

Mahalanobis Distance Cross-Correlation for Illumination-Invariant Stereo Matching

Seungryong Kim, *Student Member, IEEE*, Bumsub Ham, *Member, IEEE*, Bongjoe Kim, *Student Member, IEEE*, and Kwanghoon Sohn, *Senior Member, IEEE*

Abstract—A robust similarity measure called the Mahalanobis distance cross-correlation (MDCC) is proposed for illumination-invariant stereo matching, which uses a local color distribution within support windows. It is shown that the Mahalanobis distance between the color itself and the average color is preserved under affine transformation. The MDCC converts pixels within each support window into the Mahalanobis distance transform (MDT) space. The similarity between MDT pairs is then computed using the cross-correlation with an asymmetric weight function based on the Mahalanobis distance. The MDCC considers correlation on cross-color channels, thus providing robustness to affine illumination variation. Experimental results show that the MDCC outperforms state-of-the-art similarity measures in terms of stereo matching for image pairs taken under different illumination conditions.

Index Terms—Exposure, illumination, Mahalanobis distance, similarity measure, stereo matching.

I. INTRODUCTION

STEREO matching is a fundamental problem for many computer vision tasks such as view synthesis, autonomous navigation, and 3-D reconstruction [1]. It aims to extract 3-D scene information by finding correspondences between stereo pairs taken at different viewpoints of the same scene. Current state-of-the-art methods provide satisfactory results under the color consistency condition, that is, corresponding pixels have similar color distribution [2]. However, the color consistency assumption is often violated due to various factors including illumination source variations, non-Lambertian surfaces, vignetting, device characteristics, and image noise, resulting in a performance degradation [3].

To alleviate this problem, a number of matching methods have been proposed [3]. The seminal works proposed normalized correlation approaches [4], [5], in which each support window is normalized to reduce the effects of illumination, and the correlation similarity is computed between them. These approaches do not consider the correlation on

cross-color channels and thus, they show limited performance for stereo pairs taken under severe illumination variations. To this end, a robust similarity measure called the Mahalanobis distance cross-correlation (MDCC) is proposed. The MDCC converts pixels within each support window into the Mahalanobis distance transform (MDT) space. Then, it computes the correlation similarity between MDT vector pairs. The Mahalanobis distance exploits vectorial color information from the covariance of color channels. This means that the MDCC considers correlation similarity on each color channel and cross-color channel simultaneously, thus providing robustness for illumination variations. Furthermore, to encode spatial information even under illumination variation, the correlation similarities are aggregated in both support windows with asymmetric weight distributions based on the Mahalanobis distance. In stereo matching frameworks, the MDCC estimates more accurate disparity maps for stereo pairs taken under different illumination conditions than state-of-the-art similarity measures and even provides competitive performance under exposure changes.

A. Related Works

The MDCC aims to estimate an accurate disparity map for stereo pairs even if they are taken under different illumination conditions, and it incorporates the invariance of the Mahalanobis distance for these conditions. This section describes related works on the Mahalanobis distance and illumination robust stereo matching focusing on similarity measures.

1) *Mahalanobis Distance*: The Mahalanobis distance has been widely used as a distance measure in many computer vision problems. It measures a relative distance by taking into account the statistical characteristic of the distribution. In clustering or classification problems, the methods for distance learning have popularly focused on the Mahalanobis distance. Craw *et al.* [6] developed the Mahalanobis distance-based classifier for facial recognition to deal with the variation within samples. Habili *et al.* [7] used the Mahalanobis distance to account for density variations in color distribution for a skin-color segmentation. Weinberger and Saul [8] developed a learning method of the Mahalanobis distance for k-nearest neighbor (kNN) classification to capitalize on statistical regularity within samples. In a similar way, Xiang *et al.* [9] advocated the Mahalanobis distance in data clustering and segmentation, as it can adjust the geometrical distribution.

The Mahalanobis distance has been successfully used to establish the correspondence between feature descriptors.

Manuscript received October 21, 2013; revised February 3, 2014 and April 22, 2014; accepted June 3, 2014. Date of publication June 5, 2014; date of current version October 29, 2014. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (NRF-2013R1A2A2A01068338). This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (NIPA-2014-H0301-14-1012) supervised by the NIPA (National IT Industry Promotion Agency). This paper was recommended by Associate Editor L. Onural

The authors are with the School of Electrical and Electronic Engineering, Yonsei University, Seoul 120-749, Korea (e-mail: srkim89@yonsei.ac.kr; mimo@yonsei.ac.kr; bj1719@yonsei.ac.kr; khsohn@yonsei.ac.kr).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCSVT.2014.2329377

TABLE I
COMPARISON OF ILLUMINATION INSENSITIVE SIMILARITY MEASURES

	MI	Census	NCC	ANCC	MDCC
Type	Non-Parametric	Non-Parametric	Parametric	Parametric	Parametric
Color	Single Color	Single Color	Single Color	Single Color	Multi-Color
Spatial Structure	Non-Encoding	Encoding	Non-Encoding	Encoding	Encoding
Illumination Robustness	Entropy of Joint Probability	Intensity Ordering	Normalization	Normalization with Weight	MDT Transform
Aggregation Weight	Uniform	Uniform	Uniform	Euclidean Distance Scheme	Mahalanobis Distance Scheme

Tuytelaars and Gool [10] used it for wide baseline stereo matching. Schaffalitzky and Zisserman [11] exploited affine invariant properties of the Mahalanobis distance to measure correspondence between descriptors. Schmid and Mohr [12] measured the similarity between feature descriptors using the Mahalanobis distance. In addition, Platel *et al.* [13] exploited it to measure similarity in feature space. It is worth noting that, unlike these methods that use the Mahalanobis distance as the similarity distance between feature descriptors, the MDCC leverages it to convert each matching vector into the illumination-invariant space.

2) *Illumination Robust Stereo Matching*: The methods for illumination-invariant stereo matching can be classified into global and local approaches according to how the effect of illumination variation is handled. Global approaches aim to reduce the effect of illumination variation over the whole image. With an assumption that the illumination variation can be modeled by constant factors on each color channel, a gray world transform divides each color channel by its average value within local windows [14]. One alternative is the normalized chromaticity (NC) transform that normalizes the responses of each color channel by the overall summation [15]. However, these methods cannot simultaneously reduce the dependency on illumination-color and illumination-geometry, because the normalization cannot eliminate illumination-dependent variables [16]. Finlayson *et al.* [17] used the histogram equalization (HE) transform to provide an image representation invariant to the illumination with an assumption that the rank order is preserved under varying illumination conditions. Ogale and Aloimonos [18] proposed a contrast invariant stereo matching method by relying on multiple spatial frequency channels. Although these global approaches are suitable for global illumination change, illumination invariance is not guaranteed for local illumination change.

Local approaches aim to reduce the effect of illumination variation on each local window, and they can be classified into parametric and nonparametric methods. Parametric methods use raw color information and leverage image transforms including the Laplacian of Gaussian (LoG) transform [19], background subtraction by the bilateral (BilSub) transform [20], and normalized correlation including normalized cross-correlation (NCC) [4] and adaptive NCC (ANCC) [5]. To remove a local offset, the LoG transform constructs a residual image, which is the difference between an original image and the second-order derivative image [19].

In a similar way, the BilSub transform builds a residual image by subtracting the bilateral filtered image from an original image to remove the local offset without blurring [20]. These transform-based methods fail under severe illumination variation because they only compensate for a constant offset. The NCC normalizes each support window to have a zero mean and unit standard deviation and computes the degree of cross-correlation (CC) between normalized windows, which is robust to linear illumination variation [4]. However, this method generates blurred results at object boundaries due to an ignorance of the spatial structure. To overcome this fattening effect, Heo *et al.* [5] proposed the ANCC, which measures correlation similarity by penalizing different positions and color values. The normalized correlation methods such as NCC and ANCC provide relatively high performance compared with image transform-based methods. However, they compute correlation similarity in each color channel independently, thus ignoring correlation similarity on cross-color channels.

In contrast to parametric methods, nonparametric methods use the local order of intensities or a statistical property instead of a raw intensity value. The rank transform replaces the intensity of a center pixel with the number of pixels within a support window whose intensity is less than that of the center pixel [21]. With the rank constraint, Wang *et al.* [22] proposed the light transport constancy (LTC) method to provide the invariance to scene reflectance for non-Lambertian surfaces. The Census transform produces a bit string for the support window based on intensity comparison in such a way that a bit 1 is allocated if the intensity of the center pixel is larger than that of the neighbor pixel; otherwise, a bit 0 is allocated [23]. Although these ordering-based methods are tolerant to local illumination variation, they produce unsatisfactory performance on homogeneous or noisy regions, where the local order of intensities is indistinct [23]. Mutual information (MI) leverages the entropy of the joint probability distribution function (PDF) to measure the similarity between matching windows [24]. Kim *et al.* [25] used it as a pixelwise data cost in the maximum a posteriori on markov random field framework. However, the performance of the MI-based method largely depends on the size of the matching window, which regulates the distinctiveness of the statistical power. In addition, the method is sensitive to local variations since it is assumed that there exists a global transformation. Table I shows the summary of similarity measures insensitive to

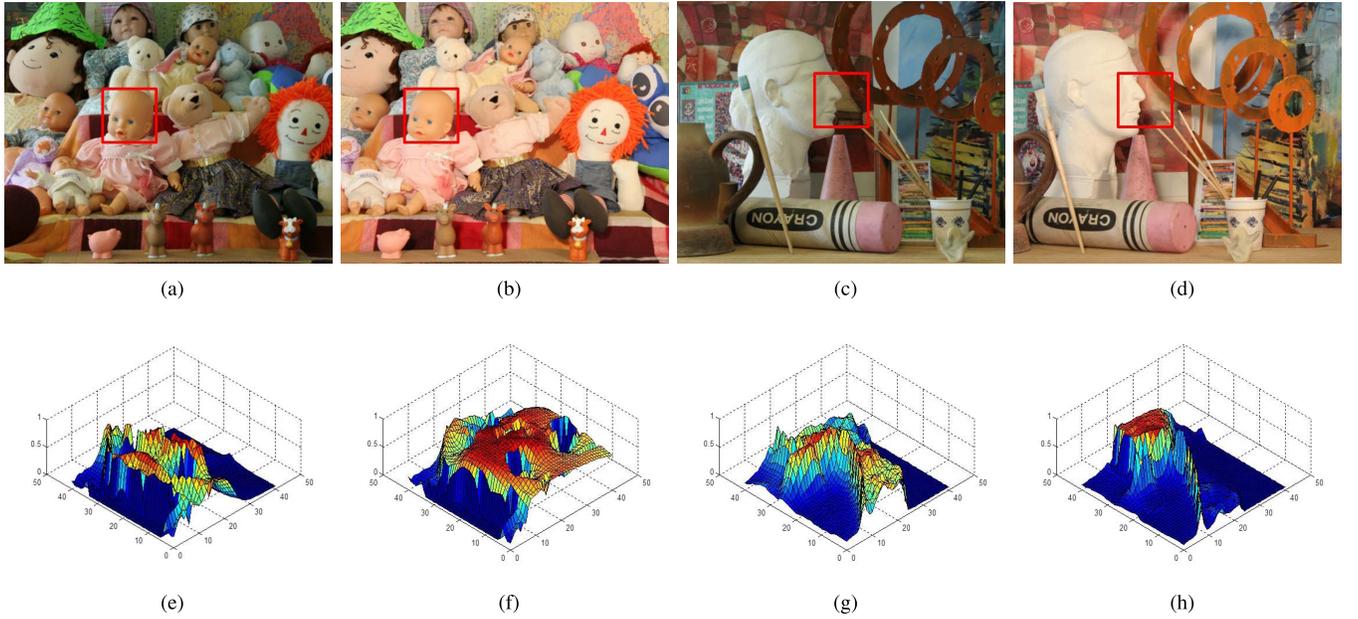


Fig. 1. Degradation of conventional weight distribution under affine illumination variations. (a) *Dolls* left image. (b) *Dolls* right image. (c) *Art* left image. (d) *Art* right image. (e)–(h) Corresponding weight distributions for indicated support windows in (a)–(d). The conventional weight distribution cannot be preserved under affine illumination variation since it is based on the Euclidean distance.

illumination variations.

