This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2513618, IEEE Transactions on Circuits and Systems for Video Technology

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2513618, IEEE Transactions on Circuits and Systems for Video Technology

Color Image Guided Boundary-inconsistent Region Refinement for Stereo Matching

Jianbo Jiao, *Student Member, IEEE,* Ronggang Wang*, *Member, IEEE,* Wenmin Wang, *Member, IEEE,* Dagang Li, *Member, IEEE,* and Wen Gao, *Fellow, IEEE*

***Abstract*—Cost computation, cost aggregation, disparity optimization and disparity refinement are the four main steps for stereo matching. While the first three steps have been widely investigated, few efforts have been taken on disparity refinement. In this letter, we propose a color image guided disparity refinement method to further remove the boundary-inconsistent regions on disparity map. First, the origins of boundary-inconsistent regions are analyzed. Then, these regions are detected with the proposed hybrid-superpixel based strategy. Finally, the detected boundary-inconsistent regions are refined by a modified weighted median filtering method. Experimental results on various stereo matching conditions validate the effectiveness of the proposed method. Furthermore, depth maps obtained by active depth acquisition devices like Kinect can also be well refined with our proposed method.**

***Index Terms*—Disparity refinement, stereo matching, boundary, Kinect.**

I. Introduction

STEREO matching has been one of the most active research areas in computer vision. Most of the stereo matching methods can be divided into four steps [1]: cost computation, cost aggregation, disparity optimization, and disparity refinement. Global and local methods are the two main categories of approaches for stereo matching. Global methods such as graph cut [2] and belief propagation [3] focus on the disparity optimization process, while local methods put great efforts on the cost aggregation, i.e., seeking robust matching window [4], [5], [6] or edge-aware weights [7]. In addition, the functions for cost computation have also drawn increasing attentions recently [4], [7].

However, few efforts have been taken on disparity refinement. Traditional refinement (or post-processing) method typically consists of left-right consistency checking, hole-filling, and a noise-removing filter. In recent years, some new methods were proposed. Mei et al. proposed a cross-region based voting method [4] to update the wrong disparities. Rhemann et al. refined the disparities by a weighted median filter [7]. Wang et al. adopted a segmentation-based mechanism [8] to handle the "fattening effects" [1]. Yang used a non-local aggregation strategy [9] to refine the disparity map. Ma et al. employed a constant time weighted median filter [10] for disparity refinement. Besides the traditional methods, Hirschmuller proposed a semi-global matching method [11] and gave a systematic postprocessing strategy [12], in which a method to distinguish occluded and mismatched pixels was proposed. Although the disparity map can be somewhat improved with the methods above, very often there are still some outliers on the refined disparity map, especially for the disparities near object boundaries.

In this letter, we propose a new disparity refinement approach to refine the false disparities near object boundaries. In our method, color images are involved to guide the whole refinement pipeline. We define the outliers near object boundary as boundary-inconsistent regions, consisting of "fattening" regions [1] and "narrowing" regions. Firstly, we investigate the origins of boundary-inconsistent regions. Then, the boundary-inconsistent regions are detected with a hybrid-superpixel based approach proposed in this letter. Finally, a modified weighted median filtering scheme is proposed to refine these detected outliers. Experimental results on disparity maps computed with various stereo matching methods and depth maps obtained with Kinect demonstrate the effectiveness of our proposed method.

The letter is organized as follows. Section 2 presents the proposed method in detail; section 3 gives the experimental results on various disparity maps and depth maps; and finally, this letter is concluded in section 4.

II. Proposed Method

Fig.1 gives the pipeline of the proposed method. The input is initial disparity map (or depth map) either computed with stereo matching methods or captured with depth sensors. Firstly, the initial disparity map is preprocessed to eliminate dark regions. Secondly, Canny operator is utilized to detect the edges in disparity map and a hybrid-superpixel method is proposed to segment the color image, both the above results are utilized to detect the boundary-inconsistent regions. Finally, a modified weighted median filtering strategy is utilized to refine the detected outliers. In the following paragraphs, why and how to detect and refine the boundary-inconsistent regions are described in detail respectively.

