


Article

Improved Cost Computation and Adaptive Shape Guided Filter for Local Stereo Matching of Low Texture Stereo Images

Hua Liu ¹, Rui Wang ², Yuanping Xia ^{1,*} and Xiaoming Zhang ³ ¹ Faculty of Geomatics, East China University of Technology, Nanchang 330013, China; liuhua@ecut.edu.cn² Ningbo Institute of Surveying and Mapping, Ningbo 315042, China; wr9098@163.com³ School of Geodesy and Geomatics, Wuhan University, No. 129, Luoyu Road, Wuhan 430079, China; xmzhang090@whu.edu.cn

* Correspondence: ypxia@ecut.edu.cn; Tel.: +86-180-7912-8080

Received: 1 February 2020; Accepted: 5 March 2020; Published: 9 March 2020



Abstract: Dense stereo matching has been widely used in photogrammetry and computer vision applications. Even though it has a long research history, dense stereo matching is still challenging for occluded, textureless and discontinuous regions. This paper proposed an efficient and effective matching cost measurement and an adaptive shape guided filter-based matching cost aggregation method to improve the stereo matching performance for large textureless regions. At first, an efficient matching cost function combining enhanced image gradient-based matching cost and improved census transform-based matching cost is introduced. This proposed matching cost function is robust against radiometric variations and textureless regions. Following this, an adaptive shape cross-based window is constructed for each pixel and a modified guided filter based on this adaptive shape window is implemented for cost aggregation. The final disparity map is obtained after disparity selection and multiple steps disparity refinement. Experiments were conducted on the Middlebury benchmark dataset to evaluate the effectiveness of the proposed cost measurement and cost aggregation strategy. The experimental results demonstrated that the average matching error rate on Middlebury standard image pairs is 9.40%. Compared with the traditional guided filter-based stereo matching method, the proposed method achieved a better matching result in textureless regions.

Keywords: stereo matching; cost measurement; adaptive shape guided filter; census transform; textureless regions

1. Introduction

Dense stereo matching is a significant research topic in the field of photogrammetry and computer vision, greatly benefiting applications like 3D reconstruction, DSM (Digital Surface Model) production, visual reality and autonomous vehicles [1–4]. A large number of efficient stereo matching algorithms have been developed in recent years, but it is still a challenging task to handle the stereo matching problem in occluded, textureless and discontinuous regions. According to the classical taxonomy method proposed by Scharstein and Szeliski [5], these existing stereo algorithms can be mainly classified into global and local approaches. Global algorithms explicitly incorporate smoothness assumption into an energy function that combines data and smoothness terms and estimate disparity by minimizing the global energy function. Belief propagation [6,7], graph cuts [8], and dynamic programming [9] are among the most commonly used global stereo matching optimization algorithms. They usually produce a more accurate disparity map than local methods but with higher computational complexity. On the other hand, local stereo matching algorithms only use the local information

within a finite support window to compute disparity. These algorithms have lower complexity and can be implemented much easier and faster. Thus, they are widely used in real-time applications. Local methods generally consist of four steps: (1) matching cost computation; (2) cost aggregation; (3) disparity computation/optimization; and (4) disparity refinement.

In terms of matching cost computation, Hirschuller and Scharstein [10,11] evaluated the performances of numerous different matching costs including parametric and nonparametric matching costs. The common parametric costs include pixel-based costs: absolute differences (AD), squared differences (SD), sampling-insensitive absolute differences, along with window-based matching costs: the sum of absolute or squared differences (SAD/SSD) and normalized cross correlation (NCC). Since most of these individual cost functions have its own strengths and weaknesses, combinations of multiple matching costs are exploited to obtain better performance. The works in [6,12,13] adopted a mixed cost computation method by combining the sum of absolute difference with gradient. Mei et al. [14] combined the absolute differences and Census transform to achieve an impressive performance. In [15–17], a combination of absolute difference, gradient and census transform or the variant versions were used for initial cost computation. The combination of multiple matching costs provides an alternative way to improve the performance of stereo matching algorithms.

Cost aggregation plays a decisive role in improving the matching efficiency and accuracy in local stereo matching algorithms. The initial matching cost is aggregated by summing or averaging over a support window to reduce image ambiguity. The naivest aggregation approach is to apply a simple low-pass filter with a fixed-size window, such as box filter or Gaussian filter, to the initial per-pixel cost. Nevertheless, the matching accuracy is determined on the fixed-size window, which easily leads to incorrect matching in textureless and discontinuous regions with fattening edges. To address this problem, a large number of cost aggregation strategies were proposed to achieve a more accurate disparity map while preserving edges. These improved aggregation methods can be categorized into two types: variable support window (VSW) and adaptive support weight (ASW). Methods based on variable support window try to find an appropriate regular or irregular support window to eliminate more outliers. Veksler [18] selected multiple windows from a number of candidates to produce smaller matching costs. Zhang et al. [19] proposed a cross-based local support window and dynamically constructed an adaptive shape region for cost aggregation. Then Mei et al. [14] modified the cross-based local support region construction by adding two additional thresholds and new constraint rules. On the contrary, the adaptive support weight approach, which was first presented by Yoon and Kweon [20], adaptively adjusts the support weights for pixels according to color similarity and spatial distance. Even though the adaptive support weight approaches achieved an outstanding performance, they suffered from high computational complexity. Therefore, several fast approximations of the bilateral filter [21,22] were then developed, but at the price of quality degradation. In order to reduce the computational complexity without quality degradation, Hosni et al. [13,23] and Rhemann et al. [24] proposed to aggregate matching cost using the guided image filter (GF). Compared with the bilateral filter, the guided image filter proposed by He et al. [25,26] can preserve the edges better and can be implemented with linear computational complexity independent of the window size. The guided filter has demonstrated great potential in stereo matching applications that have limited computation resources, especially in the real-time systems [13,27,28]. Even though the guided filter-based cost aggregation methods can achieve good performance in many cases, it is still not satisfactory enough for discontinuous regions and textureless regions since its fixed square size support window used by the traditional guided image filter.

In this paper, a local stereo matching algorithm with effective matching cost computation method and cost aggregation strategy was proposed. The proposed new matching cost combines enhanced image gradient-based matching cost with improved census transform-based matching cost and is proved to be more robust to radiometric changes. In matching cost aggregation step, we construct a cross-based adaptive shape support window for each pixel and implement a modified guided filter based on this adaptive shape region to improve the reliability of the disparity map, especially for low

texture structures. The winner-take-all strategy is then carried out to obtain the optimal disparity for each pixel. Finally, multi-constraints-based disparity refinement is applied to further improve the disparity map and eventually get sub-pixel accuracy disparity map.

The remainder of this paper is organized as follows. The proposed matching cost computation method and matching cost aggregation strategy are first described in Section 2. Section 3 presents the experimental results and discussions about the method and results. Finally, Section 4 concludes this paper.

2. Methods

The workflow of the proposed method is shown in Figure 1. Similar to a basic local stereo matching method, the proposed method consists of four steps. (1) Matching cost computation: we propose a new matching cost computation method that uses a combination of the enhanced image gradient-based cost and improved census transform-based cost. (2) Cost aggregation: first, a cross-based adaptive shape support window is constructed for each pixel. Then modified guided filter is implemented based on the constructed support window to aggregate the matching cost inside the window. (3) Disparity selection: a winner-take-all (WTA) strategy is used to find the optimal disparity for each pixel according to the aggregated cost. (4) Multi-constraints-based disparity refinement framework. In order to further detect the incorrect matching results, a multi-constraints-based disparity refinement framework is implemented. Outlier detection with left-right consistency checking process, occlusion/mismatch handling, weighted median filter, and subpixel enhancement are included in this framework.

