An improved algorithm for disparity estimation of SGM stereo matching based on edge detection

Jian Zhang¹a*, Jing Huang²b

¹School of Information, Zhejiang Sci-Tech University, Hangzhou, Zhejiang, China
²School of Information, Zhejiang Sci-Tech University, Hangzhou, Zhejiang, China
*aemail: 201920601010@mails.zstu.edu.cn, bemail: syhj_syt@163.com

*Corresponding author’s email: 201920601010@mails.zstu.edu.cn

Abstract: In the fields of depth measurement and computer vision, generating sufficiently accurate disparity maps through various stereo matching algorithms is an important foundation. The SGM semi-global matching algorithm is one of the best stereo matching algorithms, but its filling accuracy is poor in the invalid parallax area. Therefore, on the original basis, an improved algorithm of SGM stereo matching disparity filling based on edge detection is proposed. The experiment uses a standard data set to verify the algorithm. The results show that compared with the original SGM stereo matching algorithm, the paper algorithm can achieve better disparity estimation results in the invalid disparity area.

1. Introduction
In the fields of depth measurement and computer vision, the use of binocular camera imaging and stereo matching to obtain depth information is important basic work. It plays an irreplaceable role in various scenes of stereo vision, such as three-dimensional reconstruction, spatial positioning, and automatic Driving, etc. The acquisition of accurate depth information first requires an accurate disparity map, and the way to obtain the disparity map is a stereo matching algorithm. Generally, stereo matching algorithms can be divided into two categories: local stereo matching algorithms and global stereo matching algorithms. In the local stereo matching algorithm, a matching cost calculation function is generally used to calculate the matching cost, and the point with the smallest matching cost is selected as the best matching point. The typical representative is the BM stereo matching algorithm, this type of algorithm is often very fast, but due to the influence of mismatching, therefore, the accuracy of the resulting disparity map is often the worst. The global algorithm focuses on the entire image, uses a global energy function to calculate the entire image, and applies to smooth constraints to it to improve the accuracy of the parallax. The classic global stereo matching algorithms include the stereo matching algorithm based on dynamic programming¹, the stereo matching algorithm based on confidence propagation², and the stereo matching algorithm based on plane constraints³. This type of method has high parallax accuracy, but requires a lot of calculations, takes a long time, and is difficult to meet the requirements of real-time scenes such as video. At the same time, it is also difficult to deploy on some embedded systems with low computing power. Based on them, Hirschmuller⁴ combining the advantages of these two methods, a semi-global stereo matching algorithm with high disparity accuracy and good real-time performance, SGM algorithm is proposed.
The SGM algorithm is divided into four parts: cost calculation, cost aggregation, disparity calculation, and disparity optimization. However, the disparity value obtained at this time contains invalid disparity. These invalid disparity values often form invalid disparity regions of different sizes and shapes. How to quickly and accurately estimate the disparity of these regions and fill them? Each pixel of the original image (usually the left image, that is, the image taken from the left camera) has a unique disparity value corresponding to it while maintaining the real-time performance and accuracy is an issue worth studying. Xie\cite{5} based on the SGM algorithm, edge detection and linear fitting are used to optimize the disparity map, and the weak texture and occlusion areas are obtained. For sure of effect, Ko\cite{6} uses the K-means clustering method for image segmentation, and then uses the disparity initial point for cost aggregation and disparity estimation. Liu\cite{7} used the combination of image segmentation and the original disparity map to construct a new energy function and optimized this function by merging smaller segmentation blocks. Based on the above research and considering the time complexity of the algorithm, an improved disparity filling algorithm for SGM stereo matching based on edge detection is proposed. By using edge detection and Ostu algorithm, the SGM stereo matching algorithm has a better disparity filling effect. Experiments show that the algorithm proposed in this paper has high matching accuracy on object edges and low-texture areas.

