Fast Cost Computation Using Binary Information for Illumination Invariant Stereo Matching

Yong-Jun Chang, Yo-Sung Ho
School of Electrical Engineering and Computer Science
Gwangju Institute of Science and Technology (GIST)
Gwangju, Republic of Korea
{yjchang, hoyo}@gist.ac.kr

***Abstract*—Stereo matching methods are used to estimate disparity values from captured stereo images. They exploit characteristics of binocular disparity for disparity estimation. Stereo images captured under practical conditions have different illumination status, which causes disparity errors in the matching operation. To solve this problem, we can use the adaptive normalized cross correlation (ANCC) as a similarity measure that is independent from illumination factors and provides good matching results under various radiometric conditions. However, it has a very high computation complexity because of the bilateral filtered block matching operation. In this paper, we propose a new stereo matching method using binary information to reduce the computation complexity of ANCC. The proposed method uses a global mean value, instead of the bilateral filter. A census transformation is also applied in cost computation for fast block matching.**

***Index Terms*—Stereo matching, disparity map, ANCC, global mean value, census transformation**

I. INTRODUCTION

Depth information is used for making a three-dimensional (3D) image. The depth value represents a distance between the camera and the object. Therefore, we can feel a sense of reality from the captured scene using both the color image and the depth image. Depth information is also used for many applications such as a multi-view display, an augmented reality (AR), and a virtual reality (VR). There are two ways for the depth estimation.

The first method for the depth estimation is to use a depth camera. This method measures the distance between the depth camera and the object using an infrared ray. The depth value is calculated by the time-of-flight (ToF). Thus, the depth estimation using the depth camera measure the depth value fast and accurately. However, the resolution of the captured depth image is usually small because of the limitation of the depth camera. In addition, the depth camera has a difficulty with the depth estimation in the outdoors because of a sunlight.

The depth estimation using captured color images is the second method for the depth estimation. A stereo matching is one representative way for the depth estimation using captured color images [1]. Stereo matching methods use stereo images that have two different viewpoints to measure the depth value. Since stereo matching methods follow characteristics of binocular disparity, the depth value in stereo matching results is represented as a disparity value. Therefore, the disparity value means a location difference between two correspondences in stereo images.

Generally, a lot of stereo matching algorithms are implemented under ideal lighting conditions. Thus, there are no illumination differences between the left and the right viewpoint images. However, stereo images are not actually captured under those ideal conditions. Therefore, stereo images usually have different illumination status depending on the position of the light and characteristics of the camera. In this paper, we propose a fast cost computation method that is robust to various radiometric conditions and has lower cost computation time than the adaptive normalized cross correlation (ANCC) method [2], [3].

This paper is extended from a domestic conference paper that was presented in the Korean Institute of Broadcast and Media Engineers (KIBM) Fall Conference 2016 [3].

II. STEREO MATCHING USING ANCC

A. *Adaptive Normalized Cross Correlation (ANCC)*

The stereo matching using conventional similarity measures such as sum of absolute differences (SAD), sum of squared differences (SSD), and normalized cross correlation (NCC) has difficulty with the disparity estimation under different radiometric conditions. Fig. 1 represents the stereo matching result using stereo images that have different radiometric conditions each other. In Fig. 1, SAD values with regard to the color and the gradient images are used for disparity estimate. Fig. 1(c) shows the disparity map of the left viewpoint image and there are a lot of disparity errors in the disparity map.

The ANCC makes the stereo matching possible for searching the accurate disparity value under radiometric variations [2]. It eliminates several factors of the color image formation model to do the similarity measure that is independent from radiometric changes. (1) represents the color image formation model of the left viewpoint image [4].

