Open access • Journal Article • DOI:10.1109/TPAMI.2010.136

Robust Stereo Matching Using Adaptive Normalized Cross-Correlation

Yong Seok Heo, Kyong Mu Lee, Sang Uk Lee

Institutions: Seoul National University

Published on: 01 Apr 2011 - IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE Computer Society)

Topics: Computer stereo vision, Standard illuminant and Gamma correction

Robust Stereo Matching Using Adaptive Normalized Cross-Correlation

Yong Seok Heo, *Student Member*, *IEEE*, Kyoung Mu Lee, *Member*, *IEEE*, and Sang Uk Lee, *Fellow*, *IEEE*

Abstract—A majority of the existing stereo matching algorithms assume that the corresponding color values are similar to each other. However, it is not so in practice as image color values are often affected by various radiometric factors such as illumination direction, illuminant color, and imaging device changes. For this reason, the raw color recorded by a camera should not be relied on completely, and the assumption of color consistency does not hold good between stereo images in real scenes. Therefore, 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 measure that is insensitive to radiometric variations between left and right images. Unlike most stereo matching measures, we use the color formation model explicitly in our framework and propose a new measure, called the *Adaptive Normalized Cross-Correlation (ANCC)*, for a robust and accurate correspondence measure. The advantage of our method is that it is robust to lighting geometry, illuminant color, and camera parameter changes between left and right images, and does not suffer from the fattening effect unlike conventional Normalized Cross-Correlation (NCC). Experimental results show that our method outperforms other state-of-the-art stereo methods under severely different radiometric conditions between stereo images.

Index Terms—Stereo matching, color, radiometric variation, illumination, camera exposure, gamma correction.

1 Introduction

1.1 Motivation

Stereo matching aims to obtain 3D information by finding the correct correspondence between images captured from different point of views or at different times. However, finding the accurate correspondence is not an easy task; there exist a number of difficulties, such as occluded regions, textureless regions, and object boundaries. This issue has been an important area of research in the past several decades, and considerable progress has been made with respect to the problem surrounding stereo matching algorithms. Efforts toward this end have resulted in numerous stereo algorithms that perform relatively well for the images in the Middlebury database [1], [2].

These algorithms are based on a common assumption that corresponding pixels have similar color values, an assumption we refer to as *color consistency*. As a consequence, a majority of these utilize a data cost which includes simple L_1 or L_2 difference of intensities or color values of corresponding pixels. However, it should be noted that these methods do not hold good for stereo images which do not have similar corresponding color values. Nonetheless, a few studies have been performed in order to solve this problem.

• *The authors are with the Department of Electrical Engineering and Computer Science, Automation and Systems Research Institute, Seoul National University, 599 Gwanak-ro, Gwanak-gu, Seoul 151-744, Korea. E-mail: hys@diehard.snu.ac.kr, kyoungmu@snu.ac.kr, sanguk@ipl.snu.ac.kr.*

Manuscript received 31 Dec. 2008; revised 11 Nov. 2009; accepted 27 Apr. 2010; published online 20 July 2010.
Recommended for acceptance by W. Förstner.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2008-12-0891.
Digital Object Identifier no. 10.1109/TPAMI.2010.136.

In a real scenario, various factors prevent two corresponding pixels from having the same color value. One major preventing factor is a *radiometric change*, which includes *lighting geometry*, *illuminant color*, and *camera device changes* between stereo images. Different color values are obtained when the same scene is viewed under a different lighting geometry; the reason for this is that the intensity at each point is determined by the angle between the direction of the incident light and the direction of the surface normal in a Lambertian model. After fixation of the lighting geometry, the object when viewed under different illuminant colors also produces different colors because there is a change in the spectral distribution of the reflected light from the object. Furthermore, color changes can also be induced by using a camera device or setting changes such as exposure variations because there are options to vary the total amount of photon that is incident to the camera. Common situations such as these can be a reason for practical problems in stereo images such as aerial images. Having said that, one should not completely trust the raw color recorded by a camera for purposes of matching, and it remains that the color consistency assumption is no longer valid for stereo images in real scenes. Under radiometric variations such as these, the performance of most stereo matching algorithms can be severely degraded. To prove this point with an example, we show in Fig. 1 the case in which a conventional method such as the Sum of Absolute Difference (SAD) with Graph-cuts (GC) method fails, while the proposed method is more robust under severe radiometric changes between the left and right images.

However, there exists a natural color-constancy process in the human visual system that can compute colors regardless of radiometric variations and also estimate the reflectance of the object under any illumination condition

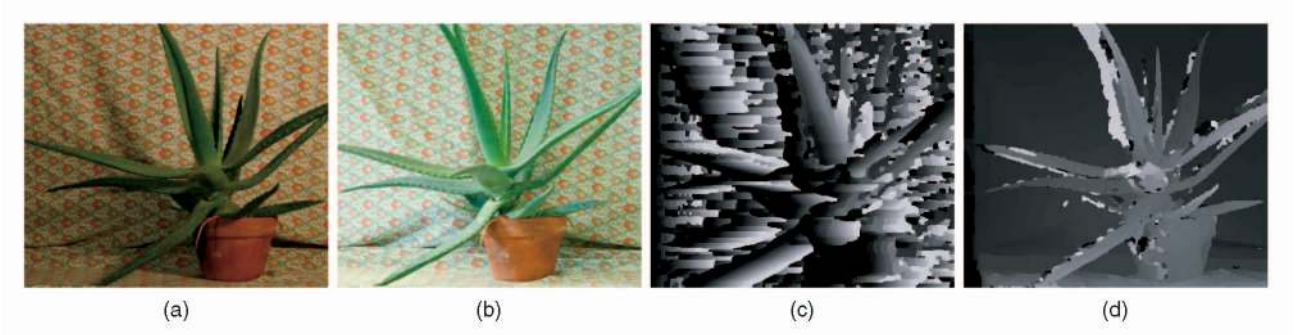

Fig. 1. Comparison of the proposed method with a conventional stereo method (SAD+GC) for the illumination-varying stereo images. (a)-(b) The left and right Aloe images with varying illumination. (c) Disparity map of SAD+GC for the input stereo image pair in (a) and (b). (d) Disparity map of the proposed method for the input stereo image pair in (a) and (b).

[3]. In practice, however, almost all existing stereo matching algorithms do not take into consideration this color-constancy process. Therefore, there is a pressing need to include this process into the matching mechanism.

In this paper, we present a new stereo matching measure that is robust in handling various radiometric changes, including local radiometric variations caused by varying lighting geometry, as well as global radiometric variations that are brought about by changes in illuminant color and camera parameter [4]. It should be noted that this paper is based on the assumption of the Lambertian world. Under the assumption, the color formation process was explicitly modeled, unlike other methods. Then, we have extracted the invariant color information from the color model and thereby propose, in this paper, a new matching measure, called *Adaptive Normalized Cross-Correlation (ANCC)*, which is robust to various radiometric changes.

1.2 Related Works

Hirschmüller and Scharstein carried out an evaluation of various cost functions for stereo matching on radiometrically different images caused by factors such as light configuration, camera exposure, gamma correction variations, and noise, etc [5]. Under various radiometric conditions, they compared Birchfield and Tomasi data cost (BT) [6], BT with Laplacian of Gaussian (LoG) filtering [7], BT with Mean filtering, BT with Rank transform [8], Normalized Cross-Correlation (NCC), and Hierarchical Mutual Information (HMI) [9] with correlation-based method, as well as the semiglobal and global method. Although BT cost is a very popular data cost which is insensitive to camera sampling, the major drawback of it is that it cannot handle radiometrically different images as it employs the use of linearly interpolated function of intensity values. Although the LoG filter calculates second-order derivatives that are insensitive to outliers [7], BT with LoG filtering itself is insufficient for use in stereo matching. BT with Rank transform [8] possesses a robust property to radiometric variations because the Rank transform is based on the principle that the rank between pixel intensities does not vary after radiometric variations. Although it is effective in handling global radiometric changes, it is weak with respect to local radiometric variations. NCC is a highly popular as well as traditional measure for matching contrast-varying images [10], [11]. It measures only the cosine of an angle between matching vectors because normalization renders the matching vectors to have zero mean and one standard deviation. Therefore, NCC is suitable only for matching affine-transformed intensity or color values and it also suffers from the fattening effect that object boundaries are not reconstructed correctly similar to the Sum of Absolute Difference (SAD) and the Sum of Squared Difference (SSD). Mutual Information (MI) has been used as a similarity measure in computer vision [9], [12], [13], [14]. For stereo matching, Kim et al. [14] proposed a pixelwise data cost in the MAP-MRF framework based on mutual information. Their work was improved by Hirschmüller [9] and extended to handle occlusion and speed up the computation by using the Hierarchical Mutual Information (HMI). The advantage of the MI-based cost is that it enables finding the correspondence between globally transformed image pairs by utilizing the joint histogram of intensities between two input images. However, local radiometric variations due to lighting geometry changes cannot be handled by MI-based cost. Because of the just-discussed reasons, Hirschmüller and Scharstein [5] concluded that all compared costs previously described fail to achieve success with respect to strong local radiometric changes caused by the light configuration changes. Besides, Wang et al. [15] 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. Ogale and Aloimonos [16] presented a contrast-robust stereo matching algorithm for local matching using multiple frequency channels. Negahdaripour [17] proposed a general linear brightness constraint for handling radiometric variations between images. The principle behind his model was that the linear transformation of the brightness between image patches was assumed, and the linear transformation parameters and the optical flow vectors were simultaneously estimated. Zhang et al. [18] presented a unified MRF framework for the estimation of illumination variation and disparity map simultaneously. Their framework sought to define Illumination Ratio Map (IRM), and assumed disparity and illumination smoothness.

A major drawback of most of the above-mentioned methods is that these utilize only raw intensity (or color) information, which can be sensitive to severe radiometric variations for stereo matching due to the fact that raw intensity depends only on the direction of the light and the surface normal and does not take into consideration the

color of surface and light and information regarding the camera parameters. Although the intensity information is applied widely in global transformations such as exposure variations, it still remains sensitive to local variations caused by changes in lighting geometry. Taking all of these factors into consideration, in this paper we intend to show that it is more efficient and robust to use invariant color information rather than the simple raw intensity (or color) information for handling various local and global radiometric variations.

2 Stereo Energy Formulation

In this paper, we define stereo matching as a minimization problem of the following energy in the MAP-MRF framework [19]:

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

where $N(p)$ is the neighborhood pixels of p and $D_p(f_p)$ is the data cost, which is a measure of 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 piecewise smooth objects. Taking these costs together, we can arrive at the optimal disparities by minimizing the total energy in (1).

From the findings of the study by Meltzer et al. [20], it can be inferred that the globally optimal disparity map that was even obtained by the powerful tree-reweighted message passing (TRW) [21] was not perfect because of the incorrect modeling of the energy functional. In order to find out correspondence, the most fundamental part is a matching cost (data cost) computation. This acted as a motivation for us to concentrate on the modeling of a more accurate, radiometric-insensitive data cost for the MAP-MRF framework in real scenarios. In order to model radiometric-insensitive data cost, the color formation process has to be taken into consideration in an explicit manner and proper color-insensitive matching information has been extracted.

