Depth Map Post-Processing for 3D-TV

Om Prakash Gangwal[1] and Robert-Paul Berretty[2]
[1]NXP Research, [2]Philips Research, High Tech Campus, Eindhoven, The Netherlands

***Abstract*—Automatically generated depth maps from stereo or monoscopic video are usually not aligned with the objects in the original image and may suffer from large area outliers. We post-process these depth maps using a novel method to first downsample the input depth map followed by filtering and upsampling applying joint-bilateral filters to produce image aligned depth maps at full resolution. Unlike a known method of applying joint-bilateral filters at full resolution depth maps, our method performs better in suppressing object details in the filtered depth map while properly aligning the object boundaries. Moreover, our method has low computational cost which enables usage in consumer products.**

I. Introduction

Image and depth/disparity maps at pixel resolution are used to create multiple views for 3D-TV [1]. Depth maps for 3D-TV require the following properties to produce high quality multiple views, 1) depth map edges must be aligned with object edges in the image, and 2) depth maps should be smooth and uniform within the objects and background.

Automatic generation of depth maps from video is a complex process. We observe that often depth maps are not aligned with objects in the image and suffer from large area outliers (i.e., large group of pixels with undesired depth values) due to the use of heuristics based methods [2] [3]. Common techniques, like median or morphological filtering, can remove outliers and smoothen the depth map but do not result in image aligned depth maps. In contrast, Joint-bilateral filters [5] are applied to create image aligned depth maps. However, they also introduce visibility of object's texture details, which is not desirable. To counter this problem, we propose a novel idea to downsample the input depth map to a lower resolution. This downsampled depth map is then filtered and upsampled using joint-bilateral filters. We show that our method produces better quality results while strongly reducing the computational cost, enabling a cost-effective implementation on consumer electronics embedded platforms.

II. Related Work

Joint-Bilateral filters offer an option to create image aligned depth maps. A bilateral filter [4] is a non-linear filter where the value of each output pixel $I_{\mathbf{p}}$ is a weighted average of pixels $I_{\mathbf{q}}$ in a neighborhood S of position $\mathbf{p}$ (see Eq. 1 and 2). The weight calculation depends on both the *spatial* distance function g and the intensity *range* function r. The function r decreases with increasing intensity differences to deliver edge stopping characteristics. The bilateral filter applied on one image (e.g., depth map D) where weights are calculated from a related image (e.g., the original image I) is called a *cross* or *joint*-bilateral filter [5].

We applied joint-bilateral filter, with g and r as Gaussian function with σ of 0.5 and 0.1 respectively, on depth maps, generated using the SSD algorithm in Middlebury software [3], for stereo images called "books" and "moebius" (see Figure 1). We varied the filter aperture from 25x25 to 137x137 to see its impact on the resulting depth map. In Figure 1c, we observe that with an aperture of 25x25, the edges of the objects in the depth map are not properly aligned, while a large aperture of 137x137 aligns the edges of the objects but has the undesirable effect that the details of the objects become visible in the depth map.

$$I'_{\mathbf{p}} = \sum_{\mathbf{q}\in S} w_{\mathbf{p},\mathbf{q}} I_{\mathbf{q}} \Big/ \sum_{\mathbf{q}\in S} w_{\mathbf{p},\mathbf{q}} \text{, where} \quad (1)$$

$$w_{\mathbf{p},\mathbf{q}} = g(\| \mathbf{p} - \mathbf{q} \|)\, r(I_{\mathbf{p}} - I_{\mathbf{q}}). \quad (2)$$

III. Our Method

We propose to use a three step approach to produce high quality depth maps at acceptable cost: 1) downsampling the input depth map, 2) filtering the depth map using a joint-bilateral filter, and, 3) upsampling the depth map using a joint-bilateral filter.

In the first step, we smooth the input depth map by first downsampling the depth map using a 2D box filter by a factor C (typically 4 or 8) in each direction. This step aids in removing local outliers within the block of CxC pixel.

