Study on semi-global matching algorithm extended for multi baseline matching and parallel processing method based on GPU

Zhaohua Liu*, **Yuxia Yang**

School of Architectural and surveying & mapping engineering, Jiangxi University of Science and Technology, Ganzhou 341000 China

Received 22 January 2014, www.tsi.lv

Abstract

This paper extended semi-global matching algorithm into multi baseline matching to improve matching reliability, especially studies kernel function optimization strategies and GPU threads' executing scheme of matching cost cube computing and aggregating, and realized its fine granularity parallel processing based on GPU. The experiment results using three UCD aerial images based on Tesla C2050 GPU showed that MVLL's semi-global optimize algorithm can improve matching effectiveness and productiveness.

Keywords: semi-global matching, multi baseline matching, dynamic programming algorithm, GPU, parallel processing

1 Introduction

Image matching is one key technology of generating DSM or DEM in photogrammetry. According to the matching strategy, it can be divided into local matching algorithm and global matching algorithm. The local matching algorithm is defined in a local window of a fixed or adaptive size, and the parallax of each pixel depends only on the pixels within the window. So its computation strategy is relatively simple, but its reliability is poor. Global matching algorithm takes full account of the compatibility, consistency and overall coordination of matching point and the points around, so it can get a more accurate parallax map through choosing a global energy function and minimizing the energy function. This strategy takes full account of the compatibility, consistency and overall coordination of matching point and the points around.

Common used global matching algorithm includes dynamic programming method, graph cut method and belief propagation method. Wherein the dynamic programming method is only carried out in the scan line direction, and can quickly achieve global optimization of the search with poor result. Graph cutting method and confidence propagation rules take full advantage of a two-dimensional constraint (horizontal direction, vertical direction) to obtain accurate dense parallax map, but the calculation is inefficient. Semi-Global Matching is an image-matching algorithm put forward by the German scholar Hirchmuller in 2005. This algorithm applies one-dimensional smoothing constraint in more than one direction to approximate a two-dimensional smoothing constraints, which cannot only get comparable results with graph cuts method and confidence spread, but also achieve higher efficiency than these algorithms. The regular structure of the process is easily mapped to the GPU, DSP and SIMD parallel processing platform [1, 2].

2 Semi-global matching algorithm and multi-baseline extended application

2.1 THE BASIC PRINCIPLES OF SEMI-GLOBAL MATCHING

As other global matching algorithm, semi-global matching algorithm generally include: matching cost calculation, matching consideration polymerization, depth assignment and optimization [3-8].

Matching cost refers to difference of similarity between two corresponding points, and is usually measured by correlation coefficient, mutual information, etc. Under normal circumstances, disparity space image is used to describe the matching cost of corresponding points in scanning line. Thus a plurality of scanning lines constitute a matching cost cubes. There are two existent space parallax image-generating programs: one is constituted by the right and left scanning lines; another is constituted by the left scanning line and the parallax range.

As previously mentioned, the matching cost aggregation is generally achieved by solving the minimum energy function. The semi-global matching algorithm reinforces the reliability of matching through the imposition of additional smoothness constraint on the global energy function, as shown in Equation (1) below, where the first representation of all the matching pixels cost; the second and third represent the punishment of the pixel and its neighbourhood pixels difference is the

* ***Corresponding author*** e-mail: lhzhufang@163.com

presence of small change and a large change in both cases, namely the smoothness constraint, apparently $P_1 < P_2$. *T(...)* refers to analysing function, if and only if its argument is true, the function value is 1, otherwise 0.

$$E(D)=\sum_p e(p,d)=\sum_p \begin{Bmatrix} c(p,d)+\sum_{q\in N_p} P_1 T(|d-d_q|=1) \\ +\sum_{q\in N_p} P_2 T(|d-d_q|>1) \end{Bmatrix}, \qquad (1)$$

For the two-dimensional image, the global minimum for the formula (1) has been proven to be NP problem (Nondeterministic Polynomial), dynamic programming algorithm is introduced to efficiently achieve the one-dimensional path minimization of the energy. Therefore in semi-global matching algorithm, Hirchmuller utilizes 8 (or 16) one-dimensional smoothness constraint to approximately fit a two-dimensional smoothing constraint:

Firstly based on the idea of dynamic programming, compute on each path in accordance with the formulas (2) and (3):

$$L_r(p_0,d)=c(p_0,d).$$

$$L_r(p,d)=c(p,d)+\min\begin{Bmatrix} L_r(p-r,d), \\ L_r(p-r,d\pm 1)+P_1 \\ \min\limits_{i=d_{\min},\ldots,d_{\max}} L_r(p-r,i)+P_2 \end{Bmatrix}, \qquad (2)$$
$$-\min_{i=d_{\min},\ldots,d_{\max}} L_r(p-r,i),$$

$$L_r(p_0,d)=c(p_0,d). \qquad (3)$$

The first item in (2) represents a matching cost of *d* endowed with *p*; the second is a point on this path containing the minimum penalty coefficient matching cost; the third item of the doesn’t influence the optimal path, solely for the purpose of preventing *L* excessively large, to make $L \le C_{\max} + P_2$.

