Graph Cuts Stereo Matching Based on Patch-Match and Ground Control Points Constraint

Xiaoshui Huang[1,2](✉), Chun Yuan[2], and Jian Zhang[1]

[1] University of Technology Sydney, Sydney, Australia
[2] Graduate School at Shenzhen, Tsinghua University, Shenzhen, China
xiaoshui.huang@student.uts.edu.au

Abstract. Stereo matching methods based on Patch-Match obtain good results on complex texture regions but show poor ability on low texture regions. In this paper, a new method that integrates Patch-Match and graph cuts (GC) is proposed in order to achieve good results in both complex and low texture regions. A label is randomly assigned for each pixel and the label is optimized through propagation process. All these labels constitute a label space for each iteration in GC. Also, a Ground Control Points (GCPs) constraint term is added to the GC to overcome the disadvantages of Patch-Match stereo in low texture regions. The proposed method has the advantage of the spatial propagation of Patch-Match and the global property of GC. The results of experiments are tested on the Middlebury evaluation system and outperform all the other PatchMatch based methods.

Keywords: Patch-Match · Graph cuts · GCP · Stereo matching · Low texture

1 Introduction

Depth reconstruction is a technique that estimates depth from images. It is a baseline technique for building 3D objects and environments in stereo vision, but conducting reconstruction from several cameras is always an challenging problem. The performance of the depth reconstruction, including such elements as accuracy and efficiency, will directly impact the quality of computer vision, graphics and multimedia applications. One solution to the depth reconstruction problem is to conduct stereo matching from image pairs for depth estimation. This is an attractive research topic which has been applied to many applications.

The existing stereo matching methods are divided into two categories: global algorithms and local algorithms [12]. Global methods [7,13,15] consider the problem as an energy function that is solved by optimization methods. Although such methods have been shown to achieve precise disparity estimation in discontinuous regions, many challenging problems remains. Local methods [17,18] estimate disparity at a given centre pixel using colour or intensity values in the

Y.-S. Ho et al. (Eds.): PCM 2015, Part II, LNCS 9315, pp. 14–23, 2015.
DOI: 10.1007/978-3-319-24078-7_2

user-defined window. However, these algorithms have weaknesses in respect of disparity discontinuity and low texture areas, because they assume that the area inside the user-defined window is locally frontal-parallel.

Significant progress has recently been made in accuracy of stereo vision. One breakthrough that has obtained good accuracy results is a global method using GC. Another is the use of 3D models [2,3,11], in which a local 3D disparity plane is estimated for each pixel and accurate photo consistency between matching pixels is measured even with large matching windows. Sub-pixel precision is achieved with this method.

Even though GC obtains good results to some extent, it can only obtain disparity accuracy in pixel level accuracy. How to integrate such a 3D model with GC to achieve sub-pixel precision accuracy is a significant research topic. While stereo with standard 1D discrete disparity labels [6,14] is directly solved by discrete optimizers such as GC [4,7] and belief propagation(BP) [8], such approaches cannot be directly used for continuous 3D labels due to the infinite label space $(a, b, c) \in R^3$.

Many papers related to this research topic has been published [2,9,13]. Patch-Match [3] is a successful method used in [2,13] to efficiently infer an accurate 3D disparity plane using spatial propagation; each pixel's candidate plane is randomly selected, refined and then propagated to neighboring pixels [3]. In [2], Patch-Match is combined with BP to create an efficient stereo matching method, PMBP, for pairwise Markov random fields(MRF). However, BP is a MRF optimization method which is considered to be a sequential optimizer, in which each node is improved individually while other conditions are kept in the current state. In contrast, GC improves all nodes simultaneously by accounting for interactions across nodes, and this global property helps GC to avoid local minima [15]. A local shared labels strategy is proposed in [16] to combine GC and Patch-Match. This method has high computation complexity because of its still large label space, though it achieves better results than PMBP [2].

Patch-Match achieves great performance in complex texture regions, but low texture regions remain a challenge for these 3D model methods because of their local propagation strategy. Methods based on GCPs have recently achieved great improvements in low texture regions, but it is a challenging problem to acquire high quality GCPs. Occlusions are another challenge for the accurate computation of visual correspondence. Occluded pixels are visible in only one image, with no corresponding pixel in other images. Non-local (NL) algorithm [17] uses non-local strategy to detect these occlusion pixels, but it needs to compute a separate disparity map for left and right image.

