

# Depth map enhancement using adaptive moving least squares method with a total variation minimization

Sang Min Yoon<sup>1</sup> · Jungho Yoon<sup>2</sup>

Received: 16 March 2015 / Revised: 16 August 2015 / Accepted: 20 August 2015 /

Published online: 2 September 2015

© Springer Science+Business Media New York 2015

**Abstract** Accurate and fast depth map acquisition and enhancement is an important issue in the area of computer vision and image processing. In this study, we present a novel method for enhancing noisy depth maps using adaptive total variation minimization, which facilitates noise smoothing and boundary sharpening for a given depth map image but without previous information. We filter the noise in the depth map with a refined total variation minimization technique. Our experimental results demonstrate that the proposed method outperforms other competitive methods in both objective and subjective comparisons of depth map enhancement and denoising.

**Keywords** Image denoising · Total variation minimization

## 1 Introduction

A depth map is defined as an image that contains information related to the distance of object surfaces from a viewpoint. Depth maps are widely used in many techniques including robotics, three-dimensional (3D) television, and interactive view interpolation. Passive stereo and active depth sensors are employed in many applications to facilitate the rapid acquisition of real-time depth maps of dynamic scenes [10, 26]. Thus, the development of an autonomous system capable of understanding the shape and location of a target object within a depth map is an active research area in the field of computer vision and image processing [10, 11, 19, 26]. Kinect, which was designed by Microsoft for computer gaming, is a popular alternative to expensive laser scanners in video surveillance, robotics, and forensics applications [27]. Kinect sensors provide depth and color images simultaneously at frame rates up to 30 fps. The

---

✉ Sang Min Yoon  
sangmin.yoon@gmail.com

Jungho Yoon  
yoon@ewha.ac.kr

<sup>1</sup> School of Computer Science, Kookmin University, 77, Jeongneung-ro, Sungbuk-gu, Seoul 136-702, South Korea

<sup>2</sup> Department of Mathematics, Ewha Womans University, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul 120-750, South Korea

integration of depth and color data yields a colored point and each frame may contain 300,000 points. The characteristics of the data captured by Kinect sensors have attracted the attention from other fields including mapping and 3D modeling. However, the high-level surface geometry must be inferred from noisy point-based data to generate 3D models for use in various applications, but connecting neighboring points obtains noisy and low-quality meshes, thereby leading to occlusions, shadowing and the generation of erroneous regions during depth estimation. Thus, to separate the layers of the acquired depth map in an effective manner, it is necessary to remove the noise and to sharpen the boundary [25].

The resolution of a depth map is lower than that of a color image because of noise degradation during the depth data acquisition process. Consequently, numerous approaches have been proposed for depth map enhancement to remove the noise and retain the layers of the given depth map. However, most of these approaches are affected by the same problems, which are caused when focusing on monoscopic color image enhancement, including spatial resolution enhancement, denoising, and sharpening. These approaches continue to produce problems when enhancing the quality of the depth map because they use the color and depth images jointly to improve the quality, or they require large numbers of training patches for learning-based depth map enhancement. Therefore, these methods are highly dependent on the quality of the color image, training patches, and applications. To overcome these drawbacks and to improve the performance of depth map enhancement without prior information about the given depth map, we propose a novel method called adaptive total variation minimization (ATVM), which facilitates both noise smoothing and boundary sharpening. The proposed method is in fact obtained by combining the moving least squares (MLS) and TV minimization methods. The MLS model provides very satisfactory results for image reconstruction but weak against outliers. In contrast, the process of minimizing TV eliminates outliers effectively since outliers make large variation values. Thus, by incorporating the TV regularizer into the MLS model the solution becomes to achieve a higher order approximation than that of the conventional TV and MLS methods. We filter the noise in the depth map using a refined total variation minimization (TV) minimization technique that uses edge-preserving and noise reduction smoothing filters [23].

