Dense Depth Map Reconstruction from Sparse Measurements Using a Multilayer Conditional Random Field Model

Francis Li, Edward Li, Mohammad Javad Shafiee, Alexander Wong, John Zelek
Department of System Design Engineering
University of Waterloo
Waterloo, Canada
Email: f27li@uwaterloo.ca

***Abstract*—Acquiring accurate dense depth maps is crucial for accurate 3D reconstruction. Current high quality depth sensors capable of generating dense depth maps are expensive and bulky, while compact low-cost sensors can only reliably generate sparse depth measurements. We propose a novel multilayer conditional random field (MCRF) approach to reconstruct a dense depth map of a target scene given the sparse depth measurements and corresponding photographic measurements obtained from stereophotogrammetric systems. Estimating the dense depth map is formulated as a maximum a posteriori (MAP) inference problem where a smoothness prior is assumed. Our MCRF model uses the sparse depth measurement as an additional observation layer and describes relations between nodes with multivariate feature functions based on the depth and photographic measurements. The method is first qualitatively analyzed when performed on data collected with a compact stereo camera, then quantitative performance is measured using the Middlebury stereo vision data for ground truth. Experimental results show our method performs well for reconstructing simple scenes and has lower mean squared error compared to other dense depth map reconstruction methods.**

***Keywords*-Dense depth map; Multilayer conditional random field; Stereophotogrammetry; 3D visualisation**

I. Introduction

Recent advancements in 3D imaging technology have seen a drastic rise in using depth sensors for research and commercial applications. Depth sensors generate depth map images where each pixel represents the distance from that point in the scene to the camera. Depth maps can help solve many computer vision problems such as segmentation, tracking, and object recognition[1]. Current high quality depth sensors such as time-of-flight cameras or high resolution structured light scanners are expensive and bulky. While recent innovations such as the Kinect (Microsoft Corp.) have reduced the cost of depth sensors, there is now a push to have depth sensors in a compact and portable medium. Potential applications include 3D scanning with mobile phones, navigating natural environments with robots, or object recognition with surveillance drones.

Currently, there are few low-cost depth sensors available in a small form factor. Existing solutions are stereophotogrammetric systems such as the Leap Motion Controller (Leap Motion Inc. [2]) and DUO (Code Laboratories Inc. [3]) that use two or more cameras to perform stereophotogrammetry and obtain depth information of the scene. However, the limitations of these systems are the small stereo baseline and low quality camera sensors; these limitations make stereo correspondence difficult and results in many non-existent or erroneous matches. Hence, these sensors are capable of only obtaining sparse depth measurements. For applications such as 3D scanning, dense depth maps are required and sparse depth measurements are insufficient for creating high quality models. Hence to have a system that is low-cost, compact, and capable of generating accurate dense depth maps is highly desirable.

The main contribution of this paper is to introduce a new method of generating dense depth maps using low-cost compact stereophotogrammetric systems. The method uses a novel multilayer conditional random field (MCRF) model for labelling the dense depth map given the observations consisting of photographic measurements (image intensity values obtained from the cameras) and sparse depth measurements (from photogrammetry). The MCRF extends the traditional CRF model by including the sparse depth measurements as an additional observation layer with missing observations due to sparsity. As well, the MCRF model uses multivariate feature functions based on the photographic and depth measurements to define unary and pairwise relationships between the observations and labels. Using the MCRF model, the dense depth map reconstruction problem is formulated as a maximum a posteriori (MAP) inference problem. The proposed MCRF model has the advantage that additional observation layers and feature functions, such as using multi-spectral measurements, can easily be incorporated into the model. While the proposed method is designed around stereophotogrammetric systems, it can be applied to any depth sensor that provides depth and photographic measurements (e.g., Kinect).

The organization of the paper is as follows: in Section II we discuss previous work in depth map reconstruction and dense depth map generation in stereo imaging. Following that is the methodology in Section III, where we begin by giving a general overview of our system and what the

expected input/outputs are. We then formulate the MAP problem and the proposed MCRF. Section IV describes our implementation on obtaining the desired dense depth map. Section V explains the experiments that we run to evaluate the proposed method and results are shown in Section VI.

II. Related Work

This section presents previous work that can be divided into two groups; the first group is methods related to depth map refinement with smoothing filters and the second is methods using probabilistic graphical models for inferring image correspondences in stereophotogrammetry.

