J. Vis. Commun. Image R. 39 (2016) 107–119

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Edge-preserving guided filtering based cost aggregation for stereo matching ☆

Shiqiang Zhu, Zhi Wang *, Xuequn Zhang, Yuehua Li

Zhejiang University, State Key Laboratory of Fluid Power and Mechatronic Systems, No. 38, Zheda Road, Hangzhou 310027, China

ARTICLE INFO

Article history:
Received 7 January 2016
Revised 27 April 2016
Accepted 23 May 2016
Available online 24 May 2016

Keywords:
Stereo matching
Adaptive support window
Cost aggregation
Guided filter

ABSTRACT

Stereo matching has been widely used in various computer applications and it is still a challenging problem. In stereo matching, the filter-based stereo matching methods have achieved outstanding performance. A local stereo matching method based on adaptive edge-preserving guided filter is presented in this paper, which can achieve proper cost-volume filtering and keep edges well. We introduce a gradient vector of the enhanced image generated by the proposed filter into the cost computation and the Census transform is adopted in the cost measurement. This cost computation method is robust against radiometric variations and textureless areas. The edge-preserving guided filter approach is proposed to aggregate the cost volume, which further proves the effectiveness of edge-preserving filter for stereo matching. The experiments conducted on Middlebury benchmark and KITTI benchmark demonstrate that the proposed algorithm produces better results compared with other edge-aware filter-based methods.

1. Introduction

Stereo matching, as a challenging problem in the research of computer vision, has been used in many applications, including 3d reconstruction, navigation of autonomous driving system, panoramic stereo imaging [1] and DoF rendering [2]. The main problem is to search for the corresponding pixels in two images. Various stereo matching methods have been presented, which are categorized as global methods and local methods [3]. The global methods usually make the smoothness assumption in the energy function, and then the disparity map is determined by minimizing the global energy. Popular implementations of global methods include dynamic programming [4], graph cut [5] and belief propagation [6]. These global algorithms usually produce poor results on the discontinuity edges and lead to streaking artifacts [6]. Most of the global optimization methods need complex global model, so they are computationally expensive and not feasible in the real-time applications. Local methods are considered to be more straightforward and simple. Compared with global methods, local methods are less time-consuming. Most of the local stereo matching methods perform the following steps: At first, the raw matching cost is computed. And then the cost aggregation is implemented in the local support window. Finally, the optimal disparity map is selected through optimization [3].

1.1. Related works

1.1.1. Cost computation

A variety of cost computation methods have been studied to compute the initial matching cost, including the sum of absolute difference, the Rank transform and the normalized cross correlation. However, these methods make the assumption that each pixel and its corresponding pixel have equal intensity values. Thus, they are not suitable for outdoor images, in which radiometric differences commonly exist [7]. Census transform encodes the local image information, and then the matching cost values are computed through Hamming distance. This method is robust under radiometric distortions [7] and many improvements have been made on it. Hirschmuller [8] proposed a stereo matching algorithm based on the mutual information (MI), which measures the similarity by utilizing the probability distribution function. But the large kernel size in the MI-based stereo matching method leads to poor performance in the object borders.

1.1.2. Cost aggregation

Cost aggregation is an essential part of local stereo matching, on which both the efficiency and accuracy largely depend. The most straightforward method that can be used is the low pass filters

☆ This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author.

E-mail address: 11325067@zju.edu.cn (Z. Wang).

http://dx.doi.org/10.1016/j.jvcir.2016.05.012
1047-3203/© 2016 Elsevier Inc. All rights reserved.

with fixed kernel size, such as box filter, Gaussian filter. However, these methods produce poor results with fatten edges. To solve this problem, many modified cost aggregation strategy have been proposed, such as the variable support window (VSW) methods [9] and adaptive support weight (ASW) methods [10].

Many VSW methods have been introduced in [9]. These methods try to build a support window with optimal size or shape that fits the region. Veksler [11] proposed a variable window approach that determines the support window by minimizing the window cost. The integral image technique is used in this algorithm to reduce the computational complexity of the cost aggregation. But the window size adjustment step is quiet time-consuming. Zhang et al. [12] presented a novel cross-based support window. This support window only consists of horizontal and vertical slices, and the cost aggregation is implemented in two directions independently.

In recent years, the ASW based methods have drawn a great deal of attention due to its outstanding performance [10]. The ASW based methods compute a weight for each pixel. Kuk-Jin and In first introduced the effectiveness of ASW based methods for stereo matching [13]. In this method, bilateral filter (BF) was used to compute the weights through the spatial distance and color dissimilarity between the corresponding pixels. After this study, various weights adjusting functions based on BF were proposed to improve the performance. However, one shortage of the bilateral filter is the high complexity in computation [14]. He et al. [15] proposed a novel algorithm called guided filter (GF) that produces high quality results and outperforms the BF.

1.2. Motivations

In local methods, those edge-preserving filters, such as BF, GF, can produce better results in stereo matching while keeping fine edges at the same time [16]. GF has shown better performance and efficiency compared with BF. But GF simply averages the pixel values without using any weighted fusion. Thus, it cannot preserve the edge well in some cases. In this paper, we incorporate an edge-preserving constraint into GF to improve the performance. Nevertheless, according to the previous research in [17], the raw BF or GF function cannot sort the ambiguity caused by those pixels with similar colors, but located at different disparity areas. A simple implementation of the GF combined with the edge-preserving factor cannot deal with the ambiguity effectively, which will be further explained in Appendix A.

To solve the problem mentioned above, adaptive support window is utilized in our proposed cost aggregation function. In the literature, several adaptive support window methods combined with GF have been presented, such as the adaptive guided filtering [16], the adaptive shape support window (ASSW) guided filtering [18], cross-based local multi-point filtering [19]. Xu et al. [18] proposed the ASSW based guided filter, which can find the optimal window with arbitrary size and shape. Cross-based local multi-point filtering, as an improved GF, determines the optimal support window based on the image information.

After surveying and analyzing the existing stereo matching algorithms, we propose an adaptive edge-preserving guided filtering (AEGF) based stereo matching algorithm. First, an enhanced image guided gradient vector is introduced into the matching cost computation. And the Census transform is adopted in the cost measurement. This measurement keeps robust against radiometric differences and textureless regions. Second, the edge-preserving guided filter (EGF) is adopted as the cost aggregation method. It can perform proper cost filtering and edge preserving, and the adaptive cross-based support window can resolve the ambiguity effectively. Third, the 'Winner-Take-All' strategy is taken to compute the raw disparity values. Finally, a multi-step post-processing method is applied to refine the disparity map. Except for stereo matching, there are other matching tasks, such as flow matching [20] and motion region matching [21]. In this study, we only consider the application to stereo matching. Comprehensive experiments have been conducted on the Middlebury benchmark [22] and KITTI benchmark [23], and the experimental results have shown the effectiveness of the proposed method.

1.3. Organization of the paper

The rest of the paper is organized as follows. Section 2 introduces the pipeline of the proposed method. The edge-preserving guided filter is introduced in Section 3. The combined cost measurement will be presented in Section 4. In Section 5, the proposed adaptive cross-based support window and overall cost aggregation method is described in detail. The multi-step post-processing is adopted in Section 6. The experiment results are given and analyzed in Section 7, and Section 8 summarizes this paper.

2. AEGF-based stereo matching method

The proposed AEGF-based stereo matching method consists of the following 5 steps: (1) preprocessing; (2) cost measurement; (3) matching cost aggregation; (4) initial disparity computation; (5) disparity refinement.

(1) Preprocessing: Due to noise and small texture, the edges of the input images are usually obscure. This will lead to disparity inconsistence and low accuracy. To improve these disadvantages, the edge-preserving guided filter is applied to preprocess the raw input images.
(2) Cost measurement: using the Census transform and gradient to compute the matching cost is proposed in this paper as the combined matching cost function usually performs better than the single method. Moreover, the enhanced image has rich gradient information for cost computation. Thus, in this paper, the combined cost measurement is adopted, which consists of the Census transform and the gradient vector of the initial image and the enhanced image.
(3) Matching cost aggregation: Many stereo matching algorithms employ adaptive support window to achieve better results. In this study, the cross-based window is adopted to construct the support window. As the cost aggregation needs repetitive computation, the cost aggregation method is proposed based on the orthogonal integral image (OII) technique [11] to accelerate this process.
(4) Initial disparity computation: The winner-take-all method is used to determine the initial disparity maps. That is, the disparity candidate which gives the minimum cost is the optimal disparity.
(5) Disparity refinement: The disparity map generated in the above steps contains many invalid matches and occlusions. Thus, a multi-step post-processing method is proposed in this paper. First, the left-to-right-consistency (LRC) check is adopted to determine the unstable and invalid pixels. Once the outliers are detected, they are corrected with the nearest valid pixel of outliers in the vertical and horizontal directions. Cross-based Occweight filtering [24] is used to correct the unstable pixels. Finally, the slanted plane smoothing [25] is used to post process the results, which is efficient and effective.

3. Edge-preserving guided filter

Inspired by GF, and weighted guided filter (WGF) [26], an edge-preserving guided filter is introduced in this section. First, we briefly describe the GF and WGF, and then present the proposed EGF.

3.1. GF and WGF

In the GF, there is an input image I. And a guidance image G is used, which could be identical to I. There is an assumption that the output $\hat{Z}$ is a linear transform of G within the support window $\Omega_{\zeta_1}(p)$ [27]:

$$\hat{Z}(q) = a_p G(q) + b_p, \quad \forall q \in \Omega_{\zeta_1}(p), \tag{1}$$

where $\Omega_{\zeta_1}(p)$ is a support window centered at p and the radius is ζ_1. a_p, b_p are two constants in the window $\Omega_{\zeta_1}(p)$. Their values are determined by minimizing the cost function $E(a_p, b_p)$, which is defined as

$$E = \sum_{q \in \Omega_{\zeta 1}(p)} \left[(a_p G(q) + b_p - I(p))^2 + \lambda a_p^2 \right], \tag{2}$$

where λ is a regularization parameter, which is similar to the range variance σ_r^2. The values of a_p and b_p are computed as

$$a_p = \frac{\mu_{G\bullet I,\zeta_1}(p) - \mu_{G,\zeta_1}(p)\mu_{I,\zeta_1}(p)}{\sigma^2_{G,\zeta_1}(p) + \lambda}, \tag{3}$$

$$b_p = \mu_{I,\zeta 1}(p) - a_p \mu_{G,\zeta 1}(p), \tag{4}$$

where $\bullet$ represents the element wise product of two matrices. $\mu_{G,\zeta 1}(p)$, $\mu_{I,\zeta 1}(p)$ and $\mu_{G\bullet I,\zeta 1}(p)$ are the mean values of the G, I and $G \bullet I$, respectively.

