

# Disparity Refinement Processor Architecture utilizing Horizontal and Vertical Characteristics for Stereo Vision Systems

Cheol-Ho Choi\*   
*Pangyo R&D Center  
Hanwha Systems Co., Ltd.  
Seongnam, Republic of Korea  
cheoro1994@hanwha.com*

Hyun Woo Oh   
*Pangyo R&D Center  
Hanwha Systems Co., Ltd.  
Seongnam, Republic of Korea  
hyunwoo.oh@hanwha.com*

**Abstract**—In embedded stereo vision systems based on semi-global matching, the matching accuracy of the initial disparity map can be degraded because of various factors. To solve this problem, weighted median-based disparity refinement hardware architectures are utilized to improve the matching accuracy. However, for the conventional hardware architectures, there is a trade-off between hardware resource utilization and refinement performance when they are implemented on a field programmable gate array (FPGA). Therefore, in this paper, we propose a hybrid max-median filter and its hardware architecture to improve the refinement performance and reduce hardware resource utilization. To evaluate the refinement performance, we used two public stereo datasets. When using the various window sizes for KITTI 2012 and 2015 stereo benchmark datasets, the proposed hardware architecture showed better matching accuracy performance compared with the conventional hardware architectures. In terms of the hardware resource utilization, when implemented on an FPGA, the proposed hardware architecture has low requirements for all types of hardware resources. That is, the proposed hardware architecture overcomes the trade-off between hardware resource utilization and refinement performance.

**Index Terms**—Stereo vision, semi-global matching, disparity refinement, hardware architecture

## I. INTRODUCTION

In embedded stereo vision systems, semi-global matching (SGM) is widely used because of its reasonable matching accuracy with reasonable hardware resource utilization [1]–[3]. In addition, the SGM can be operated in real-time because it can be designed with pipeline architecture with systolic array [4]–[6]. However, the matching accuracy of the initial disparity map in SGM can be degraded on texture-less and occluded regions [7]–[9]. Hence, various post-processing methods are used to improve the matching accuracy.

In the post-processing methods, the various methods are widely used (i.e., uniqueness function, left-right consistency check, and filtering) [10]–[13]. When using uniqueness function and left-right consistency check methods, there is an advantage in that inaccurate disparity values can be removed. However, the visual quality of disparity map is degraded because the inaccurate pixel values, called hole pixels, are

removed. Therefore, to improve the quality of disparity map, various filtering methods are used for hole-filling process.

Among the various filtering methods, the bilateral-based weighted median filter (WMF) is widely used because it provides high hole-filling performance, called refinement performance, for improving the matching accuracy [8], [9]. However, it has the drawback of large hardware resource utilization when implemented on a field programmable gate array (FPGA) [14]. Therefore, the various follow-up studies were conducted to overcome this drawback of the WMF [15], [16]. To reduce the hardware resource utilization, Chen et al. proposed the separable WMF (sWMF) [15]. The sWMF introduces a separable operation for each horizontal and vertical direction to reduce the computational complexity. However, it still has high hardware resource utilization when implemented on an FPGA. Further, Hyun et al. proposed a sparse window approach-based sWMF (ssWMF) to further reduce the hardware resource utilization [16]. Although the ssWMF reduces hardware resource utilization, its refinement performance also reduced. Thus, for the conventional hardware architectures, there exists trade-off between hardware resource utilization and refinement performance.

Therefore, in this paper, we propose a hybrid max-median filter hardware architecture to overcome the trade-off when implemented on an FPGA. The proposed hardware architecture utilizes the road environment-based disparity tendency for each horizontal and vertical direction, thereby allowing the proposed hardware architecture to achieve high refinement performance. In addition, the hardware resource utilization of the proposed hardware architecture can be reduced because it can be designed with a simple computation architecture comprising only max and median filters.

The rest of this paper is organized as follows. Section 2 briefly describes the conventional hardware architectures. Section 3 and describes the proposed max–median filter, and Section 4 presents its hardware architecture. Section 5 describes the experimental results for comparing the refinement performance and hardware resource utilization. To compare the refinement performance, we used the two public stereo

datasets. In addition, we compared the hardware resource utilization for various window sizes for the proposed and conventional hardware architectures. Finally, Section 6 provides the conclusions of this study.

## II. RELATED WORK

Weighted median-based filtering methods have the advantage of preserving edge information [17]. Hence, the WMF method is widely utilized in stereo vision systems. However, when implemented on an FPGA, the WMF requires excessive hardware resources because it involves extensive computation [14]. In addition, in terms of the software environment, the processing time increases because of the bottleneck created via the weighted computation and sorting process. Many studies were conducted to overcome this drawback by reducing the computational complexity to subsequently reduce the hardware resource and processing time requirements.

