DEPTH MAP SUPER-RESOLUTION VIA MARKOV RANDOM FIELDS WITHOUT TEXTURE-COPYING ARTIFACTS

Kai-Han Lo[1,2], Kai-Lung Hua[1], and Yu-Chiang Frank Wang[2]

[1]Dept. of CSIE, National Taiwan University of Science and Technology, Taipei, Taiwan
[2]Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan

ABSTRACT

The use of time-of-flight sensors enables the record of full-frame depth maps at video frame rate, which benefits a variety of 3D image or video processing applications. However, such depth maps are typically corrupted by noise and with limited resolution. In this paper, we present a learning-based depth map super-resolution framework by solving a MRF labeling optimization problem. With the captured depth map and the associated high-resolution color image, our proposed method exhibits the capability of preserving the edges of range data while suppressing the artifacts of texture copying due to color discontinuities. Quantitative and qualitative experimental results demonstrate the effectiveness and robustness of our approach over prior depth map upsampling works.

Index Terms— Depth Map Super-Resolution, Time-of-Flight (ToF) Sensors, Markov Random Field (MRF)

1. INTRODUCTION

In many 3D image or video processing applications, it is critical to estimate the range data such as depth map for reconstruction or synthesis purposes. Typically, one can produce depth maps either by stereoscopic matching, or using data captured by laser or range sensors. For stereoscopic matching, one requires multiple slightly displaced color images captured by different cameras for determining the disparity between them. However, such matching schemes might fail if image regions are occluded or without explicit texture information. While laser scanning allows one to accurately reconstruct depth information of a single scene, its cost and limitation to static scenes would not be preferable if 3D video processing becomes necessary. On the other hand, range sensors (i.e., depth cameras) are able to capture depth information at video rate in dynamic scenes. Such cameras typically emit infrared light and record the travel time of reflection from any object points, and thus they are called time-of-flight (ToF) sensors/cameras.

Unfortunately, depth maps captured by ToF cameras are usually with lower resolution than those of the corresponding color images. Moreover, random noise is typically presented during data acquisition due to both intrinsic physical constraints and extrinsic environmental interference, which produces disturbing artifacts for the resulting range data .

To address the above problem, depth map upsampling or super-resolution (SR) aims at improving the quality or resolution of the observed range data, using a registered high-resolution (HR) color image. For example, with the recently developed bilateral filtering techniques [1], Kopf *et al.* [2] proposed a joint bilateral upsampling framework for upsampling the depth map while preserving the edges observed from the associated HR color image. Yang *et al.* [3] also advanced joint bilateral filtering with depth hypothesis for iteratively refining the HR depth map. Although promising results were reported, color images might produce false discontinuities in range due to color or lighting variations, which would infer incorrect range results for the above methods.

Different from filtering based approaches, Diebel and Thrun approached the task of depth map SR as solving a multi-labeling optimization problem using Markov Random Fields (MRF) [4]. In recent works like [5, 6], improved depth map SR estimates were obtained by focusing on depth discontinuities when solving MRF. Nevertheless, since the above works did not particularly model the difference between color and edge discontinuities, artifacts due to texture copying tend to be presented in the above estimated outputs.

In our work, we propose a novel MRF-based depth map SR framework. The proposed framework not only provides a more effective way in modeling the data and smoothness MRF energy functions when predicting the range outputs, we further incorporate a weighting scheme which observes and fuses color and depth continuities and thus suppresses texture copying artifacts. Our experiments will later confirm that our method outperforms state-of-the-art depth map SR approaches on a variety of depth images.

2. MRF FOR DEPTH MAP SUPER-RESOLUTION

A Markov Random Field (MRF) is a graphical model describing a joint probability distribution. It consists of an undirected graph model (see Figure 1 for example) in which each node indicates a random variable and the edges determine the as-

This work was supported in part by National Science Council of Taiwan via 101-2221-E-011-138 and NSC101-2221-E-001-018-MY2.

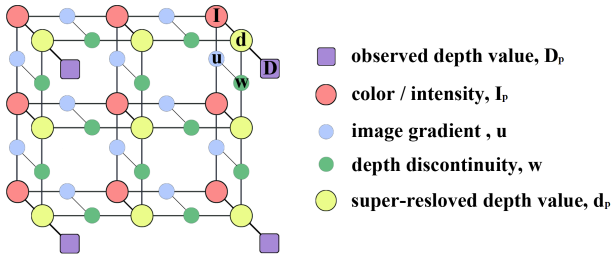

Fig. 1. MRF for depth map super-resolution.

sociated conditional dependencies. MRFs have been widely employed for several image processing tasks such as image segmentation and SR. Recently, this technique is also utilized to solve depth image SR problems [5, 6].