Most depth map refinement methods incorporate some variant of a smoothing filter to fill in the holes of the depth map. Previous work in hole filling is prominent in depth image based rendering (DIBR) techniques for 3D television applications[4], [5]. In DIBR, a single depth map is used to render a left and right warped image. Chen *et al.* [4] use an edge-aware averaging filter to perform hole filling in the depth map and warped image. Lee and Ho [5] use an adaptive smoothing filter that preserves depth edge discontinuities. There have also been work done using cross-bilateral filters where the depth and corresponding colour images are used to maintain consistency in depth maps generated from the Kinect and noisy stereophotogrammetric measurements [6], [7], [8]. Vijayanagar *et al.* [6] use a Gaussian-weighted bilateral filter based on neighbouring colour similarity, then perform a post processing refinement step based on SSID method to show decent edge preservation while smoothing over holes. However, these methods would not be able to generate accurate depth values in areas of the depth map where there are very sparse measurements and large holes.

In stereophotogrammetry, the problem of dense stereo correspondence (where each pixel is assigned a disparity value) has been studied extensively and continues to be an active area of research. Scharstein and Szeliski [9] give a detailed taxonomy of dense stereo matching algorithms and show that most stereo methods generally consist of four steps: matching cost computation, cost aggregation, disparity computation, and disparity refinement. In the past decade, dense stereo correspondence matching algorithms have seen success using probabilistic graphical models such as Markov random fields (MRFs) for cost aggregation and disparity computation. Using the MRF approach, pixels are connected to its neighbours with pairwise relations and disparities are inferred by minimizing an energy function across the image with a smoothness assumption (i.e., similar neighbouring intensity values should have similar disparities). Szeliski *et al.* [10] performed a study comparing MRF-based stereo correspondence approaches. Popular methods of optimizing the cost function in the MRF are belief propagation [11], [12] and graph cuts [13], [14]. More recent methods use different ways to optimize the cost aggregation cost either locally [15], [16], [17] or globally [18], [19]. Yoon and Kweon [16] use an adaptive support weight search window using color similarity and geometric proximity. Rhemann *et al.* [17] make use of a box and guiding filter to generate spatially smooth disparities while preserving edges. Yang [19] proposed modelling the guiding image with a tree structure and using minimum spanning tree (MST) to aggregate the matching costs. Mei *et al.* [18] expand on this concept but use a "Segment-Tree", where pixels are first segmented based on color similarity. Segmentation-based stereo is also commonly used, where a plane is fit to segments of similar pixel intensities [20], [21], [22].

The main difference between the proposed method and dense stereo correspondence matching algorithms is that the proposed method reconstructs a dense depth map from available sparse depth measurements and the corresponding photographic measurements, while the stereo correspondence matching methods attempt to infer the dense depth map directly from photographic measurements. One potential advantage of the proposed method is that, by decoupling the correspondence problem from the dense depth map reconstruction problem, one can not only reduce computational complexity but also incorporate sparse depth measurements from difference sensor sources, particularly beyond stereophotogrammetry. Again, the focus of the proposed MCRF method is to reconstruct dense depth maps from compact, low-cost sensors. Hence, in many cases the photographic measurements may be noisy and in the case of narrow baselines, much of the background has little to non-existent disparity values. This makes correspondence difficult in many dense stereo methods, especially ones using global models.

III. Methodology

As mentioned earlier, the goal is to use the sparse depth measurements and the photographic measurements to generate a dense depth map. The proposed system is shown Figure 1 and described as follows. The depth sensor generates photographic and sparse depth measurements. These measurements are the observations and input to the proposed MCRF depth reconstruction method, which infers and outputs a dense depth map. The dense depth map can then be used for example as a viewpoint in 3D reconstruction.

A. Problem Formulation and the MCRF

Given as inputs to the depth map reconstruction method are 2 sets of observations: the photographic measurements and the sparse depth measurements. Let the set of random variables X^c and X^d represent these observations respectively and $X = [X^c, X^d]$. Let Y be the set of labels representing the reconstructed dense depth map. Then, given observations X we want to find the most probable depth values in the inferred depth map Y. This is written as a

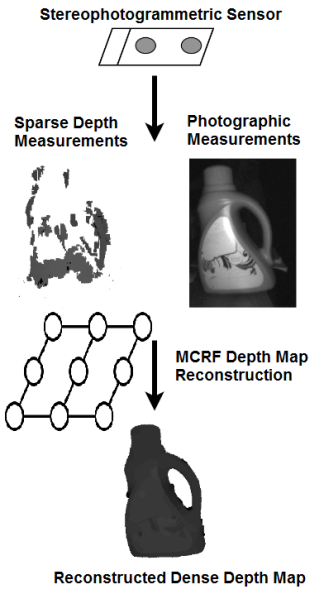

Figure 1. Proposed system flow diagram. The stereophotogrammetric sensor produces two images - the photographic and sparse depth measurements - which are then fed into the MCRF depth reconstruction method to produce a dense depth map.

