Occlusion Handling in Trinocular Stereo using Composite Disparity Space Image

Mikhail Mozerov, Ariel Amato and Xavier Roca
Computer Vision Center, Departament Informatics
Universitat Autònoma de Barcelona, Barcelona, Spain
mozerov@cvc.uab.es

Abstract

In this paper we propose a method that smartly improves occlusion handling in stereo matching using trinocular stereo. The main idea is based on the assumption that any occluded region in a matched stereo pair (middle-left images) in general is not occluded in the opposite matched pair (middle-right images). Then two disparity space images (DSI) can be merged in one composite DSI. The proposed integration differs from the known approach that uses a cumulative cost. A dense disparity map is obtained with a global optimization algorithm using the proposed composite DSI. The experimental results are evaluated on the Middlebury data set, showing high performance of the proposed algorithm especially in the occluded regions. One of the top positions in the rank of the Middlebury website confirms the performance of our method to be competitive with the best stereo matching.

Keywords: Stereo vision, graph cut techniques, trinocular stereo.

1. INTRODUCTION

Stereo is a fundamental problem for a wide variety of tasks in computer vision [1]. A numerous of approaches have been proposed to solve the problem [2-5], but still a satisfying solution has not been received, since the stereo matching problem is an ill-posed one. Recently an excellent progress has been made in stereo matching due to the effectiveness of the global optimization techniques [6-9]. Nevertheless, occlusion remains a key problem in stereo matching.

Occlusion handling is one of the most important parts of many stereo matching techniques, and ignoring the occlusion can spoil the disparity map estimation by a sensitive inaccuracy. Recently many approaches were proposed to overcome the negative effect of occlusion in stereo matching [12-16]. The last taxonomy of such approaches can be found in the work [2].The constraints that are typically used in the process of occlusion handling are: ordering constraint, uniqueness constraint visibility constraint. The visibility constraint can also be considered a parameter to be recovered for each point of the stereo image frame like the desired disparity value itself. Formally, there exist two different cases of occlusion handling: two frame stereo and multi-frame stereo. However, from the reported researches, it is difficult to recognize the principal difference between these two approaches (see e.g. [10] and [13]). Nevertheless, it is necessary to realize that the two frames stereo differs fundamentally from the multi-frame stereo. For the one baseline stereo system invisibility of some region of points in one stereo image relative another point of view is almost inevitable norm. Hence, the task of occlusion handling in this case is to localize the stereo image regions where the matching between two frames has no sense, see a zoomed region in Fig. 1. Consequently, such information can be used for the further interpolation of the dense disparity map in these particular areas, or for modification of the data term in the energy minimization problem, like it has been done in the work [10].

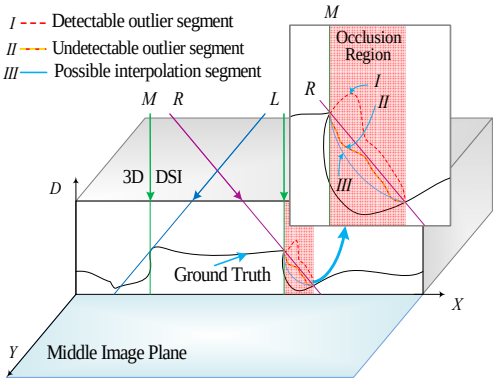


Figure 1: Occlusion region in 3D DSI.

However, even if we suppose that the right disparity map (R in Fig. 1.) is recovered perfectly in the left occluded region it is not guarantee that the occluded region can be detected. Indeed, detectable outlier in the left disparity map (I segment in Fig. 1.) makes a part of truly recovered right disparity map invisible. In contrast, for the multi baseline stereo it is feasible (except some special cases) to match every points in the middle frame with one of the neighbor frames, left or right. As a result, in general, reliable matching yields more accurate disparity map estimation than the interpolation. The problem is that without knowing the desired disparity map it is impossible to distinguish to which neighbor frame left or right the ambiguous region should correspond. Such a speculation force to use iterative approach: recover coarse disparity map and hence visibility then update data term for the further step of the algorithm iteration.

