# Research on Stereo matching

Zhengtian Lu
*the University of Melbourne*
Melbourne, Australia
zhengtianl@student.unimelb.edu.au

Yizhao Huang
*the University of Melbourne*
Melbourne, Australia
yizhaoh@student.unimelb.edu.au

***Abstract*—The task of this report is to process pairs of stereo images captured from moving cars.**

***Index Terms*—Computer vision, Stereo matching,disparity map, Winner-take-all, Semi global mathcing, SDD Search, SAD search**

## I. Introduction

There is a technique shared by all passive stereo correspondence algorithms. Compare the locations of images to see how close they are. Typically, the matching cost is calculated per pixel, taking into account any relevant variations. While the most basic matching cost would assume a constant intensity at the site of the matched picture, more complex costs may account for variations in radiometric properties and noise.

Camera factors such as slightly varied settings, vignetting, picture noise, etc., may contribute to radiometric discrepancies. However, it is not always practical to do a radiometric pre-calibration to account for all of these variations. Moreover, there might be a distinction owing to non-Lambertian surfaces, the quantity of whose reflection of light varies with the observer's perspective. The fact that this Shrink the stereo baseline to lessen the difference, but the reconstruction will lose some geometric precision in the process. The numerous impacts of real-world stereo data are shown with this example. Above is a description of a series of stereo cameras manufactured by Daimler AG and calibrated inside a moving vehicle [1].

Getting the same view at various times of day might provide different results because to variations in radiation strength or the location of the light source. Acquiring images might be time-consuming and perhaps impossible for more expansive situations. Manage the lights (eg, outdoors). parallel circumstance When comparing pictures taken from above or below. [1]

For the above reasonsthis article will cover many ways for processing stereo picture pairs acquired by moving automobiles, with the goal of increasing the model's accuracy and explaining why the results may be erroneous.This article will draw on a variety of sources and provide a variety of experimental data to show how the accuracy of the model may be improved.

## II. Various methods and results obtained

### A. WTA-SSD

In computational models of neural networks, "winner-take-all" refers to a strategy in which neurones within each layer of the network compete for activation. However, some variants, such as a soft winner-take-all, allow for several neurones to remain active by using a power function in neurones in this way, but in the traditional version, only the neurone with the largest activation stays active and all other neurones are switched off.

The SSD template matching method compares the relative brightness of each picture to a reference image to establish a level of compatibility (template and input image). The concept behind this method is to compact the reduction result of each pixel in the gap between the input image and the template.

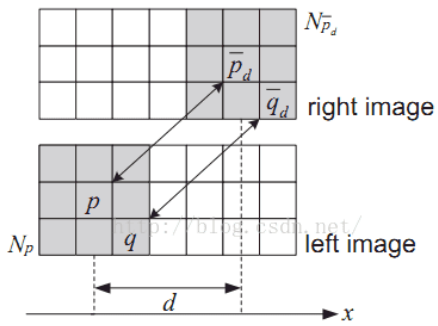



Figure 2. Reference support window ($N_p$) and target support window ($N_{\bar{p}_d}$) when $d_p = d$

Each pixel in the source picture Sourceimage(I) is compared to each pixel in the template image Templateimage(T) to determine how similar they are (R). Each matrix point's level of brightness corresponds to its degree of similarity to the template T. After that, the function minMaxLoc may be used to pinpoint the highest value in matrix R. [2](which also determines the minimum value). When comparing two photos, how do you determine if they are similar? Matching methods, or assessment criteria, may take on a variety of forms. Common examples include the normalised squared difference matching technique and the maximum likelihood matching method.

From the Fig.1, we can clearly see that use this method, comparing the left and right image. The eigenvalues of several important locations are recovered only approximatively. Because of all the background noise, its effectiveness is subpar. The frations seen TABLE1 To calculate the RMSE, we

take the standard deviation of the residuals (prediction error). Dispersion of residuals along the regression line is measured by root-mean-squared error (RMSE). What this measure does is reveal how heavily the data is clustered around the line of best fit.

