Image Restoration Algorithm Based on Double l_0-Regularization and ALM

Su Xiao

Abstract—**Although many advances have been made in the research on image restoration (IR), continuous improvement in the quality of restored clear images is a constant goal. The focus of our study is to fuse multiple priors, build a regularization model for IR and solve the model efficiently. Considering the prominent abilities of sparse priors and the l_0-norm, a gradient-sparsity prior and a transform-sparsity prior are combined to build a double l_0-regularization model for IR. For the IR model built, we propose an efficient approach derived from the augmented Lagrangian method (ALM). First, by variable splitting, the built model is equivalently converted into a minimization problem with constraints. Second, the ALM is imposed on the new constrained problem, generating several independent subproblems. Third, the subproblems are addressed individually with suitable methods to obtain their closed-form optimal solutions. After proposing the IR algorithm, we analyze it in terms of the local minimizer and convergence. In quantitative and visual experiments, the proposed algorithm is used to handle degraded gray and color images to show its advantages over some excellent algorithms and its own good properties.**

Index Terms—**image restoration, gradient-sparsity priors, transform-sparsity priors, double l_0-regularization model, augmented Lagrangian method.**

I. Introduction

FROM a mathematical point of view, IR is a largescale problem that is complicated and ill-posed. The target of IR is to obtain clear images from corresponding degraded images, so we should first determine what image degradation is. Let $y\in\mathbb{R}^N$ denote a degraded image, $x\in\mathbb{R}^M$ denote a clear image, $n\in\mathbb{R}^N$ denote Gaussian noise, and $B\in\mathbb{R}^{N\times M}$ denote an operator. Then, the degradation is cast as a linear system $y=Bx+n$. It is worth stating that the acquired digital images are originally in the form of matrices, but they are reassembled into vectors for the convenience of image processing. Many unfavorable factors, such as defocus, motions and inherent defects, often degrade image quality, which sets obstacles for applications [1], [2]. Therefore, research on IR has important theoretical significance and application value. Previous research tells us that the success of IR depends on the useful image priors. Among the prior models for IR, the most modern and successful ones are sparse prior-based regularization models, including gradient-sparsity-based regularization models and transform-sparsity-based regularization models. The former refers to the total variation (TV) model and its generalizations, and the latter refers to models that employ the l_p-norm of sparse representations as the regularization terms.

Manuscript received September 16, 2022; revised January 20, 2023. This work is supported by the Natural Science Research Project of Anhui Provincial Department of Education (Grant No. 2022AH050370).

Su Xiao is a professor of School of Computer Science and Technology, Huaibei Normal University, Huaibei, 235000, China (e-mail: csxiaos@139.com).

A. Related Work

To fix the defects of Tikhonov regularization, Rudin et al. [3] put forward the TV-model for IR. Since then, it has gradually become famous and is now one of the most successful image processing tools in the industry. Because the TV-norm therein is not differentiable, Rudin et al. suggest using a time-marching method to solve the TV-model. However, the stability constraint makes the computation converge slowly. Recently, Kamilov [4] used a parallel proximal method to solve the TV-based IR problem, obtaining a convergence speed equivalent to the fast proximal gradient method. Some splitting-technology-based methods provide possibilities for quickly solving TV-model-based IR problems. Given that the derivative space contributes to the high IR quality, Ren et al. [5] project the TV-model into the new space and use the ADMM to address the resulting model. To remove artifacts and preserve edges, Adam et al. [6] applied the second-order TV and overlapping group sparseness to IR. Li et al. [7] established an improved fractional-order TV-model to reconstruct more image details and prevent staircase artifacts. Based on the TGV, Zhang et al. [8] built a nonconvex and nonsmooth model for IR. Kongskov et al. [9] incorporated directional information into the TGV model to construct a better directional TGV model for IR. Wang et al. [10] presented a novel nonlocal TV technology based on structural similarity and established a regularization model via this technology to recover image patches.

Currently, the application of transform-sparsity priors also attracts much attention in the industry, resulting in many research results. To speed up IR, Xue et al. [11] linearly parameterized it using multidimensional filtering and wavelet transforms. For the key sparse linear coefficients, they established an unbiased estimation model to compute them. Based on a wavelet frame, Cai et al. [12] constructed a new IR model that characterizes images as piecewise functionals with singularities. By penalizing the l_2-norm and l_1-norm of framelet coefficients, the images and their singularities are estimated simultaneously. To improve the nuclear norm, Zha et al. [13] presented its weighted l_p form that can accurately utilize structural sparsity and self-similarity for IR. Under the wavelet framework, You et al. [14] employed the ADMM for a nonconvex IR model with a concave regularization. Considering the potential of nonlocal self-similarity, Ren et al. [15] proposed an overlapping nonlocal regression-based IR algorithm. To improve the performance, the authors introduced overlapping pixel groups, adjusted nonlocal regression and the split Bregman method. Makinen et al. [16] intergrated a method of calculating the noise power-spectrum into the BM3D algorithm to improve its performance in IR and other problems. Since a clear image can be represented by a sparse LR matrix, Yair et al. [17] applied multiscale WNNM (MSWNNM) and variable splitting to IR to obtain better results.

In addition to the above typical IR algorithms based on sparse priors, there are also some special algorithms with more flexible frameworks. For IR, Chierchia et al. [18] proposed a structure-tensor (ST) based IR algorithm, which can use TV or nonlocal TV as the regularizer in their built model. In the SAPG algorithm [19], the TV prior and a transform-sparsity prior are separately incorporated into the Bayesian framework to restore images while estimating parameters. The IDBP algorithm proposed by Tirer et al. [20] has a plug-and-play structure, so good image denoising methods can be employed to solve related subproblems. Similarly, the IR algorithm based on regularized similarity (RS) proposed by Kheradmand et al. [21] also used excellent external methods to address its denoising subproblem.

B. Contributions and Outline

As recalled above, most sparse regularization algorithms treat IR as a convex optimization problem with a single prior or l_1-norm-induced priors. However, theories and experiments have proven that these practices cannot provide IR with improvements. Therefore, to boost the IR, we model it as a nonconvex optimization problem with l_0-norm-induced multiple sparse priors. For the built model, a novel solution derived from ALM is presented to obtain better restoration results and eliminate more annoying visual defects. The contributions of the proposed algorithm (denoted by DRALM) are summarized as follows.

- Studies indicate that the best sparsity-inducing norm is the l_0-norm, which directly calculates the number of nonzero components [22]. The most favourable l_1-norm is just a convex relaxation of the l_0-norm, and this relaxation is optimal only under certain conditions. Therefore, we employ the sparse priors induced by the l_0-norm to promote IR.
- The gradient-sparsity priors represented by TV have strong edge-preserving abilities, and the transform-sparsity priors are good at denoising and removing artifacts. Therefore, we fuse the two kinds of sparse priors to build a novel double l_0-regularized IR model. The built model can exploit the strengths of the two kinds of sparse priors and overcome their inherent drawbacks.
- With the help of the ALM and other suitable methods, we decouple and handle the built IR model with a separable structure, thus avoiding the numerical difficulties and NP-hardness caused by the l_0-norm.
- The generated subproblems can obtain closed-form solutions without inner iterations, and subproblems of the same type and structure can be solved in parallel, so the proposed IR algorithm is very efficient.
- The fixed point of DRALM is also a local minimizer of the built IR model, and the bounded sequences obatined by DRALM converge to local minimizers.

We arrange the remaining contents as follows. In II, regularization technology and sparse prior models are reviewed. In III, based on a gradient-sparsity prior and a transform-sparsity prior, we build a double l_0-regularization model for IR. In IV, a novel IR algorithm (DRALM) is proposed and a concise analysis of it is given. In V, experiments are conducted on benchmark images to show the effectiveness of DRALM and demonstrate its superiorities through comparisons. In the last section, the conclusions of the paper are drawn.

II. Review of Regularization Technology and Sparse Prior

Because B is usually not invertible, it is obviously impossible to perform direct computations on $y{=}Bx{+}n$. Even if B is invertible, its ill-posedness only makes direct computations produce trivial solutions. Therefore, researchers have exhausted various ideas to address IR problems. At present, the standard practice is to regard IR as a given minimization problem, i.e., IR modeling.

Originally, IR was cast as a least squared problem

$$\min_x \|Bx - y\|_2^2, \tag{1}$$

where B is defined as $\|B\|_2{=}sup\frac{\|Bz\|_2}{\|z\|_2}$, and the l_2-norm $\|\cdot\|_2$ is defined as $\|z\|_2{=}(\sum_i |z_i|^2)^{\frac{1}{2}}$. The $(B^TB)^{-1}B^Ty$ obtained by handling problem (1) is not meaningful because $(B^TB)^{-1}$ does not always exist and is very ill-conditioned. Additionally, model (1) usually amplifies the noise in images, so it is not competent at IR.

A. Regularization Technology

To remedy the defects of model (1), Tikhonov et al. [23] presented the regularization model:

$$\min_x \frac{1}{2}\|Bx - y\|_2^2 + \lambda\|Ax\|_2^2, \tag{2}$$

where the operator $A{\in}\mathbb{R}^{S\times M}$ is bounded and linear; $\frac{1}{2}\|Bx-y\|_2^2$ is the fidelity-term, which punishes the difference between Bx and y to preserve consistency; $\|Ax\|_2^2$ is the crucial regularization-term, which represents an image prior and enhances the outcomes; and the parameter $\lambda{>}0$ is used to achieve a tradeoff between regularization and fidelity. When $\lambda{\to}\infty$, regularization dominates, so the restored images may have different structures from the observed images. In contrast, when $\lambda{\to}0$, the restored results tend to be blurry and noisy images.

The idea of regularization is to incorporate priors into IR in the form of regularization-terms so that the established optimization models conform to the nature of clear images. The designs of the regularization-terms are the keys to IR modeling and are major factors that affect the final restored results. For example, the regularization term $\|Ax\|_2^2$ enforces the Gibbs prior on clear images, so solving model (2) usually results in smooth images.

B. TV-Model

To promote regularization technology, the TV-based IR model

$$\min_x \frac{1}{2}\|Bx - y\|_2^2 + \lambda\|\nabla x\|_1 \tag{3}$$

is presented, where the l_1-norm $\|\cdot\|_1$ is defined as $\|z\|_1{=}\sum_i |z_i|$; and $\|\nabla\cdot\|_1$ is the famous TV-norm with $\nabla{\in}\mathbb{R}^{2M\times M}$

denoting the gradient operator. Let ∇_1 and ∇_2 be the horizontal and vertical difference operators, respectively; then, the image gradient is defined as ∇x=$[\nabla_1 x; \nabla_2 x]$.

The TV-model (3) arises from the observation that a noisy image has a larger TV than the corresponding clear image. From the perspective of statistics, the TV-model represents the Laplacian prior and is more suitable for natural images than previous models. In the perspective of image priors, the TV-model is also regarded as a sparse prior-based regularization model because the piecewise continuity makes image gradients sparse and the l_1-norm is sparsity-induced.

The TV-model was the first l_1-regularization model and is one of the most significant regularization models in the past 30 years. Currently, researchers continue to expand the boundaries of its applications.

C. Transform-Sparsity Based Prior Models

Natural images are rarely sparse, but these nonsparse images can be transformed into a set of sparse coefficients with few nonzero components. This important discovery of compressed sensing is subsequently introduced into image processing, sparking a boom in the research and use of transform-sparsity priors [24]. The tools that turn the images into sparse coefficients are known as transform matrices or dictionaries, which are generally redundant to fulfill the requirements of sparse representation. Due to the excellent properties of various wavelets, they have become the first choices for transform tools.

