

REVISITING GUIDED IMAGE FILTER BASED STEREO MATCHING AND SCANLINE OPTIMIZATION FOR IMPROVED DISPARITY ESTIMATION

Georgios A. Kordelas^{*†}, Dimitrios S. Alexiadis^{*}, Petros Daras^{*}, Senior Member, IEEE, and Ebroul Izquierdo[†], Senior Member, IEEE

^{*} Information Technologies Institute, CERTH, Thessaloniki, Greece

[†] Electronic Engineering and Computer Science department, Queen Mary, London, UK

ABSTRACT

In this paper the scanline optimization used for stereo matching, is revisited. In order to improve the performance of this semi-global technique, a new criterion to check depth discontinuity, is introduced. This criterion is defined according to the mean-shift-based image segmentation result. Additionally, this work proposes the employment of a pixel dissimilarity metric for the computation of the cost term, which is then provided to the guided image filter approach to estimate the initial cost volume. The algorithm is tested on the four images of the online Middlebury stereo evaluation benchmark. Moreover, it is tested on 27 additional Middlebury stereo pairs for assessing thoroughly its performance. The extended comparison verifies the efficiency of this work.

Index Terms— Stereo matching, Disparity map, Guided Image Filter, Scanline Optimization

1 Introduction

Stereo reconstruction is one of the most active research fields in computer vision [1]. Though mature, the task of estimating dense disparity maps from stereo image pairs is still challenging, while there is still space for improving accuracy, accelerating processing time and providing new ways of handling uniform areas, depth discontinuities and occlusions.

The work in [1] presents a complete taxonomy of approaches used for stereo disparity estimation. The categorization of the approaches is based on the following four generic steps, into which most of the stereo algorithms can be decomposed: 1. matching cost computation; 2. cost (support) aggregation; 3. disparity computation/optimization; and 4. disparity refinement.

Several metrics have been proposed in the literature for the computation of matching costs between pixels. Prevalent pixel-based cost measures include the absolute difference of image intensity values [2, 3, 4], gradient-based measures [2, 4] and non-parametric transforms such as CENSUS [3, 5] which are robust to radiometric distortions and noise. Many approaches use the combination of various cost measures in order to boost accuracy. The works in [3, 5], for example, exploit a combination of absolute intensity differences, as well

as the hamming distance of CENSUS transform coefficients. The matching cost values over all pixels and for all candidate disparities form the initial cost volume. In order to reduce matching ambiguity, the pixel-based matching costs are aggregated spatially over support regions in the initial cost volume. Performance evaluations on previous cost aggregation approaches are presented in [6] and [7]. More recently, cost aggregation approaches include guided image filter [2] and Successive weighted summation [5].

In general there are two types of approaches, global and local ones, regarding the disparity optimization. Local methods [2, 5] put emphasis on matching cost computation and cost aggregation. The final disparity map is computed by applying a simple local winner-take-all (WTA) approach independently for each pixel. Global optimization methods aim at assigning a disparity label to each pixel so that a cost function is minimized over the whole image area. Efficient techniques include Graph Cuts [8] and Belief Propagation [9]. In an additional category of approaches, the energy function is minimized on a subset of points of the stereo pair (semi-global methods), for instance along 1D paths. Such approaches, which decrease the computational complexity compared to global optimization algorithms, involve Dynamic Programming [10] or Scanline Optimization [11] techniques.

The disparity results have to be refined, since they are polluted with outliers in occluded areas, depth discontinuities and uniform areas that lack texture. Several stereo algorithms, such as [11], use segmented regions for reliable outlier handling. The work in [3] uses iterative region voting and proper interpolation to fill outliers.

This paper introduces a new criterion for the definition of the smoothness penalty terms that are used in the semi-global scanline optimization method of [11]. The definition of this criterion is based on the result of the mean-shift based image segmentation. Additionally, the estimation of the initial cost volume via the guided image filter is improved by the employment of the dissimilarity measure proposed in [12].

