Depth Image Interpolation Using Confidence-based Markov Random Field

Jae-Il Jung and Yo-Sung Ho, *Senior Member*, IEEE

Abstract — *Depth images are essential data for high-quality three-dimensional (3D) video services, but the resolution of depth images captured by commercially available depth cameras is lower than that of the corresponding color images, owing to technical limitations. A depth image up-sampling method that uses a confidence-based Markov random field is proposed for enhancing this resolution. An initial high-resolution depth image and confidence values are generated with consideration of boundaries and textures in the corresponding color images. These are used as the base for a new likelihood and prior model design. The energy function derived from this model is optimized by using a graph cut algorithm, and subsequent experiments show that the proposed algorithm provides sufficiently good up-sampled depth images compared to other state-of-the-art algorithms*[1].

Index Terms — Confidence, depth camera, depth image, interpolation, Markov random field (MRF).

I. Introduction

Three-dimensional (3D) video services are attracting considerable attention due to the commercial success of 3D movies, and multi-view and free viewpoint services are in the spotlight as the next generation of 3D video services. Such formats provide various images from different viewpoints simultaneously [1]-[3].

However, it is physically impossible in practice to capture all images. Therefore, depth image-based rendering (DIBR), which synthesizes images at desired viewpoints, has been proposed [4]. Using this concept, the Moving picture experts group (MPEG) has investigated an advanced 3D video system, and the depth image plays an important role in this system [5], [6].

Depth images represent distances between a camera and objects. They have been widely used owing to their diverse potential applications, including DIBR [7]-[9], with a number of range-measuring technologies for obtaining accurate depth images currently being explored.

To this end, depth cameras are a popular technology because they are able to capture dynamic depth images in real-time. However, although depth cameras are convenient to use and generally provide good performance, the current generation of depth cameras is limited in terms of their resolution. Moreover, recent demands require even higher image resolution, so various resolution enhancement methods have been proposed.

For this task, common up-sampling approaches initially implemented included bilinear and bicubic interpolations. However, despite being simple and reliable, the output of these approaches is erroneous with respect to depth discontinuities. Therefore, alternative algorithms for depth images, which consider not only a depth image but also its corresponding color image, have been developed.

Diebel *et al.* proposed a Markov random field (MRF) approach for integrating low-resolution depth and its corresponding high-resolution color images [10]. The MRF model integrates depth and color data and provides a probability distribution. The model is optimized using the conjugate gradient algorithm. Yang *et al.* adopted a bilateral filter for low-resolution depth images [11]. In this approach, the filter was applied to the weighted cost volume, which was followed by sub-pixel refinement. Their method refines a low-resolution depth image iteratively in terms of both its spatial resolution and depth precision.

Although the above-mentioned algorithms improve output quality in comparison with conventional up-sampling methods, their performance decreases markedly, when the resolution of depth images is significantly lower than that of the color images. In this case, edges are severely blurred, and accuracy is decreased as well.

In this paper, a relatively efficient depth image up-sampling method that uses a confidence-based MRF model is proposed to resolve these problems. Conventional MRF-based approaches use only a similar model for stereo matching, but the proposed model for depth image up-sampling is designed taking into account the properties of a depth image. After conducting depth image up-sampling with two local interpolation algorithms, the proposed method measures the confidence of each pixel. To cover inaccurate regions, new likelihood and prior functions are designed for the MRF model that consider the confidence of the initial depth image and. This model is optimized via a graph cut algorithm.

II. Markov Random Field Modeling

MRF modeling can provide a powerful solution for depth image up-sampling; a high-resolution depth image can be generated effectively from a low-resolution depth image using

[1] This research is supported by the Ministry of Culture, Sports and Tourism (MCST) and Korea Creative Content Agency (KOCCA) in the Culture Technology (CT) Research & Development Program 2012.

J. Jung and Y. Ho are with the School of Information and Communications, Gwangju Institute of Science and Technology (e-mail:jijung@gist.ac.kr and hoyo@gist.ac.kr).

