

Adaptive Variational Method for Restoring Color Images with High Density Impulse Noise

Jun Liu · Haiyang Huang · Zhongdan Huan · Haili Zhang

the date of receipt and acceptance should be inserted later

Abstract In this paper, a new variational framework of restoring color images with impulse noise is presented. The novelty of this work is the introduction of an adaptively weighting data-fidelity term in the cost functional. The fidelity term is derived from statistical methods and contains two weighting functions as well as some statistical control parameters of noise. This method is based on the fact that impulse noise can be approximated as an additive noise with probability density function (PDF) being the finite mixture model. A Bayesian framework is then formulated in which likelihood functions are given by the mixture model. Inspired by the expectation-maximization (EM) algorithm, we present two models with variational framework in this study. The superiority of the proposed models is that: the weighting functions can effectively detect the noise in the image; with the noise information, the proposed algorithm can automatically balance the regularity of the restored image and the fidelity term by updating the weighting functions and the control parameters. These two steps ensure that one can obtain a good restoration even though the degraded color image is contaminated by impulse noise with large ration (90% or more). In addition, the numerical implementation of this algorithm

is very fast by using a split algorithm. Some numerical experimental results and comparisons with other methods are provided to show the significant effectiveness of our approach.

Keywords Color image · Deblurring and denoising · Impulse noise · Variational regularization methods · EM algorithm · Split Bregman method

1 Introduction

Image restoration is an old and fundamental problem in image processing, but it continues to attract the attention of many researchers. The image degradation process is often considered as a shift-invariant model. For color images, this process could be mathematically modeled by

$$\mathbf{g} = k * \mathbf{f} + \mathbf{n},$$

where $\mathbf{g} : \Omega \subset \mathbb{R}^2 \rightarrow [0, 1]^3$ is a vector-valued function which represents the degraded color image, $\mathbf{f} : \Omega \subset \mathbb{R}^2 \rightarrow [0, 1]^3$ is the original clean image, $\mathbf{n}$ denotes random noise, $k : \mathbb{R}^2 \rightarrow \mathbb{R}$ stands for a known blur kernel, the symbol $*$ refers to the convolution operator and $k * \mathbf{f}$ represents k convoluting with each component of $\mathbf{f}$. The deconvolution, or image restoration, is to recover the latent image $\mathbf{f}$ from the given observed image $\mathbf{g}$, which is an ill-posed inverse problem.

Statistical models and variational regularization methods are the most two popular techniques for image restoration in recent years. The statistical approaches (e.g., [1–3]) are mainly based on maximum likelihood estimator (MLE) and Bayesian maximum a posteriori (MAP) method. The variational methods (e.g., [4–6]) get a latent image by minimizing of a cost functional. Most of these traditional image restoration literatures

Jun Liu, Haiyang Huang, Zhongdan Huan
School of Mathematical Sciences, Beijing Normal University, Laboratory of Mathematics and Complex Systems, Ministry of Education, Beijing, 100875, P.R. China. The corresponding author is Zhongdan Huan.

E-mail: liujn121@gmail.com (Jun Liu)
E-mail: hhywsg@bnu.edu.cn (Haiyang Huang)
E-mail: bnumath217@gmail.com (Zhongdan Huan)

Haili Zhang
Department of Mathematics, University of Florida, Gainesville, FL, 32611-8105.
E-mail: hlzhang@ufl.edu

(e.g., [4, 7, 8]) only consider the case of blurred grey scale images with a small amount of additive Gaussian noise.

However, in real applications, the images we obtain are sometimes contaminated with blur and impulse noise. Classical techniques for removing impulse noise are mainly based on median-type filters (e.g., [10, 11]), which are very effective for noisy images but may fail when the images are further degraded with blur. Due to the superiority in preserving edges, some variational based methods have emerged in recent years to deal with impulse noise. In [12, 13], Nikolova *et al.* proposed a variational framework for deblurring grey scales image with impulse noise. A significant contribution of their work is that some nonsmooth fidelity terms such as L^1 -based fidelity terms were introduced to remove impulse noise. Bar *et al.* [14] developed it by considering different Mumford-Shah functional regularizers and they extended it to the color images in [15]. Yang *et al.* [16] proposed an efficient algorithm, FTVd (fast total variation deblurring), for TVL¹ model to deblur color images with impulse noise. Comparing with the previous methods, their approaches could remove impulse noise more efficiently. However, the reconstructions are not satisfactory when the images are seriously degraded by blur and noise. In [17], Cai *et al.* introduced a Mumford-Shah two-phase method to deblur images with both Gaussian noise and impulse noise. In the first step the likely noisy data identified by a median-type filter is removed from the data set. After that the image is reconstructed from the remaining data entries. To implement the algorithm fastly, they in [18] used $L^1 - L^1$ minimization in the second step of the two-phase method. Huang *et al.* [19] used TVL²-based two-phase method to reconstruct images and offered a fast alternating minimization algorithm. Experimental results have shown that the two-phase method performs well for salt-and-pepper noise. However, it cannot get a satisfactory result for high density random-valued noise because the median-type filters fails in this case. As is shown in [17–19], the two-phase restoration result is not good when the image is corrupted by random-valued noise with noise ratio more than 55%. Of course, the two-phase method cannot well handle mixed noise such as Gaussian mixture. In [20], a statistical method is employed to recover blurred grey scale images from mixed noisy data. Essentially, we include a L^2 -based weighting fidelity term in the cost functional, which has a superior performance in removing mixed noise, especially when the level of noise is high. In fact, it could also be considered as an adaptive two-phase method.

In this work, we generalize our preliminary study (GM-TV model) to vector-valued/color images with impulse noise or other mixed noise, and present two new

models called TVAWL² and TVAWL¹ (total variation based adaptively weighting L^2 / L^1 method). These two models can restore blurred color images in the presence of high density impulse noise (with 70% density or more). The approach is formulated on the fact that probability density function (PDF) of impulse noise can be approximated as an additive noise with mixed Gaussian or two-sided exponential distributions. In contrast with existing variational models, a new adaptively weighting fidelity term is introduced in the proposed cost functional. Noise could be automatically detected by the weighting functions, so our approach can obtain an impressive reconstruction even though the image is corrupted by high density impulse noise. In addition, we introduce some control parameters to the fidelity term, which make the regularization parameter in the proposed models less sensitive than that of others. Moreover, in this study, all the proposed models are solved by splitting schemes, which could significantly speed up the numerical implementation process.

The rest of the paper is organized as follows: in section 2 we first give some basic notations which is used in this paper, then we introduce the impulse noise model and some approximations to its PDF; the expectation-maximization (EM) algorithm which is used to estimate the parameters of the approximated PDF is reviewed in section 3; in section 4, the proposed TVAWL² and TVAWL¹ models are described; section 5 contains some details about the implementation of the algorithm and experimental results; finally, we summarize our approach and conclude the paper in section 6.

2 Basic Notation and Motivation

2.1 Notation and Definition

Throughout this paper, we use boldface type (e.g. $\mathbf{f}$) for vectors or vector-valued functions. To simplify representations, we introduce here the notations used throughout the paper.

Functions:

$p, p_1, p_2 : \mathbb{R} \rightarrow \mathbb{R}^+$, which always denote PDF.

$f_\tau, g_\tau, w_\tau : \Omega \subset \mathbb{R}^2 \rightarrow [0, 1], \quad \tau = 0, 1, 2.$

$\mathbf{f}, \mathbf{g}, \mathbf{w} : \Omega \subset \mathbb{R}^2 \rightarrow [0, 1]^3,$

$\mathbf{d}_0, \mathbf{d}_1, \mathbf{d}_2 : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}^2,$

$\mathbf{d} : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}^{2 \times 3},$

$\mathbf{f} = (f_0, f_1, f_2)^T, \mathbf{g} = (g_0, g_1, g_2)^T,$

$\mathbf{w} = (w_0, w_1, w_2)^T, \mathbf{d} = (\mathbf{d}_0, \mathbf{d}_1, \mathbf{d}_2).$

Inner products and norms:

$$\langle \mathbf{f}, \mathbf{g} \rangle = \sum_{\tau=0}^2 \langle f_{\tau}, g_{\tau} \rangle = \sum_{\tau=0}^2 \int_{\Omega} f_{\tau} g_{\tau} dx,$$

$$\langle \mathbf{f}, \mathbf{g} \rangle_{\mathbf{w}} = \sum_{\tau=0}^2 \langle w_{\tau} f_{\tau}, g_{\tau} \rangle = \sum_{\tau=0}^2 \int_{\Omega} w_{\tau} f_{\tau} g_{\tau} dx,$$

$$\|\mathbf{f}\|_2^2 = \langle \mathbf{f}, \mathbf{f} \rangle = \sum_{\tau=0}^2 \int_{\Omega} f_{\tau}^2 dx,$$

$$\|\mathbf{f}\|_{2,\mathbf{w}}^2 = \langle \mathbf{f}, \mathbf{f} \rangle_{\mathbf{w}} = \sum_{\tau=0}^2 \int_{\Omega} w_{\tau} f_{\tau}^2 dx,$$

$$\|\mathbf{f}\|_1 = \sum_{\tau=0}^2 \int_{\Omega} |f_{\tau}| dx,$$

$$\|\mathbf{f}\|_{1,\mathbf{w}} = \sum_{\tau=0}^2 \int_{\Omega} w_{\tau} |f_{\tau}| dx,$$

$$|\mathbf{f}| = \sqrt{f_0^2 + f_1^2 + f_2^2},$$

$$|\mathbf{d}| = \sqrt{|\mathbf{d}_0|^2 + |\mathbf{d}_1|^2 + |\mathbf{d}_2|^2}.$$

Operators:

$$\nabla \mathbf{f} = (\nabla f_0, \nabla f_1, \nabla f_2),$$

$$\Delta \mathbf{f} = (\Delta f_0, \Delta f_1, \Delta f_2)^T,$$

$$\nabla \cdot \mathbf{d} = (\nabla \cdot \mathbf{d}_0, \nabla \cdot \mathbf{d}_1, \nabla \cdot \mathbf{d}_2)^T,$$

$$\mathbf{fg} \triangleq (f_0 g_0, f_1 g_1, f_2 g_2)^T,$$

$$\mathbf{f}^2 \triangleq (f_0^2, f_1^2, f_2^2)^T,$$

$$\frac{\mathbf{f}}{\mathbf{g}} \triangleq \left(\frac{f_0}{g_0}, \frac{f_1}{g_1}, \frac{f_2}{g_2} \right)^T,$$

$$p(\mathbf{f}) \triangleq (p(f_0), p(f_1), p(f_2))^T,$$

$$\dagger \mathbf{f} \dagger = (|f_0|, |f_1|, |f_2|)^T.$$

Discretization:

let $f_{i,j,\tau}$, $0 \leq i \leq m_1 - 1$, $0 \leq j \leq m_2 - 1$ denote the discretization of f_{τ} , $\tau = 0, 1, 2$. We also write $f_{i,j,\tau}$ in a vector form in which the i' -th element $f_{i'}$ is $f_{i,j,\tau}$ with $i' = i + j m_1 + \tau m_1 m_2$. These will be used extensively in the later sections.

2.2 Impulse Noise Model and Some Approximations to Its PDF

Two common types of impulse noise are random-valued noise and salt-and-pepper noise. Let the noise ratio be r . At each pixel the image intensity of the noisy image remains the same with probability $1 - r$ and is changed to a uniformly distributed random number with probability r , which is referred to as random-valued noise. Salt-and-pepper noise has a similar formulation except that the changed pixel intensities have only two possible values with probability r_1 and r_2 respectively. Mathematically, the process of image degradation with impulse noise can be described as

$$g_{i,j,\tau} = \mathfrak{T}[(k * \mathbf{f})_{i,j,\tau}],$$

where $\mathfrak{T}$ is an operator which represents the impulsive process. Thus for random-valued noise,

$$\mathfrak{T}[(k * \mathbf{f})_{i,j,\tau}] = \begin{cases} (k * \mathbf{f})_{i,j,\tau}, & \text{with probability } 1 - r, \\ a, & \text{with probability } r. \end{cases}$$

In the above equation a is a random number which is uniformly distributed in $[0, 1]$. Similarly, the salt-and-pepper noisy image is given by

$$\mathfrak{T}[(k * \mathbf{f})_{i,j,\tau}] = \begin{cases} (k * \mathbf{f})_{i,j,\tau}, & \text{with probability } 1 - r_1 - r_2, \\ 0, & \text{with probability } r_1, \\ 1, & \text{with probability } r_2. \end{cases}$$

Impulse noise is never additive, but it can be regarded as $\mathbf{g} = k * \mathbf{f} + \mathbf{n}$, where

$$n_{i,j,\tau} = \begin{cases} 0, & \text{with probability } 1 - r, \\ a - (k * \mathbf{f})_{i,j,\tau}, & \text{with probability } r, \end{cases} \quad (1)$$

for random-valued noise and for salt-and-pepper noise,

$$n_{i,j,\tau} = \begin{cases} 0, & \text{with probability } 1 - r_1 - r_2, \\ -(k * \mathbf{f})_{i,j,\tau}, & \text{with probability } r_1, \\ 1 - (k * \mathbf{f})_{i,j,\tau}, & \text{with probability } r_2. \end{cases}$$

Suppose $n_{i,j,\tau}$ is a realization of a random variable $\mathcal{N}$ with PDF $p(y)$, then $p(y)$ has the following properties.

