Fast Stereo Matching of High Resolution Satellite Images Using a New Tilting Technique

An-Nguyen Hong, and Dong-Min Woo

***Abstract*—** The process of stereo matching is very important in computer vision, remote sensing and photogrammetry. Especially in the area of remote sensing and photogrammetry, the matching process should be carried out in a large scaled image. The paper highlights a practical technique on stereo matching for large-scale high resolution satellite images. When dealing with large-scale high resolution satellite images, we integrate the tiling technique with the well-known dynamic programming and coarse-to-fine pyramid scheme as well as the intelligent memory management. Analyzing 350 points from an image of size of 8192 x 8192, disparity results attain an acceptable accuracy with RMS error of 0.5459. Taking the trade-off between computational aspect and accuracy, our method gives an efficient stereo matching for huge satellite image files.

***Keywords*—**stereo matching, tiling technique, disparity map, high resolution satellite image.

I. Introduction

STEREO matching is the corresponding problem in which image coordinate disparities of conjugate points from stereo images are determined. It is not only the first stage of 3D reconstruction based on the Rational Function Model [1], but is also needed for some other processes, such as image registration [2]. By coordinating computer vision, remote sensing and photogrammetry, many applications including Global Position System (GPS), environment management system, scene reconstruction, and navigator have been developed for geography and civil engineering. These applications generate a great deal of interest for the research of geo-referencing, 3D modeling, digital elevation model, and stereo matching [3][4]. Many algorithms, either region-based approaches or feature-based approaches, have been proposed to improve the results of stereo vision.

Although feature-based methods have achieved a lot in the field of computer vision in recent years, most contemporary applications, especially remote sensing and photogrammetry, require dense disparity information which is attained by region-based matching methods [5][6]. According to the recent survey [5], the region-based method employs scan-line algorithms, dynamic programming algorithms, graph-cut algorithms, and belief propagation algorithms, and there is also a need for dense disparity information. Sub-regioning method [7] produces a dense disparity map by using rectangular sub-regioning (RSR) and two-stage dynamic programming (TSDP). Others [8][9][10] have also studied this algorithm for much greater benefit. However, while most methods perform well to Middlebury benchmark and others, in tests on small-size image files, there is a lack of experience in applying them to large-scale data sets under uncontrolled conditions [11].

In this paper, we present an empirical study on stereo matching for large-scale high-resolution satellite images. Dynamic programming is used to overcome the complexity of the computation. It also enables us to find the global minimum for independent scan-line in polynomial time [12]. The fast calculation of 3D correlation coefficient cube is proposed in [7] by using rectangular sub-regioning (RSR) and two-stage dynamic programming. The computation of a dense 3D cube produces the problem of insufficient memory when dealing with large-scale images. Integrating the principle in [7] with overlapped tiling technique, we divide the huge stereo satellite images and apply the process of stereo matching in order to overcome this problem. As a result, the experimental results shows that the suggested method enables capability to perform stereo matching of large scale satellite images, whereas most methods concentrate on small size images.

II. Fast Cross-Correlation

A fast cross-correlation method [8] is proposed by using box-filtering techniques [13]. The original formula to calculate cross correlation between windows in the stereo images f and g using a disparity value of d is:

$$C(i,j,d) = \frac{\mathrm{cov}_{ij,d}(f,g)}{\sqrt{\mathrm{var}_{ij}(f)} \times \sqrt{\mathrm{var}_{ij,d}(g)}} \quad (1)$$

Where, covariance $\mathrm{cov}_{ij,d}(f,g)$ is transformed like this:

$$\mathrm{cov}_{ij,d}(f,g) = \sum_{m=i-K}^{i+K} \sum_{n=j=L}^{j+L} f_{m,n} g_{m+d,n} - W\bar{f}_{i,j}\bar{g}_{i+d,j} \quad (2)$$

Where, $f_{m,n}$ is the intensity value of an M x N image f at position (m, n) and i, j are the image row and column indices. We have the same definition for right image g. d indicates possible disparity value. $\bar{f}$ and $\bar{g}$ are the mean values within the local windows of size K x L. Fig. 1 from [8] illustrates the 3D cube of correlation coefficient between left and right image.

An-Nguyen Hong, is with the Electronics Engineering Department, Myongji University, Gyeongido, Korea 449-728.

Dong-min Woo, is with the Electronics Engineering Department, Myongji University, Gyeongido, Korea 449-728 (e-mail: dmwoo2000@gmail.com).

