

A Modified Area Based Local Stereo Correspondence Algorithm for Occlusions

Jungwook Seo and Ernie W. Hill

Department of Computer Science,
University of Manchester,
Manchester M13 9PL, UK
{seoj, ernie.hill}@cs.man.ac.uk

Abstract. Area based local stereo correspondence algorithms that use the simple 'winner takes all' (WTA) method in the optimization step perform poorly near object boundaries particularly in occluded regions. In this paper, we present a new modified area based local algorithm that goes some way towards addressing this controversial issue. This approach utilizes an efficient strategy by adding the concept of a computation skip threshold (CST) to area based local algorithms in order to add the horizontal smoothness assumption to the local algorithms. It shows similar effects to Dynamic Programming(DP) and Scanline Optimization(SO) with significant improvements in occlusions from existing local algorithms. This is achieved by assigning the same disparity value of the previous neighboring point to coherent occluded points. Experiments were carried out comparing the new algorithm to existing algorithms using the standard stereo image pairs and our own images generated by a Scanning Electron Microscope (SEM). The results show that the horizontal graphical performance improves similarly to DP particularly in occlusions but the computational speed is faster than existing local algorithms, due to skipping unnecessary computations for many points in the WTA step.

1 Introduction

The main aim of stereo vision systems is to determine depth between two or more stereo image pairs using an approach which is similar to the human vision system [4]. More recently, a variety of dense stereo correspondence algorithms have been developed for a variety of purposes such as computer vision, robot navigation, intelligent vehicles and so on [11] [13]. These dense stereo matching algorithms can be classified in two categories: namely local and global algorithms. [1] Local algorithms are broadly split into two categories: area based matching and feature based matching. The area based algorithms aggregate costs of pixels in correlated window regions without smoothness assumptions and are fast computationally but exhibit poor performance graphically. Notwithstanding, they have often been used for real time applications. [12] In the feature based matching, two classes have recently received attention: hierarchical feature matching and segmentation matching.

On the other hand global algorithms are based on iterative schemes and minimize the global cost function combining data and smoothness terms. These algorithms such as Graph Cuts(GC) [18] produce better and more accurate disparity maps but the computational costs are often too high. We have attempted to reconstruct the 3D surfaces of specimens from nano-scaled stereo image pairs generated by the SEM at slightly different eucentric tilting angles for a fast nanotechnology application. Initially, we used area based local stereo matching algorithms rather than global ones. However, the results from the existing local algorithms proved not to be accurate enough to reconstruct sufficient detail on the 3D surfaces.

Here, we propose a new modified area based local algorithm by adding the concept of a computation skip threshold (CST) to the basic steps in terms of the horizontal smoothness term. The aim of this approach is to improve graphical performance of existing area based local algorithms particularly for occlusions with speed improvement. Section 2 describes related work. Section 3 presents the new algorithm in detail. Section 4 compares the new algorithm with existing ones and shows all images used and results from experiments. Section 5 contains conclusions.

2 Related Work

The majority of area based local algorithms can be divided into 4 steps as below: [5]

1. Matching cost computations
2. Cost (support) aggregation
3. Disparity computation/optimization(WTA)
4. Disparity refinement

However, some local algorithms combine steps 1 and 2 and use a matching cost based on a correlated area (e.g. normalized cross-correlation [6]). For the matching cost commonly both the squared intensity differences (SD) [10] and the absolute intensity differences (AD) [9] have been used. Recently, there are new robust measures, such as truncated quadratics and contaminated Gaussians [7], which limit the influence of mismatches in the aggregation. However, this paper is concerned with only basic methods (AD and SD). In aggregation, window-based methods aggregate matching cost over a support region in the Disparity Space Image (DSI) [2] $C(x, y, d)$ (three-dimensional in x-y-d space) around a specific pixel of interest; the sum of absolute differences (SAD) or the sum of squared differences (SSD). Additionally, some efficient methods can be used with these techniques. For example, shiftable windows [8] can be used with a separable sliding min-filter. For disparity computation, a straightforward way to determine the best match for a point is to select the point of the other image, which shows the best similarity value within the range of disparities ($d_{min} \leq d \leq d_{max}$). This method is referred to as the 'Winner Takes All' (WTA).

