ROBUST STEREO MATCHING UNDER RADIOMETRIC VARIATIONS BASED ON CUMULATIVE DISTRIBUTIONS OF GRADIENTS

Il-Lyong Jung[1,2], *Jae-Young Sim*[3], *Chang-Su Kim*[2], *and Sang-Uk Lee*[4]

[1] Defense Agency for Technology and Quality, Seoul, Korea
[2] School of Electrical Engineering, Korea University, Seoul, Korea
[3] School of Electrical and Computer Engineering,
Ulsan National Institute of Science and Technology, Ulsan, Korea
[4] School of Electrical Engineering, Seoul National University, Seoul, Korea

E-mails: illyong@korea.ac.kr, jysim@unist.ac.kr, changsukim@korea.ac.kr, sanguk@ipl.snu.ac.kr

ABSTRACT

We propose a robust stereo matching algorithm for images captured under varying radiometric conditions, such as exposure and lighting variations, based on the cumulative distributions of gradients. The gradient operator extracts local changes in pixel values, which are less sensitive to radiometric variations than the original pixel values. Moreover, the cumulative distribution function (CDF) of gradient vectors reflects the ranks of edge strength levels, and corresponding pixels in stereo images tend to have similar ranks regardless of radiometric conditions. Therefore, we design the matching cost function based on the dissimilarity of gradient CDF values. However, since multiple pixels in an image may have the same gradient CDF value, we further constrain the correspondence matching by checking the dissimilarity of gradient orientations. Finally, to estimate an accurate disparity at each pixel, we adaptively aggregate matching costs using the color similarity and the geometric proximity of neighboring pixels. Experimental results demonstrate that the proposed algorithm provides more accurate disparities than conventional algorithms, especially under varying lighting conditions.

Index Terms— Stereo matching, 3-D image processing, radiometric variations, cumulative distribution function, and gradient-based rank matching.

1. INTRODUCTION

Recently, there has been an increasing demand for realistic multimedia applications such as three-dimensional (3-D) video. Depth maps are used to describe 3-D scenes, which are often obtained from pairs of images through stereo matching. A typical stereo matching algorithm estimates the disparity between corresponding pixels in stereo images, based on the assumption that the corresponding pixels should have similar colors [1]. However, when two images are captured under varying radiometric conditions, such as exposure and lighting, they may exhibit different color characteristics and hence conventional stereo matching algorithms cannot provide accurate disparity maps in general [2]. Figs. 1(a) and (b) show the stereo pair of 'Aloe' images with different exposure conditions, which result in a noisy disparity map in Fig. 1(c).

(a) (b) (c)

Fig. 1. Stereo matching of the 'Aloe' images under varying radiometric conditions [1]: (a) the left view with exposure level 1, (b) the right view with exposure level 2, and (c) the resulting disparity map.

Several approaches have been proposed to alleviate the effects of radiometric variations for robust stereo matching. First, as a preprocessing step of stereo matching, illumination changes can be compensated by normalizing or transforming color values [3, 4, 5]. The comprehensive color normalization [3] iteratively normalizes colors to remove the effects of global illumination and the dependency on scene geometry. The grey-edge algorithm [4] employs the average color of edge differences for the color normalization. The adaptive color transform [5] finds pseudo-corresponding pixels based on the rank matching and then transforms the color of each pixel to be consistent with that of the corresponding pixel. Second, color invariant matching costs can be employed [6, 7]. The normalized cross correlation [6] compensates for the gain and the bias in color values between stereo images. The adaptive normalized cross correlation [7] is also used, which combines the idea of the comprehensive color normalization with that of the normalized cross correlation. However, although these algorithms compensate for global exposure differences effectively, they still suffer from local radiometric changes such as lighting variations.

In this work, we propose a robust stereo matching algorithm under varying radiometric conditions, which exploits the cumulative distributions of gradients. A gradient, which represents the difference of a pixel value from its neighboring pixel values, can extract an edge feature reliably regardless of radiometric variations. Moreover, the cumulative distribution function (CDF) of gradients describes the ranks of those gradients. Note that the ranks are more robust against radiometric variations than the gradients themselves. Therefore, we

This work was supported partly by the Global Frontier R&D Program on Human-centered Interaction for Coexistence funded by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIP) (2011-0031648), and partly by the Basic Science Research Program through the NRF funded by the Ministry of Education, Science and Technology (MEST) (2012-0003908).