## 2 Related work

Many previous studies have proposed conventional two-dimensional image enhancement approaches in the field of computer vision and pattern recognition [1, 6, 23]. In particular, BM3D is very popularly used to remove the noise from a given image, but it requires the statistical variance of the noise to effectively remove the noise in prior. The conventional techniques used to enhance the contrast, sharpness, and color vividness in an image are applied directly during depth map enhancement, where local adjustments are made to increase the amount of high frequency components [21]. Previous approaches have achieved denoising and image sharpening by decreasing or increasing the high frequency components according to the local image characteristics [10, 21]. In particular, Subedar et al. [1] and Kim et al. [12] used high pass filters to enhance the depth-based sharpness, as well as depth estimation and contrast enhancement. In particular, Kinect-based depth map enhancement and denoising researches [9, 21] received numerous concerns as the preprocessing to analyze the 3D scene and human motion analysis. The KinectFusion [16] was also designed to enhance the quality of depth map using multiple depth map images. However, these previously proposed depth map enhancement algorithms used the normal light image, and the enhanced image were obtained by adding a

depth-weighted high pass-filtered color image to the original image [21]. This approach has the limitation that it cannot remove the noise from an unknown depth map image without prior information. Eisenmann and Durand [7] proposed a cross-bilateral filter, where they modified the bilateral filter and computed the edge-preserving term as a function of the depth map image. However, their method preserved edges that did not actually appear in the noisy input depth map image. Eisenmann and Durand [7] replaced the intensity value of each pixel in an image with weighted average intensity values based on nearby pixels. PDE-based denoising methods based on a variational approach of energy functional minimization have also been used for image smoothing with edge preservation. A popular variational denoising method is the TV minimizing process of Rudin-Osher-Fatemi [20]. According to previous image restoration studies, TV regularization has the effect of preserving salient edges and removing noise. However, if the variation minimizing effect is too strong, the smooth regions become flat or constant, thereby yielding a restored image that looks unnatural. This is known as the staircase effect [18] and it is primarily attributed by the fact that the TV minimization method estimates the image using a piecewise constant approximation. Thus, several variants of the TV function have been proposed to avoid the staircase effect and to obtain a higher-order approximation of the reconstructed image [3, 5, 17]. The MLS [2, 14] or kernel regression [24] methods, where the optimal fitting is expressed as a linear combination of polynomials, have been proved to be quite useful in image interpolation as well as denoising and super-resolution [22]. However, MLS based algorithms are weak against noise, since, in general, least squares methods are weak against outliers. Also, when interpolating images across edges, some artifacts (blurring or ringing) are produced into the result images.

In this study, we employ an ATVM technique that has high accuracy to preserve the details of the observed depth map. To preserve strong edges while smoothing noise, we add a TV regularization term to the moving least squares method, and use weight functions that consider the similarity of the local areas in the evaluation and the reference positions.

### 3 ATVM-based depth map enhancement

Let  $I := \{I(i, j) : i = 1, \dots, n_1, j = 1, \dots, n_2\}$  with positive integers  $n_1$  and  $n_2$ . Put  $[1, \dots, n_1] \times [1, \dots, n_2] = [X_1, X_2, X_3, \dots, X_N]$ . Then the observed depth map image  $I$  can be treated as a discrete sampling of a function at a point set  $\{X_1, \dots, X_N\}$  in a domain  $\Omega \subset \mathbb{R}^2$ , where  $N$  is the size of the image. If the given image is contaminated by noise during the image acquisition process, we may write  $I$  as  $I(X_l) = f(X_l) + \varepsilon_l$ ,  $l = 1, \dots, N$ , where  $f(X_l)$  is the value of an underlying function  $f$  and  $\varepsilon_l$  indicates the additive noise at the location  $X_l$ . The denoising method used to construct a denoised image from a depth map image is introduced below.

#### 3.1 Total variation minimization method

For a given noisy image  $I$ , the TV minimization technique [4, 15] generates a denoised image  $\hat{I}$  by solving the following minimization problem

$$\hat{I} = \underset{u}{\operatorname{argmin}} \|u - I\|_2^2 + \mu \|u\|_{TV} \quad (1)$$

where  $\|u\|_{TV} = \int_{\Omega} |\nabla u| dX$  with the gradient operator  $\nabla$ . The second term in Eq. 1,  $\|\cdot\|_{TV}$  is called the total variation norm, and the solution of the minimization problem has the property of preserving

sharp edges in images while removing noise. This is a desirable property for images because the visual quality of an image depends greatly on the preservation of edges. However, this TV scheme processes the observed image to obtain a piecewise constant image, which exhibits many false jump discontinuities and is visually unpleasant. This is mainly attributable to the fact that the TV minimization variation method approximates an image with a first-order accuracy.

