An efficient local stereo matching method based on an adaptive exponentially weighted moving average filter in SLIC space

Shan Yang | Xinyue Lei | Zhenfeng Liu | Guorong Sui

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, No. 516 JunGong Road, Shanghai, China

Correspondence
Guorong Sui, School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, No. 516 JunGong Road, Shanghai 200093, China. Email: suigr@usst.edu.cn

Funding information
National Basic Research Program of China, Grant/Award Number: 2015CB352001; Shanghai Leading Academic Discipline Project, Grant/Award Number: S30502

Abstract
Rapidly obtaining accurate dense disparity maps has been the focus of stereo matching research. At present, approaches that achieve superior disparity maps require a large amount of computation, which is not suitable for practical applications. To address this issue, this paper proposes an efficient local matching method based on an adaptive exponentially weighted moving average filter and simple linear iterative clustering segmentation algorithm. First, an effective matching cost is introduced to adaptively integrate absolute intensity difference with Census transform, which is robust against texture free and luminance variate. Following this, during the cost aggregation, the exponentially weighted moving average filter and the SLIC segmentation are combined to handle the problems of computing consumption and adaptive expansion of the cost aggregation window. Finally, the dense disparity map is obtained by a winner-takes-all approach and disparity refinement. To demonstrate its efficiency and validity, the method is quantitatively tested and compared to existing approaches on the Middlebury benchmark. The results show that it has a non-occlusion accuracy of 90.66% and an average runtime of 7.01 s on the 2014 Middlebury dataset. Compared with existing competitive methods, the proposed method achieves superior matching results with a lower time cost.

1 | INTRODUCTION

Stereo matching is at the forefront of the field of computer vision. It has wide applications, such as achieving the bokeh effect in images [1], 3D reconstruction [2], and industrial measurements [3]. Owing to the need for real-time processing and precision, fast acquisition of dense disparity maps with high accuracy has been the fundamental principle of stereo matching. To achieve this, four critical steps are considered: cost computation, cost aggregation, disparity computation, and disparity optimization [4]. Existing algorithms have been classified into global and local algorithms based on the strategy used in the disparity computation step. The global algorithms construct the global energy function through dynamic programming [5], belief propagation [6], and graph cuts [7] based on the smoothness assumption step. SGM [8, 9] transforms the 2D energy function into multiple 1D ones. These approaches have achieved a superior disparity map, albeit with a large amount of computation. The runtime of top-ranked global algorithms on the Middlebury datasets is over 120 s [10], which is not suitable for practical applications.

Moreover, local methods have received substantial attention owing to the use of the winner-takes-all strategy (WTA) for disparity computation, which significantly reduces the time complexity. The primitive approach is the box filter [4], which uses integral operations to achieve linear time. However, the fixed-size windows (FW) would destroy the disparity discontinuity and lead to edge blurring. Therefore, the local smoothness assumption has been proposed. This assumption states that the support window should fit the object boundary to improve the performance at the edges. Accordingly, methods based on variable support windows (VSW) have been researched.

In recent times, the use of convolutional neural networks (CNNs) for stereo matching to produce high quality disparity maps is increasing. Matching cost-convolutional neural network (MC-CNN) uses CNNs to get matching costs [11]. Liang [12] and Nikolaus [13] proposes an end-to-end network architecture of stereo matching. Cascade Residual Learning (CRL) [14] uses
the cascade CNN network to produce high-quality disparities. All these CNN-based approaches have yielded better results but they require a large number of training data with accurate ground truth and hardware resources for training and computation, which limits their use in low-end products.

Considering these factors, a local algorithm is chosen as the focus of this paper. This study focuses on the cost computation and aggregation steps. Cost computation provides source information for subsequent processing, which can build the foundation for excellent results. Additionally, cost aggregation is critical for preserving structure and recovering accurate disparity by producing proper support windows. An adaptively moving exponentially weighted average filter method based on simple linear iterative clustering (SLIC) space is proposed to handle the issue of balancing computation efficiency and disparity accuracy. To handle the shortage of computational efficiency in the local algorithm, the SLIC segment algorithm [15] is introduced to decrease the dimension from the pixel domain to the superpixel domain. Unlike other segment-based methods, the disparity varies in individual segments, owing to the combination of pixel-wise cost and superpixel-wise cost. To overcome the problem of insufficient information on a limited aggregation window, the adaptive exponentially weighted moving average (EWMA) filter [16] is introduced, which combines the smoothing filter and adaptive weights.

The remainder of this paper is organised as follows. Section 2 further describes the related work. Section 3 describes the proposed stereo matching method in terms of the four critical steps mentioned above. Section 4 presents the experimental approach and the evaluation results. Finally, the main conclusions are discussed in Section 5.

2 | RELATED WORKS

Since cost computation and cost aggregation are the two major areas related to our study, we briefly review the latest literature related to them in the following sections.

2.1 | Cost computation based on multiple measurement

For cost computation, researchers have proposed numerous cost functions to detect the correlation between correspondence points, including the squared intensity difference (SD), absolute intensity difference (AD), sum of absolute difference (SAD) [11], sum of squared differences (SSD) [4], normalized correlation coefficient (NCC), census transform (CT) [4, 9], gradient amplitude (GA) [17, 18], and mutual information (MI) [8]. In recent years, neural networks [11] have been increasingly applied to compute matching cost. In order to avoid the weakness of individual cost functions, multiple cost functions are combined to improve performance. Zhu [17] constructed a cost function based on truncated absolute difference (AD) and gradient. An explicit factor is used to balance the two terms. A similar equation is adopted in [18], where gradient and CT are blended with an implicit factor. However, the factor in these methods is a constant value. This study employs a variable factor to adaptively balance the shares of metrics to enhance the strength of the combination.

