

Image Smoothing via L_0 Gradient Minimization

Li Xu*

Cewu Lu*

Yi Xu

Jiaya Jia

Department of Computer Science and Engineering
The Chinese University of Hong Kong

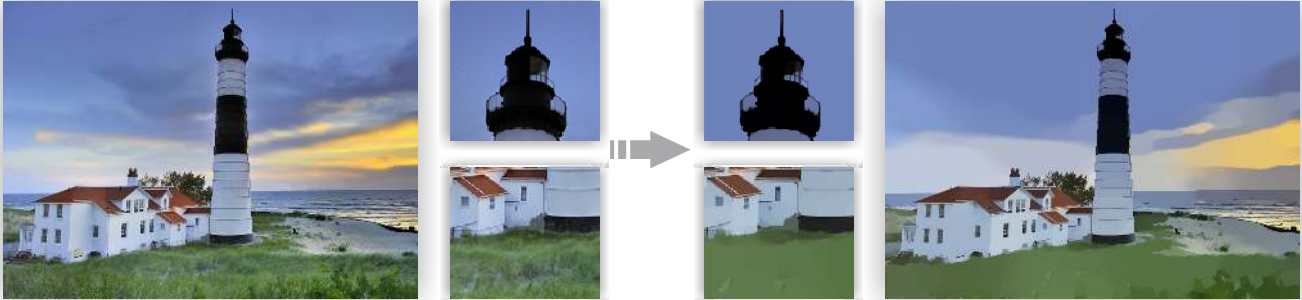


Figure 1: L_0 smoothing accomplished by global small-magnitude gradient removal. Our method suppresses low-amplitude details. Meanwhile it globally retains and sharpens salient edges. Even the high-contrast thin edges on the tower are preserved.

Abstract

We present a new image editing method, particularly effective for sharpening major edges by increasing the steepness of transition while eliminating a manageable degree of low-amplitude structures. The seemingly contradictory effect is achieved in an optimization framework making use of L_0 gradient minimization, which can globally control how many non-zero gradients are resulted in a sparsity-control manner. Unlike other edge-preserving smoothing approaches, our method does not depend on local features, but instead globally locates important edges. It, as a fundamental tool, finds many applications and is particularly beneficial to edge extraction, clip-art JPEG artifact removal, and non-photorealistic effect generation.

Keywords: image smoothing, L_0 sparsity, sharpening, filtering

Links: [DL](#) [PDF](#) [WEB](#) [VIDEO](#) [CODE](#)

1 Introduction

Photos comprise rich and well-structured visual information. In human visual perception, edges are effective and expressive stimulation, vital for neural interpretation to make the best sense of the scene. In manipulating and understanding pictures, high-level inference with regard to salient structures was intensively attended to. Research following this line embodies generality and usefulness in a wide range of applications, including image recognition, segmentation, object classification, and many other photo editing and non-photorealistic rendering tasks.

*Both authors contributed equally to this work.

ACM Reference Format

Xu, L., Lu, C., Xu, Y., Jia, J. 2011. Image Smoothing via L_0 Gradient Minimization. *ACM Trans. Graph.* 30, 6, Article 174 (December 2011), 11 pages. DOI = 10.1145/2024156.2024208
<http://doi.acm.org/10.1145/2024156.2024208>

Copyright Notice

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or direct commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701, fax +1 (212) 869-0481, or permissions@acm.org.
© 2011 ACM 0730-0301/2011/12-ART174 \$10.00 DOI 10.1145/2024156.2024208
<http://doi.acm.org/10.1145/2024156.2024208>

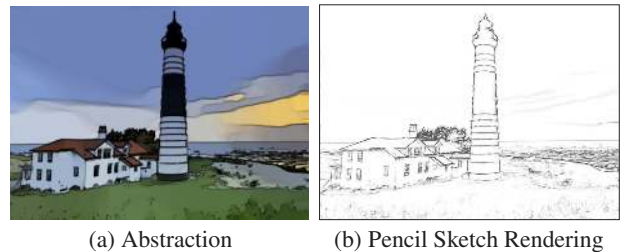


Figure 2: Our L_0 smoothing results avail non-photorealistic effect generation.

We in this paper present a new editing tool, greatly helpful for characterizing and enhancing fundamental image constituents, i.e., salient edges, and in the meantime for diminishing insignificant details. Our method relates in spirit to edge-preserving smoothing [Tomasi and Manduchi 1998; Durand and Dorsey 2002; Paris and Durand 2006; Farbman et al. 2008; Subr et al. 2009; Kass and Solomon 2010] that aims to retain primary color change, and yet differs from them in essence in focus and in mechanism. Our objective is to *globally* maintain and possibly enhance the most prominent set of edges by increasing steepness of transition while not affecting the overall acutance. It enables faithful principal-structure representation.

Algorithmically, we propose a sparse gradient counting scheme in an optimization framework. The main contribution is a new strategy to confine the discrete *number* of intensity changes among neighboring pixels, which links mathematically to the L_0 norm for information sparsity pursuit. This idea also leads to an unconventional global optimization procedure involving a discrete metric, whose solution enables diversified edge manipulation according to saliency. The qualitative effect of our method is to thin salient edges, which makes them easier to be detected and more visually distinct. Different from color quantization and segmentation, our enhanced edges are generally in line with the original ones. Even small-resolution objects and thin edges can be faithfully maintained if they are structurally conspicuous, as shown in Fig. 1.

The framework is general and finds several applications. We apply it to compression-artifact degraded clip-art recovery. High quality results can be obtained in our extensive experiments. Our method

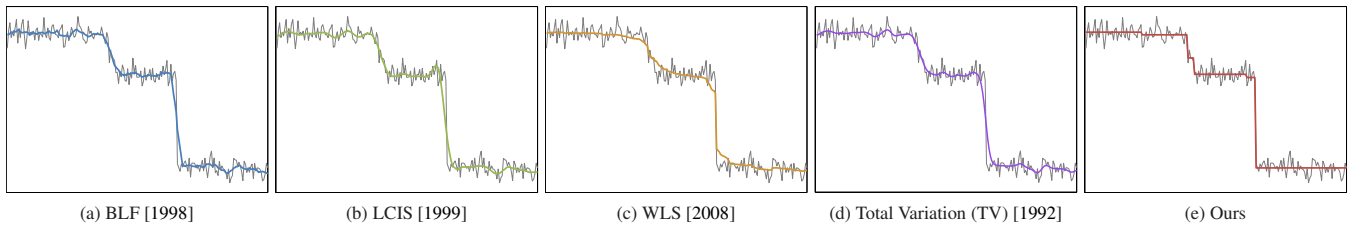


Figure 3: Signal obtained from an image scanline, containing both details and sharp edges. (a) Result of bilateral filtering. (b) Result of anisotropic diffusion used in the LCIS system. (c) Result of WLS optimization. (d) Result of TV smoothing. (e) Our L_0 smoothing result.

can also profit edge extraction, a fundamentally important operator, by effectively removing part of noise, unimportant details, and even of slight blurriness, making the results immediately usable in image abstraction and pencil sketch production, as shown in Fig. 2. In traditional layer decomposition, with an additional step to avoid structure over-enhancement, our method is applicable to detail enhancement based on separating layers, and possibly to HDR tone mapping after parameter tuning. We show several examples along with discussion of limitations that our method might cause over-sharpening for large illumination variation spanning dozens of pixels when strong smoothing is applied.

2 Background and Motivation

Edge-preserving smoothing can be achieved by local filtering, including bilateral filtering [Tomasi and Manduchi 1998], its accelerated versions [Paris and Durand 2006; Weiss 2006; Chen et al. 2007] and relatives [Choudhury and Tumblin 2003; Fattal 2009; Baek and Jacobs 2010; Kass and Solomon 2010]. Robust optimization-based approaches have also been advocated, represented by the weighted least square optimization [Farbman et al. 2008] and envelope extraction [Subr et al. 2009]. We discuss their properties using the 1D signal example (a scanline of a natural image) shown in Fig. 3.

Bilateral filtering is widely used for its simplicity and effectiveness in removing noise-like structures. This method trades off between details flattening and sharp edge preservation, as discussed in [Farbman et al. 2008]. Its result is shown in Fig. 3(a). Anisotropic diffusion [Perona and Malik 1990; Black et al. 1998] is also designed for suppressing noise while preserving important structures, which involves an edge-stopping function to prevent smoothing from crossing strong edges. The change of structures accumulates and the output would converge to a constant-value image unless being stopped halfway. One result is shown in Fig. 3(b).

