Fusion of Median and Bilateral Filtering for Range Image Upsampling

Qingxiong Yang, *Member, IEEE,* Narendra Ahuja, *Fellow, IEEE,* Ruigang Yang, *Member, IEEE,*
Kar-Han Tan, *Senior Member, IEEE,* James Davis, *Member, IEEE,* Bruce Culbertson, *Member, IEEE,*
John Apostolopoulos, *Fellow, IEEE,* Gang Wang, *Member, IEEE,*

Abstract— **We present a new upsampling method to enhance the spatial resolution of depth images. Given a low-resolution depth image from an active depth sensor and a potentially high-resolution color image from a passive RGB camera, we formulate it as an adaptive cost aggregation problem and solve it using the bilateral filter. The formulation synergistically combines the median filter and the bilateral filter; thus it better preserves the depth edges and is more robust to noise. Numerical and visual evaluations on a total of 37 Middlebury data sets demonstrate the effectiveness of our method. A real-time high-resolution depth capturing system is also developed using commercial active depth sensor based on the proposed upsampling method.**

Index Terms— **Sensor fusion, bilateral filter, weighted median filter, GPU.**

I. INTRODUCTION

Depth sensing in dynamic real-world environment is a challenging problem in computer vision. Depth from passive stereo is well-studied and is known to be reliable only when sufficient visual information is available for establishing correspondence across multiple cameras. Active depth sensors like laser range scanners can provide extremely accurate and dense 3D measurement over a large working volume but are usually very expensive and can only measure a single point at a time. Alternatively, depth sensing systems based on the time-of-flight (ToF) principle can capture the whole scene at the same time, but they are either very expensive or very limited in terms of resolution.

We presented in [54] a framework to enhance the spatial resolution of depth images (e.g., those from the Canesta sensor [2]). This approach takes advantage of the fact that a registered high-quality texture image can provide significant information to enhance the raw depth image. The depth upsampling problem is formulated in an adaptive cost aggregation framework. A *cost volume*[1] measuring the distance between the possible depth candidates and the depth values bilinearly upsampled from those captured by the active depth sensor is computed. The joint bilateral filter is then applied to the cost volume to enforce the consistence of the cost values and the color values from the registered RGB image. The high-resolution depth image is produced by taking the winner-takes-all approach on the filtered cost volume and a sub-pixel refinement afterward.

In this paper, we analyze the parameter sensitivity of this approach and the relationship between the proposed framework and the weighted median filter [55]. Actually, if the joint bilateral filter is replaced with a box filter, the proposed framework is equivalent to a median filter. Using the joint bilateral filter, the proposed framework synergistically combines the median filter and the bilateral filter so that it can better preserve the depth edges and is more robust to depth sensor noise. When the cost values are simply the absolute differences between the possible depth candidates and the captured depth values, the proposed framework is equivalent to a weighted median filter.

However, the computational complexity of this method will be linear in the depth search range. Let the number of depth hypotheses be $\mathcal{L}$ (corresponding to the depth search range), the proposed method requires $\mathcal{L}$ joint bilateral filtering.

We presented in [51] an approach to reduce the computation complexity by hierarchically upsampling the depth image. At each scale, a cost volume with up to four cost slices corresponding to four depth hypotheses will be built based on the current depth image with around 75% mis-sampled pixels. A joint bilateral filter is then applied to each slice independently to solve the depth ambiguities of the mis-sampled pixels via depth propagation. Finally, the depth hypothesis with the lowest cost at each pixel is selected as correct, and fed to the next scale for further upsampling. In this case, the number of depth hypotheses used for voting be $\mathcal{L} \leq 4$. This approach, however, is valid only for fronto-parallel surfaces.

---

[1] There is an error in Eq. (1) in [54]. Similar to Eq. (1) in [51], the quadratic cost function $(d - D_{(i)}(y, x))^2$ should be changed to a linear cost function $|d - D_{(i)}(y, x)|$.

assumption of locally planar surfaces. Instead of using only the four depth values of the neighbors, we include the mean of the four depth values for consideration and thus $\mathcal{L} \leq 5$. This simple extension of the hierarchical upsampling method in [51] turns out to be a good solution for non-fronto-parallel scenes. Using Middlebury data sets [5], our numerical evaluations show that the new hierarchical approach proposed in this paper not only reduces the computational complexity of [54] but also improves the upsampling accuracy. Our GPU implementation via CUDA shows that the proposed hierarchical upsampling method can enhance the spatial resolution of depth images obtained from Canesta sensors [2] from 160×120 to VGA size at a rate of 1200 Hertz on an Nvidia Geforce GTX 580 GPU. The CUDA implementation of the bilateral filter [3] provided by Nvidia is adopted.

II. RELATED WORK

Passive Stereo Matching. Passive stereo matching is a fundamental task in computer vision [42]. In the past decade, much of the community's effort has been focused on the specific problem of *disparity optimization*, producing a number of excellent optimization methods that have significantly advanced the state of the art. The key objective of these optimization methods is to reduce the matching ambiguities introduced by low-texture regions, and can be generally separated into two main categories: global methods and local methods. In low-texture regions, the lack of visual features makes matching a challenging problem. Global methods based on belief propagation [21], [46] or graph cuts [13] are formulated in an energy-minimization framework, where the objective is to find a disparity solution that minimizes a global energy function; global methods thus outperform local methods around low-texture regions. The computational complexity of these global methods is high, but a number of efficient approximations like hierarchical belief propagation [20] and constant space belief propagation [52] algorithms are becoming available.

Active Depth Sensors. Laser range scanners are of high accuracy and can provide high-resolution 3D measurement [9], [10], [24], [38], [41], [44]. However, they are usually very expensive and can only measure a single point at a time; thus is not suitable for dynamic environments. Active depth sensing systems based on the time-of-flight (ToF) principle [1], [30], [2], [6], [4] become popular recently because of their low cost property and real-time performance. Nevertheless, these sensors are of low-resolution thus have limited applications. Kolb *et al.* [26] gives an account of recent developments in ToF-technology and discusses the current state of the integration of this technology into various graphics-related applications.

Range Image Upsampling. Using a low-resolution depth image and a registered high-resolution camera image, Diebel and Thrun [16] design a Markov Random Field (MRF) and solve it using conjugate gradient (CG) algorithm [39]. This method gives promising spatial resolution enhancement up to $10\times$. Kopf *et al.*[27] propose to use joint bilateral filter for range image upsampling. A joint bilateral upsampling operation that can produce full resolution results from solutions

computed at low resolutions was proposed in this paper and was used in [28] for obtaining high-resolution depth images in a hybrid camera. A number of extensions of this upsampling method [27] have been proposed. Riemens *et al.* [40] choose to use the joint bilateral filter hierarchically to reduce the computational cost and Huhle *et al.* [23] replace the joint bilateral filter in [27] with a modified nonlocal means filter [15]. However, these joint upsampling methods are directly applied to the depth images based on the color information captured in the corresponding RGB image; thus they are not suitable for textured scenes (in the color image) as both the bilateral filter and the nonlocal means filter are known to be weak for texture suppression [19], [45], [48]. Direct depth filtering using information from the RGB image (via joint bilateral filter or nonlocal means filter) imposes the risk of transferring texture from the color image into the upsampled depth image.

Bilateral Filter. The bilateral filter is a robust edge-preserving filter introduced by Tomasi and Manduchi [47]. It has been used in many computer vision and computer graphics tasks, and a general overview of the applications can be found in [36]. A bilateral filter has two filter kernels: a spatial filter kernel and a range filter kernel for measuring the spatial and range distance between the center pixel and its neighbors, respectively. The two filter kernels are traditionally based on a Gaussian distribution [35], [17]. Specifically, let $I_{\mathbf{x}}$ be the color at pixel $\mathbf{x}$ and $I_{\mathbf{x}}^I$ be the filtered value, we want $I_{\mathbf{x}}^I$ to be

$$I_{\mathbf{x}}^I = \frac{\sum_{\mathbf{y} \in N(\mathbf{x})} f_S(\mathbf{x}, \mathbf{y}) f_R(I_{\mathbf{x}}, I_{\mathbf{y}}) I_{\mathbf{y}}}{\sum_{\mathbf{y} \in N(\mathbf{x})} f_S(\mathbf{x}, \mathbf{y}) f_R(I_{\mathbf{x}}, I_{\mathbf{y}})}, \tag{1}$$

where $\mathbf{y}$ is a pixel in the neighborhood $N(\mathbf{x})$ of pixel $\mathbf{x}$, and

$$f_S(\mathbf{x}, \mathbf{y}) = exp(-\frac{||\mathbf{x} - \mathbf{y}||^2}{2\sigma_S^2}) \tag{2}$$

and

$$f_R(I_{\mathbf{x}}, I_{\mathbf{y}}) = exp(-\frac{||I_{\mathbf{x}} - I_{\mathbf{y}}||^2}{2\sigma_R^2}) \tag{3}$$

are the spatial and range filter kernels measuring the spatial and range/color similarities. The parameter σ_S defines the size of the spatial neighborhood used to filter a pixel, and σ_R controls how much an adjacent pixel is down-weighted because of the color difference. A joint (or cross) bilateral filter [37], [18] is the same as the bilateral filter except that its range filter kernel $f_R(\cdot)$ is computed from another image named *guidance image*. Let J denote the guidance image, the joint bilateral filtered value at pixel $\mathbf{x}$ is

$$I_{\mathbf{x}}^J = \frac{\sum_{\mathbf{y} \in N(\mathbf{x})} f_S(\mathbf{x}, \mathbf{y}) f_R(J_{\mathbf{x}}, J_{\mathbf{y}}) I_{\mathbf{y}}}{\sum_{\mathbf{y} \in N(\mathbf{x})} f_S(\mathbf{x}, \mathbf{y}) f_R(J_{\mathbf{x}}, J_{\mathbf{y}})}. \tag{4}$$

Note that the joint bilateral filter ensures the texture of the filtered image I^J to follow the texture of the guidance image J (see Fig. 4 (a)-(b)). In this paper, we normalized the image intensity such that it ranges from $[0, 1]$. We also use normalized image coordinates so that $\mathbf{x}$ and $\mathbf{y}$ also reside in $[0, 1]$.

