# An improved MRF model for robust color guided depth up-sampling

Yi Chen, Li Yu, Shengwei Wang

*School of Electron. Inf. & Commun., Huazhong Univ. of Sci. & Tech.*
*Wuhan, Hubei, China 430074*

{chenyi, hustlyu, kadinwang}@hust.edu.cn

***Abstract*— Color guided depth up-sampling always suffers from texture copy artifacts and depth discontinuities blurring. This phenomenon is due to the depth discontinuity and color image edges at the corresponding location are not always consistent. In this paper, based on the analysis of the above two problems, we proposed an improved Markov Random Filed (MRF) model, which can reduce negative influence from pixels at the inconsistent location more effectively. Furthermore, the improved MRF model can also reduce the negative influence of noises from depth map. To determine pixels of inconsistent regions, a new concept of pixel confidence is proposed. Pixel confidence indicates the probability that pixel is at the inconsistent regions, which is embedded into the improved MRF model. The proposed method is tested on both the simulated and real datasets. The proposed method can better suppress texture copy artifact and preserve sharp depth discontinuities. Experimental results show that the proposed method also has lower mean absolutely error(MAE) than other methods in most cases.**

***Index Terms*—Depth map up-sampling, texture copy artifacts, depth discontinuities blurring, Markov Random Filed, pixel confidence**

## I. Introduction

With the rapid development of time-of-flight (ToF) cameras, depth map is widely used in 3-D scene reconstruction. However, it's difficult to achieve a precise and high-resolution depth image directly. Thus, depth up-sampling has attracted more and more attention. Many up-sampling methods have been also proposed.

A mainstream approach is color guided up-sampling which gets high resolution depth map with the guidance of the registered color(aligned) image. Joint Bilateral Up-sampling (JBU) [1] extended the Bilateral Filter (BF) [2] for depth up-sampling where the bilateral weights are based on the guidance color image. Diebel and Thrun [3] performed the restoration using Markov Random Fields (MRF) with a pairwise appearance consistency data term and an image guided smoothness term. Park *et al.* [4]proposed weighted least squares model (WLS) which regularized the depth map with a nonlocal structure regularization term to maintain fine details and structures. Yang *et al.* [5] performed color guided depth map restoration using a color guided Auto-Regressive (AR) model which take input depth map into account. Color guided depth map up-sampling may suffer from texture copy artifact and depth discontinuities blurring when depth discontinuities and the corresponding image edges are not consistent. To tackle these two problems, most existing methods focus on designing various guidance weight based on guidance color images and heuristically took the bi-cubic interpolation of the input depth map into account, for instance, the definition of the AR coefficient in the color guided AR model proposed by Yang *et al.* [5], the weight of the smoothness term in the weighted least squares model proposed by Park *et al.* [4]. However, complex guidance weight doesnt always help to improve the up-sampling quality. Noises from corrupted depth map will also be introduced to make the overall accuracy worse.

In this paper, to both handle the above two problems and reduce the noises introduced from depth map, we propose a robust MRF model, which control the effort of color guided image adaptively. The smooth term is based on a proposed new concept of pixel confidence, which indicates the probability that pixel is at the consistent regions. Color image pixels at the consistent location will be applied to restrict the depth guidance weight of smooth term, which could reduce negative influence from noises. To inconsistent regions, the guidance weight of the smooth term is only based on depth map to better avoid causing texture copy artifact. The main contributions of this paper can be summarized as follows:(1) by considering whether pixel is at the consistent location, we propose a robust MRF model. The proposed MRF model can both reduce negative influence from noisy depth map and better alleviate texture copy artifact which is caused by pixels at the inconsistent location. (2)To determine pixels which are used to constraint influence from the depth guidance weight, we propose a new concept of pixel confidence, which indicates the probability that the pixel is at the consistent regions. Color image edge pixels confidence is determined by the local edge similarity between color image and depth map. Non-edge pixels confidence is determined by the local smoothness of depth map. Experimental results show that our method can better suppress texture copy artifact and depth discontinuities blurring. Additionally, our method can reduce the noises introduced from the depth map, which can make the overall accuracy better than the state-of-art methods.

978-1-5386-0462-5/17/$31.00 ©2017 IEEE.

## II. PROPOSED METHOD

To reduce negative influence from both noise and pixels located at the inconsistent regions, we firstly determine which pixels are located at the consistent regions. The judgment indicator is the newly proposed concept of pixel confidence. Based on pixel confidence, an adaptive unified weighting framework based on MRF is proposed to tackle texture copy artifact and avoid introducing noises. Edge pixels confidence is determined by local structural similarity between color edge map and depth edge map. To non-edge pixel, pixel confidence is determined by local smoothness of depth map.