TABLE I
FRACTIONS

| fractions | 4 | 2 | 1 | 0.5 | 0.25 |
|---|---|---|---|---|---|
| a | 36.5522% | 30.2338% | 25.2275% | 14.4381% | 14.4381% |

RSME: 23.333903383393913

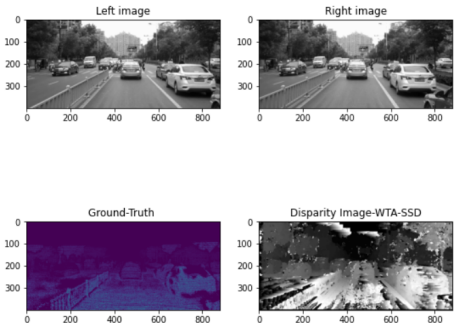



Fig. 1. The comparison under the WTA-SSD method

## B. WTA-SAD

The SAD local matching algorithm is a kind of algorithm commonly used in stereo matching algorithms. For the SAD local matching algorithm, the similarity of the gray value in the image is usually calculated based on the matching between the template and the pixel value in the original image. [3]The SAD algorithm is implemented by computing the similarity between the image and a template, which is sometimes called a window. The window consists of an image I(p,q) and a region of interest R. When doing local matching, the window is moved along each point in the image, while calculating the similarity measure L at each location. As shown in Figure 5, the similarity measure L can be expressed by the following formula:

$$SAD(u,v) = Sum|Left(u,v) - Right(u,v)| \quad (1)$$

From the Fig.2, we can clearly see that use this method, comparing the left and right image. The eigenvalues of several important locations are not imporved. Because of all the background noise still exist, its effectiveness is subpar.The frations seen TABLE2.

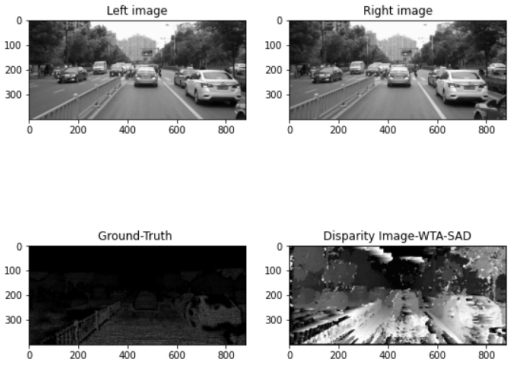



Fig. 2. The comparison under the WTA-SAD method

RSME: 25.179404239246008

TABLE II
FRACTIONS

| fractions | 4 | 2 | 1 | 0.5 | 0.25 |
|---|---|---|---|---|---|
| a | 30.0507% | 25.0837% | 19.2423% | 11.7272% | 11.7272% |

## C. WTA-NCC

The normalised cross correlation (NCC) algorithm is a matching approach that makes use of the tonal information included in a picture. It's a standard technique in the field of image processing, and it's often used to evaluate the degree to which two pictures are same. Image matching may be done in a few different ways: using grayscale, using features, or using the transform domain. With the NCC method, brightness is less of a factor in comparisons of images. Additionally, the NCC's ultimate outcome is a number between 0 and 1, making it straightforward to measure and compare. You can determine if the outcome is excellent or negative with the use of a cutoff point. The time commitment of the standard NCC comparison approach is high. Adjusting the window size and the step size of each detection may improve performance, but it still won't be fast enough for use in checking factories' output in real time. The electronic versions' small variations may assist businesses enhance product quality, lower defective-product rates, and maintain tight quality control.

In order to find a window that best fits every given pixel (px,py) in the original picture, we build a nn neighbourhood. Then, a matching window of size nn is built for the target pixel location (px+d,py), and the two windows' degree of similarity is calculated. Take into account the fact that d may take on a variety of values in this context. Before performing the NCC calculation for two images, it is recommended that the images be processed so that the optical centres of both frames are aligned on a horizontal plane and the epipolar line is also horizontal. Otherwise, the matching process will have to be performed in the epipolar direction and will require more processing power.

$$NCC(p,d)=\frac{\sum_{(x,y)\in W_p}(I_1(x,y)-\overline{I1}(p_x,p_y))\bullet(I_2(x+d,y)-\overline{I}_2(p_x+d,p_y))}{\sqrt{\sum_{(x,y)\in W_p}(I_1(x,y)-\overline{I}_1(p_x,p_y))^2\bullet\sum_{(x,y)\in W_p}(I_2(x+d,y)-\overline{I}_2(p_x+d,p_y))^2}}$$

By changing the window size we can clearly see that when window size is 10, the result is best as figure

