SPARSE DISPARITY ESTIMATION USING GLOBAL PHASE ONLY CORRELATION FOR STEREO MATCHING ACCELERATION

Takeshi Shimada, Masayuki Ikebe, Prasoon Ambalathankandy
Shinya Takamaeda-Yamazaki, Masato Motomura, and Tetsuya Asai

Graduate School of Information Science and Technology, Hokkaido University
Kita 14, Nishi 9, Kita-ku, Sapporo, 060-0814, Japan

ABSTRACT

In this study, we propose an efficient stereo matching method which estimates sparse disparities using global phase only correlation (POC). Conventionally, cost functions are to be calculated for all disparity candidates and the associated computational cost has been impediment in achieving a real-time performance. Therefore, we consider to use full image 2D phase only correlation (FIPOC) for detecting the valid disparities candidates. This would require comparatively fewer calculations for the same number of disparities. Since, the FIPOC output indicates the disparity distribution of two stereo images, we can sort the disparity candidates and choose them for sparse calculation. In our proposed method, the searchable disparity range is half of the input image size, which is much wider than that of the conventional methods. When we apply the FIPOC to naive sum of absolute difference (SAD) stereo matching method, the combined algorithm would require fewer calculations while maintaining the same accuracy. In our evaluation, the proposed method achieves 194 disparity stereo matching in 70 ms on 398 × 288 images without the need for SIMD instruction, multi-thread operation, or additional hardware while using a Intel Core i5-5257U.

Index Terms— stereo matching, disparity, POC, sparse search

1. INTRODUCTION

The field of stereo vision has gathered much attention in the last few decades. Particularly, the recent demand for lightweight stereo matching has increased due to the use of dual-cameras in smartphones and automobile onboard camera. Stereo vision requires an operation that estimates binocular disparity among objects or background for obtaining the depth information. To execute this operation, the stereo matching algorithm calculates horizontal distance between corresponding pixels in the stereo images, which is equal to the disparity. The disparity map calculation can be grouped as local methods and global methods. The local methods calculate the matching cost of local patch by template matching.The following three methods are mainly called as traditional local matching methods; the sum of absolute difference (SAD) , the sum of squared difference (SSD), and normalized cross correlation (NCC). These local matching calculations are comparatively faster than the global methods. Surveying the open literature shows that lot of research have been conducted in stereo matching to seek fast and accurate technique [1–6]. Recently, Convolutional Neural Network matching methods [5, 6] have been used and they have achieved more accurate results with GPU implementation.

On the other hand, the global method uses all the pixels in the given image to obtain global features; graph cut [7] and belief propagation [8] are listed as fundamental global methods. In these algorithms, pixels are considered as graph node, and the object connectivity calculation is executed as solving energy minimization problem. Additionally, cost aggregation can be categorized to semi-global methods, which is executed after the local calculations to optimize the disparity map by computing energy of the limited connectivity [9]. Most of the stereo matching methods take huge time; even naive SAD takes few hundred milliseconds with CPU only calculation.

To reduce the amount of calculations, block matching (BM) optimization methods have been proposed in the past [10–14]. Summed Normalized Cross-Correlation method (SNCC) [11] uses a box filter for effective acceleration, however, this results in shrinking of smaller objects. Profile shape matching [12] can also reduce the BM calculation with accuracy degradation. Recently, some methods using 1D-POC [15, 16] have been proposed for disparity optimization [15]. 1D-POC can detect disparity of one or two objects at a sub-pixel level, and it is unable to detect multiple objects. Thus, the accuracy of 1D-POC drops when compared to SAD in a usual stereo image. Yang et al. [16] use the POC for maximum and minimum disparity detection in cost aggregation step, however, it requires about 61.77 seconds for processing.

