Signal Processing: Image Communication 95 (2021) 116280

Contents lists available at ScienceDirect

Signal Processing: Image Communication

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

Efficient local stereo matching algorithm based on fast gradient domain guided image filtering

Weimin Yuan [a,*], Cai Meng [a,b], Xiaoyan Tong [a], Zhaoxi Li [a]

[a] School of Astronautics, Beihang University, Beijing 100191, China
[b] Beijing Advanced Innovation Center for Biomedical Engineering, Beihang University, Beijing 100191, China

ARTICLE INFO

Keywords:
Stereo matching
Cost aggregation
Disparity refinement
Guided image filtering

ABSTRACT

Guided image filtering (GIF) based cost aggregation or disparity refinement stereo matching algorithms are studied extensively owing to the edge-aware preserved smoothing property. However, GIF suffers from halo artifacts in sharp edges and shows high computational costs on high-resolution images. The performance of GIF in stereo matching would be limited by the above two defects. To solve these problems, a novel fast gradient domain guided image filtering (F-GDGIF) is proposed. To be specific, halo artifacts are effectively alleviated by incorporating an efficient multi-scale edge-aware weighting into GIF. With this multi-scale weighting, edges can be preserved much better. In addition, high computational costs are cut down by sub-sampling strategy, which decreases the computational complexity from $O(N)$ to $O(N/s^2)$ (s: sub-sampling ratio) To verify the effectiveness of the algorithm, F-GDGIF is applied to cost aggregation and disparity refinement in stereo matching algorithms respectively. Experiments on the Middlebury evaluation benchmark demonstrate that F-GDGIF based stereo matching method can generate more accuracy disparity maps with low computational cost compared to other GIF based methods.

1. Introduction

Stereo matching has been an active research field in computer vision. It plays an important role in many applications, including autonomous driving, 3D reconstruction, 3D tracking. Although stereo matching algorithms have made great progress, there are still challenges in the areas of low-textures regions, radiometric differences, occlusions, slanted surfaces. Large numbers of stereo matching algorithms have been presented to solve these problems [1,2].

A comprehensive taxonomy of stereo matching methods can be found in [1]. Most of stereo correspondence algorithms are following the four steps: matching cost computation, cost aggregation, disparity computation/optimization and disparity refinement. According to the optimization technology, stereo matching is divided into global and local methods. Global algorithms compute disparities by using different optimization methods, e.g., dynamic programming [1], belief propagation [3], graph cut [4,5]. The results of the global algorithm are more accurate than the local algorithm, but it also brings high computational complexity. On the contrary, local methods have lower computational complexity and can be applied in real-time applications [6]. Local methods perform averaging of matching costs [7] in a fixed size window or multiple windows and then disparities are computed by the winner-takes-all strategy (WTA). But, the local methods are limited in many practical applications because of their low accuracy. To overcome this, ASW(adaptive support weights) [7] was proposed to achieve a comparable result with global methods. In ASW, the cost aggregation can be treated as filtering the cost volume with bilateral filter (BF) [8]. Since then, several filtering based stereo matching methods have been proposed. But BF based cost aggregation method suffers from high computational complexity and gradient reversal artifacts. GIF [9] based cost aggregation or disparity refinement local stereo matching methods [10,11] can achieve better results than BF based ASW. But GIF will generate halo artifacts at some sharp edges, and has high computational costs on high-resolution images.

Inspired by the GIF improved version [12], a novel fast gradient domain guided image filter (F-GDGIF) is proposed by incorporating a multi-scale edge-aware weighting into GIF to avoid halo artifacts and sub-sampling strategy to reduce computational complexity. Then, F-GDGIF is adopted to filter cost volume in the cost aggregation [10] and raw disparity maps in the disparity refinement [11], respectively. The results on Middlebury datasets demonstrate the accuracy and efficiency of the proposed algorithm. To sum up, our main contributions are listed below:

1. By incorporating a multi-scale edge-aware weighting and sub-sampling strategy into GIF, a novel fast gradient domain guided image filter (F-GDGIF) is proposed, which alleviates halo artifacts and achieves better edge-aware performance with a faster execution time compared to GIF.

* Corresponding author.
E-mail address: yuanweimin@buaa.edu.cn (W. Yuan).

https://doi.org/10.1016/j.image.2021.116280
Received 23 April 2020; Received in revised form 31 December 2020; Accepted 6 April 2021
Available online 18 April 2021

2. The proposed F-GDGIF is employed in cost aggregation and disparity refinement, respectively. Due to its advantages of edge preserved smoothing property and low computational complexity, both experiments achieve better performance compared to GIF based stereo matching algorithms.

The rest of the paper is organized as following. Section 2 reviews the work most related to our method. The proposed method is described in Section 3. Experimental results and analyses are presented in Section 4. Concluding remarks are provided in Section 5.

2. Related works

Cost aggregation and disparity refinement as the important components determining the accuracy and efficiency of a stereo matching algorithm. Thus, a large number of cost aggregation and disparity refinement strategies have been proposed in the past decades. Here we mainly discuss the related works on these two steps.

2.1. Cost aggregation

Most cost aggregation methods can be roughly divided into the following three categories: window based methods [1,13–15], segment-tree (ST) based methods [16,17] and edge-aware filtering based methods [10,18–23].

Window based methods (fixed-size window [1]) were presented in the early stage of stereo matching. They suffer an increased error rate and blur object boundaries when the radius of the support window is enlarged. To avoid artifacts near disparity discontinuities, variable-window (VW) [13], multiple-window (MW) [14] and adaptive window (AW) [15] aggregation methods were proposed to aggregate the cost volume obtained in the matching cost computation stage. Compared to other window-based methods, AW methods achieve higher accuracy in texture-less regions and near disparity discontinuities. Segment-tree based methods also play an important role in cost aggregation. A non-local cost aggregation method (NLCA) based on the minimum spanning tree (MST) was proposed in [16], which had high precision and low computational complexity. In NLCA, each pixel is supported by other pixels through the MST structure. An improved NLCA version [17] for constructing a more reliable ST structure was proposed and achieved better results.

