Structure Selective Depth Superresolution for RGB-D Cameras

Youngjung Kim, *Student Member, IEEE*, Bumsub Ham, *Member, IEEE*, Changjae Oh, *Student Member, IEEE*, and Kwanghoon Sohn, *Senior Member, IEEE*

***Abstract*— This paper describes a method for high-quality depth superresolution. The standard formulations of image-guided depth upsampling, using simple joint filtering or quadratic optimization, lead to texture copying and depth bleeding artifacts. These artifacts are caused by inherent discrepancy of structures in data from different sensors. Although there exists some correlation between depth and intensity discontinuities, they are different in distribution and formation. To tackle this problem, we formulate an optimization model using a nonconvex regularizer. A nonlocal affinity established in a high-dimensional feature space is used to offer precisely localized depth boundaries. We show that the proposed method iteratively handles differences in structure between depth and intensity images. This property enables reducing texture copying and depth bleeding artifacts significantly on a variety of range data sets. We also propose a fast alternating direction method of multipliers algorithm to solve our optimization problem. Our solver shows a noticeable speed up compared with the conventional majorize-minimize algorithm. Extensive experiments with synthetic and real-world data sets demonstrate that the proposed method is superior to the existing methods.**

***Index Terms*— Depth upsampling, depth sensor, time of flight camera, nonconvex regularization.**

I. Introduction

PROVIDING accurate depth information of real scenes is an essential task in the fields of computer vision and image processing, and this arises in a wide range of applications, such as image-based rendering, 3DTV, and robot navigation. The 3D structure of a scene provides decisive clues, enabling overcoming limitations in several tasks such as, scene understanding [1], intrinsic image estimation [2], and visual saliency [3].

Acquiring high-quality depth information is, however, not trivial. Laser line or structured light scanners can capture accurate depth information, but their applications are limited to restricted environments, e.g., a static scene. Alternatively, depth information can be estimated through a computational approach, i.e., stereo matching algorithm. This approach computes a disparity by finding correspondences between two images and determines depth information via triangulation. The performance of stereo matching algorithms has been significantly improved, but there are still some problems for the practical application, e.g., the ill-posedness for textureless regions and the sensitivity to illumination. Using depth sensors such as Time-of-Flight (ToF) cameras is an alternative to the computational approach. This enables capturing depth information for dynamic scenes with video rate, but obtained depth images are not satisfactory due to the inherent physical limitation of ToF sensors. The resolution of the captured depth image is relatively low compared with that of the intensity image, and the captured depth image suffers from significant noise.

There have been many attempts to overcome these limitations of ToF sensors. We roughly classify the existing methods into two groups with respect to whether or not an external dictionary is used. A depth image can be modeled by a generic dictionary of local patches. For a given low resolution (LR) depth patch, the dictionary provides a suitable high resolution (HR) one to synthetically enhance the resolution of the LR depth image [4]. One can use a joint sparse coding scheme in a coupled feature space to learn statistical dependencies between intensity and depth patches [5], [6]. These methods work in general, but they usually require millions of HR/LR patch pairs in order to represent the various depth patches. Instead of relying on external dictionaries, several approaches exploit co-occurrence property between depth and intensity boundaries, i.e., intensity guided depth super-resolution [8]. The underlying assumption is that pixels having similar intensity values are likely to have similar depth. In practice, however, intensity discontinuities (or textures) do not always coincide with those of the depth image − we refer to this problem as *structure discrepancy*. The strong dependence on co-occurrence between depth and intensity images results in noticeable artifacts such as texture copying and depth bleeding. Thus, the main challenge in intensity guided depth super-resolution is to find commonly usable edges and smooth transition for visually compelling HR depth images.

This paper presents a depth upsampling framework that belongs to the second category. We propose a nonconvex model for reconstructing a HR depth image from its noisy and LR measurement. Our approach uses a simple yet effective

Manuscript received November 3, 2015; revised June 10, 2016; accepted August 3, 2016. Date of publication August 18, 2016; date of current version September 16, 2016. This work was supported by the Institute for Information and Communications Technology Promotion through the Korean Government (MSIP), High quality 2d-to-multiview contents generation from large-scale RGB+D database, under Grant R0115-15-1007. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Gene Cheung.

The authors are with the School of Electrical and Electronic Engineering, Yonsei University, Seoul 120-749, South Korea (e-mail: read12300@yonsei.ac.kr; ocj1211@yonsei.ac.kr; mimo@yonsei.ac.rk; khsohn@yonsei.ac.kr).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org., provided by the author. The material includes results for additional images. The total size of the files is 14.7 MB. Contact khsohn@yonsei.ac.kr for further questions about this work.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2016.2601262

nonlocal affinity constructed by k nearest neighbors (kNN) in a high-dimensional feature space. Rather than choosing spatially neighboring pixels, we connect pixels which are likely to have similar depth in the final estimation. A key aspect of the proposed method is that the structural discrepancy between intensity and depth images are compensated recursively, which offers sharp depth discontinuities without depth bleeding artifacts. This paper extends our work [7] in a number of ways.

- We show an extensive analysis of the proposed model and relations with other approaches.
- An intensive experimental study and comparison with other methods are presented. We apply the proposed model on three types of depth sensors for the comparison.
- We develop an alternating direction method of multipliers (ADMM) solver based on variable splitting and first order proximal approximation. Our results show that the proposed scheme is around five times faster than the conventional majorize-minimize (MM) [7] method in the case of depth upsampling.

The rest of this paper is organized as follows. Section II describes the related works. Section III presents our approach to depth upsampling. We derive an efficient numerical algorithm in Section IV. Section V analyzes the relations to the existing methods. The performance of the proposed method is then demonstrated in Section VI. Finally, we conclude the paper in Section VII.

II. Related Works

In this section, we briefly review previous works on intensity guided depth upsampling. It can be further classified into two categories: local filtering-based methods and global optimization-based methods. The first ones employ a local image filter under guidance of the HR intensity image. The second ones solve a global optimization problem specified by an objective function that consists of data and regularization terms. Both methods can be applied to a label domain, not an image domain directly.

A. Local Filtering-Based Approach

The standard approaches use explicit weighted-average operations, where the weights are decided by the affinities computed from a HR intensity image. As a pioneering work, Kopf *et al.* [8] proposed the joint bilateral filtering (JBU) method for cross-modal image enhancement. The JBU has been employed in a variety of applications including colorization, tone-mapping and depth upsampling. He *et al.* proposed an edge-preserving operator based on the guided filtering [9]. Liu *et al.* computed the intensity-guided weights by using geodesic distance and reduced computational complexity via dynamic programming [10]. One of the main drawbacks in these approaches is that they do not handle the differences in structure between intensity and depth images. The simple weighted average filters transfer intensity information to the depth image directly, resulting in notable structure changes. There have been several attempts to handle this problem. Yang *et al.* [11] proposed a cost aggregation method with an iterative JBU. In contrast to the JBU [8] that applies the filtering operation to the depth image, the upsampling is performed in a 3D cost volume. Min *et al.* [12] used a local histogram and proposed the weighted mode filtering (WMOF). Similarly, the weighted median filtering (WMEF) [13], and the nonlocal tree filtering (NWMEF) [14] were proposed to address the inherent smoothing property of the local averaging filters. Lo *et al.* integrated local gradient information of the LR depth image, and proposed the joint trilateral filtering (JTF) [15].

B. Global Optimization-Based Approach

The most relevant works to ours are the works of [16]–[24]. Diebel and Thrun [16] proposed an MRF based method with an explicit smoothness assumption on the underlying depth image. This framework provides a promising result, but it needs to solve a large sparse matrix which solely depends on the intensity appearance. Park *et al.* [18] adopted a nonlocal approach to depth upsampling. They used a linear combination of several weighting terms in the form of least squares optimization, including appearance, segmentation, edge saliency, and bicubic interpolated depth. A similar nonlocal scheme is proposed in [19] based on an adaptive autoregressive (AR) model. The constraint from the initial estimates is reused for depth upsampling. This heuristic may increase accuracy of estimation, but good initial estimates are not available for many challenging environments. Ferstl *et al.* modeled a smoothness term using a second order total generalized variation, enforcing a piecewise affine solution [20]. Ham *et al.* formulated depth upsampling as a graph-based transduction problem, and designed an indication matrix that leverages the property of a noisy depth image [21].

Beyond these convex models, there are few approaches based on nonconvex optimization. Shen *et al.* introduced a scale map to model the structure discrepancy between depth and intensity images [22]. Although this method has achieved state-of-the-art performance, explicit representations of the scale map are numerically unstable when the intensity edge does not exist, i.e., zero-gradient. Recently, Ham *et al.* [23] formulated image filtering as a nonconvex optimization problem. Hu *et al.* used the stereo view of the target depth image to compute a synthesized view matching error, which is used as a regularization term [24]. Our model differs from [22], [23], [24] in its formulation and solver. To compute affinities, we utilize the nonlocal principle established in the high-dimensional feature space. We also propose an effective and fast solver based on variable splitting and first order proximal approximation.