**Separable weighted median filter (sWMF):** The WMF method has a computational complexity of  $O(r^2)$ , where  $r$  is the radius of square filter window. To reduce the computational complexity, Chen et al. proposed the sWMF method [15]. The sWMF method has a separable computation concept that has a one-dimensional (1D) horizontal WMF (HWMF) and 1D vertical WMF (VWMF). Using the separable computation concept, the sWMF method achieves a reduced computational complexity of  $O(r)$  from  $O(r^2)$ . Hence, when implemented on an FPGA, the sWMF requires less hardware utilization than the WMF. In addition, the processing speed can be improved software and FPGA platforms.

**Sparse-window-based sWMF (ssWMF):** Although the sWMF method can reduce the hardware resource requirement compared with the WMF method, it still requires a large amount of hardware resources. Hence, Hyun et al. proposed the ssWMF method [16]. This method requires fewer pixels than the sWMF method, and hence, it requires less hardware resources when implemented on an FPGA. Therefore, a low-cost embedded stereo vision system can be designed when adopting the ssWMF. However, this method has the drawback that because only few pixels are used, the refinement performance can be degraded.

When implemented on an FPGA, the sWMF and ssWMF methods can reduce the hardware resource consumption by reducing the computational complexity. Thus, these methods are advantageous for embedded stereo vision systems, which requires low-cost characteristics. However, the experimental results of previous studies showed that the refined matching accuracy of sWMF and ssWMF methods is degraded compared to that of the WMF method [15], [16]. In other words, for the conventional methods, there exists a trade-off between hardware resource utilization and disparity refinement performance.

## III. PROPOSED METHOD

The proposed hybrid max-median filter aims to compute the refined disparity map in which incorrect matching values are corrected. To achieve this goal, the proposed method utilizes

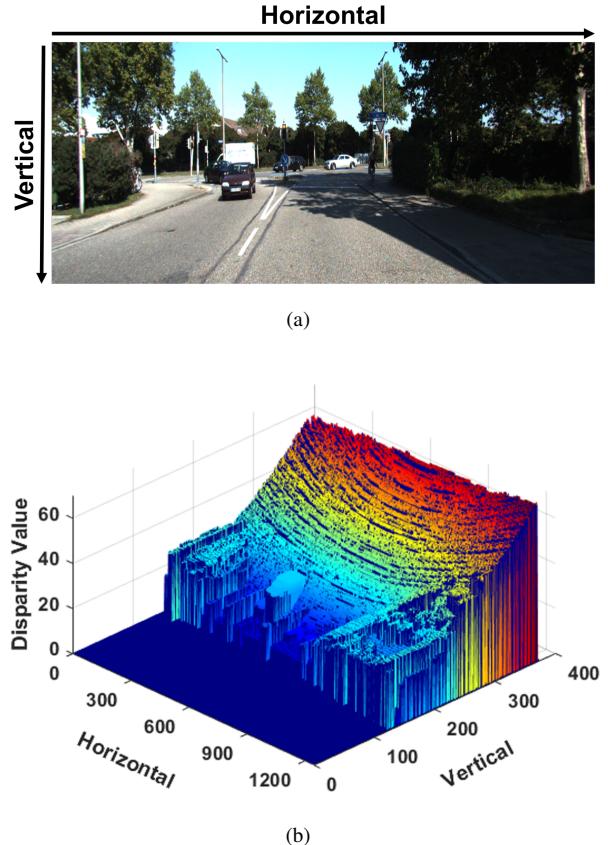


Fig. 1. Left-side stereo images and three-dimensional disparity plot: (a) left-side stereo images and (b) 3D plot for disparity values corresponding to the coordinates of the left-side stereo images

the characteristic of the disparity map based on the road environment. In the road environment-based disparity map, the disparity tendency for horizontal has no constant tendency because of various objects (e.g., pedestrians and vehicles) as shown in Fig. 1. Therefore, in the horizontal direction, there is a need to preserve the edge information for objects. In contrast, the depth value for vertical direction gradually increases from the bottom coordinate to the top coordinate [18]. In other words, the disparity value of the vertical direction increases from the top coordinate to the bottom coordinate. Therefore, there is a tendency that the disparity value gradually increases in the vertical direction from the top coordinate to the bottom coordinate, as shown in Fig. 1.

