Evaluation of Dynamic Programming among the Existing Stereo Matching Algorithms

Teo Chee Huat and Nurulfajar bin Abd Manap

Department of Electronic Engineering, Faculty of Electronics and Computer Engineering, FKEKK
Universiti Teknikal Malaysia Melaka, Melaka, Malaysia
teocheehuat@hotmail.com, nurulfajar@utem.edu.my

Abstract. There are various types of existing stereo matching algorithms on image processing which applied on stereo vision images to get better results of disparity depth map. One of them is the dynamic programming method. On this research is to perform an evaluation on the performance between the dynamic programming with other existing method as comparison. The algorithm used on the dynamic programming is the global optimization which provides better process on stereo images like its accuracy and its computational efficiency compared to other existing stereo matching algorithms. The dynamic programming algorithm used on this research is the current method as its disparity estimates at a particular pixel and all the other pixels unlike the old methods which with scanline based of dynamic programming. There will be details on every existing methods presented on this paper with the comparison between the dynamic programming and the existing methods. This can propose the dynamic programming method to be used on many applications in image processing.

Keywords: Dynamic programming, disparity, comparison, stereo matching algorithms
PACS: 07.05.Rm

INTRODUCTION

Stereo matching algorithms are the process to find the disparity or the distance from the corresponding points or pixels of a stereo pair images. There are many existing stereo matching algorithms developed by researchers on image processing field. It is an important function in computer vision technology to analyze the two dimensions, 2D and three dimensions, 3D of output based on stereo images. In developing an algorithm of stereo matching, the accuracy and speed are conditions that need to be concerned to produce a precise output of computer vision. As to get better results of disparity on stereo images, there are many stereo matching methods tried on the stereo datasets for this research such as dynamic programming, basic block matching, sub-pixel estimation, and image pyramiding. The datasets used for the experimental on the stereo matching algorithms are from the Middlebury which is a sharing network about research on stereo vision.

From the stereo matching algorithms that have tried out in this research, dynamic programming (DP) is the most preferable method to obtain a better disparity of stereo pair of datasets among the existing stereo matching algorithms. The existing stereo matching algorithms mostly will go through the standard steps like the matching cost computation, cost aggregation, disparity optimization and disparity refinement. There are various equations to be included to find the pixel-based matching cost which are the sum of absolute differences (SAD), sum of squared difference (SSD), sum of truncated absolute differences (STAD), normalized cross correlation (NCC), zero mean normalized cross correlation and many other methods. NCC and SSD are traditional matching cost. In this particular research, the SAD equation is used to get the corresponding points between the reference and target of data set. The cost of the correspondence between the reference and target data set can represent by disparity space image (DSI) [1][2][3].The DSI is formed from the matching cost values from the disparities and pixels when summing up the cost values on each matching image data set [4]. In order to find the disparity depth map, the basic blocking matching is used as it is the initial step to find the absolute difference between the pixel intensities. The basic block matching is using the sum of absolute differences (SAD) to compare the pixels found on the left and right of stereo pair data [1][2][4]. One of the example of dataset used for stereo matching experimental is the Tsukuba stereo pairs which from Middlebury. The disparity map is calculated using three sizes of window which are 3 by 3, 5 by 5 and 7 by 7 pixel of block around the reference of image data. In basic block matching, it locates the block or pixels in the range of ± 15 over columns sequence [5].

On the step of cost aggregation, there are two types of support region which are the two dimensional, 2D fixed disparity and the three dimensional, 3D with the variables of x, y and d as in equation 1. The 2D aggregation is by using square windows, shiftable windows, windows with adaptive sizes and windows with constant disparity [4] [6].

The disparity optimization can be categorized into two approaches which are the local approaches and global approaches. The local approaches use the Winner Takes All (WTA) by picking each pixel where the disparity is correlated to the minimum cost value in order to increase the signal to noise ratio (SNR) as to reduce the ambiguity. The global approach is a frame work to search for disparity function, d that minimize an energy function or global energy over the disparity computation phase like the pixel-based matching cost [1],

$$E(d) = E_{data}(d) + E_{smoot\,h} \tag{1}$$

where $E_{data}(d)$ represent the disparity function with the input of data set which minimized through the pixels on the correspondence of disparity map, d if there is similarity in intensities and maximized when the disparity map putting the pixels in correspondence which slightly differ in intensities. The $E_{smoot\,h}$ represent the conjecture of the smoothness made from the algorithm which measured from the disparity between the pixels on pixel grid [1] [2] [4].

