EXPLICIT MEASUREMENT ON DEPTH-COLOR INCONSISTENCY FOR DEPTH COMPLETION

Yifan Zuo[1,2]*, Qiang Wu*[2]*, Ping An*[1]*, Jian Zhang*[2]

1. School of Communication and Information Engineering, Shanghai University, Shanghai, China
2. School of Computing and Communications, University of Technology, Sydney, Australia

ABSTRACT

Color-guided depth completion is to refine depth map through structure light sensing by filling missing depth structure and de-nosing. It is based on the assumption that depth discontinuity and color edge at the corresponding location are consistent. Among all proposed methods, MRF-based method including its variants is one of major approaches. However, the assumption above is not always true, which causes texture-copy and depth discontinuity blurring artifacts. The state-of-the-art solutions usually are to modify the weighting inside smoothness term of MRF model. Because there is no any method explicitly considering the inconsistency occurring between depth discontinuity and the corresponding color edge, they cannot adaptively control the effect of guidance from color image when completing depth map. In this paper, we propose quantitative measurement on such inconsistency and explicitly embed it into weighting value of smoothness term. The proposed method is evaluated on NYU Kinect datasets and demonstrates promising results.

Index Terms— Depth map completion, Markov Random Field (MRF), depth recovery

1. INTRODUCTION

Acquirement of high-quality depth data is the key problem in the field of 3-D computer vision, which is required in many applications, e.g., 3DTV, 3D object modeling. Recently, structured-light depth sensors, such as the first generation of Kinect (Kinect v1), have been widely used in research and practice. However, the quality of depth maps obtained by such sensors is not satisfactory due to big holes in the regions near edges where the occlusion occurs. Moreover, the noise existing in depth measurement makes the values different from the true values. In case of poor reflection or even shadow reflection of the light patterns, missing and erroneous depth values can also be caused by absorption. Objects with darker colors, specular surfaces, or fine-grained surfaces like human hair are prime candidates for poor depth measurements [1]. Therefore, there are two major problems in depth obtained by such an imaging system, which is missing and distorted depth values.

The state-of-the-art methods in depth image completion can be grouped into two categories: non-color-guided methods [2, 3] and color-guided methods [4-13]. For non-color-guided methods, Kinect-Fusion [2] integrates noisy depth maps which are captured at various viewpoints. In contrast to the single raw Kinect depth, the fused depth has less holes and less noise. Multi-Kinect-Fusion [3] uses a multi-sensor setup of low-cost depth sensors to obtain a combined depth map that can be arbitrary positioned between the input sensors. However, these methods either have problem in capturing depth video or lack robustness due to the overlay of different infrared patters on the scene. Regarding color-guided depth completion methods, they always rely on companion color image in high quality, which uses color information for depth completion based on the fundamental assumption that the depth discontinuity and color edge at the corresponding location are consistent [7]. Color-guided completion methods can be further classified in image in-painting based methods [12-13] and super-resolution based methods [4-11]. Wang et al. [12] proposed a stereoscopic in-painting algorithm which jointly completes missing texture and depth via two pairs of RGB and depth cameras. Holes occluded by foreground are completed by minimizing a predefined energy function. Such system requires an additional pair of RGB and depth cameras to fulfill the depth completion. Super-resolution based method consists of filter-based method and global-based method which uses only one pair of RGB and depth cameras to predict missing depth information. Compared with filter-based solutions [6, 9, 10], global-based methods [4, 5, 7, 8, 11] are more robust to noise in depth map captured by sensors. Our method belongs to MRF-based methods which are major methods in the category of global-based solutions. Diebel et al. [7] modeled depth enhancement as solving a multi-labeling optimization problem via Markov Random Fields (MRF). Park et al. [5] used a non-local term to regularize depth maps to fill holes and combined with a weighting scheme which involves edge, gradient, and segmentation information extracted from color images. In addition to MRF-based methods, J. Yang et al. [8] achieved depth completion via the color-guided auto-regression model.

Although color-guided methods work well for depth completion, color guidance image might lead to texture-copy artifact as well as depth discontinuity blurring. The main problem is that the fundamental assumption of color-guided depth completion methods is not always true. That is, depth discontinuity regions on depth map do not necessarily correspond to the regions of color edge in the registered color image.

