DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-434

Depth Extraction from a Single Image Based on Block-Matching and Robust Regression

Hyeongju Jeong, Changjae Oh, Youngjung Kim, and *Kwanghoon Sohn
School of Electrical and Electronic Engineering, Yonsei University, Seoul, Korea

Abstract

Predicting scene depth (or geometric information) from single monocular images is a challenging task. This paper addresses such challenging and essentially ill-posed problem by regression on samples for which the depth is known. In this regard, we first retrieve semantically similar RGB and depth pairs from datasets using a deep convolutional activation feature. We show that our framework provides a richer foundation for depth estimation than existing hand-craft representations. Subsequently, an initial estimation is then integrated by block-matching and robust patch regression. It assigns perceptually appropriate depth values to an input query in accordance with a data-driven depth prior. A final post processor aligns depth maps with RGB discontinuities, resulting in visually plausible results. Experiments on the Make 3D and NYU RGB-D datasets show competitive results compared to recent state-of-the-art methods.

1. Introduction

Over the last decades, we have observed a massive advance in 3D-capable hardware, such as 3D TV, smartphone, and virtual reality devices. However, 3D content production still remains challenging task. They are usually produced by labor-intensive processes, which is time-consuming and expensive tasks. To make an abundant 3D contents, it is essential to convert existing 2D scene (or 2D content) into 3D. In this context, many researchers have concentrated on inferring 3D structure from a single monocular scene. However, estimating depth map from a single image is highly ill-posed, since there exist no reliable cues, e.g., stereo correspondence for the depth estimation.

In recent years, much progress has been made towards recognizing 3D scene structures from a single image. At a high level, they can be classified into two groups: semi-automatic and automatic methods. Semi-automatic methods expect a sparse depth scribble as a user interaction. The sparse scribble is propagated into the entire image in order to fill the remaining unknown pixels by modeling global interpolation. Contrary to semi-automatic methods, automatic methods extract depth information from a single monocular image without any user interaction. For example, traditional methods rely on monocular depth cues, such as shape from shading [1], and structure from motion [2]. These are attractive alternatives for estimating 3D structure, but they are unreliable due to strong modeling assumptions. Nowadays, supervised learning approaches are introduced to estimate depth map automatically. They build up a parametric model to describe the relationship between a scene and corresponding depth. It is available to provide realistic depth estimation in general environment. However, they are sensitive to varieties of training data.

After the emergence of large scale RGB-D databases, powerful nonparametric sampling methods are proposed for 2D-to-3D conversion. It is built-upon an assumption that visually similar scenes also have similar 3D structures. Instead of defining explicit parametric model, depth estimates are directly learned from visually similar scenes. The matching candidates are selected from the dataset using a high level features, such as GIST [10] and HOG [11]. Although the retrieved images have semantically similar geometric structures, they are not aligned locally with the input image. As a result, regression process is performed to obtain depth estimation from retrieved samples.

In this paper, we propose nonparametric data-driven approach for 2D-to-3D conversion. We retrieve visually similar scenes with input image using a deep convolution activation feature, which outperforms previous hand-craft representations. To transfer depth patches consisting of similar 3D structures, block-matching is executed considering nonlocal neighborhoods between input and candidate samples. And then, the matched patches are regressed in a robust manner. Finally, we adapt explicit texture removing technique to address structural inconsistency between RGB and depth. Experimental results demonstrate that our method is more superior than previous methods in the entire processes. This paper is organized as follows. Section 2 present related works. The proposed method is explained in Section 3. In Section 4, we demonstrate the effectiveness of the proposed method on MAKE 3D and NYU V2 datasets. Finally, we conclude the paper with some limitations and future works in Section 5.

2. Related Works

To date, semi-automatic methods have been regarded as more successful approach to 2D-to-3D conversion [3, 4]. Conventional methods require highly labor intensive steps, such as separating objects in individual frame, specifying proper depths, and correcting errors after final rendering. Many 2D films have been converted into 3D with this approach. However, those interactions are highly time-consuming and expensive processes. To reduce the burden of annotation process, depth scribble has been adapted as a user interaction [3,4]. Scribble-based input is simple and is generally good for smooth depth regions. In [3, 4], depth estimation is obtained from sparse scribble through optimization-based propagation [28, 30]. Usually, nonlocal neighborhoods is advocated for optimization to reduce the number of user scrib-

