

FPGA Implementation for Binocular Stereo Matching Algorithm Based on Sobel Operator

Jianhui Han¹, Zhen Wu², Lanying Li³ and Ying Ji⁴

1. The college of Computer Science and Technology,
Harbin University of Science and Technology, Harbin 150080, China

E-mail: ¹363992791@qq.com

²wuzhen_1988@163.com

Abstract

Aiming at the low accuracy of stereo matching algorithm caused by the larger gray change, an improved algorithm is proposed. Sobel operator is used to compute the gradient of pixels. Based on the gradient histogram, the adaptive thresholds are derived and the support window is determined automatically. The neighbor pixels are compared with the average of all pixels in the window instead of the center pixel to complete the census transform. The disparity image is obtained through searching the best match point in the left and right images. The hardware implementation with parallel processing is elaborately designed to improve the capability of the large data processing and computational efficiency based on FPGA. The experimental results show that the improved scheme and hardware structure can obtain disparity map with higher accuracy and stronger robustness.

Keywords: stereo matching; Sobel operator; Census transform; FPGA

1. Introduction

Binocular stereo vision is a hot issue in the field of computer vision. Stereo matching is the important segment to realize stereo vision. The accuracy of stereo matching is the bottleneck problem to restrict the development of stereo vision technology. Stereo matching algorithm can be roughly divided into global matching algorithm and partial matching algorithm. Region matching algorithm can obtain disparity information directly and has low complexity [1], therefore be widely used in the real-time stereo vision system.

Region matching algorithm is largely dependent on the establishment of matching window and the selection of similarity measurement function. The regional segmentation algorithm can obtain the better effect by selecting the shape and size of matching window dynamically. But this algorithm can't obtain better effect for complex texture images [2]. The similarity measure functions include sum of absolute differences (SAD), sum of squared differences (SSD), Rank transform, Census transform [3]. The SSD and SDA algorithm are easily affected by the uneven illumination and occlusion to cause the matching error. Rank transform algorithm is similar to filter and defines the grayscale through the matching points and gray differential in the characteristic window [4]. Zabih [5] converted level transformation algorithm to Census transform.

Although these algorithms have been able to greatly improve the performance of the region matching, these algorithms will appear parallax smooth transition phenomenon when gray level changes largely. Their complexity is difficult to meet the requirements of high-speed and real-time. This paper proposes an improved algorithm that combines with edge detection, Sobel gradient operator, gradient amplitude histogram and the improved

Census transform. In order to implement the algorithm, the paper develops the dedicated hardware parallel processing system with the aid of the FPGA [6].

2. Description of Algorithm

This paper uses the edge detection method to identify the area where gray level changes largely. Firstly, the Sobel operator is used to calculate the image pixel gradient. Secondly the threshold calculation method is obtained by analysing the gradient amplitude histogram. The size of window is determined by dynamic threshold. Lastly, this paper takes advantage of the improved Census transform to complete stereo matching.

2.1. Calculation of Gradient Amplitude

The change of pixel will impact its gradient amplitude, so the areas that gray level changes largely can be determined by calculating the gradient amplitude. Sobel is a derivative edge detection operator. It is a kind of effective calculation method of the gradient. Each pixel in the image will do the convolution operation with 3*3 template [7]. As shown in Figure 1, one affects the horizontal direction, another affects the vertical direction.

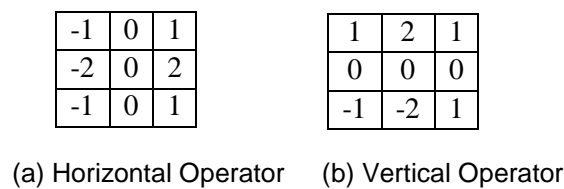


Figure 1. Sobel Operator

When sobel operator is used to calculate $P(x, y)$'s gradient, the pixel need to be regarded as the center to establish a 3*3 window. All pixels within the window need to do the convolution operation with two nuclear to obtain horizontal component G_x and vertical component G_y , respectively. As formula (1) - (4):

$$G_x = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} * \begin{pmatrix} P1 & P2 & P3 \\ P4 & P(x, y) & P5 \\ P6 & P7 & P9 \end{pmatrix} \quad (1)$$

$$G_x = -P1 + P3 - 2P4 + 2P5 - P6 + P8 \quad (2)$$

$$G_y = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & 1 \end{pmatrix} * \begin{pmatrix} P1 & P2 & P3 \\ P4 & P(x, y) & P5 \\ P6 & P7 & P8 \end{pmatrix} \quad (3)$$

$$G_y = P1 + 2P2 + P3 - P6 - 2P7 - P8 \quad (4)$$

The pixel $P(x, y)$'s gradient amplitude can be computed through the two component, as formula (5).