1. Random-valued noise.

Assume $0, -(k * \mathbf{f})_{i,j,\tau}$ are the realizations of two random variables $\mathcal{Y}_1, \mathcal{Y}_2$ with PDF $p_1(y), p_2(y)$. Actually p_1 is the Delta function δ and $p(y)$ is given by

$$p(y) = (1 - r)\delta(y) + r \int_{y-1}^y p_2(z) dz. \quad (2)$$

See Appendix A. 1 for more details about this calculation. The blur kernel k is usually assumed to satisfy nonnegativity condition $k \geq 0$ and DC-condition $\int k dx = 1$ to get a well-posed solution (see e.g.

[21]). With these two conditions, we can easily get $-1 \leq -(k * \mathbf{f})_{i,j,\tau} \leq 0$ for $0 \leq f_\tau \leq 1$. Thus $p_2(y) = 0$ for all $y \notin [-1, 0]$. In other words, $p(y)$ is a compactly supported function with support $[-1, 1]$. In addition, we have the following proposition for $p(y)$.

Proposition 1 *$p(y)$ is monotone decreasing in $(0, 1]$ and monotone increasing $[-1, 0)$; moreover, if $p_2(-0.5 + z) = p_2(-0.5 - z)$ holds for all z , then $p(y) = p(-y)$. Namely, if $p_2(y)$ is symmetric around -0.5 , then $p(y)$ is symmetric around 0 .*

Proof By (2) and the property of p_2 ,

$$p(y) = \begin{cases} r \int_{-1}^y p_2(z) dz, & -1 \leq y < 0, \\ r \int_{y-1}^0 p_2(z) dz, & 0 < y \leq 1. \end{cases}$$

The monotone property of $p(y)$ is obvious from the fact $p_2 \geq 0$. Let $0 < y \leq 1$, then $-1 \leq -y < 0$. Hence, we have

$$p(y) = r \int_{y-1}^0 p_2(z) dz = r \int_{y-0.5}^{0.5} p_2(-0.5 + z) dz,$$

$$p(-y) = r \int_{-1}^{-y} p_2(z) dz = r \int_{y-0.5}^{0.5} p_2(-0.5 - z) dz.$$

Therefore, $p(y) = p(-y)$. $\square$

2. Salt-and-pepper noise.

Similarly, for the case of salt-and-pepper noise,

$$p(y) = (1 - r_1 - r_2)\delta(y) + r_1 p_2(y) + r_2 p_2(y - 1). \quad (3)$$

For simplicity, in this paper we set $r_1 = r_2$. $p(y)$ satisfies proposition 2.

Proposition 2 *$p(y) = p(y+1)$ in $(-1, 0)$. If $p_2(-0.5 + z) = p_2(-0.5 - z)$ holds for all z , then we have $p(y) = p(-y)$.*

Proof Similar to the proof of proposition 1. $\square$

To better understand the above propositions, we give a specific example in Fig. 1, which contains the PDFs of the two kinds of impulse noise. In this experiment, we first add a Gaussian blur, with standard deviation be 3.0, to the color “lenna” image $\mathbf{f}$ (see Fig. 2), where the normalized histogram of $-k * \mathbf{f}$ is plotted in Fig. 1(a). Then the blurred image $k * \mathbf{f}$ is contaminated with random-valued noise and salt-and-pepper noise with density 60%, respectively. The normalized histogram of $g - k * \mathbf{f}$ for the two cases are plotted in Fig. 1(b) and Fig. 1(c). One could find the normalized histograms of $g - k * \mathbf{f}$ is approximately symmetric around 0. It is understandable because the histogram of $-k * \mathbf{f}$ in Fig. 1(a) is approximately symmetric around

-0.5 . We have tested lots of natural images and most of them seem to have this approximately symmetric property.

As is known in [12–15] that L^1 -based fidelity outperforms the L^2 -based one in removing impulse noise. In the following, we will interpret this by examining the approximations to the PDF of noise $p(y)$ in Fig. 2. For grey scale image, the maximum likelihood estimation of the additive white Gaussian noise $n \sim N(0, \sigma^2)$ leads to the L^2 norm based fidelity term

$$\frac{1}{2} \|g - k * f\|_2^2,$$

while the estimation for a two-sided exponential distribution would result in the L^1 norm based fidelity term

$$\|g - k * f\|_1.$$

These two special functions, Gaussian and two-sided exponential PDFs, can be considered as two different approximations to $p(y)$. From the example in Fig. 1, we can see that the two-sided exponential function has a slightly better approximation to the true PDF than the Gaussian approximation. That is why L^1 norm based fidelity could remove impulse noise more efficiently than the L^2 based one.

Based on (2), (3) and the above experiment observations, the PDF of impulse noise $p(y)$ can be approximated by

$$p(y) \approx p(y; \Theta) = \sum_{l=1}^2 \alpha_l p_l(y; \theta_l), \quad (4)$$

where $\Theta = \{\alpha_1, \alpha_2, \theta_1, \theta_2\}$ is a parameter set such that $\sum_{l=1}^2 \alpha_l = 1$, and $p_l(y; \theta_l)$ is a density function parameterized by θ_l . Based on the symmetric property discussed in the previous paragraphs, we can view $p_l(y; \theta_l)$ as a Gaussian or two-sided exponential PDF, i.e.,

$$p_l(y; \sigma_l^2) = \frac{1}{\sqrt{2\pi\sigma_l^2}} \exp\left\{-\frac{y^2}{2\sigma_l^2}\right\} \quad (5)$$

or

$$p_l(y; \sigma_l^2) = \frac{1}{2\sigma_l^2} \exp\left\{-\frac{|y|}{\sigma_l^2}\right\}. \quad (6)$$

Remark: We do not directly use (2) and (3) as the PDFs of noise but choose their approximations. This is because the cost functionals derived from (2) and (3) are extremely complex and difficult to minimize. On the other hand, numerical experimental results show that both δ and $\int_{y-1}^y p_2(z) dz$ could be well approximated by two Gaussian (or two-sided exponential) functions by choosing a small σ_1^2 or a large σ_2^2 respectively. Meanwhile, the corresponding cost functionals (ref. section

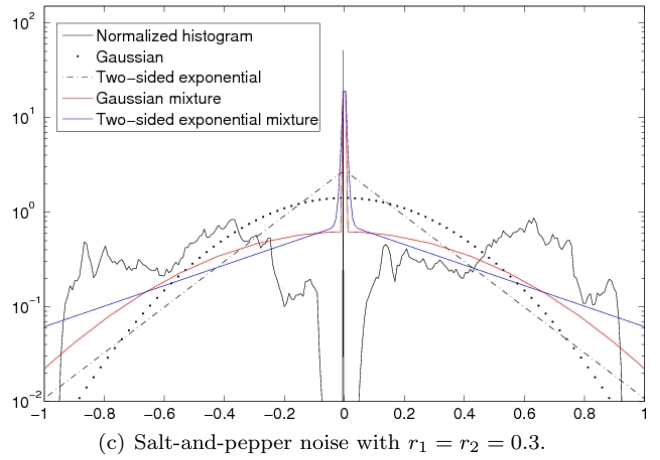
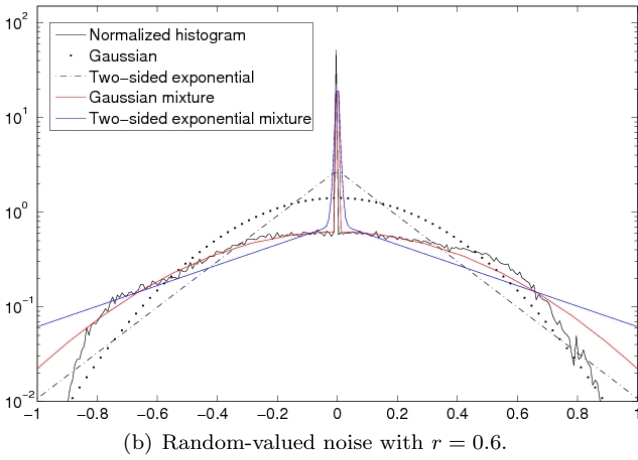
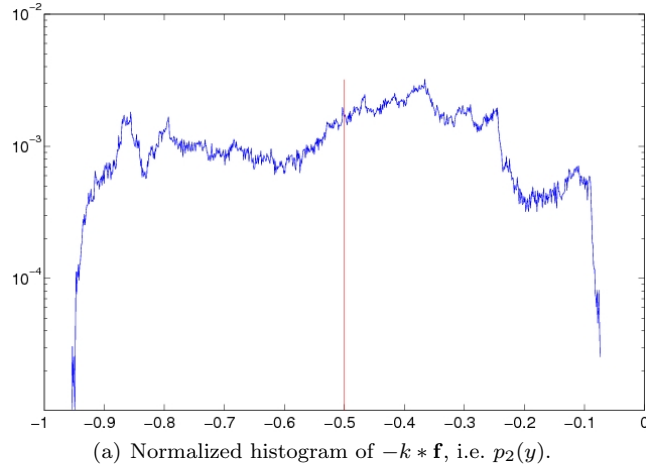


Fig. 1 Different approximations to $p(y)$ for random-valued noise and salt-and-pepper noise, respectively. Note that all the scale of y axis is logarithmic and we plot these figures by MATLAB function “semilogy”.

4) resulting from the two Gaussian and two-sided exponential functions are quadratic or half-quadratic and are easy to solve. Theoretically speaking, it is known that the Delta function can be well approximated by Gaussian function, namely,

$$\lim_{\sigma \rightarrow 0^+} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{y^2}{2\sigma^2}\right\} = \delta(y)$$

holds in the sense of distributions. However, we have not fully explored the connections between $\int_{y-1}^y p_2(z) dz$ and exponential types functions, which is left for future research.

Of course, there are many choices for such approximation, e.g., mixtures of Gaussian/uniform, two-sided exponential/uniform, two-sided exponential/Gaussian, etc. In this paper, we only choose Gaussian/Gaussian and two-sided exponential/two-sided exponential mixtures. Others can be addressed in the same manner.

The mixture models, in fact, have better approximations to the PDF of impulse noise than others, which is

evident in Fig. 1. Unfortunately, the cost functionals for the mixture models are difficult to optimize since there are many parameters and a logarithm of the sum. In the next section, we will discuss how to overcome this difficulty.

3 Parameters Estimation via the EM Algorithm

In this section, we discuss how to estimate the parameters of noise for any given data $\mathbf{n}$.

The maximum likelihood estimator is usually employed for such kind of problem. Let $M = 3m_1m_2 - 1$ and $\mathcal{N}_{i'}, i' = 0, 1, 2, \dots, M$ are random variables. Suppose all these random variables are independent and identically distributed with PDF $p(n_{i'}; \Theta)$, where Θ is an unknown parameter set. Thus $\mathcal{N} = (\mathcal{N}_0, \mathcal{N}_1, \dots, \mathcal{N}_M)$

is a random vector with PDF

$$p(\mathbf{n}; \Theta) = \prod_{i'=0}^M p(n_{i'}; \Theta).$$

We can view the noise $\mathbf{n} = (n_0, n_1, \dots, n_M)$ is a realization of $\mathcal{N}$. Then the MLE of Θ given $\mathbf{n}$ is a parameter Θ^* that maximizes the likelihood function or the log-likelihood function

$$\begin{aligned} l(\Theta; \mathbf{n}) &= \ln p(\mathbf{n}; \Theta) = \ln \prod_{i'=0}^M p(n_{i'}; \Theta) \\ &= \sum_{i'=0}^M \ln p(n_{i'}; \Theta). \end{aligned} \quad (7)$$

Actually, in the above equation $p(n_{i'}; \Theta)$ is a discretization of (4). So the likelihood function becomes

$$l(\Theta; \mathbf{n}) = \sum_{i'=0}^M \ln \sum_{l=1}^2 \alpha_l p_l(n_{i'}; \theta_l),$$

which is difficult to optimize because it contains the logarithm of the sum. To address this difficulty, we resort to EM algorithm, an efficient algorithm for mixture model parameter estimation. The basic idea of EM algorithm is that: instead of maximizing $l(\Theta; \mathbf{n})$, we turn to maximize another function $H(\Theta)$, which can be more easily maximized with the property that the values of $l(\Theta; \mathbf{n})$ do not decrease in each step of the iteration. Consequently, the main task is how to construct such a function $H(\Theta)$. We will review this [23, 22, 9, 24] in the next paragraph.

Generally speaking, we do not know in advance that a pixel in the image whether is corrupted by noise or not, especially when the image is contaminated by random-valued noise. Here we introduce a random variable $\mathcal{C}_{i'}$ to indicate the pixel at location i' is noise or not, to be more exact, if $g_{i'}$ is a noise free pixel, then let $\mathcal{C}_{i'} = 1$, else if $g_{i'}$ is contaminated by noise, we let $\mathcal{C}_{i'} = 2$. Namely, if we denote $\mathfrak{S}_1 = \{0\}$, $\mathfrak{S}_2 = \mathbb{R} \setminus \{0\}$ and $\mathcal{C} = (\mathcal{C}_0, \mathcal{C}_1, \dots, \mathcal{C}_M)$, then we have $\mathcal{C}_{i'} = 1$ when $n_{i'} \in \mathfrak{S}_1$ and $\mathcal{C}_{i'} = 2$ when $n_{i'} \in \mathfrak{S}_2$.

