Research Article

Efficient Stereo Matching Based on Pervasive Guided Image Filtering

Chengtao Zhu and Yau-Zen Chang

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

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:

$$C_L(p, d) = \min \{|V_L(p) - V_R(p + d)|, \tau\},$$

where $L_I$ and $R_I$ are the left and right images of the stereo pair and $\tau$ 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:

$$C'_L(p, d) = \sum_{q \in I_R(p)} W_{pq} \cdot C_L(q, d),$$

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

Equation (2) is a general form of window-based cost aggregation. When the aggregation weight $W_{pq}$ 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:

$$C^\text{FIF}(p, d) = \sum_{q \in I_L(p)} \prod_{L \in L_p} \exp\left(-\frac{|L_L(m) - L_L(n)|}{\sigma}\right) \cdot C_L(q, d),$$

In (3), $\sigma$ is a constant, $L_{pq}$ is the path between a pixel pair $(p, q)$, and $|L_L(m) - L_L(n)|$ 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
\[
C_{\text{FCVF}}^L(p, d) = a(p, d) \cdot I_L(q) + b(p, d),
\]
\( \forall q \in \Omega(p), \) (4)
where \( a(p, d) \) and \( b(p, d) \) are obtained by minimizing an energy function \( E(p, d) \) within the supporting window \( \Omega \). The energy function for the parameters, \( a \) and \( b \), is defined as
\[
E(p, d) = \sum_{q \in \Omega(p)} \{ [a(p, d) \cdot I_L(q) + b(p, d) - C_L(q, d)]^2 + \epsilon \cdot [a(p, d)]^2 \},
\]
where \( \epsilon \) 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
\[
a^*(p, d) = \frac{\sum_{q \in \Omega(p)} I_L(q) \cdot C_L(q, d) / |\Omega(p)| - (\sum_{q \in \Omega(p)} I_L(q) / |\Omega(p)|) \cdot \left( \sum_{q \in \Omega(p)} C_L(q, d) / |\Omega(p)| \right)}{\sum_{q \in \Omega(p)} |I_L(q)| / |\Omega(p)|}
\]
\[
b^*(p, d) = \frac{\sum_{q \in \Omega(p)} I_L(q) / |\Omega(p)|}{\sum_{q \in \Omega(p)} c_{\text{FCVF}, p} \cdot C_L(q, d)},
\]
where \( |\Omega(p)| \) is the number of pixels in the support window \( \Omega \) 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):
\[
C_{\text{FCVF}}^L(p, d) = \frac{\sum_{q \in \Omega(p)} \sum_{k \in \Omega(p) \cap \Omega(q)} \frac{1}{|\Omega(p)|} \cdot \left( 1 + \frac{L_L(p) - \mu(L_L(k)) \cdot (L_L(q) - \mu(L_L(k)))}{\sigma^2(L_L(k))} \right) \cdot C_L(q, d)}{|\Omega(p)|} \cdot \left( \sum_{q \in \Omega(p)} \frac{W_{\text{FCVF}, p} \cdot C_L(q, d)}{|\Omega(p)|} \right)
\]
where \( \mu(L_L(k)) \) is the mean values of \( L_L \) and \( \sigma^2(L_L(k)) \) is the variance of \( L_L \):  
\[
\begin{align*}
\mu(L_L(k)) &= \frac{\sum_{s \in \Omega(k)} L_L(s) / |\Omega(k)|}{|\Omega(k)|} \\
\sigma^2(L_L(k)) &= \frac{\sum_{s \in \Omega(k)} (L_L(s) - \mu(L_L(k)))^2 / |\Omega(k)|}{|\Omega(k)|}
\end{align*}
\]
where \( \omega \) is a regularization parameter to restrict the value of \( a(p, d) \) and \( b(p, d) \) are defined as a weighted version of (5):
\[
E(p, d) = \sum_{q \in \Omega(p)} \omega_{\text{GIF}, p} \cdot \{ [a(p, d) \cdot I_L(q) + b(p, d) - C_L(q, d)]^2 + \epsilon \cdot [a(p, d)]^2 \},
\]
where
\[
\omega_{\text{GIF}, p} = \omega_{\text{GIF}, (x, y, i, j)} = \frac{\omega_{l, x}}{\omega_{l, x}} \cdot \frac{\omega_{y}}{\omega_{y}},
\]
where
\[
\omega_{l, x} = \prod_{k = \min(x) + 1}^{\max(x)} \frac{f(L_L(k, j) - L_L(k - 1, j))^{1-i} - f(L_L(k, j) - L_L(k - 1, j))^{-1}}{\beta},
\]
\( i \neq x \)
\( 1, \)
\( i = x \)
\( \omega_{l, y} = \prod_{k = \min(y) + 1}^{\max(y)} \frac{f(L_L(x, k) - L_L(x, k - 1))^{1-j} - f(L_L(x, k) - L_L(x, k - 1))^{-1}}{\beta}, \)
\( j \neq y \)
\( 1, \)
\( j = y \)

