Efficient Dense Disparity Map Reconstruction using Sparse Measurements

Oussama Zeglazi[1], Mohammed Rziza[1], Aouatif Amine[2] and Cédric Demonceaux[3]

[1]*LRIT, RABAT IT CENTER, Faculty of Sciences, Mohammed V University, B.P. 1014, Rabat, Morocco*

[2]*LGS, National School of Applied Sciences, Ibn Tofail University, B.P. 241, University Campus, Kenitra, Morocco*

[3]*Le2i, FRE CNRS 2005, Arts et Métiers, Univ. Bourgogne Franche-Comté, France*

Keywords: Stereo Matching, Superpixel, Vertical Median Filter, Scanline Propagation.

Abstract: In this paper, we propose a new stereo matching algorithm able to reconstruct efficiently a dense disparity maps from few sparse disparity measurements. The algorithm is initialized by sampling the reference image using the Simple Linear Iterative Clustering (SLIC) superpixel method. Then, a sparse disparity map is generated only for the obtained boundary pixels. The reconstruction of the entire disparity map is obtained through the scanline propagation method. Outliers were effectively removed using an adaptive vertical median filter. Experimental results were conducted on the standard and the new Middlebury[a] datasets show that the proposed method produces high-quality dense disparity results.

1 INTRODUCTION

The stereo matching problem is one of the most important tasks of computer vision domain, as it is vital for various applications, such as remote sensing, 3D reconstruction, object detection and tracking, etc.

The aim of stereo matching algorithms is to estimate the depth of a scene by computing the disparity of objects between stereo pairs. There exist two classes of stereo matching algorithms: sparse and dense ones. Sparse matching algorithms are based upon key-point matching process. The resulting depth map is sparse due to the knowledge of locations with the lack of depth estimation (Schauwecker et al., 2012), (Hsieh et al., 1992). In contrast, dense algorithms estimate disparity values for every pixel of the input image. Dense algorithms can be roughly classified into global and local approaches, and can be performed in four main steps : matching cost computation, cost aggregation, disparity computation and disparity refinement (Scharstein and Szeliski, 2002). An common key step in both global and local stereo matching algorithms is the cost computation one. This latter uses different cost function based on pixel intensities as absolute intensity differences, squared intensity differences or normalized cross correlation, other ones are based on image transformations such as non-

parametric census (Zabih and Woodfill, 1994), and others examined in surveys (Hirschmuller and Scharstein, 2009), (Miron et al., 2014). Local approaches estimate each pixel′s disparity based on the aggregation of the matching costs over a local support region. The commonly used support regions are rectangular windows or their variations (Kang et al., 1995; Bobick and Intille, 1999), weighted support window (Yoon and Kweon, 2006) and adaptive support regions (Zhang et al., 2009). Global methods define an optimized energy function over all image pixels with some constraints. This energy function is minimized using various algorithms such as dynamic programming (Veksler, 2005), belief propagation (Sun et al., 2003) or graph-cuts (Kolmogorov and Zabih, 2001).

The main contribution of this work concerns dense disparity reconstruction. We propose a new method which generates dense disparity maps based on set of reliable seed points. These points are then propagated to neighboring pixels in a growing-like manner (Sun et al., 2011), (Hawe et al., 2011), (Liu et al., 2015), (Mukherjee and Guddeti, 2014).

In this paper, we present a sampling and reconstruction process to generate dense disparity maps from reliable sparse seed points. First, the proposed algorithm is initialized by sampling the reference image using SLIC superpixel method (Achanta et al., 2012). Then, a sparse disparity map is generated for only the boundary pixels obtained from the segmenta-

---

[a]http://vision.middlebury.edu/stereo/data/

tion process. Second, the propagation process is performed within superpixel graph along the scan-lines based on the color similarity. Finally, an adaptive vertical median filter is performed to tackle horizontal streaking artifacts.

The remainder of this paper is organized as follows. Section 2 describes the proposed method. In section 3, we report experimental results and the conclusive remarks are made in Section 4.

2 PROPOSED METHOD

Our algorithm calculates the depth map following the flowchart in Figure 1. Details with respect to each step are addressed in the following sub-sections.

