Research Highlights (Required)

To create your highlights, please type the highlights against each `\item` command.

It should be short collection of bullet points that convey the core findings of the article. It should include 3 to 5 bullet points (maximum 85 characters, including spaces, per bullet point.)

- A supervised classification based MRF model is proposed for depth map up-sampling.
- Low-resolution depth segmentation is used to supervise color image classification.
- Classification result is further introduced in the design of MRF energy function.
- The designed MRF energy function is optimized with a gradient descent algorithm.
- Comparisons with the state-of-the-art show the superiority of the proposed SC-MRF model.

Pattern Recognition Letters
journal homepage: www.elsevier.com

High-quality depth up-sampling via a supervised classification guided MRF model

Yiguo Qiao[a,**], Licheng Jiao[a], Xu Tang[a], Wenbin Li[b], Darren Cosker[b]

[a] *Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, International Research Center for Intelligent Perception and Computation, Joint International Research Laboratory of Intelligent Perception and Computation, School of Artificial Intelligence, Xidian University, No.2 South Taibai Road, Xi'an, Shaanxi Province 710071, China*
[b] *Centre for the Analysis of Motion, Entertainment Research and Applications (CAMERA), University of Bath, Claverton Down, Bath, BA2 7AY, UK*

ABSTRACT

In this paper, a Supervised Classification assisted Markov Random Field (SC-MRF) model is proposed for generating high-quality up-sampled depth maps. The proposed model aims to reduce depth bleeding and depth confusion artifacts that can be produced at boundary regions of the up-sampled depth maps. In the proposed model, segmentation of low-resolution (LR) depth map is first used to supervise the classification of corresponding high-resolution (HR) color image. With this supervised classification, not only can the depth edges be retained, but redundant textures in the HR color image can be omitted. The classification result is then introduced into the design of a MRF energy function, and the final up-sampled depth map is obtained by optimizing this energy function with the gradient descent algorithm. For simplicity, classical K-means clustering is adopted to segment the LR depth map into several classes, and a feature-based K-nearest neighbour (K-NN) method is utilized for the supervised classification. With the proposed SC-MRF model, interaction between depths of different classes will be strongly suppressed, meaning depth edges are well preserved. Comparisons with the state-of-the-art demonstrate the strong performance of the proposed method both visually and by quantitative evaluation.

Keyword: Depth map up-sampling; Markov random fields; supervised classification; gradient descent

1. Introduction

Depth cues are widely used across applications such as 3D movies, 3D games, depth-aided object recognition, and RGB-D image segmentation (Shao et al., 2012; Dominio et al., 2014; Mahmoudpour and Kim, 2016; Zhu et al., 2017). Due to application requirements, obtaining high-quality depth maps is of crucial importance (Kim et al., 2014; Yang et al., 2015; Yuan et al., 2017). In order to acquire HR depth maps, fusion camera devices like Microsoft Kinect and Intel RealSense have been developed (Sarbolandi et al., 2015; Fankhauser et al., 2015). Kinect pays more attention to target tracking at longer distances, while RealSense focuses on face and hand tracking at close distances. The latest Kinect v2 has a wide field of view (FoV) and captures depth maps based on time-of-flight (ToF) technology (Min-Koo et al., 2014). As low-resolution is a limitation of depth sensors, ToF makes use of a regular RGB camera and a depth sensor to capture both HR RGB images and LR depth maps, so that LR depth maps can be up-sampled under the guidance of HR GRB images (Chan et al., 2008; Choi and Jung, 2014; Kim et al., 2014; Eichhardt et al., 2017).

Many depth up-sampling works have been reported in the literature. The existing up-sampling methods can be roughly classified into two categories, static interpolation based and dynamic optimization based. Joint bilateral up-sampling (JBU) (Kopf et al., 2007) and the traditional MRF based up-sampling (Flezenszwalb and Huttenlocher, 2002) are two classical methods towards static interpolation and dynamic optimization, respectively. In JBU, the LR depth map is interpolated via a bilateral filter, which is a product of two Gaussian kernel functions determined according to pixel intensity and spacial distance, respectively. On the other hand, the MRF based up-sampling method relies on a coarse depth map initially generated by trivial interpolations such as the nearest neighbor interpolation, the bilinear interpolation, and so on. Then the precision of the

[**] Corresponding author: Tel.: +44-0225 383214;
e-mail: yiguo.qiao@bath.edu (Yiguo Qiao)

coarsely up-sampled depth map is formulated as the optimization of a designed MRF energy function. In the traditional MRF based model, only the pixel intensity and the spacial distance are taken into consideration in the design of the energy function.

