CHAPTER I

INTRODUCTION

This chapter explains the purpose of this thesis in detail. It gives the basics of disparity and depth concepts, most common disparity estimation methods, and preprocessing transforms. Furthermore, it includes the state-of-the-art works in the literature that are related to depth compression and their hardware implementations where available.

The works presented in Chapters 1 and 2, Sections 3.1 and 3.3, Chapter 4, and Section 5.3 were included in our journal submission [1].

1.1 Problem description

This work aims to introduce an area and power-efficient algorithm as well as its hardware architecture for Disparity Map (DM) compression. Such architectures are required when processing real-time, high resolution, and high frame rate video signals with limited hardware resources. Moreover, in today's System-on-Chips (SoCs) low power consumption is critical due to thermal constraints and battery life of mobile devices. Accessing DRAM may consume a significant portion of the total power in mobile SoCs. Thus, the proposed algorithm and its hardware architecture aim to reduce the total power consumption by decreasing the memory bandwidth. Depth information is essential for many computer vision applications such as augmented reality, 3D mapping, 3D movies, video surveillance, military applications, robotics, and self-driving vehicles. Depth information can be defined as the distance of the surfaces of scene objects to the viewpoint. The terms depth and disparity are related to each other. Binocular disparity describes the shift amount in coordinates of the same object between a stereo image pair [2].

The depth value depends on the disparity value and the camera calibration. Depth information can be estimated by imitating the human visual system with a

stereo camera pair. The third-dimension coordinate of an object can be extracted by using left and right images of the stereo pair, where one image will be the shifted version of the other. Disparity estimation is a process very similar to motion estimation, which is widely used in video compression [3]. Unlike the block matching for video compression, which is applied within a two-dimensional search window [4], block matching for disparity estimation implements a single dimensional search window along the horizontal axis.

In this work, horizontally rectified stereo image pairs are utilized as input for disparity estimation; i.e., it is assumed that considering disparities at x-axis only is sufficient to correctly estimate the output DM. Figure 1 shows the typical structure of a binocular stereo vision system, where the corresponding disparities of Point-1 and Point-2 are equal to:

$$Disparity(Point\text{-}1) = xL1 - xR1 \tag{1}$$

$$Disparity(Point\text{-}2) = xL2 - xR2 \tag{2}$$

Where $xL1$ and $xR1$ represent projection of Point-1 on the x-axis of left and right camera focal planes, respectively. Likewise, $xL2$ and $xR2$ represent projections of Point-2 on the x-axis. Because Point-1 is closer to the focal planes of the stereo camera pair, its disparity value will be larger when compared to the disparity value of Point-2. Depth values of Point-1 and Point-2 can be found using the following equations:

$$Depth(Point\text{-}1) = \frac{Baseline\,.\,Focal\ Length}{xL1 - xR1} \tag{3}$$

$$Depth(Point\text{-}2) = \frac{Baseline\,.\,Focal\ Length}{xL2 - xR2} \tag{4}$$

We can define the baseline as the distance between stereo camera pair lenses, whereas focal length of the lens is the distance between the lens and the image sensor when the subject is in focus.

It is important to preprocess stereo image pairs before the disparity estimation due to right and left images being exposed to different noise and brightness levels.

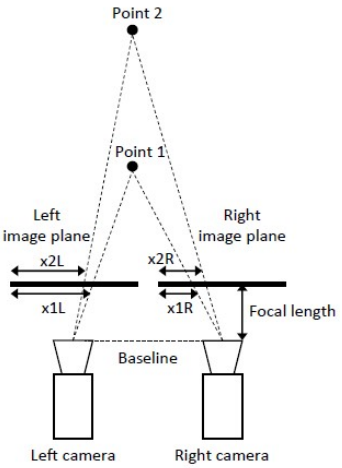


Figure 1: Binocular stereo vision system.

There are transforms presented in the literature to increase the tolerance against these negative effects.

These transforms are applied before the disparity estimation, and this work focuses on three widely used transforms in the literature: Rank Transform (RT), Census Transform (CT), and Complete Rank Transform (CRT).

RT applies a filtering window, where the center value of the filtering window gets replaced by the total number of samples within the filtering window that are smaller than the center value [5]. CT outputs a string of $(N.M-1)$ bits, where N and M are the width and the height of CT's filtering window. In CT, window locations that have smaller or equal gray-scale intensity values than the center value are flagged as '0' and remaining locations with larger intensity are flagged as '1'. Thus, the output of CT is a bit string. CRT [6] is a modified version of the RT. Unlike CT and RT, which output a single value for the center location of the filtering window, CRT outputs a window for each input disparity value. CRT is morphologically invariant; meaning that the output image is invariant to any monotonically increase in its gray-scale intensity. Figure 2 illustrates a 3 x 3-windowing example for these three transforms. Other than the local block matching methods, there are semi-global and global methods proposed in the literature for disparity estimation. Complexities of these methods increase from local methods to global methods. Semi-Global Matching (SGM) [7], global algorithms

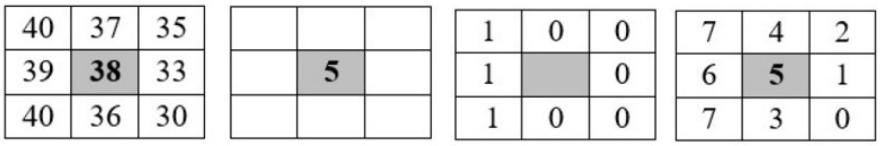


Figure 2: Preprocessing transforms applied to stereo pairs. (a) 3 x 3 Intensity window. (b) RT. (c) CT. (d) CRT.

such as Maximum Likelihood Stereo [8], Belief Propagation (BP) [9] produce more accurate DMs than the local methods. The proposed compression algorithm can be applied to DMs obtained by any disparity estimation method.