In edge-aware filtering based methods [7,10,18–21,23], cost aggregation can be regarded as a filtering process. Edge-aware filtering methods perform once for each disparity level in the process of cost aggregation. The emergence of local methods [7,10] based on different filtering methods has attracted much attention. As the earliest appeared edge-aware filtering-based method, adaptive support weight (ASW) method [7] based on bilateral filtering (BLF) [8] was proposed. However, ASW performs relatively low accuracy because of the gradient reversal artifacts and high computational complexity caused by BLF. To overcome the defects of BLF, guided image filtering (GIF) based on local linear model is proposed in [9]. GIF has better behavior near the edges and the computational complexity is independent of filter size. Inspired by ASW, GIF based cost aggregation algorithms [10, 18–23] achieved relatively satisfactory results. The accuracy of Cost-Filter [10] performed best among all the local methods at the time of proposed. GIF sometimes fail to compute an accurate weighting for the sharp edges. A Filter-size map was computed by DoG-scale to find the appropriate filtering window radius, DogGuided [18] can accurately estimate the smoothing weighting in cost aggregation. The iterative strategy GIF based cost aggregation methods can be found in [19,21]. While improving the filtering accuracy, methods based iterative strategy also introduce higher computational complexity. With two sub networks, a deep learning method DSGCA to perform GIF based cost aggregation was proposed in [20], the pre-training is much more time-consuming. A sub-sampling acceleration strategy GIF based cost volume filtering method is proposed in [23]. With this acceleration method, the efficiency of this method can be improved with no visible degradation. The limitation of [23] is that it cannot preserve sharp edges in the process of cost aggregation because of the halo artifacts in GIF.

GIF would exhibit halo artifacts near some edges such as sharp ones, which are unavoidable for most of the local filters. Obviously, the above GIF based cost aggregation methods cannot perform well in sharp edges because of the halo artifacts in GIF. To address this problem , some improved versions of GIF have been proposed [12,24] by introducing different edge-aware weightings. In [24], weighted guided image filtering (WGIF) was proposed by incorporating a single-scale weighting. On the basis of WGIF, a gradient domain GIF [12] was presented to further improve the filtering effect by incorporating a multi-scale weighting. The weighting [24] was incorporated in [22] to form WGIF based stereo matching method, the errors caused by halo artifacts are avoided to some extent by the adaptive regulation term. It is worth noting that the weighting in [22] is a single-scale factor, the weighting factor in this paper is a multi-scale factor, which can achieve much better filtering effects. An efficient depth image estimation method based on gradient weight cost aggregation is proposed in [25], the process is formulated as a local regression problem consisting of two steps, the costs of a set of points in shape-adaptive support region is computed first, then the matching costs are aggregated by gradient-based weight strategy. Compared with GIF based cost aggregation method, this method produced a more accurate disparity image for the application of wireless video sensor networks.

2.2. Refinement of disparities

As pointed out in [11], past research focused on the first three steps in stereo correspondence, whereas the disparity refinement is regarded as a small part in stereo matching methods. In fact, more and more disparity refinement algorithms [5,11,26–31] have been proposed, which demonstrate that these methods can significantly improve the accuracy of disparity results. Most of disparity refinement methods consist of the following three steps: occlusion detection, hole filling and edge-preserved smoothing filtering.

Left-and-right consistency check (LRC) [26] is defined that the matching points are regarded as occlusion region if they are not unique when computed from right-to-left and opposite direction. Due to its good performance and simplicity, LRC is widely used in many real-time disparity refinement methods. Another commonly used occlusion detection method is ordering constraint [27,28]. It is assumed that if there is no occlusion area in the scene, the relative order of matching points along the scan line is usually monotonous. The third occlusion detection methods [30] are based on the observation that image intensity edges usually caused by the orientation discontinuities and depth. After occlusion detection, the holes caused by occlusion area can be filled by surface fitting or interpolation method based on the surrounding pixels [28,29].

Median filter [5] is widely employed to improve the accuracy of disparity map. It selects intermediate value in the surrounding pixels as the final result of the central pixel. A constant time WM based disparity refinement was proposed in [11]. Due to the edge-preserved smoothing and outliers removing property of WM, the accuracy of disparity map can be greatly improved. Another filtering based disparity refinement method is Gaussian filter and its improvement [31]. The raw disparity map is estimated by taking into account the Gaussian distribution weights and the adjacent pixels. By taking into account the Gaussian distribution weights, the outliers in raw disparity map can be removed, but they also introduce high computational complexity.

3. Method

In this section, an overview of the proposed algorithm is described in the schematic flow diagram of Fig. 1. As depicted in the diagram, F-GDGIF based cost aggregation and disparity refinement are described in Fig. 1(a), (b), respectively. The details of each section will be described in the following.

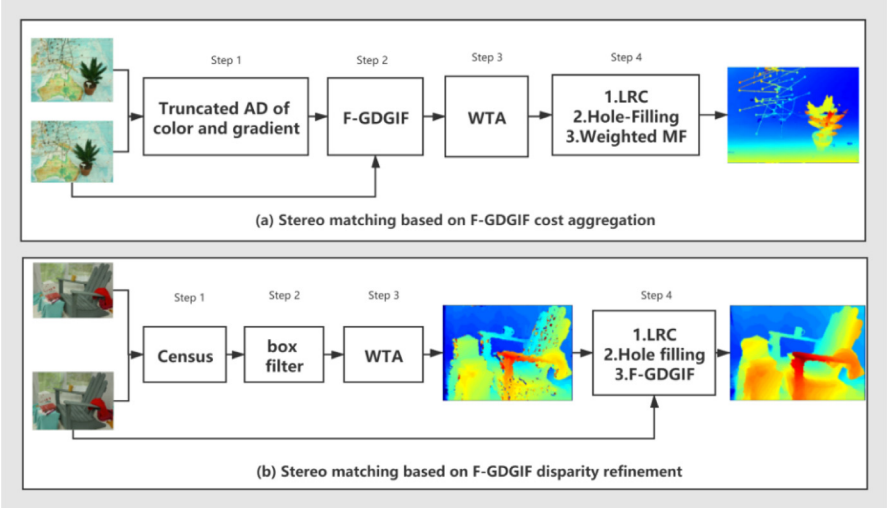

Fig. 1. Schematic flow diagram of the proposed method.