Joint geodesic up-sampling (JGU) (Liu et al., 2013) makes use of the several nearest seed pixels to interpolate the target ones (seed pixel denotes the one whose depth can be traced from the LR depth image; target pixel denotes pixel with unknown depth). Moreover, a geodesic distance is defined for the measurement of distance between two arbitrary pixels. There are also some other static interpolation based methods like the joint trilateral filtering (JTF) (Lo et al., 2017). In such methods, the LR depth map is first coarsely interpolated, and the incorrect depth edges are then refined per-pixel.

Unlike the above static interpolation methods, depth up-sampling can also be formulated as a dynamic optimization problem, such as in the anisotropic Total Generalized Variation (TGV) based method (Ferstl et al., 2013) and the SD (Static/Dynamic) filter based method (Ham et al., 2018). In TGV, a total generalized variation regularization term is included in the energy function which is then optimised using an anisotropic diffusion tensor. In SD based up-sampling, an SD filter is used to smooth the coarsely up-sampled depth map iteratively. In the construction of the SD filter, the HR color image is used as static guidance, and the smoothed depth map at each iteration is used as the dynamic guidance.

As depth bleeding and depth confusion[1] is easily produced at the boundaries of the up-sampled depth map, depth edge-preserving becomes the focus issue of depth up-sampling. The ideal situation is where the depth edges in up-sampled depth map coincide with the correlated edges in HR color image. However, both useful depth edges and unexpected image textures may be present in HR color image, thus the problem becomes how to remove the redundant image textures while preserving the useful depth edges. To solve this problem, a data-driven method has been proposed, which learns a dictionary of geometric primitives to capture the overlapping edges of RGB image and depth map(HyeokHyen Kwon et al., 2015). Besides, some segmentation based up-sampling methods have also been introduced (Park et al., 2011; Soh et al., 2012; Buyssens et al., 2015). Superpixel strategies are widely used for RGB image segmentation, so that the redundant image textures can be partially removed. Other segmentation strategies, like the patch based joint-segmentation, have also been involved (Tallón et al., 2012). These segmentation methods have a common feature that they are all local unsupervised segmentation.

In our work, a global depth supervised RGB image classification approach has been put forward. We first use the segmentation result of the LR depth map as labels to supervise the classification of the HR color image. Then, a MRF energy function is designed based on this HR classification result. By optimizing this MRF energy function, a HR depth map with clear depth boundaries can be generated. One thing worth emphasizing is that the only criteria for LR depth map segmentation is depth, and once the segmentation labels are to supervise the classification of HR color image, depth becomes the only criteria for this classification. In other words, the classification of HR color image is guided by depth so that depth edges can be preserved and redundant detail in the boundaries can be omitted.

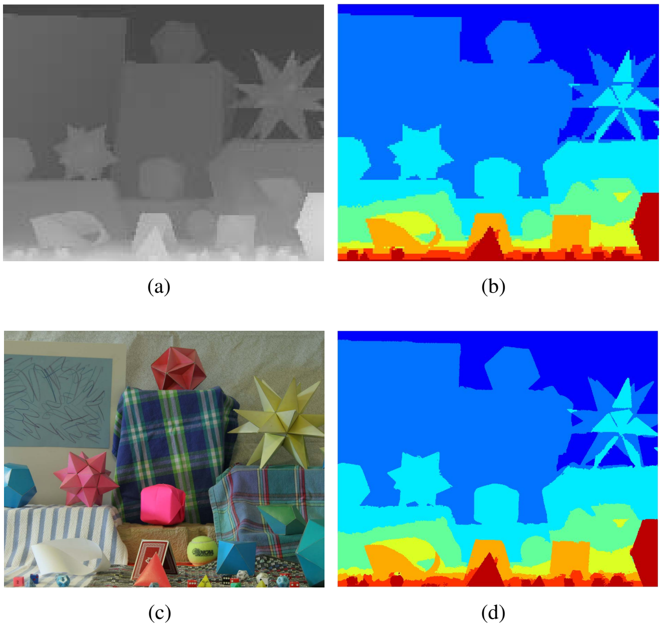

Fig. 1. Our supervised classification approach. (a) LR depth map, (b) the segmented result of (a) based on the K-means clustering, (c) HR color image and (d) the supervised classification of (c) based on the K-NN strategy, where different labels are marked in different colors.

The rest of the paper is organized as follows. In Section 2, we present the SC-MRF model. Experimental results are then provided in Section 3 and our paper is concluded in Section 4.

2. Proposed Method

The proposed SC-MRF model based depth map up-sampling mainly consists of the following four parts.

