

Mixed-Domain Edge-Aware Image Manipulation

Xian-Ying Li, Yan Gu, Shi-Min Hu, *Member, IEEE*, and Ralph R. Martin, *Member, IEEE*

Abstract—This paper gives a novel approach to edge-aware image manipulation. Our method processes a Gaussian pyramid from coarse to fine, and at each level, applies a nonlinear filter bank to the neighborhood of each pixel. Outputs of these spatially-varying filters are merged using global optimization. The optimization problem is solved using an *explicit* mixed-domain (real space and DCT transform space) solution, which is efficient, accurate, and easy-to-implement. We demonstrate applications of our method to a set of problems including detail and contrast manipulation, HDR compression, non-photorealistic rendering, and haze removal.

Index Terms—mixed-domain, edge-aware image processing, optimization-based image processing, multi-scale method.

I. INTRODUCTION

EDGE-AWARE image processing is an important technique that has received much attention in the computer graphics community. The goal is to process or filter images in some way without destroying fine scale image edges. *Anisotropic diffusion* [1] and *bilateral filtering* [2] are well-known examples of such techniques, originally devised for image smoothing, but later extended to many other applications. More recently, many other edge-aware techniques have been proposed, e.g. [3], [4], [5], [6], [7], [8]. The particular goal of this paper is to provide a simple interface for independently adjusting the *overall* appearance and *details* of input images (in a similar way to [5]), doing so in a way which both keeps fine edges, and avoids introducing unsightly artifacts. Examples of manipulations carried out using our approach are shown in Fig. 1. Unlike previous work, we use a novel mixed-domain (real space and DCT transform space) processing framework, which is fast and simple, and is readily accelerated on the GPU.

A challenging problem in edge-aware image processing, especially for algorithms which process overall appearance and detail separately, is reducing, or if possible avoiding, artifacts at image edges. The problem is illustrated using a simple 1D signal in Fig. 2 (left). Suppose we wish to amplify the detail, without losing the edge. The grey line represents an input signal. The blue line represents the ‘overall’ signal, produced by averaging of the input signal, while the red signal

Xian-Ying Li, Yan Gu, and Shi-Min Hu are with Department of Computer Science and Technology, Tsinghua University, and Beijing Engineering Research Center for Intelligent processing of Visual Media and content security, Beijing, 100084. E-mail: lee.xianying@gmail.com, shimin@tsinghua.edu.cn.

Ralph R. Martin is with the School of Computer Science & Informatics, Cardiff University, Cardiff, Wales, CF24 3AA UK. Email: ralph@cs.cf.ac.uk.

This work was supported by a Chinese National Basic Research Project (2011CB302205), a Chinese National Science and Technology Key Project (2011ZX01042-001-002), the Chinese Natural Science Foundation (61133008, 61120106007), the Chinese National High Technology Research and Development Program (2012AA011802), a UK EPSRC Travel grant, and the Tsinghua-Tencent Joint Laboratory. Xian-Ying Li was partially supported by the New PhD Researcher Award from the Chinese Ministry of Education.

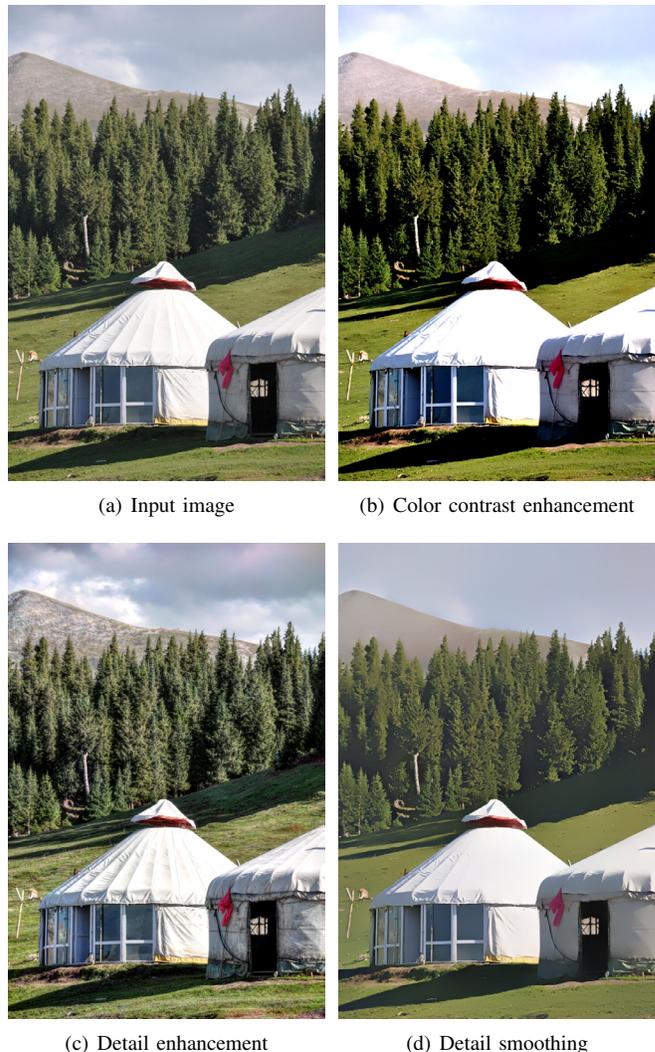


Fig. 1. An input image, and three output images produced by our method. Edges are preserved well at different scales, without unsightly artifacts such as halos, gradient reversals, or aliasing.

represents the ‘detail’, the difference between the input and overall signals. If an enhanced version of the detail is added directly back to the overall signal to give the output (the black signal), the result will overshoot at the step, causing what is often referred to in image processing as a ‘halo’ effect. Conversely, if averaging is done in such a way as to sharpen the input step, the result can even be a ‘gradient reversal’, as shown in Fig. 2 (center). Although weak halos can make image edges stand out, and indeed are deliberately introduced in *unsharp masking*, stronger halos are perceived as unwanted artifacts. Many previous works have thus carefully considered how to reduce or avoid halos [3], [5], [9].

In this paper, we advocate a new method for halo-free edge-

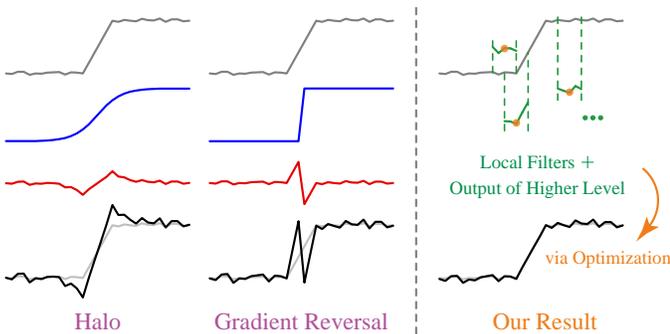


Fig. 2. A simple 1D example. Left, center: poor base-detail decomposition may cause halos or gradient reversals. Right: our artifact-free result (using $\alpha = 2.0$, $\beta = 1.0$; see later).

aware image manipulation. Our method processes a Gaussian pyramid level-by-level, from coarse to fine, in order to reduce halos at each scale. At each level, we consider a set of local filters defined on overlapping windows, and use optimization to merge their outputs with those of the previous level: the result at the current level has a detailed appearance in accordance with the local filters' output, and overall appearance in accordance with the coarser result from the next higher level. Controlling the detailed behavior of our result in this way helps to reduce halos. An enhancement result produced by our method for the same 1D signal is shown in Fig. 2 (right).

Contributions of this paper include:

- A new, rotationally invariant optimization-based formulation for edge-aware image manipulation, which avoids unsightly artifacts such as halos, gradient reversals, and aliasing.
- A direct mixed-domain solution to the resulting optimization problem, which is exact, efficient, and easy-to-implement.