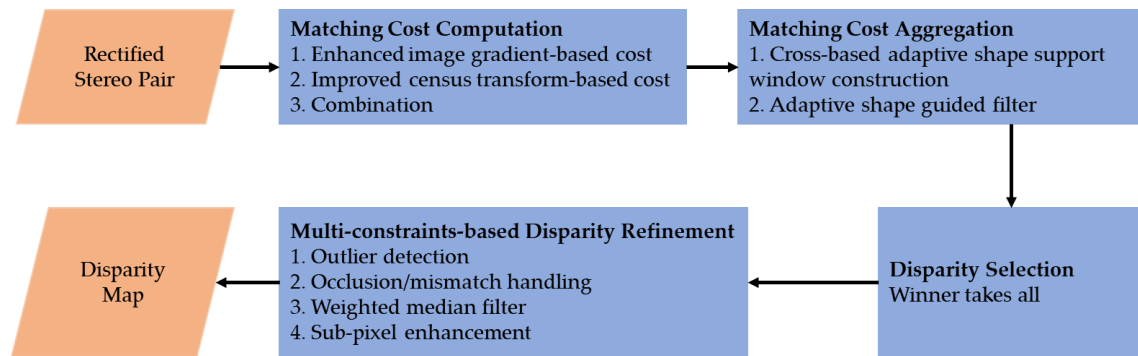


Figure 1. The workflow of the proposed method.

2.1. Matching Cost Computation

In this step, cost volume is built by computing per-pixel matching cost at all given disparity values under consideration. This cost volume is a three-dimensional array with a size of $H \times W \times D$, and H , W and D denote the height, width and disparity range of the images respectively. Although absolute difference on intensity or color channels is very simple and fast, it is too sensitive to radiometric differences and noises. Stereo matching methods using absolute difference on intensity or color have lots of errors on the disparity maps especially for outdoor images, in which radiometric changes and noises are unavoidable. By contrast, gradient similarity [29] and census transform [30] are more robust to radiometric distortion. To make the matching cost more robust to radiometric changes and noises, we propose a matching cost function that combines the enhanced image gradient-based cost with improved census transform-based cost.

In order to obtain stronger edge information, an image enhancement algorithm is first applied to the input stereo images. Here, the input images are first enhanced using Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm [31]. Then the sum of absolute derivative differences of the left and right enhanced images in the x and y directions is used as a gradient-based cost measure. The enhanced image gradient-based matching cost $C_{GRAD}^{CLAHE}(p, d)$ is computed according to Equation (1):

$$C_{GRAD}^{CLAHE}(p, d) = |\nabla_x I_L(p) - \nabla_x I_R(p - d)| + |\nabla_y I_L(p) - \nabla_y I_R(p - d)| \quad (1)$$

where I_L represents the enhanced left image and I_R represents the enhanced right image. $\nabla_x I_L(p)$ and $\nabla_y I_L(p)$ denote the gradients in x and y direction of the enhanced left image at a pixel p . Accordingly, $\nabla_x I_R(p-d)$ and $\nabla_y I_R(p-d)$ denote the gradients in x and y direction of the enhanced right image at the pixel $p-d$. d is the disparity.

Besides the gradients-based cost, the census transform-based cost is also computed using the enhanced left and right images. The original census transform is based on local intensity relations between the center pixel and its neighbor pixels. It only relies on the relative ordering of intensities and not their values, and thus it compensates for all radiometric distortions that preserving this ordering. However, Mei et al. [14] has displayed that the traditional census transform would produce wrong matches in regions with repetitive local structures in stereo matching. To address the limitations of traditional census transform, we present an improved census transform using gradients rather than intensity itself. The improved census transform in this paper is formulated as follows:

$$CTg(p) = \otimes_{q \in N(p)} \xi(|\nabla I(p)|, |\nabla I(q)|) \quad (2)$$

$$\xi(p, q) = \begin{cases} 1, & q < p \\ 0, & otherwise \end{cases} \quad (3)$$

where operator \otimes denotes a bit-wise catenation and $N(p)$ represents the neighbor pixels of anchor pixel p . q is a neighbor of p . $\nabla I(p)$ and $\nabla I(q)$ are the gradient of enhanced images at pixel p and neighbor pixel q . $|\cdot|$ means the magnitude of gradient. ξ is a function to determine the bit value as described in Equation (3). Then the Hamming distance is used to calculate the difference between the two bit-strings in left and right images and used as the improved Census transform-based matching cost.

$$C_{CTg}(p, d) = \text{Hamming}(CTg_L(p), CTg_R(p-d)) \quad (4)$$

In Equation (4), CTg_L and CTg_R are the Census transform bit strings of pixel p and $p-d$ in the left image and right image respectively, and d denotes the disparity of two pixels in the left and right images. The final combined matching cost is derived by merging the two cost components mentioned above:

$$C(p, d) = \rho(C_{GRAD}^{CLAE}, \lambda_{GRAD}) + \rho(C_{CTg}, \lambda_{CTg}) \quad (5)$$

$$\rho(c, \lambda) = 1 - \exp\left(-\frac{c}{\lambda}\right) \quad (6)$$

Here, λ is a normalizing parameter to control the influence of outliers, and the exponential function is used to normalize each cost component to the range $[0, 1]$ and ensure that the final matching cost won't severely bias to one of the matching costs. $C(p, d)$ is the final used matching cost that combines the enhanced image gradient-based matching cost (C_{GRAD}^{CLAE}) with the improved census transform-based matching cost (C_{CTg}).

Figure 2 presents the visual results of stereo matched disparity map using the proposed enhanced gradient-based matching cost and improved census transform-based matching cost on the Tsukuba of the Middlebury dataset [32]. The original gradient-based and Census transform-based matching cost are also shown in Figure 2 for the convenience of comparison. In order to make the result convincing, we use the same cost aggregation strategy as [23] and no refinement process is applied. The comparison of results obtained from the original gradient-based and enhanced gradient-based is shown in the first row. Further, it can be found that the enhanced gradient-based cost produces a better disparity map than the raw gradient-based cost at image boundaries (red circle mark). In the second row, the proposed modified census transform-based cost and the original census transform-based cost are also compared. According to the disparity map, our modified census transform-based matching cost performs visibly better than the original census transform-based matching cost at repetitive local structures (blue circle mark in Figure 2).

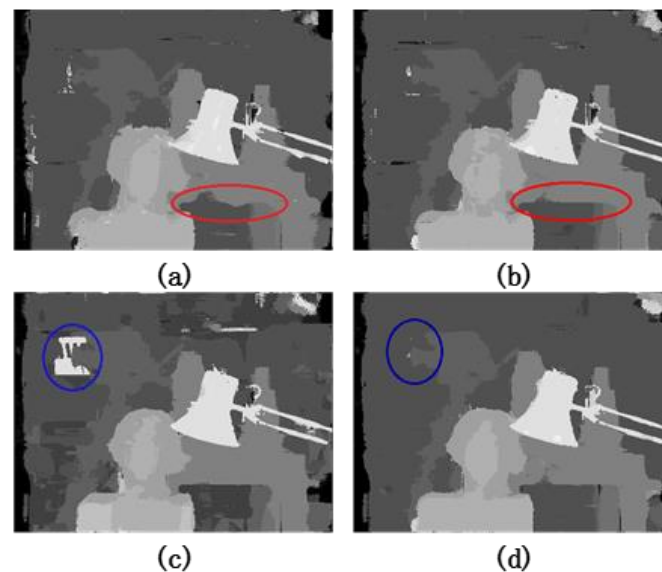


Figure 2. Initial disparity maps based on different matching costs for Tsukuba. (a) Absolute difference in image gradients; (b) proposed enhanced image gradients-based matching cost; (c) traditional census transform-based matching cost; (d) proposed improved census transform-based matching cost.

Figure 3 compares the proposed combined matching cost computation method with common individual cost methods, such as the absolute difference of intensity (AD), the absolute difference of image gradient and traditional census transform-based cost measurement. Tests are implemented on the two image pairs of the Middlebury stereo dataset (i.e., Tsukuba, Teddy). All of the disparity maps are initial stereo matching results without any post-processing and generated by the same cost aggregation method, which would be mentioned in the next section, and the winner takes all disparity computation strategy. From Figure 3, it can be found that the comprehensive performance of our proposed cost method obviously outperforms AD, gradient-based matching cost, and traditional census transform-based matching cost. AD cannot handle large textureless regions effectively, the gradient-based matching cost cannot do well with tiny boundaries, while the traditional census transform-based matching cost fails at repetitive or similar local structures.

