Robust Image Filtering Using Joint Static and Dynamic Guidance

Bumsu Han1,* Minsu Cho1,* Jean Ponce2,*

1Inria 2École Normale Supérieure / PSL Research University

Abstract

Regularizing images under a guidance signal has been used in various tasks in computer vision and computational photography, particularly for noise reduction and joint upsampling. The aim is to transfer fine structures of guidance signals to input images, restoring noisy or altered structures. One of main drawbacks in such a data-dependent framework is that it does not handle differences in structure between guidance and input images. We address this problem by jointly leveraging structural information of guidance and input images. Image filtering is formulated as a nonconvex optimization problem, which is solved by the majorization-minimization algorithm. The proposed algorithm converges quickly while guaranteeing a local minimum. It effectively controls image structures at different scales and can handle a variety of types of data from different sensors. We demonstrate the flexibility and effectiveness of our model in several applications including depth super-resolution, scale-space filtering, texture removal, flash/non-flash denoising, and RGB/NIR denoising.

1. Introduction and Background

Many tasks in computer vision and computational photography can be formulated as ill-posed inverse problems, and thus theoretically and practically, require regularization. In the classical setting, this is used to obtain a smoothly varying solution and/or ensure stability [4]. Recent work on joint regularization (or joint filtering) [10, 15, 33] provides a new perspective on the regularization process, with a great variety of applications including stereo correspondence [28, 34], optical flow [28], joint upsampling [8, 15, 20, 25], dehazing [10], noise reduction [27, 33], and texture removal [35]. The basic idea of joint regularization is that the structure of a guidance image is transferred to an input image, e.g., for preserving sharp structure transitions while smoothing the input image. It assumes that the guidance image has enough structural information to restore noisy or altered structures in the input image.

Joint regularization has been used with either static or dynamic guidance images. Static guidance regularization (e.g., [10]) provides an output image by modulating the input image with an affinity function that depends on the similarity of features in the guidance signal. This static guidance is fixed during the optimization. It can reflect internal properties of the input image itself, e.g., its gradient [15, 26], or be another signal aligned with the input image, e.g., a near infrared (NIR) image [33]. This framework determines the structure of the output image by referring to that of the guidance image only, and does not consider structural (or statistical) dependencies and inconsistencies between input and guidance images. This is problematic, especially in the case of data from different sensors, e.g., depth and color images. Dynamic guidance regularization (e.g., [35]) uses an affinity function obtained from the regularized input image. It is assumed that the affinity between neighboring pixels can be determined more accurately from already regularized images, than from the input image itself [2, 35]. This method is inherently iterative, and dynamic guidance (the regularized input image, i.e., a potential output image) is updated at every step. In contrast to static guidance regularization, dynamic guidance regularization does not use static guidance and takes into account of the properties of the input image. Data-dependent static guidance is needed to impose structures on the input image, especially when the input image is not enough in itself to pull out reliable information, e.g., joint upsampling [8, 15, 20, 25].

We present a unified framework for image filtering taking advantage of both static and dynamic guidelines. We address the aforementioned problems by fusing appropriate structures of static and dynamic guidance images, rather than unilaterally transferring structures of guidance images to the input image. To encourage comparison and future work, our source code is available at our project webpage.