The rest of this paper is organized as follows. Section 2 presents the method used for the estimation of the initial cost volume. In section 3, information regarding improvements of the scanline optimization step, is provided. Section 4 gives information on the used parameters, as well as the experimental results, while conclusions are drawn in Section 5.

The authors are grateful for support from the EU-funded IP project REVERIE under contract 287723.

2 Initial Cost Volume Computation

2.1 Matching Cost Computation

This paper exploits the approach of [4], which is inspired by the previous work of [2], in order to generate the initial cost volume. However, this paper proposes a modification regarding one of the cost terms used to estimate the final cost, leading to improved disparity results. This approach uses three different cost terms to compute the matching cost function $C_{all}(p, d)$ for a pixel p at each allowed disparity d .

The first term is based on the Gabor-Feature-Image. Let us suppose that I_l and I_r stand for the left and right images in grayscale. Their corresponding Gabor-Features-Images are $G_H(I_l(p))$ and $G_H(I_r(p))$, respectively (the parameters used for the extraction of the Gabor-Feature-Image are the same as in [4]). The cost term $C_{gab}(p, d)$ for pixel p at disparity d is given by:

$$C_{gab}(p, d) = |G_H(I_l(p)) - G_H(I_r(p-d))|, \quad (1)$$

Gradient information is also employed for the computation of a gradient based cost term. In the same sense as in (1), the gradient-based cost term is given by:

$$C_{gra}(p, d) = |\nabla_H(I_l(p)) - \nabla_H(I_r(p-d))|, \quad (2)$$

where $\nabla_H(I(p))$ denotes the gradient in horizontal direction at pixel p on image I . In [4], the third data term corresponds to the mean of the sum of absolute differences on pixel RGB values between the left and right images. However, in this work, the replacement of the third term is proposed with the dissimilarity metric on RGB values presented in [12]. Therefore the third term is given by:

$$C_{rgb}(p, d) = \sum_{c=R,G,B} \frac{D^c(p, p-d)}{3}, \quad (3)$$

where $D(p, p-d) = \min(\bar{D}(p, p-d, I_l, I_r), \bar{D}(p-d, p, I_r, I_l))$.

Finally, the combined matching cost function is expressed as:

$$C_{all}(p, d) = \alpha_1 \cdot \min(C_{gab}(p, d), T_{gab}) + \alpha_2 \cdot \min(C_{gra}(p, d), T_{gra}) + (1 - \alpha_1 - \alpha_2) \cdot \min(C_{rgb}(p, d), T_{rgb}) \quad (4)$$

where α_1, α_2 are the weight parameters and $T_{gab}, T_{gra}, T_{rgb}$ are truncation thresholds used to reject the outliers.

2.2 Cost Aggregation

In order to reduce matching ambiguity, the pixel-based matching costs $C_{all}(p, d)$ are filtered using the guided image filter [13]. I is employed as color the guidance image and $C_{all}(p, d)$ as the guided image. The filtered cost value of pixel p at disparity d is given by:

$$C'_{all}(p, d) = \sum_q W_{p,q}(I) C_{all}(q, d) \quad (5)$$

The filter weights $W_{p,q}$ are expressed as:

$$W_{p,q} = \frac{1}{|\omega|^2} \sum_{k:(p,q) \in \omega_k} \left(1 + (I_p - \mu_k)^T (\Sigma_k + \varepsilon U)^{-1} (I_q - \mu_k)\right), \quad (6)$$

where $|\omega|$ is the total number of pixels in a window ω_k centered at pixel k and ε is a smoothness parameter. Σ_k and μ_k are the covariance and the mean of pixels intensities within ω_k . I_p, I_q and μ_k are 3×1 (color) vectors, while Σ_k and the unary matrix U are of size 3×3 .

