

Dense Stereo Correspondence with Contrast Context Histogram, Segmentation-Based Two-Pass Aggregation and Occlusion Handling

Tianliang Liu, Pinzheng Zhang, and Limin Luo

Lab of Image Science and Technology (LIST), Southeast University
No.2 Sipailou, Nanjing, 210096, China
{lt1315,luckzpz,luo.list}@seu.edu.cn

Abstract. In a local and perceptual organization framework, a novel stereo correspondence algorithm is proposed to provide dense and accurate disparity maps under point ambiguity. First, the initial matching technique is based on raw matching cost obtained from local descriptor with contrast context histogram and two-pass cost aggregation via segmentation-based adaptive support weight. Second, the disparity estimation procedure consists sequentially of two steps: namely, a narrow occlusion handling and a multi-directional weighted least square (WLS) fitting for large occlusion. The experiment results indicate that our algorithm can increase robustness against outliers, and then obtain comparable and accurate disparity than other local stereo methods effectively, and it is even better than some algorithms using advanced and offline but computationally complicated global optimization based algorithms.

Keywords: Stereo vision, stereo matching, local descriptor, segmentation, parallel computing, weighted least square, large occlusion.

1 Introduction

Accurate dense stereo matching is a fundamental and crucial problem in computer vision. A comparison of current stereo matching algorithms is given on the Middlebury Stereo Pages [1]. In general, stereo vision algorithms can be classified into local and global methods [1]. In local method, an area-based cost function is carefully selected and aggregated within a certain neighborhood to obtain resulting disparity with winner-takes-all (WTA) optimization [2,3,4,5,6,7,8,9,10]. To provide a robust result in stereo matching, the family of global algorithms seeks a disparity surface minimizing a global cost function defined by making an explicit smoothness assumption [12,13,14,15,16,17]. Recently there also exists trade-off between local and global methods, such as semi-global matching [18]. The latter two families usually have high matching accuracy. But most of them are computationally expensive and need many parameters that are hard to be set. However, the local methods are generally outperformed by the global and semi-global ones in higher speed.

To resolve the point ambiguity problem in image matching, many methods have been proposed for decades. Feature-based methods match only a few points proper for matching [20,21] while filtering out ambiguous points. In general, the idea is to detect the invariant local properties of salient image corners under a class of transformations, and then establish discriminating descriptors for these corners. As a result, feature-based methods yield sparse disparity maps. This approach is comparatively robust to the point ambiguity and produces accurate results rapidly in general. However, an efficient discriminating local descriptor, which is called contrast context histogram (CCH) and adopted previously for object recognition and image matching [20], is now proposed to extract local feature from image pairs to be constructed raw and robust matching cost for dense disparity map in local stereo correspondence in our work.

The local techniques typically use some kinds of statistical correlation among color or intensity patterns in local support windows in cost aggregation step [2,5,6]. In this approach, it is implicitly assumed that all points in a support window are from the same disparity in the scene. The variable support strategies with or without segmentation information in a specific support window are proposed to compute matching costs for the state-of-art local stereo methods [2,5]. But these variable support strategies with large window size have much high computational complexity because of symmetry (left-and-right) and traverse (pixel-by-pixel).

Recently a new post-processing technique has been studied to improve stereo matching performance [23]. This approach was presented to address the disparity discontinuity problem in narrow occluded regions when the better initial disparity maps were obtained from global method (such as graph cut). It consists of two parts; namely, a greedy disparity filling and a least-squared-errors (LSE) fitting. However, if the initial results with worse quality were gotten from a simple and efficient local method other than good global method, this approach can not effectively improve the resulting disparity maps. The latter one should be modified a bit to solve new problem.

This paper proposes a novel local stereo method which employs segmentation cue and can be divided into two steps: initial matching and disparity estimation. The initial matching is on the basis of raw matching cost with the CCH descriptor and two-pass cost aggregation with segmentation-based adaptive support weight (SASW). The disparity estimation in turn consists of two parts: narrow occlusion handling and multi-directional weighted least square (WLS) fitting for the broad or large occlusion areas. By means of experimental results we demonstrate that our approach can obtain the comparable disparity maps with high quality compared to some other traditional stereo algorithms.

The remainder of this paper is organized as follows. In Section 2, the CCH-based initial matching algorithm with segmentation information is discussed. The disparity estimation is addressed as a post-processing module for some unreliable disparities in Section 3. Experimental results are shown in Section 4. At last, conclusion and our future work are given in Section 5.

2 Initial Matching

2.1 The CCH-Based Cost Initialization and Color Segmentation

The dissimilarity measure is a crucial part of the stereo correspondence in a local perspective. In this paper, before computing pixel-wise matching cost, we select a local discriminating CCH descriptor to capture the feature for each pixel robustly and efficiently [20]. The local descriptor is a histogram of the contrast values inside the local region, which features log-polar mapping. The use of log-polar transformation is introduced as a preprocessing module to recover large scale changes and arbitrary rotations, which is a nonlinear and non-uniform sampling of spatial domain. Meanwhile, the histogram of the contrast values, comparing with other dissimilarity measures, are more insensitive to image noise and intensity difference of stereo pairs.