B. Contribution

The contributions of this paper are as follows. First, it is shown that normalized correlation methods such as NCC and ANCC provide limited performance for stereo pairs that undergo affine illumination variation, and those limitations can be alleviated by the proposed similarity measure according to the invariance of the Mahalanobis distance. Second, the proposed similarity measure is designed so that it simultaneously considers correlation similarity on each color and cross-color channel, thus providing the illumination invariance and improving the matching performance. Finally, instead of a conventional weight distribution based on the Euclidean distance, a weight distribution based on the Mahalanobis distance is proposed to encode spatial information invariant to illumination variation.

The remainder of this paper is organized as follows. Section II introduces the affine illumination change model and the limitations of conventional normalized correlation methods. Section III describes the MDCC similarity measure for illumination-invariant stereo matching. Experimental results for stereo matching are given in Section IV. Finally, conclusion and suggestions for future works are given in Section V.

II. PROBLEM STATEMENT

A. Affine Transform for Illumination Change

Let us consider image pairs taken under different illumination conditions. Suppose the primary image \mathbf{I}_1 is taken under a reference illumination condition, and it is mapped to the secondary image \mathbf{I}_2 taken under an unknown illumination condition, where $\mathbf{I} = [I^R, I^G, I^B]^T$. Let $\Psi_1(\mathbf{p})$ and $\Psi_2(\hat{\mathbf{p}})$

be support windows centered at pixel $\mathbf{p} \in \mathbb{N}^2$ in \mathbf{I}_1 and corresponding pixel $\hat{\mathbf{p}}$ in \mathbf{I}_2 , respectively. The vector form for color values of pixels within each support window is defined as $\Omega_1(\mathbf{p}) = \otimes_{\mathbf{q} \in \Psi_1(\mathbf{p})} \mathbf{I}_1(\mathbf{q})^T$ and $\Omega_2(\hat{\mathbf{p}}) = \otimes_{\hat{\mathbf{q}} \in \Psi_2(\hat{\mathbf{p}})} \mathbf{I}_2(\hat{\mathbf{q}})^T$, where \otimes is the operator for the vector form.

Generally, the illumination variation of image pairs can be modeled as an affine transformation between color distributions of image pairs. The affine model for illumination change is widely used, as it deals with a wide range of imaging conditions [26]. This affine model accounts for general lighting environments such as illumination source changes, illumination direction changes, and ambient light [27]. Furthermore, it involves different elements of the color deformation such as noises and device artifacts [28]. The affine illumination variation between $\mathbf{I}_1(\mathbf{p})$ and $\mathbf{I}_2(\hat{\mathbf{p}})$ can be modeled as

$$\mathbf{I}_2(\hat{\mathbf{p}}) = \mathbf{A}_p \mathbf{I}_1(\mathbf{p}) + \mathbf{b}_p \quad (1)$$

where \mathbf{A}_p is a 3×3 full matrix for linear transformation, and \mathbf{b}_p is a translation vector.

B. Normalized Correlation Similarity Measure in Stereo Matching

To estimate the correspondence between image pairs under illumination variation, a number of methods have been proposed, and normalized correlation methods such as NCC and ANCC have shown satisfactory results. The NCC on ξ color channel ($\xi \in \Pi = \{R, G, B\}$) [4] is defined as

$$\text{NCC}_\xi(\mathbf{p}, \hat{\mathbf{p}}) = \left\langle \frac{\Omega_1^\xi(\mathbf{p}) - \bar{\Omega}_1^\xi(\mathbf{p})}{\|\Omega_1^\xi(\mathbf{p}) - \bar{\Omega}_1^\xi(\mathbf{p})\|}, \frac{\Omega_2^\xi(\hat{\mathbf{p}}) - \bar{\Omega}_2^\xi(\hat{\mathbf{p}})}{\|\Omega_2^\xi(\hat{\mathbf{p}}) - \bar{\Omega}_2^\xi(\hat{\mathbf{p}})\|} \right\rangle \quad (2)$$

where $\langle \cdot, \cdot \rangle$ is the CC operator, and $\|\cdot\|$ is the L-2 norm. $\bar{\Omega}_2^\xi(\hat{\mathbf{p}})$ is the average value of $\Omega_2^\xi(\hat{\mathbf{p}})$. Although the NCC shows

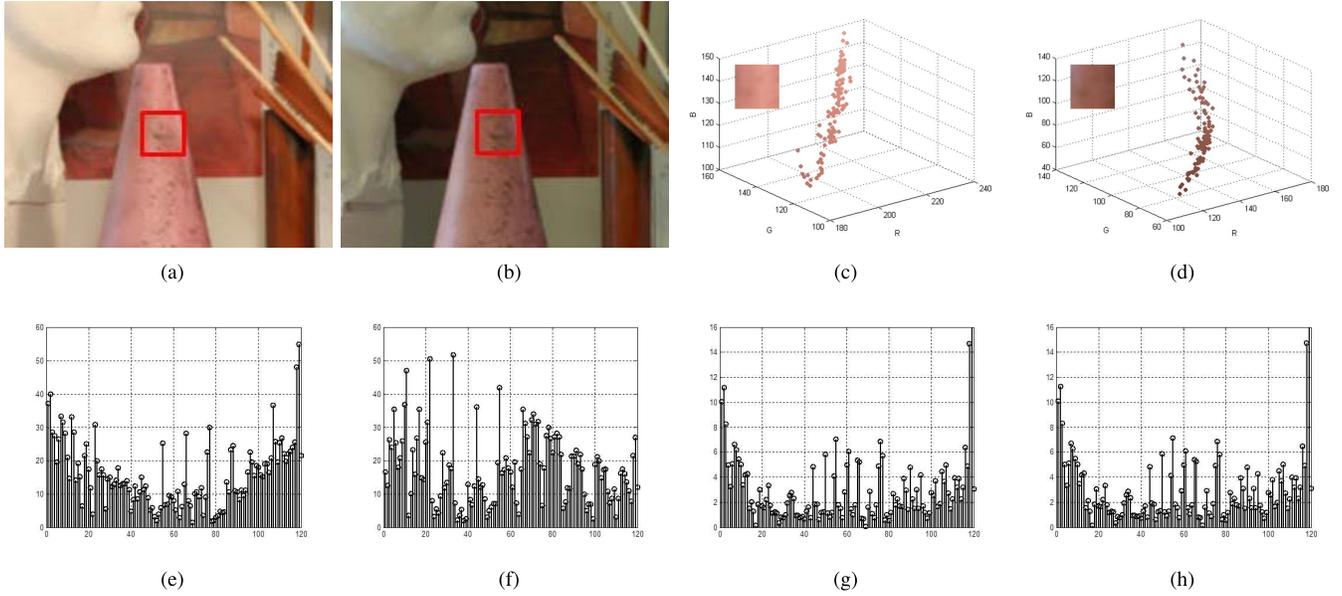


Fig. 2. Invariance of the Mahalanobis distance between color distributions of pixels within support windows and the mean color in color space under affine illumination variations. (a) and (b) *Art* image pair under affine illumination variation. (c) and (d) Color distributions for indicated support windows in (a) and (b). (e) and (f) Euclidean distance for (c) and (d). (g) and (h) Mahalanobis distance for (c) and (d). The Mahalanobis distance is preserved under affine illumination variations when compared with the Euclidean distance.

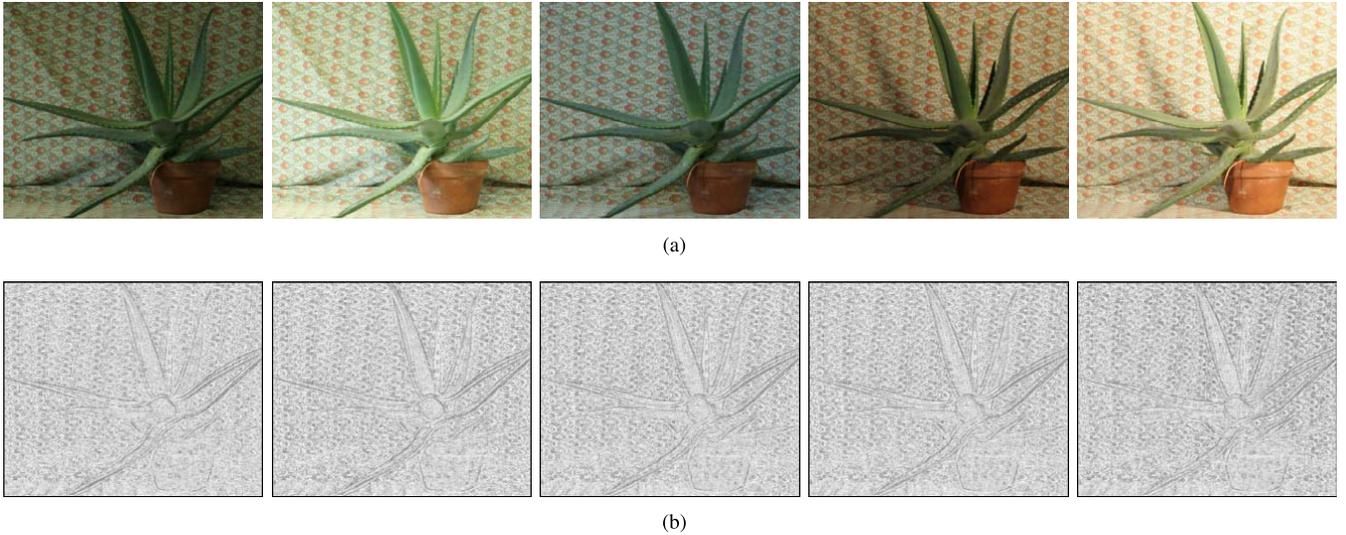


Fig. 3. Comparison of the images captured with varying illumination conditions with the MDT visualization for each image. (a) *Aloe* images captured with varying illumination conditions. (b) MDT visualization for each *Aloe* image in (a). The visualization represents the magnitude of the MDT vector.

robustness to linear variation, it produces a fattening effect. To reduce this effect, the ANCC utilizes weight distributions similar to the adaptive support-weight approach [29], defined as

$$\omega(\mathbf{q}, \mathbf{p}) = \exp\left(-\frac{\|\mathbf{q} - \mathbf{p}\|^2}{\gamma_g}\right) \exp\left(-\frac{\|\mathbf{I}(\mathbf{q}) - \mathbf{I}(\mathbf{p})\|^2}{\gamma_c}\right) \quad (3)$$

where $\omega(\mathbf{q}, \mathbf{p})$ is the affinity between pixel $\mathbf{q} \in \Psi(\mathbf{p})$ and the center pixel \mathbf{p} , computed according to the geometrical proximity and color similarity. γ_g and γ_c are used to normalize spatial and color distances, respectively. The ANCC on ξ color channel [5] is then defined as

$$\text{ANCC}_\xi(\mathbf{p}, \hat{\mathbf{p}})$$

$$= \left\langle \frac{\mathbf{W}_1(\mathbf{p})(\Omega_1^\xi(\mathbf{p}) - R_1^\xi(\mathbf{p}))}{\|\mathbf{W}_1(\mathbf{p})(\Omega_1^\xi(\mathbf{p}) - R_1^\xi(\mathbf{p}))\|}, \frac{\mathbf{W}_2(\hat{\mathbf{p}})(\Omega_2^\xi(\hat{\mathbf{p}}) - R_2^\xi(\hat{\mathbf{p}}))}{\|\mathbf{W}_2(\hat{\mathbf{p}})(\Omega_2^\xi(\hat{\mathbf{p}}) - R_2^\xi(\hat{\mathbf{p}}))\|} \right\rangle \quad (4)$$

where $\mathbf{W}(\mathbf{p})$ is a diagonal matrix, whose diagonal elements are the weight distribution $\omega(\mathbf{q}, \mathbf{p})$, and $R^\xi(\mathbf{p}) = \mathbf{W}(\mathbf{p})\Omega^\xi(\mathbf{p})/\|\mathbf{W}(\mathbf{p})\|_F$, where $\|\cdot\|_F$ is the Frobenius norm.