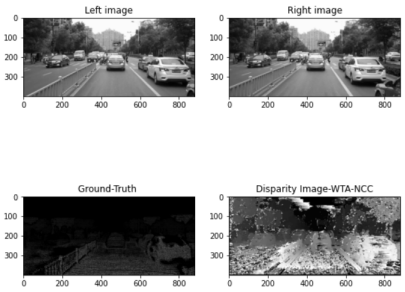



Fig. 3. The comparison under the WTA-NCC method

RSME26.9273

TABLE III
FRACTIONS

| fractions | 4 | 2 | 1 | 0.5 | 0.25 |
|---|---|---|---|---|---|
| a | 26.6450% | 23.0045% | 20.4256% | 10.4861% | 10.4861% |

### *D. Semiglobal Matching (SGM) stereo method*

The SGM is a popular method in computing stereo vision matching. The method uses fast path-wise optimizations to approximate the solution from all directions. According to Hirschmüller's [4] paper, the SGM mainly contains four steps which are matching cost calculation; cost aggregation; disparity calculation and disparity refinement. In the first step Hirschmüller applies the Mutual Information(MI) as cost function. The MI test is insensitive to changes in illumination. The entropy of each image is an important factor in determining the overall entropy of the combined image. The greater the entropy, suggests that the more detailed of the pixel grayscale. Another cost function is census transformation(CT) introduced by Zabih and Woodfill [5] which is used in this paper's experiment. CT transform the difference between the grayscale of the pixel and the grayscale of its neighbour pixels into a bit string as the final census value. CT is also insensitive to the illumination changes since it compares the relative grayscale.

In the second step, SGM needs to aggregate the cost since CT only consider the local correspondence which is very sensitive to noise and can not reflect the best disparity. [6]Most stereo matching methods use the strategy which is to finding the minimum of the global energy function. For SGM, the energy function is shown below. Where D is the disparity map, the first term is the total cost of matching pixels in the disparities of D. The second term adds a penalty for pixels in the neighborhood Np of p that change a little bit.. The third term in the differential equation penalizes inequality changes by a larger constant factor, P2, for all larger disparities. The penalty for small changes is lower in order to allow for adaptation to slanted or curved surfaces. Whenever a large change is made, there are always discontinuities that are preserved [4]. In practice, the SGM in this report aggregate four neighbors from top,down,left and right pixels within certain range.

$$E(D)=\sum_p C(p,D_p)+\sum_{q\in N_p}P_1T[|D_p-D_q|=1]+\sum_{q\in N_p}P_2T[|D_p-D_q|>1]]$$

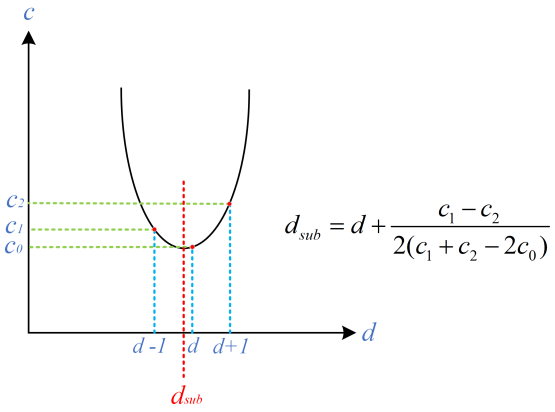



Fig. 4. Subpixel quadratic curve

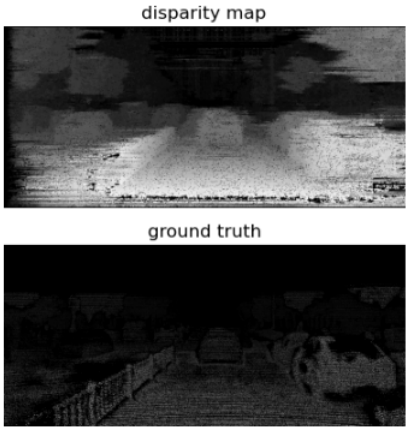
The third step disparity computation is normally done by using the winner-take-all strategy. For every pixel, choose the disparity which has the minimum cost as final disparity value. , SGM often uses subpixel to enhance the accuracy since the cross correlations can vary a great deal even at the subpixel level. The Subpixel refinement uses a quadratic curve which is fitted through the neighboring costs as the diagram shown below.