Through the 1980's there has been much interest in feature based techniques due to their efficiency but the interest has declined in the last decade because of

improvements in robust global techniques. These methods use symbolic features rather than image intensities. Venkateswar and Chellappa [14] have proposed a hierarchical feature matching algorithm utilizing four types of features: lines, vertices, edges, and surfaces. Matching starts from the highest level of the hierarchy (surfaces) to the lowest (lines). Another feature based approach is to segment the images and then match the segmented regions [16] [15]. Birchfield and Tomasi [16] segment stereo images into small planar patches for which correspondence is determined and introduce an affine transformation model with parameters as follows:

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = A \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + d \quad (1)$$

where (x_1, y_1) and (x_2, y_2) are the coordinates of corresponding points in the left and right images. The vector d defines the translation of a segment between frames and the matrix A defines the in-plane rotation, scale, and shear transformations between frames. The parameters are calculated by spatio-temporal intensity gradients [17].

Global algorithms are formulated by the energy minimization framework [19]. Those methods often skip the aggregation step and intensively work in the disparity computation step to minimize a global energy shown as follows:

$$E(d) = E_{data}(d) + \lambda E_{smooth}(d) \quad (2)$$

The data term, $E_{data}(d)$ measures how well the disparity function d agrees with the input images. The smoothness term, $E_{smooth}(d)$ encodes the vertical and horizontal smoothness assumptions.

3 A New Modified Algorithm

In this algorithm, a new Computation Skip Threshold (CST) has been added to the cost and disparity computation steps of existing local algorithms in order to determine whether or not the point in the left image should be skipped in the WTA step. In the given stereo pair images; I_L (left) and I_R (right), the pixel intensity value of each point in each epipolar scanline of the I_L will be compared with that of the neighboring previous point in the same scanline of the I_L during the first cost computation step.

$$SKIP_L(x, y) = \begin{cases} 1 & \text{if } \|I_L(x, y) - I_L(x-1, y)\| \leq \delta \\ 0 & \text{if } \|I_L(x, y) - I_L(x-1, y)\| > \delta \end{cases} \quad (3)$$

In Equation 3, δ is the Computation Skip Threshold (CST) expressed by a pixel intensity value (from 0 to 255). In the first step $SKIP_L(x, y)$ of each pixel point is calculated according to Equation 3 before the point is calculated for the AD or SD process. The SKIP function simply marks 1 or 0 in each point but has no influence in existing calculations. In the aggregation step, there is no change

in computations. For the first two steps, the new algorithm is exactly the same as existing local algorithms except marking 0 or 1 for each point by the SKIP function. During the third WTA step, if $SKIP_L(x,y)$ is 1 the point (x,y) in the left image will be skipped and then the same disparity value of the previous point will be assigned to that point. Conversely, if it is 0 then the point (x,y) will be computed as normal so that the point (x,y) is regarded as where the image intensity is sufficiently changed to be calculated.