### *A. Improved MRF by adopting pixel confidence*

Markov Random Fields (MRF) including its variants is one of major approaches, which has dominated this area for several years. According to the Hammersely Clifford theorem [6], solving MRF is equivalent to optimizing the Gibbs energy function, whose general formulation is defined as

$$\begin{aligned} D^H = \arg\min_{D} \{ & \sum_{i \in \Omega^0} (D_i - D_i^0)^2 \\ & + \lambda \sum_{i \in \Omega} \sum_{j \in N(i)} a_{i,j} (D_i - D_j)^2 \} \end{aligned} \tag{1}$$

where $D^0$ is sparse samples in the high resolution coordinate, projected from the measured low resolution depth map. $\Omega$ represents coordinate of the high resolution depth map.$\Omega^0$ represents the coordinate where depth measurement is available. $N(i)$ is the neighborhoods of $i$ in the square patch centered at $i$.

It has been proved by Liu *et al.* [7] that the weight coefficient $a_{i,j}$ is the main cause of texture copy artifact and depth discontinuities blurring. Furthermore, negative influence from noises is also introduced by influencing the guidance weight. The guidance weight $a_{i,j}$ is based on information from the color image and depth map. The WLS [4] and AR [5] model are two typical categories.

The WLS coefficient is defined as

$$a_{i,j}^c = \exp(-\frac{|i-j|^2}{2\sigma_s^2}) \exp(-\frac{\sum_{k \in C} |I_i^k - I_j^k|^2}{3 \times 2\sigma_c^2}) \tag{2}$$

The AR coefficient is defined as

$$a_{i,j} = \frac{1}{S_i} \hat{a_{i,j}^D} a_{i,j}^I \tag{3}$$

$$\hat{a_{i,j}^D} = \exp(-\frac{|\hat{D_i} - \hat{D_j}|^2}{2\sigma_1^2}) \tag{4}$$

$$a_{i,j}^I = \exp(-\frac{\sum_{k \in C} ||B_i o (P_i^k - P_j^k)||^2}{3 \times 2\sigma_2^2}) \tag{5}$$

From the above formulations, its obvious that the WLS coefficient contains of color guidance weight. While the AR coefficient contains of color guidance weight and depth guidance weight. Compared with the WLS model, depth guidance weight $\hat{a_{i,j}^D}$ can provide more information, which can better suppress texture copy artifact and depth edge blurring. However, when the depth map is corrupted by strong noises, the WLS model can perform better on the overall up-sampling accuracy.

Based on the above analysis, the improved guidance weight is expressed as

$$a_{i,j} = a_{i,j}^D \times (1 - confidence_j + a_{i,j}^I \times confidence_j) \tag{6}$$

$$a_{i,j}^I = \exp(-\frac{\sum_{k \in C} |I_i^k - I_j^k|^2}{3 \times 2\sigma_c^2}) \tag{7}$$

$$a_{i,j}^D = \exp(-\frac{|D_i - D_j|^2}{2\sigma_d^2}) \tag{8}$$

where $D$ is the interpolation of the input depth map. Notice that $0 \leq a_{i,j}^I \leq 1$, $0 \leq a_{i,j}^D \leq 1$. Apparently, our new guidance weight $a_{i,j}$ is relevant to pixel confidence. If the color pixel confidence is near to 0, the color guidance weight has little even no influence on the guidance weight $a_{i,j}$. If the color pixel confidence is near to 1, then the guidance weight is more close to $a_{i,j}^I a_{i,j}^D$. Since the color image can provide more positive information to up-sampling. Additionally, the color guidance weight $a_{i,j}^I$ can restrict the depth guidance weight, which can reduce negative influence from noises in the depth map.

### *B. The color edge pixel's confidence*

Edge pixels confidence is determined by the local structure similarity between color edge map and depth edge map. Jang *et.al* [8] proposed a novel metric named SEQM for edge maps, which models structure similarity measurement as edge map quality assessment in bi-direction evaluation. SEQM is based on the basic assumption that the edges have no deformation. Since the interpolation of input depth map doesn't satisfy it, the edge pixel confidence measurement is modeled as edge map quality assessment in one-way evaluation in this paper.