We consider $\mathcal{N}$ as the observed variable or incomplete variable and $\mathcal{Z} = (\mathcal{N}, \mathcal{C})$ as the complete variable, where $\mathcal{C}$ is the hidden variable. Next, let $p(\mathbf{n}, \mathbf{c}; \Theta)$, $p(\mathbf{c}; \Theta)$ be the PDFs of $\mathcal{Z}$ and $\mathcal{C}$ respectively. Then the conditional PDF of $\mathbf{c}$ given $\mathbf{n}$ is

$$p(\mathbf{c}|\mathbf{n}; \Theta) = \frac{p(\mathbf{n}, \mathbf{c}; \Theta)}{p(\mathbf{n}; \Theta)}. \quad (8)$$

Thus the log-likelihood function (7) for Θ given $\mathbf{n}$ becomes

$$\begin{aligned} l(\Theta; \mathbf{n}) &= \ln p(\mathbf{n}; \Theta) \\ &= \ln p(\mathbf{n}, \mathbf{c}; \Theta) - \ln p(\mathbf{c}|\mathbf{n}; \Theta) \\ &= l(\Theta; \mathbf{n}, \mathbf{c}) - l(\Theta; \mathbf{c}|\mathbf{n}). \end{aligned} \quad (9)$$

The second equality follows by equation (8) and the last one follows from the definition of the log-likelihood function.

Given parameter set Θ^ν , where ν refers to an iteration variable, we have

$$\begin{aligned} l(\Theta; \mathbf{n}) &= l(\Theta; \mathbf{n}) \sum_{\mathbf{c}} p(\mathbf{c}|\mathbf{n}; \Theta^\nu) \\ &= \sum_{\mathbf{c}} l(\Theta; \mathbf{n}) p(\mathbf{c}|\mathbf{n}; \Theta^\nu) \\ &\stackrel{(9)}{=} \sum_{\mathbf{c}} l(\Theta; \mathbf{n}, \mathbf{c}) p(\mathbf{c}|\mathbf{n}; \Theta^\nu) - \sum_{\mathbf{c}} l(\Theta; \mathbf{c}|\mathbf{n}) p(\mathbf{c}|\mathbf{n}; \Theta^\nu) \\ &\triangleq H(\Theta; \Theta^\nu) - J(\Theta; \Theta^\nu). \end{aligned} \quad (10)$$

In the above equation, $\sum_{\mathbf{c}}$ represents the sum over all the possible values of $\mathbf{c}$ and thus $\sum_{\mathbf{c}} p(\mathbf{c}|\mathbf{n}; \Theta^\nu) = 1$.

With these notations, one can get the following theorem.

Theorem 1 For every ν , if $H(\Theta^{\nu+1}; \Theta^\nu) \geq H(\Theta^\nu; \Theta^\nu)$, then $l(\Theta^{\nu+1}; \mathbf{n}) \geq l(\Theta^\nu; \mathbf{n})$.

The proof of this theorem can be found in [9]. For completeness, we give its proof in the Appendix (A. 2).

Consequently, the EM algorithm could be summarized as following:

EM algorithm.

Given the observed data $\mathbf{n}$ and initial guess parameter set Θ^0 , for $\nu = 0, 1, 2, \dots$, do

Step 1 (E-step). Compute

$$H(\Theta; \Theta^\nu) = \sum_{\mathbf{c}} \ln p(\mathbf{n}, \mathbf{c}; \Theta) p(\mathbf{c}|\mathbf{n}; \Theta^\nu). \quad (11)$$

Step 2 (M-step). Find $\Theta^{\nu+1}$, such that

$$\Theta^{\nu+1} = \arg \max_{\Theta} H(\Theta; \Theta^\nu). \quad (12)$$

The crucial procedure of EM algorithm is to calculate $H(\Theta; \Theta^\nu)$. Here we consider mixture model (4), by using Bayes's rule, $p(c_{i'}|n_{i'}; \Theta^\nu)$, $p(n_{i'}, c_{i'}; \Theta)$ in equation (11) becomes

$$\begin{aligned} p(c_{i'}|n_{i'}; \Theta^\nu) &= \frac{p(c_{i'})p(n_{i'}|c_{i'}; \Theta^\nu)}{p(n_{i'}; \Theta^\nu)} \\ &= \frac{\alpha_{c_{i'}}^\nu p_{c_{i'}}^\nu(n_{i'}; \theta_{c_{i'}}^\nu)}{\sum_{\varsigma=1}^2 \alpha_{\varsigma}^\nu p_{\varsigma}^\nu(n_{i'}; \theta_{\varsigma}^\nu)}, \end{aligned} \quad (13)$$

and

$$\begin{aligned} p(n_{i'}, c_{i'}; \Theta) &= p(c_{i'})p(n_{i'}|c_{i'}; \Theta) \\ &= \alpha_{c_{i'}} p_{c_{i'}}(n_{i'}; \theta_{c_{i'}}). \end{aligned} \quad (14)$$

Plugging (13) and (14) into equation (11) and simplifying the expression, then

$$H(\Theta; \Theta^\nu) = \sum_{l=1}^2 \sum_{i'=0}^M w^l(i'; \Theta^\nu) \ln \alpha_l + \sum_{l=1}^2 \sum_{i'=0}^M w^l(i'; \Theta^\nu) \ln p_l(n_{i'}; \theta_l), \quad (15)$$

where $\Theta = \{\alpha_1, \alpha_2, \theta_1, \theta_2\}$ and

$$w^l(i'; \Theta^\nu) = \frac{\alpha_l^\nu p_l(n_{i'}; \theta_l^\nu)}{\sum_{\varsigma=1}^2 \alpha_\varsigma^\nu p_\varsigma(n_{i'}; \theta_\varsigma^\nu)}. \quad (16)$$

More details about calculating $H(\Theta; \Theta^\nu)$ could be found in [22]. Comparing (13) and (16), we can get that $w^l(i'; \Theta^\nu)$ is a priori probability of $n_{i'} \in \mathfrak{S}_l$. This function is very important for our later approach because it could discriminate between pixels with only blur and noise.

4 The Proposed Variational Model

Now we incorporate the features of variational regularization methods and EM algorithm and present a new model that can deblur color images with impulse noise. Let us begin with some studies regarding the connections between the probability models and the regularization methods. Recall that the degraded color image $\mathbf{g}$ could be modeled by the discrete version

$$g_{i'} = (k * \mathbf{f})_{i'} + n_{i'}, \quad (17)$$

where $i' = \tau m_1 m_2 + j m_1 + i$. For all i' , assume $\mathcal{G}_{i'}$, $\mathcal{F}_{i'}$, $\mathcal{N}_{i'}$ are three different random variables, and $g_{i'}$ is a realization of $\mathcal{G}_{i'}$. Then the MAP estimator with respect to $\mathbf{f}$ is given by

$$\mathbf{f}^* = \arg \max_{\mathbf{f}} p(\mathbf{f}|\mathbf{g}) = \arg \min_{\mathbf{f}} -\ln p(\mathbf{f}|\mathbf{g}), \quad (18)$$

where

$$\mathbf{f} = (f_0, f_1, \dots, f_M), \quad \mathbf{g} = (g_0, g_1, \dots, g_M).$$

If each $f_{i'}$ is conditionally independent and identically-distributed, we can get

$$p(\mathbf{f}|\mathbf{g}) = \prod_{i'=0}^M \frac{p(g_{i'}|f_{i'})p(f_{i'})}{p(g_{i'})}.$$

Note that $p(g_{i'})$ in the above equation is a fixed constant since $g_{i'}$ is given by the observed image. Hence, the problem (18) becomes

$$\mathbf{f}^* = \arg \min_{\mathbf{f}} \left\{ -\sum_{i'=0}^M \ln p(g_{i'}|f_{i'}) - \sum_{i'=0}^M \ln p(f_{i'}) \right\}.$$

For Gaussian noise, $\mathcal{N}_{i'} \sim N(0, \sigma^2)$, we usually suppose σ^2 is a known constant. Under these conditions, the first term of (19) in the continuous setting leads to the following L²-based fidelity term

$$\frac{1}{2} \|\mathbf{g} - (k * \mathbf{f})\|_2^2 + c_1,$$

where c_1 is a constant. On the other hand, the assumption of $p(f_{i'})$ would lead to a smooth term or called regularization term. For example, the assumption

$$p(f_{i'}) \propto \exp\left\{-\frac{\lambda f_{i'}^2}{2}\right\}$$

leads to a channel-by-channel Tikhonov regularization term

$$\frac{\lambda}{2} \|\mathbf{f}\|_2^2.$$

As for impulse noise, we assume the noisy data $n_{i'} = g_{i'} - (k * \mathbf{f})_{i'}$ approximately follows a distribution with the mixture PDF (4). Inspired by EM algorithm to iteratively solve a series of minimization problems, we can establish a new model according to (19), theorem 1, (15) and (4):

$$(\mathbf{f}^{\nu+1}, \Theta^{\nu+1}) = \arg \min_{\mathbf{f}, \Theta} E(\mathbf{f}, \Theta; w_{i'}^{l,\nu}),$$

where $E(\mathbf{f}, \Theta; w_{i'}^{l,\nu})$

$$= -\sum_{l=1}^2 \sum_{i'=0}^M w_{i'}^{l,\nu} \ln p_l(g_{i'} - (k * \mathbf{f})_{i'}; \theta_l) - \sum_{l=1}^2 \sum_{i'=0}^M w_{i'}^{l,\nu} \ln \alpha_l + \lambda R(\mathbf{f}). \quad (20)$$

In the above equation,

$$w_{i'}^{l,\nu} = \frac{\alpha_l^\nu p_l(g_{i'} - (k * \mathbf{f}^\nu)_{i'}; \theta_l^\nu)}{\sum_{\varsigma=1}^2 \alpha_\varsigma^\nu p_\varsigma(g_{i'} - (k * \mathbf{f}^\nu)_{i'}; \theta_\varsigma^\nu)}, \quad (21)$$

$\alpha_2 = 1 - \alpha_1$, $R(\mathbf{f})$ is a regularization term and $\lambda > 0$ is a parameter.

In this paper, we do not focus on the smoothing term $R(\mathbf{f})$. For computational convenience, we utilize the vectorial total variation (VTV) regularization

$$\text{VTV}(\mathbf{f}) = \int_{\Omega} |\nabla \mathbf{f}| d\mathbf{x},$$

which can be found in [25]. Other regularizers [26, 27] such as the Beltrami regularizer, Mumford-Shah segmentation functional [28, 14, 15], and nonlocal regularization operator [29, 30] can also be employed for this problem.

Remark: The fidelity term in the proposed method do not couple the rgb channels since the impulse noise is usually independent of channels. However, it is easy to extend this to the coupling version by using the three dimensions PDF and considering a better channels coupled regularization such as Beltrami regularizer.

Starting from an initial guess $\mathbf{f}^0, \Theta^0$, we compute a series of minimizers

$$\mathbf{f}^1, \Theta^1, \mathbf{f}^2, \Theta^2, \dots, \mathbf{f}^{\nu+1}, \Theta^{\nu+1}, \dots$$

such that

$$(\mathbf{f}^{\nu+1}, \Theta^{\nu+1}) = \arg \min_{\mathbf{f}, \Theta} E(\mathbf{f}, \Theta; \mathbf{f}^\nu, \Theta^\nu).$$

Then plugging (5) or (6) into (20), ignoring any constant term and using the notations which is defined in section 2.1, $E(\mathbf{f}, \Theta; \mathbf{f}^\nu, \Theta^\nu)$ has the following expressions:

1. TVAWL² model. The cost functional $E(\mathbf{f}, \Theta; \mathbf{f}^\nu, \Theta^\nu)$ is defined by

$$\begin{aligned} & \frac{1}{2} \sum_{l=1}^2 \frac{\|k * \mathbf{f} - \mathbf{g}\|_{2, \mathbf{w}^{l, \nu}}^2}{\sigma_l^2} + \lambda \int |\nabla \mathbf{f}| \, d\mathbf{x} \\ & + \frac{1}{2} \sum_{l=1}^2 \langle \mathbf{w}^{l, \nu}, \mathbf{1} \rangle (\ln \sigma_l^2 - 2 \ln \alpha_l), \end{aligned} \quad (22)$$

where $\alpha_2 = 1 - \alpha_1$, and

$$\mathbf{w}^{l, \nu} = \frac{\frac{\alpha_l^\nu}{(\sigma_l^\nu)^\nu} \exp \left\{ -\frac{(k * \mathbf{f}^\nu - \mathbf{g})^2}{2(\sigma_l^\nu)^\nu} \right\}}{\sum_{\varsigma=1}^2 \frac{\alpha_\varsigma^\nu}{(\sigma_\varsigma^\nu)^\nu} \exp \left\{ -\frac{(k * \mathbf{f}^\nu - \mathbf{g})^2}{2(\sigma_\varsigma^\nu)^\nu} \right\}}. \quad (23)$$

2. TVAWL¹ model. $E(\mathbf{f}, \Theta; \mathbf{f}^\nu, \Theta^\nu)$ is given by

$$\begin{aligned} & \sum_{l=1}^2 \frac{\|k * \mathbf{f} - \mathbf{g}\|_{1, \mathbf{w}^{l, \nu}}}{\sigma_l^2} + \lambda \int |\nabla \mathbf{f}| \, d\mathbf{x} \\ & + \sum_{l=1}^2 \langle \mathbf{w}^{l, \nu}, \mathbf{1} \rangle (\ln \sigma_l^2 - \ln \alpha_l), \end{aligned} \quad (24)$$

where $\alpha_2 = 1 - \alpha_1$, and

$$\mathbf{w}^{l, \nu} = \frac{\frac{\alpha_l^\nu}{(\sigma_l^\nu)^\nu} \exp \left\{ -\frac{\|k * \mathbf{f}^\nu - \mathbf{g}\|_1}{(\sigma_l^\nu)^\nu} \right\}}{\sum_{\varsigma=1}^2 \frac{\alpha_\varsigma^\nu}{(\sigma_\varsigma^\nu)^\nu} \exp \left\{ -\frac{\|k * \mathbf{f}^\nu - \mathbf{g}\|_1}{(\sigma_\varsigma^\nu)^\nu} \right\}}. \quad (25)$$

Some symbols are needed to interpret:

$\Theta = \{\alpha_1, \sigma_1^2, \sigma_2^2\}$, where α_1, σ_l^2 are variables; ν is the iteration number; $\mathbf{w}^{1, \nu}$ and $\mathbf{w}^{2, \nu}$ are two known weighting vector-valued functions given the ν -th iterative values $\mathbf{f}^\nu$ and Θ^ν .

