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

Pangyo R&D Center
Hanwha Systems, Co., Ltd.

Cheol-Ho Choi*, Hyun Woo Oh

cheoro1994@hanwha.com*

Hanwha Systems

Contents

1. Motivation

Motivation

✓ Traditional Method-based Stereo Vision System

- In traditional method-based stereo vision systems, Semi-Global Matching (SGM) is widely used
  - High matching accuracy
  - Reasonable hardware resource utilization
  - Real-time operation (pipeline architecture design)
- When using the SGM for stereo matching, the many "hole" are occurred on texture-less and occluded regions
  - Matching accuracy performance is degraded by "hole"
- To improve matching accuracy performance, weighted median-based filters are widely used for disparity refinement are used

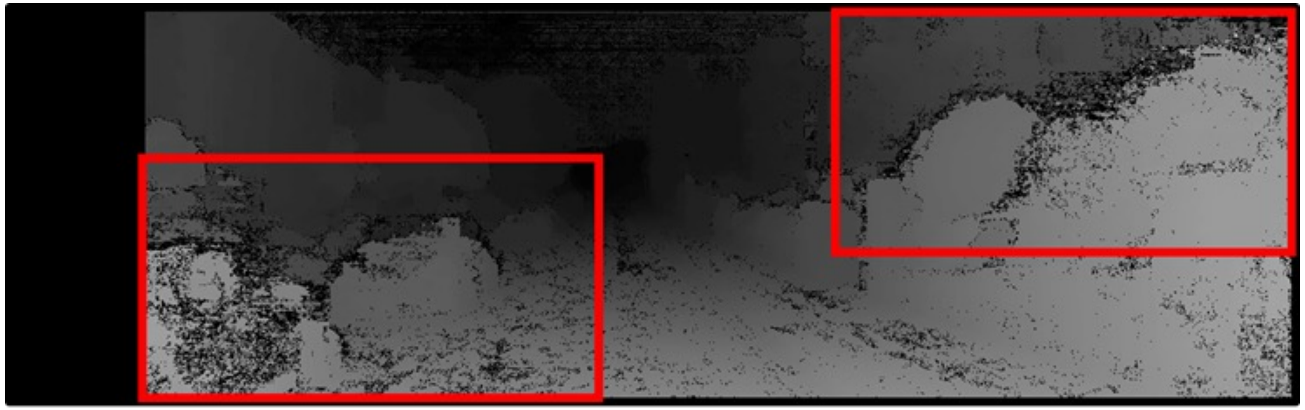

Fig. 1. Initial disparity map using semi-global matching (SGM)

Motivation

✓ Disparity Refinement Process

- weighted median filter (WMF) using bilateral weight is widely used
  - It provides high refinement performance, called hole-filling performance
- However, when implemented on an FPGA, it has drawback of requiring large hardware resource utilization
- For this reason, follow-up studies are conducted
  - Separable WMF (sWMF) [1]
    - It proposed a separable operation for each horizontal and vertical direction to reduce computational complexity
    - It still require high hardware resource utilization and its disparity refinement performance is little degraded
  - Sparse-window-approach-based WMF (ssWMF) [2]
    - It proposed a sparse-window-approach for sWMF
    - It further reduce the hardware resource utilization than sWMF
    - It still require high hardware resource utilization of block random access memory (BRAM) and its disparity refinement performance is very degraded

[1] S. Chen, et al., "sWMF: Separable weighted median filter for efficient large-disparity stereo matching," *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2017
[2] J. Hyun, et al., "Hardware-friendly architecture for a pseudo 2d weighted median filter based on sparse-window approach," *Multimedia Tools and Applications*, vol. 80, pp. 34221-34236, 2021

Motivation

✓ Disparity Tendency

- Horizontal Direction
  - It needs to **preserve the edge information** for objects
- Vertical Direction
  - Disparity value **gradually increases** from the top coordinate to the bottom coordinate
  - In other words, the depth value **gradually increases** from the bottom coordinate to the top coordinate [3]
- Therefore, we proposed hybrid max-median filter to utilize these disparity characteristics for horizontal and vertical directions