III. Proposed Model

A. Motivation

There have been many attempts to enhance the resolution of the LR depth image by using the corresponding HR intensity image, e.g., optimizing a convex objective function [16], [17] or employing weighted average filters [8], [9]. However, these methods have difficulties in dealing with the structure discrepancy between the intensity and depth images — the regularity of the depth image is determined by

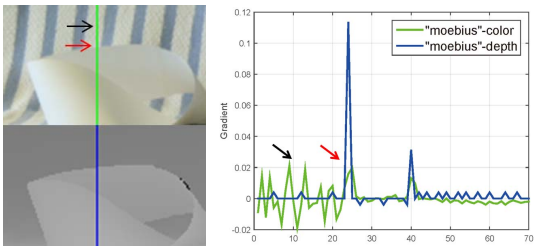


Fig. 1. Structural discrepancy between intensity and depth images: (From left to right) patches in the "*moebius*" sequence [43] and gradients along the vertical line in the two patches. The gradient distributions of depth and intensity images are different. This inconsistency causes texture copying and depth bleeding artifacts in the conventional intensity guided depth upsampling methods.

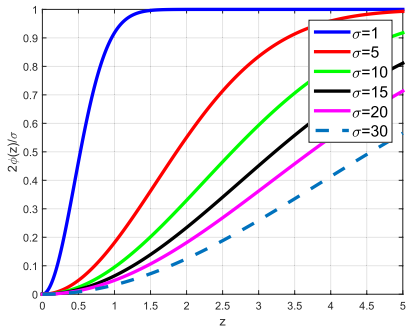


Fig. 2. Comparison of the regularization functions, ϕ with different σ: The proposed regularizer is abruptly saturated when σ is small, minimizing the edge blurring compared to the convex choices.

the structure in the intensity image only. The main inconsistency comes from the gradient magnitude variation as shown in Fig. 1. In this case, the gradient magnitude of the depth image becomes similar to that of the intensity image, resulting in a notable structure change. For example, the depth edge is blurred when the corresponding intensity image has relatively small gradient magnitude (e.g., the red arrow in Fig. 1). This behavior known as depth bleeding artifacts may blur strong depth edges. Similarly, all textural information of the intensity image is permeated into the depth image. For a textured region, (e.g., the black arrow in Fig. 1), the depth image is highly perturbed by the texture pattern (i.e., texture copying artifacts). These depth bleeding and texture copying artifacts may be reduced by cost aggregation methods [11]–[21], but they are still observed around object boundaries or in smooth surfaces.

In the following sections, we present a model to address the aforementioned problems. Our model is built upon a key observation that the gradient magnitude variation (i.e, different contrast) can be adjusted by rescaling of intensity gradients. We assume that rescaling factors can be determined implicitly by considering the underlying structure of depth images.

B. Modeling and Formulation

This section describes our model for depth upsampling. We assume that LR depth and HR intensity images are registered. Let g be the aligned sparse depth image and $\mathbf{I}$ be the intensity image in a vector form. Then, our objective is to estimate the underlying high-quality depth image u by jointly using $\mathbf{I}$ and the latent structure of u. In the following, Ω represents a set of depth points, and i or j denotes pixel indexes. T denotes the total number of pixels in HR intensity images.

1) Objective Function: We want the final estimate u to be similar to the initially observed data g and be consistent with their neighboring elements. Formally, we minimize an objective function in the form:

$$E(u) = \sum_{i\in\Omega} \frac{\lambda}{2}(u-g)_i^2 + E_s(u), \tag{1}$$

where the first and second terms denote the data consistency and regularization terms, respectively. The scalar λ is relative confidence to balance the two terms. The regularization penalty E_s is defined as:

$$E_s(u) = \sum_{p=1}^{|\mathcal{N}|} w_p^{\mathbf{I}} \phi(\mathbf{A}u)_p, \tag{2}$$

where p indexes a set of neighborhoods $\mathcal{N}$. If j is the neighborhood of i, then $(i,j)\in\mathcal{N}$. The choice of $\mathcal{N}$ will be detailed in the next section. $w_p^{\mathbf{I}} = \exp(-\|\mathbf{A}\mathbf{I}\|_p^2/\sigma_{\mathbf{I}}))$ denotes an affinity function for adjusting the strength of regularization based on the intensity structure. $\sigma_{\mathbf{I}}$ is a range bandwidth. $\mathbf{A}$ is a $|\mathcal{N}|\times T$ incidence matrix [25], defined as:

$$\mathbf{A}_{p,q} = \begin{cases} +1 & \text{if } q=i \\ -1 & \text{if } q=j \\ 0 & \text{otherwise,} \end{cases} \qquad (i,j)\in\mathcal{N}. \tag{3}$$

We introduce the regularity of the underlying depth image by Welsch's function, $\phi(z) = \sigma\left(1-\exp\left(-z^2/\sigma\right)\right)/2$. The reason of adapting such a function is twofold. First, the function ϕ satisfies $\lim_{|z|\to\infty}\phi'(z) = 0$, and thus it belongs to a robust M-estimator, which has some robustness and edge-preserving properties [26]. Second, it smoothly approximates the zero-norm, i.e., $|\cdot|_0$, that reflects the sparsity of natural depth images [5]. The regularization function is illustrated in Fig. 2. The function ϕ saturates as the neighboring pixels become dissimilar. This ensures that discontinuities in the depth image are not penalized, minimizing the *depth blurring* compared to the convex functions. Another important aspect is that ϕ is nonconvex with small values of σ, which may enhance the noise. However, the proposed regularization penalty is not only specified by the function ϕ but also by the pre-selected weight $w^{\mathbf{I}}$, derived from the high-quality intensity image [26]. This property allows us to suppress noise and visual artifacts. Note that the assumption about the intensity-depth coherence is usually achieved by optimizing a quadratic objective function [16], [17], which transfers intensity information to the depth image directly. The characteristics of the proposed model compared with other ones will be further analyzed in Section V.

2) Neighborhood System: There are several approaches to establishing neighborhoods from input data. Here, we introduce a kNN neighborhood system to define the set $\mathcal{N}$ in (2). Although the nonlocal neighborhood (NLN) system proposed in [18] and [19] is useful in maintaining fine details even

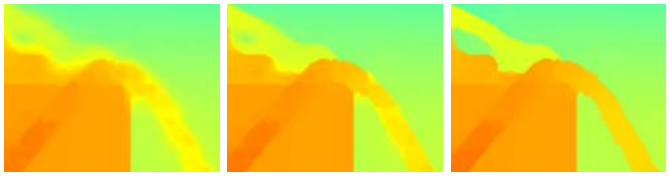


Fig. 3. Depth upsampling (×8) results on the "*reindeer*" sequence [43]: (From left to right) the proposed methods with iteration 1, 15, and 30. The proposed model gradually produces sharp depth discontinuities without noticeable artifacts.

with noisy data, it is a computationally demanding process. To overcome this difficulty, we propose an efficient connectivity strategy based on kNN search algorithm. Rather than choosing spatially neighboring pixels, the proposed neighborhood system relates each pixel i to k other pixels that represent a high probability of having similar depth. This strategy not only leads to nonlocal interaction, but also maintains a sparse affinity matrix. To this end, we connect pixels which are likely to have similar depth in the final estimation. For this initial guess, we define a feature vector $\mathbf{F}_i$ to each pixel i:

$$\mathbf{F}_i \triangleq (\alpha \mathbf{I}_i, \eta d_i, x, y)^T, \tag{4}$$

where the vector $\mathbf{I}_i$ is RGB pixel values of the intensity image, d is the bilinearly interpolated depth image from g, and the scalar values α and η control the importance of each element in the feature space. The position of i in two-dimensional space, (x, y) is also included to consider the spatial coherence. Then, for each pixel i, we simply find its kNN in this six-dimensional feature space, and such NN pairs are added to the set $\mathcal{N}$. It is worth of noting that our initial estimate may be wrong if feature space is not discriminative enough. However, the initial estimate is gradually refined through our model (please see Fig. 3).

IV. Fast Numerical Solution Using ADMM

The optimization problem of (1) can be iteratively solved by the MM algorithm [7], [23]. In the MM iteration, the next estimate u^{t+1} is obtained by solving the following linear system[1] (see Appendix for more details).

$$\left(\lambda \Lambda + \mathbf{A}^T \mathbf{W}^t \mathbf{A}\right) u^{t+1} = \lambda \Lambda g, \tag{5}$$

where Λ is diagonal matrix where $\Lambda_{ii} = 1$ for $i \in \Omega$ and 0 otherwise, $\mathbf{W}^t = \mathrm{diag}(w^{\mathbf{I}} w^{u^t})$, and $w_p^{u^t} = \exp(-(\mathbf{A}u^t)_p^2/\sigma_u)$. We note that a series of the inhomogeneous Laplacian matrices $(\mathbf{A}^T \mathbf{W}^t \mathbf{A})$ is produced, but the right-hand side is fixed at each iteration. Although the matrix $\left(\lambda \Lambda + \mathbf{A}^T \mathbf{W}^t \mathbf{A}\right)$ is invertible, it can be ill-conditioned. Thus, solving (5) generally requires huge computational cost. Several preconditioning techniques have been recently proposed [35], [36] for sparse linear systems. They, however, may not further accelerate the MM algorithm since the linear system in (5) varies in iterations. A series of preconditioners should be computed, which are not optimal and waste much time. In the following, we introduce a fast algorithm to solve our optimization problem of (1). The solver is based on variable splitting [38] and constrained optimization techniques [37], [39].

