Mutual Information-based Stereo Matching Combined with SIFT Descriptor in Log-chromaticity Color Space

Yong Seok Heo, Kyoung Mu Lee, and Sang Uk Lee
Department 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

Radiometric variations between input images can seriously degrade the performance of stereo matching algorithms. In this situation, mutual information is a very popular and powerful measure which can find any global relationship of intensities between two input images taken from unknown sources. The mutual information-based method, however, is still ambiguous or erroneous as regards local radiometric variations, since it only accounts for global variation between images, and does not contain spatial information properly. In this paper, we present a new method based on mutual information combined with SIFT descriptor to find correspondence for images which undergo local as well as global radiometric variations. We transform the input color images to log-chromaticity color space from which a linear relationship can be established. To incorporate spatial information in mutual information, we utilize the SIFT descriptor which includes near pixel gradient histogram to construct a joint probability in log-chromaticity color space. By combining the mutual information as an appearance measure and the SIFT descriptor as a geometric measure, we devise a robust and accurate stereo system. Experimental results show that our method is superior to the state-of-the art algorithms including conventional mutual information-based methods and window correlation methods under various radiometric changes.

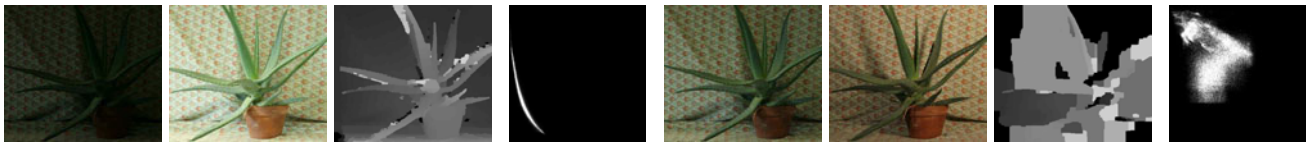
1. Introduction

The performance of stereo algorithms depends on the choice of matching cost. For images taken from radiometrically calibrated circumstances, corresponding pixels should have similar intensity values. For these images, simple matching costs such as absolute difference of intensities do not degrade the performance of stereo algorithms. In a real situation, however, image color values can be affected by radiometric variations including global intensity change (caused by camera gain and exposure or gamma correction variation) and local intensity change (caused by varying light, vignetting and non-Lambertian surface) and noise [7]. These variations often occur in general and practical settings and seriously degrade the performance of stereo. Therefore, robust matching methods to these radiometric variations are demanded and become inevitable for various applications such as the general multi-view stereo (for example, PhotoTourism [14], PhotoModeler [2], etc.) and the stereo matching of aerial images.

For this problem, mutual information (MI) can be a good candidate as a matching cost, since it can find correspondence for any globally transformed images captured from a wide variety of sensors. Owing to this advantage, MI has been successfully applied to various vision problems such as registration [16, 12, 13] and stereo matching [4, 9, 6]. Viola and Wells [16] adopted it to register MR (Magnetic Resonance) images and CT (Computed Tomography) images. For stereo matching, Egnal [4] used it to find the disparity among the local matching windows. It suffers, however, from the fixed window problem at low-texture region and discontinuities. To resolve it, Kim *et al*. [9] suggested a pixel-wise data cost based on MI in a global energy minimization framework. They approximated the MI of the whole images as a sum of pixel-wise MI using the Taylor expansion and the Parzen window technique. Hirschmuller [6] improved the work of [9] to handle the occlusion in a semi-global energy framework. Zitnick *et al*. [18] used approximative segment-wise MI for stereo matching. They defined a matching function by creating the histogram of the ratios of the corresponding pixel intensity values in case the gain difference of cameras dominates.

From the recent evaluation of various matching costs [7], however, it is shown that MI is weak at local radiometric variations, though it is the best measure at global variation. Hirschmuller and Scharstein [7] evaluated the performance and insensitivity of widely used matching costs with respect to various radiometric variations. They found that none of the matching costs are very robust to strong local variations. Heo *et al*. [5] suggested an Adaptive Normalized Cross Correlation (ANCC) measure that can handle local variation as

978-1-4244-3991-1/09/$25.00 ©2009 IEEE



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 1. (a) and (b) are the left and right images taken by different camera exposures. (c) Disparity map of stereo image (a) and (b) using conventional MI-based method. (d) Joint pdf corresponding to (c). (e) and (f) are the left and right images taken by different light configurations. (g) Disparity map of stereo image (e) and (f) using conventional MI-based method. (h) Joint pdf corresponding to (g).

well as global one. ANCC is, however, still weak at camera exposure variations.

On the other hand, most conventional MI-based methods use only pixel brightness statistics, and ignore rich spatial information to find correspondence. To increase robustness and accuracy, spatial information can be naturally incorporated. Pluim *et al*. [12] combined MI with the gradient information for registration of medical images by multiplying the MI with a gradient weight that accounted for gradient magnitude and orientation of the matching pixel. Russakoff *et al*. [13] suggested a regional mutual information to take the region information into account in MI, and applied it to medical image registration. Jeon *et al*. [8] suggested a MI-based local description method to find global correspondence between two images.

