A method for extracting 3D-element based on local cost metrics resistant to radiometric distortion for realistic media application

Kwangmu Shin✉ and Sunghoon Kim
Electronics and Telecommunications Research Institute, Daejeon, Republic of Korea
✉E-mail: kmshin@etri.re.kr

The intensity values of pairs of stereoscopic images could be affected by radiometric elements such as camera exposure, illumination direction, etc. Therefore, it is a challenging problem to get consistent and exact disparity values under a variety of real-world environment. Here, the proposed approach extracts disparity values based on a local cost metric and is resistant to global and local radiometric changes. It is designed in a way that disparity values are extracted using local uniformity of pixel values between pairs of stereoscopic images. The proposed metric-based method is defined as the resistant local cost metric (RLCM) model. As a result of the experiment, the proposed model performed better than the comparative methods under experimental conditions in which the radiometric distortion between pairs of stereoscopic images was significantly different. The average peak signal-to-noise ratio (PSNR) value of the RLCM method is 20.0 dB under the global radiometric change condition and 19.3 dB under the local condition, showing the highest value compared to the comparison group. Consequentially, it is demonstrated that the proposed method is less sensitive to various radiometric changes.

Introduction: 3D-element extraction method using stereo vision targets 2D images acquired using two cameras configured horizontally at a close distance in most cases, and extracts relative disparity values by applying the principle of binocular parallax [1]. It is common that pairs of stereoscopic images acquired in the real-world environment are exposed to conditions where the amount of illumination is not uniform due to radiometric distortion, and such conditions make it more difficult to accurately predict disparity values [2–5]. The main factors affecting the change of illumination can be largely classified into two types. First, there is an overall change that has a relatively uniform difference in intensity across the entire image between the pairs of stereoscopic images. Next, there is a local change having a dispersive and non-uniform difference in intensity. However, previous studies adapting to change of illumination have problems in several respects. First of all, previous studies generally calculate disparity values with relatively high accuracy for overall change of illumination, but the accuracy of disparity values is low for local change where it is difficult to predict relatively accurate disparity values. Next, previous studies adopt global methods and deep learning structures that can predict disparity values with relatively high accuracy in the process of selecting the optimal disparity value for each pixel unit [6–8], but the global method has a problem in that it is vulnerable to local change of illumination due to the optimization process of the entire image, and the deep learning structure has a problem of high absolute computational cost for learning.

Therefore, here, we propose a method that is resistant to global and local changes of illumination by considering the computational load in pairs of stereoscopic images under conditions where the intensity is not uniform. This method is expected to be applicable to a module that efficiently acquires 3D-element of a realistic media application system in a real-world environment. For example, it can be applied to media applications such as stereoscopic 3D OTT/OTA media service, 360 VR media service, and 3D cinema. In addition, it has been observed that various studies can improve the performance of the proposed method or expand the range of applications, and those are considered as subjects of future works. The scene graph technique, which infers relationships among objects in a scene, can contribute to the development of applications for 3D elements extracted through the proposed method [9]. A transferable network architecture for zero-shot temporal activity detection can contribute to improving the generalizability of the proposed method [10]. A novel unsupervised multi-modal machine translation approach can be applied to improve the robustness of the disparity extraction process [11]. Differentiable generative adversarial networks search for zero-shot learning can provide valuable insights into designing more flexible and adaptable models for handling unseen radiometric distortions [12].