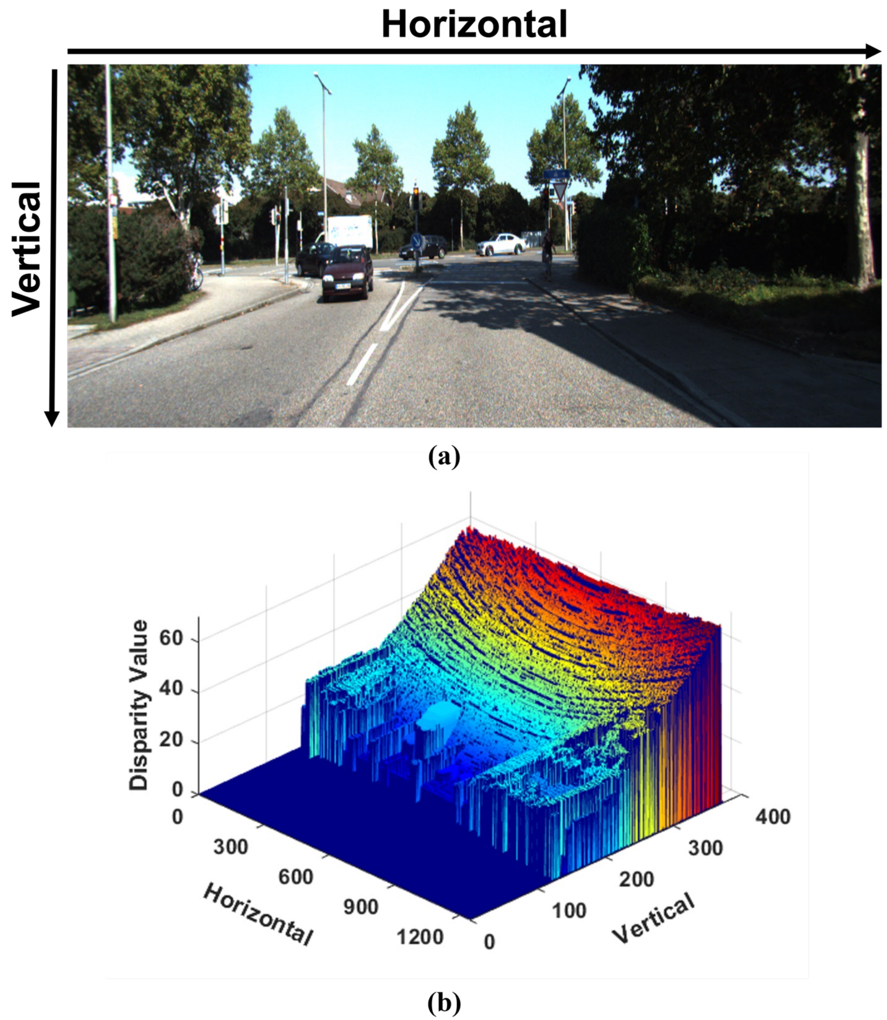

Fig. 2. (a) left-side stereo input image and (b) 3D plot for disparity values

[3] R. A. Schowengerdt, "Chapter 8-image registration and fusion," *Remote Sensing, 3rd ed.*; Academic Press: Cambridge, MA, USA, 2007

2. Proposed Method

Proposed Method

✓ Algorithm Flow

- Sub-window Generation
  - Generate the $N \times N$ sub-window
- Inner-sub-window Generation
  - **Generate the eight inner-sub-window** from $N \times N$ sub-window
- Maximum Value Selection
  - Select the **eight maximum pixel values** from eight inner-sub-windows
- Median Value Selection
  - Select the **median value** from eight maximum pixel values and center pixel value

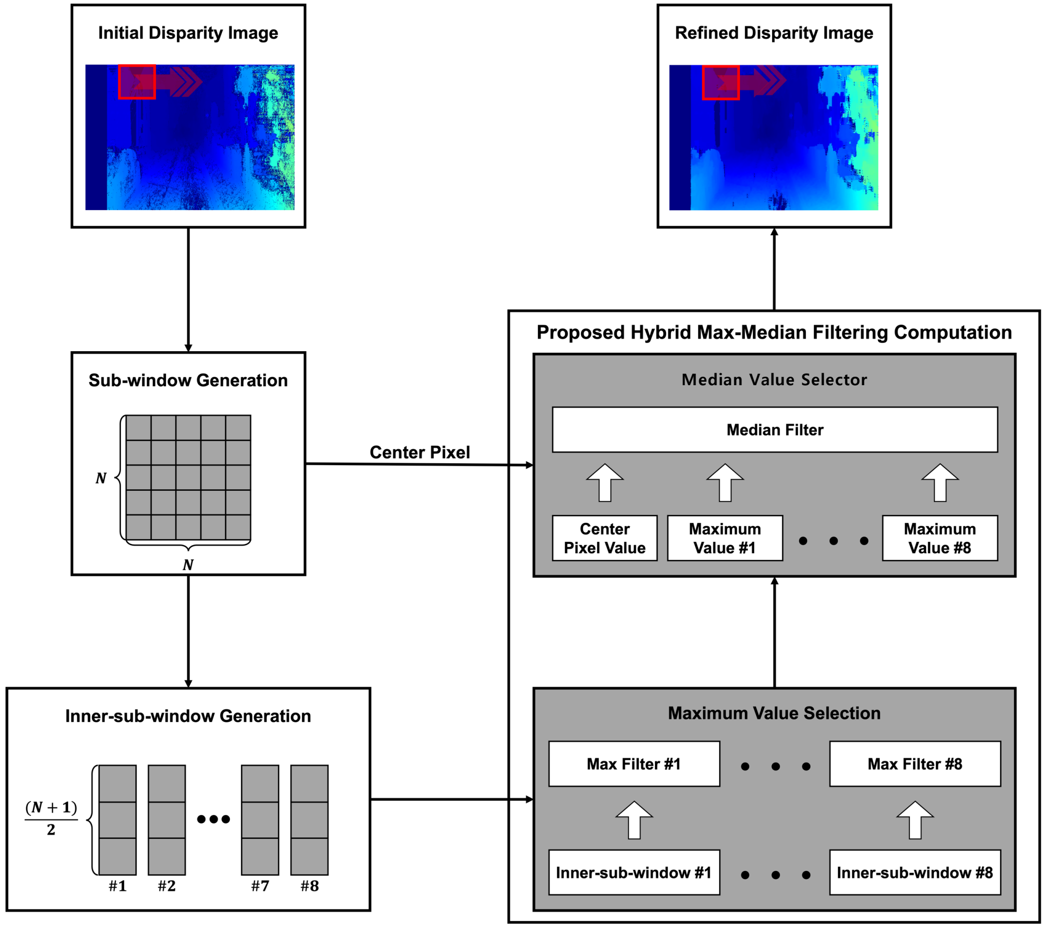

Fig. 3. Proposed method

Proposed Method

✓ Inner-sub-window Generation

- Generate each sub-window for 8-path direction
- Red indicator
    - Horizontal direction
    - Vertical direction
- Blue indicator
    - Diagonal direction
        - North-West
        - North-East
        - South-West
        - South-East