(a) Left view (b) Right view (c) Disparity map

Figure 1. Stereo matching result under the radiometric variation

$$\begin{pmatrix} R_L(p) \\ G_L(p) \\ B_L(p) \end{pmatrix} \rightarrow \begin{pmatrix} \widetilde{R_L}(p) \\ \widetilde{G_L}(p) \\ \widetilde{B_L}(p) \end{pmatrix} = \begin{pmatrix} \rho_L(p) a_L R_L^{\gamma_L}(p) \\ \rho_L(p) b_L G_L^{\gamma_L}(p) \\ \rho_L(p) c_L B_L^{\gamma_L}(p) \end{pmatrix}, \tag{1}$$

In (1), where p means a current pixel position, R_L, G_L, B_L are raw color values, and $\widetilde{R_L}, \widetilde{G_L}, \widetilde{B_L}$ are color values that are stored in the device. $\widetilde{R_L}, \widetilde{G_L}$, and $\widetilde{B_L}$ are composed of the brightness factor ρ_L, scale factors a_L, b_L, and c_L, and the gamma exponent γ_L.

The color image formation model represented in (1) has nonlinear color values. To make those color values be linear relationship, the logarithm is applied to (1). The logarithmic form of $\widetilde{R_L}$ is represented in (2).

$$R_L'(p) = \log\rho_L(p) + \log a_L + \gamma_L \log R_L(p), \tag{2}$$

Both G_L' and B_L' are also obtained in the same manner. The brightness factor $\log\rho_L(p)$ is eliminated by subtracting the average of the logarithmic color values from each logarithmic color value. The brightness factor subtracted color value is defined in (3) and it is arranged by (4) and (5).

$$R_L''(p) = R_L'(p) - \frac{R_L'(p) + G_L'(p) + B_L'(p)}{3}, \tag{3}$$

$$R_L''(p) = \log\frac{a_L}{\sqrt[3]{a_L b_L c_L}} + \gamma_L \log\frac{R_L(p)}{\sqrt[3]{R_L(p) G_L(p) B_L(p)}}, \tag{4}$$

$$R_L''(p) = \alpha_L + \gamma_L K_L(p), \tag{5}$$

In (5), where α_L is the scale factor that includes a_L, b_L, and c_L in (4) and K_L is raw color values in (4). From (5), we can check that there is no brightness factor in the equation. The remaining factors are the scale factor α_L and the gamma exponent. α_L is removed using the bilateral filtering process [5]. The bilateral filtered value is represented as follows

$$R_L'''(t) = R_L''(t) - \frac{\sum_{t \in W(p)} w(t) R_L''(t)}{Z(p)}, \tag{6}$$

where $W(p)$ is a set of pixels in the matching window that is centered at the current pixel position p, $w(t)$ is a bilateral weighting function, and $Z(p)$ is a sum of weighting values in the window. Bilateral filtered values are applied to each pixel in the window. (6) is restated as (7).

$$R_L'''(t) = \gamma_L \left(K_L(t) - \frac{\sum_{t \in W(p)} w(t) K_L(t)}{Z(p)} \right), \tag{7}$$

In (7), lighting factors except for the gamma exponent are eliminated. In order to do the independent stereo matching from the gamma exponent, the color value $R_L'''(t)$ in (7) is applied to the equation of NCC. Thus, the ANCC measure of R channel is defined as follows

$$ANCC_R(f_p) = \frac{\sum_{i=1}^{M} w_L(t_i) w_R(t_i) [R_L'''(t_i)] \times R_R'''(t_i)}{\sqrt{\sum_{i=1}^{M} |w_L(t_i) R_L'''(t_i)|^2} \times \sqrt{\sum_{i=1}^{M} |w_R(t_i) R_R'''(t_i)|^2}}, \tag{8}$$

where f_p is a disparity candidate of the current pixel p, M is a total number of pixels in the matching window, i is an index of the pixel in the matching window, w_R is the bilateral weighting value of the right viewpoint image, and R_R''' is a color value that is independent from the brightness and the scale factors. ANCC measures of both B and G are obtained in the same manner. The final cost function of ANCC is represented in (9).

$$D(f_p) = 1 - \frac{ANCC_R(f_p) + ANCC_G(f_p) + ANCC_B(f_p)}{3}, \tag{9}$$

B. Problem of ANCC

The ANCC estimates the disparity value well under different radiometric conditions. It generally estimates quite accurate disparity values in large sizes of the matching window. However, the bilateral filtering in the ANCC causes an increase of the computation complexity. The bilateral filter is applied to every pixel in the matching window as shown in (6) and (7). Therefore, it has a great effect on the implementation time depending on the size of the matching window.