Contributed Paper
Manuscript received 10/15/12
Current version published 12/28/12
Electronic version published 12/28/12.

a maximum a posteriori (MAP) framework. According to Bayes' rule, the posterior probability can be computed as

$$P(D_H \mid I_C, D_L) = \frac{P(I_C, D_L \mid D_H)P(D_H)}{P(I_C, D_L)}, \quad (1)$$

where D_H, D_L, and I_C are high- and low-resolution depth images and the corresponding high-resolution color image, respectively. $P(D_H)$ is the prior probability of the depth image D_H, and $P(I_C, D_L|D_H)$ is the conditional probability density function of I_C and D_L given D_H. $P(I_C, D_L)$ can be ignored because it does not contain unknowns; it is a constant when I_C and D_L are given. Therefore, the posterior probability is proportional to the joint distribution, and MAP is equivalently determined as

$$D_H^* = \arg\max_D \{P(I_C, D_L \mid D_H)P(D_H)\}. \quad (2)$$

Note that the first term $P(I_C, D_L|D_H)$ is the likelihood probability and the second term $P(D_H)$ is the prior probability of the depth image interpolation. In this paper, a new posterior probability is proposed, with a confidence value that supports this probability model in reducing the uncertainty of low-resolution depth images.

III. Depth Image Interpolation Method

A. Initial Depth Image and Confidence Map

Since the MRF model estimates labels based on observed data, these data are very important. In the conventional algorithm [10], only low-resolution depth images are regarded as observed data even though there is a lack of information for explicitly designing the MRF model. To supplement observed data, the proposed method generates an initial high-resolution depth image and confidence map and regards them as observed data. Two local algorithms are used complementarily to generate it: color-based and bilinear interpolation.

Bilinear interpolation is a well-known algorithm in the field of image processing. The color-based interpolation algorithm estimates depth values by considering neighbors having similar color values, as follows.

$$d_i = \frac{1}{W}\sum_{j \in N_C} f_S(i,j) d_{Rj} \quad (3)$$
$$where\ f_S(i,j) = \exp\left(-\frac{\sqrt{(i_x - j_x)^2 + (i_y - j_y)^2}}{5}\right).$$

In (3), N_c is a pixel set whose Euclidean distances to pixel i in the RGB domain are less than Th_c, and d_{Rj} is its depth value. In addition, $f_S(i, j)$ is the exponential weighting function based on the Euclidean distance in the spatial domain between two pixels i and j, and W is the sum of $f_S(i, j)$. The reference samples j are found in raster scan order from their co-located positions in the low-resolution depth image, and the number of reference samples is limited to four. Furthermore, in contrast to a bilateral filter, the ratio of color differences is not used to prevent intermediate value generation—a very important factor for obtaining high-quality depth images [12].

The results of the two interpolation methods are shown in Figs. 1(a) and 1(b). These show that whereas the color-based interpolated depth image has relatively precise values around the discontinuities, it has significantly less accurate values than the linearly interpolated depth image in the non-boundary and high-textured regions of color images.

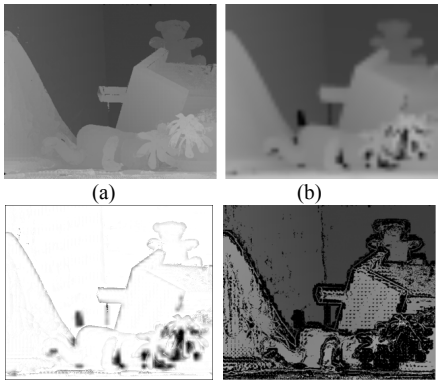

Fig. 1. (a) Color-based interpolated depth image, (b) linearly interpolated depth image, (c) confidence map, and (d) initial depth image.

A confidence map is generated by calculating the difference between the two interpolated depth images and inverting the result, as shown in Fig. 1(c). Then, depth values having high confidence (> 200) are selected. Subsequently, an initial high-resolution depth image is generated. The initial depth image also includes the depth values of the color-based interpolated depth image around its discontinuities. Fig. 1(d) shows the initial depth image, and the black regions represent holes that cannot be estimated from the two interpolated images. The initial depth image needs an additional refining process using MAP estimation.

B. Likelihood Model

The MAP estimate is similarly found by minimizing the energy function $U(D_H|I_C, D_L)$. The energy function is the sum of the likelihood energy $U(I_C, D_L|D_H)$ and the prior energy $U(D_H)$ in the MRF model for depth image interpolation.