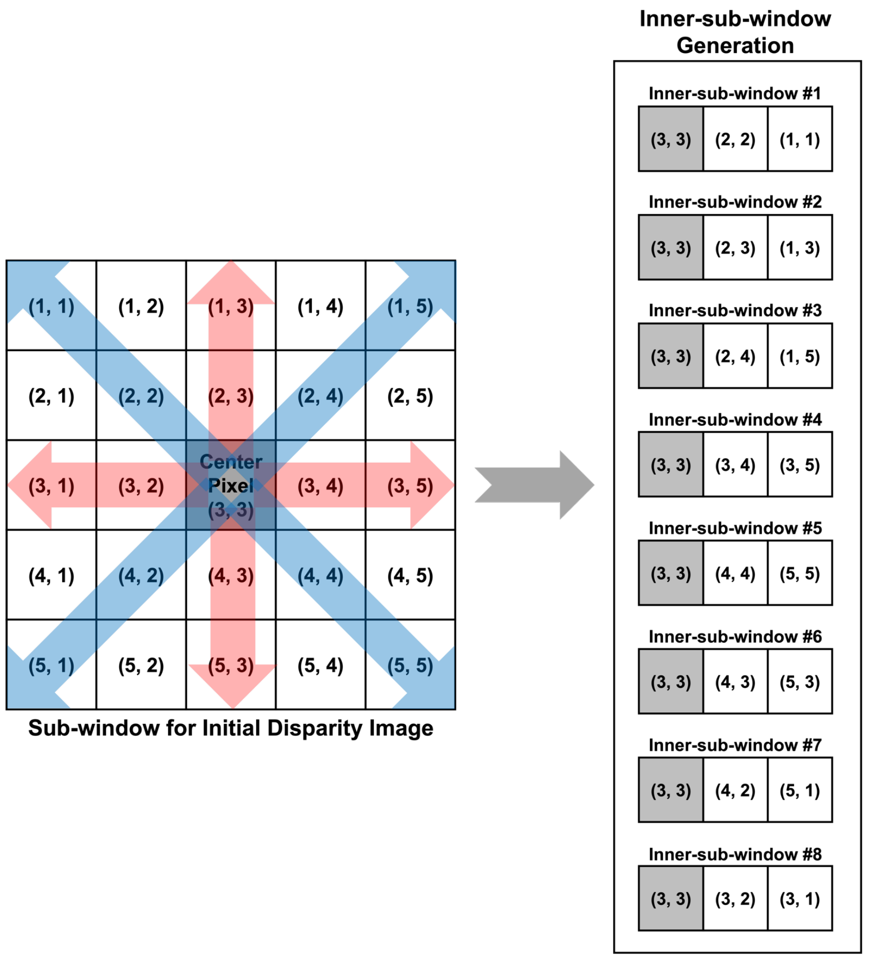

Fig. 4. Proposed inner-sub-window generation method

3. Proposed Hardware Architecture

Proposed Hardware Architecture

✓ Overall Architecture

- Inner-sub-window Generator
    - Generate N inn-sub-window from input disparity map
- Maximum Value Selector
    - Select the eight maximum pixel values and center pixel value from $N \times N$ sub-window generated by Inner-sub-window Generator module
- Median Value Selector
    - Select the median pixel value from the nine pixel values selected by Maximum Value Selector module
    - It select median pixel value as output value of refined disparity map

