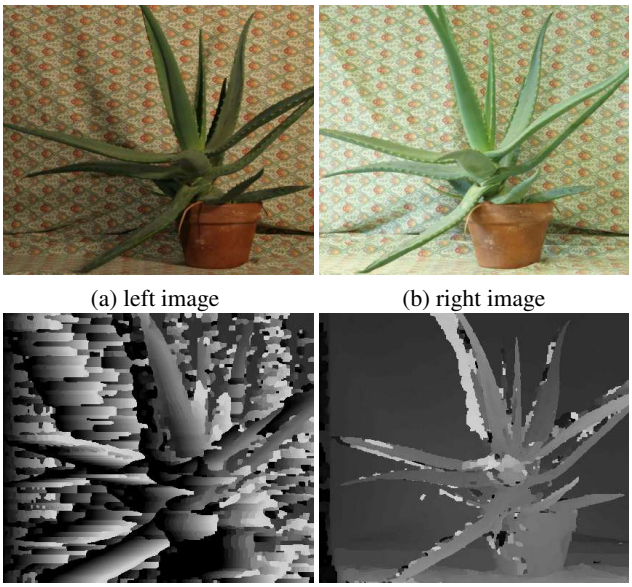
Illumination and Camera Invariant Stereo Matching

Yong Seok Heo, Kyoung Mu Lee, and Sang Uk Lee
School of EECS, ASRI, Seoul National University, 151-742, Seoul, Korea
hys@diehard.snu.ac.kr, kyoungmu@snu.ac.kr, sanguk@ipl.snu.ac.kr

Abstract

Color information can be used as a basic and crucial cue for finding correspondence in a stereo matching algorithm. In a real scene, however, image colors are affected by various geometric and radiometric factors. For this reason, the raw color recorded by a camera is not a reliable cue, and the color consistency assumption is no longer valid between stereo images in real scenes. Hence the performance of most conventional stereo matching algorithms can be severely degraded under the radiometric variations. In this paper, we present a new stereo matching algorithm that is invariant to various radiometric variations between left and right images. Unlike most stereo algorithms, we explicitly employ the color formation model in our framework and propose a new measure called Adaptive Normalized Cross Correlation (ANCC) for a robust and accurate correspondence measure. ANCC is invariant to lighting geometry, illuminant color and camera parameter changes between left and right images, and does not suffer from fattening effects unlike conventional Normalized Cross Correlation (NCC). Experimental results show that our algorithm outperforms other stereo algorithms under severely different radiometric conditions between stereo images.



(a) left image (b) right image

(c) SAD + GC (d) proposed method

Figure 1. The comparison of conventional SAD+GC method and proposed method for illumination varying stereo images. (a) and (b) are the left and right Aloe image with varying illumination. (c) is the result using SAD+GC method for images (a) and (b), and (d) is the result of proposed method for images (a) and (b).

1. Introduction

1.1. Motivation

In the last several decades, there has been considerable progress in stereo matching algorithms. To date, there are numerous stereo algorithms that perform well for the test bed stereo images provided in [1]. However, most algorithms are based on a common assumption that corresponding pixels have a similar color value called color-consistency. Meltzer *et al*. [19] showed that the globally optimal disparity map obtained by even the powerful tree reweighted message passing (TRW) was not perfect due to the incorrect modeling of the energy functional. This motivated us to study the modeling of more correct data cost for MAP-MRF framework in real situations.

In a real scene, there are many factors that prevent two corresponding pixels from having the same value. One major factor is the radiometric changes including lighting geometry and illuminant color and camera device changes between stereo images [9, 14]. The same scene viewed under a different lighting geometry produces a different color because the intensity at each point is determined by the incident light direction and surface normal direction in a Lambertian model. Fixing the lighting geometry, the object viewed under different illuminant colors also produces a different color. Moreover, a camera device or setting changes such as gamma correction and exposure also induce color changes. These situations are very common in stereo images. For this reason, the raw color recorded by a camera is not a reliable cue for matching, and the color consistency assumption is no longer valid for stereo images in a

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

real scene. Hence the performance of most stereo matching algorithms can be severely degraded under such radiometric variations. On the contrary, the human visual system has a color constancy process which is able to compute colors irrespective of radiometric variations and estimate the reflectance of the object under any illumination condition [16]. However, unlike humans, almost all current stereo matching algorithms do not consider this color constancy process.

1.2. Related works

Recently, Hirschmuller and Scharstein evaluated different cost functions for stereo matching on radiometrically different images caused by light sources, camera exposure, gamma correction and noise, etc [14]. They compared Birchfield and Tomasi data cost (BT) [3], LoG filtered BT, Mean filtered BT, BT after Rank transform [23], Normalized Cross Correlation (NCC), and Hierarchical Mutual Information (HMI) [13] under various conditions with correlation-based method, semi-global and global method. They used only the image intensity information not the color for evaluating the costs. They concluded that all compared costs were not very successful to strong local radiometric changes which were caused by the lighting position changes. Wang *et al*. [22] presented a new invariant measure called light transport constancy (LTC) based on a rank constraint for non-Lambertian surfaces. Their method required at least two stereo image pairs with different illumination conditions to be available for making use of rank constraint. NCC is a very popular and traditional measure [5] for matching contrast varying images. It measures only the cosine angle between matching vectors as normalization make the matching vector to a unit length. However, NCC is only suitable for matching affine-transformed values and also suffers from the fattening effect that object boundaries are not reconstructed correctly such as SAD and SSD. Kim *et al*. [17] suggested the pixelwise data cost based on mutual information using the Talor expansion. Ogale *et al*. [20] also presented a contrast-robust stereo matching algorithm using multiple frequency channels for local matching.