In the second step, a joint-bilateral filter is applied to filter the downsampled depth map using the corresponding downsampled (using a 2D box filter) image. The downsampling step for the image reduces the details within the objects. As the downsampled depth is smoothened using the downsampled image, containing less details, the depth within the objects uniformly smoothen by the joint-bilateral filter. In this step, a large aperture can be used to smoothen out large area outliers.

$$D^h_{\mathbf{p}} = \sum_{\mathbf{q}\in S} w_{\mathbf{p},\mathbf{q}} D^l_{\mathbf{q}} \Big/ \sum_{\mathbf{q}\in S} w_{\mathbf{p},\mathbf{q}} \text{, where} \quad (3)$$

$$w_{\mathbf{p},\mathbf{q}} = g(\| \mathbf{p} - \mathbf{q} \|)\, r(I^h_{\mathbf{p}} - I^l_{\mathbf{q}}). \quad (4)$$

In the third and final step, the filtered depth map is upsampled to the full image resolution, while keeping the edges in the depth map aligned with the image, using the joint-bilateral filter (see Eq 3 and 4). For the upsampling, the weights are calculated based on low resolution I^l and high resolution I^h images. Upsampling depth in a single-step to the final resolution using joint-bilateral filters is explained in [6]. However, we prefer using a multi-step implementation, where in each step the depth map is upsampled by a factor 2x2. Multi-step implementation allows better control for trade-off in quality versus computation cost [7]. In this step, very small aperture, e.g., 3x3, in each step to upsample the depth map is sufficient to refine the edges.

Our method effectively achieves results similar to applying joint-bilateral filters using low-pass filtered depth and low-pass filtered image at full resolution. Yet, it has a lower computation cost due to processing in the downscaled domain.

978-1-4244-2559-4/09/$25.00 ©2009 IEEE

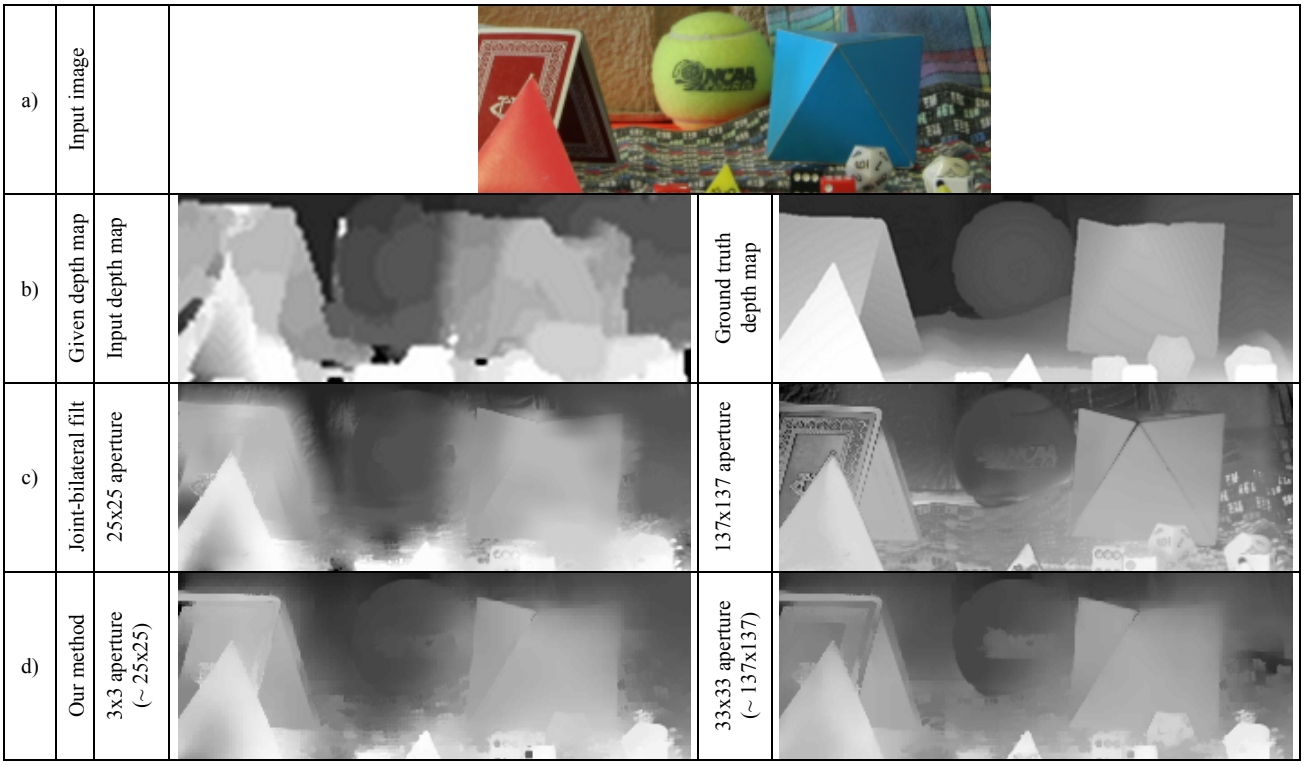

Figure. 1 Experimental results for "moebius" image parts with details (all depth maps are stretched from a luma scale of 110-200 to 0-255 for printing)