2.1 Data Sampling

Superpixel methods have been widely used for stereo matching algorithms to provide smoothness prior while strengthening all the other pixels to belong to the same 3D surface (Yamaguchi et al., 2012), (Yamaguchi et al., 2014), (Kim et al., 2015). Compared to these methods, our approach considers the boundary pixels of each superpixel region as seed points. In particular, the reference image is beforehand abstracted as set of nearly regular superpixels using the SLIC method. The obtained superpixels are therefore composed of pixels that have similar attributes, which preserves image edges. Therefore, we focus on boundary pixels that lie on the borders segment produced by the segmentation process. The sparse disparity map can be obtained using any stereo matching algorithm. However, an early erroneous disparity values lead to large disparity errors during the propagation process. Since the study of the reliability of initial disparity measurement is out the scope of this paper, we generate the initial disparity measurements using ideal disparity measurements extracted from the ground truth disparity maps. To more evaluate our method, we have also used other initial disparity measurements as the semi-global matching (SGM) method (Hirschmuller, 2008).

2.2 Propagation Process

The pixels within the same superpixel region are assumed to belong to the studied $3D$ object, as long as the scale of that superpixel is small enough (Yan et al., 2015). However, this assumption can be violated, especially, in region near depth discontinuities. Figure 2 presents a typical case from Tsukuba stereo

pair, where pixels within same superpixel (colored regions) have different depth information, although the scale of superpixel is very small. Consequently, assigning the same disparity value for all pixels in one superpixel is not effective. In order to deal with this issue, the disparity value was captured from boundary pixel of each superpixel region, then we scanned the disparity map in row-wise manner, whenever we found an unseeded pixel p, we sought for the closest seed pixels which the disparity value are already known, and which lie on the same scanline superpixel (pixel row). Then, if a unique seed pixel is found, the disparity of this latter is assigned to the unseeded pixel p. Otherwise, if it finds left and right seed pixels (p_{s_1}, p_{s_2}) with known disparity values d_1 and d_2, respectively. We assign to the unseeded pixel the disparity d using the following rule :

$$d = \begin{cases} d_1 & D_c(p_{s_1}, p) < D_c(p_{s_2}, p) \\ d_2 & otherwise \end{cases} \quad (1)$$

Where $D_c(p_{s_i}, p) = max_{i=R,G,B}|I_i(p_{s_i}) - I_i(p)|$ represents the color difference between $p_{s_i}, i \in \{1, 2\}$ and p. The disparity maps obtained after the propagation process are presented in figure 3, which shows the efficiency of the proposed method, since it produces small errors. In order to remove the remainder errors, we used the vertical median filter, which will be detailed in the next section.

2.3 Adaptive Vertical Median Filter

Since the above propagation process is performed in row-wise manner, horizontal streak-like artifacts can occur in the disparity map. Thus, including disparity information from the vertical direction can effectively reduce them. For this purpose, we incorporated the vertical disparity information by assigning to the pixel under consideration the vertical median value. Streaking artifacts can be located mainly at object boundaries. Therefore, to limit at constant vertical filter size, this enable to construct a support filter with outliers from other image structure, and then leads to erroneous disparity results. To address this issue, we opted for the construction of an adaptive vertical support filter, which only contains pixels of the same image structure. In this context, our attention was paid to the cross-based median filter described in (Stentoumis et al., 2014), which was used as a post processing method to tackle with outliers produced through the left-right consistency check method. We have also followed the color assumption in order to build vertical support filters. The vertical line of each pixel p , is defined using its both directional arms (up or bottom) using the following rules:
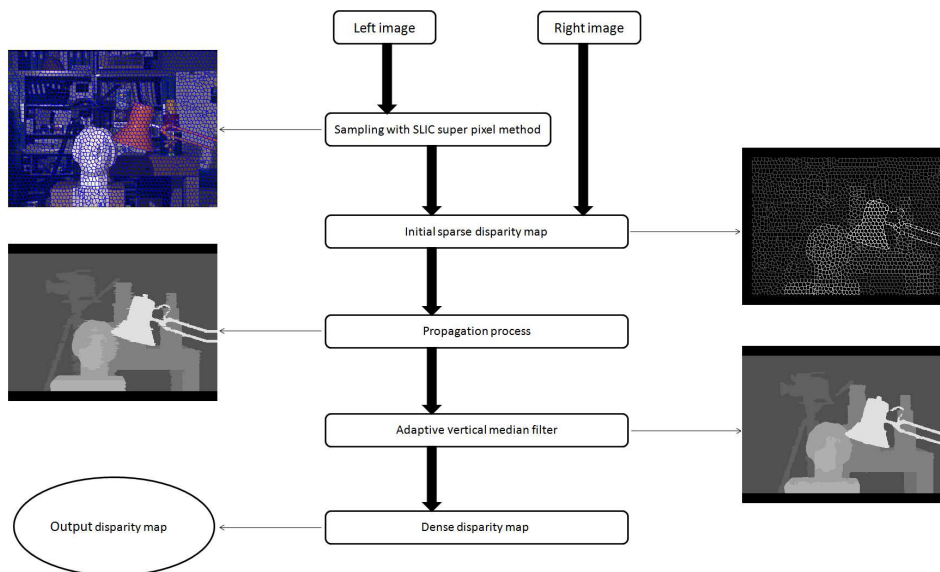
Figure 1: Flowchart of the proposed stereo matching algorithm using the "Tsukuba" stereo pair input.
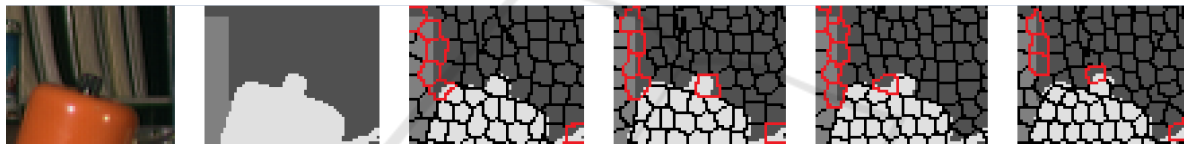