II. RELATED WORK

A. Edge-aware image processing

Edge-aware image processing for computer graphics is a challenging problem that has received much recent attention. Here we briefly review the techniques most closely related to our work.

Anisotropic diffusion [1] uses a non-linear PDE that depends on local image variation to iteratively smooth an image without blurring important features. However, being based on iteration, anisotropic diffusion and related PDE-based methods are slow, and furthermore, parameters are difficult to set [10]. Tumblin and Turk [11] show how anisotropic diffusion can be used as a basis for high dynamic range (HDR) compression, but as pointed out in [3], [9], it tends to over-sharpen image edges.

Bilateral filtering (BLF) [2] provides an alternative approach to edge-aware image smoothing. It uses a local, non-iterative, explicit, data dependent filter, whose simplicity, efficiency, and effectiveness [12], [13], [14], [15], [16] have led to its widespread use [17], [18], [19]. The survey in [20] gives an in-depth treatment. However, as discussed in [3], BLF involves a trade off between edge preservation and data smoothing. Methods relying on BLF to separate a mean surface from detail

may lead to halos at image edges if too much smoothing is applied.

Weighted least squares (WLS) and related methods attempt to avoid such halos by use of more careful *edge-aware decomposition*. WLS [3] computes the smooth (overall) component of the input image by optimizing a quadratic energy based on squared gradients with spatially-varying weights; the weights are made small at sharp image edges, to preserve them. As an alternative, Subr et al. [21] follow the ideas of empirical mode decomposition (EMD) [22]: they smoothly connect the local extrema of the input image to form a maximal envelope and a minimal envelope, and then average these two envelopes to obtain the smooth component. Subr et al.'s approach behaves better than classical EMD at sharp image edges. However, both of these methods involve solution of large sparse linear systems. Although fast numerical solvers like multi-grid methods can potentially be used, these complex methods are not easy to implement. Our algorithm avoids having to solve large linear equations.

Pyramid based methods provide a further approach to edge-aware image processing, using a multi-scale representation such as a Gaussian or Laplacian pyramid, or a wavelet decomposition. Li et al. [23] perform HDR compression by directly manipulating the wavelet coefficients with a careful correction scheme, but as noted by [5], this method can also produce halos with poor choice of parameter values. Later, Fattal [4] developed two novel data-dependent wavelets more suitable for edge-aware image applications than traditional wavelets. Most recently, Paris et al. [5] introduced the *local Laplacian filter (LLF)*, which uses a set of simple local filters to directly and ingeniously manipulate the Laplacian pyramid. However, the LLF is computationally costly because sub-pyramids must be constructed for each element of the Laplacian pyramid. Aubry et al.'s unpublished work [24] reveals that LLF is closely related to anisotropic diffusion and BLF, and provides an approximate sampling-based algorithm to accelerate LLF. However, for a desired accuracy, its ability to accelerate the method depends on the parameter settings used. Our method in this paper uses Paris et al.'s local filters, but merges them in a different, multi-scale optimization-based scheme. Unlike the above techniques, we do not directly manipulate either pyramid or wavelet coefficients.

The *domain transform* is a recent technique proposed by Gastal and Oliveira [6] for real-time edge-aware image processing. It is based on a data-dependent transform from the 5D image manifold to 2D real space that preserves geodesic distance. However, like many other fast filtering techniques, the domain transform is not rotationally invariant (i.e. filtering a rotated image and rotating a filtered image produce different results). Our approach is rotationally invariant.

B. Optimization-based image processing

Optimization is widely used in image processing, with a variety of objective functions.

Gradient domain image processing [9], [25], [26], [27], [28], [29], [30], [31] solves problems by manipulating image gradients via optimization instead of directly changing pixel

values, taking advantage of the fact that the human visual system (HVS) is more sensitive to local contrast than absolute intensities. Fattal et al. [9] utilize a Gaussian pyramid to compute a gradient attenuation function for gradient domain HDR compression without halos. Similarly, we use a multi-scale scheme to produce halo-free results. Recently, Bhat et al. [32] provided a uniform quadratic optimization framework, *Gradient Shop*, supporting a number of applications, such as saliency sharpening, pseudo re-lighting, non-photorealistic rendering, de-blocking, and colorization. However, its ability to obtain various effects relies on carefully designed spatially-varying weights acting as constraints on gradients, and as a result, the algorithm can not be accelerated using frequency-domain approaches like those in [33]. We further explore the relation of our work to gradient domain techniques later in the paper.

Quadratic energies defined on image sub-windows have also been utilized for image matting [34] and haze removal [35]. However, these methods again require the solution of large linear systems.

Other techniques such as total variation optimization [36] and L_0 smoothing [37] have considered non-quadratic energies in order to obtain outputs with ‘sparse gradients’. However, these require elaborate iterative algorithms, which are generally slow and may lack mathematical guarantees of convergence and stability [38].

We give a novel multi-scale optimization-based approach to edge-aware image manipulation. At each scale, we optimize a quadratic objective function over sub-windows to merge the outputs of a set of local edge-aware filters. Unlike WLS, coefficients of our objective function have a spatially-invariant structure, which permits a fast frequency-domain solution to the resulting optimization problem. Use of a multi-scale scheme helps to reduce halos.

III. ALGORITHM

In this section, we explain our algorithm for single-channel images only. For color images, we decompose the input image into channels (in e.g. RGB space or CIELAB space), and process each channel independently. This is a common strategy for processing color images [32], [39].

A. Overview

We express the input image as a real-valued 2D signal I .

Our algorithm allows the user to freely manipulate the *detailed* appearance and the *overall* appearance of I separately, without having to precisely specify how to decompose the image into these components. Following LLF, we use three parameters to control the output: α controls detail ($\alpha > 1$ causes detail enhancement and $\alpha < 1$ causes detail smoothing), β controls the overall appearance ($\beta > 1$ causes intensity range expansion and $\beta < 1$ causes range compression), and σ provides a threshold to determine what comprises detail.

To achieve the desired effects, we first create a Gaussian pyramid for I with l levels from bottom to top: I_0, I_1, \dots, I_{l-1} , where $I = I_0$. We then compute the output O gradually from top to bottom: $O_{l-1}, O_{l-2}, \dots, O_0$, where $O = O_0$.

In the following, we suppose that I_k and O_k share a rectangular definition domain D_k . The pixel values for I_k and O_k are denoted by $I_{k,p}$ and $O_{k,p}$ for each pixel $p \in D_k$.

For the top level, we simply scale the signal, relative to its average, to obtain the output:

$$O_{l-1,p} = \text{Avg}(I_{l-1}) + \beta(I_{l-1,p} - \text{Avg}(I_{l-1})), \quad (1)$$

where $\text{Avg}(I_{l-1})$ is the average pixel value over I_{l-1} . Note that the top level output has the desired properties for overall appearance, but lacks detail. We gradually correct the detail in the output as we proceed through the levels, as we now explain.

For each $k < l - 1$, we compute O_k while taking two concerns into account.

On one hand, we would like O_k to be like the up-sampled signal of the output from the level above: $\text{Upsample}(O_{k+1})$. However, because of the blurring effects of *both* the down-sampling *and* up-sampling, they should differ by a Gaussian convolution. Thus, we would like the following to be satisfied:

$$G * O_k = \text{Upsample}(O_{k+1}), \quad (2)$$

where G is a Gaussian kernel and $*$ is the convolution operator. In the discrete setting, we replace the Gaussian kernel by a standard 5×5 binomial kernel for both down-sampling and up-sampling [40], so G here is a 9×9 binomial kernel.