To avoid an iteration process it was proposed to use different cumulative costs for the multiple view matching [15]. This idea was adopted for occlusion handling in the work [16]: authors proposed to use acentric windows that improve algorithms with cumulative costs in the occluded regions. However such approach supposes to work with augmented costs that integrates pixelwise matching dissimilarity in a neighbourhood (of the image space or of a view sequences). Consequently, the pre-processing step smoothes edges of the resultant disparity map. This is why global optimization techniques are preferable versus local windows methods.

In this work we propose essentially new approach, which allows ridding of iterative process and in the same time avoiding negative consequences of the cumulative integration. To achieve this goal we introduce a composite 3D DSI (some researchers call it correlation volume), which is logical superposition (in contrast with cumulative summation) of two single DSI: the middle frame to the left frame DSI and the middle frame to the right frame DSI. Then a global optimization algorithm uses the prepared 3D array of the matching cost values to recover optimal disparity map. The rest of this paper is organized as follows: Section 2 describes the proposed composite DSI. Section 3 describes the full algorithm of the dense disparity map recovering. Section 4 is devoted to the experimental results. Concluding remarks are made in Section 5.

2. COMPOSITE DSI

The DSI representation is very popular in stereo matching [1, 2] due to the clear geometric interpretation of this model. Indeed, the desired disparity map should coincide with one of all the possible surfaces in the DSI. Furthermore, the global optimization approach assumes that an integral of the initial cost values plus inter-pixel smoothness term over such a surface should satisfy a chosen optimality criterion. For the trinocular stereo the 3D approach assumes that DSI has dimensions row $0 \le x \le X_{max}$, column $0 \le y \le X_{max}$, and disparity $0 \le d \le D_{max}$. All the three stereo images suppose to be rectified, each element (x, y, d) of the DSI projects to the pixel (x, y) in the middle image and to the pixel $(x-d, y)$ and $(x+d, y)$ in the left image and in the right image respectively. Let $E_{ML}(x, y, d)$ denote the DSI_{ML} cost value assigned to element (x, y, d) of the middle to left images and $E_{MR}(x, y, d)$ denote the DSI_{MR} cost value assigned to element (x, y, d) of the middle to right images matching space. All cost values (or pixelwise distance) are calculated using one of the convenient pixel-to-pixel matching metrics. In our work Euclidian distance between two compared color vectors is considered:

$$E_{ML}(x, y, d) = \left|\mathbf{I}_M(x, y) - \mathbf{I}_L(x + d, y)\right|, \qquad (1)$$
$$E_{MR}(x, y, d) = \left|\mathbf{I}_M(x, y) - \mathbf{I}_R(x - d, y)\right|,$$

where $\mathbf{I}_M(x,y)$, $\mathbf{I}_L(x,y)$ and $\mathbf{I}_R(x,y)$ are color vectors values of the middle left and the right stereo images respectively. Let us consider a very simple synthetic example: a square foreground patch on the plane background Figs. 2-4.The DSI related to the middle to left stereo matching (images a,b of Fig. 2) has an uncertainty region due to occlusion, a white triangle segment in the left part of Fig. 2 (d). A disparity map in Fig. 2. (e) obtained by a simple dynamic programming algorithm is indeed optimal, but not coincide with the ground truth of this pair. The same problem arises in the case of the middle to right stereo matching. One of possible way to solve occlusion problem in this case is to try merging two resultant disparity maps, but such a fitting might involve edge analysis or other iterative methods.

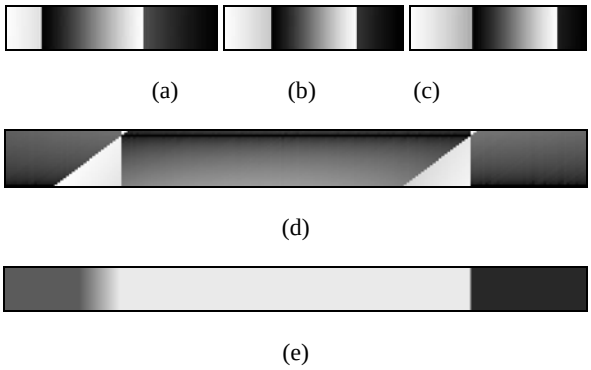


Figure 2: (a,b,c) – left middle and right synthetic images; (d) - a 2D slice of DSI_{ML}; (e) – an optimal disparity map related to DSI_{ML}.