Farbman et al. [2008] proposed a robust method with the weighted least square (WLS) measure. The optimization framework with edge preserving regularization is more flexible compared with local filtering. Its result is shown in Fig. 3(c). Another type of edge preserving regularization is total variation (TV) [Rudin et al. 1992], which is widely used to remove noise from images. It however also penalizes large gradient magnitudes, possibly influencing contrast during smoothing. One example is shown in Fig. 3(d).

Subr et al. [2009] considered local signal extremes and used edge-aware interpolation [Levin et al. 2004; Lischinski et al. 2006] to compute envelopes. A smoothed mean layer is extracted by averaging the envelopes, originated from a 1D Hilbert-Huang transform (HHT). The method aims to remove small scale oscillations. Contrarily, our method targets globally preserving salient structures, even if they are small in resolution.

Kass and Soloman [2010] used smoothed histogram to accelerate local filtering and proposed the mode-based filters. Most recently,

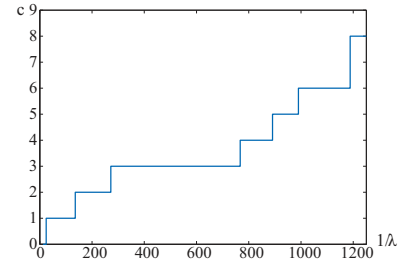


Figure 4: Correspondence between k and $1/\lambda$ in Eqs. (2) and (3). The plot is obtained by trying different λ values in Eq. (3) and by finding the corresponding k in the results after our optimization.

Paris et al. [2011] demonstrated that multi-scale detail manipulation can be achieved using a modified Laplacian pyramid with efficient classification. Our method differs from them on the overall estimation process, and on the edge enhancement behavior as exemplified in Fig. 3(e). We regard our method as complementary to prior smoothing approaches.

Finally, interactive image editing needs to select regions of interest with accurate boundaries. Graph-cut based methods [Rother et al. 2004; Li et al. 2004; Liu et al. 2009] and segmentation [Maji et al. 2011; Arbelaez et al. 2011] were employed. To efficiently propagate scribbles, geodesic distance [Criminisi et al. 2010] and diffusion distance [Farbman et al. 2010] were used, by replacing the traditional color difference, to deal with textured surfaces or those with complicated shapes. Intriguingly, user interaction can be performed more efficiently on our edge-enhanced images after removing low-amplitude structures.

2.1 1D Smoothing

We enhance highest-contrast edges by confining the number of non-zero gradients, while smoothing is achieved in a global manner. To begin with, we denote the input discrete signal by g and its smoothed result by f . Our method counts amplitude changes discretely, written as

$$c(f) = \#\{p \mid |f_p - f_{p+1}| \neq 0\}, \quad (1)$$

where p and $p+1$ index neighboring samples (or pixels). $|f_p - f_{p+1}|$ is a gradient w.r.t. p in the form of forward difference. $\#\{\}$ is the counting operator, outputting the number of p that satisfies $|f_p - f_{p+1}| \neq 0$, that is, the L_0 norm of gradient. $c(f)$ does not count on gradient magnitude, and thus would not be affected if an edge only alters its contrast. This discrete counting function is central to our method.

Note that the measure $c(f)$ alone is not functional. It is combined in our method with a general constraint – that is, the result f should be structurally similar to the input signal g – to fully exhibit the

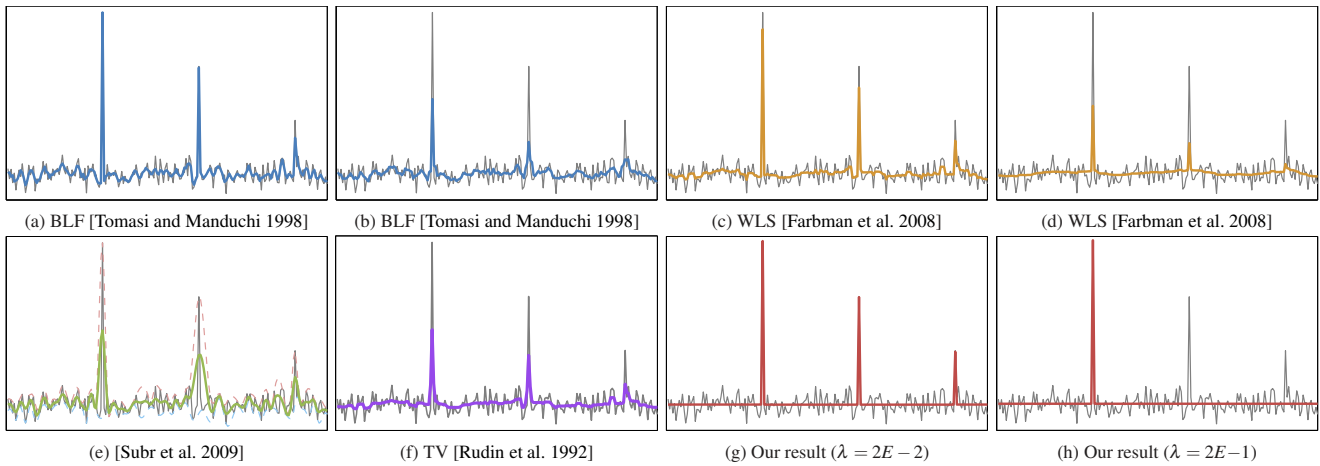


Figure 5: 1D signal with spike-edges in different scales. (a)-(b) Results of Bilateral filtering with small- and large-range filters. (c)-(d) Results of WLS optimization [Farbman et al. 2008] using weak and strong smooth parameters. (e) Result of Subr et al. [2009]. (f) Result of TV smoothing [Rudin et al. 1992]. (g)-(h) Our results. The most significant one or more spikes can be retained with different λ in Eq. (3).

competence. We express the specific objective function as

$$\min_f \sum_p (f_p - g_p)^2 \quad \text{s.t.} \quad c(f) = k. \quad (2)$$

$c(f) = k$ indicates that k non-zero gradients exist in the result. Eq. (2) is very powerful to abstract structural information. Fig. 3(e) shows the result with $k = 6$ by minimizing Eq. (2) through exhaustive search. The resulted signal flattens details and sharpens main edges. The overall shape is also in line with the original one because intensity change must arise along significant edges to reduce as much as possible the total energy. It is observed that putting edges elsewhere only raises the cost. This smoothing effect is obviously dissimilar to those of prior edge-preserving methods. A larger k yields a finer approximation, still characterizing the most prominent contrast.

As the cost in Eq. (2) stems from the quadratic intensity difference term $(f_p - g_p)^2$, it is not allowed that many pixels drastically change their color. Low-amplitude structures thus can be primarily removed in a controllable and statistical manner. Diminishing salient edges is automatically prevented. A noteworthy feature of this framework is that *no matter how k is set, no edge blurriness will be caused* due to the avoidance of local filtering and of averaging operation.

In practice, k in Eq. (2) may range from tens to thousands, especially in 2D images with different resolutions. To control it, we employ a general form to seek a balance between structure flattening and result similarity with the input, and write it as

$$\min_f \sum_p (f_p - g_p)^2 + \lambda \cdot c(f), \quad (3)$$

where λ is a weight directly controlling the significance of $c(f)$, which is in fact a smoothing parameter. A large λ makes the result have very few edges. To relate k and $1/\lambda$ presented respectively in Eqs. (2) and (3), we plot in Fig. 4 their correspondence for the example in Fig. 3. The number of non-zero gradients is monotone with respect to $1/\lambda$. We describe our 2D solver in Section 3.

Fig. 5 shows an example where three needle-like structures are small in resolution but significant in amplitude. Our results are shown in Fig. 5(g)-(h), containing complete one or more spikes by varying λ , faithfully preserving scales. Other methods also produce excellent results. They, however, attenuate the spikes in different degrees, as shown in (a)-(f). We regard our method as parallel

to these approaches. As shown later, our method can be used along with bilateral filtering to produce new smoothing effects thanks to the complementary behaviors.