For disparity refinement part, there are various approaches to improve the smoothness of the disparity map. In this research, the method of dynamic programming is chosen as the approach on the output which obtained by the basic block matching stereo correspondence algorithm. Dynamic programming is able to search for global minimum of independent scanlines when in polynomial time which approximate to about a second. It is chosen based on its accuracy when dealing at the areas of depth borders and uniform regions [7] [8].

ALGORITHM OUTLINE

Basic Block Matching

Block matching is a common method that used in finding the corresponding points in stereo matching. This particular method is used in this research as the beginning algorithm to find corresponding points from the dataset that used along the experimental of stereo matching. On the experimental, the pixel value of the target image is predicted as the corresponding pixel in the reference image where the displacement of the corresponding pixels or as the motion vector to be estimated using the block matching [9]. The block matching is used to minimize the matching errors between the block at position of (x, y) in the target image, I_t while for the position of the reference image, I_{t-1} will be $(x+u, y+v)$ where u and v is the motion vector. These variable defined can be summarized as sum of absolute difference (SAD) [9],

$$\mathrm{SAD}_{(X,Y)}(u,v) \equiv \sum_{j=0}^{P-1} \sum_{i=0}^{P-1} |I_t(x+i, y+j) - I_{t-1}(x+u+i, y+v+j)| \tag{2}$$

where P is the block size, P X P. As to minimize the $\mathrm{SAD}_{(X,Y)}(u,v)$, (a, b) is defined as the motion vector estimation to compare and obtain the SAD of each position, $(x+u, y+v)$ for the dataset. The equation shows as [9],

$$(a,b) \equiv \arg\min_{(u,v)\in Z} SAD_{(x,y)}(u,v) \tag{3}$$

where Z = {(u,v)| -B ≤ u,v ≤ B and $(x+u, y+v)$ represent the valid position of pixel in the reference image, I_{t-1} while B is an integer to search for range. From the SAD equation, the global minimum of matching error can be obtained.

Sub-pixel Accuracy

Most of the stereo matching algorithms use sub-pixel refinement or the sub-pixel accuracy after obtaining the correspondence pixels of the stereo datasets. Sub-pixel accuracy is a method that goes through discrete disparity levels that from the matching cost and also the iterative gradient descent [10]. The main purpose of the sub-pixel accuracy method is to increase the resolution of the stereo matching algorithm output from the stereo datasets. Sub-pixel is able to smooth the transition between the regions from different disparity that cause contouring effect on the images on a depth map. During the process of sub-pixel accuracy, it will focus on the minimum cost and the neighboring cost values to get the sub-pixel correction. In applying the method of sub-pixel accuracy, the normalized cross correlation (NCC) is used to compute the integral stereo images where the cross correlations at the sub-pixel location of the stereo images can be computed efficiently and the equation used for computation is shown as following as NCC (*x*,*y*,*u*,*v*) is equal to [10],

$$\frac{\sum_{(i,j)\in w} I_1(x+i,y+j).I_2(x+u+i,y+v+j)}{\sqrt[2]{I_1{}^2(x,y).I_2{}^2(x+u,y+v)}} \tag{4}$$

where the NCC can be defined as left image window at the position (*x*,*y*) while for the right image window position at (*x*+*u*, *y*+*v*). From the NCC equation, it can be substituted by using an integral stereo images which with squared of pixel values.

Dynamic Programming

On the part of disparity optimization, the global optimization algorithm that chosen for this research is using dynamic programming as this algorithm optimize energy function to be NP-hard for smoothness purpose [1]. There are two categories of global optimization such as one dimension and two dimension optimization methods. One dimension optimization is traditional method where its estimation on the disparity is focusing on a pixel that depending on other pixels on the same scanline but independent on disparity that focus on other scanlines. One dimension is not considered as a truly global optimization as its smoothness technique is only focus on horizontal direction. However, one dimension optimization is still being used by some of the researchers due to its simple implementation and its effectiveness on the disparity maps outputs.