In fact, these artifacts have been noticed for a long time, and almost all state-of-the-art methods mentioned above adopt various ways to eliminate the texture-copy and depth discontinuity blurring artifacts. But they do not explicitly evaluate the edge inconsistency between color image and depth map. Therefore, they cannot

This work was supported in part by the National Natural Science Foundation of China, under Grants U1301257, 61571285, 61422111 and the China Scholarship Council.

adaptively control the effect of guidance from color image when completing depth map.

In this paper, the main contributions are in two aspects. 1. Our method explicitly considers the inconsistency occurring between depth discontinuities and the corresponding color edges, and measuring inconsistency quantitatively; 2. Our method explicitly embeds inconsistency measurement above into weighting value of smoothness term in MRF energy function. Therefore, the proposed method is able to not only suppress texture-copy artifacts but also preserve edges better than other state-of-the-art methods.

The rest of this paper is organized as follows. Section 2 presents the proposed algorithm via Markov Random Fields with inconsistency measurement. In section 3, the experimental results are presented. Section 4 concludes this paper.

2. PROPOSED METHOD

A Markov random field, also known as a Markov network or an undirected graphical model has been widely utilized for many image processing applications and tasks. MRF formulates depth map completion as solving an optimization problem. The input includes high quality color image and low quality depth map. According to the Hammersely-Clifford theorem [14], solving MRF is equivalent to optimizing the Gibbs energy function, whose general formulation is defined as follows:

$$D = \underset{D=\{d_p\}}{\arg\min} \sum_{p\in O} E_{data}\left(d_p,D_p\right) + \lambda \sum_{p} \sum_{q\in N(p)} E_{smooth}\left(d_p,d_q\right) \qquad (1)$$

$$E_{data}\left(d_p,D_p\right) = \left(d_p - D_p\right)^2 \qquad (2)$$

$$E_{smooth}\left(d_p,d_q\right) = \omega_{pq}\left(d_p - d_q\right)^2 \qquad (3)$$

where D indicates the value set of the reconstructed depth map, d_p indicates the reconstructed value of pixel p , D_p is the observed depth value of pixel p . O is the pixel set consisting of pixels which have observed depth values. E_{data} is called the data term which maintains the consistency between the reconstructed depth value and the initial observed depth value. E_{smooth} is called the smoothness term which penalizes the differences between the reconstructed depth value and the depth values in the neighboring region. The parameter λ is used to balance the data term and smoothness term. N_p is the set of 8-connected neighboring pixels for the pixel p .

According to the MRF based depth completion framework shown in Eq.(1), a common method models $E_{data}\left(d_p,D_p\right)$ and $E_{smooth}\left(d_p,d_q\right)$ as Eq.(2) and Eq.(3) based on the assumption of Gaussian White Noise. ω_{pq} links color image to depth map which provides the guidance from color image for depth completion based on the assumed consistency between color edge and depth discontinuity (i.e. depth edge) [7]. As mentioned above, this assumption is not always true. It is the root problem of texture-copy and depth discontinuity blurring happening during depth completion because of the wrong guidance from color image. To overcome the texture-copy and depth discontinuity blurring artifacts, this paper proposed a weight $\lambda_{smooth-pq}$ to replace ω_{pq} in Eq.(3), for the first time, by introducing quantitative inconsistency measurement between color edges and depth discontinuities.

Section 2.1 proposes the quantitative measurement on the inconsistency between color edge in color image and depth discontinuity in the corresponding regions on depth map. Section 2.2 and 2.3 explicitly embed such measurement into $\lambda_{smooth-pq}$ in MRF framework to adaptively adjust MRF optimization.

2.1. Measurement on the inconsistency between color edge and depth discontinuity in the corresponding regions

Motivated by [15], the inconsistency measurement between color edge map and depth edge map is formulated as a bi-directional edge map quality assessment. Like [15], common edge map quality measurement is based on each individual edge pixel position shift against the ground truth edge pixel position. This paper is dealing with different case. Given a pair of depth map and RGB image, there is no additional information of ground truth position for each pixel. Thus, it is impossible to calculate one-to-one edge pixel position matching/checking. The consistency measurement in this paper is based on the edge map structure similarity.

