

Research Article

Adaptive Sparse Norm and Nonlocal Total Variation Methods for Image Smoothing

Qiegen Liu,¹ Biao Xiong,² and Minghui Zhang¹

¹Department of Electronic Information Engineering, Nanchang University, Nanchang 330031, China

²Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Netherlands

Correspondence should be addressed to Qiegen Liu; liuqiegen@hotmail.com

Received 28 September 2014; Accepted 4 December 2014; Published 28 December 2014

Academic Editor: Mohamed A. Seddeek

Copyright © 2014 Qiegen Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In computer vision and graphics, it is challenging to decompose various texture/structure patterns from input images. It is well recognized that how edges are defined and how this prior information guides smoothing are two keys in determining the quality of image smoothing. While many different approaches have been reported in the literature, sparse norm and nonlocal schemes are two promising tools. In this study, by integrating a texture measure as the spatially varying data-fidelity/smooth-penalty weight into the sparse norm and nonlocal total variation models, two new methods are presented for feature/structure-preserving filtering. The first one is a generalized relative total variation (i.e., GRTV) method, which improves the contrast-preserving and edge stiffness-enhancing capabilities of the RTV by extending the range of the penalty function's norm from 1 to $[0, 1]$. The other one is a nonlocal version of generalized RTV (i.e., NLGRTV) for which the key idea is to use a modified texture-measure as spatially varying penalty weight and to replace the local candidate pixels with the nonlocal set in the smooth-penalty term. It is shown that NLGRTV substantially improves the performance of decomposition for regions with faint pixel-boundary.

1. Introduction

Many practical applications of computer vision and computer graphics involve estimating spatially varying image contents from noisy original images. One important requirement of such estimation is feature-preserving filtering, which has become a fundamental tool in many applications. Unfortunately, feature-preserving filtering is inherently challenging because it is usually difficult to distinguish useful features from noise. There are a number of different feature-preserving filtering methods, which can be roughly classified into two categories, that is, the spatial filter [1–6] and variational model [7–10]. These unified formulations simply decompose a given image into structures and details by smoothing the image, simultaneously preserving or even enhancing image edges. The differences among them lie in the ways that edges are defined and this prior information is used to guide the smoothing process.

The anisotropic diffusion model [1] falls into the category of spatial filters. It employs a PDE-based formulation in

which pixel-wise spatially varying diffusivities are estimated from image gradients. The gradient of the filtered image guides the diffusion process in order to avoid smoothing edges. Bilateral filtering (BF) [2, 11] is another widely used model for removing noise from images while simultaneously performing detail flattening and edge preservation. It averages the nearby pixels by calculating weights from spatial and range domains [2] and smoothes low-contrast regions while preserving high-contrast edges. Due to its simplicity and effectiveness, BF has been successfully applied to many computational photography applications [11–14]. By taking advantage of the local linear relationship between the guidance image and the filtering output at a small neighboring window, a spatially guided filter (GF) was developed in [15]. In [3], the authors proposed a relatively simple technique to decompose an image into structures and oscillatory texture components by using a pair of nonlinear low-pass and high-pass filters. Inspired by the fact that more pixels may be included by extending the concept of neighborhood in a nonlocal way [16], smoothing at nonlocal/semilocal regions

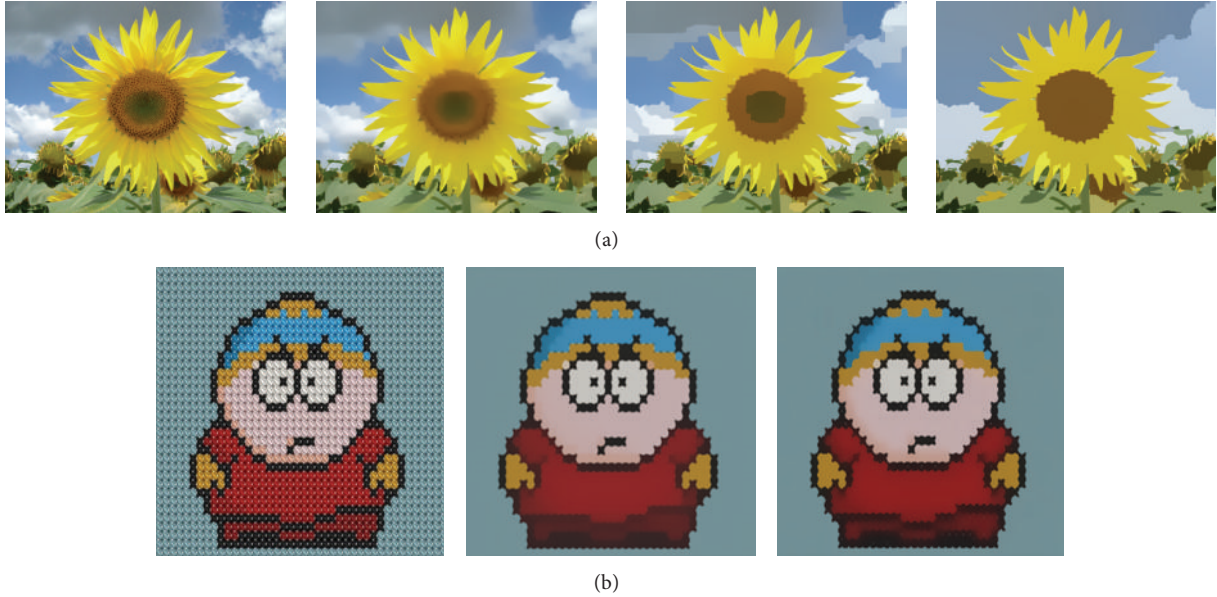


FIGURE 1: Examples of weighted total variation filtering results by using the sparse norm and nonlocal pixel sets. (a) From left to right are the input images and the local method results with p -norm of 1, 0.5, and 0.1. (b) From left to right are the input images and the local method results with p -norm of 1 and the nonlocal version of 0.1. Smaller p -values give more prominent edge sharpening results while the nonlocal method preserves the shade information well.

has drawn much attention in recent years [4–6]. The Regcovsmooth scheme developed by Karacan et al. used the second-order statistic descriptor region covariance as a similar weight to average the pixels in a squared neighborhood [4]. Ye et al. proposed a sparse norm filter (SNF) defined in square neighborhood [6]. Rousselle et al. used the variant of the nonlocal means filter for adaptive rendering [5].

Variational models are based on the energy minimization formulation with a data fidelity term and a smoothing-penalty term [7–10]. The data term measures the difference between the filtered signal s and the original signal f , while the smoothing term measures the extent to which the filtered signal is not piece-wise smooth. That is,

$$s = \arg \min_s \{G(f, s) + \lambda J(s)\}, \quad (1)$$

where λ is a nonnegative parameter controlling the weight of the smoothing term. While the design of the data fidelity term $G(f, s)$ is usually relatively simple (for instance, the squared L2 distance between the filtered signal and the original signal $\|f - s\|_2^2$ is often used), the choice of the smoothing-penalty term $J(s)$ is critical. A representative piece of work is the total variation (TV) [7], which uses L1-norm based regularization constraints to penalize large gradient magnitudes. In its original formulation, the TV model provides fairly good separations of structures from textures. Further studies extended the standard TV formulation with different norms for both data fidelity and smoothing-induced regularization terms and demonstrated that more robust norms could improve the performance of image decomposition [8–10]. The success of the weighted least square (WLS) optimization method proposed by Farberman et al. in [8] is attributed in part to

the l_p norm in the iterative reweighted least square (IRLS) framework. Recently, the L0 smoothing presented by Xu et al. [9] used l_0 term instead of l_1 term to directly measure the gradient sparsity in the context of image smoothing and achieved promising results. Later, they proposed a special method, called relative total variation (RTV) [10], dealing with textured images. The efficiency of RTV depends heavily on the defined windowed total variation (WTV) and windowed inherent variation (WIV) involving spatial information.

In summary, most existing image decomposition models aim to extract structures from textures and noise while preserving edges and each of them has its limitations. For instance, BF, WLS, and L0 smoothing work poorly for tackling images with nonuniform texture details, while RTV and Regcovsmooth may blur the edges for images containing cartoon patterns. In this work, we propose a general relative total variation (GRTV) method and its nonlocal version (i.e., NLGRTV). Motivated from the reinterpretation of L0 smoothing from the IRLS iterative-solver viewpoint, the key idea of the proposed GRTV and NLGRTV methods is to incorporate the sparse norm and nonlocal strategy into the model, taking advantage of both spatial filters and variational models. They are very different from RTV, which is based on statistical verification, in the following two senses. First, GRTV extends the norm of WTV from 1 to $[0, 1]$ and sets the norm of WIV inversely proportional to the norm of WTV. As a benefit, the amplitudes of noise-like structures will become smaller and the global edges will become more prominent and sharper when p -values are getting smaller, which is shown in Figure 1 (seen from left to right in the top line). Second, the nonlocal GRTV (NLGRTV) uses a different

smooth-penalty weight and a nonlocal sparse penalty term. By integrating the nonlocal information into the smoothing process, NLGRTV can distinguish well textures from the main structures and the information of shade is well preserved as shown in the bottom line of Figure 1. All these nice features of the proposed GRTV and NLGRTV are demonstrated later on through several applications including structure edge detection and texture enhancement, image abstraction and composition, and artifact removal.

2. Background and Motivation

In this section, two representative image smoothing techniques, L0 smoothing and RTV, will be briefly reviewed, with one focusing on the sparse and robust norm and the other emphasizing the spatially varying total variation measure. Subsequently their strengths and drawbacks are discussed in detail. Finally, the analysis and reinterpretation of the L0 smoothing model from the perspective of IRLS iterative-solver are derived, which leads to the motivation of our new method.