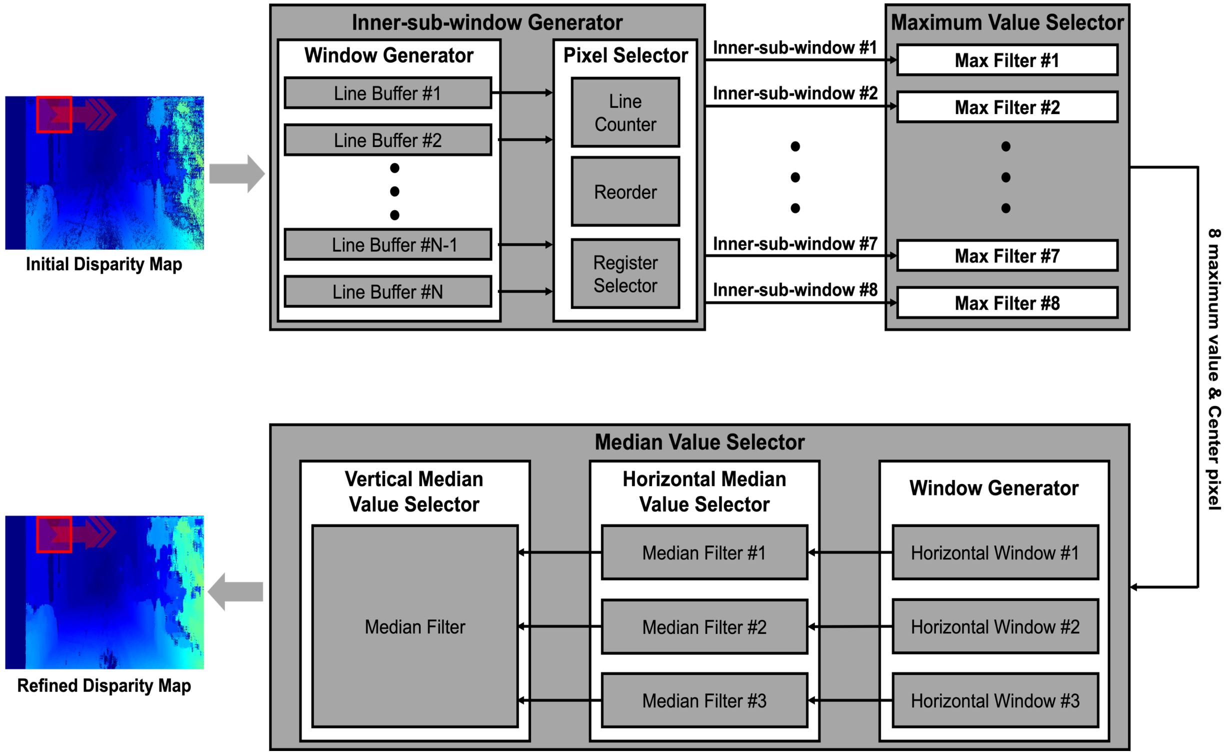

Fig. 5. Proposed disparity refinement processor architecture

Proposed Hardware Architecture

✓ Inner-sub-window Generator

- Window Generator
  - Generate $N\times N$ sub-window using line buffers based on BRAMs and registers
- Pixel Selector
  - Line Counter
    - **Count the address value** based on the line and frame valid signal (Hsync == Line valid) (Vsync == Frame valid)
  - Reorder
    - Based on result value of Line Counter module, **Reorder the parallelized input pixel values** from Window Generator module
  - Register Selector
    - Select the **corresponded pixel values**

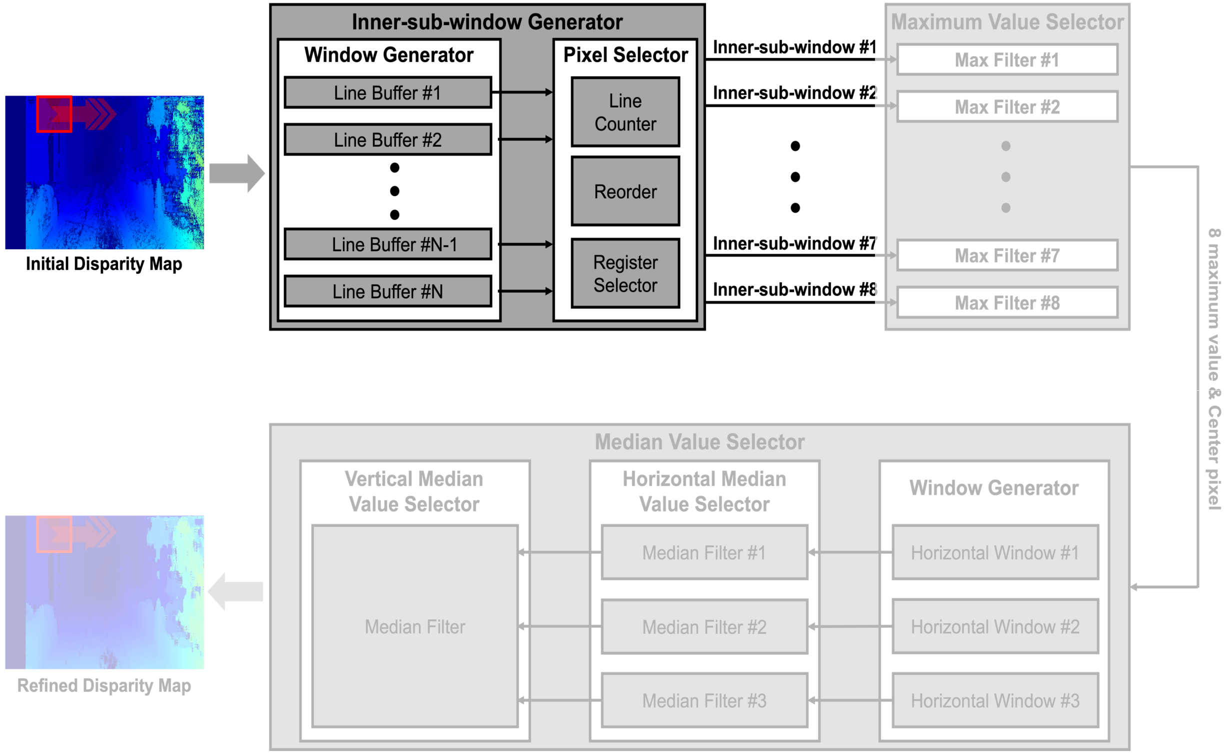

Fig. 5. Proposed disparity refinement processor architecture

Proposed Hardware Architecture

✓ Maximum Value Selector

- Max Filter
  - Select the maximum pixel value
  - It utilizes <u>pyramidal comparison architecture</u> using comparators

