Stereo Matching with Adaptive Support-Weight correlation and Graph Cuts

Limin Shi Fusheng Guo Wei Gao Zhanyi Hu

National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of sciences,
Beijing 100190, China

***Abstract*—Constructing a reliable data term and occlusion handling are two important issues for energy model based stereo method. In this paper, we at first use a 2-step adaptive support-weight correlation approach to get a reliable correlation volume. Then a pixel classification is proposed which classifies pixels into three classes: occluded, unstable and stable. For each pixel, according its class, a confidence weight is assigned. After that a new energy model is then constructed by integrating the correlation volume and the confidence weight. Finally ,through minimizing this energy using Graph cuts, a better disparity map is obtained. Experimental results on the Middlebury data set show that our proposed method has the similar good performance with the top rank Graph Cuts based algorithms listed on the Middlebury homepage.**

***Keywords*—stereo, Graph Cuts, adaptive supported window**

I. INTRODUCTION

Stereo is one of the most extensively researched topics in computer vision. In recent years stereo has made considerable progress, and many algorithms are presented. Daniel Scharstein et al. [1] category most of these algorithms into two classes: local(window-based) algorithms and global algorithms. In a stereo pair, matching pixels are recovered using disparities. Every pixel $p_1(i,j)$ in the reference image has a particular disparity d with respect to the matching pixel $p_2(i+d,j)$ in the target image. Local (window-based) algorithms, where the disparity computation at a given point depends only on intensity values within a local small window, usually make implicit smoothness assumptions by aggregating support. To obtain more accurate results on both smooth and discontinuous regions, an appropriate window should be selected adaptively for each pixel. That is, the window should be large enough to cover sufficient area in untextured regions, while small enough to avoid crossing depth discontinuities. Adaptive-window methods and multiple-window methods are two main types used to solve this ambiguity problem. Adaptive-window methods [2][3][4][5] try to find an optimal support window for each pixel. Kanade and Okutomi [2] presented a method to select an appropriate window by evaluating the local variation of intensity and disparity. But the shape of a support window is constrained to a rectangle, which is not appropriate for pixels near arbitrarily shaped depth discontinuities. Boykov et al. [3] tried to choose an arbitrarily shaped connected window. They performed plausibility hypothesis testing and computed a correct window for each pixel. Veksler [4][5] found a useful range of window sizes and shapes to explore while evaluating the window cost. However, the shapes of support windows used are not general and the method needs many user specified parameters for the window cost computation. Multiple-window methods [6][7][8] select an optimal support window among predefined multiple windows, which are located at different positions with the same shape. These methods need to perform correlation with many different windows for each pixel and retains the disparity with the smallest matching cost. The color-weighted approach proposed recently by Yoon and Kweon[9], instead of finding an optimal support window, adaptive support-weights are assigned to pixels in some large window based both on the color proximity and the spatial proximity to the pixel under consideration (the central pixel of the support window). This method can get good matching results at depth discontinuities as well as in homogeneous regions.

Global algorithms make explicit smoothness assumptions and then solve an optimization problem. Such algorithms seek a disparity assignment that minimizes a global cost function that combines data and smoothness terms. For this minimization problem, Belief Propagation[10][11][12] and Graph Cuts[13][14][15] are two of the most popular methods and they have achieved great success as their variants keep topping the comparison chart for the Middlebury datasets[1]. The BP algorithm works by passing messages around the graph defined by the four-connected image grid, and the global energy is observed empirically to converge after a certain number of iterations. Finally, for each pixel, the disparity that minimizes its energy is selected. Graph Cuts is a fast approximation algorithm for the multiple label energy minimization problem. The two most popular Graph Cuts algorithms, called the $\alpha-\beta$ swap and α expansion introduced in[13] work by repeatedly computing the global minimum of binary labeling problem in their inner loops. This process converges rapidly and results in a strong local minimum. In [16], Szeliski et al. compared these algorithms and found: α expansion could get lower energy with shorter time.

In this paper we integrate local algorithm and global algorithm to construct a robust and effective disparity estimation algorithm. Firstly, a 2-step initialization is proposed to get a reliable correlation volume by an adaptive supported window method. A confidence weight is also introduced to measure the reliability of the correlation cost for each pixel.