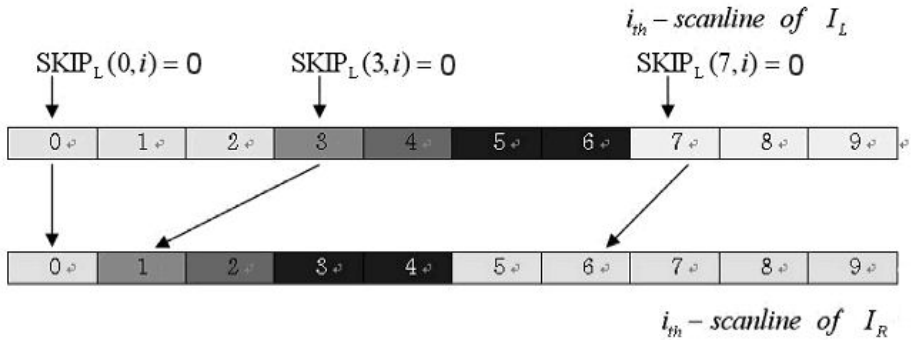


Fig. 1. An example of the i_{th} -scanline in a stereo image pair

Figure 1 illustrates how this method works in each scanline of a stereo image pair. In the cost computation step, $SKIP_L(x,i)$ of the first pixel in each scanline is forced to be 0 whereas, that of the remaining ones will be 0 only when the absolute difference between the current pixel value and the previous one in the same scanline of the I_L is bigger than CST. Thus, the value of $SKIP_L(x,i)$ for the three points $(I_L(0,i), I_L(3,i), I_L(7,i))$ where pixel intensity values are changed more than CST from the previous point is 0 and that for the others is 1. In the WTA step, only the three points of the left scanline are computed within the disparity range rather than all of the available 10 points. The disparity value of the remaining seven points will be assigned with that of the previous point of each point. For example, the disparity value of the three points $(I_L(4,i), I_L(5,i), I_L(6,i))$ will be 2 assigned by that of the point $I_L(3,i)$. In this way a large number of computations can be saved. If one assumes that the disparity range is 30 then we can skip 30×7 (points) computations but do only 30×3 (points). The aim of this approach is to compute only the points where pixel intensity values are changed in each scanline more than the CST value so that errors particularly at occlusions have been reduced with the horizontal smoothness term. Consequently, the new method produces an effect similar to that of DP and SO [1], especially for occluded regions by assigning the same disparity cost of the previous point of each point to the occluded points. In the next section, we evaluate the performance of the new algorithm with experiments.

4 Experiments and Results

All work has been done on the basis of the open sources from Scharstein and Szeliski [1] [3]. The experiments were conducted on an AMD Athlon 1.2GHz PC with 1GB RAM to compare the performance and computation speed of the new modified algorithm with those of existing fundamental local algorithms such as SAD+WTA and SSD+WTA. In all experiments, a 9x9 or 5x5 window with min-filter has been used with the commonly used images from Middlebury College [3] such as Map, Sawtooth and Tsukuba (gray level images) with the ground truth maps. The first experiment investigated the role of CST and also compared performances, particularly for occluded regions among three algorithms (Existing, New local algorithms and Dynamic Programming). Two factors such as bad pixel percentages(B) in all area and occluded regions [1] were calculated to evaluate the performance of each algorithm.

$$B = \frac{1}{N} \sum_{(x,y)} (|d_C(x,y) - d_T(x,y)| > \delta_d) \quad (4)$$

where $d_C(x,y)$ is the calculated disparity map and $d_T(x,y)$ is the ground truth map. δ_d is a disparity error tolerance, it is 1 here.

Figure 2 shows that the improved performance of the new algorithm at occluded regions has been simultaneously considered with the bad pixel percentage of the whole area in Map (284x216) and Sawtooth images (434x380). In Figure 2a while the CST increased from 5 to 15, the bad pixel percentage at occluded regions is considerably reduced from 90% in existing algorithms to 60% without increasing bad pixel errors across the image as a whole. Figure 2b also shows more than 30% decrease in error at occlusions. In both images bad pixels have been reduced around 30% but not as much as DP(see Figure 2a). This is for the reason that this algorithm adapts only the horizontal smoothness term but not

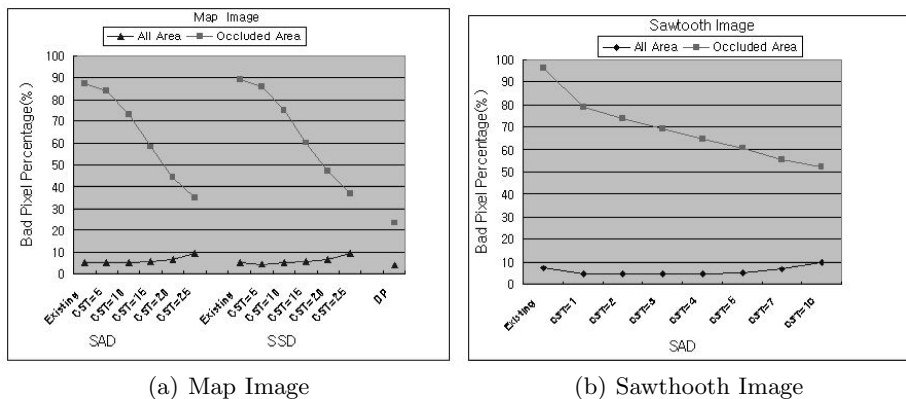


Fig. 2. Results of performance comparisons between existing and the new algorithms in Map and Sawtooth images with different CSTs