Let $D\in\mathbb{R}^{C\times M}$ denote a dictionary; then, the IR models based on the transform-sparsity priors can be uniformly written as

$$\min_x \frac{1}{2}\|Bx - y\|_2^2 + \lambda\psi(Dx), \tag{4}$$

where Dx is the sparse coefficient of image x. Since these norms play the role of sparsity induction, the choices of $\psi(\cdot)$ are normally l_p-norms ($0<p\leq 2$). Currently, the fashionable choice for $\psi(\cdot)$ is definitely the l_1-norm, that is, $\psi(Dx)$ is $\|Dx\|_1$. There are at least two reasons for employing the l_1-norm: the solutions of the l_1-minimization problems are sparse, and the l_1-minimization problems are easy convex optimization problems.

III. Presented IR Model

Let $G(Ax)$ be a regularization term; then, the regularization algorithms generally regard IR as

$$\min_x \frac{1}{2}\|Bx - y\|_2 + \lambda G(Ax). \tag{5}$$

According to the discrepancy principle [25], the equivalent constrained form of (5) is

$$\min_{x\in\Phi} G(Ax), \tag{6}$$

where Φ is defined as $\{x\text{: } \|Bx - y\|_2^2 \leq \delta,\ \forall\ x\in\mathbb{R}^M\}$, and δ is the maximum constant related to the noise variance. Compared with model (5), model (6) omits the parameter selection, but most IR algorithms still prefer model (5).

To avoid nonconvexity and simplify IR, most modern algorithms adopt single prior-based regularization models. However, studies have shown that more beneficial image priors usually mean better outcomes. Therefore, using multiple priors, we model IR as

$$\min_x \frac{1}{2}\|Bx - y\|_2^2 + \sum_i \lambda_i G_i(A_i x), \tag{7}$$

where $G_i(A_i x)$ is a regularization term and the operator $A_i\in\mathbb{R}^{C_i\times M}$ is bounded and linear.

A. Gradient-Sparsity Priors Induced by l_0-Norm

Considering the powerful edge-preserving ability of the gradient-sparsity prior represented by TV, we integrate it into the image prior system of this paper. As mentioned above, both the TV-model and its generalizations prefer using the l_1-norm to measure sparsity. From the perspective of sparse coding, the l_1-norm is the optimal convex relaxation of the l_0-norm under the premise of satisfying the RIP conditions [26]. For complicated IR problems, there are few cases that fully satisfy the RIP conditions. Therefore, we adopt the l_0-norm to induce gradient sparsity, obtaining the first regularization term $G_1(A_1 x)$=$\lambda_1\|\nabla x\|_0$ with the l_0-norm defined as

$$\|z\|_0 = \lim_{p\to 0}\|z\|_p^p = \lim_{p\to 0}\sum_i |z|^p = \#(i : z_i \neq 0). \tag{8}$$

Similar l_0-norm-induced priors are also employed by other image processing studies [27], [28], and the related results explicitly demonstrate the powerful denoising and feature-preserving abilities of the l_0-norm.

B. Transform-Sparsity Priors Induced by l_0-Norm

In addition to the gradient-sparsity prior, we incorporate a transform-sparsity prior into our model. As mentioned above, the regularization term representing the transform-sparsity prior is usually formulated as $\|Dx\|_1$. Compared to the l_1-norm, the l_0-norm can remove low amplitude structures, strengthen the salient edges globally and obtain more accurate sparse solutions. Therefore, we replace the l_1-norm with the l_0-norm, obtaining the second regularization term $G_2(A_2 x)$=$\lambda_2\|Dx\|_0$. We next discuss how to select dictionary D.

Usually, natural images have rich geometric structures and textures, so sparse coding dictionaries should be able to precisely characterize image components. According to physiological research on the human visual system, the optimal image coding tools should have the features of multiresolution, locality and directionality. Therefore, we choose the fast discrete curvelet transform (FDCT) [29] as dictionary D. It is a multiresolution, bandpass and directional analysis tool that meets the requirements of the optimal image coding tool. Moreover, the FDCT adopts directional strip-like basis functions, so it is very suitable for characterizing significant features and is competent in obtaining sparser representations with fewer basis functions.

C. Formulation of Presented IR Model

Substituting l_0-regularization terms into model (7), we obtain

$$\min_x \frac{1}{2}\|Bx - y\|_2^2 + \lambda_1\|\nabla x\|_0 + \lambda_2\|Dx\|_0, \tag{9}$$

where $\|\nabla x\|_0$ and $\|Dx\|_0$ are lower and semicontinuous, and $f(x)$=$\frac{1}{2}\|Bx-y\|_2^2$ is continuously differentiable. The Lipschitz continuous gradient of $f(x)$ is

$$\|\nabla f(x_1)-\nabla f(x_2)\|_2 \leq \mathcal{L}_f\|x_1-x_2\|_2. \tag{10}$$

In (10), the minimum value of the Lipschitz constant $\mathcal{L}_f$ is the maximum eigenvalue of the B^TB.

According to the philosophy of variable splitting [30], to make the nondifferentiable and nonconvex problem (9) solvable, we need to turn to its equivalent constrained form.

Given a minimization problem

$$\min_u g_1(u)+g_2(Au), \tag{11}$$

variable splitting converts it into

$$\begin{aligned}&\min_{u,z} g_1(u)+g_2(z)\\ &s.t. \quad z=Au\end{aligned} \tag{12}$$

with an auxiliary variable z.

Applying variable splitting, the IR model (7) is finally formulated as

$$\begin{aligned}&\min_{x,w,v}\frac{1}{2}\|Bx-y\|_2^2+\lambda_1\|w\|_0+\lambda_2\|v\|_0\\ &s.t. \quad w=\nabla x, v=Dx\end{aligned}, \tag{13}$$

where $w\in\mathbb{R}^{2M}$ and $v\in\mathbb{R}^{C}$ are auxiliary variables, and w= ∇x and v=Dx are constraints. Problem (13) is obviously equivalent to problem (9) in the sense of the feasible solutions $\{(x,\, w,\, v)$: w=∇x, v=$Dx\}$. However, compared with problem (9), problem (13) is more flexible, convenient and efficient to handle.

IV. Proposed DRALM Algorithm

Due to the l_0-norm, the IR problem (13) is a nonconvex optimization problem. Unlike convex optimization problems, the optimal solutions for nonconvex optimization problems are often difficult to obtain, which poses a great challenge to the designs of numerical computation schemes. To eliminate the dilemmas of nonconvex optimization problems, early schemes relaxed them to convex optimization problems. However, convex relaxations change the original problems, and the obtaining solutions may be trivial to the original problems.

For separable convex optimization problems, the current modish and effective optimization methods decouple them into simpler subproblems and find the appropriate methods for these subproblems. In fact, this divide-and-conquer mode is also applicable to separable nonconvex optimization problems to find satisfactory solutions [31]. Therefore, for stability and convergence, we solve the IR model in (13) by the ALM. Notably, the ALM converges without assuming that objective functions are finite or strictly convex, and it is very simple and practical for multivariate optimization problems.

A. ALM

Given a constrained minimization problem

$$\begin{aligned}&\min_x g(u)\\ &s.t.\ z=Au\end{aligned}, \tag{14}$$

Algorithm 1 ALM.

Input $\beta>0$ and $\alpha^0\in\Omega$
Set $k=0$
Do
Step 1: $u^{k+1}\leftarrow \arg\min_u \xi_\beta(u,\alpha^k)$
Step 2: $\alpha^{k+1}\leftarrow \alpha^k+\beta(Au^{k+1}-z)$
Step 3: $k=k+1$
Loop until the criteria are fulfilled
Output u^{k+1}

ALM first turns it into

$$\min_{u,\alpha}\left\{\xi_\beta(u,\alpha)=g(u)+\alpha^T(Au-z)+\frac{\beta}{2}\|Au-z\|_2^2\right\}, \tag{15}$$

where $\alpha\in\mathbb{R}^S$ is the Lagrange multiplier. Then, the ALM solves problem (15) by alternately iterating between

$$u^{k+1}=\arg\min_u \xi_\beta(u,\alpha^k) \tag{16}$$

and

$$\alpha^{k+1}=\alpha^k+\beta(Au^{k+1}-z). \tag{17}$$

Therefore, let $\Omega\in\mathbb{R}^{mn}$ denote the Euclidean space; then the ALM is summarized in **Algorithm 1**.

The optimality conditions of problem (14) are

$$\begin{cases}Au^*-z=0 & primal\ feasibility\\ 0\in\nabla g(u^*)+A^T\alpha^* & dual\ feasibility\end{cases}. \tag{18}$$

Let $u^{k+1}\in\arg\min_x \xi_\beta(u,\alpha^k)$; then, we have

$$0\in\Big\{\nabla_u\xi_\beta(u,\alpha^k)=\nabla g(u^{k+1}) + A^T\big(\alpha^k+\beta(Au^{k+1}-z)\big)\Big\} \tag{19}$$

From (19), it can be observed that the optimality conditions in (18) are satisfied with $(Hu^{k+1}$-$z)\to0$ and $\alpha^k\to\alpha^{k+1}$.

B. Solving Subproblems

Applying ALM to the IR model (13), we obtain

$$\min_{x,w,v,\alpha_1,\alpha_2}\Big\{J(x,w,v)+\alpha_1^T(\nabla x-w)+\frac{\mu_1}{2}\|\nabla x-w\|_2^2 + \alpha_2^T(Dx-v)+\frac{\mu_2}{2}\|Dx-v\|_2^2\Big\}, \tag{20}$$

where $J(x,w,v)$=$\frac{1}{2}\|Bx-y\|_2^2$+$\lambda_1\|w\|_0$+$\lambda_2\|v\|_0$; $\frac{\mu_1}{2}\|\nabla x-w\|_2^2$ and $\frac{\mu_2}{2}\|Dx-v\|_2^2$ are the penalty terms; and μ_1 and μ_2 are the penalty parameters.

Using ordinary mathematical operations, problem (20) equates to

$$\min_{x,w,v,d_1,d_2}\Big\{J(x,w,v)+\frac{\mu_1}{2}\|\nabla x-w+d_1\|_2^2 + \frac{\mu_2}{2}\|Dx-v+d_2\|_2^2\Big\}, \tag{21}$$

where d_1=$\frac{\alpha_1}{\mu_1}$ and d_2=$\frac{\alpha_2}{\mu_2}$.

The philosophy of ALM is to obtain the solution to problem (13) by solving the following independent and tractable subproblems:

$$w^{k+1}\leftarrow\arg\min_w\left\{L(x^k,w,v^k,d_1^k,d_2^k)\right\}, \tag{22}$$

$$v^{k+1} \leftarrow \arg\min_{v} \left\{ L(x^k, w^{k+1}, v, d_1^k, d_2^k) \right\}, \tag{23}$$

$$x^{k+1} \leftarrow \arg\min_{x} \left\{ L(x, w^{k+1}, v^{k+1}, d_1^k, d_2^k) \right\}, \tag{24}$$

$$d_1^{k+1} = d_1^k + (\nabla x^{k+1} - w^{k+1}), \tag{25}$$

$$d_2^{k+1} = d_2^k + (Dx^{k+1} - v^{k+1}), \tag{26}$$

where the objective function $L(x, w, v, d_1, d_2)$=$J(x, w, v)$+$\frac{\mu_1}{2}\|\nabla x - w + d_1\|_2^2$+$\frac{\mu_2}{2}\|Dx - v + d_2\|_2^2$.

1) w^{k+1} Subproblem and v^{k+1} Subproblem: The complete form of the w^{k+1} subproblem (22) is

$$w^{k+1} = \arg\min_{w} \|w - z_1^k\|_2^2 + \rho_1\|w\|_0, \tag{27}$$

where ρ_1=$\frac{2\lambda_1}{\mu_1}$ and z_1^k=∇x^k+d_1^k.

***Proposition** 1:* The solution to the w^{k+1} subproblem (27) is

$$w_i^{k+1} = \begin{cases} (z_1^k)_i & if\ (z_1^k)_i^2 > \rho_1 \\ 0 & if\ (z_1^k)_i^2 < \rho_1 \\ 0\ or\ (z_1^k)_i & if\ (z_1^k)_i^2 = \rho_1 \end{cases}, \tag{28}$$

where w_i is the ith component of w and $(z_1^k)_i$ is the ith component of z_1^k.