2.2 | Cost aggregation based on segmentation and iterative filtering

A proper aggregation strategy can reduce the computational load and can have a significant effect on large low-texture regions, repetitive regions and occlusion regions. Veksler [19] proposed multiple windows to aggregate the optimal cost. However, with this, the improvement over the box filter was limited owing to the finite number of window sizes. Mei et al. [20] improved the cross-based windows proposed by Zhang et al. [21]. This approach chooses optimal window shapes to preserve the boundaries, which significantly increases the time consumption. Considering that both the VSW filters and cross-based window filters only have binary weight values, the bilateral filter (BF) [22] was studied by adjusting weight depending on local distributions of colour and distance. Inspired by the concept of the BF, Yoon and Kweon [23, 24] proposed the adaptive approach, which replaced the Gaussian function with an exponential function to assign the weights based on the Euclidean distance of range similarity and spatial proximity. Further, in order to reduce the time consumption, Hosni [25] introduced the guided filter (GF) into the stereo matching algorithm with independent kernel size processing times. While these adaptive support weight (ASW) methods further improve the accuracy, the computational efficiency gradually decreases [26, 27].

Various methods to further increase the matching efficiency have been researched. On the one hand, the non-local solution based on the minimum spanning tree (MST) [28] has been studied. The segment tree (ST) [29] and iterative MST structure [30] are presented immediately after the MST. However, for all the MST-based solutions, it is difficult to guarantee the performance on inclined surfaces [31] owing to limited valid aggregation concerning the geodesic distance. This is particularly significant in large, low-texture regions, such as grey grounds and white walls. On the other hand, segment-based methods have received significant attention from researchers for the efficiency of down-sampling. Tombari [32] used the Mean-Shift segmentation algorithm to obtain segments. Then, it aggregated the cost from a segment and the corresponding adaptive window, which achieved better performance in highly textured regions. But errors still existed in low-textured regions owing to few identifiable correlations. To improve this problem, Chang [33] employed the SLIC segmentation algorithm and blended the cost of multiple neighbourhood segments. However, these segment-based methods have a common shortage: the disparities between segments may be discontinuous.

This paper combines the adaptive EWMA filter and SLIC segmentation algorithm in the cost aggregation step. In contrast to the segmentation-based adaptive support (SAS) algorithm [32], the proposed method performs the weighted average between segments to obtain sufficient information. Different
from the segment-based cost aggregation (SCA) algorithm [33] using a fixed weight, our approach employs adaptive weights for the EWMA filter, which efficiently extends the support window in SLIC space.

3 | METHODS

This section describes the proposed method in detail. The flowchart of the method is also shown (see Figure 1).

3.1 | Matching cost computation

The cost computation process utilises a variety of cost functions. The individual functions are straightforward and efficient, but the results are vulnerable to noise perturbation and illumination changes. Therefore, the combinations of multiple matching costs are exploited to obtain better performance.

The AD detects the luminance difference between pixels and is best suited for low-texture and flat areas, while the CT measures the relative magnitudes within neighbourhoods and has a high tolerance to radiometric distortion and highly textured regions. Therefore, the cost function combined with AD and CT is an optimal choice for robustness in a variety of scenarios.

The formula thus obtained is as follows:

$$C_p(p, d) = \gamma C_{AD}(p, d) + (1 - \gamma) C_{census}(p, d)$$

$$C_{AD}(p, d) = \frac{1}{3} \sum_{k=R_1 G B} |I^\text{Left}_k(p) - I^\text{Right}_k(p - d)|$$

$$C_{census}(p, d) = \sum H^\text{Left} \oplus H^\text{Right}$$

where $I^\text{Left}_k$ and $I^\text{Right}_k$ are the values of different channels belonging to the left and right images, respectively, and $\oplus$ is the XOR operator. $H_C$ is the binary luminance difference. The $\gamma$-value varies depending on the geometric structure to make full use of the characteristics of the AD and CT strategies, making the matching cost curve of the image more robust to textureless regions and brightness changes.

To attribute more weight to the AD-based cost $C_{AD}$ in textureless regions, where the CT cost $C_{census}$ has no discriminating power, the parameter $\gamma$ is negatively correlated with the detail strength. $\gamma$ is defined by the Laplacian function with respect to gradient $I_G$ at pixel $(i, j)$.

$$\gamma(i, j) = \exp \left( -\frac{I_G(i, j)}{\sigma_g} \right)$$

The value of $\sigma_g$ is used to control the shape of the Laplacian filter. The gradient $I_G(i, j)$ is detected with convolution operations in a discrete form.

$$I_G(i, j) = \sqrt{I^2_{GH}(i, j) + I^2_{GV}(i, j)}$$

$$I^\text{GH}_{R1 G B}(i, j) = \sum_{s=-R1}^{R1} \sum_{t=-R2}^{R2} (IJ(i+s, j+t) \cdot f_H(s, t))$$

$$I^\text{GV}_{R1 G B}(i, j) = \sum_{s=-R2}^{R2} \sum_{t=-R1}^{R1} (IJ(i, j-t) \cdot f_V(s, t))$$

Here, the $I^\text{GH}_{R1 G B}$ and $I^\text{GV}_{R1 G B}$ are the horizontal and vertical gradients, respectively, and $f_H$ and $f_V$ are filter kernels with radii $R_1$ and $R_2$. The coefficient value of the convolution kernel is negatively correlated to the radial distance. Hence, the initial values $f(x, y)$ are generated by the Gaussian function.