### 3.2 Adaptive moving least squares method with a total variation minimizing regularization term

In this section, we suggest an improved TV minimization approach, which is formulated specifically for depth map image denoising.

We employ the adapted least squares technique with the total variation regularization term in [13]. Let  $I$  be a given reference image defined on a domain  $\Omega$  and let  $X^o$  be an evaluation point in  $\Omega$ . We obtain a solution  $\hat{I}(X^o)$  as a denoised image by constructing local polynomial of degree  $m$  in  $\mathbb{R}^2$ ,  $p(X) := p_{X^o}(X)$  and evaluating  $p$  at  $X^o$ , i.e.,  $\hat{I}(X^o) := p(X^o)$ . The polynomial can be written as  $p(X) := \sum_{|\alpha|_1 \leq m} c_\alpha X^\alpha$ . For example, if  $m=2$ ,  $\alpha \in \{(0,0), (1,0), (0,1), (2,0), (1,1), (0,2)\}$ . Specifically,

the coefficients  $c_\alpha$  are obtained by minimizing the following energy functional:

$$\operatorname{argmin}_{p \in \Pi_m} \left\{ \sum_{l=1}^N \left( |p(X_l) - I(X_l)|^2 w(X^o, X_l, I) + \lambda \|\nabla p(X_l)\|_{TV} \right) \right\} \tag{2}$$

where  $\Pi_m$  is the space of bivariate polynomials of degree  $\leq m$  and  $w$  is a specialized weighting function for the denoising solution, which ensures that it obtains a result that preserves textures or repeated local features. Specifically, we use the weighting function defined as

$$w(X^o, X, I) = \exp \left\{ - \frac{\sum_{l \in S} G_a(l) |I(X^o + l) - I(X + l)|^2}{h_0^2} \right\} \tag{3}$$

where  $h_0^2$  is a small positive value and  $G_\alpha$  is a Gaussian function with standard deviation  $a$  and where  $S$  is a suitable (small) stencil for patch comparison around  $X^o$  and  $X$ . The weighting function is data adaptive and it considers the similarity of the local areas in two positions  $X^o$  and  $X$ . In our proposed method, we construct  $p$  locally in the image by solving the minimization problem in Eq. 2 for each evaluation point in  $\Omega$ . Thus, the overall approximation function  $\hat{I}$  becomes  $\hat{I}(X) := p(X) := p_X(X)$  for all  $X \in \Omega$ .

The minimization model (Eq. 2) with the  $L_1$  term can be solved using the split Bregman iteration algorithm [8]. In our method, we obtain the solution based on the following iterated steps for each  $X$ :

$$\begin{aligned} \text{step 1 : } p^{k+1}(X) &= \operatorname{argmin}_p \left\{ \sum_{l=1}^N \frac{\lambda}{2} |p(X_l) - I(X_l)|^2 w(X, X_l) \right. \\ &\quad \left. + \frac{\mu}{2} |d^k(X_l) - \nabla p(X_l) - b^k(X_l)|^2 w(X, X_l) \right\} \\ \text{step 2 : } d^{k+1}(X) &= \operatorname{shrink} \left( \nabla p^{k+1} + b^k(X_l), 1/\mu \right) \\ \text{step 3 : } b^{k+1}(X) &= b^k(X) + \nabla p^{k+1}(X) - d^{k+1}(X) \end{aligned}$$

where  $\operatorname{shrink}(x, \gamma) = \max(|x| - \gamma, 0) \cdot \operatorname{sign}(x)$ .

Without the second term in Eq. 2, the energy functional in Eq. 2 simply becomes the conventional least squares approximation, which fits data by local polynomial approximation [14]. In the previous study for Eq. 2, the TV regularization term was proved to have a better denoising property. In general, a least squares method is weak against outliers; therefore, it is not usually the best tool for denoising. However, in our method, the TV regularization term eliminates the noise very quickly, which helps the regularized method produce a better approximation of the original noise-free image.