Every pixel's confidence is firstly initialized to 0.5. Canny operator [9] is applied in color image and coarse up-sampled depth map to generate relevant edge maps. The colour edge pixels confidence is composed of the positional matching cost $C_p(s,m)$ and structural matching cost$C_q(s,m)$. $s$ denotes an edge pixel in color edge map. $m$ denotes an edge pixel in depth edge map, which is a matching candidate to $s$. $m$ is searched in the $L \times L$ patch which is centered on the same position with $s$. The size of $L$ is 3 when the up-sample factor is $2\times$ and $4\times$. When the up-sample factor is $8\times$ and $16\times$, $L$ is 5.

The positional matching cost $C_p(s,m)$ is formulated as

$$C_P(s,m) = \frac{1}{K} \sqrt{(x(s) - x(m))^2 + (y(s) - y(m))^2} \tag{9}$$

where $(x(s), y(s))$ and $(x(m), y(m))$ are the coordinates of $s$ and $m$, respectively. $K$ is the constant parameter that controls the importance of $C_p(s,m)$ in the overall matching, which is 10.

The structural matching cost $C_q(s,m)$ is based on the local edge similarity centered at $s$. $B_s$ and $B_m$ denote the $3 \times 3$

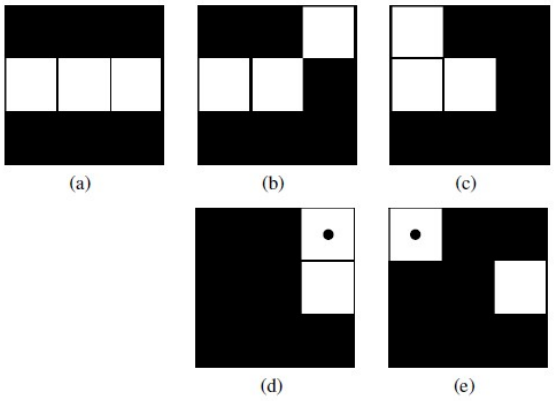


Fig. 1. (a) color edge block $B_s$. (b) depth edge block $B_{m1}$. (c) depth edge block $B_{m2}$. (d) $B_s - B_{m1}$. (e) $B_s - B_{m2}$.

blocks centered at $s$ and $m$, respectively. The difference block is formulated as

$$B_d = B_s - B_m \tag{10}$$

Pixel values in $B_s$ and $B_m$ are binary: 0 for a non-edge pixel and 1 for an edge pixel, which are depicted by black and white pixels in this work, respectively. Thus, pixel values in $B_d$ are ternary, i.e. $-1$, 0, or 1. The set of pixels in block $B_d$ are partitioned into $V^-$, $V^0$ and $V^+$. Jang proved that 1 pixels should be near to $-1$ pixels in [8]. Fig. 1 shows that depth edge block $B_{m1}$ has higher structural similarity than $B_{m2}$ does. So we need to search one to one matching between $V^+$ and $V^-$.

But not the same with [8], we allow several 1 pixels to match the same $-1$ pixels. The improved structural cost is expressed as

$$C_q(s,m) = \frac{\sum_{(v_i^-, v_j^+) \in M} w(v_i^-, v_j^+)}{16} \tag{11}$$

$$w(v^-, v^+) = H(|x(v^+) - x(v^-)| + |y(v^+) - y(v^-)|) \tag{12}$$

where $H(d)$ is a monotonic function according to the $L_1$distance between $V^+$ and $V^-$. Let $H(1) = 1$, $H(2) = 1.6$, $H(3) = H(4) = 2$. The optimal matching pairs are determined based on Hamming distance. Specifically, if $V^-$ is null, we set $w(v^-, v^+) = 2$ directly. $M$ is the set of the optimal matching pairs. If $B_s$ and $B_m$ are exactly the same, $C_q(s,m) = 0$.

Pixel confidence indicates the probability that the pixel is located at the consistent regions. Structural matching cost $C_q(s,m)$ is influenced by local structure similarity as well as noise. if the local structure of color edge map is more similar to the local structure of depth edge map , the pixel confidence should be larger and more stable. On the contrary, when the local structure similarity is small, the pixel confidence should be sensible to the structural matching cost $C_q(s,m)$ as well as noise.

Thus, the improved pixel confidence is expressed as

$$confidence = (1 - C_p(s,m)) \times \frac{1 - e^{C_q(s,m)-1}}{1 - e^{-1}} \tag{13}$$

If the structural matching cost is more close to 1, the pixel confidence should decrease more rapidly. However, when the structural matching cost is close to 0, the confidence should not decrease too rapidly. Since the structural matching cost value may be influenced slightly by noises. The improved pixel confidence can better satisfy these two conditions.

### C. Color non-edge pixel's confidence

The depth map is often corrupted by strong noises and its edges have deformation. so its edges are not reliable. Thus,the color non-edge pixels confidence measurement can't use the same algorithm as the edge pixel directly.

