

# Stereo Matching Algorithm Based On Integration Of Multi-Measurement

Qiang Wu<sup>1, a</sup>, Yuxiang Li<sup>2, b</sup>, Shinan Lang<sup>3, c</sup>

<sup>1</sup>Electronic Information and Control Engineering, Beijing University of Technology  
Beijing, 100124, China

<sup>2</sup> Electronic Information and Control Engineering, Beijing University of Technology  
Beijing, 100124, China

<sup>3</sup> Electronic Information and Control Engineering, Beijing University of Technology  
Beijing, 100124, China

<sup>a</sup>email: wuqiang@bjut.edu.cn, <sup>b</sup>email:13426224499@163.com, <sup>c</sup>email:langshinan@bjut.edu.cn

**Keywords:** Stereo matching, DSM , Census transform

**Abstract.** This paper presents a new measure method of stereo matching algorithm, which is based on image structure and pixel's feature, named CGDSM. The matching cost calculate step consist of three similarity measure factors: pixel gradient similarity measure factor, pixel feature similarity measure factor and image structural similarity measure factor. At the cost aggregate step we use adaptive window to aggregate which is based on Canny edge information. At the disparity calculate step we use WTA(winner takes all) algorithm to find the optimal disparity. At the disparity refine step we use LRC(left-right consistency) method to refine disparity. The new algorithm contain more similarity measure function from image gradient information to color domain feature information and then to the spatial structural information. Experiment shows that this algorithm enhance the accuracy of stereo matching algorithm.

## Introduction

In recent years mankind has been committed to making a computer to have human-like flexible visual function to automatically recognize and perceive surrounding environment through computer internal processing . By studying the human visual system, scholars found that using multiple cameras to simulate human eyes and comparing the difference of the image can easily obtain depth information. This process is called computer stereo vision. One of the stereo vision system is binocular stereo vision system using two cameras as information collector which has low cost, low power consumption advantages. A complete binocular vision process typically has five steps: image acquisition, camera calibration, image correction, stereo matching, depth extraction. The stereo matching process is the most complicated and critical step. By studying a large number of stereo matching algorithm, Scharstein[1] divided dense stereo matching algorithm into four steps: matching cost calculation, cost aggregation, disparity calculation and disparity refinement.

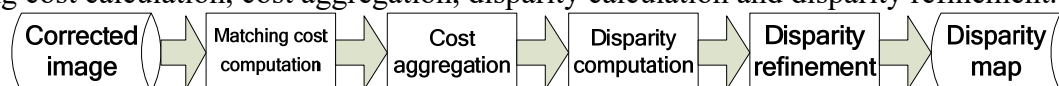


Figure 1. The process of stereo matching

The target of matching cost calculate step is to select the matched point in reference image from target image. Because of matching cost calculation is the first step of stereo matching algorithm, matching errors will continually accumulate to the subsequent step. Whether the selected cost calculate method is appropriate will directly affect the accuracy of the results. Usually when we select the matching cost, the more image information considered, the more accurate the match results will be. In the process of the matching cost calculating step, we should use similarity measure function to assess the local correlation of the matched points.

Similarity measure function is mainly divided into three categories: The first is based on color distance, such as: sum of absolute differences (SAD), absolute differences (AD).The second is

based on the correlative information, such as: Normalized Cross-Correlation(NCC). The third is based on the non-parametric conversion proposed by Zabih[2] , such as: Census Transformation (CT), Rank Transformation (RT).

This paper presents a new matching cost measure method:CGDSM, which is based on image structure and pixel's feature information. This method contains three similarity measure factors: We use Sobel operator to calculate the image gradient information. After gradient information extracted, the image edge will become apparent and the matching result at the edge region will be more accurate. We use DSM algorithm to calculate the feature information of the pixel which can enhance the accuracy of stereo match in rich feature region. We use improved three mode Census transformation to calculate the Image structure information. This improved algorithm is robust to brightness and the linear change.

We use adaptive size windows for aggregate calculation, and WTA (winner-take-all) to get an initial disparity map. In the step of disparity refinement, we use LRC to get mismatching points and use background filling method to rectify the errors.