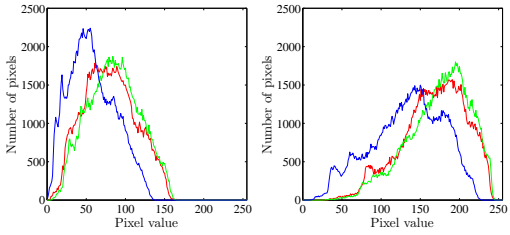


Fig. 2. The histograms of red, green, and blue color values of the stereo images in Fig. 1: (a) the left view and (b) the right view.

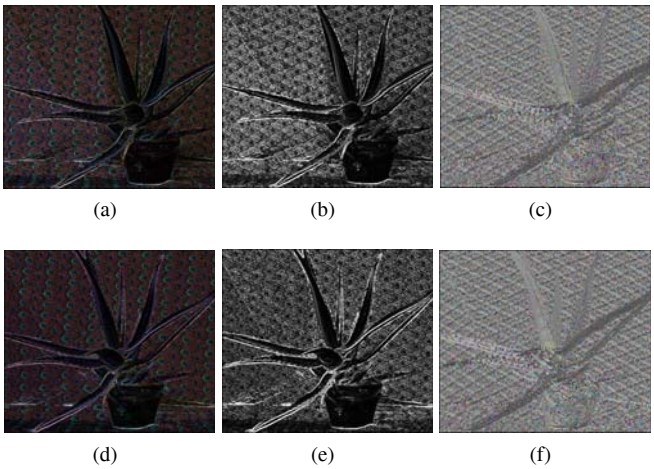


Fig. 3. The gradient features of the stereo images in Fig. 1. (a) The gradient map, (b) the gradient CDF map, and (c) the gradient orientation map of the left view. (d) The gradient map, (e) the gradient CDF map, and (f) the gradient orientation map of the right view.

design a robust matching cost function based on the CDF of gradients. Also, we incorporate gradient orientations into the cost function to avoid the ambiguity of stereo matching when multiple pixels have the same gradient rank. Experimental results demonstrate that the proposed algorithm improves the accuracy of a disparity map, as compared with conventional stereo matching algorithms [1, 5, 7], under varying exposure and illumination conditions.

This paper is organized as follows. Section 2 proposes the stereo matching algorithm using the CDF of gradients, Section 3 discusses experimental results, and Section 4 provides concluding remarks.

2. PROPOSED ALGORITHM

2.1. Cumulative Distribution of Gradients

Figs. 2(a) and (b) show the distributions of color values of the stereo images in Figs. 1(a) and (b), respectively. Although the stereo images capture the same scene, they yield different color distributions due to the different exposure levels. Consequently, conventional stereo matching algorithms using original pixel values may not find matching pixels correctly.

The gradient operator measures the changes in adjacent pixel values, and it yields large responses on object edges. Figs. 3(a) and (d) represent the three gradient maps of the red, green, and blue color components for the stereo images in Figs. 1(a) and (b), respectively, which are obtained using the Sobel operator. A bright color represents a large gradient value in each color channel. Notice that edges in each image are well preserved even under the exposure variations. In other words, Fig. 3(a) exhibits quite similar characteristics to Fig. 3(d), although Fig. 1(a) has a different exposure level than Fig. 1(b). Therefore, we employ the gradient operator to generate reliable features for stereo matching.

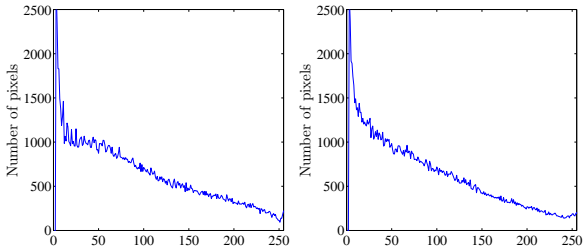


Fig. 4. The histograms of the gradient CDF values of the stereo images in Fig. 1: (a) the left view and (b) the right view.