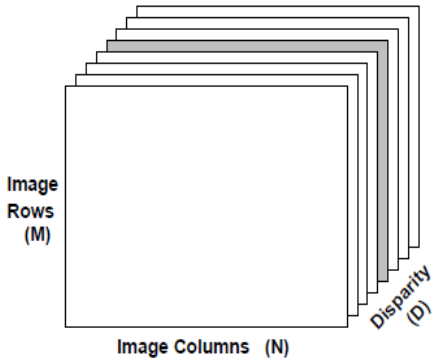

Fig. 1 An illustration of the 3D correlation coefficient volume obtained after using the fast correlation method.

III. Quad-Tree Sub-Regioning(QSR)

Because we calculate correlation for all possible disparity values, the correlation values of the whole image will cost O(MND) computation complexity. In [7],[8] Sun proposed a rectangular sub-regioning in which we just have to work with a sum of sub-regions, which are segmented from the whole image. The aim of the sub-regioning process is to segment the image so that points with a similar range of disparity value are grouped into sub-regions. In this way, the computation complexity is $\sum_{i=0}^{R-1}(M_i N_i D_i)$ working with R subimages. Obviously, the $\sum_{i=0}^{R-1}(M_i N_i D_i)$ is smaller than O(MND) especially when the disparity changes a lot within a whole image, so working with subregions requires smaller memory usage.

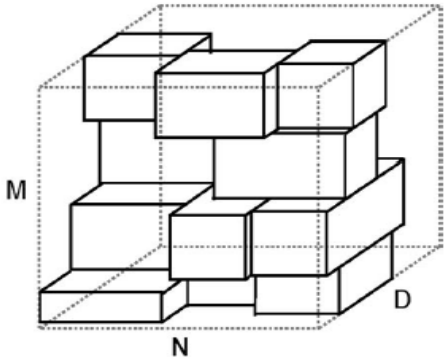

Fig. 2 Size and shape of the 3D correlation volume obtained by stacking together many smaller 3D volumes after correlation calculation using RSR

Fig.2 describes size and shape of the 3D correlation volume by using rectangular sub-regioning. However, rectangular sub-regioning is not flexible enough to accommodate the fluctuation of disparity value. It only achieves an accurate disparity map if objects in each sub-region of the scene have a maximum disparity of no more than ten pixels [9]. Our algorithm exploits quad-tree decomposition to achieve more accurate sub-regioning. By using quad-tree presentation, segmenting disparity values do not have to be stuck to only a horizontal direction or a vertical direction (see Fig. 3). The quad-tree sub-regioning method [10] is based on cost function to determine whether a region should be split or merged. Our algorithm orients to a huge size image file, so that we segment a region into small regions by using a threshold-based homogeneity check. This implementation allows us to save time and computing resources.

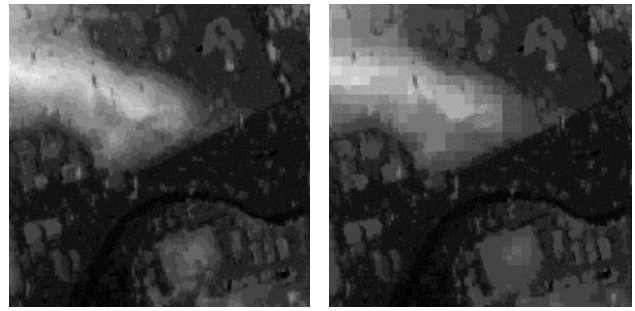

Fig. 3 Quad-tree sub-regioning - Left image is a disparity map at a particular image pyramid. Right image is the disparity map represented by quad-tree

IV. Maximum Surface in the Volume

Fig. 4 illustrates the 3D maximum surface in the volume. The algorithm is from [7] where they propose two-stage dynamic programming to find out shortest paths from a 3D correlation coefficient cube. The shortest path is representation of the most reliable disparity value of points in the image.