The method of using only intensity information is not appropriate, because intensity depends only on the light direction and surface normal direction which lacks surface and light color and camera parameter informations. Hence, color information is necessary for handling various radiometric factors. However, because most methods do not handle the color formation process explicitly to find correspondence, their performance is dependent on the radiometric variation between input stereo images. For example, Fig. 1 shows that SAD with Graph-cut (GC) method fails under severe radiometric variation while the proposed method is more robust under severe radiometric changes between the left and right images.

In this paper, we present a new stereo matching algorithm that is invariant to lighting geometry, illuminant color, and camera parameter changes. We explicitly modeled the color formation process unlike other algorithms. From this model, we extracted the invariant information and propose the invariant measure called Adaptive Normalized Cross Correlation (ANCC) to various radiometric changes.

2. Stereo Energy Formulation

We define our stereo matching as a minimization problem of the following energy :

$$\begin{aligned} E(f) &= E_{data}(f) + E_{smooth}(f), \\ E_{data}(f) &= \sum_{p} D_p(f_p), \\ E_{smooth}(f) &= \sum_{p} \sum_{q \in N(p)} V_{pq}(f_p, f_q), \end{aligned} \tag{1}$$

where $N(p)$ is the neighborhood pixels of p, and $D_p(f_p)$ is the data cost that measures the dissimilarity between pixel p in the left image and pixel $p + f_p$ in the right image. $V_{pq}(f_p, f_q)$ is the smoothness cost that favors the piece-wise smooth objects. Combining these costs, the optimal disparities can be found by minimizing the total energy in eq. (1). To make data cost radiometric-invariant, we need to model the color formation process explicitly.

3. Color Normalization Representation

There are two approaches for finding illuminant invariant representation [6] : color constancy algorithms and color invariant approaches.

Color constancy algorithms [18, 10, 16, 12, 8] attemp to separate the illumination and the reflectance components on images like the human visual system does. Retinex algorithms calculate the lightness sensations not the physical reflectances in a given image and effectively compensate for non-uniform lighting [18, 16, 12]. The gamut-mapping algorithm [10] and color-by-correlation algorithm [8] can estimate the illuminant in given images. However, because the color constancy problem is ill-posed, the estimation of the illuminant is generally not an easy task [11].

The color invariant approach [9, 7] finds the function which is independent from lighting conditions and imaging devices. Among these color invariant approaches, chromaticity normalization and gray-world assumption are commonly used methods [15]. Chromaticity normalization is usually used to remove lighting geometry effects, while gray-world assumption is used to remove illuminant color effects. However, neither chromaticity normalization nor gray-world assumption can remove both lighting geometry and illuminant color dependency simultaneously. Only a comprehensive normalization method can remove both of them iteratively [9] and non-iteratively [7].

3.1. Color image formation model

An image taken by a linear imaging device can be described by the following equation [8] :

$$h_k^x = \int_\omega E(\lambda) S^x(\lambda) Q_k(\lambda) d\lambda, \tag{2}$$

where h_k^x is the kth sensor (color channel) response at a point x in the scene, and λ is the wavelength. $E(\lambda)$ is the spectral power distribution of the incident illuminant at each wavelength, $S^x(\lambda)$ is the surface reflectance at a point x in the scene, and $Q_k(\lambda)$ is the relative spectral response of the kth sensor. We assume the Lambertian reflectance model and $Q_k(\lambda) = \delta(\lambda - \lambda_k)$, which is a Dirac delta function at the wavelength λ_k. Then eq. (2) becomes

$$h_k^x = E(\lambda_k) S^x(\lambda_k). \tag{3}$$

During image acquisition, the response of the device is linear. However, for the compression of dynamic range, image data is transformed non-linearly prior to the storage process. This is called gamma correction and it raises the value of each RGB response to a power function of a fixed exponent γ. Considering all the factors, we can represent the color image formation model at pixel p as follows [7]:

$$\begin{pmatrix} R(p) \\ G(p) \\ B(p) \end{pmatrix} \rightarrow \begin{pmatrix} \tilde{R}(p) \\ \tilde{G}(p) \\ \tilde{B}(p) \end{pmatrix} = \begin{pmatrix} \rho(p) a R(p)^\gamma \\ \rho(p) b G(p)^\gamma \\ \rho(p) c B(p)^\gamma \end{pmatrix}, \tag{4}$$

where each pixel p has its own individual brightness factor $\rho(p)$ which is dependent on the angle between the light direction and the surface normal at that point. If we change the light illumination color while fixing the lighting geometry, then the responses in three color channels change by the scale factor a, b and c, respectively.

3.2. Color image normalization

Choromaticity normalization [15] is commonly used to eliminate the effect of lighting geometry that is dependent only on the surface normal vector and the light direction in the Lambertian model. At pixel p, if we divide each of $\tilde{R}(p), \tilde{G}(p)$ and $\tilde{B}(p)$ by the averages of them, then we can obtain the $\rho(\cdot)$ independent color representation as follows.