In this thesis the performance of the proposed algorithm is evaluated on DMs that are generated by local methods. Because the proposed compression method is claimed to be low-cost, it is expected to work in conjunction with a low-cost method such as local block matching.

Local block matching algorithms divide the input frame into rectangular blocks and search for similar blocks in the reference frame. Each block from one frame (left or right) is searched in the other frame within given the search window. The block in the search window giving the minimum error value is chosen as the match. A preprocessing transform is applied to input stereo image pairs to increase the accuracy of the matching process. The error criteria in this search process can vary based on the type of the preprocessing transform. RT and CRT apply Sum of Absolute Differences (SAD). Depending on the input image resolution, increasing the block size generally leads to a better prediction accuracy, but it also increases the hardware complexity. Eq. (5) shows the SAD metric:

$$SAD = \sum_{i=1}^{M} \sum_{j=1}^{N} |I_R(i,j) - I_L(i,j)| \tag{5}$$

In this equation, I_R, I_L, i, j, M, and N represent intensity values for right and left grayscale stereo images, pixel coordinates, and maximum search window limits on x and y axes, respectively. CT uses Hamming distance for calculating the error between two bitstrings obtained from left and right images. As shown in Figure 3, which demonstrates the calculation of the Hamming distance between

two example bit strings, using Hamming distance as the error criteria requires an XOR operation between input bitstrings followed an accumulation operation.

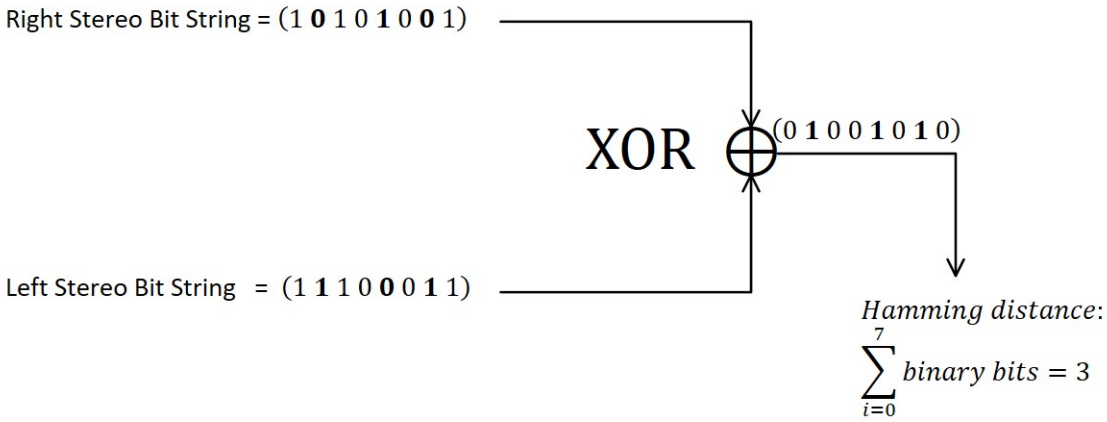


Figure 3: Calculation of Hamming distance between two bit strings of 8-bit wide.

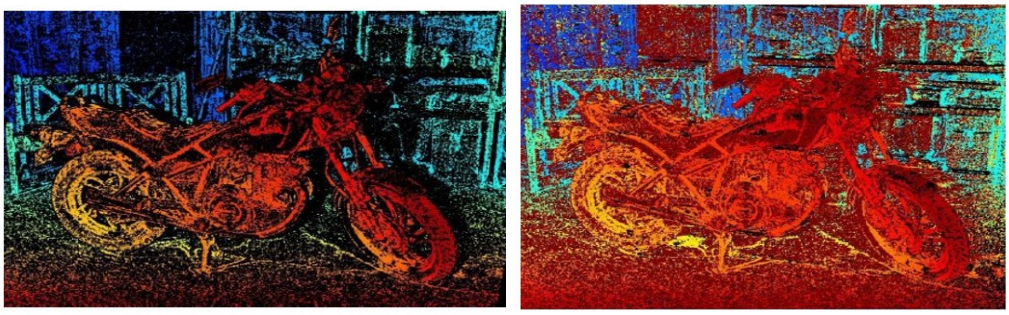
1.2 Literature review

Accessing temporal data when processing high resolution video streams leads to increased DRAM bandwidth, which is becoming the main contributor to the total dynamic power consumption of an SoC. Thus, it is crucial to compress DMs for real-time applications requiring the disparity (or depth) data of the previous frame(s) on a thermally or battery constrained device.

A wide range of papers have been reported in the literature about Multi-View extension of High Efficiency Video Coding (MV-HEVC) and 3D-HEVC [10]. These papers generally present a trade-off between coding efficiency and computational complexity [11–13]. For example, in [11], computational complexity of 3D-HEVC is significantly reduced with a slight decrease in the coding efficiency. This is accomplished by applying static decision trees, which were built with data mining and machine learning. Likewise, in [12] early termination mode decision, adaptive search range motion estimation, and fast disparity estimation approaches are presented to reduce the complexity of 3D-HEVC by compromising a small loss in PSNR and a small increase in the bitrate. In [13], data analysis and clustering based approach is used to skip depth modeling modes in 3D-HEVC where possible. HEVC can obtain much higher compression rates, but it is very costly to

2.3 Median filter

Block matching disparity estimation process is prone to errors resulting with outlier disparity values. Using a median filter, these outliers can be detected and replaced by the median value of the filtering window. Therefore, a 3 x 3 median filtering is a commonly used to enhance the quality of DMs [23]. Taking this state-of-the art processing step into account, we have smoothed DMs with a 3 x 3 median filter before compressing them with the proposed algorithm. Figure 4(a) shows a block matching-based DM in jet colormap array format, where warm colors represent higher disparities. The black pixels in the DM represent the incorrectly estimated disparity locations as if the disparity values were zero in these locations. As it can be seen in Figure 4(b), application of a 3 x 3 median filter significantly decreases these outliers and generates a better DM.