2.2 2D Formulation

In 2D image representation, we denote by I the input image and by S the computed result. The gradient $\nabla S_p = (\partial_x S_p, \partial_y S_p)^T$ for each pixel p is calculated as color difference between neighboring pixels along the x and y directions. Our gradient measure is expressed as

$$C(S) = \#\{p \mid |\partial_x S_p| + |\partial_y S_p| \neq 0\}. \quad (4)$$

It counts p whose magnitude $|\partial_x S_p| + |\partial_y S_p|$ is not zero. With this definition, S is estimated by solving

$$\min_S \left\{ \sum_p (S_p - I_p)^2 + \lambda \cdot C(S) \right\}. \quad (5)$$

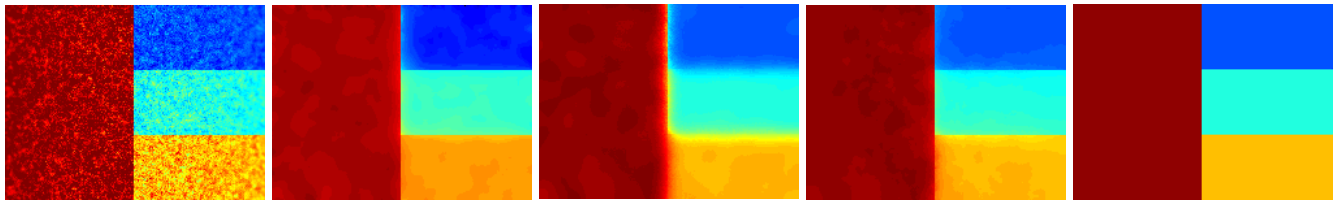
In practice, for color images, the gradient magnitude $|\partial S_p|$ is defined as the sum of gradient magnitudes in rgb . The term $\sum (S - I)^2$ constrains image structure similarity.

Before describing our solver, we use a 2D example, created by Farbman et al. [2008], to evaluate and compare smoothing performance. The color visualized input in Fig. 6(a) is a piece-wise constant image contaminated with intensive noise. (b)-(d) show results of three representative methods. Our method can globally find dominant high contrast and generate the clean result shown in (e).

3 Solver

Eq. (5) involves a discrete counting metric. It is difficult to solve because the two terms model respectively the pixel-wise difference and global discontinuity statistically. Traditional gradient decent or other discrete optimization methods are not usable.

We adopt a special alternating optimization strategy with half-quadratic splitting, based on the idea of introducing auxiliary variables to expand the original terms and update them iteratively. Wang et al. [2008] used the splitting scheme to solve a different convex problem. Our algorithm, due to the discrete nature, contains new subproblems. Both of them find their closed-form solutions. It is notable that the original L_0 -norm regularized optimization problem is known as computationally intractable. Our solver is thus



(a) Visualized input [2008] (b) [Subr et al. 2009] (c) BLF (d) WLS (e) Our result

Figure 6: Noisy image created by Farbman et al. [2008]. (a) Color visualized noisy input. (b) Result of Subr et al. [2009]. (c) Bi-lateral filtering (BLF) result ($\sigma_s = 12$, $\sigma_r = 0.45$) [Tomasi and Manduchi 1998]. (d) Result of WLS optimization ($\alpha = 1.8$, $\lambda = 0.35$) [Farbman et al. 2008]. (e) Our result.

only an *approximation*, making the problem easier to tackle and upholding the property to maintain and enhance salient structures.

We introduce auxiliary variables h_p and v_p , corresponding to $\partial_x S_p$ and $\partial_y S_p$ respectively, and rewrite the objective function as

$$\min_{S, h, v} \left\{ \sum_p (S_p - I_p)^2 + \lambda C(h, v) + \beta ((\partial_x S_p - h_p)^2 + (\partial_y S_p - v_p)^2) \right\}, \quad (6)$$

where $C(h, v) = \# \{p \mid |h_p| + |v_p| \neq 0\}$ and β is an automatically adapting parameter to control the similarity between variables (h, v) and their corresponding gradients. Eq. (6) approaches (5) when β is large enough. Eq. (6) is solved through alternatively minimizing (h, v) and S . In each pass, one set of the variables are fixed with values obtained from the previous iteration.

Subproblem 1: computing S The S estimation subproblem corresponds to minimizing

$$\left\{ \sum_p (S_p - I_p)^2 + \beta ((\partial_x S_p - h_p)^2 + (\partial_y S_p - v_p)^2) \right\} \quad (7)$$

by omitting the terms not involving S in Eq. (6). The function is quadratic and thus has a global minimum even by gradient decent. Alternatively, we diagonalize derivative operators after Fast Fourier Transform (FFT) for speedup. This yields solution

$$S = \mathcal{F}^{-1} \left(\frac{\mathcal{F}(I) + \beta (\mathcal{F}(\partial_x)^* \mathcal{F}(h) + \mathcal{F}(\partial_y)^* \mathcal{F}(v))}{\mathcal{F}(1) + \beta (\mathcal{F}(\partial_x)^* \mathcal{F}(\partial_x) + \mathcal{F}(\partial_y)^* \mathcal{F}(\partial_y))} \right), \quad (8)$$

where \mathcal{F} is the FFT operator and $\mathcal{F}(\cdot)^*$ denotes the complex conjugate. $\mathcal{F}(1)$ is the Fourier Transform of the delta function. The plus, multiplication, and division are all component-wise operators. Compared to minimizing Eq. (7) directly in the image space, which involves very-large-matrix inversion, computation in the Fourier domain is much faster due to the simple component-wise division.

Subproblem 2: computing (h, v) The objective function for (h, v) is

$$\min_{h, v} \left\{ \sum_p (\partial_x S_p - h_p)^2 + (\partial_y S_p - v_p)^2 + \frac{\lambda}{\beta} C(h, v) \right\}, \quad (9)$$

where $C(h, v)$ returns the number of non-zero elements in $|h| + |v|$. This apparently sophisticated subproblem can actually be solved quickly because the energy (9) can be spatially decomposed where each element h_p and v_p is estimated *individually*. This is the main benefit of our splitting scheme, which makes the altered problem empirically solvable. Eq. (9) is accordingly decomposed to

$$\sum_p \min_{h_p, v_p} \left\{ (h_p - \partial_x S_p)^2 + (v_p - \partial_y S_p)^2 + \frac{\lambda}{\beta} H(|h_p| + |v_p|) \right\}, \quad (10)$$

Algorithm 1 L_0 Gradient Minimization

Input: image I , smoothing weight λ , parameters β_0 , β_{\max} , and rate κ

Initialization: $S \leftarrow I$, $\beta \leftarrow \beta_0$, $i \leftarrow 0$

repeat

With $S^{(i)}$, solve for $h_p^{(i)}$ and $v_p^{(i)}$ in Eq. (12).

With $h^{(i)}$ and $v^{(i)}$, solver for $S^{(i+1)}$ with Eq. (8).

$\beta \leftarrow \kappa \beta$, $i \leftarrow i + 1$.

until $\beta \geq \beta_{\max}$

Output: result image S

where $H(|h_p| + |v_p|)$ is a binary function returning 1 if $|h_p| + |v_p| \neq 0$ and 0 otherwise. Each single term w.r.t. pixel p in Eq. (10) is

$$E_p = \left\{ (h_p - \partial_x S_p)^2 + (v_p - \partial_y S_p)^2 + \frac{\lambda}{\beta} H(|h_p| + |v_p|) \right\}, \quad (11)$$

which reaches its minimum E_p^* under the condition

$$(h_p, v_p) = \begin{cases} (0, 0) & (\partial_x S_p)^2 + (\partial_y S_p)^2 \leq \lambda/\beta \\ (\partial_x S_p, \partial_y S_p) & \text{otherwise} \end{cases} \quad (12)$$

Proof.

1) When $\lambda/\beta \geq (\partial_x S_p)^2 + (\partial_y S_p)^2$, non-zero (h_p, v_p) yields

$$\begin{aligned} E_p((h_p, v_p) \neq (0, 0)) &= (h_p - \partial_x S_p)^2 + (v_p - \partial_y S_p)^2 + \lambda/\beta, \\ &\geq \lambda/\beta, \\ &\geq (\partial_x S_p)^2 + (\partial_y S_p)^2. \end{aligned} \quad (13)$$

Note that $(h_p, v_p) = (0, 0)$ leads to

$$E_p((h_p, v_p) = (0, 0)) = (\partial_x S_p)^2 + (\partial_y S_p)^2. \quad (14)$$

Comparing Eqs. (13) and (14), the minimum energy $E_p^* = (\partial_x S_p)^2 + (\partial_y S_p)^2$ is produced when $(h_p, v_p) = (0, 0)$.

2) When $(\partial_x S_p)^2 + (\partial_y S_p)^2 > \lambda/\beta$ and $(h_p, v_p) = (0, 0)$, Eq. (14) still holds. But $E_p((h_p, v_p) \neq (0, 0))$ has its minimum value λ/β when $(h_p, v_p) = (\partial_x S_p, \partial_y S_p)$. Comparing these two values, the minimum energy $E_p^* = \lambda/\beta$ is produced when $(h_p, v_p) = (\partial_x S_p, \partial_y S_p)$. ■

With the above derivation, in this step, we compute for each pixel p the minimum energy E_p^* . Summing all of them, i.e., calculating $\sum_p E_p^*$, yields the global optimum for Eq. (10).