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp \left( -\left( \frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2} \right) \right)$$
The values of \( \sigma_x \) and \( \sigma_y \) are used to control the shape of the function. Some additional processing is necessary to make the filter sensitive to gradients similar to the Sobel operator. The values of the middle column of \( f(x,y) \) are set to zero and the values of the left column take the opposite number. Then, \( f_H \) is produced to detects the gradients in the horizontal direction and \( f_V \) takes the transpose of \( f_H \).

### 3.2 Cost aggregation

Classic cost aggregation methods are pixel-by-pixel, which leads to a significant amount of redundant computation and slow propagation of information. This paper proposes an efficient local stereo matching method using SLIC segmentation and adaptive EWMA filtering. This approach has two advantages. First, based on the weight matrix for obtaining a smooth matching cost, the EWMA filter can quickly expand the support window. Second, the SLIC segmentation can make similarly coloured or neighbouring pixels cluster together for matching. This not only makes it possible to resist the influence of noise to a certain extent, but also makes it possible to reduce its dimensionality from pixel-wise to superpixel-wise. These factors can significantly improve the computational efficiency and matching accuracy. Hence, the EWMA filter propagation in SLIC space is adopted for cost aggregation. A detailed description of the cost aggregation process is provided below.

Firstly, through the SLIC segmentation algorithm, an SLIC space is established to cluster pixels into superpixels based on the local intensity distribution. The initial cost matrix is aggregated in individual superpixels rather than a rectangle on the local intensity distribution. The initial cost matrix is set to zero and the values of the left column take the opposite number. Then, \( f_H \) is produced to detect the gradients in the horizontal direction and \( f_V \) takes the transpose of \( f_H \).

\[
C = S \times C_p
\]  

Through the mean filtering in each segment, some spurious local minima caused by noise interference disappear, while the true ones are retained [34].

SLIC segmentation algorithm [15] uses k-means to generate superpixels, which is faster and more memory efficient than other methods. Two significant principles involved here are that the space distance range is proportional to the superpixel size and that the weighted distance measurement combines normalised colour similarity and spatial closeness. Therefore, pixels with near position in colour and spatial space are efficiently clustered. The disparity maps generated from three cost aggregation strategies are shown in Figure 2. The first strategy only uses WTA on pixel cost, while the second one uses a combination of pixel cost and superpixel cost without the EWMA filter. The third one adds the EWMA filter based on the second strategy. It is noteworthy that many spurious points are removed owing to superpixel cost especially in low-texture regions, while the EWMA filter dramatically improves the performance of disparity maps owing to the cost aggregation in vast regions.

![Visual performance from various cost aggregation strategies](image)

**FIGURE 2** Visual performance from various cost aggregation strategies in (a) the initial image, (b) pixel cost, (c) combination of pixel and superpixel cost without EWMA filter, (d) combination of pixel and superpixel cost with EWMA filter. (From left to right: Adirondack, Tealby)

Secondly, cost aggregation is a critical step in maintaining the disparity boundary and smoothness. Distinct from global algorithms, the smoothness constraint is embedded into the weight matrix used for the EWMA filter. Further, to adaptively propagate the cost among specific regions according to image content, the weight matrix corresponding to the SLIC space is computed. The weight matrix is an indicator of the connection between neighbourhoods in the SLIC space. Adjacent superpixels with similar colours share prominent correlations. These correlations can be enhanced by applying the smoothness filter between them.

\[
w_{ij} = \exp \left( - \frac{\left\| \bar{x}_i - \bar{x}_j \right\|^2}{2 \times \sigma^2} + \frac{\left\| \bar{g}_i - \bar{g}_j \right\|^2}{2 \times \sigma^2_d} \right) \]

Here, \( \bar{x}_i = [L_{ix}, a_{ix}, b_{ix}] \) and \( \bar{x}_j = [L_{jx}, a_{jx}, b_{jx}] \) are the average values of pixels within superpixel \( i \) and \( j \), respectively, in CIE-Lab colour space; \( \sigma_x \) is the parameter that controls the decay rate of colour weight in Gaussian function; and \( \left\| \cdot \right\|^2 \) represents the L2 norm. Further, \( \bar{g}_i = [\bar{V}_{ix}, \bar{V}_{iy}] \), \( \bar{g}_j = [\bar{V}_{jx}, \bar{V}_{jy}] \) are the coordinates of barycenters of superpixel \( i \) and \( j \) in Euclidean space, and \( \sigma_d \) is the parameter that controls the decay rate of distance weight in Gaussian function. Based on the weight matrix between superpixels, the matching cost is passed through the EWMA...
FIGURE 3 Propagation of the EWMA in (a) the textureless region, (b) the uniform change region, and (c) the edge region (From left to right: initial image, support window after one iteration, support window after three iterations).