IV. RESULTS

To compare our results with the approach described in [5], we apply the same functions to calculate the weights as used in Section II for the joint-bilateral filter steps. In our method, we kept the aperture for the third step fixed to 3x3 for each upsampling step as its role is to refine edges. The 3x3 filter aperture for the second step in our method is chosen corresponding to the effective 25x25 aperture in the full resolution domain. The visual results for the same effective aperture using our method are shown in Figure 1d. In Figure 1d, the depth map on the left clearly has better edge alignment compared to Figure 1c-left, and the depth map on the right clearly does not contain the fine details of the objects in the depth map.

We also measured PSNR compared to the ground truth depth map for the "moebius" and the "books". Figure 2 shows the PSNR for the depth map of detailed parts of the images. We achieve better results compared to the prior-art (also for the whole depth map, not shown) in our experiments. Another major benefit of our approach is the reduced computation cost due to filtering in the downsampled domain (factor 16 to 64).

V. CONCLUSIONS

We showed a novel three step method to post-process low quality depth maps at low computational cost. The first step removes local outliers by downsampling the input depth map. The second step smoothens the depth map in the downsampled domain using joint-bilateral filters to remove large area outliers while aligning with objects in the image. The third step, upsamples the depth map using multi-step joint-bilateral filters to refine the edges in the depth map. The major advantage of this method is that it produces depth maps with strongly reduced visibility of the image object details. Furthermore, our method has low computation cost, as most of the processing is performed in the downscaled resolution, making it suitable for consumer products.

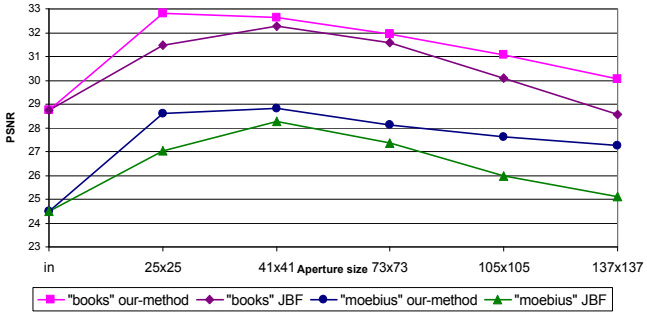

Figure. 2 PSNR compared to the ground truth depth map

REFERENCES

[1] ISO/IEC 23002-3 Auxiliary Video Data Representations Mpeg-C part 3.

[2] R-P.M. Berretty, A.K. Riemens, and P.E. Machado, "Real-time embedded system for stereo video processing for multivitew displays", *Proc of SPIE-IS&T Electronic Imaging SPIE Vol 649014*, 2007

[3] D. Scharstein, R. Szeliski, and R. Zabih, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms", *Proc. of IEEE Workshop on Stereo and Multi-Baseline Vision*, pp. 131–140, 2001. http://vision.middlebury.edu/stereo (accessed June 2008).

[4] C. Tomasi and R. Manduchi, "Bilateral Filtering for Gray and Color Images", *Proceedings of the IEEE International Conference on Computer Vision*, Bombay, India, pages 839-846, January 1998.

[5] G. Petschnigg, M. Agrawala, H. Hoppe, R. Szeliski, M. Cohen, and K. Toyama, "Digital, photography with flash and no-flash image pairs". *Proc. of the SIGGRAPH conf. ACM Trans. on Graphics, 23(3)*, 2004.

[6] J. Kopf, M. Cohen, D. Lischinski, and M. Uyttendaele. "Joint bilateral upsampling". *Proc. of the SIGGRAPH conf. ACM Trans. on Graphics, 26(3)*, 2007.

[7] A.K. Riemens, O. P. Gangwal, B. Barenbrug, R-P. M. Berretty, "Multi-step Joint Bilateral Depth Upsampling", *to appear in SPIE, VCIP 2009*.