$$\begin{pmatrix} \frac{\rho(p) a R(p)^\gamma}{\rho(p)(aR(p)^\gamma + bG(p)^\gamma + cB(p)^\gamma)} \\ \frac{\rho(p) b G(p)^\gamma}{\rho(p)(aR(p)^\gamma + bG(p)^\gamma + cB(p)^\gamma)} \\ \frac{\rho(p) c B(p)^\gamma}{\rho(p)(aR(p)^\gamma + bG(p)^\gamma + cB(p)^\gamma)} \end{pmatrix} = \begin{pmatrix} \frac{aR(p)^\gamma}{(aR(p)^\gamma + bG(p)^\gamma + cB(p)^\gamma)} \\ \frac{bG(p)^\gamma}{(aR(p)^\gamma + bG(p)^\gamma + cB(p)^\gamma)} \\ \frac{cB(p)^\gamma}{(aR(p)^\gamma + bG(p)^\gamma + cB(p)^\gamma)} \end{pmatrix}. \tag{5}$$

On the other hand, an illuminant color independent representation can be acquired by the gray-world assumption [15]. By dividing each of $\tilde{R}(p), \tilde{G}(p)$ and $\tilde{B}(p)$ by its channel mean value, we can obtain a, b and c independent expressions as follows.

$$\begin{pmatrix} \rho(p) a R(p)^\gamma \\ \rho(p) b G(p)^\gamma \\ \rho(p) c B(p)^\gamma \end{pmatrix} \rightarrow \begin{pmatrix} \frac{\rho(p) a R(p)^\gamma}{\tilde{R}_{mean}} \\ \frac{\rho(p) b G(p)^\gamma}{\tilde{G}_{mean}} \\ \frac{\rho(p) c B(p)^\gamma}{\tilde{B}_{mean}} \end{pmatrix} = \begin{pmatrix} \frac{N \rho(p) R(p)^\gamma}{\sum_{p \in I} \rho(p) R(p)^\gamma} \\ \frac{N \rho(p) G(p)^\gamma}{\sum_{p \in I} \rho(p) G(p)^\gamma} \\ \frac{N \rho(p) B(p)^\gamma}{\sum_{p \in I} \rho(p) B(p)^\gamma} \end{pmatrix},$$

$$\tilde{R}_{mean} = \frac{\sum_{p \in I} \tilde{R}(p)}{N}, \tilde{G}_{mean} = \frac{\sum_{p \in I} \tilde{G}(p)}{N}, \tilde{B}_{mean} = \frac{\sum_{p \in I} \tilde{B}(p)}{N}, \tag{6}$$

where N is the number of pixels in the image I. Note however that neither chromaticity normalization nor gray-world normalization can remove both of the dependency of lighting geometry and lighting color at the same time. Finlayson *et al*. proposed a comprehensive normalization method which combines the two normalization methods in one framework iteratively [9] and non-iteratively [7]. However, since their method needs the global image information such as channel mean value, naive application of it to stereo matching problem is not appropriate. For example, in the gray-world normalization process in eq. (6), the left and right $\tilde{R}_{mean}$ values are generally not the same, due to the view changes in stereo images. That means the true corresponding pixel values are still not the same between stereo images after comprehensive normalization. Hence, applying stereo matching algorithm to the pre-normalized images by this method does not produce satisfactory results as shown in our experiments.

4. Stereo Matching using ANCC

Let us assume that $I_L(p)$ and $I_R(p+f_p)$ are corresponding pixel values, where $I \in \{R, G, B\}$ and $I_L(p)$ is the value in the left image at pixel p and $I_R(p+f_p)$ is the value in the right image at pixel $p + f_p$ in each color channel, respectively. NCC [5] is a popular similarity measure between two pixels with neighbors that is defined by

$$NCC(f_p) = \frac{\sum_{\substack{t_L \in W_L(p), \\ t_R \in W_R(p+f_p)}} [I_L(t_L) - \bar{I}_L(p)] \times [I_R(t_R) - \bar{I}_R(p+f_p)]}{\sqrt{\sum_{t_L \in W_L(p)} |I_L(t_L) - \bar{I}_L(p)|^2} \times \sqrt{\sum_{t_R \in W_R(p+f_p)} |I_R(t_R) - \bar{I}_R(p+f_p)|^2}}, \tag{7}$$

where $\bar{I}_L(p)$ and $\bar{I}_R(p+f_p)$ are the mean values of pixels in the window $W_L(p)$ centered at p, and those in the window $W_R(p+f_p)$ centered at $p+f_p$, respectively. This NCC has invariant property to the following affine transformation of image color values [5]:

$$I_L \rightarrow \kappa_L I_L + \nu_L, \; I_R \rightarrow \kappa_R I_R + \nu_R. \tag{8}$$

Note that there are two critical problems in simply applying this NCC directly in stereo matching of general image pairs: According to eq. (4), two corresponding pixels can have different γ values, causing NCC to stop working. Thus, naively applying NCC to raw stereo images does not work well because it does not consider various radiometric changes by ρ, a, b, c and γ. Another problem is that the appearances of the supporting windows in the left and right image are not exact due to the view changes. Hence, NCC usually produces fattening effects near the object boundary similar to conventional window correlation-based matching measures such as SSD or SAD.

In this work, to apply NCC for radiometric change invariant matching, we propose a new normalization scheme that transforms the model in eq. (4) into an affine model by using the logarithm of both images and the chromaticity normalization concept [7]. Also to remove fattening effects, we employ the adaptive weight scheme using bilateral filter [21] to select the corresponding pixels between left and right windows.