(a) (b)

Figure 4: Application of a median filter on a DM. (a) Orginal DM. (b) Median filtered DM.

APPENDIX A

MIDDLEBURY STEREO DATASET

This appendix shows the used stereo pairs to generate DMs using block matching algorithm, as well as each ground truth corresponding to each pair [24]. Please notice that original images with high-resolutions as indicated in Table 3 are down-sampled into a resolution of 500 x 750 using MATLAB to produce a thesis PDF file with proper size.

(a) (b) (c)

Figure 19: Adirondack. (a) Left image. (b) Right image. (c) Ground truth DM.

(a) (b) (c)

Figure 20: Jadeplant. (a) Left image. (b) Right image. (c) Ground truth DM.

(a) (b) (c)

Figure 21: Motorcycle. (a) Left image. (b) Right image. (c) Ground truth DM.

(a) (b) (c)

Figure 22: Piano. (a) Left image. (b) Right image. (c) Ground truth DM.

(a) (b) (c)

Figure 23: Pipes. (a) Left image. (b) Right image. (c) Ground truth DM.

(a) (b) (c)

Figure 24: Playroom. (a) Left image. (b) Right image. (c) Ground truth DM.

(a) (b) (c)

Figure 25: Playtable. (a) Left image. (b) Right image. (c) Ground truth DM.

(a) (b) (c)

Figure 26: Recycle. (a) Left image. (b) Right image. (c) Ground truth DM.

(a) (b) (c)

Figure 27: Shelves. (a) Left image. (b) Right image. (c) Ground truth DM.

(a) (b) (c)

Figure 28: Vintage. (a) Left image. (b) Right image. (c) Ground truth DM.

APPENDIX B

DMs USING BLOCK MATCHING

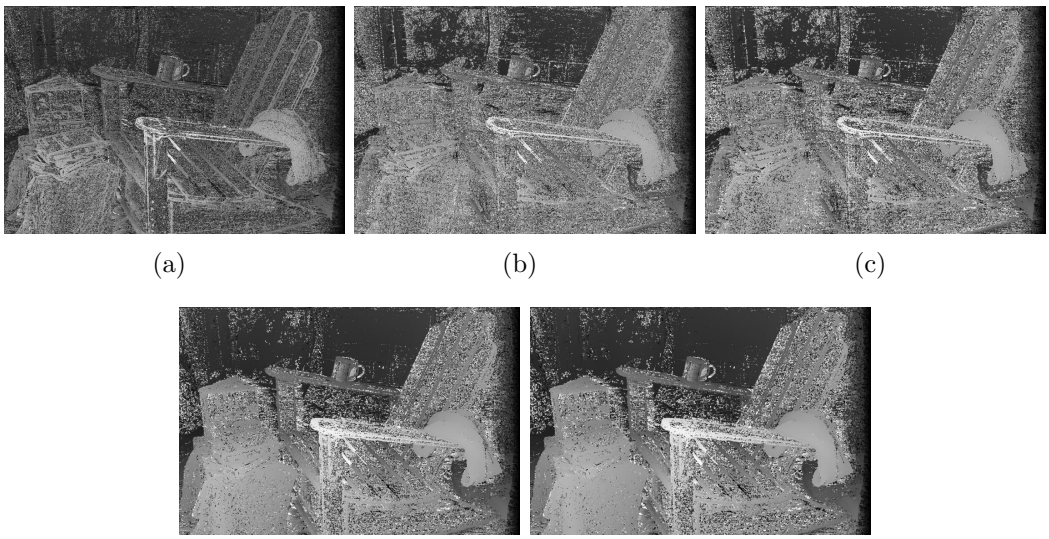
The generated DMs using block matching algorithm are displayed in this appendix with the same resolution used in Appendix A. Due to high error rate in those DMs compared to their ground truths, it is not easy for a human eye to distinguish between median filtered and unfiltered DMs. Thus, we include only postprocessed (filtered) DMs in this Appendix. Table 15 illustrates quality differences between orginal DMs dataset and generated DMs using local block matching algorithm. As it can be obviously seen from Table 15, the generated DMs are significantly noisy due to some factors which are:

1. The selected disparity range for generating DMs [0, 255] is generally lower than maximum disparity values in most ground truth DMs.

2. The selected window size for both preprocessing transforms and disparity matching is not large enough for better matching performance. Increasing window size would increase the complexity of the used cost functions such as SAD and Hamming distance.

3. The used algorithm is local, which means it does not have a specific cost function that considers multiple regions (or parameters) of each stereo image while deciding the best possible disparity value of a pixel.

