Semi-Global Depth Up-sampling for 3D Video Acquisition

Sang-Beom Lee and Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST)
123 Cheomdangwagi-ro, Buk-gu, Gwangju, Korea
{sblee, hoyo}@gist.ac.kr

***Abstract*—In this paper, we propose a new method to up-sample a depth map for 3D video acquisition. The proposed method is based on Markov random fields. The low-resolution depth values are used for data term and the high-resolution color data are exploited for smoothness term. Especially, we apply semi-global energy minimization to reduce computational time. Experimental results show that the proposed method outperforms conventional methods.**

Keywords—semi-global depth up-sampling, 3D video, depth-image-based rendering.

I. INTRODUCTION

The three-dimensional (3D) video is in the spotlight as a next-generation multimedia [1]. Owing to advances in display devices, such as stereoscopic or multi-view displays, 3D video provides users with a feeling of presence from the simulation of reality. Both at professional and consumer electronics exhibitions, companies are eager to show their new 3D products which always attract a lot of interest. Obviously, if a workable and commercially acceptable solution can be found, the introduction of 3D video will generate a huge replacement market for the current 2D video.

As a core technology of the 3D video, depth-image-based rendering (DIBR) techniques are investigated [2][3]. DIBR technique is one of the rendering processes to generate arbitrary and virtual views of a 3D scene from color images and so-called depth maps that is composed of per-pixel depth information. Using DIBR technique, the free viewpoint navigation allows one to view a distant 3D world. In order to realize DIBR technique and synthesize intermediate views at virtual viewpoints, accurate depth information is necessary.

The depth map can be obtained by several sensing techniques. They are mainly classified into two categories: passive sensor-based method and active sensor-based method. In the case of the active sensor-based method, various types of sensors, such as laser, infrared ray (IR), or light pattern, are used. Conventionally, the active sensing method has not been a wise choice to reach consumers, due to the huge cost of depth cameras. However, recently, manufacturers have introduced cheaper and smaller depth cameras, allowing wide use in 3D home game and multimedia.

Therefore, the active sensing method is spotlighted as one of the most powerful technologies in 3D content production. Nowadays, time-of-flight (TOF) cameras are widely used to obtain depth maps in real time. However, they have limitations in practical applications due to their low resolutions. Therefore, an efficient depth up-sampling technique is necessary.

Recently, filter-based depth up-sampling algorithms such as joint bilateral up-sampling (JBU) [4] and noise-aware filter for depth up-sampling (NAFDU) [5] have been proposed in order to overcome this problem. . They can reconstruct depth edges requiring only a small memory space and low complexity. However, these methods often lead to over-blurred depth regions near the depth discontinuities.

The other approach based on the Markov random field (MRF) has been presented. Diebel *et al.* proposed a MRF approach for integrating low-resolution depth and its corresponding high-resolution color images [6]. The MRF model integrates depth and color data and provides a probability distribution. Recently, many other depth up-sampling algorithms are proposed by modifying the MRF model [7][8]. However, they require significant computational and memory costs.

In this paper, we propose a new depth up-sampling method for 3D video acquisition. The main contribution of this work is that we exploit semi-global matching as an energy minimization process. Unlike other MRF-based methods, semi-global energy minimization offers a good trade-off between accuracy and runtime and is therefore well suited for many practical applications. By exploiting the low-resolution depth map and high-resolution color image, we can construct MRF model and obtain the high-resolution depth map without huge computational and memory costs.

II. RELATED WORKS

A. Filter-based Depth Up-sampling Methods

In 2007, a new approach for depth up-sampling is proposed. It is called joint bilateral up-sampling (JBU) [4]. JBU is edge-preserving and noise reducing smoothing filter by utilizing color image. The high-resolution depth value via JBU is computed by

$$\widetilde{S}_p = \frac{1}{k_p} \sum_{q_\downarrow \in \Omega} S_{q_\downarrow} \cdot f\left(\left\| p_\downarrow - q_\downarrow \right\|\right) g\left(\left\| \widetilde{I}_p - \widetilde{I}_q \right\|\right) \quad (1)$$

where $p_\downarrow$ and $q_\downarrow$ are the target pixel and neighbor pixel in the low-resolution depth map, respectively. $S_{p\downarrow}$ and $S_{q\downarrow}$ are depth values at $p_\downarrow$ and $q_\downarrow$. In addition, I_p and I_q are the associated color values at p and q in the high-resolution. k_p is a normalization factor. Notice that the JBU uses the range filter of the color image as shown in (1).