$$G(x, y) = \sqrt{G_x^2 + G_y^2} \quad (5)$$

2.2. Adaptive Matching Window

The size of window is vital for stereo matching. If the window is too small, the information will be not enough to express the characteristics of the regional. If the window is too large, it will violate the hypothesis that the disparity value is consistent. It will cause that the edge disparity is too smooth to distinguish between objects and their background. At the same time, it reduces the speed. In order to improve the matching precision, the size of matching window need to change with the change of pixel gray level.

When the pixel's gradient changes severely, the image's gray level changes relatively severely. The change of gradient value is severely in disparity discontinuous regional. The size of matching window should be set relatively smaller. As for the low texture regions, the size should be set larger. This paper adopts double threshold value to adjust the size of the window dynamically: to select two gradient amplitude threshold, G_h is higher than G_l ; size of matching window is set D_1 、 D_2 、 $D_3(D_1 > D_2 > D_3)$.

This paper calculates the gradient magnitude of each pixel by sobel operator to establish gradient amplitude histogram. The abscissa is the gradient magnitude, and the ordinate is the number of pixel points corresponding to gradient magnitude. In order to reduce the difficulty and error of selecting threshold, before drawing the gradient magnitude histogram, all pixels are need for non maxima suppression processing. That is, in the 3*3 neighborhood of current pixel, if the pixel's gradient magnitude is greater than the two adjacent pixels' along the gradient direction, it marks as 1, otherwise as 0.

Based on the gradient histogram, the threshold selection problem is actually to find out the smooth area that locates between the peak in the area where gray changes smoothly and the first peak in the region where gray changes severely. In order to find the area, the gradient amplitude of the adjacent two points need to make difference. As formula (6).

$$dif(i) = |NMS(i+1) - NMS(i)|, i = 1, 2, \dots \quad (6)$$

$NMS(i)$ is the gradient magnitude histogram after non-maximum value suppression.

The high and low threshold are set as:

$$\begin{cases} G_h = \underset{i}{\text{Arg}}(dif(i) = 0) \\ G_l = G_h / 2, i = 1, 2, \dots \end{cases} \quad (7)$$

$\underset{i}{\text{Arg}}$ represents the first pixel i that meet the $dif(i) = 0$

2.3 Improved Algorithm of Census Transform

The accuracy of matching algorithm not only depends on the establishment of matching window, also relates closely to the similarity measure function. Census transform belongs to the nonparametric similarity factor. The basic idea is to use a rectangular window to traverse the image and put the gray value of the pixel in the center of the window as a reference value. Then the remaining pixels within the window are compared with the reference value successively. If the gray value of pixel is smaller than the reference value, it is recorded as 1, otherwise recorded as 0, and then outputs a string of binary code connected bit by bit [8].

$$T_{Census}(P) = \bigotimes_{q \in W(p)} \xi(I(p), I(q)) \quad (8)$$

In the formula, $W(p)$ is the matching window as the pixel P is center, $I(p)$ is the gray value of pixel P , $I(q)$ is the gray value except P in $W(p)$, \bigotimes is bitwise connection symbol. The size of $W(p)$ is determined by comparing the gradient magnitude with the gradient threshold value of the central pixel point. The function $\xi(x, y)$ defines as:

$$\xi(x, y) = \begin{cases} 1, x < y \\ 0, x \geq y \end{cases} \quad (9)$$

However, the traditional Census transform is over-reliance on the center pixel. When the image is disturbed, the distorted center pixel will increase the chance of mismatching. This paper uses the average of all the pixel values to replace the center pixel value, and converts according to the formula (8), (9). This enhances the anti-jamming ability of Census transform.

3. Hardware Implementation

This paper takes advantage of FPGA to complete the improved stereo matching algorithm. The RTL view of algorithm module is shown in Figure 2.

