Hindawi
Mathematical Problems in Engineering
Volume 2019, Article ID 3128172, 11 pages
https://doi.org/10.1155/2019/3128172

Research Article

Efficient Stereo Matching Based on Pervasive Guided Image Filtering

Chengtao Zhu [1] **and Yau-Zen Chang** [2,3]

[1] *School of Microelectronics, Tianjin University, Tianjin 300072, China*
[2] *Department of Mechanical Engineering, Chang Gung University, Taoyuan 33302, Taiwan*
[3] *Department of Neurosurgery, Chang Gung Memorial Hospital, Taoyuan 33305, Taiwan*

Correspondence should be addressed to Yau-Zen Chang; zen@mail.cgu.edu.tw

Received 4 March 2019; Accepted 3 April 2019; Published 17 April 2019

Academic Editor: Oscar Reinoso

This paper presents an effective cost aggregation strategy for dense stereo matching. Based on the guided image filtering (GIF), we propose a new aggregation scheme called Pervasive Guided Image Filtering (PGIF) to introduce weightings to the energy function of the filter which allows the whole image pair to be taken into account. The filter parameters of PGIF are calculated as two-dimensional convolution using the bright and spatial differences between the corresponding pixels, which can be incrementally calculated for efficient aggregation. The complexity of the proposed algorithm is O(N), which is linear to the number of image pixels. Furthermore, the algorithm can be further simplified into O(N/4) without significantly sacrificing accuracy if subsampling is applied in the stage of parameter calculation. We also found that a step function to attenuate noise is required in calculating the weights. Experimental evaluation on version 3 of the Middlebury stereo evaluation datasets shows that the proposed method achieves superior disparity accuracy over state-of-the-art aggregation methods with comparable processing speed.

1. Introduction

Stereoscopic vision is less invasive and valuable for many applications, such as 3D reconstruction and environmental detection of autonomous vehicles, as it relies only on pairs of images captured from different perspectives. In stereo vision systems, stereo matching algorithms are critical for correct and accurate disparity estimation, finding corresponding pixels in matching images for each pixel in the reference image.

According to [1], dense stereo matching algorithms fall into two categories: local methods and global methods. The global methods treat disparity calculations as a minimization problem, where the objective function consists of a measurement part and a penalty part. The measurement part indicates the similarity between the slice on the pair of images and the penalty part to suppress the change in disparity. Representative global methods include Belief Propagation [2], Graph Cut [3], and Dynamic Programming [4]. These techniques require a lot of calculations and are not applicable for real-time applications.

In contrast, the local methods are popular for fast disparity calculations. Local approaches tackle the effects of light changes through local windows and are categorized into parametric [5] and nonparametric methods [6]. Local algorithms generally perform four stages [1]: (1) preliminary matching cost calculations, (2) cost aggregation over support regions, (3) disparity estimation, and (4) disparity refinement.

Cost aggregation plays a key role in the local stereo matching algorithms. Early approaches, such as in [7], achieved limited performances, especially in discontinues and occlusion regions. To accommodate the assumption that pixels are of similar disparity within the same aggregation windows [8], many adaptive weighting aggregation methods, such as in [9–11], were presented.

Recent years have seen a trend to treat the cost aggregation as image filtering. The guided image filter [12] (GIF)

is able to provide superior edge profiles without gradient-reversal artifacts, and was successfully applied to cost aggregation in [13]. In order to reduce the computational load of the GIF, [14] recommends subsampling the cost and the guidance image to calculate the coefficients. Ref. [15] proposed the Weighted Guided Image Filtering (WGIF) to improve GIF by modifying the regularization term of the energy function. Besides, [16, 17] applied WGIF for stereo matching with limited performance because of the lack of pixel information outside the fixed windows.

In order to improve the precision of stereo matching, a series of approaches with adaptive guided filters were proposed [18–20] to remove the limitation of the fixed-window formulation. Among them, [20] adaptively tunes the size of rectangular support windows based on both the intensity difference and distance between the neighboring pixel and the central pixel. However, these approaches still suffer from the loss of information outside the windows. Ref. [21] uses whole image for matching cost calculation, where the weights of aggregation are computed according to a measure of similarity between neighboring pixels in the guidance image. In addition, a new scheme called weight propagation is proposed to efficiently calculate weights.

In this work, we extend the scheme of GIF [12] for disparity cost aggregation, as suggested in [13], by introducing bilateral weights that take distance and intensity differences into account, as suggested in [21]. This approach uses the whole image for aggregation. Similar to the convolution procedure of [21], the complexity of the proposed algorithm is O(N), which is linear to the image resolution. We call our approach Pervasive Guided Image Filtering, denoted as PGIF.

The main contribution of this paper is a new cost aggregation algorithm for stereo matching which can be summarized as follows:

(1) An innovative aggregation scheme is proposed that weights the cost function of the GIF to allow for consideration of the entire image pair.

(2) We demonstrate that a constraint modification of the aggregation weight by a step function is crucial to avoid value reduction in the weight propagation process. This minor modification further improves the accuracy of disparity.

(3) The proposed aggregation algorithm can be calculated as two-dimensional convolution with complexity O(N) using the weight-propagation method of [21]. Besides, the algorithm can be further simplified into O(N/4) without sacrificing accuracy if subsampling is applied in calculating the parameters of the guided image filtering, as suggested by [14].

(4) A performance evaluation of version 3 of the Middlebury Stereoscopic Evaluation Dataset demonstrates the superiority of the proposed method to most state-of-the-art aggregation methods in terms of disparity accuracy and processing speed.

2. Aggregations of Preliminary Cost

2.1. Cost Function Definition and GIF Aggregation. In a local stereo matching procedure, a disparity map is obtained through four steps: (1) computation of the preliminary matching cost, (2) aggregation of the cost via volumetric filtering, (3) disparity selection via winner-take-all (WTA), and (4) postprocessing for disparity refinement. In the first step, the preliminary matching cost is a three-dimensional array of dissimilarity measures for every pixel within potential disparities. When a pixel at $p = (x, y)$ is assigned a disparity value d, we use the notation $C_L(p, d)$ as the preliminary cost based on the left image.