3 Color-Invariant Information

Color-constancy algorithms [3], [22], [23], [24], [25], [26], [27] attempt to separate the illumination and the reflectance components on images similar to the human visual system. The first attempt made toward this end was the Land's retinex algorithm [22] with the objective of modeling and computing the human color constancy. It is worth mentioning here that retinex algorithms can calculate only the lightness sensations and not the physical reflectance in a given image and can compensate for nonuniform lighting in an effective manner [3], [22], [23], [24]. Retinex algorithms, however, are highly sensitive as well as dependent on color variations of the adjacent objects in the images [25].

The illuminant in given images can be estimated using the gamut-mapping algorithm [26] and the color-by-correlation algorithm [27]. However, the estimation of the illuminant is not an easy task because the color constancy problem is ill-posed [28].

The color-invariant approaches [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39] find the function which is independent of lighting conditions and imaging devices. There exist two types of invariants, namely, the Lambertian-based color invariant and the invariant that is designed to deal with non-Lambertian reflectance effects such as specular highlights.

With regard to Lambertian object, Geusebroek et al. [32] derived invariant color features such as edges based on the Gaussian scale-space framework. Gevers and Smeulders [33] proposed various color-invariant features for robust object recognition. However, these invariants suffer from instabilities under certain conditions. Gevers and Stokman [34] constructed a color-invariant histogram using variable kernel density estimation to resolve these instabilities. Finlayson et al. [35] made a suggestion in their study for the removal of shadow from color images. Ebner [36] suggested an algorithm to shift colors of the input image in the direction of the gray vector according to the gray world hypothesis.

Various studies also exist in the published literature for finding invariants for non-Lambertian objects. Berwick and Lee [37] suggested utilizing a chromaticity space based on log-ratio of sensor responses for illumination pose and color and specular invariance. Tan and Ikeuchi [38] devised a method for the separation of the specular components from diffuse reflection components using the iterative logarithmic differentiation framework. Zickler et al. [39] suggested a specular-invariant by making use of the dichromatic reflectance model [40], [41]. Their method is based on the linear transformation of RGB color space (called SUV color space) with the assumption that the source-illuminant color is known [39].

Among these color-invariant approaches, the most simple and commonly used methods for Lambertian surface are the chromaticity normalization and the gray world assumption [42]. Chromaticity normalization is primarily used for the removal of the lighting geometry effects, whereas the gray world assumption is used for the removal of the illuminant color effects. However, neither of the two methods can remove the dependency of both lighting geometry and illuminant color simultaneously. A comprehensive normalization method strives to remove both of them iteratively [29] and noniteratively [31].

3.1 Color Image Formation Model

The following equation describes an image taken by a linear imaging device [27], [30]:

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

where h_k^x represents the kth sensor (color channel) response at a point x in the scene and λ is the wavelength. $E(\lambda)$ represents the spectral power distribution of the incident illuminant, $S^x(\lambda)$ represents the surface reflectance at a point x in the scene, and $Q_k(\lambda)$ stands for the spectral response of the kth sensor. As this work is based on the assumption of the Lambertian surface, the radiometric variations induced by non-Lambertian reflectance effects

such as specular highlights are not considered. For a camera sensor model, we also assume the diagonal model [43], [44], [45], from which the spectral response is approximated as the Dirac delta function. Therefore, $Q_k(\lambda) = \delta(\lambda - \lambda_k)$, which represents a Dirac delta function at wavelength λ_k. Now, (2) simply becomes

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

During the image acquisition process, the device responds in a linear fashion. However, for the compression of the dynamic range, there is a nonlinear transformation of the image data before the storage process. This process is called gamma correction and it results in the raising of the value of each RGB response to a power function of an exponent γ value depending on the camera [46], [47], [48]. Taking all of these factors into consideration, the color image formation model at pixel p can be represented as follows [31]:

$$\begin{pmatrix} R(p) \\ G(p) \\ B(p) \end{pmatrix} \rightarrow \begin{pmatrix} \tilde{R}(p) \\ \tilde{G}(p) \\ \tilde{B}(p) \end{pmatrix} = \begin{pmatrix} \rho(p)aR(p)^\gamma \\ \rho(p)bG(p)^\gamma \\ \rho(p)cB(p)^\gamma \end{pmatrix}, \tag{4}$$

where each pixel p has its own individual brightness factor $\rho(p)$ which depends on the angle between the direction of the light and the direction of the surface normal at that point. Changing illumination color while fixing the lighting geometry would result in changes in the responses in three color channels by the global scale factor a, b and c, respectively.

3.2 Color Image Normalization

As mentioned previously, chromaticity normalization [42] is commonly employed for the elimination of the effect of lighting geometry that depends only on the direction of surface normal and the direction of light 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:

$$\begin{pmatrix} \frac{3\rho(p)aR(p)^\gamma}{\rho(p)(aR(p)^\gamma + bG(p)^\gamma + cB(p)^\gamma)} \\ \frac{3\rho(p)bG(p)^\gamma}{\rho(p)(aR(p)^\gamma + bG(p)^\gamma + cB(p)^\gamma)} \\ \frac{3\rho(p)cB(p)^\gamma}{\rho(p)(aR(p)^\gamma + bG(p)^\gamma + cB(p)^\gamma)} \end{pmatrix} = \begin{pmatrix} \frac{3aR(p)^\gamma}{(aR(p)^\gamma + bG(p)^\gamma + cB(p)^\gamma)} \\ \frac{3bG(p)^\gamma}{(aR(p)^\gamma + bG(p)^\gamma + cB(p)^\gamma)} \\ \frac{3cB(p)^\gamma}{(aR(p)^\gamma + bG(p)^\gamma + cB(p)^\gamma)} \end{pmatrix}. \tag{5}$$

On the other hand, we can acquire an illuminant color-independent representation by the gray world assumption [42]. By dividing each of $\tilde{R}(p)$, $\tilde{G}(p)$, and $\tilde{B}(p)$ by its channel mean value, we can obtain the a, b, and c independent expressions as follows:

$$\begin{pmatrix} \rho(p)aR(p)^\gamma \\ \rho(p)bG(p)^\gamma \\ \rho(p)cB(p)^\gamma \end{pmatrix} \rightarrow \begin{pmatrix} \frac{\rho(p)aR(p)^\gamma}{\tilde{R}_{mean}} \\ \frac{\rho(p)bG(p)^\gamma}{\tilde{G}_{mean}} \\ \frac{\rho(p)cB(p)^\gamma}{\tilde{B}_{mean}} \end{pmatrix} = \begin{pmatrix} \frac{N\rho(p)R(p)^\gamma}{\sum_{p\in I}\rho(p)R(p)^\gamma} \\ \frac{N\rho(p)G(p)^\gamma}{\sum_{p\in I}\rho(p)G(p)^\gamma} \\ \frac{N\rho(p)B(p)^\gamma}{\sum_{p\in I}\rho(p)B(p)^\gamma} \end{pmatrix}, \tag{6}$$

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

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

where N represents the number of pixels in the image I. As mentioned earlier, it should be noted that neither chromaticity normalization nor gray world normalization can remove both the dependency of lighting geometry and illuminant color simultaneously. Finlayson et al. proposed a comprehensive normalization method that combines the two normalization methods in one framework iteratively [29] and noniteratively [31]. However, because their method requires global image information such as channel mean value, it is not appropriate to apply the method in a naive manner to the stereo matching problem. For instance, in the gray world normalization process described in (6), the left and right $\tilde{R}_{mean}$ values are mostly not the same because of the view changes characteristic of stereo images. The inference that can be drawn from this is that the true corresponding pixel values are still not the same between stereo images even after comprehensive normalization. Thus, applying the stereo matching algorithm to the prenormalized images by this method does not produce satisfactory results as shown in our experiments.

4 Stereo Matching Using ANCC

Let us make an assumption 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 [10] is a well-known similarity measure between two pixels with neighbors that is defined by the equation

$$\begin{aligned} &NCC(f_p) \\ &= \left(\sum_{\substack{t_L\in W_L(p),\\ t_R\in W_R(p+f_p)}} [I_L(t_L) - \bar{I}_L(p)] \times [I_R(t_R) - \bar{I}_R(p + f_p)] \right) \Bigg/ \\ &\quad \left(\sqrt{\sum_{t_L\in W_L(p)} |I_L(t_L) - \bar{I}_L(p)|^2} \right. \\ &\quad \left. \times \sqrt{\sum_{t_R\in W_R(p+f_p)} |I_R(t_R) - \bar{I}_R(p + f_p)|^2} \right), \end{aligned} \tag{7}$$

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 pixel p in the left image and the mean values of pixels in the window $W_R(p + f_p)$ centered at pixel $p + f_p$ in the right image, respectively. It should be mentioned here that this NCC is insensitive to the following affine transformations of image intensity values of two images [10]:

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

Simply applying this NCC directly in stereo matching of general image pairs would result in two critically important problems: According to (4), there could be a complicated nonlinear relationship between two corresponding pixels between stereo images, thereby causing NCC to cease working. Thus, applying NCC to raw stereo images in a naive fashion does not work well because the various radiometric changes caused by ρ, a, b, c, and γ are not taken into consideration. The second serious problem is that the

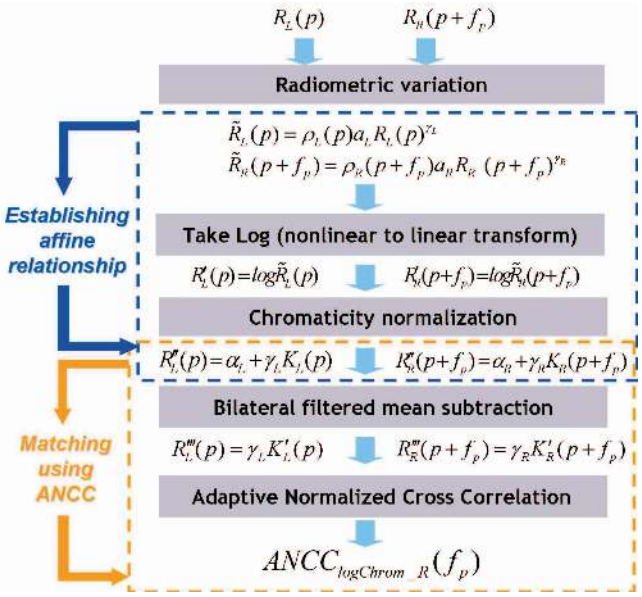

Fig. 2. Overview of the proposed algorithm.

supporting windows in the left and right images do not appear exactly because of the view changes. As a consequence, NCC usually produces a fattening effect near the object boundaries similar to conventional window-correlation-based matching measures such as SSD and SAD.

Therefore, in this work, we seek to find remedies to these two problems. First and foremost, the nonlinear relationship that exists between corresponding pixel color values because of various unknown radiometric changes is transformed into a linear one by employing *log-chromaticity* color space. Second, in order to reduce the fattening effect and increase the accuracy between matching windows, we define a modified NCC measure called the *Adaptive Normalized Cross-Correlation* which utilizes adaptive weighting scheme. This approach is described in detail in the following sections.