In general, how to construct the CCH descriptors for each pixel can be described as follows: firstly, to define a specified Log-polar mask M of the CCH descriptors, which is divided into several non-overlapping regions, R_1, R_2, \dots, R_t , by quantizing the radius and the direction in a $n \times n$ local region \mathbf{R} , as illustrated in Fig. 1. The current point p_c lies in the center of the coordinate. Then, according to the mask M with several sub-regions, we traverse each pixel p_c ignoring image borders to compute statistically positive and negative contrast histogram for each sub-region R_i . For each p in R_i , we can in turn define the two contrast histogram bins with respect to p_c as

$$H_{R_i^+} = \frac{\sum \{Diff \mid p \in R_i \text{ and } Diff \geq 0\}}{\#R_i^+} \quad (1)$$

$$H_{R_i^-} = \frac{\sum \{Diff \mid p \in R_i \text{ and } Diff < 0\}}{\#R_i^-} \quad (2)$$

where $Diff$ is the center-based intensity difference between p and p_c , $\#R_i^+$ and $\#R_i^-$ are the number of the positive and negative contrast values in the i^{th} region R_i , respectively. And then, by concatenating the values of all the contrast histogram entries from all the sub-regions into a single vector, the CCH descriptor of p_c in correspondence with its local region can be defined as follows:

$$CCH(p_c) = \{H_{R_1^+}, H_{R_1^-}, H_{R_2^+}, H_{R_2^-}, \dots, H_{R_t^+}, H_{R_t^-}\} \quad (3)$$

which can be considered as robust measurement of local intensity variations. The vector length T of this descriptor accords with the number of histogram bins.

The cost initialization module computes the initial matching cost $C(p_b, q_{m,d})$ (or $C(p_{bx}, p_{by}, d)$) between points $p_b \in I_b$ and $q_{m,d} \in I_m$ for assigning disparity hypothesis d to each pixel p_b in which the coordinates of p_b and $q_{m,d}$ are (p_{bx}, p_{by}) and $(p_{bx} - d, p_{by})$. To deal with linear lighting change and make the best use of the range that a single byte offers similarly to [6], the CCH descriptor can be normalized to a unit vector and scaled with 255. As the computed CCH descriptors are distributions represented as histograms, it is natural to calculate the correspondence scores using χ^2 distance [21]:

$$C(p_b, q_{m,d}) = \frac{1}{2} \sum_{k=1}^T \frac{(h_k(p_b) - h_k(q_{m,d}))^2}{h_k(p_b) + h_k(q_{m,d})} \quad (4)$$

where $h_k(p_b)$ and $h_k(q_{m,d})$ denote the k -bin normalized and scaled histogram at p_b and $q_{m,d}$, respectively. This matching will result in close distributions because this distance measures how unlikely it is that one distribution is drawn from the population represented by the other.

And then, we adopt color segmentation and then assume that pixels in the each segment should have similar disparity values. In our implementation, the Mean Shift algorithm [22] is used for color segmentation in CIELab space. The difference between pixel colors is measured in the CIELab color space because it provides three-dimensional representation for the perception of color stimuli similar to human color discrimination performance in short Euclidean distances [2].

2.2 The SASW-Based Two-Pass Cost Aggregation and Disparity Selection

The robust and fast support aggregation stage is also an important part in the local stereo matching. In order to reduce false matches owing to the point ambiguity and preserve efficient computation, we adopt a two-pass weighted cost aggregation with color segmentation cue. This SASW-based two-pass aggregation is inspired by the work of [5,19] and [6]. To construct the matching cost between two points p_b and $q_{m,d}$, a specific support weight, which is determined by color proximity from p_b as well as on segmentation information in monocular cue, is first assigned during the aggregation step to each point of I_b . In particular, weight $w_b(p_i, p_b)$ for point p_i belonging to I_b and close to p_b is defined as:

$$w_b(p_i, p_b) = \begin{cases} 1.0 & p_i \in S_b \\ \exp\left(-\frac{d_c(I_b(p_i), I_b(p_b))}{\gamma_c}\right) & \text{otherwise} \end{cases} \quad (5)$$

with S_b being the segment on which p_b lies, d_c being the Euclidean distance between two RGB triplets and the constant γ_c being an experimental parameter of the algorithm. Instead, the use of segmentation plays the role of an intelligent proximity criterion. It is a weight with zero value that is assigned to those points of I_b which lie too far from p_b , i.e. whose distance in the horizontal or vertical direction exceeds a certain length. As the use of segmentation in CIELab color space implies adding robustness to the support, we operate the RGB space for its convenience outside of segment in order to enforce smoothness over textured planes as well as to increase the accuracy of depth borders localization.