- Firstly, the LR depth map is intermediately up-sampled with bilateral filtering.
- Secondly, the LR depth map is segmented with K-means clustering and the result is used to supervise HR color image classification through a K-NN strategy.
- Next, a SC-MRF energy function (Geiger and Girosi, 1991; Saxena et al., 2009) is designed using the LR depth map, the HR color image, the coarsely up-sampled depth map and the supervised classification result.
- Finally, the up-sampled depth map is obtained by optimising the designed SC-MRF energy function using the Gradient Descent (GD) algorithm (Liu and Feig, 1996).

[1] Depth bleeding describes the phenomenon where the depth value of one object flows into another. Depth confusion represents the artifact that depth values of two different objects overlap one another.

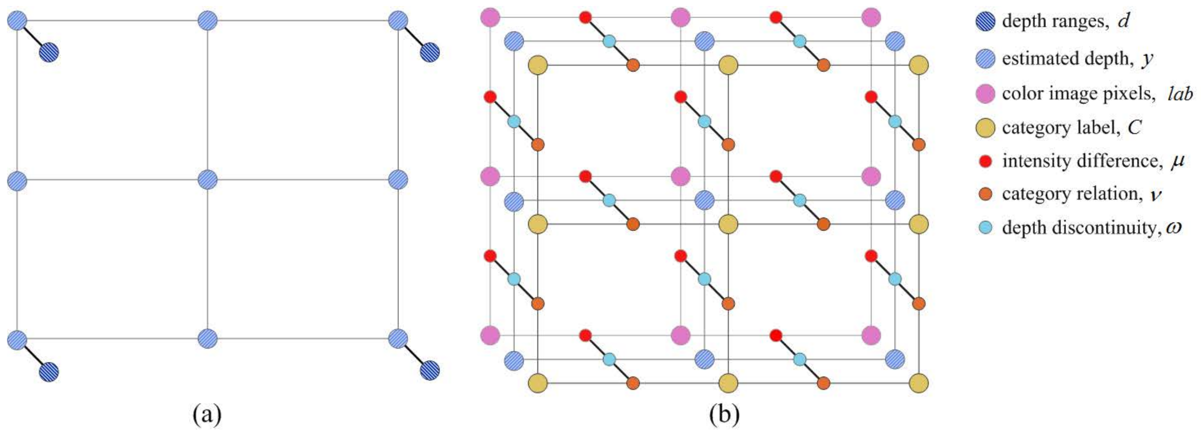

Fig. 2. The proposed MRF simulation. The data term and regularization term of the MRF model are illustrated in (a) and (b), respectively. 4 layers and 7 nodes are involved in the proposed MRF model. The estimated depth y on the up-sampled depth layer are measured by depth levels d on the LR depth layer, pixel intensities lab of the HR color image and category labels C of the classification layer. The depth discontinuity node ω is adjusted by both the intensity difference node μ and the category relation node ν.

2.1. Depth map preprocessing

To obtain the intermediate up-sampled depth map, a bilateral filter is performed on the LR depth map as Eq.(1) shows,

$$D_{iu}(p) = \frac{1}{k_p} \sum_{q_\downarrow \in N(p_\downarrow)} D_L(q_\downarrow)\, g_d(\|p-q\|)\, g_c(\|I_p - I_q\|) \qquad (1)$$

where p and q denote the pixel coordinates in the HR color image, $p_\downarrow$ and $q_\downarrow$ denote their corresponding coordinates in the LR depth map; $N(p_\downarrow)$ denotes the neighboring pixels of $p_\downarrow$; k_p is a normalization factor; I_p and I_q denote the pixel intensities of p and q; g_d and g_c are two Gaussian kernels corresponding to the spatial distance and the intensity difference, respectively.

As mentioned, depth bleeding and depth confusion are common in a coarsely up-sampled depth map. To improve the image quality by removing these artifacts, a SC-MRF energy function is designed and optimised in the following sections.

2.2. Supervised Classification

In this section, a supervised classification based MRF (SC-MRF) model is designed for maintaining the depth edges in the HR color image and suppressing the unexpected textures simultaneously. The supervised classification contains two main parts, the LR depth map segmentation and the HR color image classification under the supervision of the segmentation result.

2.2.1. LR depth map segmentation

Due to the characteristics of both simple-texture and sharp-edges, the similarity between two points on depth maps can be simply measured by Euclidean distance, which meets the assumptions of the classical K-means clustering (Honda et al., 2010; Mignotte, 2008). Besides, as an unsupervised clustering method, K-means is superior in both simplicity and speed, so that we directly use K-means for the LR depth map segmentation (Gan and Ng, 2017; Khan and Ahmad, 2004).

Firstly, a number of K_c clustering centers are selected at regular intervals over the range of $\{D_L(i)\}$, where K_c denotes the given number of classes in K-means. Secondly, samples on the LR depth map are assigned to the K_c cluster centers according to Euclidean distance as Eq.(2) shows,

$$C(p) = \underset{k}{argmin} \| D_L(p) - z_k \|, \quad s.t. \quad 1 \le k \le K_c \qquad (2)$$

where z_k denotes the kth cluster center; $C(p)$ denotes the label assigned to sample p.