In terms of the last step, the SGM in this report applies Subpixel, Consistency Check, Uniqueness Check and Median Filter. Hirschmuller [4] mentions some otherl methods for disparity refinement, including Removal of Peaks, Intensity Consistent Disparity Selection, and Discontinuity Preserving Interpolation. For Consistency Check, as the formula shown below: basically swap the left image and right image then calculate the disparity again and for each pixel A in left image disparity map, find the same pixel B in right image disparity map and check if the disparity difference is under certain threshold. The Uniqueness Check refers that if the minimum cost and the second minimum cost are too close, then consider the pixel's disparity as invalid since it can be unreliable due to noise or other factors. Median Filter is a common technique to remove outliers and reduce noises.

$$D_p=\begin{cases}D_{bp} & \text{if } |D_{bq}-D_{mq}|\le 1\\ D_{invalid} & \text{otherwise}\end{cases}$$

*E. Experiment Results and Error analysis*

1.Without Consistency Check, Uniqueness Check and Median Filter:

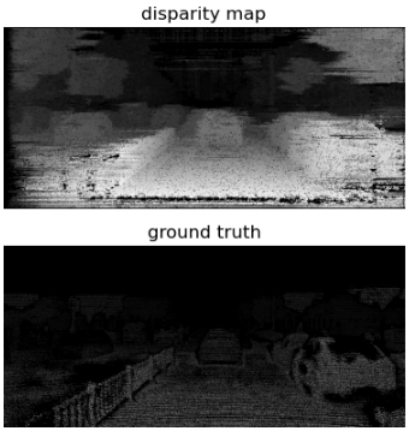



Time consume:621.4061s
RSME:4861.1920

TABLE IV
FRACTIONS

| fractions | 4 | 2 | 1 | 0.5 | 0.25 |
|---|---|---|---|---|---|
| a | 43.1879% | 31.2712% | 24.3124% | 13.8056% | 13.8056% |

2.With only Consistency Check:

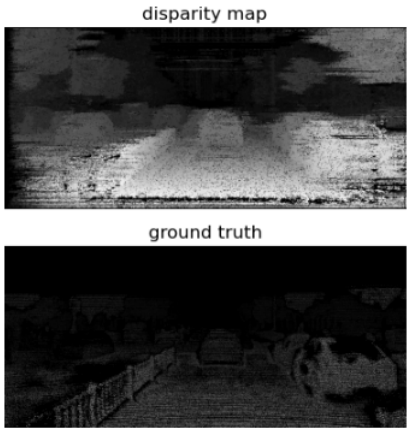



Time consume:645.7035s
RSME:4861.3131

TABLE V
FRACTIONS

| fractions | 4 | 2 | 1 | 0.5 | 0.25 |
|---|---|---|---|---|---|
| a | 44.0332% | 32.1635% | 25.2991% | 14.9344% | 14.9344% |

3.With Consistency Check and Uniqueness Check

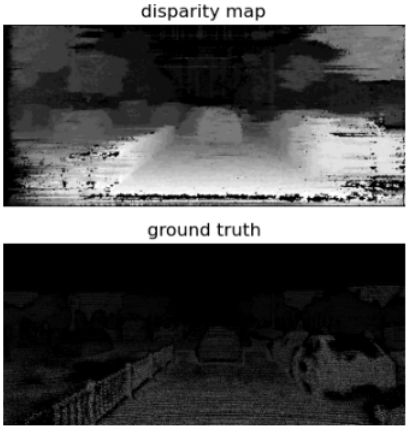



Time consume: 637.0674s
RSME: 4861.3426

TABLE VI
FRACTIONS

| fractions | 4 | 2 | 1 | 0.5 | 0.25 |
|---|---|---|---|---|---|
| a | 46.0048% | 34.5914% | 25.2991% | 17.8836% | 17.8837% |

4.With Consistency Check and Uniqueness Check and Median Filter




Time consume: 569.3676s
RSME: 4861.2035

TABLE VII
FRACTIONS

| fractions | 4 | 2 | 1 | 0.5 | 0.25 |
|---|---|---|---|---|---|
| a | 43.1649% | 30.9742% | 24.6390% | 12.8635% | 12.8635% |