Weighted Median Filter. The median filter [29] was originally introduced for reducing image noise while preserving edges. It can effectively eliminate impulsive noise and bi-exponential noise. However, it may cause edge jitter [12],

streaking [11] and remove image details [34]. The weighted median filter [14] is a more general filter. An overview of the weighted median filter is given in [55]. A nonnegative weight is assigned to each position in the filter window. The weighted median filter then sorts the samples inside the filter window, duplicates each sample to the number of the corresponding weight, and chooses the median value from the new sequence. The median filter is thus a special case of the weighted median filter with the weight being one. One class of weighted median filters like the topological median filter [43] has weights being only zero or one. This class of weighted median filters aims at excluding the contribution from disconnected features in the neighborhood of the center pixel. Other weighted median filters have less constraints on the exact weight value but its distribution. For instance, center weighted median filters [25] give more weight to the central value of each window. The original weighted median filter was limited to have "low-pass" type filtering characteristics due to the nonnegative weight constraint. To obtain "band-pass" or "high-pass" frequency filtering characteristics, this constraint is relaxed to real-value weights and named permutation weighted median filters in [7]. The fuzzy weighted median filter [33] is an extension of permutation weighted median filter with the fuzzy transformation that incorporates sample diversity into the ordering operation. Advantages have been observed in applications of DCT coded image deblocking, impulse removal, and noisy image sharpening. Polynomial weighted median filter [8] also belongs to permutation weighted median filter. Polynomial models are capable of approximating a large class of nonlinear systems with a finite number of coefficients. The use of polynomial weighting function in a weighted median filter can effectively exploit the higher order statistics of observed samples while producing filter outputs that are robust to outliers in the observation set. Median filters with other weighting functions were also proposed for different specific applications. For instance, Gaussian weighting function was used for image super-resolution [32]. In this paper, the Gaussian weighting function is also used. The main difference is that besides the spatial distance, the color similarity is also taken into account. The obtained median filter is named **bilateral weighted median filter** in the same manner as the bilateral filter [47] and the other weighted median filters [32], [8].

III. MOTIVATION

The joint bilateral filter has been demonstrated to be very effective for color image upsampling [27]. But if it is directly applied to a depth image with a registered RGB color image as the guidance image, the texture of the guidance image (that is independent of the depth information) is likely to be introduced to the upsampled depth image as shown in Fig. 1. The upsampling errors mainly reside in the texture transferring property of the joint bilateral filter as shown in Fig. 4 (a)-(b). This property also limits its application for the color-depth misalignment problem presented in Fig. 5 (a)-(d).

Meanwhile, the median filtering operation minimizes the sum of the absolute error of the given data [22], and is much more robust to outliers than the bilateral filter. We thus
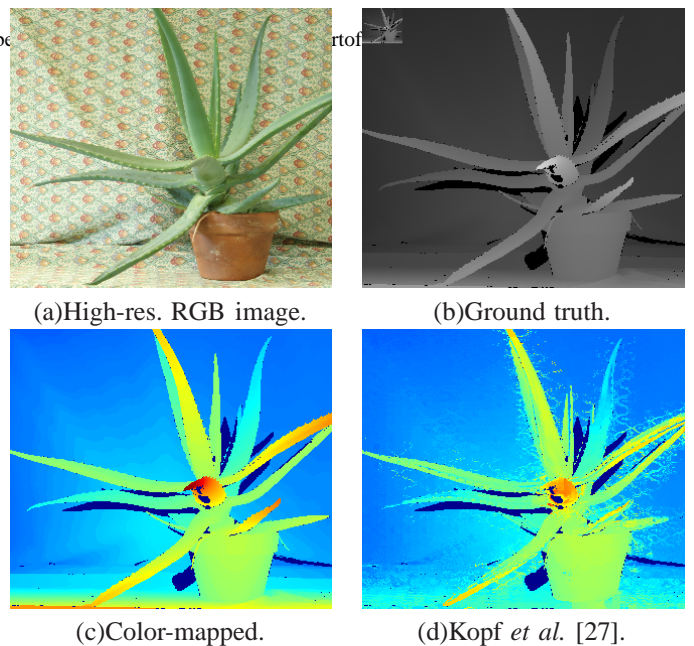

(a)High-res. RGB image. (b)Ground truth.

(c)Color-mapped. (d)Kopf *et al.* [27].

Fig. 1. A failure case of [27]. (a) is the registered high-resolution RGB image (320×278), and (b) is the ground-truth high-resolution disparity (320×278) with the input low-resolution disparity map (40×35) showing on the upper left and the corresponding colored visualized display showing in (c). (d) is the upsampled high-resolution disparity map obtained from [27]. As can be seen from (d), this method is not suitable for textured scenes due to the texture transferring property of the joint bilateral filter demonstrating in Fig. 4 (a)-(b).

focusing on the combination of the median operation with the bilateral filter so that the texture influence can be better suppressed while maintaining the edge-preserving property.

Median filter is actually a special case of the weighted median filter which is a more general filter. Weighted median filter finds the value minimizing the sum of weighted absolute error of the given data [55]:

$$\arg\min_{b} \sum_{\mathbf{y} \in N(\mathbf{x})} W(\mathbf{x}, \mathbf{y})|b - I_{\mathbf{y}}|, \qquad (5)$$

where $W(\mathbf{x}, \mathbf{y})$ corresponds to the weight assigned to a pixel $\mathbf{y}$ inside a local region centered at pixel $\mathbf{x}$. The weight is normally designed to suppress noise while maintaining signal structures such as edges and lines. The use of constant weight (*e.g.*, $W(\mathbf{x}, \mathbf{y}) = 1$) corresponds to the median filter. To preserve edges, the weights are normally designed based on existing training signals or the requirement to preserve certain image details which are supposed to be known by the designers. For instance, to remove the anisotropic artifacts of median filter as shown in Fig. 2 (b), isotropic weights can be used. Filtered result of Fig. 2 (a) using Gaussian weight is presented in Fig. 2 (c), where $W(\mathbf{x}, \mathbf{y}) = f_S(\mathbf{x}, \mathbf{y})$ is a Gaussian function defined in Eq. (2).

A limitation of the median filter is that it will round existing corners as can be seen in Fig. 3 (b). Gaussian weighted median filter cannot solved this problem as shown in Fig. 3 (c). Recall that the bilateral filter can be treated as an extension of the Gaussian filter so that the weighting for each pixel is a Gaussian function of both the spatial distance from the center pixel as well as the relative difference in

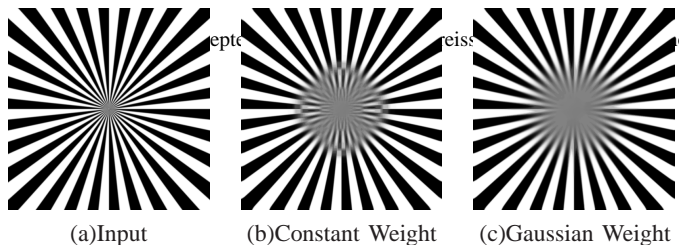(a)Input   (b)Constant Weight   (c)Gaussian Weight

Fig. 2. Visual comparison of two different weighted median filters. From (a)-(c): input image, traditional (constant weighted) median filtered image and Gaussian weighted median filtered image. Note that the anisotropic artifacts in (b) are removed from (c).

intensity/color, a bilateral weighted median filter can also be defined as an extension of Gaussian weighted median filter whose weight not only depends on the spatial distance but also the intensity/color difference. According to Eq. (1)-(3) and (5), the bilateral weighted median filter corresponds to the following minimization problem:

$$\arg\min_b \sum_{\mathbf{y} \in N(\mathbf{x})} f_S(\mathbf{x}, \mathbf{y}) f_R(I_{\mathbf{x}}, I_{\mathbf{y}})|b - I_{\mathbf{y}}|. \qquad (6)$$

The bilateral weighted median filtered image of Fig. 3 (a) is presented in Fig. 3 (d). Visual comparison to the other weighted median filtering results shows that it can better preserve image structures. The PSNR values computed against the noise-free input image are presented under Fig. 3 (a)-(d). Quantitative evaluation using PSNR metric demonstrates its robustness against noise.

(a)Input   (b)Constant   (c)Gaussian   (d)Bilateral
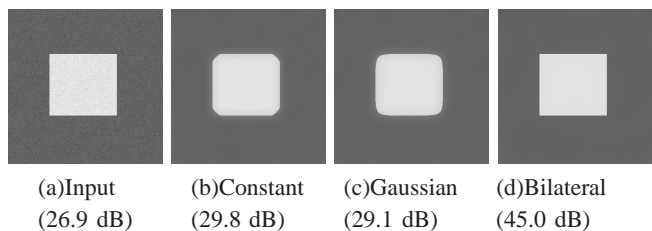(26.9 dB)   (29.8 dB)   (29.1 dB)   (45.0 dB)

Fig. 3. Image denoising using weighted median filtering. From (a)-(d): noisy input image, median filtered image, Gaussian weighted median filtered image and bilateral weighted median filtered image. The numbers under (a)-(d) are PSNR values computed against the noise-free input image (which is omitted here). Note that the bilateral weighted median filter outperforms the others both qualitatively and quantitatively.