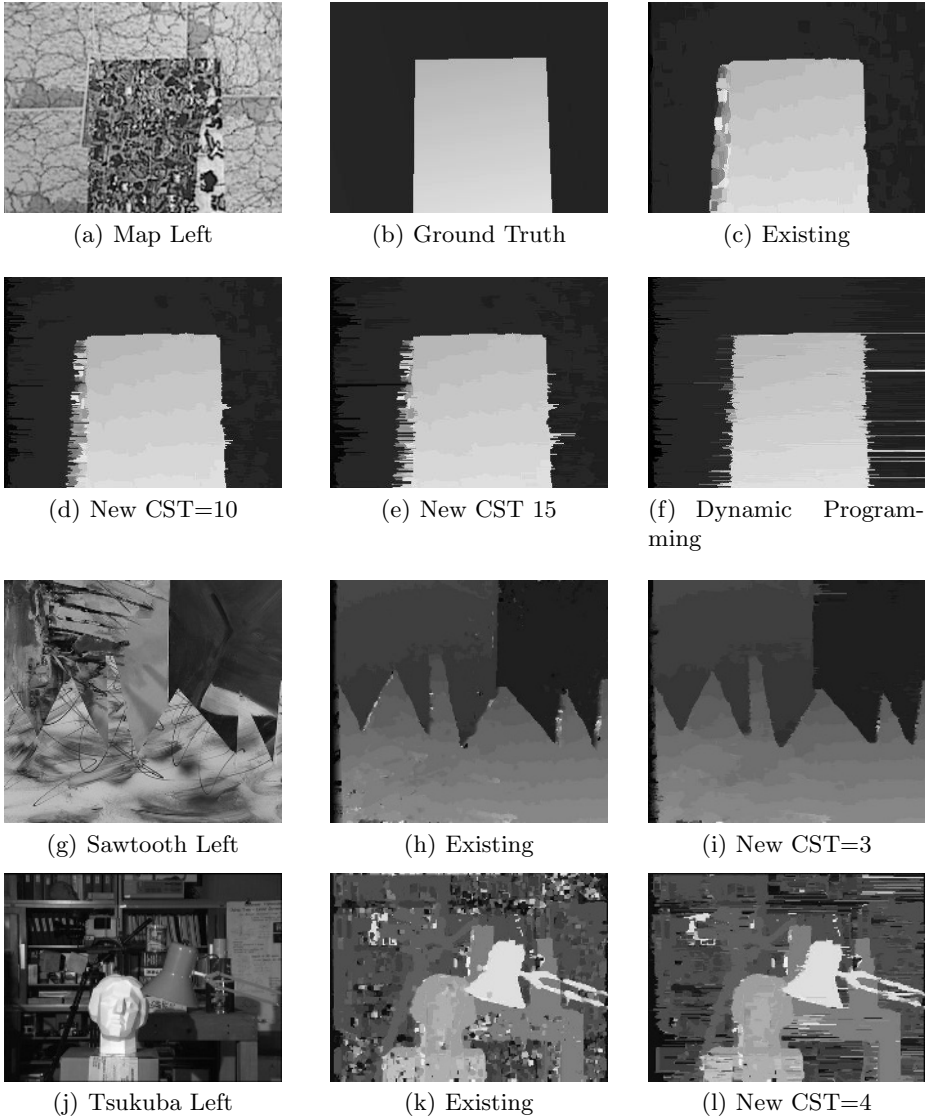


Fig. 3. Experimental images Map,Sawtooth: Window Size=9, Tsukuba: Window Size=5

the vertical one. However, from the points CST=20 in Figure 2a and CST=5 in Figure 2b the bad pixels in all areas start to be increased so that there seems to be a critical point for the CST. This problem is also the same problem of the difficulty of enforcing consistency causing the horizontal *streaks* in the disparity map of both DP and SO (see Figure 3). Using the map images, the critical CST is 15 (Figure 2a), and in the Sawtooth images the CST is 4 (Figure 2b). With