Two dimension optimization approach is smoothing the stereo images in the direction of vertically and horizontally to approximate the disparity map by using simulated annealing, continuation methods and mean-field annealing [11] [12] [13]. However, these methods are not efficient enough to optimize the equation in (1). There are two methods which compatible in optimizing the equation (1), the graph-cuts and belief propagation as these two methods able to obtain better results accordingly to ground truth data from stereo matching algorithm [14] [15]. In this paper, the dynamic programming used for the experimental results is the dynamic programming on tree due to its efficiency as the one dimension optimization. The tree graph of the dynamic programming can represent as *T (V, E)* where *V* as vertices and *E* as edges. The efficiency of the dynamic programming on tree begins with its optimization on the energy function [7],

$$E(d) = \sum_{a\in V} m(d_a) + \lambda \sum_{(a,b)\in E} s(d_a, d_b) \tag{5}$$

where a is the pixel in the left image and the d_a as the value of disparity map, *d* at the pixel of a. Assuming $m(d_a)$ is the matching penalty of relating d_a to the pixel of a where it is the absolute difference between the pixel, a in the left image and the a pixel which shifted on the right image and can be summarized as $\sum_{a\in V} m(d_a)$. Meanwhile, assume $s(d_a, d_b)$ as the smoothness penalty for the disparity of d_a and d_b to the pixel *p* and *q* and the variables can be summarized as $\sum_{(a,b)\in E} s(d_a, d_b)$.

As to get the minimum energy of equation (5), let h as the root vertex of tree, $h \in V$ and assume $z \in V$ as the number of edges the root of distance between h and z. Each node of z belongs a parent as $p(z)$ and the depth is equally to the depth of $z-1$ while if it is not a root, the minimum value of the energy in equation (5) have a sub-tree rooted at z and the edge in the middle of z and $p(z)$ can be summarized as $d_{p(z)}$ [7],

$$E_z(d_{a(z)}) = \min_{d_z \in D}(m(d_v) + s(d_z, d_{a(z)}) + \sum_{w \in C_z} E_w(d_z) \tag{6}$$

where C_z as the children set of z. while for the optimal disparity for the root node h can be represented as [7],

$$L_h = arg \min_{d_h \in D}(m(d_h) + \sum_{w \in C_h} E_w(d_h)) \tag{7}$$

where if z is a node that without children then C_z is empty and the function of E_z and L_z can be evaluated directly. Let take J as the maximum depth in the tree, the energy function of equation (5) is optimized by evaluating the functions E_z and L_z for each node z at the depth, J. After evaluation on the functions, proceed with the evaluation on the same functions for all the nodes at depth of $J-1$ due to any child w has the depth of J, this is the evaluation on E_w and L_w. Next step is to keep evaluating the function of E_z and L_z in decreasing order for the depth till it reach to the root for disparity assignment optimal computation purpose.

EXPERIMENTAL RESULTS

In this paper, there are three existing stereo matching algorithms have tried out which are the basic block matching, sub-pixel accuracy, and dynamic programming. From the three stereo matching algorithms, dynamic programming is the most efficient algorithm among the other algorithms in smoothing the disparity depth map. The effectiveness of dynamic programming in smoothing the depth map is also depending on the suitable disparity range (DR) of the stereo pair of images used. The stereo pairs of images that used in this research are Tsukuba, Teddy, Sawtooth and Venus which these stereo pairs of images are chosen from Middlebury. The results obtained are shown as Table 1, Table 2 and Figure 1.