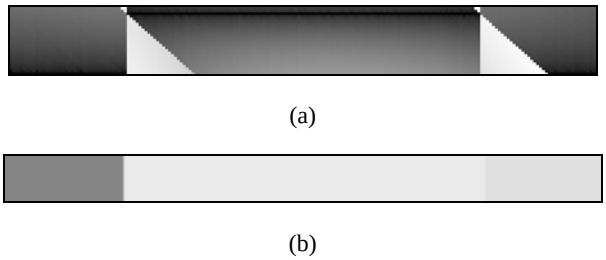


Figure 3: (a) - a 2D slice of DSI_{MR}; (b) – an optimal disparity map related to DSI_{MR}.

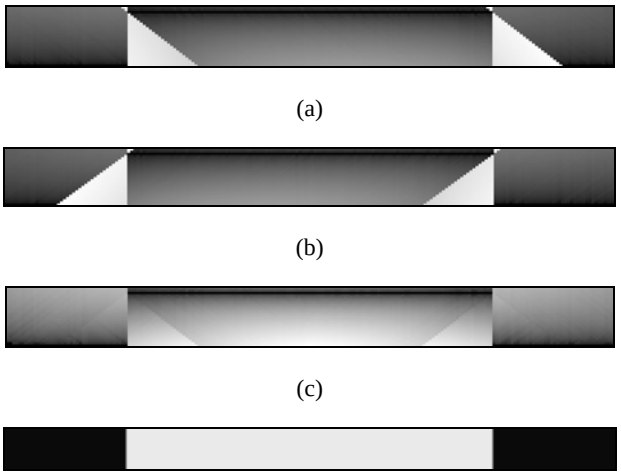
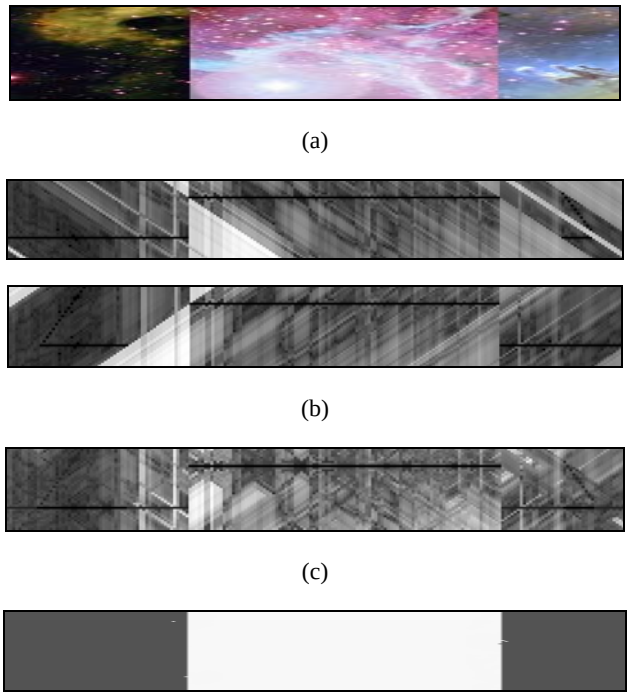


Figure 4: (a) - A 2D slice of DSI_{MR} ; (b) - a 2D slice of DSI_{ML} ; (c) - a 2D slice of composite DSI_C; (d) – an optimal disparity map related to DSI_C.

We propose a smart solution: merge two DSI before apply a global optimization algorithm (see Fig. 5.) in such a way that the resultant cost value of the composite DSI is going to be the minimum of two initial cost values of related E_{MR} and E_{ML}

$$E_C(x, y, d) = \min\left\{E_{ML}(x, y, d),\ E_{MR}(x, y, d)\right\}, \qquad (2)$$

were $E_C(x, y, d)$ is a cost function of the composite DSI.