Figure 2: Example of close-up superpixel segmentation on Tsukuba stereo pair, from left to right: Reference image, ground truth disparity map, superpixel graph on disparity map with 2000, 2200, 2400 and 2600 number of superpixels, respectively.

(1) $D_c(p_l, p) < \tau$, where p_l is a pixel lying on the vertical arm of p. $D_c(p_l, p) = max_{i=R,G,B}|I_i(p_l) - I_i(p)|$, which denotes the color difference between p_l and p, and τ is the preset threshold value.

2) $D_s(p_l, p) < L$, where $D_s(p_l, p) = |p_l - p|$, which represents the spatial distance between p_l and p, and L is the preset maximum length.

The first rule guarantees the color similarity assumption while the second one poses a limitation on the vertical line length to avoid any over-smoothed disparity results.

3 EXPERIMENTAL RESULTS

In this section, we perform an evaluation to assess the performance of the proposed stereo algorithm. Experiments were carried out on the standard and the new Middlebury datasets. The sparse disparity measurements were extracted from the ground truth and also from other disparities such as SGM disparity results (Hirschmuller, 2008). In the construction of the adaptive vertical median, the spatial and the color similarity thresholds were experimentally fixed at $L = 7$ and $\tau = 8$, respectively. Since, in our work, the number of

superpixels define the data sampling ratios. We first represent the corresponding data sampling ratios for the standard Middlebury datasets in various number of superpixels in table 1.

3.1 Reconstruction Results using Ground Truth Sparse Disparity Maps

We discuss the reconstruction accuracy with respect to different superpixel number using the ground truth

Table 1: Sampling Ratios for each superpixel number for Tsukuba, Venus, Teddy and Cones stereo pairs.