A. Origins of Boundary-inconsistent Regions

After the traditional post-processing, very often there are still many outliers around the object boundaries in the disparity map. Such outliers are mainly caused by the mismatching near

Jianbo Jiao, Ronggang Wang, Wenmin Wang and Dagang Li are with the Digital Media R&D Center, School of Electronic and Computer Engineering, Peking University Shenzhen Graduate School, China (e-mail: jianbojiao@sz.pku.edu.cn; rgwang@pkusz.edu.cn; wangwm@ece.pku.edu.cn; dgli@pkusz.edu.cn). Wen Gao is with the Digital Media R&D Center, School of Electronic and Computer Engineering, Peking University Shenzhen Graduate School, China, and also with the Natl Eng. Lab. for Video Tech., Peking University, China (e-mail: wgao@pku.edu.cn).

* Ronggang Wang is the corresponding author.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2513618, IEEE Transactions on Circuits and Systems for Video Technology

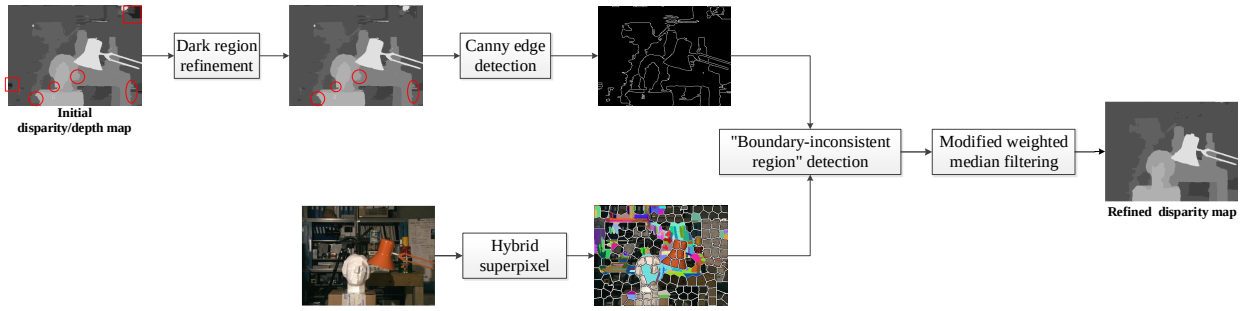

Fig. 1. Pipeline of the proposed disparity refinement method.

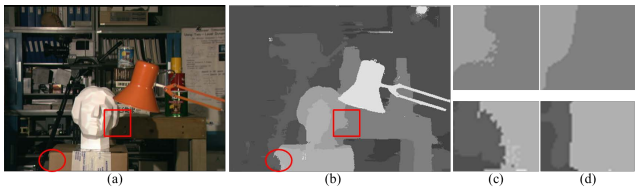

Fig. 2. Fattening region (square) and narrowing region (circle) handling. (a) Color image; (b) disparity map; (c) outlier patches; (d) refined patches.

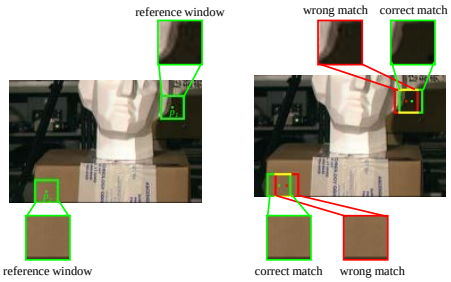

Fig. 3. Examples for fattening effect and narrowing effect. Left and right are the reference image and the target image respectively.

depth discontinuities. We define the regions containing this kind of outliers as boundary-inconsistent regions.