Filter in SLIC space.

\[ C_t(t + 1) = \epsilon \tilde{W} C_t(t) + (1 - \epsilon) C_0 \]  

(11)

For computational efficiency, the weight matrix is normalised to \( \tilde{W} \). \( C_t(t) \) is the cost matrix at \( t \) iterations, and \( \epsilon \) is the control probability. The weighted support windows following the propagation of the EWMA filter in three synthetic images are as shown in Figure 3, in which the red points represent the centre of the superpixel, and the black arrows represent the propagation direction at each iteration. Once the iterations are complete, the support window contributing to the central superpixel is constructed (marked in blue in Figures 3 and 4, and marked in grey in Figure 5b). The brighter superpixels have larger support weights). Furthermore, benefiting from the aggregation effect in SLIC-space, the propagation of the EWMA filter continues to effectively maintain its original diffusion trajectories (Figure 3) in the presence of severe Gaussian noise perturbation (\( \sigma = 64 \)). This indicates that the support window is uniformly expanding along with all directions around the central superpixel in flat regions, while the weight energy degenerates with increasing radial distance. As shown in Figure 5, a similar distribution can be observed in the edge region, the difference being that the diffusion window stops moving in the presence of rapid colour transition. Hence, the propagation strategy can quickly and accurately locate the adaptive weight window even with noise disturbance in typical scenes. As a result, adjacent segments with similar colours have homogeneous disparities, while others possess mutant disparities.

FIGURE 4 Propagation of the EWMA in (a) the textureless region, (b) the uniform change region, and (c) the edge region (From left to right: initial image corrupted with Gaussian noise (\( \sigma = 64 \)), support window after one iteration, support window after three iterations).

FIGURE 5 Propagation of the EWMA in (a) the typical regions and (b) the corresponding support window after 15 iterations.

Subsequently, for the sake of the performance of disparity computation, the superpixel-wise matching cost volume needs to be transformed. Essentially, the pixel-wise cost \( C_p^* \) is generated by the up-sampling projection of the pixel-wise cost \( C_s^* \) obtained by the EWMA filter.

\[ C_p^* = S^T \times C_s^* \]  

(12)

Finally, unlike in other methods [32, 33], the matching costs \( C \) was formed by combining the two costs from pixel-space and
SLIC-space,

\[ C(p, d) = (1 - \alpha) C^e_p(p, d) + \alpha C^f_p(p, d) \]  \hspace{1cm} (13)

Here, \( \alpha \) is the balance factor. Owing to the integration of the smooth superpixel and dependable pixel costs, the performance of the strategy is similar to that of semi-global algorithms. Consequently, the disparity preserves the discontinuity and smoothness, which outperforms other segment-based algorithms.

3.3 | Disparity selection

Following the cost combination, the final cost volume in the pixel-space is generated. Based on the characters belonging to the above-mentioned cost space, complex disparity computation methods are avoided in order to reduce the computational burden. Therefore, the disparity map is acquired only by the WTA strategy. The initial disparity map \( D \) is determined by the disparity range \( \hat{d} \) and matching cost \( C \).

\[ D(p) = \arg \min_{\hat{d} \in \hat{d}} C(p, d) \]  \hspace{1cm} (14)

3.4 | Disparity refinement

The initial disparity map obtained from the disparity selection still has errors that require detection and correction. Owing to the cost propagation in SLIC space, a mismatch often emerges as a cluster of pixels. In such cases, the commonly used median filter [8] and nearest interpolation \([35]\) approaches are not suitable. Therefore, plane fitting \([36]\) is used to refine the unstable disparity here. The left-right consistency check (LRC) is used to obtain the erroneous matching violating the coherence constraint:

\[ |D^{\text{left}}(p) - D^{\text{right}}(p - D^{\text{left}}(p))| < 1 \]  \hspace{1cm} (15)

Here, \( D^{\text{left}}(p) \) and \( D^{\text{right}}(p) \) are the disparity maps for the left and right images, respectively. Following the LRC operations, stable candidates are obtained for computing the plane parameters in segments. Furthermore, the arithmetic average error is used to measure the quality of plane fitting. Nevertheless, there are some segments with failed fitting due to the non-planarity of an object or improper segmentation. Smaller segmentation is necessary for additional fitting to further correct false disparities. Therefore, the segment size for iterative plane fitting is set to progress from large to small. Then, an optimised disparity map is produced.

4 | EXPERIMENTS

To verify the contributions of this paper, relevant experiments were designed. The proposed method is tested on a desktop computer with an Intel Core i7 3.2 GHz CPU. The implementation is done in MATLAB. The parameters in this paper are set as follows: \( \sigma_x = 5.0 \), \( \sigma_y = 2.2 \), \( \sigma_d = 1.5 \), \( \beta = 0.4 \), \( \sigma_d = 10.0 \), \( t = 15 \), and \( c = 0.9995 \). The superpixel size is set to 5 × 5. The superpixel number is adaptively calculated by dividing the pixel number with the superpixel size. As a result, the sizes of all superpixels fluctuates around a fixed size regardless of the variation in image resolution. Additionally, the Middlebury v.2 \([37]\) and the v.3 \([38]\) datasets were selected as the experimental datasets.

4.1 | Stability of the matching cost method under different luminance levels

The radiometric changes generated from fluctuations in the camera exposure as well as environmental illuminations are some of the primary factors that impact the matching precision. The 2006 dataset with third resolution utilised in the experiments included 21 stereo pairs, where each pair of images contained three different illuminances \((I_1, I_2, I_3)\), and each illumination contained three different exposures \((E_0, E_1, E_2)\). The left image was selected at exposure \(E_1\), while the right image was selected at exposures \(E_0, E_1\), and \(E_2\) under the different illuminances \(I_1, I_2, I_3\). The widely used matching cost functions were selected for comparison. This included combinations of the CT and GA, combinations of the GA and AD, the combination of the CT and AD, and the combination of the CT, AD, and GA. The results after cost aggregation \((\text{Section } 2.2)\) are presented in this section.