As depicted in Figure 1, MRF approaches depth map SR as solving a multi-labeling optimization problem. The input to the MRF consists of two set of variables: HR image pixels I and LR depth map D. More precisely, each red circle node in Figure 1, I_p, represents the pth pixel of the HR color image. On the other hand, each purple square node D_p denotes the observed LR range data for the pth pixel. The yellow circles d_p indicate the recovered HR depth map which have the same resolution as the HR color image does. The auxiliary nodes for image gradient and depth discontinuity leverage texture and depth information for upsampling the depth map.

According to the Hammersely-Clifford theorem [7], solving MRF is equivalent to optimizing the Gibbs energy function, whose general formulation is defined as follows:

$$E(L) = \sum_{p \in P} U\left(d_p, D_p\right) + \lambda_s \sum_{p,q \in N} V\left(d_p, d_q\right), \qquad (1)$$

where $L = \{L_p | p \in P\}$ indicates the label set of the reconstructed HR depth map, and N is the set of neighboring pixels for the pth pixel. $U\left(d_p, D_p\right)$ is called the data term and indicates the compatibility of the labeling with the given data, that is, the compatibility between the reconstructed depth value and the initial observed depth value. $V\left(d_p, d_q\right)$ is called the smoothness term which incorporates the notion of a piecewise smooth world and penalizes assignments that label neighboring nodes differently. Finally, the parameter, λ_s, is used to balance the data term and smoothness term. In this paper, we employ graph cut optimization algorithm proposed by Veksler et al. [8] to minimize this posterior energy function.

3. PROPOSED MRF FORMULATION

The performance of MRF-based depth map SR relies on the construction of its data term U and the smoothness term V in (1), and several distance measures have been investigated in recent works for addressing this problem. For example, Diebel and Thrun [4] considered quadratic distances, which penalizes the estimation error but results in blurring effects near depth discontinuity regions. Lu *et al.* [5] proposed a truncated absolute difference (TAD) as their distance measure, which has been shown to well preserve the depth discontinuities. However, in addition to non-differentiability, the use of truncated functions requires users to fine-tune the truncating thresholds. In other words, its performance will be sensitive to the choice of such thresholds which are expected to be significantly dependent on the image content [5].

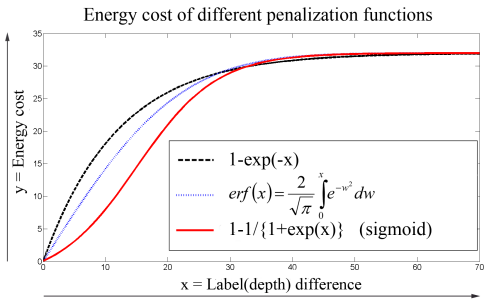

Fig. 2. Three candidate exponential functions. Note that the upper bound is 32 in this example.

To tackle the above issues, we propose to apply exponential-type functions for determining the data and smoothness terms. We will show that this choice not only better characterizes image and depth discontinuities, texture-copying artifacts will be also alleviated with a proposed texture-aware weighting scheme. We now detail our proposed MRF formulation.

3.1. Determinations of Data and Smoothness Terms

In our work, we advance exponential-type functions to suppress the influence of outliers (i.e., extreme estimation errors), which avoids the problems of non-differentiability and sensitiveness to the parameter choices as the method of TAD did. We consider three candidate functions in the exponential family: basic exponential function, error function, and sigmoid function, which are shown in Figure 2. It can be seen that exponential-type functions increase asymptotically toward an upper-bound as the input x increases. We observe that both basic exponential and error functions magnify the label difference x when it approaches 0, but the sigmoid function suppresses such a difference instead.