Motivated by [7], $s$ denotes a non-edge pixel in color edge map. Also, $t$ denotes the depth pixel in the same position. We calculate the relative smoothness $\mu(s,t)$. If (s; t) is small, the non-edge pixel is more probably located at the inconsistent regions to a certain extent.

The color non-edge pixel confidence is expressed as

$$confidence = \frac{1}{2}\mu(s,t) \tag{14}$$

$$\mu(s,t) = \frac{\sum\limits_{m \in N_t(x_{\min})} x_m}{\sum\limits_{n \in N_t(x_{\max})} x_n}, s \in \Omega \tag{15}$$

where $x_{\min}$ and $x_{\max}$ are the minimal and maximal depth values inside $N(t)$. $N(t)$ is the neighborhood of pixel $t$. $N_t(x_{\min})/N_t(x_{\max})$ denotes a small patch of radius $r_\mu \times r_\mu$ centered at the pixel whose value is $x_{\min}/x_{\max}$.

## III. Experimental Results

In this section, we present the quantitative and visual comparisons between the proposed method and other state-of-the-art methods including: the AR model [5], the WLS in [4], the image guided anisotropic total generalized variation (TGV) up-sampling in [11] and the joint geodesic up-sampling (JGU) in [12].

The proposed method is firstly tested on the simulated noisy ToF dataset from [10]. Table 1 shows the results of our method as well as other compared methods. The best results are in bold. The accuracy is measured by mean absolutely error (MAE). The results show that our methods outperforms other compared methods in most cases. Fig. 2 shows the visual comparison of $8\times$ up-sampling results.

From Table 1 and Fig. 2, we can find that the WLS can get better accuracy than the AR model in most cases. But the AR model performs better on handling the texture copy artifacts and depth discontinuities. By taking depth map into account and adjusting the color guidance weight based on pixel confidence, the proposed method can both suppress the two problems and avoid introducing noise to the high resolution depth map.

To further test our method, we also perform experiments on real ToF dataset [13] which contains devil, shark. Table 3 shows the RMSE in terms of mm. The proposed method performs better on both devil and shark.

TABLE I
QUANTITATIVE COMPARISON ON REAL TOF DATASET. THE ERROR IS CALCULATED AS MAE TO THE MEASURED GROUNDTRUTH IN MM. THE BEST RESULTS ARE IN BOLD

| | art | | | | dolls | | | | book | | | | moebius | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2× | 4× | 8× | 16× | 2× | 4× | 8× | 16× | 2× | 4× | 8× | 16× | 2× | 4× | 8× | 16× |
| JGU | 1.48 | 1.96 | 3.05 | 5.07 | 0.95 | 1.38 | 2.16 | 3.2 | 0.94 | 1.39 | 2.2 | 3.32 | 0.97 | 1.4 | 2.18 | 3.3 |
| TGV | 0.95 | 1.69 | **2.16** | 4.74 | **0.89** | 1.25 | 1.53 | 3.03 | 0.76 | 1.03 | 1.48 | 2.34 | 0.8 | **0.92** | 1.58 | 2.89 |
| AR | 1.32 | 1.86 | 3.08 | 5.47 | 1.12 | 1.39 | 2.06 | 2.89 | 0.83 | 1.06 | 1.59 | 2.74 | 1.1 | 1.41 | 2.06 | 3.06 |
| WLS | 1.4 | 1.88 | 2.74 | 4.86 | 1 | 1.36 | 1.83 | 2.28 | 0.89 | 1.25 | 1.6 | 2.28 | 0.95 | 1.33 | 1.82 | 2.72 |
| OURS | **0.94** | **1.62** | 2.53 | **4.71** | 0.94 | **1.22** | **1.48** | **2.21** | **0.71** | **0.99** | **1.44** | **2.14** | **0.79** | 1.06 | **1.49** | **2.59** |