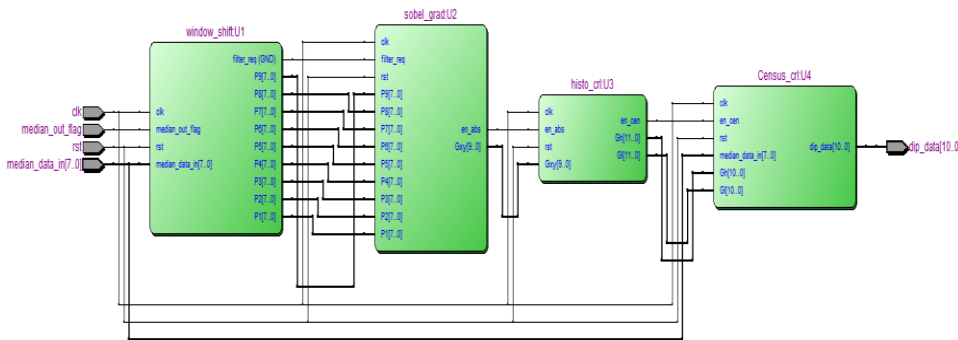


Figure 2. RTL View of Algorithm

3.1. The Design of Cache Window

The cache window is required to implement the improved algorithm. Because of the characteristics of FPGA programming, we cannot express a window of datas through an array. Therefore when using hardware to implement the algorithm, it requires a set of registers with the same size as the window to cache datas.

The paper uses the Quartus II software's shift register macro module (altshift_taps, based on RAM) to design the cache window. Altshift_taps can be configurable and have taps. Each tap outputs data at the specified location of the shift register chain. For example, for an image (256 × 256), altshift_taps module parameter is set as 8 inputs、8 outputs and 3 taps. Two adjacent taps are separated by 256 registers. The distance between two adjacent taps is set according to the image size. In engineering practice, the 3 × 3 window is often used. Shift_RAM contains two rows of data, then enters one array of data, so that you can get a 3 × 3 window data array. The shift register's simulation waveform is shown in Figure 3.

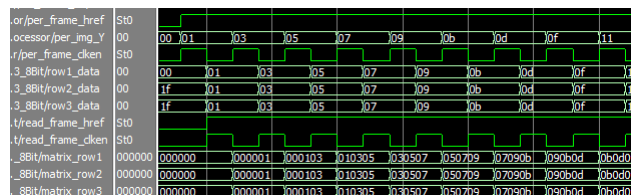


Figure 3. Simulation Waveform of Register

3.2 Sobel Gradient Operator

According to the gradient calculation formula, to calculate the gradient of a pixel needs to know about the each pixel's gray value of this point as the center of the 3*3 neighborhood. Then the horizontal and the vertical gradient operators carry out the addition, multiplication and square root arithmetic with the corresponding gradation.

The hardware implementation of Sobel gradient operator can be divided into two parts, namely image data buffer module and gradient binding module. The image data buffer module adopts Shift_Register to be achieved. The gradient combination module adopts the programmable macro module and Verilog HDL language to be achieved. Sobel gradient operator's simulation waveform is shown in Figure 4.

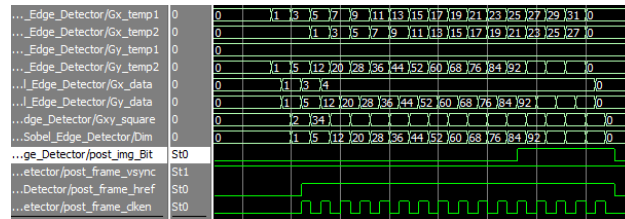


Figure 4. Simulation Waveform of Sobel Operator

3.3 Design of Adaptive Threshold

The selection of adaptive threshold is based on the gradient histogram. Therefore the image after the non maxima suppression needs to be carried out histogram statistics. By using the image (640*480*8 bit) as an example, the gradient values are concentrated between 0 and 200 by calculating. So it requires 200 registers, and each registers' width is 8 bit. These registers are used to store the number of different pixels' gradient value. Three gradient value is set as the address of these registers. To reset these registers at the beginning of each image. After entering the gradient value, the contents of register is taken out into the accumulator 1 through the address selector. Write into the register after plusing one until all pixel of the whole image finish statistics. The circuit of adptive threshold is shown in Figure 5.