In this paper, we propose a new stereo matching algorithm that is based on MI combined with SIFT descriptor [11] to cope with the two problems; weakness in local radiometric variations and lack of spatial information. Our method can find correspondence under local as well as global radiometric variations by employing log-chromaticity color space. To incorporate spatial information, we utilize the SIFT descriptor [11] in constructing the joint probability. Finally, we also combine MI as an appearance measure and SIFT descriptor distance as a geometric measure in our data cost to make it further accurate and robust.

2. Mutual Information as a Stereo Correspondence Measure in MAP-MRF Framework

In MAP-MRF framework, the disparity map f can be found by minimizing the following energy $E(f)$:

$$E(f) = E_{data}(f) + E_{smooth}(f), \tag{1}$$

where the data energy $E_{data}(f)$ and the smoothness energy $E_{smooth}(f)$ are defined by

$$\begin{aligned} E_{data}(f) &= \sum_{p} D_p(f_p), \\ E_{smooth}(f) &= \sum_{p} \sum_{q \in \mathcal{N}(p)} V_{pq}(f_p, f_q), \end{aligned} \tag{2}$$

where $\mathcal{N}(p)$ is the neighborhood pixels of the pixel p. $D_p(f_p)$ is the data cost that encodes the penalty for the dissimilarity of two corresponding pixels $I_L(p)$ and $I_R(p+f_p)$ in the left image I_L and the right image I_R, respectively. $V_{pq}(f_p, f_q)$ is the smoothness cost that penalizes the discontinuity of disparities between neighboring pixels.

Mutual information can be used as a data cost by defining

$$E_{data}(f) = -MI(I_L, I_R, f), \tag{3}$$

since maximization of mutual information implies minimization of the data cost. Given a disparity map f, $MI(I_L, I_R, f)$ between the left and right images is defined by

$$MI(I_L, I_R, f) = H(I_L) + H(I_R) - H_f(I_L, I_R), \tag{4}$$

where $H(I)$ is the entropy of image I, and $H_f(I_L, I_R)$ is the joint entropy of images I_L and I_R. Entropy and joint entropy are defined by

$$\begin{aligned} H(I) &= -\int P(i)\log(P(i))di, \\ H_f(I_L, I_R) &= -\iint P_f(i_L, i_R)\log(P_f(i_L, i_R))di_L di_R, \end{aligned} \tag{5}$$

respectively. $P(i)$ is the marginal probability of intensity i. Joint probability $P_f(i_L, i_R)$ is computed by counting the number of corresponding pixels between intensity i_L in the left image and i_R in the right image by warping the disparity. $P_f(i_L, i_R)$ can be represented by

$$P_f(i_L, i_R) = \frac{1}{N}\sum_{p} T[(i_L, i_R) = (I_L(p), I_R(p + f_p))], \tag{6}$$

where $T[\cdot]$ is one if the argument is true, zero otherwise. N is the total number of corresponding pixels in the images.

Since (4) is the mutual information of the whole images, I_L and I_R, it is difficult to use it as a data cost in an energy minimization framework. To use it as a data cost, Kim *et al*. [9] approximated the whole mutual information $MI(I_L, I_R, f)$ as the sum of the pixel-wise mutual information $mi(\cdot, \cdot)$ as follows.

$$MI(I_L, I_R, f) \approx \sum_{p} mi(I_L(p), I_R(p + f_p)). \tag{7}$$

Hence, from (2), (3) and (7), the pixel-wise data cost becomes negative of the pixel-wise mutual information as

$$D_p(f_p) = -mi(I_L(p), I_R(p + f_p)). \tag{8}$$

To compute pixel-wise mutual information, Kim *et al.* assumed that $H(I_L)$ and $H(I_R)$ are nearly constant and only approximated the joint entropy $H_f(I_L, I_R)$ as follows.

$$H_f(I_L, I_R) \approx \sum_p h(I_L(p), I_R(p+f_p)), \tag{9}$$

where $h(\cdot,\cdot)$ is a pixel-wise joint entropy that is defined by

$$\begin{aligned} &h(I_L(p), I_R(p+f_p)) = \\ &-\frac{1}{N}\log\{P_{f_0}(I_L(p), I_R(p+f_p)) \otimes G(I_L(p), I_R(p+f_p))\} \\ &\quad \otimes G(I_L(p), I_R(p+f_p)), \end{aligned} \tag{10}$$

where $G(\cdot,\cdot)$ is a 2D Gaussian function and $P_{f_0}(\cdot,\cdot)$ is a joint probability distribution using the disparity map f_0 estimated from the previous iteration. Then the pixel-wise mutual information is represented by

$$mi(I_L(p), I_R(p+f_p)) = -h(I_L(p), I_R(p+f_p)) + \theta, \tag{11}$$

where θ is a constant value.