III. Stereo Matching Using Fast Cost Computation

The global mean value and the census transformation are used for decrease of the cost computation time of the ANCC [3], [6]. The global mean value eliminates the scale factor α_L in (5), instead of the bilateral filter. The census transformation changes the pixel in the matching block to the binary value. This binary value is used for the similarity measure that is independent from the gamma exponent. It also makes the cost computation possible to be performed faster than the ANCC.

A. Global Mean Value

The bilateral filtered color value is subtracted from the color value that is represented in (6) to remove the scale factor. Since the bilateral weighting function includes the multiplication of exponential weighting values, it takes a lot of time for computing those weighting values. In addition, the bilateral filtered value must be calculated for every pixel iteratively.

The proposed method uses the global mean value of the image that is independent from the brightness factor to eliminate the scale factor. The equation of the global mean value is represented as follows

$$R_L'''(p) = R_L''(p) - \frac{\sum_{t \in W_{MN}} R_L''(t)}{MN} \quad (10)$$

where M and N are the height and the width of the input image, and W_{MN} is a set of pixels in the input image. (10) can be restated in (11) and (12).

$$R_L'''(p) = \alpha_L + \gamma_L K_L(p) - \frac{\sum_{p \in W_{MN}} (\alpha_L + \gamma_L K_L(t))}{MN} \quad (11)$$

$$R_L'''(p) = \gamma_L \left(K_L(p) - \frac{\sum_{t \in W_{MN}} K_L(t)}{MN} \right) \quad (12)$$

In (12), we can check that there is no scale factor. Once the global mean value is calculated, it can be applied to all pixels in the image without iterative computation for every pixel. Therefore, its computation time for removing the scale factor is faster than that of the bilateral filtering method.

B. *Census Transformation*

The remaining lighting factor of (12) is the gamma exponent. The ANCC uses the equation of NCC to make the independent cost function from the gamma exponent. The NCC also contains several multiplications. Therefore, it also takes time for computing the cost. The proposed method uses the binary value that is the result of the census transform to measure the similarity [6]. Fig. 2 shows an example of the census transformation. In Fig. 2, a 3×3 sized block is used for the transformation. It calculates the difference between the value of the current pixel and that of the neighboring pixel. If the difference is equal or larger than 0, the value of that neighboring pixel is set to 1. Otherwise, the pixel value is set to 0.

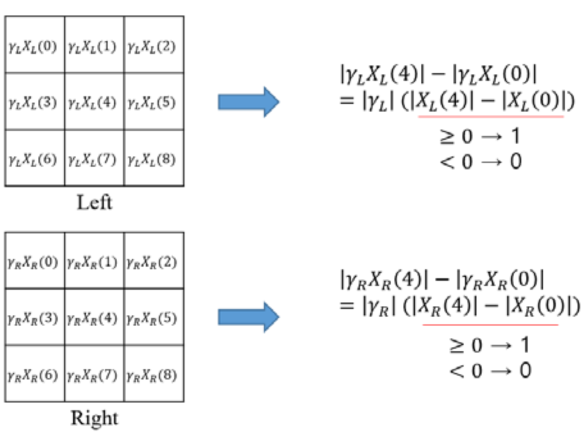


Figure 2. Example of the census transformation in the proposed method

In Fig. 2, where X_L means $K_L(p) - \frac{\sum_{t \in W_{MN}} K_L(t)}{MN}$ and X_R is also calculated in the same manner. Thus, all values in the left and the right 3×3 blocks are independent from the scale and the brightness factors. However, those still have gamma exponents. To do the block matching without the influence of the gamma exponent, the proposed method changes all pixel values of both images in Fig. 2 to absolute values. After that, the census transformation is applied to both matching blocks.