In case of color images, these methods measure the correlation similarity independently on each color channel as

$$\text{NCC}(\mathbf{p}, \hat{\mathbf{p}}) = \sum_{\xi \in \Pi} \text{NCC}_\xi(\mathbf{p}, \hat{\mathbf{p}}). \quad (5)$$

Similarly

$$\text{ANCC}(\mathbf{p}, \hat{\mathbf{p}}) = \sum_{\xi \in \Pi} \text{ANCC}_\xi(\mathbf{p}, \hat{\mathbf{p}}). \quad (6)$$

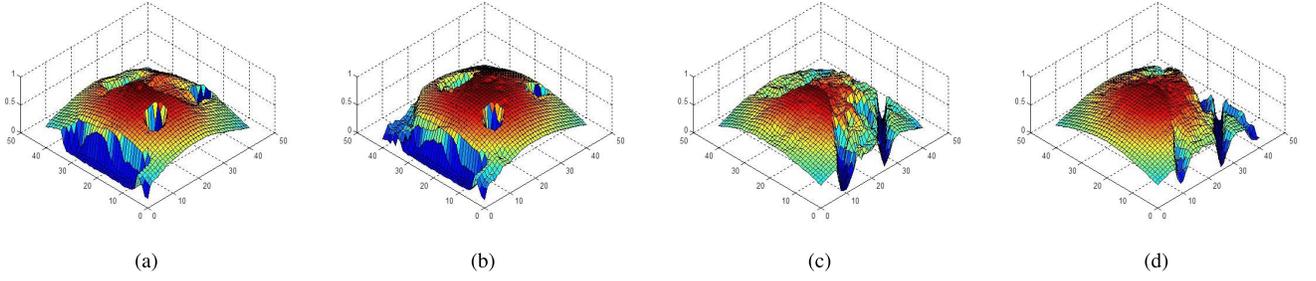


Fig. 4. Invariance of proposed weight distributions for support windows under illumination variations as shown in Fig. 1. (a) and (b) Proposed weight distribution for *Dolls* image pairs. (c) and (d) Proposed weight distribution for *Art* image pairs. Proposed weight distributions are preserved even under severe illumination variations.

For color image pairs under illumination variations, the limitations of normalized correlation methods are twofold. These methods do not consider the correlation similarity on cross-color channels. In stereo matching, color information reduces the ambiguity between similar intensity pixels and improves the distinction in local [30] and global methods [31]. However, under illumination variation, color information in these methods cannot contribute to improve the performance [3], [32]. The main reason is that the affine illumination variation causes color channels to influence each other as in (1), and thus, each color channel can be degraded by cross-color channels. That is, to estimate similarity between color image pairs under the affine variation, the similarity on cross-color channels should also be considered.

The weight distribution in ANCC does not guarantee accurate encoding of spatial information, thus degrading the performance and producing outliers. In ANCC, correlation similarities are aggregated with weight distributions on the basis of the Euclidean distance in both support windows as in (3). It is assumed that the pixel similar to a center pixel has a similar disparity value to that of the center pixel [5]. Thus, weight distributions of each support window should be preserved to provide an edge-preserving property. However, the affine illumination variation induces nonlinear transformation between weight distributions of each support window. Fig. 1 shows the degradation of conventional weight distribution under illumination variation.

III. STEREO MATCHING WITH THE MDCC SIMILARITY MEASURE

This section presents the MDCC similarity measure for illumination-invariant stereo matching. It measures the similarity using CC in the MDT space, achieving robustness to illumination variation between stereo pairs.

A. Mahalanobis Distance Transform Space

Let us assume that pixels within local support windows are degraded by identical affine illumination variation. Under these variations, the statistical characteristics between support windows $\Psi_1(\mathbf{p})$ and $\Psi_2(\hat{\mathbf{p}})$ are closely related. Denote $\Sigma_1(\mathbf{p})$ and $\Sigma_2(\hat{\mathbf{p}})$ as the covariance matrices for $\Omega_1(\mathbf{p})$ and $\Omega_2(\hat{\mathbf{p}})$, respectively. It can be shown that these corresponding covariance matrices are related as

$$\Sigma_2(\hat{\mathbf{p}}) = \mathbf{A}_p \Sigma_1(\mathbf{p}) \mathbf{A}_p^T. \quad (7)$$

In other words, the covariance matrices $\Sigma_1(\mathbf{p})$ and $\Sigma_2(\hat{\mathbf{p}})$ are similar, and the matrix \mathbf{A}_p is the similarity transformation [33].

Furthermore, the average color values computed on each color channel are related by

$$\bar{\Omega}_2(\hat{\mathbf{p}}) = \mathbf{A}_p \bar{\Omega}_1(\mathbf{p}) + \mathbf{b}_p \quad (8)$$

where $\bar{\Omega}_1(\mathbf{p})$ and $\bar{\Omega}_2(\hat{\mathbf{p}})$ are average color values of $\Omega_1(\mathbf{p})$ and $\Omega_2(\hat{\mathbf{p}})$, respectively, where $\bar{\Omega}(\mathbf{p}) = [\bar{\Omega}^R(\mathbf{p}), \bar{\Omega}^G(\mathbf{p}), \bar{\Omega}^B(\mathbf{p})]^T$. Using the statistical relationships in (7) and (8), the Mahalanobis distance between a color itself and the average color is invariant to illumination variation, as in Proposition 1.

Proposition 1: The Mahalanobis distance between $\mathbf{I}(\mathbf{q})$ for $\mathbf{q} \in \Psi(\mathbf{p})$ and $\bar{\Omega}(\mathbf{p})$ is preserved under affine illumination variation.

Proof: The Mahalanobis distance $d_M(\cdot, \cdot)$ between $\mathbf{I}_1(\mathbf{q})$ and $\bar{\Omega}_1(\mathbf{p})$ is defined as

$$d_M(\mathbf{I}_1(\mathbf{q}), \bar{\Omega}_1(\mathbf{p})) = (\mathbf{I}_1(\mathbf{q}) - \bar{\Omega}_1(\mathbf{p}))^T \Sigma_1^{-1}(\mathbf{p}) (\mathbf{I}_1(\mathbf{q}) - \bar{\Omega}_1(\mathbf{p})). \quad (9)$$

Similarly

$$d_M(\mathbf{I}_2(\hat{\mathbf{q}}), \bar{\Omega}_2(\hat{\mathbf{p}})) = (\mathbf{I}_2(\hat{\mathbf{q}}) - \bar{\Omega}_2(\hat{\mathbf{p}}))^T \Sigma_2^{-1}(\hat{\mathbf{p}}) (\mathbf{I}_2(\hat{\mathbf{q}}) - \bar{\Omega}_2(\hat{\mathbf{p}})). \quad (10)$$

By substituting (1), (7), and (8) into (10)

$$\begin{aligned} d_M(\mathbf{I}_2(\hat{\mathbf{q}}), \bar{\Omega}_2(\hat{\mathbf{p}})) &= (\mathbf{A}_p \mathbf{I}_1(\mathbf{q}) + \mathbf{b}_p - \mathbf{A}_p \bar{\Omega}_1(\mathbf{p}) - \mathbf{b}_p)^T \\ &\quad \cdot (\mathbf{A}_p \Sigma_1(\mathbf{p}) \mathbf{A}_p^T)^{-1} (\mathbf{A}_p \mathbf{I}_1(\mathbf{q}) \\ &\quad + \mathbf{b}_p - \mathbf{A}_p \bar{\Omega}_1(\mathbf{p}) - \mathbf{b}_p) \\ &= (\mathbf{I}_1(\mathbf{q}) - \bar{\Omega}_1(\mathbf{p}))^T \Sigma_1^{-1}(\mathbf{p}) (\mathbf{I}_1(\mathbf{q}) - \bar{\Omega}_1(\mathbf{p})) \\ &= d_M(\mathbf{I}_1(\mathbf{q}), \bar{\Omega}_1(\mathbf{p})). \end{aligned} \quad (11)$$

This shows that the Mahalanobis distance for pixels within corresponding support window $d_M(\mathbf{I}_2(\hat{\mathbf{q}}), \bar{\Omega}_2(\hat{\mathbf{p}}))$ is equivalent to that of reference support window $d_M(\mathbf{I}_1(\mathbf{q}), \bar{\Omega}_1(\mathbf{p}))$. ■

As shown in Fig. 2, the Mahalanobis distance is invariant to affine illumination variation between support windows, even if the color distributions of pixels within each support window are transformed. This implies that, for estimating the similarity between support windows which undergo illumination variation, it is helpful to convert pixels within support windows into its Mahalanobis distance to compensate for illumination variation.

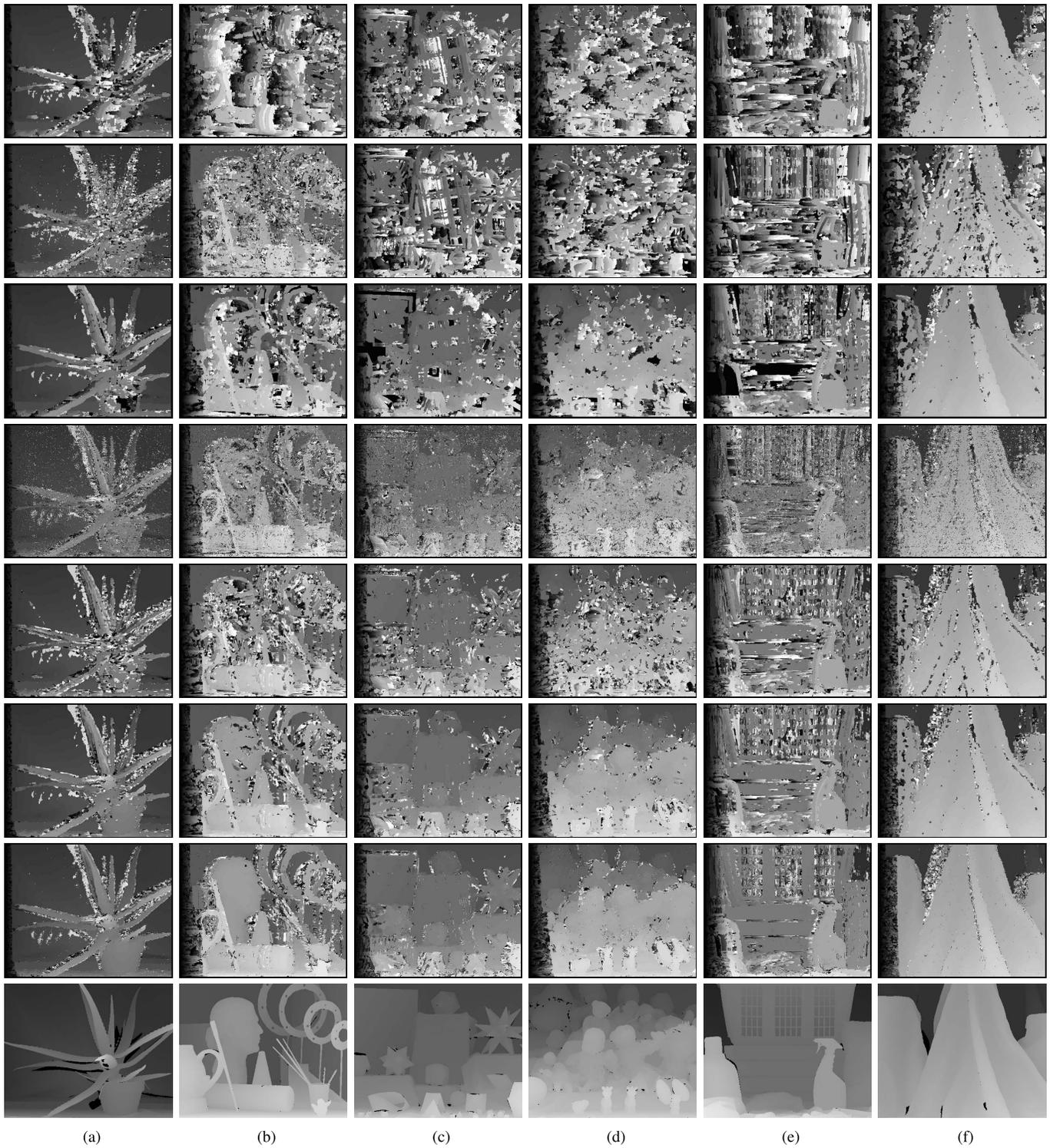


Fig. 5. Disparity maps for (a) *Aloe*, (b) *Art*, (c) *Moebius*, (d) *Dolls*, (e) *Laundry*, and (f) *Cloth4* image pairs taken under illumination combination one-third with methods (from top to bottom) including HE/CC + WTA, BilSub/CC + WTA, MI + WTA, Census + WTA, NCC + WTA, ANCC + WTA, MDCC + WTA, and ground truth.