First, we can find an equivalent formulation of (1) as follows.

$$\min_{u,v} E(u, v) = \sum_{i \in \Omega} \frac{\lambda}{2} (u - g)_i^2 + \sum_{p=1}^{|\mathcal{N}|} w_p^{\mathbf{I}} \phi(v)_p, \quad \text{s.t.} v = \mathbf{A}u \tag{6}$$

The auxiliary variable v for $\mathbf{A}u$ is introduced to separate the calculation of the data consistency term and the nonconvex regularizer. The constrained optimization problem in (6) coincides with (1) in the feasible set, $\{v = \mathbf{A}u\}$. The rationale behind this formulation is that it may be easy to solve the constrained problem rather than solving its unconstrained counterpart. We adopt the augmented Lagrangian (AL) method to approximate the problem in (6) as follows:

$$\begin{aligned} E_{\mathrm{AL}}(u, v, \gamma) = &\sum_{i \in \Omega} \frac{\lambda}{2} (u - g)_i^2 \\ &+ \sum_{p=1}^{|\mathcal{N}|} (w_p^{\mathbf{I}} \phi(v)_p + \frac{\beta}{2} (\mathbf{A}u - v - \gamma)_p^2) \end{aligned} \tag{7}$$

where γ is the augmented Lagrangian multiplier, specific for our problem. β is the parameter to penalize the difference between $\mathbf{A}u - \gamma$ and v. The ADMM iteration for solving (7), is then given by:

$$\begin{aligned} u^{t+1} &= \arg\min_u E_{\mathrm{AL}}\left(u, v^t, \gamma^t\right) \\ v^{t+1} &= \arg\min_v E_{\mathrm{AL}}\left(u^{t+1}, v, \gamma^t\right) \\ \gamma^{t+1} &= \gamma^t - \left(\mathbf{A}u^{t+1} - v^{t+1}\right). \end{aligned} \tag{8}$$

From the perspective of convergence, the AL method is same to the classical penalty decomposition (PD) [37]. However, the AL is more suitable for our depth upsampling framework. The PD method approaches the original constrained problem only if $\beta \to \infty$. The parameter β should be increased in each iteration [38], leading to a series of preconditioners similar to the MM algorithm. In contrast, the Lagrangian multiplier γ in the AL method prevents β from varying while the optimization still performs well [37].

3) The u-Subproblem: Assuming that v and γ are fixed, we minimize (7) with respect to u, which yields

$$\begin{aligned} u^{t+1} = &\arg\min_u \sum_{i \in \Omega} \frac{\lambda}{2} (u - g)_i^2 \\ &+ \sum_{p=1}^{|\mathcal{N}|} \frac{\beta}{2} \left(\mathbf{A}u - v^t - \gamma^t\right)_p^2. \end{aligned} \tag{9}$$

As the above objective is quadratic, the normal equations is linear in u and has an explicit form:

$$\left(\lambda \Lambda + \beta \mathbf{A}^T \mathbf{A}\right) u^{t+1} = \lambda \Lambda g + \beta \left(\mathbf{A}^T \left(v^t + \gamma^t\right)\right). \tag{10}$$

On the contrary to (5), the reweighting matrix $\mathbf{W}^t$ is removed in (10), but the right-hand side is changing at each iteration. Although this reversal seems trivial at first glance,

[1]The different approximations can be applied such as iterative L1 or Huber [33], but we find that a quadratic function is enough for Welsch's one and its application to depth super-resolution.

we argue that it provides a significant advantage over the MM algorithm − since β in the AL method no longer varies during iterations, the same preconditioner can be used in all iterations. In addition to this view, the linear equation (10) may be more readily solved than (5) − the reweighting matrix $\mathbf{W}^t$ inevitably increases the condition number of the linear system. Note that $\mathbf{A}^T\left(v^t+\gamma^t\right)$ to update the right-hand side can be efficiently computed by using pointwise difference.

We apply the preconditioned conjugate gradient (PCG) method [36] to update u^{t+1}. The previous solution u^t is used as an initial point for the PCG iteration. It is worth mentioning that the FFT-based solver, which is frequently used in splitting techniques [38], [39], cannot be applied to depth upsampling problem since it handles sparse data of different resolutions.

4) The v-Subproblem: Again, assuming that u and γ is fixed, we minimize (7) with respect to v, which yields

$$v^{t+1} = \arg\min_{v} \sum_{p=1}^{|\mathcal{N}|} (w_p^{\mathbf{I}}\phi(v)_p + \frac{\beta}{2}(\mathbf{A}u^{t+1} - v - \gamma^t)_p^2). \quad (11)$$

The problem of (11) is separable with respect to v_p.

$$v_p^{t+1} = \arg\min_{v_p} w_p^I\phi(v)_p + \frac{\beta}{2}(\mathbf{A}u^{t+1} - v - \gamma^t)_p^2. \quad (12)$$

Let $h(v)_p = w_p^{\mathbf{I}}\phi(v)_p$, then the v-subproblem corresponds to the proximal operator [40] of h at $(\mathbf{A}u^{t+1} - \gamma^t)_p$.[2] However, our nonconvex regularizer does not permit a closed-form solution, like to the 1D/2D shrinkage formulas for convex total variation (TV) norm [38], [39]. We minimize the v-subproblem via first order proximal approximation [40].

Proposition 1: The first order minimizer of (12) can be computed in closed form.

Proof: Since h is differentiable, its first-order approximation near some point τ_p is

$$h(v)_p \simeq h(\tau)_p + h'(\tau)_p(v-\tau)_p. \quad (13)$$

Replacing $h(v)_p$ with its linearized version (13), the proximal operator of h at τ_p is then defined as

$$\text{prox}_h(\tau_p) = \arg\min_{v_p} h'(\tau)_p v_p + \frac{\beta}{2}(v-\tau)_p^2. \quad (14)$$

Now we can get the first-order proximal approximation in closed form:

$$\text{prox}_h(\tau_p) = \tau_p - \frac{1}{\beta}h'(\tau)_p. \quad (15)$$

By letting $\tau_p = (\mathbf{A}u^{t+1} - \gamma^t)_p$, the update rule for v_p^{t+1} is given as

$$v_p^{t+1} = (1 - \frac{1}{\beta}w_p^{\mathbf{I}}w_p^{\tau})\tau_p, \quad (16)$$

where $w_p^{\tau} = \exp(-\tau_p^2/\sigma_u)$. ■

[2]We minimize the function, $h(v)_p$ with a damping term that makes v_p^{t+1} close to $(\mathbf{A}u^{t+1} - \gamma^t)_p$. This exactly corresponds to the proximal operator [40] of h at $(\mathbf{A}u^{t+1} - \gamma^t)_p$.

Algorithm 1 Fast Depth Upsampling

```
procedure UPSAMPLE(g, I, u^0)
    Compute feature vector F using (4).
    Construct N via kNN search.
    Construct A using (3).
    while ||u^{t+1} − u^t|| > τ do
        u^{t+1} ← Solve (10) using PCG iteration.
        v^{t+1} ← Update v according to (16).
        γ^{t+1} ← Update γ according to (8).
    end while
end procedure
```

We finally note that the MM iteration is actually equivalent to the gradient linearization iteration [41]. The difference is that the linearization is performed on the auxiliary variable v under the constraint, $v \approx \mathbf{A}u$, not on the primal variable u. The pseudo-code for the overall procedure is presented in Algorithm 1.

V. Analysis With Existing Approaches

In this section, we describe the relations with other approaches, and then investigate the property of the proposed method.

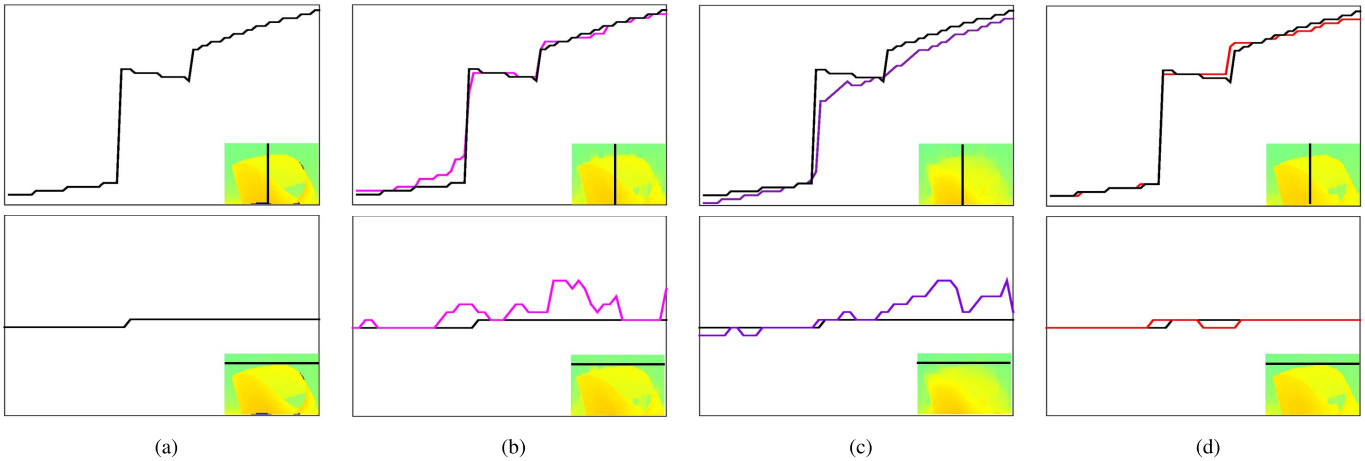
A. Comparison With Other Approaches

1) MRF-NLN Method: The MRF based method [16] is further extended in [18] by using the NLN. It utilizes several weighting terms, including intensity image, segmentation, and bicubic interpolated depth. The objective function in [18] is defined as:

$$E_{NLN}(u) = \sum_{i\in\Omega}\frac{\lambda}{2}(u-g)_i^2 + \sum_{p=1}^{|\mathcal{N}_{NLN}|} w_p^{\mathbf{c}}(\mathbf{A}u)_p^2, \quad (17)$$