3.1. Fast gradient domain guided image filtering

GIF is a well-known local linear filtering for its low computational complexity and edge-preserving smoothing property. A linear translation-variant filter process was presented in GIF, which involves guidance image I, input image p and output image q. The filter output at a pixel I is expressed as a weighted average:

$$q_i = \sum_j W_{ij}(I)p_j \tag{1}$$

Here, i and j are pixel indexes. The filter kernel W_{ij} is a function of the guidance image I. The key assumption in GIF is a local linear model between guidance image I and output image q, which can be formulated as:

$$q_i = a_k I_i + b_k, \forall i \in w_k \tag{2}$$

Here, a_k and b_k are linear coefficients which are constant in w_k. Linear coefficients are determined by minimizing the difference between the filter input p and output q:

$$a_k = \frac{\frac{1}{|w|}\sum_{i\in w_k} I_i p_i - \mu_k \overline{p_k}}{\partial_k^2 + \varepsilon} \tag{3}$$

$$b_k = \overline{p_k} - a_k \mu_k \tag{4}$$

Here $|w|$ is the total number of pixels in the w_k, $\overline{p_k}$ is the mean of p in w_k, μ_k and ∂_k^2 are the mean and variance of I in w_k, respectively. ε is a regularization parameter, which preventing a_k from being too large. After computing a_k and b_k in the whole image, the filter output can be obtained by mean filter as:

$$q_i = \frac{1}{|w|}\sum_{k:\, i\in w_k}(a_k I_i + b_k) = \overline{a_i} I_i + \overline{b_i} \tag{5}$$

The filter kernel of GIF can be expressed as:

$$w_{ij}(I) = \frac{1}{|w|^2}\sum_{k:\,(i,j)\in w_k}\left(1 + \frac{(I_i - \mu_k)(I_j - \mu_k)}{\partial_k^2 + \varepsilon}\right) \tag{6}$$

For its edge-preserved smoothing property, GIF has been widely used in many computer vision and graphics applications, such as optical flow estimation [10], detail enhancement [9], HDR compression [24], stereo correspondence [10,11] and saliency detection [32]. But there is still existing the following two defects in GIF:

(1) Halo artifacts

GIF as one of the well-known edge-preserving local filtering can avoid the gradient reversal artifacts in BLF, but it would introduce blurring near edges and exhibit 'halo artifacts' especially when smooth the sharp edges [9], as indicated by the blue arrows in Fig. 2(b). This is because the local property and fixed regularization parameter ε in GIF, which limit its filtering effect.

(2) High time consumption on high-resolution images

GIF has $O(N)$ execution time (N: the total number of pixels) for both gray-scale and color images, independent of the intensity range and the filtering kernel size. However, it will cost much more time when process high-resolution images, which limits the application of GIF in practice.

Based on the above two defects in GIF, we have made the following two improvements. To avoid the halo artifacts and achieve better edge preserved smoothing results, single channel GDGIF was proposed in [12] by incorporating a multi-scale first-order edge-aware weighting as follows:

$$\Gamma_I(p') = \frac{1}{N}\sum_{P=1}^{N}\frac{\chi(p') + \lambda}{\chi(p) + \lambda} = \frac{1}{N}\sum_{p=1}^{N}\frac{\sigma_{I,1}(p')\cdot\sigma_{I,r}(p') + \lambda}{\sigma_{I,1}(p)\cdot\sigma_{I,r}(p) + \lambda} \tag{7}$$

Here I represents the guidance image, N is the total number of pixels in I, $\sigma_{I,1}(p')\cdot\sigma_{I,r}(p')$ is the multiply of local standard variance of (3 × 3) windows and (2r + 1) × (2r + 1) windows of guidance image I. The value of λ is defined as $(0.001 * L)^2$, L is the dynamic range of guidance image. After computing edge-aware weighting, the linear coefficients a_k and b_k in GDGIF become as follows:

$$a_k = \frac{\left(\frac{1}{|w|}\sum_{i\in w_k} I_i p_i - \mu_k \overline{p_k}\right) + \frac{\varepsilon}{\Gamma_I(p')}\gamma_p{}'}{\partial_k^2 + \frac{\varepsilon}{\Gamma_I(p')}} \tag{8}$$

$$b_k = \overline{p_k} - a_k \mu_k \tag{9}$$

The value of $\gamma_p{}'$ approaches 0 if the pixels in flat regions and 1 if the pixel is at an edge, see [12] for more details. Due to the multi-scale edge-aware weighting, pixels in different areas will be assigned various weighting values: pixels in the flat region of the filtered image should have large values, while pixels in the edge area have small values. Compared to GIF, GDGIF shows a better flat area smoothed and sharped edges preserved with negligible increment on computing times.

However, both filtering processes on cost volume [10] and raw disparity image [11] are based on three channels (RGB) guidance image simultaneously instead of single channel in [9,12] separately. So in this paper the local linear model (2) needs to be extended to three channels as follows:

$$q_i = \mathbf{a}_k^T \mathbf{I}_i + b_k, \forall i \in w_k \tag{10}$$

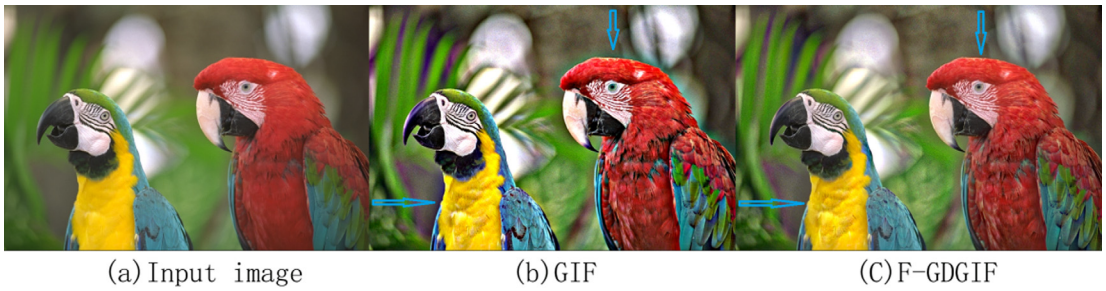

Fig. 2. The halo artifacts. The score of (b) and (c) is 18.11, 20.96 by quality metric [33]. (higher value represents better quality).

Here, $\mathbf{a}_k$ is a 3×1 vector, $\mathbf{I}_i$ is a 3×1 RGB color vector. The $\mathbf{a}_k$ and b_k for RGB color guidance image become as follows:

$$\mathbf{a}_k = (\Sigma_k + \frac{\varepsilon U}{\Gamma_I(p')})^{-1}((\frac{1}{|w|}\sum_{i \in w_k} \mathbf{I}_i p_i - \mu_k \overline{p_k}) + \frac{\varepsilon}{\Gamma_I(p')}\gamma_p{}') \tag{11}$$

$$b_k = \overline{p_k} - \mathbf{a}_k^T \mu_k \tag{12}$$

Here, Σ_k is the 3×3 covariance matrix of RGB color guidance image $\mathbf{I}$ in w_k, U is 3×3 identity matrix. Thus the filter kernel of three channels GDGIF can be expressed as:

$$w_{ij}^{GDGIF}(\mathbf{I}) = \frac{1}{|w|^2}\sum_{k:\,(i,j)\in w_k}(1 + (\mathbf{I}_i - \mu_k)^T(\Sigma_k + \frac{\varepsilon U}{\Gamma_I(p')})^{-1}(\mathbf{I}_j - \mu_k)) \tag{13}$$