Proof: We observe that problem (27) amounts to the $2M$ minimization problems

$$w_i^{k+1} = \arg\min_{w_i} \left\{ F(w_i) = \big(w_i - (z_1^k)_i\big)^2 + \rho_1\|w_i\|_0 \right\}. \tag{29}$$

According to the definition of the l_0-norm, $F(w_i)$ can be further written as

$$F(w_i) = \begin{cases} \big(w_i - (z_1^k)_i\big)^2 + \rho_1 & if\ w_i \neq 0 \\ (z_1^k)_i^2 & if\ w_i = 0 \end{cases}. \tag{30}$$

From (30), we obtain

$$\min_{w_i \neq 0} \left\{ F(w_i) = \big(w_i - (z_1^k)_i\big)^2 - \rho_1 \right\} = \rho_1. \tag{31}$$

Therefore, when $(z_1^k)_i^2$>ρ_1, $F(w_i)$ takes the minimum value ρ_1 at w_i=$(z_1^k)_i$; when $(z_1^k)_i^2$<ρ_1, $F(w_i)$ takes the minimum value $(z_1^k)_i^2$ at w_i=0; when $(z_1^k)_i^2$=ρ_1, $F(w_i)$ takes the minimum value $(z_1^k)_i^2$ or ρ_1 at w_i=0 or w_i=$(z_1^k)_i$, respectively.

The above statements indicate that

$$w_i^{k+1} = \arg\min F(w_i) = \begin{cases} (z_1^k)_i & if\ (z_1^k)_i^2 > \rho_1 \\ 0 & if\ (z_1^k)_i^2 < \rho_1 \\ 0\ or\ (z_1^k)_i & if\ (z_1^k)_i^2 = \rho_1 \end{cases} \tag{32}$$

is the solution to the w^{k+1} subproblem. ■

***Remark** 1:* The above equation indicates that w^{k+1} is computed component by component, so the w^{k+1} subproblem has a cost of $O(2M)$.

The complete form of the v^{k+1} minimization subproblem (23) is

$$v^{k+1} = \arg\min_{v} \|v - z_2^k\|_2^2 + \rho_2\|v\|_0, \tag{33}$$

where ρ_2=$\frac{2\mu_2}{\mu_2}$ and z_2^k=Dx^k+d_2^k. Since the v^{k+1} subproblem and w^{k+1} subproblem belong to the same problem, we immediately deduce that the solution to the v^{k+1} subproblem is

$$v_i^{k+1} = \begin{cases} (z_2^k)_i & if\ (z_2^k)_i^2 > \rho_2 \\ 0 & if\ (z_2^k)_i^2 < \rho_2 \\ 0\ or\ (z_2^k)_i & if\ (z_2^k)_i^2 = \rho_2 \end{cases}. \tag{34}$$

Similarly, (34) indicates that v^{k+1} is also computed component by component, so the v^{k+1} subproblem has a cost of $O(C)$.

2) x^{k+1} Subproblem: The complete form of the x^{k+1} subproblem (24) is

$$x^{k+1} \in \arg\min_{x} \Big\{ Q(x) = \|Bx - y\|_2^2 + \mu_1\|\nabla x - s_1^{k+1}\|_2^2 + \mu_2\|Dx - s_2^{k+1}\|_2^2 \Big\}, \tag{35}$$

where s_1^{k+1}=w^{k+1}-d_1^k and s_2^{k+1}=v^{k+1}-d_2^k.

***Proposition** 2:* The solution to (35) is

$$x^{k+1} = \mathcal{M}^{-1}\big(B^T y + \mu_1\nabla^T s_1^{k+1} + \mu_2 D^T s_2^{k+1}\big) \tag{36}$$

with $\mathcal{M}$=B^TB+$\mu_1\nabla^T\nabla$+$\mu_2 D^T D$.

Proof: According to proposition 16.2 in [32], we have

$$x^{k+1} \in \arg\min_{x} Q(x) \Leftrightarrow 0 \in \partial Q(x^{k+1}). \tag{37}$$

It is known that

$$\partial Q(x) = \mathcal{M}x - \big(B^T y + \mu_1\nabla^T s_1^{k+1} + \mu_2 D^T s_2^{k+1}\big). \tag{38}$$

Therefore, let 0∈$\partial Q(x^{k+1})$; then we obtain

$$x^{k+1} = \mathcal{M}^{-1}\big(B^T y + \mu_1\nabla^T s_1^{k+1} + \mu_2 D^T s_2^{k+1}\big).$$

■

***Remark** 2:* As a Parseval tight frame, the FDCT D satisfies D^TD=I (I is an identity matrix). Hence, $\mathcal{M}$ is an invertible positive definite matrix whose inverse always exists.

With the help of FFT, x^{k+1} can be computed by

$$x^{k+1} = \phi^{-1}\Big(\frac{\overline{\phi}(B)\phi(y) + \mu_1\overline{\phi}(\nabla)\phi(s_1^{k+1}) + \mu_2\overline{\phi}(D)\phi(s_2^{k+1})}{\overline{\phi}(B)\phi(B) + \mu_1\overline{\phi}(\nabla)\phi(\nabla) + \mu_2\overline{\phi}(D)\phi(D)}\Big) \tag{39}$$

where ϕ, ϕ^{-1} and $\overline{\phi}$ are the FFT, inverse FFT, and conjugate FFT, respectively.

Once the constants in (39) are precomputed, computing x^{k+1} requires only two forward FFTs and one inverse FFT per iteration. Since the computational cost of an FFT or an inverse FFT is $O(n^2 logn)$, the efficiency of exactly computing x^{k+1} is very high.

C. Description of Proposed DRALM Algorithm

According to steps (22) to (26) and their solutions, we summarize the proposed IR algorithm as **Algorithm 2**. It is observed that the closed-form solutions of all subproblems are efficiently obtained without inner iterations, and subproblems of the same type and same structure can be solved in parallel. Therefore, in general, **Algorithm 2** (i.e., the proposed DRALM algorihtm) has many advantages that

Algorithm 2 Proposed DRALM Algorithm.

Input regularization parameters λ_1, λ_2 and penalty parameters μ_1, μ_2
Initialize $x^0 = y$, $w^0 = \nabla x^0$, $v^0 = Dx^0$, and $d_1^0 = d_2^0 = 0$
For $k = 0toK$
Computing w^{k+1} according to (28)
Computing v^{k+1} according to (34)
$x^{k+1} \leftarrow \mathcal{M}^{-1}\big(B^T y + \mu_1 \nabla^T s_1^{k+1} + \mu_2 D^T s_2^{k+1}\big)$
If $\frac{\|x^{k+1}-x\|_2}{\|x\|_2} \le \tau$
Stop the iterations
EndIf
$d_1^{k+1} = d_1^k + (\nabla x^{k+1} - w^{k+1})$
$d_2^{k+1} = d_2^k + (Dx^{k+1} - v^{k+1})$
EndFor
Output x^{k+1}

enable the IR problem (13) to be solved with high efficiency. Considering the convergence and efficiency, we adopt the relative errors and the fixed number of iterations as the stopping criteria for DRALM, whichever comes first. In the relative error, the tolerance τ is a very small constant, and x is the original clear image.

D. Analysis of the Proposed DRALM Algorithm

In this subsection, we prove that the fixed point of DRALM is a local minimizer of problem (21) and discuss the convergence of the sequences obtained by DRALM under certain conditions.

***Proposition** 3:* Suppose that $\{x^*,\ w^*,\ v^*,\ d_1^*,\ d_2^*\}$ is a fixed point of DRALM; then, it is a local minimizer of problem (21).

Proof: Assume that ι_0 and ι_1 are two sets and are defined as

$$\begin{cases} \iota_0 = \{i : w^* = 0, v^* = 0\} \\ \iota_1 = \{i : w^* \neq 0, v^* \neq 0\} \end{cases}. \tag{40}$$

Then, according to the solutions to the w^{k+1} minimization subproblem and the v^{k+1} subproblem, we have

$$\begin{cases} |(\nabla x^* + d_1^*)_i| \le \sqrt{\rho_1},\ |(Dx^* + d_2^*)_i| \le \sqrt{\rho_2} & for\ i \in \iota_0 \\ (\nabla x^* + d_1^*)_i = w_i^*,\ (Dx^* + d_2^*)_i = v_i^* & for\ i \in \iota_1 \end{cases}. \tag{41}$$

Let $\{\Delta_1,\ \Delta_2,\ \Delta_3,\ \Delta_4,\ \Delta_5\}\in\Omega$ be the tiny perturbation vectors of $\{x^*,\ w^*,\ v^*,\ d_1^*,\ d_2^*\}$, respectively.

Substituting $\{x^*{+}\Delta_1,\ w^*{+}\Delta_2,\ v^*{+}\Delta_3,\ d_1^*{+}\Delta_4,\ d_2^*{+}\Delta_5\}$ into $L(x,\ w,\ v,\ d_1,\ d_2)$, we obtain

$$\begin{aligned}
&L(x^* + \Delta_1, w^* + \Delta_2, v^* + \Delta_3, d_1^* + \Delta_4, d_2^* + \Delta_5) \\
&= \frac{\|B(x^* + \Delta_1) - y\|_2^2}{2} + \lambda_1\|w^* + \Delta_2\|_0 + \lambda_2\|v^* + \Delta_3\|_0 \\
&\quad + \frac{\mu_1\|(w^* + \Delta_2) - \nabla(x^* + \Delta_1) - (d_1^* + \Delta_4)\|_2^2}{2} \\
&\quad + \frac{\mu_2\|(v^* + \Delta_3) - D(x^* + \Delta_1) - (d_2^* + \Delta_5)\|_2^2}{2} \\
&= \frac{\|Bx^* - y\|_2^2}{2} + \langle B\Delta_1, Bx^* - y\rangle + \frac{\|B\Delta_1\|_2^2}{2} \\
&\quad + \lambda_1\|w^* + \Delta_2\|_0 + \lambda_2\|v^* + \Delta_3\|_0 \\
&\quad + \frac{\mu_1\|w^* - \nabla x^* - d_1^*\|_2^2}{2} + \frac{\mu_1\|\Delta_2 - \nabla\Delta_1 - \Delta_4\|_2^2}{2} \\
&\quad + \mu_1\langle \Delta_2 - \nabla\Delta_1 - \Delta_4, w^* - \nabla x^* - d_1^*\rangle \\
&\quad + \frac{\mu_2\|v^* - Dx^* - d_2^*\|_2^2}{2} + \frac{\mu_2\|\Delta_3 - D\Delta_3 - \Delta_5\|_2^2}{2} \\
&\quad + \mu_2\langle \Delta_3 - D\Delta_3 - \Delta_5, v^* - Dx^* - d_2^*\rangle \\
&= \frac{\|Bx^* - y\|_2^2}{2} + \frac{\|B\Delta_1\|_2^2}{2} \\
&\quad + \lambda_1\|w^* + \Delta_2\|_0 + \lambda_2\|v^* + \Delta_3\|_0 \\
&\quad + \frac{\mu_1\|w^* - \nabla x^* - d_1^*\|_2^2}{2} + \frac{\mu_1\|\Delta_2 - \nabla\Delta_1 - \Delta_4\|_2^2}{2} \\
&\quad + \frac{\mu_2\|v^* - Dx^* - d_2^*\|_2^2}{2} + \frac{\mu_2\|\Delta_3 - D\Delta_3 - \Delta_5\|_2^2}{2} \\
&\quad + \mu_1\langle \Delta_2 - \Delta_4, w^* - \nabla x^* - d_1^*\rangle \\
&\quad + \mu_2\langle \Delta_3 - \Delta_5, v^* - Dx^* - d_2^*\rangle \\
&\quad + \langle \Delta_1, B^T(Bx^* - y) + \mu_1\nabla^T(\nabla x^* + d_1^* - w^*) \\
&\qquad\qquad + \mu_2 D^T(Dx^* + d_2^* - v^*)\rangle \\
&\ge \frac{\|Bx^* - y\|_2^2}{2} + \lambda_1\|w^* + \Delta_2\|_0 + \lambda_2\|v^* + \Delta_3\|_0 \\
&\quad + \frac{\mu_1\|w^* - \nabla x^* - d_1^*\|_2^2}{2} + \frac{\mu_2\|v^* - Dx^* - d_2^*\|_2^2}{2} \\
&\quad + \mu_1\langle \Delta_2 - \Delta_4, w^* - \nabla x^* - d_1^*\rangle \\
&\quad + \mu_2\langle \Delta_3 - \Delta_5, v^* - Dx^* - d_2^*\rangle \\
&\quad + \langle \Delta_1, B^T(Bx^* - y) + \mu_1\nabla^T(\nabla x^* + d_1^* - w^*) \\
&\qquad\qquad + \mu_2 D^T(Dx^* + d_2^* - v^*)\rangle.
\end{aligned} \tag{42}$$

From (35), we know that

$$\begin{aligned}
&\langle x - x^*, B^T(Bx^* - y) + \mu_1\nabla^T(\nabla x^* + d_1^* - w^*) \\
&\qquad + \mu_2 D^T(Dx^* + d_2^* - v^*)\rangle \ge 0
\end{aligned} \tag{43}$$

for all $x\in\Omega$.