With Proposition 1, the MDT space is defined as a collection of the Mahalanobis distances between the color itself and the average color. Thus, the MDT for pixel \mathbf{p} is defined as

$$\text{MDT}_1(\mathbf{p}) = \otimes_{\mathbf{q} \in \Psi_1(\mathbf{p})} d_{\mathbf{M}}(\mathbf{I}_1(\mathbf{q}), \bar{\mathbf{Q}}_1(\mathbf{p})). \quad (12)$$

In a similar way, the MDT for the pixel $\hat{\mathbf{p}}$ is defined as

$$\text{MDT}_2(\hat{\mathbf{p}}) = \otimes_{\hat{\mathbf{q}} \in \Psi_2(\hat{\mathbf{p}})} d_{\mathbf{M}}(\mathbf{I}_2(\hat{\mathbf{q}}), \bar{\mathbf{Q}}_2(\hat{\mathbf{p}})). \quad (13)$$

The MDT for the pixel \mathbf{p} is equivalent to that for corresponding pixel $\hat{\mathbf{p}}$ although the affine variation transforms

TABLE II
BAD PIXEL ERROR RATES AT UNOCCLUDED AREAS IN DISPARITY MAPS WITH VARYING THE COMBINATION OF ILLUMINATION INDEX.
BOLD RESULTS REPRESENT THE TWO LOWEST ERROR RATES AMONG SIMILARITY MEASURES

<i>Environments</i>		<i>Methods</i>						
<i>DataSet</i>	<i>illumination</i>	HE/CC	BilSub/CC	MI	Census	NCC	ANCC	MDCC
<i>Aloe</i>	1/1	20.851	27.361	17.265	18.42	16.06	13.174	12.732
	1/2	23.733	28.027	20.722	22.024	17.657	13.323	12.326
	1/3	26.754	30.464	26.102	32.539	25.637	20.856	18.737
	2/2	20.416	27.942	16.72	18.53	16.628	13.529	14.211
	2/3	26.289	29.396	24.059	28.009	22.919	18.548	15.212
	3/3	21.108	27.826	16.955	19.172	17.001	13.869	15.044
<i>Art</i>	1/1	41.317	58.613	42.733	35.923	35.022	31.482	30.992
	1/2	42.672	59.108	47.476	39.993	36.68	31.896	28.161
	1/3	49.095	62.827	56.32	53.335	48.402	44.05	35.727
	2/2	40.151	58.424	41.706	35.637	34.721	31.178	28.76
	2/3	65.066	61.9	54.81	51.233	46.543	43.627	36.819
	3/3	62.175	58.656	41.279	34.79	34.119	30.71	26.267
<i>Moebius</i>	1/1	51.054	44.167	26.471	25.438	18.774	17.109	16.259
	1/2	50.057	45.359	34.025	29.716	21.625	18.883	16.245
	1/3	52.983	49.615	43.357	41.45	30.181	28.262	24.56
	2/2	48.773	43.76	27.458	25.418	18.687	17.009	15.743
	2/3	50.836	47.672	38.991	37.299	26.953	25.081	23.216
	3/3	48.085	45.091	30.766	27.094	19.461	17.894	15.122
<i>Dolls</i>	1/1	50.415	44.782	21.188	21.669	17.044	15.72	15.156
	1/2	50.306	46.216	29.893	27.438	19.124	19.167	15.532
	1/3	56.258	52.88	30.588	33.522	27.082	27.521	21.538
	2/2	49.034	44.918	22.901	21.575	16.747	13.532	11.364
	2/3	54.49	50.961	40.887	41.924	31.596	32.735	27.811
	3/3	50.441	47.667	24.327	21.658	17.146	16.034	15.224
<i>Laundry</i>	1/1	61.784	62.375	48.824	38.843	30.404	28.067	24.86
	1/2	62.45	63.125	57.214	47.925	40.306	38.784	34.638
	1/3	64.566	64.672	58.962	55.284	46.085	42.346	39.754
	2/2	60.501	61.962	49.256	39.959	31.173	26.8	23.721
	2/3	62.016	62.842	55.486	46.78	38.503	34.547	28.317
	3/3	59.751	61.765	48.487	39.157	31.347	29.491	25.125
<i>Cloth4</i>	1/1	32.308	25.612	19.888	21.143	17.874	15.944	16.976
	1/2	33.172	26.28	23.974	31.164	22.033	20.451	17.426
	1/3	39.321	28.108	27.842	38.807	27.566	29.065	24.515
	2/2	30.808	25.334	20.315	21.188	18.013	16.14	17.571
	2/3	32.788	26.198	20.499	23.635	18.518	16.41	17.571
	3/3	32.818	26.234	20.174	21.205	18.431	16.628	15.126

the color distributions between support windows. In other words, the MDT provides illumination invariance, which is not available when using original color information. Fig. 3 shows that the illumination variation is dramatically reduced in the MDT space.

B. MDCC Similarity Measure

The MDCC similarity measure converts pixels within each support window into the MDT space and computes correlation similarity between MDT pairs.

To encode spatial information within support windows, the MDCC similarity measure uses the weight distribution around matching pixels. Because conventional weight distribution based on the Euclidean distance is degraded under the affine variation, as shown in Fig. 1, it cannot encode spatial

information correctly and causes outliers. Thus, the MDCC similarity measure employs the Mahalanobis distance instead of the Euclidean distance. Similar to Proposition 1, the Mahalanobis distance between $\mathbf{I}(\mathbf{q})$ for $\mathbf{q} \in \Psi(\mathbf{p})$ and $\mathbf{I}(\mathbf{p})$ within a support window provides illumination invariance. Then, the weight distribution based on the Mahalanobis distance $v(\mathbf{q}, \mathbf{p})$ is defined as

$$\begin{aligned}
 v(\mathbf{q}, \mathbf{p}) &= \exp\left(-\frac{\|\mathbf{q} - \mathbf{p}\|^2}{\gamma_g}\right) \exp\left(-\frac{d_M(\mathbf{I}(\mathbf{q}), \mathbf{I}(\mathbf{p}))}{\gamma_c}\right) \\
 &= \exp\left(-\frac{\|\mathbf{q} - \mathbf{p}\|^2}{\gamma_g}\right) \\
 &\quad \exp\left(-\frac{(\mathbf{I}(\mathbf{q}) - \mathbf{I}(\mathbf{p}))^T \cdot \Sigma^{-1}(\mathbf{p}) \cdot (\mathbf{I}(\mathbf{q}) - \mathbf{I}(\mathbf{p}))}{\gamma_c}\right).
 \end{aligned} \tag{14}$$

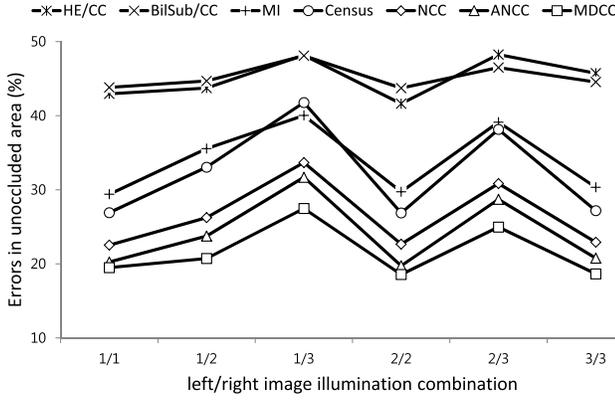


Fig. 6. Average bad pixel error rates at the unoccluded areas in disparity maps with varying the combination of illumination index. The MDCC shows the best performance with the lowest bad pixel error rates.

This weight distribution enables encoding of spatial information even with severe illumination variation, which will be described in Section III-B.

Given the MDT vector pairs, it is necessary to measure the similarity between them. For estimating corresponding pixels, proper similarity measures on the MDT space enable illumination invariance. Because the MDT itself has the effects of normalization within the support window, it is unnecessary to apply conventional normalized correlation methods between MDT pairs. In addition, simple difference measures between MDT vector pairs provide low distinctiveness, such as the sum of squared differences (SSD) and the sum of absolute differences (SAD). Thus, after transforming pixels within support windows into MDT vectors, the MDCC similarity measure computes the CC between MDT pairs with weight distributions based on the Mahalanobis distance in both support windows as shown in (15) at the bottom of this page, where $\mathbf{V}(\mathbf{p})$ is a diagonal matrix, whose diagonal elements are the weight distribution $v(\mathbf{q}, \mathbf{p})$. The maximum of the MDCC similarity measure represents the best similar candidate pixel to the reference pixel.

1) *Properties of the MDCC Similarity Measure:* In this section, the MDCC similarity measure is compared with normalized correlation methods such as NCC and ANCC, and it is verified that the MDCC similarity measure is more robust to affine illumination variation than these methods.

First, the MDCC similarity measure simultaneously considers correlation on each color and cross-color channels, thus providing illumination robustness and improving the performance. By neglecting off-diagonal components of the covariance matrix, the MDT for pixel \mathbf{p} can be approximated as

$$\text{MDT}_1(\mathbf{p}) \approx \sum_{\xi \in \Pi} \frac{\Phi_1^\xi(\mathbf{p})}{\|\Omega_1^\xi(\mathbf{p}) - \bar{\Omega}_1^\xi(\mathbf{p})\|} \quad (16)$$

where $\Phi^\xi(\mathbf{p}) = \otimes_{\mathbf{q} \in \Psi(\mathbf{p})} (I^\xi(\mathbf{q}) - \bar{\Omega}^\xi(\mathbf{p}))^2$. The denominator term in (16) can be thought of as the normalization term, which reduces an effect of illumination within support windows, similar to normalized correlation methods. After approximating the MDT for pixel $\hat{\mathbf{p}}$ in a similar way, the following relationship can be derived

$$\begin{aligned} \text{MDCC}(\mathbf{p}, \hat{\mathbf{p}}) &\approx \left\langle \sum_{\xi \in \Pi} \frac{\mathbf{V}_1(\mathbf{p})\Phi_1^\xi(\mathbf{p})}{\|\mathbf{V}_1(\mathbf{p})\|_F \cdot \|\Omega_1^\xi(\mathbf{p}) - \bar{\Omega}_1^\xi(\mathbf{p})\|}, \right. \\ &\quad \left. \sum_{\gamma \in \Pi} \frac{\mathbf{V}_2(\hat{\mathbf{p}})\Phi_2^\gamma(\hat{\mathbf{p}})}{\|\mathbf{V}_2(\hat{\mathbf{p}})\|_F \cdot \|\Omega_2^\gamma(\hat{\mathbf{p}}) - \bar{\Omega}_2^\gamma(\hat{\mathbf{p}})\|} \right\rangle. \quad (17) \end{aligned}$$

This shows that the MDCC similarity measure needs only single CC to compute the similarity between support windows; however, NCC and ANCC require correlation computation linearly with the number of color channels, as in (5) and (6). That is, the computational time of the MDCC similarity measure is reduced compared with these methods. Furthermore, the MDCC similarity measure explicitly has the effects of computing correlation similarity for each color and cross-color channel. For clarifying this property, (17) can be further decomposed into two parts according to the combination of color channel between support windows as shown in (18) at the bottom of this page.