2.1. Previous Works: L0 Smoothing and RTV Model. In [9], Xu et al. proposed an image smoothing method based on L0 gradient minimization. Let ∇s denote the gradients of s ; its energy function can be defined as follows:

$$s = \arg \min_s \left\{ \|f - s\|_2^2 + \lambda |\nabla s|_0 \right\}, \quad (2)$$

where the data term is the squared L2 distance between f and s and the smooth term is the L0 norm of ∇s . The L0 norm of a vector is the number of nonzero value, which directly measures the sparsity. Compared to the l_p ($0 < p < 1$) norm regularization such as WLS filter [6, 8], L0 smoothing can remove low-amplitude structures and globally preserve and enhance salient edges, even if they are the boundaries of very narrow objects. Like other edge-preserving smoothing approaches [1, 2, 7, 8], one main drawback associated with L0 smoothing is that the local contrast-based definition of edges may fail to capture high-frequency components that are related to fine image details or textures, because differences in the brightness values or gradient magnitudes are employed as the main cues for edge indicator at an image pixel, and this information is used to guide the smoothing process. This means that L0 smoothing cannot separate completely textured regions from the main structures as they are treated as part of the structure. Recently some advanced algorithms were proposed; for example, an L0 smoothing-L1 method was presented in [17], where a L1-defility term instead of L2-norm was used to deal with heavy noise. In [18], rather than using L0-L2 iteration algorithm to solve the model, the L0 smoothing_FCD (fused coordinate descent) algorithm iteratively repeated coordinate descent step and fusion step to approximately solve the model.

In [10], a method extracting structure from textured image was developed by employing relative total variation (RTV). For the purpose of better texture removal, the developed RTV model does not assume the type of textures in advance; instead it introduces a novel map of windowed

inherent variation (WIV). The WIV map in a region that only contains texture is generally smaller than that in a region also including structural edges. Applying WIV map as a weight of TV model in the vertical and horizontal directions yields adaptive edge-preserving and texture removal. Mathematically, the objective function can be expressed as

$$s = \arg \min_s \left\{ \|f - s\|_2^2 + \lambda \sum_i \left[\frac{D_x s(i)}{L_x s(i) + \varepsilon} + \frac{D_y s(i)}{L_y s(i) + \varepsilon} \right] \right\}, \quad (3)$$

where $D_x s(i) = \sum_{j \in R(i)} g_{i,j} |\partial_x s_j|$ and $L_x s(i) = \sum_{j \in R(i)} |g_{i,j} \partial_x s_j|$ are the weights, $g_{i,j}$ is a Gaussian weighting function [10], i and j are pixel indexes, and x and y are pixel coordinates. The second term enforces the structure part to be sparse in the gradient domain. RTV is modeled as variational formulation and its texture suppression weights are spatially variant. However, Despite having good performance in edge-preserving and texture-eliminating, its discriminative power is still limited.

2.2. Analysis and Motivation: Reinterpretation of L0 Smoothing from the IRLS Iterative-Solver. The numerical L0 gradient minimization in L0 smoothing is implemented by the alternating optimization strategy with half-quadratic splitting. Alternatively, it can also be solved by iteratively reweighted least squares (IRLS), which approximates the original minimization by iteratively solving weighted quadratic problems. That is,

$$s^{k+1} = \arg \min_s \left\{ \|f - s\|_2^2 + \lambda \sum_i \frac{|\nabla s_i|_2^2}{|\nabla s_i^k|_2^2} \right\}. \quad (4)$$

In the case of one dimension and for each pixel index i , (4) turns to

$$\begin{aligned} & s_i^{k+1} \\ &= \arg \min_s \left\{ \|f_i - s_i\|_2^2 + \lambda \left[\frac{(s_i - s_{i-1})^2}{|\nabla s_i^k|_2^2} + \frac{(s_{i+1} - s_i)^2}{|\nabla s_{i+1}^k|_2^2} \right] \right\} \\ &= \frac{f_i + \lambda \sum_{j=i-1, i+1} h_{ij}^k s_j}{1 + \lambda \sum_{j=i-1, i+1} h_{ij}^k}, \quad h_{ij}^k = \frac{1}{\|s_i^k - s_j^k\|_2^2}. \end{aligned} \quad (5)$$

The importance of rewriting (4) to (5) is that we can analyze the iterative evolution of each pixel in detail. The weight h_{ij}^k measures the similarity of pixels s_i^k and s_j^k . Generally, the pixel in the image can be classified in the flat/smooth region, edge boundary, and ‘‘high variance’’ area. For the pixel in the smooth region or edge region, there must be some similar pixels hence $\sum_{j=i-1, i+1} h_{ij}^k$ to be large, therefore $s_i^{k+1} \approx \sum_{j=i-1, i+1} h_{ij}^k s_j / \sum_{j=i-1, i+1} h_{ij}^k$. For the pixel in ‘‘high variance’’ area, there may be no pixel similar to the reference pixel and thus its update value largely depends on f_i . Totally, as the iterative progresses, pixels at some regions change/diffuse

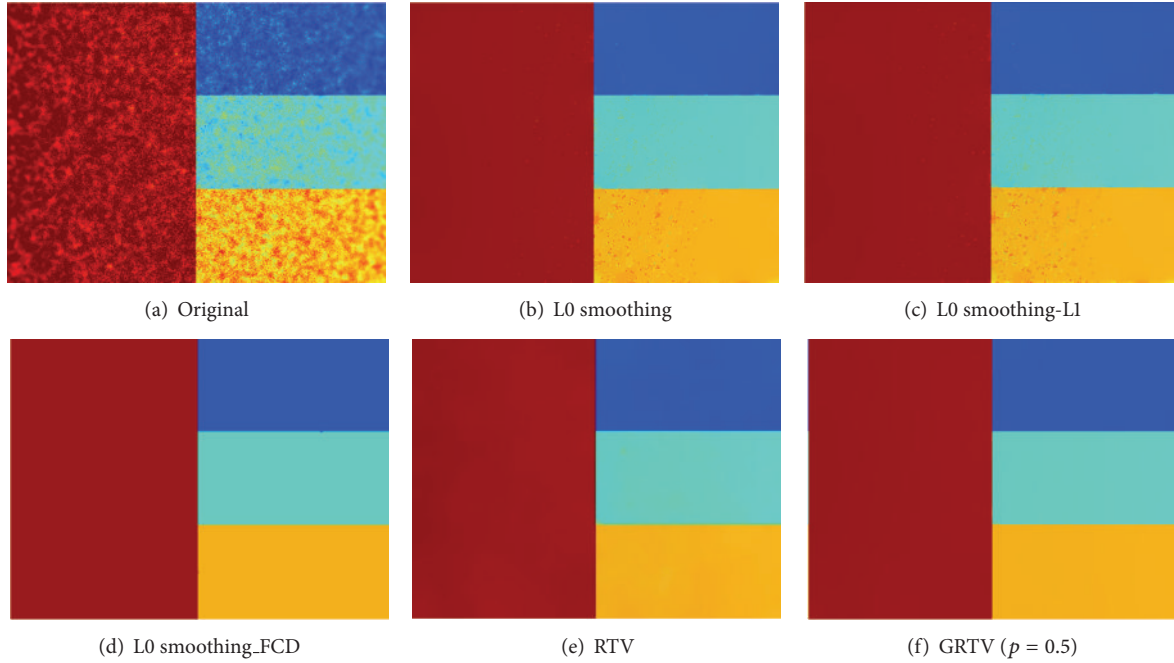


FIGURE 2: Comparison with state-of-the-art methods on image created by Farbman et al. [8]. Since L0 smoothing targets globally preserving salient structures, even if they are small in resolution, it fails to remove the noise with large magnitude. L0 smoothing produces speckle-like artifacts. Although L0 smoothing_FCD improved the noise insensitivity ability compared to that of L0 smoothing-L1 and L0 smoothing, it still retained some points near the edge, distorting the edge boundary. The result of RTV is slightly blurred. Our method performs the best to remove high-frequency noise while preserving major edges.

much faster while some others change lately. One example is illustrated in Figure 2, where L0 smoothing produces speckle-like artifacts for dealing with image corrupted by heavily noise. For better tackle with limits in (5), some points need to be fully considered for the weight term and candidate neighbor pixel sets as follows.

- (i) The weight term $\sum_{j=i-1,i+1} h_{ij}^k$ and h_{ij}^k may not be designed adaptively well in its definition. It accounts for not only the fidelity balance but also the measurement between the nearby pixels. Its functional can be better modified by spatially varying adaptation. In this sense, both L0 smoothing-L1 method and RTV can be seen as the improvements along this direction.
- (ii) The estimate depends not only on the weights but also on the candidate pixel sets. Nonlocal property is a universal prior in the community of image processing. L0 smoothing_FCD can be seen as its improvement by adding fusion step. We note that this discussion has universal guiding meaning. Recently proposed method SNF [6], Regcovsmooth [4], and adaptive NLM rendering algorithm [5] all utilized this property.

3. Approaches

In this work, our aim is to smooth a given image to suppress the textures to the most extent meanwhile retaining the edge sharpness. In general, from the viewpoint of iterative

evolution, the change of each pixel s_i^k depends on three factors: the weight of data fidelity/smoothing-penalty to the original pixel f_i and how to define the candidate pixel sets and the weight of its relation to the neighboring pixels. In the following, we exploit the sparse norm and nonlocal scheme as two bridges to achieve the purpose of both properties of edge-preserving and texture-smoothing. Specifically, the proposed local Model 1 utilizes the sparse norm of the total variation and data fidelity measure to calculate the degree of the weight of data fidelity and its relation to the nearby pixels. By incorporating the nonlocal strategy into Model 1 and modifying the weight of data fidelity, the nonlocal Model 2 is derived.

3.1. Model 1: The Local Model GRTV. One possible objective function we choose is expressed as follows:

$$s = \arg \min_s \left\{ \|f - s\|^2 + \lambda \sum_i \left[\frac{D_x^p s(i)}{L_x^{2-p} s(i) + \varepsilon} + \frac{D_y^p s(i)}{L_y^{2-p} s(i) + \varepsilon} \right] \right\}. \quad (6)$$

The proposed method jointly enhances edge and eliminates texture with $0 < p \leq 1$. Intuitively, on one hand, the operators D_x^p and D_y^p with $0 < p < 1$ have bigger edge-preserving ability than that with $p = 1$. On the other hand, L_x^{2-p} and L_y^{2-p} with $0 < p < 1$ will pose better texture-suppressing capability than that with $p = 1$.