Although the JBU can reconstruct depth edges, it has texture copying problem. The texture copying problem is that the discontinuity information of the color image influences the depth map. This problem is occurred when the discontinuity information of color and depth data is not the same. Also, it is particularly noticeable if depth sensor contains a lot of random noise.

In order to solve the texture copying problem, the noise-aware filter, shortly NAFDU, is proposed to reduce this problem [5]. NAFDU is an adaptive multi-lateral filter that blends two range filters of color and depth data. The high-resolution depth value via NAFDU is calculated by

$$\widetilde{S}_p = \frac{1}{k_p} \sum_{q \in \Omega} S_q \cdot f\left(\left\| p - q \right\|\right) \left[\alpha(\Delta\Omega) g\left(\left\| I_p - I_q \right\|\right) + (1 - \alpha(\Delta\Omega)) h\left(\left\| S_p - S_q \right\|\right) \right] \quad (2)$$

$$\alpha(\Delta\Omega) = \frac{1}{1 + e^{-\varepsilon(\Delta\Omega - \tau)}} \quad (3)$$

where α represents the blending function and $\Delta\Omega$ denotes the depth difference between the minimum and maximum depth values inside the filter kernel. The blending function is dependent to the depth difference Thus, if the depth difference is increased, α is close to 1 and NAFDU applies the JBU. Otherwise, NAFDU applies the bilateral up-sampling.

B. *MRF-based Depth Up-sampling Methods*

Diebel *et al*. proposed a depth up-sampling method by utilizing Markov random field [6]. This algorithm performs this data integration using a multi-resolution MRF model, which combines the color image and the depth map. The low-resolution depth map is sparsely distributed in the high-resolution color image and it is used for *data term*. Then, the weighting factors for color information are used for *smoothness term*. They are defined by

$$p(y \mid x, z) = \frac{1}{Z} \exp\left\{ -\frac{1}{2} (\Psi + \Phi) \right\} \quad (4)$$

$$\Psi = \sum_{i \in L} k (y_i - z_i)^2 \quad (5)$$

$$\Phi = \sum_{i} \sum_{j \in N(i)} w_{ij} (y_i - y_j)^2 \quad (6)$$

where Ψ and Φ are data term and smoothness term, respectively. In order to obtain the appropriate value for unobservable variable y, the data and smoothness terms for energy minimization are defined. L is the set of indices for which a depth measurement is available. k represents a constant weight. The data term measures the quadratic distance between the estimated depth values in the high-resolution grid y and the measured depth z.

In the case of smoothness term, $N(i)$ denotes a set of nodes adjacent to the target pixel i, the neighboring pixels j. This term also defines a weighted quadratic distance between neighboring nodes. The weighting factor w_{ij} is the crucial aspect because they make the link from the depth layer to the image layer in the MRF model.

By optimizing this MRF model using global optimization algorithm, high-resolution depth map is can be obtained.

III. PROPOSED METHOD

A. *Construction of Energy Function*

Fig. 1 shows MRF structure for the proposed depth up-sampling method. The density of image pixels is larger than that of the depth pixels. The nodes connecting the measured pixels are represented by color information. They can be used as a smoothness term for energy minimization.

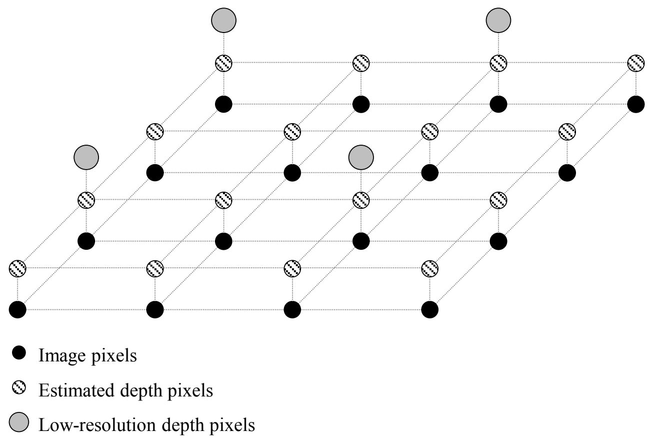


Fig. 1. MRF structure for semi-global depth up-sampling.