There are many metrics that can be used to measure the degree of matching between image patches, such as sum of squared difference and normalized cross correlation. In the following investigations, we use the basic metric, the truncated absolute difference of the gradient, for the cost:

$$C_L(p,d) = \min\left(\left|\nabla I_L(p) - \nabla I_R(p+d)\right|, \tau\right), \quad (1)$$

where I_L and I_R are the left and right images of the stereo pair and τ is a truncation threshold, normally assigned as 2, which is applied to reduce mismatch in noisy or obscured regions. This concise arrangement also reflects the effectiveness of the proposed scheme.

The general form of cost aggregation in the traditional local stereo matching algorithms can be written as a weighted sum of the preliminary matching cost:

$$C'_L(p,d) = \sum_{q\in\Omega(p)} W_{p,q} \cdot C_L(q,d), \quad (2)$$

where $C'_L(p, d)$ is the aggregated cost and q are pixels within a window Ω centered at pixel p.

Equation (2) is a general form of window-based cost aggregation. When the aggregation weight $W_{p,q}$ is set to 1, we have the simplest stereo matching algorithm, such as in [1]. The performance of the window-based approaches depends on the correct choice of the size of the support window, since the effects of external pixels are ignored.

Full Image-Guided Filtering [21] (FIF) is one of the methods using the entire image as a supporting window, which implements a scheme called weight propagation for the cost aggregation:

$$\begin{aligned} C_L^{FIGF}(p,d) &= \sum_{q\in I_L} \prod_{(m,n)\in L_{p,q}} \exp\left(-\frac{\left\|I_L(m) - I_L(n)\right\|}{\sigma}\right) \cdot C_L(q,d) \\ &= \sum_{q\in I_L} W_{p,q}^{FIGF} \cdot C_L(q,d). \end{aligned} \quad (3)$$

In (3), σ is a constant, $L_{p,q}$ is the path between a pixel pair (p, q), and $\|I_L(m) - I_L(n)\|$ is the Euclidian distance between the intensities of two pixels, m and n, along the path. Note that the guided image filtering [12] (GIF) scheme is not employed in this method.

On the other hand, Fast Cost-Volume Filtering method [13] (FCVF) achieves significant performance compared to most local stereo matching methods. In this method, the

weights are not explicitly calculated, as shown in (2). Instead, based on the principle of GIF [12], the filtered cost assumes a local linear model in terms of the guidance image I_L such that

$$C_L^{FCVF}(p,d) = a(p,d)\cdot I_L(q) + b(p,d), \quad \forall q \in \Omega(p), \tag{4}$$

where $a(p,d)$ and $b(p,d)$ are obtained by minimizing an energy function $E(p,d)$ within the supporting window Ω. The energy function for the parameters, a and b, is defined as

$$\begin{aligned} E(p,d) &= \sum_{q\in\Omega(p)} \Big\{ \left[a(p,d)\cdot I_L(q) + b(p,d) - C_L(q,d)\right]^2 \\ &\quad + \varepsilon \cdot \left[a(p,d)\right]^2 \Big\}, \end{aligned} \tag{5}$$

where ε is the normalization parameter used to limit the magnitude of $a(p,d)$, usually chosen to be 0.0001. The optimal values of the linear parameters in (5), denoted as $a^*(p,d)$ and $b^*(p,d)$, are

$$\begin{aligned} a^*(p,d) &= \frac{\left(\sum_{q\in\Omega(p)} I_L(q)\cdot C_L(q,d)\right)/|\Omega(p)| - \left(\sum_{q\in\Omega(p)} I_L(q)/|\Omega(p)|\right)\cdot\left(\sum_{q\in\Omega(p)} C_L(q,d)/|\Omega(p)|\right)}{\sum_{q\in\Omega(p)}\left[I_L(q)\right]^2/|\Omega(p)| - \left(\sum_{q\in\Omega(p)} I_L(q)/|\Omega(p)|\right)\cdot\left(\sum_{q\in\Omega(p)} I_L(q)/|\Omega(p)|\right) + \varepsilon}, \\ b^*(p,d) &= \frac{\sum_{q\in\Omega(p)} C_L(q,d)}{|\Omega(p)|} - a^*(p,d)\cdot\frac{\sum_{q\in\Omega(p)} I_L(q)}{|\Omega(p)|} \end{aligned} \tag{6}$$

where $|\Omega(p)|$ is the number of pixels in the support window Ω centered on p. This GIF-based aggregation can also be rearranged into the form of (2) by replacing the parameters of (4) with the best values of (6):

$$\begin{aligned} C_L^{FCVF}(p,d) &= \sum_{q\in\Omega(p)}\sum_{k\in\Omega(p)\cap\Omega(q)} \frac{1}{|\Omega(p)|\cdot|\Omega(k)|} \\ &\quad\cdot\left(1 + \frac{\left(I_L(p) - \mu(I_L(k))\right)\cdot\left(I_L(q) - \mu(I_L(k))\right)}{\sigma^2(I_L(k)) + \varepsilon}\right) \\ &\quad\cdot C_L(q,d) = \sum_{q\in\Omega(p)} W_{p,q}^{FCVF}\cdot C_L(q,d), \end{aligned} \tag{7}$$

where $\mu(I_L(k))$ is the mean values of I_L and $\sigma^2(I_L(k))$ is the variance of I_L :

$$\begin{aligned} \mu(I_L(k)) &= \frac{\sum_{s\in\Omega(k)} I_L(s)}{|\Omega(k)|} \\ \sigma^2(I_L(k)) &= \frac{\sum_{s\in\Omega(k)}\left[I_L(s)\right]^2}{|\Omega(k)|} - \frac{\sum_{s\in\Omega(k)} I_L(s)}{|\Omega(k)|} \\ &\quad\cdot\frac{\sum_{s\in\Omega(k)} I_L(s)}{|\Omega(k)|}. \end{aligned} \tag{8}$$

2.2. The Proposed Aggregation Scheme: Pervasive Guided Image Filtering (PGIF). According to the investigation of the last section, we have that a proper image filtering algorithm is critical for the accuracy of stereo matching. However, most GIF-based local stereo matching methods suffer from the same problem of missing information outside the supporting window.