Thirdly, the clustering centers $\{z_k\}$ are updated using pixel depths per class, with the updating formulated as Eq.(3) shows,

$$z_k = \frac{\sum_{p=1}^{n_k} D_L(p)}{n_k}, \quad s.t. \quad C(p) = k \qquad (3)$$

where n_k denotes the total number of samples in class k. The second and third steps are repeated until the program converges.

K-means clustering based LR depth map segmentation only depends on depth values of pixels. So when we use the segmentation result to supervise the HR RGB image classification, it is also depth-guided.

2.2.2. Supervised HR color image classification

A number of classification methods can be used in the proposed SC-MRF model. However, for simplicity, the classical K-nearest neighbor (K-NN) is selected (Krishnapuram and Keller, 1993; Zhang, 2012). In K-NN, the classification of each sample depends on its K-nearest neighbors - which will have already been labeled. In order to seek the K-nearest neighbors, a distance measure in feature space is therefore required.

To convert the HR color image I into feature space, both color features and the spatial features are extracted. In the proposed model, our feature space F consists of 8 feature values $\{l,\ a,\ b,\ h,\ s,\ i,\ u,\ v\}$. Among the 8 feature values, $\{l,\ a,\ b\}$ denote three color channels in CIELab space, $\{h,\ s,\ i\}$ denote three color channels in HSI space, u and v denote the horizontal and the vertical spatial coordinates, respectively.

In our constructed feature space F, distance between two arbitrary samples m and n can be calculated as in Eq.(4). Note that feature vectors should be normalized before distance calculation. By defining such a distance measure, the S nearest

Table 1. Quantitative evaluations under up-sampling factors of 2, 4 and 8 (in *MSE*). The best performance is shown in bold. Rankings of the proposed method with respect to other approaches are also provided.

Methods	Laundry			Dolls			Art			Books			Moebius			Reindeer			Avg.		
	2	4	8	2	4	8	2	4	8	2	4	8	2	4	8	2	4	8	2	4	8
JBU(Kopf et al., 2007)	1.72	2.48	3.57	1.14	1.66	2.39	3.36	5.01	6.80	1.41	2.04	3.11	1.09	1.85	2.81	2.06	2.94	3.98	2.64	3.67	4.91
MRF(Diebel and Thrun, 2006)	1.59	**1.89**	**2.68**	1.11	**1.32**	1.82	4.26	5.49	6.44	1.44	**1.86**	**2.41**	1.17	1.87	3.29	2.09	**2.69**	4.24	2.63	3.29	4.52
JGU(Liu et al., 2013)	1.92	2.46	3.29	1.29	1.71	2.37	3.59	4.66	5.73	1.47	2.14	2.89	1.44	1.97	2.61	2.71	2.81	3.65	2.61	3.48	4.55
TGV(Ferstl et al., 2013)	1.59	2.14	3.48	1.03	1.55	1.96	4.51	4.55	6.20	1.49	2.06	2.48	1.30	1.71	2.59	2.68	3.21	3.59	2.57	3.61	4.84
SD(Ham et al., 2018)	2.12	3.08	3.55	1.36	1.92	2.04	3.80	4.97	5.47	1.51	2.16	2.63	1.50	2.02	2.37	2.42	3.18	4.16	2.41	3.26	4.43
Ours	**1.57** **(1)**	1.93 **(2)**	2.93 **(2)**	**1.02** **(1)**	1.42 **(2)**	**1.78** **(1)**	**3.22** **(1)**	**4.24** **(1)**	**5.41** **(1)**	**1.32** **(1)**	1.90 **(2)**	2.42 **(2)**	**1.07** **(1)**	**1.63** **(1)**	**2.32** **(1)**	**1.96** **(1)**	2.71 **(2)**	**3.16** **(1)**	**2.00** **(1)**	**2.94** **(1)**	**4.20** **(1)**

labeling samples (seed pixels) of each unlabeled sample (target pixel) are extracted, where S denotes the pre-defined number of neighbors and lies between 5 and 10 empirically.