Similar to the conventional MRF-based method, we first construct the energy function. It is defined by

$$E(p,d) = \sum_{i \in L} \left| y_i - z_i \right| + \sum_{i \in L} \sum_{j \in N(i)} P_1 w_{ij} \mathrm{T}\left[\left| y_i - y_j \right| = 1 \right] \\ \sum_{i \in L} \sum_{j \in N(i)} P_2 w_{ij} \mathrm{T}\left[\left| y_i - y_j \right| > 1 \right] \quad (7)$$

where the first term represents the data term. The second term adds a constant penalty P_1 for all neighbor pixels j in the neighborhood $N(i)$, if the disparity slightly changes. The third term adds a larger constant penalty P_2, for all larger disparity changes. P_1 and P_2 are most important factors for the energy function. P_1 preserves slanted or curved surface, that is, homogeneous region. P_2 is a regularization factor for the discontinuity region.

B. Semi-global Depth Up-sampling

The MRF can be solved by semi-global energy minimization [9]. Its concept extends single-line dynamic stereo matching into a multi-line integration strategy. It consists of two steps. First, the data costs are calculated for all disparities *d*. Second, these calculated costs are regularized along scan lines that run across the image domain by employing an accumulative dynamic programming scheme.

The idea of Semi-global error minimization is the computation along several paths, specifically eight paths through the image. Eight paths from all directions meet at every pixel. In addition, each path contains the information about the cost for reaching a pixel with a certain disparity. For each pixel and depth values, the costs are accumulated over the paths. After all, the disparity with the lowest cost is chosen at each pixel. The energy function for semi-global energy minimization is defined by

$$E(\mathbf{p},d) = \sum_{\mathbf{r}} L_{\mathbf{r}}(\mathbf{p},d) \tag{8}$$

$$\begin{aligned} L_{\mathbf{r}}(\mathbf{p},d) = C(\mathbf{p},d) + w_r \cdot \min[&L_{\mathbf{r}}(\mathbf{p}-\mathbf{r},d), \\ &L_{\mathbf{r}}(\mathbf{p}-\mathbf{r},d-1)+P_1, \\ &L_{\mathbf{r}}(\mathbf{p}-\mathbf{r},d+1)+P_1, \\ &\min_i L_{\mathbf{r}}(\mathbf{p}-\mathbf{r},i)+P_2] - w_r \cdot \min_k L_{\mathbf{r}}(\mathbf{p}-\mathbf{r},k) \end{aligned} \tag{9}$$

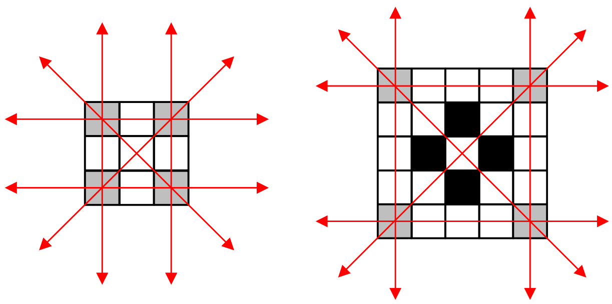
where $L_{\mathbf{r}}(\mathrm{p},d)$ represents the recursive smoothness term along a path traversed in the direction $\mathbf{r}$ of the pixel $\mathbf{p}$ at disparity *d*. $C(\mathbf{p},d)$ denotes the data term that is only available for the low-resolution depth pixels. P_1 and P_2 represent the constant penalties, respectively. The weighting factor w_r is defined by

$$w_r = \varepsilon + \exp\left\{-\frac{(R_p - R_{p-r})^2 + (G_p - G_{p-r})^2 + (B_p - B_{p-r})^2}{2\sigma^2}\right\} \tag{10}$$

where *R*, *G*, and *B* are red, green, and blue component of the color image. σ^2 represents the variance of the range filter. ε prevents that the weighting factor changes the smoothness term to zero.