On the other hand, we would like our result O_k to have the desired detail. Consider a sub-window (not necessarily rectangular) w_p of neighbours for each pixel $p \in D_k$; in practice we choose w_p to comprise those pixels inside the disk centered at p of some radius N . Within this sub-window, we use a carefully designed local filter that takes the parameters α, β, σ to provide an ideal output. The goal is that:

$$O_{k,q} = \text{LocalFilter}_{\alpha,\beta,\sigma,w_p}(I_{k,q}) + C_{k,p} \quad (3)$$

for every $p \in D_k$, and for each pixel $q \in w_p$. At first, it might seem that we should set $O_{k,q} = \text{LocalFilter}_{\alpha,\beta,\sigma,w_p}(I_{k,q})$ for each $q \in w_p$. However, this term is meant to control the detail, which is essentially determined by local differences in values, or gradients, to which human perception is particularly sensitive [32]. Thus, we may add a constant for each sub-window w_p , as doing so does not affect the gradient; the $C_{k,p}$ above are these constants.

Our local filter design follows the ideas in [5] but with certain changes to better fit our algorithm, as explained in Section III-B. Our main algorithmic contribution is an optimization framework to trade-off the two requirements in Eqs. (2) and (3) (see Section III-C), via a frequency-domain solution that is efficient and exact (see Section III-D).

B. Local filtering

For each pixel p , $\text{LocalFilter}_{\alpha,\beta,\sigma,w_p}$ is actually a *filter bank*: it leaves the center pixel value $I_{k,p}$ unchanged, but generates a new value for each of the other pixels in the window centered on p . It treats each other pixel in a two-scale manner using threshold σ . Specifically, $\text{LocalFilter}_{\alpha,\beta,\sigma,w_p}(I_{k,q})$ generates the following output for each pixel $q \in w_p$:

$$I'_{k,p,q} = I_{k,p} + \text{sign}(I_{k,q} - I_{k,p})M_{\alpha,\beta,\sigma}(|I_{k,q} - I_{k,p}|), \quad (4)$$

where $M_{\alpha,\beta,\sigma}(d)$ is a non-linear map:

$$M_{\alpha,\beta,\sigma}(d) = \begin{cases} \sigma(d/\sigma)^{1/\alpha} & \text{if } d \leq \sigma, \alpha \leq 1 \\ \tau_d \sigma(d/\sigma)^{1/\alpha} + (1 - \tau_d)d & \text{if } d \leq \sigma, \alpha > 1 \\ (\sigma^2 + \beta^2(d^2 - \sigma^2))^{1/2} & \text{if } d > \sigma \end{cases} \quad (5)$$

Here, $\tau_d = \max(0, \min(1, 100 \times (d - 0.01)))$ to avoid noise amplification while enhancing detail. Examples of $M_{\alpha,\beta,\sigma}(d)$ are shown graphically in Fig. 3. The multiple values produced for each output pixel are harmonized to a single value by our optimization step.

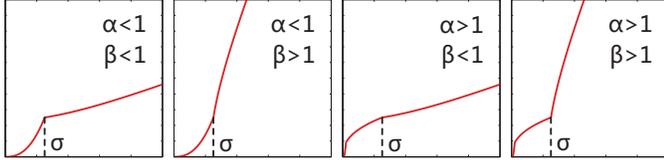


Fig. 3. Mapping function $M_{\alpha,\beta,\sigma}(d)$ for different α and β .

The only difference between our local filter in Eq. (4) and the one in [5] is that we take a hyperbolic curve instead of a straight line when $d > \sigma$. This ensures that, if the difference d is large, the mapping function $M_{\alpha,\beta,\sigma}(d)$ produces a result very close to βd (and does not differ from βd by a constant like the approach in [5]). As a result, our local filter better follows the overall appearance inherited from the top level, via Eq. (1).

C. Global optimization

We now formulate a global optimization scheme to trade-off Eqs. (2) and (3); it also reconciles the multiple outputs of Eq. (4). Our objective function has two terms.

The first measures directly the extent to which Eq. (2) is not satisfied. Its role is to control the overall appearance of the target O_k :

$$E_b(O_k) = \|G * O_k - \text{Upsample}(O_{k+1})\|^2. \quad (6)$$

For the second, we use “averaged squared differences” to express the requirement in Eq. (3):

$$E_d(O_k, C_k) = \sum_{p \in D_k} \sum_{q \in w_p} (O_{k,q} - I'_{k,p,q} - C_{k,p})^2 / W, \quad (7)$$

where $I'_{k,p,q}$ comes from Eq. (4), and $C_{k,p}$ are the constants explained earlier. W is the number of pixels in w_p .

These two terms are combined to give an overall energy to be optimized:

$$E(O_k, C_k) = E_d(O_k, C_k) + \lambda E_b(O_k). \quad (8)$$

We simultaneously determine O_k and C_k by minimizing this energy. λ is a positive parameter; simply setting $\lambda = 1$ works well in practice (for further discussion, see later).

Note that Eq. (8) is a quadratic function, with a unique minimum, so optimizing it is equivalent to solving a linear equation. However, the coefficients of Eq. (8) have a spatially-invariant structure, unlike those of an arbitrary quadratic function. This observation permits an efficient and exact frequency-domain method to optimize Eq. (8).

D. Frequency-domain solution

Suppose that O_k and C_k minimize the energy in Eq. (8). Therefore $\partial E / \partial C_{k,p} = 0$, which implies, for each $p \in D_k$:

$$C_{k,p} = \sum_{q \in w_p} (O_{k,q} - I'_{k,p,q}) / W. \quad (9)$$

By expanding the signals to cover the whole plane, we can rewrite Eq. (9) in the following form involving a convolution:

$$C_k = L * O_k - S_k, \quad (10)$$

where L is the kernel for unweighted averaging within a centered sub-window, and S_k satisfies $S_{k,p} = \sum_{q \in w_p} I'_{k,p,q} / W$.

Also, we have that $\partial E / \partial O_k = 0$, which yields:

$$O_k - T_k - L * C_k + \lambda G * (G * O_k - O'_k) = 0, \quad (11)$$

where $O'_k = \text{Upsample}(O_{k+1})$, and $T_{k,p} = \sum_{q \in w_p} I'_{k,q,p} / W$.

Note that the signals S_k and T_k are different. Intuitively, $S_{k,p}$ denotes “the average of values that pixel p hopes its neighbors have”, while $T_{k,p}$ denotes “the average of values that pixel p ’s neighbors hope pixel p has”.

Taking a Discrete Cosine Transform (DCT), the convolutions in Eqs. (10) and (11) become multiplications in the frequency-domain. Substituting Eq. (10) into Eq. (11) to eliminate C_k , and denoting the DCT transform by $\mathcal{D}(\cdot)$, we obtain a direct and exact solution:

$$O_k = \mathcal{D}^{-1} \left(\frac{\lambda \mathcal{D}(G) \mathcal{D}(O'_k) + \mathcal{D}(T_k) - \mathcal{D}(L) \mathcal{D}(S_k)}{\lambda \mathcal{D}(G)^2 + 1 - \mathcal{D}(L)^2} \right). \quad (12)$$

To extend the image signals to the whole plane as needed by DCT, we reflect signals about the image boundaries. In principle, other Fourier-related transforms could be used instead—we use DCT as it is the simplest and most common Fourier-related transform that can prevent artifacts at the boundaries (e.g., if we used DFT with cyclic extension, signal values would wrap around from one side to the opposite image edge, causing unwanted artifacts).

Use of Fourier-related transforms to speed up optimization is a standard technique (e.g. see [33] and [41]). However, unlike previous work, we have extra variables (the C_k) in our problem: we carefully eliminate these variables to obtain an explicit solution. Another feature of our solution is that we have intermediate variables (S_k and T_k), which must be calculated in real space, resulting in a *mixed-domain* computation, done partly in real space and partly in DCT transform space.