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

The parameter \( \beta \) is a constant filter factor, where the larger value corresponds to a higher weight. As \( 0 < \frac{\omega_{PGIF}(p, r)}{\omega_{PGIF}(p, 0)} < 1 \), 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:
\[
\begin{align*}
\frac{\partial E(p, d)}{\partial a(p, d)} &= 0 \\
\frac{\partial E(p, d)}{\partial b(p, d)} &= 0.
\end{align*}
\] (13)

Solving (13), the best estimate of the parameters is
\[
\begin{align*}
a^*(p, d) &= \frac{\mathcal{P}(I_L(p) \cdot C_L(p, d)) - \mathcal{P}(I_L(p)) \cdot \mathcal{P}(C_L(p, d))}{\mathcal{P}(I_L(p)) + \varepsilon} \\
b^*(p, d) &= \mathcal{P}(C_L(p, d)) - a^*(p, d) \cdot \mathcal{P}(I_L(p)),
\end{align*}
\] (14)

where
\[
\begin{align*}
\mathcal{P}(I_L(p) \cdot C_L(p, d)) &= \frac{\sum_{q \in I_L} [\omega_{PGIF}(p, q) \cdot I_L(q) \cdot C_L(q, d)]}{\sum_{q \in I_L} [\omega_{PGIF}(p, q)]} \\
\mathcal{P}(C_L(p, d)) &= \frac{\sum_{q \in I_L} [\omega_{PGIF}(p, q) \cdot C_L(q, d)]}{\sum_{q \in I_L} [\omega_{PGIF}(p, q)]} \\
\mathcal{P}(I_L(p)) &= \frac{\sum_{q \in I_L} [\omega_{PGIF}(p, q) \cdot I_L(q)]}{\sum_{q \in I_L} [\omega_{PGIF}(p, q)]} \\
\mathcal{P}(C_L(p)) &= \frac{\sum_{q \in I_L} [\omega_{PGIF}(p, q) \cdot C_L(q)]}{\sum_{q \in I_L} [\omega_{PGIF}(p, q)]} \\
\mathcal{P}(I_L(p)) &= \frac{\sum_{q \in I_L} [\omega_{PGIF}(p, q) \cdot I_L(q) \cdot I_L(q) \cdot I_L(q)]}{\sum_{q \in I_L} [\omega_{PGIF}(p, q)]}.
\end{align*}
\] (15)

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 \( \mathcal{P}(I_L(p) \cdot C_L(p, d)) \), denoted as \( Z(p, d) \), as an example:
\[
Z(p, d) = \sum_{q \in I_L} [\omega_{PGIF}(p, q) \cdot I_L(p) \cdot C_L(p, d)]
\] (16)
\[
= \sum_{i=1}^{s_1} \sum_{j=1}^{s_2} W_{lx,i} \cdot W_{ly,j} \cdot [I_L(i, j) \cdot C_L(i, j, d)],
\]
where the weight \( \omega_{PGIF} \) 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
\[
M^L(x, j, d) = \sum_{i=1}^{x-1} W_{lx,i} \cdot [I_L(i, j) \cdot C_L(i, j, d)] + I_L(x, j) \cdot C_L(x, j, d)
\] (17)
\[
M^R(x, j, d) = \sum_{i=x}^{s_2} W_{lx,i} \cdot [I_L(i, j) \cdot C_L(i, j, d)] + I_L(x, j) \cdot C_L(x, j, d).
\]

Since \( W_{lx}^H = W_{lx-1}^H \cdot W_{lx-1,x}^H \), \( M^L(x, j, d) \) can be written in recursive form:
\[
M^L(x, j, d) = \sum_{i=1}^{s_2} W_{lx,i} \cdot [I_L(i, j) \cdot C_L(i, j, d)]
\] (18)
\[
+ \sum_{i=x}^{s_2} W_{lx-1,i} \cdot [I_L(i, j) \cdot C_L(i, j, d)] + I_L(x, j) \cdot C_L(x, j, d)
\]

Similarly,
\[
M^R(x, j, d) = \sum_{i=x}^{s_2} W_{lx,i} \cdot M^R(x + 1, j, d) + I_L(x, j) \cdot C_L(x, j, d).
\] (19)

