Computer Vision Challenge
Disparity Map

Group 56
Technische Universität München

July 19, 2019

1 Introduction

In this challenge the disparity map is generated by census-algorithm based on the given images from different angles.The rotation matrix and the translation matrix are also calculated. After that the Peak Signal-to-Noise Ratio between the disparity map and the ground truth is calculated. In the end a graphical user interface is designed so that all the options and results can be easily showed.

2 Implementation

2.1 Rotation Matrix and Translation Matrix

The method implemented in homework is used to calculate the Rotation and Translation matrix.

1. Use Harris detector and normalized cross correlation (ncc) to find the corresponding points of the two images from different positions.
2. Apply eight-point algorithm to estimate the essential matrix.
3. Calculate euclidean motion (R,T) through the essential matrix.

2.2 Disparity Map

At the beginning, only the census transform [1] is used to calculate the disparity. In a evaluation by Hirschmüller and Scharstein [2], census shows the best overall results in local and global stereo matching methods.
The census transform(CT) is a non parametric summary of local spatial structure. If P is a pixel and the blocksize is 3*3, the local neighborhood that surrounds a pixel P will be mapped to a bit string that represents the set of neighboring pixels whose intensity is less than that of P (in this case an 8 bit integer). One will be written if the grayscale value of P' is smaller than that of P and otherwise 0.

$$I_{\text{census}}(u,v) = \bigotimes_{i=n} \bigotimes_{j=m} (\xi(I(u,v), I(u+i, v+j))) \tag{1}$$

The operator $\bigotimes$ denotes a bit-wise catenation, n and m are the Census window size and u , v the horizontal and vertical coordinates.

$$\xi(p_1, p_2) = \begin{cases} 0 & p_1 \leq p_2 \\ 1 & p_1 > p_2 \end{cases} \tag{2}$$

where p_1 is the center pixel and p_2 is the neighborhood pixels in the image. It should be done for every pixel of two images and the Hamming distance will be compared. The pixel inside of the search range with the minimal Hamming distance is the corresponding point of the original point. The difference in coordinates of corresponding points will be saved as disparity.
In the disparity map of terrace some points on the surface of the table and the windows are too bright so that their corresponding points can not be found in the other picture. To solve this problem the improved census transform [3] and AD-census [4] are combined and implemented. The equation of improved census transform is reformulate as equation (3).

$$I_{\text{census}}(u,v) = \bigotimes_{i=n} \bigotimes_{j=m} (\xi(std(u,v), |I(u+i, v+j) - \overline{I}(u,v)|)) \tag{3}$$

$std(u,v)$ denotes the standard deviation of the census mask. $\overline{I}(u,v)$ is the mean intensity of the mask. Given a pixel $\mathbf{p} = (x,y)$ in the left image and a disparity level d, two individual cost values $C_{\text{census}}(\mathbf{p},d)$ and $C_{AD}(\mathbf{p},d)$ are first computed. The AD-Census cost value $C(\mathbf{p},d)$ is then computed as follows:

$$\begin{aligned} C(\mathbf{p},d) = & \rho\left(C_{census}(\mathbf{p},d), \lambda_{census}\right) + \\ & \rho\left(C_{AD}(\mathbf{p},d), \lambda_{AD}\right) \end{aligned} \tag{4}$$

where $\rho(c,\lambda)$ a robust function on variable c:

$$\rho(c,\lambda) = 1 - \exp\left(-\frac{c}{\lambda}\right) \tag{5}$$

2.3 Disparity Optimization

After the disparity map is generated, several optimization steps should be applied to improve the quality of it. First of all, the "Left-Right Check" can be applied to search the bad points. But the effect of the algorithm is not obvious. To improve the quality of the images, the Gaussian filter and median filter are implemented.