To take occlusion into account, Hirschmuller [6] modified the work of [9] by considering the marginal entropy term $H(I)$ explicitly which was assumed constant in [9]. $H(I)$ is also approximated as a sum of pixel-wise marginal entropy as follows.

$$H(I) \approx \sum_p h(I(p)), \tag{12}$$

where $h(\cdot)$ is a pixel-wise marginal entropy that is defined by

$$h(I(p)) = -\frac{1}{N}\log\{P(I(p)) \otimes G(I(p))\} \otimes G(I(p)), \tag{13}$$

where $G(\cdot)$ is 1D Gaussian function and $P(I(p))$ is computed by marginalizing the joint probability $P_f(\cdot,\cdot)$ as follows.

$$\begin{aligned} P_L(I_L(p)) &= \sum_i P_f(I_L(p), i), \\ P_R(I_R(p+f_p)) &= \sum_i P_f(i, I_R(p+f_p)). \end{aligned} \tag{14}$$

Consequently, pixel-wise mutual information is defined by

$$\begin{aligned} &mi(I_L(p), I_R(p+f_p)) \\ &= h(I_L(p)) + h(I_R(p+f_p)) - h(I_L(p), I_R(p+f_p)). \end{aligned} \tag{15}$$

There are two problems in the conventional MI-based methods described above. First, they cannot handle the local radiometric variations caused by light configuration change, because they collect correspondences in the joint probability assuming that there is a global transformation.

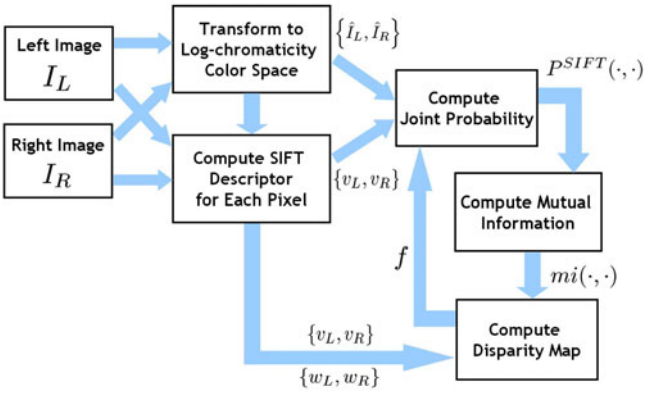


Figure 2. Overview of the proposed algorithm

Fig. 1 (a) and (b) show an example of a global variation owing to exposure change. In this case, conventional MI-based methods produce relatively good disparity map as shown in Fig. 1 (c). The corresponding joint probability (Fig. 1 (d)) clearly shows that there is a global nonlinear transformation between images.

However, when there exist local radiometric variations in images owing to the light configuration change as shown in Fig. 1 (e) and (f), the conventional MI-based methods fail to give reliable results as depicted in Fig. 1 (g). Note that as can be seen in Fig. 1 (h), in this case, the shape of the corresponding joint probability is very sparse, which means that the global relationship between the input images cannot be found. Second, they do not encode spatial information, and only gather corresponding intensities in the normalized joint probability given an estimated disparity map. If the disparity is not correct, the joint probability tends to be incorrect.

In this work, we propose a new algorithm that can cope with these two problems. Our approach is described in detail in the following section.

3. Proposed Algorithm

The whole procedure of the proposed algorithm is depicted in Fig 2. First, we transform the input color images to log-chromaticity color space in order to deal with local as well as global radiometric variations. In this color space, linear relationship between corresponding color values can be established. To incorporate spatial information in our framework, similarly to [12] we use the gradient information near the pixel. Instead of using only center pixel gradient information [12], we use a broader gradient histogram from the SIFT descriptor [11]. We compute the SIFT descriptor for both intensity and log-chromaticity color space. From the disparity map estimated in the previous iteration, we compute the joint probability. At this stage, to include the spatial gradient information, the joint probability is weighted by the distance of the SIFT descriptor. Mutual

information is computed by means of the joint probability. Finally the disparity map is computed combining the MI and the SIFT distance as our data cost in the MAP-MRF framework, and this disparity map is used to find the joint probability iteratively.

3.1. Transformation to log-chromaticity color space and discretization

As a first step, we transform the input color image I to the log-chromaticity color image $\hat{I}$ [5]. As in [5], the color model is defined by

$$I_k(p) \rightarrow \tilde{I}_k(p) = \rho(p) a_k (I_k(p))^{\gamma}, k \in \{R, G, B\}, \quad (16)$$

where we assume that the original k-channel image $I_k(p)$ at pixel p is changed to $\tilde{I}_k(p)$ by various unknown radiometric variations, and $\rho(p)$ is the brightness factor for the pixel p, a_k is the illuminant color factor, and γ is the gamma correction factor.

After transforming the observed input image $\tilde{I}_k(p)$ into the log-chromaticity color space, we can obtain $\hat{I}_k(p)$, which can be formulated as the following linear equation.

$$\hat{I}_k(p) = C_k + \gamma L_k(p), \quad (17)$$

where γ and C_k are constant for each channel k and $L_k(p)$ is an invariant color value for pixel p under radiometric variations. Log-chromaticity color space is used to establish a linear relationship between color values of input images that are affected by unknown radiometric variations. Fig. 3 shows some examples of the linear relationship in the joint probability in the log-chromaticity color space for the two images in Fig. 1 (a) and (b).

Since pixel values in the transformed image $\hat{I}_k$ have floating point numbers, we must remap those values to integer domain to make a joint probability. For that purpose, we multiply by scale factor s and round it off to get integer values. In our experiment, we set s as 1000. The final size of the joint probability matrix is determined by the maximum between the left maximum integer value $\hat{I}_{L,k,max}$ and the right maximum integer value $\hat{I}_{R,k,max}$. This process does not affect the linearity between the two log-chromaticity color images.