Furthermore, notice that the cumulative histogram of pixel values represents the ranks (or orders) of the pixel values, and the ranks of corresponding pixels in stereo images are similar to each other, regardless of radiometric variations [5]. To extend this idea, we obtain the CDFs of gradient vectors in left and right images, respectively, and take the difference of the CDF values of two pixels as the matching cost for the stereo matching.

Let R, G and B denote the random variables that represent the gradient magnitudes for red, green and blue color components, respectively. Then, the joint CDF $H_{R,G,B}(r,g,b)$ of the gradient vector (r,g,b) is given by

$$H_{R,G,B}(r,g,b) = \sum_{i=0}^{r}\sum_{j=0}^{g}\sum_{k=0}^{b} h_{R,G,B}(i,j,k), \tag{1}$$

where $h_{R,G,B}(i,j,k)$ represents the joint probability mass function of gradient magnitudes. Assuming that the three color components are independent of one another [8, 5], we set

$$h_{R,G,B}(r,g,b) = h_R(r)h_G(g)h_B(b) \tag{2}$$

where $h_R(r)$, $h_G(g)$, and $h_B(b)$ denote the marginal probability mass functions of the three components.

Figs. 3(b) and (e) illustrate the CDF maps that correspond to the gradient maps in Figs. 3(a) and (d), respectively. These CDF maps are scaled to the range of $[0, 255]$. Notice that the gradient CDF $H_{R,G,B}(r,g,b)$ represents the rank of the gradient vector (r,g,b). Thus, edge regions have larger gradient CDF values than textureless regions. Also, the gradient CDF values of corresponding pixels in stereo images tend to be similar, regardless of radiometric variations. Figs. 4(a) and (b) compare the distributions of the gradient CDF values in Figs. 3(b) and (e). Note that the stereo images yield similar distributions of the CDF values, whereas they have different distributions of the original color values in Figs. 2(a) and (b).

Let $M_0(\mathbf{p})$ and $M_1(\mathbf{p})$ denote the gradient CDF values of pixel $\mathbf{p}$ in the left and the right images, respectively. Based on the above observations, we design the gradient CDF cost $D_M(\mathbf{p},\mathbf{d})$ for stereo matching, which is given by

$$D_M(\mathbf{p},\mathbf{d}) = |M_0(\mathbf{p}) - M_1(\mathbf{p}-\mathbf{d})|, \tag{3}$$

where $\mathbf{d}$ is the disparity between pixel $\mathbf{p}$ in the left image and pixel $(\mathbf{p}-\mathbf{d})$ in the right image.

2.2. Gradient Orientation

Multiple pixels in an image may have the same gradient CDF value. This causes the ambiguity in finding matching pixels. To overcome this problem, we additionally use the gradient orientation to design the matching cost. Let $\theta_{0,R}(\mathbf{p})$, $\theta_{0,G}(\mathbf{p})$, and $\theta_{0,B}(\mathbf{p})$ denote the gradient orientations of the red, green, and blue color components at pixel $\mathbf{p}$ in the left image, and $\theta_{1,R}(\mathbf{p})$, $\theta_{1,G}(\mathbf{p})$, and $\theta_{1,B}(\mathbf{p})$ in the right image, respectively. Figs. 3(c) and (f) render the gradient orientation maps for the red, green, and blue color components together, where the angle of each gradient orientation in $[0, 2\pi]$ is linearly scaled to the range of pixel values $[0, 255]$. It is observed that corresponding pixels tend to exhibit similar gradient orientations.