2. Related Work

Static or dynamic guidance can be implicit or explicit. Implicit regularization stems from a filtering framework. The input image is filtered using a weight function that depends on the similarity of features in the guidance image
In this way, the structure of the guidance image is transferred to the input image. The bilateral filter (BF) [30], guided filter (GF) [10], and weighted median filter (WMF) [20] have been successfully adapted to static guidance regularization. Two representative filtering methods using dynamic guidance are iterative nonlocal means (INM) [2] and the rolling-guidance filter (RGF) [35]. They share the same filtering framework, but differ in that INM is for preserving textures during noise reduction, while the RGF aims at removing textures through scale-space filtering. This implicit regularization is simple and easy to implement, but the filtering formalization prevents its wide applicability. For example, it is hard to handle input images where information is sparse, e.g., in image colorization [18]. The local nature of this approach might introduce artifacts, e.g., halos and gradient reversal [10]. Accordingly, implicit regularization has been applied in a highly controlled condition, and usually employed as a pre- and/or post-processing for further applications [17, 20]. An alternative approach is to explicitly encode the regularization process into an objective functional, while taking advantage of a guidance signal. The objective functional typically consists of two parts: A fidelity term describes the consistency between input and output images, and a regularization term encourages the output image to have a similar structure to the guidance image. The weighted least-squares (WLS) framework [7] is the most popular explicit regularization method that has been used in static guidance regularization [25]. The regularization term is modeled as a weighted $l_2$ norm. Anisotropic diffusion (AD) [26] is an explicit regularization framework using dynamic guidance. In contrast to INM [2] and the RGF [35], AD updates both input and guidance images at every step; The regularization is performed iteratively with regularized input and updated guidance images. This explicit regularization enables formulating a task-specific model, with more flexibility than using implicit regularization. Furthermore, this type of regularization overcomes several limitations of implicit regularization, such as halos and gradient reversal, at the cost of global intensity shifting [7, 10].

Existing regularization methods typically apply to a limited range of applications and suffer from various artifacts. For example, the RGF is applicable to scale-space filtering only, and suffers from poor edge localization [35]. In contrast, our approach provides a unified model for many applications, gracefully handles most of these artifacts, and outperforms the state of the art in all the cases considered in the paper. Although the proposed model may look similar to WLS [7] and the RGF [35], our nonconvex objective function needs a different solver. Contrary to iteratively reweighted least-squares (IRLS) [5], we do not split a nonconvex regularizer but approximate the objective function by a surrogate (upper-bound) function.

3. Proposed Approach

3.1. Motivation and Problem Statement

There are pros and cons in regularizing images under static or dynamic guidance. Let us suppose the example of depth super-resolution, where a high-resolution (HR) color image (the guidance image) of Fig. 1 (a) is used to upsample ($\times 8$) a low-resolution (LR) depth map (the input image) of Fig. 1 (b). Regularization with static guidance reconstructs the destroyed depth edges by using the color image with high signal-to-noise ratio (SNR) [8, 15], as in the red boxes of Fig. 1 (c). However, this method has difficulties with handling differences in structure between depth and color images, transferring all the structural information of the color image to the depth map, as in the blue boxes of Fig. 1 (c). For regions of high contrast in the color image, e.g., textures, the depth is altered according to the texture pattern [3]. Similarly, the gradient of the depth map becomes similar to that of the color image for regions of low contrast in the color image, e.g., weak edges. This smoothes depth
Dynamic guidance utilizes the contents of the depth maps\(^\text{2}\), avoiding the problems of static guidance regularization, as in the blue boxes of Fig. 1 (d). The depth edges are preserved, and unwanted structures are not transferred. A limitation is that dynamic guidance only does not utilize the abundant structural information that exists in the color image. Thus, depth edges are smoothed, and even eliminated for regions of low contrast in the depth map, as in the red boxes of Fig. 1 (d). This example illustrates the fact that static and dynamic guidance complement each other, and exploiting only one of them is not sufficient to infer high quality structural information from the input image. This problem becomes even worse when input and guidance images come from various types of data and have different statistical characteristics. Our model jointly leverages the structures of static (color image) and dynamic (depth map) guidance, taking advantage of both of them, as shown in Fig. 1 (e).

3.2. Model

Given the input image \(f\), static guidance \(g\), and the output image \(u\) itself (dynamic guidance), we denote by \(f_i\), \(g_i\), and \(u_i\); the corresponding image values at pixel \(i\), with \(i\) ranging over the image domain \(\mathcal{I} \subset \mathbb{N}^2\). Our objective is to infer the structure of the input image by jointly using static and dynamic guidance. The influence of the guidance images on the input image varies spatially, and is controlled by affinity functions that measure similarities between adjacent vertices. Various features (e.g., spatial location, intensity, and textures [13, 25]) and metrics (e.g., Euclidian and geodesic distances [7, 19]) can be utilized to represent distinctive characteristics of vertices on images, and measure their similarities.