In addition to the modification made for better edge preserving result, we introduce a sub-sampling strategy. Input filtering image and the guidance image are sub-sampled by sub-sampling ratio s. This method speeds up the execution time from $O(N)$ to $O(N/s^2)$. Finally, the efficient sub-sampled filtering kernel of three channels F-GDGIF method can be expressed as:

$$w_{ij}^{F-GDGIF}(\mathbf{I}^s) = \frac{1}{|W^s|^2}\sum_{k:\,(i,j)\in w_k{}^s}(1 + (\mathbf{I}_i^s - \mu_k^s)^T(\Sigma_k^s + \frac{\varepsilon U}{\Gamma_I^s(p')})^{-1} \times (\mathbf{I}_j^s - \mu_k^s)) \tag{14}$$

where $(*)^s$ indicates the parameter is sub-sampled by sampling ratio s. The final results can be obtained by up-sampling the linear coefficients (Eq. (5)) at the same sampling ratio s. As shown in Fig. 2(c), F-GDGIF can effectively avoid halo artifacts (indicated by the blue arrows) and scored a higher value (20.96) compared to GIF (18.11).

3.2. Stereo matching based on F-GDGIF cost aggregation

In this section, F-GDGIF is adopted to study cost volume filtering in stereo matching (Fig. 1b). In order to prove the effectiveness of F-GDGIF in cost aggregation stage, the same blending matching cost function $c_d(p)$ in [10] is adopted in this section: truncated absolute difference of the color and the gradient. Then, cost aggregation is implemented with the proposed F-GDGIF algorithm

$$c_d'(p) = w^{F-GDGIF}(\mathbf{I}_s, p) \cdot c_d(p) \tag{15}$$

Here, $c_d'(p)$ is the filtered cost volume at pixel p for disparity d. The weights $w^{F-GDGIF}(\mathbf{I}_s, p)$ is expressed in Eq. (14). After the cost volume is filtered by the proposed F-GDGIF, winner-take-all (WTA) strategy is adopted for disparity selection. In disparity refinement step, LRC, hole filling and weighted median filtering [10] are employed to remove the outliers in the end.

3.3. Stereo matching based on F-GDGIF disparity refinement

F-GDGIF is adopted to study raw disparity map refinement in this section. The pipeline of F-GDGIF based disparity refinement is shown in Fig. 1(C). The initial cost volume is constructed by non-parametric matching cost function: Census Transform (CT) [5], which is robust to radiometric variations and disparity discontinuities. The CT value at a pixel has eight binary bits according to the intensity difference p and its surrounding eight pixels, as shown in Fig. 3.

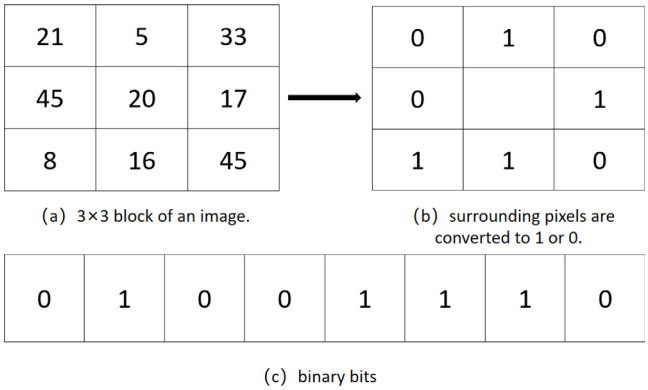

Fig. 3. Example of CT.

The CT translates the results of comparisons between the center pixel and its surrounding pixels into a bit string, which is formulated as follows:

$$cen(p) = U_{q \in N_p}\varphi(p, q) \tag{16}$$

Where N_p represents a fixed size window around centering pixel p, U indicates the concatenation operation for generating the string $cen(p)$, $\varphi(p, q)$ is a binary function defined as follows:

$$\varphi(p, q) = \begin{cases} 1, & if\ \ I(q) < I(p) \\ 0, & else \end{cases} \tag{17}$$

where $I(*)$ represents the value of pixel. Then a comparison of the Hamming distance is adopted to calculate the similarity:

$$c_d(p) = Ham\min g(cen(p), cen(pd)) \tag{18}$$

After the computation of matching cost, the obtained initial cost volume is aggregated by using regular rectangular windows, then the disparity of each pixel with the minimal cost is chosen by WTA. After the computation of raw disparity map, we first apply LRC to detect the occluded regions and hole-filling strategy to fill these detected regions, then F-GDGIF (Eq. (14)) based disparity refinement is employed to improve the accuracy of disparity map.

4. Experimental results

Experiments are conducted on Middlebury benchmark [34] to evaluate the proposed F-GDGIF method on cost aggregation and disparity refinement, respectively.

4.1. Testing database and parameters setting

To test adaptability of the proposed method, we evaluate it on the Middlebury benchmark. It consists of 15 training image pairs with

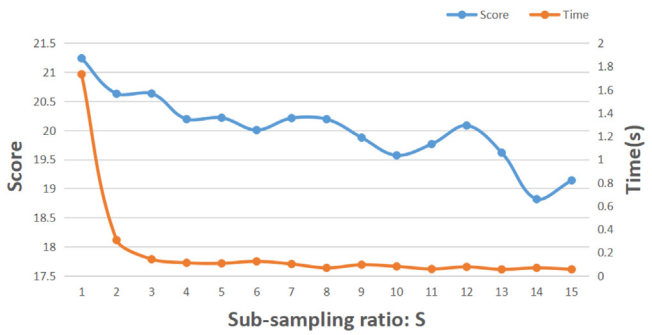

Fig. 4. The performance under different sub-sampling rate *s*.

Table 1
Parameters settings for F-GDGIF based cost aggregation.

	r	r_{median}	ϵ	λ	τ_1	τ_2	α	s
Value	9	19	$(0.01)^2$	$(0.001 * 255)^2$	7/255	2/255	0.11	3

Table 2
Parameters settings for F-GDGIF based disparity refinement.

	r_c	r_1	ϵ	λ	s
Value	5	9	$(0.01)^2$	$(0.001*255)^2$	3

corresponding ground truth disparity maps and 15 test image pairs. Each image pair is provided in three resolutions: full, half, and quarter. In our experiments, the proposed method is tested on half-resolution image pairs. The performance of each disparity map is evaluated with the percentage of absolute disparity error in non-occluded regions. The results of our method are compared with other results obtained from state-of-the-art stereo matching algorithms, which are based on GIF or GIF improved versions. All the experiments are conducted using MATLAB R2014a on a PC with Intel Core i7-3770M CPU @3.40 GHZ and 16.0 GB memories. The parameters of our methods are listed in Tables 1 and 2. To obtain convincing experimental results, all the listed parameters are consistent with parameters in [10,11], respectively.