TABLE 1. Disparity range

Stereo images	Tsukuba	Teddy	Sawtooth	Venus
Disparity Range	15-16	45	27-30	30

TABLE 2. Time taken of different set of stereo images for each stereo matching algorithm

Stereo images	Basic Block Matching	Sub-pixel Accuracy	Dynamic Programming
	Time taken (second)		
Tsukuba (DR =16)	14	19	50
Teddy (DR = 45)	29	30	164
Sawtooth (DR = 29)	25	27	113
Venus (DR = 30)	28	28	114

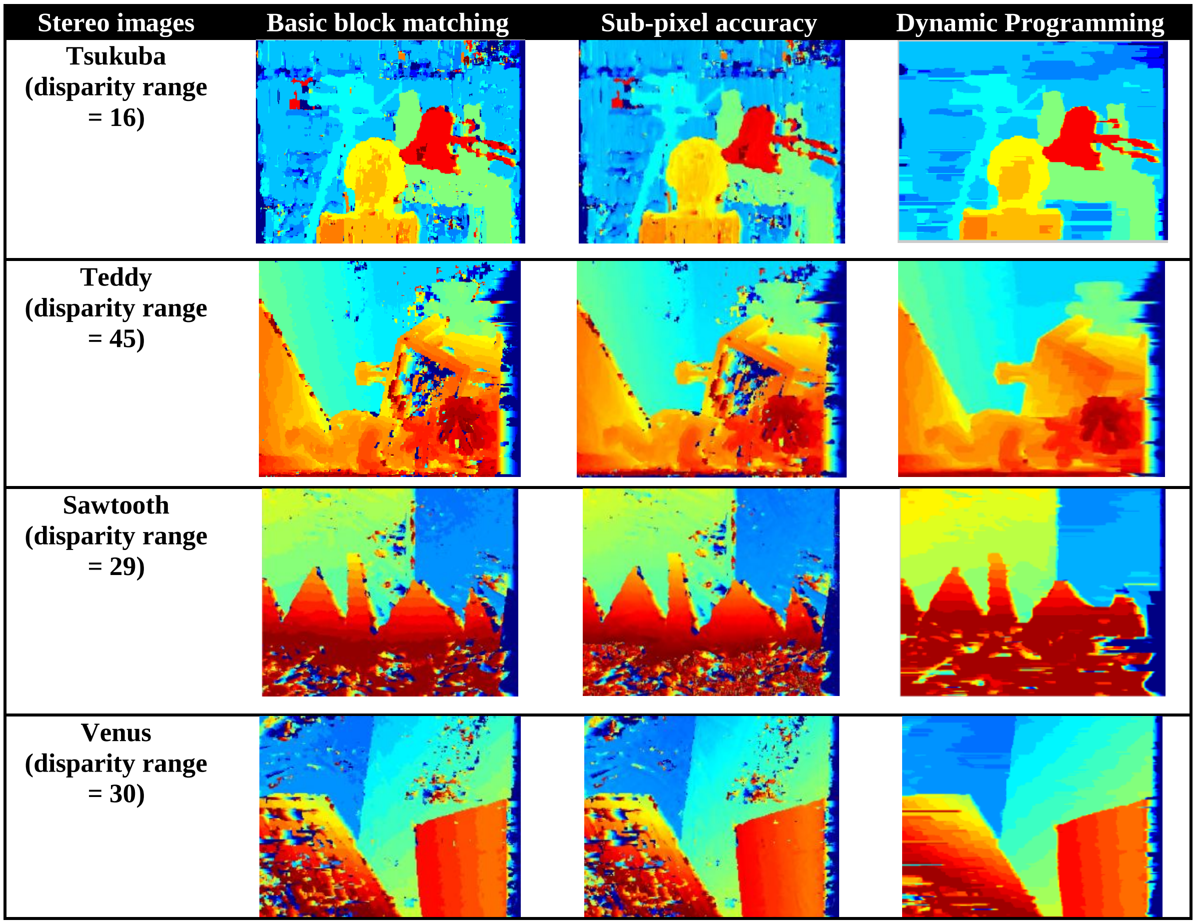

FIGURE 1. Results obtained from stereo matching algorithm

From the observation on the results of three stereo matching algorithms, the accuracy which depends on smoothness in increasing order is starting from sub-pixel accuracy followed by basic block matching and the smoothest results are obtained by the dynamic programming algorithm as shown in figure 1. The hardware use to run the simulation of the three stereo matching algorithms is the portable computer with integrated of processor of 2.5 gigahertz (GHz) and three gigabytes (GB) of installed memory. Table 1 shows on the disparity range for each dataset of stereo image where the disparity range are obtained from the experimental by applying stereo matching algorithms and observing on the output accuracy for different disparity range applied on every dataset coding. Table 2 shows the results of the time taken in second for each stereo matching algorithm on different sets of stereo images, the computation efficiency of the dynamic programming is the lowest among the stereo matching algorithms while the highest computation efficiency of stereo matching algorithms is the basic block matching where its average time taken for all the stereo images datasets is faster than the average time taken of sub-pixel accuracy algorithm. Besides that, from the results obtained in Table 2 it shows that the higher disparity range of stereo images datasets, the longer time taken to run on the stereo matching algorithms.

CONCLUSION

As a conclusion, the comparison between the three stereo matching algorithms can clearly shows that dynamic programming is the most efficient method among the others in smoothing the depth map. It is depending on the suitable disparity for different content of stereo pair of images used. Besides that, from the experimental results by using the Middlebury datasets shows that dynamic programming algorithm able to minimize the matching errors and get a better stereo matching effect compare to the basic block matching and sub-pixel accuracy algorithm. Therefore, based on the comparison among the three stereo matching algorithms it is recommended that dynamic programming may helps to get more satisfactory effect of depth map especially in removing the apparent defective stripes. The depth map that obtained from the dynamic programming algorithm is effective for most applications compared to basic block matching and sub-pixel algorithm such as applications on 3D purposes.