The main difference between TVAWL² and TVAWL¹ lies in the norm used in the first term of the cost functional. As shown in the preceding discussion and motivation in section 2.2, these are actually the results of different approximations to the PDF of noise $p(y)$.

The superiority of model (22)/(24) is that the introduction of the weighting functions $\mathbf{w}^l$ can automatically determine deblurring or denoising at each pixel. More specifically, if we let $\tilde{\mathbf{g}} = \mathbf{g} \mathbf{w}^1$ and $\tilde{\tilde{\mathbf{g}}} = \mathbf{g} \mathbf{w}^2$, then $\tilde{\mathbf{g}}$ represents a noise free blurred image while $\tilde{\tilde{\mathbf{g}}}$ stands for random noise. Our model can make the restoration $\tilde{\mathbf{g}}$ less smooth thanks to different values of σ_l^2 . That is to say, the local behaviors of denoising and deblurring could be adjusted by weighting functions $\mathbf{w}^l$ and the statistical parameters σ_l^2 , though a fixed global regularization term is utilized in our model. In the next section, we will present an experiment to further discuss the role of these two weighting functions.

5 Numerical Methods and Experimental Results

5.1 Algorithms

The simple gradient descent method can be employed to solve our problem, but it will be very slow. The main difficulty of solving the proposed minimization problem is the non-differentiability of the regularization term $\text{VTV}(\mathbf{f})$ and L^1 fidelity term in TVAWL¹. In recent years, many efficient algorithms have appeared, i.e., graph cuts [31], dual methods [32, 33], split Bregman method [34], augmented Lagrangian method [35], alternating direction method of multipliers [36], Douglas-Rachford splitting [37, 38]. Most of these algorithms are equivalent. More connections among them could be found in [35–37]. Here, we apply Chambolle's projection algorithm, which is a dual method proposed by Chambolle in [33] and then extended to vector-valued images by Bresson *et al.* in [25], to solve TVAWL² model. For TVAWL¹ problem, we resort to the split Bregman iteration. We want to mention that the proposed model can also be solved by other algorithms with some minor modifications.

First, Let us discuss the algorithm for TVAWL². To solve (22) more efficiently, we add an auxiliary variable $\mathbf{u}$ and get an approximate problem:

$$\min_{\mathbf{f}, \mathbf{u}, \Theta} E(\mathbf{f}, \mathbf{u}, \Theta), \quad (26)$$

where $E(\mathbf{f}, \mathbf{u}, \Theta)$ is defined by

$$\begin{aligned} & \frac{1}{2} \sum_{l=1}^2 \frac{\|k * \mathbf{u} - \mathbf{g}\|_{2, \mathbf{w}^{l, \nu}}^2}{\sigma_l^2} + \frac{\eta}{2} \|\mathbf{f} - \mathbf{u}\|_2^2 + \lambda \int |\nabla \mathbf{f}| \, d\mathbf{x} \\ & + \frac{1}{2} \sum_{l=1}^2 \langle \mathbf{w}^{l, \nu}, \mathbf{1} \rangle (\ln \sigma_l^2 - 2 \ln \alpha_l), \end{aligned}$$

$\eta \gg 1$ is a penalty parameter,

$$\mathbf{w}^{l, \nu} = \frac{\frac{\alpha_l^\nu}{(\sigma_l)^\nu} \exp \left\{ -\frac{(k * \mathbf{u}^\nu - \mathbf{g})^2}{2(\sigma_l^2)^\nu} \right\}}{\sum_{\varsigma=1}^2 \frac{\alpha_\varsigma^\nu}{(\sigma_\varsigma)^\nu} \exp \left\{ -\frac{(k * \mathbf{u}^\nu - \mathbf{g})^2}{2(\sigma_\varsigma^2)^\nu} \right\}}, \quad (27)$$

and $\alpha_2 = 1 - \alpha_1$. More details and theoretical results about the above split scheme can be found in [16, 25].

By applying the alternating minimization algorithm, (26) splits into three subproblems:

subproblem 1.

$$\begin{aligned} \mathbf{f}^{\nu+1} &= \arg \min_{\mathbf{f}} E(\mathbf{f}, \mathbf{u}^\nu, \Theta^\nu) \\ &= \arg \min_{\mathbf{f}} \left\{ \frac{\eta}{2} \|\mathbf{f} - \mathbf{u}^\nu\|_2^2 + \lambda \int |\nabla \mathbf{f}| \, d\mathbf{x} \right\}, \end{aligned} \quad (28)$$

subproblem 2.

$$\begin{aligned} \mathbf{u}^{\nu+1} &= \arg \min_{\mathbf{u}} E(\mathbf{f}^{\nu+1}, \mathbf{u}, \Theta^\nu) \\ &= \arg \min_{\mathbf{u}} \left\{ \frac{1}{2} \sum_{l=1}^2 \frac{\|k * \mathbf{u} - \mathbf{g}\|_{2, \mathbf{w}^{l, \nu}}^2}{(\sigma_l^2)^\nu} + \frac{\eta}{2} \|\mathbf{f}^{\nu+1} - \mathbf{u}\|_2^2 \right\}, \end{aligned} \quad (29)$$

subproblem 3.

$$\begin{aligned} \Theta^{\nu+1} &= \arg \min_{\Theta} E(\mathbf{f}^{\nu+1}, \mathbf{u}^{\nu+1}, \Theta) \\ &= \arg \min_{\Theta} \left\{ \frac{1}{2} \sum_{l=1}^2 \frac{\|k * \mathbf{u}^{\nu+1} - \mathbf{g}\|_{2, \mathbf{w}^{l, \nu}}^2}{\sigma_l^2} + \frac{1}{2} \sum_{l=1}^2 \langle \mathbf{w}^{l, \nu}, \mathbf{1} \rangle (\ln \sigma_l^2 - 2 \ln \alpha_l) \right\}. \end{aligned} \quad (30)$$

These three subproblems are all easy to solve: subproblem 1 (28) is the vectorial ROF model and we can solve it by Chambolle's projection algorithm [33, 25]; the cost functional in (29) is quadratic with respect to $\mathbf{u}$, so the associated Euler-Lagrange equation

$$\hat{k} * \left((k * \mathbf{u}) \sum_{l=1}^2 \frac{\mathbf{w}^{l, \nu}}{(\sigma_l^2)^\nu} \right) + \eta \mathbf{u} = \hat{k} * \left(\mathbf{g} \sum_{l=1}^2 \frac{\mathbf{w}^{l, \nu}}{(\sigma_l^2)^\nu} \right) + \eta \mathbf{f}^{\nu+1} \quad (31)$$

is a positive symmetric definite linear system and the conjugate gradient (CG) method can be employed to

get an approximated solution, where $\hat{k}$ is the conjugate function of k ; finally, $\Theta^{\nu+1}$ in (30) can be given explicitly by

$$\alpha_1^{\nu+1} = \frac{\langle \mathbf{w}^{1, \nu}, \mathbf{1} \rangle}{3|\Omega|}, \quad \alpha_2^{\nu+1} = 1 - \alpha_1^{\nu+1}, \quad (32)$$

$$(\sigma_l^2)^{\nu+1} = \frac{\|k * \mathbf{u}^{\nu+1} - \mathbf{g}\|_{2, \mathbf{w}^{l, \nu}}^2}{\langle \mathbf{w}^{l, \nu}, \mathbf{1} \rangle}. \quad (33)$$

Until now, we have come up with an algorithm to solve TVAWL². **Algorithm 1.**

Choose initial values $\mathbf{f}^0 = \mathbf{u}^0 = \mathbf{0}$,

$\Theta^0 = \{\alpha_1^0 = 0.5, \alpha_2^0 = 0.5, (\sigma_1^2)^0 = 10^{-4}, (\sigma_2^2)^0 = 1.0\}$, and calculate $\mathbf{w}^{l, 0}$ ($l = 1, 2$) by equation (27).

Set $\nu = 0$ do

step 1. Update $\mathbf{f}^{\nu+1}$ from (28) by Chambolle's projection inner iteration.

step 2. If $\|\mathbf{f}^{\nu+1} - \mathbf{f}^\nu\|_\infty < 10^{-2}$, end the algorithm. Else, go to the next step.

step 3. Solve $\mathbf{u}^{\nu+1}$ from (31) by CG inner iteration.

step 4. Update $\Theta^{\nu+1}$ by equations (32), (33).

step 5. Calculate $\mathbf{w}^{l, \nu+1}$ by equation (27).

step 6. $\nu = \nu + 1$.

Based on some experimental experiences, we let the number of Chambolle's projection inner iteration to be 1 and end the CG inner iteration when $\|\mathbf{u}^{\nu_1+1} - \mathbf{u}^{\nu_1}\|_\infty < 10^{-3}$, where ν_1 is the number of the CG iteration.

Now let us move on to solving TVAWL¹ problem. Other than the former model, a weighting L¹ norm in TVAWL¹ makes this problem more difficult to optimize. We mention that Chambolle's projection algorithm can still be used, but here we employ the split Bregman method [34] to solve it.

Similarly, we first add two auxiliary function $\tilde{\mathbf{d}} \in \mathbb{R}^3, \tilde{\mathbf{d}} \in \mathbb{R}^{2 \times 3}$ and get the following approximated cost functional to (24):

$$\begin{aligned} & \sum_{l=1}^2 \frac{\|\tilde{\mathbf{d}}\|_{1, \mathbf{w}^{l, \nu}}}{\sigma_l^2} + \frac{\eta_1}{2} \sum_{l=1}^2 \frac{\|\tilde{\mathbf{d}} - (k * \mathbf{f} - \mathbf{g})\|_{2, \mathbf{w}^{l, \nu}}^2}{(\sigma_l^2)^\nu} \\ & + \sum_{l=1}^2 \langle \mathbf{w}^{l, \nu}, \mathbf{1} \rangle (\ln \sigma_l^2 - \ln \alpha_l) + \lambda \int |\tilde{\mathbf{d}}| \, d\mathbf{x} \\ & + \frac{\eta_2}{2} \|\tilde{\mathbf{d}} - \nabla \mathbf{f}\|_2^2, \end{aligned} \quad (34)$$

where $\eta_1, \eta_2 \gg 1$ are two penalty parameters, $\alpha_2 = 1 - \alpha_1$ and

$$\mathbf{w}^{l, \nu} = \frac{\frac{\alpha_l^\nu}{(\sigma_l^2)^\nu} \exp \left\{ -\frac{\|\tilde{\mathbf{d}}\|_1}{(\sigma_l^2)^\nu} \right\}}{\sum_{\varsigma=1}^2 \frac{\alpha_\varsigma^\nu}{(\sigma_\varsigma^2)^\nu} \exp \left\{ -\frac{\|\tilde{\mathbf{d}}\|_1}{(\sigma_\varsigma^2)^\nu} \right\}}. \quad (35)$$