## 4 Experiments

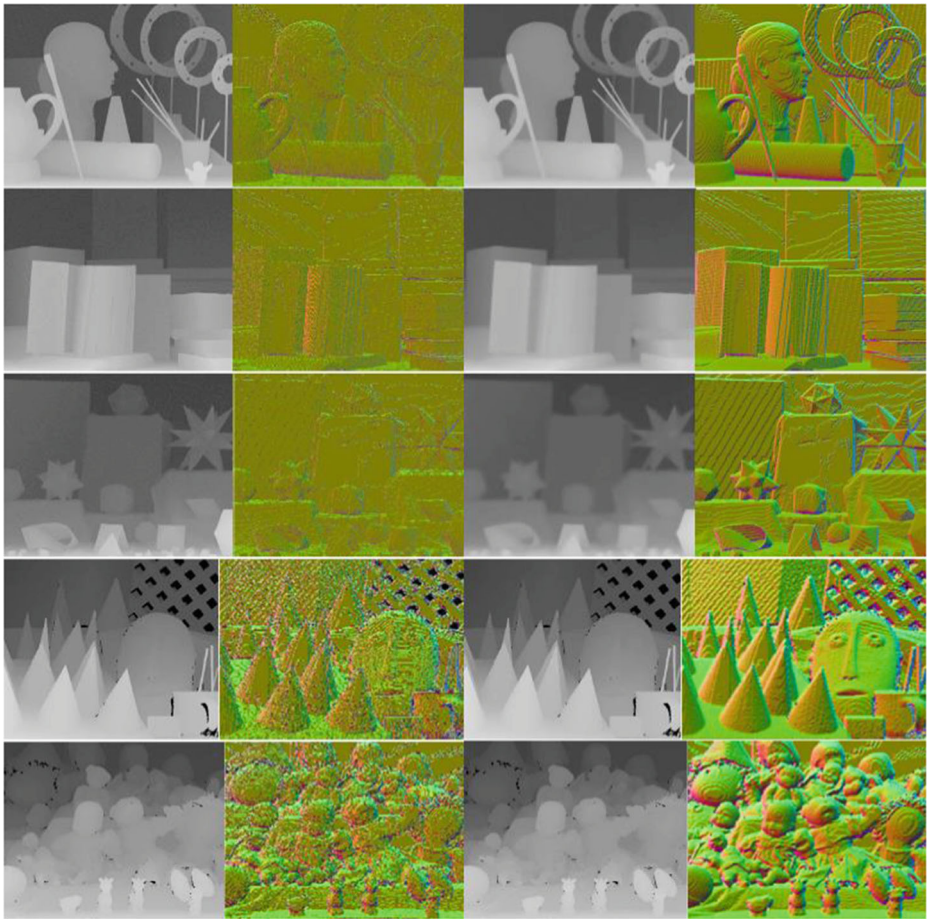
We conducted numerical experiments by using synthetic and real depth map data to evaluate the performance of the ATVM-based depth map enhancement method. In order to assess the improvement in the depth accuracy obtained with the proposed method, we tested the method using known ground truth (synthetic) data from the Middlebury stereo data set, as shown in Table 1. To generate a noisy depth map from the data sets, we added Gaussian noise with a standard deviation of 20 to the ground truth image. We used the peak signal-to-noise ratio (PSNR) which is popularly used as the qualitative measure of the engineering terms for the ratio between the maximum possible power of a signal and the power of the noise. PSNR based on the established ground truth data to quantitatively evaluate the depth map enhancement. Table 1 compares the depth map enhancement results obtained by using our approach and previous approaches, i.e., a TV-based approach and a bilateral approach. As shown in Table 1, the quantitative comparison of the depth map enhancement and denoising using bilateral denoising [7], generous TVM [20], and our approach is represented. The PSNR represents that our proposed approach is very effective to remove the noise from given data. In particular, our approach provides better noise reduction and sharpening from given noisy depth map including multiple layers.