However, the value of λ is fixed in GF, which would lead to blurred edges. To overcome this problem, WGF was proposed in Ref. [25]. The cost function in Eq. (2) is defined as

$$E = \sum_{q \in \Omega_{\zeta 1}(p)} \left[(a_p G(q) + b_p - I(p))^2 + \frac{\lambda}{\Gamma_G(p)} a_p^2 \right], \tag{5}$$

Where $\Gamma_G(p)$ is defined as

$$\Gamma_G(p) = \frac{1}{N} \sum_{q=1}^{N} \frac{\sigma^2_{G,1}(p) + \varepsilon}{\sigma^2_{G,1}(q) + \varepsilon}, \tag{6}$$

where $\sigma^2_{G,1}(p)$ is the variance of the guidance image G in the 3×3 window. ε is a constant selected as $(0.001 \times L)^2$, where L is the range of I. The weighting $\Gamma_G(p)$ represents the importance of p with respect to G. The optimal values of a_p and b_p are computed as

$$a_p = \frac{\mu_{G\bullet I,\zeta 1}(p) - \mu_{G,\zeta 1}(p)\mu_{I,\zeta 1}(p)}{\sigma^2_{G,\zeta 1}(p) + \frac{\lambda}{\Gamma_G(p)}}, \tag{7}$$

$$b_p = \mu_{I,\zeta 1}(p) - a_p \mu_{G,\zeta 1}(p). \tag{8}$$

Unfortunately, the edge-aware weighting mentioned above cannot detect edges well because the window size is too small to fully reflect the image information, and it is hard to select the proper λ.

3.2. The edge-preserving guided filter

From Eq. (1), we can obtain

$$\nabla \hat{Z}(q) = a_p \nabla G(q). \tag{9}$$

It is clear that the smoothness of $\hat{Z}$ in $\Omega_{\zeta_1}(p)$ depends on the value of a_p. If p is at a flat region, the value of a_p is 0 so that the flat region is smoothed. If p is at an edge, the value of a_p is 1 so that the edge is well preserved. Based on the observation, a first-order constraint is imposed on a_p. b_p is computed using Eq. (8). The new a_p is formulated as

$$a_p = \frac{\mu_{G\bullet I,\zeta 1}(p) - \mu_{G,\zeta 1}(p)\mu_{I,\zeta 1}(p) + \frac{\lambda}{\hat{\Gamma}_G(p)} \eta_p}{\sigma^2_{G,\zeta 1}(p) + \frac{\lambda}{\hat{\Gamma}_G(p)}}, \tag{10}$$

where $\hat{\Gamma}_G(p) = \hat{\Gamma}1_G(p)\hat{\Gamma}2_G(p)$.

$\hat{\Gamma}1_G(p)$ is defined as

$$\hat{\Gamma}1_G(p) = \frac{1}{N} \sum_{q=1}^{N} \frac{\sigma^2_{G,\zeta 1}(p) + \varepsilon}{\sigma^2_{G,\zeta 1}(q) + \varepsilon}. \tag{11}$$

$\hat{\Gamma}2_G(p)$ is defined as the Sobel gradient

$$\hat{\Gamma}2_G(p) = \sqrt{\left(\frac{\partial Z(p')}{\partial x}\right)^2_{sobel} + \left(\frac{\partial Z(p')}{\partial y}\right)^2_{sobel}}. \tag{12}$$

And η_p is defined as

$$\eta_p = \frac{1}{1 + e^{k_p \frac{\mu_{\sigma 2} - \sigma^2_{G,\zeta 1}(p)}{(\mu_{\sigma 2} - \min(\sigma^2_{G,\zeta 1}(q)))}}}, \tag{13}$$

where μ_{σ^2} is the mean value of $\sigma^2_{G,\zeta_1}(p)$ in the window $\Omega_{\zeta_1}(p)$, k_p is a constant to make sure that the value of η_p approaches 1 or 0 when p is at an edge or flat region. This paper uses $\hat{\Gamma}_G(p)$ instead of $\Gamma_G(p)$ because the window size of the filter is adaptive to fit the shape. And $\hat{\Gamma}_G(p)$ contains the local variance and the Sobel gradient. Thus, the new weighting can detect edge well and keep robust against noise. The value of $\hat{Z}(p)$ is given by

$$\hat{Z}(p) = \bar{a}_p G(p) + \bar{b}_p, \tag{14}$$

where $\bar{a}_p$ and $\bar{b}_p$ are the mean value of a_p and b_p in the window $\Omega_{\zeta_1}(p)$.

Consider the case $I = G$, if the pixel p is at flat region, the value of η_p approaches 0, then a_p is computed as

$$a_p = \frac{\sigma^2_{G,\zeta 1}(p)}{\sigma^2_{G,\zeta 1}(p) + \frac{\lambda}{\Gamma_G(p)}}, \tag{15}$$

which is similar to the a_p in the GF and WGF. This implies that in the flat region, the proposed method can achieve the same smoothness as GF and WGF.

On the other hand, if the pixel p is at edge, η_p approaches 1, then a_p is computed as

$$a_p = \frac{\sigma^2_{G,\zeta 1}(p) + \frac{\lambda}{\Gamma_G(p)}}{\sigma^2_{G,\zeta 1}(p) + \frac{\lambda}{\Gamma_G(p)}} = 1. \tag{16}$$

The value of a_p is 1 regardless of the value of λ. The value of a_p is closer to 1 than that in the GF and WGF. This implies the proposed filter preserves edges better than GF and WGF. Fig. 1 shows the filtering result of different filters. To better observe the difference, the 1 dimensional filtering result is shown in Fig. 2. It can be seen that the near the edges, the value of the proposed filter is almost the same as the input, thus, the proposed filter can preserve edges better than the GF and WGF.

Zhang et al. [28] proposed a PDE-based method to enhance the image's edges. Similar to [26], the image enhancement algorithm is defined as follows. Given the input I, the filtering output $\hat{Z}$.

The enhanced image can be defined as

$$\hat{Z}_{en} = I + k_{en}(I - \hat{Z}), \tag{17}$$

where k_{en} is a constant to boost the details and it is fixed as 4 according to [15]. However, full details of the image are enhanced.

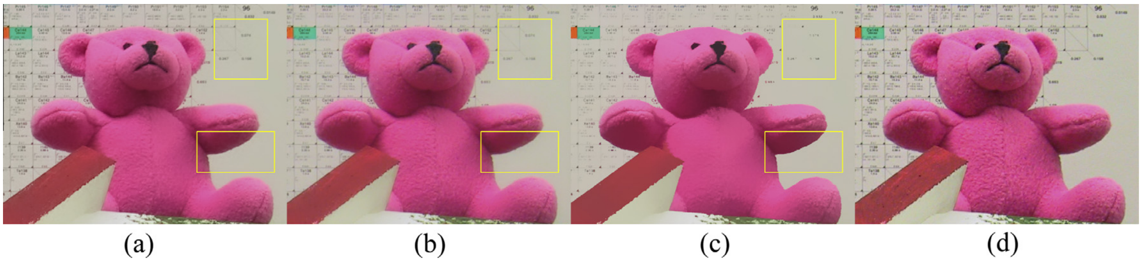


Fig. 1. Filtering results of different filters of Teddy. (a) Left image, (b) guided filter, (c) weighted guided filter, (d) proposed edge-preserving filter. As shown in the yellow rectangles, the proposed filter keeps the edges well and the tiny structures in textureless regions are smoothed out. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

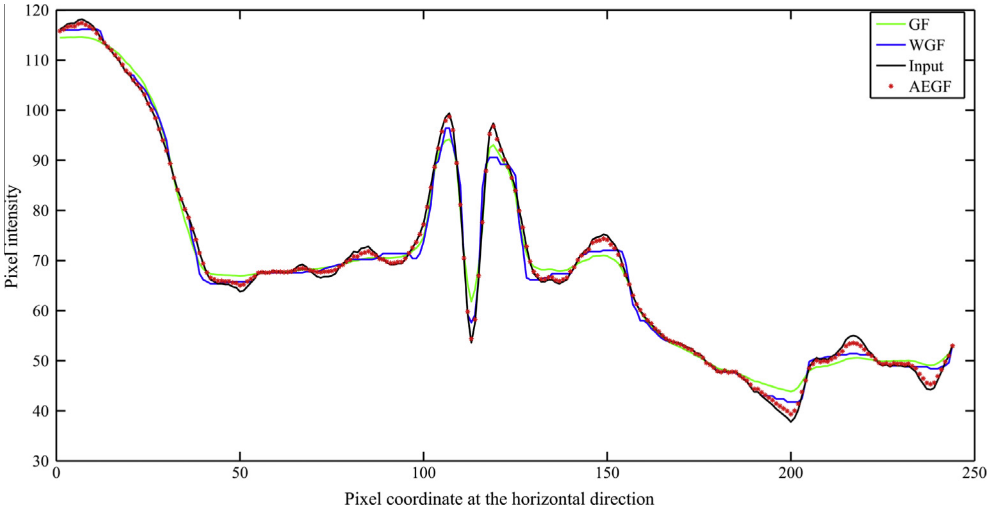


Fig. 2. 1-D illustration of different filters. The input is obtained from the middle column of the red channel of Fig. 1(a). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

To overcome this limitation, an edge-aware term η_p is used to enhance the image selectively. Thus, only the edge information is well enhanced. The new enhanced image can be expressed as follows:

$$\hat{Z}_{en} = I + k_{en}\eta_p(I - \hat{Z}). \tag{18}$$

4. Matching cost function

As discussed in Section 1.1, the combined matching cost function performs better than the single method. Thus, the combined cost measurement is adopted in this paper, which consists of the Census transform and the gradient vector of the initial image and the enhanced image. The implementation of the proposed matching cost function will be introduced in this section.