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-434

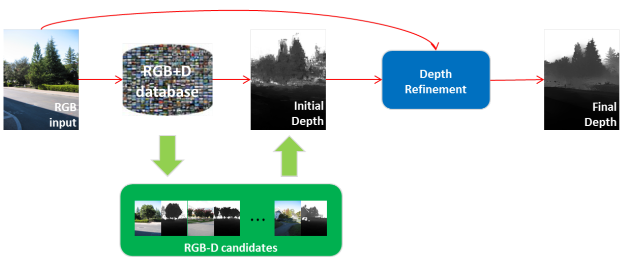

Figure 1. The overall framework of proposed method

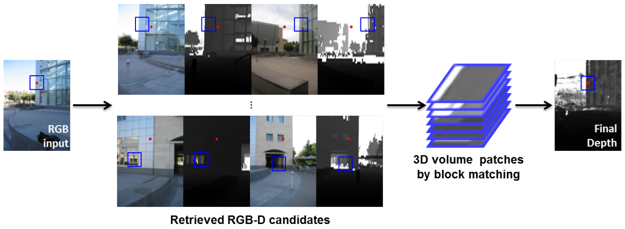
RGB input
Retrieved RGB-D candidates
3D volume patches by block matching
Final Depth

Figure 2. The illustration of grouping by block-matching for depth samples

bles. Although scribble-based 2D-to-3D conversion methods can obtain convincing depth map, they are not suitable for automatic vision-related works.

For automatic 2D-to-3D conversion, traditional methods formulated the problem of depth estimation from a single image in a various ways, such as shape-from-shading [1], and structure-from-motion [2]. Shape-from-shading method [1] estimates depth map from restricted assumption that surfaces of an image consist of uniform color and texture. However, most natural scenes do not satisfy the assumption. Structure-from motion method [2] expects an object motion for estimating a scene depth by motion parallax measured from a video/multiple images taken at a different view points. It also can not estimate depth map when no object motion exists in the scene. Recently, supervised-learning based approaches are proposed for automatic depth estimation [5, 6]. Saxena et al. [5] proposed Markov Random Field (MRF) model to estimate depth map from a single image. The MRF model trains a set of plane parameters to capture a relationship between RGB and depth using large scale of RGB-D dataset. Similarly, Wang et al. [6] train a nonlinear kernel function for the link of the image and depth. However, supervised learning methods are applicable to the pre-trained, category specific environment only.

To tackle the limitation of previous approaches, nonparametric sampling techniques are received lots of attention for estimating a plausible depth map from a monocular image [7–9, 21]. Those methods retrieve visually similar candidates with input image from RGB-D database using a high-level image feature, such as GIST [7, 9] or HOG [8]. Karsch et al. [7] proposed a depth estimation method using the candidates. By SIFT flow [12], warped depth maps is obtained from the matching candidates to align the structure of the input image. A global optimization problem is modeled to regress the warped depth candidates with a robust potential function. Although it can be applicable to arbitrary scenes, it is very time-consuming in both scene alignment and interpolation process. To efficiently compute warping functions, Choi et al. [9] adapt Patch-Match (PM) [29]. Furthermore, they design a transfer model in the depth gradient domain considering statistical invariance property of depth gradients. Poisson equation is solved under Neumann boundary conditions to reconstruct depth from depth gradient.

Konrad et al. [8] proposed nonparametric approach without warping process. The inferred depth values comes from median of retrieved depth maps. These median is performed based on an assumption that the location of some objects (i.e., sky, building, furniture) are quite consistent with the candidate images. A joint filtering is then executed to align depth boundaries to those of the input image. However, if the candidate depths are not locally consistent with input image, it might fail to estimate proper depth values from median.

In our method, we retrieve semantically similar matching candidates from RGB-D dataset using a deep convolutional feature. From the reliable matching candidates, a robust candidate fusion method is proposed considering the geometrical configuration in a whole neighborhood. Furthermore, we perform an effective refinement process, resolving the copying problem frequently occurred in conventional approaches.

3. Proposed Method

The proposed method is based on an assumption that two patches that are semantically similar also have similar geometric structure. This is reasonable, since there is co-occurrence statistics between depth and photometric discontinuities. Given a monocular query image I_0, our objective is to estimate plausible depth map based on large database which consists of RGB images and their corresponding depth maps. Our nonparametric depth estimation begins with retrieving visually similar N scenes from the database. We then perform block-matching, resulting in N depth samples for all patches in the query image. These samples are used for constructing candidate volumes which will be combined to form an initial estimate. Finally, we employ modern edge-aware filter for the refinement. The overall framework is described in Fig. 1.