2. Materials and Methods

2.1. Matching cost calculation

The algorithm in this paper uses the cost calculation method based on the Census transform to calculate the matching cost. First specify a window of size n×m, Both n and m are odd numbers. By comparing the gray value of each pixel in the window with the gray value of the center pixel of the window, the Boolean value of the comparison result is formed into a bit string, and this bit string is regarded as the center pixel of the window. Census transform value $C_u$ such as the formula (1):

$$C_u(u,v) = \bigotimes_{i=0}^{n'} \bigotimes_{j=0}^{m'} \xi(I(u,v),I(u+i,v+j))$$

$n'$ and $m'$ is the largest integer less than or equal to half of n and m, $\bigotimes$ is the bit-by-bit concatenation operation of bits, The calculation $\xi$ is the formula (2):

$$\xi(x,y) = \begin{cases} 0 & \text{if } x \leq y \\ 1 & \text{if } x > y \end{cases}$$

After obtaining the Census transform values of two pixels in the left and right images, the matching cost of the two pixels is calculated by calculating the Hamming distance, as shown in formula (3):

$$C(u,v,d) = \text{Hamming}(C_u(u,v),C_v(u-d,v))$$

among them,d represents the difference between the x-axis coordinates of the two points, that is, the parallax. $C_u$ and $C_v$ are the Census transformation values of two points. The Hamming distance calculation method is to perform the OR operation on the bit strings bit by bit, and then count the number of bits that are not 1 in the OR operation result.

2.2. Cost aggregation

We learn from the idea of a global stereo matching algorithm and adopt a global energy optimization strategy. The global energy function used is the formula (4):

$$E(D) = \sum_p C(p,D_p) + \sum_{q \in N_p} P_1 \left| |D_p - D_q| = 1 \right| + \sum_{q \in N_p} P_2 \left| |D_p - D_q| > 1 \right|$$

Among them, the first item is a data item, $C$ represents the matching cost, which represents the accumulation of the matching costs of all pixels p when the disparity map is D. The second and third terms are smoothing terms, which represent the penalty for all pixels q in the neighborhood Np of point p. $P_1$ and $P_2$ respectively penalize the case where the disparity change of the neighboring pixels of the point p is small and large. The value of $P_2$ is dynamically adjusted by formula (5):
(5)

\[ P'_p = \frac{P'_p}{I_{pq} - I_{pq}} \quad P'_p > P_p \]

\( P'_p \) needs to set the initial value, but it should be greater than \( P_p \). \( I_{pq} \) and \( I_{pq} \) are the gray value of the two points, \( p \) and \( q \). To efficiently solve the optimization problem of formula (3), a path-based cost aggregation idea is adopted, as shown in formula (6):

\[
L_c(p, d) = C(p, d) + \min \left( \frac{L_c(p - r, d)}{L_c(p - r, d - i) + P_i} \right) - \min L_c(p - r, i)
\]

Among them, \( C(p, d) \) is the matching cost value, and the second term means that the value-added to the path cost takes the smallest cost among the three penalty cases; the third term is to ensure that the new path cost value will not be too high. This article chooses an eight-way aggregation mode.

2.3. Parallax optimization

2.3.1 Sub-pixel fitting

Sub-pixel fitting is a step that almost all stereo matching algorithms have. Its purpose is to improve the parallax accuracy to the sub-pixel level. It uses a one-dimensional quadratic fitting method, such as the formula (7) shows:

\[
d_{\text{sub}} = d + \frac{c_1 - c_2}{2(c_1 + c_2 - 2c_1)}
\]

\( d_{\text{sub}} \) is the sub-pixel disparity value, \( d \) and \( c_1 \) and \( c_2 \) are the disparity value and matching the cost of integer pixels, \( c_1 \) and \( c_2 \) are the cost values of the left and right sides.

2.3.2 Consistency Check

The principle of the consistency check is to obtain the right disparity map, calculate the corresponding point in the right disparity map for each pixel of the left disparity map, and then determine whether the disparity value of these two disparity values is less than a threshold. If it exceeds the threshold, the disparity of the corresponding positions of these two points will be set to an invalid value. Among them, in order to quickly obtain the right disparity map, we directly use the matching cost of the left image to calculate the matching cost of the right image, thereby calculating the disparity map of the right image. The specific method is to calculate the corresponding pixel position of the left image as \((i + d, j)\) for the pixels \((i, j)\) of the right image according to the disparity value \(d\), and then put the pixels of the left image \((i, j + d)\) Similarly, the cost value under the disparity value \(d\) is taken out and assigned to \((i, j, d)\).