As far as we observed, the boundary-inconsistent regions can be divided into two categories: "fattening" region [1] and "narrowing" region, which are shown in Fig.2 ((b) was generated by the CostFilter[7] without its median filter post-processing). The "fattening" region (marked with red square) is a convex region with outliers spilling out of the object boundary; while the "narrowing" region (marked with red circle) is a concave one. "Fattening" effect is caused by the homogeneous background near depth boundary. As shown in Fig.3, the background near the right border of the head in the target image is occluded in the reference image, and the matching for pixel p_1 in the target image is assigned to a wrong pixel (red one), thus p_1 is assigned with a larger disparity than the ground truth. While the "narrowing" effect is caused by homogeneous foreground, such as the foreground pixel p_2 in Fig.3, which is wrongly assigned with a smaller disparity than the ground truth.

B. Boundary-inconsistent Regions Detection

We define inconsistent boundary as the edge on disparity map which is inconsistent with the border of the corresponding object in the scene. Thus the inconsistent boundary on a disparity map can be detected by checking whether its collocated edge crosses a segment in the color image.

Fig. 4. Disparity maps (a) before and (b) after dark region refinement.

1) Preprocessing: In the initial disparity map, there may be some regions with abnormal disparities, and these regions tend to affect the disparity edge detection. We name this kind of region as dark region (Fig.4(a)). These regions are corrected beforehand by the proposed preprocessing step. Firstly, we assign a disparity with low confidence if it is smaller than a threshold $\rho \cdot d_M$, where ρ and d_M represent the penalty ratio (ρ is set as 1/7 in this letter) and maximum disparity value respectively; otherwise, the disparity is assigned with high confidence. Disparity with low confidence is likely to be an outlier. Furtherly, if the disparity with low confidence is also much smaller than its neighbors, it's labeled as a hole, which is described as follows:

$$hole(p) = \begin{cases} 1, & d(p) < d(q)/2 \\ 0, & otherwise \end{cases} \tag{1}$$

Where $d(p)$ and $d(q)$ denote the disparities of pixel p and q respectively. p is q's nearest neighbor with high confident disparity. When $hole(p)$ is true, pixel p is with abnormal disparity, and those abnormal disparities make up the dark regions in disparity map.

Each detected dark region is corrected by the disparities of its neighbors. When a dark region is on the image boundary, the traditional filling method using smaller disparity (background) tends to cause "dark fill dark" effect. Thus in this case, we use the larger disparity (foreground) to fill the dark region as follows:

$$d(p^*) = \begin{cases} \min\{d(q_1), d(q_2)\}, hole(q_1) \,\|\, hole(q_2) = 0 \\ \max\{d(q_1), d(q_2)\}, hole(q_1) \,\|\, hole(q_2) = 1 \end{cases} \tag{2}$$

where q_1 and q_2 are the nearest pixels on each side of pixel p with high confident disparities (if exist), and the refined disparity is denoted by $d(p^*)$. This process is conducted in both horizontal and vertical directions. The refined disparity

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2513618, IEEE Transactions on Circuits and Systems for Video Technology

(a) (b) (c)

Fig. 5. (a) color image, (b) superpixels with SLIC, (c) segments with our method.

map is shown in Fig.4(b), from which we can see dark regions (in red circles) on the disparity map have been well corrected.

2) Hybrid Superpixel Segmentation: After the preprocessing, the remaining outliers are mainly near object boundaries. In order to refine such regions, we use the segment information from color image together with disparity edge information to detect them. Traditional color image segmentation is usually based on mean-shift segmentation [13] or other graph-based segmentation [14]. However, such algorithms have a high computational complexity. Since what we care in this case is the object boundary, a more efficient superpixel based method of SLIC [15] is used here for the segmentation. The SLIC algorithm adheres to boundaries, and it is fast and memory efficient. One example of segment results is shown in Fig.5(b). However, although most segments adhere to object boundaries, there are still some regions not well segmented. Thus, based on SLIC, we propose the following method to further refine the segmentation.

First the segments not well adhered are detected by the following voting technique. The pixels in each superpixel are divided into five bins according to their gray values: 0∼50, 51∼80, 81∼150, 151∼230, 231∼255. If the bin with the largest number of pixels contains less than 60% of the total pixels in the superpixel, this superpixel is regarded as "under-segmented". Then for the under-segmented superpixel, mean-shift [13] is utilized to further segment it into several regions.

