

Disparity estimation using Graph cuts for road applications

Mohamed El Ansari^{1,*}, Ilyas El jaafari^{2,**}, and Lahcen Koutti^{1,***}

¹LabSIV, Department of Computer Science, FSA, Ibn Zohr University, Agadir, Morocco.

²LabSIE, Department of Computer Science, FPO, Ibn Zohr University, Ouarzazate, Morocco.

Abstract. This paper proposes a new edge based stereo matching approach for road applications. The new approach consists in matching the edge points extracted from the input stereo images using temporal constraints. At the current frame, we propose to estimate a disparity range for each image line based on the disparity map of its preceding one. The stereo images are divided into multiple parts according to the estimated disparity ranges. The optimal solution of each part is independently approximated via the state-of-the-art energy minimization approach Graph cuts. The disparity search space at each image part is very small compared to the global one, which improves the results and reduces the execution time. Furthermore, as a similarity criterion between corresponding edge points, we propose a new cost function based on the intensity, the gradient magnitude and gradient orientation. The proposed method has been tested on virtual stereo images, and it has been compared to a recently proposed method and the results are satisfactory.

1 Introduction

In the field of vehicle navigation, stereo vision has mainly been applied to a large variety of applications, such as obstacle detection and tracking [1, 2], traffic sign detection and recognition [3–6], pedestrian detection and tracking [7], and so on. The key problem in stereo vision is the stereo matching problem, also called disparity estimation. By comparing information about a scene from two viewpoints, 3D information can be extracted by examining the relative positions of objects in the two stereo image [8, 9, 9, 10]. Which allows having accurate and detailed 3D representation of the environment around the Intelligent Vehicle (IV). Example of disparity estimation Methods are available in [11–15]. A taxonomy of dense disparity estimation algorithms together with a testbed for quantitative evaluation of stereo algorithms is provided by Scharstein and Szeliski [16]. It was demonstrated from [16] that Graph cuts methods [17–20] produce good results. However, they are time consuming, In order to avoid this problem in this work, the edges of the stereo images are extracted to reconstruct the scene.

The present work is devoted to road applications. Consequently, the matching algorithm will be applied to each stereo images acquired at each time. Incorporating temporal information in stereo approaches can improve the results of the matching as mentioned in [21–25].

*e-mail: m.elansari@uiz.ac.ma

**e-mail: eljaafari.ilyas@gmail.com

***e-mail: l.koutti@uiz.ac.ma

This work is an improvement of our recently proposed method [26]. Instead of using Dynamic Programming to match the edge points of the stereo images, we use the Graph Cuts algorithm to take into account the smoothness that should be between neighboring edge points. In addition, we improved the cost function used in [26] by considering the neighboring pixels of the edge points to be matched. The same idea in [26] is used to compute a disparity range for each image line. The stereo images are divided into multiple parts according to the estimated disparity ranges. Graph cuts algorithm is applied to match the edge points of each part independently. The estimated disparity ranges reduce the possible matches, which discard the false candidates and consequently improve the results. They also reduce the execution time of the applied energy minimization approach.

2 Related work

The use of sequence-frame stereo presents advantages over single-frame stereo. The temporal information may help the matching process to improve accuracy and to enforce temporal consistency of disparity maps in the case when the matching is ambiguous. In the literature, to improve the results of the stereo matching, temporal consistency has been used where different types of methods have been developed, e.g. optical flow, spatio-temporal and disparity prediction [27–31].

Optical flow methods extend the optical flow to 3D motion field and takes stereo and motion into account simultaneously [32]. Toa et al. [30] proposed a depth estimation method for non-rigid dynamic 3D scenes, in which the scene is represented by a collection of 3D piecewise planar surface patches based on color segmentation of input images. This representation is estimated by an incremental formulation. The spatial match measurement and the scene flow constraint [33, 34] are employed in the matching process. The algorithm's execution time and the accuracy of the results are limited by the image segmentation algorithm used. Hung et al. [35] proposed a depth and image scene flow estimation method to preserve motion-depth temporal consistency. Zhang et al. [34] proposed a 3D scene flow computation method, in which 3D motion model is fit to each local image region and adaptive global smoothness regularization is applied to the whole image.