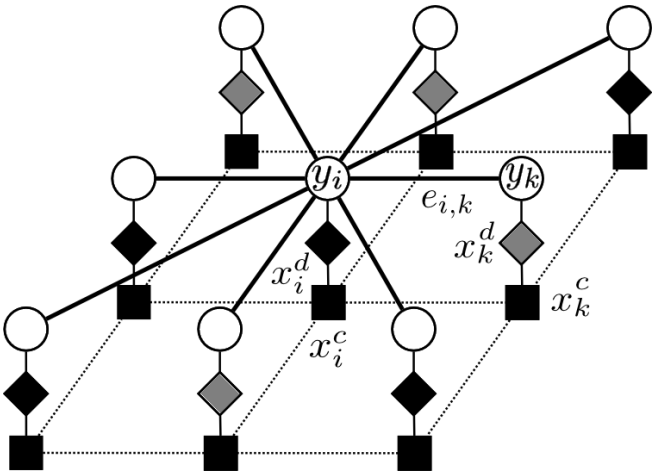

Figure 2. One unit of a repeating pattern used in the MCRF model showing the relationship between the depth map labels y_i and its neighbour y_k, depth observation X_i^d and photographic observation X_i^c. Each node represents a pixel of the photographic/depth measurements.

maximum a posteriori (MAP) problem:

$$Y^* = \arg\max_{\hat{Y}} P(Y \mid X) \tag{1}$$

where $\hat{Y}$ is the set of all possible states of Y and Y^* is desired the realization of Y which has the maximum probability.

To solve this MAP problem we propose a multilayer conditional random field (MCRF) approach. Each pixel in the dense depth map is modelled as a node in an undirected graph that follows the Markov property. The model structure is shown in Figure 2. Underneath the labels Y (shown as white circles) are the two observation layers: the depth observations X^d(diamonds) and photographic observations X^c(squares). A black-filled diamond represents that a depth observation exists for that node while a grey diamond represents a missing observation. The photographic observations are assumed to have no missing observations. The edge e_{ik} represents the influence of neighbouring nodes on the center node/pixel $y_i \in Y$ where $|Y| = n$, which is a weighted combination of its neighbours y_j, X^d, and X^c. The MCRF model extends the traditional CRF model by adding an observation layer to account for the sparse depth measurements. Using a Gibbs distribution, the resulting MCRF distribution is expressed as:

$$P(Y|X) = \frac{1}{Z(X)} exp\Big(-\psi(Y, X^d, X^c)\Big) \tag{2}$$

where $Z(X)$ is the normalization function and $\psi(\cdot)$ is some positive potential function. Comparing (1) and (2), it can be seen that the desired state Y^* that maximizes the posterior probability in (1) is the solution that minimizes the potential function:

$$Y^* = \arg\min_{\hat{Y}} \psi(Y, X^d, X^c) \tag{3}$$

The potential function ψ can be decomposed into unary and pairwise potentials ψ_u and ψ_p. The unary potential ψ_u defines the relationship between labels Y and depth measurements X^d while the pairwise potential ψ_p defines the relationship between each node in Y to its surrounding neighbours:

$$\psi(Y, X) = \alpha \sum_{i}^{n} \psi_u(y_i, X^d) + \beta \sum_{i}^{n} \sum_{j \in N_i} \psi_p(y_i, y_j, X^d, X^c) \tag{4}$$

α and β are weighting parameters and N_i is the set of neighbours of node i. The unary potential ψ_u enforces how similar the inferred states Y should be to the sparse observations X^d:

$$\psi_u(y_i, X^d) = \begin{cases} |y_i - x_i^d| & x_i^d \text{ is observed} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Because the goal is to minimize $\psi(\cdot)$, large differences between the inferred state and its observations i.e. $|y_i - x_i^d|$ are penalized. How much this penalty affects the final solution is controlled by weight parameter α in (3); if the initial depth measurements are very accurate then α should be high otherwise it should be set low.

