

# A Uniform Framework for Estimating Illumination Chromaticity, Correspondence, and Specular Reflection

Qingxiong Yang, *Member, IEEE*, Shengnan Wang, *Student Member, IEEE*, Narendra Ahuja, *Fellow, IEEE*, and Ruigang Yang, *Member, IEEE*

**Abstract**—Based upon a new correspondence matching invariant called illumination chromaticity constancy, we present a new solution for illumination chromaticity estimation, correspondence searching, and specular removal. Using as few as two images, the core of our method is the computation of a vote distribution for a number of illumination chromaticity hypotheses via correspondence matching. The hypothesis with the highest vote is accepted as correct. The estimated illumination chromaticity is then used together with the new matching invariant to match highlights, which inherently provides solutions for correspondence searching and specular removal. Our method differs from the previous approaches: those treat these vision problems separately and generally require that specular highlights be detected in a preprocessing step. Also, our method uses more images than previous illumination chromaticity estimation methods, which increases its robustness because more inputs/constraints are used. Experimental results on both synthetic and real images demonstrate the effectiveness of the proposed method.

**Index Terms**—Chromaticity, dichromatic reflection model, reflection components separation, specular reflection, stereo matching.

## I. INTRODUCTION

A COMMON assumption in many low-level vision problems is that the scene surface is Lambertian. When dealing with non-Lambertian objects, many problems have to be addressed, e.g., illumination chromaticity estimation [1]–[3], specular removal [4]–[9], and correspondence searching for non-Lambertian surfaces [10]–[13]. Usually, these problems are treated separately, with assumptions made about one factor to solve another.

Many illumination chromaticity estimation methods assume that the specular highlight regions are detected in a preprocessing step. Lee [1] introduced a method for estimating illumination chromaticity using highlights from surface parts

with at least two colors. This requires segmenting the colors of the highlights and causes problems when dealing with heavily textured surfaces. Finlayson and Schaefer [2] extended Lee's algorithm without solving the segmentation problem. Tan *et al.* [3] proposed an illumination chromaticity estimation approach for single/multicolored surfaces without using color segmentation. This method still requires the detection of rough highlight regions achieved by setting a heuristic threshold on the image intensity. Using 3-D spatial information reconstructed from a stereo image pair, Xiong and Funt [14] improved the performance of the multiresolution implementation of retinex known as McCann99 [15].

With preknowledge of the illumination chromaticity, many specular highlight removal methods can be performed. For instance, Lin and Shum [4] changed the light source direction to capture two color images with a stable camera and then, by assuming that at least one of the pixels in the two images was diffuse, the diffuse component could be extracted. With a single image that was normalized using the estimated illumination chromaticity, Tan and Ikeuchi used either a neighbor-based method [6] or a color space analyzing method [5] to recover the diffuse components. Neighbor-based methods examine the expected diffuse colors of neighboring pixels in the image. These methods either require repetitive texture or they simply work with low-textured surfaces. Color space analyzing methods analyze the distributions of image colors within a color space; they may be greatly impaired by clutter in the color space, which may be caused by a number of factors, including image noise and color blending at edges. Tan *et al.* [7] later presented a hybrid that blends both methods. An SUV color space, as proposed by Mallick *et al.* [8], separates the specular and diffuse components into S channel and UV channels, respectively, based upon knowledge of the illumination chromaticity. Mallick *et al.* [9] also used the SUV space for highlight removal by iteratively eroding the specular channel using either a single image or video sequences.

Using the estimated specular-free images or the detected highlight regions, researchers developed different correspondence searching methods for non-Lambertian surfaces. Zickler [10] further explored the SUV color space proposed in [8]. By taking the ratio of the diffuse channels, a new specular invariant was extracted, and both the diffuse channels and the invariant were then used for stereo matching, optical flow, shape from shading, and photometric stereo. Yoon [11] extracted a

Manuscript received February 15, 2010; revised June 15, 2010; accepted June 15, 2010. Date of publication June 28, 2010; date of current version December 17, 2010. This work was supported in part by Hewlett-Packard under the Open-Innovation Research program. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Mark (Hong-Yuan) Liao.

Q. Yang, S. Wang, and N. Ahuja are with the Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: qyang6@uiuc.edu; wang403@illinois.edu; n-ahuja@illinois.edu).

R. Yang is with the Department of Computer Science, University of Kentucky, Lexington, KY 40506 USA (e-mail: ryang@cs.uky.edu).

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/TIP.2010.2055573

two-band, specular-free image for correspondence matching. Correspondence searching can also be performed by rejecting the detected specular regions as outliers. In [16], specular pixels in multiview images were detected by first computing the uncertainty of depth estimates. Detected pixels were then treated as outliers when computing the similarity among pixels to reduce the effect of specular reflection. Bhat and Nayar [17] considered the likelihood of correct stereo matching by analyzing the relationship between stereo vergence and surface roughness. They also proposed a trinocular system where only two images were used at a time in the computation of depth at a point. Brelstaff and Blake [12], [13] excised specularities as a preprocessing step.

In this paper, we provide a new solution to three vision problems: illumination chromaticity estimation, correspondence searching, and specular removal. The foundation of the solution is a new matching invariant called illumination chromaticity constancy, introduced in the paper. We search for correspondence by analyzing the chromaticities of the color differences between the corresponding pixels in two camera images; we define this correspondence as chromaticity match. Each correspondence hypothesis is associated with a chromaticity match and subsequently votes for a specific illumination chromaticity hypothesis according to this match value. In this paper, we assume that the objects are stationary and that the images are captured by a moving camera under the same illuminant. For correct correspondence hypotheses, the corresponding chromaticity match values will be the same as the constant illumination chromaticity. Thus, the correct illumination chromaticity hypothesis will win with the greatest amount of votes. However, there is a significant amount of noise contributed by 1) a large number of incorrect correspondence vectors and 2) because the highlight areas are usually much smaller than the diffuse areas, and the latter contribute noise. We then apply the local smoothness constraint popularly used in the field of correspondence matching to suppress noise. Unlike previous methods, our illumination chromaticity estimation approach does not rely on the detection of the specular pixels. Moreover, since chromaticity match computed from the correct correspondence should be equal to the estimated illumination chromaticity, it can be used as a new matching invariant to match highlights. With the additional assumption that the highlights in the two images do not spatially overlap and, thus, the diffuse component of a pixel in the highlight can be extracted by finding its corresponding pixel in the other view, the diffuse reflection can be estimated too. Our experimental results demonstrate the effectiveness and robustness of our method, as more inputs/constraints are used.

## II. ILLUMINATION CHROMATICITY CONSTANCY

Surface reflection of dielectric inhomogeneous objects can be described by the dichromatic reflection model [18], which states that the light reflected from an object is a linear combination of diffuse and specular reflections. Let  $\Gamma_c$  ( $c \in \{R, G, B\}$ ) denote the fraction of the color component  $c$  present in the illumination, called illumination chromaticity. Let  $m_s(p)$  denote the

total specular reflection from a pixel  $p$  over all colors. Using the dichromatic reflection model, the color values of the pixel  $p$  in an image  $I$  taken by an RGB camera can be represented as

$$I_c(p) = I_c^{\text{diff}}(p) + I_c^{\text{spec}}(p) = I_c^{\text{diff}}(p) + m_s(p)\Gamma_c, \quad (1)$$

where  $m_s(p) = \sum_{c \in \{R, G, B\}} I_c^{\text{spec}}(p)$ , and

$$\Gamma_c = \frac{I_c^{\text{spec}}(p)}{m_s(p)}. \quad (2)$$

Assume that  $I$  and  $J$  are two images captured by the same camera from different viewing directions. Let  $p$  be a pixel in image  $I$ , and  $\bar{p} = p + D(p)$  be its corresponding pixel in  $J$  at a relative location given by the correspondence vector  $D(p)$ . We then define chromaticity match as

$$M_c(p, D(p)) = \frac{I_c(p) - J_c(p + D(p))}{\sum_{c \in \{R, G, B\}} (I_c(p) - J_c(p + D(p)))}. \quad (3)$$

$M_c(p, D(p))$  is set to zero when  $I_c(p) = J_c(p + D(p))$ . If both the objects and the light source stay still, the diffuse components of the correctly matched pixels will be cancelled, and the color difference will equal to the difference between their specular components

$$I_c(p) - J_c(p + D(p)) = (m_s(p) - m_s(p + D(p)))\Gamma_c. \quad (4)$$

Substituting (4) into (3), we can see that the chromaticity match of the correctly matched pixels is the same as the illumination chromaticity  $\Gamma_c$

$$M_c(p, D(p)) = \frac{(m_s(p) - m_s(p + D(p)))\Gamma_c}{\sum_{c \in \{R, G, B\}} ((m_s(p) - m_s(p + D(p)))\Gamma_c)} = \Gamma_c. \quad (5)$$

We refer to this property as illumination chromaticity constancy and use it to simultaneously estimate pixel correspondences and  $\Gamma_c$ , which is assumed to be constant across the scene. An estimate of  $\Gamma_c$  helps in matching pixels within specularities, and knowledge of pixel correspondences helps in estimating  $\Gamma_c$ .