Spatio temporal methods extend the spatial window used in the cost function to a spatio temporal window, which takes the spatial information of stereo image pairs and temporal information between consecutive images into account simultaneously. Zhang et al. [31] use spatial and temporal information by extending the spatial window to a spatio temporal window, spatial window used to compute the sum of squared difference (SSD) cost function, spatio temporal window to compute the sum of SSD (SSSD). Their method performs well in static scenes, but it fails to do so with dynamic scenes. Davis et al. [28] has developed a framework called space-time stereo similar to the one developed in [31] to recover shapes by studying the space-time windows, however, the method does not give good results in the dynamic scenes. Zhang et al. [36] presented a novel method for recovering depth maps from video sequence.

Disparity prediction methods use the results computed at the previous frame to compute the disparity map of the current frame. Jiang et al. A method to predict depth map between consecutive frames based on features detection, edge motion estimation and motion detection was proposed in [25, 37]. Dobias et al. [38] presented a method to transfer the depth map of the previous frame already computed to the current frame using estimated motion of the calibrated stereo rig. Cech et al. [39] predicted pixels correspondences based on the motion of pixels to compute a disparity map as well as an optical flow map between consecutive frames. In [29] an algorithm has been developed to compute both disparity maps and disparity flow

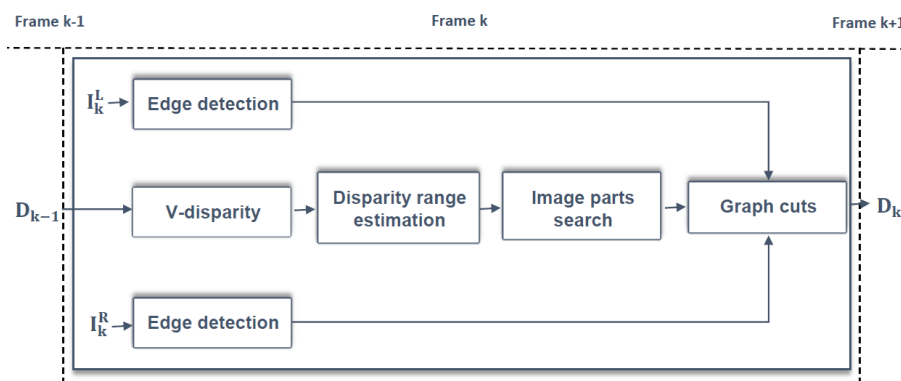


Figure 1: Main steps of the proposed algorithm.

maps in an integrated process. The disparity map generated for the current frame used to predict the disparity map for the next frame.

Other methods do not use any of the above types. El Ansari et al. [22] presented a stereo method that use temporal information in the matching process. A disparity range is deduced based on both an association algorithm and the disparity of the previous frame. The new disparity range is integrated in DP to compute the current disparity map. Mazoul et al. [24] match edges curves in the adjacent frames based on the same technique in [22]. The matched edge curves are then used to estimate a disparity ranges together with "matching control edge" points to divided the search space of the DP.

3 Proposed approach

In this section, we describe the main steps, summarized in Fig. 1, of the proposed stereo matching approach. We noted that the stereoscopic sensor used in our experiments is mounted aboard an IV and provides rectified images (i.e., corresponding pixels have the same y-coordinate). We start by mentioning the features to be matched by the proposed approach. Then, we present the constraints that the pairs of corresponding edge points should meet. Afterword, we present the proposed cost function. The last sub-section describes the Graph cuts algorithm used for the matching process.