The main advantage of the guided filter is that the computation cost is independent to the size of the selected window. This is because it can be expressed as a linear transform as follows:

$$C'_{all}(p) = \frac{1}{|\omega|} \sum_{k:p \in \omega_k} (\alpha_k I_p + b_k) \quad (7)$$

$$a_k = (\Sigma_k + \varepsilon U)^{-1} \left(\frac{1}{|\omega|} \sum_{p \in \omega_k} I_p C_{all}(p) - \mu_k \bar{C}_{all}(k) \right) \quad (8)$$

$$b_k = \bar{C}_{all}(k) - a_k^T \mu_k, \quad (9)$$

where $\bar{C}_{all}(k)$ is the mean of C_{all} in ω_k .

3 Disparity Optimization

3.1 Improved Scanline Optimization

There are multiple disparity optimization approaches. Semi-global scanline optimization is one of the most efficient methodologies [11]. It gives accurate disparity results, while at the same time has lower computational complexity when compared to global optimization methods. More specifically, this approach aggregates matching costs in 1D equally from multiple directions.

In this work four directions are considered: left to right ($r_{lr} = [+1; 0]^T$), right-to-left ($r_{rl} = [-1; 0]^T$), up-to-down ($r_{ud} = [+1; 0]^T$) and down-to-up ($r_{du} = [-1; 0]^T$) scan orders (Fig. 1(a)). For a direction $r \in \{r_{lr}, r_{rl}, r_{ud}, r_{du}\}$, the path cost for pixel p is recursively calculated from:

$$L_r(p, d) = C'_{all}(p, d) + \min \left\{ L_r(p-r, d), L_r(p-r, d \pm 1) + \pi_1, \min_i L_r(p-r, i) + \pi_2 \right\} - \min_i L_r(p-r, i), \quad (11)$$

where $i \in \{d_{\min}, \dots, d_{\max}\}$, while $p-r$ denotes the previous pixel along direction.

Parameters π_1 and π_2 are two smoothness penalty terms (with $\pi_1 \leq \pi_2$) for penalizing disparity changes between neighboring pixels. The work in [14] assumes that a depth discontinuity usually coincides with an intensity edge; hence the smoothness penalty must be relaxed along edges and enforced within low-textured areas. Therefore, it applies a symmetrical strategy so that π_1 and π_2 depend on the intensities of both left and right images. In this paper, two criteria are used to check depth discontinuity. The first criterion, similarly to [3], is based on intensity difference, which is computed as:

$$\Delta I_l(p) = \max_{c \in \{R, G, B\}} |I_l^c(p) - I_l^c(p-r)| \quad (12)$$



Fig. 1: (a) Path directions used for scanline optimization and (b) the (zoomed in) mean-shift segmentation map of the left “Teddy” image.

$$(\pi_1, \pi_2) = \left\{ \begin{array}{l} (\Pi_1, \Pi_2), \text{ if } (\Delta I_l(p) \leq P_{th} \ \& \ \Delta I_r(q) \leq P_{th}) \\ \left(\frac{\Pi_1}{1.5}, \frac{\Pi_2}{1.5} \right), \text{ if } (\Delta I_l(p) == 0 \ \& \ \Delta I_r(q) == 0) \\ \left(\frac{\Pi_1}{4}, \frac{\Pi_2}{4} \right), \text{ if } (\Delta I_l(p) \leq P_{th} \ \& \ \Delta I_r(q) > P_{th}) \text{ or } (\Delta I_l(p) == 0 \ \& \ \Delta I_r(q) \neq 0) \\ \left(\frac{\Pi_1}{4}, \frac{\Pi_2}{4} \right), \text{ if } (\Delta I_l(p) > P_{th} \ \& \ \Delta I_r(q) \leq P_{th}) \text{ or } (\Delta I_l(p) \neq 0 \ \& \ \Delta I_r(q) == 0) \\ \left(\frac{\Pi_1}{10}, \frac{\Pi_2}{10} \right), \text{ otherwise,} \end{array} \right\} \quad (10)$$

and

$$\Delta I_r(q) = \max_{c \in \{R, G, B\}} |I_r^c(q) - I_r^c(q-r)| \quad (13)$$

The second criterion, introduced in this paper, checks whether two pixels belong to the same mean-shift segment. Let us assume that after applying mean-shift segmentation to the left and right images the label images Lab_l and Lab_r , are acquired. Each segment is denoted by a specific label. The second criterion is denoted as:

$$\Delta L_l(p) = Lab_l(p) - Lab_l(p-r) \quad (14)$$

and

$$\Delta L_r(q) = Lab_r(q) - Lab_r(q-r) \quad (15)$$

According to these criteria, π_1 and π_2 are defined according to (10), where $\Pi_1 = 0.002$ and $\Pi_2 = 0.006$ are constant parameters, $P_{th} = 0.04$ is a threshold, which determines the presence of an intensity edge. The conditions to define π_1 and π_2 are examined in sequence. The first condition that evaluates to “True” is the one whose statements option will be executed.

The optimized cost volume is acquired by averaging the estimated path cost from the path directories:

$$C_f(p, d) = \frac{\sum_{r=\{r_{lr}, r_{rl}, r_{ud}, r_{du}\}} L_r(p, d)}{4} \quad (16)$$

Existing methods, such as those in [3, 14] use only intensity based criteria to check intensity discontinuity and define parameters π_1 and π_2 . The proposed approach includes an additional criterion based on mean-shift segmentation that improves the refinement results, as it is experimentally verified. The reason behind this improvement is that sometimes the first criterion denotes incorrectly a depth discontinuity due

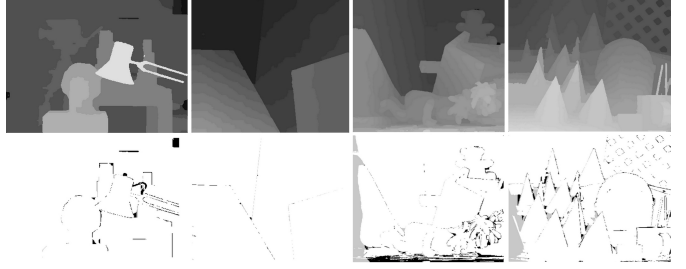


Fig. 2: Disparity maps generated with the proposed work and their corresponding disparity error maps for error threshold 1.

to edges that may exist in image areas that belong to the same depth but have some texture edges and not edges that correspond to depth discontinuity. On the contrary, mean-shift image segmentation is able to distinguish better between object texture edges and object boundaries (this fact is evident, for example, within the squared dashed region of Fig. 1(b) where texture edges have been incorporated inside larger segments). Therefore, the segmentation results are exploited in the definition of the smoothness penalties. In order to compensate for segmentation errors (include in the same segment areas with different depth) the denominator used for the definition of π_1 and π_2 is slightly increased to 1.5 for the case that the second statement of (10) is satisfied.

3.2 Occlusion Handling

The left disparity map $d_{LR}(p)$ is acquired after applying WTA to the cost volume $C_f(p, d)$, which was computed considering as reference image the left image of the stereo pair. If the right image is considered as reference image, then the right disparity map $d_{RL}(p)$ is acquired. The computation of $d_{LR}(p)$ and $d_{RL}(p)$ is fully independent. A prevalent strategy for detecting outliers is the Left-Right consistency check [14]. In this strategy, the outliers are disparity values that are not consistent between the two maps and therefore, they do not satisfy the relation:

$$|d_{LR}(p) - d_{RL}(p - d_{LR}(p))| \leq T_{LR}. \quad (17)$$

The threshold for outliers detection is set equal to $T_{LR} = 0$.

The occlusion handling strategy is kept similar to the one in [2] for computational simplicity. Therefore, an inconsistent pixel p is filled by the disparity of its closest consistent pixel. Practically, the disparity values of p 's left nearest consistent pixel p_l and p 's right nearest consistent pixel p_r are denoted as d_{p_l} and d_{p_r} , respectively. Then, the disparity value of $\min(d_{p_l}, d_{p_r})$ is assigned to p .