4.1. Chromaticity normalization

If the reflectance model is Lambertian, and the camera response function is the Dirac delta function, then the image is formed by eq. (4). When the left and right images have different lighting geometries, illuminant colors, and camera gamma functions, they can be represented as follows, respectively :

$$\begin{pmatrix} R_L(p) \\ G_L(p) \\ B_L(p) \end{pmatrix} \rightarrow \begin{pmatrix} \rho_L(p) a_L R_L^{\gamma_L}(p) \\ \rho_L(p) b_L G_L^{\gamma_L}(p) \\ \rho_L(p) c_L B_L^{\gamma_L}(p) \end{pmatrix}, \\ \begin{pmatrix} R_R(p+f_p) \\ G_R(p+f_p) \\ B_R(p+f_p) \end{pmatrix} \rightarrow \begin{pmatrix} \rho_R(p+f_p) a_R R_R^{\gamma_R}(p+f_p) \\ \rho_R(p+f_p) b_R G_R^{\gamma_R}(p+f_p) \\ \rho_R(p+f_p) c_R B_R^{\gamma_R}(p+f_p) \end{pmatrix}. \tag{9}$$

Without loss of generality, we only consider R channel value. According to eq. (9), corresponding left and right color values have a non-linear relationship due to different gamma values, γ_L and γ_R, respectively. To transform this non-linear relationship to a linear one, we take the logarithms of both images. Then, each color value can be represented by

$$\begin{aligned} &R'_L(p) = \log \rho_L(p) + \log a_L + \gamma_L \log R_L(p), \\ &R'_R(p+f_p) \\ &= \log \rho_R(p+f_p) + \log a_R + \gamma_R \log R_R(p+f_p), \end{aligned} \tag{10}$$

where $\rho(\cdot)$ term depends on each pixel position relating to the lighting direction and surface normal. We can eliminate this $\log \rho(\cdot)$ term by simply subtracting the average of the transformed color values in R, G, B channels (chromaticity normalization) that is defined by

$$\begin{aligned} \bar{I}'_L(p) &= \frac{R'_L(p)+G'_L(p)+B'_L(p)}{3} \\ &= \log \rho_L(p) + \frac{\log a_L b_L c_L}{3} + \frac{\gamma_L (\log R_L(p) G_L(p) B_L(p))}{3}. \end{aligned} \tag{11}$$

By subtracting the mean value $\bar{I}'_L(p)$ at each pixel, each color value becomes

$$\begin{aligned} &R''_L(p) = R'_L(p) - \bar{I}'_L(p) \\ &= \log \frac{a_L}{\sqrt[3]{a_L b_L c_L}} + \gamma_L \log \frac{R_L(p)}{\sqrt[3]{R_L(p) G_L(p) B_L(p)}} \\ &\triangleq \alpha_L + \gamma_L K_L(p). \end{aligned} \tag{12}$$

Similarly, the corresponding pixel value $R''_R(p+f_p)$ in the right image is represented as

$$R''_R(p+f_p) \triangleq \alpha_R + \gamma_R K_R(p+f_p). \tag{13}$$

The transformed values in other channels G and B can be similarly computed. Note that $K_L(p)$ and $K_R(p+f_p)$ are not dependent on ρ, a, b, c and γ. If the corresponding pixels are correct, $K_L(p)$ and $K_R(p+f_p)$ must be the same. And moreover, since $R''_L(p)$ and $R''_R(p+f_p)$ are the affine transformed values of $K_L(p)$ and $K_R(p+f_p)$, respectively, the matching of them is also invariant to radiometric changes by NCC.

4.2. Adaptive normalized cross correlation

To reduce the fattening effect caused by outliers, we use the weight distribution information. Each pixel t in an $m \times m$ window $W(p)$ around the center pixel p has different weights. The weight $w(t)$ is computed using bilateral filter [21] as follows :

$$w(t) = \exp\left(-\frac{\|p-t\|^2}{2\sigma_d^2} - \frac{\|I(t)-I(p)\|^2}{2\sigma_s^2}\right), \tag{14}$$

where $\|\cdot\|$ means the Euclidean distance. The first and second terms in the exponent represent the geometric distance and the color dissimilarity between the center pixel p and the pixel t in the window, respectively. This weight distribution has an edge-preserving property unlike the isotropic Gaussian weight. Bilateral filtered weighted sum $S(p)$ for the center pixel p is defined as

$$S(p) = \frac{\sum_{t \in W(p)} w(t) I(t)}{Z(p)}, \tag{15}$$

where $Z(p)$ is the normalizing constant. Note that the weights are computed for the CIELab color images instead of the raw RGB color images. Because CIELab color is perceptually similar to human visual system, the weight distributions still have reliable property after different radiometric changes due to the rank invariance [23, 6]. To reduce

the fattening effects, instead of subtracting the simple mean value as NCC, we subtract the bilateral filtered weighted sum value for each channel, thereby removing the α.

$$\begin{aligned} R'''_L(t) &= R''_L(t) - S(p) \\ &= \alpha_L + \gamma_L K_L(t) - \frac{\sum_{t\in W(p)} w(t) R''_L(t)}{Z(p)} \\ &= \alpha_L + \gamma_L K_L(t) - \frac{\sum_{t\in W(p)} w(t)(\alpha_L + \gamma_L K_L(t))}{Z(p)} \\ &= \gamma_L (K_L(t) - \frac{\sum_{t\in W(p)} w(t) K_L(t)}{Z(p)}). \end{aligned} \tag{16}$$

Let us denote the patch around pixel p in the left image as 1D vector $v_L(p)$ and the corresponding weight vector of each pixel in the window $W(p)$ is $\omega_L(p)$ that are represented as

$$\begin{aligned} v_L(p) &= (R'''_L(t_1), R'''_L(t_2), ..., R'''_L(t_M)), \\ \omega_L(p) &= (w_L(t_1), w_L(t_2), ..., w_L(t_M)), \end{aligned} \tag{17}$$

where $M = m \times m$. Similarly, the corresponding patch around $p + f_p$ in the right image is denoted as

$$\begin{aligned} v_R(p + f_p) &= (R'''_R(t_1), R'''_R(t_2), ..., R'''_R(t_M)), \\ \omega_R(p + f_p) &= (w_R(t_1), w_R(t_2), ..., w_R(t_M)). \end{aligned} \tag{18}$$

