Stereo Matching using Census Transform of Adaptive Window Sizes with Gradient Images

Jaeryun Ko and Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST),
123 Cheomdangwagi-ro, Buk-gu, Gwangju, 61005, Republic of Korea
E-mail: jrko@gist.ac.kr, hoyo@gist.ac.kr

Abstract— **The census transform in computing the matching cost of stereo matching is simple and robust under luminance variations in stereo image pairs; however, different disparity maps are generated depending on the shape and size of the census transform window. In this paper, we propose a stereo matching method with variable sizes of census transform windows based on the gradients of stereo images. Our experiment shows higher accuracy of disparity values in the area of depth discontinuities.**

I. Introduction

For 3D scene reconstruction, generating disparity maps from a pair of stereo images is the basic process and various methods for obtaining the disparity information have been researched so long. Nowadays we can get easily the depth information by using active depth cameras based on infrared sensors, but these kind of cameras have some problems and limitations of obtaining the depth distance for capturing under the outdoor conditions. We generally use stereo matching techniques by exploiting stereo images captured by the conventional color cameras.

Stereo matching is estimating the depth information of the projected view from stereo images at the different views. The main idea of this technique is based on checking the correspondences between the pair of stereo images. These techniques are separated into two classes, which one is local-based methods and another is global-based methods.

The local-based stereo matching is estimating the disparity values with the comparison of matching costs from views on the right and on the left. It is fast, simple and suitable for the real-time depth generating applications. Because of low accuracy in local-based methods, global-based methods are used for higher accuracy. It is based on the Markov random field and estimating proper disparity values by minimizing energy functions, which is pre-defined. The energy function can be minimized by using optimization algorithms such as belief propagation, graph cuts and so on.

There are many researches on various matching cost computation methods for solving kinds of threatening conditions. Absolute difference or squared difference are mostly used as the simplest one for calculating the matching cost. Normalized cross correlation (NCC) can be one of the options for better performance as a matching cost aggregation method [1]. We used the census transform as a matching cost function, which is robust under the difference of absolute intensity values on the stereo images. It is especially strong when stereo images have different luminance variations each other. The census transform can be worked with the moving windows. The results of stereo matching with the census transformed cost function depend on the window size of the transformation.

In this paper, we find out the properties on using different sizes of census transform windows with the results of stereo matching at first. Based on these properties we propose a stereo matching method with variable sizes of census transform window which is based on the stereo images of the gradients on both vertical and horizontal directions. In our experiments we compared the results of our proposed method with census transform based stereo matching of fixed window size.

II. Stereo Matching with Census Transform

A. Belief Propagation (BP)

Stereo matching using belief propagation is working on the Markov random field like Fig. 1. Each node on this field, which corresponds to each pixel on the corresponding spatial position, send the value called as message to the neighboring pixels [2].

Every $y_i, i = 1, \ldots, n$ has the matching cost for disparity candidates. m_i means the message from node y_i to node x_i. Message $m_{i,j}$ is sent from node x_i to node x_j. These nodes should be initialized through the uniform distribution as the first step of belief propagation.

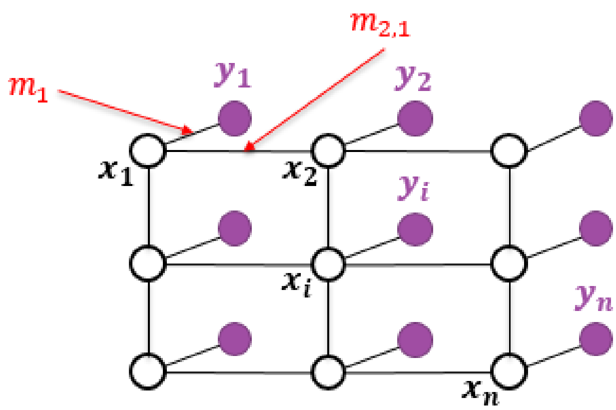

Fig. 1 Markov random field related to the image

As the second step, the updating process must iterate until all the messages converge through (1). After finishing the iteration, each belief of the corresponding pixel $b_i(x_i)$ can be calculated through (2). Finally the disparity value with corresponding pixel is determined to x_i which minimizes the belief of the corresponding pixel.

$$m_{ij}^t = \min_{x_i}[V(x_i, x_j) + D_i(x_i) + \textstyle\sum_{s\in N(i)\setminus j} m_{s,i}^{t-1}] \quad (1)$$

$$b_i(x_i) = D_i(x_i) + \sum_{i\in N(j)} m_{j,i}^t \quad (2)$$

B. *Census Transform*

Since the census transform uses the relative intensity of input images, it is robust to intensity variations of input images. It is a non-parametric transform method [3]. Let $C(P)$ be the census transform of a pixel P. $C(P)$ maps the local neighborhood of surrounding a pixel P to a bit string representing the set of neighboring pixels whose intensity is less than that of P. The census transform is defined as (3).