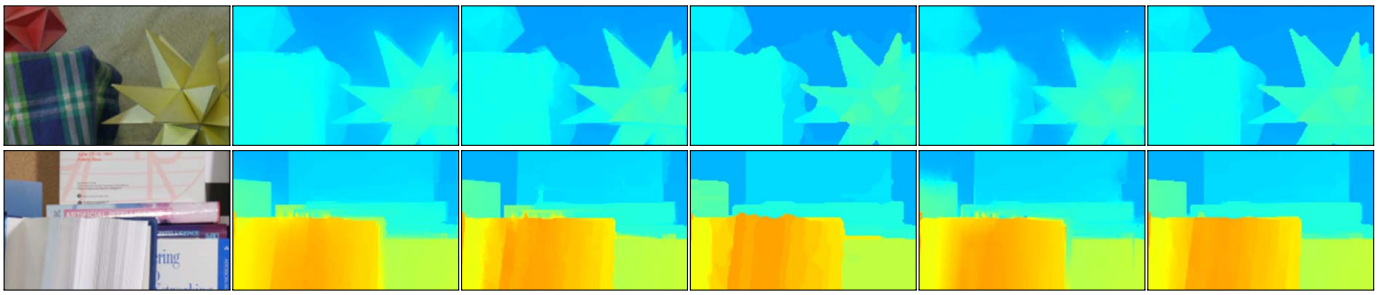
where $w^{\mathbf{c}}$ is the concatenation of weights mentioned above. $\mathcal{N}_{NLN}$ is the set of NLN. This problem is quadratic and therefore the minimizer can be computed in closed form. The MRF-NLN [18] utilizes weighting term derived from bicubic interpolated depth, which may be effective to remove depth bleeding and texture copying artifacts. This heuristic scheme may improve the performance of upsampling, but the proper weights are difficult to derive analytically when the depth measurement is highly noisy. In contrast, our model uses the robust regularizer ϕ rather than quadratic function, and this makes our scheme independent to pre-determined weights derived from the initial estimates (see Section IV). Note that the MRF-NLN scheme is similar to the first iteration of the MM algorithm [7].

The MRF-NLN also uses color segmentation to describe the weights. However, the segmentation label is not effective to handle depth bleeding artifacts (see Figs. 4 and 5). The depth bleeding problem typically appears in the regions where the foreground and the background have similar intensity values, and segmenting these regions precisely is not trivial.



(a) (b) (c) (d)

Fig. 4. Comparison of the proposed model and existing approaches on a 1D signal: (a) the ground truth, (b) MRF-NLN [18], (c) ATGV [20], and (d) the proposed method. Our model does not smooth depth discontinuities even when the corresponding intensity values are similar, and eliminates the perturbation caused by the texture pattern (see Fig. 1).



(a) (b) (c) (d) (e) (f)

Fig. 5. Visual comparison of upsampled (×8) depth images in Middlebury dataset [43]: (a) the HR color image, (b) MRF [16], (c) MRF-NLN [18], (d) NWMEF [14], (e) ATGV [20], and (f) our result. The proposed model takes both structures of intensity and depth images into consideration, preserving only necessary details and edges.

2) WMOF Method: The WMOF [12] computes mode from a (weighted) histogram generated by a set of depth measurements. This method can be defined in the minimization form:

$$E_{WMOF}(u_i) = \sum_{j \in (N_{SL}(i) \cap \Omega)} w^{\mathbf{I}}_{i,j} w^{s}_{i,j} \phi(u_i - g_j), \quad (18)$$

where $w^{\mathbf{I}}_{i,j} = \exp(-\left\|\mathbf{I}_i - \mathbf{I}_j\right\|^2/\sigma_{\mathbf{I}})$. w^s is a spatial weight to lessen the influence of distant values and $u_i \in \{1, 2, \ldots, l\}$, i.e., discrete labels with a maximum value l. $\mathcal{N}_{SL}(i)$ is semi-local neighborhoods of the pixel i, e.g., 7×7. From (18), we can find that the WMOF employs the same robust function ϕ in a local formulation. This improves the depth accuracy and represents the robustness. However, some artifacts are still observed since the WMOF considers the local structure only as in (18). It does not maintain the global consistency of the upsampled depth image. Furthermore, the WMOF requires a finite set of discrete labels as in the cost aggregation methods [11]–[21]. The computational costs grow with the number of discrete labels and the output depth image is inherently quantized. In contrast, the proposed model does not need discretization in the depth domain.

3) Relation With Gaussian Mixture Model: Let us consider the proposed regularization penalty in maximization form. It then yields the following Gaussian mixture model (GMM):

$$E^{\max}_{s}(u) = \sum_{p=1}^{|\mathcal{N}|} w^{\mathbf{I}}_{p} \psi(\mathbf{A}u)_p, \quad (19)$$

where $\psi(z) = \exp\left(-z^2/\sigma\right)$. The prior (19) on the latent depth distribution is amount to a mixture of zero-mean Gaussians. Thus, our model may be interpreted as the GMM approximation to the distribution of the HR depth image. The GMM is used in many applications, such as deblurring [30], matting [31], segmentation [32]. It usually needs an additional parameter in a estimation step. However, in our case the mixture coefficients, $w^{\mathbf{I}}_{p}$ are directly computed from high quality intensity images. Note that the expectation-maximization (EM) algorithm which commonly used to solve the GMM model, is a special case of a general class of MM optimization [34].

B. Behavioral Properties of the Proposed Method

From (5) and (10), it is straightforward to see that our model determines the regularity of the output depth image via multiplication (with reweighting (22) in the MM) or addition (with shrinkage (16) in the ADMM). The structures of intensity and depth images are jointly considered in both operations. In other words, we iteratively compensate the

inherent structure discrepancy between two images. This is a noticeable difference with traditional approaches whose affinity depends on appearances solely [16], [17]. For example, if nearby two objects share analogous intensity, the edge blurring is frequently occurred in the conventional approaches. However, our model discourages the influence of the other side in different range. These examples are illustrated in Fig. 4(top). Our model does not blur depth discontinuities even when the corresponding intensity image has low-contrast edges. Also, for the textured region in an intensity image, our model gradually eliminates the perturbation caused by the texture pattern. This property is illustrated in Fig. 4(bottom).

VI. Experimental Results

We verify the performance of the proposed model through various experiments, including real-world examples. Our model is first evaluated on Middlebury benchmark sequences [43] which provide HR intensity images and ground truth depth. The evaluation is also performed on depth images from the ToF sensor and Microsoft's Kinect 2.

A. Experimental Environment

The proposed model is implemented with Matlab using the pcg function and our ADMM solver. We use the FLANN [28] library for efficient kNN searches. All experiments are performed on Intel i7 3.4GHz CPU and 8-GB RAM. The proposed model needs six parameters, $\sigma_u = 15$ and $\sigma_I = 15$ for the bandwidths, λ for the regularization, k=16, $\alpha = \eta = 0.1$ for the kNN search algorithm. All parameters are fixed, but λ is tuned according to the amount of noise. The experiments are designed to compare the proposed method with state-of-the-art methods: WMEF [13], WMOF [12], NWMEF [14], JTF [15], MRF [16], MRF-NLN [18], ATGV [20], and JSM [22]. The first four methods are local filtering-based approaches and the others are global optimization-based ones. The results have been obtained using source codes provided by the authors, and parameters have been adjusted through intensive experiments. For the MRF [16] and JTF [15], we use our own implementation with the optimal parameter adjustment.

For quantitative evaluation, PSNR and Bad Matching Percentage ($\mathcal{BMP}$) [42] are measured between the upsampled depth image and the ground truth. $\mathcal{BMP}$ is defines as:

$$\mathcal{BMP} = \frac{1}{T}\sum_i (|u_i - \mathcal{G}_i| > \delta), \tag{20}$$

where T is the total number of pixels, $\mathcal{G}$ is the ground truth, and δ is the threshold for bad matched pixels.

B. Evaluation With Synthetic Examples

We first evaluate the proposed model with the Middlebury benchmark dataset [43]. It provides HR intensity images and corresponding depth images. The LR depth image is synthesized by downsampling the ground-truth depth.