The key parameter of this method is the sub-sampling ratio *s*, which controls the balance between the accuracy and efficiency of the algorithm. When *s* is set to 1, the algorithm does not perform sub-sampling and the efficiency is not improved. As *s* began to increase, the efficiency of the algorithm is improved, but the accuracy of results is decreased. Therefore, the proper setting of the parameter *s* is necessary for good performance. The evaluation results of *s* are shown in Fig. 4. With the metric [33], higher score value represents better quality. When *s* varies from 2 to 4, the blue curve has a higher value and the yellow curve a lower value. The best setting of *s* in our method is 3.

4.2. F-GDGIF based cost aggregation

For evaluating the performance of F-GDGIF on cost aggregation, five state-of-the-art GIF and its improved version based cost aggregation methods: Cost-Filter [10], DogGuided [18], ISM [19], DSDGA [20], IGF [21] are chosen for comparison. In addition, another three recently proposed stereo matching algorithms: AMNET [35], MTS [36], SM_AWP [37] are also compared with our algorithm. DogGuided [18] is on the basis of cost-Filter [10], guidance image in [18] is filtered by Different of Gaussian to obtain the optimal filtering size in the cost volume aggregation. ISM [19], an adaptive support weight with iterative GIF is applied to improve the result of cost aggregation, the same iterative GIF strategy in cost aggregation is also employed in [21]. In DSDGA [20], a deep learning technique is applied to perform self-guided filter based cost aggregation. An advanced deep learning method AMNet is proposed in [35]. In method [36], the disparity search range is restricted by a hierarchical representation strategy to achieve the goal of low computational complexity. Another edge aware preserved smoothing filtering: adaptive weighted BLF based cost aggregation method is proposed in [37]. The disparity maps and quantitative evaluation results of [18–21,35–37] can be obtained from Middlebury evaluation benchmark, the results of Cost-Filter [10] was evaluated by the author's source code on the Middlebury benchmark.

The disparity maps of four selected test data (Australia, Bicycle2, DjembeL, Plants) are shown in Fig. 5, it can be clearly observed that the disparity maps produced with our method are more edge-aware preserved smoothing and contained less noise compared with the results obtained from the other seven state-of-the-art stereo matching algorithms. Tables 3 and 4 show the quantitative evaluation results on Middlebury datasets, the bad pixel percentage is computed with an error threshold of two pixels. As demonstrated in Tables 3 and 4, the weighted average bad pixel percentage of our method is 31.6% and 24.0% with the threshold, the proposed method ranks at the first place among all the eight methods. In summary, our method achieves the best performance in terms of bad pixel percentage. Besides comparing the visual quality of disparity maps generated by the above eight methods, the complexity of these methods is also compared and briefly analyzed. The computational time of the proposed method and the seven comparison methods on different running environments are listed in Table 5.

Cost volume should be filtered by F-GDGIF once for each disparity level in the process of cost aggregation. So the computational complexity of cost aggregation with F-GDGIF is O(N · D/s2) for an guidance image with N pixels (D is the number of disparity levels). For a fair comparison, we set the running time of algorithm on quarter size images as a benchmark, which means that the running time on a half size images should be divided by 8 (4 times image size, 2 times disparity level), the running time on full size images is calculated by the same method. It can be observed that the proposed F-GDGIF based cost aggregation method, which is implemented on MATLAB without any parallelism or acceleration strategy is adopted, is one of the most efficient algorithms listed in Table 5. In addition, deep learning algorithms such as DSGCA [20] and AMNET [35] require training models, while the pre-training phase usually consumes much more time compared to other methods.

4.3. F-GDGIF based disparity refinement

As mentioned earlier, disparity refinement can also be regarded as a filtering process. To demonstrate the effectiveness of F-GDGIF in disparity refinement stage, we conduct a comparative experiment. In this comparative experiment, a raw disparity map is obtained by CT [38] (Fig. 6(b)), WM based disparity refinement method [11] is adopted as the benchmark in this experiment.

From the visual results in Fig. 6, F-GDGIF based disparity refinement has better edge-preserved smoothing and contains fewer outliers compared with WM [11]. The weighted average bad pixel percentage with three different thresholds (i.e., 1.0, 2.0, 4.0) in non-occ regions of all the training stereo pairs are shown in Table 6. When the threshold is set to 1.0, the weighted average bad pixel percentage of raw disparity maps (without refinement) obtained by CT is 20.49%. Refined by F-GDGIF, the weighted average bad pixel percentage of raw disparity maps is decreased by 5.58%, 4.08%, 3.69% in three different thresholds, respectively. Compared with WM [11], our proposed method reduces the bad pixel percentage by 3.61%, 1.47% and 0.58%, respectively. In addition, the running time of WM [11] and F-GDGIF based refinement method is listed in Table 7, it can be observed that our method is performed better in edge preserved with lower computational complexity.

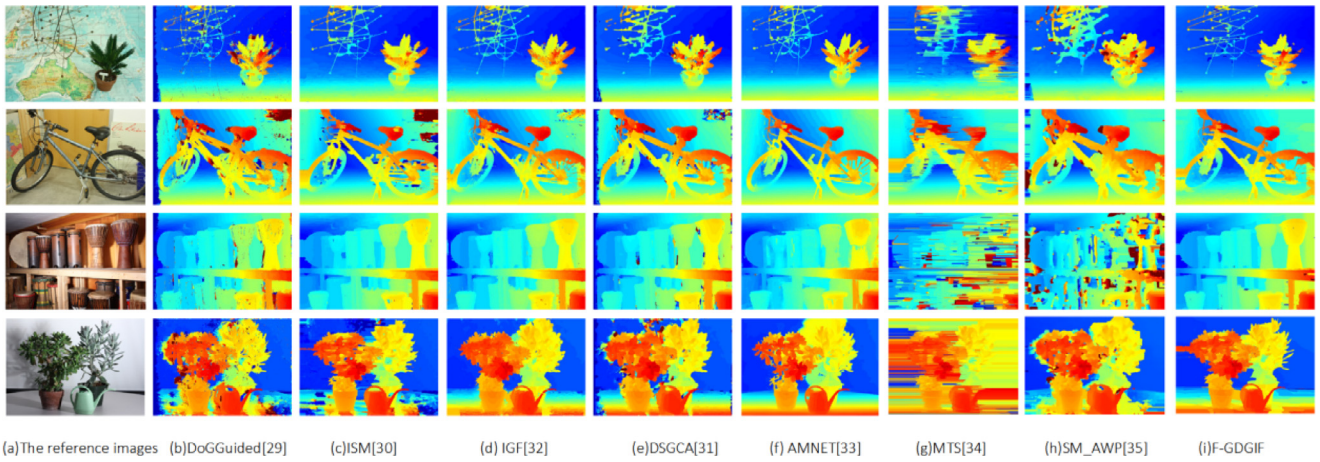

Fig. 5. The disparity map results of different methods.