4.1 Overview of Our Approach

We provide an overview of our approach in this section. Fig. 2 depicts the overview. The principle behind our novel approach is that the color formation model is explicitly modeled and incorporated into a new stereo correlation measure. To achieve this, we considered the color formation process in an explicit manner instead of using the raw color value for handling the various radiometric changes that occur between stereo images. From this color model, we propose a new data cost that is not only insensitive to the various radiometric changes but is also able to reduce the problems faced with window-based stereo methods such as the fattening effect simultaneously. In order to render the color values changed by various radiometric factors suitable for stereo matching, we carried out a proper transformation of the nonlinear relationship to the affine-transformed relationship between the corresponding pixels that can be solved effectively by the NCC framework. In addition, to ensure accurate and coherent matching, we subtracted bilateral filtered value instead of the simple window mean value for coherent normalization around window pixels. Taking into consideration these bilateral filtered weights again, we defined a new correlation measure called ANCC, using which we defined a new data cost for global energy modeling.

4.2 Log-Chromaticity Normalization

If the reflectance model is Lambertian and the spectral response function is the Dirac delta function, then the image is formed on the basis of (4). If the lighting geometries, illuminant colors, and camera parameters of the left and right images are different, they can be represented by the following equations, respectively:

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

Without loss of generality, only the red channel value is considered. According to (9), the corresponding left and right color values have a nonlinear relationship because of different gamma values, γ_L and γ_R, respectively. In order to transform this nonlinear relationship into a linear one, the logarithms of both images have to be taken. 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 is dependent on each pixel position relative to the direction of the lighting and surface normal. This $\log \rho(\cdot)$ term can be eliminated by simply subtracting the average of the transformed color values in R, G, B channels (chromaticity normalization), which is defined by the following equation:

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

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

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

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

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

We call these transformed color values in (12)-(13) *log-chromaticity* color, denoted *logChrom_R* for the red channel. The computation of the log-chromaticity values in other channels such as *logChrom_G* and *logChrom_B* can be performed similarly. It should be noted that $K_L(p)$ and $K_R(p + f_p)$ are not dependent on ρ, a, b, c, and γ. This means that they are insensitive to radiometric variations. This also implies that if the corresponding pixels are correct, then $K_L(p)$ and $K_R(p + f_p)$ must be the same. Moreover, because $R''_L(p)$ and $R''_R(p + f_p)$ are the affine-transformed values of $K_L(p)$ and $K_R(p + f_p)$, respectively, the NCC between $R''_L(p)$ and $R''_R(p + f_p)$ becomes a suitable measure for matching $K_L(p)$ and $K_R(p + f_p)$.

4.3 Adaptive Normalized Cross-Correlation

As described in the earlier text, the whole window information around matching pixels is used by the NCC in order to find the mean and standard deviation. There exist possibilities that these windows could include outliers between stereo images because of the view difference. This problem thus renders it an inaccurate measure and fails to find the accurate object boundaries as the size of window grows. In order to reduce this fattening effect that is caused by outliers, we use the weight distribution information around matching pixels. According to this, each pixel t in an $m \times m$ window $W(p)$ around the center pixel p has different weights. This weight $w(t)$ is computed by using the bilateral filter [49], [50], [51] as follows:

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

where $\|\cdot\|$ represents the euclidean distance. The first and second terms in the exponent stand for the geometric distance and the color dissimilarity between the center pixel p and the pixel t in the window, respectively. This weight distribution is different from the isotropic Gaussian weight in that the former has an edge-preserving property. The weighted sum $S(p)$ using the bilateral filter for the center pixel p is defined by the following equation:

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

where $Z(p)$ is the normalizing constant. Instead of computing the weights for the raw RGB color, weights are computed for the CIELab color because, in practice, it gives stable results for the majority of cases and the weight distributions still have stable property under different radiometric changes. In order to reduce the effect of the outliers, we subtract the bilateral filtered weighted sum value for each channel instead of subtracting the global mean value as NCC, thereby resulting in the removal of α:

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

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 $W(p)$ as $\omega_L(p)$. These are represented as

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

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

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

Now, the similarity between $v_L(p)$ and $v_R(p + f_p)$ is defined by the following equation:

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

We define (19) as Adaptive Normalized Cross-Correlation (ANCC) for the *logChrom_R* channel. ANCC for the *logChrom_G* and *logChrom_B* channels can be computed in a similar manner. It should be noted that ANCC does not vary with ρ, illuminant color (a, b, c), and camera gamma correction γ. In addition, the fattening effect can also be reduced because we incorporate the spatial weight information adaptively.

4.4 Combining Log-Chromaticity and Original RGB Colors

Although the log-chromaticity color is robust to various radiometric variations, the discriminability can be lowered as compared to the original color because the intensity information of each channel is normalized in the log-chromaticity color and the original RGB color contains the

original intensity values. We were motivated by this observation and sought to combine two color spaces such as log-chromaticity and original RGB colors in order to take advantage of their different aspects. The accuracy and robustness of the disparity maps can be enhanced by the combination of original RGB color and log-chromaticity color because both color spaces are orthogonal.

The ANCC for original RGB color, for example, R channel, is defined in a similar manner by the following equation:

$$ANCC_R(f_p) = \frac{\sum_{i=1}^{M} w_L(t_i) w_R(t_i) [\hat{R}_L(t_i)] \times [\hat{R}_R(t_i)]}{\sqrt{\sum_{i=1}^{M} \left| w_L(t_i) \hat{R}_L(t_i) \right|^2} \times \sqrt{\sum_{i=1}^{M} \left| w_R(t_i) \hat{R}_R(t_i) \right|^2}}, \tag{20}$$

where $\hat{R}_L(t_i) = R_L^o(t_i) - \frac{\sum_{t \in W(p)} w_L(t) R_L^o(t)}{Z(p)}$ and $R_L^o(\cdot)$ represents the original intensity value of the red channel in the left image. In a similar way, for the right image, $\hat{R}_R(t_i)$ can be computed. The weight distributions $w_L(\cdot)$ and $w_R(\cdot)$ are the same as those of (19). ANCC is a similarity measure that ranges from -1 to $+1$. In order to arrive at a nonnegative cost between pixel p and $p + f_p$ in the left and right images, respectively, the value of the ANCC is subtracted from $+1$. Now, by combining (19) and (20), we define our data cost $D_p(f_p)$ as follows:

$$D_p(f_p) = 1 - \left[\theta \sum_{\xi} \frac{ANCC_\xi(f_p)}{3} + (1 - \theta) \sum_{k} \frac{ANCC_k(f_p)}{3} \right], \tag{21}$$

where $\xi \in \{logChrom\_R,\ logChrom\_G,\ logChrom\_B\}$, $k \in \{R, G, B\}$. θ is a relative weighting factor between the log-chromaticity color and the original RGB color.

4.5 Global Energy Modeling

For the pairwise cost, we employed a simple truncated quadratic cost as shown by the following equation:

$$V_{pq}(f_p, f_q) = \lambda \cdot \min(\left| f_p - f_q \right|^2, V_{\max}). \tag{22}$$

By combining (21) and (22), the total energy is defined by the equation:

$$E(f) = \sum_{p} D_p(f_p) + \sum_{p} \sum_{q \in N(p)} V_{pq}(f_p, f_q). \tag{23}$$

This total energy was optimized using Graph-cuts (α-expansion) algorithm [19] in order to find the disparity map f. Note that although Graph-cuts (α-expansion) does not guarantee the global optimal solution for the energy model with a truncated quadratic smoothness prior, in practice, it can find a near-optimal solution quickly [52].

5 Experimental Results

In our experiments, we fixed all of the parameters of the proposed algorithm such that $\lambda = 1/30$, $V_{max} = 5$, $M = (31 \times 31)$, $\sigma_d = 14$, and $\sigma_s = 3.8$. For the evaluation and comparison of our method with others, we used various images such as the test bed images (Aloe, Art, Moebius, Dolls, Laundry, and Cloth4) in [1] and some aerial images. There are three different exposures (indexed as 0,1,2) and three different light sources (indexed as 1,2,3) in each data set in [1], making a total of nine different images. These data sets were used, similar to [5], to compare various data costs such as the Birchfield-Tomasi (BT) [6], BT with 5×5 Laplacian of Gaussian filtering [7] (LoG/BT), BT using Rank transform (9×9 window) [8] (Rank/BT), BT using images preprocessed by the comprehensive normalization method [29] (CN/BT), mutual-information-based data cost (MI) [9], NCC (7×7 window), and proposed ANCC. Testing was performed on all matching costs described in the earlier text for all possible combinations that could exist with respect to changes in exposure and light source. However, we avoided the case of the same illumination (or exposure) combination for left and right images. For instance, between the 1/3 (left/right illumination) and the 3/1 (left/right illumination) cases, we experimented only on the 1/3 case because the properties of 1/3 and 3/1 cases are similar, as shown in [5].

All methods are optimized in the MRF framework using Graph-cuts (GC) with truncated quadratic smoothness cost. In order to ensure that our experiment was fair, the parameters of all methods were carefully and optimally selected, similarly to how it was used in [5], using images without radiometric variations and fixed for all experiments.

We also investigated the effect of log-chromaticity color versus original color for ANCC by varying the relative weighting factor θ in (21). Therefore, we carried out our evaluations of ANCC with some different θ values in the following combinations: ANCC with original RGB color ($\theta = 0$), ANCC with log-chromaticity color ($\theta = 1$), and ANCC with two colors combined ($\theta = 0.7$).

5.1 Light Source Changes

For testing the effects of changes in light source (illumination), we set the index of exposure to 1 for all images and varied only the index of illumination from 1 to 3. Figs. 3a and 3b depict the Aloe stereo images taken under extremely different illumination condition (the left and the right images have been taken at an index of illumination of 1 and 3, respectively). Figs. 3d, 3e, 3f, 3g, 3h, 3i, 3j, 3k, and 3l show the disparity maps of test stereo algorithms for input stereo image pair in Figs. 3a and 3b, and Figs. 4 and 5 show the results of comparative test stereo algorithms carried out for the Dolls and Cloth4 images, respectively. Figs. 6a, 6b, 6c 6d, 6e, and 6f show the errors of unoccluded areas for all possible combinations of illuminations.

Changes in the light source can result in various local radiometric changes, which is one of the most difficult factors among the radiometric variations for matching problem. We can find that BT is very sensitive to these local radiometric variations. LoG/BT uses the LoG filter which is the second-order derivative of intensity. Although it has been proven that LoG/BT is more robust as compared to BT, there still remains the relatively large error with LoG/BT because the value of the derivative of intensity can vary as determined by local variations of intensity. Because of the change in the rank of intensities, Rank/BT is also affected by these local variations. There is a large error with MI when there are severe local variations in images caused by changes in the light sources because MI-based methods

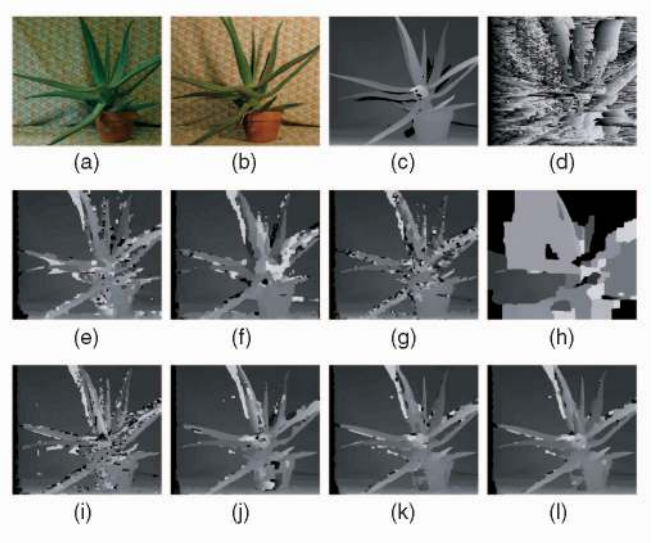

Fig. 3. Results of test stereo algorithms on Aloe image pair with varying illumination. (a) The left image with illumination(1)-exposure(1). (b) The right image with illumination(3)-exposure(1). (c) The ground truth disparity map. (d)-(l) are the disparity maps of test stereo algorithms for input stereo image pair in (a) and (b). (d) BT+GC. (e) CN/BT+GC. (f) LoG/BT+GC. (g) Rank/BT+GC. (h) MI+GC. (i) NCC+GC. (j) ANCC+GC with original RGB color. (k) ANCC+GC with log-chromaticity color. (l) ANCC+GC with colors combining log-chromaticity and original RGB colors.