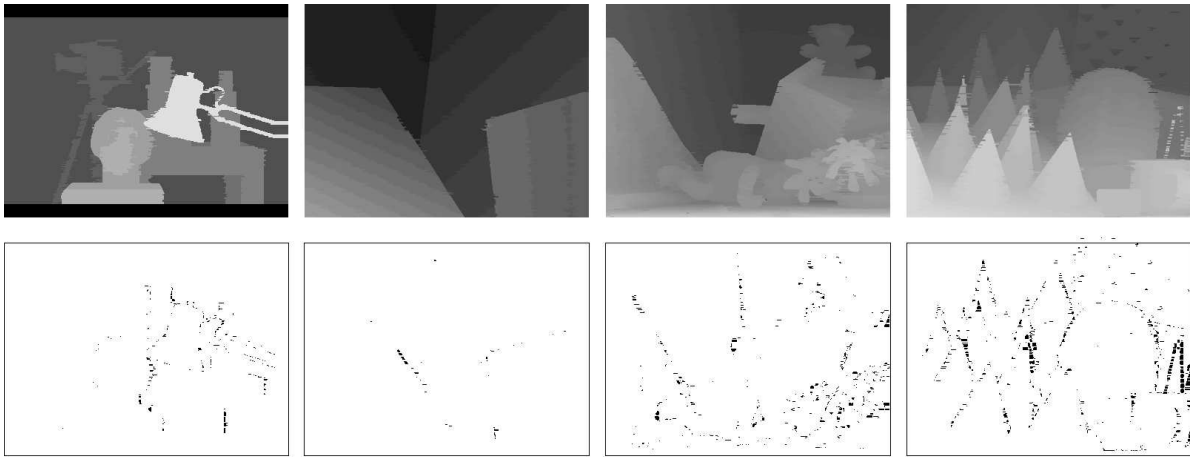| Number of superpixels	Sampling ratios (%)			
	Tsukuba	Venus	Teddy	Cones
600	15.44	12.98	12.73	12.75
800	17.87	15.06	14.81	15.08
1000	19.67	16.53	16.38	16.57
1200	21.51	18.33	17.95	18.03
1400	22.93	19.86	19.13	19.21
1600	24.50	20.86	20.55	20.71
1800	25.80	21.82	21.71	21.82
2000	27.22	23.09	22.90	23.02
2200	28.39	23.92	23.67	23.84
2400	29.06	25.00	24.74	24.97
2600	30.37	25.60	25.88	26.01

Figure 3: Propagation process results for the four Middlebury stereo pairs (Tsukuba, Venus, Teddy, Cones). From the top to the bottom: The produced disparity maps, disparity errors (marked in black) in all regions with 1 pixel error threshold.

(a) Tsukuba

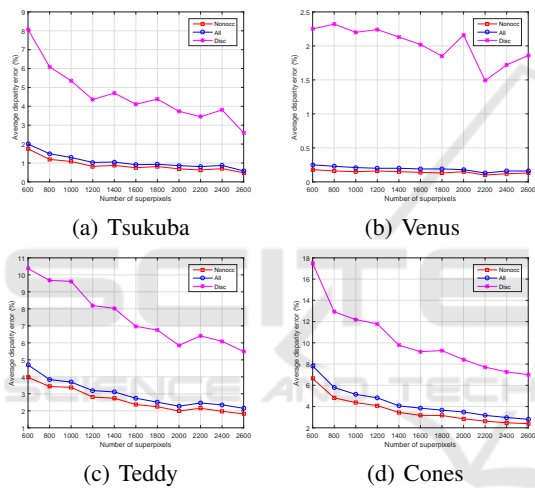(b) Venus

(c) Teddy

(d) Cones

Figure 4: The disparity results obtained with respect to superpixels number for Tsukuba, Venus, Teddy and Cones stereo pairs.

disparities. Figure 4 shows the obtained disparity errors with different superpixel numbers for the four stereo pairs. The errors are given in non-occluded regions, all and near depth discontinuities, and computed at 1 default pixel error threshold. It can be noted that the performance of the proposed approach increases with the number of superpixels. This can be explained by the fact that a small number of superpixels (i.e. large superpixel scale) can hold several objects or a small part of them. Then, in the propagation process, same disparity value may be assigned to pixels from different 3*D* objects.

we have also evaluated our algorithm over some of state of the art disparity reconstruction algorithms. Indeed, we have compared our method with the recently introduced method (Liu et al., 2015) using both its both variants Alternating Direction Method of Mul-

tipliers (ADMM) on Wavelet (WT) and Contourlet (CT), ADMM WT+CT Grid and ADMM WT+CT2-Stage, where Grid and 2-stage refer respectively to data used for the sampling and the method (Liu et al., 2015). To fairly carry out the experiments, we used the same data samples for both methods (Liu et al., 2015) and ours. Thus, we used 2600 superpixels which represent data sampling ratio of 30%, 25%, 25% and 26% for Tsukuba, Venus, Teddy and Cones stereo pairs, respectively. Since for the method (Mukherjee and Guddeti, 2014), the sampling is performed using the k-means method, if defined, we keep the same value of the k parameter (Mukherjee and Guddeti, 2014). Otherwise, we set $K = 12$, as its gives experimentally the lower disparity errors.