3.2. Joint probability using SIFT descriptor

A joint probability is computed at each channel independently by use of the estimated disparity map from the previous iteration. In this case, wrong disparity can induce an incorrect joint probability. To prevent this problem, we incorporate the spatial information in the joint probability computation step. For spatial information, we adopt the SIFT descriptor which is robust and accurately depicts local gradient information. The SIFT descriptor is computed for

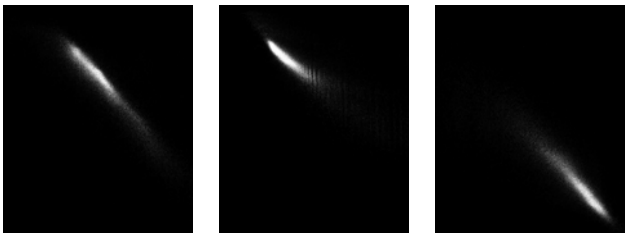


Figure 3. Linear relationship in joint probability. From left to right, R,G,B channel joint probability in log-chromaticity color space

every pixel in the log-chromaticity color space. Though the left image $\hat{I}_{L,k}$ and the right image $\hat{I}_{R,k}$ have different γ and C_k values, the SIFT descriptor distance between corresponding pixels is unaffected by this difference, since computing gradient eliminates C_k and the histogram normalization step normalizes γ during the SIFT descriptor computation step. Then, k-channel SIFT-weighted joint probability $P_k^{SIFT}(i_{L,k}, i_{R,k})$ is defined by

$$P_k^{SIFT}(i_{L,k}, i_{R,k}) = \frac{1}{Z} \sum_p u_{k,p}(f_p) \cdot T[(i_{L,k}, i_{R,k}) = (\hat{I}_{L,k}(p), \hat{I}_{R,k}(p + f_p))], \quad (18)$$

where Z is a normalization constant and $u_{k,p}(f_p)$ is defined by

$$u_{k,p}(f_p) = \exp(-\frac{\|v_{L,k}(p) - v_{R,k}(p + f_p)\|}{l}), \quad (19)$$

where $\|\cdot\|$ is the Euclidean distance, $v_{L,k}(p)$ and $v_{R,k}(p + f_p)$ are the SIFT descriptors for the pixel p in the left k-channel image and the pixel $p+f_p$ in the right k-channel image, respectively, in the log-chromaticity color space, and l is the SIFT descriptor size. Therefore, $P_k^{SIFT}(i_{L,k}, i_{R,k})$ is governed by the constraint that corresponding pixels should have similar gradient structures.

3.3. Disparity map estimation in MAP-MRF

From our SIFT-weighted joint probability in (18), the marginal probabilities for the left and right k-channel images can be computed using (14). Then the pixel-wise marginal entropies $h(\hat{I}_{L,k}(p))$, $h(\hat{I}_{R,k}(p + f_p))$ and the pixel-wise joint entropy $h(\hat{I}_{L,k}(p), \hat{I}_{R,k}(p + f_p))$ can be computed by (13) and (10), respectively. Consequently, the pixel-wise mutual information $mi_{k,p}(f_p)$ is computed from (15) for each k-channel as follows.

$$mi_{k,p}(f_p) = mi(\hat{I}_{L,k}(p), \hat{I}_{R,k}(p + f_p)) = h(\hat{I}_{L,k}(p)) + h(\hat{I}_{R,k}(p + f_p)) - h(\hat{I}_{L,k}(p), \hat{I}_{R,k}(p + f_p)). \quad (20)$$

To increase robustness, similarly to [15, 10], we also utilize the local descriptor such as SIFT in our data cost. Finally,

we define our data cost $D_p(f_p)$ as follows :

$$D_p(f_p) = \alpha_p(f_p) + \lambda \cdot \beta_p(f_p), \tag{21}$$

where λ is a weighting constant and $\alpha_p(f_p)$ and $\beta_p(f_p)$ are defined as follows :

$$\alpha_p(f_p) = \frac{1}{3}\sum_k \{-mi(\hat{I}_{L,k}(p), \hat{I}_{R,k}(p+f_p)) + \tau\}, \tag{22}$$

where τ is a constant value.

$$\beta_p(f_p) = \frac{1}{3}\sum_k \frac{|v_{L,k}(p) - v_{R,k}(p+f_p)|}{l} + \frac{|w_L(p) - w_R(p+f_p)|}{l}, \tag{23}$$

where $|\cdot|$ is the L1 norm, $v_{L,k}(p)$ and $v_{R,k}(p+f_p)$ are the SIFT descriptors for the pixel p in the left k-channel image and the pixel $p+f_p$ in the right k-channel image, respectively, in the log-chromaticity color space, $w_L(p)$ and $w_R(p+f_p)$ are the SIFT descriptors for the pixel p in the left gray (intensity) image and the pixel $p+f_p$ in the right gray (intensity) image, respectively, and l is the SIFT descriptor size.

Note that our data cost $D_p(f_p)$ consists of two terms; mutual information $\alpha_p(f_p)$ as an appearance measure and SIFT descriptor distance $\beta_p(f_p)$ as a geometric measure, and both are quite robust to linear transformation established in log-chromaticity color. At this stage, the SIFT descriptor distance is computed for both log-chromaticity color space and intensity space to fully utilize the color information.