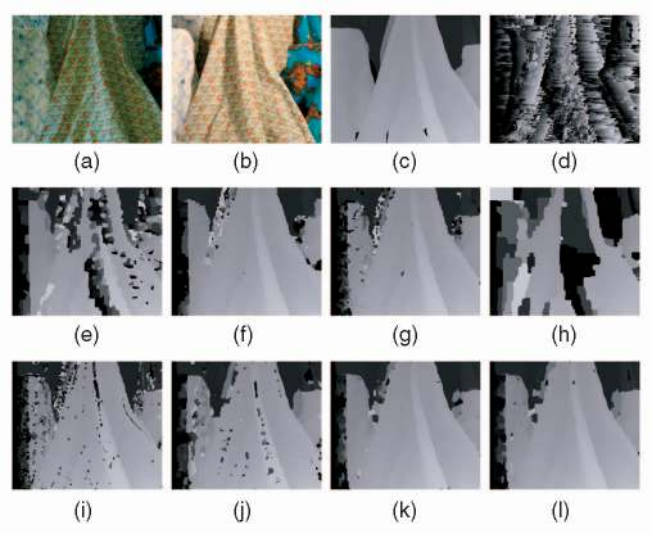

Fig. 5. Results of test stereo algorithms on Cloth4 image pair with varying illumination. (a) The left image with illumination(1)-exposure(1). (b) The right image with illumination(3)-exposure(1). (c) The ground truth disparity map. (d)-(l) are the disparity maps of test stereo algorithms for input stereo image pair in (a) and (b). (d) BT+GC. (e) CN/BT+GC. (f) LoG/BT+GC. (g) Rank/BT+GC. (h) MI+GC. (i) NCC+GC. (j) ANCC+GC with original RGB color. (k) ANCC+GC with log-chromaticity color. (l) ANCC+GC with colors combining log-chromaticity and original RGB colors.

[9], [14] account for only the global intensity changes between input images. NCC is also not strong at the local variation to input images because it assumes only the global affine-transformed difference. As described in the earlier text, CN/BT also does not produce good results because the variation in the image content between stereo images due to the view change affects color variation between corresponding pixels. From Fig. 6, it should be noted that in most test images there are steep increases of error between the (1/2) and (1/3) cases for other methods, whereas ANCC with log-chromaticity color and ANCC with combined colors show more stable results. For example, Fig. 7 demonstrates the detailed comparisons of the disparity maps of ANCC with log-chromaticity color, NCC, MI, and Rank/BT in accordance with the illumination variations for the Dolls images. The inference that can be drawn from this is that ANCC with log-chromaticity is robust at strong local variations, whereas all of the other methods are very sensitive to the changes in light sources. On the other hand, even within ANCC, ANCC with log-chromaticity shows robust property, while ANCC with original RGB color still remains weak at local variations. This shows that the log-chromaticity color is more effective in handling local variations as compared to the original RGB color. ANCC with combined colors show near-similar or slightly better results as compared to ANCC with log-chromaticity color.

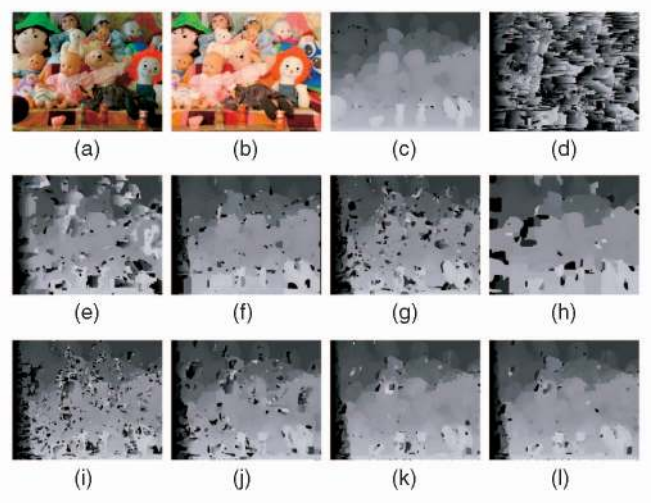

Fig. 4. Results of test stereo algorithms on Dolls image pair with varying illumination. (a) The left image with illumination(1)-exposure(1). (b) The right image with illumination(3)-exposure(1). (c) The ground truth disparity map. (d)-(l) are the disparity maps of test stereo algorithms for input stereo image pair in (a) and (b). (d) BT+GC. (e) CN/BT+GC. (f) LoG/BT+GC. (g) Rank/BT+GC. (h) MI+GC. (i) NCC+GC. (j) ANCC+GC with original RGB color. (k) ANCC+GC with log-chromaticity color. (l) ANCC+GC with colors combining log-chromaticity and original RGB colors.

5.2 Camera Exposure Changes

In order to test the effects of changes in camera exposure, we fixed the index of illumination to 1, and changed only the index of exposure from 0 to 2. Figs. 8a and 8b show the Aloe stereo images which have undergone extremely different exposure conditions (the left and the right images have been captured with an index of exposure of 0 and 2, respectively). Figs. 8d, 8e, 8f, 8g, 8h, 8i, 8j, 8k, and 8l depict the disparity maps of test stereo algorithms for input stereo image pair in Figs. 8a and 8b. In the same manner, Fig. 9 shows the results of comparative test stereo algorithms carried out for the Moebius image. Figs. 10a, 10b, 10c, 10d, 10e, and 10f show the error of unoccluded areas for all possible combinations of exposures.

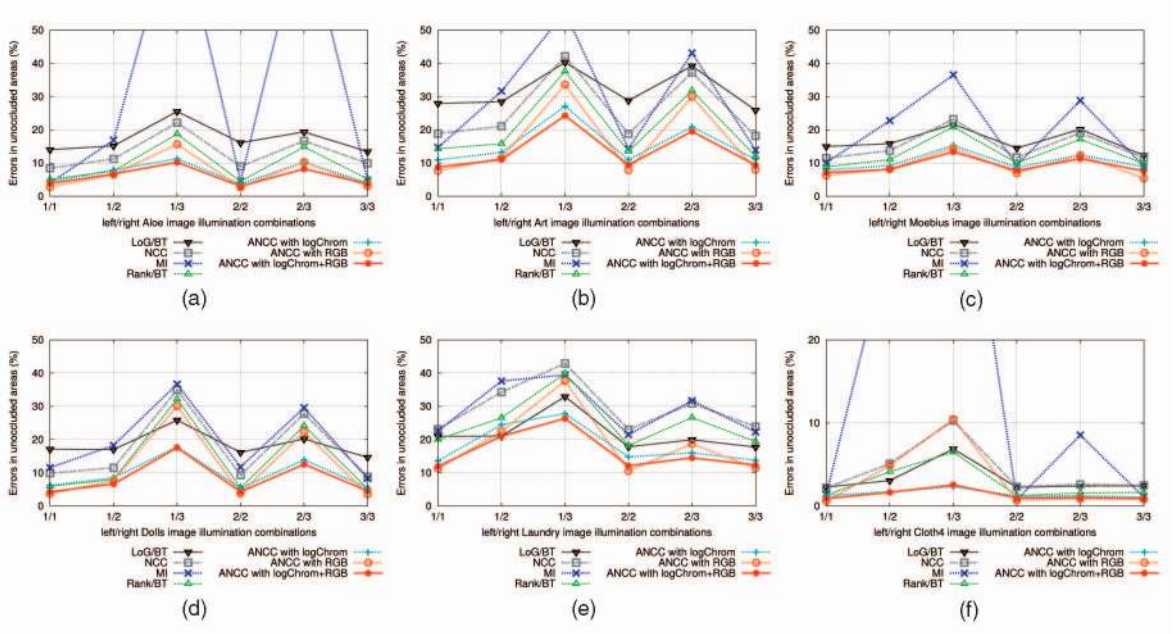

Fig. 6. Errors in the unoccluded areas of test stereo algorithms according to the different left/right image illumination combinations. "ANCC with logChrom" represents ANCC with log-chromaticity color ($\theta = 1$). "ANCC with RGB" represents ANCC with original RGB color ($\theta = 0$). "ANCC with logChrom+RGB" represents ANCC with two colors combined ($\theta = 0.7$). (a) Aloe image. (b) Art image. (c) Moebius image. (d) Dolls image. (e) Laundry image. (f) Cloth4 image.

Changes in exposure result in global changes in intensity between stereo images. In this global change, MI, NCC, Rank/BT, and ANCC put up a relatively good performance. On the contrary, BT, LoG/BT, and CN/BT are seriously affected because of these changes in exposure. The extreme changes in exposure make images either very dark or bright. These effects make image features such as edges indistinct. We can find that BT, LoG/BT, and CN/BT are more sensitive to camera exposure changes as compared to light source changes. LoG/BT is seriously influenced by these variations because the values of the derivative of intensity are smeared for very dark or bright regions. Although MI can find global relationship, the main error of MI occurs with respect to the saturated intensity region

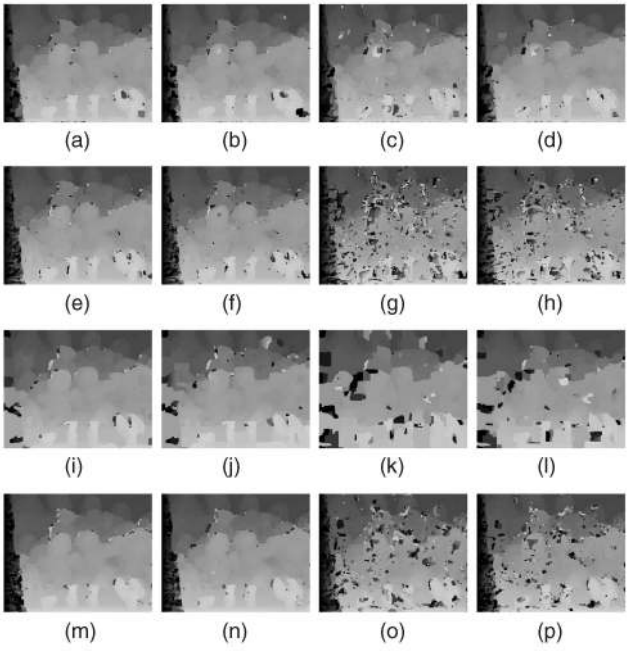

Fig. 7. Results of test stereo algorithms on Dolls image pair with varying illumination. (a)-(d), (e)-(h), (i)-(l), and (m)-(p) show the disparity maps of ANCC+GC, NCC+GC, MI+GC, and Rank/BT+GC methods, respectively, for different left/right illumination combinations. From left to right column, the (left illumination/right illumination) illumination combinations are (1/1), (1/2), (1/3), and (2/3), respectively.