Table 2 presents the disparity error in non occluded (nonocc), all and near depth discontinuities. The errors were computed at the default 1 pixel error threshold for the Middlebury database. The obtained results demonstrate the efficiency of the proposed method. Indeed, our method achieves depth reconstruction with the lower error rate for the Teddy and Cones stereo pairs in all different regions. In the case of the Venus stereo pair, the "ADMM WT+CT2-Stage" (Liu et al., 2015) method and our approach give almost exactly the same error rates.
Overall, satisfactory disparity results were achieved for all the stereo pairs. Moreover, in terms of execution time, the proposed method performs better in term of execution time. Indeed, the computational time for the four stereo pairs (Tsukuba, Venus, Teddy and Cones) are 13.1 seconds, 13.2 seconds, 11.6 seconds and 13.4 seconds, respectively.

The intermediate and final disparity results for the four stereo pairs from the standard Middlebury dataset are presented in Figure 5.

Table 2: The error percentages in different regions (nonocc, all , disc) with 1 pixel threshold.

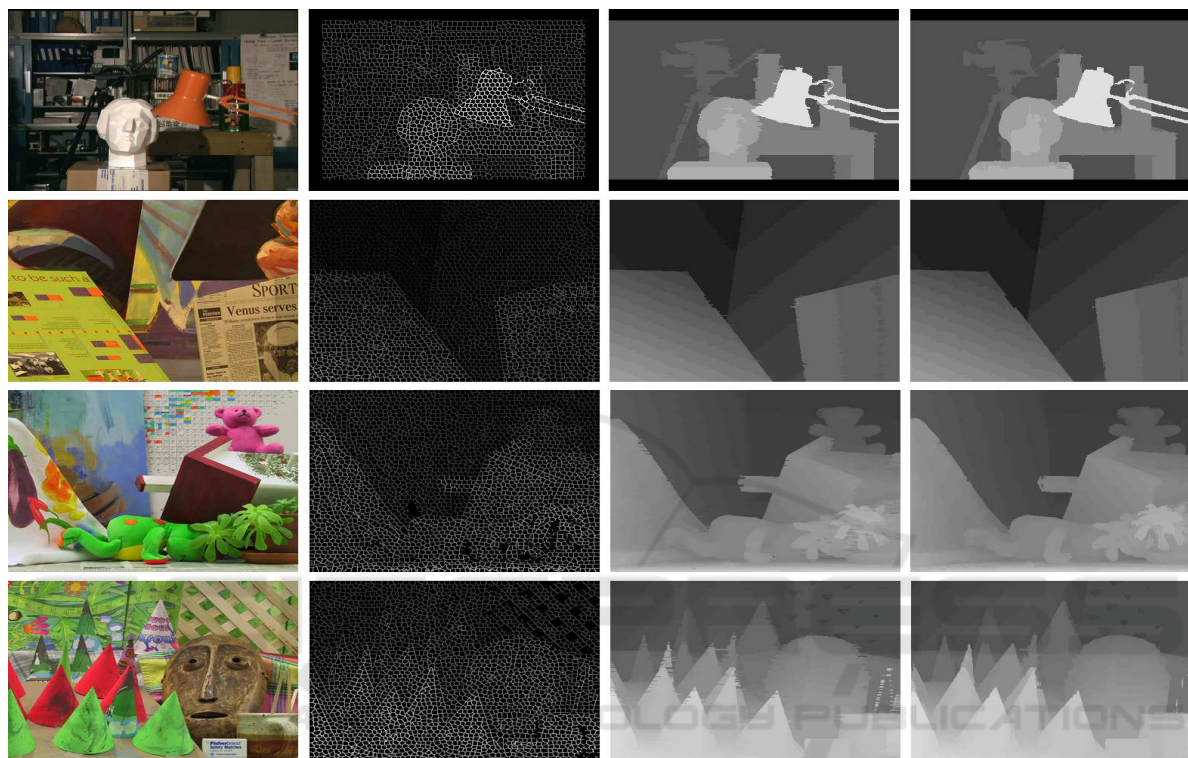| Algorithms	Tsukuba			Venus			Teddy			Cones		
	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc
Our Method	**0.48**	**0.58**	2.60	0.13	0.16	1.86	**1.82**	**2.15**	**5.49**	**2.39**	**2.80**	**7.00**
ADMM WT+CT2-Stage (Liu et al., 2015)	0.51	0.62	**2.01**	**0.08**	**0.11**	**1.19**	2.06	2.56	7.37	2.68	3.16	8.09
ADMM WT+CT Grid (Liu et al., 2015)	1.14	1.43	5.19	0.26	0.39	3.73	2.57	3.36	8.98	3.39	4.05	10.21
Method (Mukherjee and Guddeti, 2014)	2.57	2.77	10.82	2.92	3.00	20.05	11.36	12.17	18.39	13.20	14.15	21.03

Figure 5: Results on the standard Middlebury data sets. From left to right and from left to right: Reference image, Initial disparity maps, propagation results and final disparity maps.