Our alternating minimization algorithm is sketched in Alg. 1. Parameter β is automatically adapted in iterations starting from a small value β_0 , it is multiplied by κ each time. This scheme is effective to speed up convergence [Wang et al. 2008]. In our method,



Figure 7: Smoothing results and comparison (best viewed in their original resolution).

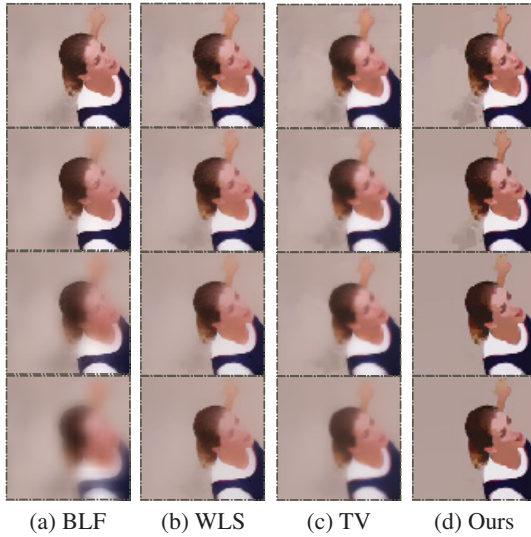


Figure 8: Close-ups of the results with smoothing strength gradually increasing from top to bottom. (a) Bilateral filtering (BLF) results. (b) Results of the WLS method. (c) Results of total variation smoothing. (d) Our results.

β_0 and β_{\max} have fixed values 2λ and $1E5$ respectively. κ that is set to 2 is a good balance between efficiency and performance. We use this value to generate most of our results. Only in Figs. 6 and 11, we set it to 1.05 to allow more iterations in optimization, leading to higher-quality results. The critical parameter λ is allowed to be adjusted to effectively control the level of structure coarseness.

20-30 iterations are generally performed in our algorithm. Most computation is spent on FFT in Eq. (8) and on pixel-wise algebraic operations in Eq. (10). Overall, it takes 3 seconds to process a single-channel 600×400 image on an Intel Core2 Duo CPU@2.13G with our Matlab implementation. The code is publicly available in the project website.

3.1 More Analysis

Relationship to compressed sensing In our results, only salient edges are preserved. The discrete counting idea conceptually relates to sparsity measure in compressed sensing [Donoho 2006]. To extract sparse descriptors from natural images, L_0 norm was also studied in sparse coding [Mairal et al. 2009]. In our work, the way to exercise the idea is new with the counting gradient mechanism. Note that trivially introducing the L_0 norm metric cannot produce reasonable results. Our approximation and development of the ef-

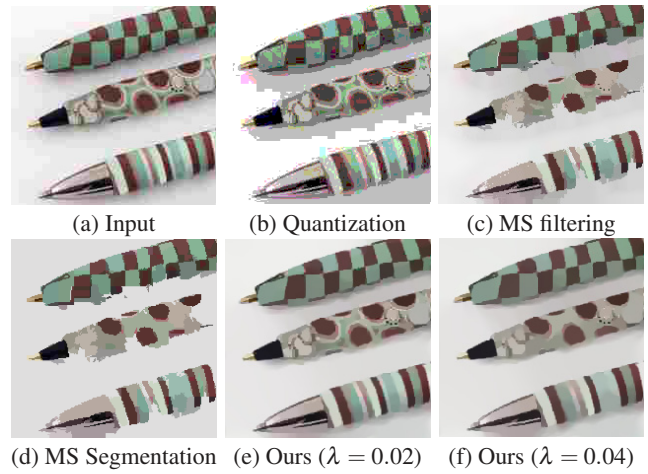


Figure 9: Result comparison with quantization and segmentation. Edges in (b)-(d) are not well maintained owing to the lack of robust boundary preservation.

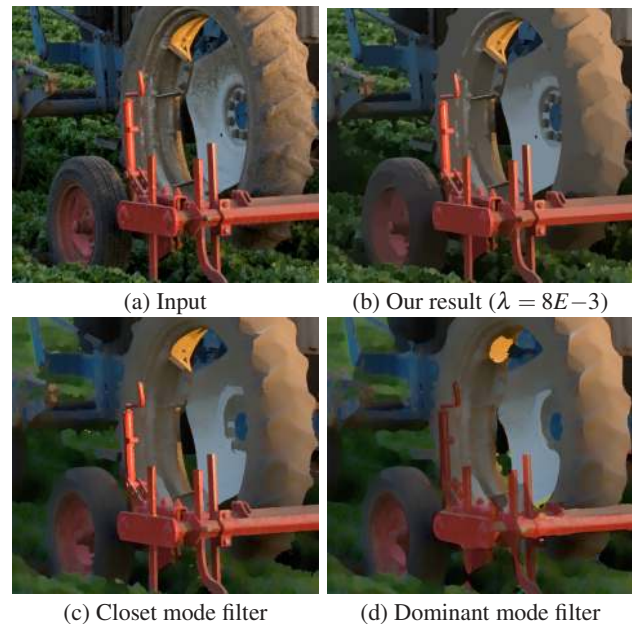


Figure 10: Comparison with mode filters [Kass and Solomon 2010].

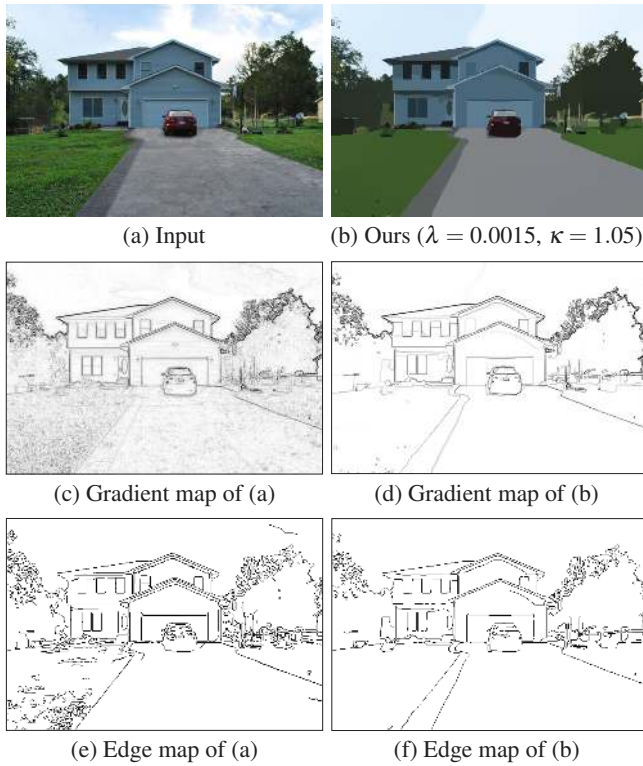


Figure 11: Edge enhancement and extraction. Our method suppresses low-amplitude details and enhances high contrast edges. The combined effect is to remove textures and sharpen main edges even if their gradient magnitudes are not significant locally.

fective algorithm to achieve sub-problem global optimization are central to the high practicality.

Difference with total variation and other regularizers Continuous L_p norm with $p = 1$ was enforced in total variation (TV) smoothing to suppress noise. In this framework, strong smoothing inevitably curtails originally salient edges to penalize their magnitudes. In our method, large gradient magnitudes are allowed by nature with our discrete counting measure.

L_p norm regularization with $0.5 \leq p \leq 1$ was also employed in [Levin et al. 2007] to model the sparsity of natural image gradients. The success of the WLS optimization attributes in part to the L_p norm in the Iterative Reweighted Least Square (IRLS) framework. Mathematically, L_p norm satisfies positive scalability constraint $\|ax\|_p^p = |a|^p \cdot \|x\|_p^p$, where a is a scalar. It yields $\|ax\|_p^p > \|x\|_p^p$ if $|a| > 1$, which implies that these norms still impose large penalties on salient gradients. On the contrary, the L_0 norm in Eq. (1) satisfies $\#\{|x| > 0\} = \#\{|ax| > 0\}$ for any non-zero a , and thus does not comply with the positive scalability constraint. This major difference leads to new smoothing behavior.

Selectively penalizing image gradients is also related to the Weak Membrane model of Blake and Zisserman [1987], which explicitly represents discontinuity and adjusts gradients only in continuous regions. Our method is dissimilar in formulation and in solver.

A natural image example is shown in Fig. 7 with comparison with other state-of-the-art approaches. More are put in our project website, produced with different parameters. Close-ups in Fig. 8 are obtained by varying smoothing strength. Our results contain globally the most salient structures in different degrees.

