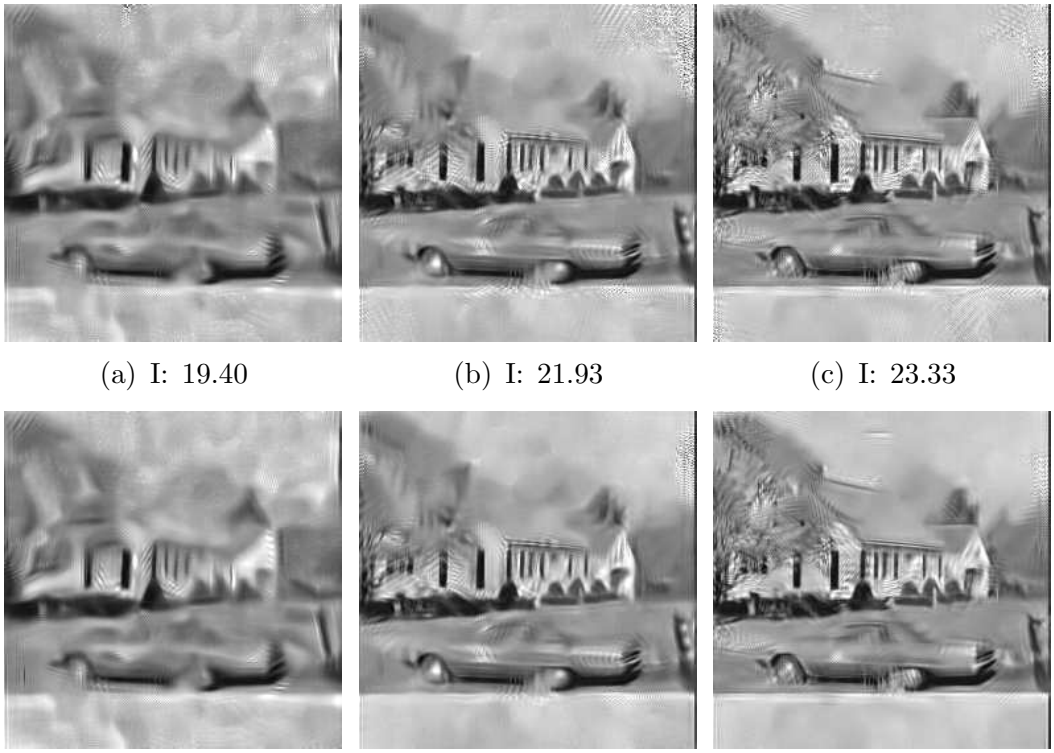
The results are shown in Fig. 3.23. In the first row (I), there are some artifacts which is obviously seen in the upper right corner. In the second row (II), the unwanted artifacts are disappeared compared with (I). It can be seen that the initialization technique of selecting a high $\lambda^{(0)}$ yields better results and higher PSNR values.

Table 3.8: The PSNR and SSIM values of the images degraded by speckle noise ($L = 8$) and the restored images by different methods. The best values are marked in bold.

Image	PSNR				SSIM			
	Noisy	IDIV	BM3DS	GSRS	Noisy	IDIV	BM3DS	GSRS
Barbara	14.51	23.46	26.68	**27.51**	0.2792	0.6235	0.7825	**0.8207**
Beauty	14.37	27.42	29.45	**29.88**	0.1626	0.7551	0.8421	**0.8700**
Boats	14.31	25.15	**27.24**	27.16	0.2559	0.6913	0.7735	**0.7778**
Car	12.75	23.06	**24.30**	23.98	0.2470	0.6017	0.6806	**0.6881**
House	13.92	26.81	**29.63**	29.59	0.1505	0.7086	0.7977	**0.8081**
Lena	14.72	25.89	**27.97**	27.96	0.2375	0.7291	0.8176	**0.8214**
SAR1	14.44	20.47	**21.74**	21.64	0.3811	0.5501	**0.6364**	0.6221
SAR2	16.43	23.06	**24.46**	24.30	0.4505	0.6929	**0.7594**	0.7485
Average	14.43	24.42	26.43	**26.50**	0.2705	0.6690	0.7612	**0.7696**