In this section, we propose a GIF-based scheme that uses the full image for the filtering and call it the Pervasive Guided Image Filtering (PGIF). Similar to (4), the scheme also approximates the filtered cost as a linear model in terms of the guidance image I_L. However, the corresponding energy function for the linear parameters is defined as a weighted version of (5):

$$\begin{aligned} E(p,d) &= \sum_{q\in I_L} \omega^{PGIF}(p,q) \\ &\quad\cdot\left\{\left[a(p,d)\cdot I_L(q) + b(p,d) - C_L(q,d)\right]^2 + \varepsilon\right. \\ &\quad\left.\cdot\left[a(p,d)\right]^2\right\}, \end{aligned} \tag{9}$$

where ε is a regularization parameter to restrict the value of $a(p,d)$ and the weight $\omega^{PGIF}(p,q)$ reflects the relative importance of the pixel $q = (i,j)$ with respect to the pixel $p = (x,y)$. Importantly, the weight is calculated as a multiplication of horizontal and vertical weighting factors, $\overline{W}_{i,x}^H$ and $\overline{W}_{j,y}^V$:

$$\omega^{PGIF}(p,q) = \omega^{PGIF}(x,y,i,j) = \overline{W}_{i,x}^H\cdot\overline{W}_{j,y}^V, \tag{10}$$

where

$$\begin{aligned} &\overline{W}_{i,x}^H \\ &= \begin{cases} \prod_{k=\min(i,x)+1}^{\max(i,x)} \exp\left(\dfrac{f\left(I_L(k,j) - I_L(k-1,j)\right)}{-\beta}\right), & i \neq x \\ 1, & i = x \end{cases} \\ &\overline{W}_{j,y}^V \\ &= \begin{cases} \prod_{k=\min(j,y)+1}^{\max(j,y)} \exp\left(\dfrac{f\left(I_L(x,k) - I_L(x,k-1)\right)}{-\beta}\right), & j \neq y \\ 1, & j = y \end{cases} \end{aligned} \tag{11}$$

and

$$f(z) = \begin{cases} 0, & \text{if } |z| < 1 \\ 1, & \text{if } |z| \geq 1. \end{cases} \tag{12}$$

The parameter β is a constant filter factor, where the larger value corresponds to a higher weight. As $0 < \overline{W}^H_{i,x}, \overline{W}^V_{j,y} \leq 1$, successive multiplication of them ensures smaller weights for longer distances. Furthermore, the step function $f(z)$ defined in (11) is introduced to avoid the loss of information when there is a significant local density difference along the path from pixel q to pixel p. By this arrangement, the values of the weightings stay within the interval $[\exp(-1/\beta), 1]$, which are proportional to density similarity.

The necessity of introducing the function $f(z)$ is illustrated in Figure 1, where all pixel values are 10, except p_3, which is 99. Assuming $\beta = 4$, we have that $\omega^{PGIF}(p_7, p_1) = \omega^{PGIF}(p_{10}, p_1) = 0.3678$. However, if $f(z)$ is not applied, we have that $\omega^{PGIF}(p_7, p_1) = \omega^{PGIF}(p_{10}, p_1) = 1.624 \times 10^{-4}$, making the intensity of p_1 invisible to both p_7 and p_{10}.

Optimal values of the linear parameters of the filter, $a(p,d)$ and $b(p,d)$, are obtained by minimizing the energy function $E(p,d)$. This can be achieved by assigning the first derivatives of it to zero:

$$\frac{\partial E(p,d)}{\partial a(p,d)} = 0$$
$$\frac{\partial E(p,d)}{\partial b(p,d)} = 0. \tag{13}$$

Solving (13), the best estimate of the parameters is

$$a^*(p,d) = \frac{\overline{\mu}(I_L(p) \cdot C_L(p,d)) - \overline{\mu}(I_L(p)) \cdot \overline{\mu}(C_L(p,d))}{\overline{\sigma}(I_L(p)) + \varepsilon} \tag{14}$$
$$b^*(p,d) = \overline{\mu}(C_L(p,d)) - a^*(p,d) \cdot \overline{\mu}(I_L(p)),$$

where

$$\overline{\mu}(I_L(p) \cdot C_L(p,d)) = \frac{\sum_{q \in I_L} \left[\omega^{PGIF}(p,q) \cdot I_L(p) \cdot C_L(p,d)\right]}{\sum_{q \in I_L} \omega^{PGIF}(p,q)}$$
$$\overline{\mu}(C_L(p,d)) = \frac{\sum_{q \in I_L} \left[\omega^{PGIF}(p,q) \cdot C_L(q,d)\right]}{\sum_{q \in I_L} \omega^{PGIF}(p,q)}$$
$$\overline{\mu}(I_L(p)) = \frac{\sum_{q \in I_L} \left[\omega^{PGIF}(p,q) \cdot I_L(q)\right]}{\sum_{q \in I_L} \omega^{PGIF}(p,q)} \tag{15}$$
$$\overline{\sigma}(I_L(p)) = \frac{\sum_{q \in I_L} \left[\omega^{PGIF}(p,q) \cdot I_L(q) \cdot I_L(q)\right]}{\sum_{q \in I_L} \omega^{PGIF}(p,q)} - \left[\overline{\mu}(I_L(p))\right]^2.$$

To reduce the computational complexity of (15), these equations can be decomposed into four one-dimensional convolutions using the principle of weight propagation, as developed in [21]. Let us take the numerator of $\overline{\mu}(I_L(p) \cdot C_L(p,d))$, denoted as $Z(p,d)$, as an example:

$$Z(p,d) = \sum_{q \in I_L} \left[\omega^{PGIF}(p,q) \cdot I_L(p) \cdot C_L(p,d)\right] = \sum_{i=1}^{s_1} \sum_{j=1}^{s_2} \overline{W}^H_{i,x} \cdot \overline{W}^V_{j,y} \cdot \left[I_L(i,j) \cdot C_L(i,j,d)\right], \tag{16}$$

where the weight ω^{PGIF} is calculated according to (10) and s_1 and s_2 are the width and height of the guidance image, respectively. Firstly, we define the left-to-right and the right-to-left weighted sums, $M^L(x,j,d)$ and $M^R(x,j,d)$, as

$$M^L(x,j,d) = \sum_{i=1}^{x-1} \overline{W}^H_{i,x} \cdot \left[I_L(i,j) \cdot C_L(i,j,d)\right] + I_L(x,j) \cdot C_L(x,j,d)$$
$$M^R(x,j,d) = \sum_{i=x+1}^{s_1} \overline{W}^H_{i,x} \cdot \left[I_L(i,j) \cdot C_L(i,j,d)\right] + I_L(x,j) \cdot C_L(x,j,d). \tag{17}$$