Same as the definition of the joint bilateral filtering, to enforce the similarity between the filtered image and a guidance image J, the range filter kernel $f_R(\cdot)$ in Eq. (6) can be computed based on J instead of I (which is the image to be filtered):

$$\arg\min_b \sum_{\mathbf{y} \in N(\mathbf{x})} f_S(\mathbf{x}, \mathbf{y}) f_R(J_{\mathbf{x}}, J_{\mathbf{y}})|b - I_{\mathbf{y}}|, \qquad (7)$$

and the resulted filter is named joint bilateral weighted median filter (**JBM**) in this paper. This type of weighted median filter has been demonstrated to be very effective for depth/disparity denoising [31] with the assumption of the availability of a registered RGB image. Fig. 4 presents an example which uses the RGB image presented in Fig. 1 (a) as the guidance to

filter the depth image presented in Fig. 1 (b). As demonstrated in Fig. 4 (c)-(d), joint bilateral weighted median filter alleviates the texture transferring problem (of the joint bilateral filter) presented in 4 (a)-(b) while maintaining edge-preserving property. It is also more robust to the misalignment problem presented in Fig. 5 (a)-(b). Note that the depth bleeding artifacts in Fig. 5 (c)-(d) are removed from Fig. 5 (e)-(f).

(a)JBF($\sigma_R = 0.05$).   (b)JBF($\sigma_R = 0.10$).

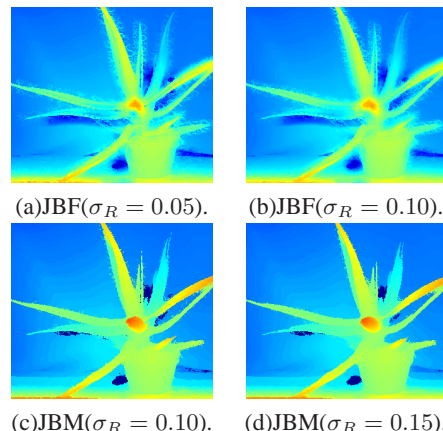(c)JBM($\sigma_R = 0.10$).   (d)JBM($\sigma_R = 0.15$).

Fig. 4. Texture transferring. (a) and (b) are the joint bilateral filtering ($\sigma_S = 0.03$) results of the disparity map in Fig. 1 (b) with the RGB image in Fig. 1 (a) as the guidance image. Note that the texture information in the RGB image will be transferred to the disparity map. (c) and (d) are the joint bilateral weighted median filtered results with respect to two different range filter kernels. Note that the bilateral weighted median filter can preserve the depth edges without picking up the texture of the RGB guidance image.

(a)RGB image.   (b)Disparity image.   (c)JBF($\sigma_R = 0.05$)

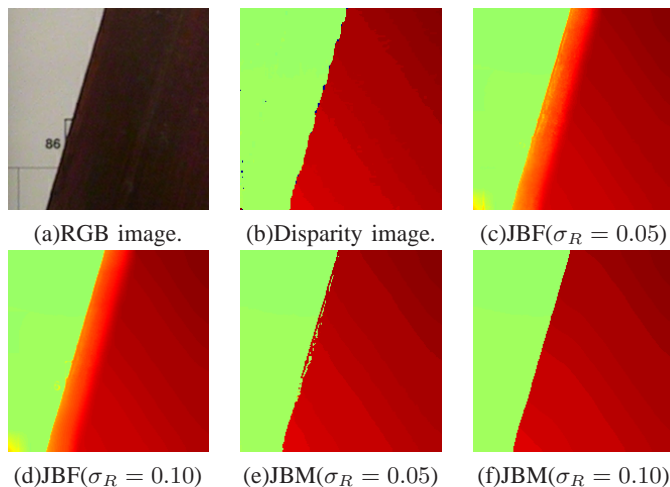(d)JBF($\sigma_R = 0.10$)   (e)JBM($\sigma_R = 0.05$)   (f)JBM($\sigma_R = 0.10$)

Fig. 5. Misalignment problem. (a) is an input RGB image and (b) is the corresponding color-mapped depth image. However, the alignment is incorrect thus the color edges in (a) are not correctly aligned with the depth edges in (b). The incorrect alignment results in noticeable depth bleeding artifacts after joint bilateral filtering as shown in (c) and (d) even when large filter kernel ($\sigma_S = 0.1$) is used. Joint bilateral weighted median filter can potentially solve this problem with the input RGB image in (a) as the guidance image as demonstrated in (e)-(f).

IV. APPROACH

In this section, we formulate our depth upsampling problem in an adaptive cost aggregation framework and solve it using bilateral filtering in Sec. IV-A. The formulation is then proved
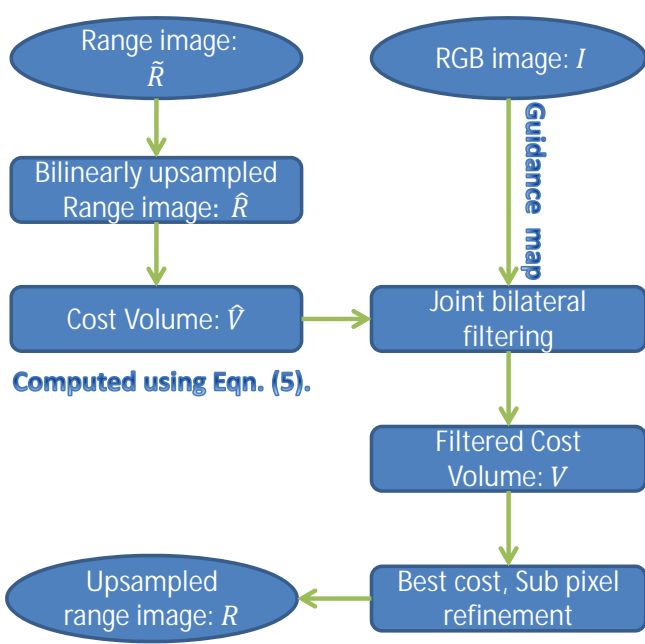
Fig. 6. Framework of our upsampling method. The low-resolution depth image $\tilde{R}$ is bilinearly upsampled to the same size as the high-resolution RGB image I, and serves as the initial depth map hypothesis $\hat{R}$. A cost volume $\hat{\mathcal{V}}$ is built according to the current depth hypotheses $\hat{R}$. Joint bilateral filtering with the high-resolution RGB image I as the guidance map is then applied to the cost volume $\hat{\mathcal{V}}$ to enforce the consistency of the depth edges and color edges. A winner-take-all and sub-pixel estimation procedure is used to produce a new depth map hypothesis based on the filtered cost volume $\mathcal{V}$.

to be a combination of the median filter and the bilateral filter in Sec. IV-B. To further improve the accuracy and reduce the computational complexity, a hierarchical upsampling method is proposed in Sec. IV-C. In this paper, we only consider the field of view shared by the active depth sensor and the passive RGB camera. We assume that the low-resolution depth image captured from the active depth sensor is registered with the high-resolution RGB image, and only the overlapped region will be processed.

A. Upsampling Using An Adaptive Cost Aggregation Framework

An overview of our depth image upsampling framework is provided in Figure 6. First, we upsample the low-resolution depth image $\tilde{R}$ to the same size as the high-resolution RGB image I. Let it be $\hat{R}$. A cost volume $\hat{\mathcal{V}}$ is then built based on $\hat{R}$. To allow large depth variations (as the current depth hypotheses are not guaranteed to be correct), the cost function should become constant when the differences become large. One such common function is the truncated linear model, where the cost increases linearly in the distance between the potential depth candidate d and the upsampled depth value $\hat{R}_{\mathbf{x}}$ at pixel $\mathbf{x}$

$$\hat{\mathcal{V}}_{\mathbf{x}}(d) = \min(\eta\mathcal{L}, |d - \hat{R}_{\mathbf{x}}|), \qquad (8)$$

where $\mathcal{L}$ is the search range and η is a constant. η is set to 0.1 in all experiments conducted in Sec. V.

Joint bilateral filtering is applied to each slice ($\hat{\mathcal{V}}(d)$) of the cost volume $\hat{\mathcal{V}}$ to enforce the consistency of the depth

edges and color edges as shown in Figure 6. The new depth image is then computed from the filtered cost volume $\mathcal{V}$ by first selecting the depth hypothesis with the minimal cost and then a sub-pixel estimation [53] afterwards. In this framework, the upsampling problem is actually formulated as an adaptive cost aggregation problem where the support-weights are computed using a bilateral filter kernel:

$$
\begin{aligned}
R_{\mathbf{x}} &= \arg\min_d \mathcal{V}_{\mathbf{x}}(d) \\
&= \arg\min_d \frac{\sum_{\mathbf{y}\in N(\mathbf{x})} f_S(\mathbf{x},\mathbf{y}) f_R(I_{\mathbf{x}},I_{\mathbf{y}}) \hat{\mathcal{V}}_{\mathbf{y}}(d)}{\sum_{\mathbf{y}\in N(\mathbf{x})} f_S(\mathbf{x},\mathbf{y}) f_R(I_{\mathbf{x}},I_{\mathbf{y}})} \\
&= \arg\min_d \sum_{\mathbf{y}\in N(\mathbf{x})} f_S(\mathbf{x},\mathbf{y}) f_R(I_{\mathbf{x}},I_{\mathbf{y}}) \hat{\mathcal{V}}_{\mathbf{y}}(d). \qquad (9)
\end{aligned}
$$

The spatial filter kernel $f_S(\cdot)$ and the range filter kernel $f_R(\cdot)$ are Gaussian distributions defined in Eq. (2) and (3).

(a) $\hat{R}$ (Input). (b) $\hat{R}$ (color-mapped). (c) $\eta = 1$ ($\sigma_R = 0.1$).