When aggregating matching costs, the original segmentation-based adaptive-weight approach computes the weighted average of adjacent matching costs, with the weights generated using both stereo images [5,19]. A similar approach is adopted to assign a weight $w_m(q_i, q_{m,d})$ to each point $q_i \in I_m$. The strategy of SASW is similar to that of traditional adaptive weight approach [2]. Under the left-and-right stereo setting with the weights being calculated, the matching cost for correspondence $(p_b, q_{m,d})$ depends on summing over the image area the

product of such weights with the above point-wise matching score normalized by the weight sum:

$$C_{osaw}(p_b, q_{m,d}) = \frac{\sum_{p_i \in N_{p_b}, q_i \in N_{q_m}} w_b(p_i, p_b) \cdot w_m(q_i, q_{m,d}) \cdot C(p_i, q_i)}{\sum_{p_i \in N_{p_b}, q_i \in N_{q_m}} w_b(p_i, p_b) \cdot w_m(q_i, q_{m,d})} \quad (6)$$

where N_{p_b} and N_{q_m} are respectively support neighbor window around p_b in base image and that of $q_{m,d}$ with respect to a disparity value d in matching image.

In this paper, we present two simplifications to the original segmentation-based algorithm with high computational complexity for achieving better performance in computational time similar to [6]. The first one is to ignore the weight term obtained from the matching image and its monocular segmentation cue. Therefore, to make it possible to compute the aggregated matching costs for different disparity hypotheses in parallel, the same weight is imposed to the same pixel when handling different disparity hypotheses. The second simplification is to approximate the weighted average of matching costs in the 2D rectangle window (i.e. $r \times r$) using a two-pass technique, in which the first pass computes the weighted average along the horizontal scanline while the second pass computing along the vertical scanline. This can further decrease the computational complexity of the aggregation approach from $O(r^2)$ to $O(r)$, which depends strongly on the window size used. Two additional steps are used to calculate the weighted averages being splitted in two separate components (horizontal and vertical). As a result, the aggregated costs are calculated in the simplified version using:

$$T^r(p_{bx}, p_{by}, d) = \frac{\sum_{u=-r}^r w(p_{bx}, p_{by}, u, 0) \cdot C(p_{bx} + u, p_{by}, d)}{\sum_{u=-r}^r w(p_{bx}, p_{by}, u, 0)} \quad (7)$$

$$C_{sasw}^r(p_{bx}, p_{by}, d) = \frac{\sum_{v=-r}^r w(p_{bx}, p_{by}, 0, v) \cdot T^r(p_{bx}, p_{by} + v, d)}{\sum_{v=-r}^r w(p_{bx}, p_{by}, 0, v)} \quad (8)$$

This cost aggregation with SASW mentioned above is a good technique for strengthening dissimilarity measure in itself. It is possible to get accurate dense matching results by performing a simple and local WTA optimization at each pixel with the proposed SASW without any complicated processes. The WTA method for the disparity of p_b in the base image can be formally defined as:

$$D_{init}(p_b) = \arg \min_{d \in R_d} C_{sasw}^r(p_{bx}, p_{by}, d) \quad (9)$$

with $R_d = [d_{min}, d_{max}]$ being the predefined range of all possible disparities. A similar approach can be adopted for the matching image I_m . After the WTA-based local optimization, coarse outliers are filtered using a 3×3 median filter.

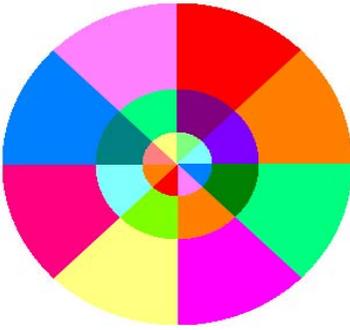


Fig. 1. Log-polar mask for the CCH

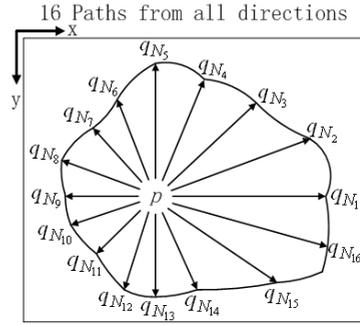


Fig. 2. WLS fitting paths in all directions

3 Disparity Estimation

3.1 Narrow Occlusion Handling

Firstly, unreliable disparities should be detected in this phase before addressing occluded regions which are small or narrow. To filter out these more erroneous matches, we apply the left-right consistency check symmetrically from stereo itself [19,23]. A threshold T_{occ} can be used for uniqueness constraint in our implementation. As mentioned above, color segmentation algorithm [22] is firstly employed for the selected base or matching image in CIELab color space in the implementation. It is more suitable for detecting outlier in segmented patch with small enough area and similar color if the segmentation is strong over-segmented.