Generally, in stereo matching algorithms, disparity candidates are to be calculated. And, the number of calculations depends on these disparity numbers. Thus, the order of arithmetic operation becomes $O(N \times D)$ (N is a number of pixels and D is the disparity range.) To overcome the limitation based on the dependency, we propose a valid disparirty selec-

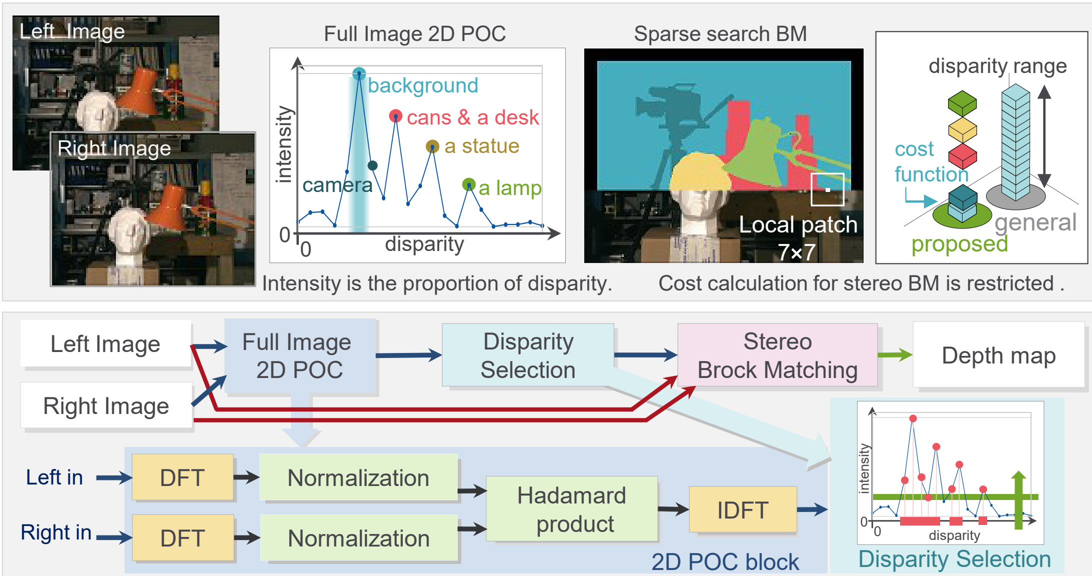


Fig. 1. An outline of our method

tion using global POC. This process can detect the disparity proportion of stereo images before local matching and achieves the reduction of calculation of the disparity estimation while maintaining the accuracy. Our selection technique can be applied to all local and semi-global methods, in this study, we have evaluated the combination using naive SAD.

Our contributions can be summarized as follows:

- Obtaining the global feature of disparities using full-image 2D POC (FIPOC).
- Sparse block matching based on valid disparities selection from FIPOC.
- High-efficiency calculation even when combined with naive SAD.

2. PROPOSED METHOD

In this section, we will discuss in detail the Full-Image 2D Phase Only Correlation (FIPOC) method which searches for valid disparity candidates in stereo vision. Figure1 shows the processing scheme of our method. It consists of three main blocks which are the FIPOC, the disparity selection, and SAD BM. In the FIPOC unit, DFT and IDFT calculations are computed for obtaining the global disparity profile; the disparity selection unit determines the valid disparities in descending order. The SAD BM unit executes sparse BM operation using 7×7 pixel blocks based on selected disparities. In the following subsections, we describe our proposed method.

2.1. Full Image Phase Only Correlation

Phase Only Correlation (POC) [17] was used to determine the degree of similarity between two local patches. And by using POC, an accurate sub-pixel stereo matching method was proposed in [4]. The basic concept of POC function can be described as follows: Given two input images $f(x, y)$ and $g(x, y)$,of size $(2N_1 + 1, 2N_2 + 1)$. The variables x and y are set so that the center is always at the origin (0). In other words, ranges of variables are $-N_1 \geq x \geq N_1$ and $-N_2 \geq y \geq N_2$. POC operation starts with a 2D Discrete Fourier Transform (2D-DFT). The output complex functions are the wavenumber space represented by $F(k_x, k_y)$ and $G(k_x, k_y)$. And they are given by Eq.(1).