The following notations are considered for the rest of the paper:

- I_k^L and I_k^R denote the left and right stereo images acquired at time k , respectively.
- D_k and D_{k-1} denote the disparity maps computed at the current frame and its preceding one, respectively.

3.1 Edge extraction

The first step consists in extracting significant features from the stereo images to be matched. In this work, we are interested in employing the edge points as features. The edges of the stereo images are extracted to reconstruct the scene in the goal to solve both the problem that large areas of pixels are similar and the time consuming of the dense reconstruction. To extract edges, we used canny edge operator [40] for the reason that it yields continuous edge curves which are vital to the proposed matching method and for its detection precision.

For the bottom part, the road is represented by an oblique line. We have only one possible disparity value for each image line. For the image line y_i , the only possible disparity is $(y_i - b)/a$, where a and b are the oblique line equation parameters. In order to take into account the uncertainty inherent to the estimation, the disparity range in this part is $[(y_i - b)/a - \alpha, (y_i - b)/a + \alpha]$ (see Fig. 2).

The line separating the two parts is L_0^k in which the disparity is equal to the disparity of the closest object, assuming the maximum displacement $L_0^k = a.(d_{k-1}^O + \Delta d_{max}) + b$. Fig. 2, illustrates how to get both disparity ranges from both top and bottom parts of the v-disparity map.

3.2.2 Image parts search

The two stereo images are divided into multiple parts according to the disparity ranges computed in the sub-section 3.2.1. In the top part of the v-disparity, we have only one disparity range that is $[0, d_{k-1}^O + \Delta d_{max} + \alpha]$. Thus, all the image lines belonging to this v-disparity part are considered as the first image part. However, in the bottom part of the v-disparity, we have a disparity range for each image line y_i which is $[(y_i - b)/a - \alpha, (y_i - b)/a + \alpha]$. We divide this v-disparity part into several image parts. Each part contains a well-defined number of image lines. The disparity range of each part is defined as follows:

- The minimum disparity value is $(y_i - b)/a - \alpha$, where y_i is the line that has the lower index i in the part.
- The maximum disparity value is $(y_i - b)/a + \alpha$, where y_i is the line that has the higher index i in the part.

Fig. 3, illustrates how to get the image parts and their disparity ranges from both top and bottom parts of the v-disparity map. The image part and its disparity range are represented with the same color. The first image part and its disparity range are deduced directly from the top part of the v-disparity. The disparity range in this part is $[0, 14]$. In this example, the bottom part of this v-disparity is divided into three image parts (part 2, part 3 and part 4). The disparity range of the part 3 for example is gotten as follows: The minimum disparity value of the first image line in this part is 9. The maximum disparity value of the last image line in this part is 22. Thus, the disparity range of this image part is $[9, 22]$.

3.3 Proposed cost function

One of the most important step of the stereo matching algorithms is the cost computation [42], which is crucial for quality of the disparity map. A given cost function is used to find similarities between corresponding features, which helps find the best candidate between the possible matches. In this work, as a similarity criterion between corresponding edge points, we propose a new cost function based on the one used in [26]. Which is defined based on the intensity, the gradient magnitude and the orientation at the edge points. Let's denote this cost function as C_{DIGO} (Difference of Intensity, Gradient and Orientation). Let e_i^L and e_i^R be two edge points belonging to two corresponding epipolar lines on the left and right images, C_{DIGO} is defined as follows:

$$C_{DIGO}(e_i^L, e_i^R) = \left\{ \left(I^L(e_i^L) - I^R(e_i^R) \right)^2 + M^L(e_i^L)^2 + M^R(e_i^R)^2 - 2 * M^L(e_i^L) * M^R(e_i^R) * \cos(\theta^L(e_i^L) - \theta^R(e_i^R)) \right\}^{1/2} \quad (1)$$