Therefore,

$$\begin{aligned}
&L(x^* + \Delta_1, w^* + \Delta_2, v^* + \Delta_3, d_1^* + \Delta_4, d_2^* + \Delta_5) \\
&\ge \frac{\|Bx^* - y\|_2^2}{2} + \lambda_1\|w^* + \Delta_2\|_0 + \lambda_2\|v^* + \Delta_3\|_0 \\
&\quad + \frac{\mu_1\|w^* - \nabla x^* - d_1^*\|_2^2}{2} + \frac{\mu_2\|v^* - Dx^* - d_2^*\|_2^2}{2} \\
&\quad + \mu_1\langle \Delta_2 - \Delta_4, w^* - \nabla x^* - d_1^*\rangle \\
&\quad + \mu_2\langle \Delta_3 - \Delta_5, v^* - Dx^* - d_2^*\rangle \\
&= \frac{\|Bx^* - y\|_2^2}{2} + \frac{\mu_1\|w^* - \nabla x^* - d_1^*\|_2^2}{2} \\
&\quad + \frac{\mu_2\|v^* - Dx^* - d_2^*\|_2^2}{2} \\
&\quad + \sum_i \Big(\lambda_1\|(w^* + \Delta_2)_i\|_0 + \lambda_2\|(v^* + \Delta_3)_i\|_0 \\
&\qquad + \mu_1\langle(\Delta_2 - \Delta_4)_i, (w^* - \nabla x^* - d_1^*)_i\rangle \\
&\qquad + \mu_2\langle(\Delta_3 - \Delta_5)_i, (v^* - Dx^* - d_2^*)_i\rangle\Big).
\end{aligned} \tag{44}$$

According to (41), (44) can be written as

$$\begin{aligned}
&L(x^* + \Delta_1, w^* + \Delta_2, v^* + \Delta_3, d_1^* + \Delta_4, d_2^* + \Delta_5) \\
&\ge \frac{\|Bx^* - y\|_2^2}{2} + \frac{\mu_1\|w^* - \nabla x^* - d_1^*\|_2^2}{2} \\
&\quad + \frac{\mu_2\|v^* - Dx^* - d_2^*\|_2^2}{2} \\
&\quad + \sum_{i\in\iota_1} \Big(\lambda_1\|(w^* + \Delta_2)_i\|_0 + \lambda_2\|(v^* + \Delta_3)_i\|_0\Big)
\end{aligned}$$

$$
\begin{aligned}
&+ \sum_{i\in\iota_0} \Big(\lambda_1\|w_i^*\|_0 + \lambda_2\|v_i^*\|_0 \\
&\qquad - \mu_1\langle(\Delta_2-\Delta_4)_i, (\nabla x^* + d_1^*)_i\rangle \\
&\qquad - \mu_2\langle(\Delta_3-\Delta_5)_i, (Dx^* + d_2^*)_i\rangle\Big). \qquad (45)
\end{aligned}
$$

When $|(\Delta_2)_i|$ and $|(\Delta_3)_i|$ are small enough, the equation

$$
\begin{cases}
\|(w^*+\Delta_2)_i\|_0 = \|w_i^*\|_0 \\
\|(v^*+\Delta_3)_i\|_0 = \|v_i^*\|_0
\end{cases}, \forall i \in \iota_1 \qquad (46)
$$

holds.

Hence, from (45), we obtain

$$
\begin{aligned}
&L(x^*+\Delta_1, w^*+\Delta_2, v^*+\Delta_3, d_1^*+\Delta_4, d_2^*+\Delta_5) \\
&\quad \geq L(x^*, w^*, v^*, d_1^*, d_2^*) \\
&\qquad + \sum_{i\in\iota_0} \Big(\lambda_1\|(\Delta_2)_i\|_0 - \mu_1\langle(\Delta_2-\Delta_4)_i, (\nabla x^* + d_1^*)_i\rangle \\
&\qquad\quad + \lambda_2\|(\Delta_3)_i\|_0 - \mu_2\langle(\Delta_3-\Delta_5)_i, (Dx^* + d_2^*)_i\rangle\Big).
\end{aligned} \qquad (47)
$$

We next demonstrate that, for $i\in\iota_0$ and sufficiently small $\{\|\Delta_j\|_\infty\}_{j=1}^5$, the equation

$$
\begin{cases}
\lambda_1\|(\Delta_2)_i\|_0 - \mu_1\langle(\Delta_2-\Delta_4)_i, (\nabla x^* + d_1^*)_i\rangle \geq 0 \\
\lambda_2\|(\Delta_3)_i\|_0 - \mu_2\langle(\Delta_3-\Delta_5)_i, (Dx^* + d_2^*)_i\rangle \geq 0
\end{cases} \qquad (48)
$$

holds.

When $(\nabla x^*+d_1^*)_i=(Dx^*+d_2^*)_i=0$, (48) immediately holds.

When $(\nabla x^*+d_1^*)_i\neq 0$ and $(Dx^*+d_2^*)_i\neq 0$, (48) holds as long as

$$
\begin{cases}
\dfrac{\lambda_1}{\mu_1|(\nabla x^* + d_1^*)_i|} \geq |(\Delta_2-\Delta_4)_i| \\
\dfrac{\lambda_2}{\mu_2|(Dx^* + d_2^*)_i|} \geq |(\Delta_3-\Delta_5)_i|
\end{cases}. \qquad (49)
$$

In summary, we conclude that there exists a small positive constant ϵ that satisfies $max(\|\Delta_1\|_\infty, \|\Delta_2-\Delta_4\|_\infty, \|\Delta_3-\Delta_5\|_\infty)<\epsilon$; then, $L(x^*+\Delta_1, w^*+\Delta_2, v^*+\Delta_3, d_1^*+\Delta_4, d_2^*+\Delta_5) \geq L(x^*, w^*, v^*, d_1^*, d_2^*)$ holds. ∎

***Remark** 3:* Since the pixel values of an image are in a fixed range (e.g., 0 to 255), the sequence $\{x^{k+1}, w^{k+1}, v^{k+1}, d_1^{k+1}, d_2^{k+1}\}$ is bounded [33]. Given that the sequence $\{x^{k+1}, w^{k+1}, v^{k+1}, d_1^{k+1}, d_2^{k+1}\}$ is bounded, then a convergent subsequence $\{x^s, w^s, v^s, d_1^s, d_2^s\}$ exists, which is verified by the experiments.

V. Experimental Results

We perform IR experiments to demonstrate the effectiveness, convergence and advantages of the proposed DRALM. The criteria for evaluating the performance of the algorithms are ISNR, SSIM, speed and visual effects. The main configurations of software and hardware for the experiments are as follows: Windows 7 OS and MATLAB platform; Core i5-3230M (2.60GHz), 8GB of memory and a SSD. The popular benchmark images shown in Fig. 1 are employed as the clear images.

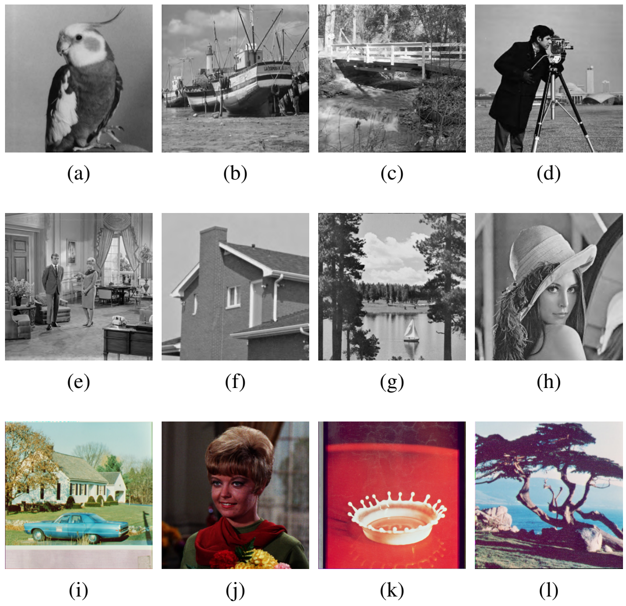


Fig. 1: Clear Images. (a) Bird, (b) Boats, (c) Bridge, (d) Cameraman (abbreviated CM), (e) Couple, (f) House, (g) Lake, (h) Lena, (i) Car, (j) Female, (k) Splash, and (l) Tree.

Fig. 2: Degraded Gray Imags. (a) to (h) GD gray images, and (i) to (p) AD gray images.

A. Restoration of Degraded Gray Images

To generate the Gaussian-degraded (GD) and average-degraded (AD) gray images in Fig. 2, the clear gray images are blurred and polluted by Gaussian noise (std=2). The two blur kernels imposed on the clear images are generated by MATLAB functions *fspecial*('Gaussian', 15, 5) and *fspecial*('average', 15), respectively. The images in Fig. 2 are severely degraded, so restoring them is a challenge

TABLE I: ISNR (dB) of Restored GD Gray Images.

Algorithms	Bird	Boats	Bridge	CM	Couple	House	Lake	Lena
SURELET	5.40	3.39	2.81	3.55	3.41	5.64	3.88	3.46
DADMM(C)	4.78	2.17	2.03	2.96	2.22	4.77	2.83	2.95
DADMM(H)	4.51	2.09	1.87	2.51	2.11	4.37	2.80	2.86
WFPS	4.38	1.76	1.57	2.38	1.68	3.96	2.57	2.77
WLPADMM	6.29	3.85	2.88	4.14	3.98	6.83	4.31	3.97
MSWNNM	4.31	2.81	2.43	2.67	2.97	4.40	2.77	3.01
SAPG(T)	6.00	3.20	2.67	3.73	3.27	6.18	3.43	3.73
SAPG(W)	3.03	1.72	1.45	1.57	1.76	3.07	2.09	2.26
IDBP	6.33	3.81	2.93	4.04	4.01	6.47	4.26	4.14
DRALM	6.66	4.50	3.57	4.44	4.62	6.77	4.93	4.27

TABLE II: SSIM of Restored GD Gray Images.