After that, outlier removal is used to cluster reliable disparities in the same color segment into groups in an iterative framework, and identify unreliable disparity based on two measurements proposed in [26]. And then, greedy disparity filling is deployed to address the unreliable disparity when the occlusion region is small or narrow. The basic assumption for the disparity filling scheme is that the disparity of an unreliable pixel is the same as that of one of its neighbors in the same color segment by using the greedy-based strategy. The algorithm can be represented in details in [23]. The binocular and the monocular image data are used sequentially. There exists a threshold s as an appropriate constraint for both image cues to fill the unreliable disparities from neighboring reliable pixels.

3.2 Large Occlusion Handling

After the narrow occlusion handling procedure, it is possible that the disparity map still has unreliable pixels which do not have a disparity value. To resolve this issue more efficiently in purely local stereo correspondence perspective, the multi-directional WLS technique is proposed in this paper.

We assume that pixels which do not have a disparity up to now are justly resolved by the WLS scheme. In a known epipolar geometry, least-square-errors (LSE) fitting [23] with only intensity cue along the corresponding horizontal

scanline is naturally selected while ignoring the boundaries of color segmentation. However, to resolve still existed larger and more unreliable disparities, we do not have to enforce only the one or two ordering direction constraint in the horizontal scanline in LSE, and should exploit sufficiently monocular cues (intensity, color and shape etc.) in all directions to pursuit perfect disparity filling.

This leads to a new idea of greedy filling unreliable disparities by means of adaptive weight WLS fitting in 1D from all directions, while looking like semi-global cost aggregation step radially and equally for each path [18]. Each 1D measured path is started from an unreliable disparity pixel p and ended in first existed disparity pixel q_{Nk} encountered in the given radial direction. This can be explained by Fig. 2. The pixels outside of a convex hull in Fig. 2 represent pixels that have a disparity while the pixels inside of the hull have no disparity. The pixels of the hull itself also represent existed disparities passed a chain of procedures above. We assume that the disparity of the pixels inside the hull varies from the range of the existed disparities on the convex hull in this example.

These are the closest disparity values that can be obtained using the greedy disparity filling scheme when being approached from not only the left and the right directions, but multi-directions in a 2D image space. Considering computation complexity, we can assume that the number of all directions with WLS is in practice not arbitrarily large, but finite positive integer K (such as, $2 < K \leq 36$). The weight in this phase can be determined similarly by adaptive support weight [2]. The reason for the weight used in the disparity filling is that the smaller the distance between them in image spatial domain is, the higher the priority of filling candidate is; while the reason for color contribution is similar.

The WLS is calculated as a function of intensity variations along specific directional paths equally, which can be defined as minimizing the total weighted intensity variations along each 1D measured path from all 1D intensity variation paths with unreliable disparity pixels. We can get

$$q_{k_i}^* = \arg \min_{k=1, \dots, K} \{f_1(q_{N1}), \dots, f_k(q_{Nk}), \dots, f_K(q_{NK})\} \tag{10}$$

where

$$f_k(q_{Nk}) = \frac{\sum_{i=0}^{M_k} w(p, q_{k_i}) \cdot (I_{L_k}(q_{k_i}) - Mean(p, k))^2}{\sum_{i=0}^{M_k} w(p, q_{k_i})} \tag{11}$$

$$Mean(p, k) = \frac{1}{M_k} \sum_{i=0}^{M_k} I_{L_k}(q_{k_i}) \tag{12}$$

with $I_{L_k}(q_{k_i})$ being the intensity of pixel q_{k_i} in the k^{th} radial direction path L_k , M_k being the length of the given k^{th} path L_k and q_{k_i} denoting the i^{th} pixel close to the current unreliable disparity pixel p in the path L_k . $w(p, q_{k_i})$ indicates the support weight between p and q_{k_i} using color similarity and spatial proximity [2]. $f_k(q_{Nk})$ represents the perceptual distance between unreliable disparity pixel p and nearest disparity existed pixel q_{k_i} in the path L_k , which is weighted and normalized from three monocular cues: intensity, color and spatial distance etc.

Then, we assign reliable and closest disparity value q_{Nk} to unreliable pixels q_{k_i} when satisfying the criteria function (10). Finally, median filter can be adopted to remove remaining irregularities and smooth the last disparity map.

4 Experiment Results

4.1 Experiment Setup on Middlebury Stereo Pairs

To verify the effectiveness of our method at present, we computed the dense disparity maps while exploiting color segmentation in local technique for the Tsukuba, Venus, Teddy and Cones from the Middlebury 's second version stereo evaluation data set [1]. The parameters were kept constant for all stereo pairs.