3.1 Retrieval of training RGB-D

Retrieving visually similar RGB-D candidates from large database is one of the most important processes in depth estimation based on nonparametric learning. To select candidate depths from the database, we retrieve similar images by making use of high-level image feature. In this configuration, visual similarity between two images is measured using the CNN descriptor, which is 4096-dimensional feature vector. In [13], deep convolutional neural network is used to classify the large number of images into the 1000 different classes. The network consists of 5 convolutional and 3 fully-connected layers. The CNN descriptor is the output of the last hidden layer. If two images produce feature activation with small difference, the neural network considers them to be similar. In [14], the authors experimentally show that the CNN descriptor outperforms SIFT feature to describe images. In this way, we extract visually similar N RGB-D candidates from the database. Let g_i denote the CNN descriptor of the image I_i in the database. The CNN descriptors for all images in the database are pre-computed for an efficient retrieval. The matching scores between input query and RGB images in the database are calcu-

DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-434

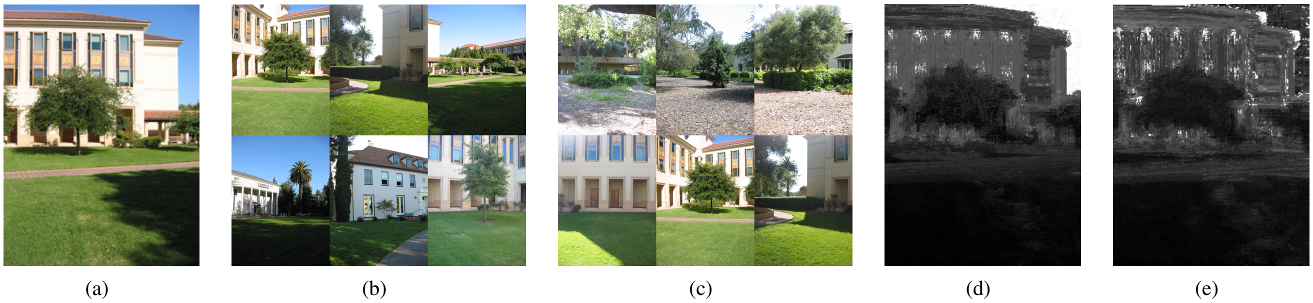

Figure 3. *Retrieved candidates by high-level feature: (a) input image, (b) Extracted candidates by the CNN descriptor, (c) Extracted candidates by GIST, (d) initial estimation using (b), (e) initial estimation using (c)*

lated using the sum of squared difference (SSD) as follows:

$$dist(I_0,I_i) = \|g_0 - g_i\|_2^2, \tag{1}$$

where g_0 is the CNN descriptor of the input image. We then select the top k matching pairs with the lowest SSD. The pairs have roughly similar geometric structure with input image. We empirically find that the CNN descriptor is more suitable to retrieve the semantically similar candidates than conventional representation, e.g., GIST [10] or HOG [11].

3.2 Depth inference via block-matching and robust regression

Although the candidate depths are roughly consistent with depth information for the query image, they are not aligned locally. We address this problem by establishing nonlocal neighborhoods between an input query and each of the retrieved images. In a nutshell, we group hypothetical depth samples via block-matching [15], resulting in N depth samples for all patches in the query image. Rather than using a patch in the spatial domain, the matching process based on nonlocal principle is implemented by computing K nearest neighbors (KNN) with $k = 1$ in the feature space. At each pixel in the input image, we find a patch in each RGB candidates using a feature vector $X(i) = (S_i, x, y)$ at a pixel i, where S_i is 128-dimensional SIFT feature [16] and (x, y) denotes the spatial coordinates of pixel i with some weight. Since the retrieved candidates have similar geometric structures, it is reasonable to find somewhat close patch with pixel i in each RGB candidate. It can be efficiently implemented by using KD-tree [17] which is one of the approximated nearest neighbors (ANN). An illustration of grouping by block-matching for depth samples is shown in Fig. 2. The red points indicate the same position for all images and blue boxes denote patches. Blue boxes in the retrieved RGB-D candidates are matched patches with the red point in the input image using KNN with $k = 1$. Since we use SIFT feature for block matching, the patches have similar appearances. We can observe that two blocks with similar appearance indeed share a common structure in a depth image.