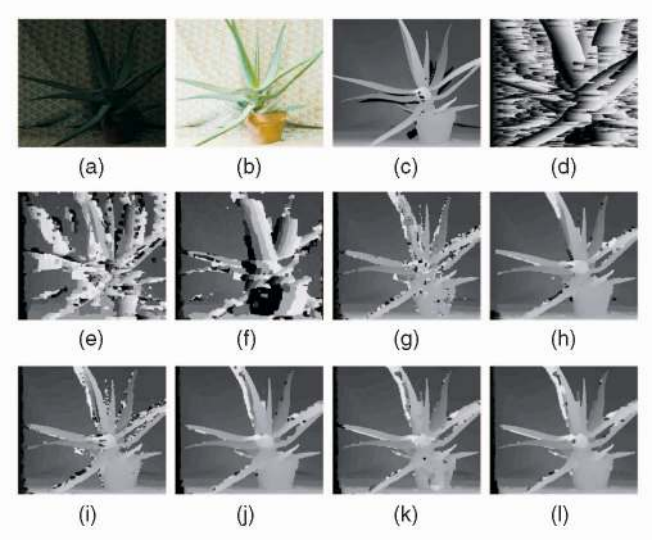

Fig. 8. Results of test stereo algorithms on Aloe image pair with varying exposure. (a) The left image with illumination(1)-exposure(0). (b) The right image with illumination(1)-exposure(2). (c) The ground truth disparity map. (d)-(l) are the disparity maps of test stereo algorithms for input stereo image pair (a)-(b). (d) BT+GC. (e) CN/BT+GC. (f) LoG/BT+GC. (g) Rank/BT+GC. (h) MI+GC. (i) NCC+GC. (j) ANCC+GC with original RGB color. (k) ANCC+GC with log-chromaticity color. (l) ANCC+GC with colors combining log-chromaticity and original RGB colors.

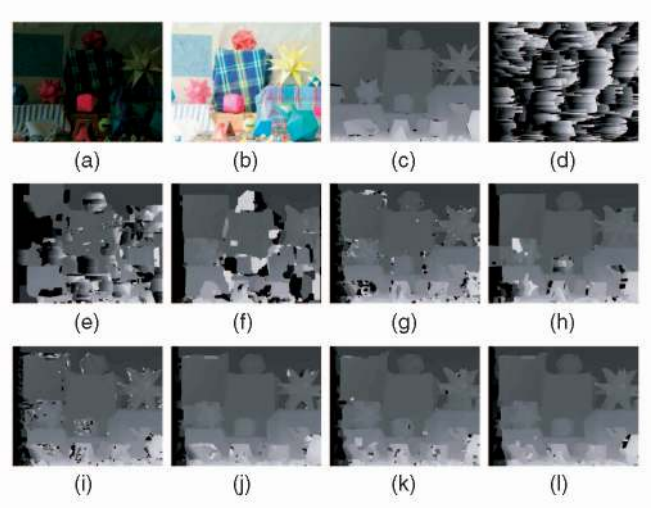

Fig. 9. Results of test stereo algorithms on Moebius image pair with varying exposure. (a) The left image with illumination(1)-exposure(0). (b) The right image with illumination(1)-exposure(2). (c) The ground truth disparity map. (d)-(l) are the disparity maps of test stereo algorithms for input stereo image pair (a)-(b). (d) BT+GC. (e) CN/BT+GC. (f) LoG/BT+GC. (g) Rank/BT+GC. (h) MI+GC. (i) NCC+GC. (j) ANCC+GC with original RGB color. (k) ANCC+GC with log-chromaticity color. (l) ANCC+GC with colors combining log-chromaticity and original RGB colors.

(e.g., very bright or dark region). Although the property of rank invariance in the case of global variation can be beneficial to Rank/BT, the main error of Rank/BT occurs due to the ambiguity of rank transform which prevails with similar rank values. It is here that the ANCC remains stable and performs better when compared with others with respect to exposure changes in most cases. However, ANCC with log-chromaticity color is somewhat unstable for the near-saturated color region, which has about (255,255,255) or (0,0,0) RGB color values. Therefore, this finding shows that the most accurate and stable results are shown by ANCC with original RGB color. ANCC with combined colors, on the other hand, shows near-similar results as ANCC with original RGB color.

5.3 Gamma Correction Changes

To test the effect of gamma variation, we set both the indexes of illumination and exposure to 1, and changed the gamma value (γ) of the right image. We transformed each color value to lie in the range $[0, 1]$. Gamma correction was then performed for each channel in accordance with the equation defined by

$$I_k(p) \rightarrow I_k(p)^{1/\gamma}, \quad k \in \{R, G, B\}. \tag{24}$$

Figs. 11a, 11b, 11c, 11d, 11e, and 11f show the error of unoccluded areas in accordance with the gamma variation of the right image. Variations in gamma correction influence the images in a global manner. As expected, the effect of variation in gamma correction is highly similar to the changes in exposure. That is the reason for MI, NCC, Rank/BT, and ANCC showing stable results to these variations. Although the performance of ANCC is stable to this gamma variation and the error is lower as compared to other methods, the error of ANCC increases slightly as the gamma value increases. This is because of the limited image representation for high gamma values.

5.4 Noise Variations

We also carried out investigations on the robustness of matching costs against noise variations. In order to test the effect of noise, we changed only the PSNR of the right image that was contaminated by the additive Gaussian noise. Fig. 12 shows the results of the comparative test according to the PSNR of the right images. The results of

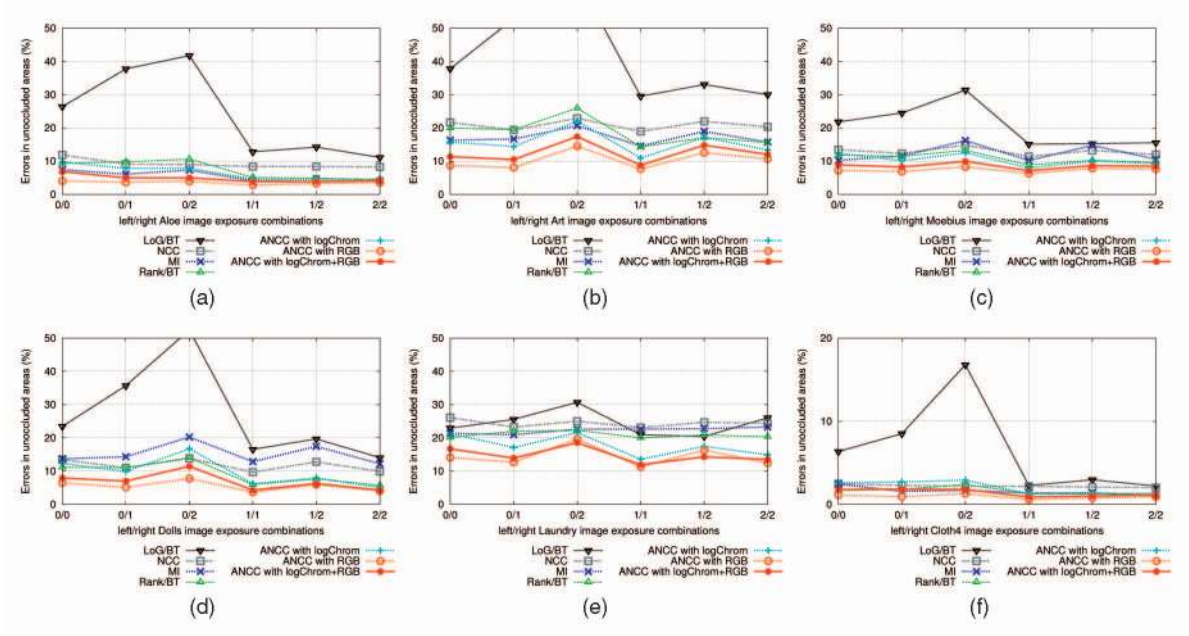

Fig. 10. Errors in the unoccluded areas of test stereo algorithms according to the different left/right image exposure combinations. "ANCC with logChrom" represents ANCC with log-chromaticity color ($\theta = 1$). "ANCC with RGB" represents ANCC with original RGB color ($\theta = 0$). "ANCC with logChrom+RGB" represents ANCC with two colors combined ($\theta = 0.7$). (a) Aloe image. (b) Art image. (c) Moebius image. (d) Dolls image. (e) Laundry image. (f) Cloth4 image.

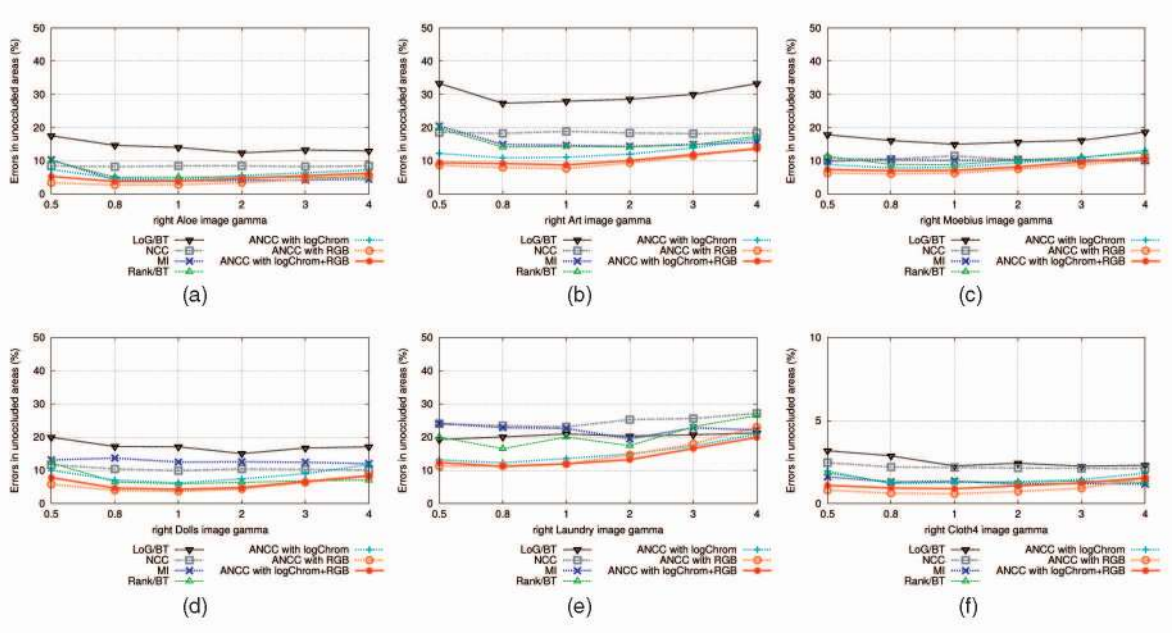

Fig. 11. Errors in the unoccluded areas of test stereo algorithms according to the gamma changes of the right image. "ANCC with logChrom" represents ANCC with log-chromaticity color ($\theta = 1$). "ANCC with RGB" represents ANCC with original RGB color ($\theta = 0$). "ANCC with logChrom+RGB" represents ANCC with two colors combined ($\theta = 0.7$). (a) Aloe image. (b) Art image. (c) Moebius image. (d) Dolls image. (e) Laundry image. (f) Cloth4 image.