Table 3
Bad pixel percentage comparison on Middlebury test data.

Data	DoG [18]	ISM [19]	IGF [21]	DSGCA [20]	AMNE [35]	MTS [36]	SM_AWP [37]	F-GDGIF
Australia	45.4	42.5	42.7	42.9	54.3	58.8	**30.7**	37.3
AustraliaP	23.6	26.4	20.1	20.9	63.7	25.2	24.0	**7.72**
Bicycle2	30.6	34.8	23.7	23.6	51.2	51.1	25.2	**16.1**
Classroom2	34.6	36.1	32.2	**30.2**	51.3	60.4	30.3	34.9
Cllassroom2E	52.5	44.5	45.6	45.5	40.6	91.3	44.9	**35.2**
Computer	28.3	34.4	28.6	27.6	39.9	48.3	38.1	**20.1**
Crusade	59.1	56.2	43.0	**42.0**	51.6	70.3	56.0	55.3
CrusadeP	53.8	52.7	37.2	**36.0**	55.9	63.4	55.8	46.6
Djembe	26.4	25.2	21.4	21.0	55.4	44.4	19.9	**9.01**
DjembeL	60.6	51.0	50.9	50.2	58.9	79.3	60.1	**47.6**
Hoops	54.7	52.4	44.7	**44.2**	57.5	71.7	51.2	52.9
Livingroom	38.3	39.7	34.7	33.3	52.7	60.9	32.1	**29.4**
Newkuba	35.5	33.3	31.9	34.6	49.5	47.4	30.2	**29.7**
Plants	44.5	38.8	37.4	38.4	58.1	79.6	40.0	**29.4**
Staircase	72.0	75.3	**47.1**	46.8	61.6	88.4	61.7	61.0
W-Avg	41.4	40.8	34.0	33.8	53.3	59.7	38.1	**31.6**

Table 4
Bad pixel percentage comparison on Middlebury training data.

Data	DoG [18]	ISM [19]	IGF [21]	DSGC [20]	CostFilte [10]	AMNET [35]	MTS [36]	SM_AW [37]	F-GDGIF
Adirondack	40.5	37.2	36.7	27.0	43.2	46.6	60.1	30.7	**19.7**
Artl	23.9	28.8	28.0	20.4	20.0	49.5	50.5	30.7	**13.2**
Jadeplant	42.5	45.5	44.6	35.0	35.8	60.8	75.3	40.7	**24.9**
Motor	30.8	32.1	31.5	26.1	25.7	39.2	49.0	21.0	**15.8**
MotorE	31.6	34.3	33.7	25.1	29.7	42.2	54.3	24.1	**19.1**
Piano	41.8	44.7	45.3	32.4	39.0	44.7	62.1	35.4	**27.6**
PianoL	58.0	53.1	52.5	45.7	63.1	46.9	79.5	52.9	**27.7**
Pipes	28.6	34.2	33.5	26.0	21.9	52.3	52.9	23.5	**15.3**
Playroom	44.6	44.6	44.4	41.8	43.5	52.4	66.7	**38.6**	42.8
Playtable	61.5	59.8	59.8	53.5	60.9	49.1	64.3	**31.5**	51.4
PlaytableP	36.4	44.5	44.4	35.1	35.9	42.7	41.9	29.6	**22.0**
Recycle	33.1	38.1	37.8	26.7	28.2	48.8	68.4	26.2	**23.7**
Shelves	54.6	51.8	52.5	54.9	53.2	**46.3**	66.8	49.1	53.9
Teddy	14.7	22.0	21.6	13.8	9.99	41.7	39.4	16.5	**8.76**
Vintage	57.5	55.0	54.9	50.8	60.7	49.7	79.9	48.4	**45.4**
W-Avg	37.0	39.5	39.1	31.3	34.4	47.3	58.6	31.1	**24.0**

Table 5
Comparison of computational time of different methods.

Algorithm	Running environments	Image size	Disparity level	Runtime (s)
ISM [19]	I7-4770 @3.2 GHZ; 8G RAM (C/C++)	Quarter	200	26.2
DSGCA [20]	I7-4770 @3.4 GHZ; GTX 1080GPU (MATLAB)	Quarter	200	42.0
DoGGuided [18]	I5 core @ 3.0 GHZ(MATLAB)	Quarter	200	1895
IGF [21]	I5 core @ 3.2 GHZ (c++, opencv)	Quarter	200	418
AMNET [35]	Python; Titan XP	Quarter	200	1.40
MTS [36]	Intel® CoreTM i7- 2600K CPU running @ 3.40 GHz	Full	800	0.31
SM_AWP [37]	MATLAB; I i7 core@2.7 GHz	Quarter	200	3.66
Costfilter [10]	i7-3770M CPU @ 3.40 GHZ (MATLAB R2014a)	Half	400	881.7
F-GDGIF	i7-3770M CPU @ 3.40 GHZ (MATLAB R2014a)	Half	400	98.5