The operation process of the proposed method is illustrated in Fig. 2. To utilize these characteristics for horizontal and vertical directions, our proposed method involves three steps: 1) sub-window generation, 2) inner-sub-window generation, and 3) max-median filtering computation. In the sub-window generation step, the  $N \times N$  sub-windows are generated. In the inner-sub-window generation step, the inner-sub-windows for the eight-path direction are generated, as illustrated in Fig. 3. Four inner-sub-windows are generated for each horizontal and vertical direction based on the center pixel of the sub-window. The purpose of generating two horizontal inner-

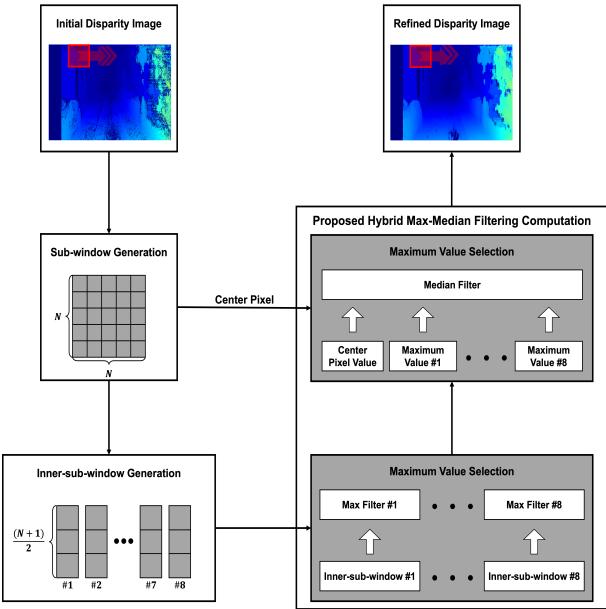


Fig. 2. Operation process of the proposed hybrid max-median filter for disparity refinement.

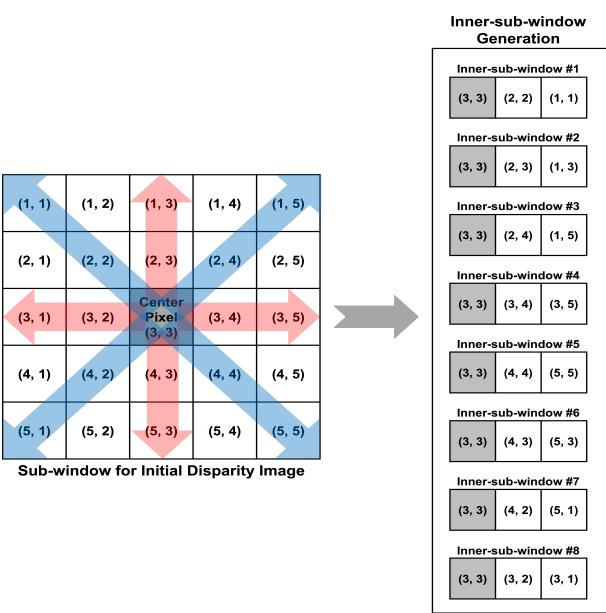


Fig. 3. Inner-sub-window generation method.

sub-windows is to extract the edge information for left and right directions based on the center pixel. The purpose of generating two vertical inner-sub-windows is to reflect the tendency of increasing disparity value from the top coordinate to the bottom coordinate based on the center pixel. The other four inner-sub-windows are generated for each diagonal direction based on the center pixel of the sub-window. In the occlusion regions, the disparity information of the target is inaccurate because the width and height of the objects may no correspond to the difference of the disparity leftwards and rightwards [19]. Therefore, four inner-sub-windows in the

diagonal direction are required to accurately perform hole-filling process in the occluded regions. In other words, hole-filling process for non-occlusion and occlusion regions can be performed more accurately by utilizing vertical, horizontal, and diagonal disparity information.

After the inner-sub-window generation step, each maximum value is selected from each inner-sub-window. The max filter is used for each inner-sub-window because the proposed method utilizes the vertical disparity tendency with the characteristic of gradually increasing disparity value from the top coordinate to the bottom coordinate. In other words, the result value of the max filter can reflect the vertical disparity tendency. After using the max filter, nine values, including eight maximum values and center pixel, are entered into the median filter. This filter selects the median value as the output value of the refined disparity map. The median filter has the advantage of preserving edge information. Hence, in the proposed method, a median filter is used after the max filter operation so that the edge information can be preserved when considering the horizontal disparity tendency.

#### IV. PROPOSED HARDWARE ARCHITECTURE

Fig. 4 illustrates the hardware architecture of the proposed max-median filter. The proposed hardware architecture consists of three modules: 1) inner-sub-window generator, 2) maximum value selector, and 3) median value selector. The inner-sub-window generator module performs inner-sub-window generation process, as shown in Fig. 3. To generate the  $N \times N$  sub-window, window generator module has  $N$  line buffers using block random access memory (BRAM). After generating the  $N \times N$  sub-window, the pixel values for each inner-sub-window are selected by using the pixel selector module. To select the corresponding pixel values for each inner-sub-window, the pixel selector module has a line counter, reorder, and register selector modules to select the appropriate pixel values for each line buffer. In terms of the line counter module, it calculates an address value by counting the entered line valid signal from the camera sensor. In terms of the reorder module, it contains demultiplexer to arrange and select the pixel values based on the address value from line value counter module. Thereafter, the register selector module selects and exports pixel values corresponding to each coordinate based on the center pixel using the pre-calculated register address values using the reorder module. After using pixel selector module, the selected pixel values are entered into the maximum value selector module. This module consists of eight max filters. Each max filter utilizes a pyramidal comparison step architecture based on the comparator. Using each max filter, eight maximum values and a center pixel are selected. Thereafter, the selected nine pixel values are entered into the median value selector module.