Analysis of Parameters

As shown in Section 3.2.4, there are many parameters in the GSRS algorithm which can be tuned. We tested to tune the parameters, for example, the patch size $P = [6 \times 6, 7 \times 7, 8 \times 8, 10 \times 10, 12 \times 12]$, the stride between patches $s = [2, 4]$, and the distance weight parameter $\gamma = [1, 2, 3, 4, 5]$. We don't discuss the tuning results of these parameters, but we provide results for two crucial parameters: the regularization parameter λ and the penalty parameter μ. In Fig. 3.24, the PSNR values of the results from varying λ and μ are plotted for the experiments of despeckling images with speckle noise $L = 1$ and $L = 8$. Fig. 3.24 shows that λ and μ exhibit consistent behaviors, which implies the stable performance of the GSRS algorithm.



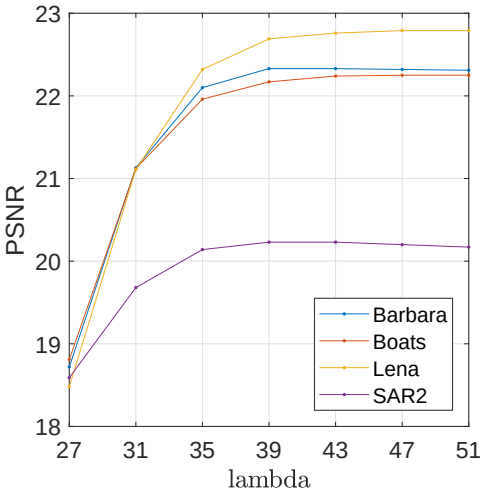
(a) I: 19.40 (b) I: 21.93 (c) I: 23.33

(d) II: 20.21 (e) II: 22.62 (f) II: 23.98

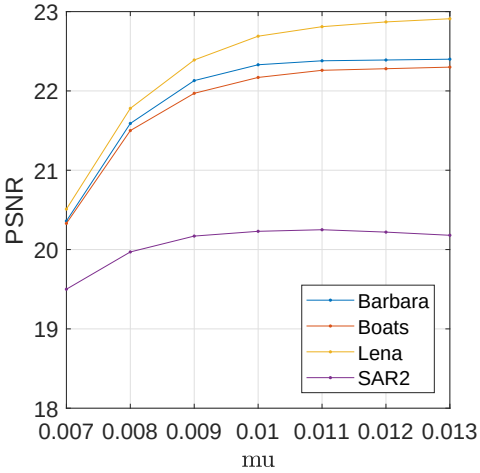
Figure 3.23: Results of GSRS from different initial conditions for the images degraded by speckle noise (From left to right: $L = 1$, $L = 4$, and $L = 8$.). The PSNR values are shown below each image. (I) $\boldsymbol{u}^{(0)} = \boldsymbol{y}$, λ fixed; (II) $\boldsymbol{u}^{(0)} = \boldsymbol{y}$, $\lambda^{(0)} = 10\lambda$.

Analysis of Convergence

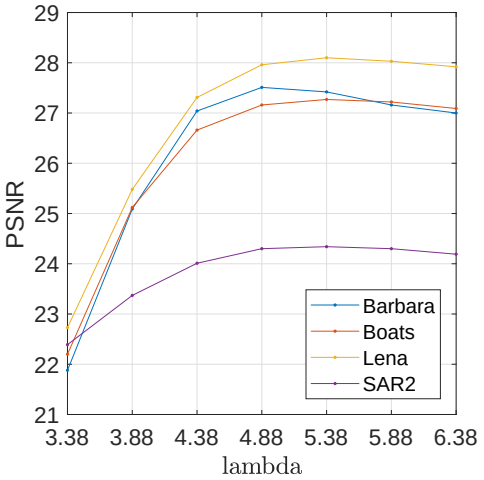
Because GSRS is a nonconvex model due to its ℓ_0 regularization term, it does not have a global optimizer. Hence the convergence of the algorithm to a global minimum cannot be guaranteed theoretically. To compensate for this, we present the convergence of the algorithm numerically. In Fig. 3.25, the PSNR values versus the iteration number are shown for different images in the despeckling experiments of Section 3.2.4. As shown, the PSNR values