$$C(P) = \otimes_{[i,j]\in D}\, \xi(P, P + [i,j]) \quad (3)$$

Symbol $\otimes$ denotes the concatenation, D means the non-parametric window around P and ξ represents transform defined by (4).

$$\xi(P, P + [i,j]) = \begin{cases} 1, if\ P > P + [i,j] \\ 0, \quad otherwise \end{cases} \quad (4)$$

Fig. 2 shows an example of the census transform. The census transform converts the relative intensity differences to 0 or 1 into the one-dimensional bit array form. Two pixels of the census transformed images are compared for the similarity using the Hamming distance, which is the number of bits that differ in the two bit strings as shown in the right side of Fig. 2. To compute the correspondences, the Hamming distance is minimized after applying the census transform.

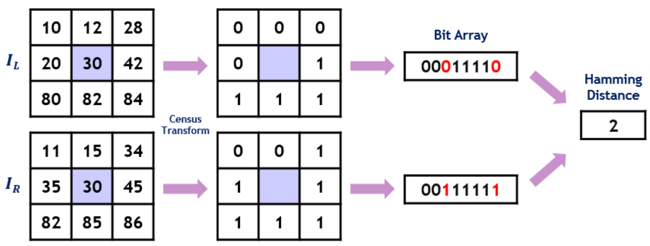

Fig. 2 Census transform from moving window of stereo image

III. ADAPTIVE WINDOW SIZES OF CENSUS TRANSFORM

The window shape of the census transform can be not only square also rectangular. Depending on the shape of rectangular windows, there are some subtle differences on the result of stereo matching with the census transform. These differences are shown in Fig. 3, applying different shape of rectangular windows to do the stereo matching with the *teddy* pair from Middlebury Stereo Dataset [4]. In the regions of red ellipses, the different results of stereo matching turned up with each 7x3 or 3x7 pixel-sized window of census transform. It seems that the result with 7x7 sized window is the composed result with the disparity maps of using each 7x3 and 3x7 census transform window.

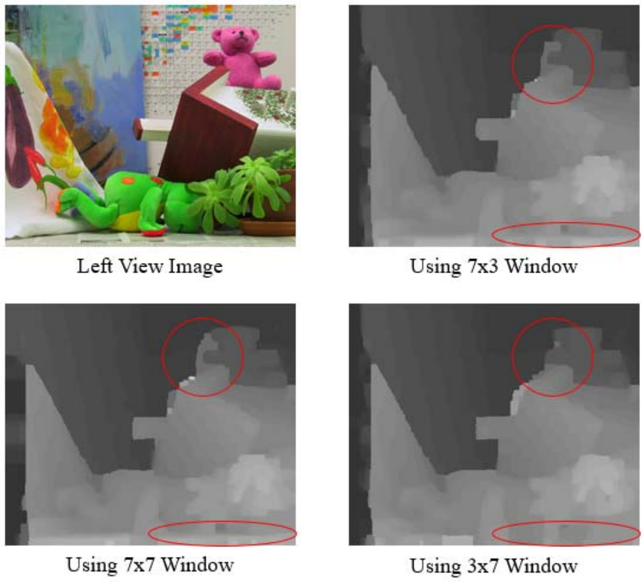

Fig. 3 The left view of *Teddy* and the disparity maps with different windows

Even though Fig. 3 is just an example of the effects of different window sizes and shapes of the census transform on the stereo matching, we can make an assumption that disparity values near the horizontal edges are plausible in the result of 7x3 sized census transform window on the stereo matching. Disparity values near the vertical edges seem to be reasonable in the matching result with 3x7 window of census transform.

Based on this assumption we propose stereo matching method with adaptive window sizes of the census transform in reference to the gradients of input stereo images. Let the left view of input stereo pair be the base image. We apply Sobel operators of both horizontal and vertical directions. [5] Then both horizontal gradient image G_x and vertical gradient image G_y are obtained like Fig. 4.