The pairwise potential ψ_p describes the influence of neighbouring state y_j on y_i based on the observations X^c:

$$\psi_p(y_j, y_i, X^c) = w_d(i,j)\Big(\lambda_d f(y_i, y_j) + \lambda_c f(X_i^c, X_j^c)\Big)(y_i - y_j) \tag{6}$$

where w_d is a Gaussian weight based on the distance between the neighbours. Here f is an inverse exponential feature function

$$f(x_1, x_2) = exp\frac{-(x_1 - x_2)^2}{\sigma^2} \quad (7)$$

where σ is the pairwise multiplier and controls the range of differences that affect the feature function. For large differences between x_1 and x_2, $f(x_1, x_2)$ is small, and vice versa. We use this function to give higher influence for homogeneous regions in the photographic measurements, where the assumption is a smooth patch in the photographic measurements corresponds to smooth depth values at that same patch. It is also used to maintain smoothness in the inferred depth map in areas where the initial depth measurements are rough . λ is a weighting parameter to define the contribution of each variable considered in the pairwise potential function. While here we only consider two feature functions (neighbouring depth and intensity values), the model can be extended to incorporate multiple features (e.g. multispectral measurements). Generally, for k number of features, the pairwise potential is:

$$\psi_p(y_i, y_j, X^1 ... X^k) = \Big(\sum_i^k \lambda_i f_i(X^i)\Big)(y_i - y_j) \quad (8)$$

IV. Implementation

This section describes the methods used to solve the MAP problem formulated in the Section III. The algorithm consists of two steps: first an initial solution is estimated based on maximum likelihood, then the MAP estimation is optimized through steepest descent.

A. Initial Estimate

Since missing observations in the sparse depth measurements effectively have values of zero, it provides no useful information and leads to slow convergence in the optimization. Therefore, an initial estimate for the missing observations is given for the dense depth map. This is done by training a Naive Bayes classifier with depth-photographic measurement pairs of the current frame. No prior information is given into the classifier and so the initial solution is a maximum likelihood estimate (MLE).

$$P(d|i) \propto P(i|d) \quad (9)$$

where d represents the depth estimate and i represents the photographic measurements respectively. A Gaussian distribution is used to fit the training features.

B. Optimization With Steepest Descent Approach

To find the disparity map labels inferred by the MCRF model, the potential function ψ is minimized with the steepest descent approach[23], [24]. For each iteration, a new state for the inferred depth map is calculated based on the previous state and unary/pairwise relations. This process is iterated until an arbitrarily chosen convergence threshold is met.

V. Experimental Setup

We evaluate our method using disparity maps obtained from stereophotogrammetry. A disparity map is inversely proportional to a depth map, therefore the two are treated the same. The depth map reconstruction algorithm was implemented in MATLAB, while stereo correspondence was done using C++ and OpenCV. Two experiments were performed to give qualitative and quantitative analysis.

A. Experiment 1

In the first experiment, data was collected using the 'DUO' stereo camera from Code Laboratories as the depth sensor in our system. This camera was used for its compactness and small form factor, with a baseline width of around 3cm between the two cameras. The camera uses infrared (IR) LEDs at 850nm wavelength for illumination and the camera takes infrared measurements of the scene to capture intensity images at 640x480 resolution. Stereo correspondence was performed using the OpenCV sum of absolute differences (SAD) block matching method [25]. Like many local correspondence algorithms, this method is efficient but the resulting disparity map has many holes in it, especially in regions with very little texture. Our dataset is composed of IR and disparity measurement pairs of selected objects and is shown in Figure 3. These disparity map and IR measurement pairs were passed into our proposed depth map reconstruction method to generate dense disparity maps. Unfortunately, no ground truth data has been created yet for this dataset, hence evaluation is done through qualitative analysis on the re-projected 3D point clouds of the disparity maps.

For comparison, we used a joint bilateral filter (JBF) that uses the photogrammetric measurements to guide the depth map smoothing. Introduced by Kopf *et al.* [26], the JBF method has been used in many applications, one of which is depth map refinement with many adaptations for this purpose [6], [7], [8], [27] Our implementation of the JBF was based on the code by Silva[28] which implements the bilateral filter [29] using a guiding image.

B. Experiment 2

For numerical analysis, the 2006 Middlebury stereo vision dataset [30] was used which provided rectified image pairs and disparity ground truths. We ran stereo correspondence on the images with the same matching algorithm used with the DUO camera and then reconstructed the dense disparity map using the proposed method. The JBF was again used for comparison, along with the Markov random field (MRF)-based global correspondence matching using expansion-move graph cuts algorithm presented in [10], [14], [31], [32]. The source code for this method, along with the datasets, is