The results ($\times 8$) on the Middlebury benchmark are shown in Fig. 5 for visual comparison. It clearly illustrates the effects of structure discrepancy between intensity and depth images. The regions of low contrast in the color image (Fig. 5(top)), inevitably smooth out depth edges. It can be seen that competing methods are not able to discriminate two different objects with similar intensity, resulting in depth bleeding artifacts. In contrast, the proposed model avoids such a problem, preserving sharp depth transitions. Similarly, the depth images from the competing methods are highly perturbed by the regions of high contrast (Fig. 5(bottom)), while those obtained from the proposed model produce sharp transitions without noticeable artifacts. Most of the existing methods cannot handle the differences in structure. That is, they do not distinguish the depth discontinuities from texture and weak edges of the intensity image. The MRF formulation [16] explicitly transfers the intensity structure to depth image, and often causes texture copying and depth bleeding artifacts. The ATGV model [20] does not eliminate these artifacts completely. On the other hand, the proposed model successfully compensates the discrepancy, and thus is free from such artifacts. Following [23], we plot the gradient distributions of upsampled depth images in Fig. 6. First order depth gradients along x- and y-axis are computed from the upsampled depth images. These results are consistent with our analysis that the proposed model can be regarded as the GMM approximation to fit the distributions of natural depth images. The results obtained from the proposed model show the most similar statistical distributions to ground truth one. The competing algorithms have a relatively low probability in large gradients, which means that they tend to smooth or eliminate depth discontinuities, and produce depth bleeding artifacts. The quantitative evaluation is summarized in Table I. The accuracy is measured by $\mathcal{BMP}$ with a threshold $\delta = 1$ and by PSNR. The top two methods yielding low $\mathcal{BMP}$ or high PSNR are marked with a shadow. As expected, our model outperforms other competing methods.

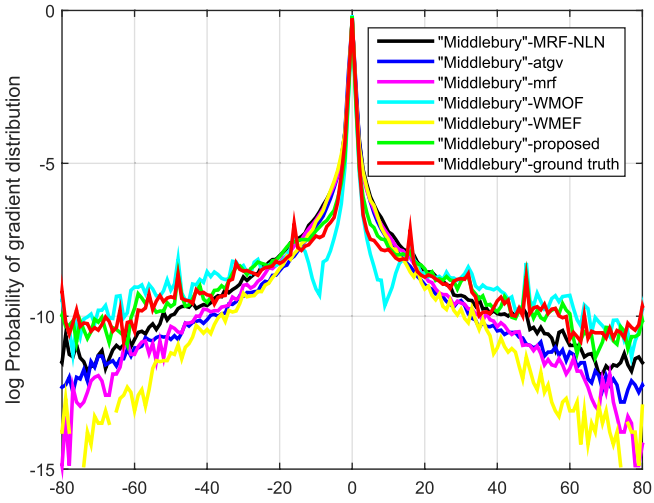


Fig. 6. Log-probability distributions of upsampled depth images from the Middlebury dataset [43]. The results from the proposed model shows a good approximation to the ground truth distribution, which contains primarily zero gradients but a few gradients have large magnitudes.

C. Evaluation With Real-World Examples

1) ToF Sensors: We use the "*Mesa Imaging SR4000*" [45] depth sensor with the "*Point Grey Flea*" color camera [12]. They provide LR depth images (176 × 144) and

TABLE I

QUANTITATIVE EVALUATION OF DEPTH UPSAMPLING METHODS ON THE MIDDLEBURY DATASET [43]

	$\mathcal{BMP}$ ($\delta = 1$)						PSNR					
	Art		*Cone*		*Moebius*		*Art*		*Cone*		*Moebius*	
	×8	×16	×8	×16	×8	×16	×8	×16	×8	×16	×8	×16
Bilinear	34.59	65.72	14.60	42.72	17.15	43.04	25.38	22.51	29.58	27.06	30.76	27.31
MRF [16]	28.37	34.08	22.43	38.46	28.27	39.05	26.58	23.73	31.71	29.15	34.14	29.03
WMEF [13]	17.98	24.77	10.69	16.26	10.52	16.88	27.33	23.95	31.86	29.59	35.38	31.37
MRF-NLN [18]	13.08	23.34	11.53	18.03	10.14	15.68	27.83	24.59	32.73	29.45	35.39	32.61
NWMEF [14]	9.71	22.64	8.10	17.82	8.13	18.60	26.85	23.11	32.26	28.67	34.62	28.49
JTF [15]	12.73	25.27	9.87	19.67	9.58	19.91	27.57	24.54	32.21	29.22	34.76	29.18
WMOF [12]	11.42	20.25	9.34	15.77	8.56	15.41	26.65	23.62	30.87	28.21	34.06	29.70
ATGV [20]	23.32	34.48	22.63	28.90	17.74	26.98	25.10	24.27	29.11	28.25	31.63	30.67
JSM [22]	22.23	30.97	15.47	22.56	16.02	25.68	25.93	23.81	30.39	27.59	33.64	29.95
Proposed method	8.38	19.41	9.23	13.94	6.74	13.94	28.38	25.57	32.59	30.74	36.29	32.89

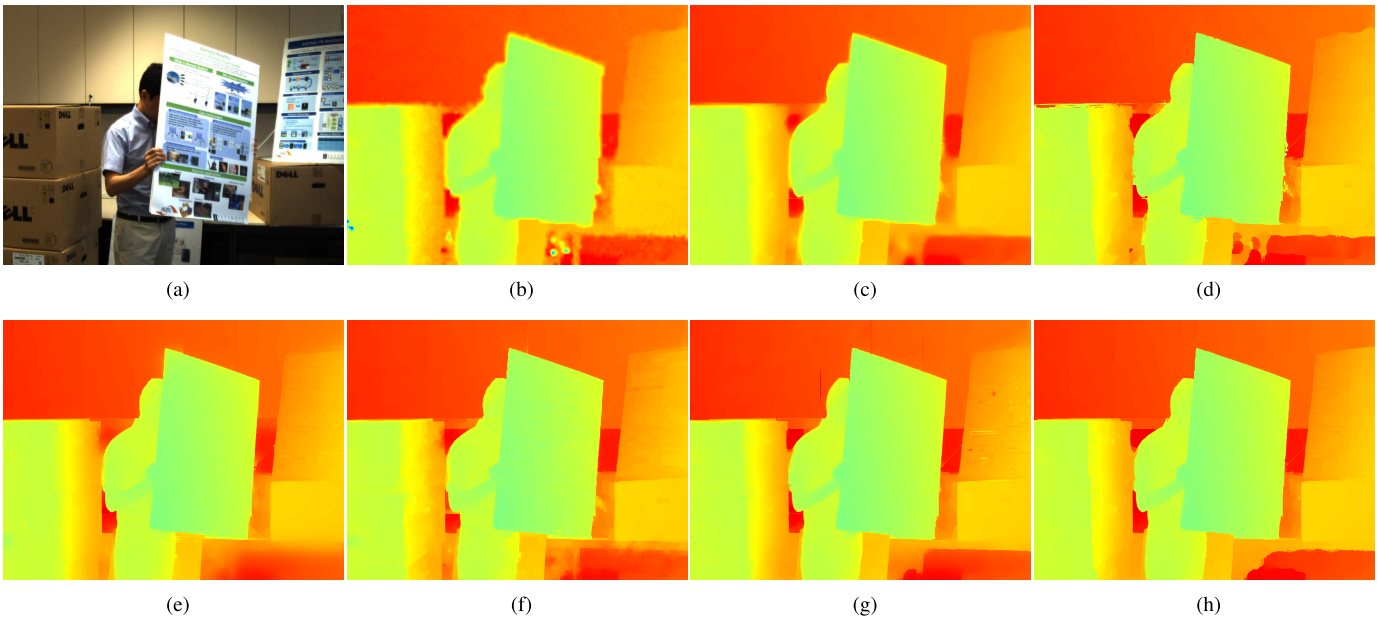
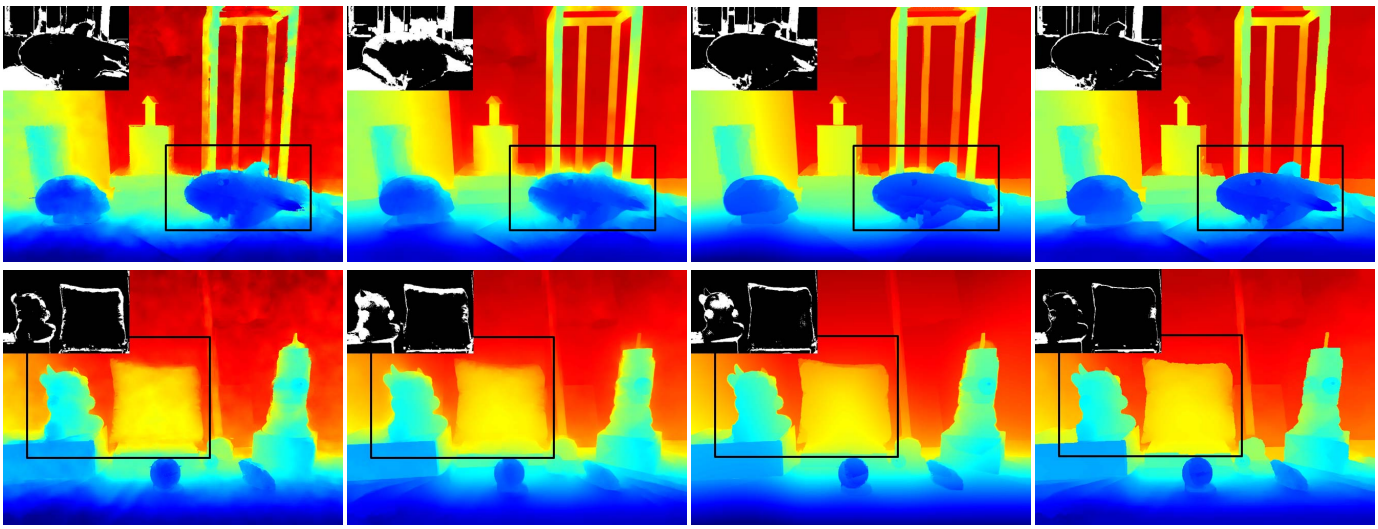