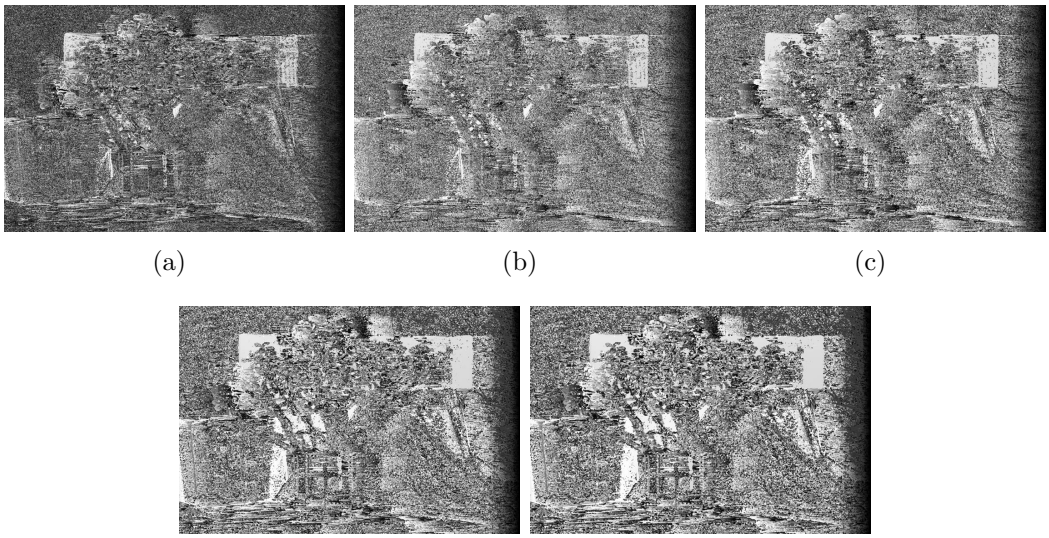
Nonetheless, this thesis has focused more on DM compression rather than the generation of DM itself. Since using a complex and global algorithm with a low-cost DM compression can be considered as a contradiction in the area of application. Moreover, generating large number of high-resolution DMs using algorithms such as SGM consumes an enormous amount of runtime in MATLAB.



(a) (b) (c)

(d) (e)

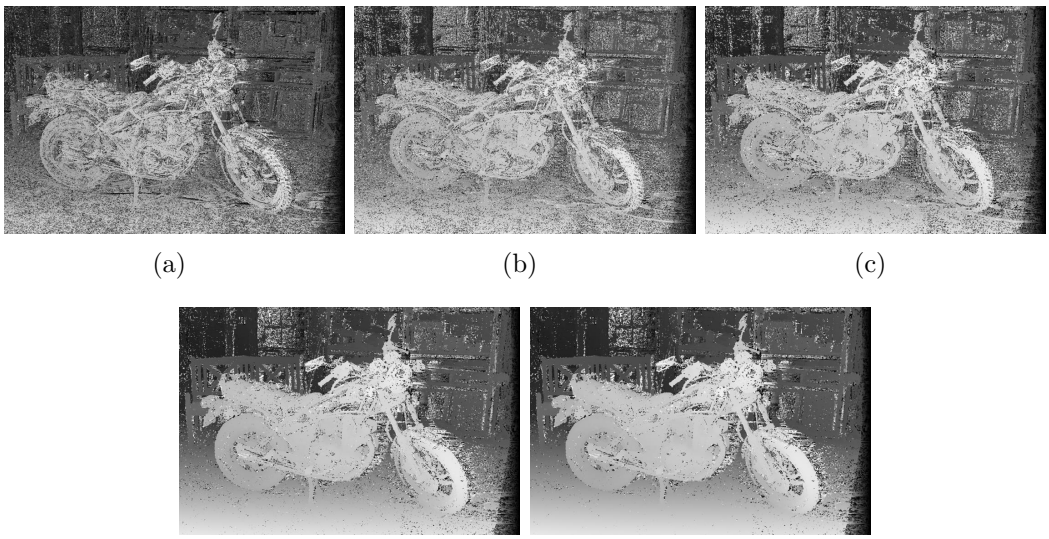
Figure 29: Adirondack. (a) 7x7 CT + 1x1 search window. (b) 7x7 RT + 5x5 search window. (c) 7x7 RT + 7x7 search window. (d) 7x7 CRT + 5x5 search window. (e) 7x7 CRT + 7x7 search window.



(a) (b) (c)

(d) (e)

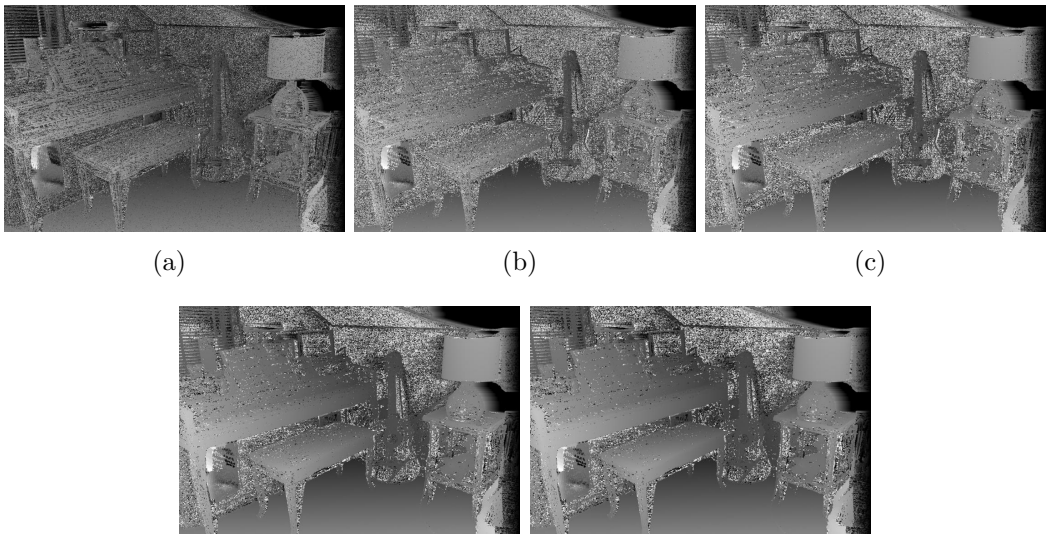
Figure 30: Jadeplant. (a) 7x7 CT + 1x1 search window. (b) 7x7 RT + 5x5 search window. (c) 7x7 RT + 7x7 search window. (d) 7x7 CRT + 5x5 search window. (e) 7x7 CRT + 7x7 search window.