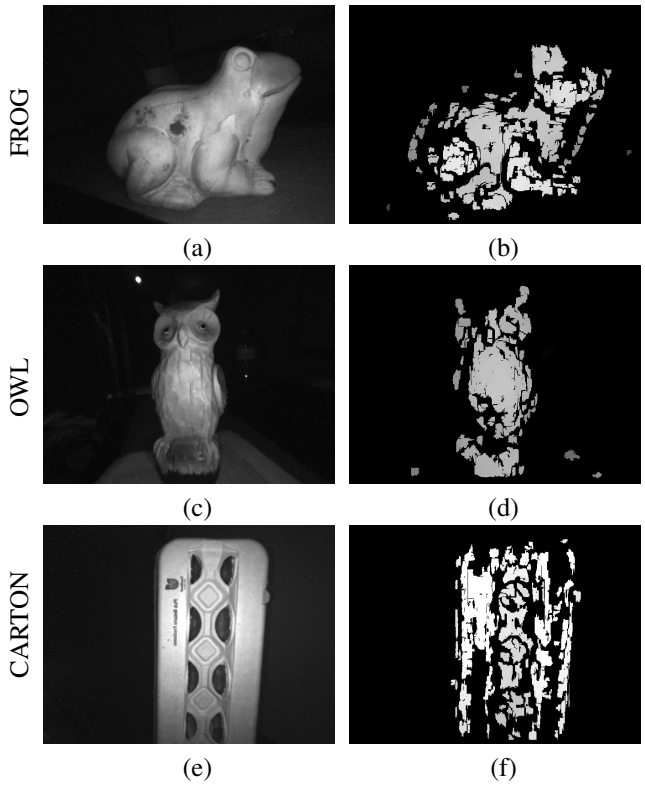

Figure 3. IR and disparity map pairs obtained from DUO camera. The IR image is taken from the left DUO camera.

available on the Middlebury stereo vision website[33]. Note that the MRF method was not used in Experiment 1 because, in general, global correspondence approaches are designed to find a disparity value for every pixel in the image and perform sub-optimally when no such correspondences exist in part of the scene (such as the background in Figure 3). The metric used to evaluate performance was mean squared error relative to the ground truth.

VI. Results

A. Experiment 1 Results

The point clouds of the scanned objects in Figure 3 are shown in Figure 4. The first column shows the point clouds of the initial sparse depth measurements, the second column shows the results after applying the JBF [26], and the last column shows the results after applying the MCRF depth reconstruction method.

Visually, the MCRF method can be seen to correctly infer the missing depth information of the scanned object. Whereas it is difficult to identify the object from the initial measurements, the additional dense points of the reconstructed point cloud makes the object surface and shape much more clear. Compared to the joint bilateral filter, the MCRF produces more reliable points on the object surface with more continuous surface transitions that reflect the surface characteristics of the object. This is seen more clearly in Figure 5, which shows a zoomed in and more detailed view of the point clouds.

Table I
Mean Squared Error of Reconstructed Disparity Maps to Ground Truth. The lowest MSE for each data-set is shown in boldface.

Data	**JBF [26]**	**MRF** [10]	**MCRF**
Aloe	920	1020	**768**
Baby	179	825	**173**
Flowerpots	2309	**2174**	2413
Plastic	1082	6526	**763**
Rocks	710	618	**493**
Wood	732	1196	**690**
AVERAGE MSE	988	2060	**883**

B. Experiment 2 Results

Table 1 shows the mean squared error results of running the proposed method (MCRF) against the MRF graph cut minimization [10] and JBF [26] methods on some of the image pairs from the 2006 Middlebury dataset [30]. The comparison to the ground truth was performed from the resultant 8-bit disparity map images. Figure 6 shows some of the resulting disparity maps and the ground truth.

The proposed method shows lower average mean squared error compared to the other two methods for most of the data. This indicates the MCRF method is able to generate comparable dense disparity maps to dense stereo methods. The reconstructed dense disparity map is also seen to be more stable for textureless regions.

VII. Conclusion

This paper introduced a new method for depth map reconstruction using a novel multilayer conditional random field. Experimental results with a compact stereophotogrammetric camera system showed improvements in the re-projected point cloud based on visual inspection. Quantitatively, the Middlebury stereo dataset was used to show our method exhibited lower mean squared error than the other methods investigated for most of the data. Currently, the performance of the proposed method depends on the quality and confidence of the initial sparse depth measurements. Future work will be to collect sparse but much more stable depth measurements through feature matching. Further future work include optimizing the rate of convergence and speed of the dense depth map reconstruction method, as well as consider including temporal information and multiview frames into the MCRF model. Finally, additional features and observations can be included into the model, such as edges or superpixels, to improve the quality of the reconstructed depth maps.