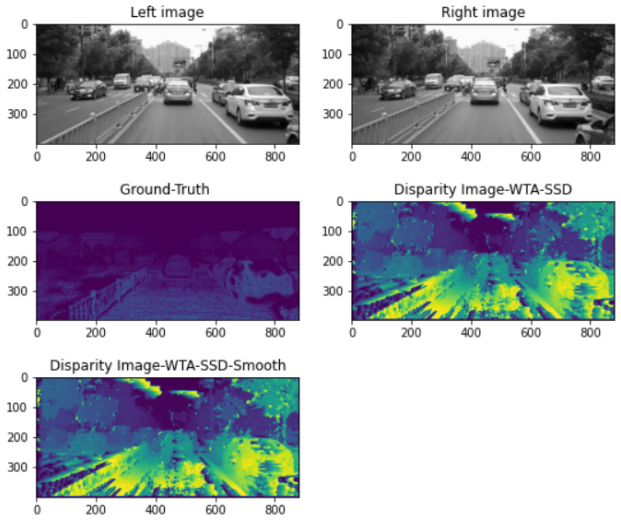
Observation: the noises reduce a lot but fraction results also reduce for the reason that Those pixels with invalid disparity also get smoothed.

## III. Further Improvement

For cost aggregation, instead of having only four paths(top, down, left, right), adding another four paths(top-left, top-right,

bottom-left, bottom-right) can improve the result accuracy. However, it will consume more time and more computational power. For disparity refinement, The removal of peaks can help remove the outliers, due to low texture, reflections, noise, and other factors. They usually show up as small patches of disparity that are very different from the surrounding disparities.

For encouraging a smooth output, we use the median filter method for both SSD and SAD method.The median filter method is a nonlinear smoothing technique that averages the grey levels of neighbouring pixels around a given place to get a single pixel's new value. In order to implement the median filter, statistical ranking principles are used. An excellent noise-cancelling tool, it makes use of nonlinear signal processing. To eliminate solitary noise points, median filtering works by swapping out the value of a given point in a digital picture or digital sequence with the median value of each point value in a neighbourhood of the point. From the figure below, we can clearly see that noise point is decreasing and the accuracy is increased.




For the accelerating the process, we use the jit. jit will let your function execute once, then optimise it based on the kind of parameter it was given. As the figure below: it can accelerate the compilation speed and efficiency. [7]

## IV. CONCLUSION

In summary, the first part of the report introduces the experiment of winner-take-all methods with different matching functions, includes SSD, SAD and NCC. Cutoff points can be used to determine whether results are excellent or negative. This method compute the similarity between the template and the image and focus to reduce the gap of them. With experiments, this method shows low root mean squared errors as well as low fraction results. In the second part of the report, the SGM stereo method has been presented. Experiments have shown that it yields a higher percentage of fractions results than the usual winner-takes-all method. Optimization of the global cost function ensures that matching is done accurately on a pixel level. The presented post filtering methods help tackling some individual problems, but also remaining individual problems. The main drawback of the implemented SGM in this report is its low speed. However, according to Hirschmuller [6], the runtime of SGM can be much faster with proper implementation which make it a prime choice for solving many practical stereo problems.

## REFERENCES


[1] H. Hirschmuller and D. Scharstein, "Evaluation of stereo matching costs on images with radiometric differences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 9, pp. 1582–1599, 2009.

[2] Y. E. Zhang, "Research of an improved snail image recognition method based on grayscale template matching," in *Applied Mechanics and Materials*, vol. 433. Trans Tech Publ, 2013, pp. 700–704.

[3] L. Di Stefano, M. Marchionni, and S. Mattoccia, "A fast area-based stereo matching algorithm," *Image and vision computing*, vol. 22, no. 12, pp. 983–1005, 2004.

[4] M. Humenberger, T. Engelke, and W. Kubinger, "A census-based stereo vision algorithm using modified semi-global matching and plane fitting to improve matching quality," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, 2010, pp. 77–84.

[5] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *European conference on computer vision*. Springer, 1994, pp. 151–158.

[6] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 328–341, 2008.

[7] S. K. Lam, A. Pitrou, and S. Seibert, "Numba: A llvm-based python jit compiler," in *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*, ser. LLVM '15. New York, NY, USA: Association for Computing Machinery, 2015. [Online]. Available: https://doi.org/10.1145/2833157.2833162