For the smoothness cost, we use a truncated quadratic cost defined by

$$V_{pq}(f_p, f_q) = \mu \cdot min(|f_p - f_q|^2, V_{max}). \tag{24}$$

The total energy (1) is minimized by the Graph-cuts expansion algorithm [3].

4. Experimental Results

To evaluate our method for various radiometric variations, similarly to [7, 5], we used the four data sets with ground truth disparity maps including Aloe, Moebius, Dolls and Art [1]. Each data set has three different camera exposures and three different configurations of the light source [1, 7]. The exposure is indexed from 0 to 2, and index 0 indicates the shortest exposure (darkest), while index 2 is the longest exposure (brightest). Light source configuration is indexed from 1 to 3, and each index has different light configuration. For the input images, for example, "L, illum(1)-exp(0)" means that "the left image with the light configuration index 1 and the exposure index 0". We evaluated our method with other matching costs including BT with Rank transformation (Rank/BT) [17], BT with LoG filtering (LoG/BT), mutual information (MI) [6], Normalized Cross Correlation (NCC) and Adaptive Normalized Cross Correlation (ANCC) [5]. For all matching costs, energy was minimized by the Graph-cuts expansion algorithm [3]. The parameters of each method were tuned individually to produce optimal results using image without radiometric variation similarly to [7]. The parameters of our method are set constant as follows: The std. dev σ of the Gaussian function in (13) and (10) is 10, $\tau = 30$, $l = 4 \times 4 \times 8 = 128$, the window size of the SIFT descriptor is 9×9, $\lambda = 0.1$, $\mu = 1.1$ and $V_{max} = 5$. The total running time of our method for most images does not exceed 8 minutes. For example, for Aloe image (size : 427×370, disparity range : 0-70), it is about 6 minutes on a PC with PENTIUM-4 2.4GHz CPU.

4.1. MI vs. SIFT

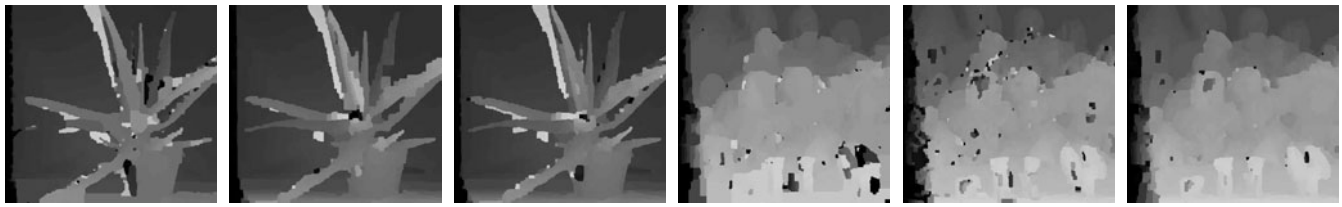
In Fig. 4, we compared the results using only MI ($\alpha_p(\cdot)$ term in (21)), using only SIFT ($\beta_p(\cdot)$ term in (21)) and using MI combined with SIFT ($D_p(\cdot)$ in (21)) in our framework. Note that MI and SIFT play complementary roles to each other, since MI is an appearance measure, while SIFT serves as a geometric measure. The MI term only suffers from the discretization error due to the floating-point values of log-chromaticity color during the construction of joint probability. In this case, SIFT helps to find correct matches and boosts accurate convergence of MI. On the other hand, SIFT term only is weak on textureless regions and blurs the boundaries. In these regions, pixel-wise MI helps it to find correct matches. Therefore, both terms are necessary to get more accurate and robust results.

4.2. Different exposures

To evaluate the effects of exposure changes, we only changed the index of exposure while fixing the index of the light configuration to 1. Fig. 5 (f)-(j) show the comparison of the performance of various matching costs for the input images in Fig. 5 (a)-(b). Similarly, Fig. 6 (f)-(j) are the results for input images in Fig. 6 (a)-(b) and Fig. 7 (f)-(j) are the results for input images in Fig. 7 (a)-(b). Fig. 8 (a)-(d) show the error ratio for the unoccluded region of each method for the left/right combination of exposure changes. Exposure changes cause a global transformation between input images. Most methods still show stable results to some extent, except for LoG/BT. Our method produces the most stable and accurate results against these global variations.

4.3. Different configurations of the light source

To evaluate the effects of the light configuration changes, we only changed the index of the light configuration while fixing the index of the exposure to 1. Fig. 5 (k)-(o) show the



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

Figure 4. (a)-(c) are results for Aloe stereo image (L, illum(1)-exp(1) and R, illum(3)-exp(1)) in Fig. 5 (c)-(d). (d)-(f) are results for Dolls stereo image (L, illum(1)-exp(1) and R, illum(3)-exp(1)) in Fig. 7 (c)-(d). (a) using only MI (error : 17.6 %). (b) using only SIFT (error : 11.97 %). (c) using MI combined with SIFT (error : 9.27 %). (d) using only MI (error : 26.45 %). (e) using only SIFT (error : 17.87 %). (f) using MI combined with SIFT (error : 11.83 %).

comparison of the performance of various matching costs for input images in Fig. 5 (c)-(d). Similarly, Fig. 6 (k)-(o) are the results for input images in Fig. 6 (c)-(d) and Fig. 7 (k)-(o) are the results for input images in Fig. 7 (c)-(d). Fig. 8 (e)-(h) show the error ratio for the unoccluded region of each method for the combination of the left/right light configuration changes. In general, different configurations of the light source cause different local variations in the input images, and this is a more difficult factor in establishing correspondences than the exposure change. MI shows very sensitive results to the local variation, since the global variation assumption is no longer valid. Note that the local correlation-based and filtering-based methods show better results than MI. Those methods, however, fail in extreme light-varying regions. Our method is still accurate and robust for these local variations.