(a) (b) (c) (d)

Figure 5: (a) - Middle image of synthetic stereo; (b) – are two 2D slices of DSI_{MR} ; and 2D slice of DSI_{ML} ; (c) – is a 2D slice of composite DSI_C; (e) – is an optimal disparity map related to DSI_C.

The advantage of such an approach is obvious, we solve a global optimization problem just once, and furthermore, the resultant solution in general coincides with ground truth and does not have to be corrected. To confirm our intuitive guess by visual representation we propose to consider another synthetic example with more rich texture on the reconstructed surface that is illustrated in Fig. 6. The black line of minimal costs coincides with ground truth, but for the DSI_{MR} and the DSI_{MR} this line is interrupted inside occlusion regions, in contrast in the line is has no uncertainty regions, see Fig. 5. (c).

One delicate moment in our method is the viewing geometry: the three camera centers lie at a single straight line perpendicular to the single viewing direction, the distance between middle left and middle right cameras are strictly equal. Such geometry is used for example in two base line stereo camera (e.g. Point Grey Research Inc.). To extend such a configuration over more general optical setup it is necessary just to preserve one line camera centers constraint. Of course, if optical axes of the cameras are not parallel the reciprocal homography of the three stereo images has to be done. The baselines inequality also can be solved in this case by a calibration process, but we have to note, that a considerable difference in the baseline values in trinocular stereo complicate a lot all calculations, due to inverse proportionality of the scene depth versus its image disparity. This is the reason, why we do not involve matching information of the left-right images pair in the reconstruction process: at the first glance the reconstruction accuracy for such pair is higher than for middle-left and middle-right, nevertheless DSI of this pair is not coincide with composite DSI.

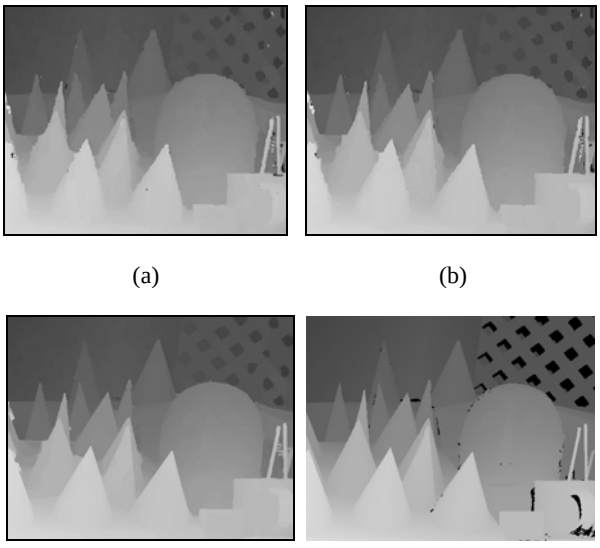
Of course, to obtain a good result with real stereo images it is necessary to use all standard steps of stereo matching, like preprocessing, global optimization and post processing. In the next section we describe our algorithm more precisely. However we have note that the solution of any 2D global optimization problem is always a tradeoff between accuracy and computational complexity. There exist a lot of papers which consider this problem, but the main goal of this work is to show the advantage of using composite DSI and we do not pretend in this paper to contribute in the field of MRF energy minimization.

3. MATCHIMG ALGORITHM

First step of our algorithm consists of calculation a composite DSI. This step was described in the previous section and based on Eqs. (1) - (2). Second step include standard preprocessing or filtering of the composite DSI. The most important part of our algorithm is the solution of the global optimization problem. It is natural to assume that the best matching of stereo images is achieved with the disparity map $P(x,y)$, which minimizes the objective function related with the composite DSI

$$O(P(x,y)) = \sum_{x,y,d \in P(x,y)} E_C(x,y,d) + \sum_{x,y,d \in P(x,y)} \sum_{i,j \in \Omega 1} S(x+i,y+j,d), \qquad (3)$$

where S is a smoothness term. Thus, we can formulate our problem as follow: find the disparity map function $P(x,y)$, which minimize the objective function in Eq. (3)



(a) (b) (c) (d)

Figure 6: (a) – Reconstruction only by using the global optimization step; (b) – reconstruction with preprocessing ; (c) – disparity map rectified with weighted median filter. (d) – ground truth of Cones.

$$P(x,y) = \arg\min_{P} O\big(P(x,y)\big). \qquad (4)$$

To solve the global optimization problem we use Graph Cut method [8], which is a modification of the very well known α-

expansion method proposed by Boykov at al. [6].