The candidate volumes built by stacking the k depth samples are then combined in a robust manner as following regression problem:

$$P_i = \arg\min_P \sum_{j=1}^{N} w_{i,j} \left\| P - P^j_{m^j(i)} \right\|_2, \tag{2}$$

where subscript i is pixel location and superscript j denotes the jth candidate. P_i is depth patch centered at i for input query, P^j is depth patch from the jth candidate and w_j denotes the weight determined by block-matching cost:

$$w_{i,j} = \exp\left(-\frac{\|S_i - S_j\|^2}{\sigma^2}\right). \tag{3}$$

The P^j is specified by a mapping function, $m^j(\cdot)$, according to the results of block-matching. The solution of (2), P_i, is Euclidean median of stacked N depth samples. There exists an extensive literature on the computation of Euclidean median [18] [19]. We solve this problem using Weiszfeld algorithm [18] [19]. The algorithm is a form of IRLS. Given the current estimate $P_i^{(t)}$, the next iterate is obtained as follows:

$$P_i^{(t+1)} = \arg\min_P \sum_{j=1}^{k} w_{i,j} \frac{\left\| P - P^j_{m^j(i)} \right\|_2^2}{\left\| P_i^{(t)} - P^j_{m^j(i)} \right\|_2}. \tag{4}$$

(4) is a least-square problem, and the minimizer is given by

$$P_i^{(t+1)} = \frac{\sum_j \mu_j^{(t)} P^j_{m^j(i)}}{\sum_j \mu_j^{(t)}}, \tag{5}$$

where $\mu_j^{(t)} = w_{i,j} / \left\| P_i^{(t)} - P^j_{m^j(i)} \right\|_2$. After processing all the patches in the query image, the obtained patch estimates can overlap and thus there are multiple estimates for each pixel. We average them to form an initial estimate.

3.3 Post-processing with joint filtering

The initial estimate has meaningful geometric information as shown in right side of Fig. 2. However, due to the outliers in candidate depth patches, the initial estimate remains still noisy. Also, it is not aligned with boundaries of the input image. To resolve the problems, we refine the initial estimate by means of joint filtering technique. In our method, the weighted median filter (WMF) [20] is employed, which accomplishes smoothing via L1 norm minimization. Traditional approaches [8], [21] usually exploit the edge-aware averaging filters (EAF) such as [22] and [23] for refinement. However, EAFs complete smoothing through L2 norm minimization, so they are not suitable for depth refinement in that

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-434

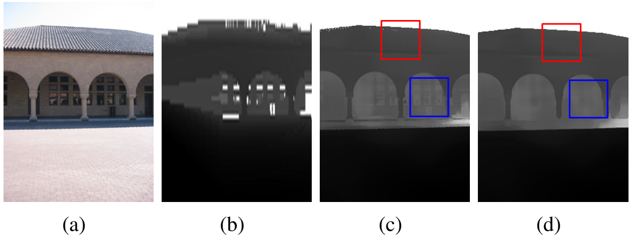

Figure 4. *Comparison of refinement process with and without texture handling: (a) input image, (b) ground truth, (c) refined depth without texture smoothing, (d) refined depth with texture smoothing*

the quadratic gives severe influence to erroneous depth hypotheses [20].

One of the problems in the joint filtering techniques is texture copying of guidance image. Thus instead of using the input image directly as a guidance image, we filter it with rolling guidance filter [24] to handle texture copying problem. By using rolling guidance filter, we can alleviate texture in the input image.

4. Experimental Results

In the experiments, we analyze the performance of the proposed method on various monocular images using both outdoor and indoor scenes. We use Make3D dataset [5] for outdoor scenes, consisting of 534 outdoor scenes and their corresponding depth maps captured by a layer scanner. For indoor scenes, database are built-up by NYU Kinect V2 dataset [25] which consists of 1449 indoor scenes by a Kinect sensor. We select 100 and 654 test images in outdoor and indoor database, respectively. In addition, we retrieve 6 RGB-D candidates from the database in all the experiments. The proposed method was implemented in Matlab and is simulated on a Single PC with Quad-core CPU 4.0GHz and 8.0GB RAM. We show the performance of the proposed method with respect to qualitative and quantitative evaluations.