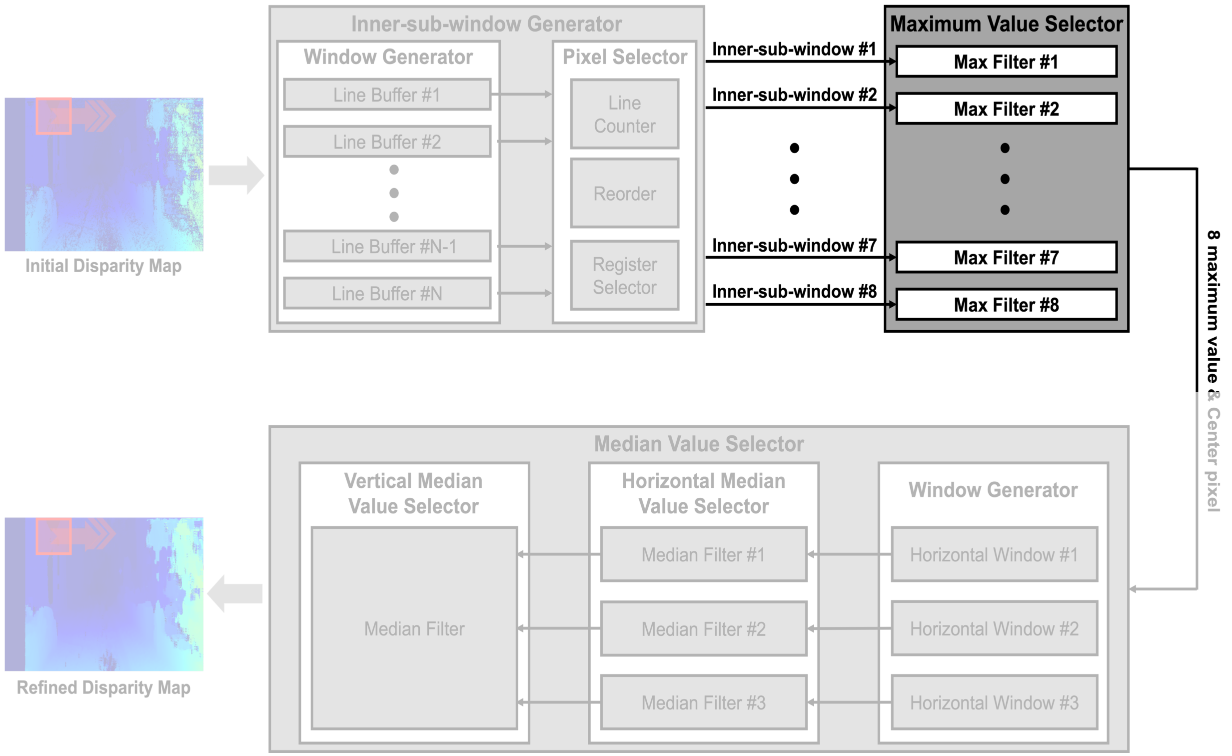

Fig. 5. Proposed disparity refinement processor architecture

Proposed Hardware Architecture

✓ Median Value Selector

- Window Generator
  - Generate the **three horizontal windows** from nine pixel values including eight maximum pixel values and center pixel value
- Horizontal Median Value Selector
  - **Select the three median pixel values** from the three horizontal windows
  - the median filter module requires 3 clocks to select the median pixel value for each horizontal window
- Vertical Median Value Selector
  - **Select the median pixel value as output value** from the selected three median pixel values from the Horizontal Median Value Selector module