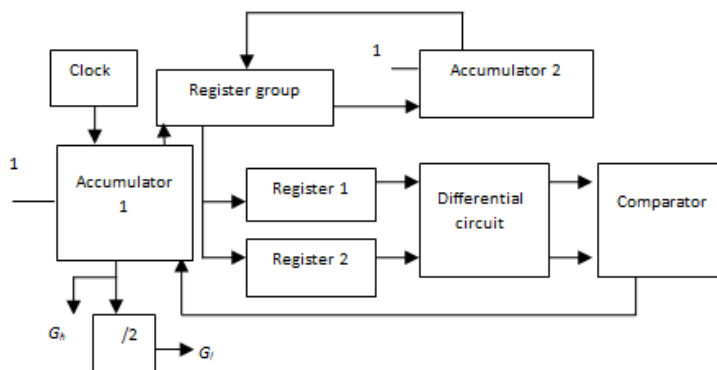


Figure 5. Circuit of Adaptive Threshold

When the image signal reach, the clock circuit will maintain the 200 system clock. During the effective period of the clock, the content of the accumulator 2 pluses one by itself since the arrival of each clock. Regard the cumulative results as register group's address, the content of the corresponding address register and content of the next address register are transferred into register 1 and register 2 respectively. The contents of register 1

and register 2 are difference according to equation (6). The comparator compares the difference results with 0. If it's 0, the comparator sends signal to stop accumulator 2. The value of accumulator 2 is the high threshold G_h , G_h is divided by 2 to obtain the low threshold G_l .

3.4 Stereo Matching

In order to improve the real-time performance of the system, the design of Census module adopts the parallel optimization method to resource exchange time. For example, there is a 7×7 local transform window. The shift register group expands to 8 lines, each column reads 8 pixels' data. When moving the window horizontally, the data of the census transform window updates once every eight clock cycles. Use the 128-bit comparator to compute two census transform windows' coding bit string (Code0, Code1). The two groups of codes are respectively obtained by Census transform. When the census transform window moving to the end of each line, it can obtain two adjacent line of pixels' value of Census transform. Therefore it needs to move the shift register group down to the two lines and restarts the next two rows of pixels' Census transform. It can be seen that after parallel input each list of data, namely after eight clock cycles, we can get two pixels' local census transform coding. However, originally it can get only one pixel's local census transform every after seven clock cycles. From the analysis of resources, the number of comparator increases by one times, but the amount of shift register only increased by $1/7$. Optimization results makes local census transform speed increase nearly one times, operation frequency/cycle (pixels) from $1/7$ up to $1/5$ (pixels/cycle). The module of Census transform is shown in Figure 6.

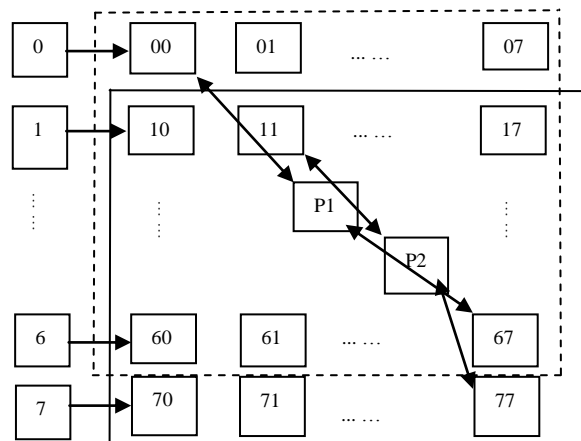


Figure 6. Module of Census Transform

4. Experimental Results Analysis

In this paper, a 4th-order low-distortion low-pass sigma-delta modulator using time sharing technology

In order to verify the validity of this algorithm, the experimental results will be uploaded to Middlebury website platform to be assessed [9-11]. This algorithm's supported windows were set up for $D_1=7$, $D_2=5$, $D_3=3$. In order to analyze the matching accuracy, we choose some similar algorithm to compare with this algorithm. The matching accuracy comparison results is shown in Table 1.

By comparison, ADCensus algorithm can achieve very low rate of mismatching, but the algorithm is complex. It needs a lot of processing. The complexity of the algorithm is higher than that of the algorithm proposed in this paper. This algorithm's mismatching

rate is lower than that of SSD + ASW and Census + ASW algorithm. By contrast, the algorithm proposed in this paper achieves better results in terms of accuracy and real-time performance.