$$D_{m,n} = \parallel F_m - F_n \parallel_2 \tag{4}$$

Algorithm 1 GD based optimization.

```
Input:
    Intermediate up-sampling depth map $D_{iu}$;
    LR depth map $D_L$;
    Depth discontinuities $\omega$;
Output:
    The final up-sampled depth map;
Extract the depth range $L$ from $D_{iu}$;
Initialize iteration $k = 0$;
Initialize $y^k = D_{iu}$;
repeat
    for each point $p$ in $y^k$ do
        Find index $x$ that satisfies $L(x) = y_p^k$;
        Calculate the gradient $g_p^k = \triangledown E(y_p^k)$;
        if $g_p^k \neq 0$ then
            $m = -g_p^k / |g_p^k|$
            repeat
                $\widetilde{g}_p^k \Leftarrow g_p^k, \widetilde{y}_p^k \Leftarrow y_p^k$;
                $y_p^k = L(x + m)$;
                $g_p^k = \triangledown E(y_p^k)$;
                $x = x + m$;
            until $g_p^k / \widetilde{g}_p^k < 0$
            if $E(\widetilde{y}_p^k) - E(y_p^k) > 0$ then
                $y_p^{k+1} = y_p^k$;
            else
                $y_p^{k+1} = \widetilde{y}_p^k$;
            end if
        else
            $y_p^{k+1} = y_p^k$;
        end if
    end for
    $k = k + 1$;
until $E$ is relatively small or $k$ is relatively large.
```

Next, the unlabeled sample will be classified into the class to which most of the S labeling samples belong. Let i denotes the unlabeled sample, $\{j_s | 1 \leq s \leq S\}$ denotes the S nearest labeling samples of i, the classification of i can be formulated as follows,

$$L(i) = \arg\max_k \sum (C(j_s) == k), \quad s.t. \quad 1 \leq s \leq S \tag{5}$$

where $L(i)$ denotes the classification label of i; $C(j_s)$ calculated by Eq.(2) denotes the class number of j_s.

After all target pixels are classified via the proposed feature based K-NN, the final classification result is obtained. The result not only assures samples with similar features can be clustered into the same class, but also separates samples into different classes only according to their depth values. The proposed supervised classification is illustrated in Fig.1, which shows the depth edges are well preserved while the redundant color textures inside the objects are effectively suppressed.

As it is inefficient to seek the K_s nearest samples in the entire labeling domain, we accelerate the program without losing accuracy by limiting the search range to a pre-specified radius.

2.3. MRF Energy Function

To improve the quality of the coarsely up-sampled depth map, a MRF model is adopted (Geiger and Girosi, 1991; Saxena et al., 2009). The proposed model consists of four layers as shown in Fig.2: the LR depth layer, the HR color image layer, the label layer resulting from the supervised classification and the HR depth layer (which is to be estimated).

Relationship between the LR depth layer and the to-be-estimated HR depth layer is used to construct the data term of the MRF energy function. Meanwhile, relationships between the HR color image layer, the label layer and the to-be-estimated HR depth layer are used to construct the regularization term. The MRF energy function is formulated as follows,

$$E = \sum_{p=1}^{n} \alpha \parallel y_p - D_L(p) \parallel + \sum_{p} \sum_{q \in N(p)} \omega_{pq} \parallel y_p - y_q \parallel \tag{6}$$

where $D_L(p)$ denotes depth value of p in the LR depth map, as introduced in Sec.2.2.1; y_p denotes depth value of p in the to-be-estimated HR depth map; y_q denotes the 4-neighbors of y_p; α is a scale parameter for balancing the data term and the regularization term; ω_{pq} denotes the weight coefficient between two neighbors p and q.

Both the pixel intensity and the category relationship are taken into consideration when constructing the weight coefficient ω as presented in Eq.(7),

$$\omega_{pq} = \exp(-\frac{\mu_{pq}}{\beta}) \cdot \nu_{pq} \tag{7}$$

Table 2. Quantitative evaluations under the up-sampling factors of 2, 4 and 8 (in *bpr*). The best performance is shown in bold. Rankings of the proposed method is also provided.

Methods	Laundry			Dolls			Art			Books			Moebius			Reindeer			Avg.		
	2	4	8	2	4	8	2	4	8	2	4	8	2	4	8	2	4	8	2	4	8
JBU(Kopf et al., 2007)	1.34	2.91	6.30	1.54	3.90	10.07	1.84	5.56	12.23	1.35	3.71	7.91	1.54	3.15	9.22	1.03	2.38	8.37	1.08	2.69	6.56
MRF(Diebel and Thrun, 2006)	3.69	4.60	7.58	2.85	4.11	9.35	1.37	4.08	13.24	1.02	3.92	8.50	1.94	4.43	9.60	1.84	3.24	8.61	2.01	3.30	7.04
JGU(Liu et al., 2013)	2.09	3.36	**6.00**	2.29	4.29	9.02	1.76	**3.69**	7.51	1.53	2.83	7.99	1.63	2.91	7.89	2.22	2.69	5.59	1.50	2.79	5.79
TGV(Ferstl et al., 2013)	1.33	3.47	7.65	1.52	3.94	**6.65**	1.92	4.76	6.84	2.05	3.15	**4.05**	1.20	3.22	**5.60**	1.52	3.88	**4.95**	1.62	3.62	7.29
SD(Ham et al., 2018)	2.18	4.75	6.14	2.98	5.61	8.91	3.46	8.65	7.15	2.31	5.65	7.26	2.27	5.32	7.82	1.89	5.49	7.00	1.52	3.16	6.70
Ours	**0.93** **(1)**	**2.88** **(1)**	6.11 **(2)**	**1.02** **(1)**	**3.09** **(1)**	8.88 **(2)**	**0.97** **(1)**	3.92 **(2)**	**6.78** **(1)**	**0.81** **(1)**	**2.59** **(1)**	5.59 **(2)**	**0.84** **(1)**	**2.86** **(1)**	7.80 **(2)**	**0.59** **(1)**	**2.20** **(1)**	6.95 **(3)**	**0.71** **(1)**	**2.30** **(1)**	**5.67** **(1)**