In the following content, we will discuss the solver of general model. After some mathematical manipulations, the smooth term in x -direction can be written as

$$\sum_i \frac{D_x^p s(i)}{L_x^{2-p} s(i) + \varepsilon} \quad (7a)$$

$$= \sum_i \sum_{j \in R(i)} \frac{g_{i,j}}{\sum_{j \in R(i)} |g_{i,j} (\partial_x s)_j|^{2-p} + \varepsilon} |(\partial_x s)_i|^p$$

$$\approx \sum_i \sum_{j \in R(i)} \frac{g_{i,j}}{L_x^{2-p} s(i) + \varepsilon} \frac{1}{|(\partial_x s)_i|^{2-p} + \varepsilon} (\partial_x s)_i^2$$

$$= \sum_i u_{x,i}^{2-p} w_{x,i}^{2-p} (\partial_x s)_i^2, \quad (7b)$$

where ε is used to prevent numerical instability. Equation (7b) is approximated by the Iteratively Reweighted Norm (IRN) approach proposed by Rodríguez and Wohlberg [19], which is closely related to the iteratively reweighted least squares (IRLS) method [20, 21] and widely used in compressed sensing and image processing fields. It has been proven that IRN is one kind of Majorization-Minimization (MM) method [19], which involves good convergence property. The weights $u_{x,i}^{2-p}$ and $w_{x,i}^{2-p}$ are, respectively,

$$u_{x,i}^{2-p} = \sum_{j \in R(i)} \frac{g_{i,j}}{L_x^{2-p} s(i) + \varepsilon} = \left(G_\sigma * \frac{1}{|G_\sigma * \partial_x s|^{2-p} + \varepsilon} \right)_i, \quad (8)$$

$$w_{x,i}^{2-p} = \frac{1}{|(\partial_x s)_i|^{2-p} + \varepsilon_s}, \quad (9)$$

where G_σ is a Gaussian filter with standard deviation σ and ε_s is a small positive number to avoid division by zero. The division in (8) is point-wise and $*$ is the convolution operator. Equation (9) indicates that $u_{x,i}^{2-p}$ at pixel point i incorporates neighboring gradient information in an isotropic spatial filter manner. Meanwhile, $w_{x,i}^{2-p}$ in (9) is only related to the pixel-wise gradient. The weights (8) and (9) pose the neighboring and the pixel constraint for the same point i . The penalty in the y -directional dimension is the same as that in x -directional dimension.

In summary, (6) can be written in a matrix form as follows:

$$v_s = \arg \min_{v_s} \left\{ (v_f - v_s)^T (v_f - v_s) + \lambda (v_s^T C_x^T U_x W_x C_x v_s + v_s^T C_y^T U_y W_y C_y v_s) \right\}, \quad (10)$$

where v_s and v_f are the vector representation of s and f , respectively. C_x and C_y are the Toeplitz matrices from the discrete gradient operators with forward difference. U_x , W_x , U_y , and W_y are diagonal matrices whose elements are defined in (8) and (9). Similar to RTV, (6) is solved by a reweighted strategy; that is, update the block variables U_x , W_x , U_y , and W_y as a function of v_s calculated from the previous iteration

and then update v_s by minimizing (10) with the last values of U_x , W_x , U_y , and W_y . Practically speaking, using $L^k = C_x^T U_x^k W_x^k C_x + C_y^T U_y^k W_y^k C_y$, the minimization of (6) is

$$(1 + \lambda L^k) v_s^{k+1} = v_f. \quad (11)$$

Since $(1 + \lambda L^k)$ is a symmetric positive definite Laplacian matrix, efficient solvers are available for it [10].

RTV is a special case of GRTV with $p = 1$. As expected, the extended range with $0 < p < 1$ will substantially improve the geometry and texture separation ability. A visual illustration is displayed in Figure 3. As the parameter p decreases, the weights $u_{x,i}^{2-p}$ and $w_{x,i}^{2-p}$ of texture and edge differ more. Figures 3(b) and 3(c) correspond to the case of $p = 1$ and $p = 0.5$, respectively. It can be observed that both filters smooth small fluctuations and meanwhile preserve edges. However, as compared in Figures 3(e) and 3(d), the estimates of the textural regions with $p = 0.5$ are bigger than with $p = 1$. At the same time the edge value with $p = 0.5$ is more stiffness.

3.2. Model 2: The Nonlocal Model NLGRTV. Recently, the nonlocal similarity property inherent in nature image has been successfully applied in many image processing tasks due to the work of nonlocal means (NLM), which averages over pixels value in nonlocal neighbor region with similar patch structure [16]. Later this idea was successfully generalized to variational framework based on nonlocal operators [22] and the nonlocal L^p ($0 < p \leq 1$) patch regression [23, 24]. In this subsection, we consider introducing the nonlocal configuration. Unlike in GRTV, in this case, the weight $H^{2-p} s(i) = \sum_{j \in R(i)} (g_{i,j} / (L_x^{2-p} s(i) + L_y^{2-p} s(i) + \varepsilon))$ is viewed as a weight of spatially varying smooth-penalty term. The objective function can be expressed as

$$s = \arg \min_s \left\{ \|f - s\|^2 + \lambda \sum_i H^{2-p} s(i) \sum_{j \in \Omega_i} |s(i) - s(j)|^p \right\}, \quad (12)$$

where Ω_i denotes a squared neighborhood set centered at pixel i and of size $(2r + 1) \times (2r + 1)$ pixels. Similar to the technique used in Section 3.2, that is, following the reweighted norm approximation, for each pixel we have

$$s_i^{k+1} = \arg \min_{s_i} \left\{ \|f_i - s_i\|^2 + \lambda H^{2-p} s(i) \times \sum_{j \in \Omega_i} \frac{1}{|s_i^k - s_j^k|^{2-p} + \varepsilon_s} (s_i - s_j)^2 \right\}. \quad (13)$$

Accordingly, the update of s_i^{k+1} is achieved by

$$s_i^{k+1} = \frac{f_i + \lambda H^{2-p} s^k(i) \sum_{j \in \Omega_i} (1 / (|s_i^k - s_j^k|^{2-p} + \varepsilon_s)) s_j}{1 + \lambda H^{2-p} s^k(i) \sum_{j \in \Omega_i} (1 / (|s_i^k - s_j^k|^{2-p} + \varepsilon_s))}. \quad (14)$$

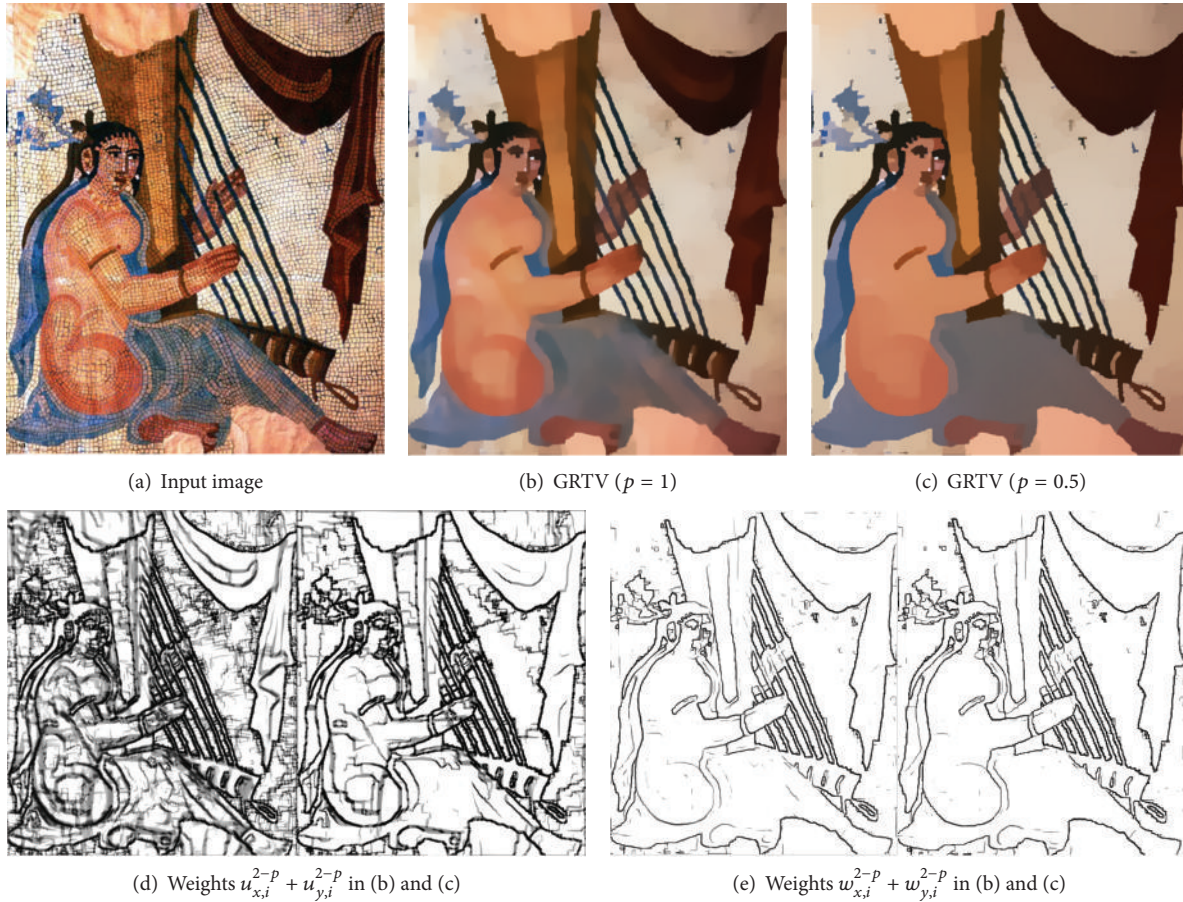


FIGURE 3: Demonstration of the weights $u_{x,i}^{2-p} + u_{y,i}^{2-p}$ and $w_{x,i}^{2-p} + w_{y,i}^{2-p}$ versus p -value.

Our alternating update algorithms GRTV and NLGRTV are sketched in Algorithm 1, where the solution variable and weight variables are updated alternately.