The left-side term and right-side term in (18) represent the correlation on each color channel and the correlation between cross-color channels, respectively. Furthermore, the left-side term corresponds to conventional normalized correlation methods in terms of the correlation similarity on each color channel. As mentioned in the preceding section, these methods provide limited performance under severe illumination variation, because they do not consider the deformation from

$$\text{MDCC}(\mathbf{p}, \hat{\mathbf{p}}) = \left\langle \frac{\mathbf{V}_1(\mathbf{p})\text{MDT}_1(\mathbf{p})}{\|\mathbf{V}_1(\mathbf{p})\|_F}, \frac{\mathbf{V}_2(\hat{\mathbf{p}})\text{MDT}_2(\hat{\mathbf{p}})}{\|\mathbf{V}_2(\hat{\mathbf{p}})\|_F} \right\rangle = \sum_{\mathbf{q} \in \Psi_{\mathbf{p}}, \hat{\mathbf{q}} \in \Psi_{\hat{\mathbf{p}}}} \frac{v_1(\mathbf{q}, \mathbf{p})v_2(\hat{\mathbf{q}}, \hat{\mathbf{p}})d_M(\mathbf{I}_1(\mathbf{q}), \bar{\mathbf{I}}_1(\mathbf{p}))d_M(\mathbf{I}_2(\hat{\mathbf{q}}), \bar{\mathbf{I}}_2(\hat{\mathbf{p}}))}{\sqrt{\sum_{\mathbf{q} \in \Psi_{\mathbf{p}}} v_1(\mathbf{q}, \mathbf{p})^2 \sum_{\hat{\mathbf{q}} \in \Psi_{\hat{\mathbf{p}}}} v_2(\hat{\mathbf{q}}, \hat{\mathbf{p}})^2}} \quad (15)$$

$$\begin{aligned} \text{MDCC}(\mathbf{p}, \hat{\mathbf{p}}) &\approx \sum_{\xi \in \Pi} \left\langle \frac{\mathbf{V}_1(\mathbf{p})\Phi_1^\xi(\mathbf{p})}{\|\mathbf{V}_1(\mathbf{p})\|_F \cdot \|\Omega_1^\xi(\mathbf{p}) - \bar{\Omega}_1^\xi(\mathbf{p})\|}, \frac{\mathbf{V}_2(\hat{\mathbf{p}})\Phi_2^\xi(\hat{\mathbf{p}})}{\|\mathbf{V}_2(\hat{\mathbf{p}})\|_F \cdot \|\Omega_2^\xi(\hat{\mathbf{p}}) - \bar{\Omega}_2^\xi(\hat{\mathbf{p}})\|} \right\rangle \\ &\quad + \sum_{\substack{\xi \in \Pi, \\ \gamma \in \Pi \setminus \xi}} \left\langle \frac{\mathbf{V}_1(\mathbf{p})\Phi_1^\xi(\mathbf{p})}{\|\mathbf{V}_1(\mathbf{p})\|_F \cdot \|\Omega_1^\xi(\mathbf{p}) - \bar{\Omega}_1^\xi(\mathbf{p})\|}, \frac{\mathbf{V}_2(\hat{\mathbf{p}})\Phi_2^\gamma(\hat{\mathbf{p}})}{\|\mathbf{V}_2(\hat{\mathbf{p}})\|_F \cdot \|\Omega_2^\gamma(\hat{\mathbf{p}}) - \bar{\Omega}_2^\gamma(\hat{\mathbf{p}})\|} \right\rangle. \quad (18) \end{aligned}$$

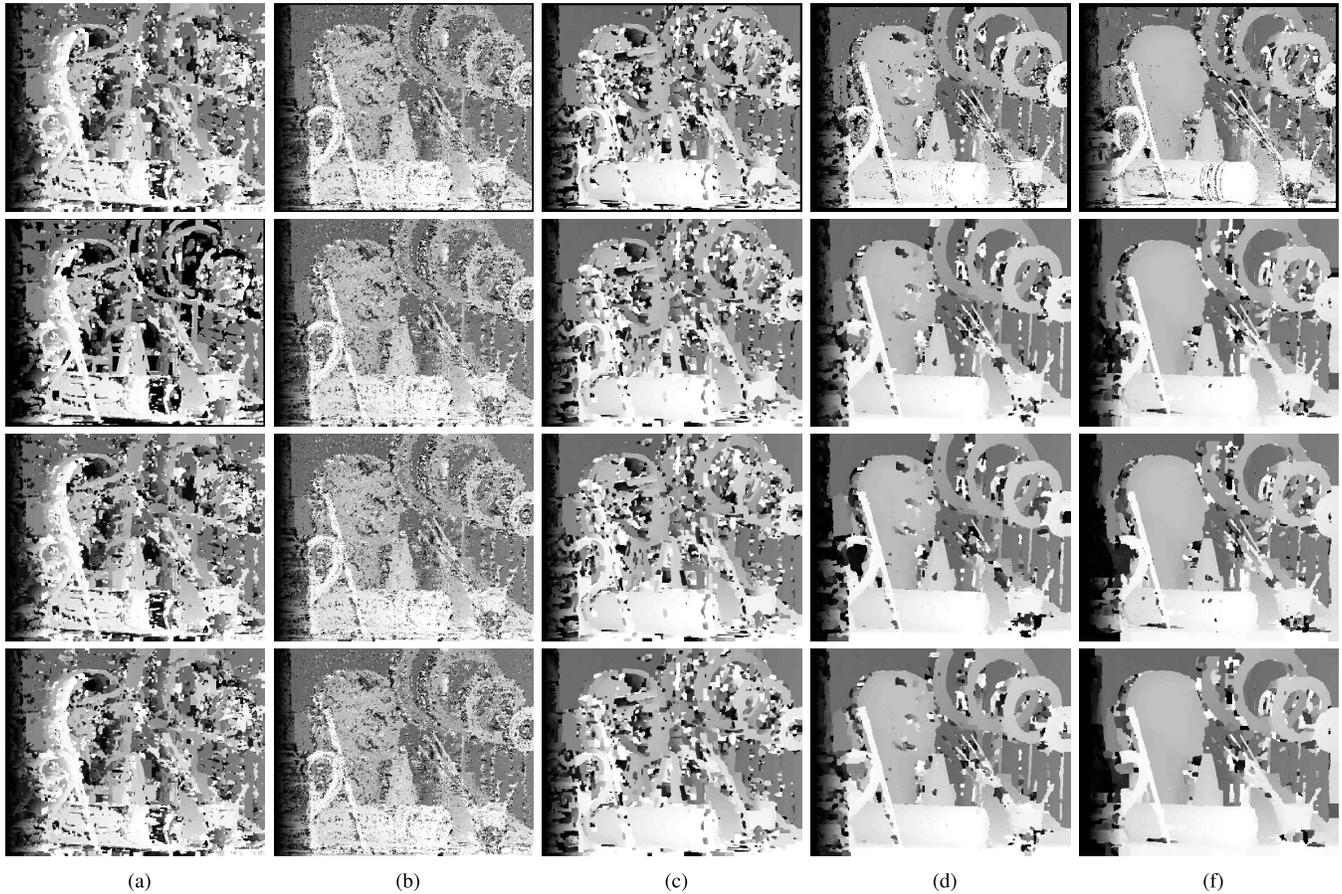


Fig. 7. Disparity maps for *Art* image pairs taken under illumination combination one-third with different optimization schemes such as (from top to bottom) ICM, BP, GC, and TRW. (a) MI. (b) Census. (c) NCC. (d) ANCC. (e) MDCC.

cross-color channels. In contrast, the MDCC similarity measure simultaneously considers the correlation on each color and cross-color channel. That is, the MDCC similarity measure provides more improved robustness than NCC and ANCC, by additionally considering the correlation similarity on cross-color channels.

Second, the MDCC similarity measure encodes spatial information invariant to the illumination variation between support windows. It employs the Mahalanobis distance to compute color difference for the weight distribution owing to the invariance of this metric under the affine deformation. Similar to Proposition 1, it can be easily derived as

$$d_M(\mathbf{I}_1(\mathbf{q}), \mathbf{I}_1(\mathbf{p})) = d_M(\mathbf{I}_2(\hat{\mathbf{q}}), \mathbf{I}_2(\hat{\mathbf{p}})). \quad (19)$$

Then, the proposed weight distribution is preserved under affine illumination variation such that

$$\begin{aligned} v_1(\mathbf{p}, \mathbf{q}) &= \exp\left(-\frac{\|\mathbf{q} - \mathbf{p}\|^2}{\gamma_g}\right) \exp\left(-\frac{d_M(\mathbf{I}_1(\mathbf{q}), \mathbf{I}_1(\mathbf{p}))}{\gamma_c}\right) \\ &= \exp\left(-\frac{\|\hat{\mathbf{q}} - \hat{\mathbf{p}}\|^2}{\gamma_g}\right) \exp\left(-\frac{d_M(\mathbf{I}_2(\hat{\mathbf{q}}), \mathbf{I}_2(\hat{\mathbf{p}}))}{\gamma_c}\right) \\ &= v_2(\hat{\mathbf{p}}, \hat{\mathbf{q}}). \end{aligned} \quad (20)$$

It shows that the proposed weight distribution enables spatial information encoding even under illumination variation. Fig. 4 shows the invariance of proposed weight distributions

between support windows taken under varying illumination conditions as shown in Fig. 1.