(a) (b) (c)

(d) (e)

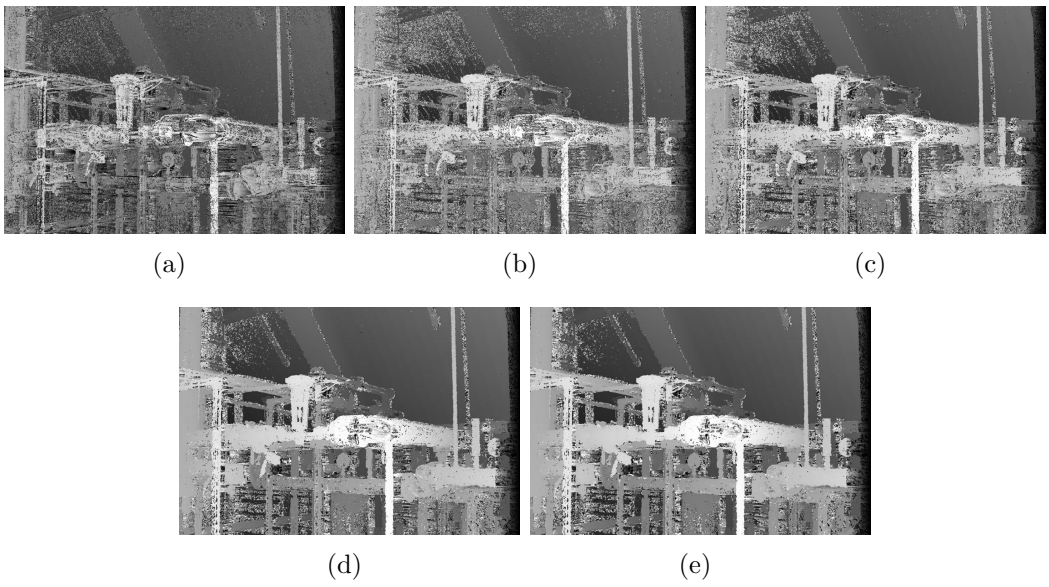
Figure 31: Motorcycle. (a) 7x7 CT + 1x1 search window. (b) 7x7 RT + 5x5 search window. (c) 7x7 RT + 7x7 search window. (d) 7x7 CRT + 5x5 search window. (e) 7x7 CRT + 7x7 search window.



(a) (b) (c)

(d) (e)

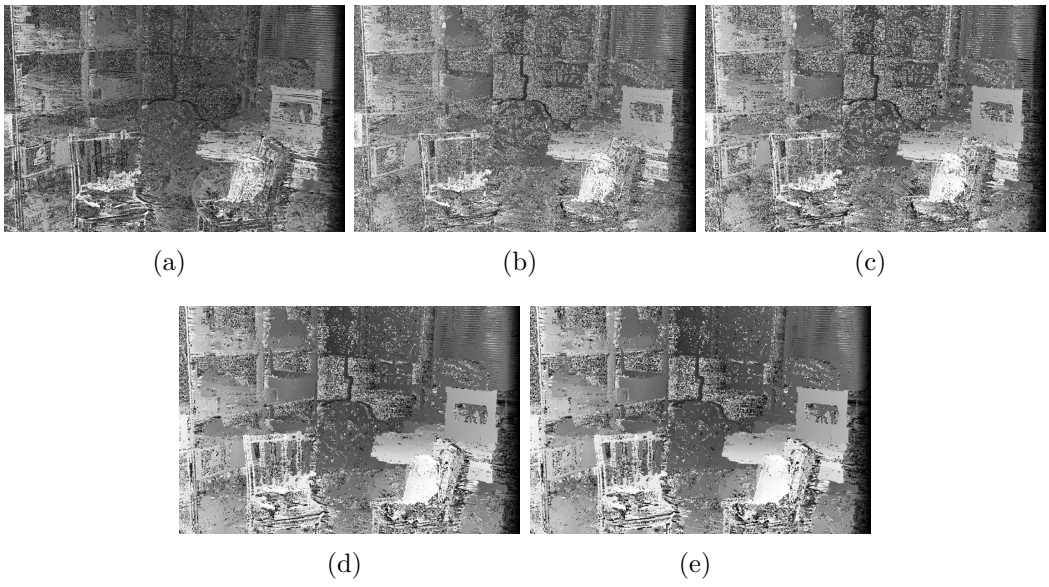
Figure 32: Piano. (a) 7x7 CT + 1x1 search window. (b) 7x7 RT + 5x5 search window. (c) 7x7 RT + 7x7 search window. (d) 7x7 CRT + 5x5 search window. (e) 7x7 CRT + 7x7 search window.



(a) (b) (c)

(d) (e)