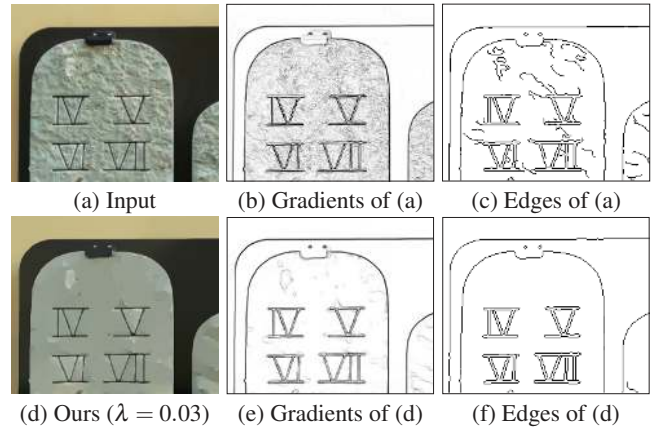


Figure 12: Smoothing for edge detection. The input image (a) contains complex structures, making edge extraction error-prone. On our smoothed image (d), primary edges can be faithfully detected using the same edge detector.

Comparison to quantization and segmentation To clarify the difference, we use the example shown in Fig. 9, where the natural image contains a very small amount of noise, common in photos. Color quantization can neither suppress noise nor accurately remove details, yielding incorrect boundaries, as shown in Fig. 9(b). Image segmentation seeks proper spatial partitioning. This set of methods are widely known as difficult to maintain fine edges, in particular for images with textures or details. Fig. 9(c) and (d) show the results of mean-shift filtering [Comaniciu and Meer 2002] and of its segmentation – formed by fusing – after trying a variety of parameters. The boundaries are not aligned with the latent edges. Our results produced with two λ values are shown in Fig. 9(e) and (f). Edges are better preserved.

Comparison to local histogram-mode filtering The method of Kass and Solomon [2010] is not based on smoothing neighboring pixels, and thus can sharpen edges while reducing details. We show in Fig. 10 a close-up comparison. The edges in our result (b) are in line with the originally salient ones due to the global optimization.

4 Applications

Our method avails several applications due to its fundamentality and the special properties in processing natural images. We apply it to edge enhancement and extraction, non-photorealistic rendering, clip-art restoration, and layer decomposition based manipulation.

4.1 Edge Enhancement and Extraction

Edge extraction from natural images is a basic manipulation tool. Structured edges can be used for natural image editing [Bae and Durand 2007] 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.

In the example shown in Fig. 11(a), the original ramp is visually distinct due to its high contrast. But the boundaries are not very sharp with overall small-magnitude gradients, making them indistinguishable from low-contrast details around. The gradient magnitude image is shown in (c), linearly enhanced for visualization. Applying the Canny edge detector to the original image produces a problematic result (e). Our method can remove statisti-

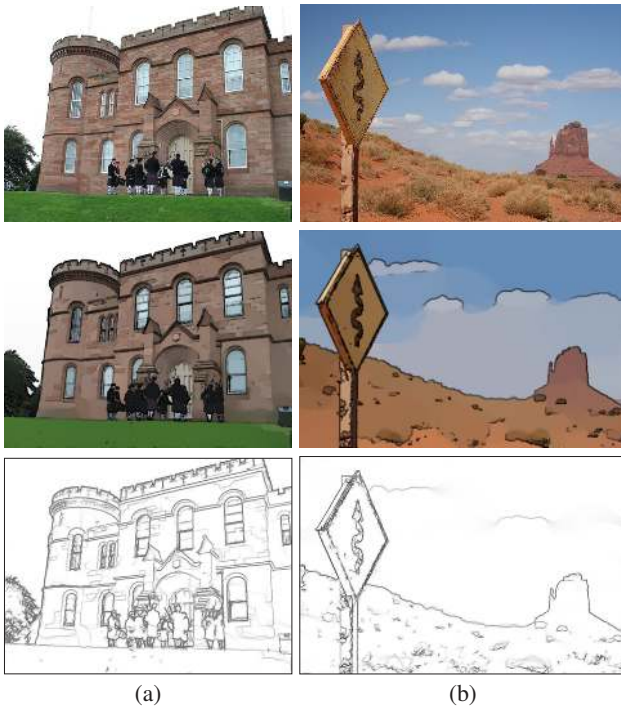


Figure 13: Image abstraction and pencil sketching results. Our method removes the least important structures.

cally insignificant details by global optimization. The remaining main structures in (d) are in the meantime slightly sharpened with notably amplified gradients. With these two main advantages, the detected edges in our result using the same operator can be much more reliable, as shown in (f).

Fig. 12(a) shows another picture containing characters together with background textures. Directly computing gradients on the input image acquires many unwanted small-amplitude structures, as shown in (b). They greatly affect binary Canny edge detector, as shown in (c). Character boundaries are broken; some are even incorrectly connected. Our smoothing result (d) has a cleaner gradient map (e). The edges in (f) computed by the same detector are much better, in accordance with human perception.

4.2 Image Abstraction and Pencil Sketching

Our smoothed image representation fits non-photorealistic abstraction with simultaneous detail flattening and edge emphasizing. Two main steps are involved in traditional methods – that is, image smoothing by mean-shift filtering [DeCarlo and Santella 2002] or bilateral filtering [Winnemöller et al. 2006], and line extraction by difference-of-Gaussian (DoG) filtering. The extracted lines are enhanced and are composed back to augment the visual distinctiveness of different regions. Our method can simultaneously achieve the two goals. After smoothing images with a large λ weight, edge detection can be directly applied.

In [2002], DeCarlo and Santella found that segmentation cannot faithfully preserve edges, and therefore employed a means to smooth them. This is unnecessary for our approach since it globally maintains edge position and magnitudes. Fig. 13 shows two abstraction results.

We create pencil sketching also based on the extracted edge maps. Two results are shown in Fig. 13. They are produced by randomly adding small sketchy lines to the extracted edges, along the tan-

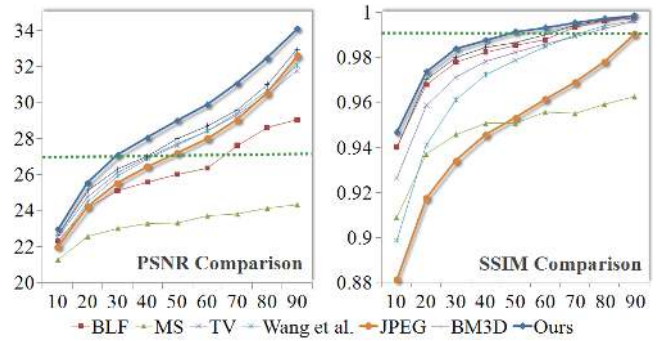


Figure 14: Quantitative evaluation of clip-art JPEG artifact removal. X-axis: JPEG quality. Left Y-axis: PSNR. Right Y-axis: SSIM values.

gent direction. The sketchy lines are with constant length and with gray levels proportional to edge magnitudes. This simple approach enhances significant edges, making structures visually pleasing.

4.3 Clip-Art Compression Artifact Removal

Our smoothing method is also advantageous for cartoon/clip-art compression artifact removal thanks to its special ability to enhance edges. We have experimented with many cartoon/clip-art images with severe compression artifacts. To cope with them, prior knowledge and a training process are required in [Wang et al. 2006]. Our smoothing method in contrast can reliably restore these degraded clip-arts without any learning procedure. The regularization weight λ in this case is set within $[0.02, 0.1]$. Note that general denoising approaches do not suit this application as the compression artifacts are strongly correlated with edges.

We evaluate a set of methods, including edge-preserving smoothing, denoising [Rudin et al. 1992; Dabov et al. 2007], segmentation, and image analogue-based method [Wang et al. 2006] for removing block-based discrete cosine transform (BDCT) artifacts on clip-art images, and show one comparison in Fig. 16.

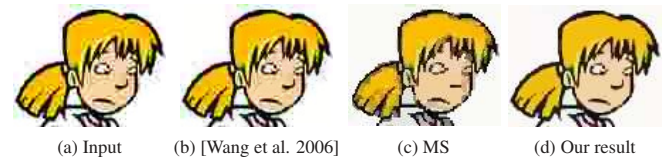


Figure 15: Cartoon restoration results of Wang et al. [2006], mean-shift (MS) segmentation, and of our method.