We minimize an objective function of the form:

\[
\mathcal{E}(u) = \sum_i c_i (u_i - f_i)^2 + \lambda \Omega(u).
\]  

It consists of fidelity and regularization terms, balanced by the regularization parameter \(\lambda\). The fidelity term helps the solution \(u\) to harmonize well with the input image \(f\) with confidence \(c_i \geq 0\). The regularization term smooths the solution \(u\), and makes it has structures similar to static and/or dynamic guidance, \(g\) and \(u\). In static guidance regularization, \(\Omega(u) = \sum_{i,j} \phi_{\mu}(g_i - g_j)(u_i - u_j)^2\) where \(\phi_{\mu}(x) = \exp(-\mu x^2)\), and \(\mu\) controls the smoothness bandwidth. In a purely equivalent dynamic guidance setting, one would take \(\Omega(u) = \sum_{i,j} \phi_{\nu}(u_i - u_j)(u_i - u_j)^2\) for some \(\nu\), and in a mixed setting: \(\Omega(u) = \sum_{i,j} \phi_{\mu}(g_i - g_j)\phi_{\nu}(u_i - u_j)(u_i - u_j)^2\). These regularizers may be hard to optimize

\[^{2}\text{Dynamic guidance is initially set to the upsampled depth map obtained from static guidance regularization as shown in Fig. 1 (c), not the LR depth map itself.}\]

Figure 2. A nonconvex regularizer \(\psi_{\nu}(x)\), its surrogate functions \(\Psi_{\nu}^1(x)\) [left], and a \(l_1\) regularizer \(\alpha |x|\) (a tight upper bound function of \(\psi_{\nu}(x)\)) [right] when \(\nu = 1\) and \(\alpha = 0.6382\). (Best viewed in color.)

\[\Omega(u) = \sum_{i,j \in \mathcal{N}} \phi_{\mu}(g_i - g_j)\psi_{\nu}(u_i - u_j),\]  

where \(\psi_{\nu}(x) = (1 - \phi_{\nu}(x))/\nu\), i.e., Welsch’s function [11], and \(\mathcal{N}\) is in our implementation, the set of image adjacencies, defined in a local 8-neighborhood system.

Nonconvex Regularizer. As shown in Fig. 2, \(\psi_{\nu}(x)\) is a nonconvex regularizer: Welsch’s function acts as a robust regularizer [11], and thus our objective function makes joint filtering robust to outliers. Inverse diffusion occurs when \(\psi_{\nu}(x)\) decreases, which enhances features having high-frequency structures (e.g., edges and corners) during regularization [6, 9, 26]. It may enhance noise as well, but the static guidance image with high SNR in our model avoids this problem. The combination of a fixed weight function and a nonconvex regularizer is common in variational approaches [37], and they are typically referred to as image- and flow-driven regularizers, respectively. Our model differs from image- and flow-driven regularization in its new regularizer and solver. It is also more flexible (e.g., we can handle sparse data of different resolutions). This leads to new applications such as joint upsampling. Note that most image- and flow-driven regularization techniques have been applied to optical flow only. Finally, we are not aware of existing joint image filters using a nonconvex regularizer.

3.3. Solver

Optimization. Let \(f = [f_i]_{N \times 1}\), \(g = [g_i]_{N \times 1}\), and \(u = [u_i]_{N \times 1}\) denote vectors representing the input image, static guidance and the output image (or dynamic guidance), respectively, where \(N = |\mathcal{I}|\) is the size of images. Let \(\mathcal{W}_g = [\phi_{\mu}(g_i - g_j)]_{N \times N}\), \(\mathcal{W}_u = [\phi_{\nu}(u_i - u_j)]_{N \times N}\), and \(\mathcal{C} = \text{diag}([c_1, \ldots, c_N])\). We can rewrite our objective function in matrix/vector form as:

\[
\mathcal{E}(u) = (u - f)^T C (u - f) + \frac{\lambda}{\mu} \|T (W - W) 1\|
\]  

where \(W = \mathcal{W}_g \circ \mathcal{W}_u\), and \(\circ\) denotes the Hadamard product of the matrices. \(1\) is a \(N \times 1\) vector, where all the
entries are 1. The diagonal entries $c_i$ of $C$ are confidence values for the pixels $i$ of the input image. Minimizing $E$ is a nonconvex optimization problem, which can be solved by the majorization-minimization algorithm (Fig. 3) as follows [16, 22, 36]:

1. **Majorization Step:** Construct a surrogate function $Q^k(u)$ of $E(u)$ such that

$$
\begin{align*}
\mathcal{E}(u) & \leq Q^k(u), \text{ for all } u, \ u^k \in \Theta, \\
\mathcal{E}(u^k) & = Q^k(u^k), \text{ for all } u^k \in \Theta,
\end{align*}
$$

where $\Theta \subset [0, 1]^N$. The nonconvexity in our objective function comes from the regularizer $\psi_{\nu}(x)$ in (2), which has a convex surrogate function $\Psi^\nu_{\nu}(x)$ defined by (see the supplementary material):

$$
\Psi^\nu_{\nu}(x) = \psi_{\nu}(y) + (x^2 - y^2)(1 - \nu \psi_{\nu}(y)),
$$

that is, the curve $x \mapsto \Psi^\nu_{\nu}(x)$ lies above the curve $\psi_{\nu}(x)$ and is tangent to it at the point $x = y [12]$, as shown in Fig. 2. The surrogate objective function $Q^k(u)$ can then be found using (5) as follows:

$$
\begin{align*}
Q^k(u) & = u^T \left[ C + \lambda L^k \right] u - 2f^T C u + f^T C f \\
& \quad - \lambda u^{kT} L^k u^k + \frac{\lambda}{\nu} \| \nu^T (W_g - W_h) I \|.
\end{align*}
$$

$L^k = D^k - W^k$ is a dynamic Laplacian matrix at step $k$, where $W^k = W_g \circ W_h$ and $D^k = \text{diag} \left( \left[ d_{x1}^k, \ldots, d_{xN}^k \right] \right)$ where $d_{xi}^k = \sum_{j=1}^N \phi_{i}(x_i - x_j) \phi_{i}(x_i - x_j)$. Note that the affinity matrix of static guidance is fixed regardless of steps, and that of dynamic guidance is iteratively updated.

2. **Minimization Step:** Obtain the next estimate $u^{k+1}$ by minimizing the surrogate function $Q^k(u)$ w.r.t. $u$ as follows:

$$
\begin{align*}
u^{k+1} & = \arg \min_{u \in \Theta} Q^k(u) = (C + \lambda L^k)^{-1} Cf.
\end{align*}
$$

The above iterative scheme decreases the value of $E(u)$ monotonically in each step, i.e.,

$$
\begin{align*}
E(u^{k+1}) & \leq Q^k(u^{k+1}) \leq Q^k(u^k) = E(u^k),
\end{align*}
$$

where the first and the second inequalities follow from (4) and (7), respectively, and it can be shown to converge to a local minimum of $E$ [31].

**Initialization.** Our solver finds a local minimum, and thus different initializations for $u^0$ (dynamic guidance at $k = 0$) may give different solutions. In our work, two initializations are used: The initial solution $u^0$ can be set to a constant vector, e.g., $u^0 = \mathbb{1}$. Note that the constant initialization,

$$
\begin{align*}
\mathcal{E}(u^0) & \leq Q^k(u^0) \leq Q^k(u^k) = E(u^k),
\end{align*}
$$

regardless of its value, makes $\mathcal{V}^0 = \mathcal{W}$. This initialization is simple, but shows a slow convergence rate as shown in Fig. 4 (a). A good initial solution accelerates the convergence of our solver. We propose to use the following regularizer to compute the initial solution $u^0$:

$$
\Omega_{\nu}(u) = \sum_{i,j \in N_C} \phi_{\nu}(g_i - g_j) \alpha |u_i - u_j|,
$$

where $\alpha$ is set to a positive constant, chosen so $\alpha|x|$ is a tight upper bound of $\psi_{\nu}(x)$ (Fig. 2). This regularizer is convex, and the global minimum of (9) is guaranteed.