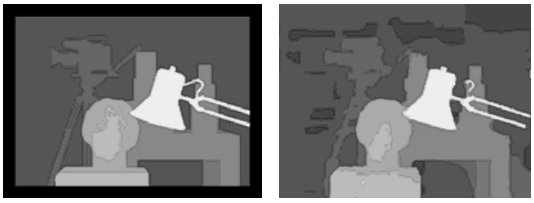
We also included an additional step for the resulting disparity map rectification. This step consist of the weighted median filtering

$$P_F(x,y) = \underset{i,j\in\Omega}{med}\left\{w_{i,j}P(x+i,y+j)\right\}; \qquad (5)$$

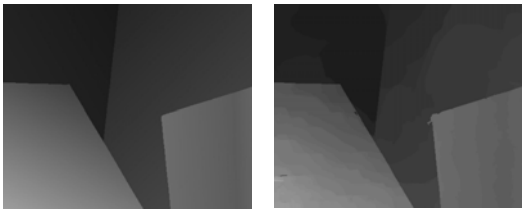
where Ω is a neighborhood region and w_{ij} is a weight factor that depends on the Euclidian distance between points (x,y) and $(x+i,y+j)$ and also depends on the Euclidian distance in the RGB color space between color vectors $\mathbf{I}_M(x,y)$ and $\mathbf{I}_M(x+i,y+j)$. The impact of each step of the algorithm is illustrated in Fig. 6.

4. EXPERIMENTAL RESULTS

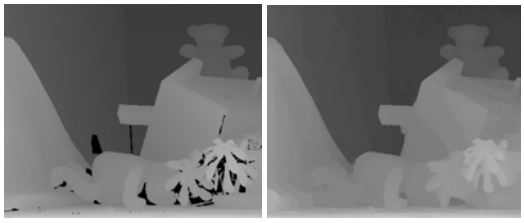
In this section experimental results are presented. In our experiments we used the stereo images benchmark from the Middlebury data set. Fig. 8. shows the corresponding disparity map obtained by our approach. The comparison with other stereo matching algorithms is sown in Table I. The first number in each column shows the error rate of the method, and the second number denotes the rank. Our method is in yellow. If the error threshold is chosen equal to 1.5 instead of traditional 1 like it is in second part of Table I, the performance of our method put the obtained result to the top of the rank. The ability of the proposed algorithm can be verified by the software that might be provided as additional material.



(a)



(b)



(c)

Figure 7: Middlebury benchmark test images: right column are the disparity maps obtained by the proposed algorithm, left column ground truth (a) – Tsukuba; (b) – Venus; (c) –Teddy.

The smoothness term in (3) is a cut linear function

$$S(x+i,y+j,d) = \lambda\begin{pmatrix}\left(\left|d(x,y)-d(x+i,y+j)\right|-1\right)\times\\ \times H\left(g-\left|d(x,y)-d(x+i,y+j)\right|\right)+1\end{pmatrix}; \qquad (6)$$

where $H()$ denotes Heaviside step function, parameters λ and g can set experimentally to optimize the output result, but as the rule of thumb we use λ is equal to mean value of the error term in (3) $\langle E_C(x,y,d)\rangle$ and the cut threshold g is equal to 3. Anyway, in [7] one can find the reason to choose one or another MRF energy representation for the objective function in (3).

Table I: The rank in Middlebury website evaluation.