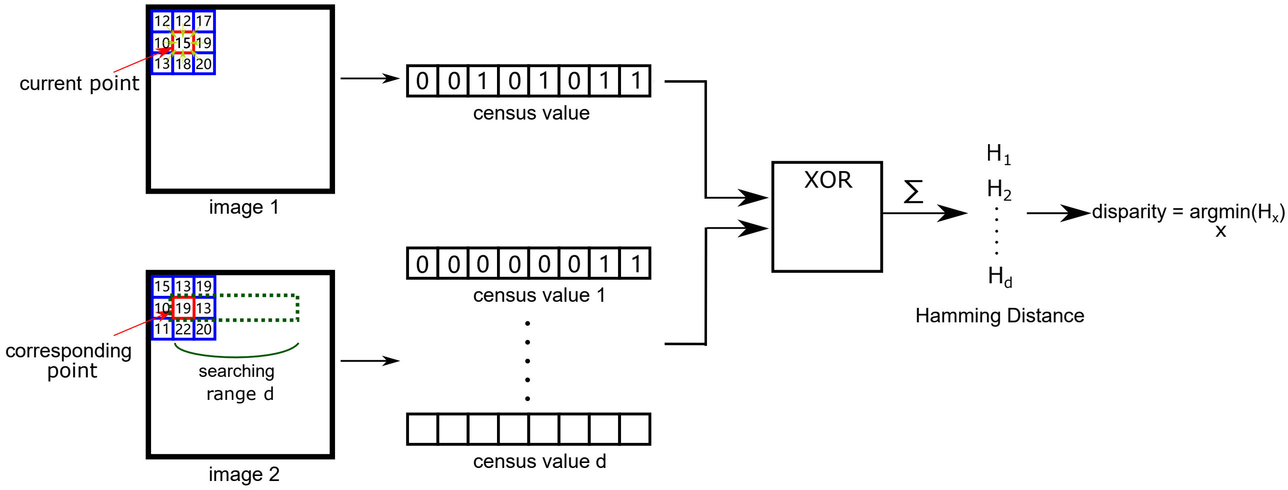

Figure 1: *Principle of census*

2.3.1 Gaussian Filter

Gaussian filtering is the process of weighted averaging of the entire image. The value of each pixel is weighted and averaged by itself and other pixel values in the neighborhood. The specific operation of Gaussian filtering is to scan each pixel in the image with a template, and replace the value of the center pixel of the template with the weighted average gray value of the pixels in the neighborhood determined by the template. It is used to remove noise from the image.

2.3.2 Median Filter

The median filter replaces the value of a point in the digital image with the median value of each point value in a neighborhood of the point, so that the pixel with a larger difference in the gray value of the surrounding pixels is changed to a value close to the surrounding pixel value, thereby isolated noise points can be eliminated. It can attenuate or eliminate the high frequency components of the Fourier space, but affect the low frequency components. Since the high-frequency component corresponds to a portion where the gray value of the edge of the region in the image has a relatively large change, the filtering can filter out the components to smooth the image. The value filtering technique can better protect the edges of the image while attenuating the noise. It is applied to remove the salt-and-pepper noise points [5].

3 Result

Peak signal-to-noise ratio (PSNR) is used to measure the quality of the disparity map. The PSNR (in dB) is defined as:

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \tag{6}$$

Here, MAX_I is the maximum possible pixel value of the image. When the pixels are represented using 8 bits per sample, this is 255. MSE is the mean squared error, which is defined as:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \tag{7}$$

After the implementation of the improved census and AD-census the reflection area is a little bit smaller(Figure 2 and 3). But the elapsed time increases from 12.57s to 34.28s and the PSNR almost doesn't change. So only the normal census is remained for the final program.

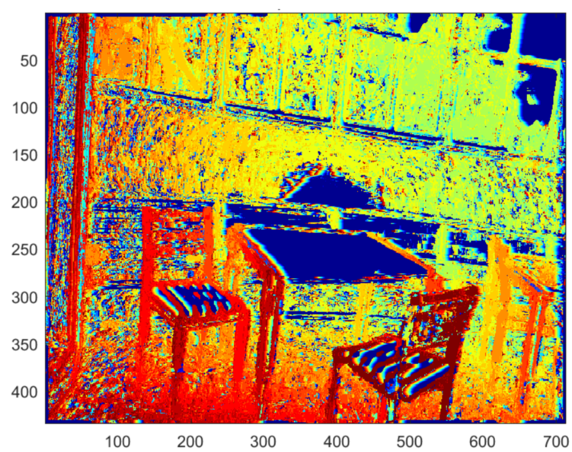

Figure 2: *Disparity map of terrace after census based on the right image*

The block size, search range(maximum disparity) and elapsed time are showed in the following table.