Extraction of disparity values based on local uniformity of intensity: The proposed cost model uses the characteristic that the difference values of intensity of each image in pair of stereoscopic images is relatively uniform based on the local region. It is necessary to confirm the objective basis through observation and analysis of the local uniformity of the difference values of intensity. The skip factor is used to implement the local uniformity of the difference values of intensity in the metric of the cost model, and is defined as a uniform distance between neighbouring pixels in each window in the both side images. Even if the intensity values of a specific region change due to a global or local change in the amount of illumination in the both side images, the difference values of intensity between neighbouring pixels in the window has a relatively uniform value based on a specific skip size. I_L represents the intensity value of the left image, I_R represents the intensity value of the right image, (i, j) represents the pixel position, s means the size of the skip, and d means the disparity value. The skip size from 1 to less than the window size is sequentially applied to calculate the ratio of pixels where the difference values of intensity is less than 5% of the maximum intensity value, and if the ratio is the highest, the skip size of the proposed cost model is decided. In addition, the local uniformity of the difference values of intensity was numerically observed in multiple stereoscopic images (Baby1, Baby2, Cloth1, Flowerpots etc.) under conditions where the amount of illumination was not uniform. The ratio of pixels in which the difference values of intensity is less than 5% of the maximum intensity value in the image is defined as r, and the higher the value, the higher the local uniformity. As a result, an average of about 93.0% of the r value in the global change of illumination and an average of about 97.4% in the local change were observed based on the skip size of 3 and the horizontal direction. The result of this observation can serve as an objective basis for the fact that the difference values of intensity between neighbouring pixels within a window in both side images has a relatively uniform value based on a specific skip size, even if the intensity value of a specific region is different.

Isotropic expansion based on the local uniformity of intensity: The proposed cost model uses vertical and diagonal components as representative components in addition to horizontal component to improve the accuracy of disparity values. This means the isotropic expansion based on the local uniformity of intensity, and is based on the fact that the local uniformity of the window-based difference values of intensity is maintained as major directional components. Experimental observation and analysis of vertical and diagonal direction components were performed under the same conditions as observation of the horizontal direction for the local uniformity of the difference values of intensity. An average of about 91.9% of the r value in the global change of illumination and an average of about 95.6% in the local change were observed based on the skip size of 3 and the vertical direction. Based on the skip size of 5, an average r value of about 88.3% in global change and an average of about 95.3% in local change was observed. In the diagonal direction, an average of about 92.0% of the r value in the global change of illumination and an average of about 95.6% in the local change were observed based on the skip size of 3. Based on the skip size of 5, an average r value of about 88.4% in global change and an average of about 95.3% in local change was observed. This experimental observation serves as an objective basis for isotropic expansion in which the window-based difference values of intensity maintain locally uniform values based on the main directional components.

Local scalling to reduce edge fattening: The proposed cost model applies local scaling based on the average pixel intensity of window to improve resistance against local change. The adoption of local scaling has the advantage of minimizing the computational load and maintaining the basic structure of this cost model. However, the uniform local scaling method has a problem that it is easily exposed to edge fattening that occurs when disparity values are calculated incorrectly near the boundary of an object. Therefore, the cost model proposed to solve this problem applies a method of adaptively performing local scaling by setting a threshold based on the cost of each direction component. Figure 1 shows