Following [34], we add two Bregman variables $\tilde{\mathbf{b}} \in \mathbb{R}^3, \tilde{\tilde{\mathbf{b}}} \in \mathbb{R}^{2 \times 3}$ to (34), then the split Bregman scheme for our problem becomes

$$\begin{cases} (\mathbf{f}^{\nu+1}, \tilde{\mathbf{d}}^{\nu+1}, \tilde{\tilde{\mathbf{d}}}^{\nu+1}, \Theta^{\nu+1}) = \arg \min_{\mathbf{f}, \tilde{\mathbf{d}}, \tilde{\tilde{\mathbf{d}}}, \Theta} E(\mathbf{f}, \tilde{\mathbf{d}}, \tilde{\tilde{\mathbf{d}}}, \Theta), \\ \tilde{\mathbf{b}}^{\nu+1} = \tilde{\mathbf{b}}^{\nu} + k * \mathbf{f}^{\nu+1} - \mathbf{g} - \tilde{\mathbf{d}}^{\nu+1}, \\ \tilde{\tilde{\mathbf{b}}}^{\nu+1} = \tilde{\tilde{\mathbf{b}}}^{\nu} + \nabla \mathbf{f}^{\nu+1} - \tilde{\tilde{\mathbf{d}}}^{\nu+1}, \end{cases}$$

where

$$\begin{aligned} E(\mathbf{f}, \tilde{\mathbf{d}}, \tilde{\tilde{\mathbf{d}}}, \Theta) = & \sum_{l=1}^2 \frac{\|\tilde{\mathbf{d}}\|_{1, \mathbf{w}^{l, \nu}}}{\sigma_l^2} + \frac{\eta_1}{2} \sum_{l=1}^2 \frac{\|\tilde{\mathbf{d}} - (k * \mathbf{f} - \mathbf{g}) - \tilde{\mathbf{b}}^{\nu}\|_{2, \mathbf{w}^{l, \nu}}^2}{(\sigma_l^2)^{\nu}} \\ & + \sum_{l=1}^2 \langle \mathbf{w}^{l, \nu}, \mathbf{1} \rangle (\ln \sigma_l^2 - \ln \alpha_l) + \frac{\eta_2}{2} \|\tilde{\mathbf{d}} - \nabla \mathbf{f} - \tilde{\tilde{\mathbf{b}}}^{\nu}\|_2^2 \\ & + \lambda \int |\tilde{\mathbf{d}}| \, \mathrm{d}\mathbf{x}. \end{aligned}$$

Applying the alternating minimization algorithm, the above minimization problem can be split into the following subproblems:

subproblem 1. $\mathbf{f}^{\nu+1} = \arg \min_{\mathbf{f}} E(\mathbf{f}, \tilde{\mathbf{d}}^{\nu}, \tilde{\tilde{\mathbf{d}}}^{\nu}, \Theta^{\nu}),$

subproblem 2. $\tilde{\mathbf{d}}^{\nu+1} = \arg \min_{\tilde{\mathbf{d}}} E(\mathbf{f}^{\nu+1}, \tilde{\mathbf{d}}, \tilde{\tilde{\mathbf{d}}}^{\nu}, \Theta^{\nu}),$

subproblem 3. $\tilde{\tilde{\mathbf{d}}}^{\nu+1} = \arg \min_{\tilde{\tilde{\mathbf{d}}}} E(\mathbf{f}^{\nu+1}, \tilde{\mathbf{d}}^{\nu+1}, \tilde{\tilde{\mathbf{d}}}, \Theta^{\nu}),$

subproblem 4. $\Theta^{\nu+1} = \arg \min_{\Theta} E(\mathbf{f}^{\nu+1}, \tilde{\mathbf{d}}^{\nu+1}, \tilde{\tilde{\mathbf{d}}}^{\nu+1}, \Theta).$

These four subproblems are all easy to optimize.

The corresponding Euler-Lagrange equation of subproblem 1

$$\begin{aligned} & \hat{k} * \left((k * \mathbf{f}) \sum_{l=1}^2 \frac{\mathbf{w}^{l, \nu}}{(\sigma_l^2)^{\nu}} \right) - \frac{\eta_2}{\eta_1} \Delta \mathbf{f} \\ & = \hat{k} * \left((\mathbf{g} - \tilde{\mathbf{b}}^{\nu} + \tilde{\mathbf{d}}^{\nu}) \sum_{l=1}^2 \frac{\mathbf{w}^{l, \nu}}{(\sigma_l^2)^{\nu}} \right) + \frac{\eta_2}{\eta_1} \nabla \cdot (\tilde{\tilde{\mathbf{b}}}^{\nu} - \tilde{\tilde{\mathbf{d}}}^{\nu}) \end{aligned} \quad (36)$$

is still linear and it can be approximately solved by many solvers (e.g. CG, algebraic multigrid (AMG) solver). If we use circular/Neumann boundary condition for the image, then the convolution can be calculated by the fast Fourier transform[9] (FFT) /discrete cosine transformation [39] (DCT).

The subproblem 2 can be done explicitly by

$$\tilde{\mathbf{d}}^{\nu+1} = \text{shrink} \circ \left(k * \mathbf{f}^{\nu+1} - \mathbf{g} + \tilde{\mathbf{b}}^{\nu}, \frac{1}{\eta_1} \right). \quad (37)$$

Here $\text{shrink} \circ : \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is an operator which has the expression

$$[\text{shrink} \circ (\mathbf{y}, \mathbf{z})]_{\tau} = \frac{y_{\tau}}{|y_{\tau}|} \max\{|y_{\tau}| - z_{\tau}, 0\}, \quad \tau = 0, 1, 2.$$

A simple proof could be found in the Appendix A. 3.

Similarly, the solution of subproblem 3 is given by

$$\tilde{\tilde{\mathbf{d}}}^{\nu+1} = \text{shrink} \left(\nabla \mathbf{f}^{\nu+1} + \tilde{\tilde{\mathbf{b}}}^{\nu}, \frac{\lambda}{\eta_2} \right), \quad (38)$$

and $\text{shrink} : \mathbb{R}^{2 \times 3} \times \mathbb{R} \rightarrow \mathbb{R}^{2 \times 3}$ is an operator such that

$$\text{shrink}(\mathbf{x}, z) = \frac{\mathbf{x}}{|\mathbf{x}|} \max\{|\mathbf{x}| - z, 0\}.$$

Finally, the optimality criteria for $\Theta^{\nu+1}$ is

$$\begin{aligned} \alpha_1^{\nu+1} &= \frac{\langle \mathbf{w}^{1, \nu}, \mathbf{1} \rangle}{3|\Omega|}, \quad \alpha_2^{\nu+1} = 1 - \alpha_1^{\nu+1}, \\ (\sigma_l^2)^{\nu+1} &= \frac{\|\tilde{\mathbf{d}}^{\nu+1}\|_{1, \mathbf{w}^{l, \nu}}}{\langle \mathbf{w}^{l, \nu}, \mathbf{1} \rangle}. \end{aligned} \quad (39)$$

As a result, we obtain an algorithm for TVAWL¹:

Algorithm 2.

Choose initial values $\mathbf{f}^0 = \mathbf{g}$, $\tilde{\mathbf{b}}^0 = \tilde{\mathbf{d}}^0 = \mathbf{0}$, $\tilde{\tilde{\mathbf{b}}}^0 = \tilde{\tilde{\mathbf{d}}}^0 = \mathbf{0}$, $\Theta^0 = \{\alpha_1^0 = 0.5, \alpha_2^0 = 0.5, (\sigma_1^2)^0 = 10^{-4}, (\sigma_2^2)^0 = 1.0\}$, and calculate $\mathbf{w}^{l, 0}$ ($l = 1, 2$) by equation (35).

Set $\nu = 0$ do

step 1. Find $\mathbf{f}^{\nu+1}$ by solving (36) using CG.

step 2. If $\|\mathbf{f}^{\nu+1} - \mathbf{f}^{\nu}\|_{\infty} < 10^{-2}$, end the algorithm. Else, go to the next step.

step 3. Update $\tilde{\mathbf{d}}^{\nu+1}$ by (37).

step 4. Update $\tilde{\tilde{\mathbf{d}}}^{\nu+1}$ by (38).

step 5. Update $\Theta^{\nu+1}$ by equation (39).

step 6. Calculate $\mathbf{w}^{l, \nu+1}$ by equation (35)

step 7. $\tilde{\mathbf{b}}^{\nu+1} = \tilde{\mathbf{b}}^{\nu} + k * \mathbf{f}^{\nu+1} - \mathbf{g} - \tilde{\mathbf{d}}^{\nu+1}.$

step 8. $\tilde{\tilde{\mathbf{b}}}^{\nu+1} = \tilde{\tilde{\mathbf{b}}}^{\nu} + \nabla \mathbf{f}^{\nu+1} - \tilde{\tilde{\mathbf{d}}}^{\nu+1}.$

step 9. $\nu = \nu + 1$, go to step 1.

5.2 Experimental results

In order to compare with other methods, the peak signal-to-noise ratio,

$$\text{PSNR} = 10 \log_{10} \frac{3m_1 m_2}{\|\mathbf{f} - \mathbf{I}\|_2^2},$$

is taken to measure the improvement of image quality, where $\mathbf{I}$ is the $m_1 \times m_2$ original image and $\mathbf{f}$ is the corresponding recovered result. The original images of size 256×256 for our synthetic experiments are shown in Fig. 2.

Some parameters which occurred in our models are selected as follows: the regularization parameter $\lambda = 5$, the penalty parameters $\eta = 100, \eta_1 = 250, \eta_2 = 100$;

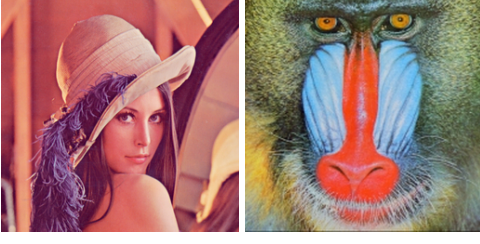


Fig. 2 The original images $\mathbf{I}$ for the synthetic experiments.

the initial values $(\sigma_1^2)^0 = 10^{-4}$ for salt-and-pepper noise and $(\sigma_1^2)^0 = 5 \times 10^{-2}$ for random-valued noise. A point worth mentioning is that we do not need to change any parameter values in all the following experiments according to different levels of noise since the parameters σ_l^2 , which indicate the levels of noise, can balance the behaviors between denoising and deblurring.

Recently, the FTVd algorithm [16] is regarded as one of the most efficient methods to solve TVL¹ problem

$$\min_{\mathbf{f}} \text{VTV}(\mathbf{f}) + \mu \|k * \mathbf{f} - \mathbf{g}\|_1.$$

And we will compare our method with other models using FTVd.

Generally speaking, images with heavy noise can be hardly restored because additive noise results in a great loss of information. But for heavy impulse noise, good reconstruction would be expected. Impulse noise with high level makes the image looks heavily blurry and noise-contaminated, however there are still some pixels with only blur. As mentioned earlier, the two adaptively weighting vectorial functions $\mathbf{w}^l$ in our method can detect them and we can still obtain a good reconstruction.

In order to better interpret the role of the weighting function $\mathbf{w}^l$, we propose an experiment in Fig.3: adding same level of salt-and-pepper noise to the same blurred image with different spatial distribution. In the first case, the noise is added such that there is no useful information in the right part of the image (see Fig.3(a)), while in the second case, the noise is added randomly to the image (see Fig.3(f)). The reconstructions with the existing variational method such as TVL¹ and our modified method TVAWL¹ are shown in the second and third column, respectively. Note that the proposed model could not do better than TVL¹ in the left half image for the first case. This is because one could choose different μ for the two parts of the image in TVL¹ to get the similar results as ours. However, it would fail in the second case because the spatial distribution of noise is random. The advantage of our models is evident in the second case, since even in this case, noise could be still detected by the weighting functions $\mathbf{w}^l$. The estimated

noise free image $\tilde{\mathbf{g}} = \mathbf{g}\mathbf{w}^1$ and noise $\tilde{\mathbf{g}} = \mathbf{g}\mathbf{w}^2$ in our models are displayed in the last two columns.

Now let us compare the proposed method with the L¹ fidelity term based methods [12, 14–18]. We test our algorithm and FTVd with a variety of blurs and different levels of impulse noise.

Fig. 4 shows the comparison of the reconstructed results obtained by TVL¹ with FTVd and our method for the case of salt-and-pepper noise. As is well known that the parameter μ is very important for TVL¹ model, so we test some values (including the suggested value in FTVds) and choose the result with the highest PSNR for comparison. In this figure, the Gaussian blurred ($\sigma = 2.0$) images with different levels of salt-and-pepper noise are shown in the first column, while the levels of noise are $r_1 + r_2 = 30\%, 60\%, 80\%, 90\%$, respectively, and the results obtained by FTVd and ours could be found in the last three columns. As can be seen from Fig. 4, there is no significant difference between the three methods in the case of low level noise (e.g. 30%). When the noise level is increased, our approach produce much better results than TVL¹. The corresponding PSNR values (the measure of image quality) and CPU time are summarized in Table 1. High PSNR values and similar CPU time can be seen for the proposed model.

For salt-and-pepper noise, the TVAWL¹ model seems to have a better performance than TVAWL². In general, it is more difficult to detect random-valued noise than salt-and-pepper noise. Our method can recover images from high-density random-valued noise and estimate the noise. Results of restoring out-of-focus blurred image (radius=7) with random-valued noise are illustrated in Fig.5. We showed the estimated noise free images and the estimated noise in the last two columns of Fig.5.

The method discussed in this paper can also be applied to image denoising and inpainting. The only difference with deblurring is that the blur kernel k in denoising/inpainting should be the delta function.

Fig.6 shows some results of denoising with TVAWL¹. Here we compare our results with those got from median filter. The corresponding PSNR values and CPU time are shown in Table 1.

A result of inpainting with the proposed method is illustrated in Fig.7. Unlike other inpainting methods, when the inpainting area satisfy certain conditions, such as the inpainting areas are corrupted by an impulsive process, we do not need to set the inpainting mask for the proposed method because the mask can be estimated by the weighting functions.