Fig. 7. Depth upsampling results with the ToF camera: (a) the HR intensity image, (b) Bilinear interpolation, (c) WMEF [13], (d) WMOF [12], (e) MRF [16], (f) MRF-NLN [18], (g) ATGV [20], and (h) the proposed method. Our results shows sharp discontinuities and smooth slanted surfaces without artifacts.

corresponding HR intensity images (1024 × 768), respectively. The depth images, obtained by the sensor, are normalized between $[0, 255]$ and are projected to the intensity image space with calibration parameters. Fig. 7 shows the upsampling results of (b) Bilinear interpolation, (c) WMEF [13], (d) WMOF [12], (e) MRF [16], (f) MRF-NLN [18], (g) ATGV [20], and (h) the proposed model. It shows that the upsampling result by bilinear interpolation contains considerable outliers, blue and green dots. While most methods reduce sensor noise relatively well, some artifacts are still observed in the upsampling results, degrading the overall quality of the depth image. The WMOF [12] shows relatively sharp depth transitions, but suffers from depth leakage artifacts. The ATGV [20] occasionally produces undesirable results in highly textured regions, e.g., the board on the box in Fig. 7(g). On the contrary, our method shows the convincing result without notable artifacts. It is superior to recover sharp depth discontinuities and slant planar surfaces.

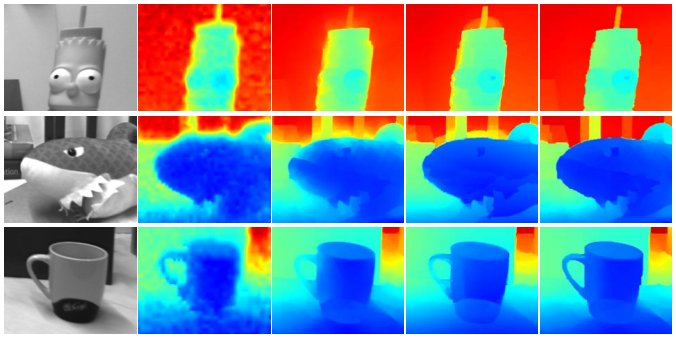
Recently, Ferstl *et al* have introduced a GRAZ dataset [20]: three intensity images (810×610), "*Books*", "*Devil*", "*Shark*" and their corresponding LR depth images (in mm, 160×120) captured by a "*PMD Nano*" ToF sensor [44]. This dataset additionally provides ground-truth depth images from a high accuracy structured light scanner, which enables performing quantitative evaluation. We normalize the depth range to a predefined depth range $[0, 255]$ for the performance evaluation with the competing methods [13], [12], [14]. The quantitative evaluation on the GRAZ dataset is shown in Table II. The $\mathcal{BMP}$ is measured with a threshold $\delta = 255 \times 0.05$. This table demonstrates that the proposed model outperforms the state of the art. Additionally, we visualize the bad matching regions of "*Devil*" and "*Shark*" sequences in Fig. 8. Our method shows much less error than others, especially in depth discontinuities.

As stated earlier, the MRF-NLN [18] uses bicubic interpolated depth to describe the weights. However, the proper weights are often not available for challenging



(a) (b) (c) (d)

Fig. 8. The bad matching regions of *Shark* and *Devil* in the GRAZ dataset [20]: (a) WMOF [12], (b) MRF-NLN [18], (c) ATGV [20], and (d) the proposed method. The proposed method is superior to other state-of-the-art methods, particularly near depth discontinuities.



(a) (b) (c) (d) (e)

Fig. 9. Influence of the depth sensing error around object boundaries: (a) Snippets of the HR intensity images from "*Devil*", "*Shark*", "*Books*", (b) Bilinear interpolation, (c) MRF-NLN [18], (d) ATGV [20], and (e) the proposed method. Our model is stable in dealing with depth discontinuities, while other methods are significantly influenced by the sensing error.

TABLE II

QUANTITATIVE EVALUATION OF DEPTH UPSAMPLING METHODS ON THE GRAZ DATASET [20]

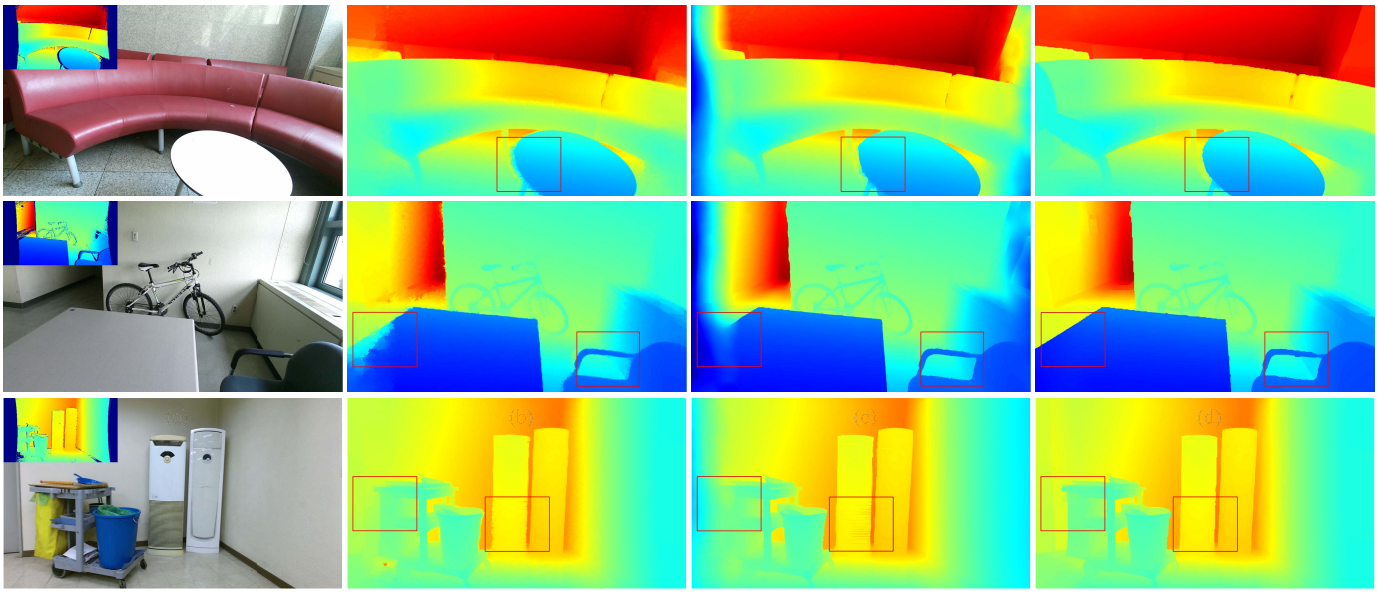
	$\mathcal{BMP}$ ($\delta = 255 \times 0.05$) / PSNR					
	Books		*Devil*		*Shark*	
Bilinear	17.96	25.74	16.63	26.34	22.34	23.93
MRF [16]	17.18	25.94	14.99	26.73	21.83	24.31
WMEF [13]	15.13	26.21	11.23	26.96	16.77	25.10
MRF-NLN [18]	16.54	26.15	13.09	27.14	19.06	24.93
NWMEF [14]	15.15	25.15	23.03	23.51	24.11	22.86
JTF [15]	16.05	25.95	12.16	27.12	16.81	24.85
WMOF [12]	13.52	26.12	11.43	27.30	14.52	25.19
ATGV [20]	14.68	26.34	10.92	27.16	15.66	25.44
JSM [22]	17.34	25.90	14.84	27.06	18.29	25.11
Proposed method	11.87	26.64	9.03	27.57	13.74	25.59

real measurements. On the contrary, our method is not specified on pre-determined weights, and thus the quality of the estimations is independent of these weights as shown in Fig. 9. The typical ToF sensor has a difficulty in measuring accurate depth information of object boundaries. That is, the background information is penetrating into the foreground or the foreground information is fattening as shown in Fig. 9(b). The MRF-NLN [18] does not properly consider the underlying structure of depth images appropriately in these cases. Although most methods handle impulsive sensor noise well, some sensing errors are propagated (see Fig. 9(c) and (d)). In contrast, the proposed method handles such errors well.