In this paper, we focus on the accuracy of depth estimation in low texture, and our work also concentrates on complex texture regions, such as regions full of slant and edge areas, which is also a significant challenge in stereo matching. We propose a new label space computation method that effectively uses the global strengths of graph cut and local depth continuous strength of Patch-Match. Our method demonstrates great ability to handle complex texture areas as well as low texture areas. As discussed above, GC and Patch-Match cannot be directly combined, so a powerful method is required to effectively decide the label space for GC in infinite R3 sub-pixel space. Our new method has the ability to use both strengths and obtains good results at relatively high speed. At the same time, GCPs are used to improve the accuracy of disparity in low texture and occlusion areas. We propose a new GCP term called soft GCPs, which is unlike other GCP methods. It varies on different kind of GCPs.

Our contributions comprise a new method that incorporates spatial propagation of Patch-Match into GC-based global optimization and a new GCP constraint term that is used to guide the process of matching in low texture and occlusion regions. The major contributions of our method are as follows:

(1) A new label space computation method is proposed. We use Patch-Match to randomly select the disparity label of each pixel and compare with its 8 neighbors; the optimal label that produces minimum pixel dissimilarity is the current label. It is quite simple, very effective and obtains relatively good results. It has the advantages of Patch-Match and effectively combines Patch-Match and GC by considering a small number of labels in such large R3 infinite label space. The new label space computation method is reasonable because depth is locally continuous in most part of the images.
(2) A new GCPs constraint term is added to GC. We define the GCPs constraint term when a pixel is GCP, or we omit the term when the pixel is not GCP. This strategy is effective because it does not need a huge amount of computation on the GCP of every pixel. The GCPs constraint term is considered to be a soft constraint.It has the ability to improve energy minimization to achieve good disparity results.

2 Proposed Method

This section describes the proposed stereo matching method. Given two input images I_L and I_R, our purpose is to estimate the disparity maps of both images.

2.1 Problem Formulation

Sufficient label space is prerequisite for conducting Graph cuts(GC). As the main contribution of this paper, we develop a effective label selection algorithm to combine Patch-Match and GC. Initially, the optimal label for each pixel is selected in a local 3×3 window size, then the whole range of pixel label forms the label space of the whole image. Due to propagation steps in obtaining optimal label for each pixel, our label computation method has the advantages of obtaining more effective label space and maintaining good matcher.

Existing global methods (i.e., graph cuts) have achieved remarkable results, but the capability of the traditional Graph cuts(GC) stereo model is limited. To reduce ambiguities in matching, additional information is required to formulate an accurate model. In this paper, the GCPs constraints(G) are integrated as additional regularization terms to obtain a precise disparity estimation(D) for a scene with low texture characteristics. According to Bayes rule, the posterior probability over D given G and D_L is

$$p(D|D_L, G) = \frac{p(D_L, G|D)p(D)}{p(D_L, G)} \tag{1}$$

where D_L is the initial disparity map conducted by Patch-Match. Due to the initial disparity map D_L and G are constant during each iteration and independent, $p(D_L, G)$ is constant. Therefore,

$$p(D|D_L, G) \propto p(D_L, G|D)p(D) \propto p(D_L|D)p(G|D)p(D) \tag{2}$$

Maximizing this posterior is equivalent to minimizing its negative log likelihood [7], therefore disparity map (D) is obtained by minimizing the following energy function:

$$E(D) = E_d + E_s + E_G \tag{3}$$

2.2 Disparity Map Computation

Label Space Computation. The method of computing label space is divided into two steps: random selection and propagation. First, we randomly select the disparity label l_i of each pixel. After selection, we form the initial label space $L = (l_1, l_2, ..., l_i, ...l_n)$, where n is the number of pixels. Second, we optimize the initial label at each pixel, using its neighbors. Our goal in the second step is to form the optimal label space $S(S \subset L)$,

$$S = \cup_{i=1...k}\{s_i\} \tag{4}$$

where k is the dimension of S and s_i is the optimal label for p_i. s_i is obtained by selecting the optimal label that has the minimum ρ (through formulation (10)) compared with its 8-connected neighbors and its own last label,

$$s_i = l(\min_{l_j \in N_i}\{\rho(p_i, p_i')\}) \tag{5}$$

where N_i is the 3×3 labels centred in pixel p_i, p_i' is the projection of p_i with disparity l_j, $l(\cdot)$ is the label in N_i that has minimum ρ.