4.1. Measurement of the raw gradient

Image gradient, as a popular matching cost function, contains rich structural image information. Almost all the methods use absolute difference (AD) to compute the computation cost for its simple implementation. So AD is adopted as the cost measurement of the raw gradient vector. The truncated absolute difference of the gradient is calculated as follows:

$$C^{raw}_{GD,dir} = \min\left(\frac{1}{3}\sum_{c \in R,G,B} \|\nabla_{dir} I^{raw}_{L,c}(p) - \nabla_{dir} I^{raw}_{R,c}(p-d)\|, \lambda^{raw}_{GD}\right), \tag{19}$$

where d is the disparity. *dir* represents the vertical or horizontal direction and λ^{raw}_{GD} is the truncated value. The superscript *raw* refers to the raw image. This process can be computed with less complexity.

4.2. Measurement of the Census transform

Hirschmuller and Scharstein [7] evaluated several kinds of stereo matching cost methods and showed that Census transform has the best performance and keeps robust against illumination variations. The Census transform in this paper is formulated as

$$cen(p) = \underset{q \in w(p)}{\otimes} \xi(D_m(p), D(p,q)), \tag{20}$$

$$\zeta(x,y) = \begin{cases} 1, & x < y \\ 0, & \text{otherwise} \end{cases}, \tag{21}$$

where $\otimes$ represents a bit-wise catenation and $w(p)$ is the support window of p. $D(p,q)$ is the Euclidean distance, and $D_m(p)$ is the mean value of all the $D(p,q)$ in the $w(p)$. The matching cost of p with disparity d is calculated as

$$C_{cen}(p,d) = \min(\text{Hamming}(cen(p), cen(p-d)), \lambda_{cen}), \tag{22}$$

where λ_{cen} is the truncated value. The kernel size of Census transform usually keeps large to keep robust (e.g., 9 × 7, 7 × 7). In this method, the window sized is selected as 5 × 7 to reduce the computation complexity without affecting the cost measurement quality.

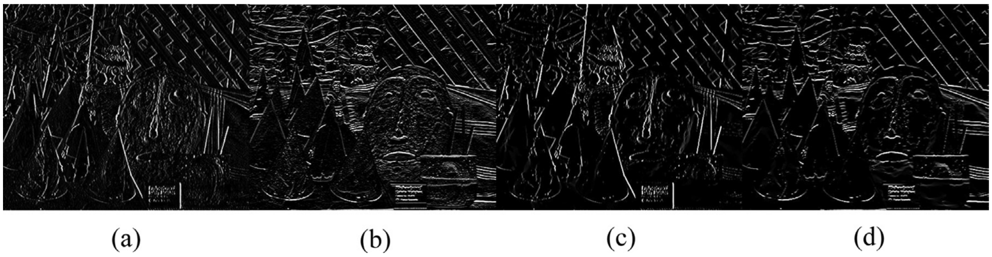


Fig. 3. Gradients of the initial image and enhanced image of Cones. (a) Horizontal gradient of the initial image. (b) Vertical gradient of the initial image. (c) Horizontal gradient of the enhanced image. (d) Vertical gradient of the enhanced image.

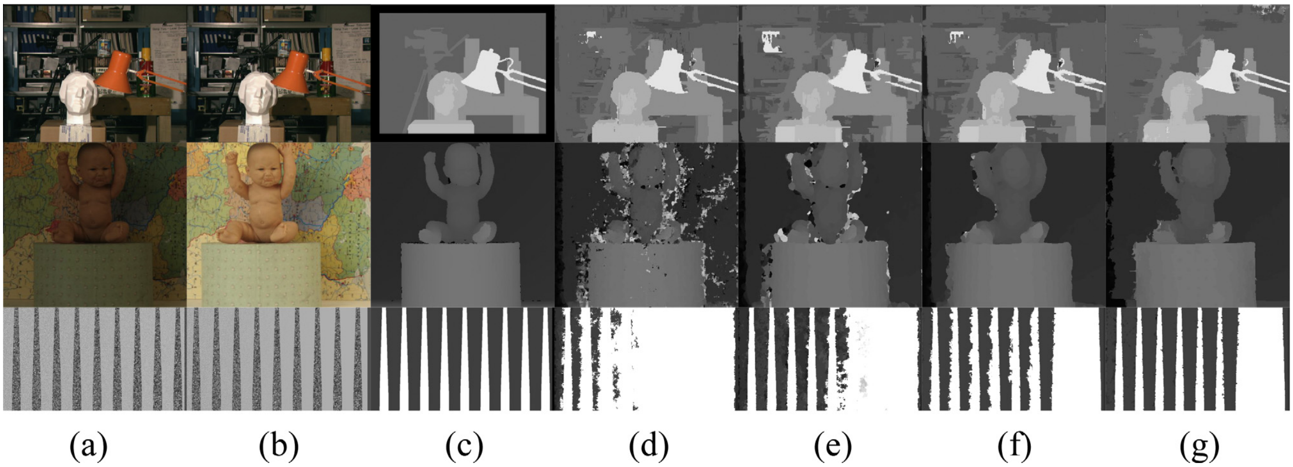


Fig. 4. Disparity results computed by different cost computation methods of Tsukuba and Baby1 stereo pairs on Middlebury dataset and foreground stereo pairs on HCI dataset with the existence of noise. (a–b) Left and right stereo images, (c) ground truth, (d) the absolute difference of gradient, (e) the Census transform, (f) the absolute difference of gradient plus Census transform, (g) the proposed method.

4.3. Measurement of the enhanced gradient

Gradient is a useful cost measurement, but the gradients of the raw image include little edge information. Thus, an enhanced gradient vector generated by the edge-preserving filter is introduced into the cost computation, which contains more edge information compared with the raw gradient vector. The gradients of the initial image and the enhanced image are shown in Fig. 3. The cost measurement is computed as

$$C^{E}_{GD,dir} = \min\left(\frac{1}{3}\sum_{c\in R,G,B}\|\nabla_{dir}I^{E}_{L,c}(p) - \nabla_{dir}I^{E}_{R,c}(p-d)\|, \lambda^{E}_{GD}\right), \qquad (23)$$

where superscript E refers to the enhanced image.

4.4. Combined matching cost

From Fig. 3, it can be found that the gradient of the raw image contains detailed local information, and the gradient of the enhanced image shows strong boundary information. Thus, the combined cost measurement is proposed in this paper, which includes both of the gradients to consider the local and global main boundaries. The combined matching cost can be obtained from three matching costs: the AD on the raw gradient vector, the Census transform, the AD on the enhanced gradient vector. This cost measurement considers both the local and global image information. A robust exponential function [13] has been utilized in the matching cost which is defined as

$$C(p,d) = 1 - e^{-\frac{\alpha C^{raw}_{GD,x} + \beta C^{raw}_{GD,y}}{(\alpha+\beta)\lambda^{raw}_{GD}}} + 1 - e^{-\frac{C_{cen}}{\lambda_{cen}}} + 1 - e^{-\frac{\eta C^{E}_{GD,x} + \delta C^{E}_{GD,y}}{(\eta+\delta)\lambda^{E}_{GD}}}, \qquad (24)$$

where $\alpha, \beta, \eta, \delta$ are weights representing the contribution of each component to the total cost. The values of three costs are scaled in the same value field. We assign different directions with different weights as they are not equally important. Fig. 4 compares the proposed combined cost with other matching cost methods. All of the disparity maps are initial matching results without post processing. To make the results more convincing, the foreground stereo pairs on HCI dataset are used in the experiment, which are more challenging as the noise exists. The results demonstrate that the proposed algorithm outperforms other methods and is more robust against noise and radiometric variations.

5. Cost aggregation

The EGF introduced in Section 3 can achieve a balance between the accuracy and computational complexity, and the filter can preserve edges well. Thus, EGF is adopted as the cost aggregation method. The disparity computation is also implemented to determine the initial disparity map. These two processes are described in detail as follows.

5.1. Cost aggregation

Cost aggregation is a crucial step in local stereo matching. Many stereo matching algorithms determine the proper support window for each pixel independently to produce accurate results. It is assumed that pixels that have similar intensities in a homogenous region are likely to have same disparity. In this paper, the cross-based window presented in [12] is adopted to construct the support region as shown in Fig. 5. The window is constructed by

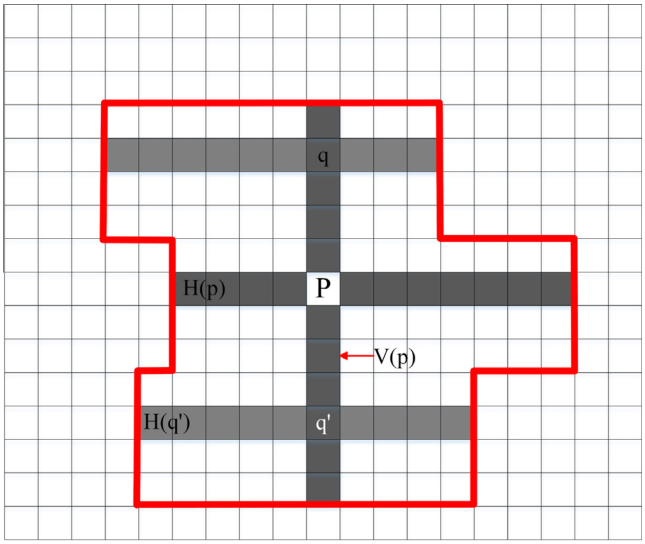


Fig. 5. The cross-based support window region at each pixel. $V(p)$ and $H(q)$ are the vertical and horizontal segments of each pixel, respectively.

expanding to four directions to form a cross-shaped skeleton support region around pixel p. Once the arm length is determined, the vertical and horizontal segments of p are constructed as

$$V(q) = \{(x,y)|\ x \in [x_p - l_v^-, x + l_v^+],\ y = y_p\}, \tag{25}$$

$$H(q) = \{(x,y)|\ y \in [y_p - l_h^-, y + l_h^+],\ x = x_p\}, \tag{26}$$

where $\{l_v^-, l_v^+, l_h^-, l_h^+\}$ represent the four arm length. And we further improve the window selection strategy by imposing an exponential threshold on the expansion in each direction to handle the textureless regions, which is computed as follows

$$\tau(L_{pq}) = \tau_{\max}\left(1 - ke^{\frac{L_{\max}-L_{pq}}{L_{\max}}}\right), \tag{27}$$

where L_{pq} and τ represent the Euclidian distance and color similarity between the pixel q and p respectively. $L_{\max}$ and $\tau_{\max}$ are two constants to control the effect of distance and color on the window selection. However, this method may fail to construct the support window in the same image structure when brightness is variable. Thus an adaptive threshold using the variance of the image is presented, which can be expressed as

$$\tau_{\max} = \frac{\sigma(I)}{s}, \tag{28}$$

where $\sigma(I)$ is the intensity deviation of image I and s is a constant which can change to adapt to different images. In many cross-based methods, the support window is determined solely on the raw image. Thus, a combined window is proposed in this study, which is the intersection of two cross windows of the raw image and enhanced image. The results of our proposed window selection method are shown in Fig. 6.