Acknowledgements

This work was supported by the Natural Sciences and Engineering Research Council of Canada, Canada Research

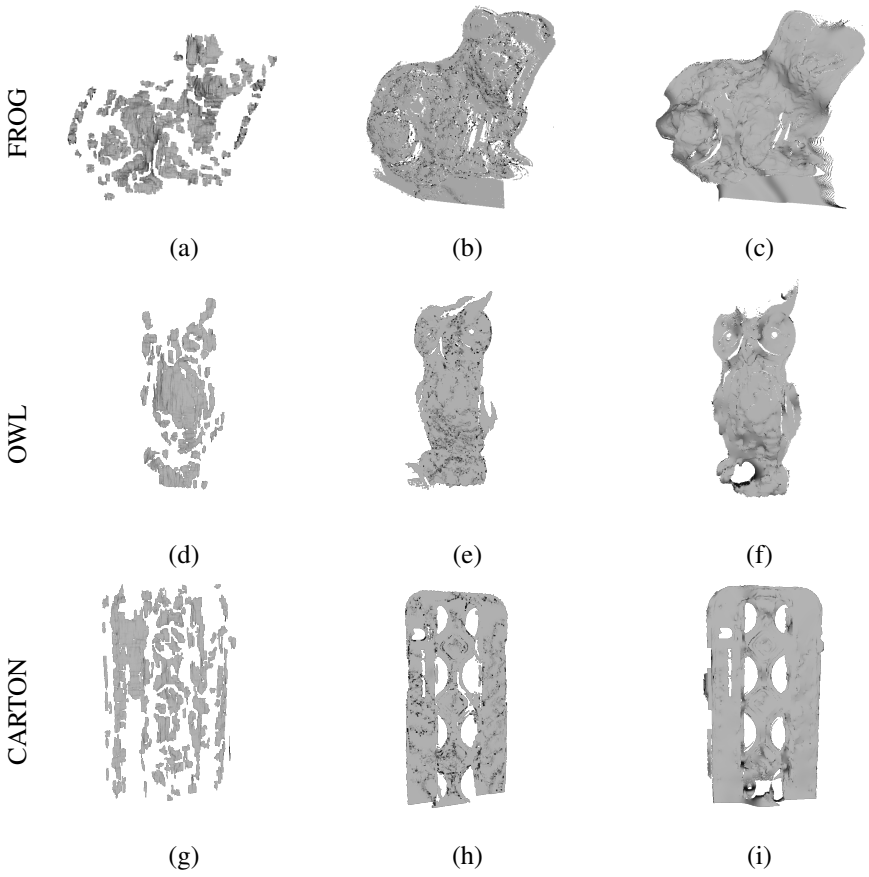

Figure 4. Resultant disparity maps projected onto point clouds. (a,d,g) initial sparse depth measurements. (b,e,h) after JBF [26] (c,f,i) after proposed MCRF depth reconstruction. Both methods are shown to be able to fill much of the missing depth information from the initial sparse depth measurements, though the results of the MCRF method appears to be smoother.

Chairs Program, the Ontario Ministry of Research and Innovation, and the Ontario Graduate Scholarships program.

REFERENCES

[1] L. Bo, X. Ren, and D. Fox, "Learning hierarchical sparse features for rgb-(d) object recognition," *The International Journal of Robotics Research*, vol. 33, no. 4, pp. 581–599, 2014.

[2] Leap motion inc. [Online]. Available: https://www.leapmotion.com/

[3] Code laboratories inc. [Online]. Available: https://duo3d.com/

[4] W.-Y. Chen, Y.-L. Chang, S.-F. Lin, L.-F. Ding, and L.-G. Chen, "Efficient depth image based rendering with edge dependent depth filter and interpolation," in *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*. IEEE, 2005, pp. 1314–1317.

[5] S.-B. Lee and Y.-S. Ho, "Discontinuity-adaptive depth map filtering for 3d view generation," in *Proceedings of the 2nd International Conference on Immersive Telecommunications*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2009, p. 8.

[6] K. R. Vijayanagar, M. Loghman, and J. Kim, "Refinement of depth maps generated by low-cost depth sensors," in *SoC Design Conference (ISOCC), 2012 International*. IEEE, 2012, pp. 355–358.

[7] M. Mueller, F. Zilly, and P. Kauff, "Adaptive cross-trilateral depth map filtering," in *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2010*. IEEE, 2010, pp. 1–4.

[8] P. Lai, D. Tian, and P. Lopez, "Depth map processing with iterative joint multilateral filtering," in *Picture Coding Symposium (PCS), 2010*. IEEE, 2010, pp. 9–12.

[9] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International journal of computer vision*, vol. 47, no. 1-3, pp. 7–42, 2002.

[10] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "A comparative study of energy minimization methods for markov random fields with smoothness-based priors," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 6, pp. 1068–1080, 2008.