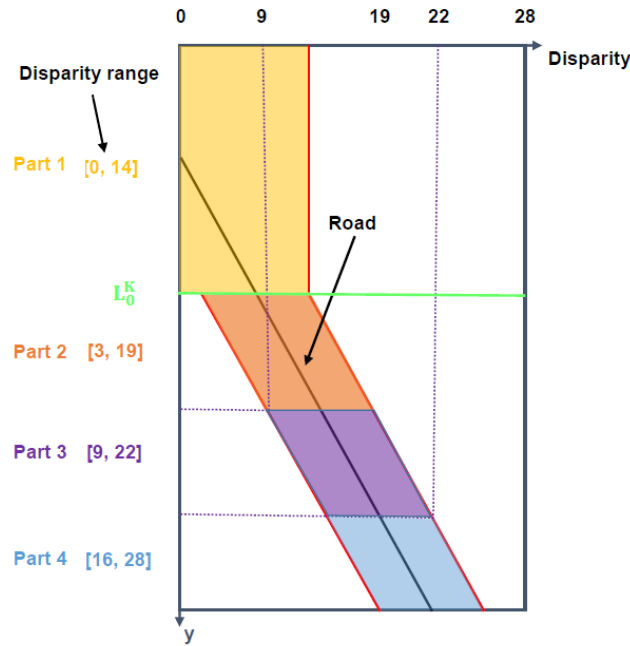


Figure 3: The image parts and their disparity ranges obtained from both top and bottom parts of the v-disparity map.

The cost function we propose take into account the considered pixel and its neighborhoods. It is defined as follows:

$$C(e_i^L(x, y), d) = \frac{1}{n} \sum_{(a,b) \in N(x,y)} C_{DIGO}(p(a, b), p(a, b - d)) \quad (2)$$

where d is the disparity, N are neighboring pixels of edge point e_i^L .

3.4 Graph cuts

As discussed in [18, 43], a graph is composed of a set V of nodes and a set E of directed edges, each with a non-negative weight. A cut on a graph is partitioning the nodes in the graph into two disjoint subsets V^S and V^T such that the source S is in V^S and the sink T is in V^T . The cost of the cut is the sum of the weights of the edges between the two partitions V^S and V^T . A minimum cut of the graph is a cut with minimal cost. The minimum cut problem can be solved by finding a maximum flow from the source S to the sink T . In [18, 43], the matching problem is considered as a minimization of an energy function (see equation 3), which is done based on graph cuts. The minimum cut corresponds to the minimum cost. The energy function of a configuration f is defined as follows:

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

This energy has four terms. The data term measures how well matched pairs fit, the occlusion term minimizes the number of occluded pixels and the last one forces neighboring pixels in the same image to have similar disparities. The data term used in [18, 43] imposes a penalty based on intensity differences of corresponding pixels. In this work, the data term is based on the cost function proposed in the section 3.3.



Figure 4: (left) Left image of the stereo pair #293 of the virtual stereo sequence and (right) its corresponding edge image.

The accurate choice of the maximum and the minimum disparity thresholds for almost any known stereo matching method is crucial to the quality of the output disparity map and the computation time. Instead of applying the energy minimization algorithm on the global image where the disparity range is very large, we apply it on each image part searched in the section 3.2.2 independently. The disparity range at each image part is very small compared to the global disparity range which improves the results and reduces the execution time. The graph cuts algorithm was applied on the edge points extracted from the stereo images.

4 Experimental results

In this section, we discuss the experimental results obtained by the proposed stereo matching approach on virtual stereo sequences. The new method has been compared with a recently proposed method [26], with the method proposed in [24] and with the same Graph cuts algorithm used in the new method without integrating temporal information. Let us refer to the proposed method as "New method", the one proposed in [26] as "Method 1", the one proposed in [24] as "Method 2" and the Graph cuts algorithm without temporal information as "Method 3". The hardware used in our experiments is a Lenovo T420 Intel(R) Core(TM) i5-5220M CPU 2.50GHz running under Windows 8.