The sign of the pixel difference in Fig. 2 is only determined by the difference between $X_L(p)$ and $X_L(t)$ (or $X_R(p)$ and $X_R(t)$). In this case, the gamma exponent does not affect the sign decision in the census transformation step. Therefore, the similarity measure without the gamma exponent is possible. The matching cost after the transformation is calculated by the Hamming distance. Fig. 3 shows how to calculate the matching cost after the census transformation.

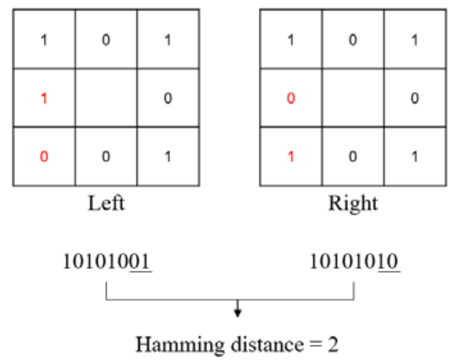


Figure 3. Hamming distance in census transformed results

In Fig. 3, both blocks represent census transformed results. From those results, bit-arrays are obtained. The hamming distance counts the number of bits that are different each other. The cost function of the proposed method is defined as follows

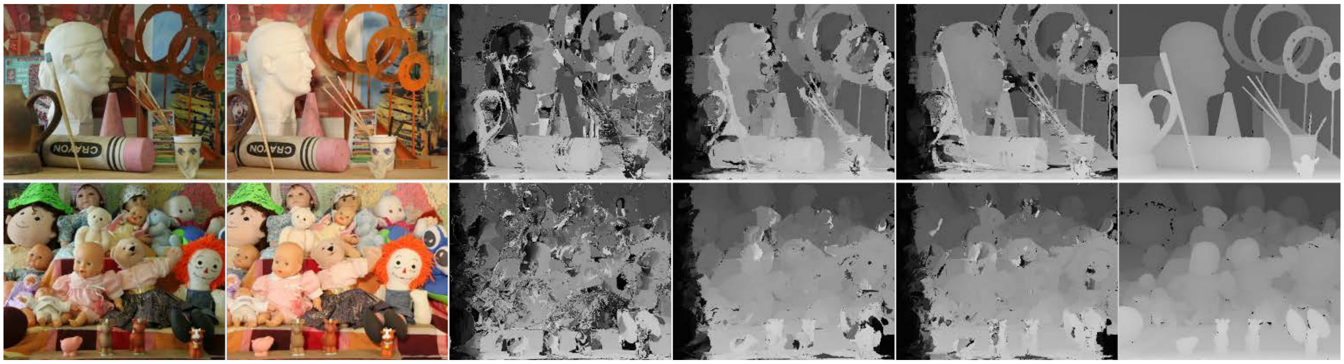
$$D_P(f_p) = (1 - \alpha) \cdot \nabla_x I(f_p) + \alpha \cdot Census(f_p) \quad (13)$$

where $\nabla_x I$ is the SAD between two correspondences that are gradient values of both input stereo images, $Census$ is the Hamming distance, and α is the constant value that balances the SAD and the Hamming distance.

IV. Experiment Result

Two types of stereo images were used for the experiment. To validate the proposed method, two different cost functions were also implemented [2], [7]. All the matching costs are aggregated by the cross-scale cost aggregation method [8].

Fig. 4 represents stereo matching results. In Fig. 4, the test image in the first row is *Art* and that of the second row is *Dolls*. Fig. 5 and Fig. 6 show error rate graphs of each test image. The x axis represents the radiometric variation between the left and the right viewpoint images. The y axis means the error rate.