In detail, using Eq. (12), O_k can be calculated with just one 2-dimensional DCT and one 2-dimensional inverse DCT, using the observations that:

- The DCTs of G and L can be efficiently calculated. G can be written as $G = G_x * G_y$, where G_x and G_y are 1-dimensional 9-tap binomial kernels in the x and y directions, giving $\mathcal{D}(G)_{x,y} = \mathcal{D}(G_x)_x \mathcal{D}(G_y)_y$. For L , we calculate each element of $\mathcal{D}(L)$ according to the formal definition of DCT; this is efficient because L has few identical non-zero elements. (These DCTs must actually be computed—although G and L are local, and the same for all images, their DCTs depend on the size of D_k .)

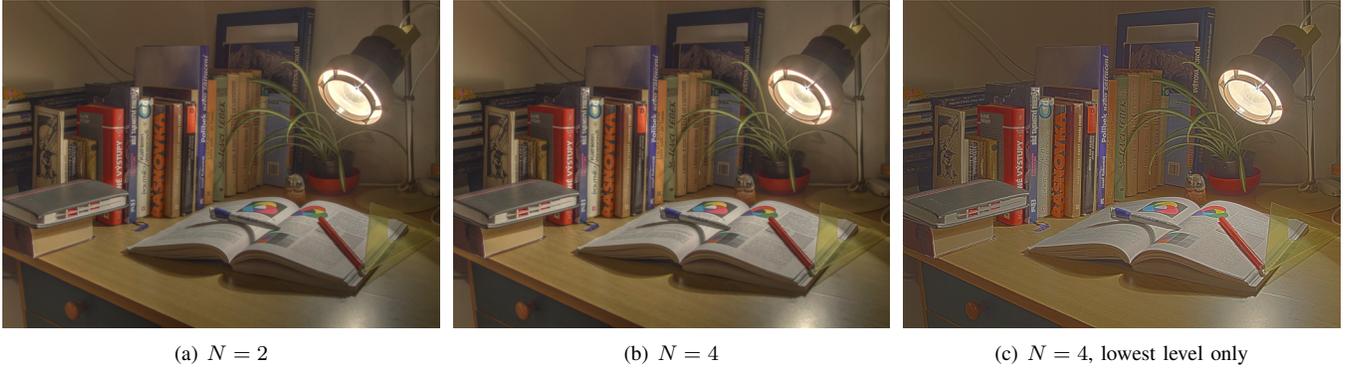


Fig. 4. HDR compression results. (a,b): $N = 2$ and $N = 4$, processing the whole pyramid. (c): $N = 4$, processing the lowest level only. All the results are with parameters $\alpha = 1.0$, $\beta = 0.1$, $\sigma = \ln(2.5)$.

- We can first calculate an intermediate signal: $U_k = \lambda G * \text{Upsample}(O_{k+1}) + T_k - L * S_k$. This is again efficient because a convolution with G or L can be calculated via 1-dimensional convolutions. The numerator in Eq. (12) equals $\mathcal{D}(U_k)$ and thus can be calculated via a single 2-dimensional DCT.

Pseudocode for our entire mixed-domain edge-aware image manipulation is summarized in Algorithm 1.

Algorithm 1 Mixed-Domain Edge-Aware Image Manipulation

Input : I , α , β , σ , λ .

Output: O .

Construct the Gaussian pyramid of I (with l levels).

Compute O_{l-1} using Eq. (1).

for k from $l-1$ to 0 **do**

 Compute $I'_{k,p,q} = \text{LocalFilter}_{\alpha,\beta,\sigma,w_p}(I_{k,q})$ for $q \in w_p$.

 Compute S_k and T_k .

$DG \leftarrow \text{DCT}(G)$, $DL \leftarrow \text{DCT}(L)$.

$U_k \leftarrow \lambda G * \text{Upsample}(O_{k+1}) + T_k - L * S_k$.

$DU_k \leftarrow \text{DCT}(U_k)$.

for all frequencies f **do**

$DO_{k,f} \leftarrow DU_{k,f} / (\lambda(DG_f)^2 + 1 - (DL_f)^2)$.

end for

$O_k \leftarrow \text{InverseDCT}(DO_k)$.

end for

$O \leftarrow O_0$.

IV. DISCUSSION

Since we use a quadratic objective function, our approach is closely related to gradient-domain image processing techniques. To see this, consider a 1×2 (or 2×1) window w with pixels q_1, q_2 . The energy term used to preserve details in this w would be:

$$E_{d,w}(O_{q_1}, O_{q_2}, C) = (O_{q_1} - I'_{q_1} - C)^2 + (O_{q_2} - I'_{q_2} - C)^2, \quad (13)$$

whose minimum given an arbitrary constant C is achieved when $C = (O_{q_1} - I'_{q_1} + O_{q_2} - I'_{q_2})/2$. The minimum value is:

$$\begin{aligned} \min_C E_{d,w}(O_{q_1}, O_{q_2}, C) &= ((O_{q_1} - I'_{q_1}) - (O_{q_2} - I'_{q_2}))^2 / 2 \\ &= ((O_{q_1} - O_{q_2}) - (I'_{q_1} - I'_{q_2}))^2 / 2. \end{aligned} \quad (14)$$

This corresponds to requiring the gradient of the output, $O_{q_1} - O_{q_2}$, to be like the desired gradient, $I'_{q_1} - I'_{q_2}$, which is the basic principle used in gradient-domain image processing [32]. However, in our approach, we choose a set of larger sub-windows providing more extensive overlap, so that our output is ‘more consistent’. Bigger windows also provide more information allowing us to produce better local detail. Fig. 4(a,b) demonstrate use of our method for HDR compression, using different sub-window sizes of $N = 2$ and $N = 4$. The smaller N produces a dull result, while the larger N better provides extra detail.

The use of a pyramid is important in our method to reduce halos. Fig. 4(c) shows the result for $N = 4$ if we only process the bottom level using our method (i.e. we simply scale I_1 following Eq. (1) to obtain O_1 , and then solve a single optimization on the bottom level to obtain O_0). Unsightly halos can be seen near the edges of the desk, as well as around the open book’s pages. These halos result from the combination of the constraints on both detailed and overall appearance. The output near an edge may be enhanced or reduced due to the detail constraint, but the output far from edges tends to be quite close to the ‘target base’, i.e. $\text{Upsample}(O_1)$. Incompatibility of the detail constraint and target base causes halos. Similar problems arise in WLS [3]. WLS simply uses the input signal as the target base, and reduces halos by applying data-dependent weights to the terms for detail constraint. Our approach does not need such weights; halos are reduced due to use of a series of carefully estimated target bases that better suit the detail constraint; these target bases are computed via a multi-scale process. The computational cost of our algorithm decreases rapidly at higher levels; processing the whole pyramid costs about 4/3 the time of processing just the bottom level. For the remaining examples in this paper, we set $N = 4$, constructing the Gaussian pyramid such that the width and height of the top level image I_{l-1} are at least 8.