The visual performance of Flowerpots is shown in Figure 6. The red marks the errors, while the blue marks the errors where the proposed method has obtained accurate results. Notably, it is clear that the variability of the exposure is inversely related to the accuracy. Table 1 presents the quantitative results in detail. Among the methods tested, the proposed approach has the fewest errors in the slant plane and edges. It also can be observed that the combination of CT and AD ranks second, which provides a solid foundation for the cost function proposed herein. Moreover, the overall average mismatching rate of the proposed method is the lowest under different illumination conditions. This indicates that the proposed approach is more robust to exposure variation.

4.2 | Effectiveness on the Middlebury datasets

4.2.1 | Quantitative comparison between existing methods and the proposed method

Accuracy and computational efficiency are two key standards of stereo matching. The Middlebury 2014 training dataset with quarter resolution was chosen to evaluate the effectiveness of the method based on these factors. This dataset includes 15 typical scenarios of stereo pairs with ground truth. Several well-known matching methods were considered for comparison. The
FIGURE 6  Visual performance from various cost functions under different illumination conditions of Flowerpots in (a) the right image (I1 E0), (b) the right image (I1 E1), and (c) right image (I1 E2) (From left to right: left image (I1 E1), right image, GA and CT, GA and AD, CT and AD, combination of GA, CT and AD, proposed method)

TABLE 1  Error matching rate of various cost computations under different illuminances in all regions

| Accuracy [%] | GA + CT | GA + AD | CT + AD | GA + CT + AD | Proposed method |
|-------------|---------|---------|---------|--------------|-----------------|
| I1          |         |         |         |              |                 |
| E0          | 76.0    | 68.5    | 78.3    | 75.9         | 80.4            |
| E1          | 83.2    | 82.4    | 81.2    | 82.9         | 83.4            |
| E2          | 81.8    | 79.0    | 82.7    | 81.7         | 83.4            |
| I2          |         |         |         |              |                 |
| E0          | 81.4    | 75.9    | 83.5    | 81.3         | 85.1            |
| E1          | 85.7    | 84.6    | 83.3    | 85.4         | 86.4            |
| E2          | 79.0    | 76.6    | 80.7    | 79.0         | 81.4            |
| I3          |         |         |         |              |                 |
| E0          | 74.2    | 67.1    | 76.2    | 74.2         | 78.8            |
| E1          | 80.9    | 80.3    | 79.0    | 80.9         | 81.4            |
| E2          | 74.6    | 72.0    | 76.5    | 74.5         | 78.1            |
| Avg.error   | 79.7    | 76.3    | 80.2    | 79.5         | 82.0            |

Bold values represent the optimal results for each comparison.

box filter [4] is the fastest local algorithm, which sets a target for the runtime. The BF [24] is an important adaptive approach in stereo matching, while the GF [25] dramatically reduces the computational burden. All the three methods were applied to the cost volume in the aggregation step. The box filter was used to evaluate the efficiency of the proposed method, while BF and GF were chosen to evaluate the quality of the proposed method.

These methods are applied in the same software environment. The disparity maps (Figure 7) show that compared with other methods, in the Adirondack, Piano, and PlaytableP datasets, the proposed method shows good performance in occluded regions. On the other hand, in the Recycle, Shelves, and Vintage datasets, the proposed method shows superior performance on slanted planes with large low-texture regions. Moreover, our method accurately restores the disparity in repetitive patterns (the back of chair in Adirondack, the keys in Piano, and the keypads in Vintage). The overall results indicate that the combination of the EWMA filter and SLIC segment can prevent the influence of occlusion and improve accuracy in slanted planes, repetitive pattern, and low-texture regions. SLIC segmentation categorises pixels in different structures and the EWMA filter propagates information from reliable regions to unreliable ones. As a consequence, the disparity of various regions is properly preserved by the proposed framework.

The detailed accuracy ratio (see Table 2) reveals that compared to the BF [24] and GF [25], the algorithm proposed herein has achieved the optimal accuracy: 85.63% overall and 90.66% in the non-occluded regions. In the discontinuity regions, the accuracy of the proposed method is 16.72% higher than that of the GF, which is the only one method whose accuracy exceeds
FIGURE 7  Comparison of the disparity maps generated from different cost functions on the 2014 Middlebury datasets (from top to bottom: Adirondack, Motorcycle, Piano, PlaytableP, Recycle, Shelves, Vintage). Error pixels in the non-occluded and occluded regions are marked in red and green, respectively.

TABLE 2  Quantitative comparison between the typical methods and the proposed method