Figure 4 shows the advantage of extending local pixel sets to nonlocal candidate data. As can be observed, both GRTV and NLGRTV improve the visual perception of the input image than that of RTV. However, since the texture patterns in the right-bottom of the original image are very similar and the magnitude is small, the GRTV with $p = 0.5$ makes their low color difference almost to zero. NLGRTV with $p = 0.5$ alleviates this drawback. Although the pixel magnitude nearby the edge boundary is small but the pixel number is large, subsequently NLGRTV still retains these structure objects in the geometry component.

3.3. Iterative Reweighted Least Square (IRLS) Framework and the Computation Cost. From the view of numerical optimization, RTV, the developed GRTV and NLGRTV fall into the category of IRLS framework. They are closely related to the WLS filter, which can be viewed as a single iteration of the IRLS framework. In WLS, the weight is $1/((\partial_x \log(s))_i)^{2-p} + \epsilon$ ($0 \leq p \leq 0.8$), which is a modification of $1/((\partial_x s)_i)^{2-p} + \epsilon$. Though the convergence of the proposed

- (1) For $k = 0$ to $K - 1$ do
- (2) update the weights according to (8) and (9)
- (3) update v_s^{k+1}/s_i^{k+1} according to (11)/(14)
- (4) End (For)

ALGORITHM 1: GRTV/NLGRTV.

GRTV and NLGRTV may not be proved strictly as in [19], our numerical experiments demonstrate good convergence behaviors. Figure 5 shows one example of the input image depicted in Figure 3(a). The intermediate structure images obtained by GRTV and NLGRTV with $p = 1$ are displayed, where we can find the proposed method quickly updates the saliency sharpened image in iterations. It indicates the effectiveness of the IRLS strategy adopted by our method. By the way, empirical experiments show that, for both GRTV and NLGRTV, the smaller the value of p and the bigger the value of r , the more iterations the algorithm needed to converge.

The computation cost of GRTV is the same as that of RTV. that is, at each iteration, the computational cost of (8) is linear with respect to problem size with $O(\sigma N)$ by

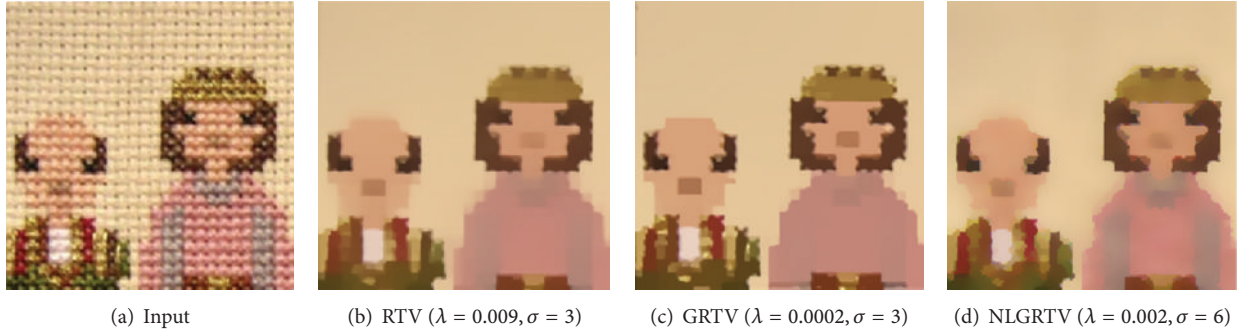


FIGURE 4: Demonstration of introducing nonlocal strategy to achieve better shade-preserving smoothing result.



FIGURE 5: Structure images in different iterations obtained by the algorithm GRTV (top line) and NLGRTV (bottom line).

taking advantage of the nice separation property of Gaussian kernel, where N is the total number of pixels in an image. As stated in [10], the step of calculating v_s is to solve a linear system with a five-point spatially inhomogeneous sparse Laplacian matrix. Fast solvers such as the multiresolution preconditioned conjugate gradient (PCG) can reach $O(N)$ complexity. In all the experiments of this paper, it usually takes 3 s to process an 800×600 color image on an Intel Corei3CPU@2.93 G with our matlab implementation. For the computation cost of NLGRTV, it differs from GRTV mainly on the updating of s_i^{k+1} in (14). Admittedly, the complexity of computing (14) is almost the same as that in SNF [6]; that is, it can be calculated in linear time $O(BN)$ by employing off-the-shelf acceleration methods [25, 26], where B is the number of bins for quantization and the pixel number N .

4. More Analysis and Comparisons

In this section, we reveal the differences between our proposed methods and recently developed L0 smoothing, RTV, and RegionCov. We also compare our methods with some existing edge-aware image smoothing methods, including BF and WLS, and some advanced improvements of L0 smoothing. Specifically, we have hand-tuned parameters carefully for these methods to achieve the best performance.

4.1. Performance on Varying Values of p and r . Parameter p controls the degree of sparse norm and r controls the search window size. The effect of varying parameters is demonstrated in Figure 6. Seen from Figure 6, it can be found from the second and third rows that both GRTV and



FIGURE 6: Effect of varying parameters. Smaller p -values give more prominent edges sharpening results and nonlocal way in favor of preserving shade information.

NLGRTV make the image more sharpen. Furthermore, the results in the third row are more smooth and robust than the second line due to the nonlocal pixel diffusion. By the way, although L0 smoothing method uses the penalty with norm $p = 0$, its smoothing result may not have visual effect due to the lack of adaptive smooth-penalty weighting term. According to the figure, it is expected that the user can obtain the specific effect by tuning parameters p and r . In the following content of this paper, we set $r = 2$ for all the experiments.

4.2. Relationship to L0 Smoothing and Comparisons with Its Improved Versions. GRTV has a close relation with L0 smoothing method; GRTV differs L0 smoothing mainly from the additional weight $w_{x,i}^{2-p}$ and $w_{y,i}^{2-p}$. The test image shown in Figure 7 contains textural sword and slightly smooth clouds with small magnitude. The result of RTV in Figure 7(c) cannot simultaneously maintain the lines on the girl's skirt and smooth the sword. With $p = 0.5$, our GRTV largely

improves this simplification ability. For this test image, the behavior of GRTV is very close to that of L0 smoothing by specifying $\lambda = 0.0004$ and $\sigma = 1$. As seen from Figure 7(d), the boundary of the cloud obtained by our method is more rich and clear. Besides, by tuning the parameters λ and σ , our method yields meaningful decomposition as depicted in Figures 7(e) and 7(f).

We can consider the relation between L0 smoothing, RTV, and GRTV from the global and local respects. L0 smoothing uses L0 gradient minimization, which can globally control how many nonzero gradients are involved to approximate prominent structure in a sparsity-control manner. It does not depend on local features but instead globally locates important edges. RTV mainly emphasizes the texture-weights, which is locally defined. Our presented method GRTV falls in between and inherits the advantages of both algorithms in an approximate way.

The flower shown in Figure 8 contains smooth branches and leaves and flower bud with small magnitude. The result of L0 smoothing, RTV, and L0 smoothing-L1 cannot balance

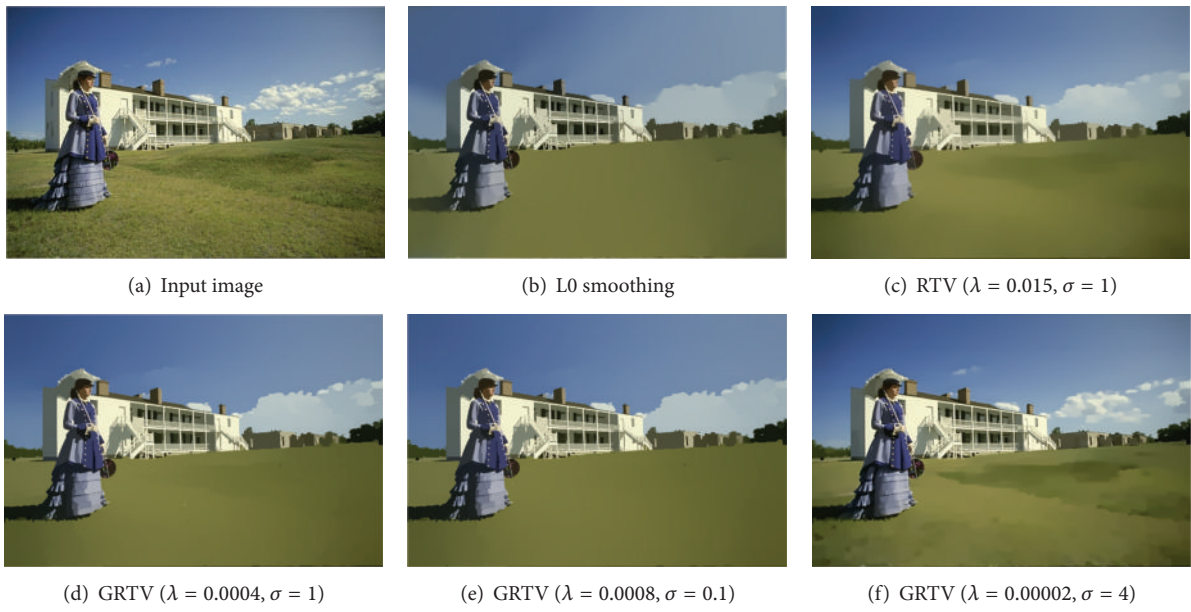


FIGURE 7: Smoothing results produced by L0 smoothing, RTV, and GRTV.



FIGURE 8: Smoothing comparison with advanced method of L0 smoothing, RTV, and GRTV.