In our objective function in (8), λ is a positive parameter to trade-off the detail constraint and the overall constraint. Using a small λ produces better detailed appearance and worse overall appearance; an extremely small λ causes color tone deviation of the output (see Fig. 5(b)). Using a large λ produces better overall appearance and worse detailed appearance, and

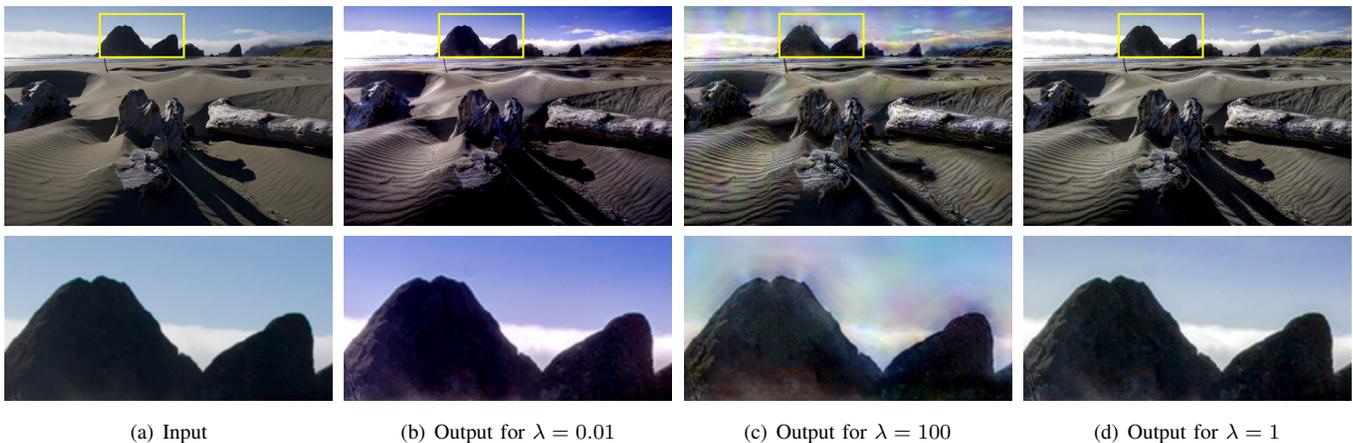


Fig. 5. Detail enhancement results with different settings of λ . Extremely small λ leads to color tone deviation (b), while extremely large λ causes ringing artifacts (c). Setting $\lambda = 1$ works well in practice (d). Other parameters are $\alpha = 4.0$, $\beta = 1.0$, $\sigma = 0.2$.

an extremely large λ causes ringing artifacts (Fig. 5(c)). We set $\lambda = 1$ for all examples in this paper, which works well.

Our method and Aubry et al.’s work [24] are two different approaches to speed up local Laplacian filters. In some ways, they are complimentary—Aubry et al. use sampling to efficiently estimate coefficients of the Laplacian pyramid, while we use an optimization-based framework to merge the local filters’ outputs level by level based on a Gaussian pyramid. However, the ability of Aubry et al.’s approximation algorithm to achieve a speed up for a desired accuracy depends on the upper cutoff frequency of the mapping function $M_{\alpha,\beta,\sigma}$ (see [24]). It is faster than our method for e.g. HDR compression (where the parameter α is typically small and therefore the upper cutoff frequency of $M_{\alpha,\beta,\sigma}$ is low), but for detail enhancement, where the parameter α is large, it would become much slower. The performance and accuracy of our method do not depend on the choice of mapping function $M_{\alpha,\beta,\sigma}$.

Our algorithm also has the benefit of rotational invariance (i.e. filtering a rotated image and rotating a filtered image produce identical results), because we used disk-shaped sub-windows. This makes our method different from many other fast filtering techniques that are not rotationally invariant (e.g. the domain transform [6]). An example is shown in Fig. 6.

V. APPLICATIONS

We now showcase several applications to demonstrate the effectiveness of our method. Further results can be found in the supplementary material.

A. Detail and contrast manipulation

Our method can be directly applied for purposes of image detail and contrast manipulation. Fig. 7 shows the results of processing a single input image (in RGB color space) with different parameter settings. In each case, edges are preserved well without halos resulting (see, for example, the edges of the flower, and the stigma of the flower). The parameter α controls the detail in the output; $\alpha < 1$ causes detail smoothing and $\alpha > 1$ causes detail enhancement. The parameter β controls the color contrast of the output; $\beta < 1$ causes contrast

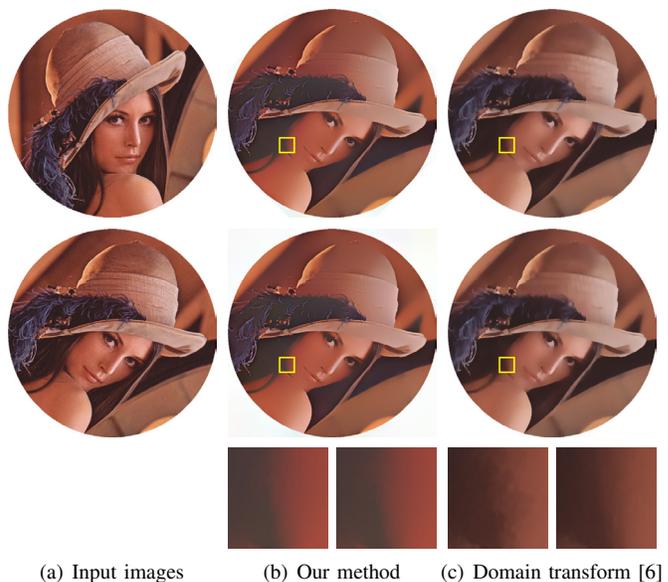


Fig. 6. Image smoothing using our method ($\alpha = 0.0$, $\beta = 1.0$, $\sigma = 0.15$) and domain transfer [6] ($\sigma_s = 30$, $\sigma_r = 0.4$). 1st row: rotating a filtered image. 2nd row: filtering a rotated image. 3rd row: close-ups of the results. Our method produces identical results as it is rotationally invariant.

reduction, while $\beta > 1$ leads to contrast enhancement. The parameter σ allows a tradeoff between detail and contrast manipulation: a larger σ increases the effects of α and decreases the effects of β . Thus, Fig. 7(c,d) show stronger detail smoothing and enhancement than Fig. 7(a,b); however, the contrast manipulation effects in Fig. 7(c,d) are weaker than in Fig. 7(a,b). σ is an important parameter which must be carefully chosen since it distinguishes details from edges. If we use $\alpha < 1$ to smooth an image with a large σ , e.g. as on the top-right of Fig. 7, some red color ‘leaks’ from the flower onto the green leaf. We suggest σ should lie within $[0.1, 0.2]$ for typical detail and contrast manipulation tasks.

In Fig. 8 we compare detail enhancement results using L_0 smoothing [37] with those from our method. Note that L_0 smoothing tends to produce smooth base components with sharp edges. Boosting from those components causes gradient

