Improvement of stereo matching algorithm based on guided filtering and Kernel Regression

Ying Shan Cao¹, Jin Gen Liu¹, Tie Xiang Wen², Xu Bi¹

¹School of Information Engineering, Wuhan University of Technology, Wuhan, Hubei, 430070, China
²Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, 518055, China

*359532404@qq.com, ljgxd@163.com, tx.wen@siat.ac.cn, 1160320250@qq.com

Abstract. Stereo matching is one of the most dynamic fields in computer vision. Though its relevant research has already stepped into a mature stage, there are still certain challenges to obtain real-time and high-precision disparity maps from stereo image pairs. This paper presents a novel local stereo matching algorithm with better performance in edge preserving. In the first stage, this paper measures matching cost through combining truncated absolute differences (TAD) of the color and gradient. In the cost aggregation stage, this paper is creatively combined the weighted guided filtering and adaptive steering kernel regression algorithm, which effectively preserves image edge and depth information. In the final stage, an adaptive steering kernel regression algorithm is employed in interpolation to refine the final disparity map. According to the Middlebury benchmark experiments, the algorithm proposed in this paper could have better performance than other local stereo matching algorithms.

1. Introduction
Stereo vision system belongs to the field of computer vision. It has a wide range of applications in 3D surface reconstruction[1], autonomous driving system navigation, virtual reality and so on. The stereo vision system consists of at least two cameras. The cameras in the system act as different perspectives for capturing the same scene, producing pairs of images with corresponding relationships[2]. Through the stereo matching algorithm, the correspondence between the pairs of images is established and a disparity map is generated [3].

The stereo matching algorithm is mainly divided into the global and local methods [4]. The global method determines the disparity value by minimizing energy function. The global methods includes dynamic programming (DP) [5], graph cut (GC) [6] and belief propagation (BP) [7]. The local methods are also known as region-based or window-based methods [8]. The global optimization methods have a large amount of computation and are less feasible in practical applications[9]. Compared with the global method, the local method has a less time-consuming. In recent years, the local method is continuously optimized, and achieved higher accuracy. The main steps of the local stereo matching algorithm as follows: 1. matching cost computation; 2. cost aggregation; 3. disparity computation; 4. post-processing and refinement [4].

Cost aggregation is the core of the local stereo matching steps. The simplest way to aggregate is to use low-pass filters in square window. However, this method tends to blur the boundaries[10]. In order to solve this problem, some improved cost aggregation method is proposed. Yoon et al.[11] proposed...
an adaptive support-weight approach. Richard et al.[12] proposed a real-time stereo matching algorithm which smoothing the matching costs with the guided filter. Zhu et al.[13] proposed an edge-preserving guided filter approach to aggregate the cost volume. Hong et al.[14] propose a new cost aggregation method based on weighted guided image filter for local stereo matching.

The purpose of the disparity map refinement is to reduce noise and improve the accuracy. Typical steps consist of occlusion detection and pixels filling or interpolation. The occlusion detection usually adopts left-right consistency check[15]. The final refinement step after interpolation usually uses median filter and bilateral filter.

This paper contributes a novel local stereo matching algorithm. There are two contributions as follow:

1. The algorithm proposed in this paper is improved on the basis of guided filtering. In the cost aggregation stage, the covariance matrix based on the local gray value is calculated by adaptive steering kernel regression algorithm[16]. The covariance matrix is used as the local weight matrix in the weighted guided filtering algorithm. This approach plays a role in edge-preserving, which shows better performance than other methods.

2. In the disparity refinement stage, the occlusion pixels are obtained by the left and right consistency detection, and the interpolation process is completed by using the adaptive steering kernel regression algorithm to obtain the disparity map.

2. The proposed algorithm

The proposed algorithm that consists of the following 5 stages: (1) In the pre-processing stage, using the adaptive steering kernel regression algorithm to filter the raw input image. This process, while effectively denoising, obtains the gradient image of the image in the horizontal and vertical directions, which will be used to calculate the matching cost. (2) In this stage, using the combined measurement to compute the matching costs. (3) In the cost aggregation stage, the covariance matrix is calculated by adaptive steering kernel regression algorithm and using the weighted guided filtering to aggregate the matching costs. (4) the ‘Winner-Take-All’ strategy is used to compute the initial disparity map. (5) In the disparity refinement process, the interpolation is performed using an adaptive steering kernel regression algorithm.