Algorithms	Bird	Boats	Bridge	CM	Couple	House	Lake	Lena
SURELET	0.8602	0.6670	0.5249	0.7125	0.6645	0.7814	0.7007	0.6898
DADMM(C)	0.8558	0.6157	0.4546	0.7149	0.5944	0.7763	0.6624	0.6725
DADMM(H)	0.8598	0.6155	0.4521	0.6986	0.5902	0.7699	0.6654	0.6738
WFPS	0.8602	0.6016	0.4319	0.6974	0.5656	0.7603	0.6587	0.6716
WLPADMM	0.8659	0.6931	0.5455	0.7403	0.6987	0.7991	0.7146	0.7115
MSWNNM	0.7808	0.6407	0.5069	0.5304	0.6416	0.7102	0.5440	0.6543
SAPG(T)	0.8643	0.6611	0.5072	0.7419	0.6547	0.7925	0.6939	0.7009
SAPG(W)	0.8179	0.5969	0.4439	0.6319	0.5793	0.7100	0.6406	0.6394
IDBP	0.8718	0.6838	0.5315	0.7180	0.6931	0.7643	0.7000	0.7205
DRALM	0.8771	0.7197	0.5934	0.7405	0.7286	0.8039	0.7210	0.7326

TABLE III: ISNR (dB) of Restored AD Gray Images.

Algorithms	Bird	Boats	Bridge	CM	Couple	House	Lake	Lena
SURELET	6.55	4.89	4.01	4.64	5.03	7.22	5.34	4.61
DADMM(C)	6.48	4.05	3.37	4.58	4.18	6.99	4.48	4.34
DADMM(H)	6.51	3.93	3.29	4.20	4.06	6.57	4.35	4.10
WFPS	6.49	3.31	2.81	3.92	3.26	6.37	3.99	4.08
WLPADMM	7.05	4.98	3.93	5.15	5.10	7.78	5.35	4.81
MSWNNM	5.15	4.15	3.50	3.45	4.48	5.74	3.26	4.04
SAPG(T)	7.07	4.76	3.93	5.30	4.71	7.40	5.31	4.96
SAPG(W)	6.15	4.24	3.33	3.34	4.42	6.19	3.96	4.15
IDBP	8.28	5.41	4.20	5.71	5.77	8.64	5.82	5.48
DRALM	8.54	6.16	4.85	5.91	6.35	9.10	6.63	5.72

TABLE IV: SSIM of Restored AD Gray Images.

Algorithms	Bird	Boats	Bridge	CM	Couple	House	Lake	Lena
SURELET	0.8457	0.6993	0.5710	0.6937	0.7102	0.7811	0.7145	0.6978
DADMM(C)	0.8562	0.6692	0.5162	0.7454	0.6746	0.7964	0.6899	0.6965
DADMM(H)	0.8606	0.6681	0.5126	0.7326	0.6712	0.7922	0.6886	0.6906
WFPS	0.8646	0.6333	0.4630	0.7212	0.6153	0.7889	0.6760	0.6895
WLPADMM	0.8323	0.6998	0.5883	0.7275	0.7126	0.7788	0.7026	0.6936
MSWNNM	0.6970	0.6395	0.5458	0.4745	0.6742	0.6746	0.4982	0.6384
SAPG(T)	0.8396	0.6933	0.5679	0.7662	0.6947	0.7741	0.7216	0.7000
SAPG(W)	0.8503	0.6851	0.5302	0.6669	0.6898	0.7596	0.6742	0.6859
IDBP	0.8863	0.7275	0.5912	0.7474	0.7414	0.7857	0.7205	0.7377
DRALM	0.8888	0.7602	0.6455	0.7825	0.7765	0.8293	0.7599	0.7553

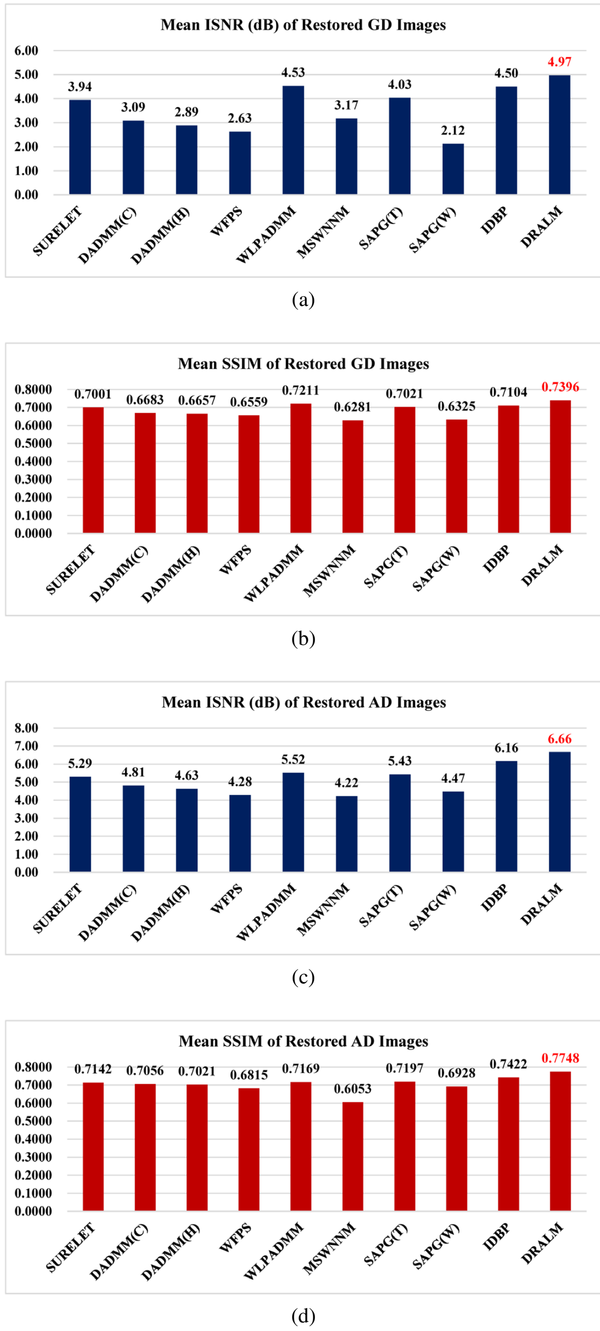


Fig. 3: Mean ISNR and SSIM of Restored Degraded Images.

and the results of their restoration are very suitable for comprehensively evaluating the performance of IR algorithms. In other words, an IR algorithm that can successfully and efficiently restore the degraded images in Fig. 2 is highly robust, stable and competitive in various applications. Therefore, our results obtained by restoring these severely degraded images are of significance. Since the pixel values of an image generally range from 0 to 255 and mapping outcomes to the above range contributes to high IR quality, per iteration, we rescale x^{k+1} to the range from 0 to 255 after computing x^{k+1}. Our DRALM competes with outstanding IR algorithms of mainstream types, including SURELET algorithm [11], DADMM(C) algorithm [5], DADMM(H) algorithm [5], WFPS algorithm [12], WLPADMM algorithm [13], MSWNNM algorithm (initialized with Wiener filter) [17], SAPG(T) algorithm (using the TV prior) [19], SAPG(W) algorithm (using a sparse wavelet prior) [19], and IDBP algorithm (using BM3D as the denoiser) [20].

Theoretically, dynamically adjusting the parameters according to the degraded images is beneficial to the restored results, but frequent adjustments undoubtedly reduce the practicability and efficiency of the proposed DRALM. Moreover, empirically speaking, these fixed parameters are not too sensitive to different degraded images. Therefore, we fix the parameters as $(\lambda_1, \lambda_2, \mu_1, \mu_2)$=(1.2, 0.9, 0.05, 0.05). For efficiency and convergence, the maximum iterations of the proposed DRALM is 200, and τ is set to 10^{-4}. All parameters and settings of the comparison algorithms are their own default values to accurately evaluate the performance.

After handling the degraded images in Fig. 2, the values of ISNR and SSIM are recorded in TABLEs I to IV and shown in Fig. 3, and the corresponding visual effects are shown in Figs. 4 and 5. The experimental results explicitly imply that DRALM successfully accomplishes IR tasks and efficiently restores clear images. In terms of ISNR and SSIM,

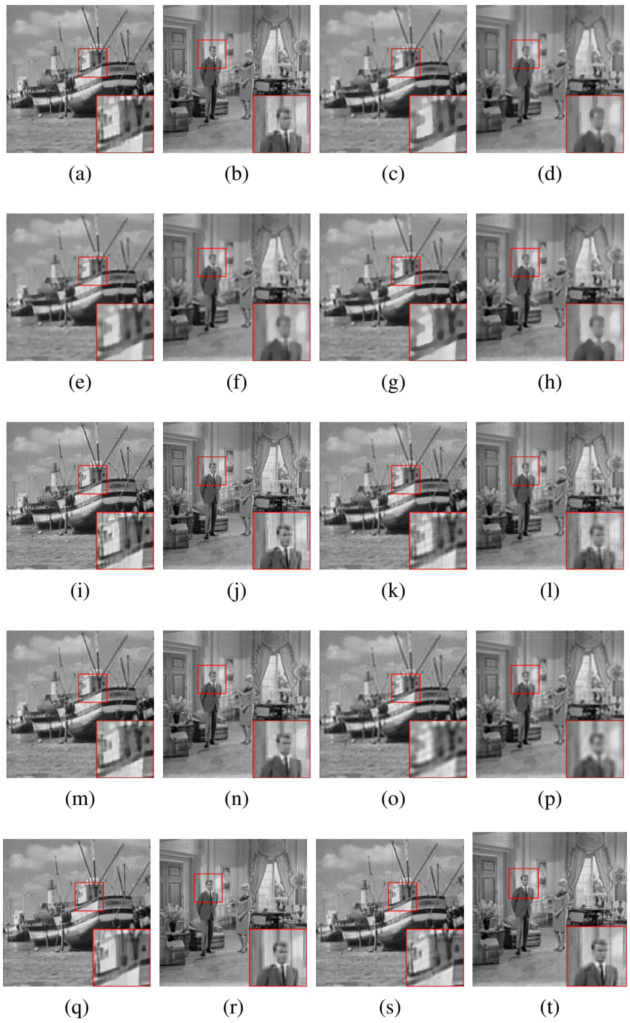


Fig. 4: Visual Examples of Restored GD Gray Images. (a) and (b) SURELET, (c) and (d) DADMM(C), (e) and (f) DADMM(H), (g) and (h) WFPS, (i) and (j) WLPADMM, (k) and (l) MSWNNM, (m) and (n) SAPG(T), (o) and (p) SAPG(W), (q) and (r) IDBP, and (s) and (t) proposed DRALM.

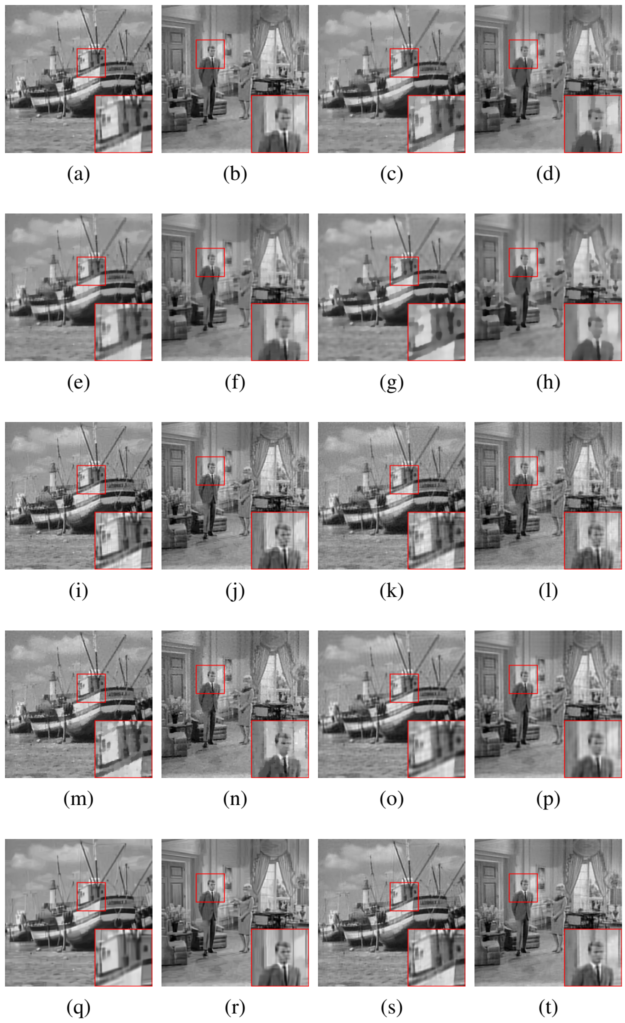


Fig. 5: Visual Examples of Restored AD Gray Images. (a) and (b) SURELET, (c) and (d) DADMM(C), (e) and (f) DADMM(H), (g) and (h) WFPS, (i) and (j) WLPADMM, (k) and (l) MSWNNM, (m) and (n) SAPG(T), (o) and (p) SAPG(W), (q) and (r) IDBP, and (s) and (t) proposed DRALM.

the data definitely indicate that DRALM outperforms the comparison algorithms. ISNR is a full-reference criterion of image quality, and SSIM focuses on image brightness, image contrast and image structure. Therefore, the numerical results of these two criteria also show that our DRALM algorithm not only has overall advantages, but also has advantages in some specific aspects. In Figs. 4 and 5, we show some of the restored images with local magnifications. Compared with other algorithms, the images obtained by DRALM illustrate more important features and fewer artifacts, which confirms the results of ISNR and SSIM. The visual comparisons indicate that double l_0-regularization balances edge sharpening and noise suppression.