After determining the support window, the cost aggregation is implemented for each pixel which needs repetitive summations. This process is computationally expensive especially when the window is variable. The orthogonal integral image (OII) technique was proposed in [12] to accelerate the cost aggregation. Based on OII, our cost aggregation algorithm is proposed. Algorithm 1 presents the steps of our proposed cost aggregation method, which is shown in Fig. 7.

Algorithm 1: The steps of our proposed cost aggregation function

Input: The initial image pairs I_L and I_R, the enhanced image I_L^E and I_R^E, raw matching cost $C_{k,d}$ of pixel k, disparity hypothesis d.

Step 1: Compute $\tau_{\max}$ using Eq. (28) and compute four arm lengths $\{l_v^-, l_v^+, l_h^-, l_h^+\}$ for pixel k in I_L and I_L^E, respectively.

Step 2: Compute $\bar{a}_k$, $\bar{b}_k$ in the cross based support window centered at pixel p, respectively. $\bar{a}_k$, $\bar{b}_k$ are the mean of a_k, b_k in window ω_k, which are expressed as follow:

$$a_k = (\psi_k + \lambda U)^{-1}\left(\frac{1}{w_k}\sum_{i\in w_k}\gamma_i C_{i,d} - \mu_k \overline{C}_{k,d}\right), \tag{29}$$

$$b_k = \overline{C}_{k,d} - a_k^T \mu_k, \tag{30}$$

where γ is a 3×1 vector representing the RGB values. u_k is the mean of γ and ψ_k is the covariance matrix of γ in window ω_k. w_k is the number of pixels in window ω_k. U is a 3×3 identity matrix. $\overline{C}_{k,d}$ is the mean of $C_{k,d}$.

Step 3: Compute the horizontal integral image $S^H(x,y)$ as

$$S^H(x,y) = S^H(x-1,y) + (\bar{a}_k^T\gamma + \bar{b}_k), \tag{31}$$

The term $S^H(x,y)$ can be computed from $S^H(x-1,y)$ with only one addition. When $x = 0, S^H(-1,y) = 0$.

Step 4: Compute the horizontal integral $E^H(x,y)$ using the horizontal integral image $S^H(x,y)$.

$$E^H(x,y) = S^H(x + l_h^+, y) - S^H(x - l_h^- - 1, y), \tag{32}$$

Step 5: Compute the vertical integral image $S^V(x,y)$ as

$$S^V(x,y) = S^V(x, y-1) + E^H(x,y), \tag{33}$$

The term $S^V(x,y)$ can be computed from $S^V(x-1,y)$ with only one addition. When $y = 0, S^V(x,-1) = 0$.

Step 6: Based on $S^V(x,y)$, the fully aggregated matching cost $C_{agg}(x,y)$ can be derived from only one subtraction.

$$C_{agg}(x,y) = S^V(x, y + l_v^+) - S^V(x, y - l_v^- - 1), \tag{34}$$

5.2. Disparity computation

As in almost all the studies, the 'Winner-Take-All' strategy is used to compute the initial disparity map. Suppose that R is the range of disparity values. Then the optimal disparity d_p is selected according to

$$d_p = \arg\min_{d\in R} C_{agg}(p,d). \tag{35}$$

6. Disparity map post-processing

The raw disparity map obtained through above steps contains lots of outliers. Thus, post-processing is needed to obtain more accurate disparity map. However, it is hard to remove those outliers using only one method, so some stereo matching algorithms use multi-step post-processing method, like the adaptiveBP [29], DoubleBP [30]. In this paper, the multi-step post-processing strat-

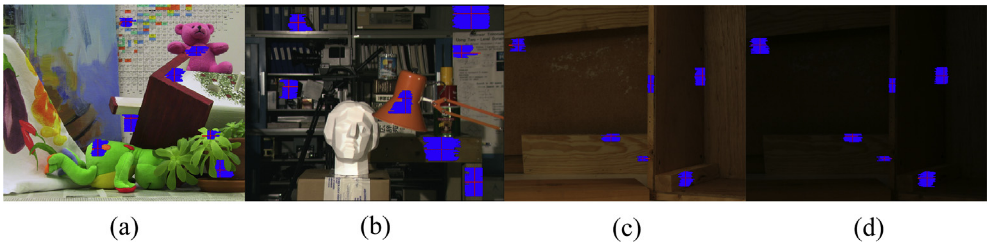


Fig. 6. Examples of the cross-based support window obtained from the proposed window selection method. (a) Teddy, (b) Tsukuba, (c–d) Wood1.

Algorithm 1 The proposed cost aggregation function

```
Input:$I_L, I_R, I_L^E, I_R^E$;
Compute $C_{k,d}, \tau_{max}, l_v^+, l_v^-, l_h^-, l_h^+$;
for $y = 1$ to $I_L.height$ do
    for $x = 1$ to $I_L.width$ do
        $a_k = (\Psi_k + \lambda U)^{-1} \times (\frac{1}{w_k} \sum \gamma_i C_{i,d} - \mu_k \bar{C}_{k,d})$;
        $b_k = \bar{C}_{k,d} - a_k^T \mu_k$;
        for $i = 0$ to $x$ do
            $S^H(x,y) = S^H(x,y) + (\bar{a}_{k_i}^T \gamma + \bar{b}_{k_i})$;
        end for
        $E^H(x,y) = S^H(x + l_h^+, y) - S^H(x - l - l_h^-, y)$;
        for $i = 0$ to $y$ do
            $S^V(x,y) = S^V(x,y) + E^H(x,i)$;
        end for
    end for
end for
```

Fig. 7. The proposed cost aggregation algorithm.

Table 1
Parameter settings of the proposed method.

Parameter	Value	Parameter	Value	Parameter	Value
λ	0.12	β	0.51	α	0.32
ε	0.001	η	0.40	s	2
λ_{cen}	35/255	δ	0.63	k_p	6
λ_{GD}^{raw}	10/255	$L_{\max}$	33	W_{cen}	5×7
λ_{GD}^{E}	12/255	$\tau_{\max}$	25/255	φ	0.3

egy is proposed to fully remove the invalid disparities, which includes LRC checking, Outlier suppression, modified Occweight filtering [24] and slanted plane smoothing [31].

6.1. Left-to-Right Consistency (LRC) check

In this paper, pixels are classified into three categories: (1) occluded pixels, (2) unstable pixels (3) stable pixels. At the beginning, the LRC check is adopted to determine the outliers in d_L. In this check, there is an assumption that the disparity of pixel $p(x,y)$ in the left image is equal to that of the corresponding pixel $q(x - \max(d_L, 0), y)$, namely

$$d_L(x,y) = d_R(x - \max(d_L, 0), y), \tag{36}$$

where $\max(d_L, 0)$ takes the larger value between $d_L(x,y)$ and 0. If this assumption does not hold, then the pixel is occluded pixel. However, the cross checking strategy may fail to detect some mismatched pixels. Thus, the correlation confidence measurement is utilized to detect the pixels, which can be expressed as

$$\left| \frac{C_1 - C_2}{C_2} \right|, \tag{37}$$

where C_1, C_2 are the best minimum the second minimum cost, respectively. If it is below a threshold φ, then the pixel is unstable pixel.

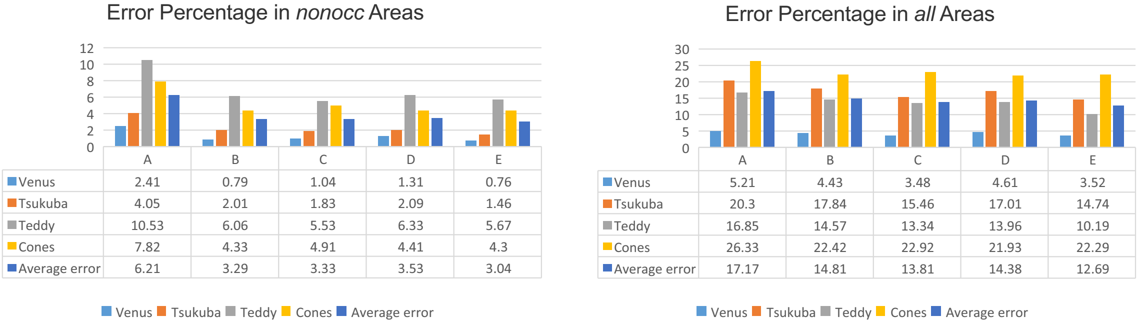
6.2. Outlier suppression

Once the outliers are detected, they are corrected with the nearest valid pixel of outliers in the vertical and horizontal directions. Suppose that the disparities of the nearest valid pixels are *dl*, *dr* respectively. The invalid pixel is updated as

$$d_p^* = \min(dl, dr), \tag{38}$$

where $\min(dl, dr)$ represents the smaller value of *dl* and *dr*. But for leftmost outliers, *dl* does not exist due to the loss of related matching information in the right image. The scan-line propagation in two directions cannot work well. Thus, we extend it to four directions by searching the most reliable pixel along four arms.

$$d_p^* = \begin{cases} dlr & \text{if only } dlr \text{ existing} \\ dud & \text{else if only } dud \text{ existing} \\ 2dr - dr' & \text{else if only } dl \text{ does not exist} \\ null & \text{else if } dlr \text{ and } dud \text{ not existing} \\ \min(dlr, dud) & \text{otherwise} \end{cases} \tag{39}$$



	A	B	C	D	E
Venus	2.41	0.79	1.04	1.31	0.76
Tsukuba	4.05	2.01	1.83	2.09	1.46
Teddy	10.53	6.06	5.53	6.33	5.67
Cones	7.82	4.33	4.91	4.41	4.3
Average error	6.21	3.29	3.33	3.53	3.04

(a)

	A	B	C	D	E
Venus	5.21	4.43	3.48	4.61	3.52
Tsukuba	20.3	17.84	15.46	17.01	14.74
Teddy	16.85	14.57	13.34	13.96	10.19
Cones	26.33	22.42	22.92	21.93	22.29
Average error	17.17	14.81	13.81	14.38	12.69

(b)

Fig. 8. Visualized performance of raw disparity maps of different cost measurements, *nonocc* and *all* mean the non-occluded and total area, respectively. A: AD on color, B: Census transform, C: AD on gradient, D: Census transform plus AD on gradient, E: The proposed cost computation method. (a) Error percentage in *nonocc* areas, (b) error percentage in *all* areas.

where *dlr* represents $\min(dl, dr)$, and *dud* is the minimum disparity value of the nearest valid pixels in the vertical direction. dr' is the second nearest reliable pixel in the right arm of outliers. In order to completely remove the outliers, this step runs for two times.