In the median value selector module, a separable concept similar to the sWMF method is employed to reduce the latency and hardware resource utilization. By adopting the separable operation for the median filter process, the hardware resource utilization can be reduced compared with the low-latency

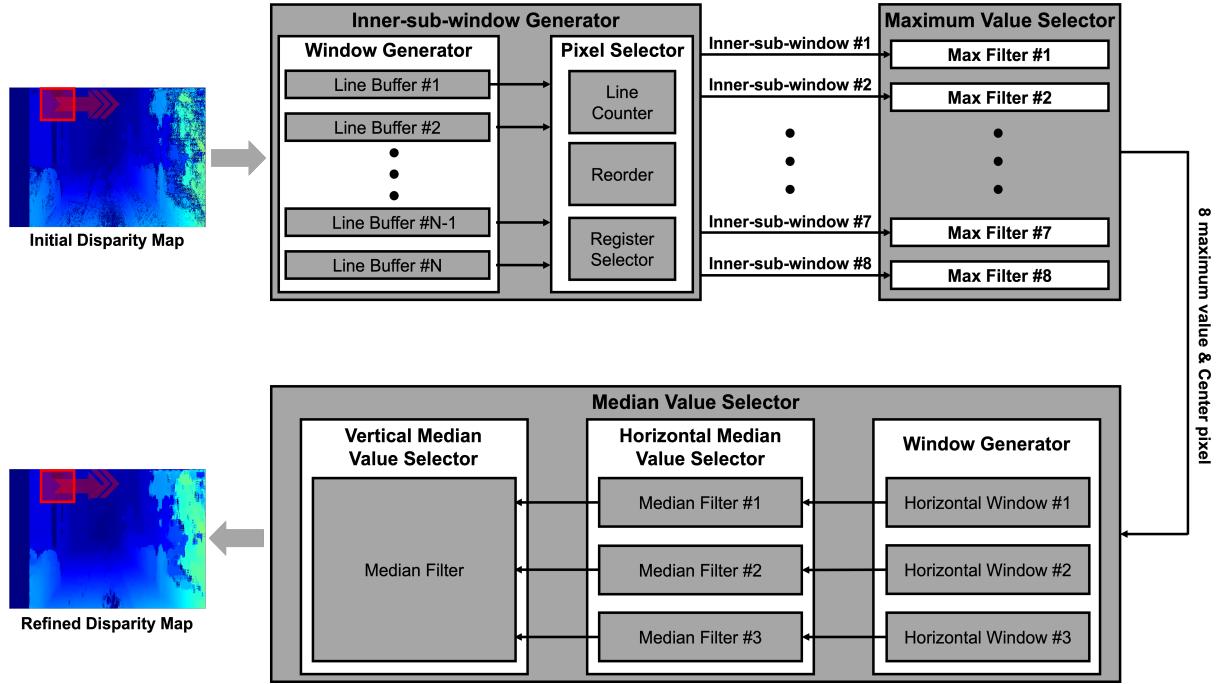


Fig. 4. Hardware architecture of the proposed hybrid max-median filter.

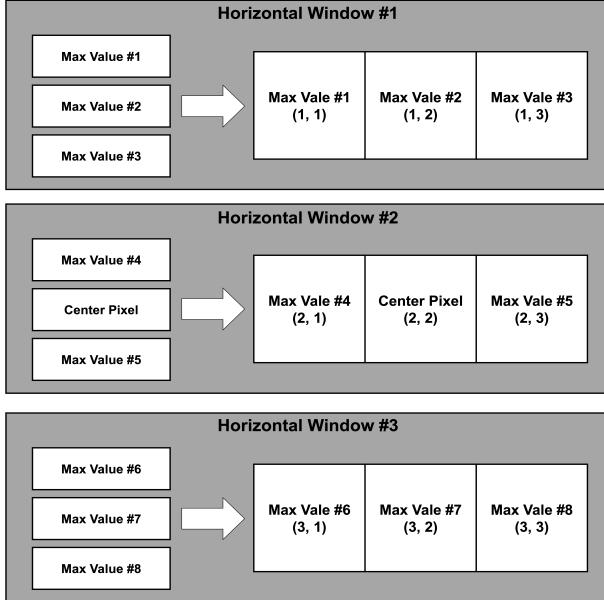


Fig. 5. Horizontal window generation method.

median filter architecture [20]. To design the median value selector module adopting separable operation concept, the horizontal window generator, horizontal median value selector, and vertical median value selector modules are employed. In the horizontal generator module, a  $3 \times 3$  window is generated using the input data, including eight maximum values and center pixel, as shown in Fig. 5. After generating this window, three horizontal median values are selected in the horizontal

median value selector module. Finally, in the vertical median value selector module, the vertical median value among the three horizontal median values is selected as the output value of the refined disparity map.