The method has been tested on the MARS/PRESCAN virtual stereo sequence [44]. The dataset contains a sequence of 512×512 stereo images and their ground truth. Fig. 4 illustrates the left stereo image #293 of the virtual sequence and its corresponding edge image. Fig. 5 depicts the disparity map computed by new method at the frame #293. We used false color to make the disparity map clear. The blue color represents the nearest 3D points and the red one represents the farthest 3D points.

The Table 1 shows the comparison between the results obtained by the four methods at the frame #293. It provides, for each method the number of correct matches (NCM), the percentage of the correct matches (PCM) and the execution time (ETime) for the frame #293.

From Table 1 we remark clearly the improvements inherent to the new method. It has matched correctly more pairs of edge points, which means that it provides less sparse disparity map compared to other ones. The NCM of the new method, Method 1, Method 2 and Method 3 are 42905, 36811, 35163 and 40623, respectively. The new method matches correctly 6094, 7742 and 2282 pairs more compared to Method 1, Method 2 and Method 3, respectively. Related to the Method 1, the new method improves the PCM by 1.31%. Compared to the Method 2, the PCM is improved by 3.11%. And compared to the Method 3, the PCM is improved by 7.89%.

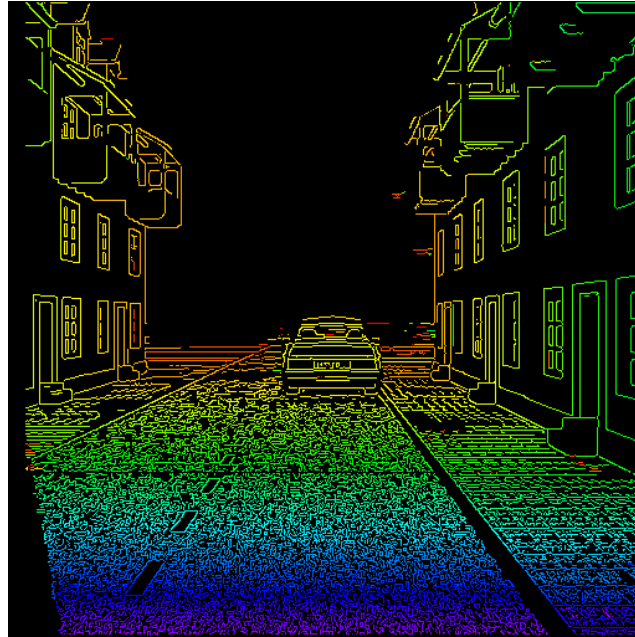


Figure 5: The disparity map computed at the frame #293 by the new method.

The results prove the success of the new approach compared to the three other methods. (Comparison between the four methods at other frames is given in Table 2).

Table 1: Summary of the results obtained by the new method, method 1, method 2 and method 3 at the frame #293.

Method	NCM	PCM (%)	ETime (ms)
New method	42905	97.54	231.13
Method 1	36811	96.23	109.71
Method 2	35163	94.43	181.72
Method 3	40623	89.65	391.84

5 conclusion

In this paper, the authors presented a new fast spatio temporal stereo matching method devoted to road applications. The main idea consists in using the disparity map obtained in the previous frame in the computation of one at the current frame. The disparity map of the preceding frame is served to compute a disparity range for each image line. The stereo images are divided into multiple parts according to the estimated disparity ranges. The optimal solution of each part is independently approximated via the Graph cuts algorithm. The search space at each image part is very small compared to the global search space which improves the results and reduces the execution time. To choose the best candidate between the possible matches, we proposed a new cost function based on the intensity, the gradient magnitude and gradient orientation. The proposed method has been tested on virtual sequences and the results are satisfactory. For future work, we will use the matching results to tackle the problem of obstacle detection and tracking.

Table 2: Summary of the results obtained by the new method, method 1, method 2 and method 3 at the frames between #290 and #299.