The energy function for the likelihood probability is modeled to measure the disagreement between estimated and observed data, and then to compensate for it. Note also that although the initial depth image still has many holes, the existing values are relatively accurate. As such, it is regarded in the MRF model as the observed depth image D_O. Therefore, the proposed algorithm compares the current estimated depth value with the co-located value in the initial depth image. This comparison is conducted only when the depth value at the current position in the initial depth image is available. Accordingly, the energy function for the likelihood probability is defined as

$$U(I_C, D_O \mid D_H) = w_L \sum_i f_L(d_i) \quad (4)$$
$$where\ f_L(d_i) = \begin{cases} 1 - \exp\left(\dfrac{-(d_i - d_{init\_i})^2}{\sigma_L}\right) & if\ d_{ci} \in D_O \\ 1 - (u[d_{O\max}] - u[d_{O\min}]) & otherwise. \end{cases}$$

As shown in (4), the likelihood energy function is the summation of the cost function f_L of the estimated depth values d_i at position i. In the cost function f_L, d_{init_i} is the depth value in the initial depth image, $u[\ ]$ is the unit step function, and w_L is a weighting factor of the likelihood energy.

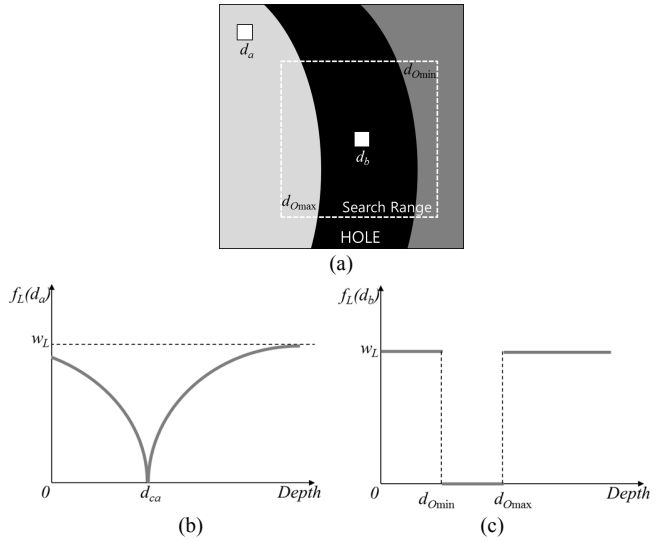

Fig. 2. Example of the likelihood energy function: (a) depth values to be estimated, d_a and d_b, and the likelihood energy functions for (b) d_a and (c) d_b.

When the initial depth image has a valid value at the current position i, f_L is used to evaluate the accuracy of pixel value matching. However, if the depth value is invalid, f_L only restricts the range of the depth values from the minimum depth value d_{Omin} to the maximum depth value d_{Omax} in a search range. Fig. 2 illustrates the properties of the likelihood energy. In Fig. 2(a), d_a is a pixel having a corresponding depth value, and d_b has no corresponding depth value in the initial depth image. Figs. 2(b) and 2(c) show examples of each energy function. While $f_L(d_a)$ penalizes differences between current and corresponding depth values, $f_L(d_b)$ allocates low penalty values to the range from the minimum depth value d_{Omin} to the maximum depth value d_{Omax} of neighbors in D_O for mobility.

C. Prior Model

The likelihood model alone is insufficient for accurately estimating depth values that do not have a corresponding depth value in the initial depth image. Therefore, we define an additional constraint based on prior knowledge that the depth values of the high-resolution depth image are piecewise continuous; hence, it is modeled to compensate for any variation in smoothness, such that

$$U(D_H) = w_p \sum_i \sum_{j \in N} f_p(d_i, d_j)$$
$$\text{where } f_p(d_i, d_j) = w_C(i, j) \cdot (d_i - d_j)^2, \tag{5}$$

where N is the set of edges in a four-connected image grid graph. Here, $w_C(i, j)$ is a weighting factor with respect to the color difference and the confidence value and is defined as

$$w_C(i, j) = conf\ (i, j)(1 - \exp\left(\frac{-\sum_{c \in \{r,g,b\}} (C_i - C_j)^2}{\sigma_p}\right) \tag{6}$$
$$\text{where } \ conf\ (i, j) = \exp(-w_p \cdot \min(conf_i, conf_j)).$$

The weighting factor is based on the Euclidean distance between pixel i and j in the RGB domain, and w_p controls the effect of the confidences. If both confidence values $conf_i$ and $conf_j$ are greater than a sufficiently large value (in this paper, this value is 250), the prior constraint can be ignored; that is, the prior model does not change the depth values of high confidence in the initial depth image.