Then the similarity between $v_L(p)$ and $v_R(p+f_p)$ is defined as

$$ANCC_R(f_p) = \frac{\sum_{i=1}^{M} w_L(t_i) w_R(t_i) [R'''_L(t_i)] \times [R'''_R(t_i)]}{\sqrt{\sum_{i=1}^{M} \left| w_L(t_i) R'''_L(t_i) \right|^2} \times \sqrt{\sum_{i=1}^{M} \left| w_R(t_i) R'''_R(t_i) \right|^2}}, \tag{19}$$

We define eq. (19) as Adaptive Normalized Cross Correlation (ANCC) for the R channel. $ANCC_G(f_p)$ and $ANCC_B(f_p)$ for the Green and Blue channel can be similarly computed. Note that ANCC does not vary with ρ, illuminant color (a, b, c) and camera gamma correction γ. Moreover, the fattening effect can also be reduced since the spatial weight information is incorporated adaptively.

4.3. Global energy modelling

ANCC is a similarity measure that ranges from -1 to $+1$. To make a non-negative cost between pixel p and $p+f_p$ in the left and right image, respectively, we subtract ANCC from +1. Now, we define our data cost $D_p(f_p)$ as follows :

$$D_p(f_p) = (1 - \frac{ANCC_R(f_p) + ANCC_G(f_p) + ANCC_B(f_p)}{3}) \cdot \tag{20}$$

For the pairwise cost, we used a simple truncated quadratic cost as

$$V_{pq}(f_p, f_q) = \lambda \cdot \min(\left| f_p - f_q \right|^2, V_{\max}). \tag{21}$$

The total energy is defined by

$$E(f) = \sum_p D_p(f_p) + \sum_p \sum_{q\in N(p)} V_{pq}(f_p, f_q). \tag{22}$$

We optimized this total energy using α-expansion algorithm [4] to find the disparity map.

5. Experimental Results

In our experiments, we fixed all the parameters of the proposed algorithm such that $\lambda = 1/50, V_{max} = 5$, and $M = (31 \times 31), \sigma_d = 14, \sigma_s = 3.8$. For implementing Graph-cuts (GC), we used the source code provided in [2]. To evaluate and compare our algorithm with others, we used various images such as the test bed images (Aloe, Art, Moebius, and Dolls) in [1, 14] and aerial images. Each data set in [1, 14] has three different exposures (indexed as 0,1,2) and three different configurations of light sources (indexed as 1,2,3), a total of nine different images. Using these data sets, we compared various stereo algorithms such as BT [3], BT using 5x5 Laplacian of Gaussian (LoG) filtered image (LOG+BT), BT (CN+BT) and Sum of Absolulte Difference (CN+SAD) using images preprocessed by the comprehensive normalization method (CN) [7], NCC (7x7 window), and ANCC. Fig. 2 shows that the proposed method produced sharp and accurate results for the illumination2 and exposure2 images.

5.1. Light source changes

First, we fixed the exposure to 1 for all images and varied only the illumination from 1 to 3. Fig. 3 (f)-(k) and Fig. 4 (f)-(k) show the comparison of the results for the test bed images [1]. Fig. 5 (a)-(d) show the error ratio of unoccluded areas for all combination of illuminations. We avoided the case of same illumination combination for left and right images. For example, among the 1/3 (left/right illumination) case and the 3/1 case, we only experimented on the 1/3 case because 1/3 and 3/1 cases have symmetric properties as shown in [14]. In our experiments, ANCC showed very stable and low error rate, while other methods were very sensitive to the lighting source changes. As described above, the stereo matching using images with comprehensive normalization (CN+BT, CN+SAD) did not produce good results.

5.2. Camera exposure changes

Similarly to the case of lighting changes, we fixed the illumination to 1, and only changed the exposure from 0 to 2 using the same data set as in section 5.1. Fig. 3 (l)-(q) and Fig. 4 (l)-(q) show the comparison results for the test bed images. Fig. 5 (e)-(h) show the error ratio of unoccluded areas for all combination of exposures. ANCC was also

stable for camera exposure changes. However, it was unstable for the saturated color region which had (255,255,255) or (0,0,0) RGB color values because the weight could not be calculated correctly for those regions. CN+BT and CN+SAD were more sensitive to camera exposure changes than light changes.

5.3. Application to aerial image

Fig. 6 shows the stereo comparison for aerial images which have been taken using one camera at different times. Generally the colors of aerial images change due to the time difference in taking pictures. Thus, the colors of aerial images can be affected by the complex factors including the lighting color and position changes. Hence, aerial images are more challenging data for stereo matching. For the test aerial images shown in Fig. 6 (a)-(b), our algorithm produced sharper and more accurate results than others as depicted in Fig. 6 (c)-(f).

6. Conclusions

In this paper, we proposed a new stereo matching algorithm that is invariant to various radiometric conditions such as lighting geometry, illuminant color and camera parameter changes between left and right images. We explicitly considered the color formation process and proposed a radiometric invariant measure called Adaptive Normalized Cross Correlation (ANCC). Our method can reconstruct disparity maps more accurately under severely different radiometric changes between stereo images and does not suffer from fattening effects as NCC.