Since $\overline{W}^H_{i,x} = \overline{W}^H_{i,x-1} \cdot \overline{W}^H_{x-1,x}$, $M^L(x,j,d)$ can be written in recursive form:

$$\begin{aligned} M^L(x,j,d) &= \sum_{i=1}^{x-2} \overline{W}^H_{i,x} \cdot \left[I_L(i,j) \cdot C_L(i,j,d)\right] \\ &+ \overline{W}^H_{x-1,x} \cdot \left[I_L(x-1,j) \cdot C_L(x-1,j,d)\right] \\ &+ I_L(x,j) \cdot C_L(x,j,d) = \overline{W}^H_{x-1,x} \\ &\cdot \left\{\sum_{i=1}^{x-2} \overline{W}^H_{i,x-1} \cdot \left[I_L(i,j) \cdot C_L(i,j,d)\right] + I_L(x-1,j) \right. \\ &\left. \cdot C_L(x-1,j,d)\right\} + I_L(x,j) \cdot C_L(x,j,d) \\ &= \overline{W}^H_{x-1,x} \cdot M^L(x-1,j,d) + I_L(x,j) \cdot C_L(x,j,d). \end{aligned} \tag{18}$$

Similarly,

$$M^R(x,j,d) = \overline{W}^H_{x,x+1} \cdot M^R(x+1,j,d) + I_L(x,j) \cdot C_L(x,j,d). \tag{19}$$

Secondly, we define the horizontal weighted sum, denoted as $M^H(x,j,d)$:

$$M^H(x,j,d) = M^L(x,j,d) + M^R(x,j,d) - I_L(x,j) \cdot C_L(x,j,d). \tag{20}$$

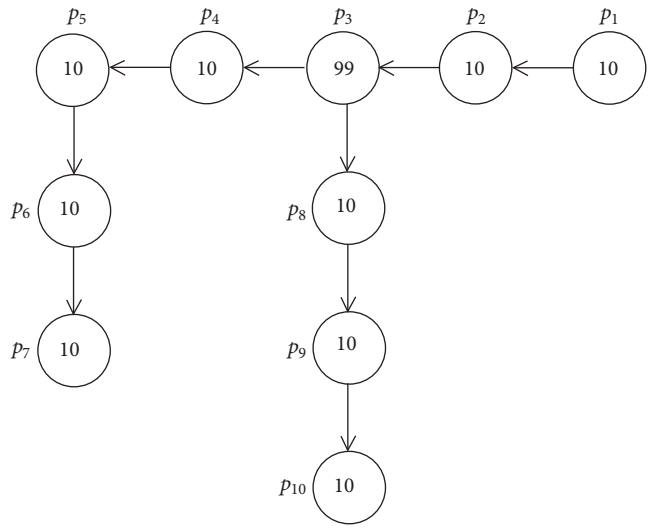


Figure 1: A case indicating the necessity of introducing a step function to avoid the effects of sudden changes in intensity.

Based on a similar procedure, we define the top-to-bottom and bottom-to-top weighted sums, $M^T(x, y, d)$ and $M^B(x, y, d)$, as

$$\begin{aligned} M^T(x, y, d) &= \overline{W}^V_{y-1,y} \cdot M^T(x, y-1, d) + I_L(x, y) \\ &\quad \cdot C_L(x, y, d) \\ &= \sum_{j=1}^{y-1} \overline{W}^V_{j,y} \cdot M^H(x, j, d) + I_L(x, y) \\ &\quad \cdot C_L(x, y, d) \end{aligned} \tag{21}$$

$$\begin{aligned} M^B(x, y, d) &= \overline{W}^V_{y,y+1} \cdot M^B(x, y+1, d) + I_L(x, y) \\ &\quad \cdot C_L(x, y, d) \\ &= \sum_{j=y+1}^{s_2} \overline{W}^V_{j,y} \cdot M^H(x, j, d) + I_L(x, y) \\ &\quad \cdot C_L(x, y, d). \end{aligned} \tag{22}$$

Hence, $Z(p, d)$ can be calculated as

$$\begin{aligned} Z(p, d) &= Z(x, y, d) \\ &= M^T(x, y, d) + M^B(x, y, d) - I_L(x, y) \\ &\quad \cdot C_L(x, y, d). \end{aligned} \tag{23}$$

In sum, we can compute $Z(p, d)$ using (23), where $M^T(x, y, d)$ and $M^B(x, y, d)$ are obtained by (21) and (22). Next, $M^H(x, j, d)$ in (21) and (22) are calculated from (20). Finally, $M^L(x, j, d)$ and $M^R(x, j, d)$ of (20) are recursively calculated using (18) and (19).

We have that all of the two-dimensional convolutions in (15) can be decomposed into one-dimensional convolution operations in four directions with computational complexity O(N), where N is the total number of pixels of the guidance image.

Once we have the optimal filter parameters $a^*(p, d)$ and $b^*(p, d)$, the aggregated matching cost volume is calculated as

$$C_L^{PGIF}(p, d) = a^*(p, d) \cdot I_L(p) + b^*(p, d). \tag{24}$$

We may substitute the parameters of (24) by the optimal values of (14) and rearrange them into the following form:

$$\begin{aligned} C_L^{PGIF}(p, d) &= \sum_{q \in I_L} \frac{\omega^{PGIF}(p, q)}{\sum_{q \in I_L} \omega^{PGIF}(p, q)} \Bigg(1 \\ &\quad + \frac{(I_L(p) - \overline{\mu}(I_L(p))) \cdot (I_L(q) - \overline{\mu}(I_L(p)))}{\overline{\sigma}(I_L(p)) + \varepsilon} \Bigg) \\ &\quad \cdot C_L(q, d) = \sum_{q \in I_L} W^{PGIF}_{p,q} \cdot C_L(q, d). \end{aligned} \tag{25}$$

The disparity map, denoted as $D_L(p)$, can then be obtained by the winner-take-all operation on the aggregated cost volume $C_L^{PGIF}(p, d)$:

$$D_L(p) = \arg\min_{d \in R} C_L^{PGIF}(p, d), \tag{26}$$

where $R = \{d_{\min}, \ldots, d_{\max}\}$ is the range of disparity values.