ACKNOWLEDGEMENTS

This research is funded by the grant from Faculty of Electronic and Computer Engineering, Universiti Teknikal Malaysia Melaka (UTeM) with project number of PJP/2013/FKEKK (16C)/S01203.

REFERENCES

1. D. Scharstein and R. Szeliski. *A taxonomy and evaluation of dense two-frame stereo correspondence algorithms*. Tech-nical Report MSR-TR-2001-81, Microsoft Research, 2001
2. M.Bleyer, M.Gelautz, 2005. *A layered stereo matching algorithm using image segmentation and global visibility constraints*, 59(3), pp.128-150.
3. Di Stefano, L., Mattoccia, S., Mola, M., 2003. *An efficient algorithm for exhaustive template matching based on normalized cross correlation*, IEEE Computer Society, pp.322-327.
4. F.Tombari, et al. 2010. *A 3D Reconstruction System Based on Improved Spacetime Stereo*, In IEEE, pp. 7-10.
5. S. Birchfield and C. Tomasi. *Depth discontinuities by pixel- to- pixel stereo*. In ICCV, pp 270–293, 1998.
6. O. Sundstr and L.Guzzella, 2009. *A Generic Dynamic Programming Matlab Function*, 1(7), pp. 1625-1630.
7. O.Veksler. *Stereo Correspondence by Dynamic Programming on a Tree*. In CVPR, Volume 2, pages 384-390,2005.
8. A.Olofsson, 2010. *Modern Stereo Correspondence Algorithms: Investigation and evaluation*, 17 June, pp.5-86.
9. Y.Chen, Y.Hung et al.2001.*Fast Block Matching Algorithm Based on the Winner-Update Strategy*. In IEEE, 10(8), pp.1212-1222
10. A.Donate,Y.Wang et al. 2008. *Efficient and accurate subpixel path based stereo matching*. In IEEE, 8(6), pp.1-4
11. D. Geiger and F. Girosi. *Parallel and deterministic al- gorithms* from MRF's: Surface reconstruction. IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(5):401–412, May 1991.
12. S. Geman and D. Geman. *Stochastic relaxation, Gibbs dis- tributions, and the Bayesian restoration of images*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 6:721–741, 1984.
13. A. Blake and A. Zisserman. *Visual Reconstruction*.MIT Press, 1987
14. V. Kolmogorov and R. Zabih. *Multi-camera scene recon- struction via graph cuts*. In ECCV02, page III: 82 ff., 2002.
15. P. Felzenszwalb and D. Huttenlocher. *Efficient belief propa- gation for early vision*. In CVPR04, pages I: 261–268, 2004.