$$F(k_x, k_y) = DFT(f(x, y)) \quad G(k_x, k_y) = DFT(g(x, y)) \tag{1}$$

The range of k_x, k_y is determined by DFT as follows: $-N_1 \geq k_x \geq N_1$,$-N_2 \geq k_y \geq N_2$. As they are complex functions, they comprise amplitude and phase components. In order to extract the phase component, the functions are divided by the amplitude. Then a Hadamard product is calculated at each wave number. Since they are complex functions, one must be conjugate. We obtain normalized cross phase spectrum $\hat{C}(k_x, k_y)$ using (Eq.(2)).

$$\begin{aligned}\hat{C}(k_x, k_y) &= \frac{F(k_x, k_y)\overline{G(k_x, k_y)}}{|F(k_x, k_y)||\overline{G(k_x, k_y)}|} \\ &= \exp\big(j(\theta_F(k_x, k_y) - \theta_G(k_x, k_y))\big)\end{aligned} \tag{2}$$

In Eq.(2) $\overline{G(k_x, k_y)}$ represents the complex conjugate of the function. The output of POC $\hat{c}(x, y)$ is obtained by performing discrete inverse Fourier transform (IDFT) on $\hat{C}(k_x, k_y)$.

$$\hat{c}(x, y) = IDFT(\hat{C}(k_x, k_y)) \tag{3}$$

(a) Tsukuba (b) Venus (c) Cones (d) Teddy

Fig. 2. Middlebury 4 stereo images [18]

.

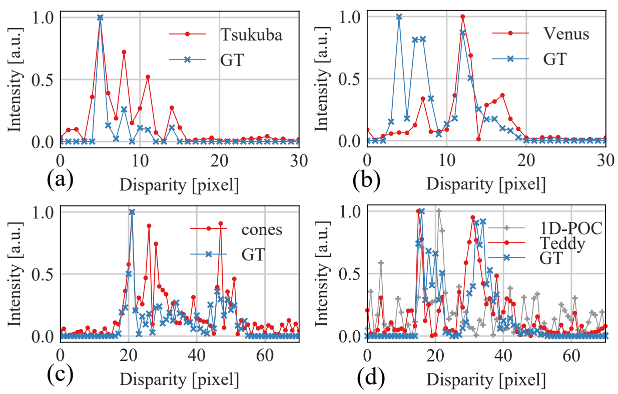


Fig. 3. Comparison with FIPOC (red), true disparity distribution (blue) and 1D-POC result (gray).

For simplicity, the processing from Eq.(1) to Eq.(3) is represented using the POC function.

$$\hat{c}(x, y) = POC(f(x, y), g(x, y)) \tag{4}$$

When the image $f(x, y)$ is described by translation of the image $g(x, y)$, the output of the POC function is a delta function with the same parallel mobility.

$$\begin{aligned} f(x, y) &= g(x - \alpha, y - \beta)) \\ POC(f(x, y), g(x, y)) &= \delta(x - \alpha, y - \beta) \end{aligned} \tag{5}$$

We propose Full Image 2D POC for stereo images by extending POC. We assume a model that is linearly independent for each object in stereo images. This assumption can be written mathematically as Eq.(6).

$$\begin{aligned} L_{full}(x, y) = \sum_{i=0}^{N} L_i(x, y) \quad R_{full}(x, y) = \sum_{i=0}^{N} R_i(x, y) \\ L_i(x, y) = R_i(x - d_i, y) \qquad d_i : \text{disparity of object.} \end{aligned} \tag{6}$$

We will now perform a POC processing with full image as the input window for POC. In this way, we can obtain an independent delta function for each object (Eq.(7)).

$$POC(L(x, y), R(x, y)) = \sum_{i=0}^{N} p_i \delta(x - d_i, y) \tag{7}$$

The height of each delta function is represented by p_i. It is worth noting that the value of p_i depends on the amount of phase in object and not on the objects area size.