978-1-4244-6587-3/10/$25.00 É2010 IEEE

Based on this initialization, a more suitable energy model is constructed, and the final disparity map is rapidly obtained by minimizing this energy model using Graph Cuts.

The rest of the paper is organized as follows: in section 2, we review the two main algorithms used in this paper LASW and Graph Cuts. Section 3 presents our algorithm. Some experimental results are reported in section 4, followed by some concluding remarks in section 5.

II. OVERVIEW OF LASW AND GRAPH CUTS

1. LASW(local adaptive suppoted window)

LASW is a window-based method for correspondence search using varying support-weights proposed by Yoon and Kweon[9]. Since the window-based stereo algorithms explicitly use the frontal parallel plane assumption, they will results in the "foreground-fattening" phenomenon when the support windows are located on depth discontinuities. To eliminate the effect of the pixels which have different disparities with the reference pixel, LASW adjusts the support-weights of the pixels in a given support window based on color similarity and geometric distance to the reference pixel. The more similar the color of a pixel is, the larger its support-weight is. In addition, the closer the pixel is, the larger the support-weight is. Therefore, the support-weight of a pixel *q* in a support window of *p* is defined using the Laplacian kernel as

$$w(p.q) = \exp(-(\frac{\Delta c_{pq}}{\gamma_c} + \frac{\Delta g_{pq}}{\gamma_p})) \qquad (1)$$

where Δg_{pq} is the Euclidean distance between p and q , Δc_{pq} is the color difference of q and p in the CIELab color space. γ_p , γ_c are two parameters which can be determined empirically.

The difference between pixel colors is measured as

$$E(p,\overline{p}_d) = \frac{\sum_{q\in N_p,\overline{q}_d\in N_{\overline{pd}}} w(p,q)w(\overline{p}_d,\overline{q}_d)e(q,\overline{q}_d)}{\sum_{q\in N_p,\overline{q}_d\in N_{\overline{pd}}} w(p,q)w(\overline{p}_d,\overline{q}_d)} \qquad (2)$$

where $\overline{p}_d$ and $\overline{q}_d$ are the corresponding pixels in the target image when the pixel *p* and *q* in the reference image have a disparity value *d*. N_x is the support window around *x*. $e(q,\overline{q}_d)$ represents the pixel-based raw matching cost computed by using the colors of *q* and $\overline{q}_d$ as.

$$e(q,\overline{q}_d) = \min(\sum_{c\in\{r,g,b\}} |I_c(q) - I_c(\overline{q}_d)|, T)$$

T=40 is the truncation value that controls the limit of the matching cost.

After the dissimilarity computation, the disparity of each pixel is simply selected by the WTA (Winner-Takes-All) method without any global reasoning as

$$d_p = \min E(p,\overline{p}_d)$$

This method can get good results at depth discontinuities as well as in homogeneous regions. Especially the performance near the depth discontinuities is much better than that of other window-based methods because it can preserve arbitrarily shaped depth discontinuities well, whereas the methods using rectangular or constrained-shaped windows cannot.

2. Graph Cuts

Like many problems in early vision, the problem of recovering an accurate disparity map can be posed as an energy minimization problem under the MRF framework. For example one of the popular Energy models is the following one which has a data term and a smooth term,

$$E(f) = E_{data}(f) + E_{smooth}(f) \qquad (3)$$

The data term $E_{data}(f)$ measures how well the disparity function *f* agrees with the input image pair. The smoothness term $E_{smooth}(f)$ encodes the smoothness assumptions made by the algorithm.

In the last few years powerful energy minimization algorithms have been developed based on graph cuts [13][14][15]. These algorithms are fast enough to be practical, and yield quite promising experimental results for stereo. All these can be done largely due to the introduction of the α-expansion algorithm by Boykov et al.[13], which can efficiently find good minima of such energy models. The α-expansion is a fast approximation algorithm for the multiple label energy minimization problem. It starts with an arbitrary labeling and performs iterative optimization cycles until the process converges. Each cycle consists of iterating over the set of labels, running the α-expansion move once for every label α. This involves finding a new labeling f' obtained by increasing the number of α labels, which is better than the current labeling *f*, ie. $E(f') \le E(f)$. The algorithm will converge when in a particular cycle, no better f' can be found. Boykov et al. also analyzed the algorithm and proved bounds as well as state necessary properties of the penalty functions under which the bounds are correct.