A close comparison between (7) and (25) indicates that the effective region of (25) extends from the local aggregation window Ω of (7) to the entire guided image I_L, where the weight of (25) includes the weight ω^{PGIF} of the energy function defined in (9).

In addition, in order to speed up the calculation, [14] suggests subsampling both the cost volume and the guidance image to calculate the linear coefficients. We can follow the same process by subsampling I_L and C_L to calculate the optimal filter parameters and then use them for normal aggregation. This will accelerate the algorithm to O(N/4).

3. Experimental Study

Extensive experiments were conducted to verify the effectiveness of the proposed scheme in calculating disparity maps. We studied five representative stereo matching algorithms and two versions of the proposed scheme:

(i) The Fast Cost Volume Filtering, denoted as FCVF [13]

(ii) The Full-Image Guided Filtering for Fast Stereo Matching, denoted as FIF [21]

(iii) The Fast Guided Image Filtering, denoted as FGIF [14]

(iv) The Adaptive Guided Image Filter, denoted as AGIF [20]

(v) The Weighted Guided Image Filtering, denoted as WGIF [17]

(vi) The proposed scheme, which is of computational complexity O(N), denoted as PGIF(N)

(vii) The proposed scheme with downsampling in calculating the linear parameters, which is of computational complexity O($N/4$), denoted as PGIF($N/4$).

It is worth noting that there is only one design parameter, β of (11), in our approach. For all of these computational experiments, we fixed its value as 4. Besides, the experiments are performed in MATLAB 2017b using Intel Core I7 3610 with 16 GB RAM.

We tested these frameworks using the Middlebury (version 3) benchmark stereo database [22], which has a collection of stereo pairs for matching performance evaluation. Among them, 15 pairs were used, from "Adirondack" to "Vintage", in the "trainingQ" image set. The Middlebury defines two measures for evaluation, including nonoccluded (non-occ), and all regions. The "error rates" to be presented below are denoted as "bad 1.0" in the benchmark, which is the percentage of "bad" pixels with a disparity error greater than 1.0. Furthermore, the "weighted average error" is calculated using different weights for different image pairs such that the image pairs "PianoL", "Playroom", "Playtable", "Shelves", and "Vintage" contribute only half of the error values. This arrangement of weights is applied, according to the remarks on the benchmark website, "to compensate for the varying difficulty of the different datasets."

Figure 2 shows the disparity maps obtained by these algorithms without disparity refinement. The corresponding error rates in the nonoccluded region and the all-region are shown in Tables 1 and 2, respectively. The erroneous pixels in the nonoccluded area are marked in red, while those in the occluded area are marked in green.

According to the results, FIF [21] performs best for Recycle and Shelves, and AGIF [20] performs best for PlaytableP in the nonoccluded (non-occ) region. Besides, FIF [21] performs best for Recycle and Shelves, and AGIF [20] performs best for MotorcycleE, Pipes, and PlaytableP in the all-region. However, PGIF(N) performs best in all other situations, including both the nonoccluded region and the all-region. It is clear that PGIF(N) has the best ranking for the weighted average error performance, which is 14.66% for the nonoccluded region and 20.49% for the all-region, while PGIF($N/4$) is just second to it, which is 15.10% for the nonoccluded region and 20.74% for the all-region, respectively.

The times required to run these stereo matching algorithms are shown in Table 3. As expected, PGIF($N/4$) only needs to calculate a quarter of the time required for PGIF(N). Furthermore, if we check the time required to run the stereo matching algorithm, we have that PGIF(N) and FIF [21] belong to the same level, while PGIF($N/4$) and FGIF [14] belong to the same level. We may conclude that PGIF($N/4$) is the best scheme when considering both the correctness of the matching and the efficiency of the calculation.

Similar to the proposed scheme, FIF [21] also uses the entire image for aggregation. However, in addition to the difference between the methods in cost-volume filtering, a step function $f(z)$, defined in (12), is introduced in the proposed scheme to reduce the effects of noise. It is important to clarify whether the performance improvement of the proposed solution is mainly due to this function. Therefore, extensive experiments were conducted on this clarification and are summarized in Figure 3, in which the same color convention as used in Figure 2 is applied. Based on these results, we have that using the step function does improve the performance of FIF [21]. Importantly, our method is still better than the modified FIF [21].

In addition, there are several parameters that may affect the performance of the algorithms under study, including the parameter r of FCVF [13], the parameter σ of FIF [21], and the parameter β of the PGIF(N). With a close look at the filter kernel of FCVF [13], included in (7), we have that the resulting disparity map is smoother with the increase of the parameter r. The parameter σ in the filter kernel of FIF [21], presented in (3), and the parameter β of the proposed scheme both have the similar behaviors.

We studied their effects on error rates through extended experimental calculations and the results are summarized in Figure 4. Based on the results, we have that the best values of these parameters are $r = 5$, $\sigma = 0.11$, and $\beta = 4$. These values were applied to the experimental calculations presented above.

4. Conclusions

Inspired by FIF [21] and FCVF [13], we propose a new GIF-based aggregation scheme for the calculation of dense disparity maps. The scheme uses the entire image for cost calculation and exploits the weight propagation [21] method for efficient computation.

We redesign the energy function of the guided image filter [12] so that we can not only preserve the original weight propagation structure [21] but also embed the effects of distance and intensity differences in the weights. This arrangement allows efficient use of the entire image when the GIF [12] scheme is used for cost aggregation.

Table 1: Comparison of the error rates in the non-occluded (non-occ) region without refinements using different matching algorithms (%).

Image Sets	FCVF[13]	FIF[21]	FGIF[14]	AGIF[20]	WGIF[17]	PGIF(N)	PGIF(N/4)
Adirondack	8.78	8.29	9.19	7.81	7.87	**6.43**	6.51
ArtL	12.22	12.12	13.21	11.75	11.89	**10.90**	11.07
Jadeplant	21.81	24.18	23.12	20.87	21.45	**20.03**	20.36
Motorcycle	9.87	11.96	10.49	9.70	9.73	**9.01**	9.54
MotorcycleE	9.72	11.80	10.03	9.27	9.62	**8.93**	9.00
Piano	16.20	15.09	15.96	15.17	15.60	**14.01**	14.27
PianoL	33.44	30.81	32.78	32.84	32.99	**30.25**	30.76
Pipes	10.45	11.40	10.74	10.05	10.16	**9.96**	10.11
Playroom	22.68	19.10	23.48	20.99	21.39	**16.62**	18.63
Playtable	41.89	41.26	40.96	41.03	41.02	**39.29**	39.47
PlaytableP	13.81	16.42	14.18	**12.93**	13.06	13.08	13.20
Recycle	10.52	**8.26**	10.72	10.13	10.39	8.38	9.03
Shelves	39.52	**32.48**	39.60	39.32	39.76	36.11	38.02
Teddy	7.18	6.80	7.30	7.00	7.09	**6.33**	6.49
Vintage	33.22	37.59	35.74	32.53	32.55	**30.05**	31.57
Weighted Average	16.47	16.56	16.90	15.84	16.06	**14.66**	15.10

Table 2: Comparison of the error rates in the all-region without refinement using different matching algorithms (%). The best score for the same image set is marked in bold.