2) Microsoft's Kinect 2: We also use the Kinect 2 camera which is recently commercialized by Microsoft. On the contrary to the Kinect 1, it uses ToF technology to obtain depth information. The depth images, however, still show holes along object boundaries, and have low resolution. The captured depth and intensity images are of size 512×424 and 1920×1080, respectively. For registration, we use the calibration toolbox provided by Microsoft's SDK. Figure 10 shows the upsampling ($\times 3.75$) results. We can see that the raw depth images contain many holes, and are not perfectly aligned to the corresponding intensity images. Most existing approaches are not particularly suitable for handling the registration error, producing spurious depth discontinuities. For example, although we perform three thousands of iterations (taking about 1 hour) for the ATGV [20], it does not properly handle registration error and large holes. In contrast, our model can successfully recover sharp discontinuities and smooth surfaces, even in the presence of outliers and missing data. This confirms that our method is versatile for both depth upsampling and completion problems.

D. Influence of the Parameters

There are six parameters in the proposed method, including the number of neighborhoods k and the bandwidth for regularizer σ_u. Here, we investigate the effect of these parameters on the resulting depth images. The parameters $\lambda = 10$ and $\sigma_I = 15$ are fixed in this experiment. We observe that the



(a) (b) (c) (d)

Fig. 10. Examples of upsampled depth images with the Kinect 2 camera: (a) HR intensity and the corresponding LR depth images, (b) MRF-NLN [18], (c) ATGV [20], and (d) the proposed method. The proposed model gives accurately localized depth boundaries and resolves ambiguities in the upsampling problem. Please see the highlighted areas.

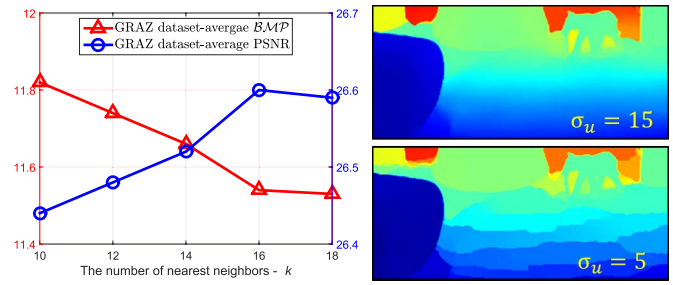


Fig. 11. Influence of the parameters on the resulting depth images: (Left) average $\mathcal{BMP}$ and PSNR according to the number of neighbors k, (Right) the upsampled depth image with different values of σ_u.

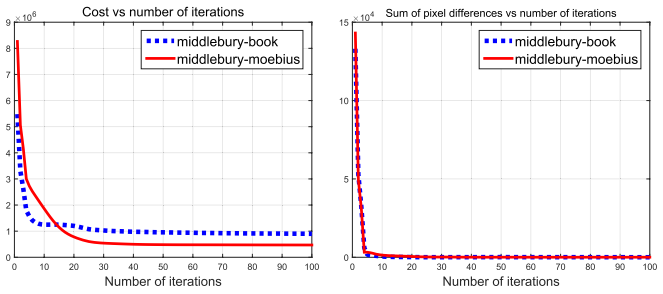


Fig. 12. The behavior of Algorithm 2: (Left) the decay of the objective function (1) according to the number of iterations, (Right) the sum of pixel difference between successive iterations, $\|u^{t+1} - u^t\|_1$. Our solver monotonically decreases the energy values and the sum of pixel difference between successive iterations. We use $\sigma_I = 15$, $\sigma_u = 15$ and k=16 for nearest neighbors.

performance improves slightly when k is increased from 10 to 16 (Fig. 11(left)). However, increasing k larger than 16 decreases the PSNR performance. In general, k(=16) works well and is kept in our implementation. The parameter σ_u determines a degree of smoothness in the upsampling result by adjusting the shape of the regularizer ϕ. If σ_u is set to be too small, the result is quantized, which is undesirable for smooth surfaces (Fig. 11(right)). This is expected since decreasing σ_u will result in the effect similar to the zero-norm minimization. Thus, σ_u should be set properly to balance the smoothness and sparseness in the result. All results in this paper are obtained with $\sigma_u = 15$.

E. Convergence and Complexity Analysis

We show the convergence behavior of our solver with $u^0 = \mathbf{0}$, a constant vector.[3] For this example, we consider the upsampling ($\times 8$) problem of the "*book*" and "*moebius*" sequence from [43]. The objective value and the sum of pixel difference, according to the number of iterations, are plotted in Fig. 12. Our solver rapidly reduces the objective value and converges to the stationary point as shown in Fig. 12(right). The per-pixel difference decreases monotonically (Fig. 12(left)), and 30-40 iterations are enough to get satisfactory results.

Empirically, we find that the MM algorithm minimizes the objective function of (1), using the smaller number of iterations than ours based on the AL method. The MM method typically requires 20 iterations for depth upsampling. However, the overall computational time used by ours is less than that used by the MM iteration, due to the property of the AL method. Table III summarizes the computational time in seconds for each minimization method. The "*moebius*" [43] and "*devil*" [20] sequences are used for this test. The direct solver in MATLAB is used for the MM. We also compare the condition numbers of the matrices $\left(\lambda\Lambda + \mathbf{A}^T\mathbf{W}^t\mathbf{A}\right)$ and $\left(\lambda\Lambda + \beta\mathbf{A}^T\mathbf{A}\right)$ arising from (5) and (10). As expected,

[3]The "warm"-initialization of u^0 using existing methods may further improve the convergence property.

TABLE III
COMPARISON OF THE COMPUTATIONAL COMPLEXITY BETWEEN THE MM AND OUR SOLVER. $\mathcal{N}_4$ DENOTES A 4-NEIGHBORHOOD SYSTEM

	Image size	Neighborhood system	iteration	factor	Time(s)	Avg. condition number	$\mathcal{BMP}$
MM [7]	370×463	$\mathcal{N}_4$ / $\mathcal{N}_{kNN}$	20	$\times 8$	9.83 / 16.97	3.3440e+10	7.92 / 6.79
	610×810			$\times 6.25$	28.83 / 58.68		10.27 / 9.14
Our solver	370×463	$\mathcal{N}_4$ / $\mathcal{N}_{kNN}$	40	$\times 8$	1.21 / 2.94	2.9771e+3	7.85 / 6.74
	610×810			$\times 6.25$	4.11 / 9.81		10.39 / 9.03

TABLE IV
COMPARISON OF THE COMPUTATIONAL COMPLEXITY BETWEEN THE PROPOSED METHOD AND OTHER OPTIMIZATION-BASED APPROACHES

	Setup	Optimization	Total time(s)
ATGV [20]	·	1456	1456
MRF-NLN [18]	89.69	6.82	96.51
Proposed method	4.58	15.61	20.19

the condition number of (10) is always smaller than that of (5) − the relative errors of upsampled depth images are proportional to the relative noise and the condition number [19].

F. Runtime Comparison

We compare the runtime of global optimization-based approaches including ours in Table IV. The LR depth image (176×144) and the HR intensity image (1024×768) [12] are used for this comparison. In [20], a similar alternating minimization algorithm is used based on primal-dual method, but the algorithm is not appropriate for a large-scale optimization problem. This algorithm requires three thousands of iterations to converge and takes 1400 seconds, which is computationally demanding. Although the MRF-NLN method [18] allows a closed-form solution, the computational cost is quite huge. It spends most of the time in the setup stage for constructing a NLN system. Our method needs 4 seconds for the kNN neighborhoods and solves the optimization problem quickly.

VII. CONCLUSION

In this paper, we have described the nonconvex model for high quality depth upsampling. We have shown that the proposed model handles the structural discrepancy between intensity and depth image. This property makes it possible to distinguish depth transitions from texture and weak edges of the intensity images, solving depth bleeding and texture copying artifacts. We have also derived a fast ADMM solver which is considerably faster than the conventional MM implementation for depth upsampling. Our model has been compared with several state-of-the-art methods through intensive experiments. Experimental results have demonstrated that our model outperforms previous ones in various datasets.

APPENDIX

We present the MM iteration [7], [23] for solving (1). For any $(\mathbf{A}u)_p$ and $(\mathbf{A}\bar{u})_p$ in $(-\infty, \infty)$, Welsch's function can be majorized by a quadratic proxy function [29]:

$$\phi(\mathbf{A}u)_p \leq \frac{\phi'(\mathbf{A}\bar{u})_p}{2(\mathbf{A}\bar{u})_p}(\mathbf{A}u)_p^2 + b \quad \text{with equality only if } (\mathbf{A}u)_p = (\mathbf{A}\bar{u})_p, \tag{21}$$

where b is an additive constant which can be ignored in the minimization process. Sequentially, by using this inequality and the fact that $w^{\mathbf{I}} > 0$, we majorize the regularization penalty $E_s(u)$ in (2) at u^t:

$$E_s(u) \leq \sum_{p=1}^{|\mathcal{N}|} w_p^{\mathbf{I}} w_p^{u^t} (\mathbf{A}u)_p^2, \tag{22}$$

where $w_p^{u^t} = \exp(-(\mathbf{A}u^t)_p^2/\sigma_u)$. Thus, the complete quadratic upper bound is obtained by adding the data consistency term as follows:

$$E(u) \leq E_{pro}\left(u \,|u^t\right) = \sum_{i \in \Omega} \frac{\lambda}{2}(u-g)_i^2 + \sum_{p=1}^{|\mathcal{N}|} w_p^{\mathbf{I}} w_p^{u^t} (\mathbf{A}u)_p^2. \tag{23}$$

Then, an iterative MM algorithm for solving (1) is given by:

$$u^{t+1} = \arg\min_u E_{pro}\left(u \,|u^t\right). \tag{24}$$

REFERENCES

[1] S. Gupta, P. Arbeláez, and J. Malik, "Perceptual organization and recognition of indoor scenes from RGB-D images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 564–571.