Figure 1 shows the given depth map image with noise and the final image after removing the noise by our proposed approach. To effectively visualize how much our proposed approach is better than given noisy depth map image, we represent the depth map with normal vector. As shown in Fig. 1, our proposed approach is superior in the complex areas that are mixed with different objects because ATVM-based denoising and enhancement approach is very efficient at retaining the edges while removing the noise around the object.

In the next experiment, we tested the performance of the ATVM-based depth map enhancement method using a real depth map obtained by Kinect sensors, which has the

**Table 1** Quantitative comparison of depth map enhancement using our proposed approach and previous approaches based on the PSNR

| Test data | Noisy depth map | Bilateral [7] | TVM [20] | Our approach |
|-----------|-----------------|---------------|----------|--------------|
| Bowling   | 17.47           | 19.75         | 19.93    | 21.72        |
| Middl     | 18.25           | 21.35         | 22.84    | 23.97        |
| Monopoly  | 18.64           | 20.12         | 22.48    | 23.63        |
| Cones     | 16.82           | 17.95         | 18.57    | 21.26        |
| Dolls     | 14.41           | 16.34         | 17.38    | 20.53        |



(a) Input noisy depth and its normal image      (b) Enhanced depth and its normal image

**Fig. 1** Depth map enhancement using our approach using normal image. **a** Input noisy depth and its normal image. **b** Enhanced depth and its normal image

resolution of  $640 \times 480$  pixels. Figure 2 represents the depth map obtained after applying our proposed ATVM-based approach. The middle column of Fig. 2 is the original depth map image from Kinect, but it is not easy to understand the shape and depth of the target object. The right column of Fig. 2 is the refined depth map using our approach. The enhanced depth map obtained using the ATVM-based approach displays the details of the scene better, compared with the input Kinect depth map image. By removing the noise and enhancing the layers of the depth map, it provides effectiveness to analyze the shape of the target objects and 3D scene. Thus, by applying ATVM-based depth map enhancement, we can separate the layers of the given image and analyzed the scene. In particular, compared to remarkable previous approach like KinectFusion [16] which also refines the depth map from Kinect using multiple depth map, the advantage of our approach is in that we use single depth map by retaining the edges and removing the noise from input depth map.

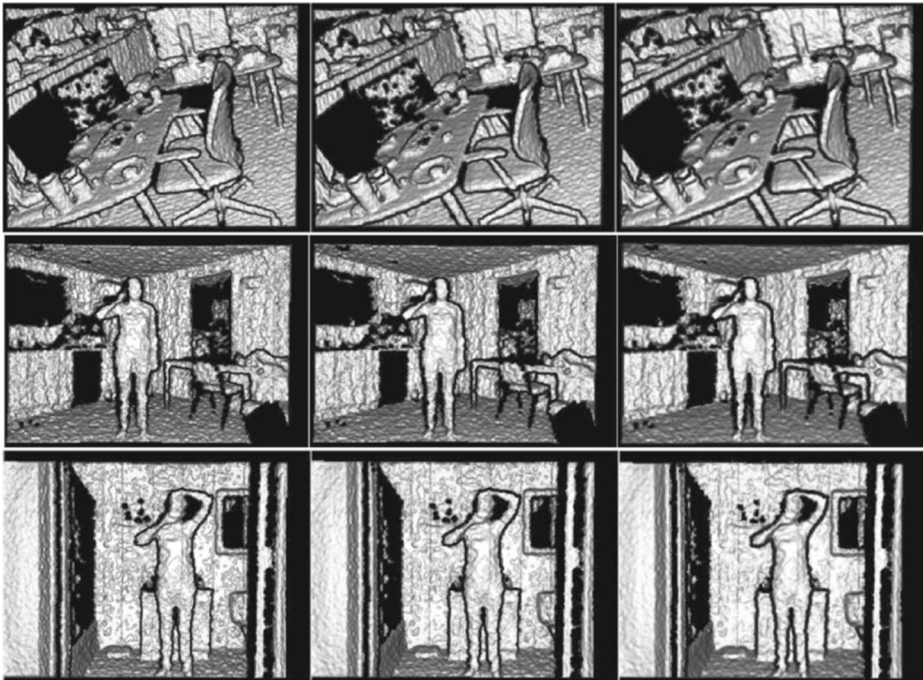




(a) Input RGB image (b) Given noisy depth from Kinect (c) Our approach

**Fig. 2** Real depth map enhancement using our proposed approach based on images captured by Kinect sensors. **a** Input RGB image. **b** Given noisy depth from Kinect. **c** Our approach