5. Conclusion

In this paper, we propose a new stereo matching algorithm based on mutual information (MI) combined with SIFT descriptor. Our method utilizes MI as an appearance measure and SIFT as a geometric cue. By transforming original colors into the log-chromaticity color space, MI can establish robust and accurate correspondence irrespective of any radiometric variations. To impose spatial information, the SIFT descriptor is employed in both the joint probability construction and data cost computation stages. Experimental results demonstrate that our proposed algorithm is quite robust and accurate to local as well as global radiometric variations.

Acknowledgements

This research was supported in part by the Defense Acquisition Program Administration and Agency for Defense Development, Korea, through IIRC (UD070007AD), and in part by the ITRC program of MKE/IITA through 3DRC (IITA-2008-C1090-0801-0018), Korea.

References

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

[2] http://www.photomodeler.com/. 1

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

[4] G. Egnal. Mutual information as a stereo correspondence measure. *Technical Report MS-CIS-00-20 University of Pennsylvania*, 2000. 1

[5] Y. S. Heo, K. M. Lee, and S. U. Lee. Illumination and camera invariant stereo matching. In *Proc. of CVPR*, 2008. 1, 4, 5

[6] H. Hirschmuller. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. PAMI*, 30(2):328–341, 2008. 1, 3, 5

[7] H. Hirschmuller and D. Scharstein. Evaluation of cost functions for stereo matching. In *Proc. of CVPR*, 2007. 1, 5

[8] H.-H. Jeon, A. Basso, and P. F. Driessen. A global correspondence for scale invariant matching using mutual information and the graph search. In *Proc. of ICME*, 2006. 2

[9] J. Kim, V. Kolmogorov, and R. Zabih. Visual correspondence using energy minimization and mutual information. In *Proc. of ICCV*, 2003. 1, 2, 3

[10] C. Liu, J. Yuen, A. Torralba, J. Sivic, and W. Freeman. Sift flow: Dense correspondence across different scenes. In *Proc. of ECCV*, 2008. 4

[11] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004. 2, 3

[12] J. P.W.Pluim, J. A. Maintz, and M. A. Viergever. Image registration by maximization of combined mutual information and gradient information. *IEEE Trans. Medical Imaging*, 19(8):809–814, 2000. 1, 2, 3

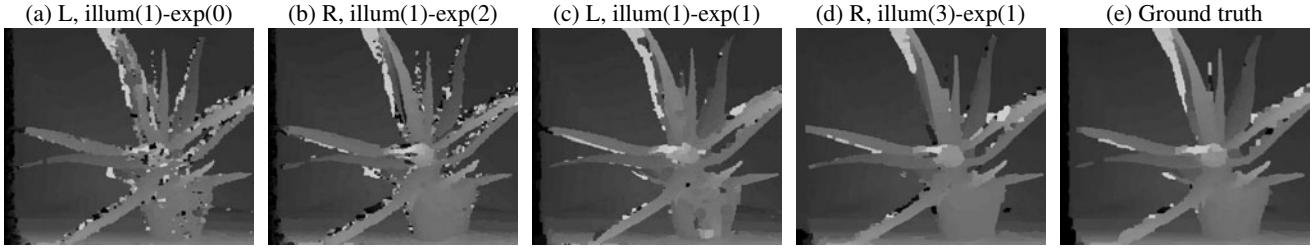
[13] D. B. Russakoff, C. Tomasi, T. Rohlfing, and C. R. Maurer, Jr. Image similarity using mutual information of regions. In *Proc. of ECCV*, 2004. 1, 2

[14] N. Snavely, S. M. Seitz, and R. Szeliski. Modeling the world from internet photo collections. *IJCV*, 80(2):189–210, 2008. 1

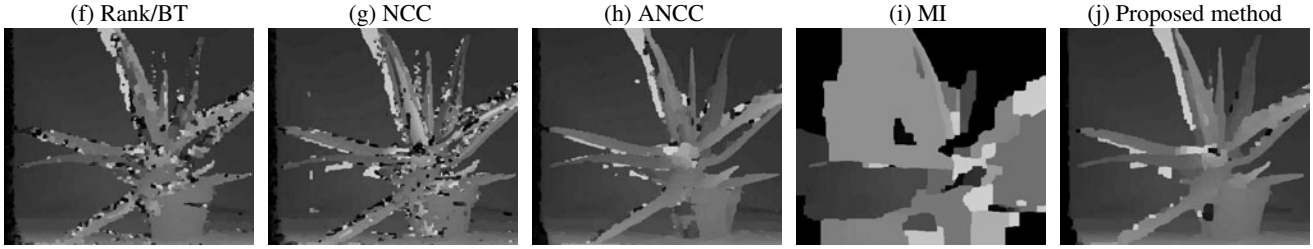
[15] E. Tola, V. Lepetit, and P. Fua. A fast local descriptor for dense matching. In *Proc. of CVPR*, 2008. 4