Method	PCM(%)										
	#290	#291	#292	#293	#294	#295	#296	#297	#298	#299	AVG
N. method	97.32	97.15	97.45	97.54	97.34	97.67	97.13	97.29	97.66	97.82	97.43
Method 1	95.69	95.78	96.14	96.23	96.15	96.09	96.17	96.03	95.40	96.51	96.01
Method 2	94.23	94.16	94.52	94.43	93.99	93.25	94.18	94.29	93.83	94.64	94.12
Method 3	89.17	89.40	90.10	89.65	89.94	89.19	88.97	89.85	90.15	89.69	89.61

References

- [1] M. Xie, L. Trassoudaine, J. Alizon, J. Gallice, *Machine Vision and Applications* **7**, 165 (1994)
- [2] I. El Jaafari, M. El Ansari, L. Koutti, A. Ellahyani, S. Charfi, *International Journal of Advanced Computer Science & Applications* **1**, 594 (2016)
- [3] S. Escalera, O. Pujol, P. Radeva, *Machine Vision and Applications* **21**, 99 (2010)
- [4] R. Timofte, K. Zimmermann, L. Van Gool, *Machine Vision and Applications* **25**, 633 (2014)
- [5] A. Ellahyani, M. El Ansari, I. El Jaafari, *Applied Soft Computing* **46**, 805 (2016)
- [6] A. Ellahyani, M. El Ansari, I. El Jaafari, S. Charfi, *International Journal of Advanced Computer Science & Applications* **1**, 686 (2016)
- [7] D.M. Gavrilu, S. Munder, *International Journal of Computer Vision* **73**, 41 (2006)
- [8] Y. Geng, *3D Research* **7**, 27 (2016)
- [9] S. Barnard, M. Fisher, *ACM Comput. Surveys* pp. 14:553–572 (1982)
- [10] A.S. Ogale, Y. Aloimonos, *International Journal of Computer Vision* **65**, 147 (2006)
- [11] X. Xiang, M. Zhang, G. Li, Y. He, Z. Pan, *Machine Vision and Applications* **23**, 1219 (2012)
- [12] T.G. Ghazouani, H., R. Zapata, *Int. J. Computational Vision and Robotics* pp. Vol. 2, No. 3, pp.237–253 (2011)
- [13] R.M.B.R.B.R. Gupta, G., B. Krishna, *Int. J. Computational Vision and Robotics* pp. Vol. 2, No. 1, pp.89–98 (2011)
- [14] S. Zhu, R. Gao, Z. Li, *Multimedia Tools and Applications* pp. 1–18 (2015)
- [15] I. El Jaafari, M. El Ansari, L. Koutti, *Signal, Image and Video Processing* pp. 1–8 (2016)
- [16] D. Scharstein, R. Szeliski, *International Journal of Computer Vision* pp. 47(1–3):7–42 (2002)
- [17] Y. Boykov, O. Veksler, R. Zabih, *IEEE Trans. Pattern Analysis and Machine Intelligence* pp. 23(11):1222–1239 (2001)
- [18] V. Kolmogorov, R. Zabih, *Int. Conf. on Computer Vision* pp. 508–515 (2001)
- [19] O. Veksler, Ph.D. thesis, Cornell University (1999)
- [20] L. Hong, G. Chen, *Segment-based stereo matching using graph cuts*, in *IEEE Conf. on Computer Vision and Pattern Recognition* (2004), pp. 74–81
- [21] M. El Ansari, S. Mousset, A. Benschair, *A new stereo matching approach for real-time road obstacle detection for situations with deteriorated visibility*, in *Intelligent Vehicles Symposium, 2008 IEEE* (IEEE, 2008), pp. 355–360