For quantitative comparison, we compressed 100 clip-art images by standard JPEG with quality values ranging from 10 to 90, and calculated the peak signal-noise ratio (PSNR) and structural similarity (SSIM) [Wang et al. 2004] values after applying different methods to restoration. The statistics are plotted in Fig. 14. Both the PSNR and SSIM scores indicate that our method performs well in removing the JPEG artifacts. Note that the compression artifacts locally shift color, which inherently reduce PSNR. The SSIM plots show that some clip-art images compressed with quality 40, after our restoration, can be structurally comparable to those with compression quality 90+. Fig. 15 shows another comparison.

4.4 Layer-Based Contrast Manipulation

Contrast expansion or compression enables detail magnification [Bae et al. 2006; Fattal et al. 2007] and HDR compression

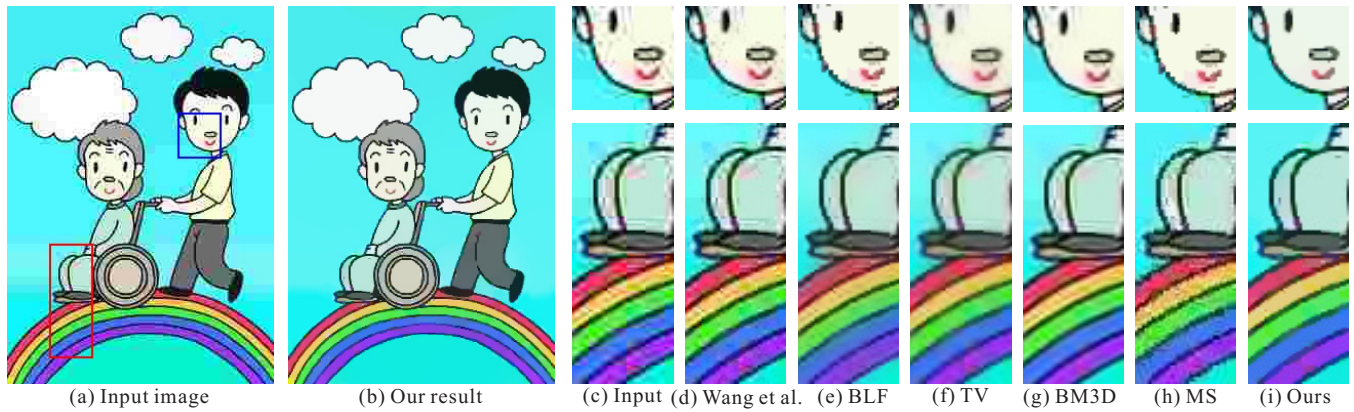


Figure 16: Cartoon JPEG artifact removal. (a) A JPEG compressed image with low quality 10. (b) Our restoration result. (c)-(g) Close-ups of the input and of results of Wang et al. [2006], BLF, TV, BM3D denoising, mean-shift segmentation (MS) and of our method.

[Durand and Dorsey 2002]. These applications can be based on layer decomposition, which separates main structures from details, so that contrast editing in one layer would not affect the others.

In this process, blurring or sharpening edges risks producing visual artifacts. If sharp edges in the base layer are blurred, “overshoot” or “undershoot” may be caused after detail magnification or attenuation. It is very difficult to roll blurred edges back to their original shapes afterwards in the base layer; so prior smoothing methods endeavored to avoid seriously blurring salient edges.

Sharpening edges, on the contrary, may cause gradient reversal. It is a less serious problem, because either gradient-domain operation followed by image reconstruction [Bae et al. 2006] or by directly modifying color [Durand and Dorsey 2002; Kass and Solomon 2010] can be employed. Since our result may contain enhanced edges, we filter images with different Gaussian kernels and selectively fuse them. We propose an optimization framework to automatically determine proper sizes of the Gaussian kernels.

Edge Adjustment Our color-adjustment objective function is

$$\min_{\sigma} \left\{ \sum_p \left((G(\sigma_p) * S) - I_p \right)^2 + \gamma \left((\partial_x \sigma_p)^2 + (\partial_y \sigma_p)^2 \right) \right\}, \quad (15)$$

where $G(\sigma_p)$ is a zero-mean Gaussian with standard deviation σ_p . It is convolved with S , which is our edge-enhanced image, to re-blur edges. $*$ is the convolution operator. In color images, Eq. (15) measures the sum of all costs in the three channels.

The first term $\left((G(\sigma_p) * S) - I_p \right)^2$ enforces the similarity between the re-blurred and original images. Its benefit is twofold. 1) For flat regions in S with details removed, Gaussian blurring S would not bring details back. 2) Sharpened edges in S will get blurred to approximate the latent ones in I . Our goal is to find a suitable Gaussian scale σ for each pixel. The second term $(\partial_x \sigma_p)^2 + (\partial_y \sigma_p)^2$ constrains scale map σ smoothness, avoiding occasional noise. γ is set to $1E-3$.

To solve (15) practically, we assign discrete values to σ_p , i.e., $\sigma \in \{0, 1/3, 2/3, \dots, 3\}$. Because we set Gaussian filter size to $6\sigma + 1$, these σ values yield a set of Gaussian filters with different integer sizes. Determining σ for each pixel thus becomes a discrete labeling problem, which can be solved using graph cuts [Boykov et al. 2001]. The similarity costs for each pixel are treated as potentials for each node in the graph. The partial derivatives are pairwise costs. More details about energy minimization can be

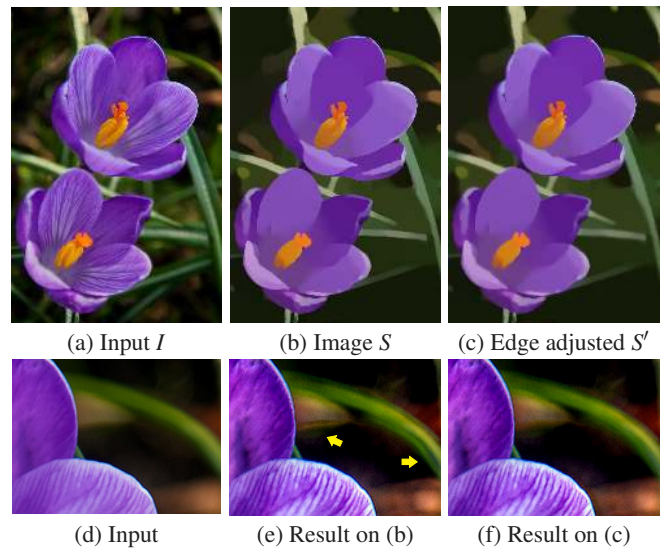


Figure 17: Edge adjustment. Our smoothing result (b) contains edges that are more steeply sloped than those in (a). Our adjustment reduces the sharpness, shown in (c). Detail magnification (2.5 \times) result in (f) with the base layer (c) is visually more compelling than that in (e).

found in [Boykov et al. 2001]. After the σ map is computed, we construct the base layer as $S' = G(\sigma_p) * S$ for each pixel p .

Fig. 17 shows an example. The out-of-focus background are sharpened (Fig. 17(b)). Taking it as the base layer and performing detail magnification produce the result shown in (e). We automatically adjust edges using the above method, and show the result S' in (c). Original edge appearance is well approximated without taking low-amplitude details back, yielding a better detail-enhancement result (f). Our edge adjustment takes less than 1 min to process a 600x800 image. It is less efficient than some other methods, e.g., bilateral filtering, when handling blurred edges.

Detail Magnification Given the input image shown in Fig. 18(a), we compute the detail magnification results by only enhancing gradients in the detail layer. Our method can satisfyingly remove low-amplitude structures from the base layer in controllable degrees, as shown in Fig. 18(b)-(c). Fig. 18(d)-(e) show the magnification results. Fig. 19 gives a comparison with other layer decomposition methods. The petals have rich textures, which are hard to be completely moved to the detail layer by previous approaches. Our

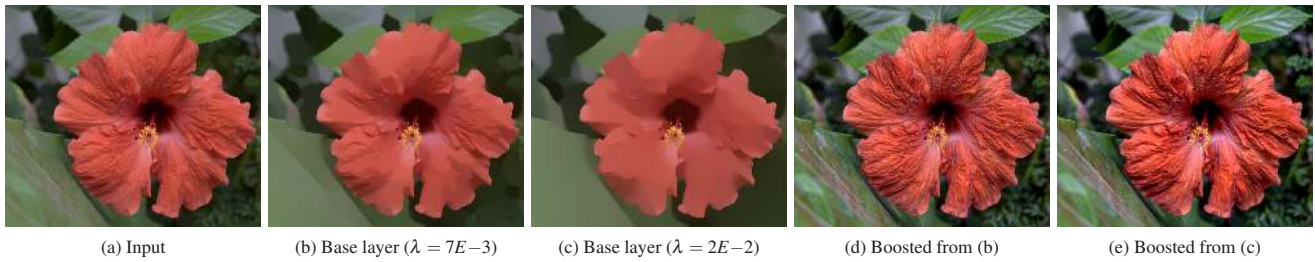


Figure 18: Base-detail separation and manipulation. (a) Input image. (b)-(c) Two base layers generated by our method. (d)-(e) Detail magnification results with (b) and (c) being the base layers respectively.