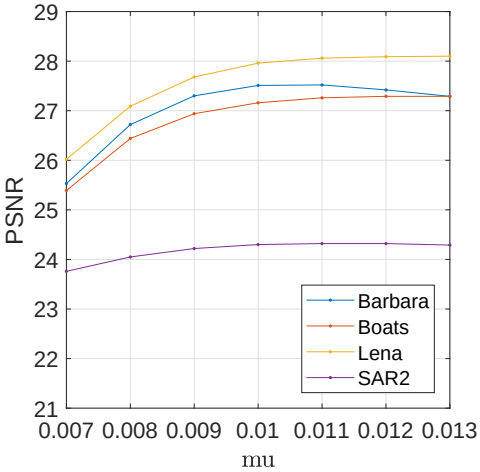
(a) speckle noise $L = 1$



(b) speckle noise $L = 1$

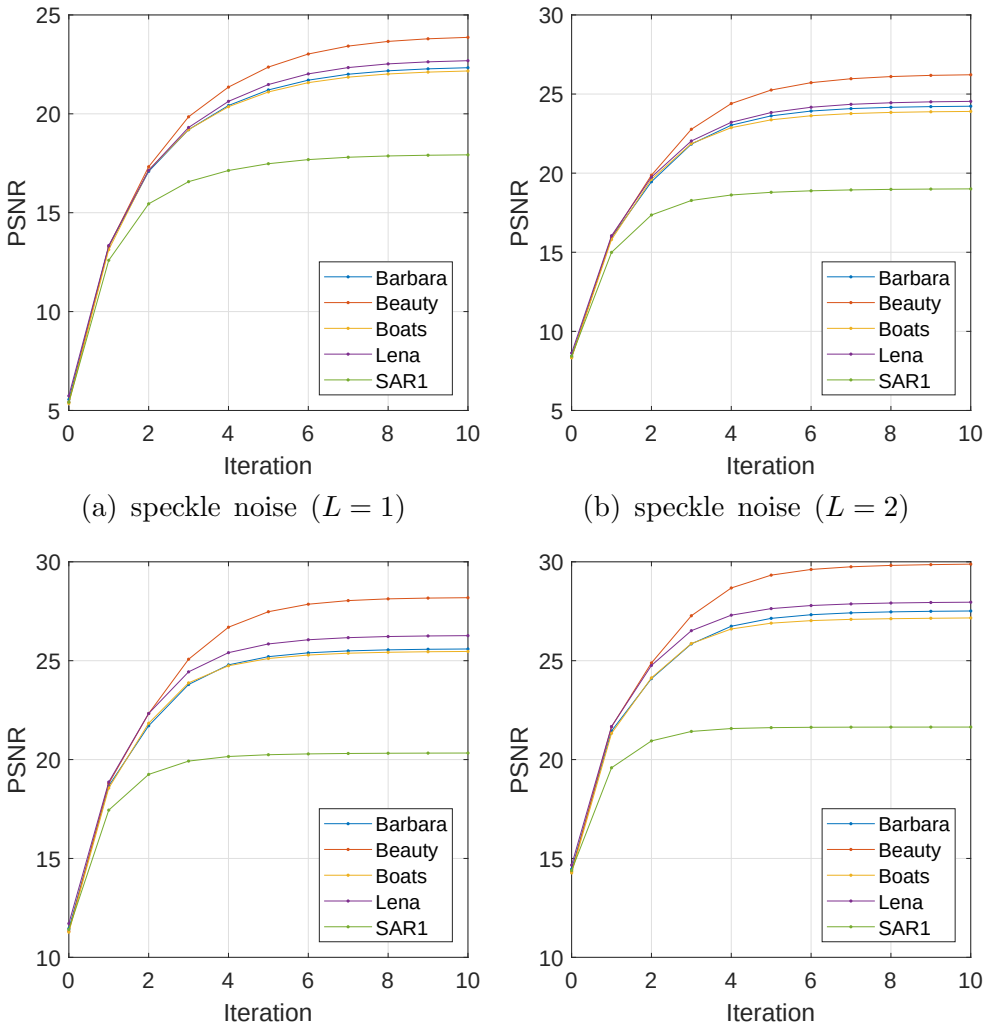


(c) speckle noise $L = 8$



(d) speckle noise $L = 8$

Figure 3.24: Plots of the PSNR values versus the parameter values of GSRS for 4 test images in the two experimental simulations. (a) and (b): PSNR values versus λ and μ, respectively, for images with speckle noise $L = 1$; (c) and (d): PSNR values versus λ and μ, respectively, for images with speckle noise $L = 8$.



(a) speckle noise ($L = 1$) (b) speckle noise ($L = 2$)

(c) speckle noise ($L = 4$) (d) speckle noise ($L = 8$)

Figure 3.25: Plots of the PSNR values versus the iteration number of GSRS for 5 test images in experiments for restoring images degraded by speckle noise with look (a) $L = 1$; (b) $L = 2$; (c) $L = 4$; (d) $L = 8$.

are monotone increasing and converge asymptotically, which demonstrate the numerical convergence of the algorithm.

Chapter 4

Conclusion

We herein propose two nonconvex variational models for restoring images degraded by non-Gaussian noise. The first model is designed to deblur and denoise images corrupted by Cauchy noise, and the second model is designed to denoise images in the presence of speckle noise. We combined a regularizer based on the GSR and a fidelity term suitable for each noise to develop the proposed models. The ADMM was utilized for the efficient implementation of the proposed algorithms. To mitigate the instability arising from the nonconvexity of our models, a novel initialization technique was introduced to obtain the desired solution. Numerical experiments demonstrated the superior performance of our algorithm compared with other methods, in terms of quantitative and qualitative measures, as well as the stable convergence.

Bibliography

[1] Alin Achim and Ercan E Kuruoglu. Image denoising using bivariate α-stable distributions in the complex wavelet domain. *IEEE Signal Processing Letters*, 12(1):17–20, 2004.