Hence we compute the dissimilarity measure of the gradient orientations for the red color component as

$$\Theta_R(\mathbf{p},\mathbf{d}) = 1 - \cos\left(\theta_{0,R}(\mathbf{p}) - \theta_{1,R}(\mathbf{p}-\mathbf{d})\right). \tag{4}$$

The dissimilarity measures $\Theta_G(\mathbf{p},\mathbf{d})$ and $\Theta_B(\mathbf{p},\mathbf{d})$ for the green and the blue color components are defined similarly. Then, we define the gradient orientation cost $D_O(\mathbf{p},\mathbf{d})$ as

$$D_O(\mathbf{p},\mathbf{d}) = \Theta_R(\mathbf{p},\mathbf{d}) + \Theta_G(\mathbf{p},\mathbf{d}) + \Theta_B(\mathbf{p},\mathbf{d}). \tag{5}$$

2.3. Matching Cost Aggregation

We define the overall matching cost $D(\mathbf{p},\mathbf{d})$ by combining the gradient CDF cost and the orientation cost together, which is given by

$$D(\mathbf{p},\mathbf{d}) = \min\{D_M(\mathbf{p},\mathbf{d}) + \zeta D_O(\mathbf{p},\mathbf{d}), \tau\} \tag{6}$$

where ζ is a weighting parameter and τ is a truncation value. Note that stereo matching is an ill-posed problem due to the inherent ambiguities in noisy and textureless regions. To overcome these ambiguities, we aggregate matching costs based on the adaptive weight scheme in [9]. We first compute the color similarity $s(\mathbf{p},\mathbf{q})$ between pixel $\mathbf{p}$ and its neighboring pixel $\mathbf{q}$. Also, we obtain the geometric proximity $e(\mathbf{p},\mathbf{q})$ based on the Euclidean distance between $\mathbf{p}$ and $\mathbf{q}$. Then, the adaptive weight $w(\mathbf{p},\mathbf{q})$ is defined as

$$w(\mathbf{p},\mathbf{q}) = \exp\left(-\left(\frac{s(\mathbf{p},\mathbf{q})}{\alpha} + \frac{e(\mathbf{p},\mathbf{q})}{\beta}\right)\right), \tag{7}$$

where α and β are user-controllable parameters. The aggregated matching cost is then given by

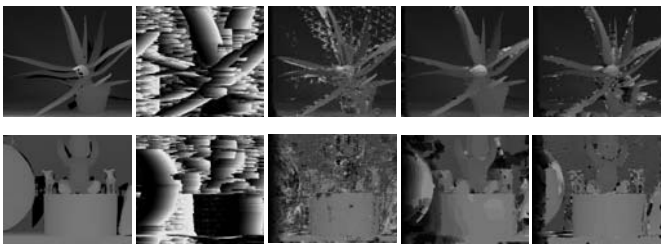
$$\bar{D}(\mathbf{p},\mathbf{d}) = \frac{\sum_{\mathbf{q}\in\mathcal{Q}} w(\mathbf{p},\mathbf{q}) w(\mathbf{p}',\mathbf{q}') D(\mathbf{q},\mathbf{d})}{\sum_{\mathbf{q}\in\mathcal{Q}} w(\mathbf{p},\mathbf{q}) w(\mathbf{p}',\mathbf{q}')} \tag{8}$$

where $\mathcal{Q}$ denotes the supporting window for the cost aggregation, $\mathbf{p}' = \mathbf{p} - \mathbf{d}$, and $\mathbf{q}' = \mathbf{q} - \mathbf{d}$. Finally, we determine the optimal disparity $\mathbf{d}(\mathbf{p})$ for each pixel $\mathbf{p}$ by

$$\mathbf{d}(\mathbf{p}) = \arg\min_{\mathbf{d}} \bar{D}(\mathbf{p},\mathbf{d}). \tag{9}$$

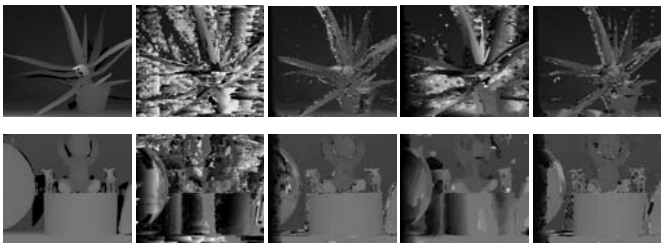
3. EXPERIMENTAL RESULTS

We evaluate the performance of the disparity map estimation on the Middlebury stereo datasets [6]: 'Aloe' and 'Baby3,' which are captured under three different exposure conditions, indexed as 0, 1 and 2, and three different lighting conditions, indexed as 1, 2 and 3. The