To show the qualitative evaluation, the experiments are performed to show the excellence of each step of the proposed method. Fig. 3 shows the performance comparison of the proposed retrieval framework with the method of [7]. By comparing Fig. 3(b) and Fig. 3(c), we can confirm that the CNN descriptor is more suitable than GIST in retrieving geometric similar candidates. It can be seen that our method provides a semantically meaningful candidates. Furthermore, we can confirm that the initial estimation using the CNN descriptor is superior to the one using GIST as shown in the Fig. 3(d) and Fig. 3(e). On the contrary, HOG descriptor employed in [8] failed to provide the meaningful candidates. In Fig. 4, we demonstrate the effectiveness of our refinement process. In the input image (Fig. 4(a)), there are highly textured regions, e.g., the roof of the building and the windows between pillars. Fig. 4(c) is the refined depth obtained using guidance image as input image directly. As a result, there exist undesirable artifacts by the textures in red and blue rectangles. On the contrary, the proposed method obtains the refined depth (Fig. 4(d)) using texture smoothed input image. As shown in Fig. 4(d), it is free from such artifacts. Fig. 5 shows the results of the proposed method and competing algorithms [5,7,8] with natural outdoor scenes. Our method estimates better results which reflect proper 3D structures. It also preserves sharp edges around image structure, whereas the competing methods lack in preservation of depth discontinuities.

Table 1. Quantitative Evaluation

	Average C	Median C	time(s)
Depth Transfer [7]	0.73	0.79	120
Konrad et.al [8]	0.80	0.86	2
Make3D [5]	0.78	0.78	25
Proposed method	0.87	0.87	22

For the quantitative evaluation, the normalized cross-covariance (C score) is measured between the estimated depth map and the ground truth. C score is calculated in n test set by the following equation:

$$C = \frac{1}{n\sigma_{d_E}\sigma_{d_G}} \sum_x (d_G[x] - \mu_{d_G})(d_E[x] - \mu_{d_E}), \tag{6}$$

where μ_d and σ_d are the mean and the standard deviation of depth map d, and d_E and d_G denote estimated depth map and ground truth depth map, respectively. Higher score indicates higher correlation to the ground truth. Table 1 summarizes the results of quantitative evaluation. The proposed method outperforms other existing methods. Fig. 6 shows C scores of the estimated depth maps with respect to other competing methods. The running times of competing algorithms and the proposed method are shown in Table 1. Our method is much faster than [7], and slightly faster than [5]. Although it is slower than [8], our method produces more accurate depth map than [8].

5. Conclusion

In this paper, we propose a depth estimation method from a single monocular scene based on nonparametric sampling approach. Based on the CNN descriptor, we can retrieve more similar candidates with the input image than conventional representation, e.g., GIST or HOG. The initial depth estimation is obtained from the candidates using block matching and robust regression algorithm which considers non-local neighborhood patches. Finally, we effectively refine the initial estimation, resolving texture copying problem frequently occurred in joint filtering methods. Experimental results show the presented method outperforms previous methods in terms of accuracy. Although our method is not faster than some competing methods, it can produce more promising results than previous methods. In future works, we will extend the proposed method to video by accelerating the algorithm.

References

[1] J. Atick, P.Griffin, and N. Redlich, ”Statistical Approach to Shape from Shading: Reconstruction of Three-dimensional Face Surfaces from Single Two-dimensional Images,” *Neural Computat.*, vol. 8, no. 6, pp. 1321-1340, 1996.

[2] R. Szeliski and P. H. S. Torr, ”Geometrically Constrained Structure from Motion: Points on Planes,” in *Proc. European Workshop 3D Struct. Multiple Images Large-Scale Environ*, pp. 171-186, 1998.

[3] O. Wang, M. Lang, and M. Gross, ”StereoBrush: Interactive 2D to 3D Conversion using Discontinuous Warps,” in *Proc. EUROGRAPHICS*, 2011.

DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-434

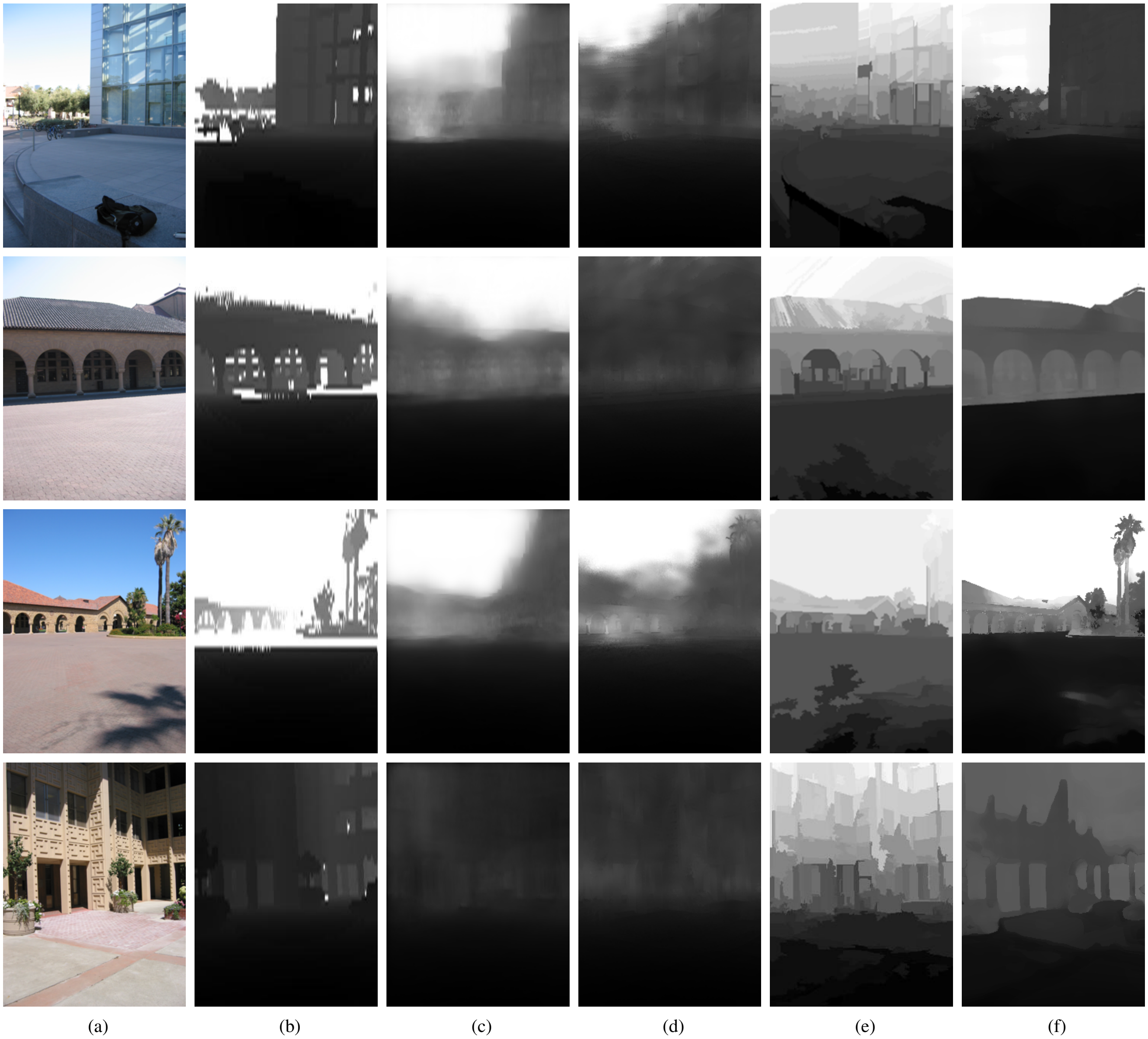

Figure 5. *Qualitative comparison (a) input image, (b) ground truth, (c) Depth Transfer [7], (d) Konrad et.al. [8], (e) MAKE3D [5], (f) our method*