(d) $\eta = 0.1$ ($\sigma_R = 0.1$). (e) $\eta = 1$ ($\sigma_R = 0.2$). (f) $\eta = 0.1$ ($\sigma_R = 0.2$).
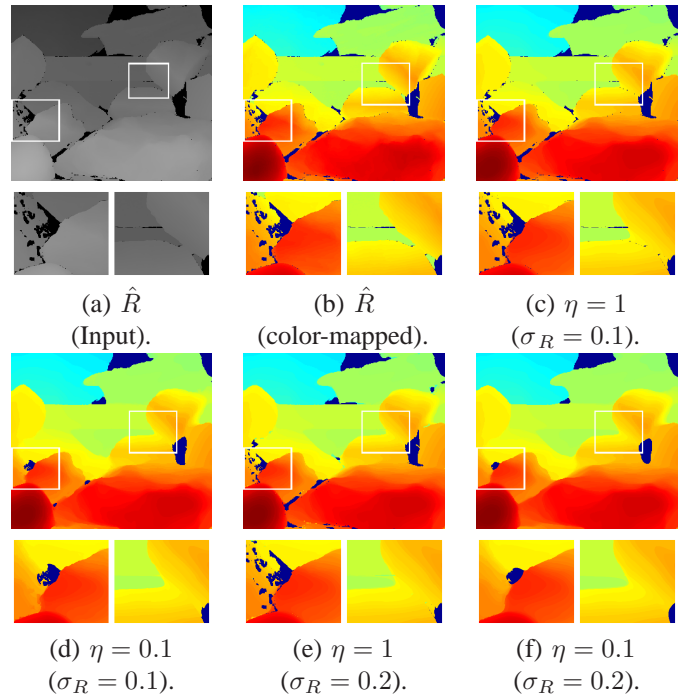
Fig. 7. Bilateral weighted median filter. (a) is an input depth image and the corresponding colored visualized display is shown in (b). (c) and (e) are obtained from the bilateral weighted median filter (equivalent to the proposed fusion framework when $\eta = 1$) and (d) and (f) are obtained from the proposed fusion framework when $\eta = 0.1$. σ_S is set to 0.03 in this experiment. Note that the tiny dark regions in (c) and (e) are removed from (d) and (f). In this example, the guidance image used in the proposed framework is the input image (a) itself. All the other experiments use an additional registered RGB image as the guidance image.

B. Bilateral Weighted Median Filter

According to Eq. (9), the main operation used in the proposed framework in Sec. IV-A is the joint bilateral filter. However, when $\eta = 1$ and $f_S(\mathbf{x},\mathbf{y}) = f_R(I_{\mathbf{x}},I_{\mathbf{y}}) = 1$, Eq. (9) becomes

$$R_{\mathbf{x}} = \arg\min_d \sum_{\mathbf{y}\in N(\mathbf{x})} |d - \hat{R}_{\mathbf{y}}|. \qquad (10)$$

Because the median minimizes the sum of absolute error of the given data [22], the above equation actually computes the median values of image $\hat{R}$. As a result, Eq. (9) is a combination of the median operation with the joint bilateral filter when the bilateral filter kernel $f_S(\mathbf{x}, \mathbf{y}) f_R(I_\mathbf{x}, I_\mathbf{y})$ is not a constant value.

Eq. (9) indeed corresponds to a bilateral weighted median filter when $\eta = 1$. The computational complexity of the brute-force implementation of the bilateral weighted median filter is linear in the filter kernel size. However, Eq. (9) allows the bilateral weighted median filter to be computed independent of the filter kernel size when constant time bilateral filters [50], [49] are employed.

In practice, we use $\eta < 1$ to further reject potential outliers as shown in Fig. 7. Note that the tiny dark regions in the input image in Fig. 7 (b) are preserved in (c) and (e) when $\eta = 1$. However, these regions are very likely to be noise when presented in a depth image, thus it is reasonable to use a small η to reduce the contribution from depth noise. Fig. 7 (d) and (f) present the filtered images obtained from the proposed filtering framework in Eq. (9) when $\eta = 0.1$. As can be seen, most of the tiny dark regions in the input image are removed. In this example, the guidance image used in the proposed framework is the input image itself. All the other experiments use an additional registered RGB image as the guidance image.

C. Hierarchical Upsampling

By using the color information provided in the registered RGB images, the method presented in Section IV-A is able to maintain correct depth edges after upsampling, which is the main contribution of the paper. However, the computational cost may be high. The main computation resides in the use of joint bilateral filtering. Let the number of depth hypotheses used to create the cost volume $\hat{\mathcal{V}}$ be $\mathcal{L}$ as defined in Eq. (8), $\mathcal{L}$ joint bilateral filtering processes will be required.

In this section, we present a hierarchical approach to reduce the computational complexity. We start from the lowest resolution (the resolution of the original depth image $\check{R}$), and hierarchically estimate the depth values for the mis-sampled pixels. An example is shown in Figure IV-B (a) with the assumption that the original low-resolution depth image $\check{R}$ has only 9 pixels (blue squares in Figure IV-B (a)). We increase the resolution of $\check{R}$ and show the new depth image in Figure IV-B (b). Let $\check{R}^{(0)}$ denote this new depth image. Meanwhile, the high-resolution RGB image I will be downsampled to a new RGB image $I^{(0)}$ so that $I^{(0)}$ has the same spatial resolution as $\check{R}^{(0)}$. We do not apply any low-pass filter to the RGB image prior to downsampling. Otherwise, the color edges are likely to be blurred and the depth edges will not be preserved in the upsampled depth image. The white squares in $\check{R}^{(0)}$ in Figure IV-B (b) are mis-sampled pixels (without depth values) and blue squares are pixels with depth values captured from the range sensor. We next propagate the depth values from the blue squares to the white squares sequentially. For every white pixel $\mathbf{x}$ which has four blue neighbors (as shown in Figure IV-B (c)), we estimate its depth value by selecting one

from the four depth values of its four blue neighbors based on the framework presented in Section IV-A. Specifically, let $\vec{d}_\mathbf{x}$ denote a vector comprising of the four depth values of the four blue neighbors and $\mathcal{B}$ denote the collection of all pixels with depth values (blue squares in Figure IV-B), the depth value of the white square is computed as follows:

$$R_\mathbf{x}^{(0)} = \arg \min_{d \in \vec{d}_\mathbf{x}} \sum_{\mathbf{y} \in N(\mathbf{x})} \lambda(\mathbf{y}) f_S(\mathbf{x}, \mathbf{y})$$
$$f_R(I_\mathbf{x}^{(0)}, I_\mathbf{y}^{(0)}) \min \left(\eta \mathcal{L}, |d - \check{R}_\mathbf{y}^{(0)}| \right), \qquad (11)$$

where $\mathcal{L}$ and η are defined in Eq. (8), and

$$\lambda(\mathbf{y}) = \begin{cases} 1 & \text{if } \mathbf{y} \in \mathcal{B}, \\ 0 & \text{else.} \end{cases}$$

The depth values of the white squares will be estimated using Eq. (11), which requires four joint bilateral filtering processes and a winner-take-all procedure. We show the depth estimation result in Figure IV-B (d). Pixels with depth values propagated from blue pixels are presented as green squares. This is the end of the first depth propagation.

We next change all the green squares to blue, and start a new depth propagation process as shown in Figure IV-B (e)-(f). Finally, only a part of the boundary pixels remain to be white as shown in Figure IV-B (g). We thus start final propagation using Eq. (11). The boundary pixels has up to three blue neighbors, so the length of $\vec{d}$ is less or equal to three. The final upsampled depth image $R^{(0)}$ ($i = 0$) is presented in Figure IV-B (h). We next go up to a higher resolution as shown in Figure IV-B (i), and estimate the depth values for the new white squares (mis-sampled pixels) until the RGB image I's spatial resolution is reached. If the resolution of the high-resolution RGB image is different from the resolution of the final upsampled depth image, it may result in either a whole line of mis-sampled pixels on the right or/and on the bottom at some scale. In this case, we simply add an extra round of depth propagation to estimate the depth values of these mis-sampled pixels.

Confining the depth hypothesis to only the depth values of four neighbors with sampled depth values is valid for fronto-parallel surfaces. However, when this assumption is violated, the performance can be greatly deteriorated. For instance, a floor surface perpendicular to the image plane. A synthetic non-fronto-parallel scene is presented in Figure 9. The image resolution is enhanced $64\times$ in this example. The upsampling result using only four depth hypotheses is presented in Figure 9 (d). Obviously, it results in quantization errors as the correct depth value of a white square may not be presented in the depth values of its four blur neighbors.