(a) (b) (c) (d) (e)

Fig. 5. Comparison of the disparity maps for the 'Aloe' and 'Baby3' datasets under different exposure conditions: (a) ground truth, (b) SAD [1], (c) ANCC [7], (d) ACT [5], and (e) the proposed algorithm. In this test, the exposure levels for the left and the right images are 0 and 2, respectively.



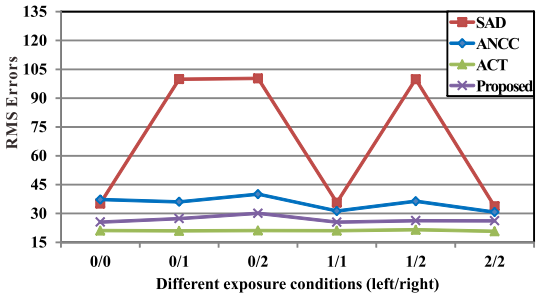
(a) (b) (c) (d) (e)

Fig. 6. Comparison of the disparity maps for the 'Aloe' and 'Baby3' datasets under different lighting conditions: (a) ground truth, (b) SAD [1], (c) ANCC [7], (d) ACT [5], and (e) the proposed algorithm. In this test, the lighting levels for the left and the right images are 1 and 3, respectively.

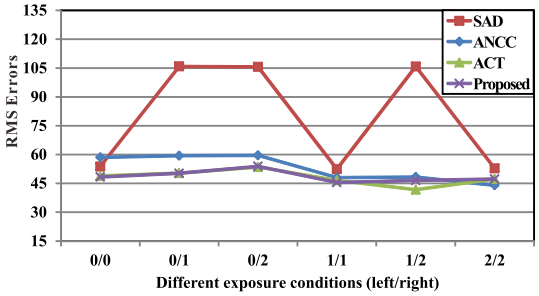
'Aloe' and 'Baby3' datasets have the spatial resolutions of 427×370 and 437×370, respectively. Each dataset contains multiple views, from which two views are selected: view 1 as the left view and view 5 as the right view. In all experiments, the weighting parameter ζ and the truncation value τ in (6) are fixed to 0.033 and 20, and the user-controllable parameters α and β in (7) are fixed to 5 and 9.5, respectively. The size of the support window $\mathcal{Q}$ in (8) is set to 19×19.

We compare the resulting disparity maps of the proposed algorithm with those of the conventional algorithms: the sum of absolute differences (SAD) [1], the adaptive normalized cross correlation (ANCC) [7], and the adaptive color transform (ACT) [5]. The adaptive weight scheme [9] is employed in all the conventional algorithms as well for a fair comparison. Note that ACT is used as a preprocessing step of the stereo matching to alleviate the effect of varying radiometric conditions, whereas ANCC and the proposed algorithm modify the matching costs directly.

Fig. 5 shows the stereo matching results under the varying exposure conditions: exposure levels 0 for the left image and 2 for the right image. This simulates a global change of brightness. SAD yields the worst performance. ANCC yields better performance than SAD, but it still provides low quality disparity maps due to the different exposure levels. Both ACT and the proposed algorithm compensate for the exposure variations effectively and provide comparable disparity maps. In case of 'Aloe,' ACT yields a slightly better performance than the proposed algorithm. However, it is noted that ACT demands significantly higher complexity than the proposed algorithm, since it uses the computationally heavy graph-cut optimization. On the contrary, the proposed algorithm employs the simple

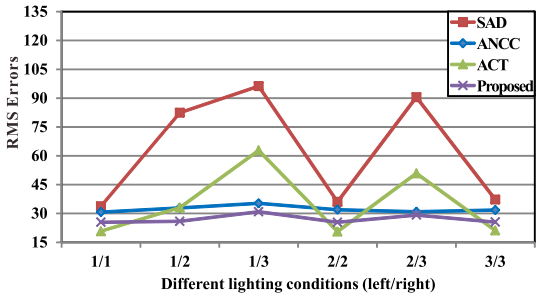


(a)