2.1. Matching cost function

Compared to using a single method to calculate the matching cost, a combination of methods can often achieve better performance. Image gradients serve as a popular matching cost function which contain rich structural image information. The gradient image is:

$$\nabla G = \left( \nabla I_x^2 + \nabla I_y^2 \right)^{1/2}$$

(1)

where $\nabla I_x, \nabla I_y$ are the gradient value of the horizontal and vertical directions respective. The TAD of the gradient is generally determined as:

$$C_{\text{gradient}}(p) = \min \left( |\nabla G_L(p) - \nabla G_R(p-d)|, T_g \right)$$

(2)

where $d$ are disparity values, $C_{\text{gradient}}(p)$ is the matching cost of the image gradient. $T_g$ is the gradient truncation value. The TAD of color is computed as:

$$C_{\text{color}}(p) = \min \left( \frac{1}{3} \sum_{C_{r,g,b}} |I_{L,C}(p) - I_{R,C}(p-d)|, T_c \right)$$

(3)

where $I_{L,C}(p)$ is the color intensity value of pixel $p$ that is the left image of one of the channels. Correspondingly, $I_{R,C}(p-d)$ is the color intensity value in the right image. $T_c$ is a threshold for color truncation.

The combination of matching costs is expressed as:


\[ C(p) = \alpha \cdot C_{\text{color}}(p) + (1 - \alpha) \cdot C_{\text{gradient}}(p) \]  (4)

where \( \alpha \) is a constant with a range of \([0,1]\).

2.2. Cost aggregation

Cost aggregation is the most important step in the stereo matching algorithm. This stage is mainly divided into two steps: 1. The covariance matrix based on the local gray value is calculated and used as the local weight matrix. 2. The weighted guided filter is used to aggregate the matching cost.

In the adaptive kernel regression, \( K_{H^d}^{steer}(x_d - x) \) is data-adapted kernel. It is determined by the pixel spatial position and pixel intensity, thus effectively protecting the boundary. \( H_d^{\text{steer}} \) called steering matrices. It can be expressed by:

\[ H_d^{\text{steer}} = h \mu_d C_d^{-\frac{1}{2}} \]  (5)

where \( C_d \) is covariance matrix based on differences in the local gray-values. It can be calculated from:

\[ C_d = \gamma_d U_{\theta_d} \Lambda_d U_{\theta_d}^T \]  (6)

where \( U_{\theta_d} \) is the rotation matrix, \( \Lambda_d \) is the expansion matrix and \( \gamma_d \) determine the scale parameters. Thus, the covariance matrix determines the shape of the local adapted kernel. These local adapted kernels can protect the image edges because it uses large windows in the smooth region and small windows in the depth discontinuities region. It is effective to improve the effect of guided filter that using the covariance matrix to calculate the weighted kernel.

Let \( G = [G_R, G_G, G_B]^T \) be the guidance color image, the matching cost to be filtered is \( z(x_d) \), the image obtained after filtering is \( z'(x_d) \). It is assumed that in each disparity \( d \), for the center point is \( k \), and the support window \( \omega_k \) with radius is \( r \), \( z'(x_d) \) can be expressed as:

\[ z'(x_d) = a_k \cdot G(x_d) + b_k \]  (7)

where \( a_k \) is a 3×1 coefficient vector, and \( b_k \) is scalar. The weighted guided filtering is an energy function minimization problem, \( a_k, b_k \) can be obtained by the following energy function:

\[
\arg \lim_{a_k,b_k} \left\{ E = \sum_{x \in a_k} (a_k \cdot G(x_d) + b_k - z(x_d))^2 + \frac{\tilde{\lambda}}{W_G(k)} \cdot a_k^2 \right\}
\]  (8)

where \( W_G(k) \) is a local weights matrix and \( \tilde{\lambda} \) is a regulation parameter. The parameters \( a_k, b_k \) are obtained based on solution of \( \nabla E = 0 \).