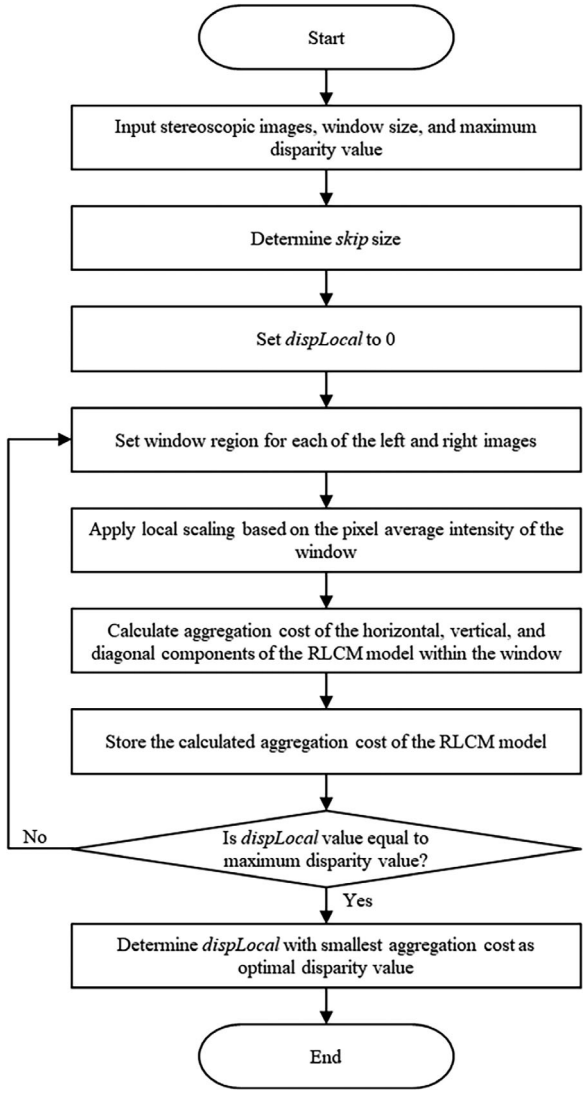

Fig. 1 *The algorithm of the resistant local cost metric model.*

the flow chart of the algorithm of the proposed RLCM model sequentially from data input to optimal variance output. The detailed process of the RLCM model, which isotropically expands based on the local uniformity of intensity and applies local scaling, is represented procedurally.

Equations (1)–(3) represents the process of adding local characteristics by applying the average value of the intensity in the corresponding window based on the horizontal component, and $\overline{C_H}$ means the aggregation cost of the horizontal component. C_{HL} and C_{HR} represent the difference values of intensity based on windows in the left image and the right image, respectively, and m_{PL} and m_{PR} represent average values of intensity within windows in the left and right images, respectively. $I_L(i,j+s)$ represents the intensity value of a pixel moved from (i, j) by the s with the horizontal direction in the window of the left image and $I_R(i,j+d+s)$ means the intensity value of a pixel moved by the d and s with the horizontal direction based on (i, j) in the window of the right image.

$$C_{HL} = |I_L(i, j) - I_L(i, j+s)| \ . \tag{1}$$

$$C_{HR} = |I_R(i, j+d) - I_R(i, j+d+s)| \ . \tag{2}$$

$$\overline{C_H} = \begin{cases} \sum\limits_{(i,j)\in P} \left(C_{HL} - C_{HR}\frac{m_{PL}}{m_{PR}} \right) \\ \sum\limits_{(i,j)\in P} \left(C_{HR} - C_{HL}\frac{m_{PR}}{m_{PL}} \right) \end{cases}, \quad \begin{matrix} \text{if} & C_{HL} \geq C_{HR} \\ \text{otherwise} & C_{HL} < C_{HR} \end{matrix} \tag{3}$$

Equations (4)–(6) represents the process of adding local characteristics by applying the average value of the intensity in the corresponding window based on the vertical component, and $\overline{C_V}$ means the aggregation cost of the vertical component.

Table 1. The comparison results of peak signal-to-noise ratio values under the global radiometric change (dB).

Methods	Flowpots	Rocks1	Wood1	Baby1	Baby2	Cloth1
LSAD	14.4	19.5	15.4	19.3	18.6	20.6
LSSD	14.7	19.7	15.6	19.3	19.2	20.6
ZSAD	9.5	10.4	10.6	10.6	11.6	17.7
ZSSD	9.5	10.4	10.7	10.6	11.7	18.1
ZNCC	17.3	20.9	18.6	20.8	21.1	20.6
ANCC	14.9	19.0	17.4	20.2	18.1	18.8
RLCM	17.6	21.0	18.9	21.5	20.4	20.7

$$C_{VL} = |I_L(i, j) - I_L(i+s, j)| \ . \tag{4}$$

$$C_{VR} = |I_R(i, j+d) - I_R(i+s, j+d)| \ . \tag{5}$$

$$\overline{C_V} = \begin{cases} \sum\limits_{(i,j)\in P} \left(C_{VL} - C_{VR}\frac{m_{PL}}{m_{PR}} \right) \\ \sum\limits_{(i,j)\in P} \left(C_{VR} - C_{VL}\frac{m_{PR}}{m_{PL}} \right) \end{cases}, \quad \begin{matrix} \text{if} & C_{VL} \geq C_{VR} \\ \text{otherwise} & C_{VL} < C_{VR} \end{matrix} \tag{6}$$