## V. EXPERIMENTAL RESULTS

To evaluate the refinement performance of the proposed hardware architecture, two types of public stereo datasets are used to compare the mean error rate (MER) index. In the MER performance evaluation, a smaller value in the MER index indicates higher matching accuracy. For fair performance comparison, experiments were conducted for sub-window sizes from  $5 \times 5$  to  $21 \times 21$ . For performance evaluation on the same initial disparity map, the disparitySGM built-in function of MATLAB R2022b tool was used. To perform the evaluation under the same experimental conditions as previous studies, we set the DisparityRange and UniquenessThreshold parameters in the disparitySGM built-in function to  $[0, 128]$  and 5, respectively [16]. In addition, we used public evaluation code provided by the KITTI benchmark to compute the MER index performance.

### A. KITTI Stereo Benchmark

Table I lists the MER performance results under non-occlusion and occlusion conditions when using the KITTI 2012 and 2015 stereo benchmark datasets [21], [22]. When using KITTI 2012 stereo benchmark dataset, the WMF showed the best MER performance for the  $9 \times 9$  window size. In the case of the  $9 \times 9$  window size, the MER performance of the WMF was 17.77% and 19.66% under the non-occlusion and occlusion conditions, respectively. In terms of the performance

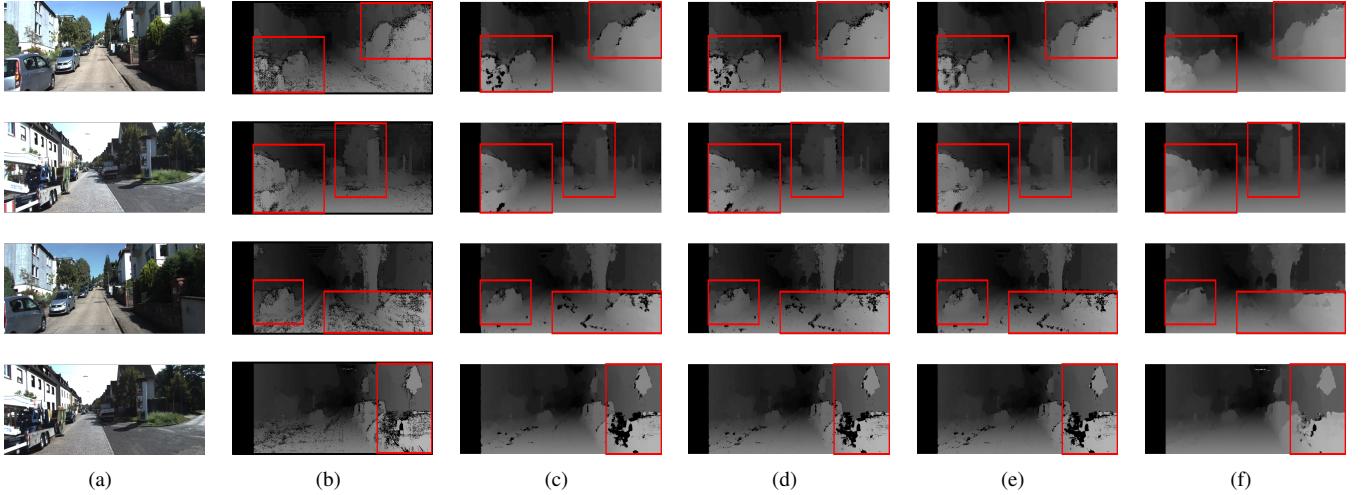


Fig. 6. Experimental results using KITTI 2012 and 2015 stereo benchmark datasets: (a) left-side input image, (b) initial disparity map using semi-global matching, (c) weighted median filter (WMF), (d) separable WMF (sWMF), (e) sparse-window approach-based sWMF (ssWMF), and (f) proposed method.

TABLE I

MEAN ERROR RATE (MER) INDEX PERFORMANCE OF THE PROPOSED METHOD AND CONVENTIONAL METHODS WHEN USING THE KITTI 2012 AND 2015 STEREO BENCHMARK DATASETS.