From formulation (4) and (5), we can obviously find that S is usually smaller than L, which is rational in reality. The new method obtains label space by propagation which is using the continuous property of neighbours' labels.

GC Settings Data Term. To measure photo-consistencies, we use a data term that was recently proposed by [18]. In the term, we use the 3D model to represent disparity. A pixels disparity d_i can be represented as $d_i = a_i x + b_i y + c_i$. Therefore, the objective of finding the best disparity for each pixel is to search its disparity plane $f_p = (a_i, b_i, c_i) \in R^3$. We find the best $f_p$though an energy method (e.g., GC) by minimizing $E(f_p)$. With the disparity f_p, we project the pixel $q = (q_x, q_y)^T$ in the left image to the right image by using the projection function:

$$\phi_{pq} = q - (a_p q_x + b_p q_y + c_p, 0)^T \tag{6}$$

Here, we only suppose it has x-axis disparity, because we use the Middlebury dataset to test our algorithm. If there is y-axis disparity, the $\phi_{pq} = q - (0, a_p q_x + b_p q_y + c_p)^T$; if there is both axis disparity, this function becomes

$$\phi_{pq} = q - (a_{px} q_x + b_{px} q_y + c_{px}, a_{py} q_x + b_{py} q_y + c_{py})^T \tag{7}$$

The data term of pixel p in the left image is defined as:

$$E_d = \sum_{q \in W_p} \omega(p, q) \rho(q, q') \tag{8}$$

Where W_p denotes a window centred on pixel p, pixel q' is the projection of pixel q, $q' = \phi_{pq}$ .The weight implements the adaptive support window proposed in [18], and defined as

$$\omega(p, q) = e^{-\|I^q - I^p\|_1 / \gamma} \tag{9}$$

The function $\rho(q, q')$ represents the pixel dissimilarity degree between q and q'. It is calculated by the following formulation,

$$\rho(q, q') = (1 - \alpha) \cdot \parallel I_q - I_{q'} \parallel_1 + \alpha \parallel \bigtriangledown I_q - \bigtriangledown I_{q'} \parallel_1 \tag{10}$$

where I is the intensity of the pixel and $\bigtriangledown I$ is the gradient of the pixel. In our experiment, we use Middlebury datasets, so there is only disparity on x-axis, thus $\nabla I = \bigtriangledown_x I$. If there is disparity on the y-axis, $\nabla I = \bigtriangledown_y I$; if on both axes, we define $\nabla I = sqrt(\bigtriangledown_x I \cdot \bigtriangledown_x I + \bigtriangledown_y I \cdot \bigtriangledown_y I)$.

Smooth Term. For the smooth term, a second-order regularization, curvature-based term [11] is used. It defined as

$$E_s(f_p, f_q) = max(\omega_{pq}, \varepsilon) min(S(f_p, f_q), \tau) \tag{11}$$

where ω_{pq} is a parameter which is defined in (9), it help to have a more robust smooth term. $S(f_p, f_q)$ estimates the penalty of the discontinuity between label f_p and label f_q in the same pixel. The penalty function is defined as

$$S(f_p, f_q) = |d_p(f_p) - d_p(f_q)| + |d_q(f_p) - d_q(f_q)| \tag{12}$$

GCP Constraint Term. GCPs are usually referred as highly reliable matched points that are used to guide the local structure of pixels. They improve the performance of GC by dealing with the problem of matching ambiguities in repetitive or low-texture areas. Because obtaining precision GCPs is a difficult task, GCPs constraint term is considered as a soft constraint. The GCPs constraint term is defined as

$$E_G = \lambda_G \sum_{p \in I_L} \lambda_{sg} \Psi(f_p, \hat{f}_p) \tag{13}$$

where $\Psi(f_p, \hat{f}_p)$ penalizes the disparity assignment that diverges from the Patch-Match. λ_G is a parameter that controls the GCPs constraint term and λ_{sg} controls GCPs term when the current pixel is GCP and when is not GCP,

$$\lambda_{sg} = \begin{cases} 0, & \text{if the pixel is not GCP} \quad (14) \\ 1, & \text{if the pixel is GCP} \quad (15) \end{cases}$$

In this paper, our robust penalty function $\Psi(f_p, \hat{f}_p)$ is derived from the Total Variance model [1] as

$$\Psi(a,b) = -In(1-\eta)exp(\frac{-\mid a-b \mid}{r} + \eta) \tag{16}$$

where r and η are parameters that control the sharpness and upper-bound of the robust function.