Then, the cost of the matching in each direction is formed by adding the total number of matching cost, such as in (4):

$$S(p,d)=\sum_r L_r(p,d). \qquad (4)$$

After computing matching cost of all pixels, the depth assignment or disparity map is a simple process: each point on the reference image parallax corresponding matching cost minimum parallax value $d_P = \min_d S(p,d)$; each pixel in the reference image corresponding to the parallax $d_m = \min_d S(e_{mb}(q,d),d)$.

Although optimization is not focus of matching algorithm, but in some cases can significantly improve the quality of matches, commonly used optimization method include: sub-pixel, consistency detection and median filtering.

2.2 MULTI-BASELINE EXPANSION METHOD

In multi-baseline matching, exterior orientation elements of the image are used as a link between the multiple images to undertake the matching process. But the parallax is no longer is the only parameter of the point of the same name in the search process, and the intersection point in the Object Space is only one [10]. So we must use elevation instead of parallax, namely (x,y,Z) instead of (x,y,d). The corresponding match measure is rewritten as a function of elevation value. For a given image point in the reference image and its approximate elevation value, you can define all single stereo pair matching measure function into a single framework, such as: traditional dual-like the matching correlation coefficient match measure in the multi-baseline matching pattern is rewritten as SNCC, where the NCC match measure the correlation coefficient of the two images based on elevation variable [11]:

$$SNCC(p_0,Z)=\frac{1}{n}\sum_{i=1}^{n} NCC_i(p_0,Z), \qquad (5)$$

$$NCC_i(p_0,Z)=\frac{\sum_{s\in W}(I_0(s)-\bar{I}_0)\times(I_i(s_i(Z))-\bar{I}_i)}{\sqrt{\sum_{s\in W}(I_0(s)-\bar{I}_0)^2}\sqrt{\sum_{s\in W}(I_i(s_i(Z))-\bar{I}_i)^2}}, \qquad (6)$$

$$\bar{I}_0=\frac{1}{M\times N}\sum_{s\in W} I_0(s),\ \bar{I}_i=\frac{1}{M\times N}\sum_{s\in W} I_0(s_i(Z)).$$

$s_i(Z)$ refers to the corresponding point in the *i* search image computed according to the exterior parameter and DSM.

To integrate the multi-baseline matching algorithm into the overall framework of the semi-global optimization, we should construct a matching cost cube according to the elevation search range, the length and width of the matching area. As shown in Figure 1, wherein each cube element indicates that, the matching cost of corresponding point on the multiple images.

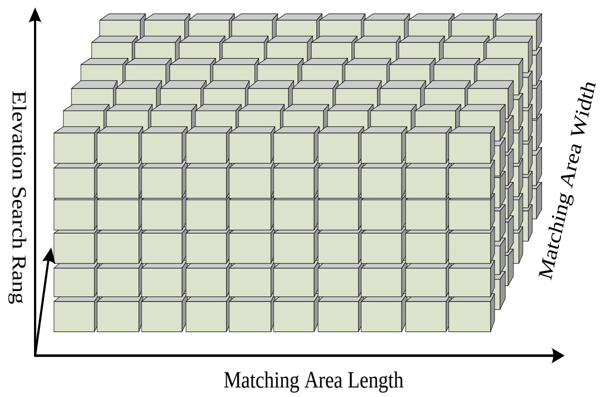

FIGURE 1 The matching area

After generating a matching cost cube, undertake multi-path optimization in the matching region of the X-coordinate direction and Y coordinate directions respectively based on the idea of semi-global optimization. The above method does not change the semi-global matching algorithm overall framework, so can take advantage of the GPU parallel processing.

3 GPU parallel processing technology

So-called GPU parallel processing, in essence, is the co-processing of the GPU-CPU, namely, to integrate GPU and CPU to form the synergistic mode on the hardware; In accordance with the Compute Unified Device Architecture defined (CUDA) programming model to achieve the synergy of GPU and CPU. The CPU is responsible for executing sequential code, while the GPU is responsible for intensive parallel computing, so that the CPU and GPU perform their duties to improve processing efficiency [9-10]. In the process of programming, CUDA allows the programmer to define a C-like language core (kernel) function to achieve the GPU parallel processing.