IV. Experimental Results

Experiments were performed on Middlebury datasets, and the proposed MRF model was optimized using a graph cut algorithm [13]. The ground-truth depth images were down-sampled by factors of 4 and 8 for each axis, and then up-sampled using the proposed algorithm. Throughout test runs, the algorithm parameters were set to constant values: $\{Th_c, w_L, w_p\} = \{10, 15, 13\}$. The search range was set to 16.

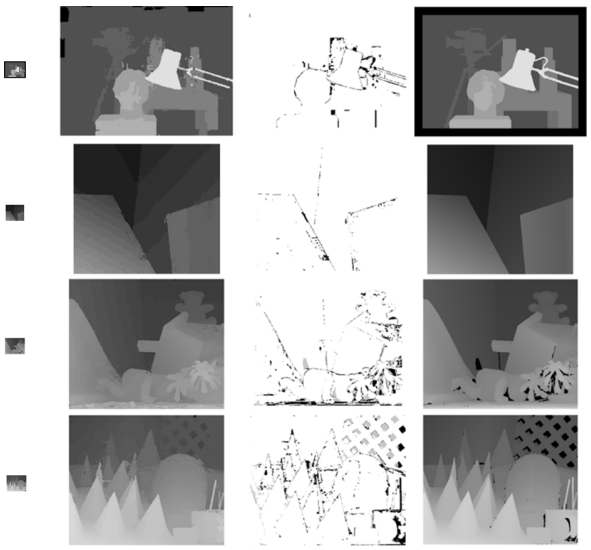

Fig. 3. Input (a scaling factor of 8) depth images, up-sampled depth images, corresponding error maps, and ground-truth depth images for Tsukuba, Teddy, Venus, and Cones.

Fig. 3 shows the interpolation results for the datasets for a down-sampling factor of 8. The first column contains the input depth images, and the second column shows the results. The third column shows the corresponding error maps between the up-sampling results and the ground truth in the last column. Despite the large down-sampling factor, the proposed interpolation method efficiently reconstructed depth borders.

For an objective evaluation, the proposed algorithm was compared with two traditional interpolation algorithms, the bilinear and bicubic algorithms, and two state-of-the-art algorithms, Diebel's [10] and Yang's [11]. Table I summarizes the proportion of bad pixels according to the

different down-sampling factors. Estimating bad pixels is a standard measure of depth image accuracy in which an estimated depth is considered correct if it is within the ground-truth depth ±1. In the table, the proposed algorithm offers more accurate depth values than the others do, especially for severely down-sampled depth images.

TABLE I
COMPARISON OF PROPORTION OF BAD PIXELS

%	Tsukuba		Teddy		Venus		Cone	
Factor	4	8	4	8	4	8	4	8
Bilinear	8.64	14.98	11.04	18.89	1.63	3.33	14.04	23.61
Bicubic	7.96	13.03	10.42	17.33	1.35	2.77	12.81	22.27
Diebel's	5.12	9.68	8.33	14.50	1.24	2.69	7.52	14.40
Yang's	2.56	6.95	5.95	11.50	0.42	1.19	4.76	11.00
Proposed	1.62	2.81	5.01	7.33	0.52	1.02	5.59	8.78

These results confirm that the proposed likelihood and prior models based on confidence are well defined in the energy functions. The accuracy of the proposed algorithm makes it possible to use it in practical applications. For instance, Fig. 4 shows the reconstructed models using the up-sampling results (a scaling factor of 8). The proposed algorithm supports synthesis of various images from different viewpoints without considerable visual distortions, even though the resolution of input depth images is very small. Optimizing the parameters in the MRF model is expected to lead to further improvements.

Fig. 4. 3D mesh models using the proposed up-sampling method.