Canny operator [16] is applied in color image and coarsely interpolated depth map to generate relevant edge maps. Due to low quality interpolated depth map by scattered interpolation method, the positions of corresponding edge pixels on color edge map and interpolated depth edge map are not consistent strictly. In this paper, inconsistency measurement is casted as a MRF optimization problem. For each edge position on reference edge map, it will search the best consistency in a neighboring region around the corresponding position on the target edge map. This implies that if the edge maps between color image and depth map are consistent, the position shift of each edge pixel should be small and it should only moves to a closely nearby region. Moreover, the shift including strength and orientation in a nearby region should happen consistently. These two constrains are solved in a MRF framework through its data term and smoothness term respectively.

$$L = \underset{L=\{l_p\}}{\arg\min} \sum_{p\in ref} C\left(p,p+l_p\right) + \mu\cdot \sum_{p\in ref} \sum_{k\in N(p)} V\left(l_p,l_k\right) \qquad (4)$$

where $C\left(p,p+l_p\right)$ is the data term of the MRF model. p represents the position of edge pixel in the reference edge map. l_p stands for the displacement so $p+l_p$ represents a position of edge pixel q which is in a neighboring region corresponding to the coordinate of p in target edge map. In our work, the size of neighboring region is 7×7 . Data term $C(p,q)$ matches the reference edge pixel p against target edge pixel q . Given p , if certain target pixel q in neighboring region of p is not an edge pixel in target edge map, it is regarded as definite inconsistency. In that case, $C(p,q)$ is assigned to the maximum inconsistency value (i.e. 1 in our work). Otherwise, this inconsistency is measured on two blocks where edge pixel p and edge pixel q are the center positions of these blocks respectively. In this paper, the size of block is 3×3. $E_p = \left\{e_{p1},e_{p2},\ldots,e_{pM}\right\}$ and $E_q = \left\{e_{q1},e_{q2},\ldots,e_{qN}\right\}$ are defined to represent the sets of edge pixels in these two blocks respectively (excluding p and q). M and N are the number of edge pixels inside these two sets. Thus, the inconsistency measurement between p and q is regarded as a matching problem between two data sets E_p and E_q . This matching problem is sorted out via Bipartite graph matching [17] which is more robust than MAD (mean of absolute difference) and Euclidean distance. The

Bipartite graph $G\left(E_p,E_q,W\right)$ is defined, where E_p and E_q are vertices in Bipartite graph and W represents the link between vertices whose weight is defined as $\phi(i,j)$ which is a monotonic function that returns a positive penalty for local structural matching.

$$\phi(i,j)=f\left(\left|i^x-j^x\right|+\left|i^y-j^y\right|\right) \tag{5}$$

where $f(0)=0, f(1)=1, f(2)=1.6$ and $f(x)=2$ when $x>2$. i,j are vertices in Bipartite graph, i^x,i^y are the coordinate of i .

Bipartite matching [17] is employed to enforce one-to-one matching between edge pixel data sets above. That is, it assures any edge pixel in E_p/E_q matches only one edge pixel in E_q/E_p , leaving $|M-N|$ unmatched pixels. Unmatched pixels represent the potential structure differences between edge pixel sets E_p/E_q . Bipartite matching is used to define the inconsistency measurement term $C(p,q)$ in Eq.(4) as,

$$C(p,q)=\left(\sum_{(p_s,q_s)\in E'_{pq}}\phi(p_s,q_s)/2+|M-N|\right)/8 \tag{6}$$

$E'_{pq}=\left\{(p_1,q_1),(p_2,q_2),\ldots(p_r,q_r)\right\}$ is edge pixel pair sets selected by Bipartite graph matching [17]. $\phi(p_s,q_s)$ is the weight of the link between edge pixel p_s and edge pixel q_s and $s=\{1,2,3\ldots r\}$. Therefore, $\sum_{(p_s,q_s)\in E'_{pq}}\phi(p_s,q_s)$ is the matching cost of Bipartite matching [17] mentioned above. Through normalization, the range of data term $C(p,q)$ is ensured in [0, 1].

$V\left(l_p,l_k\right)$ is the smoothness term in Eq.(4), which gives a penalty when adjacent edge pixels have different displacements as,

$$\underset{k\in N(p)}{V\left(l_p,l_k\right)}=\begin{cases}0, & l_p=l_k;\\ 1, & l_p\neq l_k;\end{cases} \tag{7}$$

$V\left(l_p,l_k\right)$ takes the connectivity of adjacent edge pixels into account, which means that connectivity of adjacent edge pixels is encouraged to maintain in the solution of Eq.(4).

In Eq.(4), μ is a balance factor between data term and smoothness term. It is set to 0.1 in this paper. $N(p)$ is the set of 8-connected neighboring pixels of p . Graph cut [18] is adopted to solve Eq.(4) MRF problem. The output of data term C computed by optimized displacements L represents the inconsistency between reference edge map and target edge map.