III. OUR ALGORITHM

In the above energy model, data term usually plays an essential role. If the cost distribution of a data term is uninformative, the unreliable cost measurement will make the optimization problematic[17]. Most of stereo methods based on Graph Cuts only use pixel dissimilarity to compute the data term where strong matching ambiguity in textureless areas will occur and erroneous matching arises due to noise. In addition, occlusion handling is also an important issue affecting the matching accuracy. In this section, we construct a more reasonable energy model by integrating a reliable data term and an occlusion handling scheme which can be optimally minimized by Graph Cuts.

Our algorithm includes the following two main steps:

1) Initialization

In this step, we are to build a reliable correlation volume for each pixel in the reference image. In addition, a pixel classification is processed, which classifies the pixels into three classes: occluded, unstable and stable. For each pixel, according its class, a confidence weight is assigned.

2) Energy construction and optimization

Based on the above correlation volume and occluding labeling information a more reasonable energy model is constructed. The final disparity map is obtained by using Graph cuts to minimize the energy.

In the sequel, these two issues will be elaborated.

1. initialization

The main objective of the first step is to initialize a reliable correlation volume. We design a two-step initialization algorithm as follows.

a) Raw Initialization

To obtain a reliable correlation volume on both smooth and discontinuous regions, a variant of LASW approach is used as follows.

$$E(p,\overline{p}_d)=\frac{\sum_{q\in N_p,\overline{q}_d\in N_{\overline{pd}}} w(p,q)e(q,\overline{q}_d)}{\sum_{q\in N_p} w(p,q)}$$

where $e(q,\overline{q}_d)$ is the Birchfield and Tomasi's pixel dissimilarity[18]. The width of window is 9, and the parameters γ_p , γ_c are all 5. Compared to equation (2), here we only use the support-weight in the reference image. Even though Yoon and Kweon claimed that combing the support-weights in both support windows in the reference and target images is better than that only using the support-weight in the reference image, in our experiments, the support-weight in reference image is only used. Because we found that using only the support weight in the reference image can obtain better correspondence especially near discontinuous regions since the combined support-weight will generate the ambiguity of the foreground and background. An illustrative example is shown in Fig. 1. Suppose the black region is foreground, and the white region enclosed by green line is background. The supported window regions are enclosed by the red line. Combined support-weight will enlarge the effect of the foreground, and make the pixel p in the reference image match the correspond pixel q in the target image more harder.

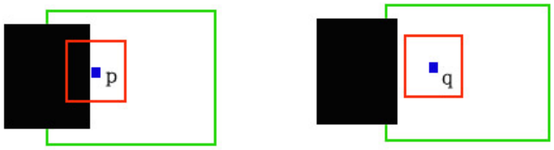

Fig.1. a synthetic stereo image pair.

Using the left image, then the right image as the reference image, we get the two correlation volumes. Disparity maps for both left image and right image can be obtained by selecting the best disparity by WTA(Winner-Takes-All). Then the mutual consistency check is carried out to detect occluded pixels. For a pixel p in the left image, if it dose not satisfy the following mutual consistency relation

$$D_L(p)=D_R(p-D_L(p))$$

It is marked occluded. D_L and D_R are the disparity maps of left image and right image respectively. If p satisfies this mutual consistency relation, we will classify it to stable or unstable according its correlation volume. Assuming that the cost for the best disparity value is $C_p(d_1)$ and the cost for the second best disparity value is $C_p(d_2)$, the correlation confidence is defined as

$$\frac{C_p(d_2)-C_p(d_1)}{C_p(d_2)}$$

If it is above a threshold $\alpha=0.4$, the pixel is regarded stable, otherwise unstable. Fig.2. shows the result of pixel classification on Tsukuba data set. Compared to the ground truth, we can see that most of the stable pixels have the accurate disparities in the disparity map, and most of the occluded pixels are precisely in the occluded region.