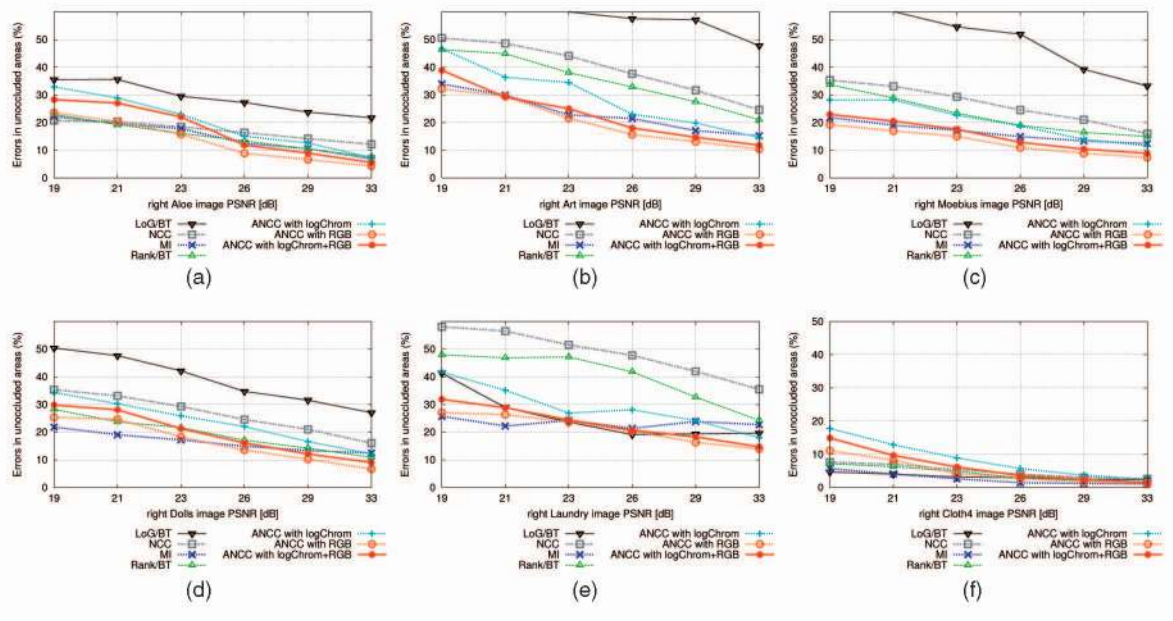

Fig. 12. Errors in the unoccluded areas of test stereo algorithms according to the PSNR changes of the right image. "ANCC with logChrom" represents ANCC with log-chromaticity color ($\theta = 1$). "ANCC with RGB" represents ANCC with original RGB color ($\theta = 0$). "ANCC with logChrom+RGB" represents ANCC with two colors combined ($\theta = 0.7$). (a) Aloe image. (b) Art image. (c) Moebius image. (d) Dolls image. (e) Laundry image. (f) Cloth4 image.

noise variations for LoG/BT are highly sensitive because the LoG operation is the second-order derivative in which the effect of noise is considerable. There is considerable effect of noise for NCC also. In most cases, Rank/BT is more robust when compared to LoG/BT and NCC with respect to noise variations. However, it is also affected severely when the noise level is high because the noise can change the rank transform. MI, on the other hand, shows highly robust and accurate results to noise. ANCC with original RGB color also shows robust property to noise, which is comparable to MI. However, ANCC with log-chromaticity color poses problems to strong noise because the log operation can increase the noise factors. ANCC with the combined colors yields a compromise between noise sensitivity and illumination invariance.

5.5 Application to Aerial Image

Aerial images are captured using only one camera at different time intervals. There are changes in the colors of aerial images because of the different time intervals while

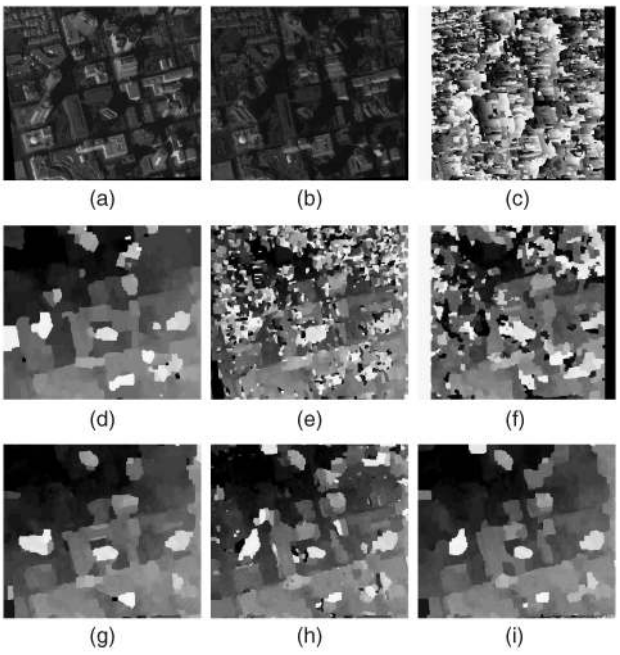

Fig. 13. Results of test stereo algorithms on aerial image pair. (a)-(b) The left and right aerial image pair. (c) BT+GC (d) MI+GC. (e) NCC+GC. (f) Rank/BT+GC. (g) ANCC+GC with original RGB color. (h) ANCC+GC with log-chromaticity color. (i) ANCC+GC with colors combining log-chromaticity and original RGB colors.

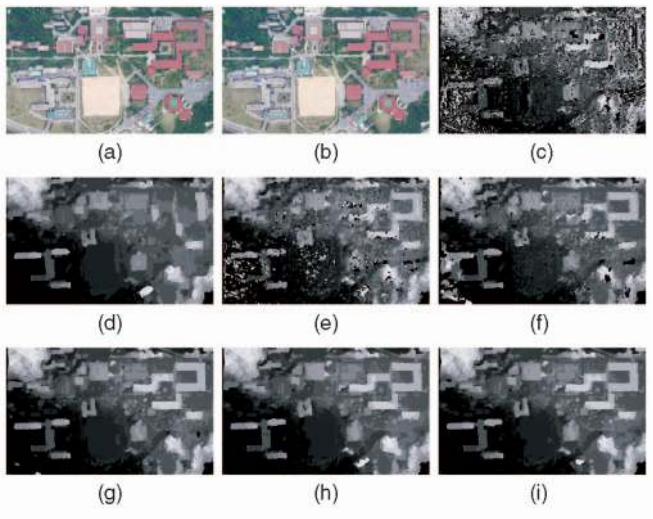

Fig. 14. Results of test stereo algorithms on aerial image pair. (a)-(b) The left and right aerial image pair. (c) BT+GC (d) MI+GC. (e) NCC+GC. (f) Rank/BT+GC. (g) ANCC+GC with original RGB color. (h) ANCC+GC with log-chromaticity color. (i) ANCC+GC with colors combining log-chromaticity and original RGB colors.

taking pictures. The colors of the aerial images are affected by complex factors such as changes in lighting color and position. Therefore, aerial images offer more challenges for stereo matching. The performances of various matching costs for aerial images are compared in Figs. 13 and 14. For instance, Figs. 13a and 13b are stereo images of a complicated urban area which has severe illumination and view changes. NCC and Rank/BT produced noisy and inaccurate disparity maps in these images. The result of MI shows ambiguous building boundary. The proposed ANCC produced highly accurate disparity map with sharp building boundaries. For the other aerial images such as Figs. 14a and 14b, our method also produced more accurate results as compared to other methods.

On the other hand, ANCC with log-chromaticity color produced more accurate building boundaries as compared to ANCC with original RGB color because there exist severe local radiometric variations in these aerial images. ANCC with two colors combined produced near-similar and less noisy results as compared to ANCC with log-chromaticity color.

6 Discussion

6.1 Effects of Adaptive Weight in ANCC

We carried out an investigation of the effects of the adaptive weight in the ANCC framework. The adaptive weight is used at two parts in the ANCC, namely, the adaptive mean computation part (16) and the adaptive correlation computation part (19). The effect of each of the two parts is investigated and analyzed separately in the following text:

- **Effect of adaptive weight for mean computation of each window**: The effect of adaptive weight for mean computation of each window in ANCC ($\theta = 1$) was investigated. In ANCC, in order to find the coherent mean value, we subtracted the adaptive mean value instead of global mean value. In order to evaluate this, we carried out an additional test on ANCC with a global mean approach that sets all weights $w(\cdot)$ in (16) equally. We then launched a comparison of the ANCC using adaptive mean approach and the ANCC using global mean approach. Fig. 15 clearly shows the benefits of the adaptive mean approach over the simple global mean approach. (a)-(d) in Fig. 15 is the result of the global mean approach, whereas (e)-(h) is the result of the adaptive mean approach. For stereo images without radiometric variations ((a) and (e) and (c) and (g) in Fig. 15), the main error comes from the geometric variations such as viewpoint changes. The greater the change in the viewpoint, the greater the

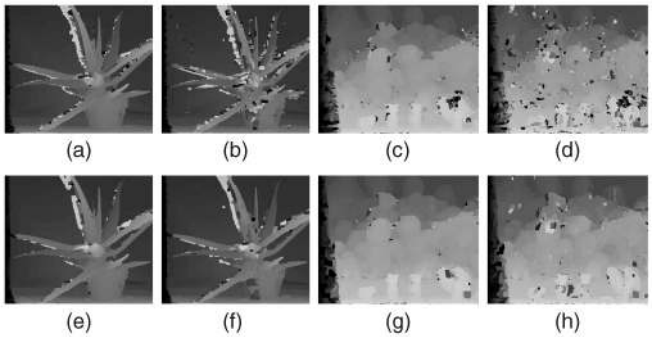

Fig. 15. The effects of adaptive weight for mean computation of each window. (a)-(d) show results for the global mean approach, while (e)-(h) show the proposed adaptive mean approach. (a) and (e) and (c) and (g) are results for Aloe and Dolls images without radiometric variation. (b) and (f) and (d) and (h) are results for Aloe and Dolls images with illumination variations. (a) err: 4.7 percent. (b) err: 15.8 percent. (c) err: 7.7 percent. (d) err: 27.8 percent. (e) err: 4.5 percent. (f) err: 11.3 percent. (g) err: 7.3 percent. (h) err: 18.6 percent.