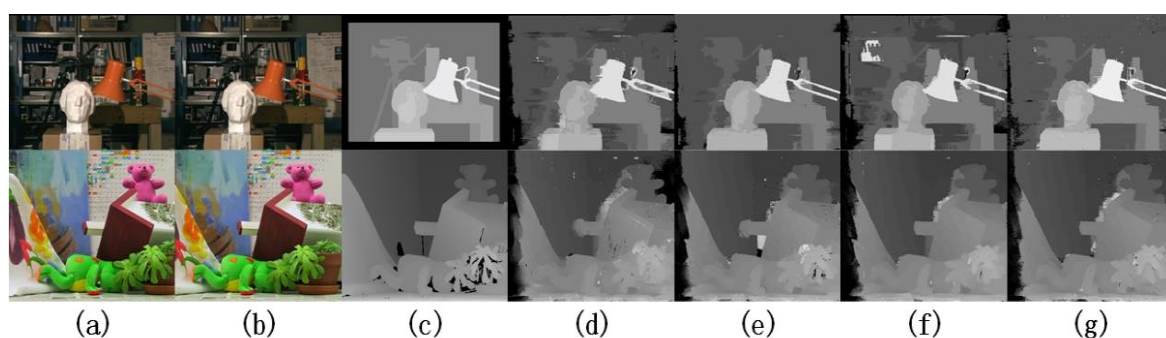


Figure 3. Comparison of different matching cost for Tsukuba and Teddy. (a) Left image; (b) right image; (c) ground truth map; (d) absolute difference of image channels; (e) absolute difference of image gradients; (f) traditional census transform-based matching cost; (g) proposed combined matching cost.

2.2. Matching Cost Aggregation Method

Per-pixel-based matching cost is sensitive to noise. Aggregating the matching cost over a local support window is an effective way to improve the accuracy and robustness of local stereo matching methods. Matching cost aggregation can be viewed as filtering on the cost volume. Given a cost volume C , the aggregated matching cost at pixel i with disparity d can be computed according to:

$$C'_{i,d} = \sum_j W_{ij} C_{j,d} \quad (7)$$

where $C_{j,d}$ is the matching cost of pixel j with disparity d . pixel j is a neighbor of pixel i . W_{ij} is the weight coefficient of the pixel j . $C'_{i,d}$ is the aggregated matching cost at pixel i with disparity d .

As an efficient and effective edge-preserving image filter, guided image filter has been successfully adopted in local stereo matching algorithms, achieving commendable disparity maps [23,33,34]. After applying guided filter using guidance image G , the kernel W_{ij} in Equation (7) can be expressed as:

$$W_{ij}(G) = \frac{1}{|\omega|^2} \sum_{k:(i,j) \in \omega_k} \left(1 + \frac{(G_i - \mu_k)(G_j - \mu_k)}{\sigma_k^2 + \varepsilon} \right) \quad (8)$$

where i and j represent pixel indexes. G_i and G_j are pixel values of guidance image at pixel i and pixel j . μ_k and σ_k^2 are the mean and variance of kernel window ω_k in guided image G , reflecting the statistical characteristics of pixels inside the support window. $|\omega|$ is the number of pixels in the window ω_k with a fixed size $r \times r$. ε is a smooth parameter.

The matching cost aggregation implicitly assumes that the disparity of the pixels inside the aggregation window are the same, which means that the support window should only contain neighbor pixels that have the same disparity as the center pixel. However, the support window with fixed window size in this traditional guided filter can hardly adapt to objects with different sizes in the scene and guarantee that the pixels inside the window have the same disparity, resulting in inappropriate aggregation results due to pixels with different disparities are involved especially in discontinuous regions. Besides, larger support windows are preferred in textureless regions because more accurate mean value (μ_k) and variance value (σ_k^2) can be estimated and thus improve the matching cost aggregation performance. In summary, the fixed size window guided filtering is not suitable for matching cost aggregation of textureless and discontinuous regions.

In order to better aggregate matching cost in discontinuous and textureless regions, adaptive shape support window is required. In general, the pixels with similar intensities within a constrained area are more likely captured from the same image structure, and thus have similar disparities. According to this assumption, Zhang et al. [19] and Mei et al. [14] proposed to construct cross-based support window. Cross-based support regions are constructed by expanding each pixel p to its neighbor pixels that have similar intensities with p in the horizontal and vertical directions, expressing as $V(p)$ and $H(p)$:

$$V(p) = \left\{ (x, y) \mid x \in [x_p - l_v^-, x_p + l_v^+], y = y_p \right\} \quad (9)$$

$$H(p) = \left\{ (x, y) \mid y \in [y_p - l_h^-, y_p + l_h^+], x = x_p \right\} \quad (10)$$

where $\{l_v^-, l_v^+, l_h^-, l_h^+\}$ are the four arm lengths. Then the support region $S(p)$ is generated by merging all of the pixels lying on the horizontal direction for each pixel q which belongs to the vertical direction $V(p)$, as illustrated in Figure 4.

$$S(p) = \bigcup_{q \in V(p)} H(q), \quad (11)$$

One of the core issues in constructing the cross-based adaptive shape support window is how to design proper rules to expand the pixel p to its neighbors. Zhang et al. [19] used color similarity $D_c(p, q)$ and constant color similarity threshold to construct the cross-based support window. This method cannot perform well in depth-discontinuous regions and low-texture regions at the same time. In order to perform better in both depth discontinuous and low-texture regions, Mei et al. [14] used both color similarity $D_c(p, q)$ and spatial distance $D_s(p, q)$ to construct the support window. Two color similarity thresholds were set according to two spatial distance thresholds, as illustrated in Figure 5a. If the spatial distance is smaller than threshold L_1 , a larger color similarity threshold τ_1 is

used. If the spatial distance is between L_1 and L_2 , a smaller color similarity threshold τ_2 is used. We further improve the method in Mei et al. [14] by calculating the color similarity threshold for each pixel adaptively. As illustrated in Figure 4, if we want to construct the adaptive support region of pixel p , the color difference $D_c(p, q)$ and the spatial distance $D_s(p, q)$ between pixel p and q are first calculated. According to the spatial distance $D_s(p, q)$, the pixel q is labeled as textured region pixel or textureless region pixel. If $D_s(p, q)$ is larger than spatial distance threshold d_{Lim} , the pixel q is labeled as textureless region pixel. Otherwise pixel q is labeled as textured region pixel. In the richly-textured regions, the color similarity thresholds should be larger and decrease with the increase of the spatial distance. In the textureless regions, the color similarity thresholds can be smaller and should decrease with the increase of the spatial distance. Based on these findings, we adaptively calculate the color similarity threshold for each pixel according to the following two rules:

$$\text{Rule 1 } \tau^{large}(D_s(p, q)) = -\frac{\tau_1}{L_1} \times D_s(p, q) + \tau_1, \text{ if } D_s(p, q) \leq d_{Lim}.$$

$$\text{Rule 2 } \tau^{small}(D_s(p, q)) = -\frac{\tau_2}{L_2} \times D_s(p, q) + \tau_2, \text{ otherwise.}$$

In the above rules, L_1 is a relatively small spatial distance constant and τ_1 is a relatively large color similarity constant for richly-textured regions. L_2 is a relatively large spatial distance constant and τ_2 is a relatively small color similarity constant for low texture regions. $\tau^{large}(D_s(p, q))$ and $\tau^{small}(D_s(p, q))$ are the adaptively calculated color similarity threshold for richly-textured regions and textureless regions accordingly. Rule 1 is a restricted condition which ensures that limited area is included in the richly-textured regions. While Rule 2 is fulfilled to ensure as many points from the same depth as possible are included in the textureless regions. The adaptively calculated color similarity threshold is illustrated in Figure 5b.