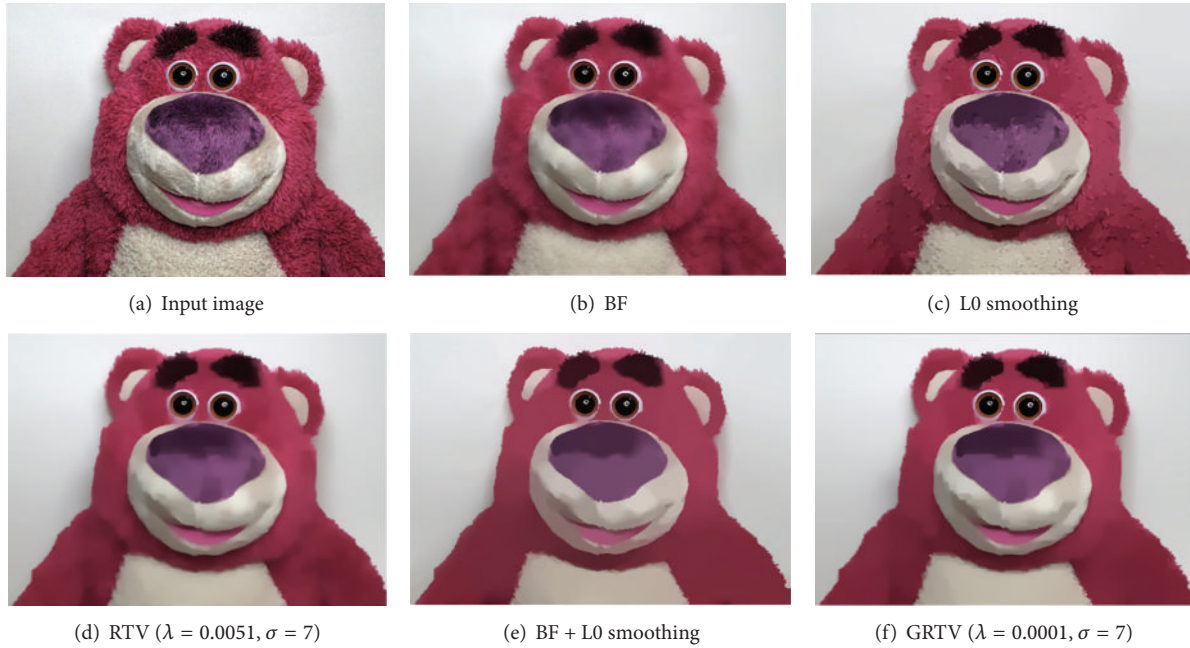


FIGURE 9: Smoothing comparison with combined BF and L0 smoothing.

detail smoothing and edge enhancing. Since L0 smoothing_FCD uses the fuse technique, that is, a group of variables rather than a single variable to be updated, it may cause overfitting. On the other hand, our method GRTV obtains a visually satisfactory result for better preserving the contrast around edges than the original RTV filter.

As stated in [9], although L0 smoothing cannot efficiently tackle texture images, it still produces novel effects by applying first the local filtering BF and then the global filter L0 smoothing. Figure 9 shows the example reproduced in [9], where BF yields overblurring result and L0 smoothing maintains fluff textures. However, combining them in a consecutive way yields an image with sharpen prominent edges, as shown in Figure 9(e). In Figure 9(d) RTV gives a slightly blurred image. Finally, we can observe that our result with $p = 0.5$ in Figure 9(f) sharpens the large-scale salient edges more than that of RTV.

4.3. Relationship to RTV and Comparisons with Related State-of-the-Art Methods. Essentially, both RTV and our methods use a weight function which is very similar to the fast cartoon + texture method by Buades et al. [3]. In [3], they introduced the relative reduction rate of local total variation as a weight function to measure to what extent the considered point belongs to a textured region. One main drawback of this filter is that its output result is directly computed by weighted average of the original image and a low-pass filtered image. However, RTV and our methods work in an iterative scheme that estimates the weights adaptively.

When visually evaluating the performance of standard decomposition method, one important criterion is to consider how strong the remaining parts of one component on the image of another component. For a part of the image

“Barbara,” we can see in Figures 10(b) and 10(c) that Subr et al., 2009 [27] and Buades et al., 2010 [3] method cannot completely eliminate the texture from the table cover. RTV also provides similar results [10]. The proposed method NLGRTV with $p = 1$ gives well-separated results and exhibits contrast-preserving feature, by eliminating the texture from the cartoon (see Figure 10(e)). Lowering the value of p to be 0.5, its output image changes to be more sharpening, as presented in Figure 10(f).

In the following, some examples with ample nonuniform and anisotropic texture are shown to exhibit the flexibility of our method. Note that the TV model, BF, and WLS filters were used in natural image smoothing and do not have effective terms to tackle textures. L0 smoothing also has limitation in dealing with the “structure + texture” images, since it usually cannot suppress the texture component. Figure 11(a) shows the image “Bishapur_zan,” in which the main structures are surrounded by the background formed by many tiles with salient but fine boundaries. From Figures 11(d) to 11(f), it can be observed that our method NLGRTV with varying parameters $p = 1, 0.5, 0.1$ can attain satisfied results with different styles. By tuning the regularization parameter, NLGRTV well preserves long coherent and prominent edges, meanwhile eliminating the outliers and irregular patterns.

Figure 12(a) shows another input image and the result from RTV method is presented in Figure 12(b). Despite making use of local signed gradients and exhibiting special properties, RTV fails to preserve some fine structures and blurs main edges compared to that of our GRTV method with $p = 0.5$ shown in Figure 12(c). Additionally, the decomposition of GRTV preserves the contrast between the girl and the background better. Our edge-preserving texture-smoothing filtering method NLGRTV with $p = 0.5$ improves

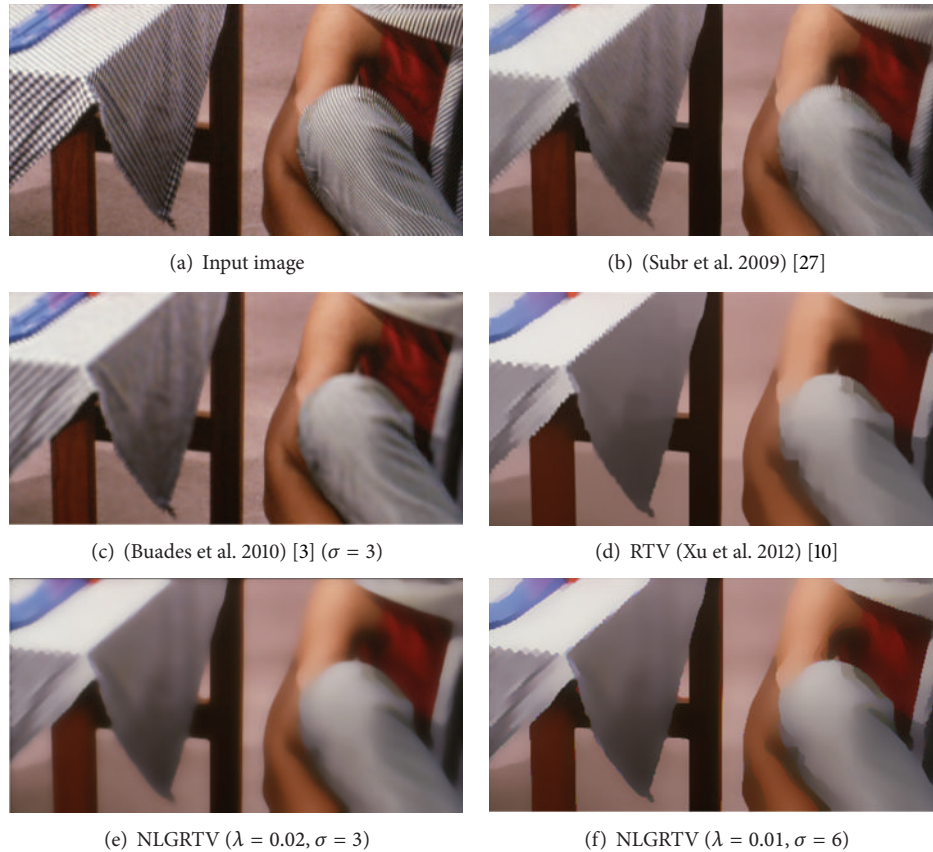


FIGURE 10: Comparison with the extracted structures from the image, Barbara image, by different methods. Our method in (e) extracts the structure from the texture well while preserving shading information.

the visibility of input image than the other two algorithms. It can effectively smooth the white dots around the edges that still remain in Figures 12(b) and 12(c) while keeping edges sharp.

4.4. Comparison with Nonlocal-Based Methods. Recently, the emergence of new nonlocal-based filters which have provided new insights and attained potential performances for image smoothing problem has been witnessed. Sparse norm filter (SNF), Regcovsmooth, and our NLGRTV all use the nonlocal neighborhood pixels. Specifically, the latter two approaches focus on weight adaption to achieve texture-structure separation. The Regcovsmooth method built on the patch-based nonlocal region covariance descriptor similarity framework, considering the region covariance matrices as image features to estimate the similarity between two image patches. Instead, our method NLGRTV lies on the nonlocal pixel-difference framework. An illustration is depicted in Figure 13. As can be observed, the nonlocal candidate scheme makes both Regcovsmooth and NLGRTV better preserve shading information than RTV. Besides, our method with $p = 1$ better preserves the edge geometries such as the regions in eye and cheek. Note that the search neighborhood size in Regcovsmooth is 21×21 while ours is only 5×5 .

5. Applications

The incorporation of sparse norm and nonlocal strategy in variational model largely improves the ability of the existing models. By varying values of p and r as shown in Figure 6, the good separation to various types of meaningful cartoon and texture patterns under different parameter values indicates the diverse usage of the proposed method. For instance, with smaller values of p and r , the method will prefer to enable applications of enhancement, image abstract/rendering, and vectorization. On the other hand, it may be better to handle images containing complex patterns and for the applications of detail enhancement, HDR compression, and haze removal. We show a few applications in image editing by using the property of effectively eliminating texture, meanwhile preserving/enhancing structure.

5.1. Structure and Texture Enhancement. Edge extraction from images is usually the basic preprocessor to natural image editing (Bae and Durand 2007 [28]) and high-level structure inference. High quality results that are continuous, accurate, and thin are generally very difficult to produce due to high susceptibility of edge detectors to complex structures and inevitable noise. Our method is able to suppress low-amplitude details, which remarkably stabilizes the extraction process. One example is shown in Figure 14; the boundaries



FIGURE 11: Smoothing results on the marble mosaic image. NLGRTV is processed under $\sigma = 2$.



