Depth map upsampling using depth local features

Y.S. Kang, S.B. Lee and Y.S. Ho

A depth map upsampling method using depth local features is proposed. Depth discontinuity and depth variance are extracted from a low-resolution depth map and a colour image as the depth local features. They are incorporated into an energy function for the Markov random field (MRF)-based depth upsampling. The high-resolution depth map is obtained by optimising the energy function using belief propagation. The experimental results show that the proposed method outperforms other depth upsampling approaches in terms of the bad pixel rate.

Introduction: Three-dimensional (3D) video provides immersive sense using images from dense multiple viewpoints. However, since the capturing of the dense multiple viewpoints is difficult due to the mechanical limitation, depth maps are required to synthesise intermediate viewpoint images. Thus, acquiring high-quality depth maps is important in 3D video and related applications. In recent years, depth sensors based on the time-of-flight technique have been used to acquire the depth map of the scene. Although they generate the depth map in real-time, low-resolution output is one of the critical drawbacks. To overcome this low-resolution output, depth map upsampling methods such as filter-based [1, 2] and Markov random field (MRF)-based [3] approaches have been proposed. However, the filter-based methods can cause an over-smooth problem at the depth discontinuity regions, and also the MRF-based method can occur with error propagation during the optimisation process. In this Letter, a new depth map upsampling method to reduce such problems is presented.

Depth upsampling using the depth local features: In this Letter, a high-resolution colour image, a low-resolution depth map and an upsampled depth map are denoted by I, D_L and D_U, respectively. Since the proposed method is based on the MRF model, an energy function E is defined as (1). The first term represents a data term where y means each pixel in D_U and z indicates the mapped pixels to I from D_L. i and j represent the pixel positions. L is a set of pixels that have z_i values. The second term is the smoothness term between two pixels i and j, where $N(i)$ means four-neighbouring pixels around the ith pixel. k and w_{ij} are the weights for the data and the smoothness terms, respectively. They control the influences of each term. By optimising (1), D_U is acquired:

$$E = \sum_{i \in L} k(y_i - z_i)^2 + \sum_{i} \sum_{j \in N(i)} w_{ij}(y_i - y_j)^2 \tag{1}$$

The proposed method extracts and utilises the depth local features to overcome the depth error problem at the depth discontinuity regions. Fig. 1 shows the procedure of the proposed method. By using I and D_L, a depth discontinuity map D_D is generated, and the depth variance at the ith pixel is calculated. They are incorporated into w_{ij} as (2), where $w_{d,ij}$ is the depth discontinuity based colour similarity weight and $w_{v,i}$ is the depth variance weight:

$$w_{ij} = w_{d,ij} w_{v,i} \tag{2}$$

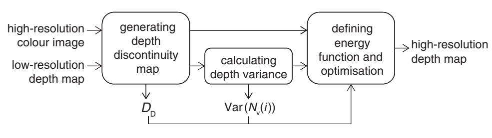

Fig. 1 *Procedure of proposed depth map upsampling*

First, D_D is generated since $w_{d,ij}$ is defined according to D_D. For the I shown in Fig. 2*a* and the corresponding D_L, the generation of D_D is as follows. After finding the colour edges shown in Fig. 2*b*, we decide whether each edge pixel is the depth discontinuity or not. For this decision, a temporarily upsampled depth map D_{TU} of D_L by bilinear interpolation is used, as shown in Fig. 2*c*. Then, D_D is generated by applying (3) to the edge pixels in the colour edge map. In (3), $N_D(i)$ is a rectangular window based on the ith pixel and th_D is a threshold for the depth discontinuity:

$$D_D(i) = \begin{cases} 1 & \text{if } \left| \min_{i \in N_D(i)} D_{TU}(i) - \max_{i \in N_D(i)} D_{TU}(i) \right| > th_D \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

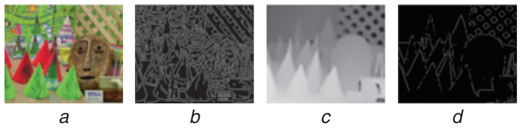

Fig. 2 *Depth discontinuity map generation*

a Colour image
b Colour edges by Canny edge detection
c Temporarily upsampled depth map by bilinear interpolation
d Depth discontinuity map

By using D_D, $w_{d,ij}$ is defined for three cases as (4). The first case is that the ith pixel and $N(i)$ are not related to the depth discontinuity. In this case, only the colour similarity is used to define $w_{d,ij}$ as *Case I* in (4). The second case is that the ith pixel is on the depth discontinuity, as shown in Fig. 3*a*. At the depth discontinuity, more neighbouring pixels are used to carefully measure the colour similarity since there are pixels with blurred colour between two different objects. As indicated in *Case II* in (4) and Fig. 3*a*, m pixels along each neighbouring direction are used. In the last case shown in Fig. 3*b*, the neighbour of the ith pixel is located on the depth discontinuity. To avoid the possibility that the pixels over the depth discontinuity affect the current pixel, a large constant C_{trunc} for truncation is used as *Case III* in (4) to make $w_{d,ij}$ near to zero:

$$w_{d,ij} = \begin{cases} \exp\left(-\dfrac{\sum_{RGB}(I(i) - I(j))^2}{2\sigma_{ij}^2} \right) & \textit{Case I} \\ \exp\left(-\dfrac{\sum_{RGB}\left(I(i) - \frac{1}{m}\sum_{p=1}^{m} I(j_m) \right)^2}{2\sigma_{ij}^2} \right) & \textit{Case II} \\ \exp(-C_{trunc}) & \textit{Case III} \end{cases} \tag{4}$$