In the CCH descriptor, we adopt three levels in the quantization of the distance and eight intervals in the quantization of the orientation under the log-polar coordinate system to generate the mask M with $3 \times 8 = 24$ non-overlapping regions, as shown in Fig. 1. Hence, the dimensions of the CCH descriptor T are $2 \times 3 \times 8 = 48$. And the definition of the distance and orientation is similar to that of them in the paper [20]. The color segmentation is obtained by running the Mean Shift algorithm using high speed version in CIELab space with a constant set of parameters (spatial radius $\delta S = 3$, range radius $\delta R = 3$, minimum region size $minR = 35$). For what means the variable support for the base image

Table 1. Quantitative evaluation of the proposed algorithm, comparing the percentage of "bad pixels" in non-occluded regions (R_{O-}), all regions except for unknown pixels (R_A), and regions near depth discontinuities (R_D). In each column, our result and some best of them are in bold and italic print, respectively. The overall performance measure is displayed in the 2th column, in which the average rank are over all latter 12 columns while subscript numbers being the relative ranks similar to the website [1].

Algorithm	Rank	Tsukuba			Venus			Teddy			Cones		
		R_{O-}	R_A	R_D									
CooptRegion [16]	3.3 ₁	<i>0.87</i>	<i>1.16</i>	<i>4.6</i>	0.11	0.21	1.54	5.16	8.31	13.0	2.79	<i>7.18</i>	8.01
AdaptingBP [13]	3.5 ₂	1.11	1.37	5.79	<i>0.10</i>	0.21	<i>1.44</i>	4.22	7.06	11.8	<i>2.48</i>	7.92	7.32
AdaptOvrSegBP [14]	11.6 ₇	1.69	2.04	5.64	0.14	<i>0.20</i>	1.47	7.04	11.1	16.4	3.60	8.96	8.84
AdaptDispCalib [4]	13.8 ₁₀	1.19	1.42	6.15	0.23	0.34	2.50	7.80	13.6	17.3	3.62	9.33	9.72
C-SemiGlob [18]	15.0 ₁₂	2.61	3.29	9.89	0.25	0.57	3.24	5.14	11.8	13.0	2.77	8.35	8.20
SO+borders [19]	15.0 ₁₃	1.29	1.71	6.83	0.25	0.53	2.26	7.02	12.2	16.3	3.90	9.85	10.2
CostAggr+occ [3]	17.2 ₁₆	1.38	1.96	7.14	0.44	1.13	4.87	6.80	11.9	17.3	3.60	8.57	9.36
SegmentSupport [5]	17.3 ₁₇	1.25	1.62	6.68	0.25	0.64	2.59	8.43	14.2	18.2	3.77	9.87	9.77
AdaptWeight [2]	20.7 ₂₀	1.38	1.85	6.90	0.71	1.19	6.13	7.88	13.3	18.6	3.97	9.79	8.26
2OP+occ [17]	26.8 ₂₇	2.91	3.56	7.33	0.24	0.49	2.76	10.9	15.4	20.6	5.42	10.8	12.5
Our method	27.6 ₂₈	1.74	2.11	9.23	0.41	0.94	3.97	8.08	14.3	19.8	7.07	12.9	16.3
FastAggreg [8]	28.0 ₂₉	1.16	2.11	6.06	4.03	4.75	6.43	9.04	15.2	20.2	5.37	12.6	11.9
GC+occ [12]	28.2 ₃₀	1.19	2.01	6.24	1.64	2.19	6.75	11.2	17.4	19.8	5.36	12.4	13.0
AdaptPolygon [10]	30.6 ₃₃	2.29	2.88	8.94	0.80	1.11	3.41	10.5	15.9	21.3	6.13	13.2	13.3
TensorVoting [11]	32.4 ₃₅	3.79	4.79	8.86	1.23	1.88	11.5	9.76	17.0	24.0	4.38	11.4	12.2
RealTimeGPU [6]	32.8 ₃₆	2.05	4.22	10.6	1.92	2.98	20.3	7.23	14.4	17.6	6.41	13.7	16.5
CostRelax [9]	33.7 ₃₇	4.76	6.08	20.3	1.41	2.48	18.5	8.18	15.9	23.8	3.91	10.2	11.8
TreeDP [15]	36.7 ₃₉	1.99	2.84	9.96	1.41	2.10	7.74	15.9	23.9	27.1	10.0	18.3	18.9