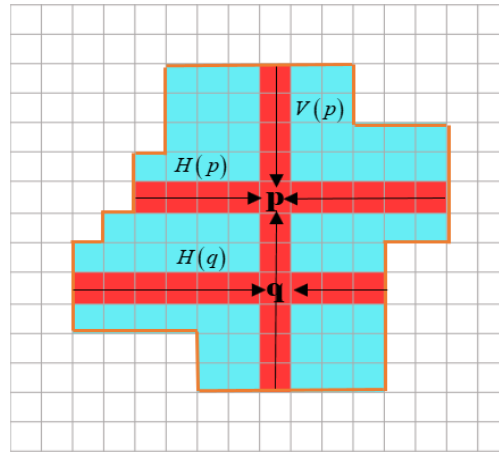


Figure 4. The cross-based support region construction for cost aggregation.

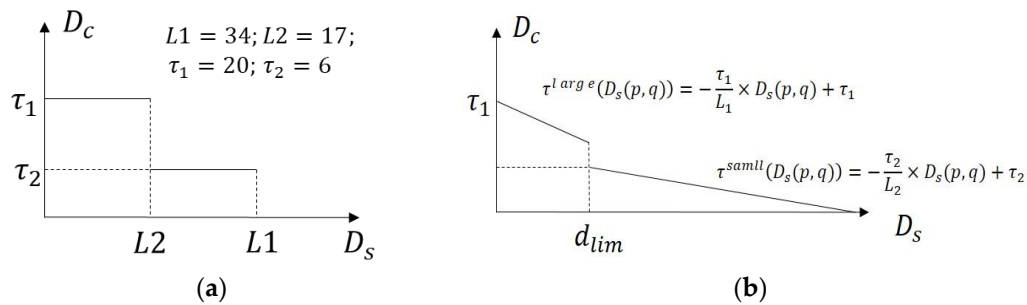


Figure 5. The spatial distance thresholds and color similarity thresholds proposed by Mei et al. [14] (a) and the adaptive color similarity threshold proposed by this paper (b).

The support regions generated by these three approaches are presented in Figure 6, and we can find that more valid pixels are included in the support window generated by our proposed method in large low texture regions compared with the other two methods.

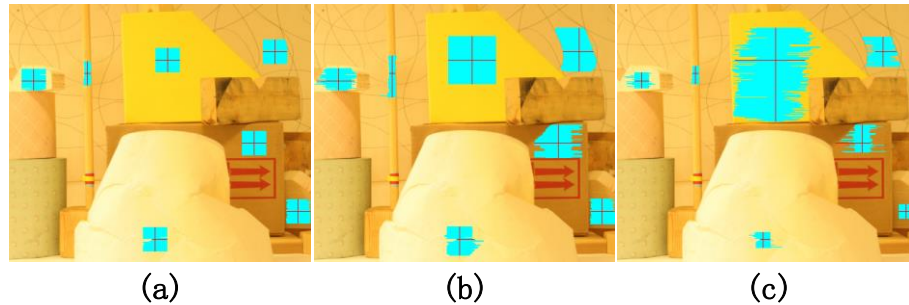


Figure 6. Examples of cross-based windows on the Lampshade1 image via various construction approaches. (a) Zhang et al. [19]; (b) Mei et al. [14]; (c) proposed construction approach.

After the adaptive shape support region constructed, the matching cost can be aggregated using guided image filter over the support region. The guided filter kernel in Equation (8) is designed for fixed square size window. In order to apply guided filter with adaptive shape support window to the cost volume, the guided filter should be modified accordingly. The modified weight kernel for adaptive shape support window is defined as:

$$W_{ij}(G) = \frac{1}{|N_i|} \sum_{k \in N_i} \left(\frac{1}{|N_k|} \sum_{j \in N_k} \left(1 + \frac{(G_i - \mu_k)(G_j - \mu_k)}{\sigma_k^2 + \varepsilon} \right) \right), \quad (12)$$

where $|N_i|$ and $|N_k|$ represent the pixel numbers of adaptive shape support regions N_i and N_k respectively. G_i and G_j are obtained directly from the guidance image. μ_k and σ_k^2 are calculated from the pixels of guidance image G that inside the support region. Thus, the weight coefficient for each neighbor can be compute. To speed up the cost aggregation process, we use an orthogonal integral image technique [19] to aggregate costs in the adaptive shape regions horizontally and vertically respectively.

2.3. Disparity Selection

The initial disparity for each pixel can be directly selected using winner-take-all (WTA) strategy as defined by:

$$d_p = \arg \min_{d \in D} C'(p, d), \quad (13)$$

where $C'(p, d)$ represents the aggregated matching cost volume obtained by the cost aggregation, and D denotes the set of all allowed candidate disparities. The disparity d_p for a specific pixel p is obtained by choosing the disparity that has the minimum aggregated matching cost.

2.4. Disparity Refinement by Multi-Constraints-Based Methods

The initial disparity maps obtained by WTA strategy still have many occluded and mismatched pixels. In this section, multi-constraints-based disparity refinement methods that consist of outlier detection, occlusion/mismatch interpolation, weighted median filter, and subpixel refinement are adopted to handle the disparity errors.

Outlier Detection: In order to find out the outliers in the left and right disparity maps, a left-right consistency check is applied. A pixel p is labeled as an outlier if it violates the following constraint:

$$|d_L(p) - d_R(p - d_L(p))| < 1, \quad (14)$$

where $d_L(p)$ and $d_R(p - d_L(p))$ are the disparities of pixel p and $p - d_L(p)$ in the left and right disparity maps respectively. Then outliers are further categorized into occlusions and mismatches according to the technique proposed by Hirschmuller [35] to better interpolate the disparity of the outlier pixels using different methods in the interpolation step.

Occlusion/mismatch interpolation: In this step, we adopt different interpolation strategies to interpolate the disparities of detected occlusion and mismatch pixels. For an occluded pixel p , a valid non-occluded disparity value from background region is required since occluded areas normally locate on the background. Therefore, we extract the nearest reliable pixels in eight different directions and select the pixel with the lowest disparity value for interpolation. Otherwise, the holes due to mismatches are processed by selecting the pixels with the most similar color value.

Weighted median filter: A weighted median filter [36] is usually implemented following behind the interpolation process to smooth outliers and streak-like artifacts while preserving the object boundaries. In this paper, only the invalid pixels are filtered with bilateral weights, which is computed as:

$$W^{BF} = \exp\left(-\frac{|p-q|^2}{\sigma_s^2}\right) \exp\left(-\frac{|I(p)-I(q)|^2}{\sigma_c^2}\right), \quad (15)$$

where $I(p)$ and $I(q)$ represent the intensity values of the pixel p and q . Parameters σ_s and σ_c adjust the spatial distance and color similarity, respectively. The filter assigns higher weights to pixels spatially close and similar in color.

Subpixel refinement: Finally, a subpixel estimation approach based on quadratic polynomial interpolation is performed to reduce the discontinuities caused by discrete disparity levels [35]. For each pixel p , its optimal sub-pixel disparity value d_{sub} is determined by the following formula:

$$d_{sub} = d - \frac{C'(p, d_+) - C'(p, d_-)}{2(C'(p, d_+) + C'(p, d_-) - 2C'(p, d))}, \quad (16)$$

where d is the discrete depth with the minimal cost, $d_- = d - 1$, and $d_+ = d + 1$. $C'(p, d)$, $C'(p, d_+)$, and $C'(p, d_-)$ denote the aggregated costs with the corresponding disparities, respectively. Finally, a 3×3 median filter is adopted to remove a few spikes.

3. Results and Discussion

In this section, we evaluate the performances of our proposed cost computation measurement and cost aggregation strategy. The rectified stereo image pairs from the Middlebury benchmark dataset [32] are used as experimental data. The experimental parameters are given in Table 1, and they are kept constant for all tests. The percentage of bad pixels of the estimated disparities over the stereo pairs was served as evaluating criterion. And the disparity error threshold was set to 1.0 pixel.

Table 1. Parameter settings of the proposed algorithm.