The segmentation result with the proposed hybrid superpixel method is shown in Fig.5(c), in which the colorful regions are the "under-segmented" superpixels further refined with mean-shift.

In addition, to find out the most suitable superpixel size, we perform a trial on the Middlebury dataset [16]. The number of pixels within each superpixel versus the average error percentage of disparity is shown in Fig.6. We can see that superpixel size has little influence on the result, and in our experiments we set the superpixel size at 600 which gives the minimum error percentage.

We also detect the edges in the disparity map with Canny edge detector [17]. We label a disparity edge as an inconsistent boundary if its collocated edge is not aligned with the corresponding segment border (crosses a segment) in color image.

3) Inconsistent Regions Detection: Boundary-inconsistent region is the outlier region lies on one side of the inconsistent boundary. For "fattening" region, it lies on the foreground side [1], while for "narrowing" region it lies on the background side. Thus a judgment is demanded to decide on which side the boundary-inconsistent region lies. Here, we label it by means of "checking-window", as shown in Fig.7. At each point

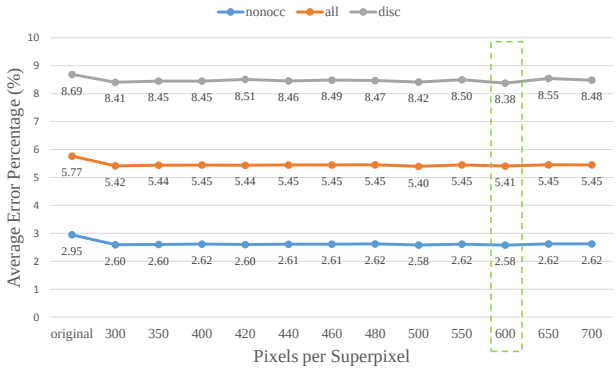

Fig. 6. Determination on superpixel size.

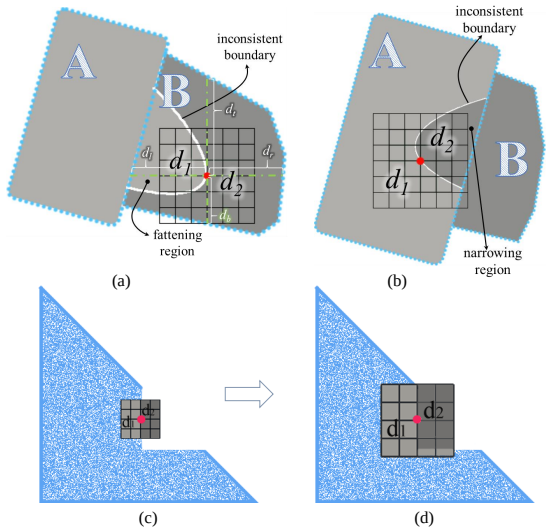

Fig. 7. (a) Fattening region detection and (b) narrowing region detection. d_1 and d_2 are two areas in the checking-window with different disparities. (c) and (d): size changing of the adaptive checking-window.

(red point) along the inconsistent boundary, a square checking-window is formed to detect the boundary-inconsistent region. The radius r (half of the width) of the checking-window is set as the minimum distance from the window center to each border of the segment in four directions:

$$r = \min\{d_l, d_r, d_t, d_b\} \tag{3}$$

where d_l, d_r, d_t, d_b denote the distances in left, right, top and bottom directions respectively (denoted by the green lines in Fig.7(a)). Fig.7 illustrates the complete detection process. Two correct segments, foreground segment A and background segment B are circled with blue dotted line in both Fig.7(a) and Fig.7(b), and the "fattening" effect and "narrowing" effect occur on region A in Fig.7(a) and Fig.7(b) respectively. The inconsistent boundary which crosses region B in Fig.7(a) and A in Fig.7(b) is detected firstly. Then checking-window is applied on the inconsistent boundary. The checking-window is divided into two sides by the inconsistent boundary, and normally the boundary-inconsistent region lies on the side with smaller area (i.e., d_1 in Fig.7(a) or d_2 in Fig.7(b)), since in most cases, outliers region occupies less area of one segment. If not all the pixels on inconsistent boundary give the same conclusion, the majority conclusion is set as the final one.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2513618, IEEE Transactions on Circuits and Systems for Video Technology

If the area of d_1 equals to that of d_2 (Fig.7(c)) exactly, the size of the checking-window will increase until the areas of d_1 and d_2 are different, as shown in Fig.7(d).