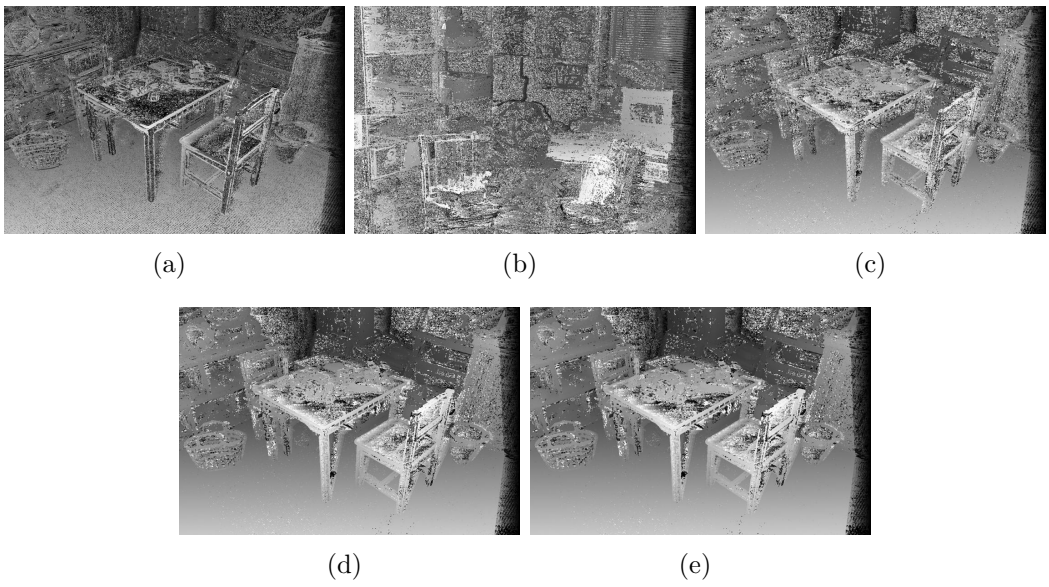
Figure 33: Pipes. (a) 7x7 CT + 1x1 search window. (b) 7x7 RT + 5x5 search window. (c) 7x7 RT + 7x7 search window. (d) 7x7 CRT + 5x5 search window. (e) 7x7 CRT + 7x7 search window.



(a) (b) (c)

(d) (e)

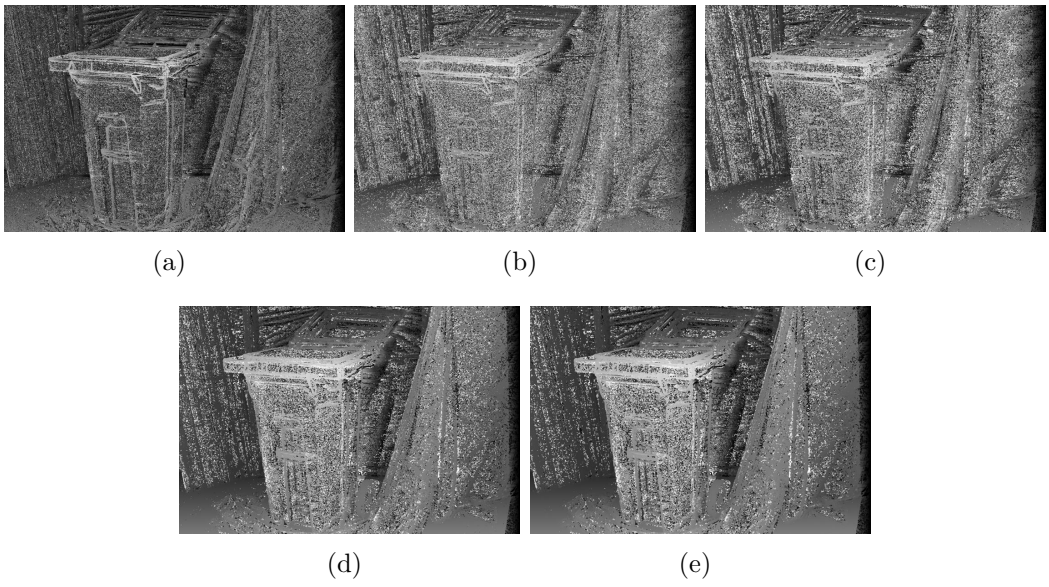
Figure 34: Playroom. (a) 7x7 CT + 1x1 search window. (b) 7x7 RT + 5x5 search window. (c) 7x7 RT + 7x7 search window. (d) 7x7 CRT + 5x5 search window. (e) 7x7 CRT + 7x7 search window.



(a) (b) (c)

(d) (e)

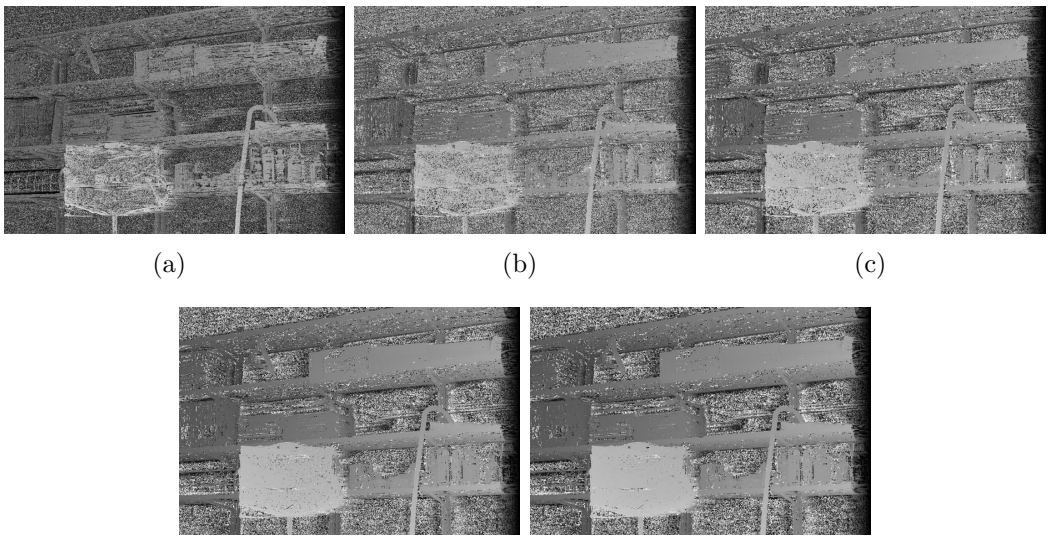
Figure 35: Playtable. (a) 7x7 CT + 1x1 search window. (b) 7x7 RT + 5x5 search window. (c) 7x7 RT + 7x7 search window. (d) 7x7 CRT + 5x5 search window. (e) 7x7 CRT + 7x7 search window.