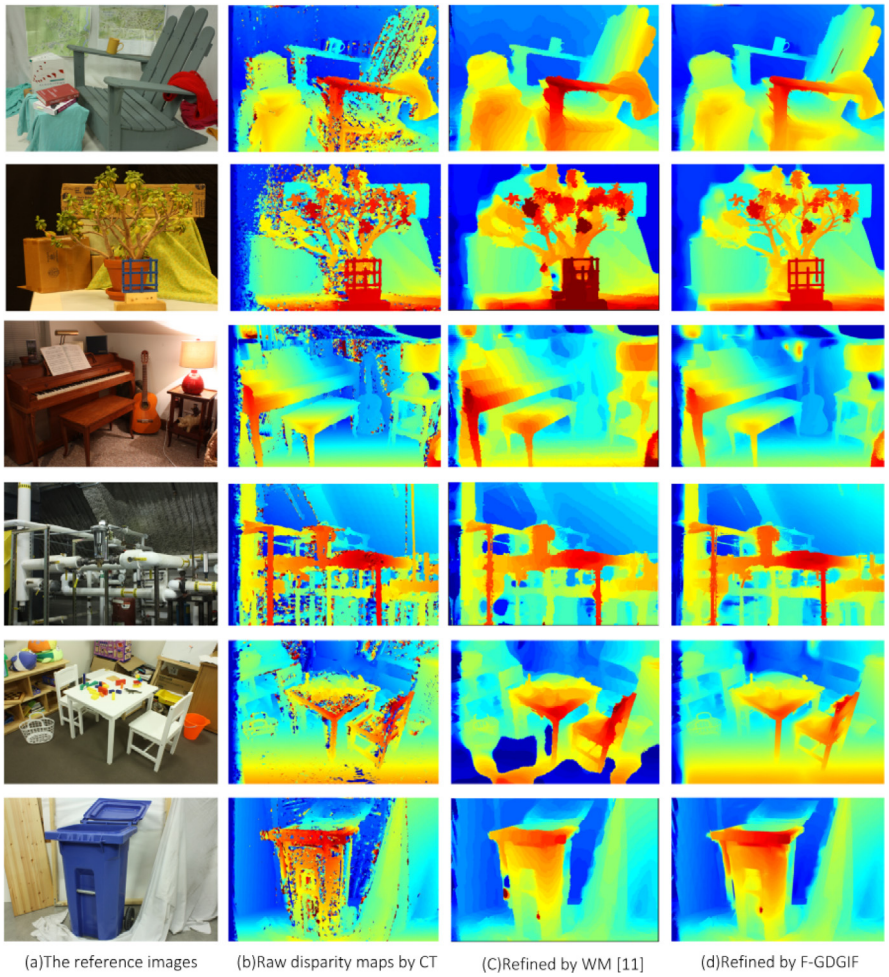

Fig. 6. The disparity map results of different methods.

Table 6
Bad pixel percentage comparison with three different thresholds.

Data	CT	CT + WM	CT + F-GDGIF	CT	CT + WM	CT + F-GDGIF	CT	CT + WM	CT + F-GDGIF
	$\sigma = 1.0$			$\sigma = 2.0$			$\sigma = 4.0$		
Adirondack	12.22	8.71	**5.67**	6.79	3.43	**2.77**	4.72	1.75	**1.62**
Artl	18.45	15.70	**14.63**	15.06	12.61	**11.86**	11.76	9.37	**9.11**
Jadeplant	25.98	22.76	**20.77**	20.12	17.58	**16.66**	15.65	13.37	**13.00**
Motorcycle	11.05	10.41	**8.70**	7.07	6.47	**5.96**	4.99	3.60	**3.38**
MotorcycleE	10.16	9.98	**8.30**	6.61	6.11	**5.63**	4.63	3.55	**3.40**
Piano	19.90	16.27	**14.76**	16.60	12.44	**11.66**	13.24	8.58	**8.21**
PianoL	33.98	27.66	**25.02**	30.02	23.41	**22.59**	25.94	19.48	**19.12**
Pipes	15.66	16.58	**15.94**	13.37	14.62	**12.89**	11.11	12.11	**11.8**
Playroom	28.43	29.73	**20.10**	22.00	17.91	**14.87**	17.09	12.29	**10.94**
Playtable	37.05	28.97	**25.99**	30.36	22.37	**21.20**	25.92	16.93	**15.86**
PlaytableP	21.6	27.86	**15.49**	16.53	12.81	**8.54**	12.52	4.81	**4.01**
Recycle	14.91	14.39	**7.64**	9.66	6.42	**5.00**	6.39	3.56	**3.05**
Shelves	43.09	40.17	**31.33**	34.34	29.35	**25.52**	27.15	21.61	**19.24**
Teddy	11.47	9.93	**8.32**	8.96	7.39	**6.61**	5.70	4.85	**4.32**
Vintage	35.78	31.42	**29.77**	27.93	27.85	**24.75**	23.39	21.88	**19.87**
W-Avg	20.49	18.52	**14.91**	15.44	12.83	**11.36**	12.04	8.93	**8.35**

Table 7
The computational time of comparative experiment.

Method	Environment	Runtimes (s)
CT+WM [11]	i7-3770M CPU @3.40 GHZ (MATLAB R2014a)	744
CT+F-GDGIF	i7-3770M CPU @3.40 GHZ (MATLAB R2014a)	276

5. Conclusion

A novel fast gradient domain guided filtering (F-GDGIF) is proposed in this paper. Compared to GIF, F-GDGIF can effectively avoid halo artifacts by incorporating multi-scale edge-aware weighting and show low computational complexity by introducing sub-sampling strategy. The

proposed F-GDGIF is applied to cost aggregation and disparity refinement steps in stereo matching algorithms, respectively. The proposed method has been tested on the challenging Middlebury benchmark. Experimental results demonstrate that the F-GDGIF based cost aggregation method outperformed other state-of-art GIF and its improved version cost aggregation methods in terms of the weighted average bad pixel error. The F-GDGIF based disparity refinement method also achieves good performance with low computational cost. Recently, GIF improvement versions [39,40] based on deep learning have been proposed. This will be the focus of our future research on cost aggregation and disparity refinement in stereo matching.

CRediT authorship contribution statement

Weimin Yuan: Conceptualization, Methodology, Validation, Writing - original draft, Supervision. **Cai Meng:** Review and editing . **Xiaoyan Tong:** Writing . **Zhaoxi Li:** Review and editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grants No. 61873010, 91748201), the National Key Research and Development Program of China (Grant No. 2019YFB1311703), Tianjin Key Research and Development Project under Grant No. 19YFSLQY00050.