To effectively visualize the differentiation between our approach and previous approaches like bilateral and TV methods, Fig. 3 shows the noise removed depth maps which are captured from Kinect. Especially, depth map enhancement and denoising using our approach keeps the separation of the layers and remove the noise in a flat layer. It can be used for layer separation by removing the noise from Kinect.



(a) Bilateral approach [15] (b) Generous TVM approach [16] (c) Our approach

**Fig. 3** Quantitative comparison of depth enhancement and denoising using our approach and previous approaches. **a** Bilateral approach [7]. **b** Generous TVM approach [20]. **c** Our approach

## 5 Conclusion

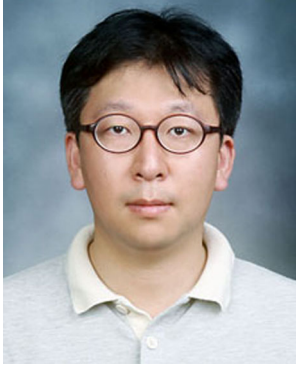
In this study, we proposed a novel depth map enhancement approach based on ATVM. Our method employs a moving least squares method combined with TV minimization, to retain the edges and to remove the noise from input depth images. The moving least squares method facilitates rapid denoising, which allows us to obtain a sufficiently smooth approximation. The TV-based depth map denoising and deblurring approach exhibits robust performance in reducing the noise while retaining the edges in the depth map. Experiments using real/synthetic images demonstrated that our ATVM-based depth map enhancement method satisfied our objectives. By enhancing the resolution of the depth map, the proposed scheme retained the benefits of the TV minimization method and preserved geometric information. In particular, the proposed ATVM performed well in maintaining the details of the target object while reducing the noise, but without requiring prior information.

**Acknowledgments** S.M. Yoon was supported by the ICT R&D program of MSIP/IITP, Korea (B0101-15-1347), A Study on the Key Technology of Optical Modulation and Signal Processing for Implementation of 400 Gb/s Optical Transmission. S.M. Yoon was also supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2014R1A1A1002890). Jungho Yoon was supported by NRF20151009350 (Science Research Center Program) and 2009-0093827 (Priority Research Centers Program) through the National Research Foundation of Korea.