C. Boundary-inconsistent Regions Refinement

In our method, we propose a new modified weighted median (MWM) filter based on binocular guided color images to refine the boundary-inconsistent region. The proposed MWM filter is inspired by the constant time weighted median (CTWM) filter [10]. Compared to the traditional median filter ($O(r^2)$, where r is the filter size), CTWM is much more efficient ($O(1)$). In CTWM filter, guided filter [18] weights were used as the filter weights [10]. However, for disparity refinement, there are two images from different views of one scene, thus more information can be utilized. In this letter, we use both of the two views as the guided color images, like the symmetric guided filter mentioned in [7]. In our MWM filter, symmetric guided filter weights [7] are used as the filter weights, so that the information from both views can be utilized, especially the edge information. The process is shown in Fig. 8, in which (a) and (b) are the disparity maps before and after MWM filtering. First, (a) is divided into several slices, in which each slice represents a disparity level, and the content is just the part with that disparity level in (a), as shown in (b). Then for each disparity slice, MWM filter is applied on the boundary-inconsistent regions, as shown in (c). Finally, the filtered slices are combined together to obtain the refined disparity map (d).

III. Experimental Results

In this section, our proposed method is tested on disparity/depth maps obtained with various methods, and also compared with other state-of-the-art disparity refinement methods.

A. Test on Disparity Maps from Stereo Matching

1) Experimental Settings: – Local Methods. The framework of [7] is utilized for the local methods, which consists of a truncated AD-gradient cost measure, a guided filter cost aggregation, and an initial post-processing composed of a left-right check and a weighted median filter. But the cost measure and aggregation are replaced by different strategies in two sub-experiments:

In the first sub-experiment, we fix the cost aggregation and use five most widely used cost functions for the cost computation: absolute difference (AD), gradient, census, truncated AD-gradient [7], and AD-census [4];

In the second sub-experiment, we fix the cost function and choose the following five representative aggregation methods for the cost aggregation: box filter, guided filter [7], cross region [5], information permeability [6], and domain transform aggregation [19].

– Global Methods. We also test our refinement scheme on global methods. The traditional graph-cut based stereo matching [2] and stereo matching using belief propagation [3] are selected in this letter. Related parameters are set as the same as in these methods.

Test image pairs (Tsukuba, Venus, Teddy, Cones) from the Middlebury dataset [16] are used in the above experiments.

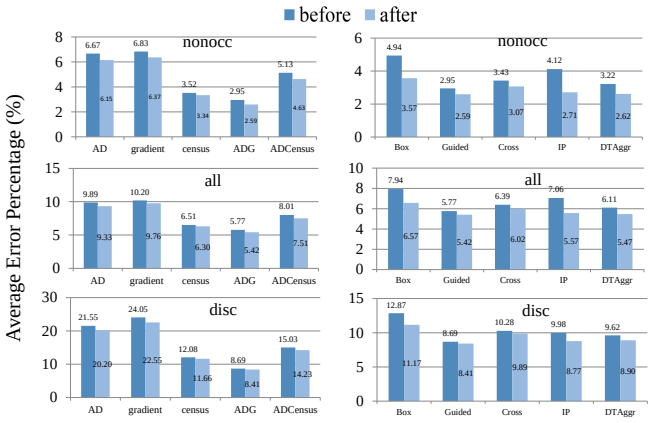

Fig. 9. Experimental evaluation on different cost measures and aggregation methods.