FIGURE 12: Structure extraction result comparison with RTV, GRTV, and NLGRTV.

of the original ramp in Figure 14(a) are not very sharp with overall small-magnitude gradients and it is difficult to distinguish them from low-contrast details around. The results of L0 smoothing and our method are shown in Figures 14(b) and 14(c). Their corresponding gradient magnitude

images are shown in Figures 14(d), 14(e), and 14(f) by being linearly enhanced for visualization. As can be observed, our result is more faithful and the boundary is clearer.

Figure 15 shows an example for detail/texture enhancement. The results of WLS, L0 smoothing, GF, and NLGRTV

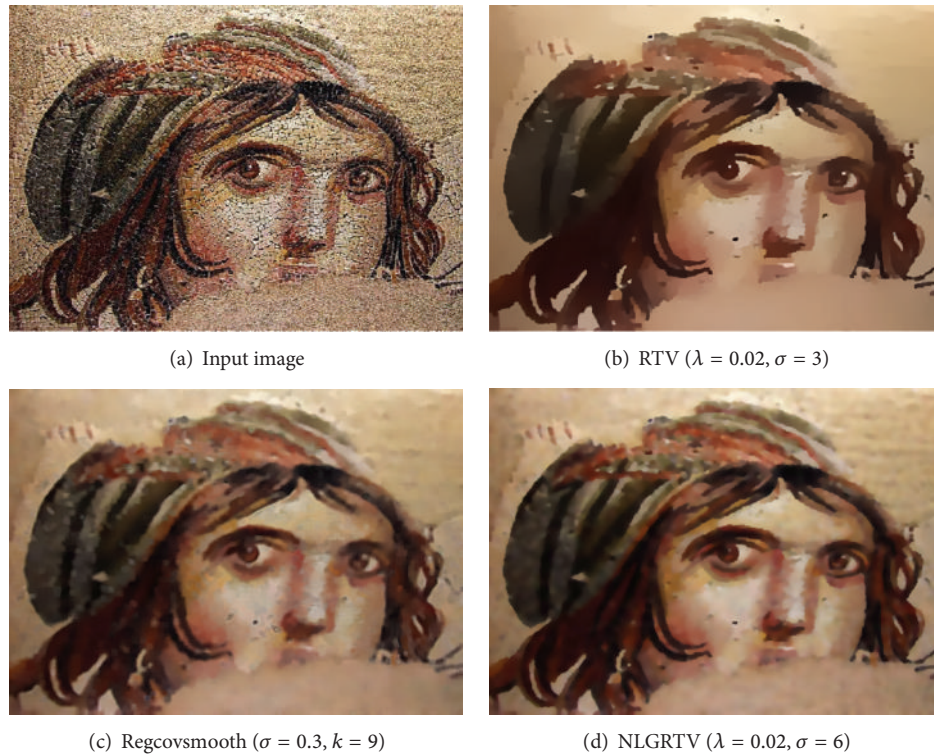


FIGURE 13: Comparison with RTV and Regcovsmooth on the Gypsy girl mosaic image.

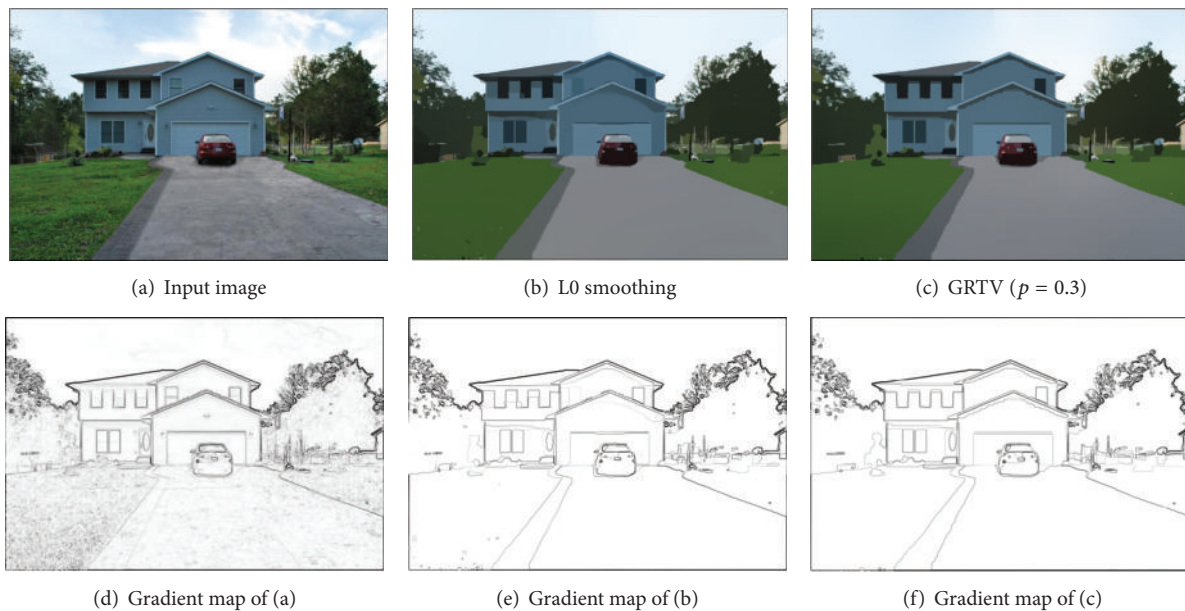


FIGURE 14: Edge enhancement and extraction. Our result is more faithful and the boundary is clearer.

are provided. As can be observed from enhancements of the small petal on the upper region of the image, the edge-aware filters WLS and L0 smoothing lead to gradient reversal artifacts due to their edge sharpening results on their corresponding base layers. GF alleviates this drawback because of its local linear model nature. Since NLGRTV

employs a nonlocal scheme, it also attains visual pleasure estimate. Figure 16 shows another example. As illustrated in [10] and in Figures 16(b) and 16(c), both RTV and GRTV can nicely separate the cloth patterns from the image. Then the magnification effect can be created followed by enhancing the texture contrast and adding this layer back to the base layer.

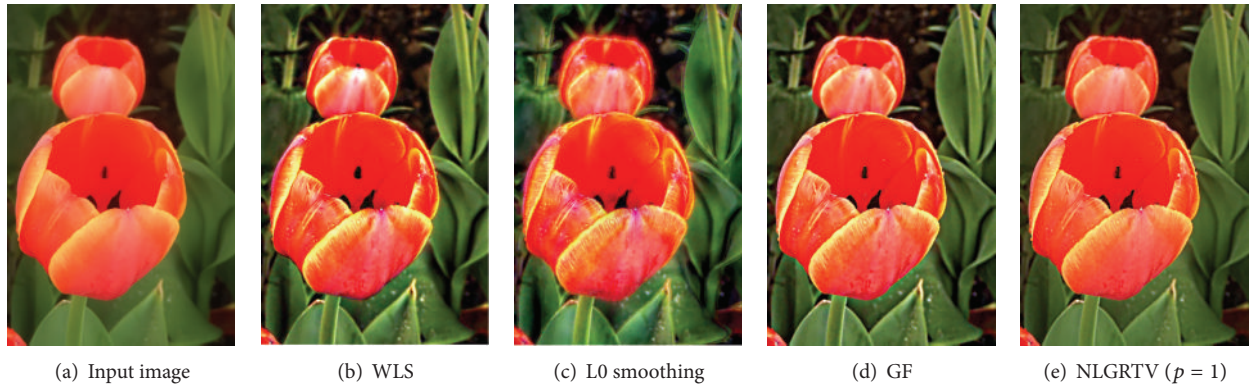


FIGURE 15: Detail enhancement. The detail layer is boosted by 5.

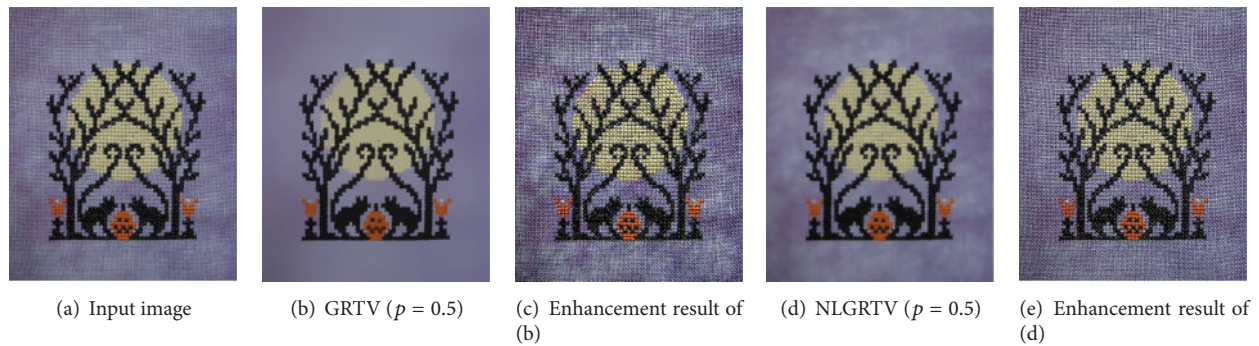


FIGURE 16: Texture enhancement. The detail layer is boosted by 2.

One main drawback is that the white smog is also magnified. NLGRTV alleviates the shortcoming as it can preserve the main slow-varying shape of white smog in the structure part and thus avoid being magnified.

5.2. Image Composition. Mosaics images, paintings, and drawings sometimes cannot be directly used in image composition because the source and target textures are incompatible. Our proposed GRTV and NLGRTV methods can largely attenuate this shortcoming by first smoothing the original input with our structure-preserving filter and then performing composition between the resulting base layers with another image. As shown in Figure 17, the well-smoothed image by GRTV enables the composed image in Figure 17(e) to look more visually pleasing than Figure 17(d). Besides, by tuning parameters, the nonlocal method NLGRTV even obtains one composed image containing the shade information inherent in the original image.

Another composition example is displayed in Figure 18, in which the input image was partly analyzed in Figure 4. There the local smoothed methods RTV and GRTV cannot well preserve the subtle color information in the boy. Figure 18(b) presents the result of Regcovsmooth, from which we find that the image is blurred with the purpose of diminishing the knitted textiles. This makes the composition image in Figure 18(e) look a little indistinct. The results in Figure 18(c) and 18(f) indicate that our method can achieve

the goal of joint edge-preserving and texture-diminishing elegantly and hence yield the composed image to be more impressive.