|                  | Box filter [4] (7.44 s) | BF [24] (162.8 s) | GF [25] (15.55 s) | Proposed method (7.01 s) |
|------------------|-------------------------|-------------------|-------------------|--------------------------|
|                  | All  | Nocc | Disc | All  | Nocc | Disc | All  | Nocc | Disc | All  | Nocc | Disc |
| Adirondack       | 88.14| 92.39| 41.69| 91.13| 95.11| 47.74| 91.26| 95.23| 47.95| 94.63| 97.47| 63.72|
| ArtL.            | 71.08| 80.89| 39.94| 78.47| 88.23| 47.50| 76.47| 86.77| 43.76| 78.05| 89.27| 42.43|
| Jadeplant        | 60.05| 74.18| 14.74| 67.40| 81.27| 22.92| 68.30| 83.79| 18.60| 72.39| 84.59| 33.23|
| Motorcycle       | 84.42| 89.16| 47.96| 85.74| 92.00| 37.66| 86.62| 92.62| 40.47| 90.37| 95.20| 53.24|
| MotorcycleE      | 84.62| 89.29| 48.79| 86.34| 92.43| 39.54| 86.98| 92.95| 41.14| 91.52| 96.02| 56.86|
| Piano            | 79.57| 84.15| 51.28| 83.17| 87.95| 32.75| 82.51| 87.01| 35.13| 88.82| 91.68| 58.77|
| PianoL.          | 70.43| 74.22| 30.52| 69.73| 73.71| 27.78| 70.32| 73.92| 32.30| 75.59| 77.68| 53.55|
| Pipes            | 70.88| 83.17| 15.34| 76.79| 88.82| 22.37| 76.63| 88.84| 21.42| 78.99| 90.61| 26.47|
| Playroom         | 73.12| 80.84| 29.17| 76.02| 83.76| 31.93| 76.06| 84.93| 25.61| 82.72| 89.16| 46.08|
| Playtable        | 74.34| 79.81| 31.89| 80.14| 85.27| 40.29| 79.96| 85.54| 36.61| 89.25| 93.91| 52.98|
| PlaytableP       | 79.93| 85.73| 34.23| 85.33| 90.51| 44.42| 85.00| 90.95| 38.07| 91.91| 94.94| 68.01|
| Recycle          | 87.23| 90.40| 48.90| 89.40| 93.11| 44.56| 89.47| 93.34| 42.52| 94.22| 96.41| 67.73|
| Shelves          | 68.12| 70.03| 43.85| 68.68| 71.45| 33.59| 68.53| 70.98| 37.40| 76.69| 78.14| 58.28|
| Teddy            | 87.33| 89.08| 73.09| 89.19| 92.82| 59.69| 88.81| 92.95| 55.13| 93.11| 95.30| 75.26|
| Vintage          | 76.10| 79.26| 42.18| 76.93| 80.65| 36.96| 78.94| 82.46| 41.18| 86.25| 89.49| 51.45|
| Avg. error       | 77.02| 82.84| 38.24| 80.30| 86.47| 37.98| 80.39| 86.82| 37.15| 85.63| 90.66| 53.87|
50%. The computations required also show that its runtime is far lower than that of others, even competing with that of the box filter.

The proposed method is further compared with related local algorithms on the Middlebury 2014Q. Bad 1.0 is set as the error threshold in non-occluded regions. SAS [32], SCA [33], and the proposed method are implemented with the SLIC algorithm. GF_EMA [39] employed the EWMA filter and guided filter for smoothing in the cost volume in pixel space.

The quantitative results are shown in Table 3. On the one hand, the proposed method performs well with higher accuracy and computational efficiency among the four algorithms. Compare to SAS and SCA, the running time is reduced by 97.94% and 96.38% respectively. On the other hand, our method is more stable with a standard deviation of 66.17% lower than GF_EMA.

4.2.2 Quantitative comparison between competitive methods and the proposed method

To further evaluate the potential of the proposed method in terms of efficiency and effectiveness, it is compared with competitive methods, including MSCS [40], LocalExp [41], DoG-Guided [42], and INTS [43], on the Middlebury 2014 dataset. These algorithms are the latest evolution of the classic local and global techniques, such as GF, SGM, and graph cuts. MSCS is a global disparity estimator that combines MeshStereo and cross-scale cost filtering for faster computation. LocalExp is a state-of-the-art algorithm that combines slanted patch matching and improved graph cuts with MC-CNN-acrt embedded. DoG-Guided is a local stereo matching algorithm that employs an adaptive GF, while INTS is a stereo matching algorithm based on a semi-global approach.

The quantitative results are shown in Table 4. The average precision percentage is listed, where the subscript numbers indicate the ranks. Bad 2.0 is set as the error threshold. It is easy to conclude that the proposed method ranks second among all the five algorithms considered. Although MSCS ranks first, it only achieves 1.08% higher accuracy in non-occluded regions and 0.59% in all regions as compared to the proposed method. However, in terms of time efficiency, MSCS is ten times slower than the proposed method. This is due to the convergence of the global energy function in the former approach. Similarly, the proposed method is 120 times faster than LocalExp, although they share the same average rank. Both DoGGuided and INTS are less efficient than the proposed method in terms of both accuracy and time consumption. These comparisons indicate that the proposed method not only achieves higher accuracy, but also significantly reduces the running time. This demonstrates that the present work enables extensive improvement in stereo matching performance.

4.2.3 Quantitative comparison on Middlebury v.2 and v.3 datasets

To verify the stability and robustness of the algorithm, four well-known algorithms, including SWSS [44], MGM [45], MSCS [40], and CS-PM [46], are tested on Middlebury v.2 datasets (31 stereo pairs with third resolution) and Middlebury v.3 datasets (28 stereo pairs with quarter resolution). Here, SWSS and CS-PM are local algorithms whereas MGM and MSCS are global algorithms.

The quantitative results are shown in Tables 5 and 6. Bad 1.0 and bad 2.0 in non-occluded regions are measured for v.2 and v.3 datasets, respectively. The proposed method achieves the first rank along with MSCS. However, although MSCS has a lower matching error than our method, its execution time is more than 10 times as long as ours. Furthermore, our method is faster and more accurate than SWSS, MGM, and CS-PM. The result demonstrates that the proposed method performs well in various scenarios in terms of stability and robustness.
Impact of different parameter settings

Different combinations of parameters can produce different disparity qualities. Therefore, to obtain better results, it is necessary to explore the effects of typical parameters. Based on the theory presented above, the number of iterations and the size of a superpixel are two key parameters. The experiments with various parameter values are conducted on the 2014 training datasets.