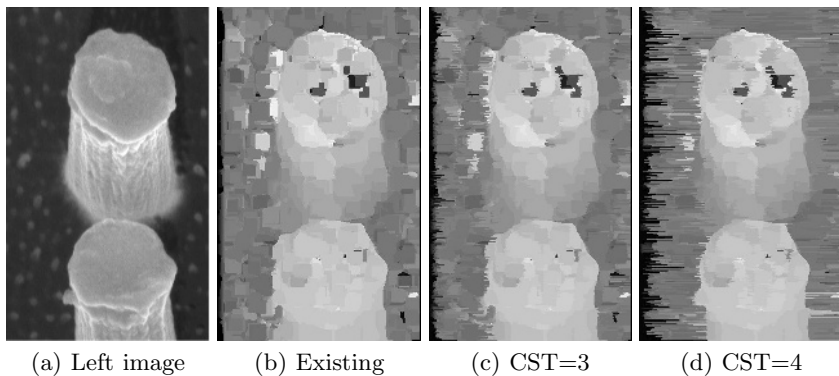


Fig. 4. Results from a stereo image pair generated by the SEM using a 21x21 window with min filter

Table 1. Results of computation time taken in both images as disparities are increased

Map Image				Sawtooth Image			
Dispa- -rity	Improv. (Sec)	Exist. (Sec)	New (Sec)	Dispa- -rity	Improv. (Sec)	Exist. (Sec)	New (Sec)
30	0	1.53	1.53	20	0	2.6	2.6
50	0.02	2.56	2.54	40	0.1	3.9	3.8
70	0.04	3.61	3.57	60	0.2	7.9	7.7

an appropriate CST value, the performance of the new algorithm for occluded regions will be between the existing local algorithm shown in the left hand side of Figure 2a and DP shown in the most right hand side.

In the second experiment, a comparison of computation speed between the existing and the new local algorithms in both images has been executed with different disparities. Table 1 shows results of computation time taken. In both images the computation time of the new algorithm is faster than the existing one for all cases. Moreover, in each image the improvement of the computation speed slightly increased as the range of disparities increased. For example, in the Sawtooth image of Table 1 for a disparity range of 20, it is the same speed and for a disparity range of 60, 0.2 seconds faster. Figure 3 shows disparity maps generated by the above experiments. Figure 3c is the map from the existing SSD with a 9x9 adaptive window. It shows bad performance at occluded regions of the left hand side of the object. From Figure 3d, the errors at occluded regions start to be reduced by the new algorithm with CST=10. In Figure 3e with CST=15 shows the best performance. As the value of CST is increased, errors in occluded regions significantly reduced so that the results are going towards those generated by DP (see Figure 3f). In the sawtooth images errors in the left hand side of the sawtooth have been reduced with CST=3(around 30% improvement)as shown

Figure 3g–i. Also, Tsukuba shows also graphically better results in Figure 3j–l with CST=4.

Figure 4 shows another example using smaller objects magnified at a smaller scale. A nano-scaled stereo image pair in high magnification (30,000) is generated by the SEM, using eucentric tilting within slightly different angles ($-5\ 5$). It also shows better performance at the occluded regions. Figure 4b is the disparity map generated by the existing SAD algorithm with a 21×21 window and min-filter. As a result, many errors at occlusions have been generated in the left hand side of each object. In Figure 4c when CST=3, the errors in left hand side of the upper object can be handled and when CST=4 even errors in the left hand side of the bottom object are also handled as shown Figure 4d.

5 Future Work and Conclusion

From all of the experiments, the new modified local algorithm apparently shows approximately a 30% improvement especially in occluded regions due to the added horizontal smoothness term and is moreover faster than existing local ones. The effect of it is similar to that of SO with no occlusion cost necessary and with the lack of the vertical smoothness terms. But both algorithms are different in that SO utilizes DP algorithms to compute the global minimum so as to take much more time than our algorithm. The computing time of the new algorithm is considerably improved when the range of disparities and the size of the images are quite large.