### Matching Cost Function

**Sobel edge detection:** The matching precision in depth discontinuity area is one of the problems faced by stereo matching algorithm. Usually the image depth discontinuity area is the junction between foreground and background, which belongs to the edge portion of the image. The edge detection technique based on Sobel operator is the underlying technology of image processing. Sobel operator contains two 3x3 matrix, respectively the horizontal and vertical templates shown in figure 2.

$$G_x = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \qquad G_y = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

a) A filter based on x direction      b) A filter based on y direction

Figure 2. Sobel operator

After obtaining horizontal and vertical gradient value, the gradient of the image is calculated by the following formula.

$$G = \sqrt{G_x^2 + G_y^2} \tag{1}$$

The following figure shows the gradient image detected by Sobel operator. It is obvious that the edge regions of the image has been enhanced.

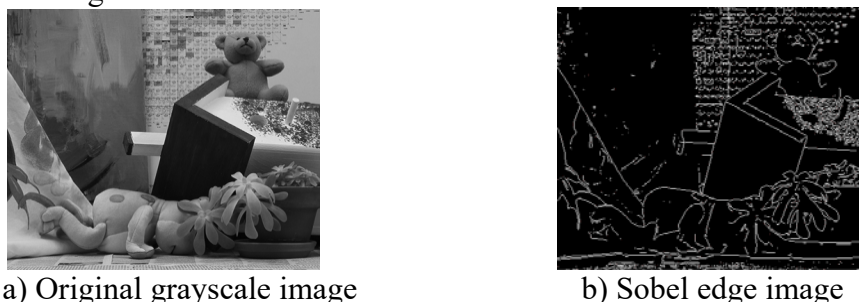


Figure 3. Sobel edge detection

**DSM Cost Computing Operator:** Yoon[3] proposed a pixel feature similarity measurement method in 2007 named DSM(Distinctive Similarity Measure). This method take pixel feature into consideration when calculating the matching cost. Yoon thinks that the matching accuracy not only depend on the similarity of two matching points but also depend on the pixel's feature.

The DSM algorithm consists of two parts, one is extracting pixel's feature, the other is establishing pixel's similarity model.

1. Extracting pixel's feature:  
The pixel's feature is calculated as follows:

$$D(p) = \min_{q_N \in N_p, q_N \neq p} f(p, q_N) \quad (2)$$

$D(p)$  represents a pixel's feature,  $f(p, q_N)$  represents the different between  $p$  and  $q_N$ ,  $N_p$  indicates the size of the supporting window. For the corrected image, the supporting window is one-dimensional. If the disparity search range is  $d \in [d_{\min}, d_{\max}]$ , The  $N_p$  calculate formula is as follows:

$$N_p = \{p + d \mid d_{\min} - d_{\max} \leq d \leq d_{\max} - d_{\min}\} \quad (3)$$

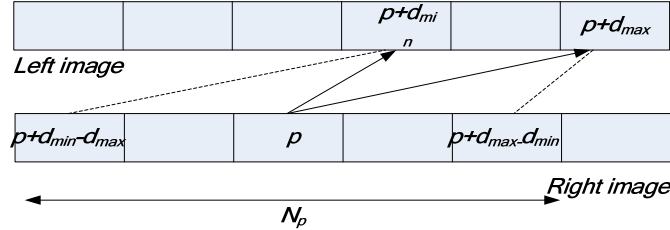


Figure 4. The search range of pixel feature

## 2. Establishing pixel's similarity model:

The similarity measure formula based on pixel's feature is shown as follows:

$$C_{DSM}(p_l, p_r) = \frac{f(p_l, p_r)}{D(p_l) \times D(p_r)} \quad (4)$$

$C_{DSM}(p_l, p_r)$  indicate the results of similarity measurement,  $f(p_l, p_r)$  indicate a basic similarity measurement,  $D(p_l)$  and  $D(p_r)$  respectively present the pixel's features in refer image and target image. The more similar and obvious of two pixels and respective feature, the smaller the similarity measure results will be. In this paper, we use TAD algorithm to calculate the  $f(p_l, p_r)$  and  $D(p_l)$ . Generally the threshold value in TAD equal to 33.

$$f(p_l, p_r) = TAD(p_l, p_r) \quad (5)$$

**Improved Census Transformation:** Traditional Census transformation treat supporting window's central pixel as reference pixel, then obtain the binary value of each pixel in the neighborhood, which is compared with the reference value. If this value is greater than reference value, its binary value is 1, otherwise its binary values is 0. Finally using Hamming distance measurement judge the similarity of two binary code set. If the central pixel is  $p$ ,  $q_N$  represents the remaining pixels in the supporting window  $N$ .  $B_{q_N}$  indicate a binary value of a single pixel after transform.  $T_{C_p}$  represent a binary code set of pixels within the supporting window. Census transformation of point  $p$  as shown in formula 6.