Recently, Cai *et al.* [17, 18] proposed a two-phase method to deblur grey scale image with impulse noise

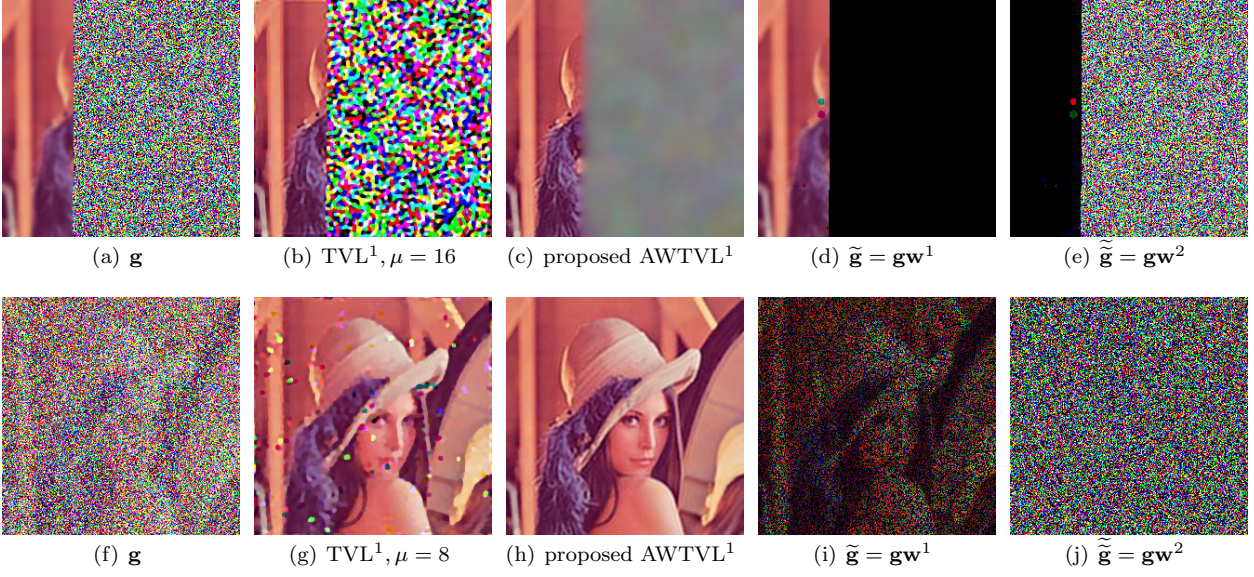


Fig. 3 Explaining the role of the weighting functions $\mathbf{w}^1, \mathbf{w}^2$. (a),(f) blurred images corrupted by the same noise with different spatial distributions; (b),(g) restored by FTVD; (c),(h) restored by the proposed method; (d),(i) the estimated noise free image; (e),(j) the estimated noise.

Table 1 PSNR (dB) values and CPU time (s) in the experiments (Fig. 4~Fig. 6).

Observed	PSNR (dB)				CPU time (s)			
	Median filter	TVL ¹	TVAWL ²	TVAWL ¹	Median filter	TVL ¹	TVAWL ²	TVAWL ¹
Fig.4(a)	-	28.84	29.82	30.21	-	15.66	14.77	16.89
Fig.4(e)	-	27.44	29.75	30.01	-	17.97	16.81	17.87
Fig.4(i)	-	23.55	28.47	29.39	-	27.16	24.90	25.20
Fig.4(m)	-	19.87	23.57	27.92	-	43.53	26.23	28.22
Fig.5(a)	-	25.13	-	28.61	-	26.47	-	38.87
Fig.5(f)	-	21.86	-	26.96	-	31.83	-	43.31
Fig.6(a)	30.44	-	-	41.28	0.095	-	-	3.12
Fig.6(b)	23.61	-	-	35.66	0.34	-	-	4.38
Fig.6(c)	18.01	-	-	29.09	0.42	-	-	5.41
Fig.6(d)	6.77	-	-	21.65	0.60	-	-	5.86

and obtained good results. With our notations, the two-phase method actually minimizes the following functional

$$\|k * f - g\|_{1,\chi} + \beta R(f),$$

where $R(f)$ is a regularization term, $\beta > 0$ is a parameter and the characteristic function χ is estimated by the median-type filters, which is determined by

$$\chi(x) = \begin{cases} 0, & (k * f)(x) \text{ is degraded by noise,} \\ 1, & \text{else.} \end{cases}$$

Now we compare experimental results of the two-phase method and our models. It appears that the presented algorithms have the similar results to the two-phase method for salt-and-pepper noise. An interpretation is given as follows: in TVAWL¹ model (24), if we fix $\mathbf{w}^{1,\nu} =$

χ , $\mathbf{w}^{2,\nu} = 0$ and let all the parameters α_l, σ_l^2 be constants, then it is equivalent to the two-phase method; in fact, $\mathbf{w}^{1,\nu}$ can be considered as a smooth version of χ , and both $\mathbf{w}^{l,\nu}$ and χ play the same role of adjusting the fidelity term, but in different ways— χ eliminates the likely noisy pixels from the fidelity term and $\mathbf{w}^{l,\nu}$ gives different weights to the noisy and only blurred pixels; moreover, salt-and-pepper noise is easy to detect and thus both two methods can produce good reconstructions. However, there is no good detector for random-valued noise with high noise ratio. Compared with the two-phase method, the main superiority of our approach is that it can provide much better restorations for the case of high density random-valued noise. This is because the noise detection and deblurring in the proposed algorithm are alternately implemented and the

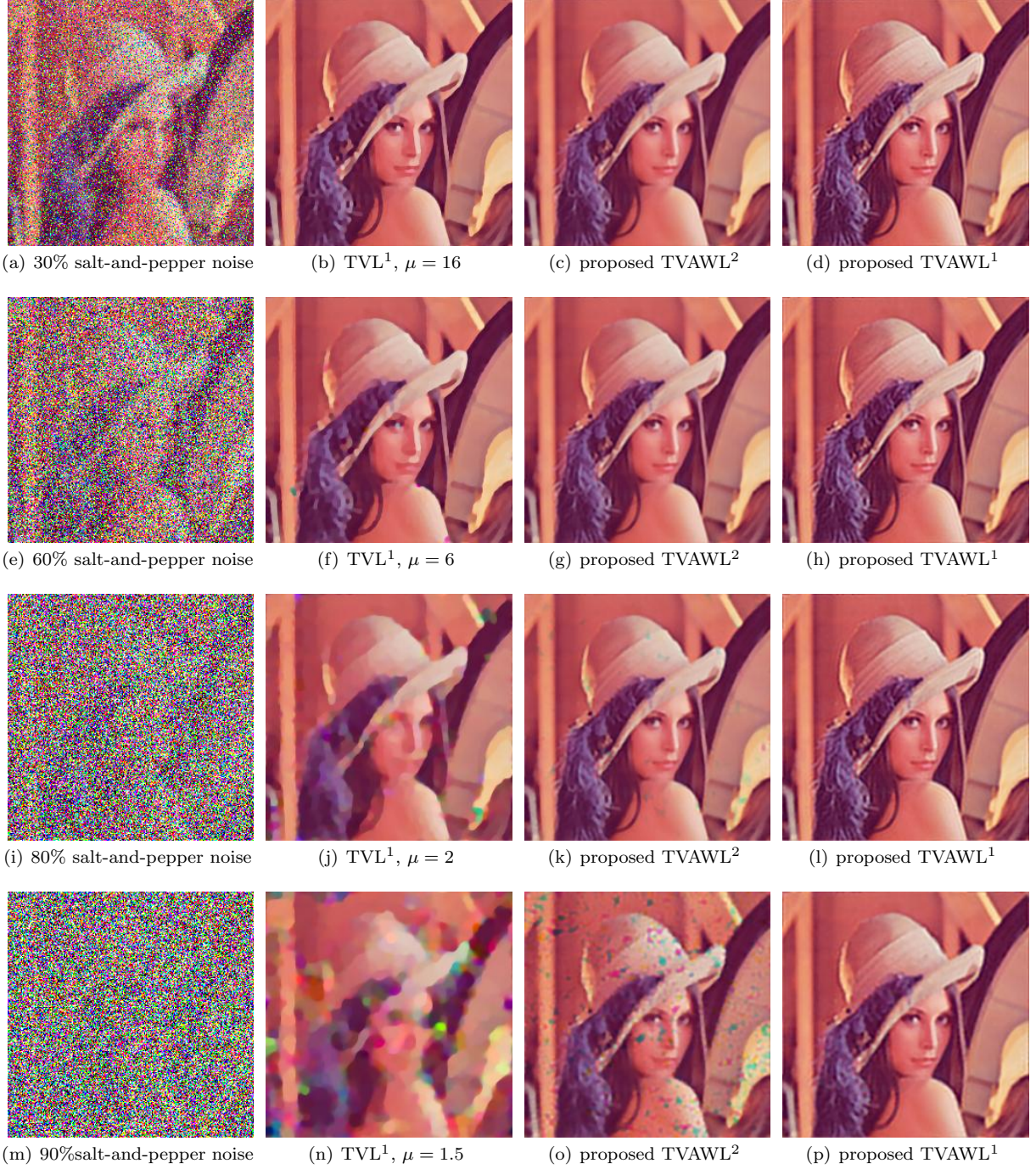


Fig. 4 Comparison. First column: $\mathbf{g}$, blurred (Gaussian, $\sigma = 2.0$) images corrupted by different levels ($r_1 + r_2 = 30\%, 60\%, 80\%, 90\%$) salt-and-pepper noise. Second column: restored by TVL^1 with FTVd. Third column: restorations with TVAWL^2 . Fourth column: restorations with TVAWL^1

random-valued noise can be better identified. So, in the following, more attentions will be paid to the random-valued noise.

To be fair, we test our algorithms with grey scale images and the ending conditions of the CG solver in the two-phase method and our algorithms are all set to

$$\text{be } \frac{\|f^{\nu_1+1} - f^{\nu_1}\|_2}{\|f^{\nu_1+1}\|_2} < 10^{-5}.$$

In the experiments, we first add some out of focus blur with radius 7 to the “lenna” grey scale image, which is then further corrupted by adding different levels random-valued noise.

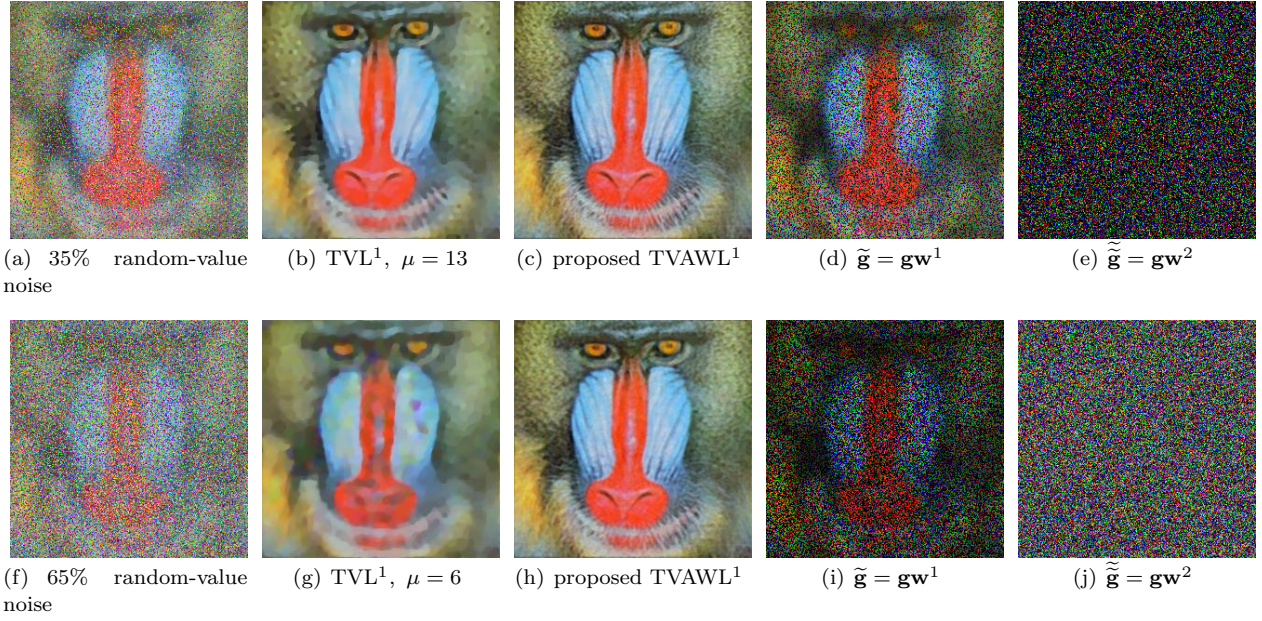


Fig. 5 Comparison. (a),(f) the blurred (out-of-focus blur, radius=7) image with random-valued noise, $r = 35\%, 65\%$; (b),(g) restoration with FTVd; (c),(h) restoration with the proposed method; (d),(i) the estimated noise free image $\tilde{\mathbf{g}} = \mathbf{g}\mathbf{w}^1$; (e),(j) the estimated noise $\tilde{\mathbf{g}} = \mathbf{g}\mathbf{w}^2$.