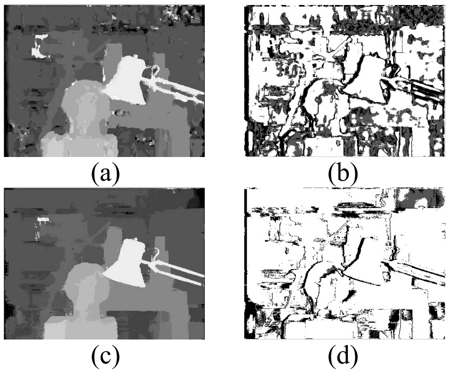

Fig. 2. The disparity maps((a),(c)) and pixel classifications((b),(d)) from raw initialization (top row)and refined initialization(below row) .in (c),(d), the black pixels are occluded, the gray pixels are unstable, and the other are stable.

b) Refined initialization

The disparity maps obtained from the above step could still contain many erroneous matches, consequently the resultant correlation volume is not reliable. Since now we have already some information about occlusion and matching stability, we can use it to refine our initial correlation volume. Support-weighted window based method is once again adopted as:

$$C_p^d=\frac{\sum_{q\in N_p} w(p,q)C_q(d)}{\sum_{q\in N_p} w(p,q)}$$

where $C_q(d)$ is the correlation cost of q with disparity d in the correlation volume obtained in the above step, and the width of the window is set to 35. We also change the weight definition as follows

$$w\ (p,q)=\lambda_q\cdot\exp(-(\frac{\Delta c_{pq}}{\gamma_c}+\frac{\Delta g_{pq}}{\gamma_p}))$$

γ_p , γ_c are 18 and 1 respectively. If q is occluded, $\lambda_q=0.01$. If q is unstable, λ_q =0.5, otherwise λ_q =1. The goal of using λ_q here is to propagate information from the stable pixels to the unstable and occluded pixels.

Having obtained the refined correlation volumes and the refined disparity maps for both the left image and the right image, the pixel classification operation is invoked again. And the confidence weight is assigned to each pixel according its class

$$CF_q = \begin{cases} 0.001, & q \text{ is occluded} \\ 0.2, & q \text{ is unstable} \\ 1.0, & q \text{ is stable} \end{cases}$$

Fig. 2 shows the disparity maps and the pixel classification of Tsukuba data set from the raw initialization and the refined initialization. After the refined operation, the accuracy is improved substantially. In Fig. 3, we also demonstrate the disparity maps from this initialization on the other three data sets used in our experiments.

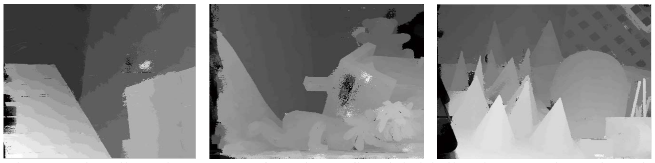

Fig. 3. The disparity maps after the refined initialization on the other three data sets used in our experiments.

We should point out that this initialization step is an effective one and is the main contribution of our work. On the one hand window-based local methods generally use some kind of statistical correlation between color or intensity patterns in local support windows. By using the local support windows, the image ambiguity is reduced effectively while the discriminative power of the similarity measure is increased, which can help us to construct more reliable data terms for energy model. On the other hand, the occlusion information and correlation confidence are also very useful for a more reasonable energy model construction.

Concerning the computational overload of our 2-step initialization, since the supported-weight correlation can be computed in parallel, we thought it is not a too serious issue in practice.

2. *Energy construction and optimization*

In this step a global optimization is taken to optimize the disparities map of the reference image(left image) from the re-initialization step. Based on the more reliable correlation volume and confidence weight from the re-initialization, we firstly construct an energy model. Then the Graph Cuts is used to minimize this energy and get the final disparity map. Here we use the traditional energy model as equation (3) which has data term and smooth term. The data term and smooth term are defined as follows respectively

$$E_{data}(f) = \sum\nolimits_{p \in I} D_p(f_p)$$

$$E_{smooth}(f) = \sum\nolimits_{p,q \in N} V_{pq}(f_p, f_q)$$

where I is the reference image, and N is the set of pairs of adjacent pixels. For a pixel p, its data term is the product of its correlation volume and confidence weight

$$D_p(f_p) = CF_p \cdot C_p^{f_p}$$

The purpose of introducing confidence weight into the data term is also to help to propagate information from the stable pixels to the unstable and occluded pixels.