$$B_{q_N} = \begin{cases} 0 & I_{q_N} \leq I_p \\ 1 & I_{q_N} > I_p \end{cases}, T_{C_p} = \bigcup_{q_N \in N_p} B_{q_N} \quad (6)$$

In 1950, Hamming proposed a new measurement to evaluate the similarity of the two pixels after Census transform. The smaller of the Hamming distance, the higher similar the two pixels are. The Census similarity measure formula based on the Hamming distance is shown as follows:

$$C_{Census}(p_l, p_r) = T_{C_{p_l}} \otimes T_{C_{p_r}} \quad (7)$$

$\otimes$  means XOR operation,  $T_{C_{p_l}}$  and  $T_{C_{p_r}}$  present the binary set of refer image and target image calculated by Census transform.

In reality, photographing process will inevitably introduce salt or pepper noise which can lead to center pixel distorted. So the traditional Census transform could easily bring about mismatch.

In this paper, we adopt an improved Census transform to get the structural information which is proposed by Bernhard Froba[4]. Improved Census transformation use the mean value of the pixels in the supporting window as refer pixel instead of using center pixel. The improved algorithm will alleviate the mismatching phenomenon caused by noise to a certain extent.

If  $I_{mean}$  means the pixel gray value, improved Census transform formula is shown as follows:

$$B_{q_N} = \begin{cases} 0 & I_{q_N} \leq I_{mean} \\ 1 & I_{q_N} > I_{mean} \end{cases}, T_{C_p} = \bigcup_{q_N \in N_p} B_{q_N} \quad (8)$$

**Matching Cost Function:** We synthetically consider the gradient information, the pixel's structural information and the color domain information then propose a new matching cost calculation named CGDSM. The new matching cost function includes the following three aspects: image gradient information extracted by Sobel operator, pixel structural information calculated by improved Census transformation, color domain information described by DSM algorithm. The presented matching cost calculate formula is as follows:

$$C_{CGDSM}(p, d) = C_{Census}(p, d) + \alpha C_{DSM}(p, d) + \beta C_{grad}(p, d) \quad (9)$$

$\alpha$  and  $\beta$  are weight parameters, usually  $\alpha$  is 0.5 and  $\beta$  is 0.01. When the two pixels matched, the final result calculated by GDSM algorithm will be minimum.

### Cost Aggregation

**Canny Edge Detection:** Canny edge detect method is a basic edge detection algorithm proposed by John F. Canny[5] in 1986. Canny edge detection has four steps. First, Canny algorithm calculate the gradient information and get the real edge information by finding the local maximum gradient. Then using double threshold determine the edge belongs to which kind of edge, the strong edge or the weak edge. Finally connect the weak edges to the strong edges forming a closed edge image.

In this paper, we use Sobel gradient image as original gradient image and then use eight-neighbor non-maxima suppression algorithm suppress the gradient image. After the non-maxima suppression step, gradient width refined into one pixel wide and retain the complete edge information. It requires two threshold in Canny algorithm to reduce the number of false edges.  $T_l$  indicate the lower threshold and  $T_h$  indicate the higher threshold. We believe that the edge belongs to strong edge if its gray value higher than  $T_h$ , the edge belongs to weak edge if its gray value between  $T_h$  and  $T_l$ , the edge will be discarded if its gray value lower than  $T_l$ .

Tsukuba' edge image shown as below:

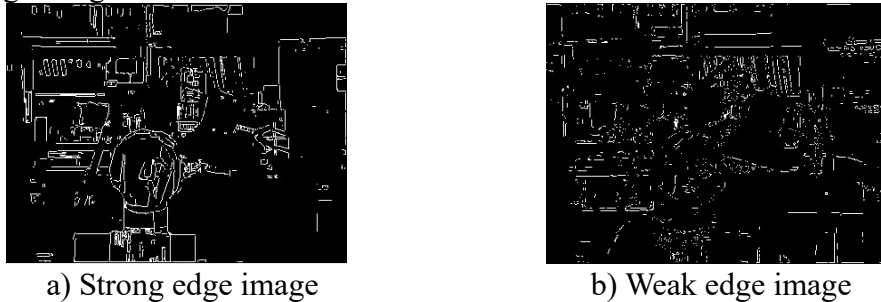


Figure 5. Canny edge detection

**Adaptive Window Cost Aggregation Based On Canny Edge:** Traditional cost aggregation module commonly use fixed window. Using this kind of window will greatly reduce the accuracy of the aggregated results in weak feature region and edge region, even if the matching speed is very fast. To solve this problem, we use one of the three sizes of adaptive windows selected by Canny edge information to aggregate the cost. At the Strong edge region we choose the smallest size of the window, at the weak edge region we choose medium size of the window, at the non-edge region we choose the biggest size of the window. Assuming three window size are  $a_1, a_2, a_3 (a_1 > a_2 > a_3)$ , the following equation shows the adaptive window cost aggregate method.