In terms of speed, we run each algorithm ten times and record the mean time in TABLEs V and VI. Although DRALM is not the fastest, its speed is on the same order of magnitude as the fastest algorithms, so the practicality of DRALM does not decrease. Moreover, for DRALM, more time is also an inevitable tradeoff to pay when adopting double l_0-induced priors to promote restoration quality.

In summary, the proposed DRALM has a full range of competitive advantages over other algorithms. Regardless of what type of degraded gray images are processed, the overall performance of DRALM is stable, and it also performs better in the consistency of the evaluation results of various criteria. This should be attributed to our well-designed IR model and corresponding efficient solution.

Below, we verify the convergence of DRALM by the evolutions of relative errors $\frac{\|x^{k+1}-x\|_2}{\|x\|_2}$. As revealed by the curves in Figs. 6 and 7, there exists a positive integer K and a sufficiently small positive constant τ such that $\frac{\|x^{k+1}-x\|_2}{\|x\|_2}\leq\tau$ holds when $(k{+}1)\geq K$, that is, $\{x^{k+1}\}$ converge to x with sufficient iterations. It is noteworthy that the first several iterations converge quickly and approach the optimization target, so even if the stopping criteria are relaxed, the DRALM can still obtain satisfactory solutions.

B. Restoration of Degraded Color Images

To generate the degraded color images shown in Fig. 8, the clear color images are blurred by *fspecial*('Gaussian', 15, 5) and *fspecial*('average', 15), and polluted by Gaussian

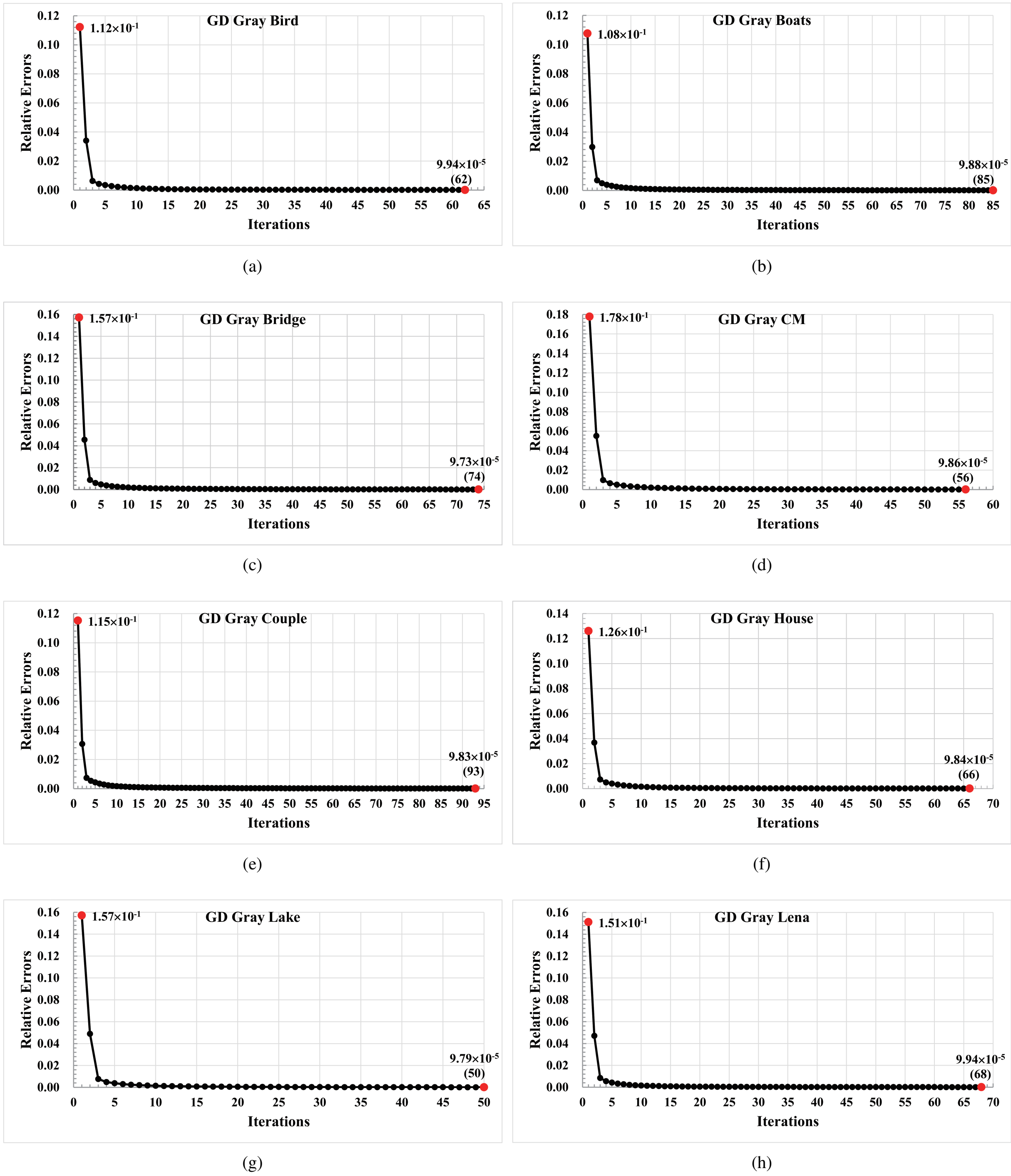


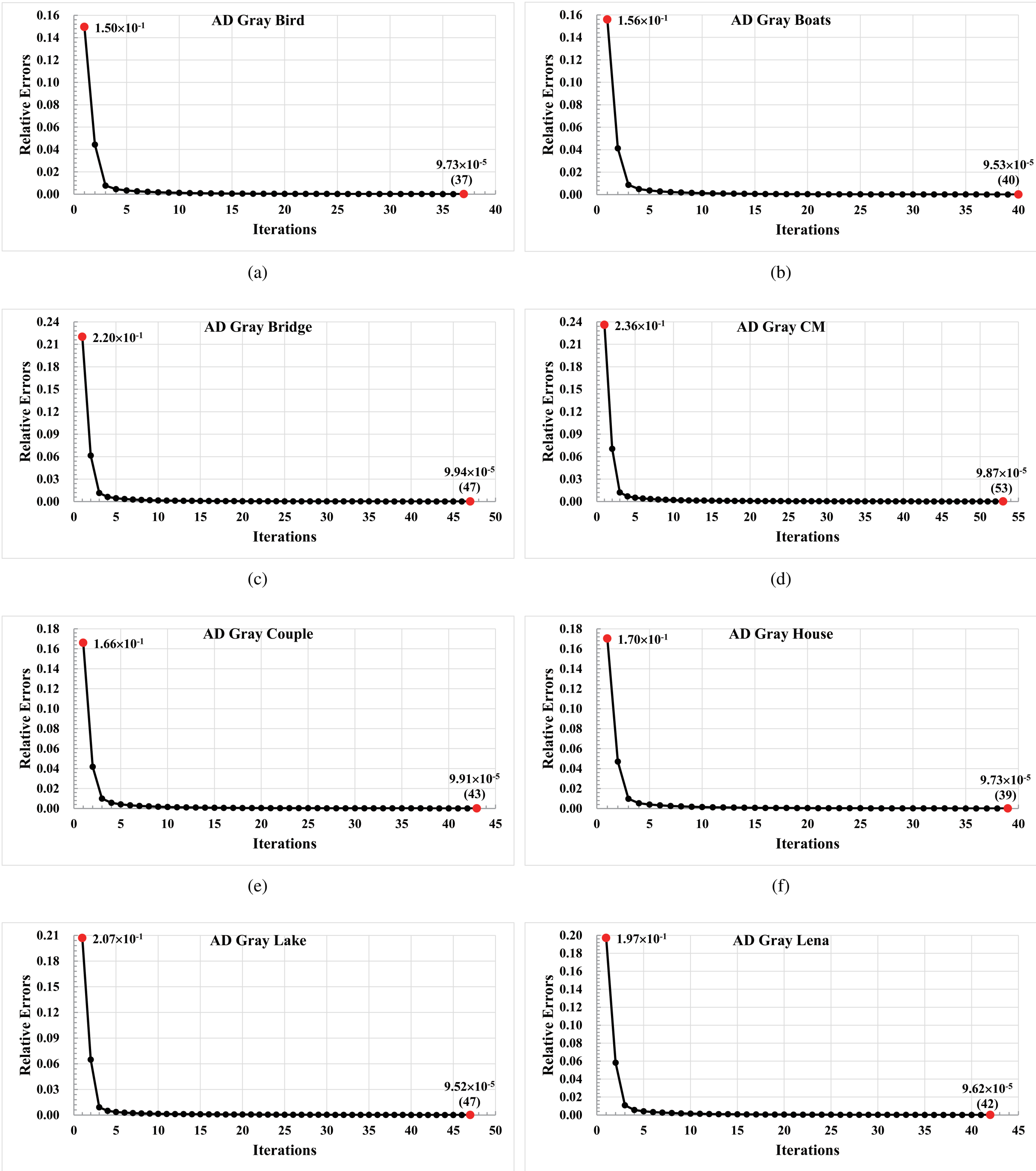
Fig. 6: Evolutions of Relative Errors When Dealing with GD Gray Images.The red dots represent the first iteration and last iteration, respectively; the values next to the red dots are the relative errors; the numbers in parentheses represent the total iterations.

TABLE V: Mean Time (seconds) Needed for Handling GD Gray Images.

Algorithms	Bird	Boats	Bridge	CM	Couple	House	Lake	Lena
SURELET	0.68	2.24	2.27	0.70	2.24	0.68	2.25	0.66
DADMM(C)	2.10	13.76	13.87	2.12	13.80	2.11	13.75	2.11
DADMM(H)	0.73	4.50	4.56	0.69	4.43	0.73	4.40	0.72
WFPS	45.42	292.35	326.47	34.82	302.20	49.16	224.99	44.10
WLPADMM	666.17	3468.36	2403.62	1236.21	2694.16	982.74	2657.11	740.05
MSWNNM	523.01	2277.08	2260.51	520.00	2048.05	515.29	2231.43	495.89
SAPG(T)	27.87	172.74	174.16	27.94	174.76	27.93	179.12	28.30
SAPG(W)	15.69	54.14	58.18	16.44	56.54	16.21	54.30	16.18
IDBP	23.61	103.26	120.06	23.15	102.63	23.90	100.69	23.56
DRALM	6.63	41.87	40.65	6.82	48.85	8.52	26.05	6.41

TABLE VI: Mean Time (seconds) Needed for Handling AD Gray Images.