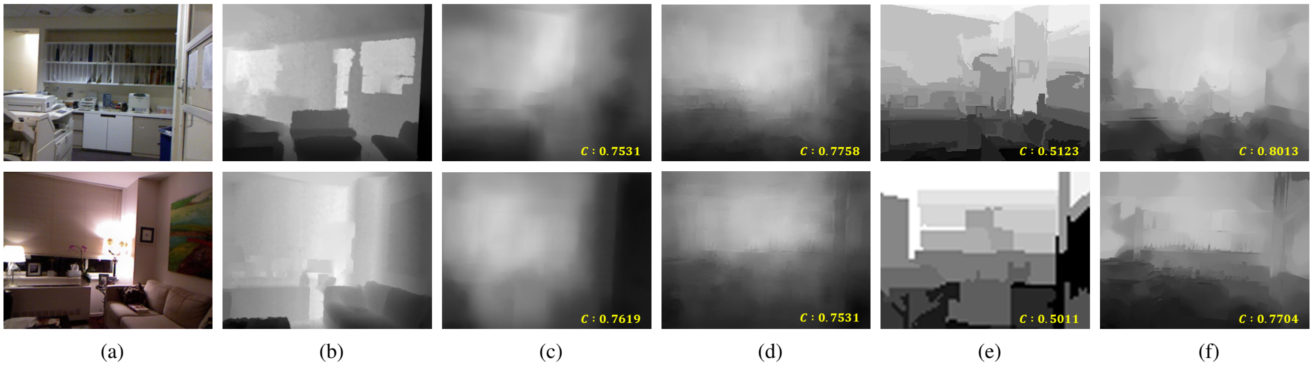

Figure 6. *Quantitative comparison (a) input image, (b) ground truth, (c) Depth Transfer [7], (d) Konrad et.al. [8], (e) MAKE3D [5], (f) our method*

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-434

[4] H. Yuan, S. Wu, P. Cheng, P. An, and S. Bao, ”Nonlocal Random Walks Algorithm for Semi-Automatic 2D-to-3D Image Conversion,” *IEEE Signal Porcessing Letters*, vol. 22, no. 3, pp. 371-374, 2015.

[5] A. Saxena, M. Sun, and A. Ng, ”MAKE3D: Learning 3D Scene Structure from a Single Still Image,” *IEEE Transaction on Patern Analysis Machine, Intelligence* (TPAMI), vol. 31, no. 5, pp. 824-840, 2009.

[6] Y. Wang, R. Wang, and Q. Dai, ”A parametric model for describing the correlation between single color images and depth maps,” *IEEE Signal Processing Letters*, vol. 21, no. 7, pp. 800-803, 2014.

[7] K. Karsch, C. Liu, and S. Kang, ”Depth Extraction from Video using Non-parametric Sampling,” in *Proc. European Conerence on Computer Vision* (ECCV), pp. 775-788, 2012.

[8] J. Konrad, M. Wang, and P. Ishwar, ”Learning-Based, Automatic 2D-to-3D Image and Video Conversion,” *IEEE Transaction on Image Processing* (TIP), vol. 22, no. 9, pp. 3485-3496, 2013.

[9] S. Choi, D. Min, B. Ham, Y. Kim, C. Oh, and K. Sohn, ”Depth Analogy: Data-Driven Approach for Single Image Depth Estimation using Gradient Samples,” *IEEE Transaction on Image Processing* (TIP), vol. 24, no. 12, pp. 5953-5966, 2015.

[10] A. Oliva, and A. Torralba, ”Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope,” *International Journal of Computer Vision* (IJCV), vol. 42, no. 3, pp. 145-175, 2001.

[11] N. Dalal and B. Triggs, ”Histograms of Oriented Gradients for Human Detection,” in *IEEE Computer Vision and Pattern Recognition* (CVPR), vol. 1, pp. 886-893, 2005.

[12] C. Liu, J. Yuen, and A. Torralba, ”SIFT Flow: Dense Correspondence across Scenes and Its Applications,” *IEEE Transaction Pattern Analysis Machine Intelligence* (TPAMI), vol. 33, no. 33, pp. 978-994, 2011.

[13] A. Krizhevsky, I. Sutskever, and G. Hinton, ”ImageNet Classification with Deep Convolutional Neural Networks,” *Advances in Neural Information Processing Systems* (NIPS), pp. 1097-1105, 2012.

[14] P. Fischer, A. Dosovitskiy, and T. Brox, ”Descriptor Matching with Convolutional Neural Networks: a Comparison to SIFT,” *ArXiv e-prints*, abs/1405.5769, 2014.

[15] S. Ourselin, A. Roche, S. Prima, and N. Ayache, ”Block Matching: A General Framework to Improve Robustness of Rigid Registration of Medical Images,” in *Medical Image Computing and Computer-Assisted Intervention,* S. L. Delp, A. M. Digioia, and B. Jarmaz, Eds. Berlin, Germany: Springer-Verlag, vol. 1935, pp. 557-566, 2000.

[16] D. Lowe, ”Distinctive Image Features from Scale-Invariant Keypoints,” *International Journal of Computer Vision* (IJCV), vol. 60, no. 2, pp. 91-110, 2004.

[17] A. Vedaldi and B. Fulkerson, ”VLFeat: An Open and Portable Library of Computer Vision Algorithms,” http://www.vlfeat.org/, 2008.