Before calculating the matching cost function through the census transform window for assigning the disparity value to a pixel P, we compares the values between the corresponding pixels on G_x and G_y. If the horizontal gradient is larger than the vertical gradient on the corresponding pixel P, that means most edges near P are likely to be vertical, then we choose a portrait window, i.e. 7x3 sized window of the census transform for the matching cost computation on P. When the vertical gradient is larger than the horizontal gradient, we choose a landscape window of the census transform. Otherwise we use a square window for the census transform. D_p, D_l and D_s denotes the portrait, the landscape and the square windows of the census transform in order. This idea is simple expressed in (5).

$$D = \begin{cases} D_p, if\ G_x(P) > G_y(P) \\ D_l, if\ G_x(P) < G_y(P) \\ D_s, otherwise \end{cases} \quad (5)$$

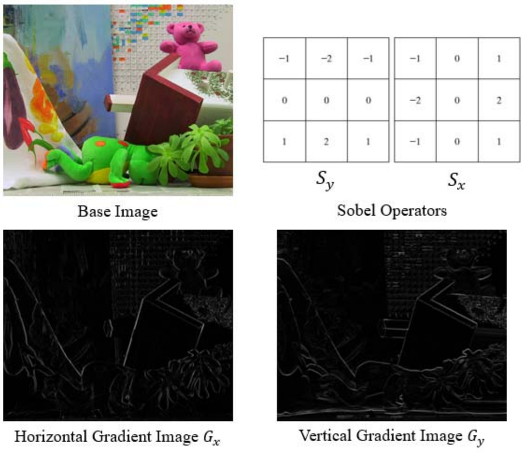

Fig. 4 The left view image as the base image and the gradient images from applying Sobel operators

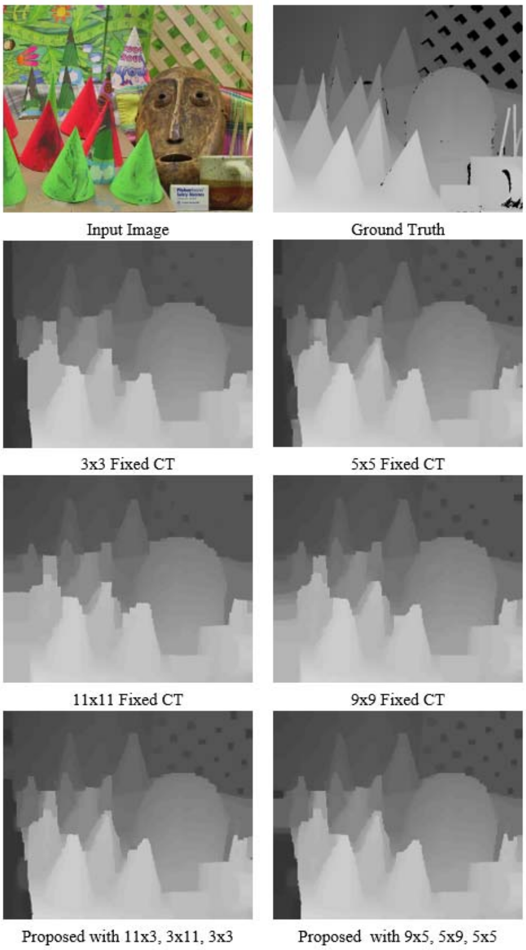

Fig. 5 The disparity maps of *Cones* under various window sizes of the census transform and our proposed method

IV. Experiment Results

To compare the proposed method with the global stereo matching which exploits the census transform and the belief propagation, we used four stereo image datasets, *Cones, Teddy, Tsukuba,* and *Venus* which are free for use [4].

The window sizes of census transform are fixed as 3x3, 5x5, 9x9, and 11x11 for the conventional global stereo matching. In order to test our proposed method, we set two cases for adaptive window sizes. One is composed with 3x11, 11x3, and 3x3 depending on the gradient images. Another is made up with 5x9, 9x5, and 5x5 as well. We applied hierarchical belief propagation as the matching cost optimization for all the test cases commonly [6]. There are no pre-processing or post-processing for the disparity map improvement in our experiments.

The disparity maps of *Cones* generated from the stereo matching with various fixed-size windows of the census transform and the proposed method are shown in Fig. 5. In the results of 3x3 and 11x11, the diamond-shaped disparities near the top right corner of the image are quite missing. In our proposed method we can find get better results in the area of top right corner with some diamond-shaped disparities. The results of our proposed method shows better performance near around some peaks of cones.