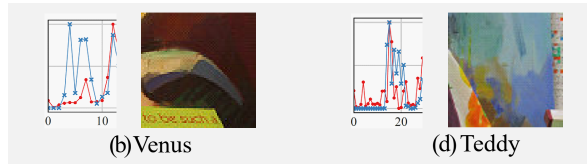


Fig. 4. Discussion points of output.

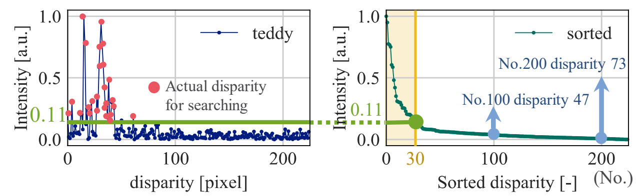


Fig. 5. Disparity sorting method of our algorithm.

We now analyse our FIPOC method's performance using the Middlebury 4 stereo images (Fig.2) [18]. The FIPOC results and ground truths are shown in Fig.3. The 1D-POC result has a larger noise floor. Characteristic of the FIPOC function, as expected, the results do indicate that the disparity distribution of stereo images and its disparity range are half of the image size. However, we confirm that there are some unmatched distributions between the ground truths and the FIPOC results. Figure 4 shows the distributions and corresponding image regions. In FIPOC, the disparity distribution is detected by the phase of the images. In this case, since low frequency components have less edges which are equal to the phases, height of the POC results are lower than the ground truth. And, disparity of the low frequency components can be detected as values which are much larger than the noise level.

2.2. Sparse Disparity Searching Stereo Matching Method

Using FIPOC results, we realize valid disparities selection. Figure 5 (See also Fig.1) shows the selection scheme of our sparse block matching. First, disparity sorting is performed based on the intensity of obtained FIPOC result. In our method, since each intensity reflects the disparity distribution, the disparity with less intensity seems to be excluded from the valid disparities group. Thus, we are able to reduce the number of block matching. We choose SAD as calculation cost function of our stereo matching, because the SAD is a benchmark for stereo matching methods. And, it is worth noting that our disparity selection can be applied to other stereo matching methods.

3. EXPERIMENT

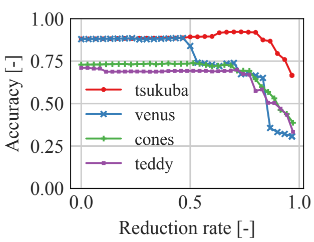
Our experimental evaluation has two objectives. First, we evaluate and verify how the accuracy and speed are affected by our disparity reduction method (Fig.6). Second, we compare

Table 1. Comparison with other real time stereo algorithms based on research of B.Tippetts [20] published in 2016

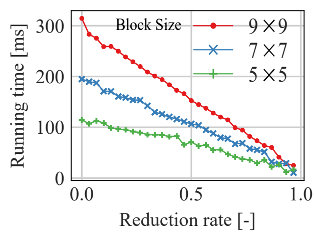
Alg	t(ms)	W×H (disp)	Mde/s	CLK	normalized Mde/s★	Avg.Acc(%)	Tsukuba(%)
RTCensus [14]	77.6	450×375(60)	130.5*	2.0(GHz)	69.60	92.00	93.75
SADLR [13]	109.6	512 × 512(48)	114.8*	3.2(GHz)	38.27	N/A	N/A
ProfShape [12]	16	384×288(16)	110.5	2.8(GHz)	126.29	78.04	90.42
RTDP [19]	18	384×288(16)	98.3*	2.8(GHz)	27.44	N/A	N/A
SNCC [11]	140	450 ×375(60)	77.1	3.0(GHz)	82.24	93.01	93.92
Distinct SAD [10]	25	320 ×240(16)	49.1	800(MHz)	196.4	N/A	90.68
Naive SAD	180	384×288(30)	18.4	2.7(GHz)	21.81	77.26	86.06
Ours+SAD	70	384×288(192)	**302.8**	2.7(GHz)	**358.87**	76.69	90.42
Ours+SAD+WM	175.6	384×288(192)	120.9	2.7(GHz)	**143.29**	81.82	93.99

* Asterisk indicates using SIMD operation. According to B.Tippetts report [20] , SIMD instruction achieves $3-4\times$ speed up.