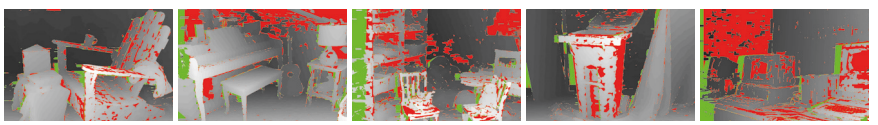
Image Sets	FCVF[13]	FIF[21]	FGIF[14]	AGIF[20]	WGIF[17]	PGIF(N)	PGIF(N/4)
Adirondack	10.09	10.22	10.53	9.57	9.68	8.39	**8.38**
ArtL	21.90	21.96	22.93	21.81	22.02	21.64	**21.22**
Jadeplant	34.51	36.88	35.60	34.09	34.46	**33.45**	33.49
Motorcycle	13.50	15.80	14.16	13.59	13.64	**13.23**	13.47
MotorcycleE	13.68	15.63	13.78	**13.24**	13.55	13.25	13.09
Piano	19.99	19.49	19.81	19.34	19.45	**18.29**	18.38
PianoL	36.43	34.44	35.87	36.04	36.14	**33.91**	34.22
Pipes	21.45	22.66	21.63	**21.17**	21.30	21.52	21.44
Playroom	30.61	27.78	31.45	29.07	29.34	**25.38**	27.18
Playtable	44.65	44.13	43.86	43.98	44.00	**42.63**	42.65
PlaytableP	17.29	19.81	18.35	**15.92**	16.39	17.55	17.37
Recycle	12.30	**10.22**	12.50	12.42	12.72	10.50	11.21
Shelves	40.14	**33.62**	40.29	39.85	40.27	36.89	38.66
Teddy	12.59	12.44	12.72	12.48	12.57	**11.94**	12.09
Vintage	37.18	40.99	39.55	36.38	36.39	**34.01**	35.41
Weighted Average	21.74	22.05	22.20	21.30	21.51	**20.49**	20.74

Table 3: Comparison of the time required to run the stereo matching algorithms (s). The smallest value for the same image set is marked in bold.

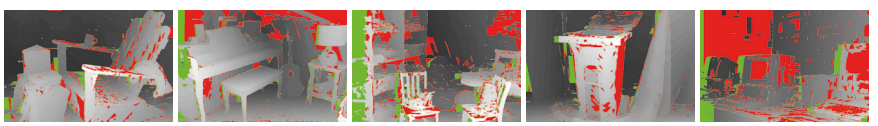
Methods	Adirondack	Piano	Playroom	Recycle	Vintage
FCVF[13]	16	12	17	13	37
FIF[21]	31	24	33	26	74
FGIF[14]	**4**	**4**	**5**	**4**	**10**
AGIF[20]	55	48	57	49	106
WGIF[17]	18	14	20	16	41
PGIF(N)	33	26	35	27	77
PGIF(N/4)	8	6	9	7	17

(a) Color images selected from the benchmark: Adirondack, Piano, Playroom, Recycle, and Vintage

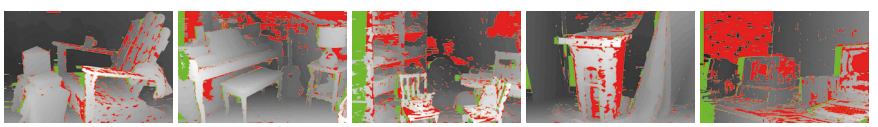
(b) Ground truth disparity maps



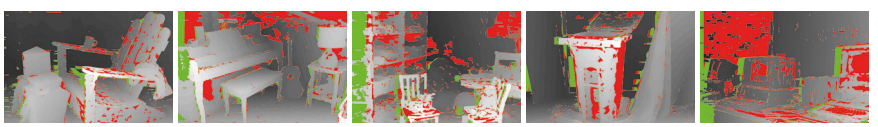
(c) Disparity maps generated by FCVF [13]



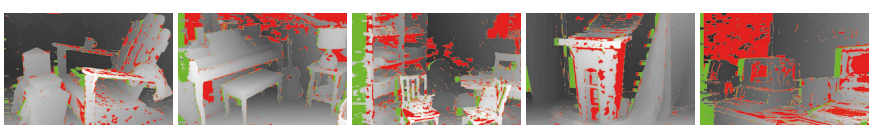
(d) Disparity maps generated by FIF [21]



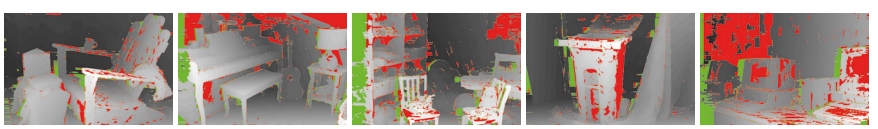
(e) Disparity maps generated by FGIF [14]



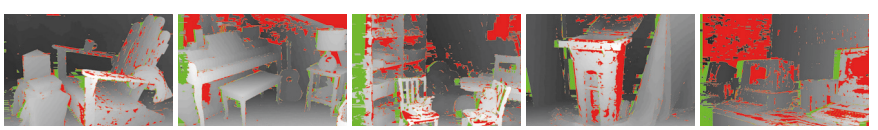
(f) Disparity maps generated by AGIF [20]



(g) Disparity maps generated by WGIF [17]