We use the methods in [9,17] to obtain GCPs, and the GCPs in I_L are obtained from several local stereo-matching algorithms. These methods include the left-right check procedure, which helps to find pixels which are occluded.

GC Optimization. We use the Fusion-move [8,11]method for optimization, which is an extension of the α-expansion algorithm [6]. In Fusion-move, an arbitrary values for each pixel is allowed instead of a fixed value for each optimization, and each optimization can be conducted by parallelization.

Optimization is conducted iteratively. The label space is computed using our label space computation method described in Sect. 3.2. The fusion-move method is then used to optimize the energy formulation (3). After each iteration, we update the last label space into a new one, which becomes the initialization of the next iteration. The optimization is continued until the energy converges or iteration times reach the maximum.

3 Experiments

The proposed method is evaluated on the Middlebury datasets and an explanation of the results is given. In addition, we compare the performance of our algorithm with GCPs and without GCPs. The performance of the proposed method is analysed.

Experiment Setting. The following settings is used throughout the experiments. We use a PC with an I5 CPU (2.50 GHz $\times$ 4 cores) and 16G memory. The parameters of our algorithm are set as $(\alpha, \tau, \varepsilon, \lambda_G, \gamma, \eta) = (0.9, 1.0, 0.01, 15, 2, 0.005)$. All labels of last label in formulation (5) is set into 0 when the first iteration of GC.

3.1 Evaluation on the Middlebury Datasets

Table 1 is our selected algorithms in the Middlebury evaluation database, there are all Patch-Match(PM) related methods ([7,16,17]). Our algorithm achieves

Table 1. Quantitative comparison of methods on the Middlebury datasets.

Algorithm	Avg. Rank	Tsukuba all	Venus all	Teddy all	Cones all	Average error
GC-LSL [13]	51.4	2.73	0.36	3.77	7.37	4.19
PMBP [2]	47.6	2.21	0.49	8.57	6.64	4.46
PM-PM [16]	42.2	3.23	0.30	8.27	6.43	4.52
PMF [10]	44.5	2.04	0.49	5.87	6.80	4.06
PM [3]	47.7	2.33	0.39	8.16	7.80	4.59
PM-Huber [5]	48.8	0.22	5.56	5.56	6.69	4.56
Ours	**40.4**	2.19	**0.21**	**3.11**	**4.32**	**3.73**

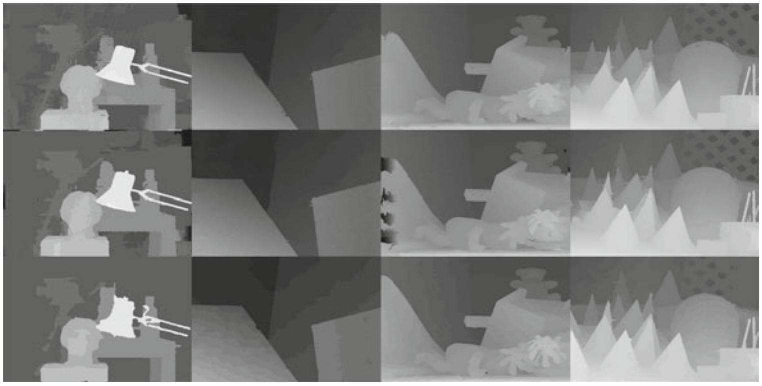

Fig. 1. The results of disparity maps of GC-LSL(row 1), PMBP(row 2), and the proposed method(row 3). The datasets are Tsukuba, Venus, Teddy, Cones, which is displayed in a row for each method.

the current best average error (3.73) and average rank (40.4) of all the PM based methods on threshold 1, and the proposed method ranked better than the others. In addition, our algorithm gets best results on the Venus, Teddy and Cones datasets.