**3.4. Properties**

**Convergence.** We show the convergence rate of (7) as $k$ increases, and observe its behavior with different initializations ($u^0 = \mathbb{1}$ and $u^0 = u_{\mathcal{I}}$, a global minimum of (1) using $\Omega = \Omega_{\mathcal{I}}$). Figure 4 shows how (a) the energy and (b) the intensity differences (i.e., $\|u^k - u^{k+1}\|_1$) evolve at each step given the input image in (c). Our solver converges in fewer steps with the $\mathcal{I}$ initialization ($u^0 = u_{\mathcal{I}}$) than with the constant one ($u^0 = \mathbb{1}$), with faster overall speed, despite the overhead of the $\mathcal{I}$ minimization. On this example, our solver with the constant and $\mathcal{I}$ initializations converges in 30 and 7 steps (Fig. 4 (d) and (e)), each of which takes 45 and 20 seconds, respectively. Although our solver with $u^0 = \mathbb{1}$ converges more slowly, the per-pixel intensity difference decreases monotonically, and 5 steps are typically enough to get satisfactory results in both cases.

It should be noted that most filtering methods, except the recently proposed RGF [35], if they are applied repeatedly, eventually converges to a trivial solution, i.e., a constant signal [35], regardless of whether they have implicit or explicit forms (e.g., BF [30] and AD [26]). In contrast, repeatedly solving the linear system of (7) still gives a meaningful solution in the steady-state.

**Scale Adjustment.** There are two approaches to incorporating scale information in image regularization [7, 26, 35].

---

3 A range of intensity values is normalized such that they exist between 0 and 1.

4 In case of a color image, the linear system is solved in each channel.
Let us define diagonal matrices $\mathcal{A}$ and $\mathcal{A}'$ as

$$\mathcal{A} = \lambda(\mathcal{C} + \lambda \mathcal{D}^*)^{-1} \mathcal{D}^*,$$

and

$$\mathcal{A}' = (\mathcal{C} + \lambda \mathcal{D}^*)^{-1} \mathcal{C},$$

such that $\mathcal{A} + \mathcal{A}' = \mathcal{I}$. By multiplying the left- and right-hand sides of (10) by $(\mathcal{C} + \lambda \mathcal{D}^*)^{-1}$, we obtain

$$\mathbf{u}^* = (\mathcal{I} - \lambda(\mathcal{C} + \lambda \mathcal{D}^*)^{-1} \mathcal{D}^* \mathcal{P}^{-1}(\mathcal{C} + \lambda \mathcal{D}^*)^{-1} \mathcal{C} f)
\quad (13)$$

$$= (\mathcal{I} - \mathcal{A})(\mathcal{I} - \mathcal{A} \mathcal{P})^{-1} f = S \mathbf{f},$$

where $\mathcal{P} = \mathcal{D}^{*-1} \mathcal{W}^*$ and

$$S = (\mathcal{I} - \mathcal{A})(\mathcal{I} - \mathcal{A} \mathcal{P})^{-1} = (\mathcal{I} - \mathcal{A}) \sum_{n=0}^{\infty} \mathcal{A}^n \mathcal{P}^n. \quad (14)$$

That is, $S$ is defined as a weighted average of all matrices $\mathcal{P}^n$, $n = 0, \ldots, \infty$. Note that $\mathcal{P}^n$ is the $n^{th}$ order transition probability of the random walker, and its element $p_{ij}^n$ represents the probability that the random walker at the vertex $j$ arrives at the vertex $i$ after $n$ time transitions [24]. As $n$ increases, the random walker can travel far away, and we expect to see coarser structures. Thus, $\sum_{n=0}^{\infty} \mathcal{A}^n \mathcal{P}^n$ considers all paths between two vertices at all scales ($n = 0, \ldots, \infty$), and each $\mathcal{P}^n$ is modulated by the weight $\mathcal{A}^n$ that is controlled by the regularization parameter $\lambda$ as in (11). By increasing $\lambda$, the random walker can travel to a distant vertex more easily. This indicates that the regularization parameter has a similar role to the standard deviation in the Gaussian kernel.

4. Applications

Our model is applied to depth super-resolution, scale-space filtering, texture removal, flash/non-flash denoising, and RGB/NIR denoising. Additional results are available in the supplementary material.