★ Mde/s was normalized taking CLK and SIMD into consideration. Concretely, the clock speed were normalized to 3.2GHz. If using SIMD, Mde/s was divided by 3. Our Mde/s is measured using tsukuba image. Kernel size is 7. 192 disparity candidates decreased to 10 by using FIPOC. Avg.ACC(%) is average accuracy of 4 images.Tsukuba(%) is Tsukuba accuracy. In this table, bad threshold is 1 pixel, and all region is measured.



(a)



(b)

Fig. 6. (a)Accuracy vs Reduction rate. (b) Running time vs Reduction rate. The reduction rate is the ratio of the number reduced from the first valid disparities. First valid disparities are 30 in Tsukuba and Venus image and 70 in Cones and Teddy image. The accuracy is measured by 4 stereo images in bad threshold 2 pixels, and all region. Running time is measured in Tsukuba image.

our implementation with other real-time stereo vision systems (Table 1, Fig.7).

Our method was implemented in C++ on CPU. The code was executed on Intel Core i5-5257U (2.70GHz) CPU without any multicore, multithread and SIMD operations.

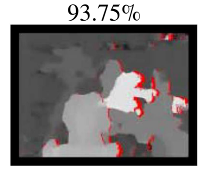
Figure 6 shows our first experiment and results are presented in terms of accuracy and speed. We observe that our disparity reduction technique preserves the accuracy and accelerates the matching. In terms of speed, given that our FIPOC method requires 4 ms, and this is much less than the block matching operation time.

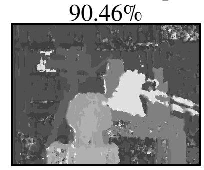
We compared our stereo matching method with other real-time implementation in terms of their efficiency and accuracy. We used two methods: ours + SAD and ours + SAD + weighted median filter (WMF) [21]. WMF was used for post filtering. In efficiency, an index of Mde/s is used in real-time implementation. Mde/s stands for millions of disparity


(a) RTCensus 93.75%


(b) Profile shape 90.46%


(c) SNCC 93.92%


(d) DistinctSAD 90.68%


(e) Ours+SAD 90.42%

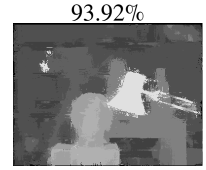

(f) Ours+SAD+WM 93.99%

Fig. 7. The output results of Tsukuba image.

evaluations per second and is described as follows:

$$Mde/s = \frac{W \times H \times D}{t \times 10^6} \tag{8}$$

Table 1 shows that ours+SAD has highest Mde/s in real-time methods in Mde/s and in normalized Mde/s, because our technique has wide disparity detection. In accuracy comparison of 4 images, our result has less accuracy, because our system accuracy depends on the SAD matching accuracy. However, in Tsukuba image comparison (Fig.7), our + SAD + WM output is comparable to other accurate method.

4. CONCLUSION

In this paper, we introduce sparse disparity search method for stereo block matching. Our full image 2D POC is able to search possible disparity widely and quickly. Our stereo matching method is more efficient than other real time methods.

Acknowledgement This work was partly supported by Toshiaki Yoshida, CTO at Acutelogic Corporation.

5. REFERENCES

[1] Zheng Gu, Xianyu Su, Yuankun Liu, and Qican Zhang, "Local stereo matching with adaptive support-weight, rank transform and disparity calibration," *Pattern Recognition Letters*, vol. 29, no. 9, pp. 1230 – 1235, 2008.

[2] L. Ma, J. Li, J. Ma, and H. Zhang, "A modified census transform based on the neighborhood information for stereo matching algorithm," in *2013 Seventh International Conference on Image and Graphics*, July 2013, pp. 533–538.