Dataset Type	Window Size	MER (%)							
		Methods (Non-Occlusion Condition)				Methods (Occlusion Condition)			
		WMF	sWMF [15]	ssWMF [16]	Proposed	WMF	sWMF [15]	ssWMF [16]	Proposed
KITTI 2012 [21]	5×5	18.2143	18.6557	19.1182	15.1707	20.0922	20.5225	20.9746	17.1172
	9×9	17.7743	18.0694	18.7314	13.6956	19.6617	19.9498	20.5969	15.6760
	13×13	17.8572	17.9641	18.9748	13.0475	19.7431	19.8472	20.8350	15.0426
	17×17	18.2973	18.0814	19.6769	12.7410	20.1734	19.9620	21.5213	14.7166
	21×21	18.9869	18.3367	20.8203	12.5686	20.8475	20.2117	22.6387	14.5743
KITTI 2015 [22]	5×5	22.7569	23.1292	23.7470	19.3041	24.1061	24.4718	25.0787	20.7115
	9×9	22.3964	22.6435	23.2954	17.3801	23.7518	23.9947	24.6349	18.8220
	13×13	22.5073	22.5134	23.4517	16.4204	23.8608	23.8669	24.7885	17.8795
	17×17	22.9633	22.5696	23.4817	15.8713	24.3089	23.9221	24.8163	17.3405
	21×21	23.6811	22.8059	23.6448	15.5413	24.9959	24.2987	24.9266	17.0167

improvement, the proposed method improved the matching accuracy under the non-occlusion and occlusion conditions by 22.95% and 20.27%, respectively, compared with the WMF. For the sWMF, the best MER performance was observed for  $13 \times 13$  window size. In this case, the MER of the sWMF was 17.96% and 19.85% under the non-occlusion and occlusion conditions, respectively. For the ssWMF, the best MER performance was observed for the  $9 \times 9$  window size. In terms of the performance improvement, the proposed method showed better MER performance than sWMF and ssWMF methods. In addition, the proposed method showed better MER performance than the conventional methods for all window sizes.

When using the KITTI 2015 stereo benchmark dataset, The conventional methods showed MER performance similar to the experimental results for KITTI 2012 stereo benchmark dataset. For the  $9 \times 9$  window size, the proposed method had an MER performance of 17.38% and 18.82% under the non-occlusion and occlusion conditions, respectively. In terms of performance improvement, the proposed method improved the

matching accuracy by 22.40% and 23.24%, and 25.39% under the non-occlusion, respectively, compared with the WMF, sWMF, and ssWMF methods. In the occlusion condition, the proposed method improved the matching accuracy by 20.76%, 21.56%, and 23.60%, respectively, compared with the WMF, sWMF, and ssWMF methods. In addition, the proposed method showed better MER performance than the conventional methods for all window sizes.

Fig. 6 shows the left-side input image, initial disparity map using SGM, and refined disparity maps using conventional and proposed methods. As shown in Fig. 6.(b), it can be visually confirmed that the initial disparity map has a large amount of hole regions due to the influence of the noise components in the edge area of the object and the asphalt road. Conversely, when using the conventional methods, it can be visually confirmed that the hole regions are significantly reduced in the edge of the objects and the asphalt road as shown in Fig. 6.(c)-(e). However, although the disparity refinement process was performed using the conventional methods, it is difficult to say that the hole regions are greatly reduced compared with

the initial disparity map. When the proposed method is used, as shown in Fig. 6.(f), it can be visually confirmed that the hole regions are greatly reduced in the edge of the objects and the asphalt road compared with the refined disparity maps using the conventional methods. Therefore, the proposed method can improve not only numerical performance but also visual quality of the disparity map for the user's point of view.

### B. Hardware Resource Utilization

To compare the hardware resource utilization under fair condition, the proposed hardware architecture must be synthesized on the same FPGA platform. Therefore, we used Xilinx XC7K325T FPGA for a fair comparison with sWMF and ssWMF architectures. The WMF architecture was not included in the comparison because previous studies have already shown that a large number of hardware resource are required when the WMF is implemented on the FPGA [15], [16].

Table II lists the synthesis results of the sWMF, ssWMF, and proposed hardware architectures. For fair comparison, the proposed and conventional hardware architectures had working frequency of 148.5 MHz, a disparity range of 128, and an image resolution of 1080p. The ssWMF architecture requires less utilization of the slice look-up table (LUT) and slice register than the sWMF architecture. However, in the ssWMF architecture, the utilization of the BRAM is similar to that of the sWMF architecture. In comparison, the proposed hardware architecture requires less hardware resource utilization than the sWMF and ssWMF architectures. In terms of BRAM, the sWMF and ssWMF architectures require a large number of resources because they require a dual-port BRAM to obtain and store the bilateral weight values and the pixel values of the guided image, which are necessary to compute the output value of the refined disparity map. In contrast, the proposed hardware architecture requires less BRAM than the sWMF and ssWMF architectures. The proposed hardware architecture only uses single-port BRAM because it does not require bilateral weight values for computing the output value of the refined disparity map. Hence, the BRAM utilization of the proposed hardware architecture is less than that of the sWMF and ssWMF architectures. In terms of the slice LUT and slice register, the proposed hardware architecture requires fewer resources because it does not involve a weight computation process.