4.1. Depth Super-Resolution

Parameter Settings. In our model, input and guidance images, $f$ and $g$, are set to sparse depth and HR images, respectively, where $c_3 = 1$ if the pixel $i$ of the sparse depth map $f$ has valid data, and otherwise, 0. The constant initialization is used, and the bandwidths and the step index are fixed to all experiments ($\mathbf{u}^0 = 1$, $\mu = 60$, $\nu = 30$, and $k = 10$). The regularization parameter $\lambda$ is set to 0.1 for synthetic examples, and set to 5 for real-world examples due to huge amounts of noise. Other results for the comparison have been obtained from source codes provided by the authors, and all the parameters have been carefully set through extensive experiments for the best performance. For the quantitative comparison, the bad matching errors (BMEs) for all regions and regions near depth discontinuities are measured as $O_{\text{all}} = \sum \{ |u^*_i - u^{gt}_i| > \delta \} / N$ and $O_{\text{disc}} = \sum \{ m_i |u^*_i - u^{gt}_i| > \delta \} / M$, respectively, where $\delta$ is a depth error tolerance [29]. $u^*_i \in \mathbf{u}^*$ and $u^{gt}_i \in \mathbf{u}^{gt}$ represent estimated and ground truth depth maps, respectively. $m$ is a binary mask where $m_i = 1$ if the pixel $i$ belongs to the regions near depth discontinuities, and otherwise, 0, and $M = \|m\|_1$. 

Figure 4. An example of (a) energy evolution and (b) a sum of intensity difference between successive steps, i.e., $\|\mathbf{u}^k - \mathbf{u}^{k+1}\|_1$, given (c) the input image. Our model monotonically converges, and $\sum_{n=0}^{\infty} \mathcal{A}^n \mathcal{P}^n$ considers all paths between two vertices at all scales ($n = 0, \ldots, \infty$), and each $\mathcal{P}^n$ is modulated by the weight $\mathcal{A}^n$ that is controlled by the regularization parameter $\lambda$ as in (11). By increasing $\lambda$, the random walker can travel to a distant vertex more easily. This indicates that the regularization parameter has a similar role to the standard deviation in the Gaussian kernel.
Figure 5. Visual comparison of upsampled depth maps on a snippet of the books sequence in the Middlebury test bed [29]. In contrast to static guidance regularization such as (d) GF [10] and (e) WMF [20], (g) joint static and dynamic guidance model interpolates LR depth maps by considering structures of color and depth images both, preserving sharp depth transitions.

Table 1. Average BMEs of Upsampled Depth Maps on the Middlebury Test Bed [29]

<table>
<thead>
<tr>
<th>Method</th>
<th>BME (10&lt;sup&gt;-5&lt;/sup&gt;)</th>
<th>Std. (10&lt;sup&gt;-5&lt;/sup&gt;)</th>
<th>Disc. (10&lt;sup&gt;-5&lt;/sup&gt;)</th>
<th>Std. (10&lt;sup&gt;-5&lt;/sup&gt;)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilinear Int.</td>
<td>15.98±2.89</td>
<td>38.63±5.17</td>
<td>7.05±1.70</td>
<td>12.05±6.30</td>
</tr>
<tr>
<td>GF [10]</td>
<td>19.85±11.2</td>
<td>35.40±6.96</td>
<td>7.08±2.42</td>
<td>12.05±6.30</td>
</tr>
<tr>
<td>Park et al. [25]</td>
<td>14.81±5.97</td>
<td>22.65±3.89</td>
<td>7.08±3.41</td>
<td>12.05±6.30</td>
</tr>
<tr>
<td>TGV [8]</td>
<td>12.34±6.40</td>
<td>22.73±6.08</td>
<td>7.08±3.41</td>
<td>12.05±6.30</td>
</tr>
<tr>
<td>WMF [20]</td>
<td>9.84±5.48</td>
<td>19.88±7.47</td>
<td>7.08±3.41</td>
<td>12.05±6.30</td>
</tr>
<tr>
<td>Ours</td>
<td>7.08±3.42</td>
<td>12.05±6.30</td>
<td>7.08±3.41</td>
<td>12.05±6.30</td>
</tr>
</tbody>
</table>

Table 2. Quantitative Comparison of Upsampled Depth Maps from Constant and l<sub>1</sub> Initializations on the Middlebury Test Bed [29]