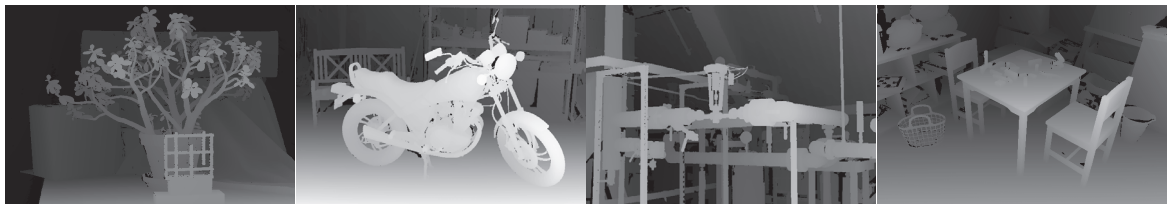
(h) Disparity maps generated by PGIF(N)



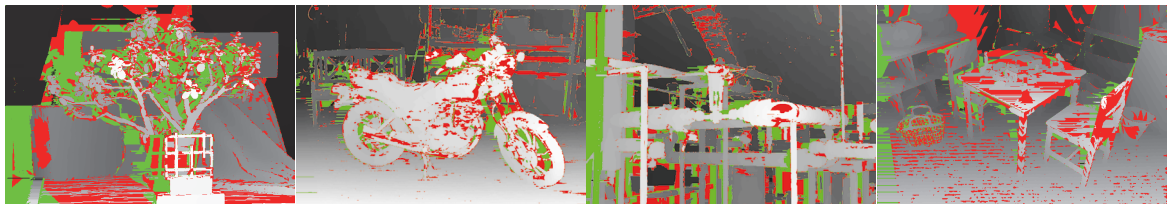
(i) Disparity maps generated by PGIF($N/4$)

Figure 2: Comparison of disparity maps obtained by different algorithms without refinement. The erroneous pixels in the nonoccluded area are marked in red, while those in the occluded area are marked in green.

(a) Color images selected from the benchmark (from left to right): Jadeplant, Motorcycle, Pipes, and PlaytableP

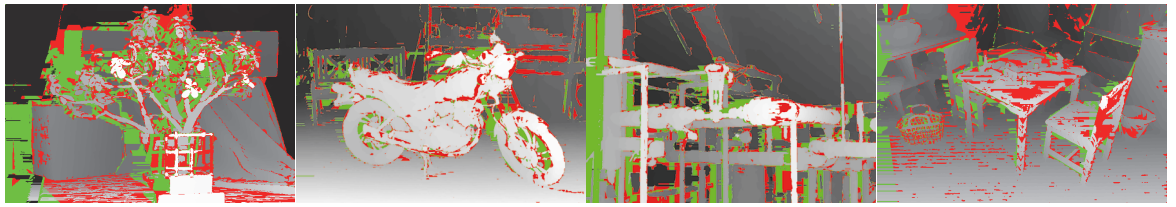


(b) Ground truth disparity maps



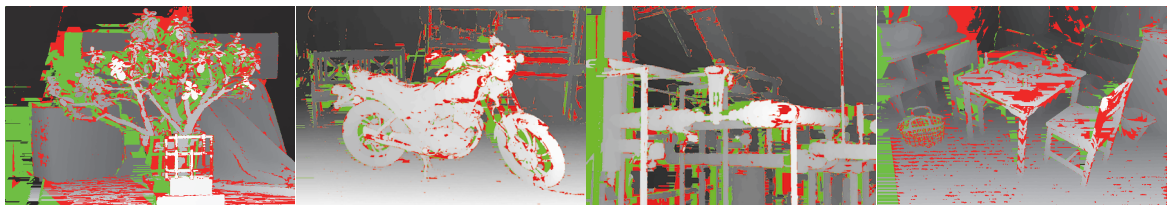
36.88 % 15.80 % 22.66 % 19.81 %

(c) Error rates in the all-region using the original FIF [21]



34.99 % 13.61 % 22.37 % 17.61 %

(d) Error rates in the all-region with the step function $f(z)$ in applying FIF [21]



33.45 % 13.23 % 21.52 % 17.55 %

(e) Error rates in the all-region using PGIF(N)

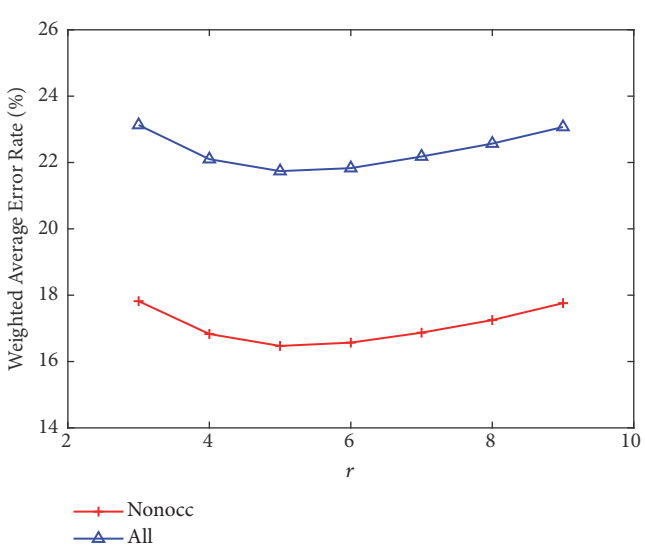
FIGURE 3: Performance comparison of error rates in the all-region of disparity maps using different schemes. These experiments were performed to verify the importance of using a step function (the same color convention as in Figure 2 is applied here).

Besides, there is only one design parameter of the proposed scheme. We also suggest the introduction of a step function in the calculation of weights to attenuate noise.

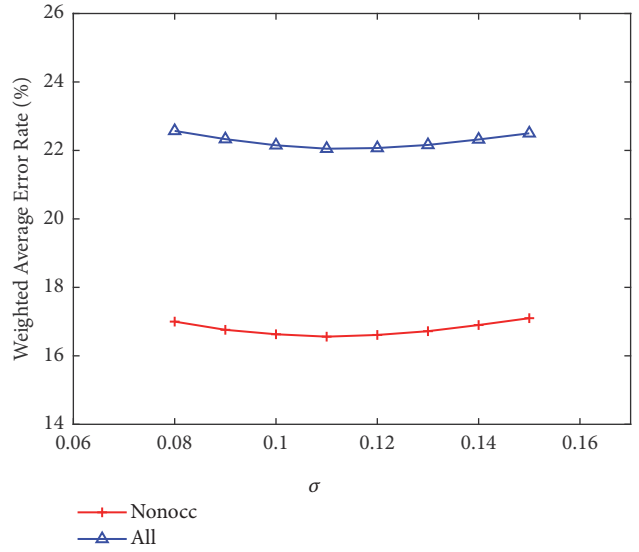
A performance evaluation using the Middlebury (version 3) benchmark stereo database [22] shows that the proposed solution provides superior disparity accuracy and comparable processing speed compared to representative aggregation methods. The experimental results also justify the use of a step function when calculating the aggregate weights.