IV. EXPERIMENTAL RESULTS

In this section, the stereo matching performance of the MDCC and other global or local methods is compared using standard Middlebury data sets including the *Aloe*, *Art*, *Moebius*, *Dolls*, *Laundry*, and *Cloth4* image pairs [34]. Each data set consists of color image pairs taken under three different illumination conditions indexed from 1 to 3 and three different exposure conditions indexed from 0 to 2. To evaluate robustness to illumination variation, stereo pairs were selected according to the index of illumination, for example, illumination combination a/b was defined as an index of illumination varying from a to b. In the experiments, the parameters of the MDCC were fixed (15×15 window, $\gamma_g = 392$, and $\gamma_c = 62.7$). The MDCC was compared with several state-of-the-art methods: CC with histogram equalization transform (HE/CC) [17], CC with BilSub transform (BilSub/CC) [18], Census (7×7 window) [23], MI-based method (MI) (11×11 window, $binsize = 20$) [24] on the gray channel, and NCC (9×9 window) [4], ANCC (31×31 window, $\gamma_g = 392$, and $\sigma_s = 28.8$) [5] on RGB color channels. The parameters of each method were set with reference to the original works. HE/CC and BilSub/CC were also compared since the MDCC

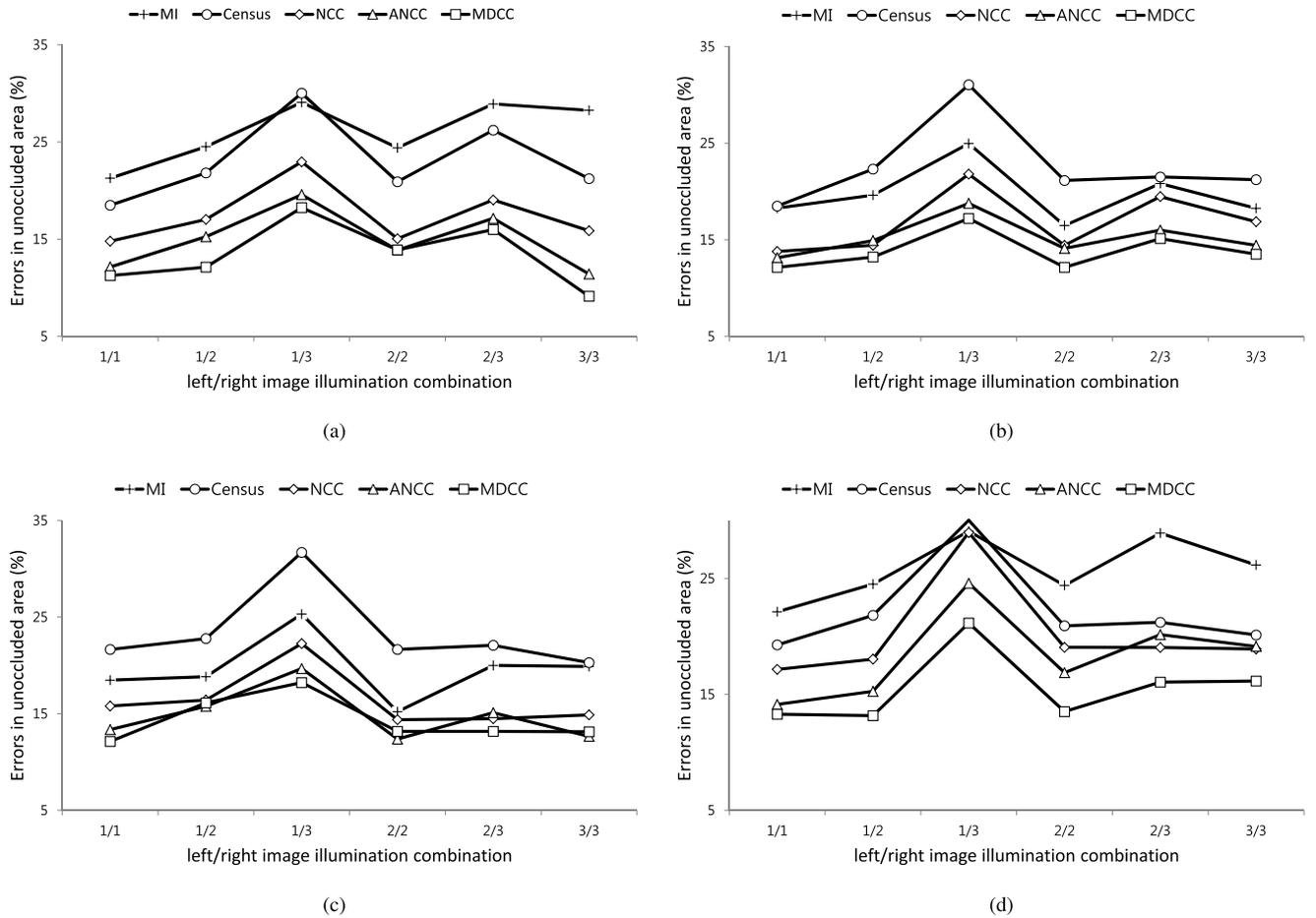


Fig. 8. Average bad pixel error rates at unoccluded areas in disparity maps with varying the combination of illumination index. (a) ICM. (b) BP. (c) GC. (d) TRW.

consists of CC with the Mahalanobis distance transform (MDT/CC).

To evaluate the effects of similarity measures only, a stereo matching framework was constructed with cost computation using similarity measures and simple winner-take-all (WTA) optimization [1]. In addition, the performances of similarity measures were also evaluated using varying optimization schemes such as iterated conditional modes (ICMs) [35], belief propagation (BP) [36], graph-cut (GC) [37], and tree-reweighted max-product message passing (TRW) [38], instead of WTA. To evaluate the robustness of spatial structure encoding in similarity measures, the disparity maps were estimated by varying window sizes and weight distributions. Furthermore, the performance of the MDCC was also evaluated under exposure variation.

The evaluation criterion is the bad pixel error rate in nonoccluded areas ε of the disparity map defined as

$$\varepsilon(\%) = 100 \times \frac{1}{N_{\mathcal{I}_{\text{noccl}}}} \sum_{\mathbf{p} \in \mathcal{I}_{\text{noccl}}} \begin{cases} 1, & \text{if } |\mathbf{D}_E(\mathbf{p}) - \mathbf{D}_G(\mathbf{p})| \geq 1 \\ 0, & \text{otherwise} \end{cases} \quad (21)$$

where $\mathcal{I}_{\text{noccl}}$ is the set of all nonoccluded pixels, $N_{\mathcal{I}_{\text{noccl}}}$ is the number of pixels within $\mathcal{I}_{\text{noccl}}$, and $\mathbf{D}_E(\mathbf{p})$ and $\mathbf{D}_G(\mathbf{p})$ are the estimated and ground truth disparity at pixel \mathbf{p} , respectively.

A. Illumination Change

1) *Comparison of Similarity Measures With Local (WTA) Optimization:* To evaluate the robustness for illumination variation, the disparity maps were estimated with different similarity measures for stereo pairs taken under varying illumination condition from 1 to 3 with fixed exposure condition. Fig. 5 shows the disparity maps of the MDCC and other similarity measures for the illumination combination 1/3. Table II shows the bad pixel error rates for all possible combinations of illumination.

Because the HE transform is based on the global rank order, its performance is severely degraded under local illumination variation dispersing the global rank order. The BilSub transform is also sensitive to severe illumination variation because it only compensates for a constant offset. The performance of the MI-based method is degraded under local variations, since it is assumed that there are global variations. The Census transform estimates relatively accurate disparity maps since it uses both the intensity order and spatial structure. However, it provides poor results on homogeneous regions with an indistinct order of pixels, as shown in the results of *Art* images. The normalized correlation methods such as NCC and ANCC perform well when compared with other methods. The NCC

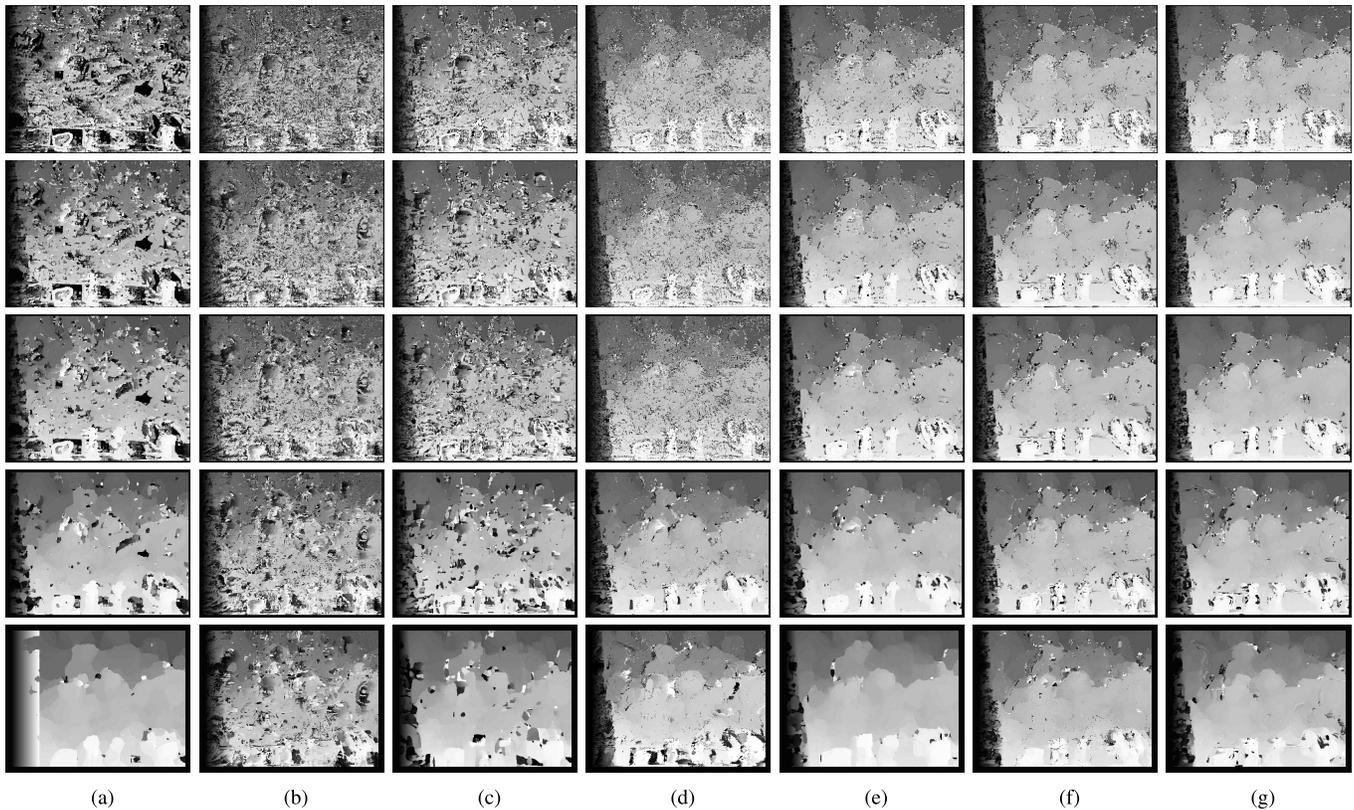


Fig. 9. Disparity maps for *Dolls* image pairs taken under illumination combination one-third with varying window size (from top to bottom) 5×5 , 7×7 , 9×9 , 15×15 , and 31×31 . (a) MI + WTA. (b) Census + WTA. (c) NCC + WTA. (d) ANCC + WTA. (e) MDCC + WTA without weight. (f) MDCC + WTA with conventional weight. (g) MDCC + WTA with proposed weight.

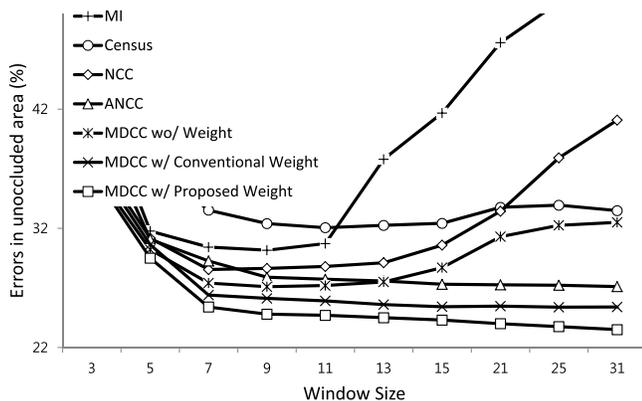


Fig. 10. Average bad pixel error rates at unoccluded areas in disparity maps while varying the window size. The MDCC shows the best performance with the lowest bad pixel error rates.

estimates relatively accurate disparity maps under illumination variation. However, its disparity maps contain large errors in the boundary regions because it does not encode the spatial structure. The ANCC improves matching performance using weight distributions in both support windows compared with the NCC. However, the discrimination power of the ANCC is limited under severe illumination variation such as illumination combination 1/3 as it ignores correlation between cross-color channels, similar to NCC. Furthermore, the ANCC

produces many outliers owing to the degradation of weights under illumination variation. Note that the ANCC was computed on the RGB color space, without a log-chromaticity normalization for fair evaluation, and these results are consistent with other evaluation results [3], [5]. The MDCC outperforms other similarity measures in most illumination combinations. Because the MDCC computes the correlation similarity between the illumination-invariant MDT spaces with weight distribution based on the Mahalanobis distance, it estimates accurate disparity maps and reduces the outliers under severe illumination variation. However, like all other methods, the performance of the MDCC weakens on regions in which the color distribution is unstable such as highly textured regions in *Laundry* images. Fig. 6 shows average bad pixel error rates over all stereo pairs with varying illumination combinations. The MDCC shows the best performance with the lowest error rates compared with other similarity measures. It shows remarkably improved performance in severe illumination variation such as illumination combination 1/3 or 2/3. Furthermore, the MDCC provides the highest performance for stereo pairs taken under consistent illumination conditions.