<table>
<thead>
<tr>
<th>Method</th>
<th>BME (10&lt;sup&gt;-5&lt;/sup&gt;)</th>
<th>Std. (10&lt;sup&gt;-5&lt;/sup&gt;)</th>
<th>Disc. (10&lt;sup&gt;-5&lt;/sup&gt;)</th>
<th>Std. (10&lt;sup&gt;-5&lt;/sup&gt;)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilinear Int.</td>
<td>10.05±4.76</td>
<td>7.53±3.54</td>
<td>6.30±5.46</td>
<td>7.53±3.54</td>
</tr>
<tr>
<td>Park et al. [25]</td>
<td>7.23±3.52</td>
<td>7.07±3.36</td>
<td>6.86±5.33</td>
<td>7.07±3.36</td>
</tr>
<tr>
<td>Ours</td>
<td>7.05±3.41</td>
<td>7.08±3.38</td>
<td>6.86±5.33</td>
<td>7.07±3.38</td>
</tr>
</tbody>
</table>

Table 3. BMEs of Upsampled Depth Maps on the Graz Data Set [8]

<table>
<thead>
<tr>
<th>Method</th>
<th>BME (10&lt;sup&gt;-5&lt;/sup&gt;)</th>
<th>Std. (10&lt;sup&gt;-5&lt;/sup&gt;)</th>
<th>Disc. (10&lt;sup&gt;-5&lt;/sup&gt;)</th>
<th>Std. (10&lt;sup&gt;-5&lt;/sup&gt;)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilinear Int.</td>
<td>16.21</td>
<td>17.60</td>
<td>15.83±1.99</td>
<td>15.83±1.99</td>
</tr>
<tr>
<td>GF [10]</td>
<td>19.65</td>
<td>20.68</td>
<td>17.82±4.10</td>
<td>17.82±4.10</td>
</tr>
<tr>
<td>Park et al. [25]</td>
<td>11.83</td>
<td>13.98</td>
<td>11.84±2.14</td>
<td>11.84±2.14</td>
</tr>
<tr>
<td>TGV [8]</td>
<td>8.29</td>
<td>9.84</td>
<td>5.17</td>
<td>5.17</td>
</tr>
<tr>
<td>WMF [20]</td>
<td>8.36</td>
<td>9.81</td>
<td>5.17</td>
<td>5.17</td>
</tr>
<tr>
<td>Ours</td>
<td>9.91</td>
<td>8.09</td>
<td>12.71</td>
<td>12.71</td>
</tr>
</tbody>
</table>

Synthetic Examples. We have synthesized the LR depth map by ×8 downsampling a ground truth image from the Middlebury benchmark data set [29]: Tsukuba, Venus, Teddy, Cones, Art, Books, Dolls, Laundry, Moebius, and Reindeer, and used the corresponding color image as the HR intensity image. Table 1 summarizes average BMEs with the tolerance δ = 1. O<sub>all</sub> has been measured only, except when the ground truth index map of discontinuous regions m is available. Our model outperforms other regularization methods, especially around depth discontinuities. Figure 5 gives quantitative results, and clearly shows the different behavior between static guidance and joint static and dynamic guidance models. For example, the gradient of the depth map becomes similar to that of the color image in static guidance methods [8, 10, 20], which tends to eliminate or smooth depth boundaries, and causes jagged artifacts. This can be further verified by observing statistical distributions of upsampled depth maps as shown in Fig. 6. Table 2 compares average BMEs and processing time of the constant and l<sub>1</sub> initializations by varying the number of steps. This table shows that 1) our solver with the l<sub>1</sub> initialization convergences faster than that with the constant one. For example, the l<sub>1</sub> and constant initializations converge with 5 and 30 steps, respectively, 2) both initializations give almost the same error at the convergence, and 3) the l<sub>1</sub> initialization takes less time than the constant initialization for the convergence.

Real-World Examples. Recently, Ferstl et al. [8] have introduced a benchmark data set where they provide both LR depth maps captured by ToF sensor and highly accurate ground truth depth maps acquired from using structured light. We have performed a quantitative evaluation using this data set [8] in Table 3. The BMEs are computed by setting the error tolerance to 5% of a pre-defined depth range (0~255). This experiment demonstrates that the proposed method outperforms the state of the art.