The results of recovering blurred images with random-valued noise by the two-phase method [18] and our algorithm are shown in Fig. 8 and the PSNR values and corresponding CPU time are summarized in Table 2. We can see from the figure and table that our method is much better than the two-phase method. As is reported in [17,18], the two-phase method can not work efficiently if the ratio of random-valued noise exceeds 55%, but our method can still give good results even if the ratio is as high as 75%.

Remark: In this experiment, the ending condition of CG in our method is more rigorous than those of the previous ones, thus it is more CPU time-consuming than the previous implementations.

From all the above experiments, one can conclude that the proposed method is superior than other models in recovering images from impulse noise.

6 Conclusion and Discussion

We proposed a novel approach of reconstructing color images with blur and impulse noise. The main idea is that at each pixel, the impulse noise is viewed as a realization of the sum of two Gaussian or two two-sided exponential PDFs, then we formulate a functional which contains two adaptively weighting functions and some statistical control parameters. Compared with the existing variational methods, the introduced weighting functions can identify impulse noise more efficiently,

and the algorithm could locally adjust denoising and deblurring by the introduced parameters of PDFs and these weighting functions. Our experimental results had shown that the quality of the restored images by the proposed method are better than the existing models such as TVL¹ (both under random-valued noise and salt-and-pepper noise) and two-phase method [18] (under random-valued noise). Our method can give good restorations for random-valued noise with noise ratio as high as 75%, which can not be well handled by the existing variational methods.

In section 5 we have shown the connections between the proposed algorithms and the two-phase method. In fact, the model in this study is an extension of the traditional ones. For example, if $\alpha_1 = \alpha_2 = 0.5$ and $\sigma_1^2 = \sigma_2^2$ are fixed constants, then the weighting functions $\mathbf{w}^1 = \mathbf{w}^2 = 0.5$. In this case, the cost functional in our method, after ignoring some constant terms, would reduce to the existing ones.

Our method is based on TV regularization and EM algorithm, and it is well known that EM algorithm has a local convergence. Thus, the proposed algorithms partly depend on the initial value Θ^0 . It is a good idea to set the initial value $\mathbf{w}^{l,0}$ to the results of the first phase in two-phase method [17–19]. In addition, we choose an approximation to $p(y)$ in E.q. (2) to avoid implemental difficulties and obtain some good experimental result. However, a natural alternative is to consider (2) directly, and how to address this difficulty is our future work directions.

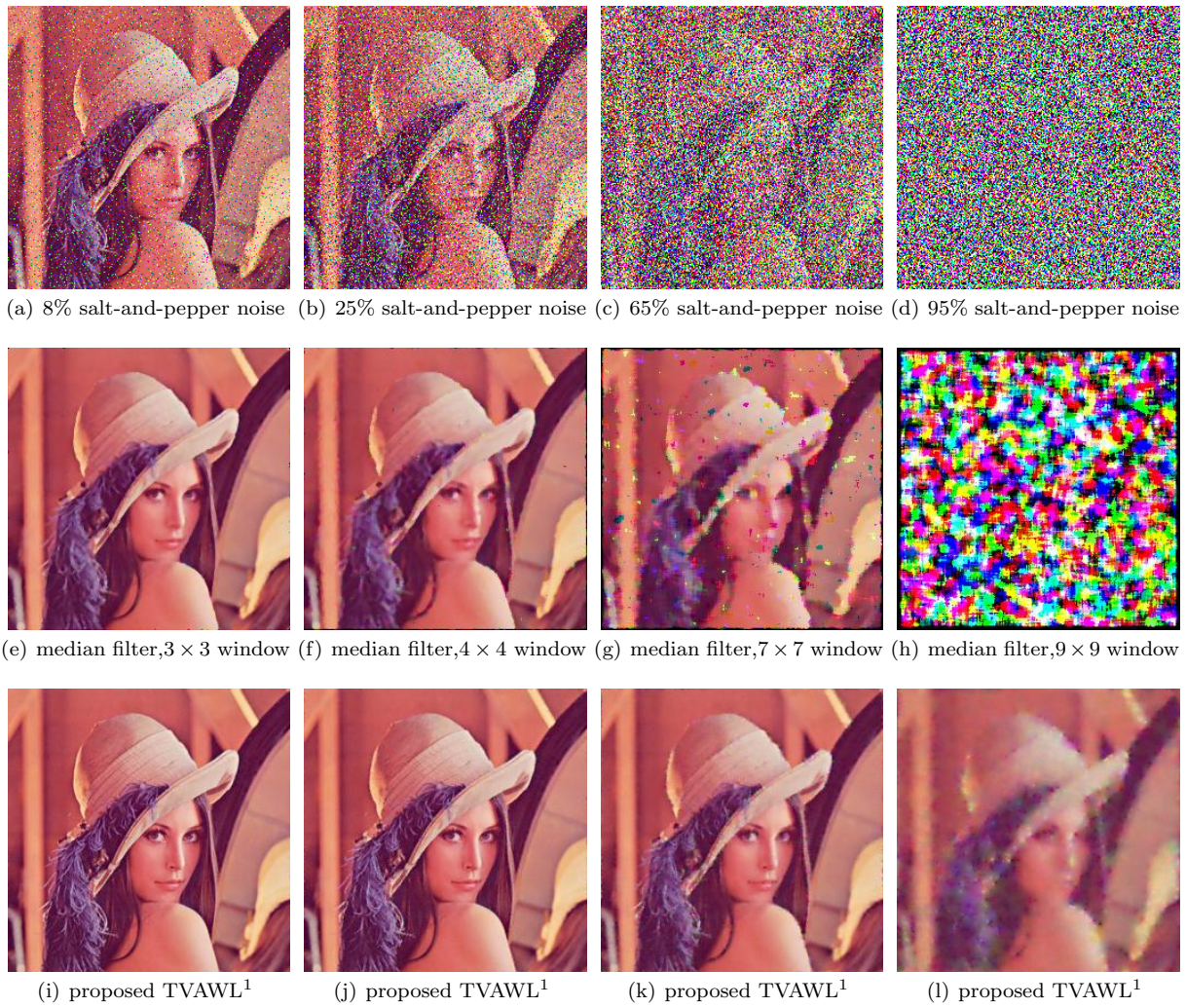


Fig. 6 Some results of denoising. First row: noisy image. Second row: denoising with MATLAB function *medfilt2* channel by channel. Third row: denoising with the proposed method.



Fig. 7 Automatic inpainting.

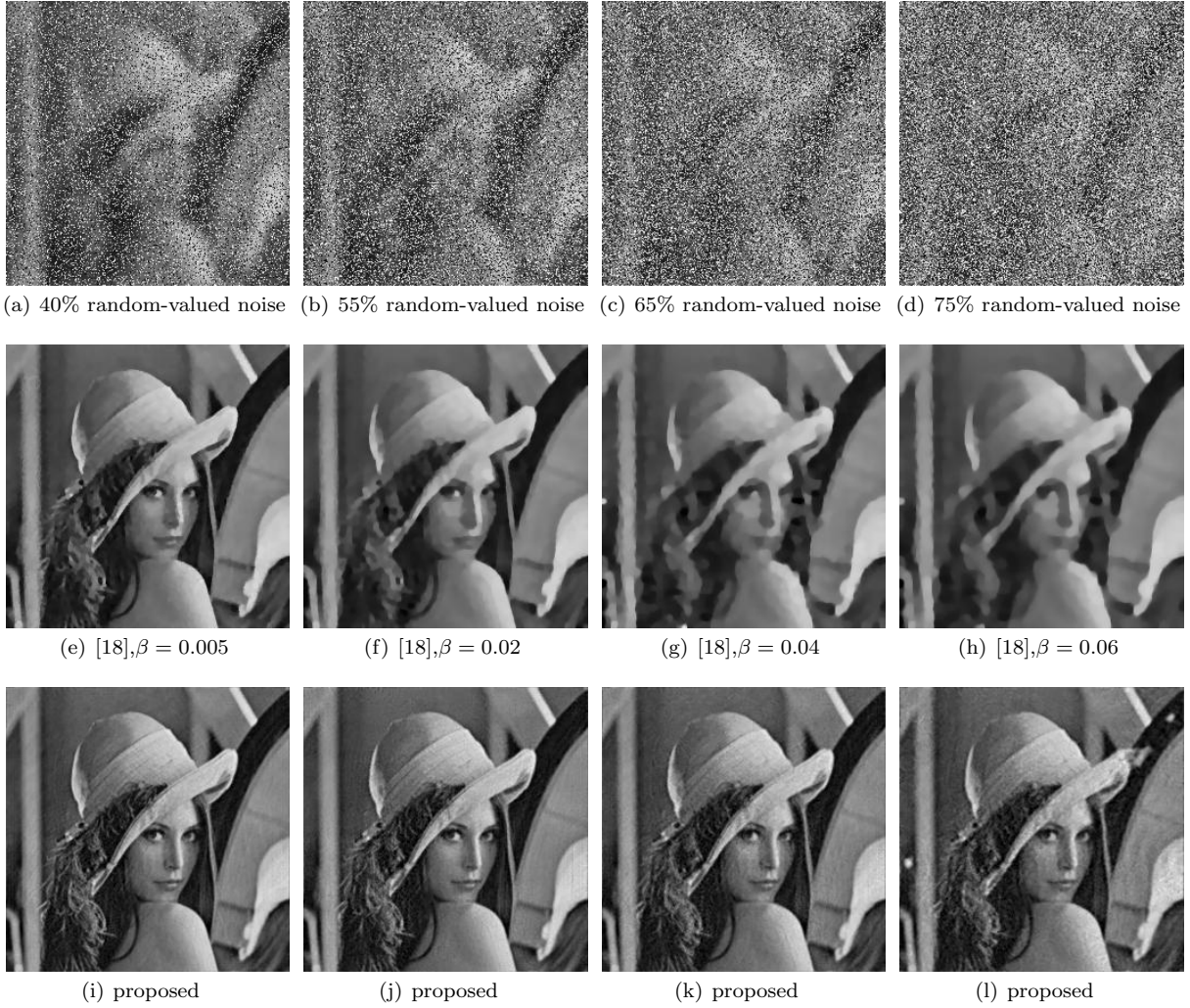


Fig. 8 Some restorations of two-phase method [18] and the proposed model under random-valued noise. First row: noisy images. Second row: restorations with two-phase method. Third row: restorations with TVAWL².

Table 2 Comparing with PSNR values and CPU time of two-phase method [18] and TVAWL² under random-valued noise (Fig. 8).

Noise density r	PSNR (dB)				CPU time (s)			
	40%	55%	65%	75%	40%	55%	65%	75%
Two-phase[18]	29.37	27.14	23.83	23.67	64.3	74.7	117.2	132.2
TVAWL ²	32.32	30.34	29.11	25.48	39.3	46.9	65.7	95.9

The proposed method can be directly extended to blind deblurring for impulse noise and also can be extended to nonlocal version. We have noticed that the weighting functions $\mathbf{w}^{l,\nu}$ play a role of classifying the impulse noise according to the different variances σ_l^2 , which is similar to the region based image segmentation in which the pixels are clustered according to the different means. Thus with the proposed framework, many better image segmentation method such as globally convex segmentation [40] can be employed in our method. Finally, we mention that the proposed variational framework can be applied to the regularization

term. If we assume $|\nabla \mathbf{f}|$ at each x obeys a distribution which can be approximated by the mixture model, then one can get a regularizer which has the similar form as Mumford-Shah functional regularization. We do not plan to discuss much about these, which are left for further study.

Acknowledgements We would like to thank Dr. Jian-Feng Cai for his kind offer the source codes of [17, 18] and the authors of [16] for their open source codes FTVd, we also thank the reviewers for their valuable comments. The research has been supported by National Science Foundation of China (NSFC, No. 10531040).

Appendix

A. 1:

Considering the following random event,

$$N = \begin{cases} Y_1, & \text{when } B \text{ occurs,} \\ Y, & \text{when } C \text{ occurs,} \end{cases}$$

where B and C are mutually exclusive events. According to the equation (1), $P(B) = 1 - r$, $P(C) = r$, and we get

$$\begin{aligned} P(N) &= P(N \cap B) + P(N \cap C) \\ &= P(B)P(N|B) + P(C)P(N|C) \\ &= (1 - r)P(N|B) + rP(N|C), \end{aligned}$$

Let a be the realization of a random variable $\mathcal{Y}_3$ with PDF $p_3(y)$, then

$$P\{\mathcal{N} < y\} = (1 - r)P\{\mathcal{Y}_1 < y\} + rP\{\mathcal{Y}_2 + \mathcal{Y}_3 < y\}. \quad (40)$$

Note that $\mathcal{Y}_2$ and $\mathcal{Y}_3$ are independent, thus following (40), we have

$$p(y) = (1 - r)\delta(y) + r(p_2 * p_3)(y).$$

Moreover,

$$p_3(y) = \begin{cases} 1, & y \in [0, 1], \\ 0, & \text{else,} \end{cases}$$

and thus

$$(p_2 * p_3)(y) = \int_{-\infty}^{\infty} p_2(z)p_3(y - z) dz = \int_{y-1}^y p_2(z) dz.$$

A. 2:

Theorem 1 For every ν , if $H(\Theta^{\nu+1}; \Theta^\nu) \geq H(\Theta^\nu; \Theta^\nu)$, then $l(\Theta^{\nu+1}; \mathbf{n}) \geq l(\Theta^\nu; \mathbf{n})$.