6.3. Cross-based Occweight filtering

This step is proposed to correct the unstable pixels. Wang and Zhang [24] proposed a Occweight method to update the pixel with the most likely disparity in the square window. The adaptive cross-based window is proposed as described in Section 5 because the fixed square window fails to keep robust under radiometric illuminations. In the support window centered at p, the weight for each pixel is computed as

$$w(p,q)=\begin{cases}\exp\left(-\frac{\Delta c_{pq}}{\varphi_c}-\frac{\Delta s_{pq}}{\varphi_r}\right), & \text{if } q \text{ is stable}\\ 0, & \text{else}\end{cases} \tag{40}$$

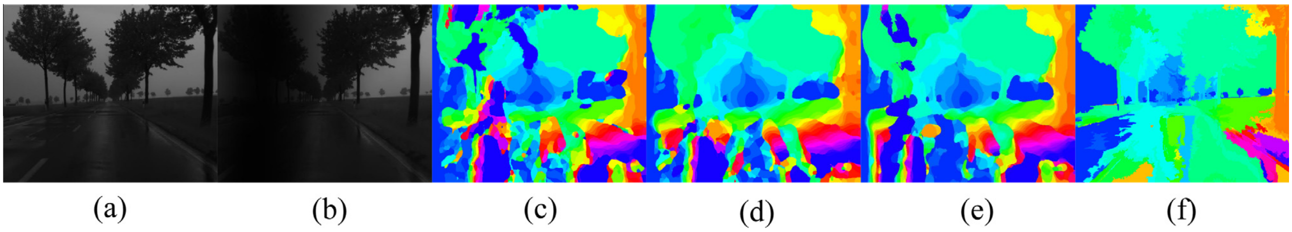


Fig. 9. Disparity results computed by different cost computation methods of Rain Blur stereo pairs on HCI dataset. (a–b) Left and right stereo images, (c) GDxy, (d) Census, (e) CG, (f) the proposed method.

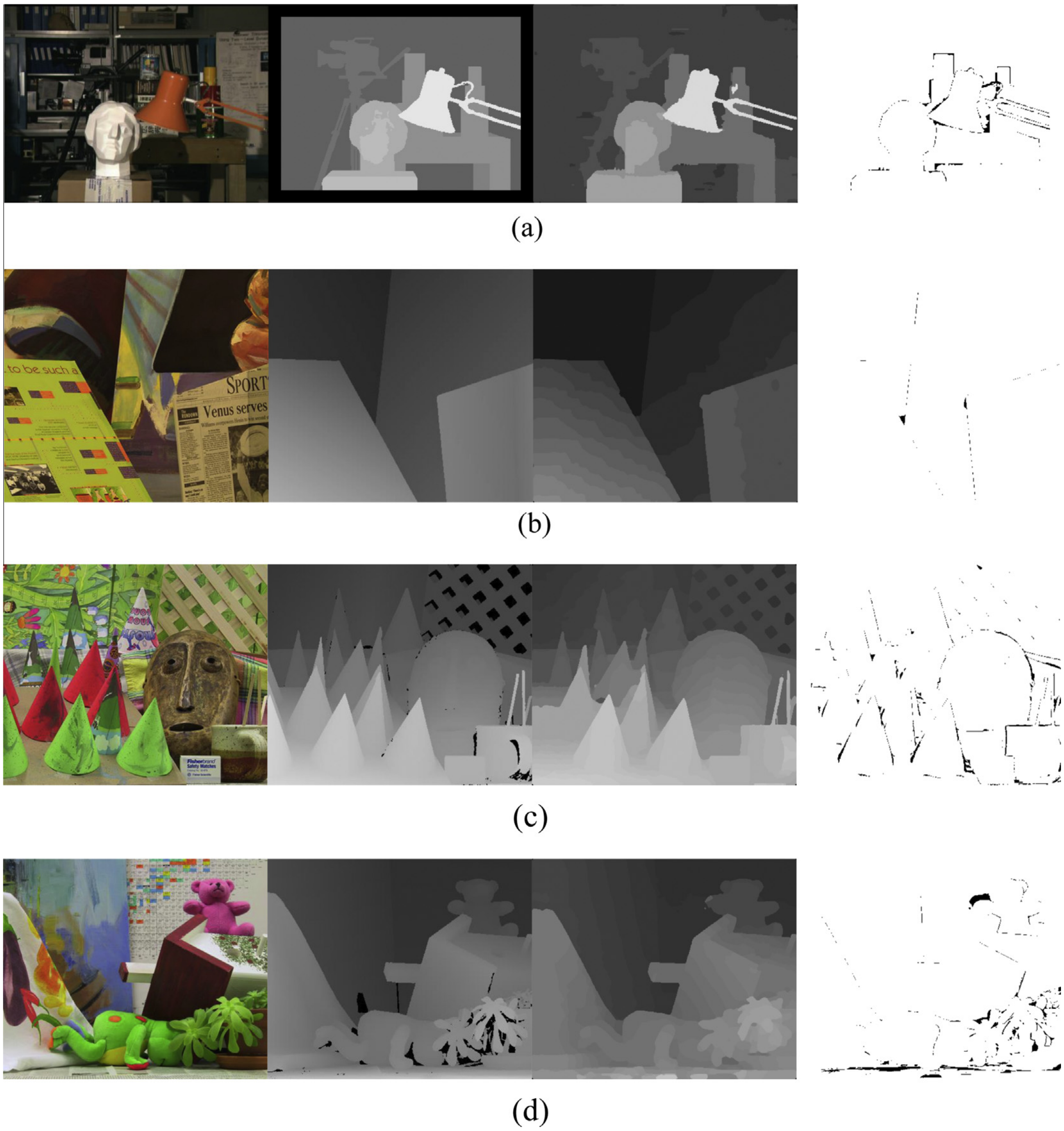


Fig. 10. Our results of the Middlebury dataset: (a) Tsukuba, (b) Venus, (c) Cones, (d) Teddy. From the left image to the right image: original image, ground truth maps, results of the proposed algorithm, error maps.

where Δc_{pq} represents the color similarity between p and q, and Δs_{pq} is the spatial distance between the coordinates of p and q. Then the updated disparity is expressed as

$$d_p^* = \arg\max_{d \in R} \left(\sum_{q \in W_p} w(p,q) \times f(q,d) \right). \tag{41}$$

$f(q,d) = 1$ if $d(q) = d$, else $f(q,d) = 0$, where W_p represents the set of pixels within the support window centered at p.

6.4. Slanted plane smoothing

Slanted plane models for stereo were first introduced by Birchfield and Tomasi [31]. Then, several slanted-plane algorithms were proved successful for stereo matching [29,32]. Yamaguchi et al. [25] proposed a slanted-plane MRF model which infers segmentations as well as boundaries. It achieved good performance. Slanted plane smoothing was proposed in [33] that jointly solves segments, occlusion and estimated disparity map. Based on this idea, we use slanted plane smoothing to post process the remaining invalid pixels after the post-processing steps above.

7. Experimental results

In this paper, comprehensive experiments have been conducted to evaluate the proposed stereo matching method. First, experiments with different cost measurements are carried out to evaluate the advantage of our cost computation method. Next we will compare with other stereo matching algorithms and analyze the adaptability of the proposed method.

7.1. Parameter setting

In this paper, we focus on two benchmarks: the Middlebury benchmark and the KITTI benchmark. The Middlebury benchmark provides more than 30 pairs of stereo images. The experiments are conducted not only on the four standard pairs (Tsukuba, Venus, Teddy, Cones), but also on the other 26 pairs of images for a more comprehensive analysis. We also carry out the experiments on the KITTI benchmark to test the adaptability of our algorithm. The performance is evaluated by computing the average percentage of erroneous pixels. The experimental parameter setting is shown in Table 1. In order to make the results more convincing, the parameters keep the same for all data sets.

7.2. Evaluation of the proposed cost computation method

At first, the performance of different cost measurements is discussed, including the absolute difference of color (AD), the absolute difference of gradient (GDxy) and the Census transform. Then, experiments are conducted on the four stereo pairs (i.e. Tsukuba, Venus, Cones, Teddy) to verify the effectiveness of the proposed cost measurement. Fig. 8 shows the experiment results. It is demonstrated that our combined algorithm outperforms the traditional cost measurement.

The AD, Census and GDxy methods are commonly used in the cost computation. However, these operators have some drawbacks: the AD cannot handle large textureless regions effectively, the Census fails to process image with repetitive local structure, and the tiny boundaries is hard for GDxy. Recently, some combined cost computation algorithms have been presented to absorb the advantage of each operator mentioned above, like AD plus GDxy, Census plus GDxy(CG). In this paper, our cost computation method consists of the Census transform, the absolute difference of the gradient of the raw images and the enhanced images. Because the Census transform is robust against radiometric differences. The gradient-based algorithm improves the accuracy in non-occluded areas. And the enhanced gradient-based method can handle the discontinuity region effectively.

Table 2
Error percentage in non-occlusion regions.