In order to deal with horizontal artifacts that are produced from this simple occlusion filling scheme, a bilateral filter is used to smooth the filled regions. The bilateral filter weights are given by:

$$W_{p,q} = \frac{1}{k} \cdot \exp\left(-\left(\frac{\Delta s_{p,q}}{\gamma_s} + \frac{\Delta c_{p,q}}{\gamma_c}\right)\right), \quad (18)$$

where k is a normalization factor, $\Delta s_{p,q}$ and $\Delta c_{p,q}$ denote the proximity distance and the color similarity between pixels p , q and γ_s, γ_c are constant parameters that adjust the spatial and color similarity. The parameters of the bilateral filter are set as in [2]: $\gamma_s = 9$, $\gamma_c = 0.1$ and the window size is 19×19 .

4 Experimental Results

4.1 Set of optimum parameters

The parameters used for the experiments are the same for all tested stereo pairs. The size of the window ω_k in Section 2 is 19×19 . The rest of the parameters used for the computation of the initial cost volume in Section 2 are defined as: $\{\alpha_1, \alpha_2, T_{gab}, T_{gra}, T_{rgb}, \varepsilon\} = \{0.20, 0.75, 0.015, 0.007, 0.028, 0.0001\}$. The parameters used for the mean-shift segmentation in Section 3.1 are the spatial radius, which is set equal to 3 and the feature space radius, which is set equal to 3. The selection of these strict values ensures that the segmentation map will be of high reliability, meaning that most likely a segment will not overlap with a depth discontinuity, and this fact is verified also in [15] and [16].

4.2 Middlebury Online Stereo evaluation

The proposed algorithm is evaluated on the Middlebury online stereo evaluation benchmark (reference period: February 2014). The disparity results of the proposed framework accompanied with the disparity error maps, as extracted by the Middlebury evaluation system, are visualized in Fig. 2. Errors in non-occluded and occluded regions are marked in black and gray respectively. The ranking results in Table 1, for absolute threshold equal to 1, indicate that the proposed method is 13th out of 149 methods that are included in the Middlebury Stereo Evaluation. This is an important achievement bearing in mind the reduced computational complexity of this algorithm and the very basic technique used for the occlusion handling. Moreover, Table 1 shows that this approach gives superior disparity results than [2, 4], which also exploit the guided image filter.

In more detail, the proposed method ranks: 58th for the ‘‘Tsukuba’’ image pair, 8th for the ‘‘Venus’’ image pair, 26th for the ‘‘Teddy’’ image pair and 12th for the ‘‘Cones’’ image pair. From Table 1, it is also obvious that the proposed approach enhances significantly the previous works that are based on the Guided Image Filter. In order to prove how the proposed work fosters the disparity results, Table 1 also includes the disparity results (‘‘No Criterion’’ row) using the scanline optimization without the proposed criterion to check depth discontinuity and the disparity results (‘‘Intensity Diff.’’ row) using the difference of intensities as in [2, 4] instead of the dissimilarity measure exploited in this work.

Algorithm	Rank	Error Rate(All)				Av.E
		Tsuk.	Venus	Teddy	Cones	
Proposed	13	2.01	0.30	10.4	7.71	4.39
No Criterion	14	1.82	0.34	10.8	7.82	4.99
Intensity Diff.	19	1.96	0.31	10.5	8.11	5.06
Gabor[4]	23	2.30	0.35	10.5	7.60	5.03
CostFilt.[2]	37	1.85	0.39	11.8	8.24	5.55

Table 1: The rankings in the Middlebury benchmark.

Error%	$\Delta d > 1$	$\Delta d > 1$	$\Delta d > 2$	$\Delta d > 2$
	Visible	All	Visible	All
Proposed	7.76	12.22	5.45	8.62
Inf. Perm.[5]	7.98	14.15	6.46	10.34
CostFilter[2]	8.40	15.06	6.80	11.82

Table 2: The error results for the extended stereo datasets.