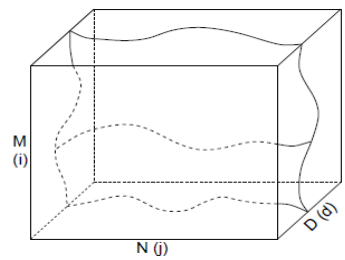

Fig. 4 The illustration of the 3D maximum-surface which gives the maximum accumulation of values in the 3D cross correlation coefficient volume

The first stage is to calculate the maximum summation of C(i, j, d) in the vertical direction for each slide from top to bottom of the 3D volume.

$$Y(i,j,d) = C(i,j,d) + \max_{t:|t|\leq p} Y(i-1,j,d+t) \quad (3)$$

Where, p determines the number of local values that need to be checked and

$$Y(0, j, d) = C(0, j, d) \quad (4)$$

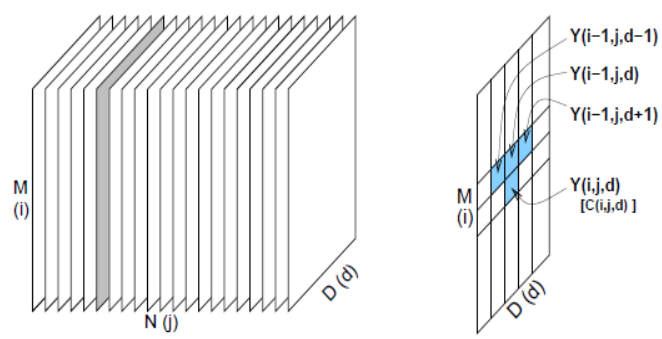

Fig. 5 Obtaining the Y(i,j,d) volume. Left shows the 3D volume Y with vertical slice in grey. Right illustrates the positions of the Y values at each iteration

The second stage obtains the disparity map:

- Starting from the bottom slice of the 3D volume Y, obtain the shortest-path from left to right (similar to stage 1 but in a different direction).
- The sum of values along the obtained path gives the maximum value, which is also the maximum summation value along the whole 3D surface.
- Move from the bottom slice of Y upwards.
- Calculating the disparity, mask out the values which are more than p position away from the shortest-path of the previous slice.

V. Integration of Overlapped Tiling Technique

In the field of stereo matching, particularly region-based matching, Normalized Cross Correlation is most widely used for its simplicity and reliability. However, it is sensitive to significant scale, rotation and shearing between the images to be correlated. Although fast cross correlation and maximum 3D surface show their efficiency on computing, they still suffer from heavy memory loading when applying the computation for the whole large-scale high resolution satellite image. This situation inspired us to integrating the current algorithm to the tiling technique. By dividing the huge size image into sub-images and refining the usage of memory, we could solve the problem of large-scale satellite image with least memory resource. Fig. 6 gives an illustration of our overlapped tiling process.

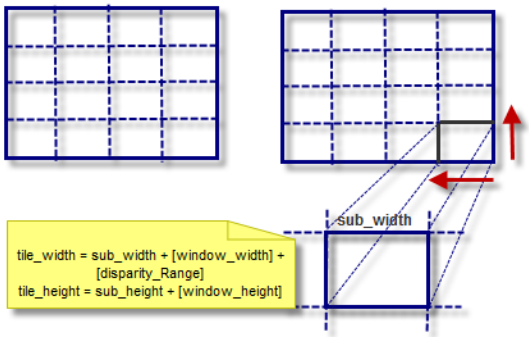

Fig. 6 Overlapped tiling technique with added boundary for mask-filtering and disparity_range

Because of mask-filtering which is a rectangular window, to calculate the correlation between two images, we used overlapped tiling, which creates ghost region around each tile that contains a copy of the element of the neighbor tiles needed for the computation [14]. The primary part of a sub-image will have an added boundary depending on the position of the sub-image in the whole image.

VI. Experimental Result

In this section, we present an experiment with large-scale high resolution images of Daejeon (Korea). The size of each image is 11004 x 11004 of pixels. The matching process needs stereo images epipolar-resampled so we perform piece_wise linear epipolar resampling with the system we developed at Myongji University. After that, an epipolar image pair is cropped to the size of 8192 x 8192 of pixels

Fig. 7 is the reference image and target image of Daejon area (South Korea). Their epipolar images are presented in Fig. 8. The cropped images of those epipolar images are in Fig. 9.