References

[1] D. Scharstein, R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, Int. J. Comput. Vis. 47 (1) (2002) 7–42.
[2] H. Hirschmuller, D. Scharstein, Evaluation of Stereo Matching Costs on Images with Radiometric Differences, IEEE Computer Society, 2009.
[3] M.G. Mozerov, J.V.D. Weijer, Accurate stereo matching by two step energy minimization, IEEE Trans. Image Process. 24 (3) (2015) 1153–1163.
[4] T. Taniai, Y. Matsushita, Y. Sato, T. Naemura, Continuous 3D label stereo matching using local expansion moves, IEEE Trans. Pattern Anal. Mach. Intell. 40 (11) (2018) 27252739.
[5] H.R. Affendi, I. Haidi, Literature survey on stereo vision disparity map algorithms, J. Sens. 2016 (2016) 1–23.
[6] X. Mei, X. Sun, M. Zhou, H. Wang, X. Zhang, S. Jiao, On building an accurate stereo matching system on graphics hardware, in: Proc. IEEE Int. Conf. Comput. Vis. Workshops, Nov. 2012, pp. 467–474.
[7] K.J. Yoon, I.S. Kweon, Adaptive support-weight approach for correspondence search, IEEE Trans. Pattern Anal. Mach. Intell. (2006).
[8] C. Tomasi, R. Manduchi, Bilateral filtering for gray and color images, in: Computer Vision, 1998. Sixth International Conference on, IEEE, 1998.
[9] Kaiming He, Jian Sun, Xiaoou Tang, Guided image filtering, IEEE Trans. Pattern Anal. Mach. Intell. 35 (6) (2013) 1397–1409.
[10] C. Rhemann, A. Hosni, M. Bleyer, et al., Fast cost-volume filtering for visual correspondence and beyond, 2011.
[11] Z. Ma, K. He, Y. Wei, et al., Constant time weighted median filtering for stereo matching and beyond, in: Proceedings of the 2013 IEEE International Conference on Computer Vision, IEEE, 2013.
[12] F. Kou, W. Chen, C. Wen, et al., Gradient domain guided image filtering, IEEE Trans. Image Process. 24 (11) (2015) 1.
[13] O. Veksler, Fast variable window for stereo correspondence using integral images, in: Proc. Conf. Computer Vision and Pattern Recognition, 2003, pp. 556–561.
[14] Heiko Hirschmüller, P.R. Innocent, J. Garibaldi, Real-time correlation-based stereo vision with reduced border errors, Int. J. Comput. Vis. 47 (1–3) (2002) 229–246.
[15] J. Lu, G. Lafruit, F. Catthoor, Anisotropic local high-confidence voting for accurate stereo correspondence, in: Image Processing: Algorithms and Systems, Vol. VI, DBLP, San Jose, California, USA, 2008, 2008.
[16] Q. Yang, A non-local cost aggregation method for stereo matching, in: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE, 2012.
[17] X. Mei, X. Sun, W. Dong, et al., Segment-tree based cost aggregation for stereo matching, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2013.
[18] Masamichi Kitagawa, Ikuko Shimizu, Radim Sara, [IEEE 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA) - Nagoya, Japan] 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA) - High accuracy local stereo matching using DoG :258-261..
[19] Rostam Affendi Hamzah, Fauzan Kadmin, Saad Hamid, Fakhar Ghani, Haidi Ibrahim, Improvement of stereo matching algorithm for 3D surface reconstruction, Signal Process., Image Commun. 65 (2018) 165–172.
[20] Williem, In Kyu Park, Deep self-guided cost aggregation for stereo matching, Pattern Recognit. Lett. 112 (2018) 168–175.
[21] Rostam Affendi Hamzah, Haidi Ibrahim, A.H. Abu Hassan, Stereo matching algorithm based on per pixel difference adjustment, iterative guided filter and graph segmentation, J. Vis. Commun. Image Represent. 42 (2017) 145–160.
[22] G. Hong, B. Kim, A local stereo matching algorithm based on weighted guided image filtering for improving the generation of depth range images, Displays 49 (sep.) (2017) 80–87.
[23] G.S. Hong, J.K. Park, B.G. Kim, Near real-time local stereo matching algorithm based on fast guided image filtering, in: 2016 6th European Workshop on Visual Information Processing (EUVIP), IEEE, 2016.
[24] Z. Li, J. Zheng, Z. Zhu, et al., Weighted guided image filtering, IEEE Trans. Image Process. 24 (1) (2015) 120–129.
[25] G.S. Hong, B.G. Kim, T.J. Kim, et al., Efficient depth map estimation method based on gradient weight cost aggregation strategy for distributed video sensor networks, Int. J. Distrib. Sens. Netw. (2014).
[26] Q. Yang, L. Wang, R. Yang, et al., Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling, IEEE Trans. Pattern Anal. Mach. Intell. 31 (3) (2009) 492–504.
[27] C. Silva, J. Santos-Victor, Intrinsic Images for Dense Stereo Matching with Occlusions, in: Proc. European Conf. Computer Vision, 2000, pp. 100-114.
[28] M.Z. Brown, D. Burschka, G.D. Hager, Advances in computational stereo. IEEE Trans. Pattern Anal. Mach. Intell., 25(8) 993-1008.
[29] S.D. Cochran, G. Medioni, 3-d surface description from binocular stereo, IEEE Trans. Pattern Anal. Mach. Intell. 14 (10) (1992) 981–994.
[30] S. Birchfield, C. Tomasi, Multiway cut for stereo and motion with slanted surfaces, in: Proc. Int'l Conf. Computer Vision, Vol.1, 1999, pp. 489-495.
[31] K.R. Vijayanagar, M. Loghman, J. Kim, Real-time refinement of kinect depth maps using multi-resolution anisotropic diffusion, Mob. Netw. Appl. 19 (3) (2014) 414–425.
[32] Y. Ding, J. Xiao, J. Yu, Importance filtering for image retargeting, in: Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2011, pp. 89-96.
[33] A.K. Moorthy, A.C. Bovik, A two-step framework for constructing blind image quality indics, IEEE Signal Process. Lett. 17 (5) (2010) 513–516.
[34] D. Scharstein, H. Hirschmuller, Y. Kitajima, G. Krathwohl, N. Nei, X. Wang, P. Westling, High-resolution stereo datasets with subpixel-accurate ground truth, in: German Conference on Pattern Recognition, 2014, pp. 31–42.
[35] Xianzhi Du, Mostafa El-Khamy, Jungwon Lee, Amnet: Deep atrous multiscale stereo disparity estimation networks, 2019, arXiv:1904.09099v1 [cs.CV].
[36] R. Brandt, N. Strisciuglio, N. Petkov, et al., Efficient binocular stereo correspondence matching with 1-d max-trees, Pattern Recognit. Lett. (2020).
[37] Siti Safwana Abd Razak, Mohd Azlishah Othman, Ahmad Fauzan Kadmin, Siti safwana abd razak mohd azlishah othman ahmad fauzan kadmin the effect of adaptive weighted bilateral filter on stereo matching algorithm, Int. J. Eng. Adv. Technol. (2019).
[38] R. Zabih, J. Woodfifill, Non-parametric local transforms for computing visual correspondance, in: Proceedings of the European Conference of Computer Vision, Stockholm, Sweden, 1994, pp. 151–158.
[39] H. Wu, S. Zheng, J. Zhang, K. Huang, Fast end-to-end trainable guided filter, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, http://dx.doi.org/10.1109/CVPR.2018.00197.
[40] H. Yin, Y. Gong, G. Qiu, Fast and efficient implementation of image filtering using a side window convolutional neural network, Signal Process. (2020) 107717.