## III. ESTIMATE ILLUMINATION CHROMATICITY VIA MATCHING

We first formulate the illumination chromaticity estimation problem as a maximum likelihood (ML) estimation problem that infers the illumination chromaticity given the two input color images and the correspondence vectors. Using Bayes' rule, the optimal solution  $\Gamma$  is given by

$$\arg\max_{\Gamma} P(\Gamma|I, J, D) = \arg\max_{\Gamma} \frac{P(I, J|\Gamma, D)P(\Gamma, D)}{P(I, J, D)} \quad (7)$$

where  $\Gamma$  is the constant illumination chromaticity vector (the same for every pixel because the illumination is assumed to be chromatically uniform),  $D$  is the correspondence vectors, and  $I$  and  $J$  are the camera images. Let  $H_I$  denote the set of highlights

in image  $I$ , and  $p$  denote a pixel in  $I$ . Assuming that the observation noise at different pixels follows an independent identical distribution (i.i.d.)

$$P(I, J|\Gamma, D) \propto \prod_{p \in H_I} \exp(-F(I, J, \Gamma, p, D(p))) \quad (8)$$

where  $D(p)$  represents the correspondence vectors for  $p \in H_I$  and  $F(I, J, \Gamma, p, D(p))$  is the cost function that measures the difference between the estimated value of the illumination chromaticity  $\Gamma$  and the chromaticity match  $M(p, D(p))$  estimated from the individual pixel correspondences as defined in (3), all of which should ideally be the same as  $\Gamma$  (we dropped the subscript  $c$  for simplicity). In practice, both i.i.d. and not i.i.d. noise may be present in images. One type of noise that is reasonably well modeled as i.i.d. is sensor noise (e.g., shot noise); noise that occurs due to certain environmental effects that cause spatial bursts is obviously not i.i.d. Our approach handles only i.i.d. noise. Also, we consider only low values of  $F$  because the others arise from wrong correspondences. To this end, we limit the  $F$  value of interest to 1, using the truncated squared difference as the cost

$$F(I, J, \Gamma, p, D(p)) = \min \left( \frac{(\Gamma - M(p, D(p)))^2}{2\sigma_p^2}, 1 \right) \quad (9)$$

where  $\sigma_p < 1$  is the standard deviation of the pdf of  $(I, J|\Gamma, D)$  as in (8). In this paper, we set  $\sigma_p = 0.01$ .

Since  $D$  (representing 3-D structure) and  $\Gamma$  (representing light source) are independent

$$P(\Gamma, D) = P(\Gamma)P(D). \quad (10)$$

In the absence of any knowledge of the light source, we will assume that  $\Gamma$  is uniformly distributed in  $[0, 1]$ , thus,  $P(\Gamma) = 1$ . For the given input images  $I$  and  $J$  and the given correspondence vectors  $D$ ,  $P(I, J, D) = 1$  and  $P(D) = 1$ .

Taking logarithm on both sides of (7), and using (8), the optimal  $\Gamma$  is given by

$$\operatorname{argmax}_{\Gamma} \log(P(\Gamma|I, J, D)) \quad (11)$$

$$= \operatorname{argmax}_{\Gamma} \sum_{p \in H_I} (-F(I, J, \Gamma, p, D(p))) \quad (12)$$

$$= \operatorname{argmax}_{\Gamma} \sum_{p \in H_I} (1 - F(I, J, \Gamma, p, D(p))). \quad (13)$$

To obtain the value of  $\Gamma$ , we compute votes for discretized values of  $\Gamma$ . We define the vote distribution  $V(\Gamma)$  as

$$V(\Gamma) = \sum_{p \in H_I} V(\Gamma, p, D(p)) \quad (14)$$

$$= \sum_{p \in H_I} (1 - F(I, J, \Gamma, p, D(p))). \quad (15)$$

We select the  $\Gamma$  with the largest vote value as the solution. The vote distribution is computed for each channel separately, and the peak of the vote distributions is accepted as the correct illumination chromaticity.

The values of  $D$  have not been constrained at all so far. Because we do not know where the highlights are, we must consider all pixels as candidates. Further, since we do not know where the pixel correspondences are either, we must consider all possible pairs. If images  $I$  and  $J$  are taken from calibrated cameras, we can restrict the correspondences of a pixel in  $I$  to lie along its epipolar line in  $J$ , that is, we can rewrite (15) as

$$V(\Gamma) = \sum_{p \in I} \sum_{D(p) \in E(p)} (1 - F(I, J, \Gamma, p, D(p))) \quad (16)$$

where  $E(p)$  denotes the epipolar line of  $p$ . Alternatively, if the cameras are not calibrated, we will need to search across all possible values through the maximum possible.

In our experiments (Section V), we searched for a pixel's correspondence within a  $300 \times 300$  window. Thus, there is a significant amount of noise contributed by 1) a large number of incorrect correspondence vectors, and 2) because the highlight areas are usually much smaller than the diffuse areas, and the latter contribute noise.

To suppress the noise, a local smoothness assumption (commonly used in the field of correspondence matching) is adopted, which enforces similarity between chromaticity estimates at a pixel and those at other pixels within a surrounding window, called the *support window*. Let  $N(p)$  be the support window at pixel  $p$  in image  $I$  and  $q \in N(p)$  be corresponding pixels in the support windows  $N(p)$ . Then, the local smoothness assumption implies that chromaticity match  $M(p, D(p)) = M(q, D(p))$  if the correspondence vector  $D(p)$  is correct. Using this constraint, we rewrite the vote distribution in (16) as

$$V(\Gamma) = \sum_{p \in I} \sum_{D(p) \in E(p)} C_p (1 - F(I, J, \Gamma, p, D(p))) \quad (17)$$

where

$$C_p = \prod_{q \in N(p)} \exp \left( -\frac{(M(p, D(p)) - M(q, D(p)))^2}{2\sigma_p^2} \right) \quad (18)$$

measures the consistency of the chromaticity matchers inside the support window.

There are other constraints that can be used to suppress the noise. Examples include: 1) the three channels of the computed chromaticity match should be of the same sign because the illumination chromaticity will never be negative; and 2) in (3), if  $|I_c(p) - J_c(p + D(p))|$  is smaller than a threshold (set to 10), the corresponding chromaticity match will be dropped to avoid quantization noises.

Our framework, as described in this paper, is based upon binocular matching. However, the framework can be easily extended to multiview matching. We simply need to perform binocular matching on every image pair and accumulate their vote distributions.

#### IV. STEREO MATCHING AND HIGHLIGHT REMOVAL

In stereo vision, the corresponding pixels in the stereo images are assumed to have the same texture/color values. In this paper, we refer to this phenomenon as texture constancy (**TC**). More specifically, let  $I$  and  $J$  be the left and right camera images,  $p$  be