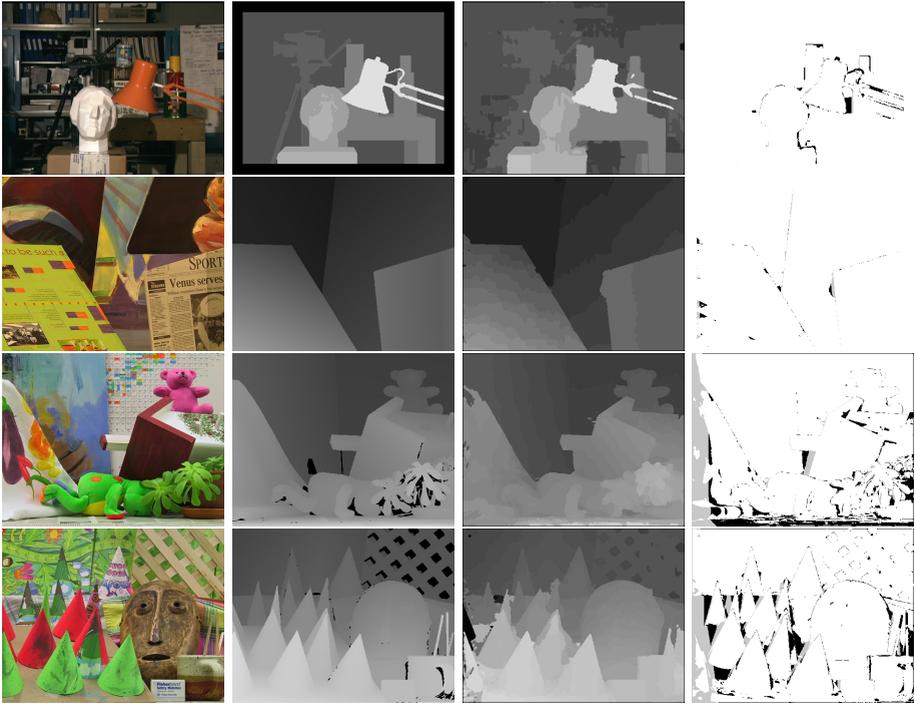


Fig. 3. Dense disparity results for the Tsukuba, Venus, Teddy and Cones stereo pairs: base images (first column), ground truth (second column), our results (third column) and bad pixels (last column)

in stereo pairs, the size of support window r is set to 51; and the parameter γ_c is equal to 15 in the two-pass cost aggregation stage.

For parameters in Section 3, the parameter T_{occ} in the symmetrical occlusion detection module is set to 2 in order to consider appropriately for part slanted object surfaces. In outlier removing [23], the first criterion (i.e., the ratio of occlusion in the segment) is set to be $O/S \geq 0.75$, where O and S are the numbers of pixels without disparity values and the area of the segment. The second criterion is chosen to be "if the percentage of the same disparity is smaller than 0.05%", pixels with a disparity value are still set to unreliable pixels. Meanwhile the threshold s is set to 5 for greedy disparity filling. K is equal to 36 and the parameters for the adaptive weight in WLS fitting are set by default values [2].

4.2 Quantitative and Qualitative Evaluation

The comparative results measured for each pair are summarized in Table 1 in terms of the percentage of bad matching pixels with the error tolerance $\delta = 1.0$. The Middlebury's second version stereo evaluation is measured based on known ground truth data. We cannot list total 49 algorithms including ours (as of July 2008) for lacking enough space; some other details can be found in the website [1].