Parameter	Value	Parameter	Value
λ_{GRAD}	25	λ_{CTg}	15
τ_1	30	L_1	31
τ_2	6	L_2	80
d_{Lim}	9	ϵ	0.01^2

3.1. Evaluation of the Robustness to the Illumination and Exposure of Our Cost Computation Method

To verify the effectiveness of our proposed cost computation method, we used six pairs of stereo images with ground truth: Aloe, Baby1, Bowling1, Cloth1, Flowerpots, and Rocks 1, provided by Middlebury 2006 datasets [32] which have three illuminations (1, 2, 3) and three different exposure settings (0, 1, 2). Figure 7 shows the left image of the aloe data under three different illuminations

(no exposure variation) and with three different exposure settings (no illumination change). In this paper, three widely used cost computation methods were considered for comparison, including a function combining the sum of absolute difference (SAD) and gradient [6], a function combining absolute difference (AD) with census transform [14] and a combination of absolute difference, gradient, and Census transform [15]. To highlight the preference of cost computation function, all disparity maps were computed with the same cost aggregation algorithm proposed in Section 2.2 and no further refinement was applied.

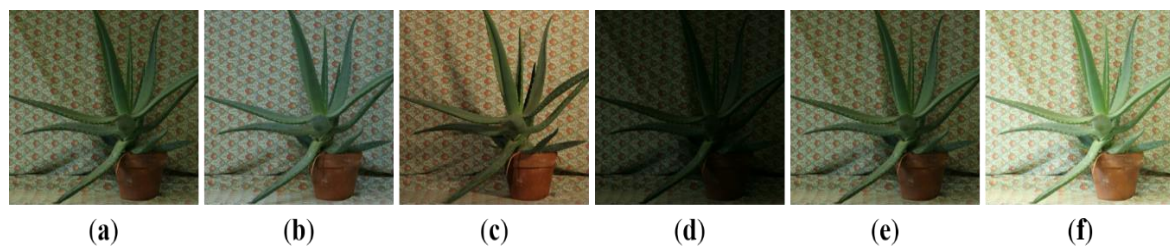


Figure 7. The left image of the aloe data under three different illuminations (with the same exposure 1) and with three different exposure settings (under the same Illumination 1). (a) Illumination 1; (b) Illumination 2; (c) Illumination 3; (d) Exposure 0; (e) Exposure 1; (f) Exposure 2.

Figures 8 and 9 show two pairs of the left and right images, the ground truth maps, and the disparity maps obtained by our proposed cost function and other three common combined cost methods under various illuminations and with different exposure settings, respectively. In Figure 8, the left images are captured under Illumination 1, the right images are captured under Illumination 3, and both are with the same Exposure 1. In Figure 9, the left images are taken with Exposure 0, while the right images are taken with Exposure 2. Furthermore, both the left image and right image are under the same Illumination 2. Comparing Figure 8g with Figure 8d–f, the disparity maps in Figure 8g are better than the disparity maps in Figure 8d–f, which indicates that the proposed cost computation method is more robust to illumination variation for these textureless stereo images. Comparing Figure 9g with Figure 9d,f, the disparity maps in Figure 9g are better than the disparity maps in Figure 9d,f, which indicates that the proposed cost computation method is more robust to exposure variation than cost combined SAD with gradient and cost combined AD, gradient and census. The disparity maps in Figure 9g and the disparity maps in Figure 9d have no significant differences.

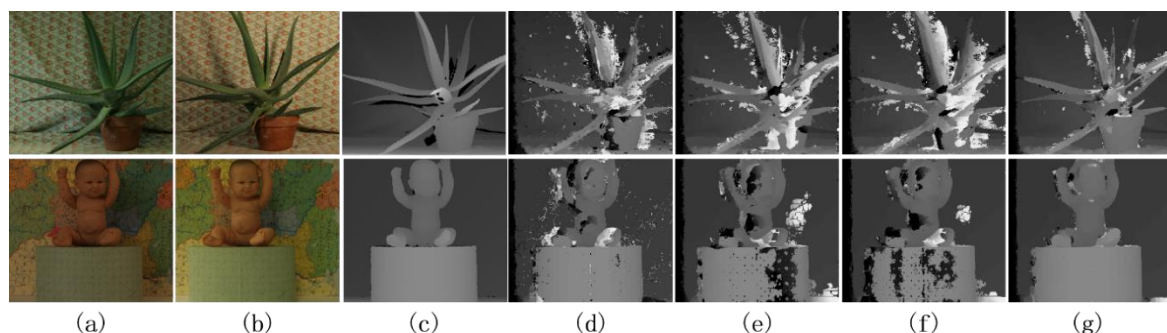


Figure 8. Disparity maps under different illumination conditions of Aloe and Baby1. (a) Left image under Illumination 1; (b) right image under Illumination 3; (c) ground truth; (d) cost combined sum of absolute difference (SAD) with gradient; (e) cost combined absolute difference (AD) with census transform; (f) combination of AD, gradient and census; (g) the proposed cost function.

The average percentage of bad pixels of disparity maps via these four different cost measurements under three radiometric conditions are shown in Tables 2 and 3. As a reference, we also computed the disparity maps without any illumination changes (Illumination 1) and exposure changes (Exposure 1) and the result is shown in Table 4. In Table 2, the proposed cost computation method achieved smallest

error matching rate in five of six pairs of stereo images and achieved the best average error matching rate, which indicates that the proposed cost computation method is more robust to illumination variation than the comparing methods. In Table 3, the proposed cost computation method also achieved smallest error matching rate in five of six pairs of images and achieved the best average error matching rate, which indicates that the proposed cost computation method is more robust to exposure variation than the comparing methods. In Table 4, the proposed cost computation method achieved smallest error matching rate in three of six pairs of the stereo images and achieved the best average error matching rate, which indicates that the proposed cost computation method is comparable to state-of-the-art cost computation methods in situations without radiometric changes. From the results in Tables 2–4, we can conclude that our proposed cost function is less sensitive to lighting changes and exposure changes. This is mostly because the absolute difference is too sensitive to the image intensity variations, which weaken the accuracy of the other three approaches. Additionally, our proposed cost method is more robust under different exposure configurations than different illumination settings. One possible reason is that the change of exposure is regarded as a global linear transformation, while changing illumination results in local radiometric differences.

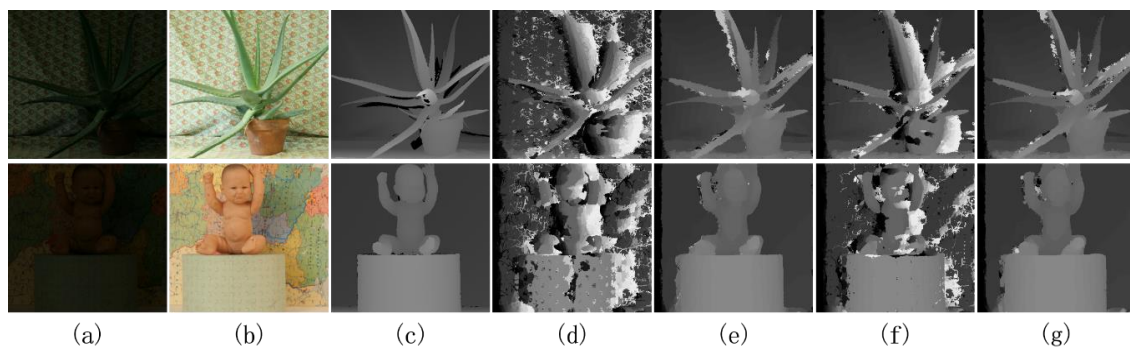


Figure 9. Disparity maps with different exposures of Aloe and Baby1. (a) Left image with Exposure 0; (b) right image with Exposure 2; (c) ground truth; (d) cost combined SAD with gradient; (e) cost combined AD with census transform; (f) combination of AD, gradient and census; (g) the proposed cost function.

Table 2. Error matching rate of various cost computation under different illumination.

Algorithms	Aloe	Baby1	Bowling1	Cloth1	Flowerpots	Rocks1	Avg
SAD + Grad	32.175	16.882	40.9	10.829	53.528	27.238	30.259
AD + Cen	32.274	25.055	46.147	13.212	56.0	18.732	31.903
AD + Grad + Cen	37.149	23.175	46.658	12.69	72.106	32.375	37.359
Proposed	22.034	11.115	26.946	11.333	34.185	13.849	19.910

Table 3. Error matching rate of various cost computation with different exposures.