Relaxing the fronto-parallel surfaces assumption to only planar surfaces, the quantization errors can be greatly reduced by adding all possible depth values. However, this will greatly increase the computational complexity. Nevertheless, our experiment shows that adding as few as one depth hypothesis at each mis-sampled pixels is usually enough. Specifically, instead of using only the four depth values of the neighbors, we include the mean of the four depth values for consideration.
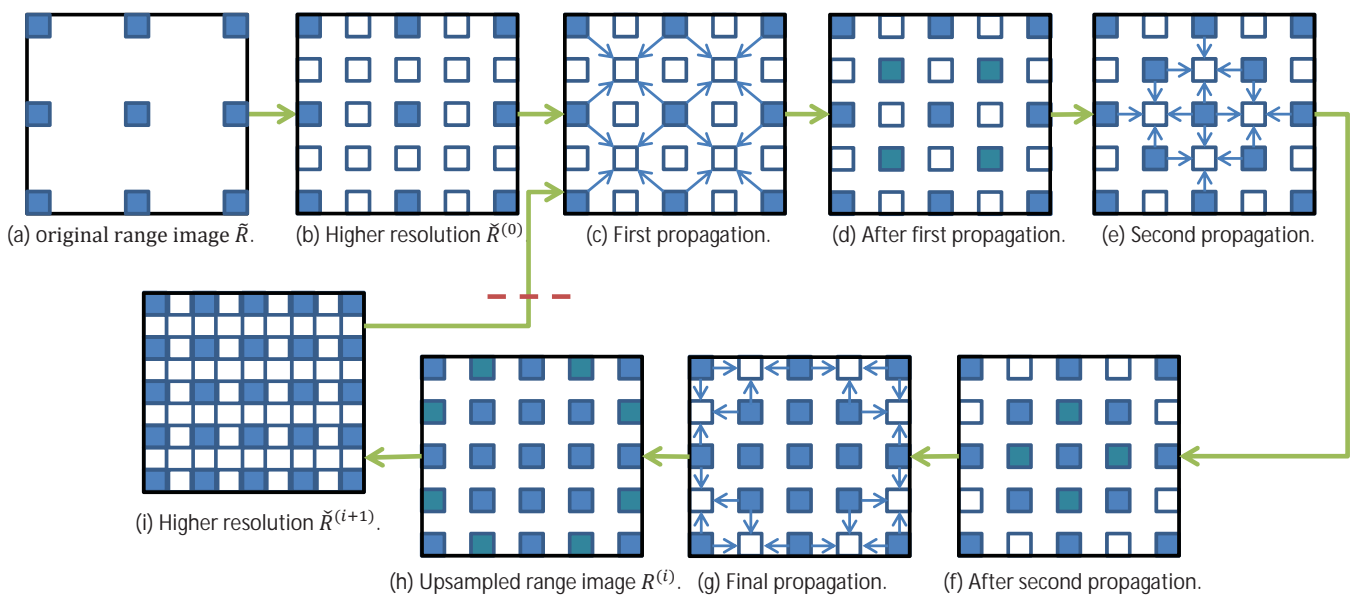
Fig. 8. An overview of our hierarchial upsampling approach. At each scale $i \geq 0$, we increase the spatial resolution of the current depth image (when $i = 0$, see (a)), such that only about 25% of the new depth image $\check{R}^{(i)}$ contains depth estimates (blue squares), and the rest without depth estimates (white squares) as shown in (b). We then propagate the depth values from the blue squares to white squares sequentially in three steps as shown from (c)-(g). (c)-(d) depict the first depth propagation step. We first locate white squares with four blue neighbors. The depth value of each white square are then selected from the depth values of its blue neighbors based on the framework presented in Section IV-A. We next change the white squares with estimated depth values to blue. The second step shown in (e)-(f) is the same as the first step: propagate depth values from blue squares to white squares. In the final step, we estimate the depth values for the white boundary pixels as shown in (g). (h) shows the result after depth propagation: $R^{(i)}$. We then go to the next scale and do the depth propagation again until the spatial resolution of the RGB image I is reached.
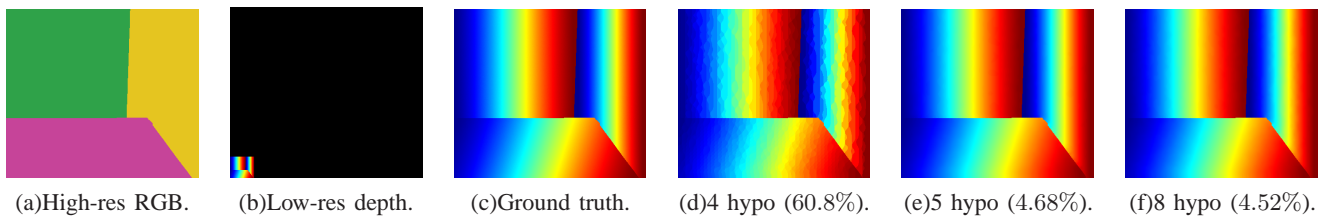

Fig. 9. A synthetic non-fronto-parallel scene. (a) is the original high-resolution RGB image (432×384); (b) to (f) are color-mapped depth images: (b) low-resolution depth image (54×48); (c) ground-truth high-resolution depth image (432×384); (d) high-resolution depth image obtained from our method with 4 depth hypotheses; (e) high-resolution depth image obtained from our method with 5 depth hypotheses; (f) high-resolution depth image obtained from our method with 8 depth hypotheses. The spatial resolution is enhanced $64\times$. The percentage of bad pixels in (d)-(f) are 60.8%, 4.68% and 4.52%, respectively.

The upsampled depth image using these five depth hypotheses is presented in Figure 9 (e). The percentage of bad pixels [2] in the upsampled depth image is 4.68%. If we add another three depth hypotheses, the upsampling accuracy does not improve much as shown in Figure 9 (f). The percentage of bad pixels in the upsampled depth image in Figure 9 (f) is 4.52%. Hence, in all the experiments conducted, we set the number of depth hypothesis to be five.

V. EXPERIMENTS

In this section, we present experiments on real images to demonstrate the effectiveness and robustness of our method. Sec. V-A numerically evaluate the performance of our method using Middlebury data sets. These data sets are obtained using structured light and there may be holes remaining in areas where structured light fails (due to the lack of illumination

[2]If the disparity error of a pixel is larger than 1, it is treated as a bad pixel.

codes to begin with). In this paper, pixels inside the holes are treated as outliers and are left unchanged.

Our experiments demonstrate that the proposed upsampling method is robust to parameter σ_R and σ_S (in Eq. (2) and (3), respectively), and that the proposed upsampling methods outperform existing depth upsampling methods [27], [23], [40] on average. For simplicity, in the rest of the paper, we will refer to our upsampling method in Section IV-A as **JBMU**, and the hierarchical method in Section IV-C as **HJBMU**. Section V-B presents visual evaluation on real indoor images. A high-resolution depth sensing system using Canesta sensor [2] is developed based on our HJBMU method. This system can upsample a 160×120 depth image to 640×480 at a rate of 1200 Hertz on an Nvidia Geforce GTX 580 GPU with very few visible errors.
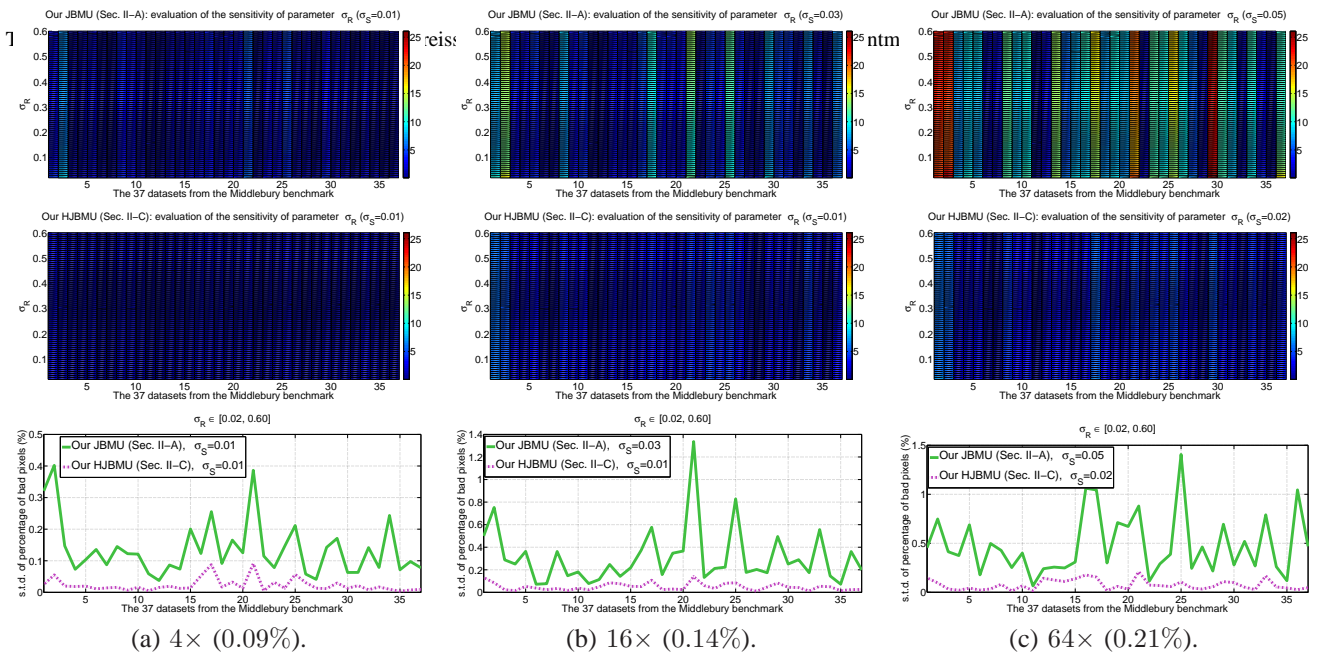
Fig. 10. Sensitivity of parameter σ_R when the spatial resolution is enhanced $4\times$, $16\times$ and $64\times$, respectively. (a)-(c) evaluate the performance when σ_S is a constant and $\sigma_R \in [0.02, 0.60]$. The color maps in the first two rows correspond to the percentage of bad pixels obtained from our JBMU and HJBMU upsampling methods. Red means higher error rate, and blue lower. The last row presents the standard deviations of our upsampling methods. The maximum standard deviations of the percentage of bad pixels from our HJBMU method are placed under the curves, which demonstrate that our method is robust to parameter σ_R for all the 37 Middlebury data sets[5].
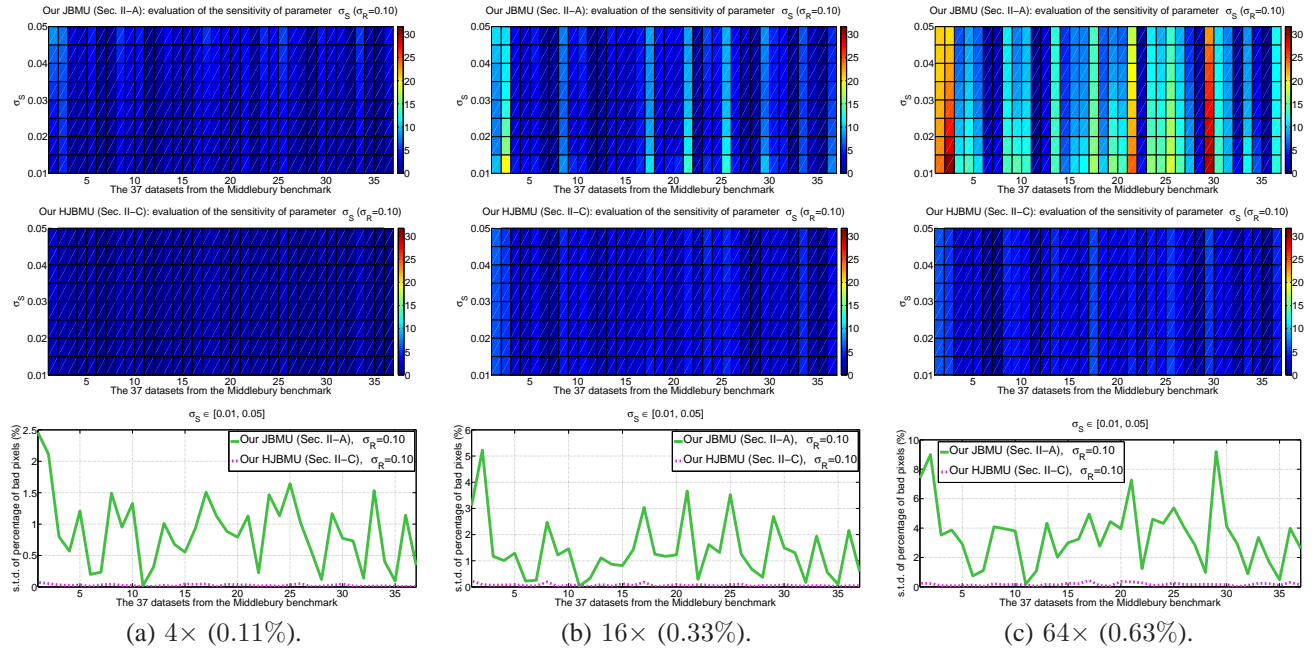