2) *Comparison of Similarity Measures With Global Optimization*: To evaluate the effects of similarity measures with global optimization schemes, the similarity measures were combined with ICM, BP, GC, and TRW. The parameters were fixed for all methods [5], [35]. Fig. 7 shows the disparity maps estimated by several similarity measures

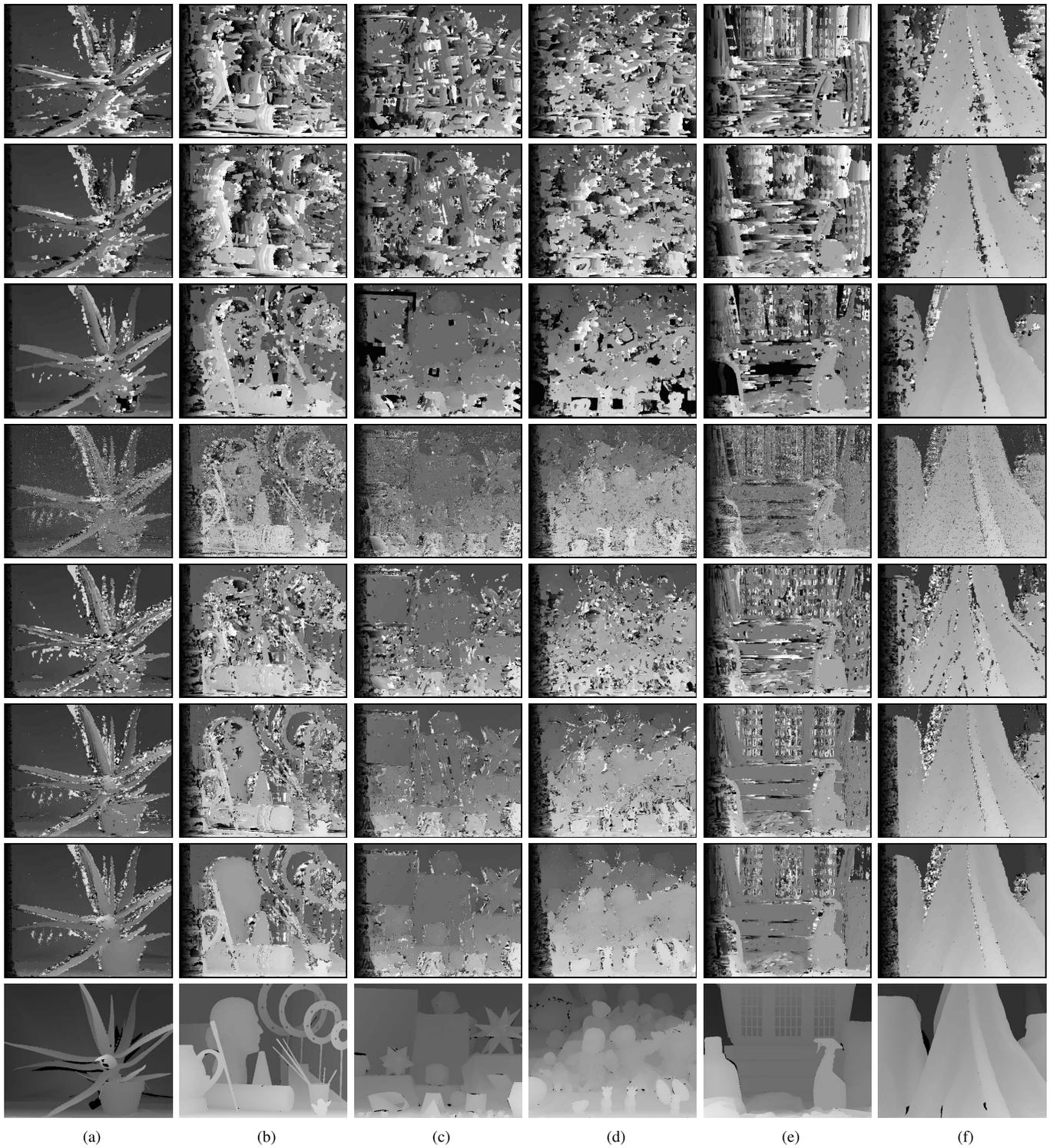


Fig. 11. Disparity maps for (a) *Aloe*, (b) *Art*, (c) *Moebius*, (d) *Dolls*, (e) *Laundry*, and (f) *Cloth4* image pairs taken under exposure combination 0/2 with methods (from top to bottom) including HE/CC + WTA, BilSub/CC + WTA, MI + WTA, Census + WTA, NCC + WTA, ANCC + WTA, MDCC + WTA, and ground truth.

including MI, Census, NCC, ANCC, and MDCC combined with different optimization schemes for Art images taken under illumination combination 1/3. Fig. 8 shows average pixel error rates for overall image pairs while varying the combination of illumination indexes, similarity measures, and

optimization schemes. Although the performances of disparity estimation vary with the optimization scheme, the MDCC shows competitive performance compared with state-of-the-art similarity measures, similar to the results of WTA optimization.

TABLE III
 BAD PIXEL ERROR RATES AT UNOCCLUDED AREAS IN DISPARITY MAPS WITH VARYING THE COMBINATION OF EXPOSURE INDEX.
 BOLD RESULTS REPRESENT THE TWO LOWEST ERROR RATES AMONG SIMILARITY MEASURES

<i>Environments</i>		<i>Methods</i>						
<i>DataSet</i>	<i>Exposure</i>	HE/CC	BilSub/CC	MI	Census	NCC	ANCC	MDCC
<i>Aloe</i>	0/0	26.564	26.502	18.251	18.377	13.734	10.869	11.125
	0/1	26.476	24.533	21.468	17.515	13.199	11.807	10.253
	0/2	26.306	23.787	21.055	17.563	13.223	11.667	11.126
	1/1	26.26	24.361	14.265	15.163	13.06	10.174	9.122
	1/2	26.041	23.491	14.775	15.163	13.061	10.761	9.423
	2/2	25.796	23.457	12.086	14.839	13.116	10.374	8.123
<i>Art</i>	0/0	59.71	56.403	44.385	36.719	32.793	29.171	25.124
	0/1	59.643	56.061	47.673	35.369	32.256	31.993	27.223
	0/2	59.861	57.658	53.285	38.608	34.365	34.159	28.522
	1/1	59.575	55.613	39.778	32.923	32.022	28.482	25.223
	1/2	59.755	57.041	45.63	35.798	33.792	31.757	28.122
	2/2	59.555	56.71	41.16	34.355	32.787	29.636	25.123
<i>Moebius</i>	0/0	48.707	42.682	30.414	26.367	16.815	14.367	14.529
	0/1	48.43	41.686	36.238	24.596	15.98	17.898	14.81
	0/2	47.876	43.596	40.429	26.227	17.556	20.315	18.214
	1/1	48.054	41.167	23.624	22.438	15.774	14.109	15.513
	1/2	47.587	42.457	30.356	23.706	16.932	17.008	14.085
	2/2	47.155	41.697	24.509	23.399	16.713	15.211	14.117
<i>Dolls</i>	0/0	47.824	42.959	23.671	21.843	15.278	12.905	16.124
	0/1	47.575	42.396	25.781	20.732	14.475	16.337	14.236
	0/2	47.453	46.656	32.855	22.678	15.511	19.651	16.247
	1/1	47.415	41.782	18.544	18.669	14.044	12.72	12.124
	1/2	47.209	45.454	26.419	20.11	14.492	15.95	13.324
	2/2	47.098	44.489	20.788	19.111	14.155	13.169	12.117
<i>Laundry</i>	0/0	59.539	59.776	49.076	39.019	27.404	25.558	20.214
	0/1	59.225	59.615	49.3	37.633	37.306	29.106	23.217
	0/2	58.845	60.476	57.594	39.707	43.085	31.509	24.824
	1/1	58.784	59.375	46.363	35.843	28.173	25.067	21.344
	1/2	58.262	59.945	51.633	37.612	35.503	30.305	25.236
	2/2	58.062	59.928	48.615	37.177	28.374	27.031	24.244
<i>Cloth4</i>	0/0	29.592	24.109	21.542	19.512	14.874	12.796	11.235
	0/1	29.603	22.896	20.853	18.945	19.033	13.239	13.723
	0/2	29.412	25.356	23.809	19.277	24.566	12.311	12.451
	1/1	29.308	22.612	16.868	18.163	15.013	12.944	11.242
	1/2	29.118	24.934	18.349	18.371	15.518	13.327	11.236
	2/2	28.843	24.479	17.86	18.378	15.431	12.765	11.255

3) *Comparison of Spatial Information Encoding*: To evaluate the robustness of spatial structure encoding in similarity measures, the disparity maps were estimated while varying the window size in each similarity measure for Dolls image pairs as shown in Fig. 9. Furthermore, to evaluate the effect of proposed weight distribution, the disparity maps were estimated by the MDCC while varying the weight distribution including the uniform, conventional, and proposed weight. Fig. 10 shows the average bad pixel error rates over all stereo pairs while varying the window size.

The MI-based method and the NCC with large window size generate degraded boundaries as the fattening effects because they do not consider the spatial structure. Because the Census transform and the ANCC encode the spatial information within support windows, they provide improved performance

as window size increases. However, the Census transform produces many outliers owing to inconsistency of ordering within a support window. Since the ANCC aggregates correlation similarity with the weight distribution is based on the Euclidean distance, the results are degraded owing to production of many outliers under severe illumination variation. This weight distribution cannot correctly encode spatial information and produces outliers as shown in Fig. 9(f), because it is also degraded under illumination variation. In contrast, the MDCC computes correlation similarity with the weight distribution based on the Mahalanobis distance, thus encoding spatial information within support windows invariant to illumination variation. This enables the estimation of more accurate disparity maps than conventional weight distributions as shown in Fig. 9(g). Note that, without the weight distribution,

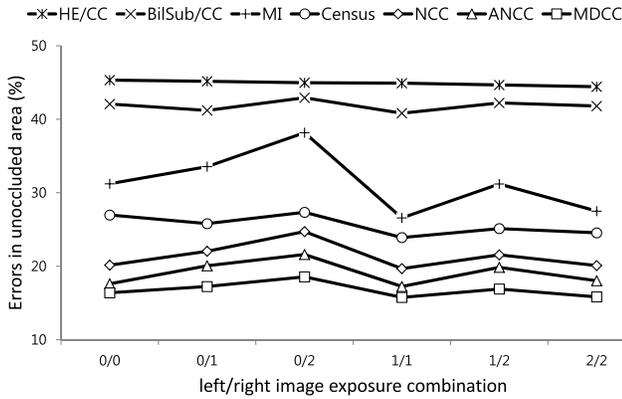


Fig. 12. Average bad pixel error rates at unoccluded areas in disparity maps with varying the combination of exposure index. The MDCC shows the best performance with the lowest bad pixel error rates.

the MDCC induces fattening effects similar to NCC, as shown in Fig. 9(e), as it does not consider the spatial structure. Furthermore, the MDCC shows relatively stable and competitive matching performance that is largely invariant to window size.

B. Exposure Change

The MDCC provides competitive performance under exposure changes because these variations can also be approximated using the affine model. For evaluating robustness to exposure changes, the index of exposure was changed from 0 to 2 with fixed illumination. Fig. 11 shows an example for the disparity maps estimated by the MDCC and other similar measures for stereo pairs taken under exposure condition 0/2. Table III shows average bad pixel error rates for all possible exposure combinations.