GPU parallel processing based on Semi-global matching algorithm focuses on compiling the "kernel" function of matching cost and the rational organization of GPU threads to run "kernel" function.

3.1 THE GPU PARALLEL COMPUTING OF MATCHING COST CUBE GENERATION

In the process of generating matching cost cube, each cube element calculation process is completely independent, with a high degree of parallelism, which is suitable for GPU grained parallel processing. The corresponding GPU threads organizational scheme: each thread is responsible for calculating points with the same plane coordinates and elevation coordinates in the search range; thread block and thread grid are divided in accordance with the two-dimensional plane coordinates of the matching region.

Taking into account that the calculation frequently reads the image data and the exterior orientation data, and if you do not optimize but directly read and write from the GPU global memory, will inevitably affect the GPU's performance due to the memory reading and writing latency. We consider these two data access characteristics and the memory structure of the GPU and use the shared memory and texture memory optimization respectively.

Each thread reads the same exterior orientation elements data in the calculation process, and the data is relatively small, so use shared memory for optimization, that is, a one-time elements of exterior orientation data read from the global memory to the high-speed shared memory all the threads within the entire thread block, to improve data access efficiency.

In the calculation process, the image data read by each thread is different. There will be some re-read, but after the projection transformation, the data's regular structure is destroyed so it is difficult to meet the combined access conditions. Taking into account that the array structure of the original image is easily mapped into texture, use texture memory to save the data of the original image. The GPU threads use texture fetch function to read the original image data.

3.2 GPU PARALLEL COMPUTING OF MATCHING COST CUBE'S AGGREGATION

In the calculation process of Polymerization match cost cube, the calculation for each pixel is no longer independent, but need to use this match cost of a point, which makes the traditional block mode in accordance with the image matrix no longer applicable. However, can according to scan row (or column), as is shown in Figure 2.

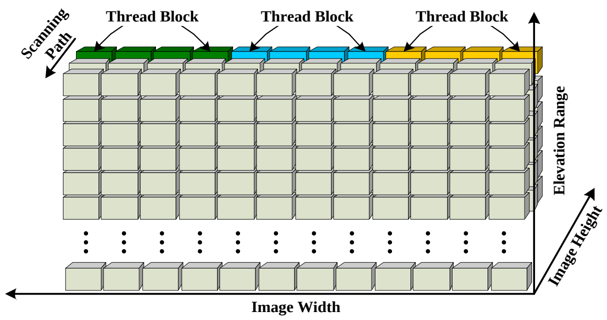

FIGURE 2 The Polymerization match

In the calculation of the three directions, each row of the image is divided into N segments, and N = ImageWidth/PixelAmount. Each segment consists of a PixelAmount*DisparityRange thread blocks to be calculated. In the calculation process for each thread loop is executed ImageLength times to scan the entire image. ImageWidth represents the image width and ImageLength represents image height. PixelAmount represents the number of pixels for each thread block. DisparityRange represents elevation search range.

In addition, in order to further speed up the process, in a scanning process, simultaneously calculate matching cost polymerization of a plurality of directions.

4 Experiment and analysis

4.1 EXPERIMENT

In order to verify the advantages of GPU parallel processing, use large degree of overlap UCD area array image of Hubei Baoying to do multi-baseline MVLL match and GPU parallel processing experiments. The experimental images are shown in Figure 3, wherein (B) is the reference image, (A) and (C) as the search image, the image size is 7500 * 11500 pixels, a pixel size of 9.0 microns, and a focal length of 101.4000 mm.

(A) search image 1 (B) reference image 1 (C) search image 2

FIGURE 3 Experiment image

Experimental platform is a personal desktop supercomputer developed by the United States AMAX PSC-2N, the specific configuration is as follows: two Intel E5620 2.40GHz CPU, 2*8 = 16 cores; 6*4 = 24GB system memory; 2TB hard drive and three NVIDIA Tesla C2050 GPU parallel accelerators. The Tesla C2050 parallel Accelerator configuration parameters: memory capacity of 3.0GB, Memory Interface 384bit GPU processing unit number 448, the frequency of the GPU processing is 1.15GHz.

4.2 RESULTS AND ANALYSIS

First carry out an adjustment processing of POS data to obtain a precise exterior orientation elements of the image; then select the central portion of the reference image, and set the elevation of the search range from -5.0 m to 25.0 m, the search step 0.5 m, DSM grid interconnection interval of 0.5 meters, matching window of 7x7 pixels to carry out the half global optimization MVLL matching experiment. As shown in Figure 4, 800x800 size DSM data generated for the match, the building outline basically preserved. The experiment proves the effectiveness of the semi-global optimization of multi-baseline MVLL match algorithm.