2.3.3 Uniqueness constraint

The parallax of each pixel should satisfy a principle: the cost value corresponding to this parallax must be the only one with the smallest cost value among all the candidate parallaxes. In actual operation, if the two minimum-cost values of a certain pixel are less than a threshold, the disparity value of this pixel is invalidated.

2.3.4 Remove small connected domains

Some mismatched pixels are often connected into small areas. In order to determine these pixels, the breadth-first area tracking algorithm is used to find these connected domains, and all pixels in the connected domains whose area exceeds a certain threshold are invalidated.

2.3.5 Median Filtering

As a typical image smoothing algorithm, median filtering is used to remove isolated points in the disparity map while filling in small disparity holes.
2.4. Parallax filling
The disparity filling part is a disparity invalid region estimation algorithm based on edge detection and the Ostu algorithm. The flowchart of the algorithm is shown in the Figure 1:

![Figure 1. Parallax filling flowchart](image)

2.4.1 Edge detection
The edge detection uses the canny algorithm. First, perform image grayscale and median filtering on the original image to avoid the influence of noise on the edge detection of the image, and then use the Sobel operator to calculate the horizontal gradient Gx and the vertical gradient Gy, as shown in formulas (8) and (9):

\[
G_x = \begin{bmatrix}
-1 & 0 & +1 \\
-2 & 0 & +2 \\
-1 & 0 & +1 \\
\end{bmatrix} \ast I
\]
\[
G_y = \begin{bmatrix}
-1 & -2 & -1 \\
0 & 0 & 0 \\
+1 & +2 & +1 \\
\end{bmatrix} \ast I
\]

(8)

(9)

Among them, the constant matrix is the Sobel operator, and I is the image window. From this, the size and direction of the gradient of each pixel can be obtained, as shown in the formula (10), (11):

\[
G = \sqrt{G_x^2 + G_y^2}
\]
\[
\theta = \arctan\left(\frac{G_y}{G_x}\right)
\]

(10)

(11)

After obtaining the gradient size and direction of each pixel, traverse each pixel to determine whether the gradient size of the pixel is the local maximum in its neighborhood in the gradient direction of the pixel. If not, set the value of the pixel to 0.
Finally, set two thresholds $t_{min}$ and $t_{max}$. Then all pixels are traversed, the pixels with gradient size less than $t_{min}$ are set to 0, and the pixels with gradient size greater than $t_{max}$ are determined edge pixels. For the pixels between the two, if there is a connection with the determined edge pixel, keep it; otherwise, set it to 0.

2.4.2 Neighborhood detection
After obtaining the disparity map after disparity optimization, all its pixels are traversed, and if it is an invalid disparity pixel, then disparity filling is performed. For a pixel that needs to be filled with parallax, the nearest neighbors in eight directions are used to form a set of parallax points, as shown in Figure 2:

![Figure 2. Schematic diagram of invalid parallax point obtaining neighborhood value](image)

They are black squares represent invalid parallax pixels, and white squares represent valid parallax pixels. For the obtained parallax point set, if there are edge pixels less than a certain threshold $T_h$, these points are eliminated, otherwise, the Ostu algorithm is used to divide these points into the front scenic spot set and the background point set, and then the front scenic spot set is eliminated.

The specific method is to first arrange all the pixels according to the parallax value from small to large, calculate the mean value and select a certain dividing point from the middle, divide it into two groups $C_1$ and $C_2$, and the number of pixels in each group are $n_1$ and $n_2$, respectively, and the mean values within the group are $m_1$ and $m_2$, respectively. Then there is the formula (12) when calculating the variance between classes:

$$
\sigma^2 = \frac{1}{n_1} (m_1 - m) + \frac{1}{n_2} (m_2 - m)^2
$$

2.4.3 Disparity estimation
First, two methods for judging invalid points of parallax as occluded points are proposed: one is to judge invalid pixels through the previous parallax optimization operation, and the other is to judge by parallax checking method. The specific method is that if the matching pixel point of a left image pixel $p$ on the right image is point $q$, the disparity between the two is $d$, and if the disparity value of point $q$ on the right disparity map is $d'$, pass $d'$, finds the matching $p'$ on the left disparity map and obtains the disparity value $d_1$ of this point. If $d_1 > d$, then point $p$ is the occlusion point.