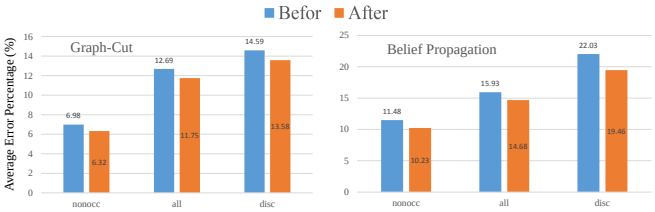

Fig. 10. Refinement results on Graph cut and Belief propagation.

Metrics of "nonocc", "all", and "disc" are calculated for the four image pairs, representing non-occluded region, all region, and the region near discontinuities, respectively.

2) Results and Analysis: Fig.9 gives the average error percentage of local methods over the four Middlebury pairs before and after applying the proposed disparity refinement method. The left side and right side represent different cost measures and different aggregations respectively. It is worth noting that all of these test cases have been already post-processed by a weighted median filter with bilateral weights [7] beforehand. As shown in Fig.9, all the disparity maps are improved over all of the three metrics by our proposed method. Experimental results on global methods are shown in Fig.10, which also validate the effectiveness of our method.

The above experimental results show that even with the state-of-the-art cost measures, aggregation strategies, or post-processing methods, there are still some outliers in the disparity map. And these outliers, especially those near disparity boundaries, are well refined by the proposed refinement method. An average of 10% improvement can be obtained over all the test cases, which also indicates the importance of disparity refinement for stereo matching.

3) Comparison with Other Refinement Methods: We also compare our proposed method with state-of-the-art disparity refinement methods: CTWM [10], FattenHandl [8], Vote [4], jointBF [20] and jointTF [21]. The initial disparity maps are generated by [7], without its disparity refinement step. The above methods are used to refine the initial disparity maps respectively. The visual results and the average error percentage results over the Middlebury dataset are shown in Fig. 11 and Table. I respectively. We can see that our proposed method performs the best among all these methods. Besides,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2513618, IEEE Transactions on Circuits and Systems for Video Technology

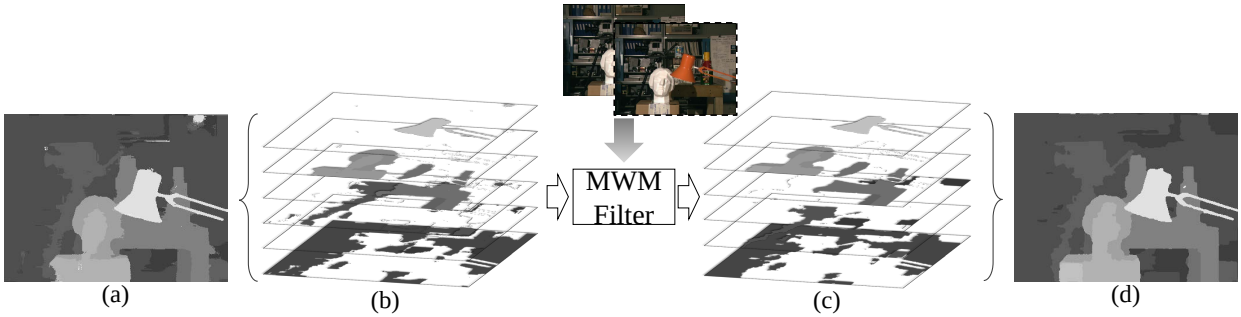

Fig. 8. Modified weighted median filtering process. (a) original disparity map, (b) disparity slices, (c) filtered disparity slices, (d) refined disparity map.