Algorithm	mSGM-LDE	ARW	GF	AEGF	MC-CNN
	[35]	[34]	[37]		[36]
Tsukuba	2.46	2.85	2.45	1.32	1.06
Venus	1.05	0.73	1.16	0.26	0.21
Cones	3.58	3.81	3.23	2.01	1.87
Teddy	7.72	7.05	7.18	5.21	4.80
Barn1	0.89	0.97	0.77	0.51	0.55
Barn2	1.21	1.14	1.02	0.35	0.38
Bull	0.45	0.75	0.42	0.41	0.35
Poster	1.38	1.64	1.55	1.38	0.79
Sawtooth	0.96	1.21	1.21	1.02	0.83
Art	10.24	8.68	8.93	7.58	5.51
Books	10.36	9.45	9.32	8.05	7.28
Dolls	6.29	5.69	5.37	5.36	3.95
Laundry	14.18	13.61	12.89	14.32	10.27
Moebius	9.33	8.52	9.12	8.71	7.52
Reindeer	6.57	6.82	7.25	3.57	2.87
Aloe	6.25	5.77	6.83	5.25	4.75
Baby1	4.83	4.21	3.82	3.18	3.14
Baby2	5.94	5.25	5.78	4.15	3.83
Baby3	6.85	5.16	5.35	4.34	3.92
Bowling1	14.35	14.18	14.97	13.33	11.58
Bowling2	10.87	9.95	9.15	7.13	6.57
Cloth1	0.74	0.85	0.73	0.49	0.85
Cloth2	3.69	3.16	3.52	2.88	3.12
Cloth3	3.29	2.53	1.49	1.31	1.35
Cloth4	1.83	1.93	1.85	1.54	1.27
Flowerpots	14.21	14.72	13.53	10.39	8.29
Lampshade1	14.15	14.15	14.21	11.43	10.36
Lampshade2	18.36	18.61	18.71	16.98	14.68
Rocks1	3.83	3.94	3.84	2.35	2.21
Rocks2	2.66	2.88	2.33	2.44	2.05
Average error	6.28	6.01	5.93	4.90	4.21

To demonstrate the advantage of our cost computation method, a quantitative comparison is made with three cost measurements (i.e. AD, Census and GDxy). Besides, the combined cost measurement is also taken into account. As shown in Fig. 8, the proposed cost measurement performs better than other methods. Furthermore, to test the robustness of our method, experiments with more challenging HCI dataset are conducted. The Rain Blur stereo pairs are used in the experiment. The visual comparisons are shown in Fig. 9. From Fig. 9, it can be found that in the more challenging environment, our proposed method still outperforms other cost measurements.

7.3. Evaluation on Middlebury data set

Recently, a variety of ASW-based cost aggregation algorithms have been proposed, such as the guided filter based method, bilateral filter based method. They indeed achieve good performance. In this study, comparisons are made not only with the GF-based method, but also with other methods presented recently: the ARW method [34], the mSGM-LDE method [35] and MC-CNN [36] method. First, we carry out the experiment on the four stereo image pairs. The disparity results are shown in Fig. 10. Then, experiments are conducted on 30 pairs of stereo image including the four image pairs mentioned above. Various parameter settings are tried to get the optimal results. The results are determined by computing the percentages of invalid pixels in non-occluded areas with error threshold 1. The quantitative results are shown in Table 2 and the visual comparisons are shown in Fig. 11. In our experiment, the Middlebury 2005 and 2006 training datasets

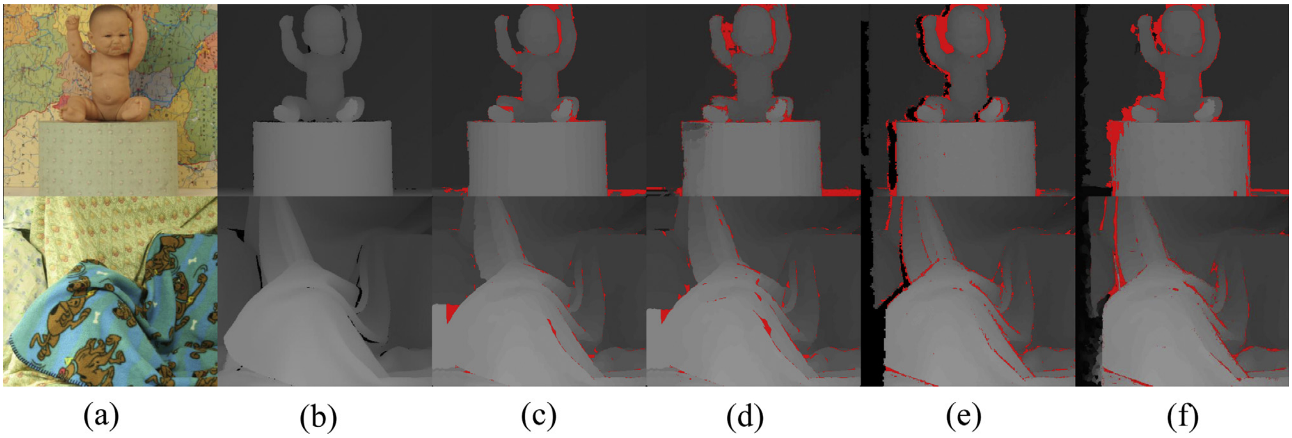


Fig. 11. Disparity results of Baby1 and Cloth3 stereo pairs. The bad pixels in non-occluded areas are labeled in red. (a) Original image, (b) ground truth maps, (c) the proposed method, (d) CostFilter [37], (e) ARW [34], (f) mSGM-LDE [35]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

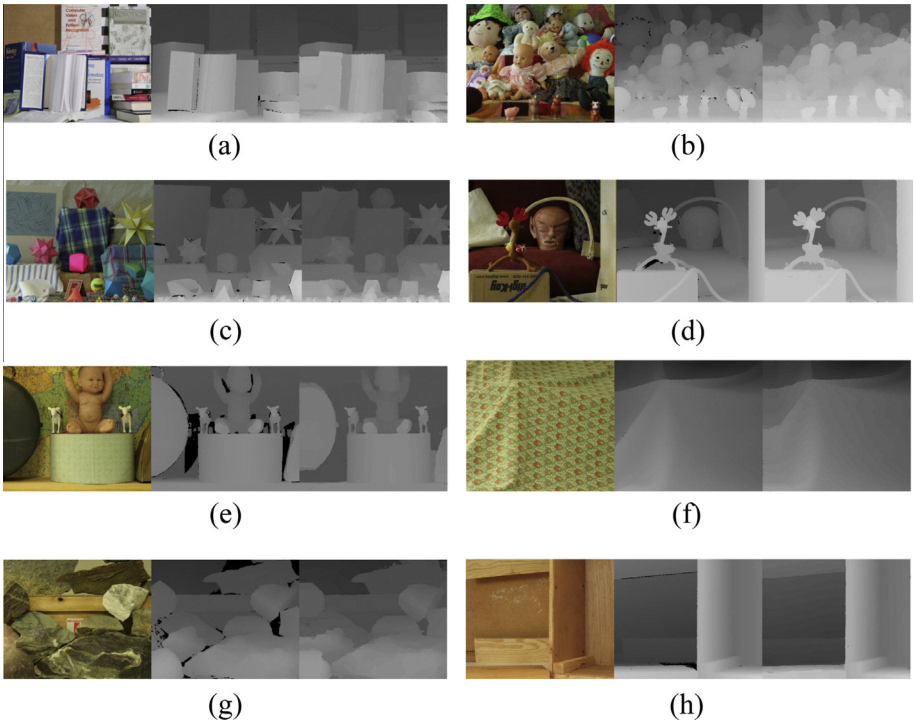


Fig. 12. The Middlebury 2005 datasets and 2006 datasets. Left to right: Original image, ground truth maps, left disparity maps. (a) Books, (b) Dolls, (c) Moebius, (d) Reindeer, (e) Baby3, (f) Cloth1, (g) Rocks1, (h) Wood1.

are also used for visual evaluation, which is illustrated in Fig. 12. The results will be analyzed as follows.

First, from Fig. 10, the AEGF preserves the edges well because of the edge-preserving filter. Most errors are at occluded regions, which is a problem for almost all the local stereo matching methods including AEGF. As can be found in Table 2, our proposed AEGF method outperforms the GF-based method in most stereo pairs, which shows the effectiveness of the edge-preserving filter in cost aggregation. The average percentage of error of the proposed method decreases by 1.03% comparing with GF. It also demonstrates that the proposed algorithm achieve better performance than the ARW and mSGM-LDE in non-occluded areas. The average percentage of error is decreased by 1.38% comparing with mSGM-LDE and 1.11% comparing with ARW. It can also be found in the visual comparison in Fig. 10 that AEGF produces less error pixels. Compared with MC-CNN, our results are better on Cloth1, Cloth2. Moreover, the MC-CNN method needs pre-training and is much more time-consuming, which makes it impractical for application.

7.4. Evaluation on KITTI data sets

Most of the stereo matching algorithms only evaluate their methods on the standard Middlebury Benchmark. However, in this experiment, we also evaluate the proposed method on the KITTI datasets to prove the adaptability of our method. The results are reported in Table 3 and the visual comparisons are illustrated in Fig. 13. In Table 3, four indicators are evaluated. “Out-Noc”: percentage of bad pixels in non-occluded regions. “Out-All”: percentage of bad pixels in total. “Avg-Noc”: average disparity error in non-occluded regions. “Avg-All”: average disparity error in total

Table 3
Evaluation results on KITTI benchmark with error threshold 3.

Method	Out-Noc	Out-All	Avg-Noc	Avg-All	Runtime	Environment
MC-CNN [36]	2.61%	3.84%	0.8 px	1.0 px	100 s	Nvidia GTX Titan (CUDA, Lua/Torch7)
Our method	4.81%	6.12%	1.2 px	1.8 px	31.1 s	1 cores @ 3.2 GHz (C/C++)
ATGV [38]	5.02%	6.88%	1.0 px	1.6 px	6 min	>8 cores @ 3.0 GHz (Matlab + C/C++)
ARW [34]	5.20%	6.87%	1.2 px	1.5 px	4.6 s	1 core @ 3.5 GHz (C/C++)
DLP	5.28%	7.21%	1.2 px	2.0 px	60 s	8 cores @ >3.5 GHz (C/C++)
mSGM-LDE [35]	6.01%	8.22%	1.4 px	2.4 px	55 s	2 cores @ 2.5 GHz (C/C++)
OCV-SGBM [8]	7.64%	9.13%	1.8 px	2.0 px	1.1 s	1 core @ 2.5 GHz (C/C++)
ELAS [39]	8.24%	9.96%	1.4 px	1.6 px	0.3 s	1 core @ 2.5 GHz (C/C++)
Deep-Raw [40]	8.93%	11.07%	3.9 px	4.9 px	1 s	1 core @ 2.5 GHz (C/C++)
CrossScaleGF [41]	9.03%	11.21%	2.1 px	3.4 px	140 s	1 core @ 3.0 GHz (C/C++)
GF(Census) [41]	11.65%	13.76%	4.5 px	5.6 px	120 s	1 core @ 3.0 GHz (C/C++)
CostFilter [37]	19.99%	21.08%	5.0 px	5.4 px	4 min	1 core @ 2.5 GHz (Matlab)