Equations (7)–(9) represents the process of adding local characteristics by applying the average value of the intensity in the corresponding window based on the diagonal component, and $\overline{C_D}$ means the aggregation cost of the diagonal component.

$$C_{DL} = |I_L(i, j) - I_L(i+s, j+s)| \ . \tag{7}$$

$$C_{DR} = |I_R(i, j+d) - I_R(i+s, j+d+s)| \ . \tag{8}$$

$$\overline{C_D} = \begin{cases} \sum\limits_{(i,j)\in P} \left(C_{DL} - C_{DR}\frac{m_{PL}}{m_{PR}} \right) \\ \sum\limits_{(i,j)\in P} \left(C_{DR} - C_{DL}\frac{m_{PR}}{m_{PL}} \right) \end{cases}, \quad \begin{matrix} \text{if} & C_{DL} \geq C_{DR} \\ \text{otherwise} & C_{DL} < C_{DR} \end{matrix} \tag{9}$$

As a result, $\overline{RLCM}$, the final cost model that provides reliability in the pairs of stereoscopic images under conditions where the intensity is not uniform is expressed as follows:

$$\overline{RLCM} = \overline{C_H} + \overline{C_V} + \overline{C_D}. \tag{10}$$

Experimental results: In this experiment, the proposed RLCM model was evaluated using the Middlebury stereo image dataset [1] through experiments under various conditions. Test-bed images with various features such as Baby1, Baby2, Cloth1, Flowerpots, Rocks1, and Wood1 were used. The comparison group includes zero-mean sum of squared differences (ZSSD) using local domain-based features, locally scaled sum of squared differences (LSSD), locally scaled sum of absolute differences (LSAD), and zero-mean sum of absolute differences (ZSAD). Also, based on the global domain, zero-mean normalized cross correlation (ZNCC) and adaptive normalized cross correlation (ANCC) are included [13]. The accuracy measurement method consisted of objective and subjective quality evaluation. In objective quality evaluation, the peak signal to noise ratio method and the structural similarity method are used.

Compared to the comparison group under the global radiometric change condition, the average value of the structural similarity index measure (SSIM) index of the RLCM method was 0.819, which is the highest value. Compared to 0.787 of the ANCC method, the proposed model has a high value of about 4%. Compared to the ANCC method, which can be directly compared in terms of the resistance of the radiometric change, the proposed model shows high accuracy of the disparity value at the boundary or inside the object.

Table 1 shows the comparison results of the PSNR values under the global radiometric change condition. Compared to the comparison group, the average value of the PSNR of the RLCM method was 20.0 dB, which is the highest value. Compared to 18.1 dB of the ANCC method, the proposed model has a high value of about 10%.

Figure 2 shows the results of the RLCM model and comparison group under the global radiometric change condition. Figures 2a and 2b show

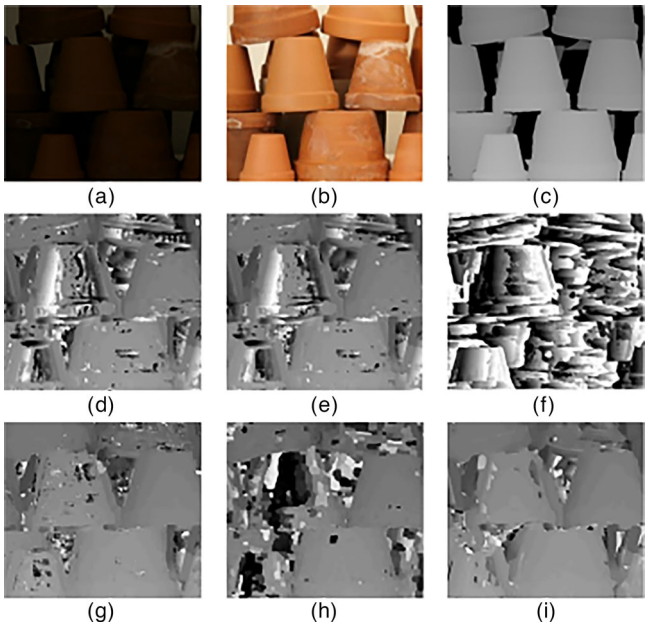

Fig. 2 *The comparison results of the resistant local cost metric model under the global radiometric change (Flowerpots): (a) left image. (b) right image. (c) ground truth. (d) locally scaled sum of absolute differences. (e) locally scaled sum of squared differences. (f) zero-mean sum of squared differences. (g) zero-mean normalized cross correlation. (h) adaptive normalized cross correlation. (i) resistant local cost metric.*