Under the assumption of piecewise-constant surfaces, the smooth cost should decrease at depth edges. The color difference between neighboring pixels p and q is used to decide the amount of the decrease of the cost since the color edges are likely to coincide with the depth edges. So for a neighboring pixel pair p, q, the smooth term is defined as

$$V_{pq}(f_p, f_q) = d_{pq} \cdot \min(| f_p - f_q |, 5)$$

Where

$$d_{pq} = \begin{cases} 25 - g(p,q), & \text{if } g(p,q) < 20 \\ 1, & \text{else} \end{cases}$$

$$g(p,q) = \sum_{c \in \{r,g,b\}} | I_c(p) - I_c(q) | / 3$$

The final optimal disparity map is obtained using α expansion by minimizing the energy.

IV. EXPERIMENTS

The performance of our proposed algorithm is evaluated on the Middlebury data set. Fig. 4 shows the ground truths and the disparity maps by our algorithm.

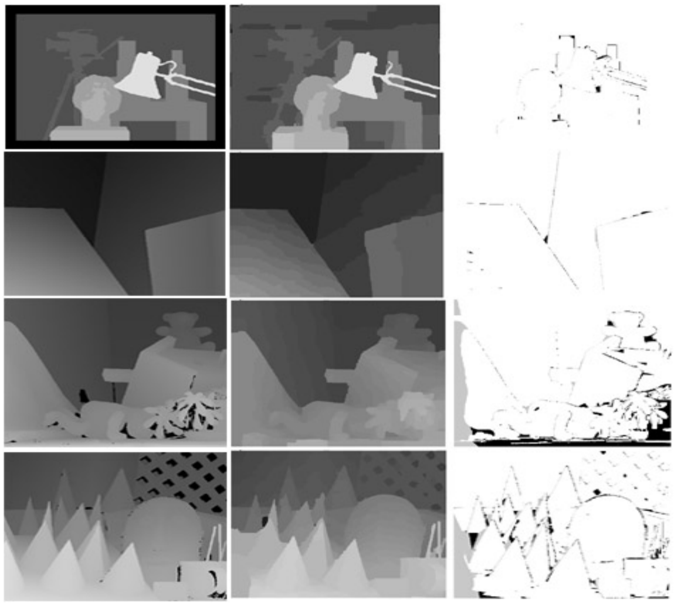

Fig. 4. The disparity maps by our algorithm for the four different standard test sets compared to the ground truth. Left column: ground truth. Middle column: our results. Right column: the bad pixels(black) detected by comparing our results with the ground truth.

Our algorithm is also compared with the top three Graph cuts based algorithms listed on the Middlebury homepage. The result on each data set is computed by measuring the percentage of pixels with an incorrect disparity estimate(error threshold is 1 pixel). This measure is computed for three subsets of the image.

- The subset of nonoccluded pixels, denoted as “nonocc”.
- The subset of the pixels near the occluded areas, denoted as ”disc”.
- The subset of the pixel being either nonoccluded or half-

occluded, denoted as ”all”.

Table I illustrates the performance of our algorithm and the other three Graph cuts based algorithms. For all the data sets, our algorithm performs well, especially in some data sets, we take the first or the second place, which demonstrates that our algorithm is robust and effective in disparity estimation.