As suggested in [9, 10], it is preferable to quantize the label estimation difference when solving image synthesis problems. When upsampling the depth map near image/depth discontinuity regions, it is inevitably to have a large number of estimated depth map outputs which are slightly different from the ground-truth ones. As a result, it would be desirable to suppress such errors unless they occur exactly along the edges. Hence a function that suppresses the label difference when it is small is a better choice in this case. Based on the above observations, we select sigmoid functions for penalizing the label differences in our MRF-based depth map SR framework. More precisely, we determine the data and smoothness terms as follows:

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

Fig. 3. An example of texture-copying artifacts in a part of the image *Venus*. (a) Color image, (b) ground-truth depth map, (c) depth map estimated by Lu *et al.* [5], and (d) our result. The up-sampling factor is 4.

$$U\left(d_p, D_p\right) = 1 - \left(\frac{1}{1+\exp\left(\mu\left(\left|d_p - D_p\right| - t_x\right)\right)}\right) - t_y, \quad (2)$$

$$V\left(d_p, d_q\right) = \omega_{p,q}\left(1 - \left(\frac{1}{1+\exp\left(\mu\left(\left|d_p - d_q\right| - t_x\right)\right)}\right) - t_y\right), \quad (3)$$

where μ controls the slope of the sigmoid function, $\omega_{p,q}$ is a weighting constraint for alleviating the texture-copying artifacts (discussed later in Section 3.2), and (t_x,t_y) are the offsets to adjust the error cost to be zero when there is no label estimation error.

3.2. Texture-Aware Scheme for Weighting Constraint

In MRF, the smoothness term V calculates the negative log likelihood of the prior and is used to penalize regularity violations. We determine this term in (3), which preserves the consistency of the estimated depth output close to each other, while taking the image/depth discontinuity into consideration by a weight constraint $w_{p,q}$. We note that, while prior methods [4, 5] typically assumed that the color edges are consistent with the depth discontinuities. However, this might not always hold as examples illustrated in Figure 3. Comparing Figure 3(a) and (b), it is clear that regions with the same depth might possess different colors. Texture-copying artifacts will be produced if not properly handling image and depth discontinuity during depth map SR (as shown in Figure 3(c)). To address this issue, we first evaluate the discontinuity level of depth map for each pixel and determine the value of $w_{p,q}$ accordingly. To be more particular, we calculate the range information of a pixel p around its neighbors by:

$$P_{range} = \max_{p \in W_p}(p) - \min_{p \in W_p}(p), \quad (4)$$

where W_p is the reference window centered at p. The binary map is hence obtained by:

$$P_{edge} = \begin{cases} 0\,, & P_{range} \le \sigma \\ 1\,, & P_{range} > \sigma, \end{cases} \quad (5)$$

where σ is a pre-defined threshold depending on the scale of depth variations. Afterward, this binary map is upsampled to the target resolution by nearest neighbor interpolation. Finally, the weight $w_{p,q}$ is calculated as:

$$\omega_{p,q} = \begin{cases} \exp\left(-\frac{\Delta I_{p,q}}{\gamma}\right), & P_{edge} = 1 \\ \alpha_p \exp\left(-\frac{\Delta I_{p,q}}{\gamma}\right) & \\ \quad + (1-\alpha_p)\exp\left(-\frac{P_{range}}{\gamma}\right), & P_{edge} = 0 \end{cases} \quad (6)$$

where

$$\alpha_p = \frac{P_{range}}{\max_{q \in P}\{Q_{range}\}}. \quad (7)$$

In (6), γ is a constant and $\Delta I_{p,q}$ is the maximum color difference of adjacent pixels across RGB channels.

From (6), we see that when P_{edge} equals one (i.e., there is a relatively large range for the block centered at p), it implies the existence of discontinuity in depth. Thus, the weight $w_{p,q}$ will be negatively correlated to the value $\Delta I_{p,q}$. In other words, if the location of the pixel p shows a strong edge in color, the resulting smoothness term will be weighted by a small value, which encourages the depth discontinuity to be aligned with such color edges. On the other hand, when P_{edge} approaches zero, this would avoid texture copying by assigning a larger value for the second term of (6) when determining $w_{p,q}$. From the above discussion, it can be seen that our proposed MRF framework will be able to better preserve the color and depth discontinuities, while not suffering from the artifacts of texture copy as prior MRF-based approaches did. Moreover, our MRF will not be sensitive to the pre-defined threshold σ due to the proposed blending weighting scheme when P_{edge} approaches zero.