We are closely compare our method to the related methods GC-LSL [13] and PMBP [2]. From the Table 1, we can see that our results achieve better accuracy compared to GC-LSL and PMBP on the four datatasets on threshold 1. In threshold 0.75, our algorithm can also obtain better results on the last three datasets(Venus, Teddy, Cones) than the others(The results can be seen on supplement materials). Because of our GCPs constraint term, low texture regions obtain guided information when they are finding matching pixels. Also, our new space computing method combine the strengths of PatchMatch with GC. The both strengths contribute the improvement of accuracy. Figure 1 illustrates the results evaluated on the Middlebury system. Row 3 shows our disparity results, from which we can clearly see that our results looks smoother and less artificial.

Fig. 2. Example regions that better than other methods. GC-LSL (col 1), PMBP (col 2), and the proposed method (col 3).

Figure 2 is the detailed comparison with GC-LSL and PMBP on Teddy and Tsukuba datasets. From the comparison results, we can see that our results (col 3) in the red rectangular region clearly obtain better results, demonstrating that our method achieves good results in both complex texture regions(row 1) and low texture regions(row 2).

The proposed method achieves better accuracy primarily because we find one optimal label space in each iteration of GC. It is a new way to compute label space, which illustrates that it is effective to conduct expansion steps in GC. Also, an additional GCP constraint term is added to our energy formulation which contributes to the improvements in low texture regions. As we discussed above, our algorithm can not obtain the best results compared to our related methods under threshold 0.75 in the first dataset of Middlebury (Tsukuba). It is because our GCPs are computed using some fast pixel level accuracy methods, while existed sub-pixel level accuracy methods have much slow speed and are not robust to get a same disparity in sub-pixel level accuracy. Such GCPs may have inaccurate points in sub-pixel level. It influences the accuracy. In real application, if we can use high accuracy device to obtain sparse high level accuracy GCPs, this drawback will be overcome.

3.2 Disparity Results with and Without GCPs

Our method contains two main contributions to stereo matching. One is a new graph cut combination with Path-Match, and the other is the introduction of GCPs into our new global method. In this section, we test them separately.

Figure 3 shows the disparity result of our algorithm both with and without considering GCPs. From the results, we can see that GCPs contributes significantly to improvement in the accuracy of the final disparity map. Table 2 gives the quantitative evaluation on Middlebury evaluation system. From the table, we can see that the disparity error of Teddy without considering GCPs is 6.3, while the disparity error considering the GCPs constraint term guide the

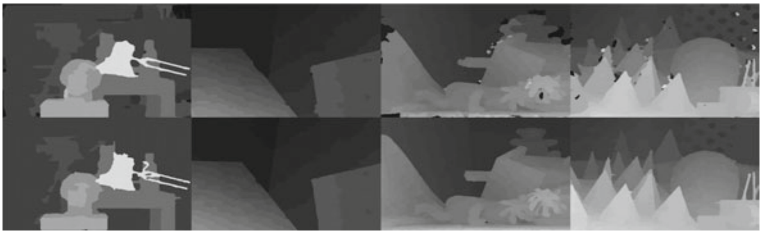

Fig. 3. Disparity map without GCPs(row 1) and with GCPs(row 2).

Table 2. Quantitative comparison of disparity with and without GCPs

Algorithm	Avg.	Tsukuba			Venus			Teddy			Cones			Average
	Rank	ncc	all	disc	ncc	all	disc	ncc	all	disc	ncc	all	disc	error
without GCPs	47.6	2.09	2.37	9.90	0.31	0.35	2.63	2.88	6.3	8.6	2.97	5.75	6.2	5.37
with GCPs	25.6	1.63	2.17	8.71	0.15	0.19	2.13	1.91	2.29	5.47	1.32	2.02	3.6	2.64

disparity computation is 2.29, which contributes to the improved results. The performance in other three Middlebury datasets (Tsukuba, Venus and Cones) has different extend improvement in its disparity results.

3.3 Performance of Proposed Method

Graph cuts generate the highest computation cost step in our algorithm. The proposed algorithm obtains higher computation speed than GC-LSL, because more effective label space is obtained for the GC. The proposed method iterates six times in each experiment which costs about 15 min. From their papers [2,13], we can find that, GC-LSL requires about 70 min when CPUs are used and PMBP need 16 min either.