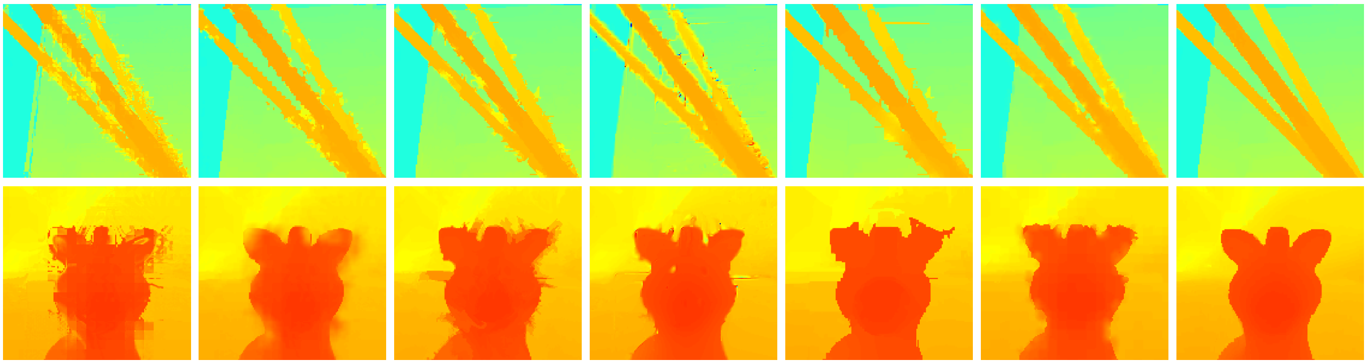

Fig. 3. Visual results under different up-sampling methods. Top: 4 **times up-sampling results of** *Art*. **Bottom:** 8 **times up-sampling results of** *Doll*. **From left to right are JBU method, MRF method, JGU method, TGV method, SD method, our method and the ground truth.**

where β is a pre-defined constant, selected empirically to be of value 10 in our experiments; μ_{pq} and ν_{pq} denote the intensity difference and the category relationship between p and q, respectively. The formulations of μ_{pq} and ν_{pq} are as follows,

$$\mu_{pq} = (l_p - l_q)^2 + (a_p - a_q)^2 + (b_p - b_q)^2 \tag{8}$$

$$\nu_{pq} = \begin{cases} 1, & C(p) = C(q) \\ \gamma, & C(p) \neq C(q) \end{cases} \tag{9}$$

where l, a, b, are the three color channels in CIELab space as mentioned; $C(p)$ denotes the category number of p; γ is a positive decimal that is very close to zero.

Based on the proposed SC-MRF model 1) neighbors belonging to the same class and with similar colors are closely linked, 2) neighbors belonging to the same class but with distinct colors are weakly linked, and 3) neighbours falling in different classes are intensely suppressed. By this means, depth edges can be effectively preserved.

2.4. GD based Optimization

Many optimization methods may be employed to optimise our framework, such as the GD method, the conjugate gradient (CG) method or the Newton method. For simplicity, classical GD method is employed here (Liu and Feig, 1996). Algorithm1 shows the procedure of the GD based optimization, in which the coarsely up-sampled depth map is used as the initial state.

3. Experimental Results

To investigate the performance of the proposed method, the algorithm is implemented using Matlab on a PC with Core Duo 3.20GHz CPU and 4.0G RAM.

We test our method on all data sets from Middlebury 2005, 2006 and 2014[2]. Both a HR color image and the corresponding HR depth map are included in each data set. Any anti-aliasing low-pass filter can be used for the HR depth map degradation, so that the to-be-up-sampled LR depth map can be obtained. Without loss of generality, the nearest filter is selected in our experiments. The original HR depth map is then used as ground truth to verify the up-sampling result.