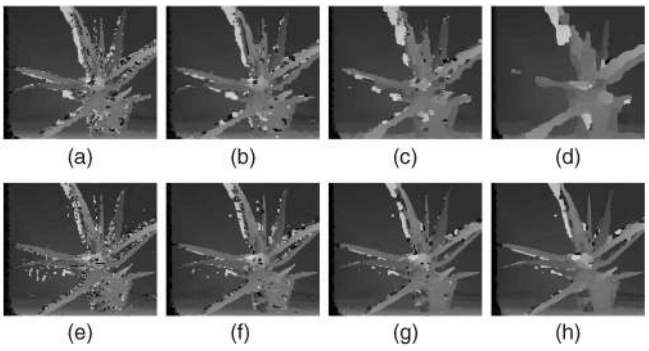

Fig. 16. The effects of adaptive weight in correlation computation. (a)-(d) show the results of ANCC without adaptive correlation weight, whereas (e)-(h) show the results of ANCC with adaptive correlation weight. From left to right columns, the window sizes are 7×7, 11×11, 21×21, 31×31, respectively. (a) err: 15.48 percent. (b) err: 18.17 percent. (c) err: 20.95 percent. (d) err: 26.27 percent. (e) err: 15.93 percent. (f) err: 12.86 percent. (g) err: 11.42 percent. (h) err: 11.36 percent.

change in the mean value. This creates some problems for window matching in the global mean approach such as the original NCC framework. For instance, the region of the front stem of the Aloe and region of the feet for the Dolls are prone to errors with respect to the simple global mean approach. However, this problem can be reduced by using the adaptive mean approach. On the other hand, for stereo images that have illumination variations ((b) and (f) and (d) and (h) in Fig. 15), our adaptive mean approach also shows more robust properties as compared to the simple global mean approach because the adaptive mean approach has the function of suppressing outliers. Thus, we find that the adaptive mean approach is more accurate as well as robust even for illumination variations against the simple global mean approach.

- **Effect of adaptive weight for correlation computation**: In order to investigate the effect of adaptive weight for the correlation computation part, we additionally tested the nonadaptive correlation approach which sets all weights $w(\cdot)$ in (19) equally. As shown in Fig. 16, we compared the adaptive correlation approach with the nonadaptive correlation approach by varying window size as 7×7, 11×11, 21×21, and 31×31. Fig. 16 shows the comparative results for the Aloe image pair with illumination variations as well as the effects and benefits of adaptive weight in the correlation computation. The finding of this comparison is that without the adaptive weight for the correlation (Figs. 16a, 16b, 16c, and 16d), the error is increased and the object boundaries are blurred as the size of the window increases. On the contrary, our adaptive weight correlation scheme (Figs. 16e, 16f, 16g, and 16h) is effective in producing sharp and accurate results as the size of the window increases. Thus, it is here that the role of adaptive weight for correlation computation comes into play, which is to avoid the fattening effect even for large window matching.

6.2 Effect of Only Similarity Measure

In order to see the effect of only the proposed similarity measure ANCC, we performed some experiments to obtain disparity maps using the Winner-Takes-All (WTA) approach instead of Graph-cuts. The WTA is the local method that takes the best label for each pixel using only data cost volume. Fig. 17 demonstrates a part of the results using WTA method for each matching cost for the Aloe images. Figs. 17a, 17b, 17c, 17d, and 17e are the results for the Aloe stereo image pair without radiometric variation, whereas Figs. 17f, 17g, 17h, 17i, and 17j are the results for the Aloe stereo image pair with illumination variation. It can be noted that the ANCC has the least error when compared to other methods. Furthermore, ANCC with WTA shows very accurate results similar to ANCC with Graph-cuts. From this finding, it can be inferred that ANCC with global optimization such as Graph-cuts also shows highly accurate and robust properties.

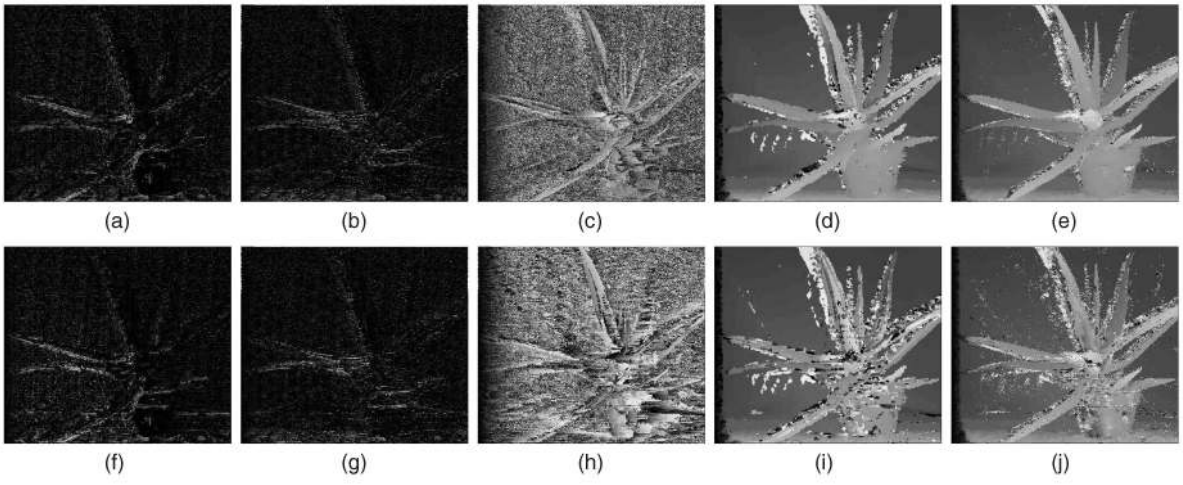

Fig. 17. (a) LoG/BT+WTA (err: 96.8 percent). (b) Rank/BT+WTA (err: 96.5 percent). (c) MI+WTA (err: 80.1 percent). (d) NCC+WTA (err: 11.7 percent). (e) ANCC+WTA (err: 9.8 percent). (f) LoG/BT+WTA (err: 97.7 percent). (g) Rank/BT+WTA (err: 97.3 percent). (h) MI+WTA (err: 96.7 percent). (i) NCC+WTA (err: 26.5 percent). (j) ANCC+WTA (err: 22.9 percent). (a)-(e) are the results for Aloe stereo image pair without radiometric variations. (f)-(j) are the results for Aloe stereo image pair with illumination variations.

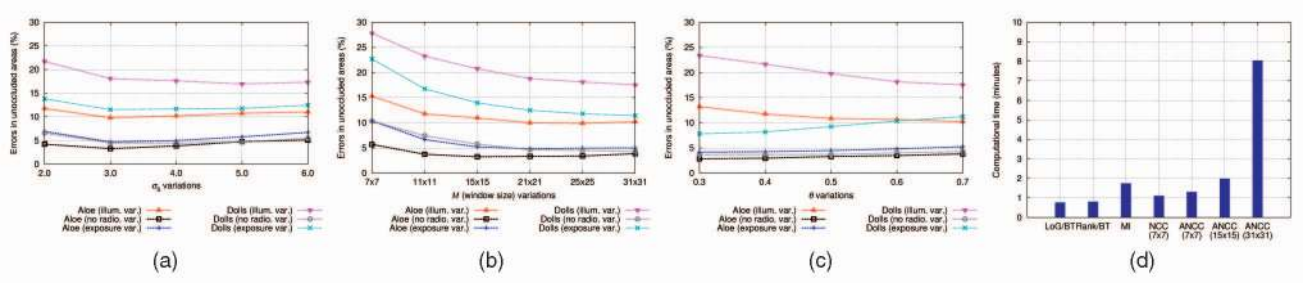

Fig. 18. Performance of the proposed algorithm according to (a) σ_s, (b) window size M, and (c) θ. (d) Computational time.

6.3 Performance Assessment Based on Parameter Values and Comparison of Computational Time

In this section, the performance of ANCC according to parameter variations and the comparison of computational time are investigated. ANCC includes following four parameters: σ_s, σ_d, window size M, and θ. To investigate the effect of each parameter independently, we varied each parameter value while fixing others as described in Section 5. Figs. 18a, 18b, and 18c show the changes in the performance of ANCC according to the variation of three parameters, σ_s, M, and θ, respectively, for Aloe and Dolls images under different illuminations (left illum.(1) and right illum.(3)), different exposures (left exp.(0) and right exp.(2)), and no radiometric change. We observe from Fig. 18a that ANCC is not so sensitive to the variation of σ_s value. Note that since σ_d is dependent on the window size M, we only tested for M. As shown in Fig. 18b, when M was larger than "15×15," it gave nearly constant performance. The parameter θ determines the relative weighting between the log-chromaticity color and the RGB color. So, the larger θ is, the more log-chromaticity color is emphasized. Empirically, the values of θ between 0.6 and 0.7 yield an appropriate level of compromise between the log-chromaticity color and the RGB color.

Fig. 18d compares computational time (including the Graph-cuts optimization) of various methods for the Aloe data set (image size: 427×370, disparity range: 0-70) using a PC with an Intel Pentium-IV 2.4GHz CPU. The numbers in the parentheses indicate window size. The computational time of ANCC quadratically increases with the window size. However, this computational time increase can be reduced by using parallel processors such as GPUs, because most of the time is spent in computing correlations and weights of pixels in windows.

7 Conclusions

From a practical perspective, it is a requirement for the stereo matching algorithm to be both robust to radiometric variations as well as accurate. However, in general, the achievement of both these requirements is not trivial. Although methods such as Rank transform, mutual information, and NCC have achieved robustness to some extent, they are unable to achieve high accuracy. In addition, although the original color or intensity remains sufficient for global radiometric variations, it is still weak at local variations. As a remedy to these problems, we have proposed in this paper a new stereo matching measure that is robust to various radiometric variations, such as local and global radiometric variations, and at the same time accurate. For the purposes of dealing with various radiometric variations, we took the color formation model into consideration in an explicit manner instead of using raw color or intensity values. From the original color, we have extracted color-insensitive information called *log-chromaticity* color. Further, for increasing the accuracy, we have proposed a new measure called *Adaptive Normalized Cross-Correlation*. Our method produces promising results that are quite robust to various kinds of radiometric changes.

It would be fair to comment on certain limitations of our method, which are as follows: The method cannot handle the multiple illumination conditions as well as non-Lambertian reflectance objects. These problems would be resolved in future work.

It is worth noting that our matching measure can be effectively applied to other vision problems that require correspondence and matching such as feature matching, registration, and visual tracking. Also, it can be extended to solve the multiple-view stereo problems which can be impacted by severe radiometric variations between input stereo images.

Acknowledgments

The authors would like to thank the anonymous reviewers for their constructive comments. 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 of the Ministry of Information and Communication.

References

[1] http://vision.middlebury.edu/stereo/, 2010.
[2] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," *Int'l J. Computer Vision*, vol. 47, no. 1, pp. 7-42, 2002.
[3] 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*, vol. 6, no. 7, pp. 965-976, July 1997.
[4] Y.S. Heo, K.M. Lee, and S.U. Lee, "Illumination and Camera Invariant Stereo Matching," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2008.
[5] H. Hirschmüller and D. Scharstein, "Evaluation of Cost Functions for Stereo Matching," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2007.
[6] S. Birchfield and C. Tomasi, "A Pixel Dissimilarity Measure that Is Insensitive to Image Sampling," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 4, pp. 401-406, Apr. 1998.
[7] H. Hirschmüller, P. Innocent, and J. Garibaldi, "Real-Time Correlation-Based Stereo Vision with Reduced Border Errors," *Int'l J. Computer Vision*, vol. 47, nos. 1-3, pp. 229-246, 2002.