However, the best performance requires an optimized CST value for each image pair as DP also requires a smoothness weight(λ). Fortunately, from our experiments the optimized CST value of all images was 3 or 4 except the map images. It is for the same reason that in DP map images, which are well textured and only have two planar regions, are required for a high value for the smoothness weight(λ) as an input parameter in equation 2 while other images that have many objects at different depth levels are required for smaller values for the parameter [1]. Therefore, the CST value actually works in a similar way to the smoothness weight(λ). If CST is 0, the result of the new algorithm is exactly the same as that of the existing local algorithms. From CST=1 disparity maps of local algorithms start to become smooth up to an optimized CST point (mostly 4) in gray level images without losing the quality of the whole disparity map with speed improvements. This issue of more clearly finding the optimized CST value will be tackled in future work. Finally, we conclude that the new local algorithm adopting the horizontal smoothness can be applied to all kinds of local algorithms, which use the WTA, for a wide range of applications.

References

1. R. Szeliski D. Scharstein: *A taxonomy and evaluation of dense twoframe stereo correspondence algorithms*. IJCV, Vol 47 pp. 7-42, 2002.
2. S. S. Intille A. F. Bobick: *Large occlusion stereo*. IJCV, 33(3):181-200, 1999.

3. <http://www.middlebury.edu/stereo>.
4. D. Marr: *Vision*. W. H. Freeman and Company, New York, 1982.
5. D. Scharstein: *View Synthesis Using Stereo Vision*. volume 1583 of Lecture Notes in Computer Science (LNCS).Springer-Verlag, 1999.
6. M. J. Hannah R. C. Bolles, H. H. Baker: *The JISCT stereo evaluation*. DARPA Image Understanding Workshop, pages 263-274, 1993.
7. M. J. Black P. Anandan: *A framework for the robust estimation of optical flow*. ICCV, pages 231-236, 1993.
8. M. J. Black and P. Anandan: *A stereo matching algorithm with an adaptive window: theory and experiment*. In Proc. Image Understanding Workshop, pp. 383-389.
9. T. Kanade: *Development of a video-rate stereo machine*. In Image Understanding Workshop, pages 549-557, 1994.
10. T. Kanade L. Matthies, R. Szeliski: *Kalman filter-based algorithms for estimating depth from image sequences*. IJCV, 3:209-236, 1989.
11. D. Gavrilu H. Sunyoto, W. Mark: *A Comparative Study of Fast Dense Stereo Vision Algorithms*. IEEE Intelligent vehicles symposium, Parma, Italy, 2004.
12. J.M. Garibaldi H. Hirschmuller, P.R Innocent: *Real-Time Correlation-Based Stereo Vision with Reduced Border Errors*. IJCV, Vol. 47(1/2/3), pp. 229-246, 2002.
13. D. Murray and J. Little: *Using Real-Time Stereo Vision for Mobile Robot Navigation*. Autonomous Robots, Vol. 8 pp. 161-171, 2000.
14. V. Venkateswar and R. Chellappa: *Hierarchical Stereo and Motion Correspondence Using Feature Groupings*. Int'l J.Computer Vision, vol. 15, pp. 245-269, 2000.
15. S. Randriamasy and A. Gagalowicz: *Region Based Stereo Matching Oriented Image Processing*. Proc. Computer Vision and Pattern Recognition, pp. 736-737, 1991.
16. S. Birchfield and C. Tomasi: *Multiway Cut for Stereo and Motion with Slanted Surfaces*. Proc. Int'l Conf. Computer Vision, vol. 1 pp. 489-495, 1999.
17. J. Shi and C. Tomasi: *Good Features to Track*. Proc. Computer Vision and Pattern Recognition, 1994.
18. O. Veksler Y. Boykov and R. Zabih: *Fast approximate energy minimization via graph cuts*. IEEE TPAMI, 23(11):1222.1239, 2001.
19. D. Terzopoulos. *Regularization of inverse visual problems involving discontinuities*. IEEE TPAMI, 8(4):413.424, 1986.