[18] E. Weiszfeld, ”Sur le point par lequel le somme des distances de *n* points donnes est minimum,” *Tohoku Math. J.,* vol. 43, pp. 355-386, 1937.

[19] G. Xue and Y. Ye, ”An Efficient Algorithm for Minimizing a Sum of Euclidean Norms with Applications,” *SIAM Journal on Optimization* vol. 7, pp. 1017-1036, 1997.

[20] Z. Ma, K. He, and J. Sun, ”Constant Time Weighted Median Filtering for Stereo Matching and Beyond,” in *IEEE International Conference on Computer Vision* (ICCV), 2013.

[21] J. Konrad, M. Wang, and P. Ishwar, ”2D-to-3D Image Conversion using 3D Examples from the Internet,” in emphProc. SPIE, vol. 8288, 2012.

[22] K. He, J. Sun, and X. Tang, ”Guided Image Filtering,” *IEEE Transaction Pattern Analysis Machine Intelligence* (TPAMI), vol. 35, no.6, pp. 1397-1409, 2013.

[23] C. Tomasi, and R. Manduchi, ”Bilateral Filtering for Gray and Color Images,” *IEEE International Conference on Computer Vision* (ICCV), pp. 839-846, 1998.

[24] Q. Zhang, X. Shen, L. Xu, and J. Jia, ”Rolling Guidance Filter,” *Proc. European Conference on Compututer Vision* (ECCV), pp. 815-830, 2014.

[25] N. Silberman, D. Hoiem, and R. Ferus, ”Indoor Segmentation and Support Inference from RGBD Images,” in *Proc. European Conference on Compututer Vision* (ECCV), pp. 746-760, 2012.

[26] Y. Horry, K.-I Anjyo, and K. Arai, ”Tour into the Picture: Using a Spidery Mesh Interface to make Animation from a Single Image,” in *Proc. of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 225-232, 1997.

[27] L. Zhang, G. Dugas-Phocion, J.-S. Samson, and S. M. Seitz, ”Singleview Modelling of Free-Form Scenes,” *The Journel of Visualization and Computer Animation*, vol. 13, no. 4, pp. 225-235, 2002.

[28] A. Levin, D. Lischinski, and Y. Weiss, ”Colorization using Optimization,” *ACM Transactions on Graphics* (TOG), vol. 23, no. 3, pp. 689-694, 2004.

[29] C. Barnes, E. Shechtman, A. Finkelstein, and D. Goldman, ”PatchMatch: A Randomized Correspondence Algorithm for Structural Image Editing,” *ACM Transactions on Graphics* (TOG), vol. 28, no. 3, p. 24, 2009.

[30] P. Lee, Y. Wu, ”Nonlocal Matting,” *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), 2011.

Author Biography

Hyeongju Jeong (S'15) received the B.S. degree in elcectrical and electronic engineering from Yonsei University, Seoul, Korea, in 2014, where he is currently pursuing the joint M.S. and Ph.D. degrees in electrical and electronic engineering. His current research interests include variational method and optimization, both in theory and applications in computer vision and image processing.

Changjae Oh (S'13) received the B.S. and M.S. degree in electrical and electronic engineering from Yonsei University, Seoul, Korea, in 2011 and 2013, respectively. He is currently pursuing the Ph.D. degree at Yonsei University. His current research interests include 3D computer vision, 3D visual fatigue assessment, and image segmentation.

Youngjung Kim (S'14) received the B.S. degree in electrical and electronic engineering from Yonsei University, Seoul, Korea, in 2013, where he is currently pursuing the joint M.S. and Ph.D. degrees in electrical and electronic engineering. His current research interests include variational method and optimization, both in theory and applications in computer vision and image processing

Kwanghoon Sohn (M'92 SM'12) received the B.E. degree in electronic engineering from Yonsei University, Seoul, Korea, in 1983, the M.S.E.E. degree in electrical engineering from the University of Minnesota, Minneapolis, MN, USA, in 1985, and the Ph.D. degree in electrical and computer engineering from North Carolina State University, Raleigh, NC, USA, in 1992. He was a Post-Doctoral Fellow with the MRI Center, Medical School of Georgetown University, Washington, DC, USA, in 1994. He was a Visiting Professor with Nanyang Technological University, Singapore, from 2002 to 2003. He is currently a Professor in the School of Electrical and Electronic Engineering, Yonsei University. His research interests include 3D video processing, computer vision, and video communication. He is a a senior member of IEEE and a member of SPIE.