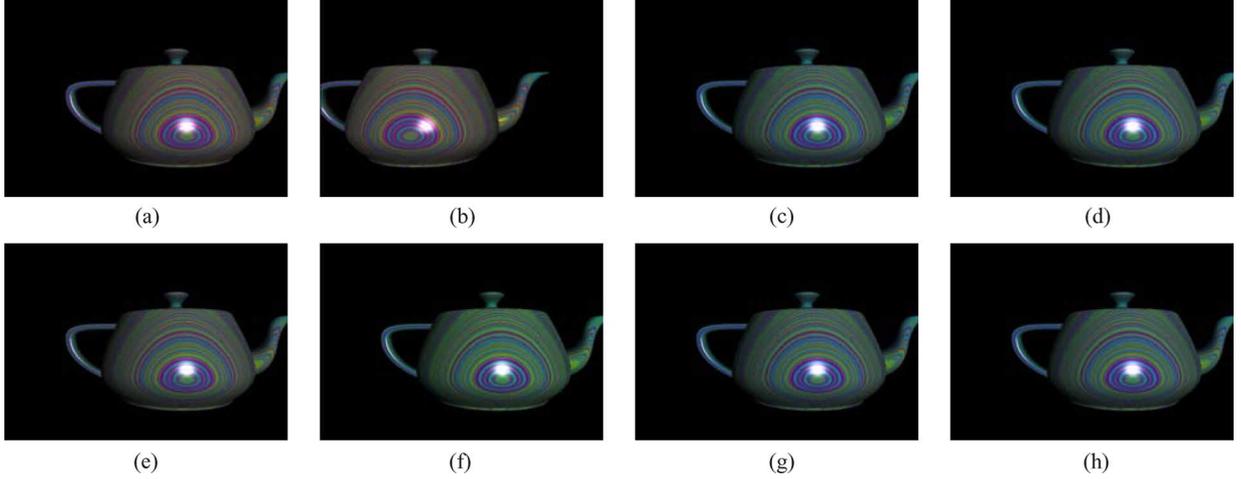


Fig. 1. Illumination chromaticity estimation for the synthetic images. (a) and (b) Input images. (c) Normalized image using the ground-truth illumination chromaticity. (d) and (e) Normalized images using the methods presented in [3] and [14], respectively. (f)–(h) Normalized images using our method with three support window sizes:  $1 \times 1$ ,  $3 \times 3$ , and  $5 \times 5$ . Note that (d), (g), and (h) are visually very similar to the ground truth in (c). The blue angle numbers in parentheses under (d)–(h) are the angular errors of the estimated illumination chromaticity. The images are gamma corrected for better illustration. (a) Left. (b) Right. (c) Ground truth. (d) [3] ( $0.04^\circ$ ). (e) [14] ( $3.35^\circ$ ). (f)  $1 \times 1$  ( $2.9^\circ$ ). (g)  $3 \times 3$  ( $0.1^\circ$ ). (h)  $5 \times 5$  ( $0.007^\circ$ ).

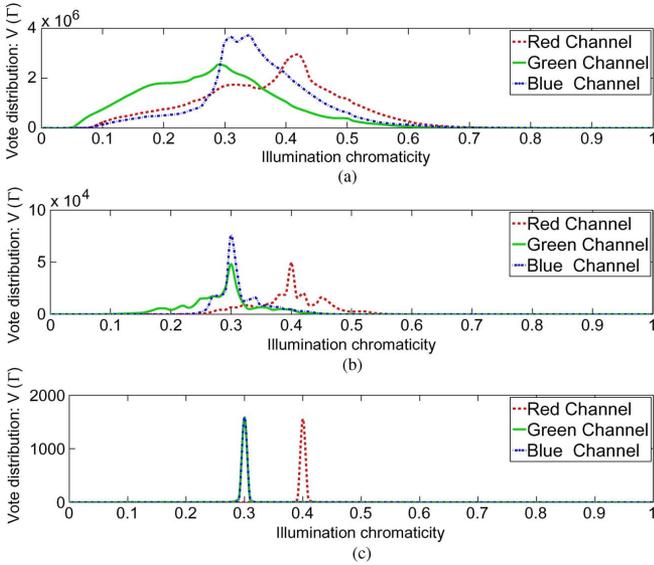


Fig. 2. Illumination chromaticity vote distributions for the synthetic images in Fig. 1. (a)–(c) Illumination chromaticity vote distributions with three support window sizes:  $1 \times 1$ ,  $3 \times 3$ , and  $5 \times 5$ . The blue angle numbers in parentheses under (a)–(c) are the angular errors of the estimated illumination chromaticity. (a)  $1 \times 1$  ( $2.9^\circ$ ). (b)  $3 \times 3$  ( $0.1^\circ$ ). (c)  $5 \times 5$  ( $0.007^\circ$ ).

a pixel in image  $I$ ,  $\bar{p} = p + D(p)$  be its corresponding pixel in  $J$  at a relative location given by the correspondence vector  $D(p)$ , and  $N(p)$  be the support window for pixel  $p$ . The dissimilarity between the pixels  $p$  and  $\bar{p}$  is then measured by aggregating raw matching costs inside the support window  $N(p)$

$$E(p, D(p)) = \sum_{q \in N(p)} \sum_{c \in \{R, G, B\}} |I_c(q) - J_c(q + D(p))|. \quad (19)$$

Given pixel  $p$ , the correspondence hypothesis ( $D(p)$ ) that corresponds to the smallest dissimilarity values ( $E(p, D(p))$ ) is

accepted as correct. For a rectified stereo image pair, the corresponding pixels lie in the same horizontal scanline, and the shifted amount is called the disparity in stereo vision. The correspondence vectors are then represented as a 2-D image called disparity map. The disparity value ( $z$ ) and the depth value ( $d$ ) of a pixel are related by the product of the focal length ( $f$ ) and the baseline ( $b$ ) between the two cameras

$$z \cdot d = f \cdot b. \quad (20)$$

**TC** is invalid for specular pixels because the highlight shifts as the camera moves, that is,  $I_c(p) \neq J_c(p + D(p))$  when either pixel  $p$  in image  $I$  or its correspondence  $p + D(p)$  in image  $J$  is specular. Nevertheless, Section III shows that for specular pixels, chromaticity match computed from the correct correspondence is the same as the estimated illumination chromaticity:  $M_c(p, D(p)) = \Gamma_c$ . This property is defined as illumination chromaticity constancy (**ICC**) in Section II and can be used to match pixels inside specular highlights. However, **ICC** is invalid for diffuse pixels.

**TC** and **ICC** can be integrated in some manner to match both diffuse and specular pixels. Let

$$H(q, D(p)) = \sqrt{\frac{1}{3} \sum_{c \in \{R, G, B\}} (M_c(q, D(p)) - \Gamma_c)^2}. \quad (21)$$

The integration is obtained by redefining the dissimilarity in (19) as

$$E(p, D(p)) = \sum_{q \in N(p)} H(q, D(p)) \cdot \sum_{c \in \{R, G, B\}} |I_c(q) - J_c(q + D(p))| \quad (22)$$

when the three channels of  $M_c(q, D(p))$  are all positive or all negative and  $H(q, D(p)) > 0.1$  (to exclude pixels that have high confidence to be diffuse).

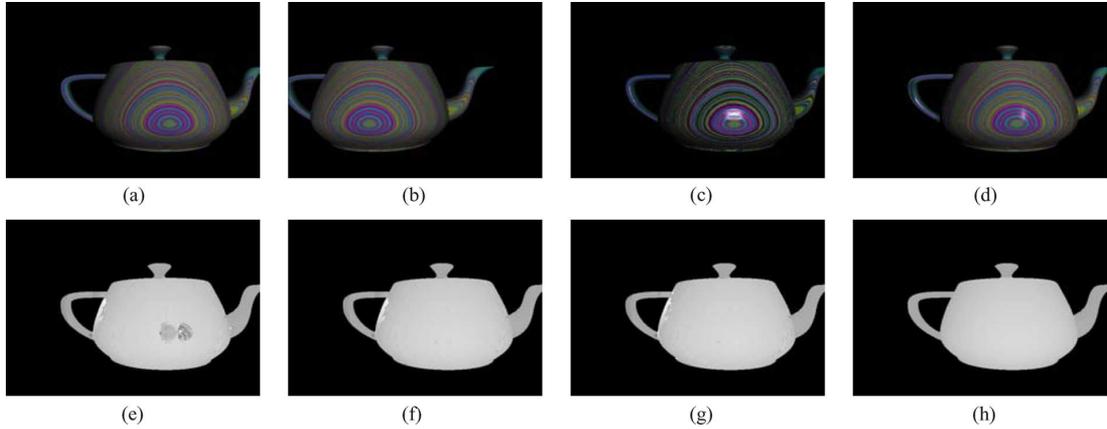


Fig. 3. Stereo matching and highlight removal for the synthetic images. (a) and (b) Ground-truth diffuse reflection of the input images presented in Fig. 1(a) and (b). (c) and (d) Extracted diffuse reflections using the method presented in [6] and our method, respectively. (e)–(g) Estimated disparity maps with the ground truth in (h). (e) Disparity map obtained from standard stereo matching method (using TC). (f) Disparity map obtained from standard stereo matching method but uses the ground-truth diffuse reflections presented in (a) and (b) as input. (g) Disparity map estimated from the proposed stereo matching method which integrates TC and ICC. The percentage of the bad pixels in (e)–(g) are 5.41%, 4.29%, and 4.26%. The images are gamma corrected for better illustration. (a) G.T. left diffuse. (b) G.T. right diffuse. (c) Diffuse from [6]. (d) Our diffuse. (e) Disparity (TC). (f) Disparity (TC + a + b). (g) Disparity (TC + ICC). (h) G.T. disparity.