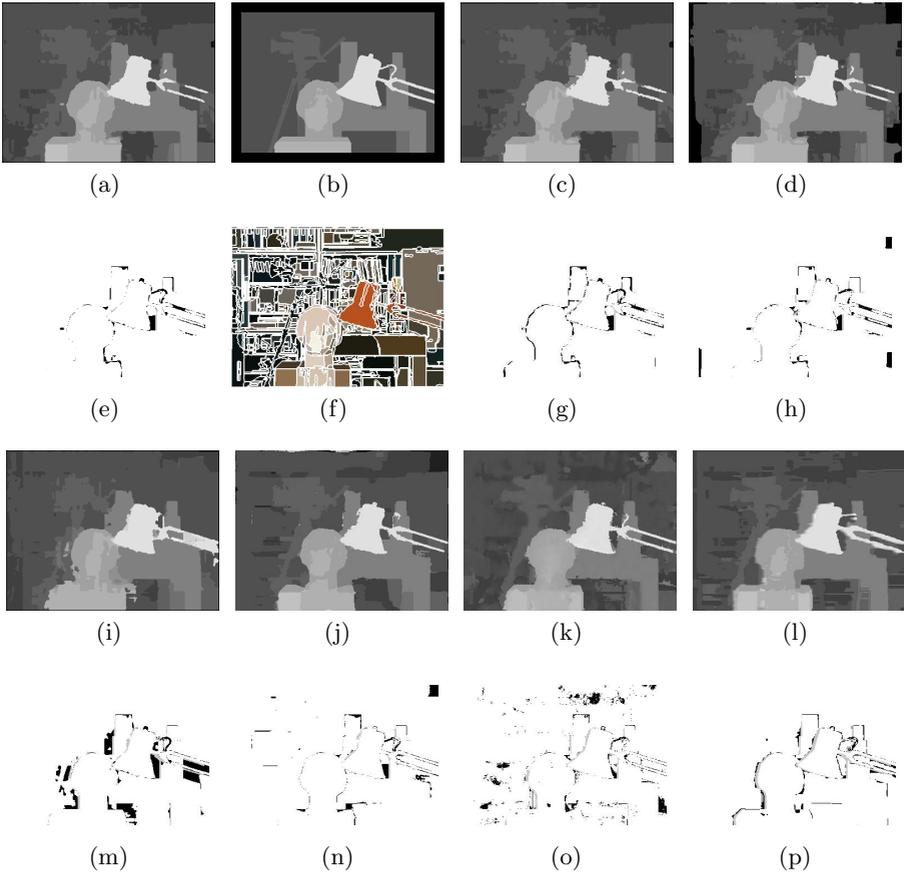


Fig. 4. Some parts of our algorithm contribute to the robustness of the disparity maps while comparing with several well-known local algorithms on the famous Tsukuba stereo image pair. (a and e) our final result and its bad pixels, (b and f) ground truth and segmentation result, (c and g) the result and its bad pixels replacing our disparity estimation by the default disparity estimation [23], (d and h) our initial result and its bad pixels, (i and m) the result and its bad pixels replacing our raw CCH-based matching cost by non-truncated SSD, (j and n) the result and its bad pixels via AdaptPolygon [10], (k and o) the result and its bad pixels via TensorVoting [11], (l and p) the result and its bad pixels via RealTimeGPU [6].

As it is clear from the Table 1 and the website, the rank of our algorithm with low computation cost is currently the 28th top of overall 49 algorithms in the evaluation. Our overall results in matching precision are apparently improved on the whole than some local methods, such as FastAggreg [10], TensorVoting [11], RealTimeGPU [6] and CostRelax [9] et al., and some advanced global ones such as GC+occ [12] and TreeDP [15] et al. However, our results are a bit worse than the other state-of-the-art methods in overall performance measure, such as CooptRegion [16], AdaptingBP [13], AdaptDispCalib [4], C-SemiGlob [18],

SO+borders [19], SegmSupport [5] and 2OP+occ [17] et al. As can be seen from the table, the proposed approach is comparably good among the purely local methods on standard stereo benchmarks.

Meanwhile, the proposed method is less expensive than other local methods in computational complexity. For some state-of-art local methods, such as Adapt-DispCalib, AdaptWeight, SO+borders and SegmSupport, the support window selected with too large value in cost aggregation will introduce very expensive complexity being the dominant processing time in the overall computation time; however, our method with parallel computing ability can generate efficiently comparative or equivalent result. But the complexity in two other modules will be increased a bit in computation time. The reasons may be listed as follows. Firstly, the Log-polar transformation should be run on several non-uniform sampling sub-regions to retrieval the local feature in the raw matching cost. Secondly, the disparity maps for both views should be doubly obtained from initial matching to check left-right consistency symmetrically in the narrow occlusion handling. Finally, the WLS fitting with K directions should be deployed to each unreliable disparity pixel in the large occlusion area. Fortunately, comparing with the obvious increase of matching precision, the little additional computational cost is negligible.

To compare visually and understand clearly our discussed algorithm, Fig. 3 shows the actual dense disparity results in our experiment. As can be seen clearly from the figure, the proposed approach can produce dense and accurate piecewise smooth disparity maps. Fig. 4 shows the disparity maps from some parts of our algorithm replaced by other traditional and similar module, while comparing with several previously known local algorithms, to illustrate how they complement each other to achieve robust disparity estimation. Especially, our algorithm can handle large occlusion effectively while comparing our results (a and e) with the results (c and g) by the default disparity estimation [23] without multi-directional weighted large occlusion handling in Fig. 4.

5 Conclusion

This paper presents a new and simple stereo approach with the CCH descriptor, SASW-based two-pass cost aggregation and multi-directional WLS fitting in a local perspective to generate more reliable and accurate disparity maps under point ambiguity effectively and efficiently. The stereo correspondence roughly consists of two steps sequentially: initial matching and disparity estimation. The CCH descriptor in the cost initialization, color segmentation and variable support weight in the two-pass cost aggregation are combined to obtain reliable and initial disparity maps; and then disparity estimation via narrow occlusion handling and multi-directional WLS fitting is designed to improve the stereo matching performance.

The advantages and shortcomings of the underlying design mechanisms in our method are discussed and analyzed through experimental evaluations conducted for the Middlebury data sets quantitatively and qualitatively. The experimental

results show that the proposed algorithm has higher matching precision and better robustness when compared with some part of standard stereo benchmarks.

In our future, we plan to observe this technique with more robust and other dissimilarity measure as raw pixel-wise matching cost, resegment strategy for large segments and more robust post-processing with reduced border errors while preserving higher processing speed.