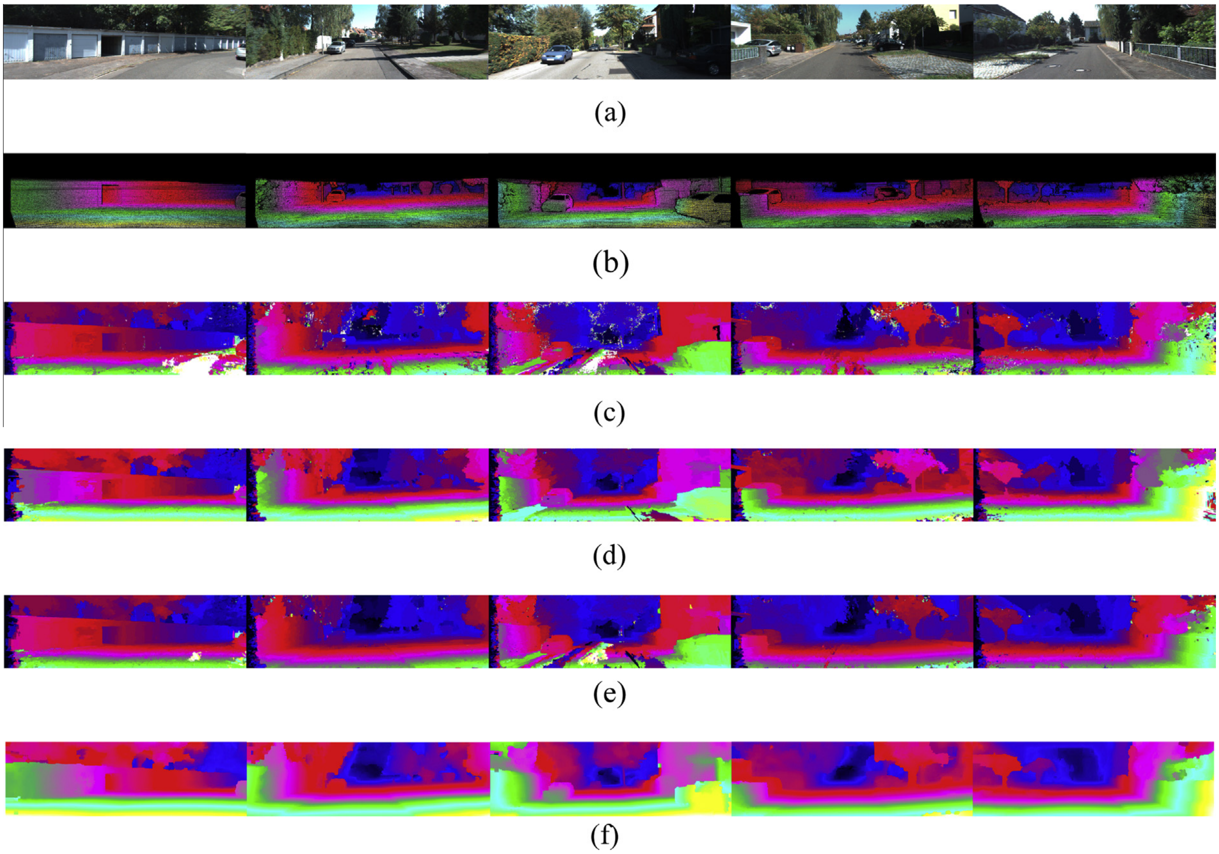


Fig. 13. Disparity results on KITTI dataset (frames #000023, #000024, #000032, #000065, #000085). (a) Left image, (b) ground truth, (c–f) disparity results of different methods. (c) mSGM-LDE [35], (d) ARW [34], (e) CostFilter [37], (f) our proposed method.

[23]. From Table 3, we can see that our proposed algorithm performs better than other GF-based method. The reason is that in the KITTI dataset, the images are captured in the real-world conditions. Thus, there exist many textureless regions and variable brightness conditions are inevitable. As the support window in GF-raw, GF-Census and cross-scale GF is fixed and not robust to environment changes. On the contrary, the proposed method AEGF uses the edge-preserving filter to keep the edges and adapts to the image structure. It can also be found in Fig. 13 that the disparity map generated by our method is more smoothing and contains less black areas.

Moreover, we evaluate the cost aggregation methods and post processing approaches using the training set of the KITTI dataset. In this experiment, we compare different cost aggregation and post processing methods while other steps keep the same. The results are reported in Table 4. Gray image model and RGB image model represent using Gray model and RGB model in the EGF for the cost aggregation step respectively. *Occ* and *spm* represent Occweight filtering and slanted plane model in the post processing step respectively.

Table 4
Evaluation of different cost aggregation and post processing methods.

Algorithm	Average error percentage
Gray image model	7.25
Gray image model + *Occ*	6.49
Gray image model + *Occ* + *spm*	5.87
RGB image model	6.63
RGB image model + *Occ*	5.54
RGB image model + *Occ* + *spm*	4.93

From the above experimental results, it can be shown the proposed stereo matching method is effective. To ensure our method

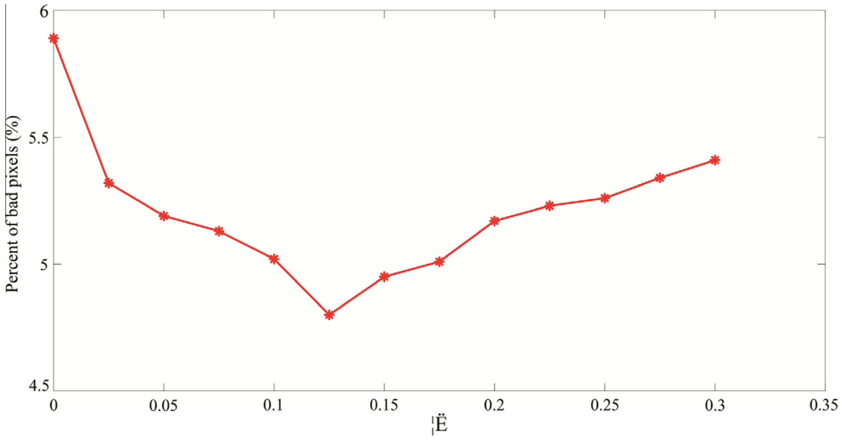


Fig. 14. Performance comparison of different λ on Middlebury dataset.

works well as described above, the pre-processing is necessary. And the combined cost computation method and multi-step post processing are required. The refinement is taken based on the qualities of the outliers. Moreover, appropriate parameters are essential for the edge-preserving filter and the cost computation. In this study, we conducted some experiments by varying the regularization parameter λ to test the robustness of our method. The result is shown in Fig. 14. The result shown in Fig. 14 demonstrates that proper λ should be chosen to obtain good performance. When the value of λ is too small, the proposed filter turns into the guided filter, the percent of bad pixels increases. Without a proper λ, the performance of the proposed method will be degraded.

7.5. Computational complexity analysis

The complexity of the overall method can be computed by analyzing each step. Suppose that R_i is the resolution of the image and D is the range of the disparity candidates. The complexity of each step is described as: $O(k_1k_2R_iD + k_3R_iD + k_4R_iD)$ for cost computation; $O(L_{\max}R_i)$ for support window adjusting; $O(R_iD)$ for cost aggregations; $O(R_bD)$ for post processing. In these expressions, k_1 is the number of gradient directions; R_b is the number of bad pixels. k_2, k_3 are the number of color channel of the image; k_4 ($O(k_4)$ = $O(N^2)$) is the complexity of the Census transform. The runtime of each step of the proposed method in the KITTI dataset is reported in Table 5. We can see from Table 5 that most of the processing time is consumed by the cost computation and cost aggregation, which is similar to other stereo matching methods. We can also see from Table 3 that the average processing time of the MC-CNN method is 100 s using NVIDIA GTX Titan for parallel computation. But that of the proposed method is 31.1 s which only uses 1 core without any acceleration.

8. Conclusion

In this paper, an edge-preserving guided filter (EGF) is presented, which extends the GF by introducing an edge-aware term. Furthermore, based on EGF, we propose a cost aggregation method using adaptive edge-preserving guided filter (AEGF). The AEGF method can achieve proper cost volume filtering as well as edge-preserving. The experimental results demonstrated that our proposed local stereo matching algorithm can produce accurate results as well as keep linear computational complexity. The results also show that our method can offer accurate performance in both indoor and outdoor scenes. In the future work, this study could be extended by using parallel method to achieve real-time performance. We will evaluate different edge-preserving filters to further understand their influence on the performance of proposed stereo matching method.

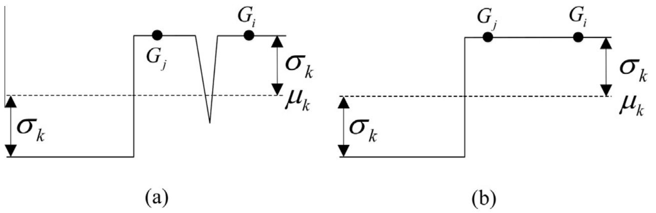


Fig. 15. Two cases of a step edge.

Acknowledgments

The authors would like to thank the Hangzhou Civic Significant Technological Innovation Project of China (No: 20131110A04), Hangzhou Civic Significant Technological Innovation Project of China (No: 20142013A56) for supporting this work.

Appendix A

As described in [17], the GF filter fails to sort the depth ambiguity. The reason will be explained as follows. We take the gray-scale guidance image as example. The kernel function, which determines the weight for each pixel within the local support window, can be expressed as

$$W_{ij}(G) = \frac{1}{|w|^2} \sum_{k:(i,j)\in w_k} \left(1 + \frac{(G_i - \mu_k)(G_j - \mu_k)}{\sigma_k^2 + \varepsilon_{gf}}\right), \tag{42}$$

where $|w|$ is the pixel number in w_k. μ_k is the mean value of guidance image G in the window w_k and σ_k are the variance of G in the window w_k. ε_{gf} denotes the smoothness parameter. To consider why GF fails to handle the depth ambiguity, let us consider two cases shown in Fig. 15.

Table 5
Runtime of each step of the proposed algorithm.