Finally, we relate the correct correspondences to the diffuse components of the specular pixels by assuming that the highlights in the two images do not spatially overlap. Hence, the diffuse component of a pixel in the highlight can be extracted by finding its corresponding diffuse pixel in the other view.

One problem associated with our method is that the method is invalid for specular pixels that are saturated. Hence, the saturated pixels are treated as outliers in this paper. Specifically, if only two images are used, the matching cost of the saturated pixels will be set to zero ( $E(p, D(p)) = 0$ ) for all possible depth hypothesis/correspondence vectors  $D(p)$ , and the depth information of the nonsaturated pixels can then be propagated to the saturated pixels during depth optimization process. However, if more images are used, we may simply obtain the depth values of the saturated pixels from the depth estimates on other images viewed from other directions/positions, as the highlight moves as the view direction/position changes. Additionally, the stereo reconstruction quality of our method decreases as the areas of the overlapping highlights increase. This is because the color differences of the specular pixels and their correspondences inside the overlapping highlights will be small, and the precision of chromaticity match [ $M_c(p, D(p))$ , (3)] will decrease due to quantization error (8-b images are used in our experiments). But note that our stereo matching method does not require the correspondence of a specular pixel to be diffuse. In fact, our method requires that their color difference be large enough to avoid the quantization noise arising from the computation of chromaticity match. However, if the color difference is very small, TC will be valid for both specular and diffuse pixels.

These difficulties can be greatly reduced if images captured from many viewing directions/positions are available. From these images, coarse estimates can be obtained from each selected stereo pair and then fused to give coherent estimates. In this paper, the coarse depth maps are fused using the efficient method presented in [19]. The fused depth maps are then used to remove highlights. Each pixel  $p$  in each image is projected to the other images using its depth value obtaining by the colors

of its correspondences  $\bar{p}_i$  in the other images. If the luminance of  $\bar{p}_i$  is much larger than the luminance of  $p$  (if the difference is larger than a constant, 10 in our experiments),  $\bar{p}_i$  is treated as a specular pixel. The median value of the colors of all the correspondences  $\bar{p}_i$  (each color band is processed separately) believed to be diffuse are then selected as the correct diffuse reflection of pixel  $p$ . We use the median values to make sure that the diffuse reflections are consistent when viewed from different directions/locations.

## V. EXPERIMENTAL RESULTS

<sup>1</sup> To evaluate our method, we conduct experiments on a synthetic data set and several real images captured by a Sony DFW-X700 camera with gamma correction off. To quantitatively evaluate the proposed illumination chromaticity estimation method for real scenes, we compare the results with the average value of the image chromaticity of a white reference image captured by the same camera. Specifically, we cast light on white printing paper and then captured it using the same camera under the same setting. We also compare our results with the methods presented in [3] and [14]. For highlight removal, we compare our results with images captured with polarizing filters over the camera and the light source. Comparison with the highlight removal method presented in [6] is also provided. [6] assumes that the illumination chromaticity is known, thus, in our experiments, the measured illumination chromaticity (measured with a white reference) is used.

### A. Synthetic Data Set

Fig. 1 visually compares our illumination chromaticity estimation method with [3] and [14]. The blue angle numbers in parentheses under (d)–(h) are the angular errors of estimated illumination chromaticities, which numerically prove that with a  $5 \times 5$  support window, our method can obtain the most accurate estimates. Fig. 2 presents the corresponding vote distributions

<sup>1</sup>Figs. 1, 3, 4, 5, 6, 7, 9, and 10 are gamma corrected for better illustration.

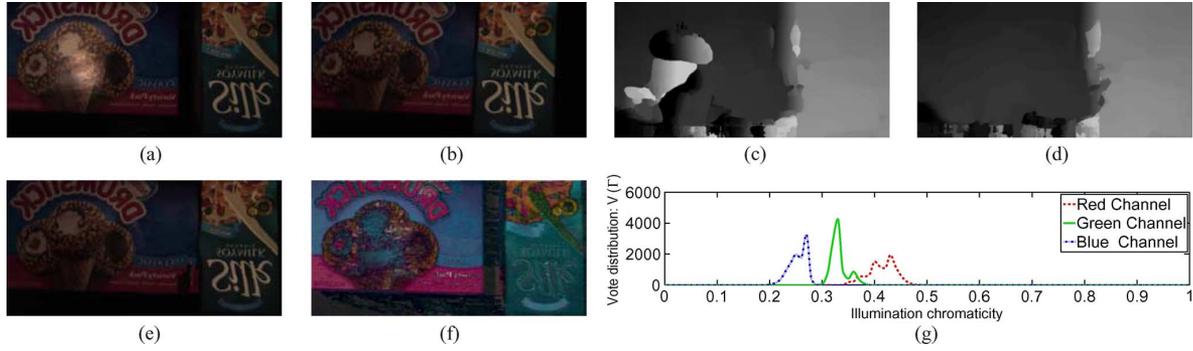


Fig. 4. Binocular stereo matching. (a) and (b) Input images. (c) Disparity map estimated from standard stereo matching method using TC. (d) Disparity map estimated from the proposed method which integrates TC and ICC. (e) and (f) Recovered diffuse reflection using our method and [6], respectively. (g) Illumination chromaticity vote distribution obtained using a  $5 \times 5$  support window. The images are gamma corrected for better illustration. (a) Left. (b) Right. (c) Disparity (TC). (d) Disparity (TC + ICC). (e) Diffuse. (f) Diffuse from [6]. (g)  $5 \times 5$  ( $2.20^\circ$ ).