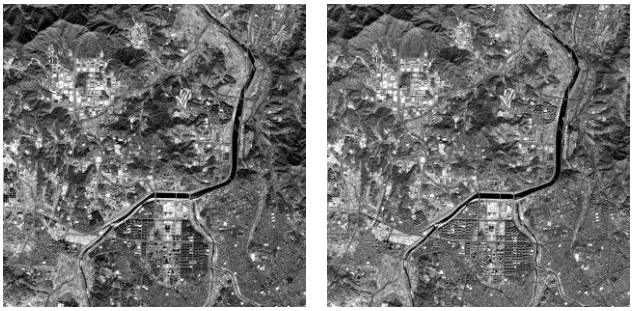

Fig. 7 Reference (left) image and Target (right) image of Daejon area (South Korea)

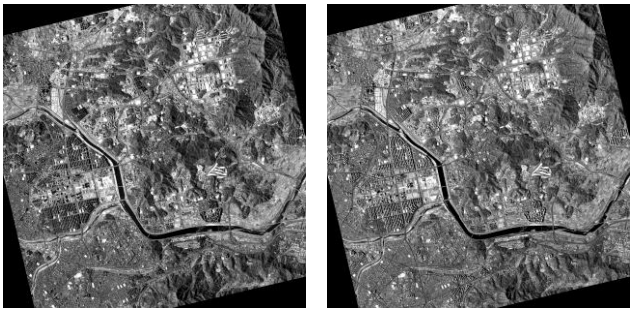

Fig. 8 Reference (left) epipolar image and Target (right) epipolar image of Daejon area (South Korea).

Fig. 9 Cropped Reference (left) epipolar image and Target (right) epipolar image of Daejon area (South Korea).

We use 4 levels of pyramid scheme (from level 3 to level 0) to perform the stereo matching process. Then we use (8 x 8) tiling pieces for level 3, 2, 1 and (16 x 16) tiling pieces for level 0. That means, for instance, in final level 0, we divide image with size (8192 x 8192) into (16 x 16) sub images, then add boundary for each sub image, so that we have (16 x 16) overlapped tiling pieces. At level 3, the disparity range [-3, 10] spreads from -3 (disparity towards the left) to 10 (disparity toward the right). That is proportional to [-6, 20], [-12, 40], [-24, 80] at level 2, 1, and 0 respectively. The sizes of the mask filtering window are (3 x 3) for level 3, (5 x 5) for level 2, (9 x 9) for level 1, (17 x 17) for level 0. The threshold values of disparity homogeneity in Quad-tree Sub-regioning are 2, 4, 8 for level 2, 1, and 0 respectively.

Analyzing 350 points scattered over an image, we get acceptable accuracy. The disparity result attains an RMS of around 0.5459. Table 1 presents the results of experiment in

more detail. Fig. 10 is the final disparity map obtained at pyramid level 0.

TABLE 1
EXPERIMENT RESULTS

Image size (pixel x pixel)	8192 x 8192
RMS of disparity result	0.5459
Number of tiling pieces	16 x 16
Subimage size (pixel x pixel)	512 x 512
Disparity range (pixel)	104 [-24, 80]
Window size (pixel x pixel)	17 x 17
Maximum overlapped tiling piece width (pixel)	(512 + 2 x 17 + 104) = 650
Maximum overlapped tiling piece height (pixel)	(512 + 2 x 17) = 546
Memory for processing a sub image (MB)	512 x 512 x 104 x 4 = 109
Memory for processing an overlapped tiling piece (MB)	650 x 546 x 104 x 4 = 148

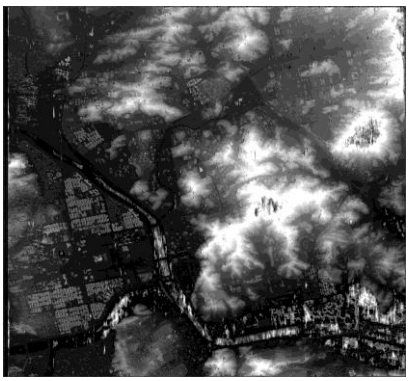

Fig. 10 Final disparity map at level 0 (8192 x 8192)

An image of size 8192 x 8192 will require 28GB (8192 x 8192 x 104 x 4) of memory to calculate a float cube of 3D correlation coefficient. It is obviously an excess requirement. Dividing the image into (16 x 16) sub images needs just around 109 MB (512 x 512 x 104 x 4) for each sub image. For the boundary of overlapped tiling pieces, the memory requirement also does not exceed 148 MB for one piece. However, as a region-based approach with cross correlation calculation, we also have a problem with texture-less region. Other features which can cause program to generate erroneous results are a repeating pattern and occlusion. Other approaches are said to have tried solving these problems. Graph-cut based stereo matching using image segmentation with symmetrical treatment of occlusions and a novel interest-point matching algorithm for high resolution satellite images can be efficiently used for these problems.