4.3 Extended Comparison

Many of methods present comparative results on just the four well-known stereo pairs from the Middlebury stereo database, which are mentioned in Section 4.2. However, evaluation on limited data is not adequate to assess the overall performance of an algorithm, since the average error rates of the best performing techniques are quite close to each other. Therefore, except for the four stereo pairs from the Middlebury online stereo evaluation benchmark, evaluation is performed on two additional Middlebury datasets in order to assess more efficiently the performance of the proposed improvements. The 2005 and 2006 datasets, presented in [17], include 27 stereo pairs with their ground truth. The error percentage is measured for both non-occluded and all regions. Table 2 shows the results for the percentage of erroneous pixels having 1 or 2 disparity level difference with respect to ground truth. The results regarding the rest of methods in Table 2 are copied from the very recent work of [5]. Obviously, the proposed work gives better results than the rest of the methods that are evaluated in [5]. The improvement is more evident for the case of all regions and $\Delta d > 1$.

5 Conclusion

In this paper, we propose the exploitation of the pixel dissimilarity measure introduced in [12], which replaces the difference of pixels intensities. This replacement improves that disparity results. Additionally, the optimization of the initial cost volume is performed using a semi-global matching method, where a new criterion is introduced for the definition of the smoothness penalty terms that improves the disparity results. Extended experimental results on multiple stereo pairs prove the efficiency of the proposed approach regarding the disparity estimation problem. Another advantage point of this method is that it is compatible for optimization on the GPU and therefore can be exploited in Real-Time applications.

Future work could focus on the development of an efficient technique for the occlusion handling approach.

6 References

- [1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense twoframe stereo correspondence algorithms," *IJCV*, vol. 47, pp. 7–42, 2002.
- [2] A. Hosni, C. Rhemann, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," *TPAMI*, vol. 35, pp. 504–511, 2013.
- [3] X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang, and X. Zhang, "On building an accurate stereo matching system on graphics hardware," *ICCV Workshop on GPU in Computer Vision Applications*, 2011.
- [4] H. Liu, Y. Liu, S. OuYang, C. Liu, and X. Li, "A novel method for stereo matching using gabor feature image and confidence mask," *VCIP*, pp. 1–6, 2013.
- [5] C. Cigla and A. A. Alatan, "Information permeability for stereo matching," *Image Communication*, 2013.
- [6] F. Tombari, S. Mattoccia, L. Di Stefano, and E. Addimanda, "Classification and evaluation of cost aggregation methods for stereo correspondence," *CVPR*, pp. 1–8, 2008.
- [7] M. Gong, R.G. Yang, W. Liang, and M.W. Gong, "A performance study on different cost aggregation approaches used in real-time stereo matching," *IJCV*, vol. 75, pp. 283–296, 2007.
- [8] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," *ICCV*, vol. 2, pp. 508–515, 2001.
- [9] A. Klaus, M. Sormann, and K. Karner, "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure," *ICPR*, pp. 15–18, 2006.
- [10] J. Kim, K. Lee, B. Choi, and S. Lee, "A dense stereo matching using two-pass dynamic programming with generalized ground control points," *CVPR*, pp. 1075–1082, 2005.
- [11] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *TPAMI*, vol. 30, pp. 328–341, 2008.
- [12] S. Birchfield and C. Tomasi, "Depth discontinuities by pixel-to-pixel stereo," *IJCV*, vol. 35, pp. 269–293, 1999.
- [13] K. He, J. Sun, and X. Tang, "Guided image filtering," *ECCV*, pp. 1–14, 2010.
- [14] S. Mattoccia, F. Tombari, and L. D. Stefano, "Stereo vision enabling precise border localization within a scan-line optimization framework," *ACCV*, pp. 517–527, 2007.
- [15] T. Liu, P. Zhang, and L. Luo, "Dense stereo correspondence with contrast context histogram, segmentation-based two-pass aggregation and occlusion handling," *IEEE PSVIT*, 2009.
- [16] L. Di Stefano, F. Tombari, and S. Mattoccia, "Segmentation-based adaptive support for accurate stereo correspondence," *IEEE PSVIT*, 2007.
- [17] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," *CVRP*, pp. 1–8, 2007.