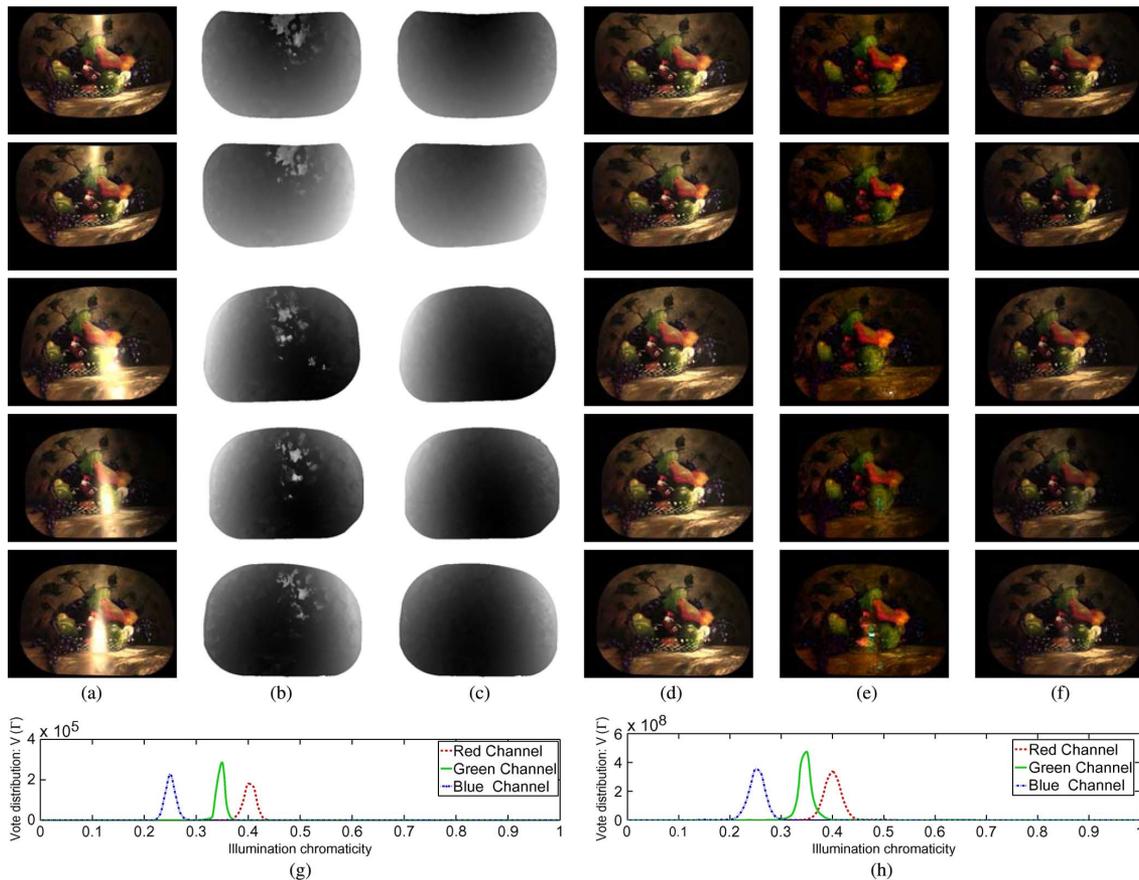


Fig. 5. Multiview stereo matching. (a) Five of the input images. (b) Depth map obtained from standard method (using TC). (c) Depth map obtained from our method (integrating TC and ICC). (d) Diffuse reflection obtained from our method [using the depth map in (c)]. (e) Diffuse reflection estimated using the method presented in [6]. (f) Measured diffuse reflection. (g) Illumination chromaticity vote distribution using the first two images in (a). (h) Illumination chromaticity vote distribution using all the five images in (a). The angular errors of (g) and (h) are  $2.27^\circ$  and  $1.97^\circ$ , respectively. The images are gamma corrected for better illustration. (a) Input. (b) Depth (TC). (c) Depth. (d) Diffuse. (e) Diffuse [6]. (f) G.T. Diffuse. (g)  $5 \times 5$  (using two images,  $2.27^\circ$ ). (h)  $5 \times 5$  (using five images,  $1.97^\circ$ ).

for our method with three support window sizes:  $1 \times 1$ ,  $3 \times 3$  and  $5 \times 5$ . Note that the vote values decrease dramatically from window size  $1 \times 1$  to  $3 \times 3$  due to resulting ability to identify inconsistent correspondences (with dissimilar local chromaticity matches).

Fig. 3 presents the results for the depth estimation and highlight removal methods using the stereo pair presented in either Fig. 1(a), (b) or Fig. 3(a), (b). Fig. 3(a) and (b) are the

ground-truth diffuse reflections, while (c) and (d) are the estimated diffuse reflections using the method presented in [6] and our method. The disparity map obtained with our method [Fig. 3(g)] is used to compute the diffuse reflection in (d). As can be seen, [6] is invalid for this data set, as the highlights in (c) are not removed. Also, the colors in (c) are different from the ground truth in (a). Our results is better than [6], as can be seen in (d). However, some of the specularities are not removed due to the

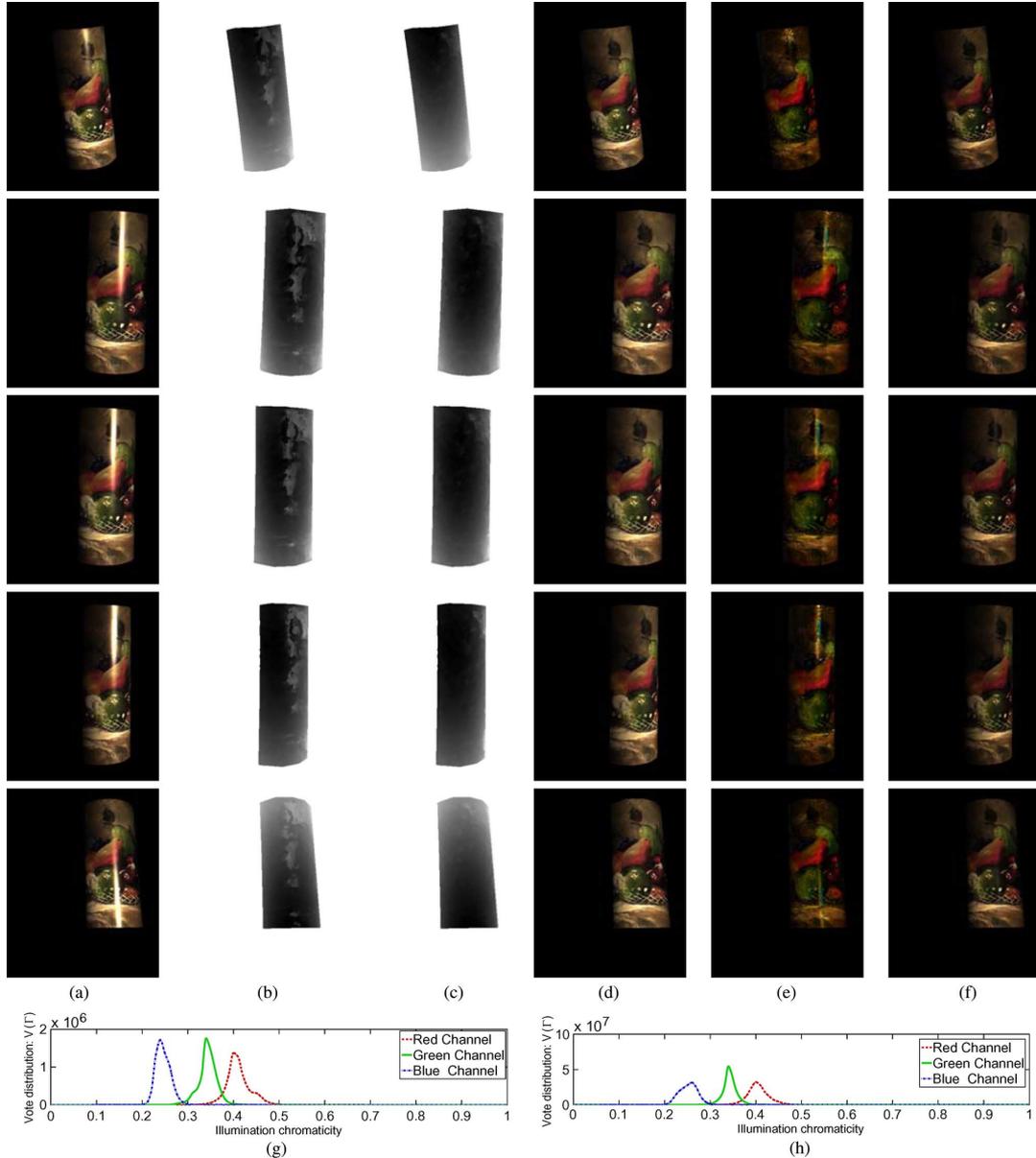


Fig. 6. Multiview stereo matching. (a) Five of the input images. (b) Depth map obtained from standard method (using TC). (c) Depth map obtained from our method (integrating TC and ICC). (d) Diffuse reflection obtained from our method [using the depth map in (c)]. (e) Diffuse reflection estimated using the method presented in [6]. (f) Measured diffuse reflection. (g) Illumination chromaticity vote distribution using the first two images in (a). (h) Illumination chromaticity vote distribution using all the five images in (a). The angular errors of (g) and (h) are  $2.81^\circ$  and  $1.24^\circ$ , respectively. The images are gamma corrected for better illustration. (a) Input. (b) Depth (TC). (c) Depth. (d) Diffuse. (e) Diffuse [6]. (f) G.T. Diffuse. (g)  $5 \times 5$  (using two images,  $2.81^\circ$ ). (h)  $5 \times 5$  (using five images,  $1.24^\circ$ ).

violation of the assumption that the highlights in the two images should not spatially overlap. Nevertheless, our stereo matching method is robust to the violation of this assumption, as can be seen in (g). Visual comparison with the ground-truth disparity map in (h) shows that the estimated disparity values of the specular pixels are correct in (g). (e) is the disparity map obtained from the standard stereo matching method (using TC). The disparity map in (f) is also estimated using the standard method, but the input images are free of specularities, which means that the ground-truth diffuse reflections in (a) and (b) are used as input. Fig. 3(e), (f) shows that this standard method is invalid for specular highlights. (g) is the disparity map estimated from the proposed stereo matching method which integrates TC and ICC.

Note that (g) is visually very similar to (f). Let a pixel be a bad pixel if the difference of the estimated disparity value and the ground truth is larger than 1 [20]. The percentages of bad pixels in (e)–(g) are 5.41%, 4.29%, and 4.26%, which shows that our method improves the reconstruction accuracy of this synthetic data set.