Fig. 11. Sensitivity of parameter σ_S when the spatial resolution is enhanced $4\times$, $16\times$ and $64\times$, respectively. (a)-(c) evaluate the performance when σ_R is a constant and $\sigma_S \in [0.01, 0.05]$. The color maps in the first two rows correspond to the percentage of bad pixels obtained from our JBMU and HJBMU upsampling methods. Red means higher error rate, and blue lower. The last row presents the standard deviations of our upsampling methods. The maximum standard deviations of the percentage of bad pixels from our HJBMU method are placed under the curves, which demonstrate that our method is robust to parameter σ_S for all the 37 Middlebury data sets[5].

A. Numerical Evaluation Using Middlebury Data Sets

In this section, we numerically compare Kopf's method [27] and its extensions [23], [40] with our JBMU and HJBMU method. A total of 37 data sets from Middlebury benchmark [5] are used. Each data set contains a disparity map obtained from structured light and a registered RGB image. The disparity map is considered as the ground truth for quantitative analysis. We downsample the ground-truth disparity map to obtain the low-resolution disparity map, which is then fed to the upsampling methods for spatial enhancement. To avoid

(a) Resolution is enhanced $4\times$.   (b) Resolution is enhanced $16\times$.   (c) Resolution is enhanced $64\times$.
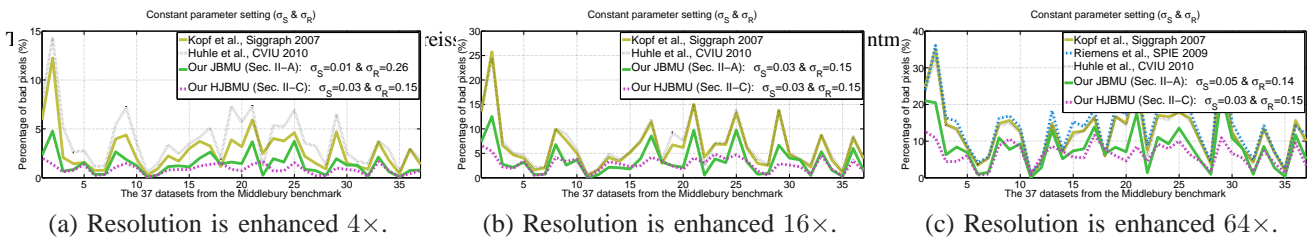
Fig. 12. Percentage of bad pixels obtained with constant parameter setting for all the 37 Middlebury data sets [5]. From top to bottom: spatial enhancement at $4\times$, $16\times$, and $64\times$.

(a) Resolution is enhanced $4\times$.   (b) Resolution is enhanced $16\times$.   (c) Resolution is enhanced $64\times$.
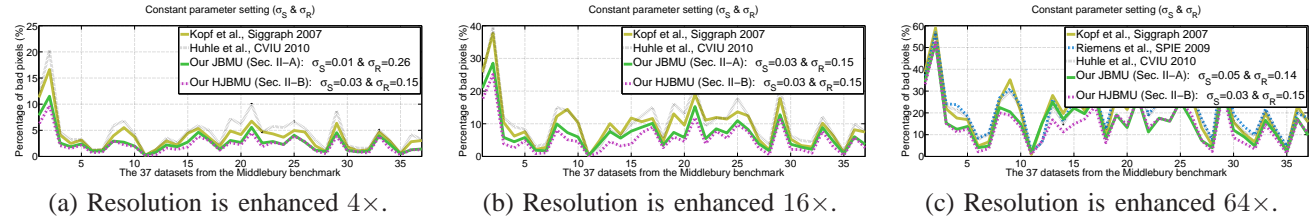
Fig. 13. Percentage of bad pixels obtained with constant parameter setting for all the 37 Middlebury data sets [5] when an anti-aliasing filter is used. From top to bottom: spatial enhancement at $4\times$, $16\times$, and $64\times$.

blurring depth edges, we adopt the depth downsampling scheme used by Middlebury benchmark [5]: no low-pass filter will be applied to the depth image prior to downsampling. The upsampled disparity map is compared with the ground truth for quantitative evaluation. The same criteria as the Middlebury benchmark is used. Specifically, we use the percentage of bad pixels. A pixel is considered to be bad pixel if the difference of the upsampled disparity map and the ground truth is larger than one disparity. The larger the percentage of bad pixels, the lower the performance.

To analyze the sensitivity of parameter σ_R and σ_S that are used to control the behavior of joint bilateral filter, we first set σ_S to a constant and then change σ_R from 0.02 to 0.60 to collect the error rate of our JBMU and HJBMU upsampling method. The experimental results are presented in Fig. 10. We next set σ_R to a constant and let σ_S vary between 0.01 and 0.05 to analyze the sensitivity of parameter σ_S. The results are reported in Fig. 11. The spatial resolution was enhanced $4\times$, $16\times$ and $64\times$ in both experiments. The colors in Fig. 10-11 correspond to the percentage of bad pixels. Red means higher percentage and blue lower. The upsampling accuracy of our HJBMU method appears to be better than JBMU method as most of its colors are blue. The last rows in Fig. 10 and 11 report the standard deviations of the upsampling errors of the two proposed methods. Let $\sigma_R \in [0.02, 0.60]$ (for 8-bit image, it means that $\sigma_R \in [5, 153]$), the maximum standard deviation of the percentage of bad pixels among the 37 Middlebury data sets [5] obtained using our HJBMU method is 0.09, 0.14 and 0.21 for spatial enhancement at $4\times$, $16\times$ and $64\times$, respectively. Apparently, the performance of HJBMU method is very robust with respect to σ_R. Let $\sigma_S \in [0.01, 0.05]$, in Fig. 11, the maximum standard deviations of our HJBMU method is 0.11, 0.33 and 0.63, respectively. This numerical comparison shows that our HJBMU method is more sensitive to σ_S than σ_R. This is also true to our JBMU method as can be seen from the green curves in the last rows in Fig. 10 and 11.

Finally, we need to choose a constant parameter setting for all the 37 Middlebury data sets [5]. For each upsampling method, we locate the σ_R values corresponding to the minimum error rate at each data set (in Fig. 10) and use the mean as the best estimate. So is the σ_S parameter. The errors obtained from these constant parameter settings are summarized in Fig. 12. As can be seen, our methods (green and pink curves) outperform the other upsampling methods [27], [40], [23] on average. The black dash curves in Fig. 12 (a)-(c) correspond to the upsampling method presented in [23], which replace the joint bilateral filter in [27] with a modified nonlocal means filter. Its performance is actually very close to [27] (the yellow curves). The performance of the nonlocal means based upsampling method presented in [23] is similar to the joint bilateral upsampling method as a noise-free high-resolution RGB image will be used to guide the upsampling of the low-resolution depth image. Meanwhile, the main contribution of the hierarchical joint bilateral upsampling method presented in [40] is to reduce the computational cost of the original joint bilateral upsampling method [27] thus its performance (the blue curve in Fig. 12 (c)) is also close to [27]. Note that the error rates of all the upsampling methods at the 11th data set are very close to each other. This is mainly because the depth variances of this data set is very smooth, and thus very few depth edges.

The low-resolution disparity maps used in all the above experiments were downsampled from the corresponding high-resolution disparity maps without anti-aliasing pre-filtering. In practice, some depth cameras do come with anti-aliasing filters. To simulate the behavior of this type of cameras, a low-pass filter was applied to the high-resolution disparity maps before downsampling, and the upsampling errors obtained with constant parameter settings are summarized in Fig. 13. The low-pass filter blurs the depth edges and deteriorates the upsampling accuracy.

(a)Huhle *et al.* [23].