V. CONCLUSION

The resolution of depth images captured by depth cameras is normally smaller than that of the corresponding color image owing to technical limitations. In this paper, an MRF model-based depth image interpolation method was proposed. To use observed data more efficiently, the proposed method generates the initial depth image using two local interpolation methods, and subsequently defines a confidence map. Based on the initial depth image and confidence values, the proposed MRF model was designed and optimized via a graph cut algorithm. Experimental results showed that the proposed algorithm exhibits better performance for depth image interpolation than conventional methods. The proposed method can be useful in various commercial applications that use low-resolution depth cameras.

REFERENCES

[1] A. Kubota, A. Smolic, M. Magnor, M. Tanimoto, T. Chen, and C. Zhang, "Multi-view imaging and 3DTV (Special Issue Overview and Introduction)," *IEEE Signal Process. Mag.*, vol. 24, no. 6, pp. 10-21, Nov. 2007.

[2] P. Merkle, K. Muller, and T. Wiegand, "3D video: acquisition, coding, and display," *IEEE Trans. Consum. Electron.,* vol. 56, no. 2, pp. 946-950, May 2010.

[3] M. Tanimoto, "Overview of free viewpoint television," *Signal Process.-Image Commun.*, vol. 21, no. 6, pp. 454-461, July 2006.

[4] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," *in Proceedings of SPIE Stereoscopic Displays and Virtual Reality Systems XI*, vol. 5291, no. 2, pp. 93-104, May 2004.

[5] A. Smolic and D. McCutchen, "3DAV exploration of video-based rendering technology in MPEG," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 3, pp. 348-356, March 2004.

[6] A. Smolic, K. Mueller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, and T. Wiegand, "3D video and free viewpoint video-technologies, applications and MPEG standards," *in proceedings of IEEE International Conference on Multimedia and Expo,* pp. 2161-2164, July 2006.

[7] W. Jang and Y. Ho, "Efficient disparity map estimation using occlusion handling for various 3D multimedia applications," *IEEE Trans. Consum. Electron.,* vol. 57, no. 4, pp. 1937-1943, Nov. 2011.

[8] S. Lee, I. Shin, and Y. Ho, "Gaze-corrected view generation using stereo camera system for immersive videoconferencing," *IEEE Trans. Consum. Electron.,* vol. 57, no. 3, pp. 1033-1040, Aug. 2011.

[9] E. Lee and Y. Ho, "Generation of multi-view video using a fusion camera system for 3D displays," *IEEE Trans. Consum. Electron.,* vol. 56, no. 4, pp. 2797-2805, Nov. 2010.

[10] J. Diebel and S. Thrun, "An application of markov random fields to range sensing," *in proceedings of Advances in neural information processing systems*, vol. 18, pp. 291-298, 2006.

[11] Q. Yang, R. Yang, J. Davis, and D. Nister, "Spatial-depth super resolution for range images," *in proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, June 2007.

[12] K. Oh, A. Vetro, and Y. Ho, "Depth coding using a boundary reconstruction filter for 3-D video systems," *IEEE Trans. Circuits Syst. Video Technol.,* vol. 21, no. 3, pp. 350-359, March 2011.

[13] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222-1239, Nov. 2001.

BIOGRAPHIES

Jae-Il Jung received his B.S. degree in electronics and communications engineering in 2005, and M.S degree in information display engineering from Hanyang University, Korea, in 2007. He is pursuing a Ph.D. degree in the School of Information and Communications at the Gwangju Institute of Science and Technology (GIST). His research interests include digital signal processing, video coding, 3D broadcasting, hologram, and solution optimization.

Yo-Sung Ho (M'81-SM'06) received both B.S. and M.S. degrees in electronic engineering from Seoul National University, Korea, in 1981 and 1983, respectively, and Ph.D. degree in Electrical and Computer Engineering from the University of California, Santa Barbara, in 1990. He joined the Electronics and Telecommunications Research Institute (ETRI), Korea, in 1983. From 1990 to 1993, he was with Philips Laboratories, Briarcliff Manor, New York, where he was involved in development of the advanced digital high-definition television (AD-HDTV) system. In 1993, he rejoined the technical staff of ETRI and was involved in development of the Korea direct broadcast satellite (DBS) digital television and high-definition television systems. Since 1995, he has been with the Gwangju Institute of Science and Technology (GIST), where he is currently a professor in the School of Information and Communications. His research interests include digital image and video coding, image analysis and image restoration, advanced coding techniques, digital video and audio broadcasting, 3D television, and realistic broadcasting.