### B. Real Data Sets

We next present our experimental results on real data sets. Fig. 4 provides the experimental results with a real stereo image pair. The illumination chromaticity vote distribution with a  $5 \times 5$  support window is presented in Fig. 4(g), and the angular errors are  $2.20^\circ$  [blue angle number in parentheses under (g)]. The

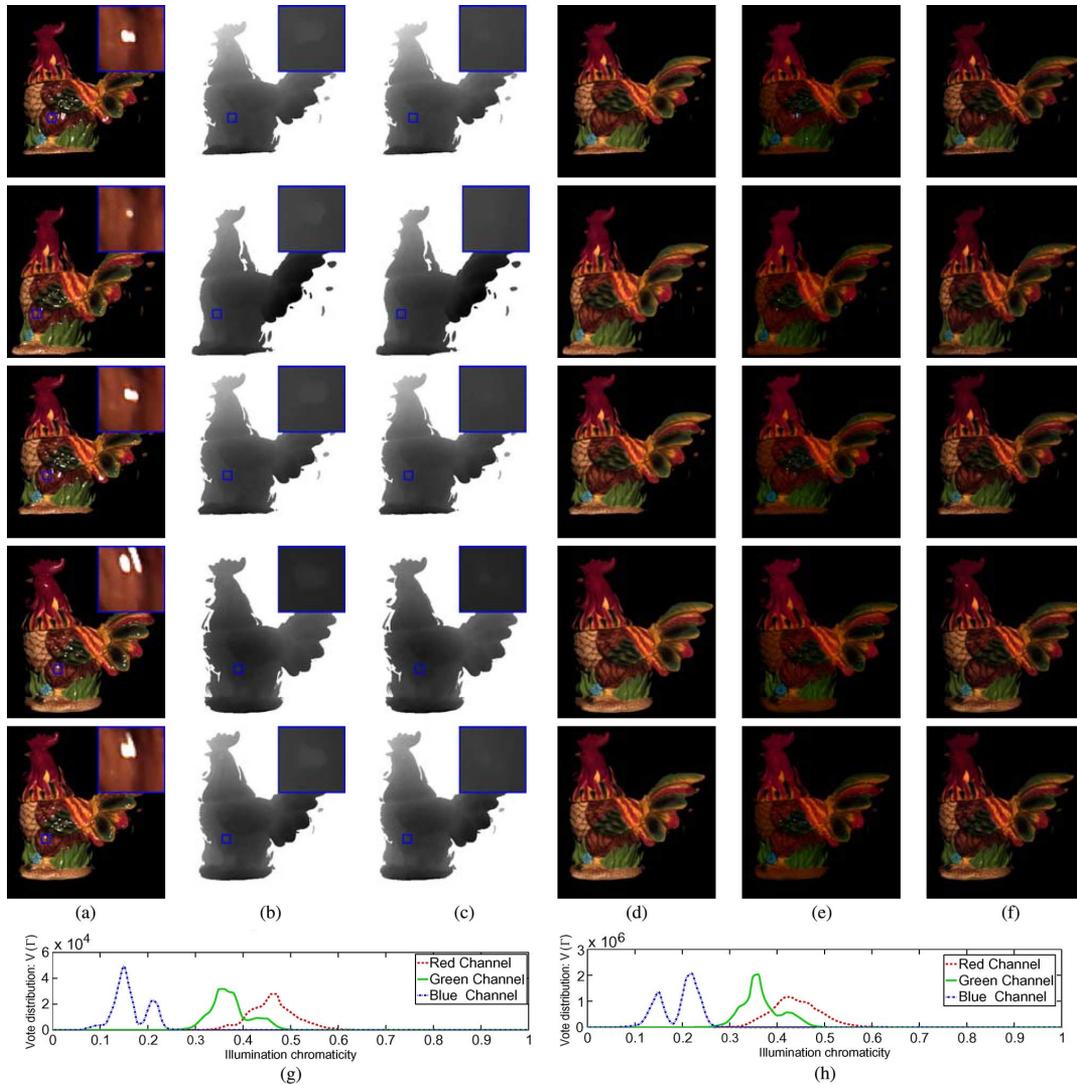


Fig. 7. Multiview stereo matching. (a) Five of the input images. (b) Depth map obtained from standard method (using TC). (c) Depth map obtained from our method (integrating TC and ICC). (d) Diffuse reflection obtained from our method [using the depth map in (c)]. (e) Diffuse reflection estimated using the method presented in [6]. (f) Measured diffuse reflection. (g) Illumination chromaticity vote distribution using the first two images in (a). (h) Illumination chromaticity vote distribution using all the five images in (a). The angular errors of (g) and (h) are  $12.97^\circ$  and  $5.74^\circ$ , respectively. The images are gamma corrected for better illustration. (a) Input. (b) Depth (TC). (c) Depth. (d) Diffuse. (e) Diffuse [6]. (f) G.T. Diffuse. (g)  $5 \times 5$  (using two images,  $12.97^\circ$ ). (h)  $5 \times 5$  (using five images,  $5.74^\circ$ ).

estimated disparity map presented in (d) shows that our method is able to remove the matching errors in (c) due to specular highlights. The estimated diffuse reflections presented in (e) and (f) show that our highlight removal method is more robust than the single-view-based highlight removal method presented in [6], as the highlights in (f) are not removed. Also, Fig. 4(f) shows that [6] is invalid for neutral pixels, and the estimated diffuse colors are incorrect. However, our highlight removal method obtained incorrect diffuse colors around the half-occluded regions due to the incorrect correspondences. This problem can be greatly reduced when more images are used. Figs. 5–7 present the experimental results with multiple images. Specifically, Figs. 5 and 6 used 21 images, and Fig. 7 used 15 images.

The angular errors of the estimated illumination chromaticity of the three real data sets (Figs. 5–7) using only two of the input images are  $2.27^\circ$ ,  $2.81^\circ$  and  $12.97^\circ$ . The last data set (Fig. 7) has large angular error, as the highlights are very sparse and

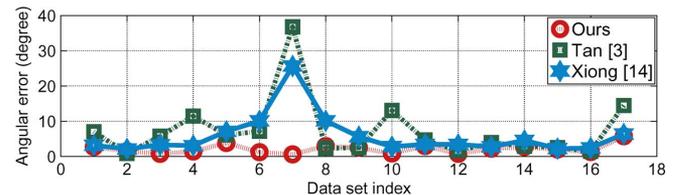


Fig. 8. Quantitative comparison of the illumination chromaticity estimation methods. The red curve corresponds to the angular difference between the measured illumination chromaticity (measured using a white reference) and the estimated illumination chromaticity using our method. The green and blue curves correspond to the methods presented in [3] and [14], respectively. As can be seen, our method is generally more robust than [3] and [14] for these data sets.

most of them are saturated. However, using more images can reduce the estimation error. The angular errors of the three real data sets using five of the input images [presented in Fig. 5(a), Fig. 6(a), and Fig. 7(a)] are  $1.97^\circ$ ,  $1.24^\circ$ , and  $5.74^\circ$ . As can be