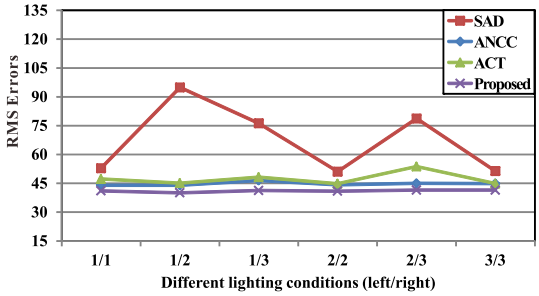


(b)

Fig. 7. Comparison of the RMS errors of the disparity maps under different exposure conditions: (a) 'Aloe' and (b) 'Baby3.'



(a)



(b)

Fig. 8. Comparison of the RMS errors of the disparity maps under different lighting conditions: (a) 'Aloe' and (b) 'Baby3.'

winner-takes-all approach in (9).

Fig. 6 compares the disparity maps under the different lighting conditions: lighting levels 1 for the left image and 3 for the right image. This simulates locally varying brightness. SAD does not find the disparity accurately. In contrast to the exposure variation experiment, ACT provides worse performance than ANCC and the proposed algorithm under the local lighting variations. This is because ACT transforms pixel colors based on the global transform. We see that the proposed algorithm outperforms all the conventional algorithms.

In Fig. 7 and Fig. 8, we quantitatively compare the root mean squared (RMS) errors of the estimated disparity maps. The RMS errors are computed against the ground truth disparity maps. We see that the proposed algorithm yields smaller RMS errors than the conventional algorithms especially under varying lighting conditions, since the gradient CDF values are less sensitive to both global and local radiometric variations than original pixel values.

4. CONCLUSIONS

We proposed a robust stereo matching algorithm under varying radiometric conditions, based on the CDFs of gradient vectors. We designed the matching cost that measured the dissimilarity of the gradient CDF values between matching pixels. Moreover, we further improved the matching accuracy by employing the gradient orientation data. The matching costs were then aggregated based on the color similarity and the geometric proximity among neighboring pixels, and the disparity of each pixel was estimated by the winner-takes-all approach. Experimental results demonstrated that the proposed algorithm reliably estimated disparity maps from stereo images, which were captured under different exposure or lighting conditions.

5. REFERENCES

[1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithm," *Int. J. Comput. Vision*, vol. 47, pp. 7–42, Apr. 2002.

[2] A. Gijsenij, T. Gevers, and J. van de Weijer, "Computational color constancy: survey and experiments," *IEEE Trans. Image Proc.*, vol. 20, pp. 2475–2489, Sept. 2011.

[3] G. D. Finlayson, B. Schiele, and J. L. Crowley, "Comprehensive colour image normalization," in *Proc. ECCV*, June 1998, pp. 475–490.

[4] J. van de Weijer, T. Gevers, and A. Gijsenij, "Edge based color constancy," *IEEE Trans. Image Proc.*, vol. 16, pp. 2207–2214, Sept. 2007.

[5] I.-L. Jung, T.-Y. Chung, J.-Y. Sim, and C.-S. Kim, "Consistent stereo matching under varying radiometric conditions," *IEEE Trans. Multimedia*, vol. 15, pp. 56–69, Jan. 2013.

[6] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *Proc. IEEE CVPR*, July 2007, pp. 1–8.

[7] Y. S. Heo, K. M. Lee, and S. U. Lee, "Robust stereo matching using adaptive normalized cross-correlation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 4, pp. 807–822, Apr. 2011.

[8] P. E. Trahanias and A. N. Venetsanopoulos, "Color image enhancement through 3-D histogram equalization," in *Proc. ICPR*, Aug. 1992, pp. 545–548.

[9] K.-J. Yoon and I. S. Kweon, "Adaptive support-weight approach for correspondence search," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, pp. 650–656, Apr. 2006.