References

1. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. Jour. Computer Vision (IJCV)* 47(1/2/3), 7–42 (2002), <http://vision.middlebury.edu/stereo/>
2. Yoon, K.J., Kweon, I.S.: Adaptive support-weight approach for correspondence search. *IEEE Trans. PAMI* 28, 650–656 (2006)
3. Min, D.B., Sohn, K.: Cost aggregation and occlusion handling with WLS in stereo matching. *IEEE Trans. IP* 17(8), 1431–1442 (2008)
4. Gu, Z., Su, X.Y., Liu, Y.K., Zhang, Q.C.: Local stereo matching with adaptive support-weight, rank transform and disparity calibration. *Pattern Recognition Letters (PRL 2008)* 29, 1230–1235 (2008)
5. Tombari, F., Mattoccia, S., Di Stefano, L.: Segmentation-based adaptive support for accurate stereo correspondence. In: Mery, D., Rueda, L. (eds.) *PSIVT 2007*. LNCS, vol. 4872, pp. 427–438. Springer, Heidelberg (2007)
6. Gong, M.L., Yang, R.G., Wang, L., Gong, M.W.: A performance study on different cost aggregation approaches used in real-time stereo matching. *Int. Jour. Computer Vision (IJCV)* 75(2), 283–296 (2007)
7. Yoon, K.J., Kweon, I.S.: Stereo matching with the distinctive similarity measure. In: *Proc. Int. Conf. on Computer Vision (ICCV 2007)*, pp. 1–7 (2007)
8. Tombari, F., Mattoccia, S., Di Stefano, L., Addimanda, E.: Near real-time stereo based on effective cost aggregation. In: *Proc. Int. Conf. on Pattern Recognition (ICPR 2008)* (2008)
9. Brockers, R., Hund, M., Mertsching, B.: Stereo vision using cost-relaxation with 3D support regions. In: *Image and Vision Computing New Zealand (IVCNZ 2005)* (2005)
10. Lu, J.B., Lafruit, G., Catthoor, F.: Anisotropic local high-confidence voting for accurate stereo correspondence. In: *Proc. SPIE*, vol. 6812 (2008)
11. Mordohai, P., Medioni, G.: Stereo using monocular cues within the tensor voting framework. *IEEE Trans. PAMI* 28(6), 968–982 (2006)
12. Kolmogorov, V., Zabih, R.: Computing visual correspondence with occlusions using graph cuts. In: *Proc. Int. Conf. on Computer Vision (ICCV 2001)*, pp. 508–515 (2001)
13. Klaus, A., Sormann, M., Karner, K.: Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In: *Proc. Int. Conf. on Pattern Recognition (ICPR 2006)*, vol. 3, pp. 15–18 (2006)
14. Taguchi, Y., Wilburn, B., Zitnick, C.L.: Stereo reconstruction with mixed pixels using adaptive over-segmentation. In: *Proc. Int. Conf. on Computer Vision and Pattern Recognition (CVPR 2008)*, pp. 2720–2727 (2008)
15. Veksler, O.: Stereo correspondence by dynamic programming on a tree. In: *Proc. Int. Conf. on Computer Vision and Pattern Recognition (CVPR 2005)*, pp. 384–390 (2005)

16. Wang, Z.F., Zheng, Z.G.: A region based stereo matching algorithm using cooperative optimization. In: Proc. Int. Conf. on Computer Vision and Pattern Recognition (CVPR 2008), pp. 887–894 (2008)
17. Woodford, O.J., Torr, P.H.S., Reid, I.D., Fitzgibbon, A.W.: Global stereo reconstruction under second order smoothness priors. In: Proc. Int. Conf. on Computer Vision and Pattern Recognition (CVPR 2008), pp. 2570–2577 (2008)
18. Hirschmuller, H.: Stereo processing by semiglobal matching and mutual information. *IEEE Trans. PAMI* 30(2), 328–341 (2008)
19. Mattoccia, S., Tombari, F., Di Stefano, L.: Stereo vision enabling precise border localization within a scanline optimization framework. In: Yagi, Y., Kang, S.B., Kweon, I.S., Zha, H. (eds.) ACCV 2007, Part II. LNCS, vol. 4844, pp. 517–527. Springer, Heidelberg (2007)
20. Huang, C.R., Chen, C.S., Chung, P.C.: Contrast context histogram—an efficient discriminating local descriptor for object recognition and image matching. *Pattern Recognition (PR 2008)* 41(10), 3071–3077 (2008)
21. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. *IEEE Trans. PAMI* 24(4), 509–522 (2002)
22. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. *IEEE Trans. PAMI* 24(5), 603–619 (2002)
23. Oh, J.D., Ma, S.W., Kuo, C.-C.J.: Stereo matching via disparity estimation and surface modeling. In: Proc. Int. Conf. on Computer Vision and Pattern Recognition (CVPR 2007), pp. 1696–1703 (2007)