The inconsistency is measured based on reference edge map against target edge map. Thus, the measurement results will be different when swapping these two edge maps. In this work, the two edge maps are color edge map and depth discontinuity (edge) map. When color edge map is regarded as the reference edge map for inconsistency measurement, it can be observed that inconsistent positions detected reflect the texture copy happening areas. On the other hand, when depth discontinuity (edge) map is regarded as the reference edge map, it is observed that inconsistent positions reflect the depth discontinuity blurring happening areas.

2.2. Alignment of inconsistency maps

After bi-direction evaluation, there are two inconsistency maps C_{color} , C_{depth} as well as two displacement maps L_{color} , L_{depth} for an image pair. They represent the inconsistency measurement and displacement when color edge map or depth edge map are the reference edge map respectively. Before embedding the inconsistency measurement values into MRF based depth completion framework, these two inconsistency maps must be consolidated to each other.

As mentioned before, coarsely interpolated depth map may shift the position of edge pixels a bit from their true locations. On the other hand, the position of edge pixel on color edge map is more precise because of high quality of color image. Through the solution of MRF problem, Eq(4), mentioned above with depth edge map as the reference edge map, the displacement between each depth edge pixel p and its color edge pixel q is $L_{depth}(p)$. Consequently, the true location of depth edge pixel p supposes to be more close to $p+L_{depth}(p)$ when $C_{depth}(p)\neq 1$ which excludes the case of definite inconsistency that represents no corresponding pixel on color edge map for p . Therefore, the C_{depth} is adjusted as,

$$\begin{aligned}&\underset{p'=p+L_{depth}(p)}{C'_{depth}}(p')=\min_{p=p'-L_{depth}(p)}C_{depth}(p)\quad if\ C_{depth}(p)\neq 1\\ &C'_{depth}(p)=C_{depth}(p)\qquad\qquad\qquad otherwise\end{aligned} \tag{8}$$

Eq.(8) defines that if there are more than one pixel p mapping to the same pixel p', the best mapping with the lowest cost is adopted. Otherwise, the proposed method maintains the positions and values of the rest mappings unchanged from C_{depth} to C'_{depth} .

Once two inconsistency maps C'_{depth} and C_{color} are aligned, a confidence map α_p is defined as below, taking two directions of evaluation into account. It describes the final inconsistency status between color edge map and depth discontinuity (edge) map, which is embedded into MRF based depth completion framework i.e. Eq.(1) (see Section 2.3).

$$\alpha_p=\max\left(C'_{depth}(p),C_{color}(p)\right) \tag{9}$$

2.3. Improved MRF by considering inconsistency measurement

To simplify the explanation in the follows, Eq.(3) is updated below,

$$E_{smooth}\left(d_p,d_q\right)=\lambda_{smooth-pq}\left(d_p-d_q\right)^2 \tag{10}$$

where $\lambda_{smooth-pq}$ is to replace ω_{pq} in Eq.(3). Generally speaking, guidance information for depth completion task can be derived from two sources. One is from registered color image, and the other is from original depth map. Based on the confidence map α_p computed in Eq.(9), this paper combines it to generate a new guidance image to compute the weighting value $\lambda_{smooth-pq}$. $\lambda_{smooth-pq}$ is constructed as below.

$$\lambda_{smooth-pq}=e^{-\frac{\left(\nabla_{color}{}^{pq}\cdot(1-\alpha_{pq})+\nabla_{depth}{}^{pq}\cdot\alpha_{pq}\right)^2}{2\delta^2}} \tag{11}$$

where $\nabla_{color}{}^{pq}$ and $\nabla_{depth}{}^{pq}$ represent color difference and depth difference between position p and its neighboring pixel q in guided color image and coarsely interpolated depth map respectively. δ controls decay rate of exponential function in Eq.(11). In our work, applying "max" operation is better than mean operation when integrating α_p and α_q together which is expressed as $\alpha_{pq}=\max\left(\alpha_p,\alpha_q\right)$. It is also observed that when the corresponding color edge map is more consistent with depth edge map,