Acknowledgements

This research was supported in part by the Defense Acquisition Program Administration and Agency for Defense Development, Korea, through the Image Information Research Center & Technology under the contract UD070007AD, and in part by the ITRC program by Ministry of Information and Communication.

References

[1] http://vision.middlebury.edu/stereo/. 1, 5

[2] http://www.adastral.ucl.ac.uk/ vladkolm/software.html. 5

[3] S. Birchfield and C. Tomasi. A pixel dissimilarity measure that is insensitive to image sampling. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20(4):401–406, 1998. 2, 5

[4] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001. 5

[5] O. Faugeras, B. Hotz, H. Mathieu, T. Viéville, Z. Zhang, P. Fua, E. Théron, L. Moll, G. Berry, J. Vuillemin, P. Bertin, and C. Proy. Real time correlation based stereo: algorithm implementations and applications. *INRIA Technical Report*, 1993. 2, 3

[6] G. Finlayson, S. Hordley, G. Schaefer, and G. Y. Tian. Illuminant and device invariant colour using histogram equalisation. *Pattern Recognition*, 38(2):179–190, 2005. 2, 4

[7] G. Finlayson and R. Xu. Illuminant and gamma comprehensive normalisation in log rgb space. *Pattern Recognition Letters*, 24(11):1679–1690, 2003. 2, 3, 4, 5

[8] G. D. Finlayson, S. D. Hordley, and P. M. Hubel. Color by correlation : A simple, unifying framework for color constancy. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(11):1209–1221, 2001. 2, 3

[9] G. D. Finlayson, B. Schiele, and J. L. Crowley. Comprehensive colour image normalization. In *Proc. of European Conference on Computer Vision*, 1998. 1, 2, 3

[10] D. A. Forsyth. A novel algorithm for color constancy. *International Journal of Computer Vision*, 5(1):5–24, 1990. 2

[11] B. Funt, K. Barnard, and L. Martin. Is machine colour constancy good enough? In *Proc. of European Conference on Computer Vision*, 1998. 2

[12] B. Funt, F. Ciurea, and J. McCann. Retinex in matlab. In *Proc. of Eighth Color Imaging Conference: Color Science, Systems and Applications*, 2000. 2

[13] H. Hirschmuller. Accurate and efficient stereo processing by semi-global matching and mutual information. In *Proc. of Computer Vision and Pattern Recognition*, 2005. 2

[14] H. Hirschmuller and D. Scharstein. Evaluation of cost functions for stereo matching. In *Proc. of Computer Vision and Pattern Recognition*, 2007. 1, 2, 5

[15] R. Hunt. The reproduction of color. *fifth ed. Fountain Press*. 2, 3

[16] D. J. Jobson, Z. Rahman, and G. A. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Trans. Image Processing*, 6(7):965–976, 1997. 2

[17] J. Kim, V. Kolmogorov, and R. Zabih. Visual correspondence using energy minimization and mutual information. In *Proc. of International Conference on Computer Vision*, 2003. 2

[18] E. H. Land. An alternative technique for the computation of the designator in the retinex theory of color vision. *Nat. Acad. Sci.*, 83(1):3078–3080, 1986. 2

[19] T. Meltzer, C. Yanover, and Y. Weiss. Globally optimal solutions for energy minimization in stereo vision using reweighted belief propagation. In *Proc. of International Conference on Computer Vision*, 2005. 1

[20] A. S. Ogale and Y. Aloimonos. Robust contrast invariant stereo correspondence. In *Proc. of IEEE Conf. on Robotics and Automation*, 2004. 2

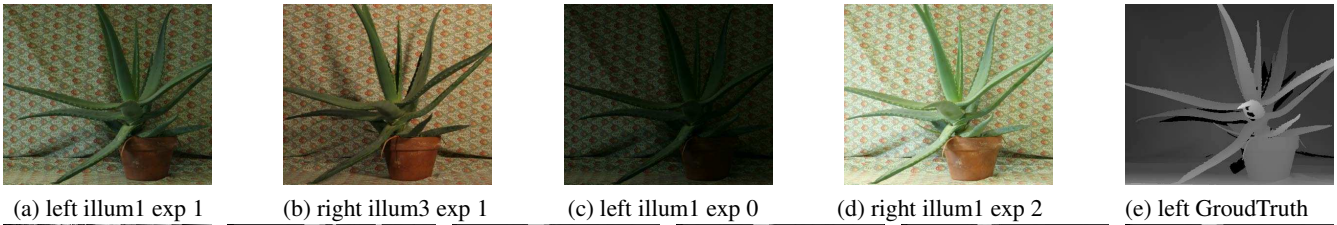
[21] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proc. of International Conference on Computer Vision*, 1998. 4

[22] L. Wang, R. Yang, and J. E. Davis. Brdf invariant stereo using light transport constancy. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29(9):1616–1626, 2007. 2

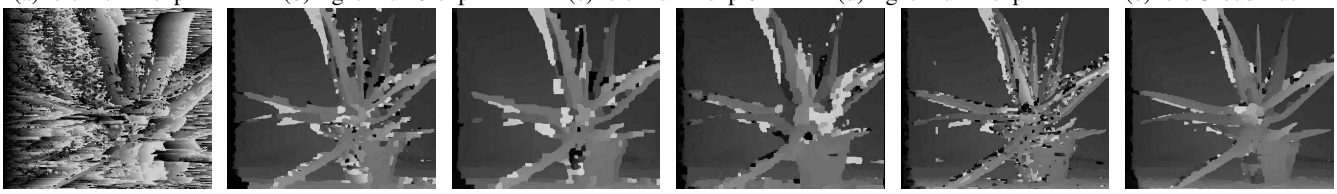
[23] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. In *Proc. of European Conference on Computer Vision*, 1994. 2, 4