(b)Riemens *et al.* [40].
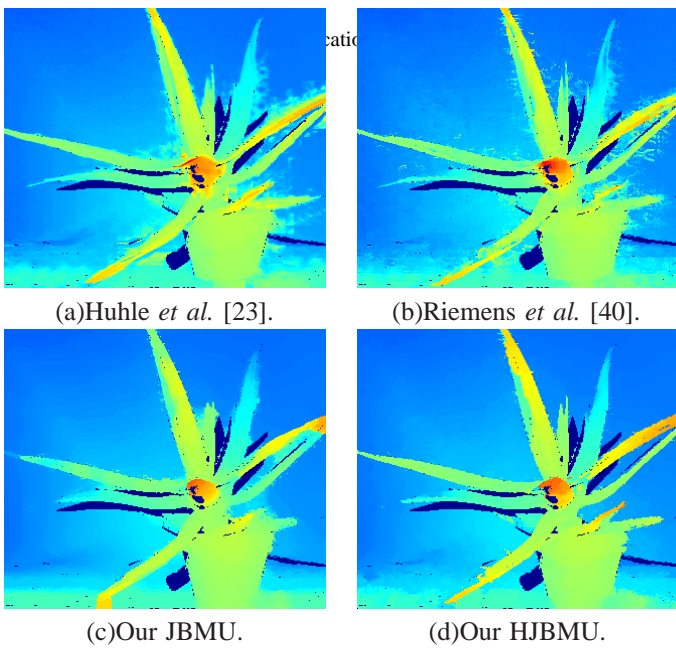
(c)Our JBMU.

(d)Our HJBMU.

Fig. 14. Upsampling results on a textured Middlebury data set. (a) to (d) are the upsampled disparity maps obtained using the upsampling methods presented in [23], [40], our JBMU method, and our HJBMU method, respectively. As can be seen, Kopf's method [27] (see Fig. 1 (d)) and its extensions [23], [40] are not suitable for textured scenes as the joint bilateral filter and the nonlocal means filter are weak for texture suppression. The spatial resolution is enhanced $64\times$ in this example.

B. Visual Evaluation Using Real Indoor Data Sets

We present visual evaluation on the presented upsampling methods in this section. We start from Middlebury data sets. Fig. 1 (d) and Fig. 14 present experimental results on the textured Middlebury data set in Fig. 1 (a). As can be seen, Kopf's method [27] and its extensions ([23] and [40]) are not suitable for textured scenes as the joint bilateral filter and the nonlocal means filter are weak for texture suppression. We next present results obtained from a real-time high-resolution depth sensing system based on our HJBMU method. Two frames extracted from a video captured from our system is presented in Figure 15. The first row in Figure 15 (a) and (b) are high-resolution (640×480) RGB images and low-resolution (160×120) depth images and brightness images captured by Canesta sensor [2]. Our system can upsample the depth image and brightness image at a rate of 1200 Hertz (on an Nvidia Geforce GTX 580 GPU) with very few visible errors as shown in the last row of Figure 15 (a) and (b).

VI. CONCLUSION

We have presented a robust range image upsampling method based on weighted median filtering. It extends our previous upsampling methods presented in [54] and [51] for non-fronto-parallel scenes. Additionally, we have given detailed analysis of its parameter sensitivity and discussed the relationship between the proposed framework and the weighted median filter. Using adaptive weights computed from a registered high-resolution RGB image based on a joint bilateral filter operation, the proposed weighted median filter based up-sampling method can upsample 160×120 depth image to
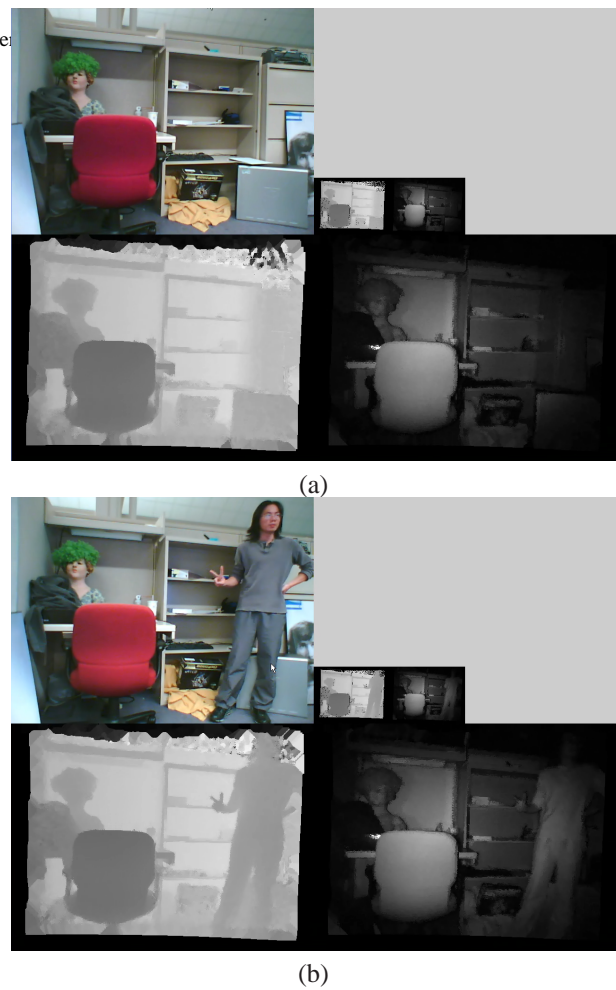

(a)

(b)

Fig. 15. Frames captured using our real-time depth upsampling system. The first rows in (a) and (b) are high-resolution (640×480) RGB images and low-resolution (160×120) depth images and brightness images captured by Canesta sensor [2]. The last rows in (a) and (b) present the upsampling results. The depth edges are well-preserved in the upsampled images, e.g., the details of the fingers in (b).

VGA size at a rate of 1200 Hertz with few noticeable errors. The effectiveness and robustness of proposed method have been demonstrated on several real data sets, including 37 Middlebury data sets.

REFERENCES

[1] 3dv systems, z-cam.
http://en.wikipedia.org/wiki/ZCam.
[2] Canesta inc.
http://en.wikipedia.org/wiki/Canesta.
[3] Cuda c/c++ sdk code samples.
http://developer.nvidia.com/cuda/cuda-cc-sdk-code-samples.
[4] Fotonic, fotonic-b70.
http://www.fotonic.com/content/News-And-Press/Default.aspx.
[5] Middlebury stereo benchmark.
http://vision.middlebury.edu/stereo/data.
[6] Pmd technologies, pmd s3.
http://www.pmdtec.com/.
[7] G. R. Arce and J. Paredes. Image enhancement and analysis with weighted medians. *Nonlinear Image Processing*, pages 27–68, 2001.
[8] K. E. Barner and T. C. Aysal. Polynomial weighted median filtering. *IEEE Trans. Signal Processing*, 54(2):635–650, 2006.

[9] J. Batlle, E. Mouaddib, and J. Salvi. Recent progress in coded structured light as a technique to solve the correspondence problem: A survey. *Pattern Recognition*, 31(7):963–982, 1998.

[10] P. Besl. *Active Optical Range Imaging Sensors, in Advances in Machine Vision, chapter 1*, pages 1–63. 1989.

[11] A. C. Bovik. Streaking in median filtered images. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 35(4):493–503, 1987.

[12] A. C. Bovik, T. S. Huang, and D. C. Munson, Jr. The effect of median filtering on edge estimation and detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 9(2):181–194, 1987.

[13] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(11):1222–1239, 2001.

[14] D. R. K. Brownrigg. The weighted median filter. *Commun. ACM*, 27(8):807–818, 1984.

[15] A. Buades, B. Coll, and J. Morel. Nonlocal image and movie denoising. *International Journal of Computer Vision*, 76:123–139, 2008.

[16] J. Diebel and S. Thrun. An application of markov random fields to range sensing. In *Neural Information Processing Systems*, pages 291–298, 2005.

[17] F. Durand and J. Dorsey. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Transactions on Graphics*, 21(3):257–266, 2002.

[18] E. Eisemann and F. Durand. Flash photography enhancement via intrinsic relighting. *ACM Transactions on Graphics*, 23(3):673–678, 2004.

[19] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski. Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Transactions on Graphics*, 27(3):67:1–67:10, 2008.

[20] P. Felzenszwalb and D. Huttenlocher. Efficient belief propagation for early vision. *International Journal of Computer Vision*, 70(1):41–54, 2006.

[21] W. T. Freeman, E. Pasztor, and O. T. Carmichael. Learning low-level vision. *International Journal of Computer Vision*, 40(1):25–47, 2000.

[22] P. Huber. *Robust statistics*. Wiley, New York, 1981.

[23] B. Huhle, T. Schairer, P. Jenke, and W. Straíer. Fusion of range and color images for denoising and resolution enhancement with a non-local filter. *Computer Vision and Image Understanding*, 114:1336–1345, 2010.

[24] R. Jarvis. A perspective on range finding techniques for computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5(2):122–139, 1983.

[25] S. J. Ko and Y. H. Lee. Center weighted median filters and their applications to image enhancement. *IEEE Transactions on Circuits and Systems*, 38(2):984–993, 1991.

[26] A. Kolb, E. Barth, R. Koch, and R. Larsen. Time-of-flight cameras in computer graphics. *Computer Graphics Forum*, pages 141–159, 2009.

[27] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele. Joint bilateral upsampling. *ACM Transactions on Graphics*, 26(3):673–678, 2007.

[28] Feng Li, Jingyi Yu, and Jinxiang Chai. A hybrid camera for motion deblurring and depth map super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–12, 2008.

[29] Tukey. l.W. Exploratory data analysis. *Addison-Wesley, Reading, Mass.*, 1974.

[30] Swiss Ranger SR-2 MESA Imaging AG. The swiss center for electronics and microtechnology.
http://www.mesa-imaging.ch/.

[31] M. Mueller, F. Zilly, and P. Kauff. Adaptive cross-trilateral depth map filtering. In *3DTV Conference - DTV-CON*, pages 1–4, 2010.