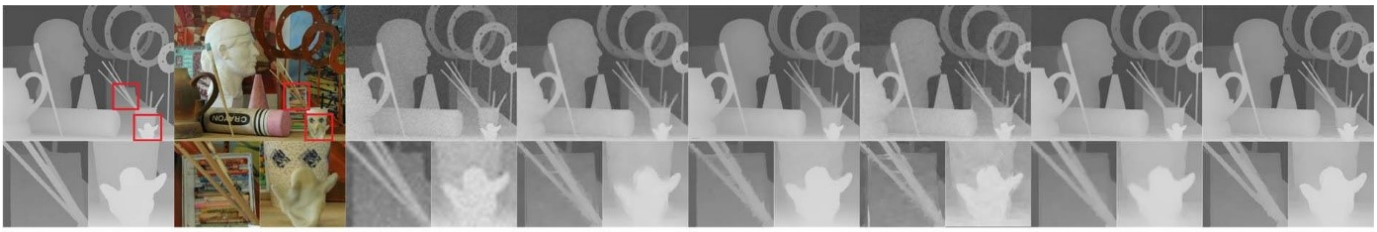


Fig. 2. 8 up-sampling of Art on the Middlebury dataset [10]. (a) The groundtruth depth map. (b) The corresponding color image. (c) The bicubic interpolation of the noisy low resolution depth map. The result obtained by (d) the WLS [4], (e) the TGV upsampling [11], (f) the JGU [12], (g) the AR model [5] and (h) our method. Regions in red boxes are highlighted.

The quantitative and qualitative comparisons show that the proposed method has promising performances both on the simulated and real datasets. Especially, the proposed method can better suppress texture copy artifact and depth edge blurring. Additionally, when the input low resolution depth map contains strong noises and structure missing, the proposed method still can smooth noises and keep complete structure.

TABLE II
QUANTITATIVE COMPARISON ON REAL TOF DATASET. THE ERROR IS CALCULATED AS **MAE** TO THE MEASURED GROUNDTRUTH IN **mm**. THE BEST RESULTS ARE IN BOLD

| | JGU | TGV | AR | WLS | OURS |
|---|---|---|---|---|---|
| Devil | 16.28 | 14.97 | 15.59 | 15.38 | **14.76** |
| Shark | 17.25 | 15.53 | 16.28 | 15.99 | **15.12** |

## IV. CONCLUSION

In order to suppress texture copy artifact as well as depth discontinuities blurring more efficiently and directly, a new concept of pixel confidence is proposed to determine which pixels are likely to cause negative influence. The pixel confidence stands for the probability that the pixel is located at the consistent regions. In the following steps, the pixel confidence is embedded into MRF model to control the effort of color guidance weight adaptively. By reducing negative influence from pixels located at the inconsistent regions and restricting the depth guidance weight, the proposed method can suppress texture copy artifact more efficiently and keep sharp depth discontinuities. The future work will focus on how to ensure pixel confidence more accurate by efficiently filling missing structure and smoothing noises.

## ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (NSFC) (No. 61231010), National High Technology Research and Development Program (No. 2015AA015903).

## REFERENCES

[1] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," in *Proc.ACM Trans.Graph*, 2007, p. 96.
[2] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc.ICCV,Jan.*, 1998, pp. 839–846.
[3] J. Diebel and S. Thrun, "An application of markov random fields to range sensing," in *NIPS*, 2006, pp. 291–298.
[4] J. Park, H. Kim, Y.-W. Tai, M. S. Brown, and I. Kweon, "High quality depth map upsampling for 3d-tof cameras," in *Proc.ICCV*, Nov.2011, pp. 1623–1630.
[5] J. Yang, X. Ye, K. Li, C. Hou, and Y. Wang, "Color-guided depth recovery from rgb-d data using an adaptive autoregressive model," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3443–3458, Aug. 2014.
[6] J. M. Hammersley and P. Clifford, "Markov fields on finite graphs and lattices," 1971.
[7] W. Liu, X. Chen, J. Yang, and Q. Wu, "Variable bandwidth weighting for texture copy artifacts suppression in guided depth upsampling," *IEEE Trans. Circuits Syst. Video Technol*, 2016.
[8] W. D. Jang and C. S. Kim, "Seqm: Edge quality assessment based on structural pixel matching," in *Visual Communications and Image Processing (VCIP)*. IEEE Conference on, 2012, pp. 1–6.
[9] J. Canny, "A computational approach to edge detection," *Pattern Analysis and Machine Intelligence IEEE Transactions on*, no. 6, pp. 679–698, 1986.
[10] D. Scharstein and C. Pal, "Learning conditional random fields for stereo," in *CVPR*, 2007, pp. 1–8.
[11] D. Ferstl, C. Reinbacher, R. Ranftl, M. Rüther, and H. Bischof, "Image guided depth upsampling using anisotropic total generalized variation," in *Proc.ICCV*, 2013, pp. 993–1000.
[12] M.-Y. Liu, O. Tuzel, and Y. Taguchi, "Joint geodesic upsampling of depth images," in *Proc.CVPR*, 2013, pp. 169–176.
[13] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from rgbd images," *Proc.ECCV*, pp. 746–760, 2012.