Visual results under different up-sampling rates are presented in Fig.3, from which we can see that the MRF based method (Flezenszwalb and Huttenlocher, 2002) can generate HR depth maps with very smooth boundaries. However, this is prone to over-smoothing. On the other hand, the JGU based method (Liu et al., 2013), the TGV based method (Ferstl et al., 2013), and SD (static/dynamic) filter based method (Ham et al., 2018) can provide relatively sharp boundaries. However, burrs are brought in to the results. In comparison with these state-of-the-art methods, the proposed SC-MRF based method outputs quite clear depth boundaries with very little visible burrs, which improves the quality of the up-sampled result significantly.

Indexes of mean-square-error (*MSE*) and bad pixel rate (*bpr*) are introduced for quantitative evaluations. Bad pixels are those whose values deviate from the ground truth by more than one

[2] The data sets can be found at http://vision.middlebury.edu/stereo/data/

disparity. According to different up-sampling rates of 2, 4 and 8, *MSE* and *bpr* evaluations on 6 random data sets, as well as average *MSE* and *bpr* evaluations on all data sets are provided in Table 1 and 2. The proposed approach yields lower accuracy on 'Reindeer' under an up-sampling factor of 8. We believe that there are foreground regions having very similar but discontinuous depth values to the background. This may cause inaccurate segmentation of the LR depth map, and lead to an inaccurate classification of the HR RGB image further. This issue ultimately affects the up-sampling results. In general, the quantitative evaluations indicate the stability of the proposed method.

4. Conclusions

We present a novel SC-MRF model for depth map up-sampling. In the proposed model, HR color image is classified at pixel-level under the supervision of the segmentation of LR depth map. The supervised classification result is further used to design the MRF energy function. The final up-sampled depth map is generated by optimizing the energy function through a GD algorithm. Due to the supervised classification, the proposed SC-MRF model obtains high-quality up-sampled depth map with both smooth homogeneous regions and clear depth edges. Experimental results indicate that the proposed method performs well in both visual and quantitative evaluations. Future work includes accelerating the method and developing a semi-supervised classification based up-sampling model.

Acknowledgments

The authors would like to thank Middlebury, for distributing their data sets. This work was additionally supported by the Major Research Plan of National Natural Science Foundation of China Number 91438201, the National Natural Science Foundation of China Number 61573267, the InnovateUK grant Number 104541 and the EPSRC grant Number EP/M023281/1.

References

Buyssens, P., Daisy, M., Tschumperlé, D., Lézoray, O., 2015. Superpixel-based depth map inpainting for rgb-d view synthesis, in: 2015 IEEE International Conference on Image Processing (ICIP), pp. 4332–4336. doi:10.1109/ICIP.2015.7351624.

Chan, D., Buisman, H., Thebalt, C., , Thrun, S., 2008. A noise-aware filter for real-time depth upsampling. in Workshop on M2SFA2, ECCV .

Choi, O., Jung, S.W., 2014. A consensus-driven approach for structure and texture aware depth map upsampling. IEEE Transactions on Image Processing 23, 3321–3335. doi:10.1109/TIP.2014.2329766.

Diebel, J., Thrun, S., 2006. An application of markov random fields to range sensing. Adv. Neural Inf. Process. Syst. 18, 291 – 298.

Dominio, F., Donadeo, M., Zanuttigh, P., 2014. Combining multiple depth-based descriptors for hand gesture recognition. Pattern Recognition Letters 50, 101 – 111. Depth Image Analysis.

Eichhardt, I., Chetverikov, D., Jankó, Z., 2017. Image-guided tof depth upsampling: a survey. Machine Vision and Applications 28, 1–16. doi:10.1007/s00138-017-0831-9.

Fankhauser, P., Bloesch, M., Rodriguez, D., Kaestner, R., Hutter, M., Siegwart, R., 2015. Kinect v2 for mobile robot navigation: Evaluation and modeling, in: 2015 International Conference on Advanced Robotics (ICAR), pp. 388–394. doi:10.1109/ICAR.2015.7251485.

Ferstl, D., Reinbacher, C., Ranftl, R., Ruether, M., Bischof, H., 2013. Image guided depth upsampling using anisotropic total generalized variation. Proc. ICCV , 993–1000.

Flezenszwalb, P.F., Huttenlocher, D.P., 2002. Efficient belief propagation for early vision. IEEE Trans. Pattern Anal. Mach. Intell 24, 603 – 619.

Gan, G., Ng, M.K.P., 2017. k-means clustering with outlier removal. Pattern Recognition Letters 90, 8 – 14.