4 Conclusion

A new method stereo matching method is proposed that obtains accuracy results in both the complex texture regions and low texture regions. The new algorithm integrates the Patch-Match stereo method into the graph cuts optimization method. In addition, a new GCPs constraint term is added to energy formulation of GC to obtain more accurate results in low texture regions.

Acknowledgement. This work is supported by the High Technology Development Program of China(863 Program), under Grant No.2011AA01A205, National Significant Science and Technology Projects of China, under Grant No.2013ZX01039001-002-003; by the NSFC project under Grant Nos. U1433112, and 61170253.

References

1. Barnes, C., Shechtman, E., Goldman, D.B., Finkelstein, A.: The generalized patch-match correspondence algorithm. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part III. LNCS, vol. 6313, pp. 29–43. Springer, Heidelberg (2010)
2. Besse, F., Rother, C., Fitzgibbon, A., Kautz, J.: PMBP: patchmatch belief propagation for correspondence field estimation. Int. J. Comput. Vis. **110**, 1–12 (2013)
3. Bleyer, M., Rhemann, C., Rother, C.: Patchmatch stereo-stereo matching with slanted support windows. In: BMVC, vol. 11, pp. 1–11 (2011)
4. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Trans. Pattern Anal. Mach. Intell. **26**(9), 1124–1137 (2004)
5. Heise, P., Klose, S., Jensen, B., Knoll, A.: Pm-huber: Patchmatch with huber regularization for stereo matching. In: IEEE International Conference on Computer Vision (ICCV), pp. 2360–2367. IEEE (2013)
6. Kolmogorov, V., Zabih, R.: Computing visual correspondence with occlusions using graph cuts. In: IEEE International Conference on Computer Vision (ICCV), vol. 2, pp. 508–515. IEEE (2001)
7. Kolmogorov, V., Zabin, R.: What energy functions can be minimized via graph cuts? IEEE Trans. Pattern Anal. Mach. Intell. **26**(2), 147–159 (2004)
8. Lempitsky, V., Rother, C., Blake, A.: Logcut - efficient graph cut optimization for markov random fields. In: International Conference on Computer Vision, pp. 1–8 (2007)
9. Liu, J., Li, C., Mei, F., Wang, Z.: 3D entity-based stereo matching with ground control points and joint second-order smoothness prior. Vis. Comput. **31**, 1–17 (2014)
10. Lu, J., Yang, H., Min, D., Do, M.N.: Patchmatch filter: efficient edge-aware filtering meets randomized search for fast correspondence field estimation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1854–1861. IEEE (2013)
11. Olsson, C., Ulén, J., Boykov, Y.: In defense of 3D-label stereo. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1730–1737. IEEE (2013)
12. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int. J. Comput. Vis. **47**(1–3), 7–42 (2002)
13. Taniai, T., Matsushita, Y., Naemura, T.: Graph cut based continuous stereo matching using locally shared labels. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1613–1620, June 2014
14. Wang, L., Yang, R.: Global stereo matching leveraged by sparse ground control points. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3033–3040. IEEE (2011)
15. Woodford, O., Torr, P., Reid, I., Fitzgibbon, A.: Global stereo reconstruction under second-order smoothness priors. IEEE Trans. Pattern Anal. Mach. Intell. **31**(12), 2115–2128 (2009)
16. Xu, S., Zhang, F., He, X., Shen, X., Zhang, X.: PM-PM: patchmatch with potts model for object segmentation and stereo matching. IEEE Trans. Image Process. **24**(7), 2182–2196 (2015)
17. Yang, Q.: A non-local cost aggregation method for stereo matching. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1402–1409, June 2012
18. Yoon, K.J., Kweon, I.S.: Adaptive support-weight approach for correspondence search. IEEE Trans. Pattern Anal. Mach. Intell. **28**(4), 650–656 (2006)

Springer

http://www.springer.com/978-3-319-24077-0

Advances in Multimedia Information Processing -- PCM 2015
16th Pacific-Rim Conference on Multimedia, Gwangju, South Korea, September 16-18, 2015, Proceedings, Part II
Ho, Y.-S.; Sang, J.; Ro, Y.M.; Kim, J.; Wu, F. (Eds.)
2015, XXIII, 683 p. 357 illus. in color., Softcover
ISBN: 978-3-319-24077-0