TABLE I. Performance Comparison of the Proposed Method

Algorithm	Tsukuba			Venus			Teddy			Cones		
	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc
GC+Occ	1.19	2.01	6.24	1.64	2.19	6.75	12.0	11.2	18.8	5.36	12.4	13.0
GC+SegmBorder	1.47	1.82	7.86	0.19	0.31	2.44	4.25	5.55	10.9	4.99	5.78	8.66
MultiResGC	0.90	1.32	4.82	0.45	0.84	3.32	6.46	11.8	17.0	4.34	10.5	10.7
Proposed method	0.92	1.34	4.97	0.31	0.89	4.16	8.73	14.0	21.1	3.37	9.60	8.86

V. Conclusion

In this paper, we proposed an energy model based global optimization algorithm for stereo matching. Firstly, in the initialization step, an adaptive support-weight correlation is used to compute the cost volume. In addition, a pixel classification is introduced to indicate which pixel is occluded or has a reliable/unreliable correlation cost. And a confidence weight is assigned to each pixel according its class. Based on this initialization, a more reasonable energy model is constructed. By optimally minimizing the energy using Graph Cuts, the disparity map is obtained. Experimental results on the Middlebury data set show that our proposed method has the similar good performance with the top rank Graph Cuts based algorithms listed on the Middlebury homepage.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under the grant No (60835003, 60973005).

References

[1] D. Scharstein and R. Szeliski, “A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms,” Int’l J. Computer Vision, vol. 47, no. 1, pp. 7-42, 2002.

[2] T. Kanade and M. Okutomi, “A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiments,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 16, no. 9, pp. 920-932, Sept. 1994.

[3] Y. Boykov, O. Veksler, and R. Zabih, “A Variable Window Approach to Early Vision,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 12, pp. 1283-1294, Dec. 1998.

[4] O. Veksler, “Stereo Correspondence with Compact Windows via Minimum Ratio Cycle,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 12, pp. 1654-1660, Dec. 2002.

[5] O. Veksler, “Fast Variable Window for Stereo Correspondence using Integral Images,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 1, pp. 556-561, 2003.

[6] A. Fusiello, V. Roberto, and E. Trucco, “Efficient Stereo with Multiple Windowing,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 858-863, 1997.

[7] A.F. Bobick and S.S. Intille, “Large Occlusion Stereo,” Int’l J. Computer Vision, vol. 33, no. 3, pp. 181-200, 1999.

[8] S.B. Kang, R. Szeliski, and C. Jinxjang, “Handling Occlusions in Dense Multi-View Stereo,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 1, pp. 103-110, 2001.

[9] K.-J.Yoon and I.-S.Kweon, “Locally Adaptive Support-Weight Approach for Correspondence Search”, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 28, no. 4,pp.650-656, Apr. 2006.

[10] P.F. Felzenszwalb and D.P. Huttenlocher, “Efficient Belief Propagation for Early Vision,” Proc. IEEE CS Conf. Computer Vision and Pattern Recognition, vol. 1, pp. 261-268, 2004.

[11] J. Sun, N.-N. Zheng, and H.-Y. Shum. “Stereo matching using belief propagation.” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 25, No. 7, pp. 787-800, 2003.

[12] Q. Yang, L. Wang and R. Yang, et.al., “ Stereo Matching with Color-Weighted Correlation, Hierarchical Belief Propagation, and Occlusion Handling,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31, No. 3, pp. 1-13, Mar. 2009.

[13] Y. Boykov, O. Veksler, and R. Zabih, “Fast Approximate Energy Minimization via Graph Cuts,” IEEE Trans. Pattern Analysis an Machine Intelligence, vol. 23, No. 11, pp. 1222-1239, Nov. 2001.

[14] V. Kolmogorov and R. Zabih, “What Energy Functions Can be Minimized via Graph Cuts?” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 26, no. 2, pp. 147-159, Feb. 2004.

[15] Oliver Woodford, Philip Torr, lan Reid and Ansrew Fitzgibbon, “Global Stereo Reconstruction Under Second-Order Smoothness Priors”, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31, No. 12, pp. 2115-2128, Jan. 2009.

[16] R. Szeliski, R.Zabin, D.Scharstein et.al., “A Comparative Study of Energy Minimization Methods gor Markov Random Fields with Smoothness-Based Priors”, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 30, No. 6, pp. 1068-1080, Jun. 2008.

[17] Guofeng Zhang, Jiaya Jia, Tien-Tsin Wong, Hujun Bao, “Reconvering Consistent Video Depth Maps via Bundle Optimization”, Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 1-8, 2008.

[18] Birchfield and C. Tomasi, “A Pixel Dissimilarity Measure That Is Insensitive to Image Sampling,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, pp. 401-406, 1998.