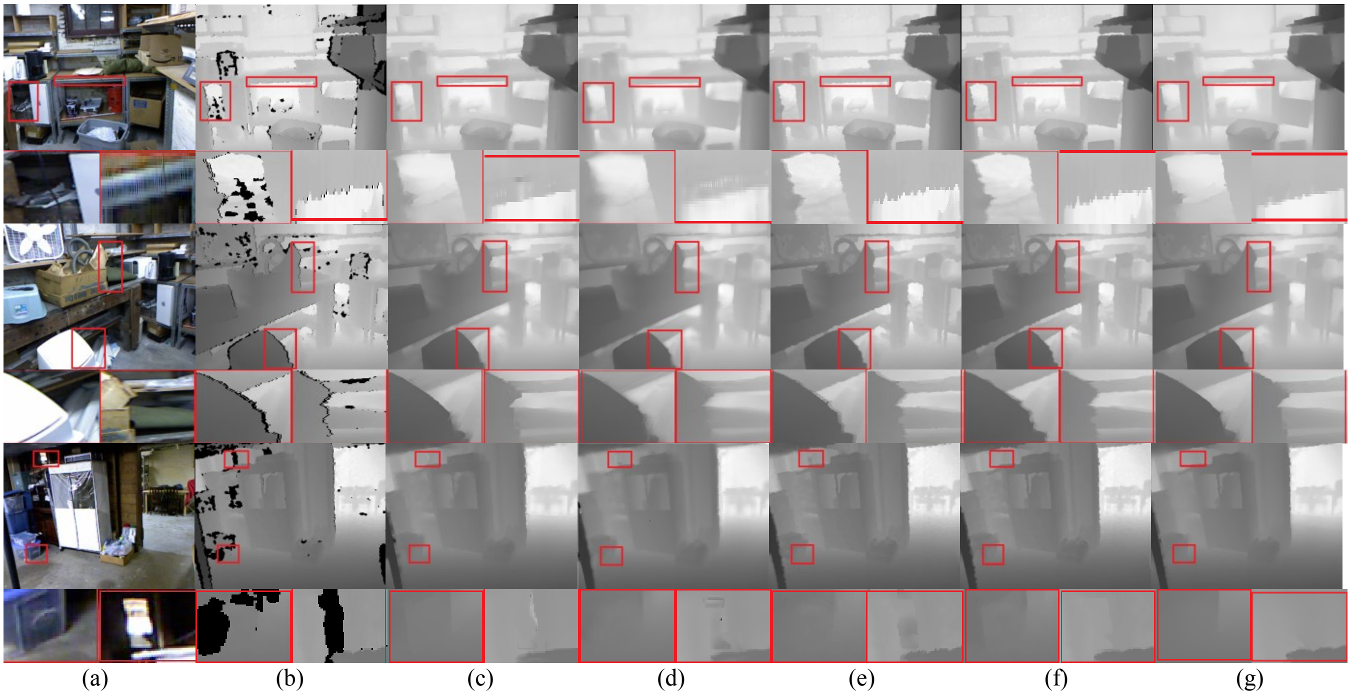

Fig. 1 Depth map completion results. (a) RGB Images, (b) Registered raw depth maps from Kinect v1, Depth map completion using (c) AR [8], (d) MLS [9], (e) JBU [10], (f) Colorization [11] and (g) Our results. Note the high-lighted regions.

$\nabla_{color}{}^{pq}$ plays more important role in computing the weighting value $\lambda_{smooth-pq}$.

The scenario discussed above is on the regions around edge pixels. When the depth incompleteness happens on the smooth regions (windows centered at p,q) where there are no edge pixels on either color image or depth map, Eq.(11) cannot satisfy such case because the previous guidance information is based on the presence of edge pixels and their relation between color image and depth map. In this paper, it is updated as Eq.(12) for this special case, where the guidance information for depth completion is from depth map only to thoroughly overcome texture copy artifact. In our work, we also see that a larger δ is needed to suppress noise in these smooth regions.

$$\lambda_{smooth-pq} = e^{-\frac{\left(\nabla_{depth}{}^{pq}\right)^2}{2\delta_{larger}{}^2}} \qquad (12)$$

Based on the analysis above, the proposed method can preserve depth edges and prevent texture-copy artifacts efficiently by adaptively controlling the guidance from color image for depth completion.

3. EXPERIMENTAL RESULTS

The proposed method is evaluated on NYU Kinect v1 datasets [19]. The comparison performance against the state-of-the-art methods is demonstrated.

3.1. Parameters setting

All the edge maps are computed through Canny operator. The dual thresholds setting are [0.04, 0.12] and [0.03 0.07] for color and depth respectively. λ is set to 5 for the whole datasets. δ and δ_{larger} is fixed to 2 and 4 respectively.