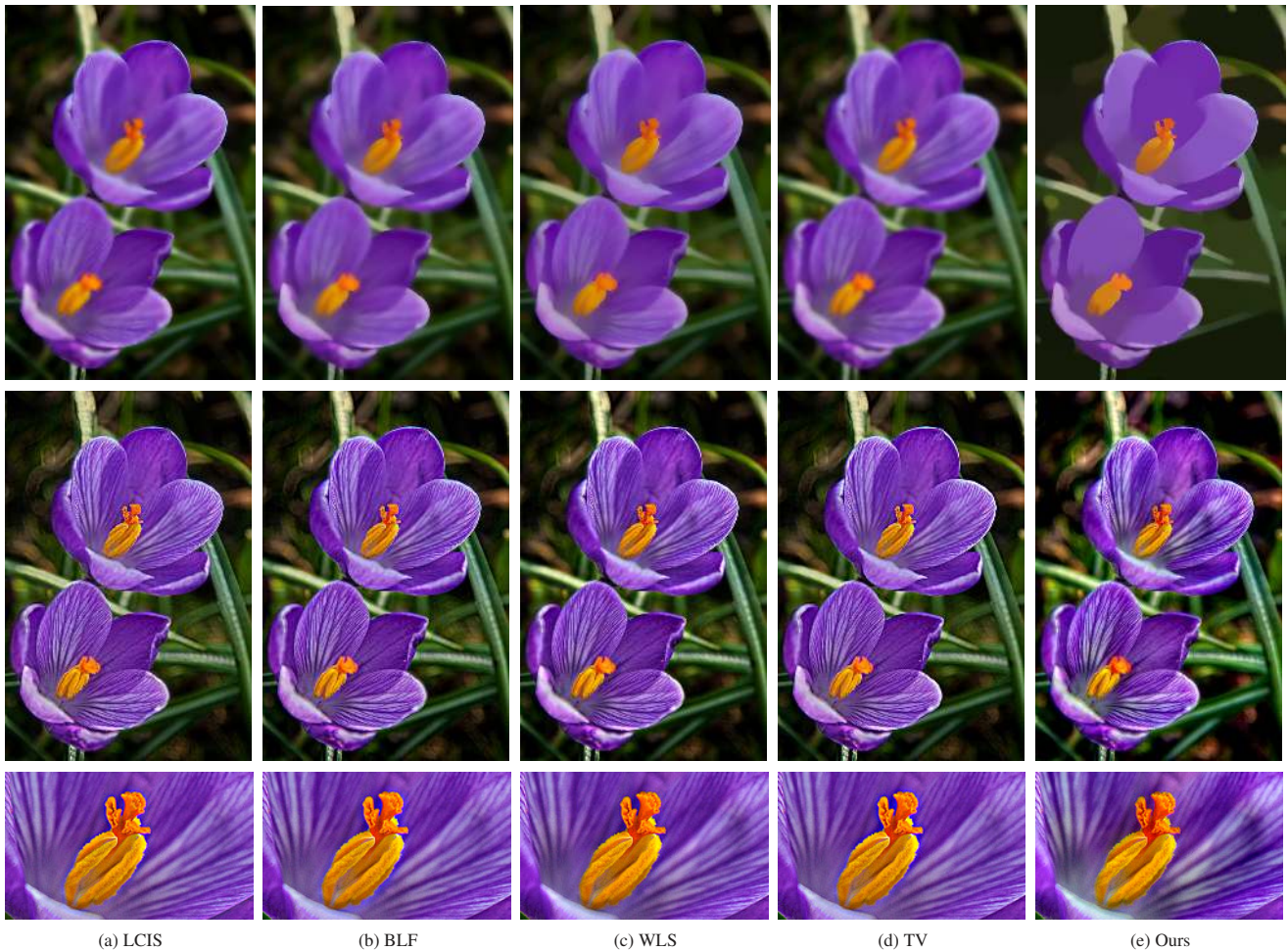


Figure 19: Base-detail separation and manipulation. From top to bottom: the base layers, the detail enhanced results ($2.5\times$), and close-ups. Parameters: LCIS ($n = 1000$, $k = 0.25$ for diffuse $Kval$, $t = 1$ for diffuse $Tstep$), BLF ($\sigma_s = 4$, $\sigma_r = 0.2$), WLS ($\alpha = 1.2$, $\lambda = 0.8$), TV ($\lambda = 2E-2$), and ours ($\lambda = 3E-2$).

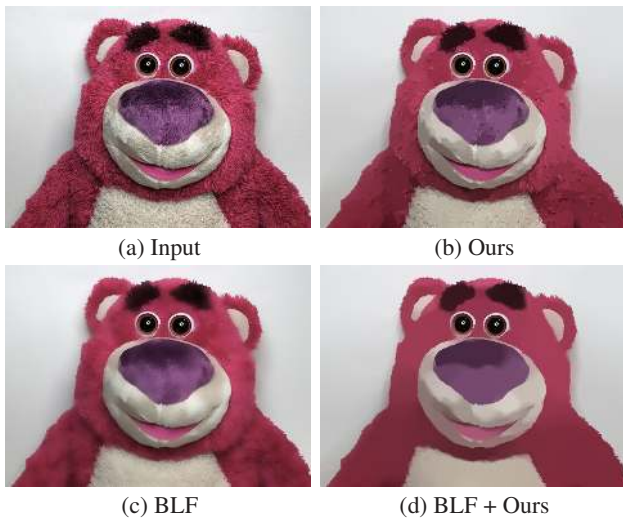


Figure 20: Tone mapping results.



(a) Tone mapping ($\lambda = 0.4$) (b) Tone mapping ($\lambda = 0.07$)

Figure 21: An excessively large λ causes unnatural reflection in (a) in tone mapping. Result (b) is produced with a more appropriate λ .



(a) Input

(b) Ours

(c) BLF

(d) BLF + Ours

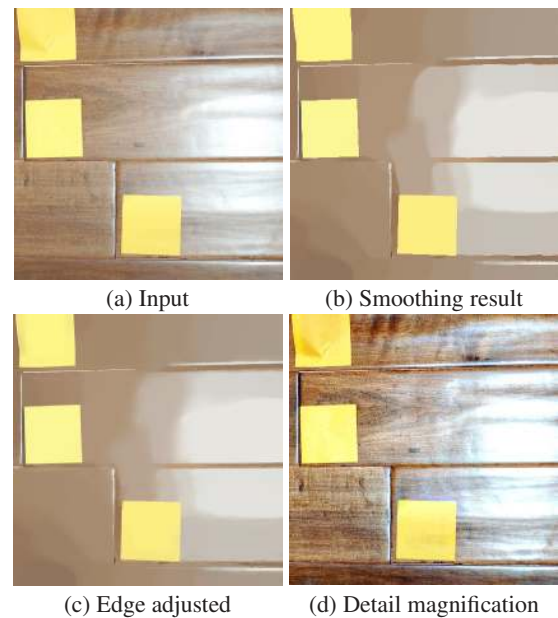
Figure 22: Novel smoothing effect to remove small-resolution structures that are with large amplitudes. We propose applying bilateral filtering and our method consecutively to accomplish it.

result contains nearly no low-amplitude edges, and no blurriness is caused. More results are in the project website.

Tone Mapping HDR tone mapping is another popular application that can be achieved by decomposing an HDR image into a piece-wise smooth base layer conveying most of the energy and a detail layer [Tumblin and Turk 1999; Durand and Dorsey 2002; Choudhury and Tumblin 2003; Li et al. 2005; Farbman et al. 2008]. The base layer is then nonlinearly mapped to a low dynamic range and is re-combined with the detail layer. The base layer is required to preserve sharp discontinuities to avoid halos [Tumblin and Turk 1999] and be smooth enough for reasonable contrast maintenance in range compression, as discussed in [Choudhury and Tumblin 2003].

In the tone mapping framework of Durand and Dorsey [2002], we use our smoothing method for layer decomposition, which is applied to the logarithmic HDR images. One result is shown in Fig. 20. Structures are preserved or enhanced in the tone mapped image.