Semi-global energy minimization is very powerful in terms of computational time compared with other optimization methods. However, if the scale factor of up-sampling is larger, single operation cannot cover the entire pixel. Fig. 2 shows empty pixels with single operation. If the scale factor is 2 as shown in Fig. 2(a), whole pixels are calculated by semi-global energy minimization. On the other hand, in the case of scale factor 4, there exist empty pixels that are not affected by the low-resolution depth values as shown in Fig. 2(b).

Therefore, iterative process is required to fill out those empty pixels. The second iteration process is as follows. From the high-resolution depth map obtained by the first iteration, data term is calculated except for empty pixels. Then, semi-global energy minimization is performed to fill out empty pixels. Although the minimum number of iteration is two, we set to half of scale factor in the proposed method.



(a) 2× up-sampling (b) 4× up-sampling

Fig. 2. Empty pixels of single operation.

IV. Experimental Results

In order to evaluate the proposed algorithm, we tested using three Middlebury datasets with nearest neighbor down-sampling [10]: *Art*, *Books*, and *Moebius*. They include ground truth depth data at 1376×1088 resolution. The proposed method is compared with the JBU, NAFDU, and MRF-based method using belief propagation [10]. In order to evaluate the up-sampled depth maps, the bad pixel rate (BPR), whose absolute difference is greater than 1, was used. The constant penalties P1 and P2 are fixed as 1 and 5, respectively. The variance of the range filter σ^2 is set to 25.5.

Table 1 provides a quantitative evaluation of the conventional and proposed methods. According to the results, the proposed algorithm outperforms the conventional algorithms in most cases. Especially, if the scale factor is 2, the results of the proposed method were the best.

TABLE I. Bad Pixel Rates (%)

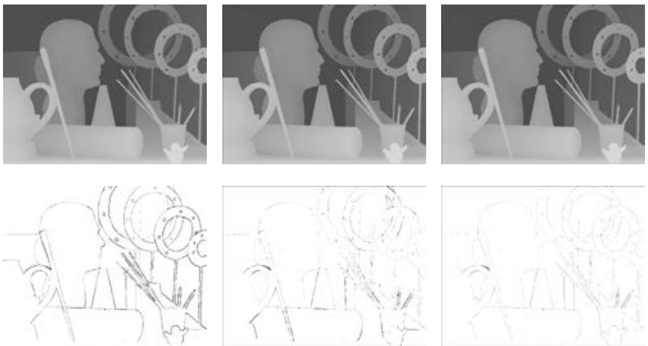
Methods	Art		Books		Moebius	
	2×	4×	2×	4×	2×	4×
JBU	2.631	4.060	1.061	2.024	1.623	2.905
NAFDU	2.630	4.052	1.057	2.009	1.621	2.898
MRF	1.104	1.385	1.106	1.657	1.601	2.049
Proposed	0.710	2.095	0.719	2.354	0.779	2.320
Rank	**1**	**2**	**1**	**4**	**1**	**2**

Table 2 represents the average running time of each method. The complexity of the proposed method is almost half of the MRF-based method. In other words, since the proposed semi-global energy minimization is much faster than the conventional MRF-based method, it is more suitable for practical applications.

TABLE II. COMPUTATIONAL TIME (ms)

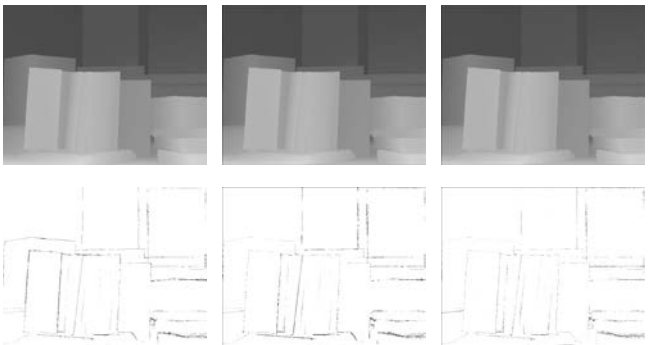
Scale	JBU	NAFDU	MRF	**Proposed**
2×	396	611	29,147	**15,966**
4×	721	919	56,174	**31,823**

Fig. 3 through Fig. 5 present the up-sampled depth maps (row 1) and error maps with respect to BPR (row 2) with a scale factor of 2. From the figures, we noticed that the proposed method reduces the depth errors, especially near the depth discontinuity, compared with the conventional methods.