Algorithms	Aloe	Baby1	Bowling1	Cloth1	Flowerpots	Rocks1	Avg
SAD + Grad	52.51	50.672	46.434	50.178	87.562	79.773	61.188
AD + Cen	16.173	11.118	20.022	11.096	41.021	15.329	19.127
AD + Grad + Cen	31.012	30.182	31.374	13.543	77.590	44.218	37.987
Proposed	15.205	10.658	22.782	11.060	29.834	14.094	17.272

Table 4. Error matching rate of various cost computation without radiometric changes.

Algorithms	Aloe	Baby1	Bowling1	Cloth1	Flowerpots	Rocks1	Avg
SAD + Grad	12.409	12.009	26.122	9.619	20.697	10.598	15.242
AD + Cen	13.61	11.811	23.859	10.475	22.676	12.766	15.866
AD + Grad + Cen	15.349	12.350	24.563	11.236	21.832	12.586	16.319
Proposed	14.478	9.749	18.663	11.085	18.644	12.008	14.104

3.2. Evaluation of Adaptive Shape Guided Filter on the Middlebury Benchmark Dataset

In this section, we chose 21 stereo pairs in Middlebury 2006 dataset for evaluation. To verify the effectiveness of our proposed algorithm, we evaluated both the local stereo matching method with the original guided image filter (GF) [23] and our proposed stereo matching method. The parameters of cost aggregation with an original guided filter such as the window size and smooth parameter are set according to [23]. We adopted the same cost computation method in both the original guided image filter algorithm and the proposed algorithm, which was proved to be more robust than other combined cost computation methods in Section 3.1. Further, to obtain more accurate disparity maps, we also used the same multi-constraints-based disparity refinement to exclude as many outliers as possible.

The stereo matching results of six pairs of textureless stereo images (Lampshade1, Lampshade2, Midd1, Midd2, Monopoly, and Plastic) are shown in Figure 10. Comparing the disparity maps in Figure 10e matched by our proposed method to the disparity maps in Figure 10c matched by the traditional guided filter algorithm, the disparity maps matched by our proposed method are much smoother in the textureless regions and preserves the edges better. Comparing the error maps of the proposed method in Figure 10f to that of the traditional guided filter algorithm in Figure 10e, the error pixels of our proposed method is much less than that of the traditional guided filter. The results presented in Figure 10 indicates that our proposed algorithm performs better than the original guided filter in large low texture images.

The percentage of bad pixels in the matching results of all 21 stereo pairs matched by our proposed method and traditional guided filter algorithm are shown in Table 5. The traditional guided filter algorithm performs much better (better than one percent of bad pixels) than our proposed method in 7 of the 21 stereo pairs (Aloe, Baby1, Baby2, Bowling1, Bowling2, Cloth2, and Wood1). In 8 of the 21 stereo pairs (Baby3, Cloth1, Cloth3, Cloth4, Flowerpots, Rocks1, Rocks2, and Wood2), our proposed method achieved comparable bad pixel rate to the traditional guided filter algorithm. For the six pairs of low texture stereo images, the proposed method achieved a much better result than the traditional guided algorithm. Although the error pixels have increased in 8 of the 21 image pairs, our algorithm is obviously more effective for textureless images. Especially in textureless image pairs Midd1 and Midd2, our proposed algorithm obviously recovers more correct disparity in low texture background (the third and fourth row in Figure 10). The error rate of Midd1 data is reduced by 23.8% and the error rate of Midd2 data is reduced by about 19%.

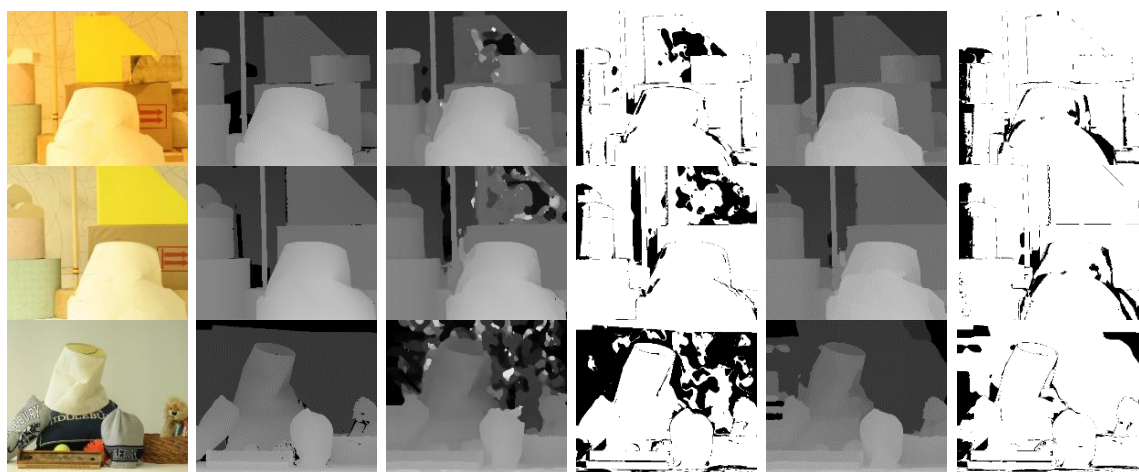


Figure 10. Cont.

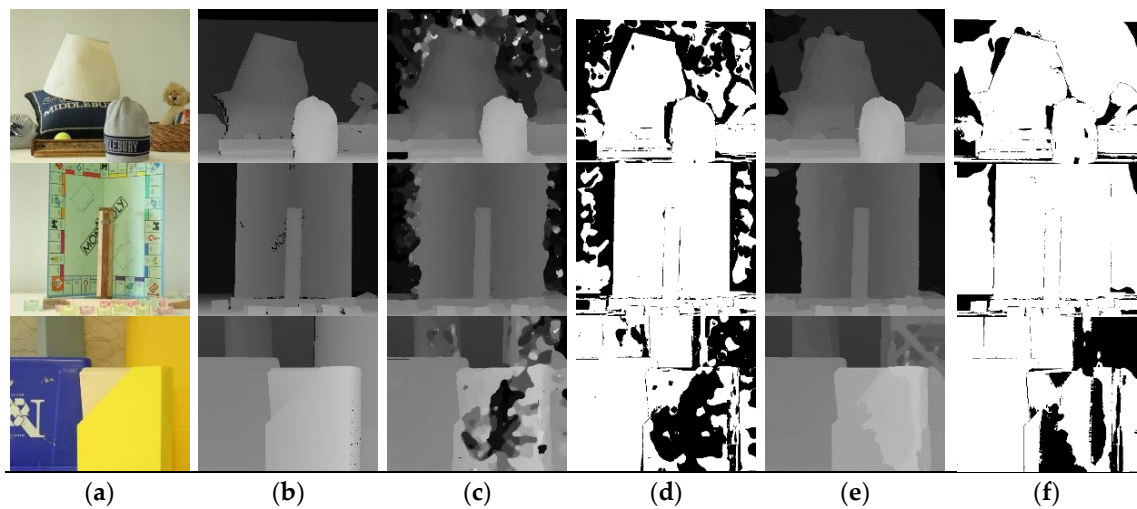


Figure 10. Comparison of the original cost aggregation method with our proposed method for the Middlebury dataset. From top to bottom: Lampshade1, Lampshade2, Midd1, Midd2, Monopoly, Plastic. (a) Left images; (b) ground truth maps; (c) results of the local stereo matching method based on traditional guided filter; (d) error maps for the method based on traditional guided filter; (e) results of the proposed method; (f) error maps of the proposed method.

Table 5. Percentage of bad pixels for the Middlebury 2006 dataset.