Data Availability

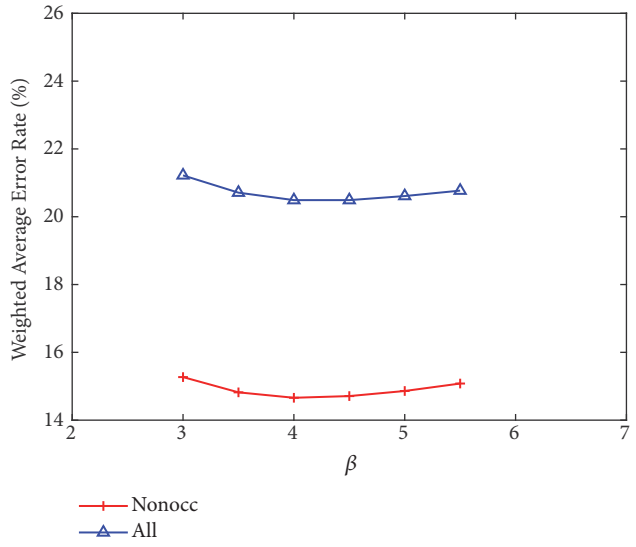
The dataset used to support the findings of this study is included in the article, which is cited at relevant places within the text as [22].



(a) The effect of parameter r on the weighted average error rate of FCVF [13]



(b) The effect of parameter σ on the weighted average error rate of FIF [21]



(c) The effect of parameter β on the weighted average error rate of the PGIF(N)

Figure 4: Effects of parameter values on the weighted average error rate.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research is supported by the National Natural Science Foundation of China, under Grant no. 61471263; the Natural Science Foundation of Tianjin, China, under Grant no. 16JCZDJC31100; the Ministry of Science and Technology, ROC, under Grant nos. MOST 106-2221-E-182-033 and 107-2221-E-182-078; and Chang Gung Memorial Hospital, Taiwan, under Grant no. CORPD2H0011.

References

[1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, no. 1, pp. 7–42, 2002.

[2] J. Sun, N. Zheng, and H. Shum, "Stereo matching using belief propagation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 7, pp. 787–800, 2003.

[3] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, 2001.

[4] L. Wang, R. Yang, M. Gong, and M. Liao, "Real-time stereo using approximated joint bilateral filtering and dynamic programming," *Journal of Real-Time Image Processing*, vol. 9, no. 3, pp. 447–461, 2014.

[5] H. Hirschmüller and D. Scharstein, "Evaluation of stereo matching costs on images with radiometric differences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 9, pp. 1582–1599, 2009.

[6] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *Computer Vision, ECCV94*, pp. 151–158, Springer, Berlin, Germany, 1994.

[7] O. Veksler, "Fast variable window for stereo correspondence using integral images," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 556–561, Madison, WI, USA, 2003.

[8] K.-J. Yoon and I. S. Kweon, "Adaptive support-weight approach for correspondence search," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 650–656, 2006.

[9] A. Hosni, M. Bleyer, M. Gelautz, and C. Rhemann, "Local stereo matching using geodesic support weights," in *Proceedings of the 16th IEEE International Conference on Image Processing, ICIP 2009*, pp. 2093–2096, Cairo, Egypt, November 2009.

[10] C. Stentoumis, L. Grammatikopoulos, I. Kalisperakis, and G. Karras, "On accurate dense stereo-matching using a local adaptive multi-cost approach," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 91, no. 91, pp. 29–49, 2014.

[11] X. Sun, X. Mei, S. Jiao, M. Zhou, Z. Liu, and H. Wang, "Real-time local stereo via edge-aware disparity propagation," *Pattern Recognition Letters*, vol. 49, pp. 201–206, 2014.

[12] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1397–1409, 2013.

[13] A. Hosni, C. Rhemann, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 2, pp. 504–511, 2013.

[14] G. Hong, J. Park, and B. Kim, "Near real-time local stereo matching algorithm based on fast guided image filtering," in *Proceedings of the 2016 6th European Workshop on Visual Information Processing (EUVIP)*, pp. 1–5, Marseille, France, October 2016.

[15] Z. Li, J. Zheng, Z. Zhu, W. Yao, and S. Wu, "Weighted guided image filtering," *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 120–129, 2014.

[16] G.-S. Hong, M.-S. Koo, A. Saha, and B.-G. Kim, "Efficient local stereo matching technique using weighted guided image filtering (WGIF)," in *Proceedings of the IEEE International Conference on Consumer Electronics, ICCE 2016*, pp. 484-485, Las Vegas, NV, USA, January 2016.

[17] G.-S. Hong and B.-G. Kim, "A local stereo matching algorithm based on weighted guided image filtering for improving the generation of depth range images," *Displays*, vol. 49, pp. 80–87, 2017.

[18] Q. Yang, P. Ji, D. Li, S. Yao, and M. Zhang, "Fast stereo matching using adaptive guided filtering," *Image and Vision Computing*, vol. 32, no. 3, pp. 202–211, 2014.

[19] S. Zhu, Z. Wang, X. Zhang, and Y. Li, "Edge-preserving guided filtering based cost aggregation for stereo matching," *Journal of Visual Communication and Image Representation*, vol. 39, pp. 107–119, 2016.

[20] S. Zhu and L. Yan, "Local stereo matching algorithm with efficient matching cost and adaptive guided image filter," *The Visual Computer*, vol. 33, no. 9, pp. 1087–1102, 2017.

[21] Q. Yang, D. Li, L. Wang, and M. Zhang, "Full-image guided filtering for fast stereo matching," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 237–240, 2013.

[22] D. Scharstein, R. Szeliski, and H. Hirschmüller, "Middlebury stereo vision," http://vision.middlebury.edu/stereo/, 2015.