4.3.1 Effects of varying the superpixel size

The size of the superpixel k-value not only impacts the accuracy but also influences the computation efficiency. Since the superpixel size is inversely proportional to the superpixel number, the larger the superpixel size, the lesser is the computational load. The results of different superpixel sizes (varying from approximately $2 \times 2$ to $15 \times 15$) on the average mismatching ratio and time consumption is shown for all regions (see Figure 8a). It is noteworthy that for up to a moderate superpixel size, such as $7 \times 7$, the accuracy oscillates within a small range. Furthermore, the good performance is achieved (average accuracy of 90.66% and runtime of 6.94 s) when the superpixel size $k$ is $5 \times 5$.

4.3.2 Effects of the number of iterations

It is clear that the number of iterations directly impacts the support window of the EWMA filter. The greater the number of iterations, the wider the adaptive window. Meanwhile, the runtime increases. Here, the number of iterations was varied from 5 to 35, and the resulting impact on the average mismatching ratio and time consumption is shown for all regions (Figure 8b). The graph shows that in the non-occluded region, good performance is achieved (average accuracy of 90.01% and runtime of 6.71 s) with the number of iterations $t$ is 15.

5 CONCLUSION

This study aims to provide a local stereo matching method to more efficiently obtain an accurate disparity map. This is achieved by focusing on the cost computation and aggregation steps of stereo matching to quickly obtain the cost volume with appropriate support windows. By this approach, a novel algorithm is proposed based on an adaptive EWMA filter and SLIC segmentation algorithm. In the cost computation step, an adaptive strategy based on AD and CT is introduced. Then, during cost aggregation, the EWMA filter and the SLIC segmentation are integrated to tackle the problems of computing efficiency and increase smooth information acquisition. Experiments on the Middlebury 2006 dataset demonstrate that the proposed approach is robust against radiometric variations. The average matching error rate on 2014 Middlebury standard image pairs is 90.66%, which is superior to several representative local methods. Furthermore, the average runtime is 7.01 s, which is comparatively lower than that of other commonly used methods. Furthermore, the proposed method performs well on Middlebury v.2 and v.3 datasets.

The efficiency and effectiveness of the proposed approach show considerable potential for practical applications, such as the bokeh effect and auto focus in smartphones. However, the approach proposed in this paper has a weak discriminatory disparity in large flat areas with no valid information. Adding proper embodying constraints into EWMA is a possible solution that can be explored in future work. Furthermore, the software environment in which the proposed method is implemented is MATLAB. For practical applications, hardware acceleration could be adopted to reduce the runtime.

ACKNOWLEDGEMENT

This work was supported by the National Basic Research Program of China (2015CB352001) and the Shanghai leading academic discipline project (S30502).

AUTHOR CONTRIBUTIONS

S.Y.: Formal analysis; investigation; methodology; software; validation; writing-original draft; writing-review and editing. X.L.: Investigation; writing-original draft. Z.L.: Investigation; writing-original draft. G.S.: Project administration; supervision; writing-review and editing