To further compare the hardware resource utilization and verify the operation in a real environment, the proposed hardware architecture was synthesized on a Xilinx XC7K325T FPGA. For comparison for a small window size, we set the window size as  $13 \times 13$  for the ssWMF and proposed hardware architectures. Only the ssWMF architecture was used for comparison as it requires fewer hardware resources than the experimental results. Therefore, for the  $13 \times 13$  window size, the low-cost characteristic of the architectures can be judged from the fact that the proposed architecture involves less resource utilization than the ssWMF architecture. Fig. 7 shows the hardware resource utilization for the proposed



Fig. 7. Resource utilization of the proposed hardware architecture and ssWMF architecture for the  $13 \times 13$  window size.

TABLE II  
SYNTHESIS RESULTS OF THE PROPOSED ARCHITECTURE AND CONVENTIONAL ARCHITECTURES.

Window Size	Architecture	Resource Type		
		LUT	Register	BRAM
$41 \times 41$	ssWMF [16]	9,737	5,349	63
	Proposed	3,242	4,436	21
$39 \times 39$	sWMF [15]	12,200	15,813	55
	Proposed	2,757	3,840	20
$37 \times 37$	ssWMF [16]	8,211	4,832	57
	Proposed	2,438	3,422	19

hardware architecture and ssWMF hardware architecture. To compare the hardware resource utilization under fair conditions, both architectures had an operation frequency of 148.5 MHz and image resolution of 1080p. For all resource types, the proposed hardware architecture requires less resources than the ssWMF architecture. In terms of reduction percentage, the proposed hardware architecture reduces the slice LUT, slice register, and BRAM utilization by 60.44%, 5.15%, and 66.67%, respectively, compared with the ssWMF architecture.

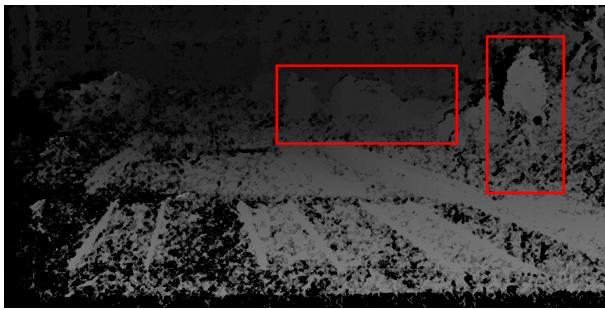
To verify the operation in real environment, our proposed hardware architecture was implemented on Xilinx FPGA Virtex-7 XC7V2000T-FLG1925-2. Fig. 8 shows the initial disparity map using SGM and refined disparity map using proposed hardware architecture. To acquire the initial disparity map, we used a stereo camera with a resolution of  $1280 \times 720$  and YUV-422 format. In terms of the implementation constraint, an operation clock frequency of the proposed hardware architecture was set to 74.25 MHz for synchronizing the stereo camera. As shown in Fig. 8(a), the initial disparity map has many hole regions. These hole regions are reduced compared to the initial disparity map when using the proposed hardware architecture for disparity refinement process, as shown in Fig. 8(b).

## VI. CONCLUSIONS

Herein, a hybrid max–median filter and its hardware architecture are proposed for the disparity refinement process to



(a)



(b)

Fig. 8. Field programmable gate array (FPGA)-based experimental results in real-environment: (a) initial disparity map by using SGM and (b) refined disparity map by using the proposed hardware architecture.

compute the refined disparity map. To evaluate the refinement performance of the proposed method, we used two public stereo datasets. For all window sizes, the proposed method exhibited better refinement performance than the conventional methods for the KITTI 2012 and 2015 stereo benchmark datasets. A comparison of hardware resource utilization showed that the proposed hardware architecture requires less resources in terms of slice LUT, slice register, and BRAM because it does not perform any weight computation based on an exponential equation. In addition, the feasibility of using the proposed hardware architecture in real environment for a working of 74.25 MHz was confirmed. Therefore, the proposed method can be used for embedded stereo vision systems that require a low-cost characteristic and high matching accuracy.

In the future work, we will verify the refinement performance of the proposed hardware architecture by conducting additional experiments on Cityscapes and DrivingStereo datasets. In addition, we will conduct the experiments on performance evaluation based on various disparity range or input image resolution. Thereafter, based on the experimental results, it plans to conduct experiments on infrared stereo cameras as well as YUV or RGB-based stereo cameras used for autonomous driving scenarios.