(a) Aloe (b) Art (c) Moebius (d) Dolls

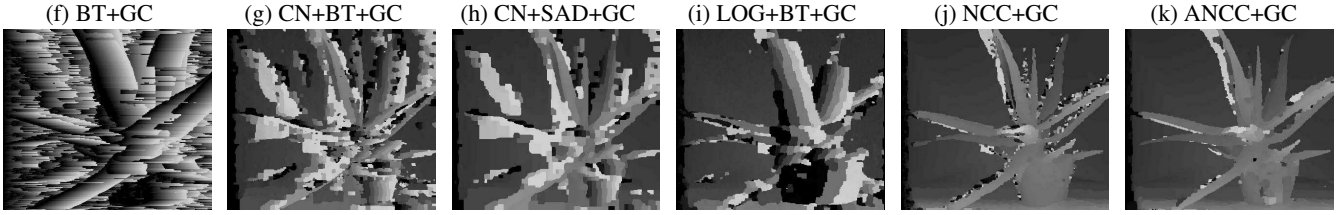
Figure 2. Results of the proposed method for illum 2 exp 2 left and right images .



(a) left illum1 exp 1 (b) right illum3 exp 1 (c) left illum1 exp 0 (d) right illum1 exp 2 (e) left GroudTruth



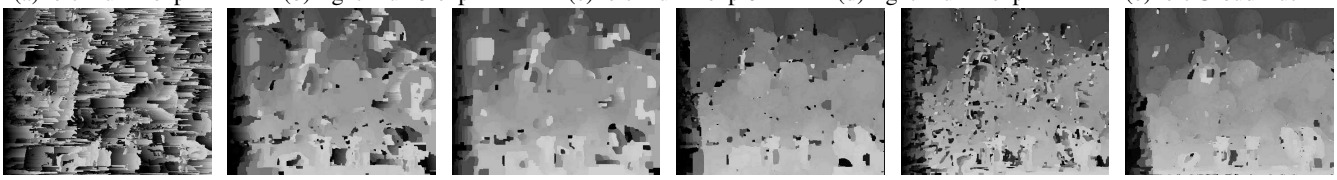
(f) BT+GC (g) CN+BT+GC (h) CN+SAD+GC (i) LOG+BT+GC (j) NCC+GC (k) ANCC+GC



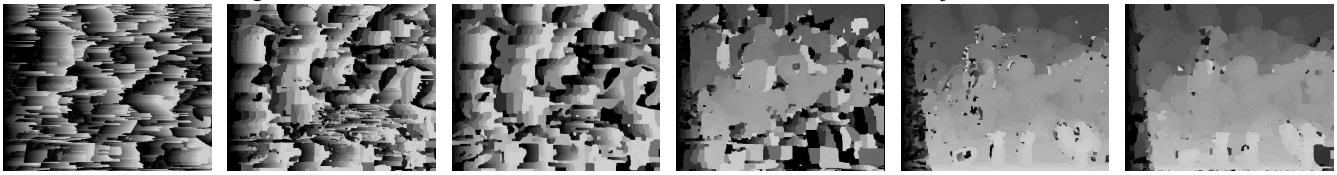
(l) BT+GC (m) CN+BT+GC (n) CN+SAD+GC (o) LOG+BT+GC (p) NCC+GC (q) ANCC+GC

Figure 3. Results of test stereo algorithms on Aloe image pair with varying illumination and camera exposure. (f) - (k) are the results using the image pair (a) and (b). (l) - (q) are the results using the image pair (c) and (d). (e) is the ground truth disparity map for the left image.

(a) left illum1 exp 1 (b) right illum3 exp 1 (c) left illum1 exp 0 (d) right illum1 exp 2 (e) left GroudTruth