Error Threshold = 1 (Error Threshold...) — Sort by nonocc / Sort by all / Sort by disc

Algorithm	Avg. Rank	Tsukuba ground truth nonocc	Tsukuba ground truth all	Tsukuba ground truth disc	Venus ground truth nonocc	Venus ground truth all	Venus ground truth disc	Teddy ground truth nonocc	Teddy ground truth all	Teddy ground truth disc	Cones ground truth nonocc	Cones ground truth all	Cones ground truth disc	Average Percent Bad Pixels
CoopRegion [41]	3.6	**0.87** 1	**1.16** 1	**4.61** 1	0.11 2	0.21 2	1.54 3	5.16 8	8.31 5	13.0 7	2.79 5	7.18 2	8.01 6	4.41
AdaptingBP [17]	3.7	1.11 8	1.37 4	5.79 9	**0.10** 1	0.21 3	**1.44** 1	4.22 3	7.06 3	11.8 4	**2.48** 1	7.92 4	7.32 3	4.23
DoubleBP [35]	4.5	0.88 3	1.29 2	4.76 3	0.13 4	0.45 8	1.87 6	3.53 2	8.30 4	**9.63** 1	2.90 6	8.78 11	7.79 4	4.19
YOUR METHOD	4.8	0.99 7	1.30 3	5.29 7	0.20 9	0.38 7	2.27 10	5.03 6	**6.37** 1	11.7 3	2.67 2	**4.03** 1	7.16 2	3.95
OutlierConf [42]	5.3	0.88 2	1.43 6	4.74 2	0.18 8	0.26 4	2.40 11	5.01 5	9.12 7	12.8 6	2.78 4	8.57 7	**6.99** 1	4.60
SubPixDoubleBP [30]	7.2	1.24 13	1.76 16	5.98 10	0.12 3	0.46 9	1.74 5	**3.45** 1	8.38 6	10.0 2	2.93 7	8.73 10	7.91 5	4.39
AdaptOwrSegBP [33]	11.8	1.69 25	2.04 23	5.64 8	0.14 5	**0.20** 1	1.47 2	7.04 17	11.1 9	16.4 13	3.60 13	8.96 10	8.84 12	5.59
SymBP+occ [7]	12.6	0.97 5	1.75 14	5.09 5	0.16 6	0.33 5	2.19 8	6.47 11	10.7 8	17.0 17	4.79 27	10.7 23	10.9 22	5.92

Error Threshold = 1.5 (Error Threshold...) — Sort by nonocc / Sort by all / Sort by disc

Algorithm	Avg. Rank	Tsukuba ground truth nonocc	Tsukuba ground truth all	Tsukuba ground truth disc	Venus ground truth nonocc	Venus ground truth all	Venus ground truth disc	Teddy ground truth nonocc	Teddy ground truth all	Teddy ground truth disc	Cones ground truth nonocc	Cones ground truth all	Cones ground truth disc	Average Percent Bad Pixels
YOUR METHOD	3.2	0.99 8	1.30 3	5.29 9	0.10 2	0.19 2	1.33 3	2.19 4	**3.38** 1	6.40 3	**1.65** 1	**2.85** 1	**4.91** 1	2.55
CoopRegion [41]	4.7	**0.87** 1	**1.16** 1	**4.61** 1	0.11 7	0.20 4	1.54 8	3.00 6	5.02 4	8.80 7	2.33 6	6.37 3	6.89 8	3.41
AdaptingBP [17]	4.8	1.08 9	1.33 5	5.79 10	0.10 4	0.20 3	1.42 5	**1.91** 1	3.63 2	**6.05** 1	2.18 5	7.06 6	6.52 7	3.11
DoubleBP [35]	4.8	0.88 3	1.29 2	4.76 4	0.11 5	0.42 13	1.47 6	2.03 2	5.46 6	6.38 2	1.98 2	7.50 10	5.92 3	3.18
SubPixDoubleBP [30]	5.6	0.92 4	1.33 4	5.00 6	0.11 5	0.42 12	1.47 6	2.03 3	5.59 6	6.57 4	2.02 3	7.56 11	6.04 4	3.26
OutlierConf [42]	5.9	0.88 2	1.43 7	4.74 3	0.16 11	0.22 5	2.17 12	2.95 5	5.73 7	8.42 6	2.05 4	7.27 7	5.58 2	3.47
AdaptOwrSegBP [33]	9.3	1.26 16	1.51 8	4.69 2	0.12 8	**0.16** 1	1.38 4	5.05 16	8.01 9	12.3 13	2.51 11	7.59 12	7.20 11	4.31