[16] P. Viola and W. M. Wells III. Alignment by maximization of mutual information. *IJCV*, 24(2):137–154, 1997. 1

(a) L, illum(1)-exp(0) (b) R, illum(1)-exp(2) (c) L, illum(1)-exp(1) (d) R, illum(3)-exp(1) (e) Ground truth



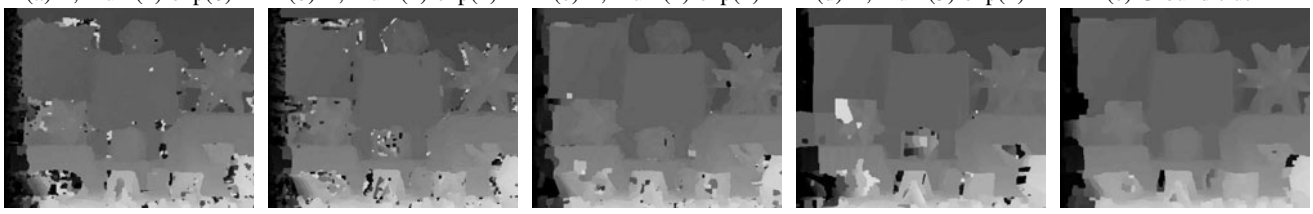
(f) Rank/BT (g) NCC (h) ANCC (i) MI (j) Proposed method



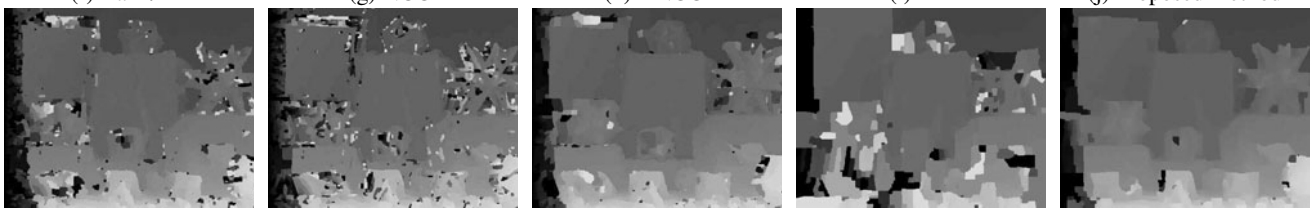
(k) Rank/BT (l) NCC (m) ANCC (n) MI (o) Proposed method

Figure 5. Results of test stereo matching costs on Aloe image pair with varying camera exposure and light configurations. (f) - (j) are the results from the image pair (a) and (b). (k) - (o) are the results from the image pair (c) and (d). (e) is the ground truth disparity map.

(a) L, illum(1)-exp(0) (b) R, illum(1)-exp(2) (c) L, illum(1)-exp(1) (d) R, illum(3)-exp(1) (e) Ground truth