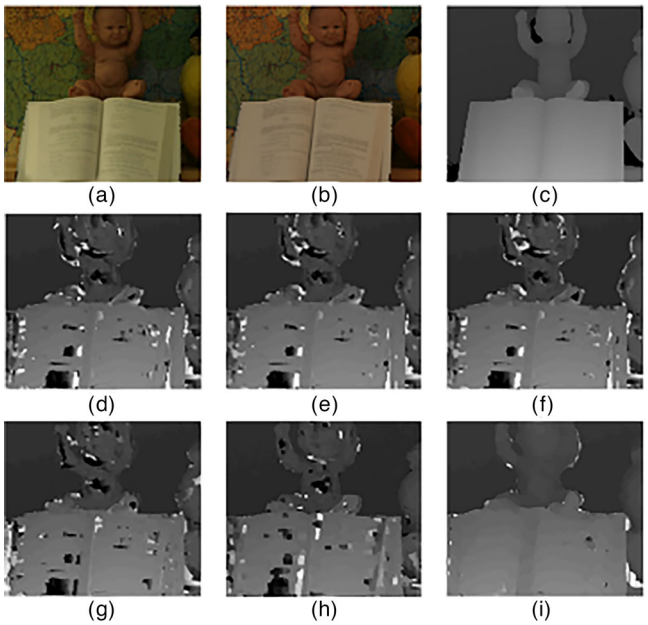

Fig. 3 *The comparison results of the resistant local cost metric model under the local radiometric change (Baby2): (a) left image. (b) right image. (c) ground truth. (d) locally scaled sum of absolute differences. (e) locally scaled sum of squared differences. (f) zero-mean sum of squared differences. (g) zero-mean normalized cross correlation. (h) adaptive normalized cross correlation. (i) resistant local cost metric.*

Table 2. The comparison results of peak signal-to-noise ratio values under the local radiometric change (dB).

Methods	Flowpots	Rocks1	Wood1	Baby1	Baby2	Cloth1
LSAD	10.8	18.9	15.3	19.1	18.0	20.6
LSSD	11.0	19.0	15.6	19.1	18.3	20.6
ZSAD	10.2	14.6	14.9	18.4	17.9	20.5
ZSSD	10.4	14.7	15.2	18.3	18.2	20.6
ZNCC	11.5	19.6	16.3	19.3	18.4	20.6
ANCC	12.1	18.7	16.5	19.8	17.8	18.8
RLCM	15.1	19.9	17.7	20.9	21.7	20.6

the left and right input images, respectively, and 2c is the ground truth image that serves as a criterion for subjective evaluation. In order to apply the global radiometric change difference between the left and right images, the camera exposure level is set differently for each image. Figure 2d–i shows the disparity images of the comparison group and the RLCM model. By subjective comparison with the ground truth, it can be confirmed that the disparity images of the RLCM model show higher accuracy than the methods of the comparison group.

Compared to the comparison group under the local radiometric change condition, the average value of the SSIM index of the RLCM method was 0.769, which is the highest value. Compared to 0.749 of the ANCC method, the proposed model has a high value of about 2%.

Table 2 shows the comparison results of the PSNR values under the local radiometric change condition. Compared to the comparison group, the average value of the PSNR of the RLCM method was 19.3 dB, which is the highest value. Compared to 17.3 dB of the ANCC method, the proposed model has a high value of about 12%.