[32] A. Nasonov and A. Krylov. Fast super-resolution using weighted median filtering. In *International Conference on Pattern Recognition*, pages 2230 – 2233, 2010.

[33] Y. Nie and K. E. Barner. The fuzzy transformation and its applications in image processing. *IEEE Trans. Image Processing*, 15(4):910–927, 2006.

[34] Ari Nieminen, Pekka Heinonen, and Yrjo Neuvo. A new class of detail-preserving filters for image processing. *IEEE Trans. Pattern Anal. Mach. Intell.*, 9(1):74–90, 1987.

[35] S. Paris and F. Durand. A fast approximation of the bilateral filter using a signal processing approach. *International Journal of Computer Vision*, 81:24–52, January 2009.

[36] S. Paris, P. Kornprobst, J. Tumblin, and F. Durand. Bilateral filtering: Theory and applications. *Foundations and Trends in Computer Graphics and Vision*, 4(1):1–73, 2009.

[37] G. Petschnigg, M. Agrawala, H. Hoppe, R. Szeliski, M. Cohen, and K. Toyama. Digital photography with flash and no-flash image pairs. *ACM Transactions on Graphics*, 23(3):664–672, 2004.

[38] D. Poussart and D. Laurendeau. *3-D Sensing for Industrial Computer Vision, in Advances in Machine Vision, chapter 3*, pages 122–159. 1989.

[39] W. H. Press. *Numerical recipes in C: the art of scientific computing*. Cambridge University Press, New York, 1988.

[40] A. K. Riemens, O.P. Gangwal, B. Barenbrug, and R-P.M Berretty. Multistep joint bilateral depth upsampling. In *SPIE Visual Communication and Image Processing*, volume 7257, pages 336–347, 2009.

[41] J. Salvi, J. Pagès, and J. Batlle. Pattern codification strategies in structured light systems. *Pattern Recognition*, 37(4):827–849, 2004.

[42] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1/2/3):7–42, April-June 2002.

[43] H. Senel, R. Peters, and B. Dawant. Topological median filters. *IEEE Trans. Image Processing*, 11(2):89–104, 2002.

[44] T. C. Strand. Optical three-dimensional sensing for machine vision. *Optical Engineering*, 24(1):33–40, 1985.

[45] K. Subr, C. Soler, and F. Durand. Edge-preserving multiscale image decomposition based on local extrema. *ACM Transactions on Graphics*, pages 147:1–147:9, 2009.

[46] J. Sun, N. Zheng, and H. Y. Shum. Stereo matching using belief propagation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(7):787–800, 2003.

[47] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *IEEE International Conference on Computer Vision*, pages 839–846, 1998.

[48] L. Xu, C. Lu, Y. Xu, and J. Jia. Image smoothing via l0 gradient minimization. *ACM Transactions on Graphics*, 30(6):174:1–174:12, 2011.

[49] Q. Yang. Recursive bilateral filtering. In *European Conference on Computer Vision*, pages 399–413, 2012.

[50] Q. Yang, K.-H. Tan, and N. Ahuja. Real-time o(1) bilateral filtering. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 557–564, 2009.

[51] Q. Yang, K.-H. Tan, B. Culbertson, and J. Apostolopoulos. Fusion of active and passive sensors for fast 3d capture. In *IEEE International Workshop on Multimedia Signal Processing*, pages 69–74, 2010.

[52] Q. Yang, L. Wang, and N. Ahuja. A constant-space belief propagation algorithm for stereo matching. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1458–1465, 2010.

[53] Q. Yang, L. Wang, R. Yang, H. Stewenius, and D. Nister. Stereo matching with color-weighted correlation, hierachical belief propagation and occlusion handling. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(3):492–504, 2009.

[54] Q. Yang, R. Yang, J. Davis, and D. Nistér. Spatial-depth super resolution for range images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 291–298, 2007.

[55] L. Yin, R. Yang, M. Gabbouj, and Y. Neuvo. Weighted median filters: A tutorial. *IEEE Trans. Circuits and Systems II: Analog and Digital Signal Processing*, 43(3):157–192, 1996.
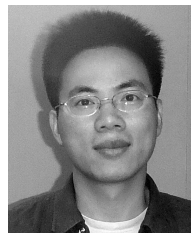
Qingxiong Yang received the BE degree in Electronic Engineering & Information Science from University of Science & Technology of China in 2004 and the PhD degree in Electrical & Computer Engineering from University of Illinois at Urbana-Champaign in 2010. He is an assistant Professor in the Computer Science Department at City University of Hong Kong. His research interests reside in computer vision and computer graphics. He is a recipient of the best student paper award at MMSP 2010 and the best demo award at CVPR 2007.

Narendra Ahuja received the B.E. degree with honors in electronics engineering from the Birla Institute of Technology and Science, Pilani, India, in 1972, the M.E. degree with distinction in electrical communication engineering from the Indian Institute of Science, Bangalore, India, in 1974, and the Ph.D. degree in computer science from the University of Maryland, College Park, USA, in 1979. From 1974 to 1975 he was Scientific Officer in the Department of Electronics, Government of India, New Delhi. From 1975 to 1979 he was at the Computer Vision Laboratory, University of Maryland, College Park. Since 1979 he has been with the University of Illinois at Urbana-Champaign where he is currently Donald Biggar Willet Professor in the Department of Electrical and Computer Engineering, the Beckman Institute, and the Coordinated Science Laboratory. His current research is focused on extraction and representation of spatial structure in images and video; integrated use of multiple image-based sources for scene representation and recognition; versatile sensors for computer vision; and applications including visual communication, image manipulation, and information retrieval. He is a fellow of IEEE, ACM, AAAI, AAAS, IAPR, and SPIE.

Ruigang Yang received the MS degree in Computer Science from Columbia University in 1998 and the PhD degree in Computer Science from the University of North Carolina, Chapel Hill, in 2003. He is an associate professor in the Computer Science Department, University of Kentucky. His research interests include computer vision, computer graphics, and multimedia. He is a recipient of the US National Science Foundation (NSF) Faculty Early Career Development (CAREER) Program Award in 2004.

Kar-Han Tan received the B.Sc. degree from the National University of Singapore, Singapore, the M.S. degree from the University of California, Los Angeles, and the Ph.D. in computer science from the University of Illinois at Urbana-Champaign, Urbana, IL.

While with the University of Illinois at Urbana-Champaign, he was a Beckman Graduate Fellow. He is currently a Senior Research Scientist with the Mobile and Immersive Experience Lab (MIXL), Hewlett-Packard Laboratories, Palo Alto, CA, where he is working on 3-D capture and display technologies as well as next-generation remote collaboration systems. He contributes actively to the research community and has received best paper awards for his work. Prior to joining Hewlett-Packard, he was Manager of Algorithms Group, EPSON R&D, where he led the invention of View Projection, a technique that enables one-touch setup of light displays on arbitrary surfaces. He coinvented multi-flash imaging at Mitsubishi Electric Research Lab (MERL) and the Virtual Structures algorithm at the University of California, Los Angeles, which is widely recognized today as one of the fundamental techniques for mobile robot formation control.

James Davis is an Associate Professor in Computer Science at the University of California, Santa Cruz. He received his PhD from Stanford University in 2002, and was previously a senior research scientist at Honda Research Institute. His existing research expertise is in computer graphics, machine vision, and sensing systems for building digital models of the real world, work that has resulted in over 80 peer-reviewed publications, patents, and invited talks, received best paper awards at ICRA 2003, ICCV 2009, and an NSF CAREER award. His research has been commercialized by companies including Sony and PrenticeHall. He serves as the Faculty Director of the Center for Entrepreneurship at UCSC, and sits on advisory councils for a handful of startups and nonprofits.
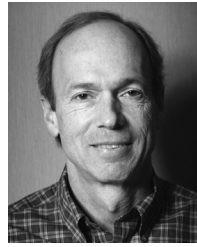
Bruce Culbertson received the M.A. degree in mathematics from the University of California, San Diego, and the M.S. degree in computer and information science from Dartmouth College, Hanover, NH.

He joined Hewlett-Packard in 1983 and has worked at Hewlett-Packard Laboratories, Palo Alto, CA, since 1984. He manages researchers with expertise in computer vision and graphics who have applied those disciplines primarily to multimedia communications for the last decade. Prior to that, he worked on voice and data networks, computer design, computer-aided design of digital logic, reconfigurable computers, and defect-tolerant computer design. He has served on the program committees of IEEE CVPR and IEEE ICCV. His research interests are in computer vision and multimedia.

John Apostolopoulos received the B.S., M.S., and Ph.D. degrees from the Masachusetts Institute of Technology (MIT), Cambridge.
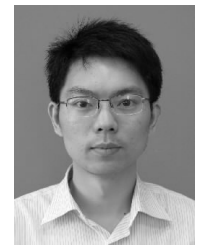
He is a Distinguished Technologist and the Director of the Mobile and Immersive Experience Lab (MIXL) with Hewlett-Packard Laboratories, Palo Alto, CA. He also teaches and conducts joint research at Stanford University, Stanford, CA, where he is a Consulting Associate Professor of Electrical Engineering, and he is a frequent Visiting Lecturer with MIT. His research interests include immersive communication and improving the reliability, fidelity, scalability, and security of multimedia communications over wired and wireless packet networks.

Dr. Apostolopoulos was the recipient of Emmy Award Certificate for his contributions to the U.S. Digital TV standard. He was named one of the world's top 100 young (under 35) innovators in science and technology (TR100) by MIT Technology Review in 2003 and was the recipient of a number of Best Paper awards.

Gang Wang is an Assistant Professor in Electrical and Electronic Engineering at the Nanyang Technological University. He is also a Research Scientist of the Advanced Digital Science Center. He received the B.S. degree from Harbin Institute of Technology, China, in 2005 and the Ph.D. degree from the University of Illinois at Urbana-Champaign, Urbana. His research interests include computer vision and machine learning.