4. EXPERIMENTAL RESULTS

To evaluate the performance of our method, we consider images from the Middlebury stereo dataset [10], which provides HR color and depth image pairs with resolution about 300×400 pixels. We downsample the depth map into lower resolution ones, and use different upsampling factors for performing depth map SR. We use the same parameter setting for all images in our experiments, and we have $\lambda_s = 1.0$, $\mu = 0.2$, $\sigma = 10$, $\gamma = 20$. Besides qualitatively evaluating the output SR depth maps by visual comparisons, we also apply the bad pixel (BP) percentage as the metric for assessing the performance. As suggested in [9, 10], this is achieved by scaling the output depth map into a particular range in terms of depth and determining the number of depth pixels which differ from the ground-truth ones by 1.

We consider the approaches of bicubic interpolation, Diebel *et al.* [4], JBU [2], NAFDU [11], Lu *et al.*[5], and a recent work of Kim *et al.*[6] for both qualitative and quantitative comparisons. We list the BP rates of different images in Table 1. From this table, we can see that our method achieved the lowest or comparable error rates as others did. To visualize our results for alleviating texture-copying artifacts, we show examples produced by different methods in Figure 4.

Table 1. Performance comparisons of error rate for *Venus*, *Teddy*, and *Cones* with upsampling factors 4 and 8.

Algorithm	Venus						Teddy						Cones					
	8×			4×			8×			4×			8×			4×		
	nonocc.	all	disc.	nonocc.	all	disc.	nonocc.	all	disc.	nonocc.	all	disc.	nonocc.	all	disc.	nonocc.	all	disc.
Bicubic	1.47	1.86	20.52	0.66	0.92	9.25	11.82	12.61	41.88	6.27	6.95	23.12	15.33	16.04	44.28	8.34	8.93	25.08
Diebel *et al.* [4]	2.10	2.44	7.52	0.85	1.16	3.93	15.95	15.99	31.59	7.46	8.17	18.02	11.99	13.84	30.06	6.98	8.10	19.82
JBU [2]	1.10	1.61	12.44	0.40	0.72	5.62	11.75	12.68	36.40	5.75	6.61	20.08	12.93	14.74	33.34	6.43	7.54	19.18
NAFDU [11]	0.98	1.46	12.51	0.41	0.69	5.74	11.45	12.37	36.04	5.64	6.48	19.75	12.50	14.31	32.67	6.24	7.31	18.78
Lu *et al.* [5]	0.98	1.36	7.29	0.24	0.31	3.27	13.73	14.93	31.58	5.14	5.60	14.47	10.95	12.67	23.55	3.73	4.51	10.07
Kim *et al.* [6]	0.60	0.74	5.83	0.17	0.30	2.31	9.80	**10.47**	**25.63**	5.33	6.20	16.86	8.45	9.46	22.89	4.86	5.33	14.62
Ours	**0.39**	**0.49**	**4.74**	**0.12**	**0.16**	**1.67**	**9.49**	10.77	26.39	**3.35**	**3.69**	**9.59**	**6.92**	**8.16**	**18.36**	**3.15**	**3.82**	**9.43**

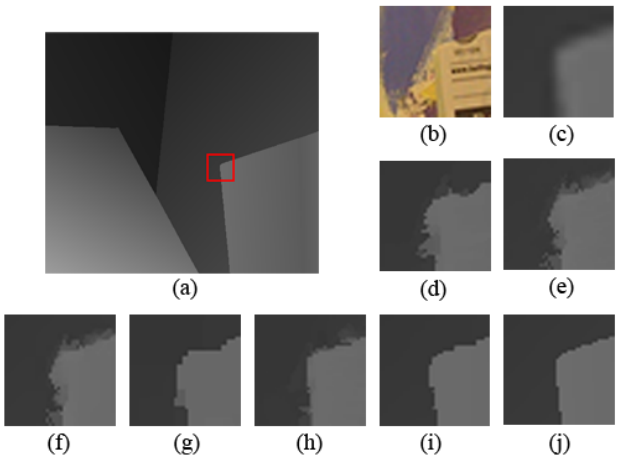

Fig. 4. An example region of *Venus* with similar color but inconsistent depths. Ground-truth depth map with the example region (squared in red) and the corresponding color image are shown in (a) and (b), respectively. Depth maps produced by (c) Bicubic (BP % = 0.92), (d) Diebel *et al.* [4] (1.16), (e) JBU [2] (0.72), (f) NAFDU [11](0.69), (g) Lu *et al.* [5] (0.31), (h) Kim *et al.* [6] (0.30), (i) ours (0.16), and (j) the ground truth.