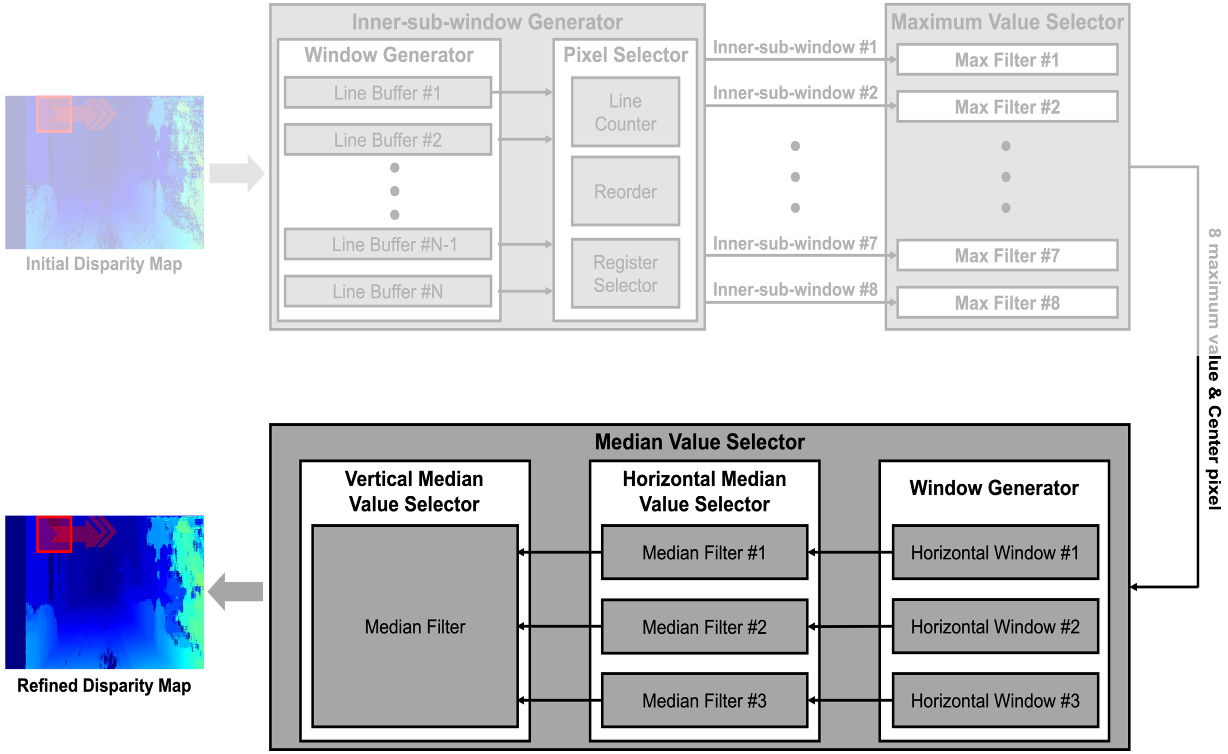

Fig. 5. Proposed disparity refinement processor architecture

4. Experimental Results

Experimental Results

✓ **KITTI Stereo Dataset**

- KITTI 2012 stereo dataset: 195 images
- KITTI 2015 stereo dataset: 200 images
- Figure 6 Explanation
  - Fig. 6(a) : Left-side stereo image
  - Fig. 6(b) : Initial disparity map using SGM
  - Fig. 6(c) : Refined disparity map using the WMF
  - Fig. 6(d) : Refined disparity map using the sWMF
  - Fig. 6(e) : Refined disparity map using the ssWMF
  - Fig. 6(f) : Refined disparity map using the proposed method

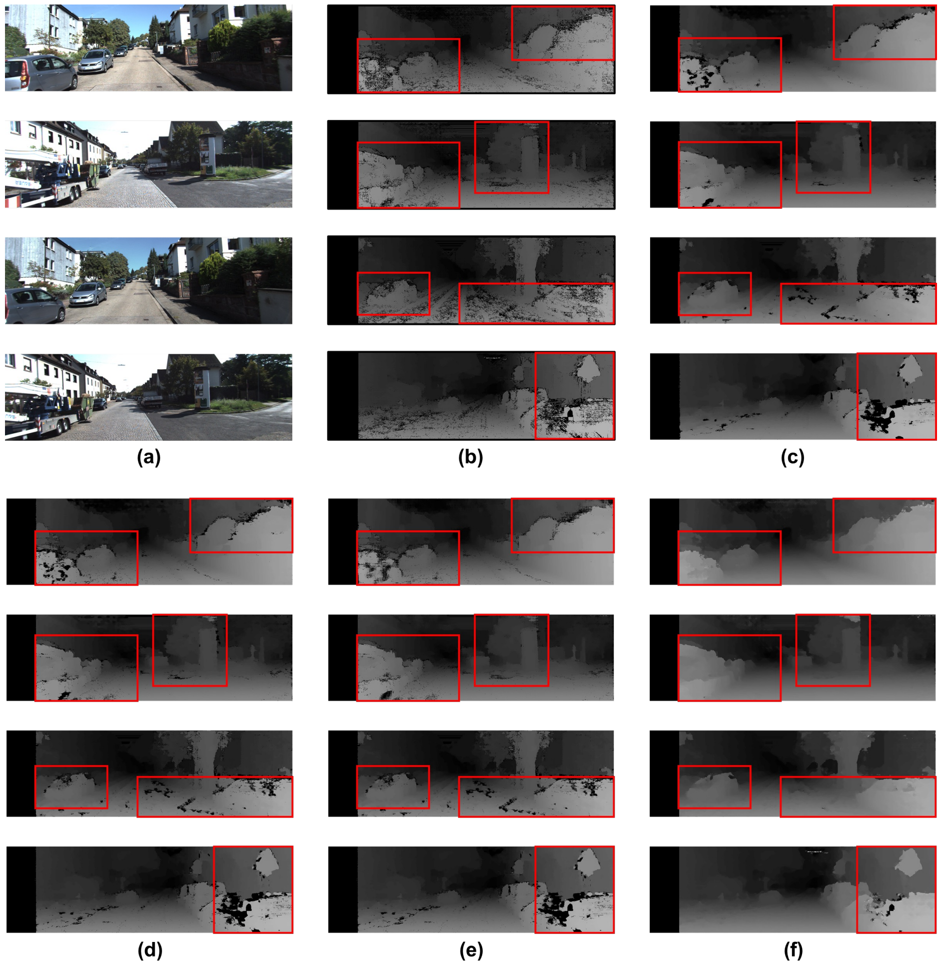

Fig. 6. Experimental results using KITTI 2012 and 2015 stereo benchmark datasets

Experimental Results

✓ **KITTI 2012 and 2015 Stereo Dataset**

Dataset Type	Window Size	MER (%)							
		Methods (Non-Occlusion Condition)				Methods (Occlusion Condition)			
		WMF	sWMF	ssWMF	Proposed	WMF	sWMF	ssWMF	Proposed
KITTI 2012	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	5 × 5	22.7569	23.1292	23.7470	19.0341	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

Table 1. Mean Error Rate (MER) performance of the proposed and conventional methods using the KITTI 2012 and 2015 stereo benchmark datasets