Step	Preprocessing	Cost computation	Cost aggregation	Disparity computation	Disparity refinement
Runtime (s)	2.2	13.7	10.9	0.5	3.8

As shown in Fig. 15, there exists a sharp edge between i and j, but the weights computed in Eq. (42) for the pixel pairs are the same as all the parameters are equivalent. Thus, the GF function cannot distinguish these two cases. Even with the proposed edge-preserving filter, the depth ambiguity still cannot be solved. The filter kernel of the proposed filter which is given by

$$W'_{ij} = \frac{\partial \hat{Z}_i}{\partial I_j}. \tag{43}$$

The derivative gives

$$\frac{\partial \hat{Z}_i}{\partial I_j} = \frac{1}{|w|} \sum_{k \in w_i} \left(\frac{\partial a_k}{\partial I_j} (G_i - \mu_{G,\xi_1}(k)) + \frac{\partial \mu_{I,\xi_1}(k)}{\partial I_j} \right), \tag{44}$$

where

$$\frac{\partial \mu_{I,\xi_1}(k)}{\partial I_j} = \frac{1}{|w|} \delta_{k \in w_j}, \tag{45}$$

$$\begin{aligned} \frac{\partial a_k}{\partial I_j} &= \frac{1}{\sigma^2_{G,\zeta_1}(k) + \frac{\lambda}{\hat{\Gamma}_G(k)}} \left(\frac{\partial \mu_{G\bullet I,\xi_1}(k)}{\partial I_j} - \mu_{G,\xi_1}(k) \frac{\partial \mu_{I,\xi_1}(k)}{\partial I_j} \right) \\ &= \frac{1}{\sigma^2_{G,\zeta_1}(k) + \frac{\lambda}{\hat{\Gamma}_G(k)}} \left(\frac{1}{|w|} G_j - \frac{1}{|w|} \mu_k \right) \delta_{k \in w_j}, \end{aligned} \tag{46}$$

where $\delta_{k \in w_j}$ is one when j is in the window w_k and is zero otherwise. Putting Eq. (45) and Eq. (46) into Eq. (44), it gives

$$W'_{ij} = \frac{1}{|w|^2} \sum_{k:(i,j) \in w_k} \left(1 + \frac{(G_i - \mu_k)(G_j - \mu_k)}{\sigma_k^2 + \frac{\lambda}{\hat{\Gamma}_G(k)}} \right). \tag{47}$$

The filter kernel of the edge-preserving filter is similar to that of the GF filter. Thus, the edge-preserving filter cannot handle the depth ambiguity as well. So the adaptive edge-preserving filter is proposed to solve this problem. Because the cross-based method determines the support window based on the image structures.

References

[1] L.E. Gurrieri, E. Dubois, Depth consistency and vertical disparities in stereoscopic panoramas, J. Electron. Imaging 23 (2014). 011004–011004.

[2] Q. Wang, Z. Yu, C. Rasmussen, J. Yu, Stereo vision-based depth of field rendering on a mobile device, J. Electron. Imaging 23 (2014). 023009–023009.

[3] D. Scharstein, R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, Int. J. Comput. Vision 47 (2002) 7–42.

[4] A.F. Bobick, S.S. Intille, Large occlusion stereo, Int. J. Comput. Vision 33 (1999) 181–200.

[5] Y. Boykov, O. Veksler, R. Zabih, Fast approximate energy minimization via graph cuts, The Proceedings of the Seventh IEEE International Conference on Computer Vision, vol. 1, 1999, pp. 377–384.

[6] Q. Yang, Stereo matching using tree filtering, IEEE Trans. Pattern Anal. Mach. Intell. 37 (2015) 834–846.

[7] H. Hirschmuller, D. Scharstein, Evaluation of stereo matching costs on images with radiometric differences, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2009) 1582–1599.

[8] H. Hirschmuller, Stereo processing by semiglobal matching and mutual information, IEEE Trans. Pattern Anal. Mach. Intell. 30 (2008) 328–341.

[9] F. Tombari, S. Mattoccia, L.D. Stefano, E. Addimanda, Classification and evaluation of cost aggregation methods for stereo correspondence, in: IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008, 2008, pp. 1–8.

[10] A. Hosni, M. Bleyer, M. Gelautz, C. Rhemann, Local stereo matching using geodesic support weights, in: 2009 16th IEEE International Conference on Image Processing (ICIP), 2009, pp. 2093–2096.

[11] O. Veksler, Fast variable window for stereo correspondence using integral images, Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, 2003. pp. I-556–I-561.

[12] K. Zhang, J. Lu, G. Lafruit, Cross-based local stereo matching using orthogonal integral images, IEEE Trans. Circ. Syst. Video Technol. 19 (2009) 1073–1079.

[13] Y. Kuk-Jin, K. In, So, adaptive support-weight approach for correspondence search, IEEE Trans. Pattern Anal. Mach. Intell. 28 (2006) 650–656.

[14] S. Zhu, D. Cao, Y. Wu, S. Jiang, Edge-aware dynamic programming-based cost aggregation for robust stereo matching, J. Electron. Imaging 24 (2015). 043016–043016.

[15] K. He, J. Sun, X. Tang, Guided image filtering, IEEE Trans. Pattern Anal. Mach. Intell. 35 (2013) 1397–1409.

[16] Q. Yang, P. Ji, D. Li, S. Yao, M. Zhang, Fast stereo matching using adaptive guided filtering, Image Vis. Comput. 32 (2014) 202–211.

[17] D. Chen, M. Ardabilian, L. Chen, A fast trilateral filter-based adaptive support weight method for stereo matching, IEEE Trans. Circ. Syst. Video Technol. 25 (2015) 730–743.

[18] Y. Xu, Y. Zhao, M. Ji, Local stereo matching with adaptive shape support window based cost aggregation, Appl. Opt. 53 (2014) 6885–6892.

[19] J. Lu, K. Shi, D. Min, L. Lin, M.N. Do, Cross-based local multipoint filtering, IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2012 (2012) 430–437.

[20] L. Liu, W. Lin, Y.P. Zhong, Traffic flow matching with clique and triplet cues, in: 2015 IEEE 17th International Workshop on Multimedia Signal Processing (MMSP), 2015, pp. 1–6.

[21] W. Lin, Y. Mi, W. Wang, J. Wu, J. Wang, T. Mei, A diffusion and clustering-based approach for finding coherent motions and understanding crowd scenes, IEEE Trans. Image Process. 25 (2016) 1674–1687.

[22] Middlebury Stereo Datasets. <http://vision.middlebury.edu/stereo/data>.

[23] The KITTI Vision Benchmark. <http://www.cvlibs.net/datasets/kitti/eval_stereo_flow.php?benchmark=stereo/>.

[24] W. Wang, C. Zhang, Local disparity refinement with disparity inheritance, in: 2012 Symposium on Photonics and Optoelectronics (SOPO), 2012, pp. 1–4.

[25] K. Yamaguchi, T. Hazan, D. McAllester, R. Urtasun, Continuous Markov random fields for robust stereo estimation, in: A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, C. Schmid (Eds.), Proceedings, Part V Computer Vision – ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7–13, 2012, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 45–58.

[26] Z. Li, J. Zheng, Z. Zhu, W. Yao, S. Wu, Weighted guided image filtering, IEEE Trans. Image Process. 24 (2015) 120–129.

[27] A. Levin, D. Lischinski, Y. Weiss, A closed-form solution to natural image matting, IEEE Trans. Pattern Anal. Mach. Intell. 30 (2008) 228–242.

[28] C. Zhang, W. Lin, W. Li, B. Zhou, J. Xie, J. Li, Improved image deblurring based on salient-region segmentation, Signal Process.: Image Commun. 28 (2013) 1171–1186.

[29] A. Klaus, M. Sormann, K. Karner, Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure, Proceedings of the 18th International Conference on Pattern Recognition, vol. 03, IEEE Computer Society, 2006, pp. 15–18.

[30] Y. Qyngxiong, W. Liang, Y. Ruigang, H. Stewenius, D. Nister, Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling, IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2006 (2006) 2347–2354.

[31] S. Birchfield, C. Tomasi, Multiway cut for stereo and motion with slanted surfaces, The Proceedings of the Seventh IEEE International Conference on Computer Vision, vol. 1, 1999, pp. 489–495.

[32] W. Zeng-Fu, Z. Zhi-Gang, A region based stereo matching algorithm using cooperative optimization, in: IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008, 2008, pp. 1–8.

[33] K. Yamaguchi, D. McAllester, R. Urtasun, Efficient joint segmentation, occlusion labeling, stereo and flow estimation, in: D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars (Eds.), Computer Vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V, Springer International Publishing, Cham, 2014, pp. 756–771.

[34] S. Lee, J.H. Lee, J. Lim, I.H. Suh, Robust stereo matching using adaptive random walk with restart algorithm, Image Vis. Comput. 37 (2015) 1–11.

[35] V.D. Nguyen, D.D. Nguyen, S.J. Lee, J.W. Jeon, Local density encoding for robust stereo matching, IEEE Trans. Circ. Syst. Video Technol. 24 (2014) 2049–2062.

[36] Z.a.Y. LeCun, Computing the stereo matching cost with a convolutional neural network, in: IEEE International Conference on Computer Vision, 2015, pp. 1592–1599.

[37] A. Hosni, C. Rhemann, M. Bleyer, C. Rother, M. Gelautz, Fast cost-volume filtering for visual correspondence and beyond, IEEE Trans. Pattern Anal. Mach. Intell. 35 (2013) 504–511.

[38] R. Ranftl, T. Pock, H. Bischof, Minimizing TGV-based variational models with non-convex data terms, in: A. Kuijper, K. Bredies, T. Pock, H. Bischof (Eds.), Proceedings of the Scale Space and Variational Methods in Computer Vision: 4th International Conference, SSVM 2013, Schloss Seggau, Leibnitz, Austria, June 2–6, 2013, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 282–293.

[39] A. Geiger, M. Roser, R. Urtasun, Efficient large-scale stereo matching, in: R. Kimmel, R. Klette, A. Sugimoto (Eds.), Computer Vision – ACCV 2010: 10th Asian Conference on Computer Vision, Queenstown, New Zealand, November 8–12, 2010, Revised Selected Papers, Part I, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 25–38.

[40] Z. Chen, X. Sun, L. Wang, Y. Yu, C. Huang, A deep visual correspondence embedding model for stereo matching costs, IEEE International Conference on Computer Vision (ICCV) 2015 (2015) 972–980.

[41] K. Zhang, Y. Fang, D. Min, L. Sun, S. Yang, S. Yan, Q. Tian, Cross-scale cost aggregation for stereo matching, IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2014 (2014) 1590–1597.