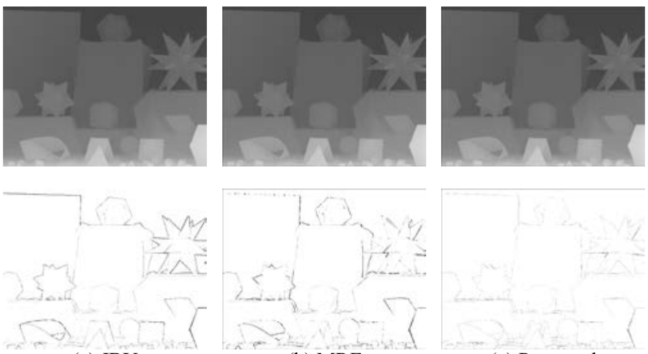
(a) JBU (b) MRF (c) Proposed

Fig. 3. 2× up-sampling results for '*Art*'.



(a) JBU (b) MRF (c) Proposed

Fig. 4. 2× up-sampling results for '*Books*'.



(a) JBU (b) MRF (c) Proposed

Fig. 5. 2× up-sampling results for '*Moebius*'.

V. CONCLUSIONS

In this paper, we have proposed a new depth up-sampling method in order to obtain a high-resolution depth map. Compared with the conventional algorithms, the proposed method exploited semi-global energy minimization for fast operation. Experimental results have shown that we efficiently reduce the computational and memory costs while preserving reliable depth quality. As a result, the proposed method will be beneficial to various practical applications.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2013-067321)).

REFERENCES

[1] A. Smolic, K. Müller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, and T. Wiegand, "3D Video and Free Viewpoint Video – Technologies, Applications and MPEG Standards," *International Conference on Multimedia and Expo*, pp. 2161–2164, July 2006.

[2] A. Redert, M. O. Beeck, C. Fehn, W. IJsselsteijn, M. Pollefeys, L. Van Gool, E. Ofek, I. Sexton, and P. Surman, "ATTEST: Advanced Three-dimensional Television System Techniques," *International Symposium on 3D Data Processing*, pp. 313–319, March 2002.

[3] C. Fehn, "Depth-Image-Based Rendering (DIBR), Compression and Transmission for a New Approach on 3-D TV," *SPIE Conference on Stereoscopic Displays and Virtual Reality Systems*, vol. 5291, pp. 93–104, Jan. 2004.

[4] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," *ACM Transactions on Graphics*, vol. 26, no. 3, pp. 1–5, July 2007.

[5] D. Chan, H. Buisman, C. Theobalt, and S. Thrun, "A Noise-aware Filter for Real-time Depth Upsampling," *ECCV Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications*, pp. 1–12, Oct. 2008.

[6] J. Diebel and S. Thrun, "An Application of Markov Random Fields to Range Sensing," *Neural Information Processing Systems*, pp. 291–298, Dec. 2005.

[7] J. Park, H. Kim, M. S. Brown, and I. Kweon, "High Quality Depth Map Upsampling for 3D-TOF Cameras," *International Conference on Computer Vision*, pp. 1623–1630, 2011.

[8] D. Kim and K. Yoon, "High-quality Depth Map Up-sampling Robust to Edge Noise of Range Sensors," *International Conference on Image Processing*, pp. 553–556, 2012.

[9] H. Hirschmüller, "Stereo Processing by Semiglobal Matching and Mutual Information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 328–341, Mar. 2008.

[10] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient Belief Propagation for Early Vision," *International Journal of Computer Vision*, vol. 70, no. 1, pp. 41–54, May 2006.