(a) Left image (b) Right image (c) Intensity+gradient [7] (d) ANCC [2] (e) Our method [3] (f) Ground truth

Figure 4. Stereo matching results

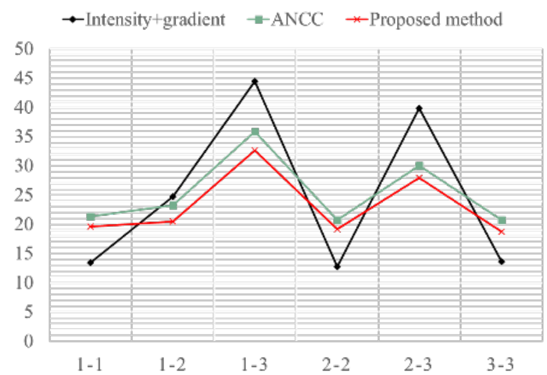


Figure 5. Error rate comparison (Art)

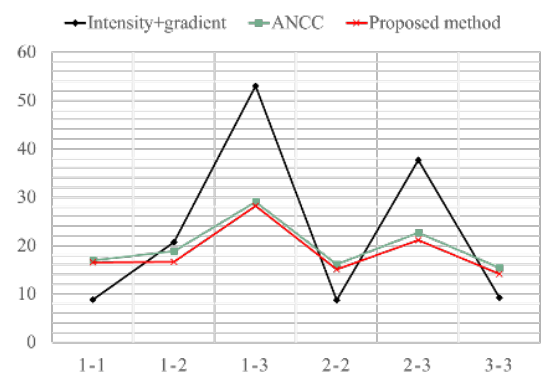


Figure 6. Error rate comparison (Dolls)

In Fig. 5 and Fig. 6, error rates of the proposed method have the most stable results compared with other results. We also compared the cost computation time among those algorithms. Table 1 represents the cost computation time comparison. In Table 1, the proposed method is faster than the ANCC.

Table 1. COST COMPUTATION TIME COMPARISON

Algorithms	**Time (Sec.)**
Intensity + gradient [7]	0.4
ANCC [2]	121.29
Proposed method [3]	40.84

V. CONCLUSION

The adaptive normalized cross correlation (ANCC) estimates the disparity value well under different radiometric conditions. However, it has a high complexity problem. In order to improve this problem, the proposed method uses the global mean value and the census transformation. The global mean value removes the scale factors from the brightness factor subtracted color value. The census transformation is also used for the similarity measure that is independent from the gamma exponent. The matching cost between two census transformed results is calculated by the Hamming distance. The final cost function is composed of the Hamming distance value and the SAD of input gradient images. As a result, the proposed method has faster and more stable results than the ANCC method.

REFERENCES

[1] Y. J. Chang and Y. S. Ho, “Complexity reduction for local stereo matching method using modified SAD algorithm,” *The Korean Society of Broad Engineers Fall Conference*, pp. 218-228, Nov. 2011.

[2] Y. S. Heo, K. M. Lee, and S. U. Lee, “Illumination and camera invariant stereo matching,” *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, Aug. 2008.

[3] Y. J. Chang and Y. S. Ho, “Robust stereo matching to radiometric variation using binary information of census transformation,” *The Korean Institute of Broadcast and Media Engineers Fall Conference*, pp. 1-2, Nov. 2016.

[4] G. D. Finlayson and S. D. Hordley, “Illuminant and gamma comprehensive normalisation in log RGB space,” *Pattern Recognition Letters*, vol. 24, no. 11, pp. 1679-1690, July 2003.

[5] C. Tomasi and R. Manduchi, “Bilateral filtering for gray and color images,” *IEEE Sixth International Conference on Computer Vision*, pp. 839-846, Jan. 1998.

[6] R. Zabih and J. Woodfill, “Non-parametric local transforms for computing visual correspondence,” *European Conference on Computer Vision*, pp. 151-158, May 1994.

[7] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz, “Fast cost-volume filtering for visual correspondence and beyond,” *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1402-1409, June 2011.

[8] K. Zheng, Y. Fang, D. Min, L. Sun, S. Yang, S. Yan, and Q. Tian, “ Cross-scale cost aggregation for stereo matching,” *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1590-1597, June 2014.