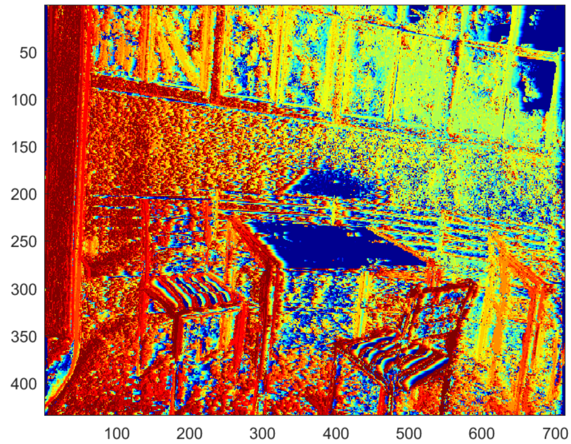

Figure 3: *Disparity map of terrace after AD-census based on the right image*

	Motorcycle	Playground
Blocksize	25*25	7*7
Search Range	250	20
Elapsed time(s)	1756	7.9318
PSNR (dB)	13.0225	32.2488

	Sword	Terrace
Blocksize	25*25	11*11
Search Range	370	16
Elapsed time(s)	2361	8.0693
PSNR (dB)	8.88	30.2707

R and T of motorcycle:

$$R = \begin{bmatrix} 0.9982 & -0.0487 & -0.0341 \\ 0.0485 & 0.9988 & -0.0080 \\ 0.0344 & 0.0063 & 0.9994 \end{bmatrix}, T = \begin{bmatrix} -0.1229 \\ 0.0805 \\ -0.9892 \end{bmatrix} \tag{8}$$

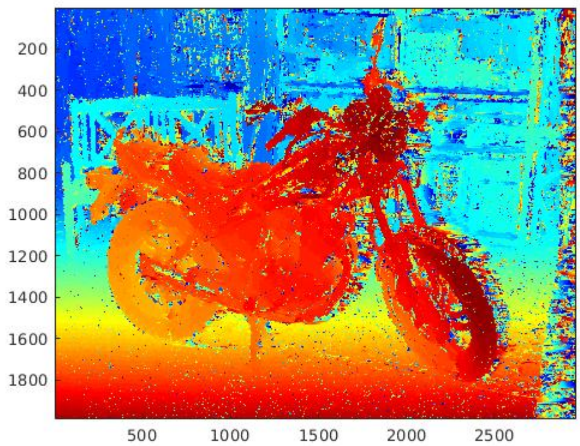

Figure 4: *Disparity map of motorcycle*

R and T of playground:

$$R = \begin{bmatrix} 0.9405 & -0.0603 & 0.3344 \\ -0.0638 & 0.9354 & 0.3479 \\ 0.3337 & 0.3485 & -0.8759 \end{bmatrix}, T = \begin{bmatrix} -0.1732 \\ -0.1798 \\ 0.9683 \end{bmatrix} \tag{9}$$

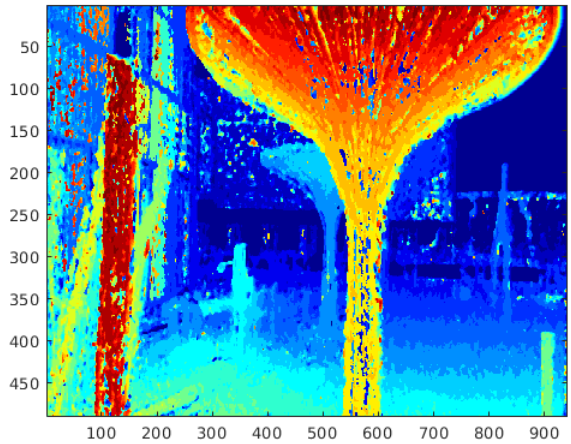

Figure 5: *Disparity map of playground*

R and T of sword:

$$R = \begin{bmatrix} 0.9946 & -0.0951 & -0.0423 \\ 0.0954 & 0.9954 & -0.0066 \\ 0.0414 & -0.0106 & 0.9991 \end{bmatrix}, T = \begin{bmatrix} 0.0277 \\ 0.0642 \\ 0.9976 \end{bmatrix} \tag{10}$$

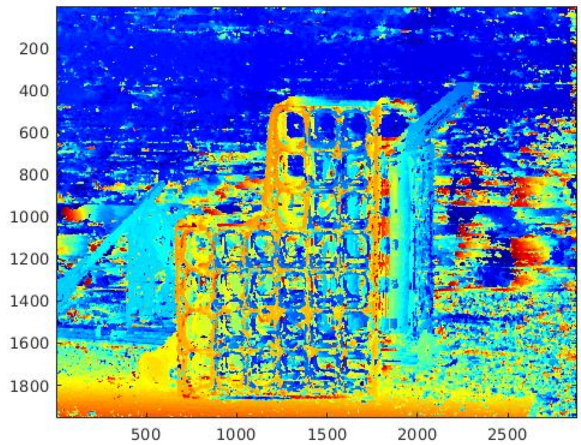

Figure 6: *Disparity map of sword*

R and T of terrace:

$$R = \begin{bmatrix} 1.0000 & -0.0053 & -0.0036 \\ 0.0053 & 1.0000 & 0.0020 \\ 0.0036 & -0.0020 & 1.0000 \end{bmatrix}, T = \begin{bmatrix} -0.2717 \\ -0.2248 \\ 0.9358 \end{bmatrix} \tag{11}$$

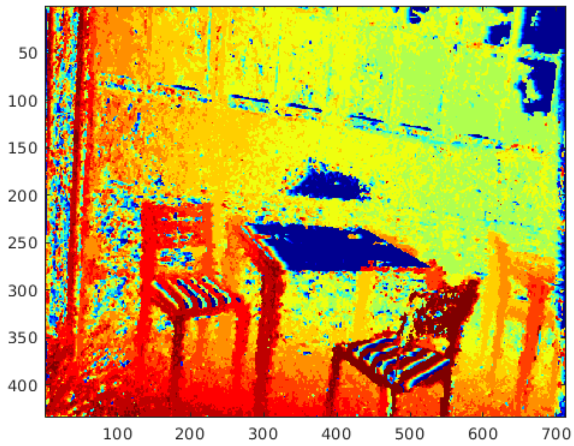

Figure 7: *Disparity map of terrace*

4 Outlook

Several questions still remain to be answered, e.g. the reflection problem in "terrace" images. Because the intensity of certain area are the same, therefore the census value of the points are also the same, which makes it difficult to find the right corresponding points.
In addition, in the picture of the sword. Due to the unique mesh (fence) structure of the object. The spacing of the fence is too wide relative to the width of the fence itself, and the fence structure is highly repeatable. As a result, once the corresponding point is not found in the correct position, it will be searched in the next similar pattern, resulting in a huge error.

References

[1] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," pp. 151–158, Springer-Verlag, 1994.

[2] H. Hirschmuller and D. Scharstein, "Evaluation of stereo matching costs on images with radiometric differences," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, pp. 1582 – 1599, 10 2009.

[3] X. Luan, F. Yu, H. Zhou, X. Li, D. Song, and w. Bingwei, "Illumination-robust area-based stereo matching with improved census transform," pp. 194–197, 05 2012.

[4] X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang, and X. Zhang, "On building an accurate stereo matching system on graphics hardware.," [4], pp. 467–474.

[5] T. Huang, G. Yang, and G. Tang, "A fast two-dimensional median filtering algorithm," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 27, pp. 13–18, Feb. 1979.