Proof

$$\begin{aligned} &\because J(\Theta^\nu; \Theta^\nu) - J(\Theta^{\nu+1}; \Theta^\nu) \\ &= -\sum_{\mathbf{c}} (l(\Theta^{\nu+1}; \mathbf{c}|\mathbf{n}) - l(\Theta^\nu; \mathbf{c}|\mathbf{n}))p(\mathbf{c}|\mathbf{n}; \Theta^\nu) \\ &= -\sum_{\mathbf{c}} \ln\left(\frac{p(\mathbf{c}|\mathbf{n}; \Theta^{\nu+1})}{p(\mathbf{c}|\mathbf{n}; \Theta^\nu)}\right)p(\mathbf{c}|\mathbf{n}; \Theta^\nu) \\ &\geq -\ln \sum_{\mathbf{c}} p(\mathbf{c}|\mathbf{n}; \Theta^{\nu+1}) = -\ln 1 = 0. \end{aligned} \quad (41)$$

Here the inequality is given by the convexity of function $-\ln x$.

$$\begin{aligned} &\therefore l(\Theta^{\nu+1}; \mathbf{n}) - l(\Theta^\nu; \mathbf{n}) \\ &= (H(\Theta^{\nu+1}; \Theta^\nu) - H(\Theta^\nu; \Theta^\nu)) \\ &\quad + (J(\Theta^\nu; \Theta^\nu) - J(\Theta^{\nu+1}; \Theta^\nu)) \geq 0. \end{aligned} \quad (42)$$

□

A. 3:

Proposition 3 if $\tilde{\mathbf{d}}^{\nu+1} =$

$$\arg \min_{\tilde{\mathbf{d}}} \left\{ E(\mathbf{f}^{\nu+1}, \tilde{\mathbf{d}}, \tilde{\mathbf{d}}^\nu, \Theta^\nu) = \sum_{l=1}^2 \frac{\|\tilde{\mathbf{d}}\|_{1, \mathbf{w}^{l, \nu}}}{(\sigma_l^2)^\nu} + \frac{\eta_1}{2} \sum_{l=1}^2 \frac{\|\tilde{\mathbf{d}} - (k * \mathbf{f}^{\nu+1} - \mathbf{g}) - \tilde{\mathbf{b}}^\nu\|_{2, \mathbf{w}^{l, \nu}}^2}{(\sigma_l^2)^\nu} \right\},$$

then

$$\tilde{\mathbf{d}}^{\nu+1} = \text{shrink} \circ \left(k * \mathbf{f}^{\nu+1} - \mathbf{g} + \tilde{\mathbf{b}}^\nu, \frac{1}{\eta_1} \right).$$

Proof : Suppose $\tilde{\mathbf{d}} = (\tilde{d}_0, \tilde{d}_1, \tilde{d}_2)$. Let $\mathbf{z} = \sum_{l=1}^2 \frac{\mathbf{w}^{l, \nu}}{(\sigma_l^2)^\nu}$, and $\mathbf{y} = k * \mathbf{f}^{\nu+1} - \mathbf{g} + \tilde{\mathbf{b}}^\nu$, with $\mathbf{z} = (z_0, z_1, z_2)$, $\mathbf{y} = (y_0, y_1, y_2)$, then

$$\begin{aligned} E(\tilde{\mathbf{d}}) &= \|\tilde{\mathbf{d}}\|_{1, \mathbf{z}} + \frac{\eta_1}{2} \|\tilde{\mathbf{d}} - \mathbf{y}\|_{2, \mathbf{z}}^2 \\ &= \sum_{\tau=0}^2 \int z_\tau |\tilde{d}_\tau| dx + \frac{\eta_1}{2} \sum_{\tau=0}^2 \int z_\tau (\tilde{d}_\tau - y_\tau)^2 dx, \end{aligned}$$

$$\frac{\delta E}{\delta \tilde{d}_\tau} = \frac{z_\tau \tilde{d}_\tau}{|\tilde{d}_\tau|} + \eta_1 z_\tau (\tilde{d}_\tau - y_\tau).$$

A minimizer of E satisfies $\frac{\delta E}{\delta \tilde{d}_\tau} = 0$, please note $z_\tau > 0$, which indicates

$$\left(\frac{1}{|\tilde{d}_\tau^{\nu+1}|} + \eta_1 \right) \tilde{d}_\tau^{\nu+1} = \eta_1 y_\tau. \quad (43)$$

If $|y_\tau| \geq \frac{1}{\eta_1}$, then taking $|\cdot|$ for two sides of (43), we get

$$|\tilde{d}_\tau^{\nu+1}| = |y_\tau| - \frac{1}{\eta_1}.$$

Take the above expression to (43), and immediately yield

$$\tilde{d}_\tau^{\nu+1} = \frac{y_\tau}{|y_\tau|} \left(|y_\tau| - \frac{1}{\eta_1} \right).$$

Else, $|y_\tau| < \frac{1}{\eta_1}$,

$$\begin{aligned} E(\tilde{\mathbf{d}}) &= \sum_{\tau=0}^2 \int z_\tau |\tilde{d}_\tau| dx + \frac{\eta_1}{2} \sum_{\tau=0}^2 \int z_\tau (\tilde{d}_\tau^2 + y_\tau^2 - 2\tilde{d}_\tau y_\tau) dx \\ &\geq \sum_{\tau=0}^2 \int z_\tau |\tilde{d}_\tau| dx + \frac{\eta_1}{2} \sum_{\tau=0}^2 \int z_\tau (\tilde{d}_\tau^2 + y_\tau^2 - 2|\tilde{d}_\tau| |y_\tau|) dx \\ &\geq \sum_{\tau=0}^2 \int z_\tau |\tilde{d}_\tau| dx + \frac{\eta_1}{2} \sum_{\tau=0}^2 \int z_\tau (\tilde{d}_\tau^2 + y_\tau^2 - \frac{2}{\eta_1} |\tilde{d}_\tau|) dx \\ &= \frac{\eta_1}{2} \sum_{\tau=0}^2 \int z_\tau (\tilde{d}_\tau^2 + y_\tau^2) dx, \end{aligned}$$

thus $\tilde{d}_\tau^{\nu+1} = 0$.

In summary, $\tilde{d}_\tau^{\nu+1} = \frac{y_\tau}{|y_\tau|} \max\{|y_\tau| - \frac{1}{\eta_1}, 0\}$. □

References

1. R. Lagendijk, A. Tekalp, and J. Biemond, "Maximum likelihood image and blur identification: a unifying approach," *Opt. Eng.*, vol. 29, no. 5, pp. 422-435, 1990.
2. A. Jalobeanu, L. Blanc-Feraud, and J. Zerubia, "An adaptive Gaussian model for satellite image deblurring," *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1469-1478, 2005.
3. Q. Shan, J. Jia, and A. Agarwala, "High-quality Motion Deblurring from a Single Image," *ACM Trans. Graph. (SIGGRAPH)*, vol. 27, pp. 1-10, 2008.
4. L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, pp. 259-268, 1992.
5. L. Rudin, and S. Osher, "Total variation based image restoration with free local constraints," *IEEE ICIP 1994*, vol. 1, pp. 31-35, 1994.
6. W. Zhao, and A. Pope, "Image restoration under significant additive noise," *IEEE Signal. Proc. Lett.*, vol. 14, no. 6, pp. 401-404, 2007.
7. C. Vogel, and M. Oman, "Iterative methods for total variation denoising," *SIAM J. Sci. Comput.*, vol. 17, no. 1, pp. 227-238, Jan. 1996.
8. C. Vogel, and M. Oman, "Fast, robust total variation-based reconstruction of noisy, blurred images," *IEEE Trans. Image Process.*, vol. 7, no. 6, pp. 813-824, 1998.
9. C. Vogel, "Computational methods for inverse problems," *SIAM*, pp. 53-54, 2002.
10. H. Eng and K. Ma, Noise adaptive soft-switching median filter, *IEEE Trans. Image Process.*, vol. 10, pp. 242-251, 2001.
11. T. Chen and H. Wu, Space variant median filters for the restoration of impulse noise corrupted images, *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, Vol. 48, pp. 784-789, 2001.
12. M. Nikolova, "A variational approach to remove outliers and impulse noise," *J. Math. Imag. Vis.*, vol. 20, pp. 99-120, 2004.
13. R. Chan, C. Ho, and M. Nikolova, "Salt-and-pepper noise removal by median-type noise detectors and edge-preserving regularization," *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1479-1485, Oct. 2005.
14. L. Bar, N. Sochen, and N. Kiryati, "Image deblurring in the presence of impulsive noise," *Int. J. Comput. Vis.*, vol. 70, pp. 279-298, 2006.
15. L. Bar, A. Brook, N. Sochen, and N. Kiryati, "Deblurring of color images corrupted by impulsive noise," *IEEE Trans. Image Process.*, vol. 16, no. 4, pp. 1101-1111, 2007.
16. J. Yang, Y. Zhang, and W. Yin, "An Efficient TVL1 Algorithm for Deblurring Multichannel Images Corrupted by Impulsive Noise," *SIAM J. Sci. Comput.*, vol. 31, no. 4, pp. 2842-2865, 2009.
17. J. Cai, R. Chan, and M. Nikolova, "Two-phase approach for deblurring images corrupted by impulse plus Gaussian noise," *Inv. Prob. Imag.*, Vol. 2, pp. 187-204, 2008.
18. J. Cai, R. Chan, and M. Nikolova, "Fast two-phase image deblurring under impulse noise," *J. Math. Imaging Vis.*, DOI 10.1007/s10851-009-0169-7, 2009.
19. Y. Huang, M. Ng, and Y. Wen, "Fast image restoration methods for impulse and gaussian noises removal," *IEEE Sig. Process. Lett.*, Vol. 16, pp. 457-460, 2009.
20. J. Liu, Z. Huan, H. Huang, and H. Zhang, "An adaptive method for recovering image from mixed noisy data," *Int. J. Comput. Vis.*, vol. 85, no. 2, pp. 182-191, 2009.
21. T. Chan, and J. Shen, "Image processing and analysis-variational, PDE, wavelet, and stochastic methods," *SIAM Publisher*, pp. 207-243, 2005.
22. J. Bilmes, "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models," <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.28.613>, 1998.
23. G. J. McLachlan, and T. Krishnan, "The EM algorithm and extensions," *John Wiley & Sons Inc.*, 2007.
24. R. Redner, and H. Walker, "Mixture Densities, Maximum Likelihood and the EM Algorithm," *SIAM Rev.*, vol. 26, no. 2, pp. 195-239, Apr., 1984.
25. X. Bresson, T. F. Chan, Fast dual minimization of the vectorial total variation norm and applications to color image processing, *Inverse Problems and Imaging*, vol. 2, no. 4, pp. 455-484, 2008.
26. P. Blomgren, and T. F. Chan, "Color TV: total variation methods for restoration of vector-valued images," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 304-309, 1998.
27. D. Tschumperle, "PDEs based regularization of multivalued images and applications," Ph.D. dissertation, Univ. Nice-C Sophia Antipolis, Sophia Antipolis, France, 2002.
28. D. Mumford, and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Comm. Pure Appl. Math.*, vol. 42, pp. 577-685, 1989.
29. G. Gilboa, and S. Osher, "Nonlocal linear image regularization and supervised segmentation," *Multiscale Model. Simul.*, vol. 6, pp. 595-630, 2007.
30. G. Gilboa, and S. Osher, "Nonlocal operators with applications to image processing," *Multiscale Model. Simul.*, vol. 7, pp. 1005-1028, 2008.
31. J. Darbon, M. Sigelle, A fast and exact algorithm for total variation minimization, *Pattern Recognition and Image Analysis*, vol. 3522, pp. 351-359, 2005.
32. T. Chan, G. Golub, and P. Mulet, "A nonlinear primal-dual method for total variation-based image restoration," *SIAM J. Sci. Comput.*, vol. 20, pp. 1964-1977, 1999.
33. A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imaging Vis.*, vol. 20, pp. 89-97, 2004.
34. T. Goldstein, and S. Osher, "The split Bregman method for L1 regularized problems," *UCLA CAM Report* 08-29, 2008.
35. X. Tai, and C. Wu, "Augmented Lagrangian method, dual methods and split Bregman iteration for ROF model," *UCLA CAM Report* 09-05, 2009.
36. E. Esser, "Applications of Lagrangian-based alternating direction methods and connections to split Bregman," *UCLA CAM Report* 09-31, 2009.
37. S. Setzer, "Operator splittings, Bregman methods and frame shrinkage in image processing," *Int. J. Comput. Vis.*, 2009, accepted.
38. J. Douglas, and H. H. Rachford, "On the numerical solution of heat conduction problems in two and three space variables," *T. Am. Math. Soc.*, vol. 82, no. 2, pp. 421-439, 1956.
39. K. Michael, H. Chan, and W. Tang, "A fast algorithm for deblurring models with Neumann boundary conditions," *SIAM J. Sci. Comput.*, vol. 21, no. 3, pp. 851-866, 1999.
40. X. Bresson, S. Esedoglu, P. Vanderghynst, J. Thiran, and S. Osher, "Fast global minimization of the active contour/snake model," *J. Math. Imag. Vis.*, vol. 28, pp. 151-167, 2007.