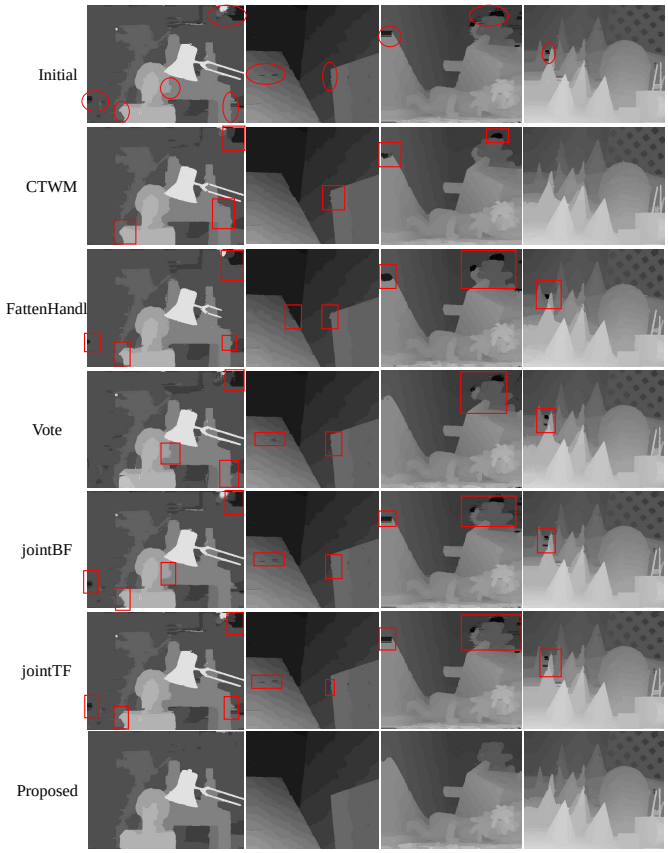

Fig. 11. Comparison with other refinement methods.

TABLE I
COMPUTATIONAL TIME (S) AND AVERAGE ERROR PERCENTAGE (%) COMPARISON

	CTWM	FattenHandl	Vote	jointBF	jointTF	Proposed
Time	3.45	8.85	4.12	0.14	0.11	4.57
Error Percentage	4.95	5.40	5.11	5.61	5.43	4.52

the computational time for image Tsukuba (384×288 with disparity level 16) is shown in Table. I. The computational complexity of our method is on the similar level with those of CTWM, FattenHand1 and Vote, and the MWM filtering process of our method has a high parallelism, which can be easily accelerated on GPU-like platform.

B. Test on Depth Maps from Kinect

Kinect is an active 3D scanner composed of a laser device, a laser feeder, and an RGB camera. Depth map can be acquired by laser scanning. However, there are many outliers in the initial depth maps as shown in Fig.12.

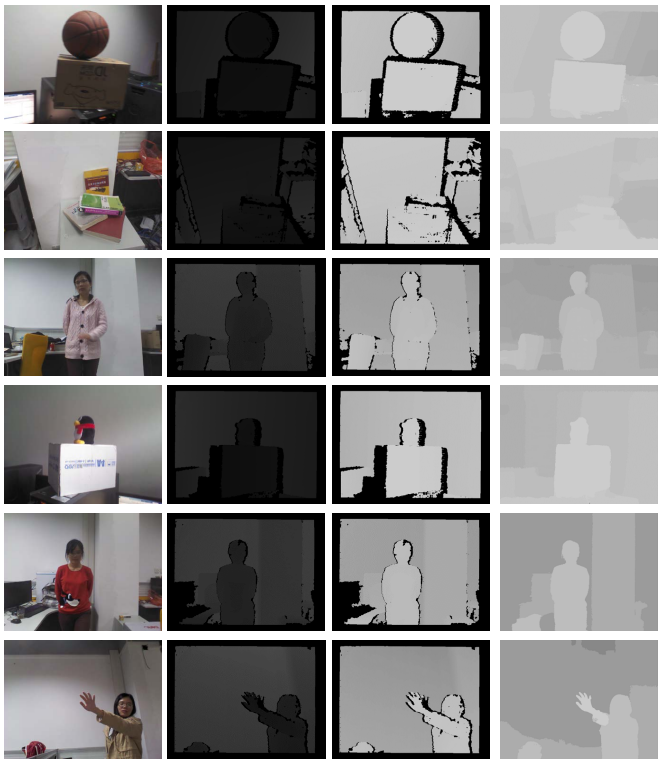

Fig. 12. Experimental results on depth maps obtained with Kinect.