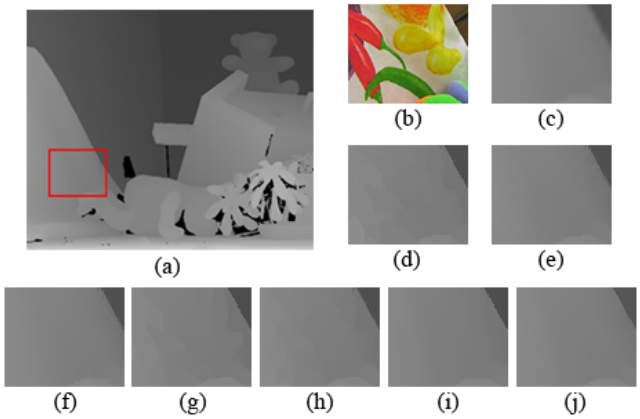

Fig. 5. An example region of *Teddy* with the same depth but different color information. Ground-truth depth map with the example region and the corresponding color image are shown in (a) and (b), respectively. Depth maps produced by (c) Bicubic (BP % = 6.95), (d) Diebel *et al.* [4] (8.17), (e) JBU [2] (6.61), (f) NAFDU [11] (6.48), (g) Lu *et al.* [5] (5.60), (h) Kim *et al.* [6] (6.20), (i) ours (3.69), and (j) the ground truth.

By examining Figures 4(d), (g), and (h), texture-copying artifacts were produced in the depth maps due to similar color presented in the region of interest (ROI). Our result in Figure 4(i) was robust to such effects and thus was closest to the ground truth. Another example is shown in Figure 5, in which a ROI with the same depth contains different color information. Comparing the results shown in this figure, we see that our proposed MRF framework with texture-aware weighting schemes alleviated this problem, and thus our output was the best among different approaches.

5. CONCLUSION

We presented a novel learning-based depth map SR framework, which is able to synthesize a HR depth map given its LR version and a corresponding HR color image. By advancing sigmoid functions in the proposed formulation, our MRF suppresses extreme estimation errors while not magnifying those near depth discontinuities. By incorporating a texture-aware weighting constraint into our proposed framework, artifacts of texture copying can be significantly alleviated. Our experiments confirmed that our approach achieved promising results, and it was shown to quantitatively and qualitatively outperformed state-of-the-art depth map SR works.

6. RELATION TO PRIOR WORK

When applying MRF for up-sampling depth maps, one needs to determine data and smoothness terms for estimating or synthesizing the depth map output. Although metrics like l_2-norm or truncated absolute distance have been investigated in [4, 5], their lack of robustness in suppressing extreme estimation errors or ad-hoc threshold selection would limit the performance. In our proposed MRF work, we advance sigmoid functions for addressing the above issues. We further address the problem of texture-copying, which is not explicitly studied in recent works [4, 5, 6]. The presence of such artifacts is due to discontinuity or inconsistency in images or depths, and we introduce a texture-aware weighting scheme for alleviating such artifacts. This scheme allows us to calculate the smoothness term of MRF based on color and depth image inputs, and thus improved SR performance can be achieved.

7. REFERENCES

[1] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *IEEE ICCV*, 1998, pp. 839–846.

[2] J. Kopf, M. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," in *ACM SIGGRAPH*, 2007.

[3] Q. Yang, R. Yang, J. Davis, and D. Nister, "Spatial-depth super resolution for range images," in *IEEE CVPR*, 2007, pp. 1–8.

[4] J. Diebel and S. Thrun, "An application of Markov random fields to range sensing," in *MIT Press NIPS*, 2005, pp. 291–298.

[5] J. Lu, D. Min, R. S. Pahwa, and M. N. Do, "A revisit to MRF-based depth map super-resolution and enhancement," in *IEEE ICASSP*, 2011, pp. 985–988.

[6] D. Kim and K. Yoon, "High quality depth map up-sampling with consideration for edge noise of range sensors," in *IEEE ICIP*, 2012, pp. 553–556.

[7] J. M. Hammersley and P. Clifford, "Markov fields on finite graphs and lattices," 1971.

[8] O. Veksler, A. Delong, and A. Mueller, "Code of multi-label optimization," `http://vision.csd.uwo.ca/code/`.

[9] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," in *IJCV*, 2001, pp. 7–42.

[10] "Middlebury stereo," `http://vision.middlebury.edu/stereo/`.

[11] D. Chan, H. Buisman, C. Theobalt, and S. Thrun, "A noise-aware filter for real-time depth upsampling," in *ECCV Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications*, 2008.