Experimental Results

✓ KITTI 2012 and 2015 Stereo Dataset

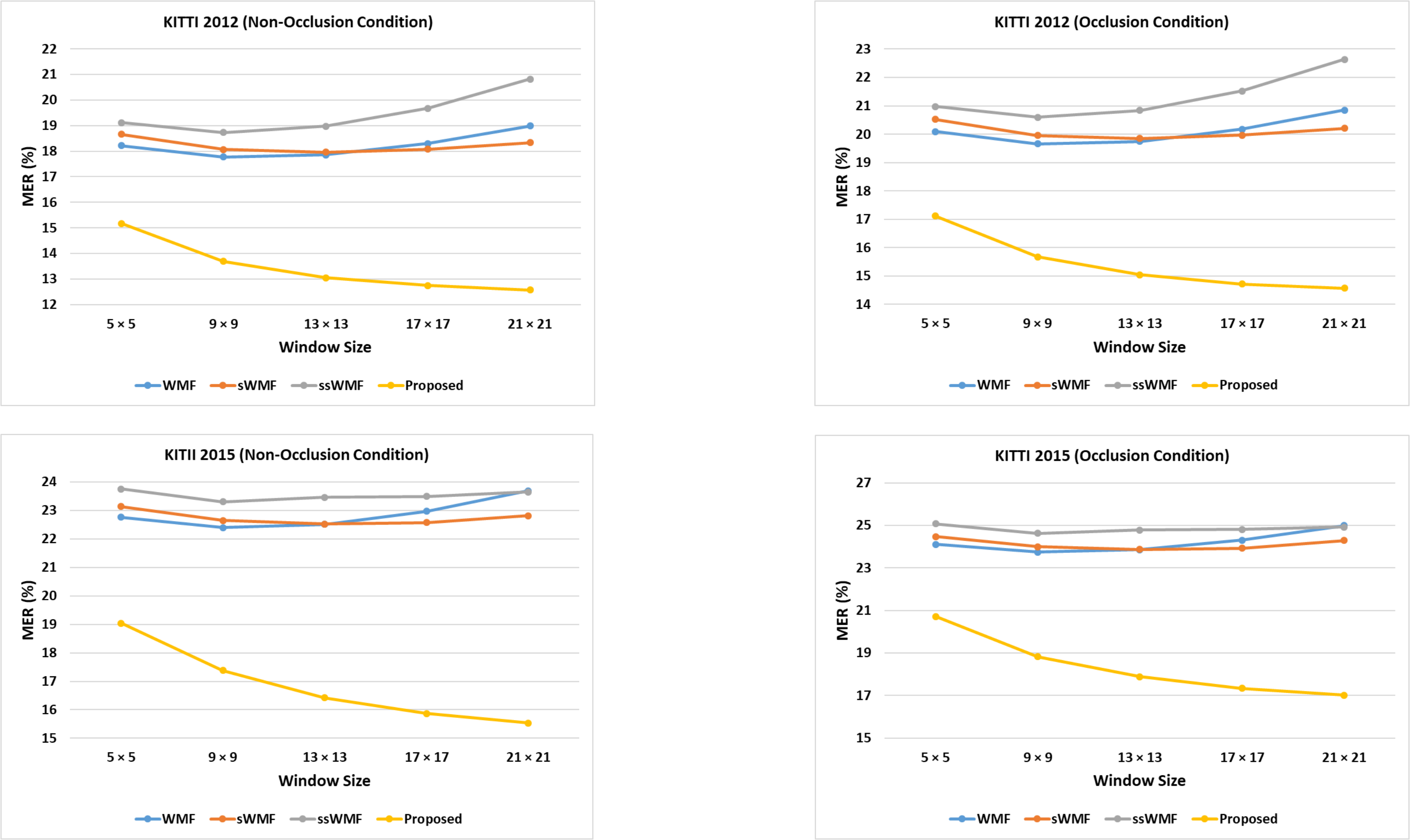

Fig. 7. Experimental results when using the KITTI 2012 and 2015 stereo benchmark datasets

Experimental Results

✓ Cityscapes Dataset

- Dataset collected from various German cities (e.g., Berlin and Munich)
- It has 1525 stereo test images
- Refinement performance
    - In all window size, the proposed method showed better refinement performance than the conventional methods
    - In the 13 × 13 window size, the proposed method showed best refinement performance