Algorithms	Bird	Boats	Bridge	CM	Couple	House	Lake	Lena
SURELET	0.69	2.26	2.36	0.69	2.25	0.72	2.26	0.68
DADMM(C)	2.12	13.86	13.94	2.08	13.96	2.13	13.87	2.10
DADMM(H)	0.73	4.22	4.18	0.69	4.17	0.68	4.04	0.69
WFPS	49.29	297.75	333.39	43.50	308.55	54.61	309.66	50.44
WLPADMM	288.92	1474.67	1154.09	730.66	1039.44	520.62	1668.75	469.99
MSWNNM	524.91	1962.18	1972.77	416.33	1948.89	459.18	2012.80	456.85
SAPG(T)	27.65	186.99	189.59	28.39	179.44	28.05	178.98	30.23
SAPG(W)	17.36	74.58	83.08	20.42	83.88	19.72	70.11	20.09
IDBP	23.77	104.21	108.61	23.91	104.97	24.50	102.71	23.75
DRALM	3.61	20.61	24.81	5.19	22.96	3.85	24.71	4.12



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 7: Evolutions of Relative Errors When Dealing with AD Gray Images. The red dots represent the first iteration and last iteration, respectively; the values next to the red dots are the relative errors; the numbers in parentheses represent the total iterations.

TABLE VII: ISNR (dB) of Restored GD Color Images.

Algorithms	Car	Female	Splash	Tree	Mean Val.
ST-TV	2.11	2.49	3.18	3.01	2.70
ST-NLTV	2.39	2.61	3.46	2.68	2.79
RS	4.79	4.41	6.02	4.80	5.01
newBM3D	3.52	3.84	5.62	4.21	4.30
DRALM	5.57	5.31	8.76	6.08	6.43

TABLE VIII: SSIM of Restored GD Color Images.

Algorithms	Car	Female	Splash	Tree	Mean Val.
ST-TV	0.5860	0.6395	0.7216	0.5114	0.6146
ST-NLTV	0.5861	0.6401	0.6742	0.5241	0.6061
RS	0.6942	0.7243	0.8076	0.6503	0.7191
newBM3D	0.6986	0.7455	0.8485	0.6489	0.7354
DRALM	0.7691	0.7954	0.8829	0.7210	0.7921

noise (std=2). The images in Fig. 8 are severely degraded, so restoring them is a challenge and the results of their restoration are very suitable for comprehensively evaluating the performance of IR algorithms. In other words, an IR algorithm that can successfully and efficiently restore the degraded images in Fig. 8 is highly robust, stable and

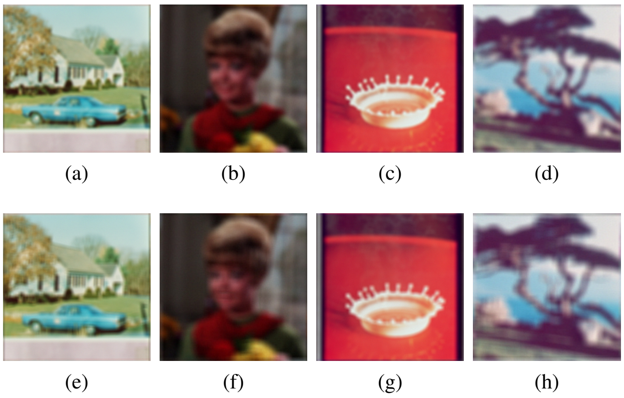


Fig. 8: Degraded Color Images. (a) to (d) GD color images, and (e) to (h) AD color images.

TABLE IX: ISNR (dB) of Restored AD Color Images.

Algorithms	Car	Female	Splash	Tree	Mean Val.
ST-TV	3.37	3.46	4.95	3.94	3.93
ST-NLTV	3.86	3.77	5.40	4.53	4.39
RS	5.96	5.57	7.39	6.69	6.40
newBM3D	5.41	4.91	7.83	5.97	6.03
DRALM	7.79	7.40	12.15	8.69	9.01

TABLE X: SSIM of Restored AD Color Images.

Algorithms	Car	Female	Splash	Tree	Mean Val.
ST-TV	0.6237	0.6451	0.7975	0.5508	0.6543
ST-NLTV	0.6350	0.6583	0.7937	0.5772	0.6661
RS	0.6943	0.6988	0.7518	0.6638	0.7022
newBM3D	0.7315	0.7431	0.8581	0.6834	0.7540
DRALM	0.8208	0.8285	0.8990	0.7829	0.8328

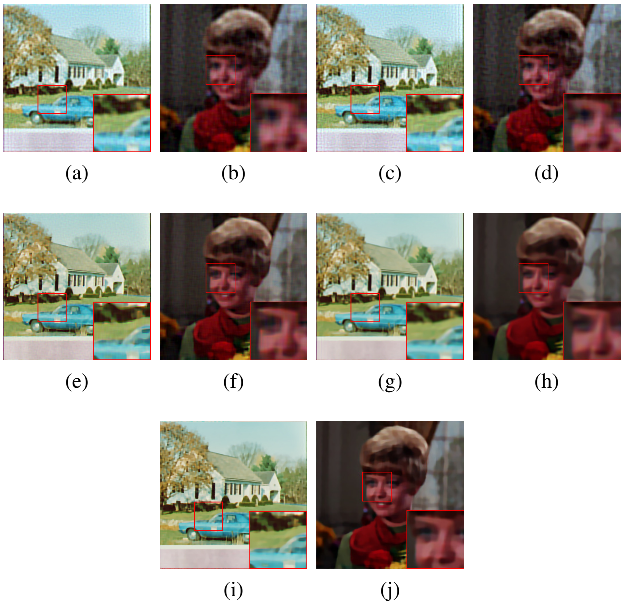


Fig. 9: Visual Examples of Restored GD Color Images. (a) and (b) ST-TV, (c) and (d) ST-NLTV, (e) and (f) RS, and (g) and (h) proposed DRALM.

competitive in various applications. Therefore, our results obtained by restoring these severely degraded images are of significance. Considering the differences between gray images and color images, few gray IR algorithms (including the comparison algorithms in the gray IR experiments) have been extended to color images. Therefore, when restoring

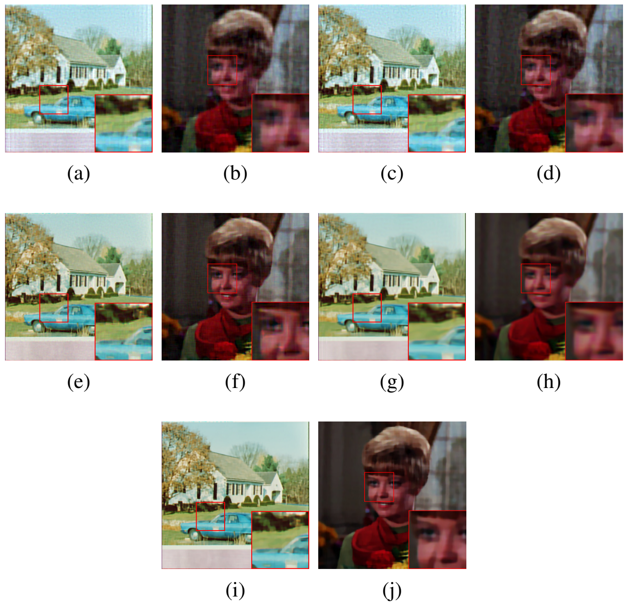


Fig. 10: Visual Examples of Restored AD Color Images. (a) and (b) ST-TV, (c) and (d) ST-NLTV, (e) and (f) RS, and (g) and (h) proposed DRALM.

TABLE XI: Mean Time Needed (seconds) for Handling GD Color Images.

Algorithms	Car	Female	Splash	Tree
ST-TV	635.13	228.50	658.62	178.52
ST-NLTV	2052.25	556.21	1873.10	498.61
RS	154.89	31.64	163.27	35.41
newBM3D	28.34	7.63	38.48	7.36
DRALM	57.36	17.52	61.38	14.10

TABLE XII: Mean Time (seconds) Needed for Handling AD Color Images.

Algorithms	Car	Female	Splash	Tree
ST-TV	652.94	209.00	675.03	193.80
ST-NLTV	2086.24	567.61	2002.52	570.55
RS	151.59	31.46	142.65	31.74
newBM3D	25.88	8.08	26.51	6.84
DRALM	47.45	14.94	52.21	12.99

degraded color images, our DRALM competes with outstanding color IR algorithms of mainstream types, including the ST-TV algorithm [18], ST-NLTV algorithm [18], RS algorithm [21] and newBM3D algorithm [16]. For color images, we stack their R(ed), G(reen), and B(lue) channel images together to form single channel images like gray images. In this way, the DRALM is able to directly handle color images. It can also reduce the difficulty and improve the speed, and avoid processing the R(ed), G(reen) and B(lue) channel images one by one.

Theoretically, dynamically adjusting the parameters according to the degraded images is beneficial to the restored results, but frequent adjustments undoubtedly reduce the practicability and efficiency of the proposed DRALM. Moreover, empirically speaking, these fixed parameters are not too sensitive to different degraded images. Therefore, we fix the parameters of DRALM as $(\lambda_1, \lambda_2, \mu_1, \mu_2)$=(2.8, 5.5, 0.7, 0.3). For efficiency and convergence, the maximum iterations of DRALM is 100, and τ is set to 10^{-3}. The ST algorithm,

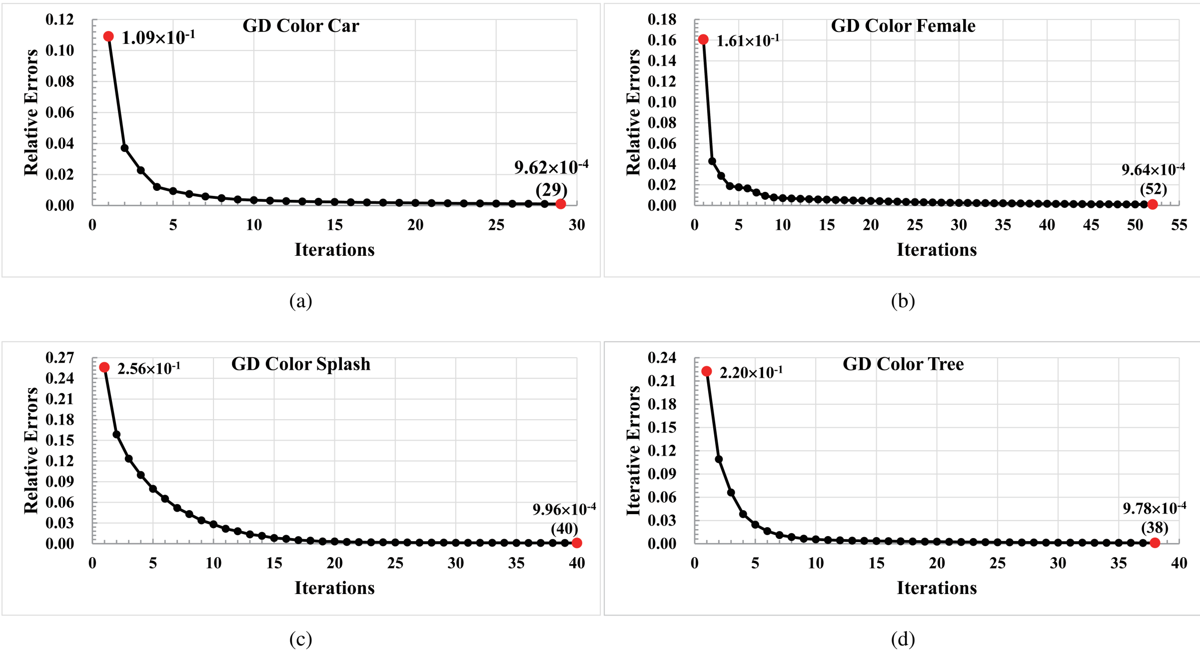


Fig. 11: Evolutions of Relative Errors When Dealing with GD Color Images. The red dots represent the first iteration and last iteration, respectively; the values next to the red dots are the relative errors; the numbers in parentheses represent the total iterations.