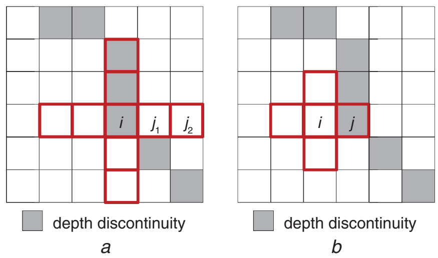

Fig. 3 *Relationship between current pixel and depth discontinuity*

a Current pixel is on depth discontinuity
b Neighbouring pixel is on depth discontinuity

For the next, the depth variance weight $w_{v,i}$ is defined based on the current pixel and its neighbouring region as (5). The depth variance $\mathrm{Var}(N_v(i))$ at the ith pixel is calculated by using the depth values in $N_v(i)$ of D_{TU}. According to (5), $w_{v,i}$ approaches zero and makes w_{ij} very small when $\mathrm{Var}(N_v(i))$ is large:

$$w_{v,i} = \exp\left(-\frac{\mathrm{Var}(N_v(i))}{\sigma_v^2} \right) \tag{5}$$

After defining $w_{d,ij}$ and $w_{v,i}$, E is optimised by using the belief propagation algorithm [4] and D_U is generated. Since the w_{ij} in E reflects the depth discontinuity and the depth variance to the smoothness term, the power of the influence of the smoothness term is increased in the depth homogeneous regions and decreased near the depth discontinuity regions. Also, w_{ij} reduces the depth error near the depth discontinuity regions.

Experimental results: To test the proposed method, four test image sets, ‘Cones’, ‘Teddy’, ‘Venus’ and ‘Tsukuba’, provided by the Middlebury Stereo were used [5]. The ground-truth depth maps were downsampled by factors of 2, 4 and 8 with the nearest neighbour method. During the experiments, the parameters were fixed as follows: $k = 1$, $th_D = 6$, $m = 5$ and $C_{trunc} = 10$. The length of one side of the square windows N_D and N_v is $(s + 1)$, where s means the scaling factor.

The proposed method is compared with the joint bilateral upsampling (JBU) [1], the noise-aware filter for depth upsampling (NAFDU) [2] and the MRF-based depth upsampling [3]. Fig. 4 shows the upsampled depth maps and the error maps of ‘Teddy’ for the scaling factor of 4. The depth error near the depth discontinuity regions and also the depth homogeneous regions were reduced as a result of the proposed method.

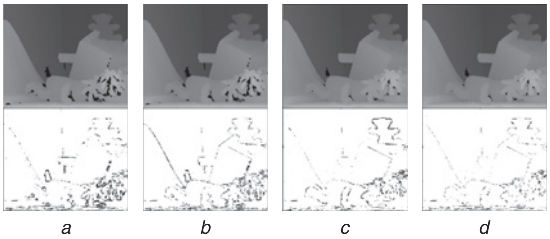

Fig. 4 *Upsampled depth maps and error maps*
a JBU
b NAFDU
c MRF-based depth upsampling
d Proposed method

For an objective evaluation, the bad pixel rate (BPR) was calculated with respect to the ground-truth depth maps. The comparison results are summarised in Table 1 for each scaling factor. The proposed method provided outstanding results in terms of the BPR for all the scaling factors and all the tested images. Moreover, the BPR increment of the proposed method according to the increasing scaling factor is relatively small compared with the other methods.

Table 1: BPR (%)

Scaling factors	Images	Upsampling methods			
		JBU	NAFDU	MRF	Proposed
2	Cones	3.94	3.88	4.26	2.73
	Teddy	3.99	4.35	4.25	3.39
	Venus	0.35	0.46	0.41	0.29
	Tsukuba	4.96	5.21	6.33	4.90
4	Cones	6.42	7.13	6.33	4.43
	Teddy	7.29	8.08	6.72	5.50
	Venus	0.71	0.94	0.67	0.49
	Tsukuba	9.32	10.10	8.10	6.07
8	Cones	11.85	14.39	12.32	9.15
	Teddy	12.36	13.43	13.37	11.07
	Venus	2.40	2.22	1.54	1.22
	Tsukuba	15.76	17.94	12.46	9.10

Conclusion: In this Letter, a depth map upsampling method using depth local features is proposed in order to generate an upsampled depth map of high-resolution. From the experimental results, it has been confirmed that the proposed method can effectively reduce the depth error and also decrease the error increment according to the increasing scaling factors compared with the other depth upsampling approaches.

Acknowledgment: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (2013-067321).

30 November 2013
doi: 10.1049/el.2013.3956
One or more of the Figures in this Letter are available in colour online.

Y.S. Kang, S.B. Lee and Y.S. Ho (*Gwangju Institute of Science and Technology, 123 Cheomdangwagi-ro, Buk-gu, Gwangju, Republic of Korea*)

E-mail: yunsuk@gist.ac.kr

References

1 Kopf, J., Cohen, M.F., Lischinski, D., and Uyttendaele, M.: ‘Joint bilateral upsampling’, *ACM Trans. Graph.*, 2007, **26**, (3), pp. 1–5
2 Chan, D., Buisman, H., Theobalt, C., and Thrun, S.: ‘A noise-aware filter for real-time depth upsampling’. Proc. ECCV Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications, Marseille, France, October 2008, pp. 1–12
3 Diebel, J., and Thrun, S.: ‘An application of markov random fields to range sensing’, *Adv. Neural Inf. Process. Syst.*, 2006, **18**, pp. 291–298
4 Flezenszwalb, P.F., and Huttenlocher, D.P.: ‘Efficient belief propagation for early vision’, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, **24**, (4), pp. 603–619
5 Scharstein, D., and Szeliski, R.: ‘A taxonomy and evaluation of dense two-frame stereo correspondence algorithms’, *Int. J. Comput. Vis.*, 2002, **47**, (1), pp. 7–42