3.2. Experimental results on NYU Kinect datasets

The proposed method is compared with state-of-the-art methods: AR [8], MLS [9], JBU [10] and Colorization [11]. Fig.1 shows the depth completion results of three datasets which have rich texture in color, challenging the basic color-guided depth completion assumption.

For the details, the second and the forth rows in fig.1 illustrate the performance on preserving edges. Moreover, the second and the sixth rows show the performance on suppressing texture-copy artifacts. From these enlarge regions, it is shown that the existing methods [8, 9, 10, 11] have texture-copy artifacts on the different degrees. In term of preserving edges, AR [8] performs best in these existing methods. Through comparison, it is shown that the proposed method demonstrates the best depth completion performances which have the best results in not only edge preserving but also texture-copy artifacts suppressing.

4. CONCLUSION

This paper proposes a color-guided depth completion method in MRF framework. The key contribution is to explicitly measure the inconsistency between color edge map and the depth discontinuity (edge) map and embed it into MRF framework. It relaxes the assumption in color-guided depth completion methods. And it eliminates texture-copy and depth discontinuities blurring artifacts. Experimental results on the NYU Kinect datasets prove the improved performance of the proposed method.

5. REFERENCES

[1] J. Cho, S. Kim, Y. Ho, and K. Lee, “Dynamic 3d human actor generation method using a time-of-flight depth camera,” IEEE Transactions on Consumer Electronics, vol. 54, pp. 1514-1521, 2008.

[2] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, and A. Fitzgibbon, “Kinectfusion: Real-time 3d reconstruction and interaction using a moving depth camera,” 24th annual ACM symposium on User interface software and technology, ser. UIST ’11, pp. 559-568, 2011.

[3] M. Schmeing, E. Krauskopf, J. Xiaoyi, "Real-time depth fusion using a low-cost depth sensor array," 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON),pp. 1-4, 2014.

[4] J. Shen, S.-C.S Cheung, "Layer Depth Denoising and Completion for Structured-Light RGB-D Cameras," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.1187-1194, 2013.

[5] J. Park, H. Kim, Y. Tai, M. Brown, I. Kweon, "High-Quality Depth Map Upsampling and Completion for RGB-D Cameras," IEEE Transactions on Image Processing, vol. 23, no. 12, pp. 5559-5572, 2014.

[6] Q. Yang, R. Yang, J. Davis, and D. Nister, "Spatial-depth super resolution for range images," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1-8, 2007.

[7] J. Diebel and S. Thrun, "An application of markov random fields to range sensing," Advances in neural information processing systems, vol. 18, pp. 291–295, 2006.

[8] J. Yang, X. Ye, K. Li, C. Hou, Y. Wang, "Color-Guided Depth Recovery From RGB-D Data Using an Adaptive Autoregressive Model," IEEE Transactions on Image Processing, vol. 23, no. 8, pp. 3443-3458, 2014.

[9] N. K. Bose and N. A. Ahuja, “Superresolution and noise filtering using moving least squares,” IEEE Trans. Image Process., vol. 15, no. 8, pp. 2239–2248, Aug. 2006.

[10] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, “Joint bilateral upsampling,” Graphics, ACM Transaction on, vol. 26, no. 3, pp. 96, 2007.

[11] A. Levin, D. Lischinski, and Y. Weiss, "Colorization using optimization," ACM Trans. Graphics, vol. 23, no.3, pp. 689–694, 2004.

[12] L. Wang, H. Jin, R. Yang, and M. Gong, "Stereoscopic inpainting: Joint color and depth completion from stereo images," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1-8, 2008.

[13] L. Torres-Mendez and G. Dudek, "Reconstruction of 3d models from intensity images and partial depth," in Proceedings of the National Conference on Artificial Intelligence (AAAI), pp. 117-122, 2004.

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

[15] W.D. Jang, C.S. Kim, "SEQM: Edge quality assessment based on structural pixel matching," Visual Communications and Image Processing (VCIP), IEEE Conference on, pp. 1-6, 2012.

[16] J. Canny, “A Computational Approach to Edge Detection,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 8, no. 6, pp. 679–698, 1986.

[17] H.W. Kuhn, “The Hungarian method for the assignment problem,” Naval Research Logistics Quarterly, vol. 2, no. 1, pp. 83-97, 1955.

[18] Y. Boykov, O. Veksler, and R. Zabih, "Fast Approximate Energy Minimization via Graph Cuts," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 23, no. 11, pp. 1222-1239, 2001.

[19] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from rgbd images," In Proc. of European Conf. on Computer Vision (ECCV), 2012.