Table 1. Results of Comparison

Threshold=1		ADCensus	Census+ASW	SSD+ASW	Proposed algorithm
Tsuku	occ	1.07	5.23	6.51	3.26
	all	1.48	7.07	8.43	3.96
	disc	5.73	24.1	19.7	21.8
Venus	occ	0.09	3.74	10.5	1.3
	all	0.25	5.16	12	1.57
	disc	1.15	11.9	32.7	31.3
Teddy	occ	4.1	16.5	15.7	6.02
	all	6.22	24.8	24.1	12.2
	disc	10.9	32.9	32.8	36.3
Cones	occ	2.41	10.6	14.1	3.06
	all	7.25	19.8	23.1	9.75
	disc	6.95	26.3	24.7	28.9

The algorithm proposed in this paper runs on both PC and FPGA to process the same image (640*480). The comparison result is shown in Table 2.

Table 2. Results of Comparison

Processing platform	Experimental environment	Processing time of one frame
PC	OpenCV	2.68s
FPGA	EP2C35F672	32.4ms

The matching results are compared by processing the Teddy image in Middlebury platform. The comparison results of SSD+ASW, Census+ASW and the proposed algorithm is shown in Figure 7 - 10.



Figure 7. Image of Teddy



Figure 8. Result of Proposed Algorithm



Figure 9. Result of Census + ASW

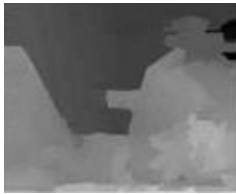


Figure 10. Result of SSD + ASW

It can be seen from Figure 8 that the disparity map of this algorithm is more clear than that of Census+ASW and SSD+ASW algorithm. And it expresses more information accurately.

5. Conclusion

This paper proposes an improved stereo matching algorithm based on Sobel gradient operator and implements this algorithm with the aid of the FPGA. The experimental results show that the matching effect of the algorithm can obtain the more dense disparity map. FPGA is helpful for improving the real-time property of the algorithm effectively. But the parallax edge is not accurate due to occlusion problems. These aspects will be improved in the subsequent research.

Acknowledgement

Supported by Scientific Research Fund of Heilongjiang Provincial Education Department (NO: 12531107).

References

- [1] L. Wang, S. B. Kang and H. Y. Shum, "Cooperative segmentation and stereo using perspective space search", Proceeding of Asian Conf on Computer Vision. Jeju Island, (2004), pp. 366-371.
- [2] V. Pratibha, "Enhanced stereo matching technique using image gradient for improved search time", International Journal of Computer Science Issues, vol. 8, no. 5, (2011), pp. 483-486.
- [3] L. Slimane, "FPGA implementation for stereo matching", Proceedings of 2013 and Information Conference, (2013), pp. 448-454.
- [4] B. Xiangzhi, Z. Fugen and B. Xue, "Fusion of infrared and visual images through region extraction by using multi scale center-surround top-hat transform", Optics Express, vol. 19, (2011), pp. 8444-8457.
- [5] K. Sunqil, "Near-real-time stereo matching method using temporal and spatial propagation of reliable disparity", Optical Engineering, (2014).
- [6] K. J. Yoon and I. S. Kweon, "Locally adaptive support-weight approach for visual correspondence search", Proceeding of IEEE Conf on Computer Vision and Pattern Recognition. San Diego, (2005), pp. 924-931.
- [7] C. Hua, R. E. Hong and Z. J. Qing, "The Extraction of Bamboo's Inner and Outer Contours Using Sobel Operator to Ant Colony Algorithm", Journal of Harbin University of Science and Technology, (2014), pp. 23-26.
- [8] T. P. Wu, S. K. Yeung and J. Jia, "Quasi-dense 3D reconstruction using tensor-based multiview stereo", Proceedings of the Conference on Computer Vision and Pattern Recognition, San Francisco, CA, (2010), pp. 1482-1489.

- [9] D. Scharstein, R. Szeliski and R. Zabih, "A taxonomy and evaluation of dense two frame stereo correspondence algorithms", Proceeding of IEEE Workshop on Stereo and Multi-Baseline Vision, **(2001)**, pp. 131-140.
- [10] W. Pei and W. Fang, "A local stereo matching algorithm based on region growing", Communications in Computer and Information Science, **(2012)**, pp. 459-464.
- [11] Y. M. Ji and Z. Zhen, "Research of WAP real-time image browsing platform based on DSP", Journal of Harbin University of Science and Technology, **(2010)**, pp. 34-39.