Then perform parallax filling. After collecting the surrounding effective disparity values, for all the collected disparity values, according to whether it is an occluded point, the average value or the median value is taken as the estimated disparity value of the point.
3. Results & Discussion

In order to test the effect of the algorithm in this paper, we select four groups of images from the standard stereo images provided in the stereo matching algorithm test platform Middle-bury to test the algorithm. The four sets of images (Cloth3, Cone, Reindeer, Wood2) are shown in Figure 3.

![Figure 3](image)

At the same time, in order to reflect the accuracy of the disparity filling, we use the standard disparity map as the basis to show the average accuracy of the disparity filling, as shown in Table 1:

| algorithm | Cloth3 | Cone     | Reindeer | Wood2 |
|-----------|--------|----------|----------|-------|
| SGM       | 0.8735 | 0.9138   | 0.7889   | 0.9442|
| Algorithm | 0.8912 | 0.9283   | 0.8025   | 0.9493|

In order to observe the time cost of the algorithm in this paper, we have made statistics on the time required for the algorithm to run. It should be pointed out that the experimental platform of this experiment is the PC platform installed with the Windows 10 operating system, the processor is AMD R7-4800H, the Visual Studio 2017 is used as the program code design and operation platform, and the version of the opencv computer vision library used The number is 3.4.1. The program runs in a single-threaded mode and does not use multi-threaded programming or GPU acceleration technology to improve the calculation speed of the algorithm. Table 2 shows our test results.
Table 2. Calculation speed of stereo matching (unit: second)

| Algorithm | Cloth3 | Cone | Reindeer | Wood2 |
|-----------|--------|------|----------|-------|
| SGM       | 2.513  | 0.612| 2.696    | 2.553 |
| Algorithm | 2.796  | 0.701| 3.116    | 2.934 |

On the whole, because the influence of the edge area on the disparity estimation is considered, the invalid disparity point is filled with a more scientific neighborhood value when the disparity is filled, so that the algorithm in this paper will not significantly increase the original algorithm time compared to the original SGM algorithm. In the case of cost, the filled disparity map has a more accurate effect in the edge area.

4. Conclusions
This paper proposes an improved algorithm for the disparity filling algorithm in the SGM stereo matching algorithm based on edge detection. Based on the original SGM algorithm, classify the different situations of invalid disparity points, then take the edge features of the image into account, and finally use the maximum between-class variance method for grouping and finally obtain the final filling value. We conduct comparative experiments on the four sets of standard images provided by Middle-bury. Experimental results show that compared with the original SGM algorithm, this algorithm better preserves the edges and improves the filling accuracy without significantly increasing the time cost of the algorithm.

Acknowledgments
This work was supported by the funded project "Key R&D Program of Zhejiang Province" (2021C01048).

References
[1] Veksler O. (2005) Stereo correspondence by dynamic programming on a tree. J. Proc Cvpr IEEE, 2: 384-390.
[2] De H, Pf F. (2006) Efficient belief propagation for early vision. J. International Journal of Computer Vision, 70(1): 41-54.
[3] Chen H., Wang L., Liu G. (2020) Review of the stereo matching algorithm. J. High tech communication, 30 (2): 157-165
[4] Hirschmuller H.(2008) Stereo Processing by Semiglobal Matching and Mutual Information.J. IEEE. Transactions on Pattern Analysis and Machine Intelligence,30(2):328-341.
[5] Xie Y,Zeng S,Chen Let al. (2014)A Novel Disparity Refinement Method Based on SemiGlobal Matching Algorithm.C. International Conference on Data Mining:1135-1142.
[6] Stentoumis C，Karkalou E，Karras G C et al. (2014) A review and evaluation of penalty functions for Semi-Global Matching. C. International Conference on Intelligent Computer Communication and Processing: 167-172.
[7] Ko E, Ho Y. (2016) Disparity Map estimation using semi-global matching based on image segmentation. C. Asia Pacific Signal and Information Processing Association Annual Summit and Conference: 1-4.