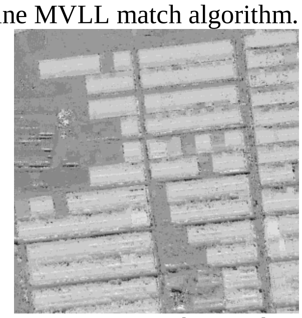

FIGURE 4 Matching result

Make a statistic of CPU and GPU processing time of different matching points (unit: ms), and calculates the acceleration efficiency, as shown in Table 1. It can be seen: through GPU parallel processing computational efficiency can be greatly improved, which is beneficial to the popularization and application of the algorithm.

TABLE 1 CPU and GPU processing time

Matching points	**CPU processing time**	**GPU processing time**	**Acceleration**
200x200	305292	9139	**33.4**
400x400	1223372	19274	**63.5**
800x800	**4896547**	**50397**	**97.2**

5 Conclusions and prospect

This paper firstly analysed the basic principles of the semi-global matching algorithm, and then expanded it to the multi-baseline matching of the remote sensing image. Further improved the matching reliability, at the same time, retained the regular structure of the algorithm; then studied the GPU fine-grained parallel processing technology of the expanded algorithm. Focus on the nuclear function optimization strategies of the polymerization procedure and thread organization scheme. Finally, take advantage of Tesla C2050 GPU parallel accelerator card to process three UCD Aerial Images in semi-global optimization MVLL-multi-baseline-matching, the result proves the effectiveness and efficiency of the algorithm.

This paper is only a research of a single GPU, while the experimental platform PSC-2N personal supercomputer equipped with three Tesla C2050-GPU parallel accelerator cards. In the follow-up study, build the GPU cluster, and further develop the coarse-grained multi-block GPU paralleling between the accelerator cards.

Acknowledgement

The paper is funded by Science and Technology Plan Project of Education Department of Jiangxi Province (NO.Gjj11474).

References

[1] Boykow Y, Veksler O, Zabih R 2001 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**(11) 1222-39

[2] Sun J, Zheng N N, Shum H Y 2003 *IEEE Transcations on Pattern Analysis and Machine Intelligence* **25**(&) 787-800

[3] Hirschmuller H 2008 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **30** 328-341

[4] Hirschmuller H, Scholten F, Hirzing G 2005 Stereo Vision Based Reconstruction of Huge Urban Areas from an Airborne Pushbroom Camera (HRSC) *The 27th DAGM Symposium, Vienna, Austria* 58-66

[5] Hirschmuller H 2008 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **30** 328-41

[6] Ernst I, Hirschmuller H 2008 *The 4th International Symposium on Visual Computing (ISVC08) USA* 228-39

[7] Zhu K, Butenuth M, d'Angelo P 2010 Computational Optimized 3D Reconstruction System For Airborne Image Sequences *The international society for photogrammetry and remote sensing* (ISPRS 2010), CANADA http://www.isprs.org/proceedings/XXXVIII/part1/

[8] Heinrichs M, Rodehorst V, Hellwich O 2007 Efficient Semi-Global Matching for Trinocular Stereo *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* (PIA07) 185-190

[9] *NVIDIA. CUDA 2.0 for WINDOWS CUDA 2.0 Programming guide [EB/OL]* http:// developer.download.nvidia.com/compute/cuda/2.0/doc/NVIDIA_CUDA_Programm- ing_guide_2.0.pdf 2008-06-07

[10] *NVIDIA. CUDA 2.0 for WINDOWS CUDA 2.0 Reference Manual [EB/OL]* http:// developer.download.nvidia.com/compute/cuda/2.0/docs CudaReferenceMannual_2.0.pdf 2008-06-12

Authors

Zhaohua Liu, born on August 7, 1977, Shanxi, China

Current position, grades: Master of Photogrammetry and Remote Sensing, Associate Professor and postgraduate supervisor in Jiangxi University of Science and Technology
University studies: Jiangxi University of Science and Technology
Scientific interests: Remote Sensing Image Processing
Publications : 16 Papers
Experience: an expert in Remote sensing image processing. In recent years, the study of parallel processing method based on GPU and Semi-Global matching algorithm extended for Multi Baseline matching.

Yuxia Yang, born on September 3, 1986, Hebei, China

Current position, grades: Post-graduate students of School of Architectural and survey & mapping engineering, Jiangxi University of Science and Technology
University studies: Jiangxi University of Science and Technology
Scientific interests: Remote Sensing Image Processing
Publications: 3 papers
Experience: She is a parallel algorithm for lovers. Her works are devoted to parallel processing based on GPU.