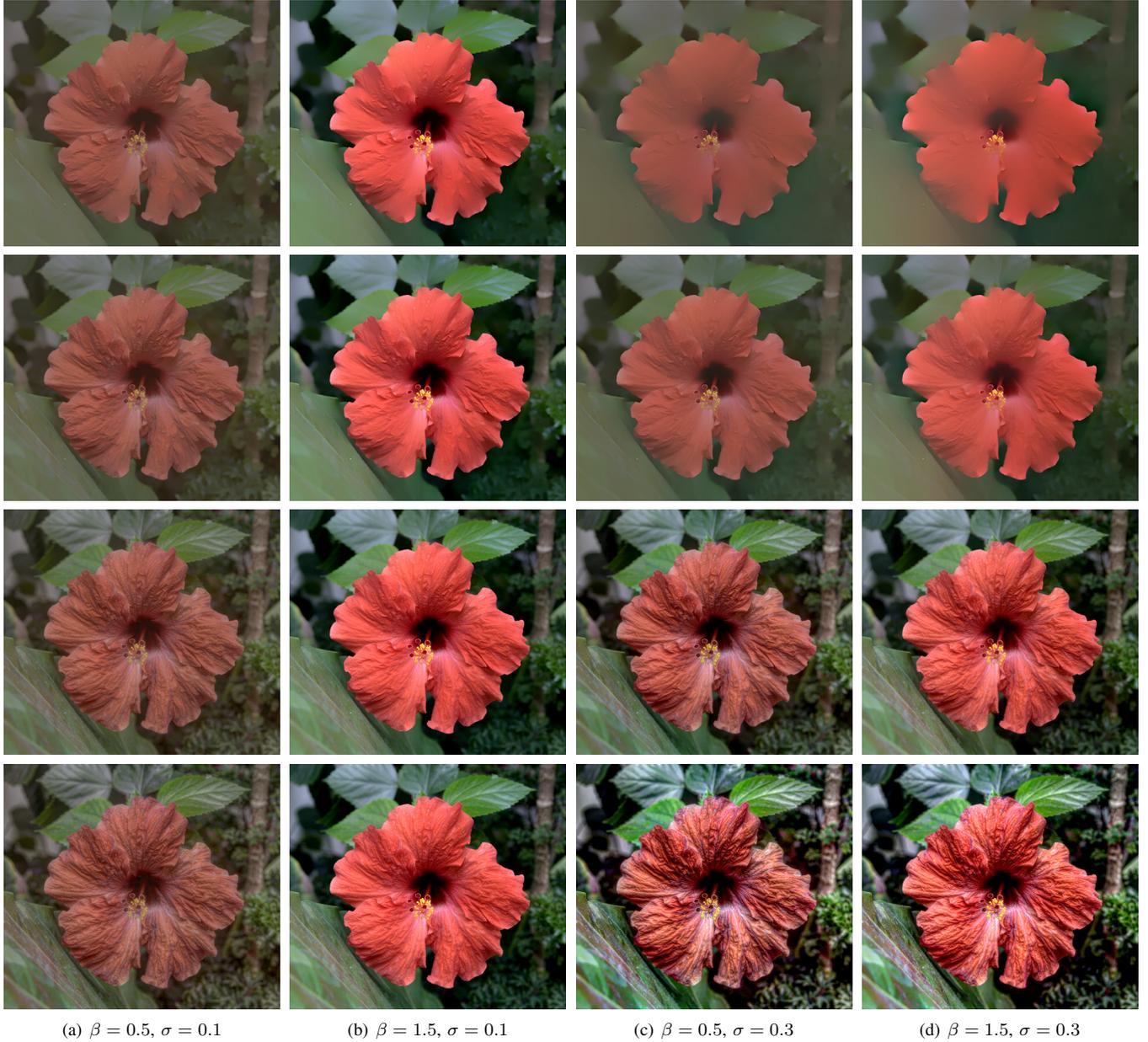


Fig. 7. Detail and contrast manipulation (in RGB color space). Top to Bottom: $\alpha = 0.3, 0.6, 1.5, 4.0$. (a,b): contrast reduction and enhancement with a normal threshold $\sigma = 0.1$. (c,d): contrast reduction and enhancement with an extreme threshold $\sigma = 0.3$.

reversals as discussed in Section I. Our method does not produce such artifacts.

B. HDR compression

HDR compression, or tone mapping, is concerned with compressing the intensity range of an HDR image while keeping details. Our implementation follows previous work, i.e. we first compute the luminance channel using a linear combination of RGB values: $L_i = (20r_i + 40g_i + b_i)/61$ [17], and then process the logarithm of the luminance $\ln(L_i)$ [9], [11] using our method to compress the range without reducing detail, setting $\beta < 1$ and $\alpha \geq 1$. Given the output $\ln(L_o)$, we offset and scale $\ln(L_o)$ to make its maximum 0 and its minimum $-\ln(100/(\max(L'_i) - \min(L'_i)))$, where

L'_i is the sub-signal of L_i obtained by throwing away the top and bottom 0.5% values [5]. Finally, we convert $\ln(L_o)$ back to L_o , and compute the output RGB values. We set $(r_o, g_o, b_o) = (r_i L_o / L_i, g_i L_o / L_i, b_i L_o / L_i)$, then gamma correct each channel with an exponent of $1/2.2$ [5], and finally clamp values to $[0, 1]$ for display.

In our experiments, we set $\beta = 0.1, \sigma = \ln(2.5)$ with $\alpha = 1$ to give ‘photorealistic’ output and $\alpha = 4$ to give ‘exaggerated’ output. It would also be possible to post-process our results using interactive techniques like those in [42] but doing so is beyond the scope of this paper.

We compared our method with several previous methods (a BLF-based method [17], WLS [3], and LLF [5]) in Fig. 9, by setting parameters for ‘exaggerated’ effects. Observe that the BLF-based method causes edge halos (e.g. see close-ups

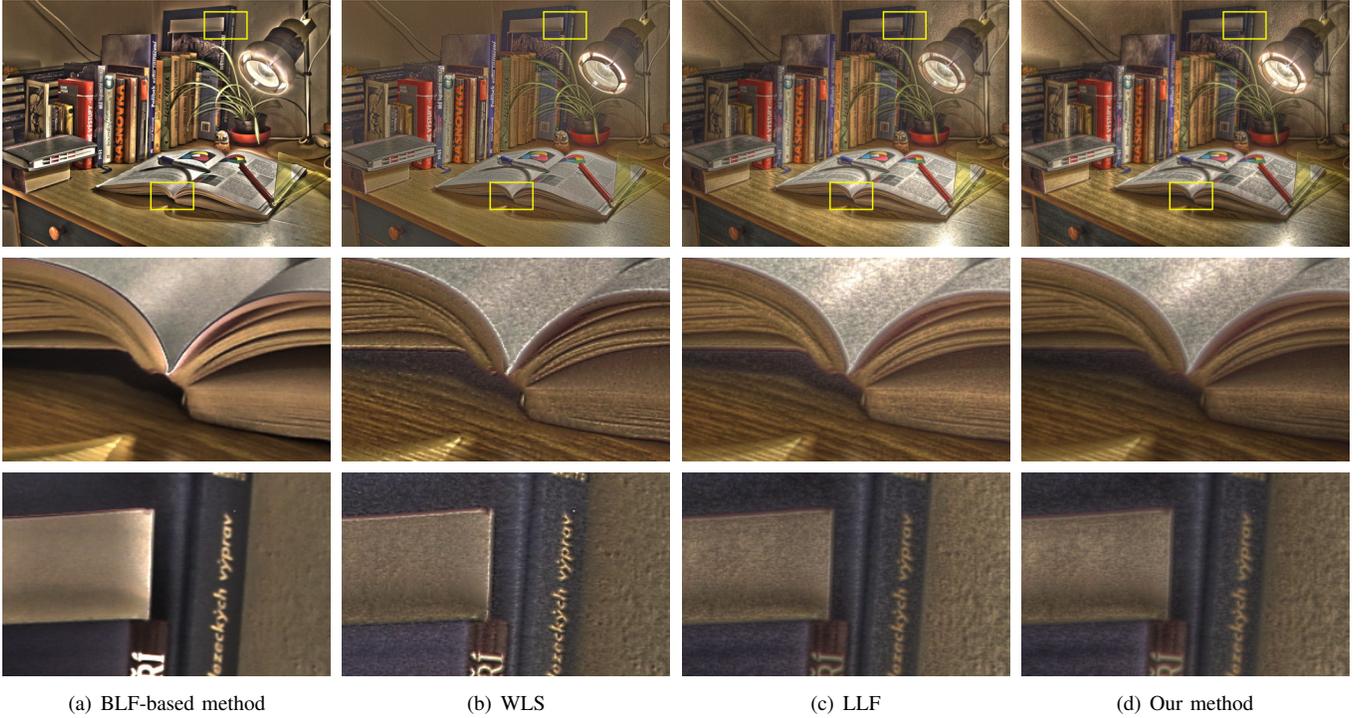


Fig. 9. Comparison of our method with previous HDR compression methods. (a): BLF [17]. (b): Weighted Least Squares [3]. (c): Local Laplacian Filter [5]. (d): our method; $\alpha = 4.0$, $\beta = 0.1$, $\sigma = \ln(2.5)$.