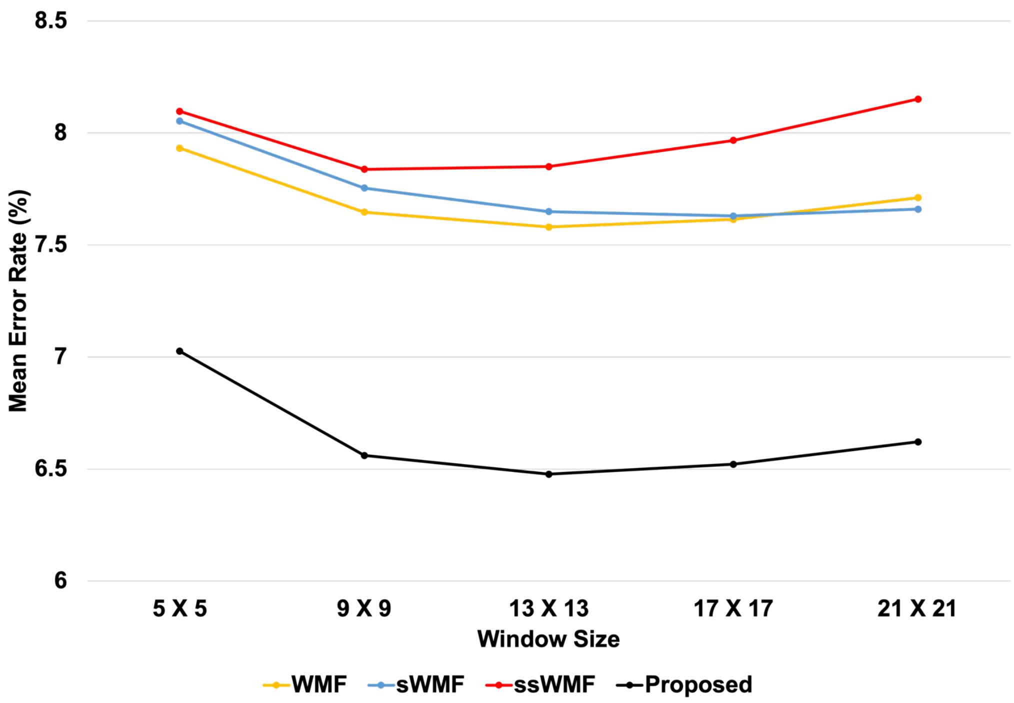

Fig. 8. Experimental results using Cityscapes dataset

Experimental Results

✓ Hardware Resource Utilization

- Synthesis Condition
  - Vivado version : 2020.2
  - Target FPGA board : Xilinx XC7K325T
  - Operation frequency : 148.5 MHz
  - Resolution : FHD (1080p)
  - Disparity range : [0 128]

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

Table 2. Synthesis results of the proposed architecture and conventional architectures

Experimental Results

✓ Hardware Resource Utilization

- When implemented on 13 × 13 window size, the proposed disparity refinement processor required less hardware resource utilization than the ssWMF architecture
  - LUT : 2040 → 773 (62.11% ↓)
  - Register : 1516 → 1265 (16.56% ↓)
  - BRAM : 21 → 7 (66.67% ↓)

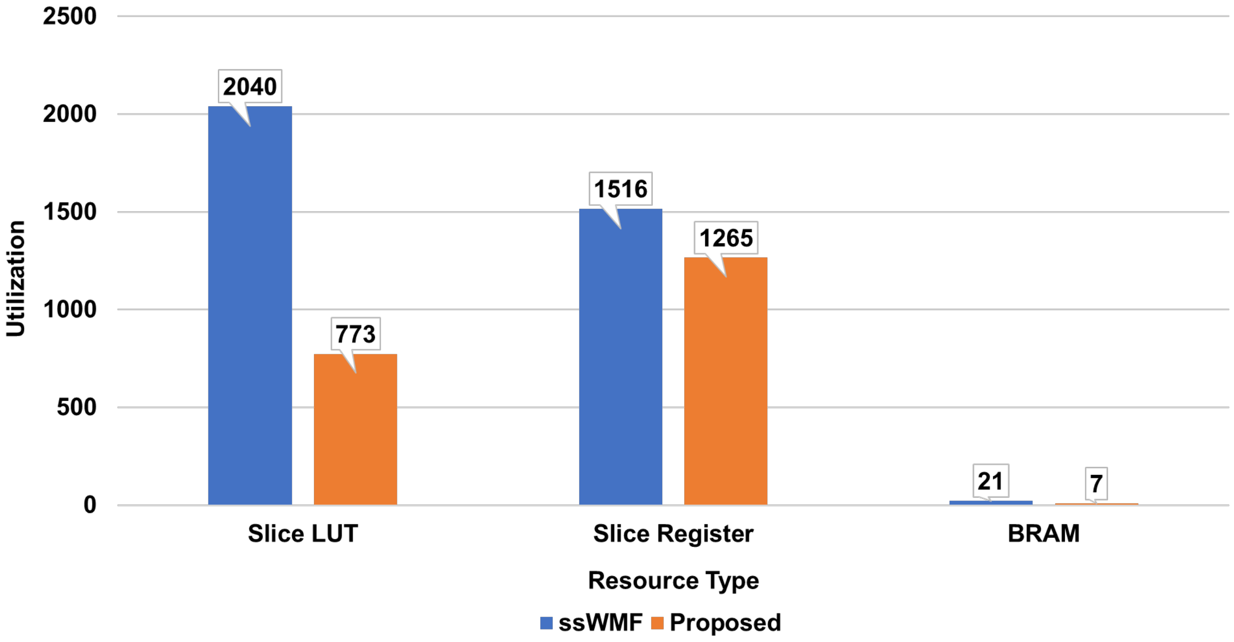

Fig. 9. Resource utilization of the proposed hardware architecture and ssWMF architecture for the 13 × 13 window size

Experimental Results

✓ Implementation Result

- Implementation Condition
  - Target FPGA: Xilinx Virtex-7 XC7V2000T
  - Resolution: 1280 × 720 (HD)
  - Video format: YUV-422
  - Operation frequency: 74.25 MHz
  - Window size: 13 × 13

(a) (b) (c)

Fig. 10. Implementation results: (a) left-side input image, (b) initial disparity map, and (c) refined disparity map

5. Conclusion

Conclusion and Future Work

✓ Conclusion

- We proposed a disparity refinement processor architecture based on hybrid max-median filter
- When using the KITTI and Cityscapes stereo datasets, proposed architecture showed better refinement performance than the conventional architectures → High performance characteristic
- When synthesized on Xilinx FPGA board, proposed architecture required fewer hardware resource utilization than the conventional architectures → Low-cost characteristic
- It can be used for embedded stereo vision systems that requires low-cost and high-performance characteristics

✓ Future Work

- We will verify the refinement performance by conducting additional experiments on DrivingStereo dataset.
- We will conduct the experiments on performance evaluation and hardware resource utilization comparison based on various disparity range or input image resolution.
- We plans to conduct experiments on infrared stereo cameras.
    - Infrared stereo camera : QuantumRed by Hanwha Systems, Co., Ltd.

Thank you for listening