[2] Q. Chen and V. Koltun, "A simple model for intrinsic image decomposition with depth cues," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 241–248.

[3] C. Lang, T. V. Nguyen, H. Katti, K. Yadati, M. Kankanhalli, and S. Yan, "Depth matters: Influence of depth cues on visual saliency," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2012, pp. 101–115.

[4] O. M. Aodha, N. D. F. Campbell, A. Nair, and G. J. Brostow, "Patch based synthesis for single depth image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2012, pp. 71–84.

[5] M. Kiechle, S. Hawe, and M. Kleinsteuber, "A joint intensity and depth co-sparse analysis model for depth map super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1545–1552.

[6] H. Kwon, Y.-W. Tai, and S. Lin, "Data-driven depth map refinement via multi-scale sparse representation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 159–167.

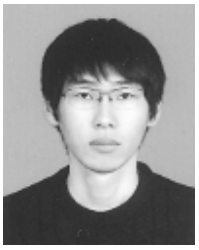
[7] Y. Kim, S. Choi, C. Oh, and K. Sohn, "A majorize-minimize approach for high-quality depth upsampling," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 392–396.

[8] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," *ACM Trans. Graph.*, vol. 26, no. 3, pp. 96–100, Jul. 2007.

[9] K. He, J. Sun, and X. Tang, "Guided image filtering," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2010, pp. 1–14.

[10] M.-Y. Liu, O. Tuzel, and Y. Taguchi, "Joint geodesic upsampling of depth images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 169–176.

[11] Q. Yang *et al.*, "Fusion of median and bilateral filtering for range image upsampling," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4841–4852, Dec. 2013.
[12] D. Min, J. Lu, and M. N. Do, "Depth video enhancement based on weighted mode filtering," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1176–1190, Mar. 2012.
[13] Z. Ma, K. He, Y. Wei, J. Sun, and E. Wu, "Constant time weighted median filtering for stereo matching and beyond," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 49–56.
[14] Q. Yang, "Stereo matching using tree filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 4, pp. 834–846, Apr. 2015.
[15] K.-H. Lo, Y.-C. F. Wang, and K.-L. Hua, "Joint trilateral filtering for depth map super-resolution," in *Proc. IEEE Vis. Commun. Image Process.*, Nov. 2013, pp. 1–6.
[16] J. Diebel and S. Thrun, "An application of Markov random fields to range sensing," in *Proc. NIPS*, vol. 18. Dec. 2005, pp. 291–298.
[17] B. Ham, D. Min, and K. Sohn, "A generalized random walk with restart and its application in depth up-sampling and interactive segmentation," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2574–2588, Jul. 2013.
[18] J. Park, H. Kim, Y.-W. Tai, M. S. Brown, and I. S. Kweon, "High-quality depth map upsampling and completion for RGB-D cameras," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5559–5572, Dec. 2014.
[19] J. Yang, X. Ye, K. Li, C. Hou, and Y. Wang, "Color-guided depth recovery from RGB-D data using an adaptive autoregressive model," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3443–3458, Aug. 2014.
[20] D. Ferstl, C. Reinbacher, R. Ranftl, M. Ruether, and H. Bischof, "Image guided depth upsampling using anisotropic total generalized variation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 993–1000.
[21] B. Ham, D. Min, and K. Sohn, "Depth superresolution by transduction," *IEEE Trans. Image Process.*, vol. 24, no. 5, pp. 1524–1535, May 2015.
[22] X. Shen, Q. Yan, L. Xu, L. Ma, and J. Jia, "Multispectral joint image restoration via optimizing a scale map," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 12, pp. 2518–2530, Dec. 2015.
[23] B. Ham, M. Cho, and J. Ponce, "Robust image filtering using joint static and dynamic guidance," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 4823–4831.
[24] W. Hu, G. Cheung, X. Li, and O. Au, "Depth map super-resolution using synthesized view matching for depth-image-based rendering," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops*, Jul. 2012, pp. 605–610.
[25] L. Grady, "Random walks for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1768–1783, Nov. 2006.
[26] M. J. Black, G. Sapiro, D. H. Marimont, and D. Heeger, "Robust anisotropic diffusion," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 421–432, Mar. 1998.
[27] R. Ranftl, K. Bredies, and T. Pock, "Non-local total generalized variation for optical flow estimation," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2014, pp. 439–454.
[28] M. Muja and D. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," in *Proc. Int. Conf. Compu. Vis. Theory Appl.*, 2009, pp. 311–340.
[29] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. Image Process.*, vol. 6, no. 2, pp. 298–311, Feb. 2010.
[30] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 787–794, 2006.
[31] J. Wang and M. F. Cohen, "An iterative optimization approach for unified image segmentation and matting," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2. Dec. 2005, pp. 936–943.
[32] C. Rother, V. Kolmogorov, and A. Blake, "Interactive foreground extraction using iterated graph cuts," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 309–314, Aug. 2004.
[33] P. Ochs, A. Dosovitskiy, T. Brox, and T. Pock, "On iteratively reweighted algorithms for nonsmooth nonconvex optimization in computer vision," *SIAM J. Image Sci.*, vol. 8, no. 1, pp. 331–372, Feb. 2015.
[34] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *Amer. Statist.*, vol. 58, no. 1, pp. 30–37, 2004.
[35] D. Krishnan and R. Szeliski, "Multigrid and multilevel preconditioners for computational photography," *ACM Trans. Graph.*, vol. 30, no. 6, Dec. 2011, Art. no. 177.
[36] D. Krishnan, R. Fattal, and R. Szeliski, "Efficient preconditioning of laplacian matrices for computer graphics," *ACM Trans. Graph.*, vol. 32, no. 4, Jul. 2013, Art. no. 142.
[37] J. Nocedal and S. Wright, *Numerical optimization*. 2nd ed. New York, NY, USA: Springer-Verlag, 2006.
[38] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM J. Imag. Sci.*, vol. 1, no. 3, pp. 248–272, Aug. 2008.
[39] M. V. Afonso, J.-M. Bioucas-Dias, and M. A. T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2345–2356, Sep. 2010.
[40] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 123–231, 2013.
[41] M. Nikolova and R. H. Chan, "The equivalence of half-quadratic minimization and the gradient linearization iteration," *IEEE Trans. Image Process.*, vol. 16, no. 6, pp. 1623–1627, Jun. 2007.
[42] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, nos. 1–3, pp. 7–42, Apr. 2002.
[43] *Middlebury Benchmark*, accessed on Jan. 5, 2015. [Online]. Available: http://vision.middlebury.edu/stereo/
[44] *PMD Technology*, accessed on Jan. 5, 2015. [Online]. Available: http://www.pmdtec.com/
[45] *MESA Imaging*, accessed on Feb. 15, 2015. [Online]. Available: http://www.mesa-imaging.ch/
[46] *PointGrey*, accessed on Feb. 2, 2015. [Online]. Available: http://www.ptgrey.com/



Youngjung Kim (S'14) received the B.S. degree in electrical and electronic engineering from Yonsei University, Seoul, South Korea, in 2013, where he is currently pursuing the joint M.S. and Ph.D. degrees in electrical and electronic engineering. His current research interests include variational method and optimization, both in theory and applications in computer vision and image processing.

Bumsub Ham (M'14) received the B.S. and Ph.D. degrees from the School of Electrical and Electronic Engineering, Yonsei University, Seoul, South Korea, in 2008 and 2013, respectively. From 2014 to 2016, he was a Post-Doctoral Research Fellow with the Willow Team of INRIA Rocquencourt, École Normale Supérieure de Paris, and the Centre National de la Recherche Scientifique. Since 2016, he has been an Assistant Professor with the School of Electrical and Electronic Engineering, Yonsei University. His current research interests include computer vision, computational photography, and machine learning, in particular, graph matching and superresolution, both in theory and applications.

Changjae Oh (S'13) received the B.S. and M.S. degrees in electrical and electronic engineering from Yonsei University, Seoul, South Korea, in 2011 and 2013, respectively, where he is currently pursuing the Ph.D. degree. His current research interests include 3D computer vision, 3D visual fatigue assessment, and image segmentation.

Kwanghoon Sohn (M'92–SM'12) received the B.E. degree in electronic engineering from Yonsei University, Seoul, South Korea, in 1983, the M.S.E.E. degree in electrical engineering from the University of Minnesota, Minneapolis, MN, USA, in 1985, and the Ph.D. degree in electrical and computer engineering from North Carolina State University, Raleigh, NC, USA, in 1992. He was a Senior Member of the Research Staff with the Satellite Communication Division, Electronics and Telecommunications Research Institute, Daejeon, South Korea, from 1992 to 1993, and a Post-Doctoral Fellow with the MRI Center, Medical School of Georgetown University, Washington, DC, USA, in 1994. He was a Visiting Professor with Nanyang Technological University, Singapore, from 2002 to 2003. He is currently a Professor with the School of Electrical and Electronic Engineering, Yonsei University. His research interests include 3D video processing, computer vision, and video communication.