It is notable that the quality of tone mapping results can be affected by parameter tuning in our layer decomposition. It is possible for



(a) Input

(b) Smoothing result

(c) Edge adjusted

(d) Detail magnification

Figure 23: Wide illumination transition. Using a large λ in our method makes textures be removed; the result is shown in (b). Based on it, layer-based detail enhancement is achieved, shown in (d).

some results to present visual artifacts when the smoothing weight λ is not appropriately set. One example is shown in Fig. 21(a), where the floor is with a blocky reflection. It is caused by applying strong smoothing, which sends structures excessively to the detail layer, making smooth gradients be flattened. Using a smaller weight can dampen the problem, as illustrated in the result (b).

5 Discussion and Limitations

We have presented a well-principled and powerful smoothing method based on the mechanism of discretely counting spatial changes, which can remove low-amplitude structures and globally preserve and enhance salient edges, even if they are boundaries of very narrow objects.

As our system does not use spatial filtering or averaging, it can be regarded as complementary to previous local approaches. Interestingly, when combined with local filtering, our method can produce novel effects. For the example shown in Fig. 22, applying our method alone remains part of the fluff texture because it is with high amplitude. Only bilaterally filtering the image contrarily blurs main boundaries under strong smoothing. We propose first applying bilateral filtering, which lowers the amplitudes of noise-like structures more than those of long coherent edges, followed by our method to globally sharpen prominent edges. Result in (d) only contains large-scale salient edges, profiting main structure extraction and understanding.

Limitations Over-sharpening is sometimes unavoidable in challenging circumstances to remove details. In the example shown in Fig. 23, strong illumination variation spans many pixels. To remove textures, our method may produce an over-sharpened result, as exemplified in (b). This result, however, can still be used in detail magnification. After edge adjustment as described in Sec. 4.4 and taking the result as the base layer, we magnify only details. The final result is shown in Fig.23(d). The aforementioned parameter tuning for tone mapping is another limitation.

Acknowledgements

We thank Michael S. Brown for his help in making the video, and the following people and flicker users for the photos used in the paper: John McCormick, conner395, cyber-seb, T-KONI, Remi Longva, dms_a_jem. This work is supported by a grant from the Research Grants Council of the Hong Kong SAR (project No. 413110).

References

- ARBELAEZ, P., MAIRE, M., FOWLKES, C., AND MALIK, J. 2011. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 898–916.
- BAE, S., AND DURAND, F. 2007. Defocus magnification. *Comput. Graph. Forum* 26, 3, 571–579.
- BAE, S., PARIS, S., AND DURAND, F. 2006. Two-scale tone management for photographic look. *ACM Trans. Graph.* 25, 3, 637–645.
- BAEK, J., AND JACOBS, D. E. 2010. Accelerating spatially varying gaussian filters. *ACM Trans. Graph.*
- BLACK, M. J., SAPIRO, G., MARIMONT, D. H., AND HEEGER, D. 1998. Robust anisotropic diffusion. *IEEE Transactions on Image Processing* 7, 3, 421–432.
- BLAKE, A., AND ZISSERMAN, A. 1987. *Visual reconstruction*. The MIT Press.
- BOYKOV, Y., VEKSLER, O., AND ZABIH, R. 2001. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* 23, 11, 1222–1239.
- CHEN, J., PARIS, S., AND DURAND, F. 2007. Real-time edge-aware image processing with the bilateral grid. *ACM Trans. Graph.* 26, 3, 103.
- CHOUDHURY, P., AND TUMBLIN, J. 2003. The trilateral filter for high contrast images and meshes. In *Rendering Techniques*, 186–196.
- COMANICIU, D., AND MEER, P. 2002. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 24, 5, 603–619.
- CRIMINISI, A., SHARP, T., ROTHER, C., AND PÉREZ, P. 2010. Geodesic image and video editing. *ACM Trans. Graph.* 29, 5, 134.
- DABOV, K., FOI, A., KATKOVNIK, V., AND EGIAZARIAN, K. O. 2007. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing* 16, 8, 2080–2095.
- DECARLO, D., AND SANTELLA, A. 2002. Stylization and abstraction of photographs. *ACM Trans. Graph.* 21, 3, 769–776.
- DONOHU, D. 2006. Compressed sensing. *IEEE Transactions on Information Theory* 52, 4, 1289–1306.
- DURAND, F., AND DORSEY, J. 2002. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Trans. Graph.* 21, 3, 257–266.
- FARBMAN, Z., FATTAL, R., LISCHINSKI, D., AND SZELISKI, R. 2008. Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Trans. Graph.* 27, 3.
- FARBMAN, Z., FATTAL, R., AND LISCHINSKI, D. 2010. Diffusion maps for edge-aware image editing. *ACM Trans. Graph.*
- FATTAL, R., AGRAWALA, M., AND RUSINKIEWICZ, S. 2007. Multiscale shape and detail enhancement from multi-light image collections. *ACM Trans. Graph.* 26, 3, 51.
- FATTAL, R. 2009. Edge-avoiding wavelets and their applications. *ACM Trans. Graph.* 28, 3.
- KASS, M., AND SOLOMON, J. 2010. Smoothed local histogram filters. *ACM Trans. Graph.* 29, 4.
- LEVIN, A., LISCHINSKI, D., AND WEISS, Y. 2004. Colorization using optimization. *ACM Trans. Graph.* 23, 3, 689–694.
- LEVIN, A., FERGUS, R., DURAND, F., AND FREEMAN, W. T. 2007. Image and depth from a conventional camera with a coded aperture. *ACM Trans. Graph.* 26, 3, 70.
- LI, Y., SUN, J., TANG, C.-K., AND SHUM, H.-Y. 2004. Lazy snapping. *ACM Trans. Graph.* 23, 3, 303–308.
- LI, Y., SHARAN, L., AND ADELSON, E. H. 2005. Compressing and companding high dynamic range images with subband architectures. *ACM Trans. Graph.* 24, 3, 836–844.
- LISCHINSKI, D., FARBMAN, Z., UYTENDAELE, M., AND SZELISKI, R. 2006. Interactive local adjustment of tonal values. *ACM Trans. Graph.* 25, 3, 646–653.
- LIU, J., SUN, J., AND SHUM, H.-Y. 2009. Paint selection. *ACM Trans. Graph.* 28, 3.
- MAIRAL, J., BACH, F., PONCE, J., SAPIRO, G., AND ZISSERMAN, A. 2009. Non-local sparse models for image restoration. In *ICCV*, 2272–2279.
- MAJI, S., VISHNOI, N., AND MALIK, J. 2011. Biased normalized cuts. In *CVPR*.
- PARIS, S., AND DURAND, F. 2006. A fast approximation of the bilateral filter using a signal processing approach. In *ECCV (4)*, 568–580.
- PARIS, S., HASINOFF, S. W., AND KAUTZ, J. 2011. Local laplacian filters: Edge-aware image processing with a laplacian pyramid. *ACM Trans. Graph.*
- PERONA, P., AND MALIK, J. 1990. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* 12, 7, 629–639.
- ROTHER, C., KOLMOGOROV, V., AND BLAKE, A. 2004. "grab-cut": interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.* 23, 3, 309–314.
- RUDIN, L., OSHER, S., AND FATEMI, E. 1992. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena* 60, 1-4, 259–268.
- SUBR, K., SOLER, C., AND DURAND, F. 2009. Edge-preserving multiscale image decomposition based on local extrema. *ACM Trans. Graph.* 28, 5.
- TOMASI, C., AND MANDUCHI, R. 1998. Bilateral filtering for gray and color images. In *ICCV*, 839–846.
- TUMBLIN, J., AND TURK, G. 1999. Lcis: A boundary hierarchy for detail-preserving contrast reduction. In *SIGGRAPH*, 83–90.
- WANG, Z., BOVIK, A. C., SHEIKH, H. R., AND SIMONCELLI, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* 13, 4, 600–612.
- WANG, G., WONG, T.-T., AND HENG, P.-A. 2006. Deringing cartoons by image analogies. *ACM Trans. Graph.* 25, 4, 1360–1379.
- WANG, Y., YANG, J., YIN, W., AND ZHANG, Y. 2008. A new alternating minimization algorithm for total variation image reconstruction. *SIAM J. Imaging Sciences* 1, 3, 248–272.
- WEISS, B. 2006. Fast median and bilateral filtering. *ACM Trans. Graph.* 25, 3, 519–526.
- WINNEMÖLLER, H., OLSEN, S. C., AND GOOCH, B. 2006. Real-time video abstraction. *ACM Trans. Graph.* 25, 3, 1221–1226.