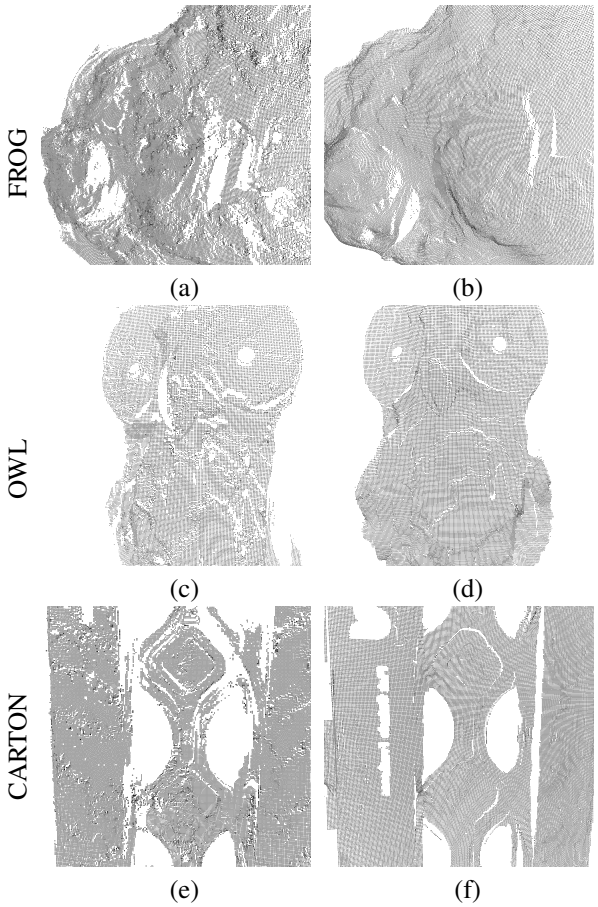

Figure 5. Resultant disparity maps projected onto point clouds. (a,d,g) initial sparse depth measurements. (b,e,h) after JBF [26] (c,f,i) after proposed MCRF depth reconstruction. Both methods are shown to be able to fill much of the missing depth information from the initial sparse depth measurements, though the results of the MCRF method appears to be smoother.

[11] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 2. IEEE, 2001, pp. 508–515.

[12] K.-L. Tang, C.-K. Tang, and T.-T. Wong, "Dense photometric stereo using tensorial belief propagation," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 132–139.

[13] J. Sun, N.-N. Zheng, and H.-Y. Shum, "Stereo matching using belief propagation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 7, pp. 787–800, 2003.

[14] 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.

[15] K. Zhang, Y. Fang, D. Min, L. Sun, S. Yang, S. Yan, and Q. Tian, "Cross-scale cost aggregation for stereo matching," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014, pp. 1590–1597.

[16] K.-J. Yoon and I. S. Kweon, "Adaptive support-weight approach for correspondence search," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 650–656, 2006.

[17] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 3017–3024.

[18] X. Mei, X. Sun, W. Dong, H. Wang, and X. Zhang, "Segment-tree based cost aggregation for stereo matching," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 313–320.

[19] Q. Yang, "A non-local cost aggregation method for stereo matching," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 1402–1409.

[20] H. Tao, H. S. Sawhney, and R. Kumar, "A global matching framework for stereo computation," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 1. IEEE, 2001, pp. 532–539.

[21] Y. Taguchi, B. Wilburn, and C. L. Zitnick, "Stereo reconstruction with mixed pixels using adaptive over-segmentation," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