[3] Kenji Takita, Takafumi Aoki, Yoshifumi Sasaki, Tatsuo Higuchi, and Koji Kobayashi, "High-accuracy subpixel image registration based on phase-only correlation," *IEICE transactions on fundamentals of electronics, communications and computer sciences*, vol. 86, no. 8, pp. 1925–1934, 2003.

[4] Kenji Takita, Takafumi Aoki, Yoshifumi Sasaki, Tatsuo Higuchi, and Koji Kobayashi, "High-accuracy subpixel image registration based on phase-only correlation," *IEICE transactions on fundamentals of electronics, communications and computer sciences*, vol. 86, no. 8, pp. 1925–1934, 2003.

[5] Jure Zbontar and Yann LeCun, "Stereo matching by training a convolutional neural network to compare image patches," *Journal of Machine Learning Research*, vol. 17, no. 1-32, pp. 2, 2016.

[6] Sergey Zagoruyko and Nikos Komodakis, "Learning to compare image patches via convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4353–4361.

[7] Vladimir Kolmogorov and Ramin Zabin, "What energy functions can be minimized via graph cuts?," *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, no. 2, pp. 147–159, 2004.

[8] Jonathan S Yedidia, William T Freeman, and Yair Weiss, "Generalized belief propagation," in *Advances in neural information processing systems*, 2001, pp. 689–695.

[9] Heiko Hirschmuller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. IEEE, 2005, vol. 2, pp. 807–814.

[10] Luigi Di Stefano, Massimiliano Marchionni, and Stefano Mattoccia, "A fast area-based stereo matching algorithm," *Image and Vision Computing*, vol. 22, no. 12, pp. 983 – 1005, 2004, Proceedings from the 15th International Conference on Vision Interface.

[11] Nils Einecke and Julian Eggert, "A two-stage correlation method for stereoscopic depth estimation," in *Digital Image Computing: Techniques and Applications (DICTA), 2010 International Conference on*. IEEE, 2010, pp. 227–234.

[12] Beau J Tippetts, Dah-Jye Lee, James K Archibald, and Kirt D Lillywhite, "Dense disparity real-time stereo vision algorithm for resource-limited systems," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 10, pp. 1547–1555, 2011.

[13] Wannes Van Der Mark and Dariu M Gavrila, "Real-time dense stereo for intelligent vehicles," *IEEE Transactions on intelligent transportation systems*, vol. 7, no. 1, pp. 38–50, 2006.

[14] Martin Humenberger, Christian Zinner, Michael Weber, Wilfried Kubinger, and Markus 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, Special issue on Embedded Vision.

[15] Alfonso Alba, Edgar Arce-Santana, Ruth M Aguilar-Ponce, and Daniel U Campos-Delgado, "Phase-correlation guided area matching for realtime vision and video encoding," *Journal of real-time image processing*, vol. 9, no. 4, pp. 621–633, 2014.

[16] Xiao Yang, Xiaobo Chen, and Juntong Xi, "Block based dense stereo matching using adaptive cost aggregation and limited disparity estimation," in *Instrumentation and Measurement Technology Conference (I2MTC), 2017 IEEE International*. IEEE, 2017, pp. 1–6.

[17] Joseph L Horner and Peter D Gianino, "Phase-only matched filtering," *Applied optics*, vol. 23, no. 6, pp. 812–816, 1984.

[18] ," vision.middlebury.edu/stereo.

[19] Sven Forstmann, Yutaka Kanou, Jun Ohya, Sven Thuering, and Alfred Schmitt, "Real-time stereo by using dynamic programming," in *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW'04. Conference on*. IEEE, 2004, pp. 29–29.

[20] Beau Tippetts, Dah Jye Lee, Kirt Lillywhite, and James Archibald, "Review of stereo vision algorithms and their suitability for resource-limited systems," *Journal of Real-Time Image Processing*, vol. 11, no. 1, pp. 5–25, 2016.

[21] Qi Zhang, Li Xu, and Jiaya Jia, "100+ times faster weighted median filter (wmf)," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2830–2837.