4.2. Scale-Space Filtering and Texture Removal Parameter Settings. For scale-space filtering, the input image f is guided by itself (g = f). In texture removal, the
static guidance image is set to a Gaussian-filtered version of the input image, \( g = G_\sigma f \) where \( G_\sigma \) is the Gaussian kernel with standard deviation \( \sigma \). The regularization parameter \( \lambda \) and \( \sigma \) vary according to the scale. The bandwidths and the step index are fixed to all experiments (\( \mu = 5, \nu = 40, \) and \( k = 5 \)) in both applications.

**Scale-Space Filtering.** A scale-space representation can be obtained by repeatedly applying the regularization method while varying the regularization parameter \( \lambda \). Figure 7 shows examples of the scale-space constructed by (a) and (b) WLS [7], (c) RGF [35], and (d) our model. The WLS [7], a representative of static guidance regularization, alters the scale of structures by varying the regularization parameter. It suffers from global intensity shifting [10] (Fig. 7(a)) or does not preserve structural information at coarse scales (Fig. 7(b)). This could be alleviated by dynamic guidance regularization as in the RGF [35]. However, the RGF does not use the structure of the input image, and the scale is controlled by isotropic Gaussian kernels, which leads to poor boundary localization at coarse scales (Fig. 7(c)). In contrast, our model uses the structures of input and desired output images, and the scale depends on the regularization parameter, providing well localized boundaries even at coarse scales. Moreover, it is robust to global intensity shifting (Fig. 7(d)). The scale-space representation meets two criteria: causality and immediate localization. Causality means that any feature at a coarse scale must possess a cause at a finer scale [26]. Immediate localization means that object boundaries should be sharp and coincide well with the meaningful boundaries at each scale. We have empirically found that our model meets the causality cond-

![Figure 7. Examples of the scale-space representation obtained by (a) WLS [7] (from left to right) \( \lambda = 5 \times 10^3, 3 \times 10^3, 2 \times 10^2, \mu = 40 \), (b) WLS [7] (from left to right) \( \lambda = 50, 300, 2000, \mu = 5 \), (c) RGF [35] (from left to right) \( \sigma_r = 4, 10, 50, \sigma_s = 0.05, k = 5 \), (d) our model \( \{u^0 = u_1 \} \); (from left to right) \( \lambda = 200, 1200, 3000 \).](image)

![Figure 8. Evaluation of edge localization on the BSDS300 [23] (top), and examples of the gradient magnitude averaged over the scale-space (bottom). Given (a) input image, the scale-space is constructed by (b) WLS [7] \( \{u^0 = 1\} \), (c) RGF [35] \( \{u^0 = 1\} \), and (d) our model \( \{u^0 = 1\} \), by varying scale parameters, i.e., \( \lambda \leq 2 \times 10^5 \) in WLS [7], \( \sigma_s \leq 50 \) in RGF [35], and \( \lambda \leq 2 \times 10^5 \) in our model.](image)
tant structures to be preserved, e.g., corners.

### 4.3. Other Applications

Our model can be applied to joint image restoration tasks. We have applied it to RGB/NIR and flash/non-flash denoising problems as shown in Figs. 10 and 11. In RGB/NIR denoising, color image $f$ is regularized with the flash NIR image $g$. Similarly, the non-flash image $f$ is regularized with flash image $g$. Since there exist structural dissimilarities between static guidance and input images ($g$ and $f$), the results might have artifacts and unnatural appearance. For example, static guidance regularization such as GF [10] cannot deal with a gradient reversal in flash NIR images [33], resulting in smoothed edges. Our model handles the structural differences between images, and shows performance comparable to the state of the art [33].

### 5. Discussion

We have presented a joint filtering framework that is widely applicable to computer vision and computational photography. Contrary to static guidance methods, we leverage dynamic guidance images as well, and they can exploit the structural information of the input image. Although our model does not have a closed-form solution, it converges rapidly to a local minimum. The simple and flexible formulation of our framework makes it applicable to a great variety of applications.

Acknowledgments. The authors would like to thank Francis Bach for helpful discussions. This work was supported in part by the ERC grant VideoWorld and the Institut Universitaire de France.
References