Figure 3 shows the results of the RLCM model and comparison group under the local radiometric change condition. In order to apply the local radiometric change difference between the left and right images, the illumination level is set differently for each image. Figure 3d–i shows the disparity images of the comparison group and the RLCM model. By subjective comparison with the ground truth, it can be confirmed that the disparity images of the RLCM model show higher accuracy than the methods of the comparison group.

Conclusion: The proposed RLCM model has an approach for extracting disparity values based on local cost metrics that is resistant to local and global radiometric changes. This method is particularly strong in local change having a dispersive and non-uniform difference in intensity. It is difficult to predict a relatively accurate disparity values compared to the global change of illumination, even if various optimization methods such as belief-propagation and graph-cut are used. Experimental results showed that the proposed model performed better than the comparison group under experimental conditions with differences in the radiometric distortion between pairs of stereoscopic images. It is demonstrated that the average PSNR value of the RLCM method is improved by about 10% under the global radiometric change condition and about 12% under the local condition compared to the ANCC method. The exploration of potential applications where the proposed method can be applied and performance evaluation based on large-scale datasets are being considered as future works.

Author Contributions: Kwangmu Shin: Conceptualization, Formal analysis, Investigation, Methodology, Software, Writing—original draft, Writing—review and editing. Sunghoon Kim: Project administration, Supervision, Writing—review and editing

Acknowledgements: This work was supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government. [23ZH1200, The research of the basic media**.**contents technologies]

Conflict of Interest: The authors declare no conflict of interest.

Data Availability Statement: Research data are not shared.

Received: *29 March 2023* Accepted: *6 August 2023*
doi: 10.1049/ell2.12916

References

1 Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vision* **47**(1), 7–42 (2002)

2 Hirschmuller, H., Scharstein, D.: Evaluation of stereo matching costs on images with radiometric differences. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(9), 1582–1599 (2009)

3 Jung, I.L., Chung, T.Y., Sim, J.Y., Kim, C.S.: Consistent stereo matching under varying radiometric conditions. *IEEE Trans. Multimedia* **15**(1), 56–69 (2013)

4 Zhou, X., Boulanger, P.: Radiometric invariant stereo matching based on relative gradients. In: *2012 19th IEEE International Conference on Image Processing (ICIP)*. Orlando, FL, pp. 2989–2992 (2012)

5 Shin, K.M., Kim, D.K., Chung, K.D.: Visual stereo matching combined with intuitive transition of pixel values. *Multimed. Tools Appl.* **75**(23), 15381–15403 (2016)

6 Zhang, C., Li, Z., Cheng, Y., Cai, R., Chao, H., Rui, Y.: MeshStereo: A global stereo model with mesh alignment regularization for view interpolation. In: *2015 IEEE International Conference on Computer Vision (ICCV)*. Santiago, Chile, pp. 2057–2065 (2015)

7 Yoon, K.: Stereo matching based on nonlinear diffusion with disparity-dependent support weights. *IET Comput. Vision* **6**(4), 306–313 (2012)

8 Laga, H., Jospin, L.V., Boussaid, F., Bennamoun, M.: A survey on deep learning techniques for stereo-based depth estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**(4), 1738–1764 (2022)

9 Chang, X., Ren, P., Xu, P., Li, Z., Chen, X., Hauptmann, A.: A comprehensive survey of scene graphs: Generation and application. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**(1), 1–26 (2023)

10 Zhang, L., Chang, X., Liu, J., Luo, M., Li, Z., Yao, L., Hauptmann, A.: TN-ZSTAD: Transferable network for zero-shot temporal activity detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**(3), 3848–3861 (2023)

11 Li, M., Huang, P.Y., Chang, X., Hu, J., Yang, Y., Hauptmann, A.: Video pivoting unsupervised multi-modal machine translation. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**(3), 3918–3932 (2023)

12 Yan, C., Chang, X., Li, Z., Guan, W., Ge, Z., Zhu, L., Zheng, Q.: ZeroNAS: Differentiable generative adversarial networks search for zero-shot learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**(12), 9733–9740 (2022)

13 Heo, Y.S., Lee, K.M., Lee, S.U.: Robust stereo matching using adaptive normalized cross-correlation. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(4), 807–822 (2011)