## REFERENCES

- [1] J. Toledo, M. Lauer, and C. Stiller, “Real-time stereo semi-global matching for video processing using previous incremental information,” *Journal of Real-Time Image Processing*, pp. 1–12, 2022.
- [2] J. Wang, Z. Li, L. Yao, S. Chen, and F. Wu, “Low-resource hardware architecture for semi-global stereo matching,” in *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2019, pp. 1–4.
- [3] C.-H. Choi, Y. Kim, J. Ha, and B. Moon, “Haar filter hardware architecture for the accuracy improvement of stereo vision systems,” in *2021 18th International SoC Design Conference (ISOCC)*. IEEE, 2021, pp. 401–402.
- [4] Z. Lu, J. Wang, Z. Li, S. Chen, and F. Wu, “A resource-efficient pipelined architecture for real-time semi-global stereo matching,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 2, pp. 660–673, 2021.
- [5] L. F. Cambuim, L. A. Oliveira Jr, E. N. Barros, and A. P. Ferreira, “An fpga-based real-time occlusion robust stereo vision system using semi-global matching,” *Journal of Real-Time Image Processing*, vol. 17, no. 5, pp. 1447–1468, 2020.
- [6] L. F. Cambuim, J. P. Barbosa, and E. N. Barros, “Hardware module for low-resource and real-time stereo vision engine using semi-global matching approach,” in *Proceedings of the 30th Symposium on Integrated Circuits and Systems Design: Chip on the Sands*, 2017, pp. 53–58.
- [7] Y. Xie, S. Zeng, and L. Chen, “A novel disparity refinement method based on semi-global matching algorithm,” in *2014 IEEE International Conference on Data Mining Workshop*. IEEE, 2014, pp. 1135–1142.
- [8] P. Yao and J. Feng, “Ensemble learning with advanced fast image filtering features for semi-global matching,” *Machine Vision and Applications*, vol. 32, no. 4, p. 83, 2021.
- [9] P. Yao and J. Feng, “Stacking learning with coalesced cost filtering for accurate stereo matching,” *Journal of Visual Communication and Image Representation*, vol. 78, p. 103169, 2021.
- [10] S. A. Fahmy, “Generalised parallel bilinear interpolation architecture for vision systems,” in *2008 International Conference on Reconfigurable Computing and FPGAs*. IEEE, 2008, pp. 331–336.
- [11] Q. Yang, “Hardware-efficient bilateral filtering for stereo matching,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 5, pp. 1026–1032, 2013.
- [12] M. Humenberger, C. Zinner, M. Weber, W. Kubinger, and M. Vincze, “A fast stereo matching algorithm suitable for embedded real-time systems,” *Computer Vision and Image Understanding*, vol. 114, no. 11, pp. 1180–1202, 2010.
- [13] J. Ding, X. Du, X. Wang, and J. Liu, “Improved real-time correlation-based fpga stereo vision system,” in *2010 IEEE International Conference on Mechatronics and Automation*. IEEE, 2010, pp. 104–108.
- [14] C. Ttofis, C. Kyrkou, and T. Theocharides, “A low-cost real-time embedded stereo vision system for accurate disparity estimation based on guided image filtering,” *IEEE Transactions on Computers*, vol. 65, no. 9, pp. 2678–2693, 2015.
- [15] S. Chen, X. Zhang, H. Sun, and N. Zheng, “swmf: Separable weighted median filter for efficient large-disparity stereo matching,” in *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2017, pp. 1–4.
- [16] J. Hyun, Y. Kim, J. Kim, and B. Moon, “Hardware-friendly architecture for a pseudo 2d weighted median filter based on sparse-window approach,” *Multimedia Tools and Applications*, vol. 80, pp. 34 221–34 236, 2021.
- [17] L. Yin, R. Yang, M. Gabbouj, and Y. Neuvo, “Weighted median filters: a tutorial,” *IEEE Transactions on circuits and systems II: analog and digital signal processing*, vol. 43, no. 3, pp. 157–192, 1996.
- [18] R. A. Schowengerdt, “Chapter 8—image registration and fusion,” *Remote Sensing*, 3rd ed.; Academic Press: Cambridge, MA, USA, 2007.
- [19] D. Akimov, A. Shestov, A. Voronov, and D. Vatolin, “Occlusion refinement for stereo video using optical flow,” in *2012 International Conference on 3D Imaging (IC3D)*. IEEE, 2012, pp. 1–8.
- [20] V. Kumar, A. Asati, and A. Gupta, “Low-latency median filter core for hardware implementation of 5 × 5 median filtering,” *IET Image Processing*, vol. 11, no. 10, pp. 927–934, 2017.
- [21] A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? the kitti vision benchmark suite,” in *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012, pp. 3354–3361.
- [22] M. Menze and A. Geiger, “Object scene flow for autonomous vehicles,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3061–3070.