(a) (b) (c)

(d) (e)

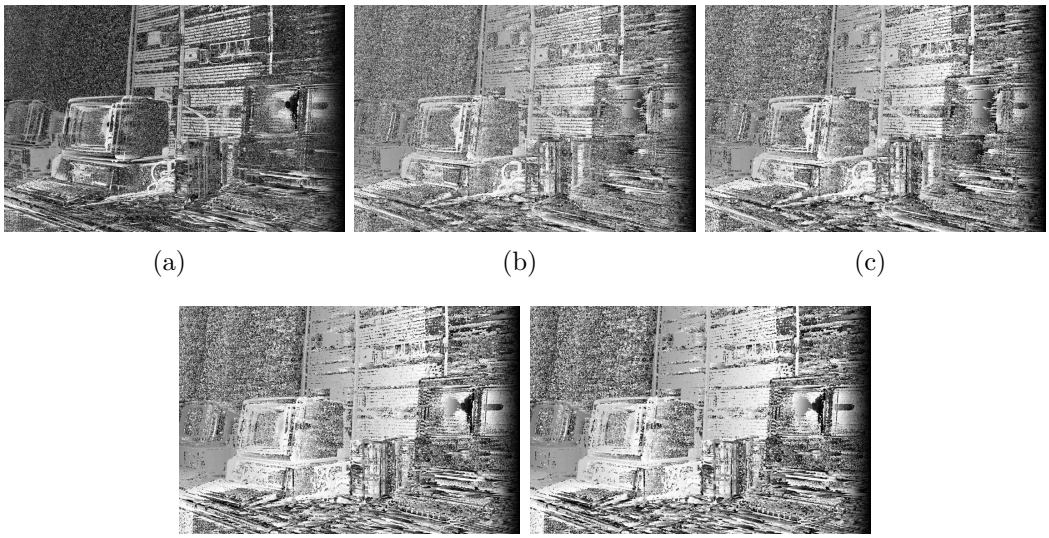
Figure 36: Recycle. (a) 7x7 CT + 1x1 search window. (b) 7x7 RT + 5x5 search window. (c) 7x7 RT + 7x7 search window. (d) 7x7 CRT + 5x5 search window. (e) 7x7 CRT + 7x7 search window.



(a) (b) (c)

(d) (e)

Figure 37: Shelves. (a) 7x7 CT + 1x1 search window. (b) 7x7 RT + 5x5 search window. (c) 7x7 RT + 7x7 search window. (d) 7x7 CRT + 5x5 search window. (e) 7x7 CRT + 7x7 search window.



(a) (b) (c)

(d) (e)

Figure 38: Vintage. (a) 7x7 CT + 1x1 search window. (b) 7x7 RT + 5x5 search window. (c) 7x7 RT + 7x7 search window. (d) 7x7 CRT + 5x5 search window. (e) 7x7 CRT + 7x7 search window.

Table 15: Different quality scores of the generated DMs using local block matching algorithm compared to orginal ground truth DMs.

Preprocessing transform	Search window	Median filter	PSNR (dB)	RMSE	*Accuracy (%)	**Accuracy (%)
7 x 7 RT	5 x 5	No	9.93	103.47	26.05	29.75
7 x 7 RT	7 x 7	No	10.27	100.13	31.36	34.91
7 x 7 CRT	5 x 5	No	11.30	90.88	41.26	45.21
7 x 7 CRT	7 x 7	No	11.57	88.75	44.13	48.00
7 x 7 CT	1 x 1	No	9.48	110.52	20.46	24.43
7 x 7 RT	5 x 5	Yes	10.82	96.09	28.89	32.98
7 x 7 RT	7 x 7	Yes	10.84	95.54	33.12	36.88
7 x 7 CRT	5 x 5	Yes	11.57	88.92	42.13	46.19
7 x 7 RT	7 x 7	Yes	11.75	87.44	44.67	48.63
7 x 7 CT	1 x 1	Yes	10.83	99.54	26.10	30.74

* Considers absolute disparity differences that are larger than 5 as incorrect.
** Considers absolute disparity differences that are larger than 10 as incorrect.

REFERENCES

[1] M. Ghanim, O. Tasdizen, and H. F. Ugurdag, "An efficient algorithm for disparity map compression based on spatial correlations and its low-cost hardware architecture." Manuscript submitted for publication, Oct., 2021.

[2] N. Qian, "Binocular disparity and the perception of depth," *Neuron*, vol. 18, no. 3, pp. 359–368, 1997. https://doi.org/10.1016/S0896-6273(00)81238-6.

[3] I. Hamzaoglu, O. Tasdizen, and E. Sahin, "An efficient H.264 intra frame coder system," *IEEE Trans. Consum. Electron.*, vol. 54, no. 4, pp. 1903–1911, 2008. https://doi.org/10.1109/TCE.2008.4711252.

[4] O. Tasdizen, A. Akin, H. Kukner, and I. Hamzaoglu, "Dynamically variable step search motion estimation algorithm and a dynamically reconfigurable hardware for its implementation," *IEEE Trans. Consum. Electron.*, vol. 55, no. 3, pp. 1645–1653, 2009. https://doi.org/10.1109/TCE.2009.5278038.

[5] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, pp. 151–158, 1994. https://doi.org/10.1007/bfb0028345.

[6] O. Demetz, D. Hafner, and J. Weickert, "The complete rank transform: A tool for accurate and morphologically invariant matching of structures.," in *Proc. Br. Mach. Vis. Conf. (BMVC)*, pp. 50.1–50.12, 2013. https://doi.org/10.5244/C.27.50.

[7] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, 2007. https://doi.org/10.1109/TPAMI.2007.1166.