5.3. Image Abstraction. Nonphotorealistic abstraction with simultaneous detail simplifying and edge emphasizing usually gives a nonrealistic look of images. The conventional methods mainly consist of the edge-preserving smooth step and line extraction step, which is often accomplished by BF (Winnemöller et al. 2006 [13]) and the difference of Gaussian (DoG) filtering, respectively. Finally, the enhanced extracted lines are composed back to augment the visual distinctiveness of different regions. As discussed in Section 4.2 and shown in Figure 19, both L0 smoothing and our method can simultaneously achieve the two goals. By setting larger regularization parameter λ to smooth the image, the resulting structure component can be directly applied on edge detection.

5.4. Artifact Removal. When cartoon-like images such as the clip-art were heavily compressed, this usually causes severe compression artifacts. In [9], Xu et al. found that the general denoising approaches even with the state-of-the-art BM3D [29] cannot achieve satisfying output and their developed method L0 smoothing gives appear performance. In fact, as can be expected, this artifact that strongly correlated with edges can be viewed as a special nonuniform texture and hence can be properly coped by our texture-smoothing and

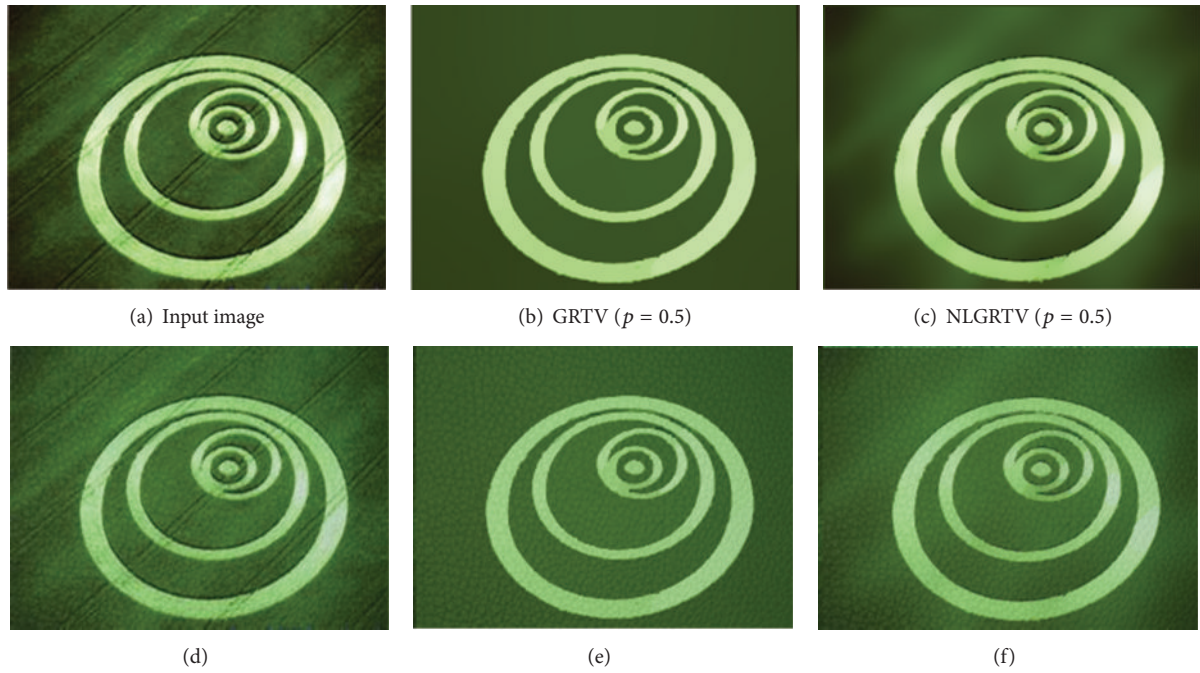


FIGURE 17: Image composition example. (a) Input; (b), (c) structure image by GRTV and NLGRTV; (d), (e), and (f) composed images obtained with the original and the structure images (a), (b), and (c) in top row.

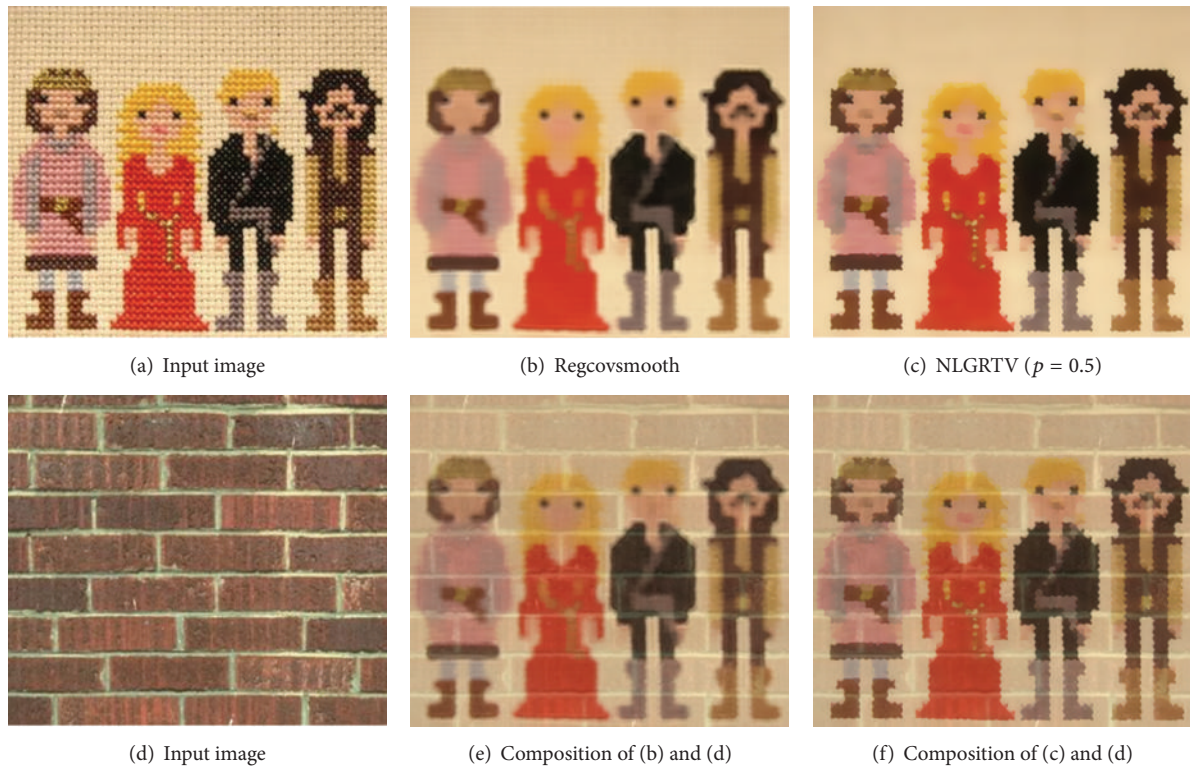


FIGURE 18: Another composition example. (a) Input; (b), (c) structure image by Regcovsmooth and NLGRTV; (e), (f) composed images obtained by (b) and (d) and (c) and (d).

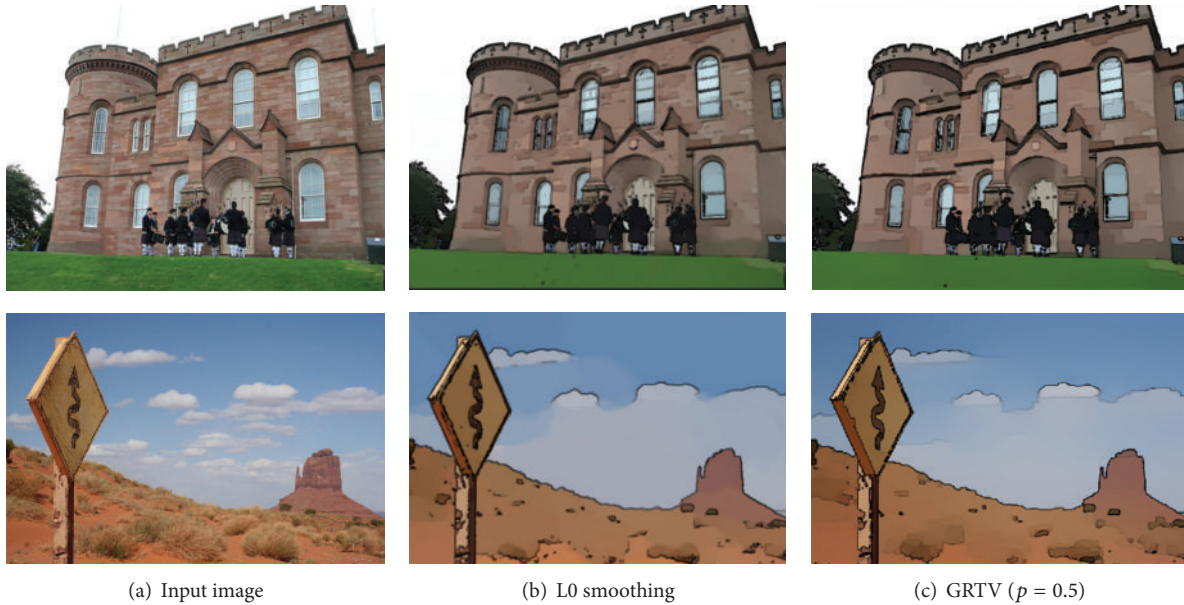


FIGURE 19: Image abstraction results. Our model effectively removes texture from structure and the extracted structure can be used to obtain a nonphotorealistic rendering of the scene.