ORCID

Shan Yang https://orcid.org/0000-0002-3057-8810

REFERENCES

1. Jonathan, T.B., et al.: Fast bilateral-space stereo for synthetic defocus. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, pp. 4466–4474 (2015)
2. Hamzah, R.A., et al.: Improvement of stereo matching algorithm for 3D surface reconstruction. Signal Process. Image Commun. 65, 165–172 (2018)
3. Zhao, X.W., et al.: Regional fuzzy binocular stereo matching algorithm based on global correlation coding for 3D measurement of rail surface. Optik 207, 164488. (2020)
4. Daniel, S., Richard, S.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int. J. Comput. Vis. 47(1/2/3), 7–42 (2002)
5. Ingemar, J.C., Sunita, L.H., Satis, B.R.: A maximum likelihood stereo algorithm. Comput. Vis. Image Underst. 63(3), 542–567 (1996)
6. Chuang, P., Yi, Q.L., Huang, D.P.: Novel belief propagation algorithm for stereo matching with a robust cost computation. IEEE Access 7, 29699–29708 (2019)
7. Boykov, Y.: Graph cuts and efficient N-D image segmentation. Int. J. Comput. Vis. 70(2), 109–131 (2006)
8. Heiko, H.: Stereo processing by semiglobal matching and mutual information. IEEE Trans. Pattern Anal. Mach. Intell. 30(2), 328–341 (2008)
9. Simon, H., Reinhard, K.: Iterative semi-global matching for robust drifter assistance systems. In: Asian Conference on Computer Vision, Okinawa, Japan, pp. 465–478 (2013)
10. Yuan, T., et al.: Segment-based disparity refinement with occlusion handling for stereo matching. IEEE Trans. Image Process. 28(8) 3885–3897 (2019)
11. Jure, Z., Yann, L.C.: Computing the stereo matching cost with a convolutional neural network. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Boston, pp. 1592–1599 (2015)
12. Li, Q., et al.: Learning deep correspondence through prior and posterior feature constancy. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, pp. 2811–2820 (2017)
13. Nikolaus, M., et al.: A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, pp. 4040–4048 (2016)
14. Pang, J.H., et al.: Cascade residual learning: A two-stage convolutional neural network for stereo matching. In: IEEE international Conference on Computer Vision Workshops, Venice, pp. 878–886 (2017)
15. Radhakrishna, A., et al.: SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell 34(11), 2274–2282 (2012)
16. Perry, M.B.: The exponentially weighted moving average. Wiley Encylopedia of Operations Research and Management Science, John Wiley & Sons, Hoboken (2010)
17. Zhu, H.M., Yin, J.H., Yuan, D.: SVCC: Segmentation volume combined with cost volume for stereo matching. IET Comput. Vis. 11(8), 733–745 (2017)
18. Liu, H., et al.: Improved cost computation and adaptive shape guided filter for local stereo matching of low texture stereo images. Appl. Sci. 10(5), 1869 (2020)
19. Veksler, O.: Fast variable window for stereo correspondence using integral images. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Madison, WI, USA, pp. 556–561 (2003)
20. Mei, X., et al.: On building an accurate stereo matching system on graphics hardware. In: Proceedings of IEEE international Conference on Computer Vision Workshops, Barcelona, Spain, pp. 467–474 (2011)
21. Zhang, K., Lu, J.B., Gauthier, L.: Cross-based local stereo matching using orthogonal integral images. IEEE Trans. Circuits Syst. Video Technol. 19(7), 1073–1079 (2009)
22. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: Fifth International Conference on Computer Vision, Prague, Czech Republic, pp. 839–846. (1998)
23. Yoon, K.J., Jeong, Y.K., Kweon, I.S.: Support aggregation via non-linear diffusion with disparity-dependent support-weights for stereo matching. In: Asian Conference on Computer Vision, Berlin, Heidelberg, pp. 25–36 (2009)
24. Yoon, K.J., Kweon, I.S.: Adaptive support-weight approach for correspondence search. IEEE Trans. Pattern Anal. Mach. Intell. 28(4), 650–656 (2006)
25. Hosni, A., et al.: Fast cost-volume filtering for visual correspondence and beyond. IEEE Trans. Pattern Anal. Mach. Intell 35(2), 504–511 (2013)
26. Toftis, C., Theoharidou, T.: High-quality real-time hardware stereo matching based on guided image filtering. In: Design, Automation, and Test in Europe Conference and Exhibition, New York, USA, pp. 1–6 (2014)
27. Chen, Y., et al.: Real-time hardware stereo matching using guided image filter. In: Proceedings of 2016 International Great Lakes Symposium on VLSI, Boston, MA, USA, pp. 18–20 (2016)
28. Yang, Q.X.: A non-local cost aggregation method for stereo matching. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, pp. 1402–1409 (2012)
29. Mei, X., Sun, X., Dong, WM.: Segment-tree based cost aggregation for stereo matching. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Portland, pp. 313–320 (2013)
30. Yao, P., et al.: Iterative color-depth MST cost aggregation for stereo matching. In: Proceedings of IEEE International Conference on Multimedia and Expo, Seattle, WA, USA pp. 1–6 (2016)
31. Xu, J.X., et al.: Texture discrimination-enforced matching cost computation and smoothness-weighted cost regularization for stereo matching. J. Electron. Imaging 28(5), 053025 (2019)
32. Tombari, F., Mattoccia, S., Stefano, F.D.: Segmentation-based adaptive support for accurate stereo correspondence. In: Proceedings of Second Pacific Rim Symposium, Santiago, Chile, pp. 427–438 (2007)
33. Chang, D.F., et al.: Accurate and fast segment-based cost aggregation algorithm for stereo matching. In: IEEE 19th International Workshop on Multimedia Signal Processing, Luon, pp. 1–6 (2017)
34. Tony, B., Bartomeu, C., Jean, M.M.: A non-local algorithm for image denoising. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Boston (2005)
35. Lee, S., Hong, H.: Near-real-time stereo matching method using both cross-based support regions in stereo views. Opt. Eng. 57(2), 023103 (2018)
36. Samprit, C., Ali, S.H.: Influential observations, high leverage points, and outliers in linear regression. Stat. Sci. 1(3), 408–412 (1986)
37. Heiko, H., Daniel, S.: Evaluation of cost functions for stereo matching. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, pp. 1319–1399 (2007)
38. Daniel, S., Heiko, H., York, K., et al.: High-resolution stereo datasets with subpixel-accurate ground truth. In: 36th German Conference on Pattern Recognition, Muenster, Germany, pp. 31–42 (2014)
39. Zhu, C.T., Chang, Y.Z.: Stereo matching for infrared images using guided filtering weighted by exponential moving average. IET Image Process. 14(5), 830–837 (2020)
40. Yao, P., et al.: MSCS: MeshStereo with cross-scale cost filtering for fast stereo matching. IET Comput. Vis. 12(6), 908–918 (2018)
41. Tatsunori, T., et al.: Continuous 3D label stereo matching using local expansion moves. IEEE Trans. Pattern Anal. Mach. Intell. 40(11), 2725–2739 (2018)
42. Masamichi, K., Ikako, S., Radim, S.: High accuracy local stereo matching using DoG scale map. In: Fifteenth IAPR International Conference on Machine Vision Applications, Nagoya, Japan, pp. 258–261 (2017)
43. Xu, H., Zhang, Y.J., Yue, Z.K.: Image-guided non-local dense matching with three-steps optimization. In: ISPRS Annals of the Photogrammetry, Prague, Czech Republic, pp. 67–74 (2016)
44. Li, X.S., et al.: Efficient methods using slanted support windows for slanted surfaces. IET Comput. Vis. 10, 384–391 (2016)
45. Gabriele, F., Carlo, D.F., Emic, M.: MGM: A significantly more global matching for stereovision. In: Proceedings of British Machine Vision Conference, Swansea, pp. 1–12 (2015)
46. Zhang, K., et al.: Cross-Scale cost aggregation for stereo matching. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, pp. 1590–1597 (2014)