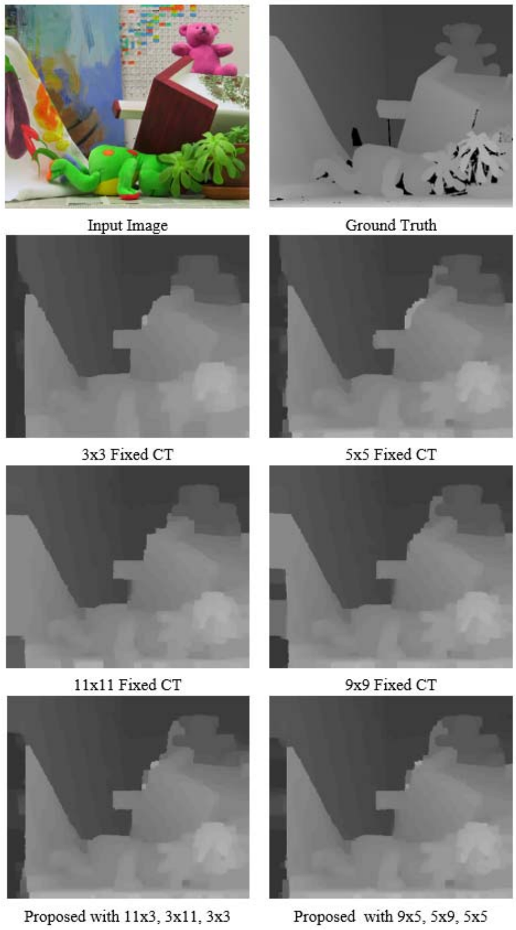

Fig. 6 Disparity maps of *Teddy* under various conditions same as Fig. 5

Fig. 6 shows the results on *Teddy* pair under the same conditions. It is not easy to distinguish a very difference between the conventional methods and the proposed ones but we can check it better by measuring in bad pixel rates (BPR).

The bad pixel rates on all the region or near the depth discontinuities are shown in Table I and Table II respectively. Lower rates of bad pixels mean that it is better performance because it represents the amount of bad pixels comparing the ground truth and the results of disparity maps. Our proposed methods are slightly worse than the results with fixed windows of the census transform on all area. On the other hand, the proposed methods perform better than others on the depth discontinuities.

TABLE I
BAD PIXEL RATES ON ALL AREA OF DISPARITY MAPS

BPR ALL (%)	Test Images			
	Teddy	*Cones*	*Tsukuba*	*Venus*
Fixed 3x3	21.68%	16.85%	4.59%	2.87%
Fixed 11x11	17.59%	15.34%	5.40%	2.84%
Proposed 3x11, 11x3, 3x3	18.78%	15.75%	4.61%	2.89%
Fixed 5x5	20.05%	15.77%	4.63%	3.04%
Fixed 9x9	18.02%	14.31%	5.09%	2.88%
Proposed 5x9, 9x5, 5x5	18.71%	14.88%	4.67%	3.13%

TABLE II
BAD PIXEL RATES ON DEPTH DISCONTINUITIES OF DISPARITY MAPS

BPR D_DISC (%)	Test Images			
	Teddy	*Cones*	*Tsukuba*	*Venus*
Fixed 3x3	34.06%	20.23%	18.08%	17.88%
Fixed 11x11	29.97%	20.36%	21.57%	17.51%
Proposed 3x11, 11x3, 3x3	24.61%	17.54%	16.04%	17.00%
Fixed 5x5	27.79%	17.42%	15.65%	17.45%
Fixed 9x9	27.37%	18.50%	18.20%	16.82%
Proposed 5x9, 9x5, 5x5	25.10%	17.50%	15.89%	18.20%

V. CONCLUSIONS

The census transform as a matching cost function is generally used for solving the difficulties of matching the correspondences under the luminance variation between stereo images. Global stereo matching with the census transform and the belief propagation shows good performances, but different disparity maps are generated depending on the size and shape of the census transform windows. In order to get better performance in the area of depth discontinuities, we proposed global stereo matching method with adaptive window sizes of the census transform, referring to the gradients of stereo images in both horizontal and vertical directions. In our experiments even if the accuracy of disparity values on all the area is slightly worse, we got better accuracy on the depth discontinuities.

ACKNOWLEDGEMENT

This work was supported by the 'Civil-Military Technology Cooperation Program' grant funded by the Korea government.

REFERENCES

[1] S. Patil, J. S. Nadar, J. Gada, S. Motghare and S. S. Nair, "Comparison of Various Stereo Vision Cost Aggregation Methods," International Journal of Engineering and Innovative Technology, vol. 2, pp. 222-226, 2013.

[2] S. Yedidia, W. T. Freeman, and Y. Weiss, "Understanding belief propagation and its generalizations," Exploring Artificial Intelligence in the New Millenium, pp. 239-269, 2003.

[3] R. Zabih, J. Woodfill, "Non-parametric Local Transforms for Computing Visual Correspondence," European Conference on Computer Vision, pp. 151-158, 1994.

[4] D. Scharstein, and R. Szeliski "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," International Journal of Computer Vision, vol 47, no. 1-3, pp. 7-42, 2002.

[5] I. Sobel, G. Feldman, "A 3x3 isotropic gradient operator for image processing," A Talk at the Stanford Artificial Project, pp. 271-272, 1968.

[6] P. Felzenszwalb, D. Huttenlocher, "Efficient Belief Propagation for Early Vision," IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp 261-268, 2004

[7] Y. Chang, Y. Ho, "Modified SAD Using Adaptive Window Sizes for Efficient Stereo Matching," International Conference on Embedded Systems and Intelligent Technology, pp. 9-11, 2014.