[2] Alin Achim, Anastasios Bezerianos, and Panagiotis Tsakalides. Novel bayesian multiscale method for speckle removal in medical ultrasound images. *IEEE transactions on medical imaging*, 20(8):772–783, 2001.

[3] Manya V Afonso, José M Bioucas-Dias, and Mário AT Figueiredo. Fast image recovery using variable splitting and constrained optimization. *IEEE transactions on image processing*, 19(9):2345–2356, 2010.

[4] Gilles Aubert and Jean-Francois Aujol. A variational approach to removing multiplicative noise. *SIAM journal on applied mathematics*, 68 (4):925–946, 2008.

[5] Saurav Basu, Thomas Fletcher, and Ross Whitaker. Rician noise removal in diffusion tensor mri. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 117–125. Springer, 2006.

[6] Md Imamul Hassan Bhuiyan, M Omair Ahmad, and MNS Swamy. Spatially adaptive wavelet-based method using the cauchy prior for denois-

ing the sar images. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(4):500–507, 2007.

[7] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

[8] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 60–65. IEEE, 2005.

[9] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation*, 4(2):490–530, 2005.

[10] Jian-Feng Cai, Stanley Osher, and Zuowei Shen. Linearized bregman iterations for frame-based image deblurring. *SIAM Journal on Imaging Sciences*, 2(1):226–252, 2009.

[11] Emmanuel J Candès and Carlos Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on pure and applied Mathematics*, 67(6):906–956, 2014.

[12] Tony F Chan and Jianhong Shen. Variational image inpainting. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 58(5):579–619, 2005.

[13] S Grace Chang, Bin Yu, and Martin Vetterli. Adaptive wavelet thresholding for image denoising and compression. *IEEE transactions on image processing*, 9(9):1532–1546, 2000.