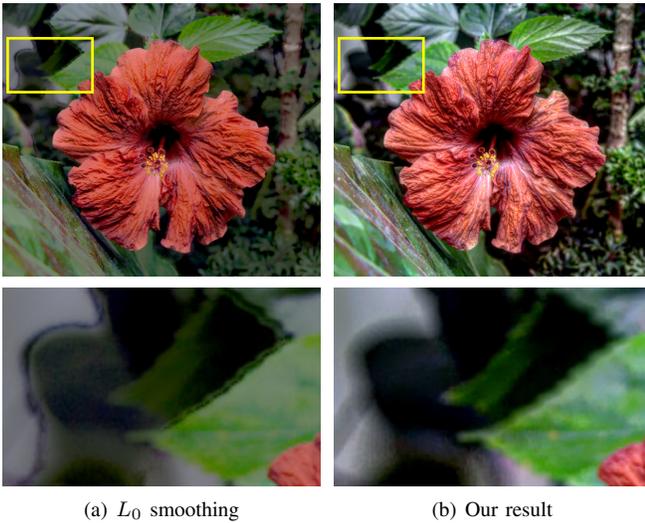


Fig. 8. Detail enhancement. (a): using L_0 smoothing; $\lambda = 0.015$; $3\times$ detail boosting. (b): our result; $\alpha = 4.0$, $\beta = 1.0$, $\sigma = 0.3$.

of the edges of the book's pages, the corner of the shadow on the desk, and the white square on the blue book). WLS largely reduces halos, but some remain visible. In contrast, our method produces halo-free output like the state-of-the-art LLF method. This is unsurprising, since both methods use very similar local filters. However, our approach merges those local filters' multiple outputs via a novel optimization framework, with an exact and efficient mixed-domain solution, unlike LLF.

Further 'photorealistic' and 'exaggerated' results using our method are shown in Fig. 10.

C. Non-photorealistic abstraction

The edge-aware smoothing ability of our method is well suited to the needs of image abstraction [39], [43]. In Fig. 11, we process an input image (in RGB color space) with parameters $\alpha = 0$, $\beta = 1$, $\sigma = 0.15$ to obtain a highly smoothed image, and then overlay this smoothed image with its DoG (Difference of Gaussian) edges to give a non-photorealistic abstraction effect.

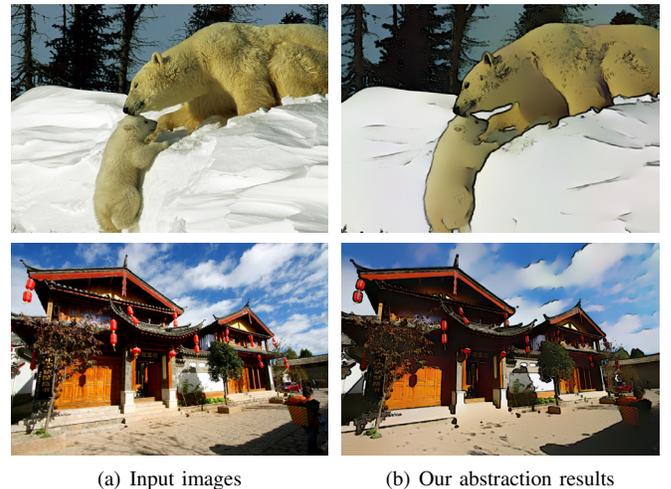


Fig. 11. Image Abstraction (use RGB color space). (a): input images. (b): our abstraction results; $\alpha = 0.0$, $\beta = 1.0$, $\sigma = 0.15$.



Fig. 10. HDR compression. (a): ‘photorealistic’ rendition with parameters $\alpha = 1.0, \beta = 0.1, \sigma = \ln(2.5)$. (b): ‘exaggerated’ rendition with parameters $\alpha = 4.0, \beta = 0.1, \sigma = \ln(2.5)$.

D. Haze removal

Our method can also be applied to *joint image filtering*, i.e. to process an image with the edge information provided by another reference image, denoted R . We simply replace the local filters in Eq. (4) by:

$$I'_{k,p,q} = I_{k,p} + (I_{k,q} - I_{k,p}) \frac{M_{\alpha,\beta,\sigma}(|R_{k,q} - R_{k,p}|)}{|R_{k,q} - R_{k,p}|} \quad (15)$$

We demonstrate haze removal as an application of this idea, in Fig. 12. Following [35], we first use the dark channel prior to roughly estimate a haze transmission map. We then use our method with the joint local filters in Eq. (15) to smooth the transmission map, taking the input hazy image as the reference image. As the reference is a color image, we use Euclidean distance in CIELAB color space for the term $|R_{k,q} - R_{k,p}|$ in Eq. (15).

VI. IMPLEMENTATION

We have implemented our method on both CPU and GPU. In the CPU case, we used the FFTW 3.3 library [44] to compute DCTs. Our implementation takes less than 3.2 seconds to process each channel of the 800×1200 image in Fig. 1, using a single thread on an Intel Core 750, 2.66GHz, with sub-window size $N = 4$. Our GPU implementation used NVIDIA CUFFT [45] for DCTs on an NVIDIA GeForce 560 GTX. The processing time was reduced to under 280 milliseconds.

Our algorithm is easy to program, since it involves only DCTs (widely supported by libraries) and a few pixel-wise operations, rather than complex solvers for large sparse linear systems (e.g. multi-grid methods) or complicated signal decomposition.



Fig. 12. Haze removal (use CIELAB color space for reference images). (a): input hazy images. (b): our haze-free results; $\alpha = 0.0, \beta = 1.0, \sigma = 20$.

VII. SUMMARY AND FUTURE WORK

This paper has presented a novel approach to edge-aware image manipulation, controlled by several intuitive parameters directly affecting the detailed and overall appearance of the output. Our method processes a Gaussian image pyramid level-by-level from coarse to fine. At each level, a set of local filters operates on overlapping sub-windows, and a spatial-domain optimization problem is solved in the frequency-domain to merge the local filters’ outputs. This frequency-domain solution is exact, fast, and easy-to-implement. The results are halo-free, and can match the state-of-art results in

several applications.

Our mixed-domain method is best suited to rectangular images; others can be padded to a rectangle at a slightly increased computational cost. A minor drawback of our method is that we can currently only support two-scale manipulation (detail versus overall appearance), as we have to process the whole pyramid to reduce halos. Currently we simply set the parameter $\lambda = 1$; if λ is too small or too large artifacts result. It would be interesting in future to study how to adaptively choose λ based on the values of σ, α, β , to obtain the best trade-off between detail and overall appearance.

We believe that the idea of mixing spatial-domain and frequency-domain processing will have other useful applications in image processing. In particular, it would be interesting to adapt our frequency-domain solver to other more general optimization problems (e.g. we could use a linear transform instead of the offsetting model in Eq. (7), like in [34]). We also hope to generalize our optimization approach to other spatially-varying local filters for further applications such as up-sampling and deblurring. Finally, extending our approach to video by taking temporal consistency into account is a further problem of interest.

ACKNOWLEDGEMENTS

We thank the reviewers for helpful comments. Many source images here come from other researchers and we thank them for making their data available. In particular, we thank Shanshan Lu for sharing the input image in Fig. 1(a) and Sylvain Paris for sharing the results in [5] to enable a comparison to be made.