5. CONCLUSION

We propose a method to handle occlusion in stereo matching using trinocular stereo. The main advantage of the approach is that we solve a global optimization problem just once, and the resultant solution does not have to be corrected in the occluded regions. Three stereo images are used instead of two, thus competition with two image matching is not correct. Anyway, accumulation of information in a short baseline stereo does not automatically leads to more accurate results. Probably it is the reason why there are not a lot of announced comparisons with the Middlebury data set for more than two images. At least we are not aware about such evaluations.

6. ACKNOWLEDGEMENTS

This work has been supported by EC grant IST-027110 for the HERMES project and by the Spanish MEC under projects TIC2003-08865 and DPI-2004-5414. M. Mozerov acknowledges the support of the Ramon y Cajal research program, MEC, Spain.

7. REFERENCES

[1] D. Scharstein and R. Szeliski, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” *International Journal of Computer Vision,* 47(1/2/3):pp. 7-42, 2002.

[2] Y. Ohta and T. Kanade, “Stereo by intra – and intra-scanline search using dynamic programming,” *IEEE Trans.on PAMI*, 7(2): 139-154, 1985.

[3] S. Roy and I. J. Cox. “A maximum-flow formulation of the N-camera stereo correspondence problem,” *Proc. Int’l Conf. Computer Vision*, :492-499, 1998.

[4] H. Zhao. “Global optimal surface from stereo,” *Proc. Int’l Conf. Pattern Recognition*, 1:101-104, 2000.

[5] C. L. Zitnik and T. Kanade. A cooperative algorithm for stereo matching and occlusion detection. IEEE Trans. Pattern Analysis and Machine Intelligence, 22(7): 675-684, 2000.

[6] Y. Boykov, O. Veksler, and R. Zabih. “Fast approximate energy minimization via graph cuts,” *IEEE Trans.on PAMI*, 23(11): 1222-1239, 2001.

[7] V. Kolmogorov and R. Zabih. “What energy functions can be minimized via graph cuts,” *IEEE Trans.on PAMI*, 26(2): 147-159, 2004.

[8] J. Sun, N.N. Zheng, and H.Y. Shum, “Stereo matching using belief propagation,” *IEEE Trans.on PAMI*, 25(7):787-800, 2003.

[9] Pedro F. Felzenszwalb and Daniel P. Huttenlocher. “Efficient Belief Propagation for Early Vision,” *International Journal of Computer Vision*, 70(1):41-45, 2006.

[10] J. Sun, Y. Li , S.B. Kang, H-Y. Shum, “Symmetric stereo matching for occlusion handling,” *CVPR05*, (2):399-406, 2005.

[11] Y. Wei, L. Quan. “Asymmetrical occlusion handling using graph cut for multi-view stereo,” *CVPR05*, (2):902-909, 2005.

[12] Y. Wei, L. Quan. “Asymmetrical occlusion handling using graph cut for multi-view stereo,” *CVPR05*, (2):902-909, 2005.

[13] M.-A. Drouin, M. Trudeau, S. Roy. “Geo-consistency for wide multi-camera stereo,” CVPR05, (2):351-358, 2005.

[14] M.-A. Drouin, M. Trudeau, S. Roy. “Geo-consistency for wide multi-camera stereo,” CVPR05, (2):351-358, 2005.

[15] *****Proc. in IEEE Conference on* (CVPR*****),, *****.

[16] S.B. Kang, R. Szeliski, J. Chai, “Handling occlusions in dense multi-view stereo,” *Proc. in IEEE Conference on Computer Vision and Pattern Recognition* (CVPR 2001), pp. I-103-110, 2001.