[14] Dai-Qiang Chen and Li-Zhi Cheng. Spatially adapted total variation model to remove multiplicative noise. *IEEE Transactions on Image Processing*, 21(4):1650–1662, 2011.

[15] Davide Cozzolino, Sara Parrilli, Giuseppe Scarpa, Giovanni Poggi, and Luisa Verdoliva. Fast adaptive nonlocal sar despeckling. *IEEE Geoscience and Remote Sensing Letters*, 11(2):524–528, 2013.

[16] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, Aug 2007. ISSN 1057-7149. doi: 10.1109/TIP.2007.901238.

[17] Philip J Davis. *Circulant matrices*. Wiley, 1979.

[18] Charles-Alban Deledalle, Loïc Denis, and Florence Tupin. Iterative weighted maximum likelihood denoising with probabilistic patch-based weights. *IEEE Transactions on Image Processing*, 18(12):2661–2672, 2009.

[19] Charles-Alban Deledalle, Florence Tupin, and Loïc Denis. Patch similarity under non gaussian noise. In *2011 18th IEEE International Conference on Image Processing*, pages 1845–1848. IEEE, 2011.

[20] Charles-Alban Deledalle, Loïc Denis, and Florence Tupin. How to compare noisy patches? patch similarity beyond gaussian noise. *International journal of computer vision*, 99(1):86–102, 2012.

[21] Meng Ding, Ting-Zhu Huang, Si Wang, Jin-Jin Mei, and Xi-Le Zhao. Total variation with overlapping group sparsity for deblurring images under cauchy noise. *Applied Mathematics and Computation*, 341:128–147, 2019.

[22] Weisheng Dong, Lei Zhang, Guangming Shi, and Xiaolin Wu. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. *IEEE Transactions on Image Processing*, 20(7): 1838–1857, 2011.

[23] Weisheng Dong, Lei Zhang, Guangming Shi, and Xin Li. Nonlocally centralized sparse representation for image restoration. *IEEE transactions on Image Processing*, 22(4):1620–1630, 2012.

[24] Yiqiu Dong and Tieyong Zeng. A convex variational model for restoring blurred images with multiplicative noise. *SIAM Journal on Imaging Sciences*, 6(3):1598–1625, 2013.

[25] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image processing*, 15(12):3736–3745, 2006.

[26] Jing Fang, Shuaiqi Liu, Yang Xiao, and Hailiang Li. Sar image denoising based on texture strength and weighted nuclear norm minimization. *Journal of Systems Engineering and Electronics*, 27(4):807–814, 2016.

[27] Wensen Feng, Hong Lei, and Yang Gao. Speckle reduction via higher order total variation approach. *IEEE Transactions on Image Processing*, 23(4):1831–1843, 2014.

[28] Alessandro Foi. Noise estimation and removal in mr imaging: The variance-stabilization approach. In *2011 IEEE International symposium on biomedical imaging: from nano to macro*, pages 1809–1814. IEEE, 2011.

[29] Simon Foucart. Hard thresholding pursuit: an algorithm for compressive sensing. *SIAM Journal on Numerical Analysis*, 49(6):2543–2563, 2011.

[30] B Roy Frieden. A new restoring algorithm for the preferential enhancement of edge gradients. *JOSA*, 66(3):280–283, 1976.

[31] Geoffrey Grimmett and Dominic Welsh. *Probability: an introduction*. Oxford University Press, 2014.

[32] Dongdong Guan, Deliang Xiang, Xiaoan Tang, and Gangyao Kuang. Sar image despeckling based on nonlocal low-rank regularization. *IEEE Transactions on Geoscience and Remote Sensing*, 57(6):3472–3489, 2018.

[33] Milan Holickỳ. *Introduction to probability and statistics for engineers*, pages 68–69. Springer Science & Business Media, 2013.

[34] Yu-Mei Huang, Lionel Moisan, Michael K Ng, and Tieyong Zeng. Multiplicative noise removal via a learned dictionary. *IEEE Transactions on Image Processing*, 21(11):4534–4543, 2012.