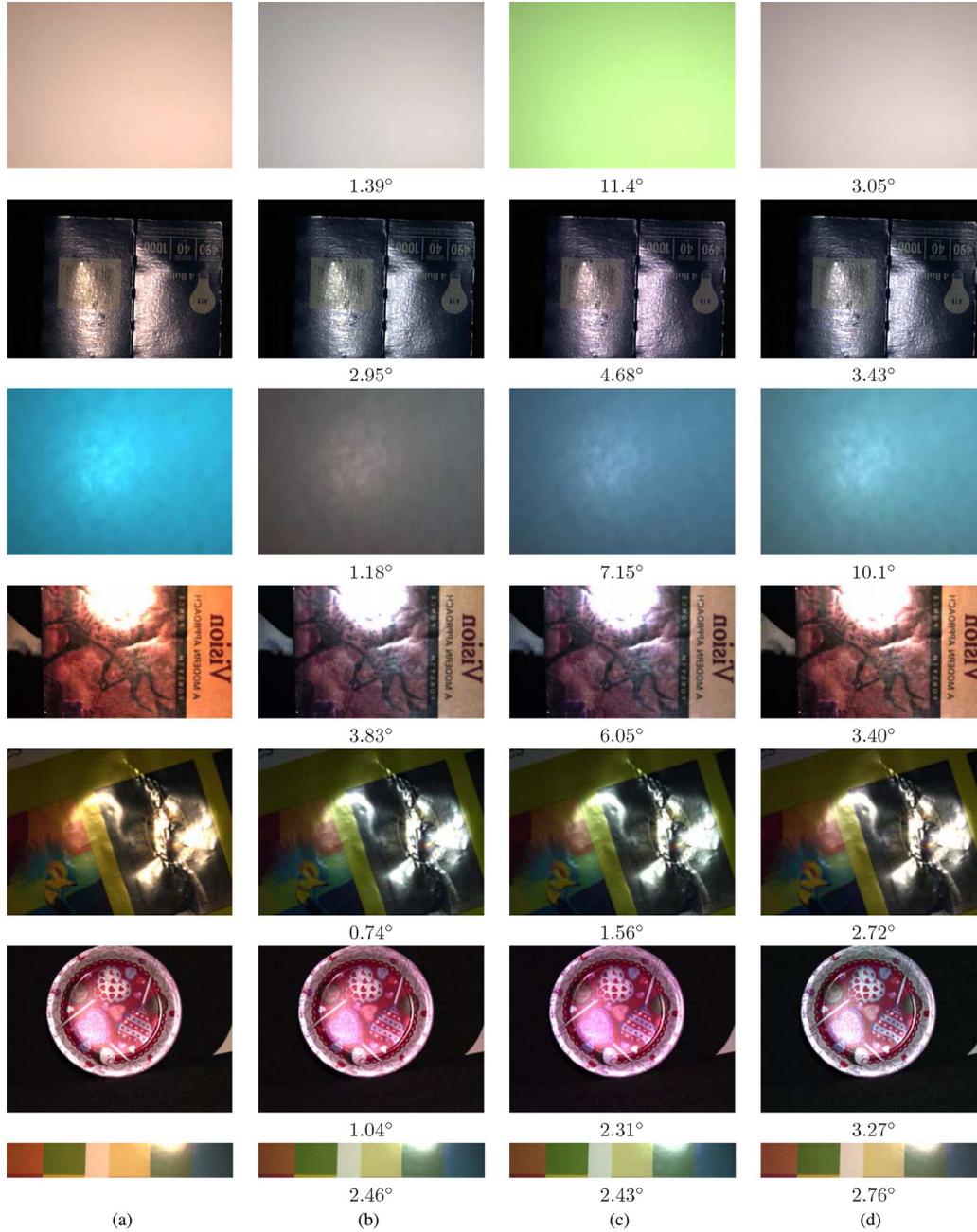


Fig. 9. Illumination chromaticity estimation for real scenes with corresponding angular errors. From (a) to (d): reference images, normalized images using our method, normalized images using [3], and normalized images using [14]. The images are gamma corrected for better illustration. (a) Input. (b) Ours. (c) [3]. (d) [14].

seen, the estimation error drops with the increasing number of input images.

Visual comparison of the depth maps presented in Fig. 5(b), (c) and Fig. 6(b), (c) shows that our stereo matching can greatly improve the reconstruction accuracy for specular pixels. However, since the specular highlights in Fig. 7(a) are very sparse, the depth maps estimated using our method [Fig. 7(c)] are only slightly better than the standard method [Fig. 7(b)].

Fig. 5(d)–(f), Fig. 6(d)–(f), and Fig. 7(d)–(f) visually compare the estimated diffuse reflections using our method, the method presented in [6], and the measured diffuse reflection. Visual comparison shows that our method outperforms [6], as

the colors of estimated diffuse reflections using [6] are incorrect [see Fig. 5(e), Fig. 6(e), and Fig. 7(e)], and [6] is invalid for saturated pixels.

To evaluate the robustness of the proposed illumination chromaticity estimation method, we conduct experiments on a total of 17 real data sets. Besides the four data sets presented in Figs. 4–7), another 13 real data sets are provided in Figs. 9 and 10. The illumination chromaticities of the real illuminants are grouped into five different sets using the image chromaticities of the white reference:  $\{0.13158, 0.40248, 0.46829\}$ ,  $\{0.38781, 0.32666, 0.28822\}$ ,  $\{0.44792, 0.31771, 0.23437\}$ ,  $\{0.61730, 0.36693, 0.01197\}$  and  $\{0.40070, 0.33177, 0.26676\}$ . We calculate the estimation errors by comparing the chromaticity

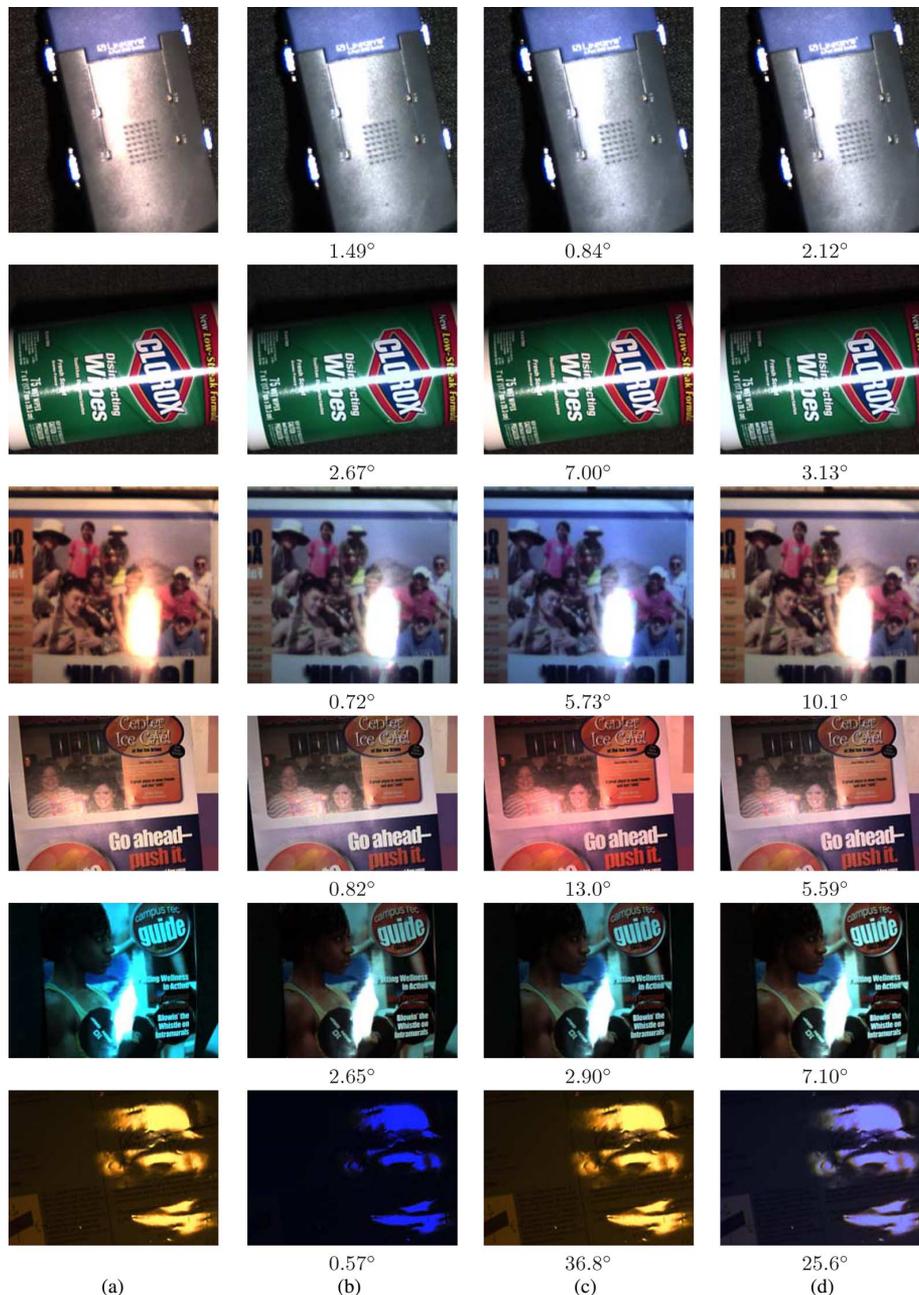


Fig. 10. Illumination chromaticity estimation for real scenes with corresponding angular errors. From (a) to (d): reference images, normalized images using our method, normalized images using [3], and normalized images using [14]. The images are gamma corrected for better illustration. (a) Input. (b) Ours. (c) [3]. (d) [14].