VII. CONCLUSIONS

We have presented a new method to perform stereo matching for large-scale satellite images. Taking advantage of dynamic programming from an original algorithm, we experimented with large-scale images which have huge size. By integrating a tiling technique, we can solve the problem of memory excess. We have also exploited a quad-tree sub-regioning algorithm to improve the results of stereo matching. The quad-tree sub-regioning is implemented in an efficient way with threshold-based sub-regioning. In addition, because of the memory requirement for processing a big image, the algorithm should release memory right after the processing of each sub part is finished. However, with a quite large range of disparity (about 104), our algorithm still has a time consuming problem, as it wants to compute a dense disparity map. Other disadvantages, normally from the region-based approaches, include getting lost in a texture-less region, or producing error result in a repetitive and occluded region. All these kinds of problem have been addressed by many researchers. However, they solved problems with small sized benchmark images. We believe in the near future, more researches on all these kinds of problems will be done for huge size images. Integrating existing techniques to tiling, and seeing if we can solve the problem in a short time may also be an interesting direction.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education(No. NRF-2012R1A1A2004950).

REFERENCES

[1] Vincent Tao and Yong Hu. 3D Reconstruction Methods Based on the Rational Function Model. Photogrammetric Enginerring & Remote Sensing, Vol. 68, No. 7, July 2002, pp. 705-714

[2] Barbara Zitová, Jan Flusser. Image registration methods: a survey. Image and Vision Computing 21 (2003) 977-1000 http://dx.doi.org/10.1016/S0262-8856(03)00137-9

[3] Gaurav Gupta and M.S. Rawat. Region growing stereo matching method for 3D building reconstruction. Int. J Computational Vision and Robotics, Vol. 2, No. 1, 2011

[4] Hae-Yeoun Lee, Taejung Kim, Wonkyu Park, Heung Kyu Lee. Extraction of digital elevation models from satellite stereo images through stereo matching based on epipolarity and scene geometry. Image and Vision Computing 21 (2003) 789- 706 http://dx.doi.org/10.1016/S0262-8856(03)00092-1

[5] Nalpantidis Lazaros, Georgios Christou Siraloulis & Antonios Gasteratos (2008): Review of Stereo Vision Algorithm: From Software to Hardware, International Journal of Optomechatronics, 2:4, 435-462 http://dx.doi.org/10.1080/15599610802438680

[6] Zhen Xiong and Yun Zhang. A Novel Interest-Point-Matching Algorithm for High-Resolution Satellite Images. IEEE Transaction on Geoscience and Remote Sensing, Vol. 47, No. 12, December 2009

[7] Changming Sun. Fast Stereo Matching Using Rectangular Subregioning and 3D Maximum-Surface Techniques. International Journal of Computer Vision. Vol. 47, no. 1/2/3, pp. 99-117, May 2002

[8] Changming Sun. Multi-Resolution Rectangular Subregioning Stereo Matching Using Fast Correlation and Dynamic Programming Techniques. CMIS Report No. 98/246. December 1998

[9] Arturo Donate, Xiuwen Liu, and Emmanuel G. Collins, Jr. Efficient Path-Based Stereo Matching With Subpixel Accuracy. IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics, Vol. 41, No. 1, February 2011 http://dx.doi.org/10.1109/TSMCB.2010.2049839

[10] Carlos Leung, Ben Appleton, Changming Sun. Iterated dynamic programming and quadtree subregioning for fast stereo matching. Image and Vision Computing 26 (2008) 1371-1383 http://dx.doi.org/10.1016/j.imavis.2007.11.013

[11] Andreas Geiger, Martin Roser, and Raquel Urtasun. Efficient Large-Scale Stereo Matching. 2010. ACCV'10 Proceeding of the 10th Asian conference on Computer vision - Volume Part 1. Pages 25-38

[12] Daniel Scharstein, Richard Szeliski. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. International Journal of Computer Vision 47 (1/2/3), 7-42, 2002.

[13] M. J. McDonnell, "Box-filtering techniques", Computer Graphics and Image Processing, vol. 17, pp. 65-70, 1981 http://dx.doi.org/10.1016/S0146-664X(81)80009-3

[14] Jia Guo, Ganesh Bikshandi, Basilio B. Fraguela, Maria J. Garzarán, David Padua. Programming with Tiles. PPoPP '08, Proceeding of the 13th

ACM SIGPLAN Symposium on Principles and practice of parallel programming, pp. 111-122, 2008.

[15] Daniel Scharstein, Richard Szeliski. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. International Journal of Computer Vision 47 (1/2/3), 7-42, 2002.