[35] Moshe Idan and Jason L Speyer. Cauchy estimation for linear scalar systems. *IEEE transactions on automatic control*, 55(6):1329–1342, 2010.

[36] Ercan E Kuruoglu, William J Fitzgerald, and Peter JW Rayner. Near optimal detection of signals in impulsive noise modeled with a symmetric/spl alpha/-stable distribution. *IEEE Communications Letters*, 2(10): 282–284, 1998.

[37] Friederike Laus, Fabien Pierre, and Gabriele Steidl. Nonlocal myriad filters for cauchy noise removal. *Journal of Mathematical Imaging and Vision*, 60(8):1324–1354, 2018.

[38] Shujun Liu, Guoqing Wu, Xinzheng Zhang, Kui Zhang, Pin Wang, and Yongming Li. Sar despeckling via classification-based nonlocal and local sparse representation. *Neurocomputing*, 219:174–185, 2017.

[39] Su Liu, Gong Zhang, and Yeo Tat Soon. An over-complete dictionary design based on gsr for sar image despeckling. *IEEE Geoscience and Remote Sensing Letters*, 14(12):2230–2234, 2017.

[40] Su Liu, Gong Zhang, and Wenbo Liu. Group sparse representation based dictionary learning for sar image despeckling. *IEEE Access*, 7: 30809–30817, 2019.

[41] Artur Loza, David Bull, Nishan Canagarajah, and Alin Achim. Non-gaussian model-based fusion of noisy images in the wavelet domain. *Computer Vision and Image Understanding*, 114(1):54–65, 2010.

[42] Jian Lu, Lixin Shen, Chen Xu, and Yuesheng Xu. Multiplicative noise removal in imaging: An exp-model and its fixed-point proximity algorithm. *Applied and Computational Harmonic Analysis*, 41(2):518–539, 2016.

[43] Jin-Jin Mei, Yiqiu Dong, Ting-Zhu Huang, and Wotao Yin. Cauchy noise removal by nonconvex admm with convergence guarantees. *Journal of Scientific Computing*, 74(2):743–766, 2018.

[44] Hanwool Na, Myeongmin Kang, Miyoun Jung, and Myungjoo Kang. An exp model with spatially adaptive regularization parameters for multiplicative noise removal. *Journal of Scientific Computing*, 75(1):478–509, 2018.

[45] Michael K Ng, Raymond H Chan, and Wun-Cheung Tang. A fast algorithm for deblurring models with neumann boundary conditions. *SIAM Journal on Scientific Computing*, 21(3):851–866, 1999.

[46] João P Oliveira, José M Bioucas-Dias, and Mário AT Figueiredo. Adaptive total variation image deblurring: a majorization–minimization approach. *Signal processing*, 89(9):1683–1693, 2009.

[47] Stanley Osher, Zuoqiang Shi, and Wei Zhu. Low dimensional manifold model for image processing. *SIAM Journal on Imaging Sciences*, 10(4): 1669–1690, 2017.

[48] Sara Parrilli, Mariana Poderico, Cesario Vincenzo Angelino, and Luisa Verdoliva. A nonlocal sar image denoising algorithm based on llmmse wavelet shrinkage. *IEEE Transactions on Geoscience and Remote Sensing*, 50(2):606–616, 2011.

[49] Isabel Rodrigues, Joao Sanches, and Jose Bioucas-Dias. Denoising of medical images corrupted by poisson noise. In *2008 15th IEEE International Conference on Image Processing*, pages 1756–1759. IEEE, 2008.

[50] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

[51] Hamidreza Sadreazami, M Omair Ahmad, and MNS Swamy. Ultrasound image despeckling in the contourlet domain using the cauchy prior. In *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 33–36. IEEE, 2016.

[52] Federica Sciacchitano, Yiqiu Dong, and Tieyong Zeng. Variational approach for restoring blurred images with cauchy noise. *SIAM Journal on Imaging Sciences*, 8(3):1894–1922, 2015.