Algorithm	Aloe	Baby1	Baby2	Baby3	Bowling1	Bowling2
GF	7.407	2.575	5.534	5.981	7.94	12.184
Proposed	8.626	4.092	10.635	6.197	14.636	14.794
Algorithm	Cloth1	Cloth2	Cloth3	Cloth4	Flowerpots	Lampshade1
GF	2.96	8.613	3.94	8.393	12.405	11.223
Proposed	3.225	10.418	4.332	8.454	12.696	9.54
Algorithm	Lampshade2	Midd1	Midd2	Monopoly	Plastic	Rocks1
GF	15.729	37.653	35.381	22.803	32.666	4.183
Proposed	8.57	13.857	16.27	7.335	25.724	4.968
Algorithm	Rocks2	Wood1	Wood2	Avg (all)		
GF	3.587	3.829	0.965	11.712		
Proposed	3.973	8.574	0.484	9.4		

3.3. Evaluation of the Influences of Parameter Settings

Selecting proper parameters is very important in local stereo matching methods. In this section, we explore the influence of different parameter settings. The main parameters in the cost computation step are regularization parameters λ_{GRAD} and λ_{CTg} , which play an important role to adjust the proportion of two costs in the combined cost function. We chose five textureless images as test images and tuned the parameter individually with the rest of the parameters remaining constant. Figure 11a,b show the quantitative influence of parameters λ_{GRAD} and λ_{CTg} to disparity estimation in detail. From this figure, we can find that disparity results are stable with regularization parameters λ_{GRAD} and λ_{CTg} varying from 15 to 45. When λ_{GRAD} is set to 25 and λ_{CTg} is set to 15, the overall performance is relatively better, so λ_{GRAD} is recommended to be set to 25 and λ_{CTg} is recommended to be set to 15.

In the step of cross-based support window construction, the color similarity threshold τ_1 and arm length threshold L_1 were used to adjust the size of the support window for the richly textured regions. These two parameters have no significant influence on the disparity accuracy for most images, as shown in Figure 11c,d, since we set a small distance threshold ($d_{Lim} = 9$) to distinguish the textureless region from the textured region. For the Lampshade2 data set, the parameter settings ($\tau_1 = 30$, $L_1 = 31$) improved the disparity accuracy by about 2% in the discontinuous regions. The experimental results are more affected by tuning the color similarity threshold τ_2 and arm length threshold L_2 , which were

used to determine the size of the adaptive support window in the large textureless regions. In this paper, we set the color similarity threshold τ_2 to 6 and arm length threshold L_2 to 80 respectively to obtain good accuracy. In addition, the disparity results degraded gradually with the increasing distance threshold d_{Lim} , so we selected a small distance threshold which was set to 9. The smooth parameter of the modified guided filter ε was set according to [23].

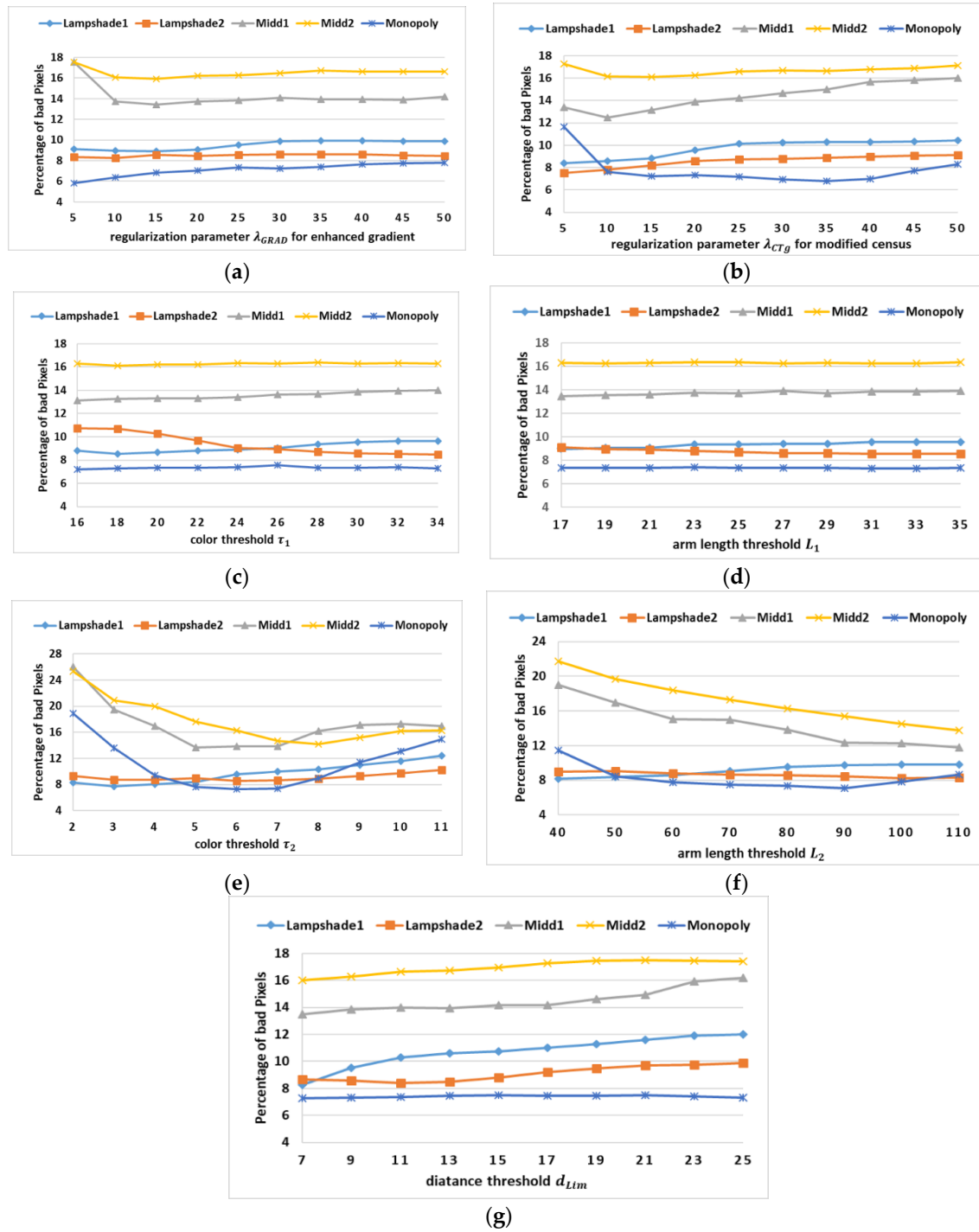


Figure 11. The experimental results on different parameter settings using the Middlebury textureless images. (a) Regularization parameter λ_{GRAD} for enhanced gradient-based matching cost; (b) regularization parameter λ_{CTg} for modified census transform-based matching cost; (c) color similarity threshold τ_1 ; (d) arm length threshold L_1 ; (e) color similarity threshold τ_2 ; (f) arm length threshold L_2 ; (g) distance threshold d_{Lim} .

4. Conclusions

It is challenging to handle occluded, textureless and discontinuous regions in stereo matching. In order to obtain better disparity maps in large textureless regions, this paper proposed a local stereo matching method using efficient combined matching cost measurement and adaptive shape guided filter. A matching cost computation method that combines enhanced gradient-based matching cost with improved census transform-based matching cost, which is more robust against exposure variations and illumination changes as well as textureless areas, is proposed. Besides, cross-based adaptive shape support region is constructed for each pixel using adaptively calculated color similarity threshold and an adaptive shape guided filter based on this cross-shaped support region is implemented to aggregate matching cost and thus improve the accuracy of disparity estimation for large low texture regions. Experiments were conducted using Middlebury benchmark dataset to validate the proposed methods. The experimental results indicate that our proposed stereo matching method can produce more accurate disparity maps for large low texture regions. The average percentage of bad pixels of our proposed method is about 9.40%, which is much lower compared with the original guided filter-based method.

Author Contributions: Conceptualization, H.L. and R.W.; methodology, H.L. and R.W.; software, R.W.; validation, H.L., R.W. and X.Z.; formal analysis, X.Z.; writing—original draft preparation, H.L. and R.W.; writing—review and editing, H.L., R.W., Y.X. and X.Z.; supervision, Y.X.; project administration, Y.X.; funding acquisition, Y.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China, grant number 41962018.