As the depth map is too "dark" and hard to see the content, we first convert the depth to disparity by "disparity=255-depth" (except depth=0), and scale it by 0.8 for better view. And since only one color view can be acquired by Kinect, the guided image in the MWM filter is set to single color image. Then our proposed method is applied on the disparity scales derived from the Kinect dataset. The visual results before and after using our method are shown in Fig.13. We can see that both the dark regions and boundary-inconsistent regions are well refined by our method.

IV. CONCLUSION

In this letter, we proposed a new disparity refinement method. Dark regions were removed with the proposed pre-processing step, boundary-inconsistent regions were detected

by utilizing color image segmentation results, and a hybrid superpixel method is also proposed to improve the segmentation. A modified weighted median filter was proposed to correct the boundary-inconsistent regions. With our method, the outliers in disparity maps, especially outliers near object boundaries were well refined. Experimental results on disparity maps computed with various stereo matching methods demonstrated the effectiveness of our method. In addition, the depth maps captured with Kinect-like depth sensors were also greatly improved with our method.

ACKNOWLEDGEMENT

Thanks to National Science Foundation of China 61370115, 61402018, China 863 project of 2015AA015905, Shenzhen Peacock Plan and Fundamental Research Project for funding.

REFERENCES

[1] D. Scharstein, R. Szeliski, and R. Zabih, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," in *Proc. IEEE Workshop SMBW*, 2001, pp. 131–140.

[2] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.

[3] P. Felzenszwalb and D. Huttenlocher, "Efficient belief propagation for early vision," in *Proc. IEEE Conf. CVPR*, 2004, pp. 261–268.

[4] X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang, and X. Zhang, "On building an accurate stereo matching system on graphics hardware," in *Proc. IEEE Workshop ICCV*, 2011, pp. 467–474.

[5] K. Zhang, J. Lu, and G. Lafruit, "Cross-based local stereo matching using orthogonal integral images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 7, pp. 1073–1079, Jul. 2009.

[6] C. Cigla and A. A. Alatan, "Information permeability for stereo matching," *Signal Processing: Image Communication*, vol. 28, no. 9, pp. 1072–1088, Oct. 2013.

[7] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," in *Proc. IEEE Conf. CVPR*, 2011, pp. 3017–3024.

[8] Y.-C. Wang, C.-P. Tung, and P.-C. Chung, "Efficient disparity estimation using hierarchical bilateral disparity structure based graph cut algorithm with a foreground boundary refinement mechanism," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 5, pp. 784–801, May 2013.

[9] Q. Yang, "A non-local cost aggregation method for stereo matching," in *Proc. IEEE Conf. CVPR*, 2012, pp. 1402–1409.

[10] Z. Ma, K. He, Y. Wei, J. Sun, and E. Wu, "Constant time weighted median filtering for stereo matching and beyond," in *Proc. IEEE Conf. ICCV*, 2013, pp. 49–56.

[11] H. Hirschmuller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in *Proc. IEEE Conf. CVPR*, 2005, pp. 807–814.

[12] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2008.

[13] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.

[14] P. Felzenszwalb and D. Huttenlocher, "Efficient graph-based image segmentation," *Intl J. Computer Vision*, vol. 59, no. 2, pp. 167–181, Sept. 2004.

[15] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[16] D. Scharstein and R. Szeliski, "Middlebury stereo vision page [online]," *Available: http://vision.middlebury.edu/stereo*.

[17] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[18] K. He, J. Sun, and X. Tang, "Guided image filtering," in *Proc. ECCV*, 2010, pp. 1–14.

[19] C. C. Pham and J. W. Jeon, "Domain transformation-based efficient cost aggregation for local stereo matching," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 7, pp. 1119–1130, Jul. 2013.

[20] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, and K. Toyama, "Digital photography with flash and no-flash image pairs," *ACM Trans. on Graphics*, vol. 23, no. 3, pp. 664–672, Aug. 2004.

[21] K. H. Lo, Y. C. F. Wang, and K. L. Hua, "Joint trilateral filtering for depth map super-resolution," in *Proc. IEEE Conf. VCIP*, 2013, pp. 1–6.