- [22] M. El Ansari, S. Mousset, A. Bensrhair, *Pattern Recognition Letters* **31**, 1226 (2010)
- [23] M. El Ansari, S. Mousset, A. Bensrhair, G. Bebis, *Temporal consistent fast stereo matching for advanced driver assistance systems (ADAS)*, in *Intelligent Vehicles Symposium (IV), 2010 IEEE* (IEEE, 2010), pp. 825–831
- [24] A. Mazoul, M. El Ansari, K. Zebbara, G. Bebis, *Pattern Analysis and Applications* **17**, 211 (2014)
- [25] J. Jiang, J. Cheng, B. Chen, X. Wu, *Neurocomputing* **142**, 335 (2014)
- [26] I. El Jaafari, M. El Ansari, L. Koutti, A. Mazoul, A. Ellahyani, *Neurocomputing* **194**, 24 (2016)
- [27] L. Koutti, I.E. Jaafari, M.E. Ansari, *Temporal consistent stereo matching approach for road applications*, in *2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA)* (2016), pp. 1–6
- [28] J. Davis, D. Nehab, R. Ramamoorthi, S. Rusinkiewicz, *IEEE Trans. Pattern Analysis and Machine Intelligence* pp. 27(2):1–7 (2005)
- [29] M. Gong, *Enforcing temporal consistency in real-time stereo estimation*, in *European Conference on Computer Vision* (2006), pp. 564–577
- [30] H. Tao, H.S. Sawhney, R. Kumar, *Dynamic depth recovery from multiple synchronized video streams*, in *IEEE Int. Conf. on Computer Vision and Pattern Recognition* (2001), p. .
- [31] L. Zhang, B. Curless, S.M. Seitz, *Spacetime stereo: shape recovery for dynamic scenes*, in *IEEE Int. Conf. on Computer Vision and Pattern Recognition* (2003), pp. 367–374
- [32] S. Vedula, P. Rander, R. Collins, T. Kanade, *Pattern Analysis and Machine Intelligence*, *IEEE Transactions on* **27**, 475 (2005)
- [33] S. Vedula, S. Baker, P. Rander, R. Collins, T. kanade, *Three-dimensioal scene flow*, in *IEEE Int. Conf. on Computer Vision and Pattern Recognition* (1991), pp. 722–729
- [34] Y. Zhang, C. Kambhamettu, *Integrated 3d scene flow and structure recovery from multi-view image sequences*, in *IEEE Int. Conf. on Computer Vision and Pattern Recognition* (2000), pp. 674–681
- [35] C. Hung, L. Xu, J. Jia, *Int. J. Comput. Vision* pp. 271–292 (2013)
- [36] G. Zhang, J. Jia, T. Wong, H. Bao, *IEEE Trans. Pattern Analysis and Machine Intelligence* pp. 974–988 (2009)
- [37] J. Jiang, J. Cheng, B. Chen, pp. 333–342 (2013)
- [38] M. Dobias, R. Sara, *Real-time global prediction for temporally stable stereo*, in *IEEE International Conference on Computer Vision Workshops* (2011), pp. 704–707
- [39] J. Cech, J. Riera, R. Horaud, *Scene flow estimation by growing correspondence seeds*, in *IEEE Conference on Computer Vision and Pattern Recognition* (2011), pp. 3129–3136
- [40] J. Canny, *IEEE Trans Pattern Anal Mach Intell* pp. 679–698 (1986)
- [41] M. Hariti, Y. Ruichek, A. Koukam, *A voting stereo matching method for real-time obstacle detection*, in *Robotics and Automation, 2003. Proceedings. ICRA '03. IEEE International Conference on* (2003), Vol. 2, pp. 1700–1704 vol.2, ISSN 1050-4729
- [42] K.r. Bae, B. Moon, *Multimedia Tools and Applications* pp. 1–16 (2016)
- [43] V. Kolmogorov, P. Monasse, P. Tan, *Image Processing On Line* **4**, 220 (2014)
- [44] *Mars/prescan virtual stereo images*, <http://stereodatasets.wvandermark.com/> (2006), accessed: 2016