The HE/CC and BilSub/CC are sensitive to severe exposure changes since they cannot handle local variations of exposure. Compared with the experiments involving illumination changes, the MI-based method shows competitive performance under global variation. However, it shows limited performance due to the decrease in statistical power under exposure changes. Although the Census transform estimated relatively accurate disparity maps, errors still exist in homogenous regions because of the ambiguity of the rank order. NCC and ANCC show limited performance under severe exposure variation such as exposure combination 0/2. However, the MDCC demonstrates more stable performance with respect to the exposure variation in most cases, as shown in Fig. 12.

C. Computational Complexity

The computational time of the MDCC was compared with that of other similarity measures with WTA optimization for the *Aloe* images of 427×370 size and 70 disparities, which was calculated using an average of 20 runs. The window size of similarity measures was fixed to 15×15 for fair evaluation. The experiments ran on an Intel(R) Core(TM) i7-3770 CPU at 3.40 GHz. Fig. 13 shows normalized computational times, which are relative ratios of processing time with that of

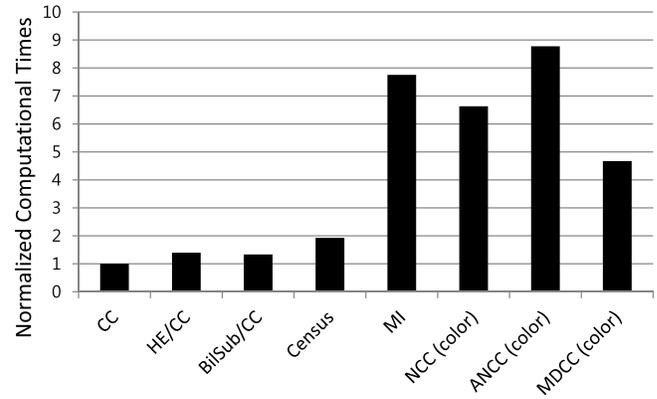


Fig. 13. Computational time performances of different similarity measures.

simple CC. Compared with the NCC, the computational complexity of the ANCC increases because of the computation of the weight distribution. To measure the correlation similarity between color image pairs, the complexity of NCC and ANCC increases linearly with the number of color channels since they independently compute the correlation on each color channel. In contrast, the MDCC computes the similarity with CC between the MDT vector pairs only. Thus, the computational complexity of the MDCC decreases compared with that of normalized correlation methods.

V. CONCLUSION

A robust similarity measure called MDCC has been proposed for illumination-invariant stereo matching. The MDCC uses the Mahalanobis distance instead of raw color information. It converts pixels within each support window into the MDT space and computes the similarity using the CC between MDT vector pairs. The MDCC aggregates the correlation similarity with a weight distribution on the basis of the Mahalanobis distance, encoding spatial information invariant to illumination changes. Furthermore, by computing correlation similarity on each color channel and cross-color channels simultaneously, the MDCC improves robustness and matching performance. Experimental results have shown that the MDCC outperforms state-of-the-art similarity measures for color image pairs under different illumination conditions. In addition, the MDCC provided excellent performance even under exposure changes.

In further work, the MDCC will be applied to address other correspondence problems taken under illumination variations, such as optical flow, visual tracking, and local feature matching.

REFERENCES

- [1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, nos. 1–3, pp. 7–42, May 2002.
- [2] K. Zhang, J. Lu, and G. Lafuit, "Cross-based local stereo matching using orthogonal integral images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 7, pp. 1073–1079, Jul. 2009.
- [3] H. Hirschmüller and D. Scharstein, "Evaluation of stereo matching costs on images with radiometric differences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 9, pp. 1582–1599, Sep. 2009.

- [4] O. Faugeras *et al.*, "Real-time correlation-based stereo: Algorithm, implementations and applications," INRIA, Rocquencourt, France, Tech. Rep. RR-2013, 1993.
- [5] Y. S. Heo, K. M. Lee, and S. U. Lee, "Robust stereo matching using adaptive normalized cross-correlation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 4, pp. 807–822, Apr. 2011.
- [6] I. Craw, N. Costen, T. Kato, and S. Akamatsu, "How should we represent faces for automatic recognition?" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 8, pp. 725–736, Aug. 1999.
- [7] N. Habili, C. C. Lim, and A. Moini, "Segmentation of the face and hands in sign language video sequences using color and motion cues," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 8, pp. 1086–1097, Aug. 2004.
- [8] K. Q. Weinberger and L. Saul, "Distance metric learning for large margin nearest neighbour classification," *J. Mach. Learn. Res.*, vol. 10, pp. 207–244, Dec. 2009.
- [9] S. Xiang, F. Nie, and C. Zhang, "Learning a Mahalanobis distance metric for data clustering and classification," *Pattern Recognit.*, vol. 41, no. 12, pp. 3600–3612, Dec. 2008.
- [10] T. Tuytelaars and L. V. Gool, "Wide baseline stereo matching based on local, affinely invariant regions," in *Proc. Brit. Mach. Vis. Conf., BMVC*, 2000, pp. 412–422.
- [11] F. Schaffalitzky and A. Zisserman, "Viewpoint invariant texture matching and wide baseline stereo," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2001, pp. 636–643.
- [12] C. Schmid and R. Mohr, "Local grayvalue invariants for image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 5, pp. 530–535, May 1997.
- [13] B. Platel, E. Balmachnova, L. M. J. Florack, and B. M. Romeny, "Top-points as interest points for image matching," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 418–429.
- [14] M. D. Fairchild, *Color Appearance Models*, 2nd ed. Chichester, U.K.: Wiley, 2005.
- [15] R. W. G. Hunt, *The Reproduction of Color*, 5th ed. Amenia, NY, USA: Fountain Press, 1995.
- [16] G. Finlayson, B. Schiele, and J. L. Crowley, "Comprehensive colour image normalization," in *Proc. Eur. Conf. Comput. Vis.*, 1998, pp. 475–490.
- [17] G. Finlayson, S. Hordley, G. Schaefer, and G. Y. Tian, "Illuminant and device invariant colour using histogram equalisation," *Pattern Recognit.*, vol. 38, no. 2, pp. 179–190, Feb. 2005.
- [18] A. S. Ogale and Y. Aloimonos, "Robust contrast invariant stereo correspondence," in *Proc. IEEE Int. Conf. Robot. Autom.*, Apr. 2005, pp. 815–824.
- [19] H. Hirschmüller, P. Innocent, and J. Garibaldi, "Real-time correlation-based stereo vision with reduced border errors," *Int. J. Comput. Vis.*, vol. 47, nos. 1–3, pp. 229–246, Apr. 2002.
- [20] A. Ansar, A. Castano, and L. Matthies, "Enhanced real-time stereo using bilateral filtering," in *Proc. Int. Symp. 3D Data Process., Vis. Transmiss.*, 2004, pp. 455–462.
- [21] D. Bhat and S. Nayar, "Ordinal measures for image correspondence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 4, pp. 415–423, Apr. 1998.
- [22] L. Wang, R. Yang, and J. E. Davis, "BRDF invariant stereo using light transport constancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 9, pp. 1616–1626, Sep. 2007.
- [23] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *Proc. IEEE Conf. Eur. Conf. Comput. Vis.*, May 1994, pp. 151–158.
- [24] G. Egnal, "Mutual information as a stereo correspondence measure," Dept. Comput. Inform. Sci., Univ. Pennsylvania, Philadelphia, PA, USA, Tech. Rep. MS-CIS-00-20, 2000.
- [25] J. Kim, V. Kolmogorov, and R. Zabih, "Visual correspondence using energy minimization and mutual information," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2003, pp. 1033–1040.
- [26] X. Song, D. Muselet, and A. Tremeau, "Affine transforms between image space and color space for invariant local descriptors," *Pattern Recognit.*, vol. 46, no. 8, pp. 2376–2389, Aug. 2013.
- [27] Y. Tsin, R. Collins, V. Ramesh, and T. Kanade, "Bayesian color constancy for outdoor object recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Dec. 2001, pp. 1132–1139.
- [28] K. V. Sande, T. Gevers, and C. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1582–1596, Sep. 2010.
- [29] K. Yoon and I. S. Kweon, "Adaptive support-weight approach for correspondence search," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
- [30] K. Muhlmann, D. Maier, J. Hesser, and R. Manner, "Calculating dense disparity maps from color stereo images, an efficient implementation," *Int. J. Comput. Vis.*, vol. 47, no. 1, pp. 79–88, 2002.
- [31] M. Bleyer, S. Chambon, and M. Gelautz, "Evaluation of different methods for using colour information in global stereo matching approaches," in *Proc. Int. Soc. Photogrammetry Remote Sensing, ISPR*, 2008, pp. 415–422.
- [32] M. Bleyer and S. Chambon, "Does color really help in dense stereo matching," in *Proc. Int. Symp. 3D Data Process., Vis. Transmiss.*, 2010, pp. 1–8.
- [33] M. Hazewinkel, *Covariance Matrix, Encyclopedia of Mathematics*. New York, NY, USA: Springer-Verlag, 2001.
- [34] (2010). *Middlebury Stereo*. [Online]. Available: <http://vision.middlebury.edu/stereo/>
- [35] R. Szeliski *et al.*, "A comparative study of energy minimization methods for Markov random fields," in *Proc. IEEE Eur. Conf. Comput. Vis.*, Jun. 2006, pp. 16–29.
- [36] M. F. Tappen and W. T. Freeman, "Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2003, pp. 900–907.
- [37] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.
- [38] V. Kolmogorov, "Convergent tree-reweighted message passing for energy minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1568–1583, Oct. 2006.



Seungryong Kim (S'12) received the B.S. degree in electrical and electronic engineering from Yonsei University, Seoul, Korea, in 2012, where he is currently working toward the joint M.S. and Ph.D. degrees.

His research interests include local feature descriptor, image matching, machine learning, and 3-D computer vision.



Bumsub Ham (M'14) received the B.S. and Ph.D. degrees in electrical and electronic engineering from Yonsei University, Seoul, Korea, in 2008 and 2013, respectively.

He was with Yonsei University as a Post-Doctoral Researcher from 2013 to 2014. He is currently with the WILLOW Team, INRIA Rocquencourt, École Normale Supérieure, Paris, France, and the Centre National de la Recherche Scientifique, Paris, as a Post-Doctoral Fellow. His research interests include variational methods and geometric partial differential equations, both in theory and applications in computer vision and image processing, particularly regularization, stereovision, superresolution, and HDR imaging.

Dr. Ham received the Honor Prize in the 17th Samsung Human-Tech Prize in 2011 and the Grand Prize in the Qualcomm Innovation Fellowship in 2012.



Bongjoe Kim (S'05) received the B.S., M.S., and Ph.D. degrees from the School of Electrical and Electronic Engineering, Yonsei University, Seoul, Korea, in 2005, 2007, and 2013, respectively.

He is with Samsung Electronics Company, Ltd., Suwon, Korea. His research interests include color image processing and 3-D computer vision.



Kwanghoon Sohn (M'92–SM'12) received the B.E. degree in electronic engineering from Yonsei University, Seoul, Korea, in 1983; the M.S.E.E. degree in electrical engineering from University of Minnesota, Minneapolis, MN, USA, in 1985; and the Ph.D. degree in electrical and computer engineering from North Carolina State University, Raleigh, NC, USA, in 1992.

He was a Senior Research Staff Member with the Satellite Communication Division, Electronics and Telecommunications Research Institute, Daejeon, Korea, from 1992 to 1993 and a Post-Doctoral Fellow with the MRI Center, Medical School of Georgetown University, Washington, DC, USA, in 1994. He was a Visiting Professor with Nanyang Technological University, Singapore, from 2002 to 2003. He is currently a Professor with the School of Electrical and Electronic Engineering, Yonsei University. His research interests include 3-D image processing, computer vision, and image communication.

Dr. Sohn is a member of the International Society for Optics and Photonics.