[53] Mu-Ga Shama, Ting-Zhu Huang, Jun Liu, and Si Wang. A convex total generalized variation regularized model for multiplicative noise and blur removal. *Applied Mathematics and Computation*, 276:109–121, 2016.

[54] Jianing Shi and Stanley Osher. A nonlinear inverse scale space method for a convex multiplicative noise model. *SIAM Journal on imaging sciences*, 1(3):294–321, 2008.

[55] Wuzhen Shi, Congcong Chen, Feng Jiang, Debin Zhao, and Weizheng Shen. Group-based sparse representation for low lighting image enhancement. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 4082–4086. IEEE, 2016.

[56] Gabriele Steidl and Tanja Teuber. Removing multiplicative noise by douglas-rachford splitting methods. *Journal of Mathematical Imaging and Vision*, 36(2):168–184, 2010.

[57] Fawwaz Ulaby, M Craig Dobson, and José Luis Álvarez-Pérez. *Handbook of radar scattering statistics for terrain*. Artech House, 2019.

[58] Yu Wang, Wotao Yin, and Jinshan Zeng. Global convergence of admm in nonconvex nonsmooth optimization. *Journal of Scientific Computing*, 78(1):29–63, 2019.

[59] Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, et al. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[60] Hyenkyun Woo and Sangwoon Yun. Alternating minimization algorithm for speckle reduction with a shifting technique. *IEEE Transactions on Image Processing*, 21(4):1701–1714, 2011.

[61] Hyenkyun Woo and Sangwoon Yun. Proximal linearized alternating direction method for multiplicative denoising. *SIAM Journal on Scientific Computing*, 35(2):B336–B358, 2013.

[62] Hua Xie, Leland E Pierce, and Fawwaz T Ulaby. Statistical properties of logarithmically transformed speckle. *IEEE Transactions on Geoscience and Remote Sensing*, 40(3):721–727, 2002.

[63] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma. Image super-resolution via sparse representation. *IEEE transactions on image processing*, 19(11):2861–2873, 2010.

[64] Jing-Hua Yang, Xi-Le Zhao, Jin-Jin Mei, Si Wang, Tian-Hui Ma, and Ting-Zhu Huang. Total variation and high-order total variation adaptive model for restoring blurred images with cauchy noise. *Computers & Mathematics with Applications*, 77(5):1255–1272, 2019.

[65] Junfeng Yang and Yin Zhang. Alternating direction algorithms for ℓ_1-problems in compressive sensing. *SIAM journal on scientific computing*, 33(1):250–278, 2011.

[66] Jian Zhang, Debin Zhao, and Wen Gao. Group-based sparse representation for image restoration. *IEEE Transactions on Image Processing*, 23(8):3336–3351, 2014.

[67] Jian Zhang, Siwei Ma, Yongbing Zhang, and Wen Gao. Image deblocking using group-based sparse representation and quantization constraint prior. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 306–310. IEEE, 2015.

[68] Daniel Zwillinger. *CRC standard mathematical tables and formulae*. Chapman and Hall/CRC, 2002.

국문초록

영상 복원 문제에서, 영상의 비국지적인 정보를 활용하는 최근의 다양한 접근 방식은 국지적인 특성을 활용하는 기존 방법과 비교하여 크게 개선되었다. 따라서, 우리는 비가우시안 잡음 영상을 복원하기 위해 영상 그룹 희소 표현에 기반한 두 가지 변분법적 모델을 제안한다. 제안된 모델은 각각 코시 잡음과 스펙클 잡음 영상을 복원하도록 설계되었다. 효율적이고 안정적인 성능을 달성하기 위해, 교대 방향 승수법과 새로운 초기화 기술이 사용된다. 실험 결과는 제안된 방법이 시각적인 인식과 수치적인 지표 모두에서 다른 방법보다 우수함을 나타낸다.

주요어휘: 영상 복원, 영상 잡음 제거, 영상 블러 제거, 스펙클 잡음, 코시 잡음, 그룹 희소 표현, 교대 방향 승수법
학번: 2014-21191