TABLE I  
SUMMARY OF THE QUANTITATIVE EVALUATION OF THE PROPOSED ILLUMINATION CHROMATICITY ESTIMATION APPROACH FOR REAL DATA SETS

Algorithm	Mean angular error	Minimum angular error	Maximum angular error	Std. dev. of angular error
Ours	1.98°	0.57°	5.74°	1.34°
[3]	7.31°	0.84°	36.8°	8.64°
[14]	5.75°	2.12°	25.6°	5.72°

estimates with those of the white reference. The angular errors are shown in Fig. 8, and a summary of the experimental results is provided in Table I. The results from the methods presented in [3] and [14] are also included. The error rates demonstrate

that our method is generally robust, and our method appears better than [3] and [14] on average for these data sets.

## VI. CONCLUSION

A new invariant called illumination chromaticity constancy for matching highlights between images is introduced in the paper. Algorithms are presented that use this invariant for three vision problems: illumination chromaticity estimation, correspondence searching, and specularity removal. In relation to previous approaches, the most significant advantage of the presented method is that we have related the correct correspondence vectors to both the illumination chromaticity and the diffuse components of the specular pixels and that we have presented an attempt to estimate these properties in a

uniform framework. Additionally, our method does not require detecting the specular highlights. However, if the illuminant and the object surface have the same chromaticity, the proposed correspondence matching and highlight removal method will fail, as every possible chromaticity match will be equal to the illumination chromaticity. Nevertheless, the proposed illumination chromaticity estimation method has no problem under this condition. Also, our framework assumes chromatic surfaces and is invalid for grayscale objects.

## REFERENCES

- [1] H. C. Lee, "Method for computing the scene-illuminant chromaticity from specular highlights," *COLOR*, pp. 303–308, 1992.
- [2] G. Finlayson and G. Schaefer, "Solving for colour constancy using a constrained dichromatic reflection model," *Int. J. Comput. Vis.*, vol. 42, pp. 127–144, 2001.
- [3] R. Tan, K. Nishino, and K. Ikeuchi, "Illumination chromaticity estimation using inverse-intensity chromaticity space," in *Proc. IEEE Comput. Soc. Comput. Vis. Pattern Recognit.*, 2003, vol. I, pp. 673–680.
- [4] S. Lin and H. Y. Shum, "Separation of diffuse and specular reflection in color images," in *Proc. IEEE Comput. Soc. Comput. Vis. Pattern Recognit.*, 2001, pp. 341–346.
- [5] R. Tan and K. Ikeuchi, "Reflection components decomposition of textured surfaces using linear basis functions," in *Proc. IEEE Comput. Soc. Comput. Vis. Pattern Recognit.*, 2005, vol. I, pp. 125–131.
- [6] R. Tan and K. Ikeuchi, "Separating reflection components of textured surfaces using a single image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 178–193, Feb. 2005.
- [7] P. Tan, L. Quan, and S. Lin, "Separation of highlight reflections on textured surfaces," in *Proc. IEEE Comput. Soc. Comput. Vis. Pattern Recognit.*, 2006, pp. 1855–1860.
- [8] S. P. Mallick, T. E. Zickler, D. J. Kriegman, and P. N. Belhumeur, "Beyond lambert: Reconstructing specular surfaces using color," in *Proc. IEEE Comput. Soc. Comput. Vis. Pattern Recognit.*, 2005, vol. II, pp. 619–626.
- [9] S. P. Mallick, T. Zickler, P. N. Belhumeur, and D. J. Kriegman, "Specularity removal in images and videos: A pde approach," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 550–563.
- [10] T. Zickler, S. P. Mallick, D. J. Kriegman, and P. N. Belhumeur, "Color subspaces as photometric invariants," in *Proc. IEEE Comput. Soc. Comput. Vis. Pattern Recognit.*, 2006, pp. 2000–2010.
- [11] K. J. Yoon and I. S. Kweon, "Correspondence search in the presence of specular highlights using specular-free two-band images," in *Proc. Asian Conf. Comput. Vis.*, 2006, pp. 761–770.
- [12] A. Blake, "Specular stereo," in *Proc. IJCAI*, 1985, pp. 973–976.
- [13] G. Brelstaff and A. Blake, "Detecting specular reflections using lambertian constraints," in *Proc. Int. Conf. Comput. Vis.*, 1988, pp. 297–302.
- [14] W. Xiong and B. Funt, "Stereo retinex," *Image Vis. Comput.*, vol. 27, pp. 178–188, 2009.
- [15] B. Funt, F. Ciurea, and J. McCann, "Retinex in matlab," *J. Electron. Imag.*, pp. 48–57, 2004.
- [16] S. Lin, Y. Li, S. Kang, X. Tong, and H. Y. Shum, "Diffuse-specular separation and depth recovery from image sequences," in *Proc. Eur. Conf. Comput. Vis.*, 2002, pp. 210–224.
- [17] D. Bhat and S. Nayar, "Stereo in the presence of specular reflection," in *Proc. Int. Conf. Comput. Vis.*, 1995, pp. 1086–1092.
- [18] S. Shafer, "Using color to separate reflection components," *Color Res. App.*, vol. 10, pp. 210–218, 1985.
- [19] P. Merrell, A. Akbarzadeh, L. Wang, P. Mordohai, J. M. Frahm, R. Yang, D. Nistér, and M. Pollefeys, "Real-time visibility-based fusion of depth maps," *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 1–8, 2007.
- [20] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, pp. 7–42, 2002.



**Qingxiang Yang** (S'07–M'07) received the B.E. degree from University of Science and Technology of China (USTC), Hefei, China, in 2004, the M.Sc. degree in computer science from University of Kentucky, Lexington, KY, in 2007, and is currently pursuing the Ph.D. degree in the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL.

His research interests reside in computer vision and graphics, especially on stereo vision, reflectance, illumination, and color.



**Shengnan Wang** (S'10) received the B.E. degree from University of Science and Technology of China (USTC), Hefei, China, in 2007, the M.Sc. degree in computer science from the University of Wisconsin, Madison, WI, in 2009, and is currently pursuing the Ph.D. degree in the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL.

Her research interests reside in computer vision and graphics.



**Narendra Ahuja** (S'79–M'79–SM'85–F'92) received the Ph.D. degree from the University of Maryland, College Park, MD, in 1979.

He is currently a Professor in the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory (CSL), University of Illinois at Urbana-Champaign, Urbana, IL, and a full-time faculty member in the Beckman Institute Artificial Intelligence Group, University of Illinois at Urbana-Champaign. His research interests are next generation cameras, 3-D computer vision, video analysis, image analysis, pattern recognition, human computer interaction, image processing, image synthesis, and robotics.

Dr. Ahuja has received numerous honors which include the Donald Biggar Willet Professorship, University of Illinois at Urbana-Champaign College of Engineering (1999), on Incomplete List of Teachers Ranked Excellent by Their Students (2002), University of Illinois at Urbana-Champaign Campus Award for Guiding Undergraduate Research-Honorable Mention (1999), Emanuel R. Piore Award, IEEE (1999), SPIE Technology Achievement Award (1998). He is a Fellow of the ACM, the AAAS, the AAAI, the SPIE, and the IAPR. He is a Beckman associate, University of Illinois at Urbana-Champaign Center for Advanced Study (1990–1991 and 1998), and received the University of Illinois at Urbana-Champaign University Scholar Award (1985), and the Presidential Young Investigator Award (1984).

Dr. Ahuja has received numerous honors which include the Donald Biggar Willet Professorship, University of Illinois at Urbana-Champaign College of Engineering (1999), on Incomplete List of Teachers Ranked Excellent by Their Students (2002), University of Illinois at Urbana-Champaign Campus Award for Guiding Undergraduate Research-Honorable Mention (1999), Emanuel R. Piore Award, IEEE (1999), SPIE Technology Achievement Award (1998). He is a Fellow of the ACM, the AAAS, the AAAI, the SPIE, and the IAPR. He is a Beckman associate, University of Illinois at Urbana-Champaign Center for Advanced Study (1990–1991 and 1998), and received the University of Illinois at Urbana-Champaign University Scholar Award (1985), and the Presidential Young Investigator Award (1984).



**Ruigang Yang** (S'02–M'03) received the M.S. degree in computer science from Columbia University, New York, in 1998, and the Ph.D. degree in computer science from the University of North Carolina, Chapel Hill, NC, in 2003.

He is currently an Associate Professor in the Computer Science Department, University of Kentucky, Lexington, KY. His research interests include computer graphics, computer vision, and multimedia.

Dr. Yang is a recipient of the U.S. NSF CAREER Award in 2004 and is a Member of the IEEE Com-

puter Society and ACM.