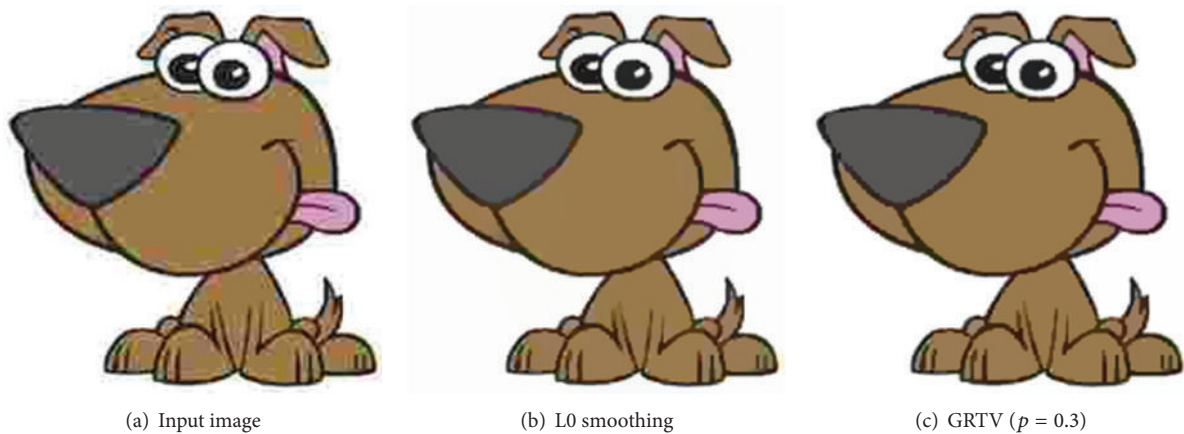


FIGURE 20: Clip-art compression artifact removal example.

edge-preserving filters. Two examples are shown in Figures 20 and 21, where one input image contains acute edges and the other one in contrast has smooth edges. In Figure 20 GRTV outperforms RTV mainly that it better lowers the amplitudes of noise-like structures on the animal's claws. In Figure 21, the nonlocal scheme NLGRTV corrects the locally shifted color nearby the edges.

6. Conclusions and Discussions

In this paper, models can effectively eliminate texture; meanwhile preserving/enhancing structure was proposed. The image decomposition results were improved by assigning with different norms for the spatially varying smooth-penalty weights and regularization terms. The proposed model can be applied to images with high noise levels and complex content and thus possesses advantages such as contrast-preserving,

edge-preserving/sharpening, and data-driven scale selection. Additionally, the proposed model allows a high level of randomness due to the nonlinear formulation. With the aid of iteratively reweighted least squares (IRLS) technique, the original nonlinear problem was transformed to the alternative iterations of weights updating and the least square solver that are much easier to solve quickly.

Experimental results demonstrate the effectiveness and robustness of the proposed method. With smaller p -norm and neighbor size, the method will prefer the piece-wise image. On the other hand, nonlocal sampling strategy enables it better to handle images containing complex patterns. Particularly, our method GRTV can enhance the edge boundary by increasing steepness of transition as displayed in Figures 7, 8, and 9. This property may be in favor of the image segmentation such as the variational active contour/snake models [30, 31]; therefore our developed weight can be incorporated to their weighted TV-norm for better segmentation.

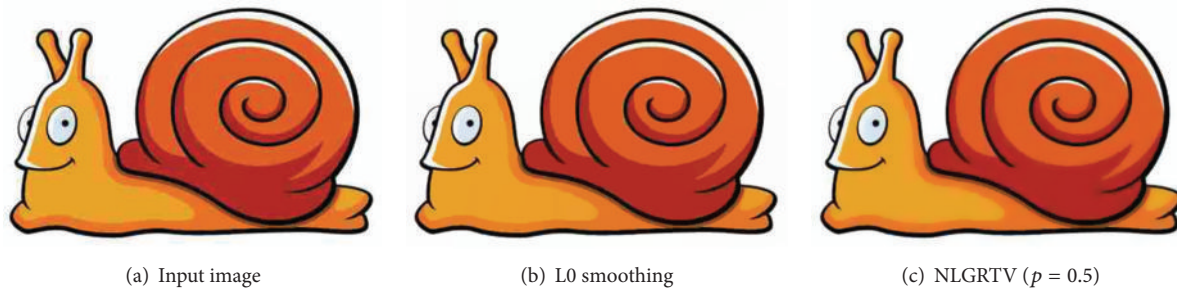


FIGURE 21: Another Clip-art compression artifact removal example.

An important issue that needs further investigation in NLGRTV is replacing the current WIV map by other texture measurements as adaptive varying smooth-penalty weight. As stated in methods RTV and Regcovsmooth, there is no texture/feature measurement which is perfect and it may misinterpret part of structures as texture when it is statically similar in shape and scale. State-of-the-art texture measurements discussed in [3, 32] can be considered to be integrated into our framework. This is an interesting direction for future work that may lead to substantial improvement.

Conflict of Interests

The authors of the paper do not have a direct financial relation that might lead to a conflict of interests for any of the authors.

Acknowledgments

This work was partly supported by the National Natural Science Foundation of China under Grant nos. 61362001, 62162084, 61340025, 61261010, 81120108012, and 51165033, the Science and Technology Department of Jiangxi Province of China under Grant nos. 20133BDH80026, 20121BBE50023, and 20132BAB211030, Postdoctoral Research Funds (2014M551867), Jiangxi Advanced Projects for Postdoctoral Research Funds (2014KY02), and the Basic Research Program of Shenzhen Jc201104220219A.

References

- [1] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.
- [2] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proceedings of the 6th IEEE International Conference on Computer Vision*, pp. 839–846, Bombay, India, January 1998.
- [3] A. Buades, T. M. Le, J.-M. Morel, and L. A. Vese, "Fast cartoon + texture image filters," *IEEE Transactions on Image Processing*, vol. 19, no. 8, pp. 1978–1986, 2010.
- [4] L. Karacan, E. Erdem, and A. Erdem, "Structure-preserving image smoothing via region covariances," *ACM Transactions on Graphics*, vol. 32, no. 6, 2013.
- [5] F. Rousselle, C. Knaus, and M. Zwicker, "Adaptive rendering with non-local means filtering," *ACM Transactions on Graphics*, vol. 31, no. 6, article 195, 2012.
- [6] C. Ye, D. Tao, M. Song, D. W. Jacobs, and M. Wu, "Sparse norm filtering," <http://arxiv.org/abs/1305.3971>.
- [7] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1–4, pp. 259–268, 1992.
- [8] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," *ACM Transactions on Graphics*, vol. 27, no. 3, 2008.
- [9] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via L_0 gradient minimization," *ACM Transactions on Graphics*, vol. 30, no. 6, 2011.
- [10] L. Xu, Q. Yan, Y. Xia, and J. Jia, "Structure extraction from texture via relative total variation," *ACM Transactions on Graphics*, vol. 31, no. 6, article 139, 2012.
- [11] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," *ACM Transactions on Graphics*, vol. 21, no. 3, pp. 257–266, 2002.
- [12] R. Fattal, M. Agrawala, and S. Rusinkiewicz, "Multiscale shape and detail enhancement from multi-light image collections," *ACM Transactions on Graphics*, vol. 26, no. 3, 2007.
- [13] H. Winnemöller, S. C. Olsen, and B. Gooch, "Real-time video abstraction," *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 1221–1226, 2006.
- [14] Z. Su, X. Luo, Z. Deng, Y. Liang, and Z. Ji, "Edge-preserving texture suppression filter based on joint filtering schemes," *IEEE Transactions on Multimedia*, vol. 15, no. 3, pp. 535–548, 2013.
- [15] K. He, J. Sun, and X. Tang, "Guided image filtering," in *Proceedings of the 11th European Conference on Computer Vision [ECCV '10]*, pp. 1–14, Crete, Greece, September 2010.
- [16] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 2, pp. 60–65, June 2005.
- [17] C.-T. Shen, F.-J. Chang, Y.-P. Hung, and S.-C. Pei, "Edge-preserving image decomposition using L_1 fidelity with L_0 gradient," in *Proceedings of the SA '12 SIGGRAPH Asia*, 2012.
- [18] X. Cheng, M. Zeng, and X. Liu, "Feature-preserving filtering with L_0 gradient minimization," *Computers and Graphics*, vol. 38, no. 1, pp. 150–157, 2014.
- [19] P. Rodríguez and B. Wohlberg, "Efficient minimization method for a generalized total variation functional," *IEEE Transactions on Image Processing*, vol. 18, no. 2, pp. 322–332, 2009.
- [20] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 600–616, 1997.

- [21] B. D. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Transactions on Signal Processing*, vol. 47, no. 1, pp. 187–200, 1999.
- [22] G. Gilboa and S. Osher, "Nonlocal operators with applications to image processing," *Multiscale Modeling & Simulation*, vol. 7, no. 3, pp. 1005–1028, 2008.
- [23] K. N. Chaudhury and A. Singer, "Non-local euclidean medians," *IEEE Signal Processing Letters*, vol. 19, no. 11, pp. 745–748, 2012.
- [24] K. N. Chaudhury and A. Singer, "Non-local patch regression: robust image denoising in patch space," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '13)*, pp. 1345–1349, Vancouver, Canada, May 2013.
- [25] Q. Yang, K.-H. Tan, and N. Ahuja, "Real-time O(1) bilateral filtering," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR '09)*, pp. 557–564, Miami, Fla, USA, June 2009.
- [26] E. S. L. Gastal and M. M. Oliveiray, "Adaptive manifolds for real-time high-dimensional filtering," *ACM Transactions on Graphics*, vol. 31, article 33, 2012.
- [27] K. Subr, C. Soler, and F. Durand, "Edge-preserving multiscale image decomposition based on local extrema," *ACM Transactions on Graphics*, vol. 28, no. 5, pp. 147–147, 2009.
- [28] S. Bae and F. Durand, "Defocus magnification," *Computer Graphics Forum*, vol. 26, no. 3, pp. 571–579, 2007.
- [29] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.
- [30] X. Bresson, S. Esedoglu, P. Vanderghenst, J.-P. Thiran, and S. Osher, "Fast global minimization of the active contour/snake model," *Journal of Mathematical Imaging and Vision*, vol. 28, no. 2, pp. 151–167, 2007.
- [31] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung, "Saliency filters: contrast based filtering for salient region detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 733–740, June 2012.
- [32] G. Gilboa, N. Sochen, and Y. Y. Zeevi, "Variational denoising of partly textured images by spatially varying constraints," *IEEE Transactions on Image Processing*, vol. 15, no. 8, pp. 2281–2289, 2006.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