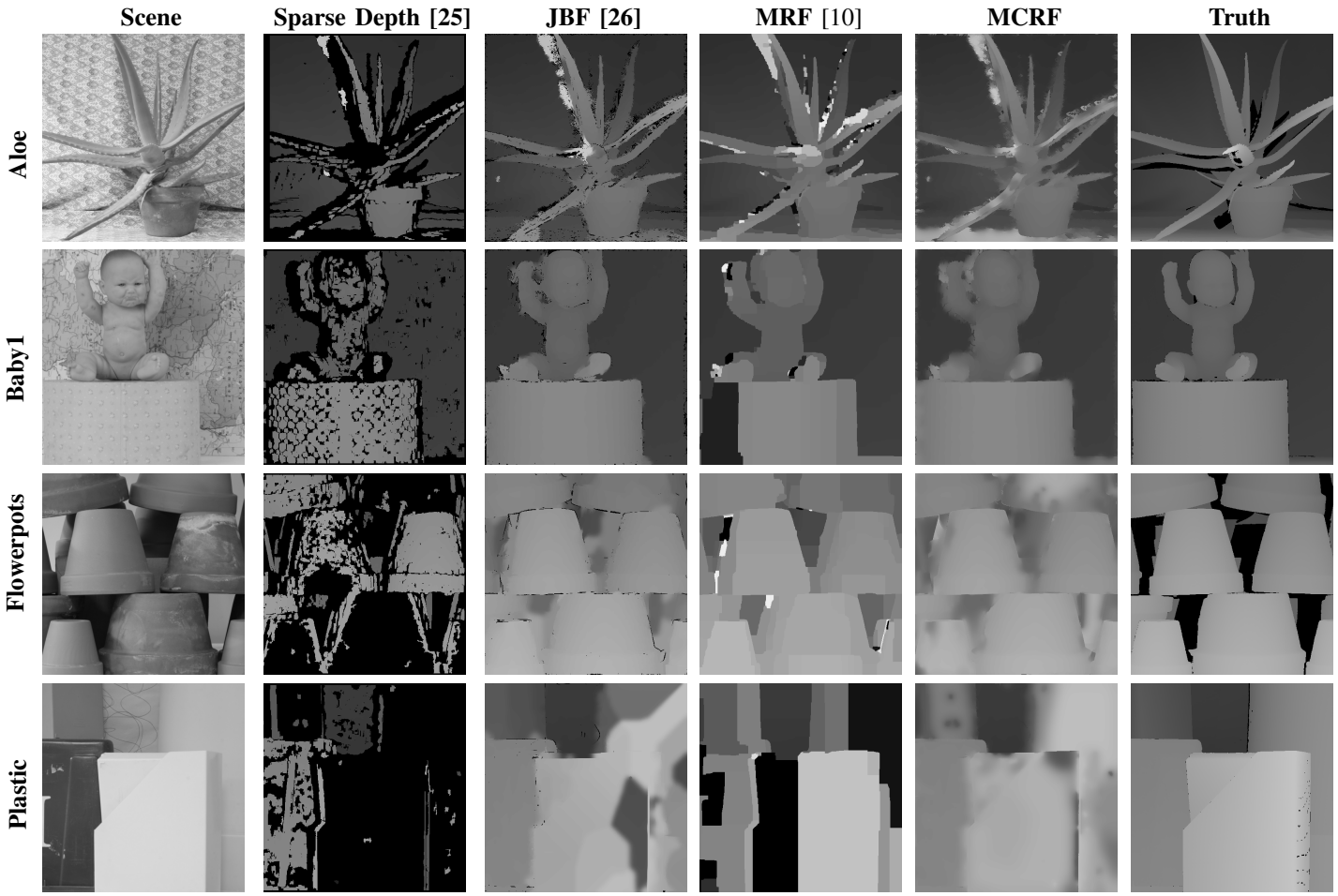

Figure 6. Example reconstructed depth maps for Middlebury data-set using the tested methods. The ground truth depth map and the initial sparse depth measurements using the SAD cost [25] are also shown for context

[22] M. Gerrits and P. Bekaert, "Local stereo matching with segmentation-based outlier rejection," in *Computer and Robot Vision, 2006. The 3rd Canadian Conference on*. IEEE, 2006, pp. 66–66.

[23] P. Rosenbloom, "The method of steepest descent," in *Proc. of Symp. in Applied Math*, vol. 6, 1956, pp. 127–176.

[24] X. Wang, "Method of steepest descent and its applications," *IEEE Microwave and Wireless Components Letters*, vol. 12, pp. 24–26, 2008.

[25] Opencv. [Online]. Available: http://opencv.org/

[26] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," in *ACM Transactions on Graphics (TOG)*, vol. 26, no. 3. ACM, 2007, p. 96.

[27] O. P. Gangwal and R.-P. Berretty, "Depth map post-processing for 3d-tv," in *Consumer Electronics, 2009. ICCE'09. Digest of Technical Papers International Conference on*. IEEE, 2009, pp. 1–2.

[28] Silva, "Joint bilateral filter," http://www.mathworks.com/matlabcentral/fileexchange/27468-joint-bilateral-filter, 2010.

[29] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Computer Vision, 1998. Sixth International Conference on*. IEEE, 1998, pp. 839–846.

[30] D. Scharstein and R. Szeliski. (2006) 2006 stereo datasets with ground truth. [Online]. Available: http://vision.middlebury.edu/stereo/data/scenes2006/

[31] V. Kolmogorov and R. Zabin, "What energy functions can be minimized via graph cuts?" *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 2, pp. 147–159, 2004.

[32] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 9, pp. 1124–1137, 2004.

[33] D. Scharstein. (2012) Source code for mrf minimization. [Online]. Available: http://vision.middlebury.edu/MRF/code/