## References

1. Balanna P, Suvarna K, Sudhakar K (2013) Improved depth conception with sharpness augmentation for stereo video. *Int J Comput Electron Res* 2(3):183
2. Bose NK, Ahuja NA (2006) Super-resolution and noise filtering using moving least squares. *IEEE Trans Image Process* 15(8):2239–2248
3. Caselles V, Chambolle A, Novaga M (2007) The discontinuity set of solutions of the TV denoising problem and some extensions. *Multiscale Model Simul* 6(3):879–894
4. Chambolle A (2004) Total variation minimization and applications. *J Math Imaging Vision* 20:89–97
5. Chan T, Esedoglu S, Mulet P (2007) Image decomposition combining staircase reduction and texture extraction. *J Vis Commun Image Represent* 18(6):464–486
6. Danielyan A, Katkovnik V, Egiazarian K (2012) BM3D frames and variational image deblurring. *IEEE Trans Image Process* 21(4):1715–1728
7. Eisenmann E, Durand F (2004) Flash photography enhancement via intrinsic relighting. In *Proceeding of SIGGRAPH*, p 673–678
8. Goldstein T, Osher S (2009) The split Bregman method for  $l_1$  regularized problems. *SIAM J Image Sci* 2(2): 323–343
9. Hu J, Hu R, Wang Z, Gong Y, Duan M (2013) Kinect depth map based enhancement for low light surveillance image. In *Proceeding of ICIP*, p 1090–1094
10. Jung S-W (2013) Enhancement of image and depth map using adaptive joint trilateral filter. *IEEE Trans Circuits Syst Video Technol* 23(2):258–269
11. Jung S-W, Ko S-J (2012) Depth map based image enhancement using color stereopsis. *IEEE Signal Process Lett* 19(5):303–306
12. Kim S-M, Cha J, Ryu J, Lee (2006) Depth video enhancement for haptic interaction using a smooth surface reconstruction. *Inst Electron Inf Commun Eng E89-D*:37–44
13. Lee Y, Lee S, Yoon J (2014) A framework for moving least squares method with total variation minimizing regularization. *J Math Imaging Vision* 48:566–582
14. Levin D (1998) The approximation power of moving least-squares. *Math Comput* 67(224):1517–1531
15. Needell D, Ward R (2012) Total variation minimization for stable multidimensional signal recovery. *SIAM J Numer Anal* 50(3):1162–1180
16. Newcombe RA, Izadi S, Hilliges O, Molyneaux D, Kim D, Davison AJ, Kohli P, Shotton J, Hodges S, Fitzgibbon AW (2011) KinectFusion: real-time dense surface mapping and tracking. In *Proceeding of ISMAR*, p 127–136
17. Nikolova M (2000) Local strong homogeneity of a regularized estimator. *SIAM J Appl Math* 61(2):633–658
18. Nikolova M (2004) Weakly constrained minimization: application to the estimation of images and signals involving constant regions. *J Math Imaging Vision* 21:155–175
19. Rana PK, Ma Z, Taghia J, Flierl M (2013) Multi-view depth map enhancement by variational bayes inference estimation of Dirichlet mixture models. In *Proceeding of ICASSP*
20. Rudin L, Osher S, Fatemi E (1992) Nonlinear total variation based noise removal algorithms. *Physica D* 60: 259–268
21. Stefanoski N, Bal C, Lang M, Wang O, Smolic A (2013) Depth estimation and depth enhancement by diffusion of depth features. In *Proceeding of ICIP*, p 1247–1251
22. Subedar MM, Karam LJ (2010) Increased depth perception with sharpness enhancement for stereo video. In *Proc. SPIE-IS&T Electronic Imaging*
23. Swenson D (2011) Intensity-constrained total variation regularization for image denoising and deblurring. UC Merced: Applied Mathematics
24. Takeda H, Farsui S, Milanfar P (2007) Kernel regression for image processing and reconstruction. *IEEE Trans Image Process* 16(2):349–366
25. Tepper M, Sapiro G (2013) Fast  $L_1$  smoothing splines with an application to Kinect depth data. In *Proceeding of ICIP*, p 504–508
26. Zhang Q (2012) Reconstruction of intermediate view based on depth map enhancement. *J Multimed* 7(6): 415–419
27. Zhang Z (2012) Microsoft kinect sensor and its effect. *IEEE MultiMed* 19(2):4–10



**Sang Min Yoon** received B.E. and M.E. degrees in electronics engineering from the Korea University, Seoul, Korea, in 2000 and 2002, respectively, and a Dr.-Ing. degree in Informatik, TU Darmstadt, Darmstadt, Germany in 2010. Then, he joined the Digital Human Research Center, Advanced Industrial Science and Technology (AIST), Tokyo, Japan, as post-doctoral researcher. He was with Samsung Advanced Institute of Technology (SAIT) as a researcher in the area of face recognition and video surveillance from 2002 to 2005. Currently, he has been working at the school of computer science at Kookmin University, South Korea, as an assistant professor since 2012. His main interests are human motion analysis and understanding, 3D model retrieval, and the human computer interface (HCI).



**Jungho Yoon** received the Ph.D. degree in 1998 from the University of Wisconsin, Madison, USA. From 1998 to 2000, he held a position of Postdoctoral Research Associate at the Department of Mathematics of University of South Carolina, USA. From 2000 to 2001, he was a Postdoctoral Research Associate at the School of Mathematical Sciences of Tel-Aviv University, Israel. From 2001?2002, he was a Visiting Assistant Professor at the Department of Mathematics of Arizona State University, USA. He is currently a Professor at the Department of Mathematics of Ewha Womans University, South Korea. His research interests include image processing, data approximation, subdivision and wavelet.