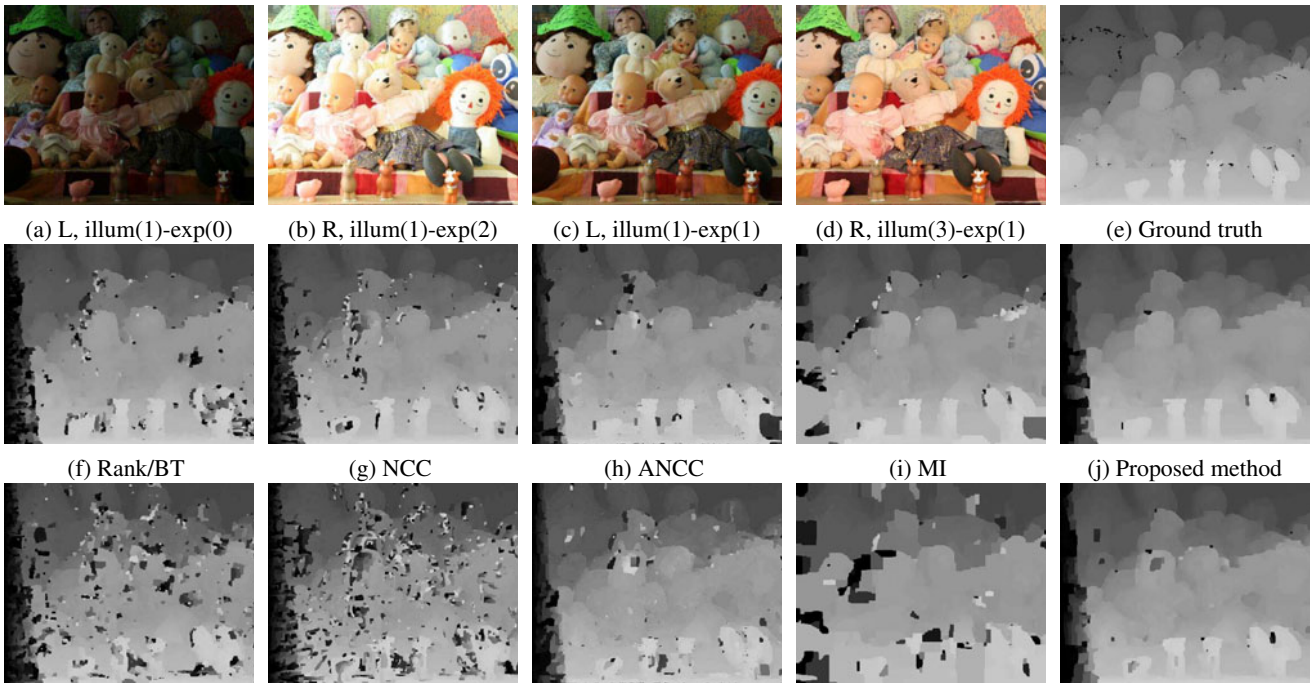


(f) Rank/BT (g) NCC (h) ANCC (i) MI (j) Proposed method



(k) Rank/BT (l) NCC (m) ANCC (n) MI (o) Proposed method

Figure 6. Results of test stereo matching costs on Moebius image pair with varying camera exposure and light configurations. (f) - (j) are the results from the image pair (a) and (b). (k) - (o) are the results from the image pair (c) and (d). (e) is the ground truth disparity map.

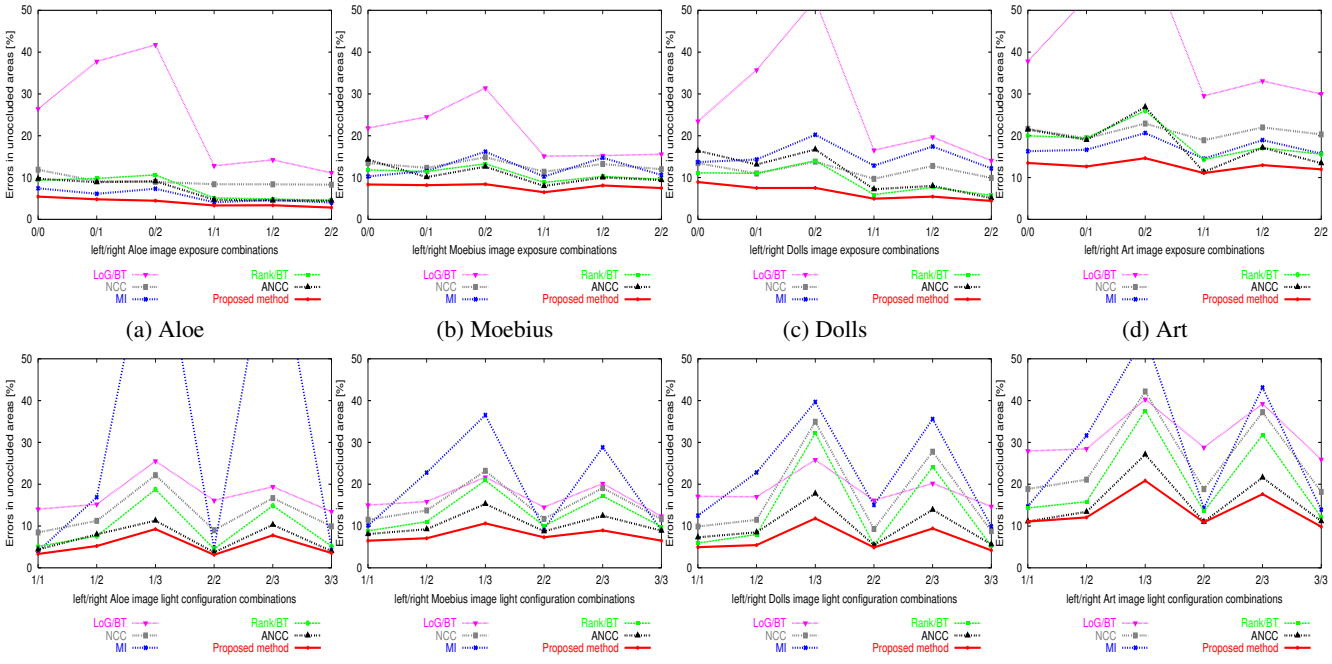


(a) L, illum(1)-exp(0) (b) R, illum(1)-exp(2) (c) L, illum(1)-exp(1) (d) R, illum(3)-exp(1) (e) Ground truth

(f) Rank/BT (g) NCC (h) ANCC (i) MI (j) Proposed method

(k) Rank/BT (l) NCC (m) ANCC (n) MI (o) Proposed method

Figure 7. Results of test stereo matching costs on Dolls image pair with varying camera exposure and light configurations. (f) - (j) are the results from the image pair (a) and (b). (k) - (o) are the results from the image pair (c) and (d). (e) is the ground truth disparity map.



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

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

Figure 8. Quantitative comparisons for exposure and light configuration changes. (a)-(d) are the comparisons for exposure changes. (e)-(h) are the comparisons for light configuration changes.

[17] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. In *Proc. of ECCV*, 1994. 5

[18] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High-quality video view interpolation using a layered representation. In *Proc. of SIGGRAPH*, 2004. 1