Therefore \( W^{\text{WG}} \) can be defined as:

\[
W^{\text{WG}} = \sum_{k \in a_k} \left\{ \frac{1}{|\omega_k|} \cdot \frac{1}{|\omega_{\gamma_d}|} \cdot \left[ 1 + (G(x_d) - \mu_k)^T \left( \Sigma_k + \tilde{\lambda} \cdot w_G(k)^{-1} \right)^{-1} (G(y_d) - \mu_k) \right] \right\}^{-1}
\]  (9)

It can be seen from the above equation that the pixel \( y_d \) has a larger weight, which in the neighborhood of \( x_d \) has a similar color intensity. The intensity of the weight change is controlled by the local weight matrix \( w_G(k) \), which is a 3×3 matrix defined by:
\(w_G(k) = \begin{pmatrix} w_G^R(k) & 0 & 0 \\ 0 & w_G^C(k) & 0 \\ 0 & 0 & w_G^G(k) \end{pmatrix}\) \(10\)

\(w_G^C(k)\) is the weight kernel of one of the color channels, which is represented by the covariance matrix calculated above. We calculate the covariance matrix of the guided image \(G\) in each color channel, and take the determinant of the covariance matrix:

\[w_G^C(k) = \frac{1}{\det(C_G^C(k)) + \epsilon}\] \(11\)

where \(\epsilon\) is a constant to prevent \(w_G(k)\) from being a singular matrix.

After the cost aggregation, the aggregated cost \(z'(x_d)\) at the disparity \(d\) is obtained.

### 2.3. Disparity computation

After the cost aggregation, for each pixel \(x_d\), using the ‘Winner-Take-All’ strategy. That is, the minimum value of all the disparity \(d\) is selected to obtain a disparity map, which is expressed as:

\[D(x) = \arg \lim_{d \in [d_{\text{min}}, d_{\text{max}}]} \{z'(x_d)\}\] \(12\)

### 2.4. Post-processing

The disparity map obtained by cost aggregation contains outliers. In order to obtain a high-precision disparity map, further processing is required. This part consists of two main steps: the left and right consistency check and the interpolation process.

The pixel will be determined as an outlier if it satisfies the following inequality:

\[|D(x) - D_R(x - D(x))| \geq d_l\] \(13\)

where \(D_R(x)\) is the disparity map obtained from the right image, and \(d_l\) is the disparity values.

We use adaptive steering kernel regression algorithm for interpolating, still adopt the Gaussian kernel.

\[K_{\text{steer}}^{\text{new}(x_d - x)} = \frac{\sqrt{\det(C_d)}}{2\pi h^2 \mu_d^2} \exp \left\{ -\frac{(x_d - x)^T C_d (x_d - x)}{2 h^2 \mu_d^2} \right\}\] \(14\)

Interpolation can generate streak-line artifacts, hence the iterative process of this method is used to smooth the disparity map.

### 3. Experimental results

For the proposed algorithm, four sets of standard stereo image pairs in the Middlebury stereo dataset were used to evaluate it, namely Tsukuba, Venus, Cones and Teddy. The method for evaluating the performance of the algorithm is the bad pixels percentage, which are evaluated from three aspects: all, non-occluded and depth discontinuities area. Defined as follows:

\[B = \frac{1}{N} \sum |D_e(x,y) - D_g(x,y)| > \varphi\] \(15\)

where \(D_e(x,y)\) is the disparity map obtained by the proposed algorithm, and \(D_g(x,y)\) is the ground-truth disparity map. \(\varphi\) is the threshold, which is taken as 1 in this paper. \(N\) is the total number of pixels. Compared with the AdaptWeight[11], CostAggr[12] and the WGF[14] algorithm, the proposed algorithm shows better performance and reduces the bad pixel percentage of the disparity map. The comparison result shows in Table 1.
Table 1. The bad pixel percentage results of Middlebury evaluation of the established algorithms.

| Algorithm       | Tsukuba | Venus | Teddy | Cones |
|-----------------|---------|-------|-------|-------|
|                 | All     | Disc  | Non-all | All     | Disc  | Non-all | All     | Disc  | Non-all | Average |
| Proposed algorithm | 1.64    | 7.44  | 1.25   | 0.38   | 2.41  | 0.20    | 10.22   | 14.85 | 5.63    | 8.21    | 2.70    | 5.21    |
| WGF [14]        | 1.79    | 6.55  | 1.37   | 0.48   | 2.40  | 0.21    | 8.51    | 7.69  | 2.29    | 11.2    | 15.5    | 5.98    | 5.33    |
| CostAggr[12]    | 1.85    | 7.61  | 1.51   | 0.39   | 2.42  | 0.20    | 11.8    | 6.16  | 16      | 8.24    | 7.66    | 2.71    | 5.54    |
| AdaptWeight[11] | 1.85    | 6.90  | 1.38   | 1.19   | 6.13  | 0.71    | 13.3    | 18.6  | 7.88    | 9.79    | 8.26    | 3.97    | 6.67    |

Figure 1 shows the disparity map generated by the proposed algorithm. By visually comparing the disparity map obtained by the proposed algorithm with the ground-truth disparity map, it can be seen that the matching percentage rate is high, and the matching effect in the depth discontinuous region is better.