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, Jul. 1990.
- [2] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *ICCV '98: Proceedings of the Sixth International Conference on Computer Vision*. Washington, DC, USA: IEEE Computer Society, 1998, pp. 839–846.
- [3] 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, pp. 67:1–10, 2008.
- [4] R. Fattal, "Edge-avoiding wavelets and their applications," *ACM Transactions on Graphics*, vol. 28, no. 3, pp. 22:1–10, 2009.
- [5] S. Paris, S. W. Hasinoff, and J. Kautz, "Local laplacian filters: edge-aware image processing with a laplacian pyramid," *ACM Transactions on Graphics*, vol. 30, no. 4, pp. 68:1–12, 2011.
- [6] E. S. L. Gastal and M. M. Oliveira, "Domain transform for edge-aware image and video processing," *ACM Transactions on Graphics*, vol. 30, no. 4, pp. 69:1–11, 2011.
- [7] Z. Su, X. Luo, and A. Artusi, "A novel image decomposition approach and its applications," *The Visual Computer*, 2012, to appear.
- [8] P. Thevenaz, D. Sage, and M. Unser, "Bi-exponential edge-preserving smoother," *IEEE Transactions on Image Processing*, vol. 21, no. 9, pp. 3924–3936, Sep. 2012.
- [9] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," *ACM Transactions on Graphics*, vol. 21, no. 3, pp. 249–256, 2002.
- [10] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*. Springer-Verlag, 2006.
- [11] J. Tumblin and G. Turk, "LCIS: A boundary hierarchy for detail-preserving contrast reduction," in *SIGGRAPH '99: Proc. 24th annual conference on Computer graphics and interactive techniques*. New York, NY, USA: ACM, 1999, pp. 83–90.
- [12] B. Weiss, "Fast median and bilateral filtering," *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 519–526, 2006.
- [13] J. Chen, S. Paris, and F. Durand, "Real-time edge-aware image processing with the bilateral grid," *ACM Transactions on Graphics*, vol. 26, no. 3, pp. 103:1–9, 2007.
- [14] S. Paris and F. Durand, "A fast approximation of the bilateral filter using a signal processing approach," *Int. J. Computer Vision*, vol. 81, no. 1, pp. 24–52, 2009.
- [15] Q. Yang, K.-H. Tan, and N. Ahuja, "Real-time $O(1)$ bilateral filtering," in *IEEE CVPR*, 2009, pp. 557–564.
- [16] J. Baek and D. E. Jacobs, "Accelerating spatially varying gaussian filters," *ACM Transactions on Graphics*, vol. 29, no. 6, pp. 169:1–10, 2010.
- [17] 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.
- [18] P. Choudhury and J. Tumblin, "The trilateral filter for high contrast images and meshes," in *ACM SIGGRAPH 2005 Courses*. New York, NY, USA: ACM, 2005.
- [19] 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, pp. 51:1–9, 2007.
- [20] S. Paris, P. Kornprobst, J. Tumblin, and F. Durand, "Bilateral filtering: Theory and applications," *Foundations and Trends in Computer Graphics and Vision*, vol. 4, no. 1, pp. 1–73, 2009.
- [21] 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. 147:1–9, 2009.
- [22] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.
- [23] Y. Li, L. Sharan, and E. H. Adelson, "Compressing and companding high dynamic range images with subband architectures," *ACM Transactions on Graphics*, vol. 24, no. 3, pp. 836–844, 2005.
- [24] M. Aubry, S. Paris, S. W. Hasinoff, J. Kautz, and F. Durand, "Fast and robust pyramid-based image processing," *Technical Report*, pp. 1–11, 2011.
- [25] P. Pérez, M. Gangnet, and A. Blake, "Poisson image editing," *ACM Transactions on Graphics*, vol. 22, no. 3, pp. 313–318, 2003.
- [26] A. Levin, A. Zomet, S. Peleg, and Y. Weiss, "Seamless image stitching in the gradient domain," in *ECCV '04: Proceedings of the Eighth European Conference on Computer Vision*. Springer-Verlag, 2004, pp. 377–389.
- [27] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen, "Interactive digital photomontage," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 294–302, 2004.
- [28] A. Agarwala, "Efficient gradient-domain compositing using quadrees," *ACM Transactions on Graphics*, vol. 26, no. 3, pp. 94:1–10, 2007.
- [29] A. Agrawal and R. Raskar, "Gradient domain manipulation techniques in vision and graphics," in *ICCV 2007 Courses*. Washington, DC, USA: IEEE Computer Society, 2007.
- [30] J. McCann and N. S. Pollard, "Real-time gradient-domain painting," *ACM Transactions on Graphics*, vol. 27, no. 3, pp. 93:1–7, 2008.
- [31] Z.-F. Xie, R. Lau, Y. Gui, M.-G. Chen, and L.-Z. Ma, "A gradient-domain-based edge-preserving sharpen filter," *The Visual Computer*, vol. 28, no. 12, pp. 1195–1207, 2012.
- [32] P. Bhat, C. L. Zitnick, M. Cohen, and B. Curless, "Gradientshop: A gradient-domain optimization framework for image and video filtering," *ACM Transactions on Graphics*, vol. 29, no. 2, pp. 10:1–14, 2010.
- [33] P. Bhat, B. Curless, M. Cohen, and C. L. Zitnick, "Fourier analysis of the 2d screened poisson equation for gradient domain problems," in *ECCV '08: Proceedings of the Eighth European Conference on Computer Vision*. Springer-Verlag, 2008, pp. 114–128.
- [34] A. Levin, D. Lischinski, and Y. Weiss, "A closed-form solution to natural image matting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 228–242, Feb. 2008.
- [35] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," in *IEEE CVPR*, 2009, pp. 1956–1963.
- [36] 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.
- [37] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via L_0 gradient minimization," *ACM Transactions on Graphics*, vol. 30, no. 6, pp. 174:1–12, 2011.

- [38] S. J. Wright, "Sparse optimization methods," in *Conference on Advanced Methods and Perspectives in Nonlinear Optimization and Control*, 2010.
- [39] H. Winnemöller, S. C. Olsen, and B. Gooch, "Real-time video abstraction," *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 1221–1226, 2006.
- [40] P. J. Burt and E. H. Adelson, "The laplacian pyramid as a compact image code," *IEEE Transactions on Communication*, vol. 31, no. 4, pp. 532–540, Apr. 1983.
- [41] S. Cho and S. Lee, "Fast motion deblurring," *ACM Transactions on Graphics*, vol. 28, no. 5, pp. 145:1–7, 2009.
- [42] D. Lischinski, Z. Farbman, M. Uyttendaele, and R. Szeliski, "Interactive local adjustment of tonal values," *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 646–653, 2006.
- [43] D. DeCarlo and A. Santella, "Stylization and abstraction of photographs," *ACM Transactions on Graphics*, vol. 21, no. 3, pp. 769–776, 2002.
- [44] FFTW, "<http://www.fftw.org/>."
- [45] CUFFT, "<http://developer.nvidia.com/cufft>."



Ralph R. Martin obtained his PhD in 1983 from Cambridge University. Since then he has been at Cardiff University, as Professor since 2000, where he leads the Visual Computing research group. He is also a Guest Professor at Tsinghua and other universities in China, and Director of Scientific Programmes of the One Wales Research Institute of Visual Computing. His publications include over 200 papers and 12 books covering such topics as solid modelling, surface modelling, reverse engineering, intelligent sketch input, mesh processing, video processing, computer graphics, vision based geometric inspection, and geometric reasoning. He is a Fellow of the Learned Society of Wales, the Institute of Mathematics and its Applications, and the British Computer Society. He is on the editorial boards of several journals, including *Computer Aided Design*, *Computer Aided Geometric Design*, and *Geometric Models*.



Xian-Ying Li received his bachelor's degree in computer science in 2008 from Tsinghua University, Beijing. He is currently a PhD candidate at Tsinghua University. He received a gold medal at the International Mathematical Olympiad in 2004. His research interests include computer graphics, geometric modeling, image processing, interpolation and approximation, and computational origami.



Yan Gu is an undergraduate student in the Department of Computer Science and Technology of Tsinghua University, Beijing. His research interests include computer graphics, geometric modeling, and rendering.



Shi-Min Hu received the PhD degree from Zhejiang University in 1996. He is currently a professor in the Department of Computer Science and Technology at Tsinghua University, Beijing. His research interests include digital geometry processing, video processing, rendering, computer animation, and computer-aided geometric design. He is associate Editor-in-Chief of *The Visual Computer* (Springer), and on the editorial boards of *Computer-Aided Design* and *Computer & Graphics* (Elsevier). He is a member of the IEEE and ACM.