Secondly, we define the horizontal weighted sum, denoted as \( M^H(x, j, d) \):
\[
M^H(x, j, d) = M^L(x, j, d) + M^R(x, j, d) - I_L(x, j) \cdot C_L(x, j, d).
\] (20)
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

\[
M_T(x, y, d) = \sum_{j=1}^{y-1} W_{j,y} \cdot M_H(x, j, d) + M_T(x, y - 1, d) + I_L(x, y) \cdot C_L(x, y, d)
\]

\[
M_B(x, y, d) = \sum_{j=y+1}^{s} W_{j,y} \cdot M_H(x, j, d) + M_B(x, y + 1, d) + I_L(x, y) \cdot C_L(x, y, d)
\]

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

\[
Z(p, d) = Z(x, y, d)
\]

\[
= M_T(x, y, d) + M_B(x, y, d) - I_L(x, y) \cdot C_L(x, y, d)
\]

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, \( M_T(x, j, d) \) and \( M_B(x, j, d) \) 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

\[
C_{PGIF}^L(p, d) = a^*(p, d) \cdot I_L(p) + b^*(p, d)
\]

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

\[
C_{PGIF}^L(p, d) = \sum_{q \in I_L} \omega_{PGIF}(p, q) \left( 1 + \frac{(I_L(p) - \mu(I_L(p))) \cdot (I_L(q) - \mu(I_L(p)))}{\sigma(I_L(p)) + \epsilon} \right) \cdot C_L(q, d) = \sum_{q \in I_L} W_{p,q} \cdot C_L(q, d).
\]

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

\[
D_L(p) = \arg\min_{d \in R} C_{PGIF}^L(p, d)
\]