Figure 1. The disparity maps of the proposed algorithm by using Tsukuba, Venus, Teddy, and Cones: (a) Left images, (b) Right images, (c) ground-truth images, (d) result for the proposed algorithm
4. Conclusion and future works
The local stereo matching algorithm proposed in this paper mainly improves the cost aggregation and disparity refinement. In the cost aggregation, this paper creatively combines the weighted guided filtering and adaptive steering kernel regression algorithm. In the disparity refinement, an iterative adaptive control kernel regression algorithm is used for interpolation and smoothing. The proposed method shows results in edge-awareness and denoising. In the experiment, the higher precision disparity map is obtained when compared with other local algorithms. In the future, we will develop and implement parallel programming of the proposed algorithm, which enables it to run on multiple graphics processing units (GPUs).

References
[1] Hamzah, R. A., Ibrahim, H., Hassan, A.H.A.. (2016) Stereo matching algorithm for 3D surface reconstruction based on triangulation principle. In: International Conference on Information Technology, Information Systems and Electrical Engineering. Yogyakarta. pp. 119–124.
[2] Hamzah, R. A., Ibrahim, H.. (2016)Literature survey on stereo vision disparity map algorithms. J. Sensors, 2016(2):1-23.
[3] Hansard, M., Lee, S., Choi, O.. (2012) Time-of-Flight Cameras: Principles, Methods and Applications. Springer Publishing, New York.
[4] Scharstein, D., Szeliski, R.. (2002) A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision, 47: 7–42.
[5] Bobick, A.F., Intille, S.S.. (1999) Large occlusion stereo. International Journal of Computer Vision, 33: 181–200.
[6] Boykov, Y., Veksler, O., Zabih, R.. (2001) Fast approximate energy minimization via graph cuts. Pattern Analysis & Machine Intelligence. 23: 1222-1239.
[7] Yang, Q.. (2015) Stereo matching using tree filtering. IEEE Trans Pattern Anal Mach Intell. 37(4):834-46
[8] Hamzah, R. A., Kadmin, A.F.. (2018) Improvement of stereo matching algorithm for 3D surface reconstruction. Signal Processing: Image Communication, 65 (2018): 165–172.
[9] Kordelas, G., Alexiadis, D., Dara, P.. (2015) Enhanced disparity estimation in stereo images. Image and Vision Computing. 35: 31-49.
[10] Yang, Q., Ji, P., Li, D.. (2014) Fast stereo matching using adaptive guided filtering. Image and Vision Computing. 32: 202-211.
[11] Yoon, K.J., Kweon, I.S.. (2005) Locally Adaptive Support-Weight Approach for Visual Correspondence Search. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. California. pp: 924-931.
[12] Rhemann, C., Hosni, A., Bleyer, M., Rother, C.. (2011) Fast cost-volume filtering for visual correspondence and beyond. In: IEEE Conference on Computer Vision and Pattern Recognition. California. Pp: 3017–3024.
[13] Zhu, S.Q., Wang, Z.. (2016) Edge-preserving guided filtering based cost aggregation for stereo. Journal of Visual Communication and Image Representation. 39: 107-119.
[14] Hoog, G.S., Kim, B.G.. (2017) A local stereo matching algorithm based on weighted guided image filtering for improving the generation of depth range images. Displays. 49: 80-87.
[15] Yang, Q., Wang, L. Yang, R.. (2009) Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling. IEEE Transactions on Pattern Analysis and Machine Intelligence. 31: 492-504.
[16] Takeda, H., Farsiu, S., Milanfar, P.. (2007) Kernel Regression for Image Processing and Reconstruction. IEEE Transactions on Image Processing. 16: 349 – 366.