Geiger, D., Girosi, F., 1991. Parallel and deterministic algorithms from mrfs: surface reconstruction. IEEE Transactions on Pattern Analysis and Machine Intelligence 13, 401–412. doi:10.1109/34.134040.

Ham, B., Cho, M., Ponce, J., 2018. Robust guided image filtering using nonconvex potentials. IEEE Transactions on Pattern Analysis and Machine Intelligence 40, 192–207. doi:10.1109/TPAMI.2017.2669034.

Honda, K., Notsu, A., Ichihashi, H., 2010. Fuzzy pca-guided robust *k*-means clustering. IEEE Transactions on Fuzzy Systems 18, 67–79. doi:10.1109/TFUZZ.2009.2036603.

HyeokHyen Kwon, Yu-Wing Tai, Lin, S., 2015. Data-driven depth map refinement via multi-scale sparse representation, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 159–167. doi:10.1109/CVPR.2015.7298611.

Khan, S.S., Ahmad, A., 2004. Cluster center initialization algorithm for k-means clustering. Pattern Recognition Letters 25, 1293 – 1302.

Kim, J., Jeon, G., Jeong, J., 2014. Joint-adaptive bilateral depth map upsampling. Signal Processing: Image Communication 29, 506 – 513.

Kopf, J., Cohen, M.F., Lischinski, D., Uyttendaele, M., 2007. Joint bilateral upsampling. ACM Transaction on Graphics 26.

Krishnapuram, R., Keller, J.M., 1993. A possibilistic approach to clustering. IEEE Transactions on Fuzzy Systems 1, 98–110. doi:10.1109/91.227387.

Liu, L.K., Feig, E., 1996. A block-based gradient descent search algorithm for block motion estimation in video coding. IEEE Transactions on Circuits and Systems for Video Technology 6, 419–422. doi:10.1109/76.510936.

Liu, M.Y., Tuzel, O., Taguchi, Y., 2013. Joint geodesic upsampling of depth images. CVPR .

Lo, K.H., Wang, Y.C.F., Hua, K.L., 2017. Edge-preserving depth map upsampling by joint trilateral filter. IEEE Transactions on Cybernetics PP, 1–14. doi:10.1109/TCYB.2016.2637661.

Mahmoudpour, S., Kim, M., 2016. The effect of depth map up-sampling on the overall quality of stereopairs. Displays 43, 9 – 17.

Mignotte, M., 2008. Segmentation by fusion of histogram-based *k*-means clusters in different color spaces 17, 780–7.

Min-Koo, K., Dae-Young, K., Kuk-Jin, Y., 2014. Adaptive support of spatial-temporal neighbors for depth map sequence up-sampling. IEEE Signal Processing Letters 21.

Park, J., Kim, H., Tai, Y.W., Brown, M.S., Kweon, I., 2011. High quality depth map upsampling for 3d-tof cameras, in: 2011 International Conference on Computer Vision, IEEE. pp. 1623–1630.

Sarbolandi, H., Lefloch, D., Kolb, A., 2015. Kinect range sensing: Structured-light versus time-of-flight kinect. Computer Vision and Image Understanding 139, 1 – 20.

Saxena, A., Sun, M., Ng, A.Y., 2009. Make3d: Learning 3d scene structure from a single still image. IEEE Transactions on Pattern Analysis and Machine Intelligence 31, 824–840. doi:10.1109/TPAMI.2008.132.

Shao, F., Jiang, G., Yu, M., Chen, K., Ho, Y.S., 2012. Asymmetric coding of multi-view video plus depth based 3-d video for view rendering. IEEE Transactions on Multimedia 14, 157–167. doi:10.1109/TMM.2011.2169045.

Soh, Y., Sim, J.Y., Kim, C.S., Lee, S.U., 2012. Superpixel-based depth image super-resolution 8290, 117 – 126. doi:10.1117/12.909848.

Tallón, M., Derin Babacan, S., Mateos, J., Do, M.N., Molina, R., Katsaggelos, A.K., 2012. Upsampling and denoising of depth maps via joint-segmentation, in: 2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO), pp. 245–249.

Yang, Y., Gao, M., Zhang, J., Zha, Z., Wang, Z., 2015. Depth map super-resolution using stereo-vision-assisted model. Neurocomputing 149, 1396 – 1406.

Yuan, L., Jin, X., Li, Y., Yuan, C., 2017. Depth map super-resolution via low-resolution depth guided joint trilateral up-sampling. Journal of Visual Communication and Image Representation 46, 280 – 291.

Zhang, S., 2012. Nearest neighbor selection for iteratively knn imputation. Journal of Systems and Software 85, 2541 – 2552.

Zhu, J., Rizzo, J.R., Fang, Y., 2017. Learning domain-invariant feature for robust depth-image-based 3d shape retrieval. Pattern Recognition Letters .