[8] I. J. Cox, S. L. Hingorani, S. B. Rao, and B. M. Maggs, "A maximum likelihood stereo algorithm," *Comput. Vis. Image Underst.*, vol. 63, no. 3, pp. 542–567, 1996. https://doi.org/10.1006/cviu.1996.0040.

[9] Z. Xie, S. Chen, and G. Orchard, "Event-based stereo depth estimation using belief propagation," *Front. Neurosci.*, vol. 11, p. 535, 2017. https://doi.org/10.3389/fnins.2017.00535.

[10] G. Tech, Y. Chen, K. Müller, J. R. Ohm, A. Vetro, and Y. K. Wang, "Overview of the multiview and 3D extensions of high efficiency video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 35–49, 2015. https://doi.org/10.1109/TCSVT.2015.2477935.

[11] M. Saldanha, G. Sanchez, C. Marcon, and L. Agostini, "Fast 3D-HEVC depth map encoding using machine learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 3, pp. 850–861, 2019. https://doi.org/10.1109/TCSVT.2019.2898122.

[12] Q. Zhang, M. Chen, X. Huang, N. Li, and Y. Gan, "Low-complexity depth map compression in HEVC-based 3D video coding," *EURASIP J. Image Video Process.*, vol. 2015, no. 1, pp. 1–14, 2015. https://doi.org/10.1186/s13640-015-0058-5.

[13] H. Hamout and A. Elyousfi, "Fast depth map intra coding for 3D video compression-based tensor feature extraction and data analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 7, pp. 1933–1945, 2019. https://doi.org/10.1109/TCSVT.2019.2918770.

[14] S. Wang, L. Yu, and S. Xiang, "A low complexity compressed sensing-based codec for consumer depth video sensors," *IEEE Trans. Consum. Electron.*, vol. 65, no. 4, pp. 434–443, 2019. https://doi.org/10.1109/TCE.2019.2929586.

[15] L. F. Lucas, K. Wegner, N. M. Rodrigues, C. L. Pagliari, E. A. da Silva, and S. M. de Faria, "Intra-predictive depth map coding using flexible block partitioning," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4055–4068, 2015. https://doi.org/10.1109/TIP.2015.2456509.

[16] S. Lee and A. Ortega, "Adaptive compressed sensing for depthmap compression using graph-based transform," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, pp. 929–932, 2012. https://doi.org/10.1109/ICIP.2012.6467013.

[17] I. Tabus and P. Astola, "Sparse prediction for compression of stereo color images conditional on constant disparity patches," in *Proc. True Vision—Capture, Transmiss. Display 3D Video. (3DTV) Conf.*, pp. 1–4, 2014. https://doi.org/10.1109/3DTV.2014.6874766.

[18] M. Zamarin and S. Forchhammer, "Lossless compression of stereo disparity maps for 3D," in *Proc. IEEE Int. Conf. Multimed. Expo Work. (ICMEW)*, pp. 617–622, 2012. https://doi.org/10.1109/ICMEW.2012.113.

[19] M. J. Weinberger, G. Seroussi, and G. Sapiro, "The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS," *IEEE Trans. Image Process.*, vol. 9, no. 8, pp. 1309–1324, 2000. https://doi.org/10.1109/83.855427.

[20] O. Palaz, H. F. Ugurdag, O. Ozkurt, B. Kertmen, and F. Donmez, "RImCom: raster-order image compressor for embedded video applications," *J. Signal Process. Syst.*, vol. 88, no. 2, pp. 149–165, 2017. https://doi.org/10.1007/s11265-016-1211-9.

[21] M. E. Papadonikolakis, A. P. Kakarountas, and C. E. Goutis, "Efficient high-performance implementation of JPEG-LS encoder," *J. Real-Time Image Process.*, vol. 3, no. 4, pp. 303–310, 2008. https://dx.doi.org/10.1007/s11554-008-0088-7.

[22] T. Inatsuki, M. Matsuura, K. Morinaga, H. Tsutsui, and Y. Miyanaga, "An FPGA implementation of low-latency video transmission system using lossless

and near-lossless line-based compression," in *Proc. IEEE Int. Conf. Digit. Signal Process. (DSP)*, pp. 1062–1066, 2015. https://doi.org/10.1109/ICDSP.2015.7252041.

[23] K. Mühlmann, D. Maier, J. Hesser, and R. Männer, "Calculating dense disparity maps from color stereo images, an efficient implementation," in *Proc. IEEE Work. Stereo Multi-Baseline Vision (SMBV)*, vol. 47, pp. 30–36, 2001. https://doi.org/10.1109/SMBV.2001.988760.

[24] D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nešić, X. Wang, and P. Westling, "High-resolution stereo datasets with subpixel-accurate ground truth," in *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, pp. 31–42, 2014. https://doi.org/10.1007/978-3-319-11752-2_3.

[25] M. Grabner, "jpeg_encode.zip," *MATLAB File Exchange Center*, 2021. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/46424-jpegls_encode-zip. Accessed: Nov. 20, 2021.

[26] S. Garg, "Lossless compression software for camera raw images." [Online]. Available: http://imagecompression.info/lossless/, 2015. Accessed: Nov. 20, 2021.

[27] Xilinx, "Xilinx synthesis technology (XST) user guide." [Online]. Available: https://shorturl.at/glpsH, n.d. Accessed: Nov. 20, 2021.

[28] T. Schmitz, "The rise of serial memory and the future of DDR," *Xilinx UltraScale Devices. White Pap.*, pp. 1–9, 2015. [Online]. Available: http://www.xilinx.com/support/documentation/white_papers/wp456-DDR-serial-mem.pdf. Accessed: Nov. 20, 2021.

[29] H. Li and K. N. Ngan, "Learning to extract focused objects from low dof images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 11, pp. 1571–1580, 2011. https://doi.org/10.1109/TCSVT.2011.2129150.