[8] R. Zabih and J. Woodfill, "Non-Parametric Local Transforms for Computing Visual Correspondence," *Proc. European Conf. Computer Vision,* 1994.
[9] H. Hirschmüller, "Stereo Processing by Semiglobal Matching and Mutual Information," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 30, no. 2, pp. 328-341, Feb. 2008.
[10] 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," Technical Report RR-2013, INRIA, 1993.
[11] M. Goesele, N. Snavely, B. Curless, H. Hoppe, and S.M. Seitz, "Multi-View Stereo for Community Photo Collections," *Proc. IEEE Int'l Conf. Computer Vision,* 2007.
[12] P. Viola and W.M. Wells, III, "Alignment by Maximization of Mutual Information," *Int'l J. Computer Vision,* vol. 24, no. 2, pp. 137-154, 1997.
[13] G. Egnal, "Mutual Information as a Stereo Correspondence Measure," Technical Report MS-CIS-00-20, Computer and Information Science, Univ. of Pennsylvania, 2000.
[14] J. Kim, V. Kolmogorov, and R. Zabih, "Visual Correspondence Using Energy Minimization and Mutual Information," *Proc. IEEE Int'l Conf. Computer Vision,* 2003.
[15] L. Wang, R. Yang, and J.E. Davis, "BRDF Invariant Stereo Using Light Transport Constancy," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 29, no. 9, pp. 1616-1626, Sept. 2007.
[16] A.S. Ogale and Y. Aloimonos, "Robust Contrast Invariant Stereo Correspondence," *Proc. IEEE Int'l Conf. Robotics and Automation,* 2004.
[17] S. Negahdaripour, "Revised Definition of Optical Flow: Integration of Radiometric and Geometric Cues for Dynamic Scene Analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, no. 9, pp. 961-979, Sept. 1998.
[18] J. Zhang, L. McMillan, and J. Yu, "Robust Tracking and Stereo Matching under Variable Illumination," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2006.
[19] Y. Boykov, O. Veksler, and R. Zabih, "Fast Approximate Energy Minimization via Graph Cuts," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 11, pp. 1222-1239, Nov. 2001.
[20] T. Meltzer, C. Yanover, and Y. Weiss, "Globally Optimal Solutions for Energy Minimization in Stereo Vision Using Reweighted Belief Propagation," *Proc. IEEE Int'l Conf. Computer Vision,* 2005.
[21] V. Kolmogorov, "Convergent Tree-Reweighted Message Passing for Energy Minimization," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 28, no. 10, pp. 1568-1583, Oct. 2006.
[22] E.H. Land and J.J. McCann, "Lightness and Retinex Theory," *J. Optical Soc. of Am. A,* vol. 61, no. 1, pp. 1-11, 1971.
[23] B. Funt, F. Ciurea, and J. McCann, "Retinex in Matlab," *Proc. Eighth Color Imaging Conf. Color Science Systems and Applications,* 2000.
[24] R. Kimmel, M. Elad, D. Shaked, R. Keshet, and I. Sobel, "A Variational Framework for Retinex," *Int'l J. Computer Vision,* vol. 52, no. 1, pp. 7-23, 2003.
[25] D.H. Brainard and B.A. Wandell, "Analysis of the Retinex Theory of Color Vision," *J. Optical Soc. of Am. A,* vol. 3, no. 10, pp. 1651-1661, 1986.
[26] D.A. Forsyth, "A Novel Algorithm for Color Constancy," *Int'l J. Computer Vision,* vol. 5, no. 1, pp. 5-24, 1990.
[27] 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,* vol. 23, no. 11, pp. 1209-1221, Nov. 2001.
[28] B. Funt, K. Barnard, and L. Martin, "Is Machine Colour Constancy Good Enough?" *Proc. European Conf. Computer Vision,* 1998.
[29] G.D. Finlayson, B. Schiele, and J.L. Crowley, "Comprehensive Colour Image Normalization," *Proc. European Conf. Computer Vision,* 1998.
[30] G.D. Finlayson and S.D. Hordley, "Color Constancy at a Pixel," *J. Optical Soc. of Am. A,* vol. 18, no. 2, pp. 253-264, 2001.
[31] G.D. Finlayson and R. Xu, "Illuminant and Gamma Comprehensive Normalisation in Log RGB Space," *Pattern Recognition Letters,* vol. 24, no. 11, pp. 1679-1690, 2003.
[32] J.M. Geusebroek, R. van den Boomgaard, A.W.M. Smeulders, and H. Geerts, "Color Invariance," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 12, pp. 1338-1350, Dec. 2001.
[33] T. Gevers and A.W.M. Smeulders, "Color Based Object Recognition," *Pattern Recognition,* vol. 32, no. 1, pp. 453-464, 1999.
[34] T. Gevers and H. Stokman, "Robust Histogram Construction from Color Invariants for Object Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 26, no. 1, pp. 113-118, Jan. 2004.
[35] G.D. Finlayson, S.D. Hordley, and M.S. Drew, "Removing Shadows from Images," *Proc. European Conf. Computer Vision,* 2002.
[36] M. Ebner, "Color Constancy Using Local Color Shifts," *Proc. European Conf. Computer Vision,* 2004.
[37] D. Berwick and S.W. Lee, "A Chromaticity Space for Specularity, Illumination Color- and Illumination Pose-Invariant 3D Object Recognition," *Proc. IEEE Int'l Conf. Computer Vision,* 1998.
[38] R.T. Tan and K. Ikeuchi, "Separating Reflection Components of Textured Surfaces Using a Single Image," *Proc. IEEE Int'l Conf. Computer Vision,* 2003.
[39] T. Zickler, S.P. Mallick, D.J. Kriegman, and P.N. Belhumeur, "Color Subspaces as Photometric Invariants," *Int'l J. Computer Vision,* vol. 79, no. 1, pp. 13-30, 2008.
[40] S.A. Shafer, "Using Color to Separate Reflection Components," *COLOR Research and Applications,* vol. 10, no. 4, pp. 210-218, 1985.
[41] G.J. Klinker, S.A. Shafer, and T. Kanade, "A Physical Approach to Color Image Understanding," *Int'l J. Computer Vision,* vol. 4, no. 1, pp. 7-38, 1990.
[42] R.W.G. Hunt, *The Reproduction of Color,* fifth ed. Fountain Press, 1995.
[43] G.D. Finlayson, M.S. Drew, and B.V. Funt, "Diagonal Transforms Suffice for Color Constancy," *Proc. IEEE Int'l Conf. Computer Vision,* 1993.
[44] G.D. Finlayson, M.S. Drew, and B.V. Funt, "Spectral Sharpening: Sensor Transformations for Improved Color Constancy," *J. Optical Soc. of Am. A,* vol. 11, no. 5, pp. 1553-1563, 1994.
[45] H.Y. Chong, S.J. Gortler, and T. Zickler, "The von Kries Hypothesis and a Basis for Color Constancy," *Proc. IEEE Int'l Conf. Computer Vision,* 2007.
[46] P.E. Debevec and J. Malik, "Recovering High Dynamic Range Radiance Maps from Photographs," *Proc. ACM SIGGRAPH,* 1997.
[47] T. Mitsunaga and S.K. Nayar, "Radiometric Self Calibration," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 1999.
[48] M.D. Grossberg and S.K. Nayar, "Modeling the Space of Camera Response Functions," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 26, no. 10, pp. 1272-1282, Oct. 2004.
[49] C. Tomasi and R. Manduchi, "Bilateral Filtering for Gray and Color Images," *Proc. IEEE Int'l Conf. Computer Vision,* 1998.
[50] K. Yoon and I.S. Kweon, "Adaptive Support-Weight Approach for Correspondence Search," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 28, no. 4, pp. 650-656, Apr. 2006.
[51] J. Shi and J. Malik, "Normalized Cuts and Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 8, pp. 888-905, Aug. 2000.
[52] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "A Comparative Study of Energy Minimization Methods for Markov Random Fields with Smoothness-Based Priors," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 30, no. 6, pp. 1068-1080, June 2008.

Yong Seok Heo received the BS degree in electrical engineering and the MS degree in electrical engineering and computer science from Seoul National University, Korea, in 2005 and 2007, respectively. He is currently working toward the PhD degree in electrical engineering and computer science at Seoul National University. His research interests include stereo matching, 3D reconstruction, and computational photography. He is a student member of the IEEE.

Kyoung Mu Lee received the BS and MS degrees in control and instrumentation engineering from Seoul National University (SNU), Korea, in 1984 and 1986, respectively, and the PhD degree in electrical engineering from the University of Southern California (USC), Los Angeles, in 1993. He was awarded the Korean Government Overseas Scholarship during his PhD courses. From 1993 to 1994, he was a research associate in the Signal and Image Processing Institute (SIPI) at USC. He was with Samsung Electronics Co. Ltd. in Korea as a senior researcher from 1994 to 1995. In August 1995, he joined the Department of Electronics and Electrical Engineering of Hong-Ik University, and worked as an assistant and associate professor. Since September 2003, he has been with the Department of Electrical Engineering and Computer Science at Seoul National University as a professor, where he leads the Computer Vision Laboratory. His primary research is focused on statistical methods in computer vision that can be applied to various applications including object recognition, segmentation, tracking, and 3D reconstruction. He has received several awards, in particular, the Most Influential Paper over the Decade Award by the IAPR Machine Vision Application in 2009, the ACCV Honorable Mention Award in 2007, the Okawa Foundation Research Grant Award in 2006, and the Outstanding Research Award by the College of Engineering of SNU in 2010. He served as an Editorial Board member of the *EURASIP Journal of Applied Signal Processing*, and is an associate editor of the *Machine Vision Application Journal*, the *IPSJ Transactions on Computer Vision and Applications*, and the *Journal of Information Hiding and Multimedia Signal Processing*. He has (co)authored more than 100 publications in refereed journals and conferences including TPAMI, IJCV, CVPR, ICCV, and ECCV. He is a member of the IEEE.

Sang Uk Lee received the BS degree from Seoul National University, Korea, in 1973, the MS degree from Iowa State University, Ames, in 1976, and the PhD degree from the University of Southern California, Los Angeles, in 1980, all in electrical engineering. From 1980 to 1981, he was with General Electric Company, Lynchburg, Virginia, working on the development of digital mobile radio. From 1981 to 1983, he was a member of the technical staff at M/A-COM Research Center, Rockville, Maryland. In 1983, he joined the Department of Control and Instrumentation Engineering, Seoul National University, as an assistant professor, where he is now a professor in the department of electrical engineering and computer science. He is also affiliated with the Automation and Systems Research Institute and the Institute of NewMedia and Communications, both at Seoul National University. He was the president of the Korean Institute of Communication Science in 2005. His current research interests are in the areas of image and video signal processing, digital communication, and computer vision. He served as editor-in-chief for the *Transactions of the Korean Institute of Communication Science* from 1994 to 1996. He was an associate editor for the *IEEE Transactions on Circuits and Systems for Video Technology* from 2002 to 2005, and was on the Editorial Board of the *EURASIP Journal of Applied Signal Processing* from 2003 to 2004. He is currently on the Editorial Board of the *Journal of Visual Communication and Image Representation*. He is a member of Phi Kappa Phi. He is a fellow of the IEEE.