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 \( \Omega \) of (7) to the entire guided image \( I_L \), where the weight of (25) includes the weight \( \omega_{PGIF} \) 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, $\beta$ 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 $\sigma$ of FIF [21], and the parameter $\beta$ 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 $\sigma$ in the filter kernel of FIF [21], presented in (3), and the parameter $\beta$ 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 (%).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Adirondack</td>
<td>8.78</td>
<td>8.29</td>
<td>9.19</td>
<td>7.81</td>
<td>7.87</td>
<td><strong>6.43</strong></td>
<td>6.51</td>
</tr>
<tr>
<td>ArtL</td>
<td>12.22</td>
<td>12.12</td>
<td>13.21</td>
<td>11.75</td>
<td>11.89</td>
<td><strong>10.90</strong></td>
<td>11.07</td>
</tr>
<tr>
<td>Jadeplant</td>
<td>21.81</td>
<td>24.18</td>
<td>23.12</td>
<td>20.87</td>
<td>21.45</td>
<td><strong>20.03</strong></td>
<td>20.36</td>
</tr>
<tr>
<td>Motorcycle</td>
<td>9.87</td>
<td>11.96</td>
<td>10.49</td>
<td>9.70</td>
<td>9.73</td>
<td><strong>9.01</strong></td>
<td>9.54</td>
</tr>
<tr>
<td>MotorcycleE</td>
<td>9.72</td>
<td>11.80</td>
<td>10.03</td>
<td>9.27</td>
<td>9.62</td>
<td><strong>8.93</strong></td>
<td>9.00</td>
</tr>
<tr>
<td>Piano</td>
<td>16.20</td>
<td>15.09</td>
<td>15.96</td>
<td>15.17</td>
<td>15.60</td>
<td><strong>14.01</strong></td>
<td>14.27</td>
</tr>
<tr>
<td>PianoL</td>
<td>33.44</td>
<td>30.81</td>
<td>32.78</td>
<td>32.84</td>
<td>32.99</td>
<td><strong>30.25</strong></td>
<td>30.76</td>
</tr>
<tr>
<td>Pipes</td>
<td>10.45</td>
<td>11.40</td>
<td>10.74</td>
<td>10.05</td>
<td>10.16</td>
<td><strong>9.96</strong></td>
<td>10.11</td>
</tr>
<tr>
<td>Playtable</td>
<td>41.89</td>
<td>41.26</td>
<td>40.96</td>
<td>41.03</td>
<td>41.02</td>
<td><strong>39.29</strong></td>
<td>39.47</td>
</tr>
<tr>
<td>PlaytableP</td>
<td>13.81</td>
<td>16.42</td>
<td>14.18</td>
<td>12.93</td>
<td>13.06</td>
<td>13.08</td>
<td>13.20</td>
</tr>
<tr>
<td>Recycle</td>
<td>10.52</td>
<td>8.26</td>
<td>10.72</td>
<td>10.13</td>
<td>10.39</td>
<td>8.38</td>
<td>9.03</td>
</tr>
<tr>
<td>Shelves</td>
<td>39.52</td>
<td>32.48</td>
<td>39.60</td>
<td>39.32</td>
<td>39.76</td>
<td>36.11</td>
<td>38.02</td>
</tr>
<tr>
<td>Teddy</td>
<td>7.18</td>
<td>6.80</td>
<td>7.30</td>
<td>7.00</td>
<td>7.09</td>
<td><strong>6.33</strong></td>
<td>6.49</td>
</tr>
<tr>
<td>Vintage</td>
<td>33.22</td>
<td>37.59</td>
<td>35.74</td>
<td>32.53</td>
<td>32.55</td>
<td><strong>30.05</strong></td>
<td>31.57</td>
</tr>
<tr>
<td>Weighted Average</td>
<td>16.47</td>
<td>16.56</td>
<td>16.90</td>
<td>15.84</td>
<td>16.06</td>
<td><strong>14.66</strong></td>
<td>15.10</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Adirondack</td>
<td>10.09</td>
<td>10.22</td>
<td>10.53</td>
<td>9.57</td>
<td>9.68</td>
<td>8.39</td>
<td><strong>8.38</strong></td>
</tr>
<tr>
<td>ArtL</td>
<td>21.90</td>
<td>21.96</td>
<td>22.93</td>
<td>21.81</td>
<td>22.02</td>
<td>21.64</td>
<td><strong>21.22</strong></td>
</tr>
<tr>
<td>Jadeplant</td>
<td>34.51</td>
<td>36.88</td>
<td>35.60</td>
<td>34.09</td>
<td>34.46</td>
<td><strong>33.45</strong></td>
<td>33.49</td>
</tr>
<tr>
<td>Motorcycle</td>
<td>13.50</td>
<td>15.80</td>
<td>14.16</td>
<td>13.59</td>
<td>13.64</td>
<td><strong>13.23</strong></td>
<td>13.47</td>
</tr>
<tr>
<td>Piano</td>
<td>19.99</td>
<td>19.49</td>
<td>19.81</td>
<td>19.34</td>
<td>19.45</td>
<td><strong>18.29</strong></td>
<td>18.38</td>
</tr>
<tr>
<td>PianoL</td>
<td>36.43</td>
<td>34.44</td>
<td>35.87</td>
<td>36.04</td>
<td>36.14</td>
<td><strong>33.91</strong></td>
<td>34.22</td>
</tr>
<tr>
<td>Playroom</td>
<td>30.61</td>
<td>27.78</td>
<td>31.45</td>
<td>29.07</td>
<td>29.34</td>
<td><strong>25.38</strong></td>
<td>27.18</td>
</tr>
<tr>
<td>Playtable</td>
<td>44.65</td>
<td>44.13</td>
<td>43.86</td>
<td>43.98</td>
<td>44.00</td>
<td><strong>42.63</strong></td>
<td>42.65</td>
</tr>
<tr>
<td>PlaytableP</td>
<td>17.29</td>
<td>19.81</td>
<td>18.35</td>
<td><strong>15.92</strong></td>
<td>16.39</td>
<td>17.55</td>
<td>17.37</td>
</tr>
<tr>
<td>Recycle</td>
<td>12.30</td>
<td><strong>10.22</strong></td>
<td>12.50</td>
<td>12.42</td>
<td>12.72</td>
<td>10.50</td>
<td>11.21</td>
</tr>
<tr>
<td>Shelves</td>
<td>40.14</td>
<td><strong>33.62</strong></td>
<td>40.29</td>
<td>39.85</td>
<td>40.27</td>
<td>36.89</td>
<td>38.66</td>
</tr>
<tr>
<td>Teddy</td>
<td>12.59</td>
<td>12.44</td>
<td>12.72</td>
<td>12.48</td>
<td>12.57</td>
<td><strong>11.94</strong></td>
<td>12.09</td>
</tr>
<tr>
<td>Vintage</td>
<td>37.18</td>
<td>40.99</td>
<td>39.55</td>
<td>36.38</td>
<td>36.39</td>
<td><strong>34.01</strong></td>
<td>35.41</td>
</tr>
<tr>
<td>Weighted Average</td>
<td>21.74</td>
<td>22.05</td>
<td>22.20</td>
<td>21.30</td>
<td>21.51</td>
<td><strong>20.49</strong></td>
<td>20.74</td>
</tr>
</tbody>
</table>

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.

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