Table 3: Percentage of erroneous disparities in non-occluded regions for the Middlebury training set.

Algorithms	0.75 px threshold nonocc	1 px threshold nonocc	2 px threshold nonocc
Our method	**18.10**	**3.00**	**2.05**
ADMM WT+CT2-Stage (Liu et al., 2015)	29.11	5.92	3.20
Method (Mukherjee and Guddeti, 2014)	23.73	10.94	7.92

We carried out one further experiment on one of the widely used dataset for stereo matching , the 2014 Middlebury datastes. This benchmark is divided into two parts, training and testing sets which contain 15 stereo pairs each. The training set is available in three resolutions, for which the ground truth disparity maps are provided. The testing set uses an evaluation platform to recorded the results. In our experiments, we used 15 images in quarter resolution from training set. We also used the same number of superpixels, which is set to 2600 as the first experiment. Figure 6 gives the corresponding percentage of samples for each pair.
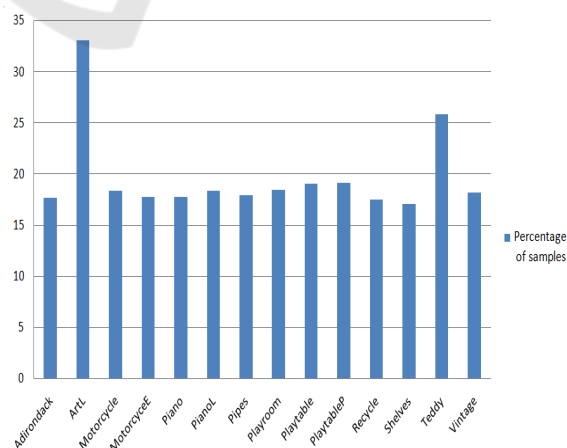
Table 3 presents the mean errors in non-occluded

Figure 6: Percentage of samples for each pair from the new Middlebury training set using sampling with 2600 super-pixels.

Figure 7: Results on the new Middlebury training set for (a) ArtL. (b) Piano. (c) Playtable. (d) Shelves. From left to right: Reference image, disparity maps produced by (Mukherjee and Guddeti, 2014), (Liu et al., 2015) and ours, respectively. Finally ground truth disparity maps.

Table 4: Percentage of erroneous disparities in non-occluded regions for the Middlebury 2014 training set.

Algorithms	1 px threshold nonocc	2 px threshold nonocc
SGM (Liu et al., 2015)	12.47	9.44
Our method	13.40	9.77

regions computed at 3 different pixels error thresholds.

The obtained results demonstrate that the proposed method gives the lowest disparity error rate in all pixel error threshold. Finally, figure 7 shows examples of disparity results from the new Middlebury 2014 datasets for Method (Mukherjee and Guddeti, 2014), ADMM WT+CT2-Stage (Liu et al., 2015) and ours, respectively.

3.2 Reconstruction Results using other Initial Disparity Maps

To more evaluate the efficiency of the proposed method, we studied the performance of reconstruction using the SGM algorithm (Hirschmuller, 2008). We opted for the SGM method as it is the best studied approach in-between local and global matching. We kept the same configuration set used previously for the Middlebury 2014 datasets, and we changed the initial disparity seeds by the ones derived from the SGM algorithm using the same sampling ratios in figure 6.

Table 4 presents the disparity results obtained using the SGM algorithm and the reconstruction ones. according to results reported in table 4, we note that the proposed approach enables to reconstruct images based on the SGM method without losing information. For instance, using 2 pixels error threshold, the mean error changed slightly from 9.44 to only 9.77. This suggest the efficiency of the proposed method when dealing with other initial disparity maps rather than the ground truth.

4 CONCLUSION

In this paper, we presented a new stereo matching algorithm based on SLIC superpixel sampling. The estimation of dense disparity maps is based on superpixel boundaries disparities. The reconstruction of the dense disparity map from the boundaries disparities was performed using the scanline propagation technique. Streaking artifacts were effectively addres-

sed using an adaptive vertical median filter. Experimental results conducted on the Middlebury datasets have demonstrated the accuracy and the efficiency of the proposed method.