$$C_{aggr}(p, d) = \frac{1}{N} \sum_{q_w \in W} C_{CDSM}(q_w, d), W \in a_1, a_2, a_3 \quad (10)$$

$C_{aggr}(p, d)$  is the new matching cost after aggregation,  $N$  indicates the number of pixels in supporting window.

## Disparity Computation

We Apply WTA (winner-takes-all) query strategy to get the best initial disparity map . The formula of WTA strategy is as follows:

$$D(x, y) = \arg \min_{d \in S_d} C_{agg}(x, y, d) \quad (11)$$

$S_d = \{d_{\min}, \dots, d_{\max}\}$  are disparity value for all possible.

## Disparity Refinement

In the disparity refinement step, we apply LRC algorithm to detect all the mismatching points. LRC algorithm described as follow:

$$d(p) - d(q) > \delta \quad (12)$$

$d(p)$  and  $d(q)$  mean the disparity of pixel  $p$  and  $q$ .  $p$  and  $q$  are a pair of matching points.  $\delta$  represent the mismatching threshold equal to 1. When the disparity of two matching points greater than the threshold, we believe that these two pixels are mismatching points.

Those mismatching points usually belong to occluded areas which is caused by the foreground objects blocking the view of the background area . So we directly use background disparity value to fill the occluded areas.

## RESULTS AND EVALUATION OF ALGORITHMS

The matching effect of stereo matching algorithm is usually assessed by the network evaluation platform <http://vision.middlebury.edu>. This platform is raised and maintained by Scharstein et al. Scharstein use mismatching rate as algorithm evaluation criteria, which can be computed as follows:

$$B = \frac{1}{Num_{(x,y)}} \sum (|d_c(x, y) - d_t(x, y)| > \delta_d) \quad (13)$$

The following graphics show the final disparity image calculated by CGDSM and CDSM:

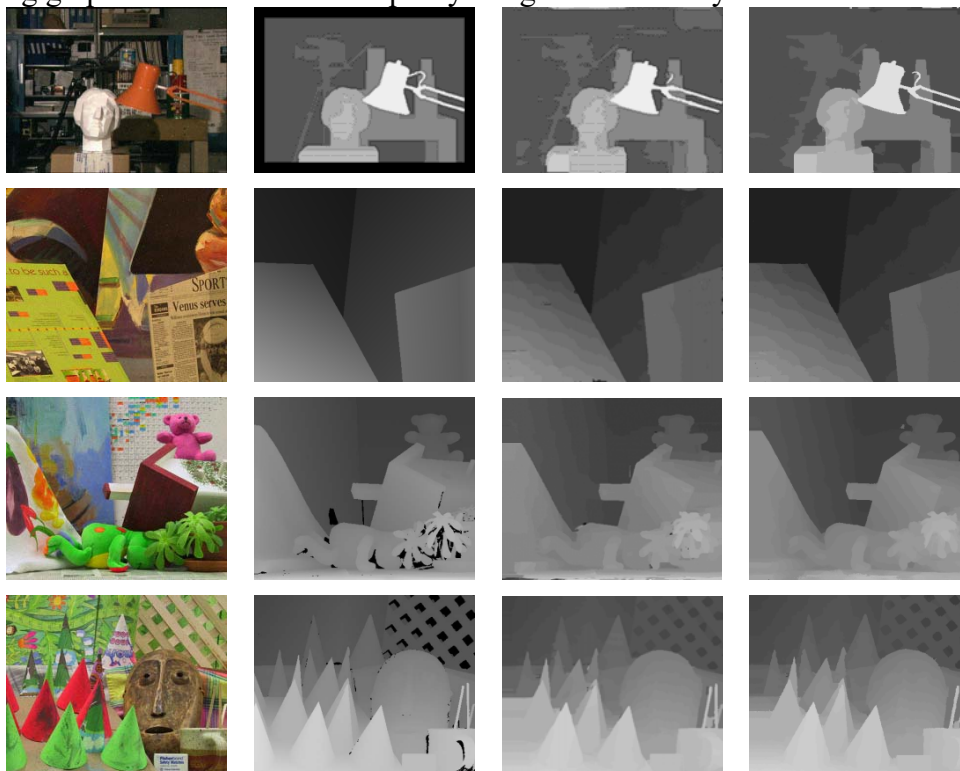


Figure 6. The experimental results of traditional CDSM and CGDSM algorithm

Table I. The Performance Comparison Of CDSM And CGDSM Algorithm

| Algorithm | Tsukuba |      |      | Venus  |      |      | Teddy   |      |      | Cones   |      |      | Avg. (%) |
|-----------|---------|------|------|--------|------|------|---------|------|------|---------|------|------|----------|
|           | non-occ | all  | disc | nonocc | all  | disc | non-occ | all  | disc | non-occ | all  | disc |          |
| DSM       | 1.21    | 1.75 | 6.39 | 0.35   | 0.44 | 1.32 | 5.45    | 11   | 17.1 | 2.31    | 9.91 | 8.83 | 5.51     |
| CGDSM     | 1.19    | 2.25 | 6.36 | 0.15   | 0.38 | 1.15 | 5.32    | 10.4 | 13.5 | 2.11    | 7.31 | 5.33 | 4.62     |

Table 1 and figure 6 show the comparison of disparity map between the DSM algorithm and CGDSM algorithm based on stereo matching algorithm criteria and describes the false match rate and the average error rate in non-occluded region , whole image and discontinuous region

From the experimental result and comparison chart, we can prove that multi-measure combined algorithm CGDSM effectively reduce the false matching rate and improve the accuracy of stereo matching algorithm. The mismatching rate of our algorithm is 4.62%, which is smaller than the traditional DSM algorithm .

## CONCLUSION

In this paper we propose a new stereo matching algorithm which is based on the image structure and pixel feature (CGDSM). The core of the presented algorithm is the improvement of matching cost calculation. The matching cost is consist of three separate similarity measure function : gradient information extracted by Sobel operator , DSM pixel feature information and local structure information described by improved Census transformation.

Through standard image provided by image library verify that the new matching algorithm has greatly improved the matching accuracy compared with a single measure stereo matching algorithm. The following figure shows the algorithm flowchart. The follow-up work is to transplant the new algorithm to DSP for the purpose of real-time performance.

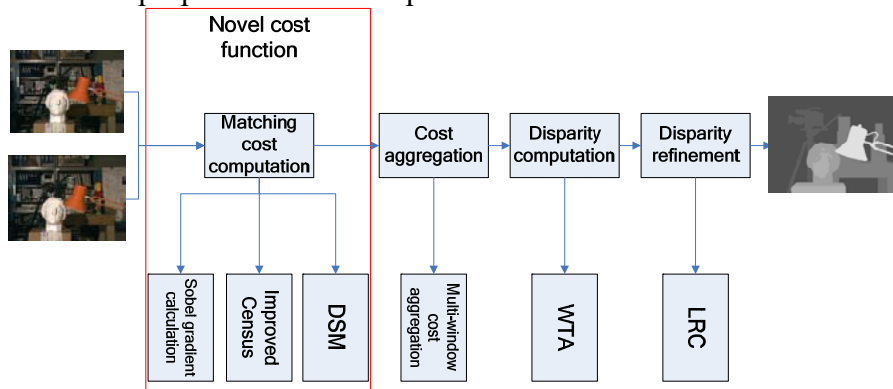


Figure 7. The algorithm flowchart

## References

- [1] D. Scharstein, R. Szeliski and R. Zabih. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. Stereo and Multi-Baseline Vision. Kauai. 2001: 131-140.
- [2] R. Zabih and J. Woodfill. Non-Parametric Local Transforms for Computing Visual Correspondance. Computer Vision.1994. 1994: 151-158.
- [3] Y. Kuk-Jin and K. In-So. Stereo Matching with the Distinctive Similarity Measure. Computer Vision. Rio de Janeiro. 2007. 2007:1-7.
- [4] Bernhard Froba,Andreas Erns,“Face Detection with the Modified Census,” IEEE FGR‘04,pp.1-6,2004
- [5] J. Canny. A Computational Approach to Edge Detection. Pattern Analysis and Machine Intelligence. 1986. PAMI-8(6): 679-698.