Acknowledgments: The authors would like to thank the anonymous reviewers for their valuable comments which helped to improve this paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Shen, S. Accurate multiple view 3D reconstruction using patch-based stereo for large-scale scenes. *IEEE Trans. Image Process.* **2013**, *22*, 1901–1914. [[CrossRef](#)] [[PubMed](#)]
2. Stentoumis, C.; Grammatikopoulos, L.; Kalisperakis, I.; Petsa, E.; Karras, G. A Local Adaptive Approach for Dense Stereo Matching in Architectural Scene Reconstruction. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2013**, *XL-5/W1*, 219–226. [[CrossRef](#)]
3. Howard, A. Real-time stereo visual odometry for autonomous ground vehicles. In Proceedings of the IEEE/Rsj International Conference on Intelligent Robots and Systems, Nice, France, 22–26 September 2008; pp. 3946–3952.
4. D’Angelo, P.; Reinartz, P. Semiglobal Matching Results on the ISPRS Stereo Matching Benchmark. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2011**, *XXXVIII-4/W19*, 79–84.
5. Scharstein, D.; Szeliski, R. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *Int. J. Comput. Vision* **2002**, *47*, 7–42. [[CrossRef](#)]
6. Klaus, A.; Sormann, M.; Karner, K. Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure. In Proceedings of the International Conference on Pattern Recognition, Hong Kong, China, 20–24 August 2006; pp. 15–18.
7. Sun, J.; Shum, H.Y.; Zheng, N.N. Stereo Matching Using Belief Propagation. In Proceedings of the European Conference on Computer Vision, Berlin, Germany, 28–31 May 2002; pp. 510–524.
8. Kolmogorov, V.; Zabih, R. Computing Visual Correspondence with Occlusions using Graph Cuts. In Proceedings of the Eighth IEEE International Conference on Computer Vision. ICCV 2001, Vancouver, BC, Canada, 7–14 July 2001; Volume 2, pp. 508–515.
9. Meerbergen, G.V.; Vergauwen, M.; Pollefeys, M.; Gool, L.V. A hierarchical stereo algorithm using dynamic programming. In Proceedings of the Stereo and Multi-Baseline Vision, Kauai, HI, USA, 9–10 December 2001; pp. 166–174.

10. Hirschmuller, H.; Scharstein, D. Evaluation of Cost Functions for Stereo Matching. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–8.
11. Hirschmuller, H.; Scharstein, D. Evaluation of Stereo Matching Costs on Images with Radiometric Differences. *IEEE Trans. Pattern Anal.* **2009**, *31*, 1582–1599. [[CrossRef](#)] [[PubMed](#)]
12. Hosni, A.; Bleyer, M.; Gelautz, M. Secrets of adaptive support weight techniques for local stereo matching. *Comput. Vis. Image Und.* **2013**, *117*, 620–632. [[CrossRef](#)]
13. Hosni, A.; Bleyer, M.; Rhemann, C.; Gelautz, M.; Rother, C. REal-time local stereo matching using guided image filtering. In Proceedings of the 2011 IEEE International Conference on Multimedia and Expo, Barcelona, Spain, 11–15 July 2011.
14. Mei, X.; Sun, X.; Zhou, M.; Jiao, S.; Wang, H.; Zhang, X. On building an accurate stereo matching system on graphics hardware. In Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), Barcelona, Spain, 6–13 November 2011; pp. 467–474.
15. Zhu, S.; Yan, L. Local stereo matching algorithm with efficient matching cost and adaptive guided image filter. *Vis. Comput.* **2016**. [[CrossRef](#)]
16. Jiao, J.; Wang, R.; Wang, W.; Dong, S.; Wang, Z. Local Stereo Matching with Improved Matching Cost and Disparity Refinement. *IEEE Multimed.* **2014**, *21*, 16–27. [[CrossRef](#)]
17. Hamzah, R.A.; Ibrahim, H.; Abu Hassan, A.H. Stereo matching algorithm based on per pixel difference adjustment, iterative guided filter and graph segmentation. *J. Vis. Commun. Image Represent.* **2017**, *42*, 145–160. [[CrossRef](#)]
18. Veksler, O. Fast variable window for stereo correspondence using integral images. In Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Madison, WI, USA, 18–20 June 2003.
19. Zhang, K.; Lu, J.; Lafruit, G. Cross-Based Local Stereo Matching Using Orthogonal Integral Images. *IEEE Trans. Circuits Syst. Video Technol.* **2009**. [[CrossRef](#)]
20. Yoon, K.J.; Kweon, I.S. Adaptive support-weight approach for correspondence search. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 650–656. [[CrossRef](#)] [[PubMed](#)]
21. Tomasi, C.; Manduchi, R. Bilateral filtering for gray and color images. In Proceedings of the Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271), Bombay, India, 7 January 1998.
22. Yang, Q.; Tan, K.H.; Ahuja, N. Real-time O(1) bilateral filtering. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009.
23. Hosni, A.; Rhemann, C.; Bleyer, M.; Rother, C.; Gelautz, M. Fast Cost-Volume Filtering for Visual Correspondence and Beyond. *IEEE Trans. Pattern Anal.* **2013**, *35*, 504–511. [[CrossRef](#)] [[PubMed](#)]
24. Rhemann, C.; Hosni, A.; Bleyer, M.; Rother, C.; Gelautz, M. Fast Cost-Volume Filtering for Visual Correspondence and Beyond. In *IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2011; pp. 3017–3024.
25. He, K.; Sun, J.; Tang, X. Guided Image Filtering. *IEEE Trans. Pattern Anal.* **2013**, *35*, 1397–1409. [[CrossRef](#)] [[PubMed](#)]
26. He, K.; Sun, J.; Tang, X. *Guided Image Filtering*; Daniilidis, K., Maragos, P., Paragios, N., Eds.; Springer: Berlin, Germany, 2010; pp. 1–14.
27. Yang, C.; Li, Y.; Zhong, W.; Chen, S. Real-Time Hardware Stereo Matching Using Guided Image Filter. In Proceedings of the 2016 International Great Lakes Symposium on VLSI (GLSVLSI), Boston, MA, USA, 18–20 May 2016.
28. Ttofis, C.; Theodoridis, T. High-Quality Real-Time Hardware Stereo Matching Based on Guided Image Filtering. In *Design, Automation, and Test in Europe Conference and Exhibition*; IEEE: New York, NY, USA, 2014.
29. De-Maeztu, L.; Villanueva, A.; Cabeza, R. Stereo matching using gradient similarity and locally adaptive support-weight. *Pattern Recogn. Lett.* **2011**, *32*, 1643–1651. [[CrossRef](#)]
30. Zabih, R.; Woodfill, J. *Non-Parametric Local Transforms for Computing Visual Correspondence*; Eklundh, J., Ed.; Springer: Berlin, Germany, 1994; pp. 151–158.
31. Zuiderveld, K. Contrast Limited Adaptive Histogram Equalization. In *Graphics Gems*; Academic Press Professional, Inc.: San Diego, CA, USA, 1994; pp. 474–485.
32. Scharstein, D.; Szeliski, R. Middlebury Stereo Evaluation—Version 3. Available online: <http://vision.middlebury.edu/stereo/data/> (accessed on 5 November 2017).

33. Yang, Q.; Ji, P.; Li, D.; Yao, S.; Zhang, M. Fast stereo matching using adaptive guided filtering. *Image Vis. Comput.* **2014**, *32*, 202–211. [[CrossRef](#)]
34. Hong, G.; Kim, B. A local stereo matching algorithm based on weighted guided image filtering for improving the generation of depth range images. *Displays* **2017**, *49*, 80–87. [[CrossRef](#)]
35. Hirschmüller, H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**. [[CrossRef](#)] [[PubMed](#)]
36. Ma, Z.; He, K.; Wei, Y.; Sun, J.; Wu, E. Constant Time Weighted Median Filtering for Stereo Matching and Beyond. In Proceedings of the 2013 IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 1–8 December 2013. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).