REFERENCES

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Ssstrunk, S. (2012). Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282.

Bobick, A. F. and Intille, S. S. (1999). Large occlusion stereo. *International Journal of Computer Vision*, 33(3):181–200.

Hawe, S., Kleinsteuber, M., and Diepold, K. (2011). Dense disparity maps from sparse disparity measurements. In *2011 International Conference on Computer Vision*, pages 2126–2133.

Hirschmuller, H. (2008). Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):328–341.

Hirschmuller, H. and Scharstein, D. (2009). Evaluation of stereo matching costs on images with radiometric differences. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(9):1582–1599.

Hsieh, Y. C., McKeown, D. M., and Perlant, F. P. (1992). Performance evaluation of scene registration and stereo matching for cartographic feature extraction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):214–238.

Kang, S. B., Webb, J. A., Zitnick, C. L., and Kanade, T. (1995). A multibaseline stereo system with active illumination and real-time image acquisition. In *Proceedings of IEEE International Conference on Computer Vision*, pages 88–93.

Kim, S., Ham, B., Ryu, S., Kim, S. J., and Sohn, K. (2015). *Robust Stereo Matching Using Probabilistic Laplacian Surface Propagation*, pages 368–383. Springer International Publishing, Cham.

Kolmogorov, V. and Zabih, R. (2001). Computing visual correspondence with occlusions using graph cuts. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 2, pages 508–515 vol.2.

Liu, L. K., Chan, S. H., and Nguyen, T. Q. (2015). Depth reconstruction from sparse samples: Representation, algorithm, and sampling. *IEEE Transactions on Image Processing*, 24(6):1983–1996.

Miron, A., Ainouz, S., Rogozan, A., and Bensrhair, A. (2014). A robust cost function for stereo matching of road scenes. *Pattern Recognition Letters*, 38:70–77.

Mukherjee, S. and Guddeti, R. M. R. (2014). A hybrid algorithm for disparity calculation from sparse disparity estimates based on stereo vision. In *2014 International Conference on Signal Processing and Communications (SPCOM)*, pages 1–6.

Scharstein, D. and Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47:7–42.

Schauwecker, K., Klette, R., and Zell, A. (2012). A new feature detector and stereo matching method for accurate high-performance sparse stereo matching. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5171–5176.

Stentoumis, C., Grammatikopoulos, L., Kalisperakis, I., and Karras, G. (2014). On accurate dense stereo-matching using a local adaptive multi-cost approach. {*ISPRS*} *Journal of Photogrammetry and Remote Sensing*, 91:29 – 49.

Sun, J., Zheng, N.-N., and Shum, H.-Y. (2003). Stereo matching using belief propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(7):787–800.

Sun, X., Mei, X., Jiao, S., Zhou, M., and Wang, H. (2011). Stereo matching with reliable disparity propagation. In *2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, pages 132–139.

Veksler, O. (2005). Stereo correspondence by dynamic programming on a tree. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 384–390 vol. 2.

Yamaguchi, K., Hazan, T., McAllester, D., and Urtasun, R. (2012). Continuous markov random fields for robust stereo estimation. In *Proceedings of the 12th European Conference on Computer Vision - Volume Part V*, ECCV'12, pages 45–58, Berlin, Heidelberg. Springer-Verlag.

Yamaguchi, K., McAllester, D., and Urtasun, R. (2014). *Efficient Joint Segmentation, Occlusion Labeling, Stereo and Flow Estimation*, pages 756–771. Springer International Publishing, Cham.

Yan, J., Yu, Y., Zhu, X., Lei, Z., and Li, S. Z. (2015). Object detection by labeling superpixels. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5107–5116.

Yoon, K.-J. and Kweon, I. S. (2006). Adaptive support-weight approach for correspondence search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4):650–656.

Zabih, R. and Woodfill, J. (1994). Non-parametric local transforms for computing visual correspondence. In *Proceedings of the Third European Conference on Computer Vision (Vol. II)*, ECCV '94, pages 151–158, Secaucus, NJ, USA. Springer-Verlag New York, Inc.

Zhang, K., Lu, J., and Lafruit, G. (2009). Cross-based local stereo matching using orthogonal integral images. *IEEE Trans. Circuits Syst. Video Techn.*, 19(7):1073–1079.