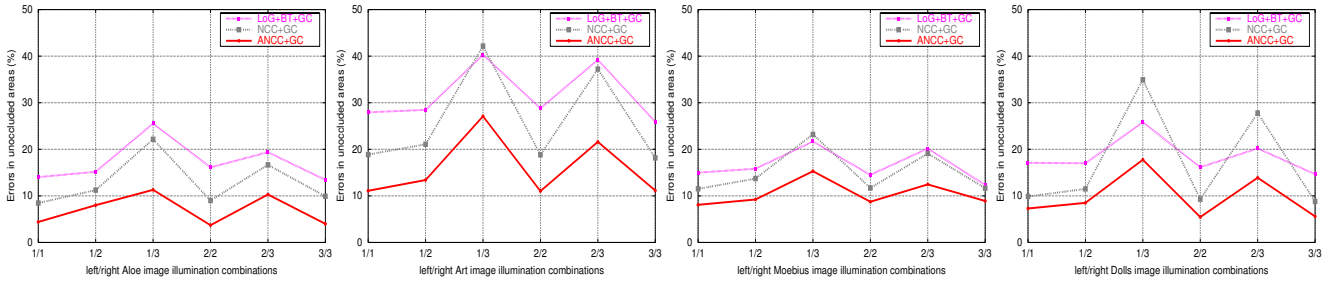


(f) BT+GC (g) CN+BT+GC (h) CN+SAD+GC (i) LOG+BT+GC (j) NCC+GC (k) ANCC+GC

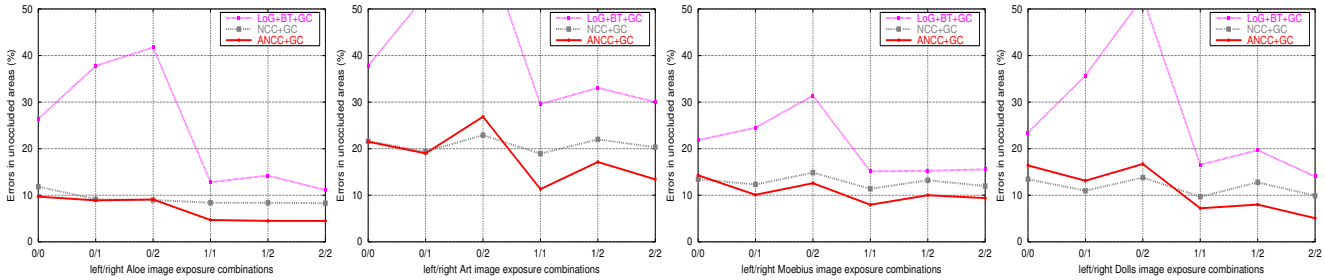


(l) BT+GC (m) CN+BT+GC (n) CN+SAD+GC (o) LOG+BT+GC (p) NCC+GC (q) ANCC+GC

Figure 4. Results of test stereo algorithms on Dolls image pair with varying illumination and camera exposure. (f) - (k) are the results using the image pair (a) and (b). (l) - (q) are the results using the image pair (c) and (d). (e) is the ground truth disparity map for the left image.

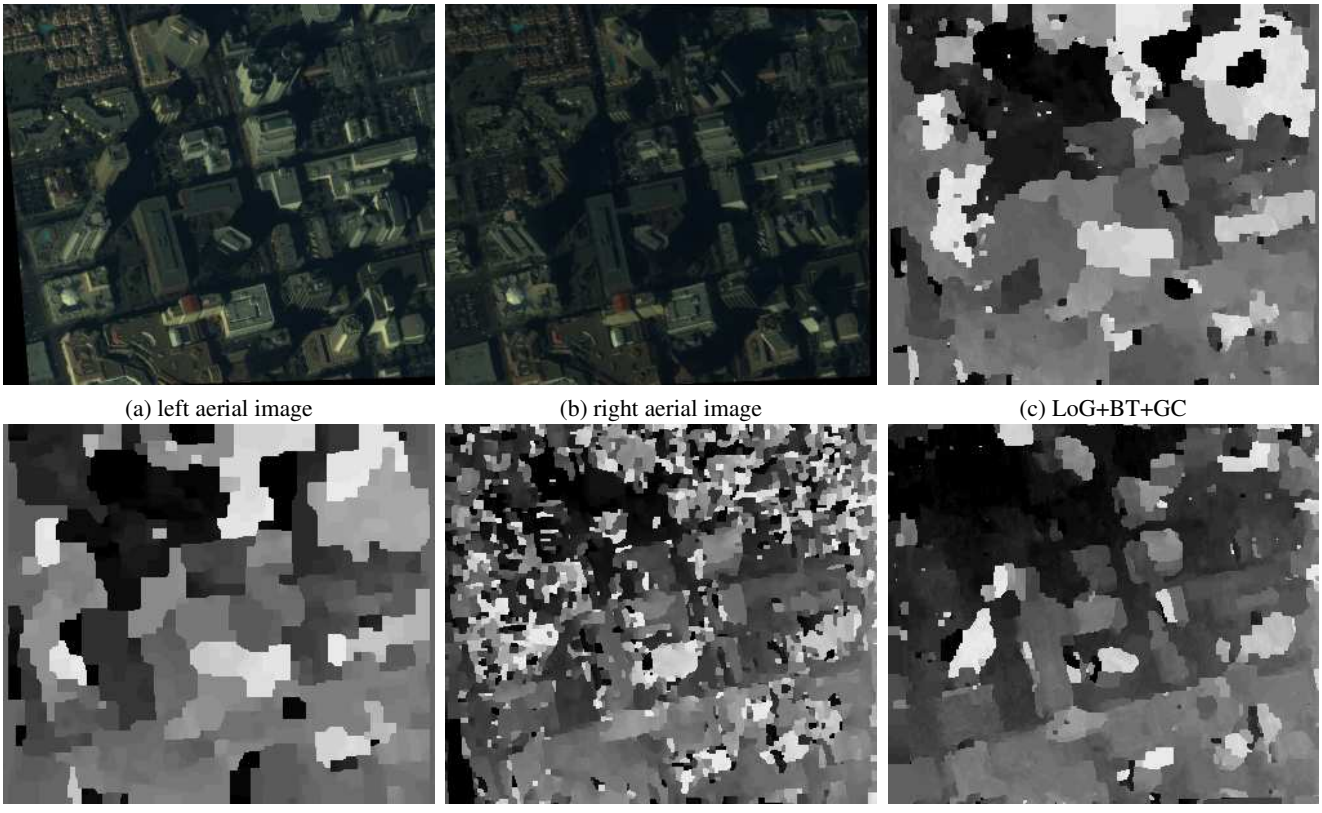


(a) Aloe (b) Art (c) Moebius (d) Dolls



(e) Aloe (f) Art (g) Moebius (h) Dolls

Figure 5. Quantitative comparison for illumination and exposure changes. (a)-(d) are the comparison results for the illumination changes. (e)-(h) are the comparison results for the exposure changes.



(a) left aerial image (b) right aerial image (c) LoG+BT+GC

(d) LoG+SAD+GC (e) NCC+GC (f) ANCC+GC

Figure 6. Results of stereo algorithms on aerial image pair. (c)-(f) are the stereo results for the input image pair (a) and (b)