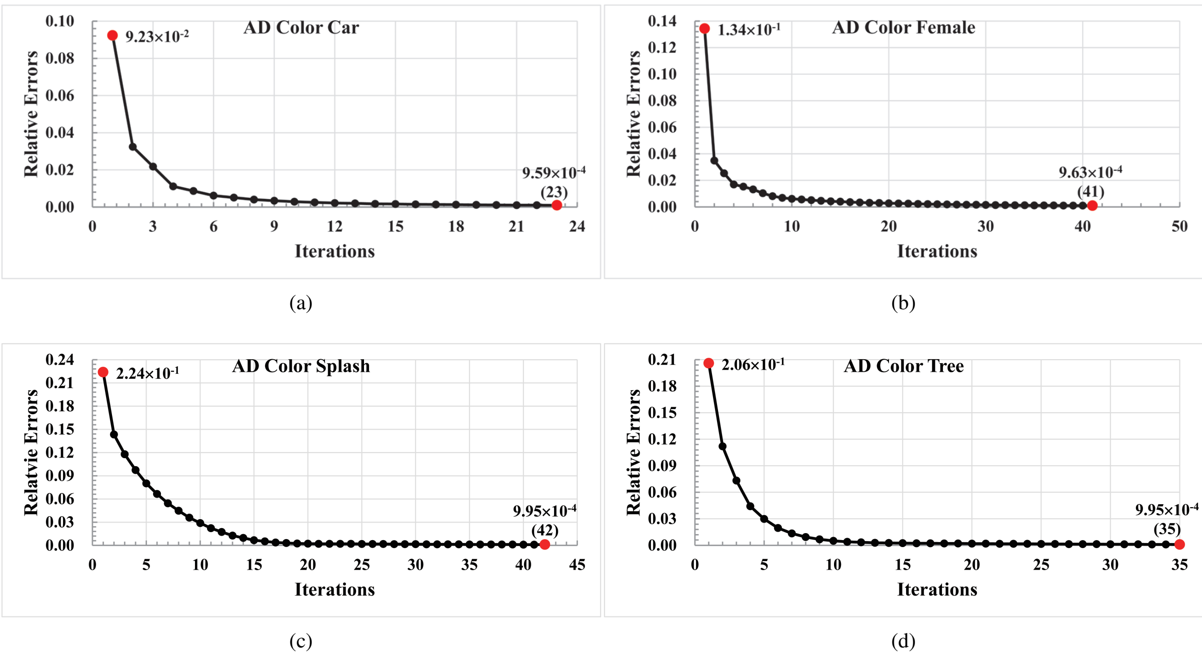


Fig. 12: Evolutions of Relative Errors When Dealing with AD Color Images. The red dots represent the first iteration and last iteration, respectively; the values next to the red dots are the relative errors; the numbers in parentheses represent the total iterations.

RS algorithm and newBM3D algorithm adopt their own default settings and parameters to ensure true evaluations.

The results in Figs. 9 and 10 and TABLEs VII to X explicitly show the effectiveness of the proposed DRALM and its advantages over other algorithms in ISNR, SSIM and visual effects. Similarly, in colr IR experiments, the numerical results of these two criteria also show that our DRALM algorithm not only has overall advantages, but also has advantages in some specific aspects. Visually, compared to other algorithms, the images obtained by DRALM have more salient features and fewer artifacts, which is consistent with the performance of DRALM in ISNR and SSIM.

In terms of speed, we take the mean time for each algorithm to process each degraded image ten times as the final results and record them in TABLEs XI and XII. As shown by the data, our DRALM is significantly faster than the ST algorithm and RS algorithm but slightly slower than BM3D algorithm. Therefore, the proposed DRALM also has strong competitiveness in color IR, especially considering its superiorities under other criteria.

In color IR experiments, we also employ evolutions of relative errors to verify the convergence of DRALM. As shown by the curves in Figs. 11 and 12, the relative errors decrease continuously with iterations, which indicates that with $(k+1)\geq K$ and $\frac{\|x^{k+1}-x\|_2}{\|x\|_2}\leq\tau$, $\{x^{k+1}\}$ obtained by DRALM converge to x. It is also noteworthy that the first several iterations converge quickly and approach the optimization target, so even if the stopping criteria are relaxed, the proposed DRALM can still obtain satisfactory solutions in color IR.

VI. Conclusion

We built a novel model with double l_0-regularization for IR and proposed an efficient solution to solve the model. The proposed DRALM algorithm first obtains several subproblems by equivalent decomposition and then uses effective methods to obtain their closed-form solutions. Subsequently, the convergence of DRALM is analyzed. In the experiment, Our DRALM is applied to IR, and its effectiveness, convergence and advantages are verified by the results. Moreover, the experimental results also show the advantages of l_0-induced sparse priors and multiple-prior models. In the fut ure, we will apply DRALM to image denoising and inpainting.

References

[1] K. Salhi, E. M. Jaara, M. T. Alaoui, and Y T. Alaoui, “Color-texture image clustering based on neuro-morphological approach,” *IAENG International Journal of Computer Science*, vol.46, no.1, pp.134-140, 2019.

[2] A. Desiani, Erwin, B. Suprihatin, S. Yahdin, A. I. Putri and F. R. Husein, “Bi-path architecture of CNN segmentation and classification method for cervical cancer disorders based on pap-smear images,” *IAENG International Journal of Computer Science*, vol.48, no.3, pp.782-791, 2021.

[3] D. N. H. Thanh, V. B. S. Prasath, L. M. Hieu and S. Dvoenko, “An adaptive method for image restoration based on high-order total variation and inverse gradient,” *Signal Image Video P.*, vol.14, no.6, pp.1189-1197, 2020.

[4] U. S. Kamilov, “A parallel proximal algorithm for anisotropic total variation minimization,” *IEEE Trans. Image Process.*, vol.26, no.2, pp.539-548, 2017.

[5] D. Ren, H. Zhang, D. Zhang and W. Zuo, “Fast total-variation based image restoration based on derivative alternated direction optimization methods,” *Neurocomputing*, vol.170, pp.201-212, 2015.

[6] T. Adam and R. Paramesran, “Hybrid non-convex second-order total variation with applications to non-blind image deblurring,” *Signal Image Video P.*, vol.14, no.1, pp.115-123, 2020.

[7] D. Li, X. Tian, Q. Jin and K. Hirasawa, “Adaptive fractional-order total variation image restoration with split Bregman iteration,” *Isa Trans.*, vol.82, pp.210-222, Nov. 2018.

[8] H. Zhang, L. Tang, Z. Fang, C. Xiang and C. Li, “Nonconvex and nonsmooth total generalized variation model for image restoration,” *Signal Process.*, vol.143, pp.69-85, 2018.

[9] R. D. Kongskov, Y. Dong and K. Knudsen, “Directional total generalized variation regularization,” *Bit.*, vo.59, no.4, pp.903-928, 2019.

[10] W. Wang, F. Li and M. K. Ng, “Structural similarity-based nonlocal variational models for image restoration,” *IEEE Trans. Image Process.*, vol.28, no.9, pp.4260-4272, 2019.

[11] F. Xue, F. Luisier and T. Blu, “Multi-Wiener SURE-LET deconvolution,” *IEEE Trans. Image Process.*, vol.22, no.5, pp.1954-1968, 2013.

[12] J. Cai, B. Dong and Z. Shen, “Image restoration: a wavelet frame based model for piecewise smooth functions and beyond,” *Appl. Comput. Harmon. A.*, vol.41, no.1, pp.94-138, 2016.

[13] Z. Zha, X. Zhang, Y. Wu, Q. Wang, X. Liu, L. Tang and X. Yuan, “Non-convex weighted l_p nuclear norm based ADMM framework for image restoration,” *Neurocomputing*, vol.311, pp.209-224, 2018.

[14] J. You, Y. Jiao, X. Lu and T. Zeng, “A nonconvex model with minimax concave penalty for image restoration,” *J. Sci. Comput.*, vol.78, no.2, pp.1063-1086, 2019.

[15] C. Ren, X. He and T. Q. Nguyen, “Adjusted non-local regression and directional smoothness for image restoration,” *IEEE Trans. Multimedia*, vol.21, no.3, pp.731-745, 2019.

[16] Y. Makinen, L. Azzari and A. Foi, “Collaborative filtering of correlated noise: exact transform-domain variance for improved shrinkage and patch matching,” *IEEE Trans. Image Process.*, vol.29, pp.8339-8354, 2020.

[17] N. Yair and T. Michaeli, “An edge driven wavelet frame model for image restoration,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Salt Lake City, 2018, pp.3165-3174.

[18] G. Chierchia, N. Pustelnik, B. Pesquet-Popescu and J. Pesquet, “A nonlocal structure tensor-based approach for multicomponent image recovery problems,” *IEEE Trans. Image Process.*, vo.23, no.12, pp.5531-5544, 2014.

[19] A. F. Vidal, V. De Bortoli, M. Pereyra and A. Durmus, “Maximum likelihood estimation of regularisation parameters in high-dimensional inverse problems: an empirical Bayesian approach part i: methodology and experiments,” *SIAM J. Imaging Sci.*, vol.13, no.4, pp.1945-1989, 2020.

[20] T. Tirer and R. Giryes, “Image restoration by iterative denoising and backward projections,” *IEEE Trans. Image Process.*, vol.28, no.3, pp.1220-1234, 2018.

[21] A. Kheradmand and P. Milanfar, “A general framework for regularized, similarity-based image restoration,” *IEEE Trans. Image Process.*, vol.23, no.12, pp.5136-5151, 2014.

[22] M. Xu, D. Hu, F. Luo, F. Liu, S. Wang and W. Wu, “Limited-angle X-ray CT reconstruction using image gradient l_0-norm with dictionary learning,” *IEEE Trans. Radiat. Plasma Med. Sci.*, vol.5, no.1, pp.78-87, 2021.

[23] Y. Park, L. Reichel, G. Rodriguez and X. Yu, “Parameter determination for Tikhonov regularization problems in general form,” *J. Comput. Appl. Math.*, vol.343, pp.12-25, 2018.

[24] B. K. Gunturk and X. Li, *Image Restoration: Fundamentals and Advances*. CRC Press: Boca Raton, 2018.

[25] D. Calvetti, M. Pragliola, E. Somersalo, and A. Strang, “Sparse reconstructions from few noisy data: analysis of hierarchical Bayesian models with generalized gamma hyperpriors,” *Inverse Probl.*, vol.36, no.2, pp.025010, 2020.

[26] Y. Ding, D. Kunisky, A. S. Wein and A. S. Bandeira, “The average-case time complexity of certifying the restricted isometry property,” *IEEE Trans. Inform. Theory*, vol.67, no.11, pp.7355-7361, 2021.

[27] M. Kaur, D. Singh, V. Kumar and K. Sun, “Color image dehazing using gradient channel prior and guided L_0 filter,” *Inform. Sciences*, vol.521, pp.326-342, 2020.

[28] P. Duan, X. Kang, S. Li, P. Ghamisi and J. A. Benediktsson, “Fusion of multiple edge-preserving operations for hyperspectral image classification,” *IEEE Trans. Geosci. Remote*, vol.57, no.12, pp.10336-10349, 2019.

[29] M. Arif and G. Wang, “Fast curvelet transform through genetic algorithm for multimodal medical image fusion,” *Soft Comput.*, vol.24, no.3, pp.1815-1836, 2020.

[30] N. G. Polson, J. G. Scott and B. T. Willard, “Proximal algorithms in statistics and machine learning,” *Statist. Sci.* , vol.30, no.4, pp.559-581, 2015.

[31] P. Jain and P. Kar, “Non-convex optimization for machine learning,” *Found. Trends Mach. Le.*, vol.10, no.3-4, pp.142-336, 2017.

[32] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory In Hilbert Spaces*. New York: Springer, 2017.

[33] S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng and L. Zhang, “Weighted nuclear norm minimization and its applications to low level vision,” *Int. J. Comput. Vis.*, vol.121, no.2, pp.183–208, 2017.