

Special Section on CAD &amp; Graphics 2019

## Intrinsic color correction for stereo matching

Qing Ran, Wenjing Zhao, Jieqing Feng\*

State Key Laboratory of CAD&amp;CG, Zhejiang University, Hangzhou 310058, China



### ARTICLE INFO

#### Article history:

Received 9 March 2019

Revised 24 April 2019

Accepted 18 May 2019

Available online 21 May 2019

#### Keywords:

Color correction

Intrinsic decomposition

Stereo matching

### ABSTRACT

The improvement of color similarity between stereo images can bring better performance to stereo matching algorithms. For this purpose, we present a color correction method to alleviate the color discrepancy between a pair of stereo images, so that the color appearance of one image, i.e., the target image, is consistent with the other image, i.e., the source image. Our method starts with decomposing both the target image and the source image into two intrinsic layers, i.e., the reflectance layer and the shading layer, using intrinsic decomposition. The purpose of intrinsic decomposition is to distinguish and then process different color discrepancies caused by different factors (shading and reflectance) separately and appropriately. Then, a practical and effective consistent segmentation algorithm, which applied to the original stereo images, is proposed to establish the region correspondences. Next, luminance correction method and color correction method, based on the local region correspondences, are adopted to correct the shading layer and the reflectance layer in the target image, respectively, making them as similar as possible to those of the source image. Eventually, the two corrected layers of the target image are combined to yield the final corrected image. The experimental results demonstrate that the proposed method, which using intrinsic decomposition to handle color discrepancy caused by different factors, not only enhances the visual color similarity between the stereo images but also improves the accuracy of their stereo matching.

© 2019 Elsevier Ltd. All rights reserved.

## 1. Introduction

Stereo matching, the purpose of which is the inference of disparity maps from stereo images, is of great importance in many areas, such as autopilot, robotics and augmented reality, etc. Most stereo matching methods depend on computing a matching cost to measure the similarity between two pixels in two stereo images. Typically, the matching cost is computed under the assumption that a pair of stereo images is radiometrically similar, i.e., the corresponding pixels in two images have similar colors. However, in practice, the colors of corresponding pixels may be quite different due to various inevitable factors, such as camera device changes, different illumination conditions or failure to meet the diffuse surface conditions. Although some robust matching cost functions, such as mutual information [1,2], census transform [3], and recent matching cost from deep learning [4], account for color discrepancy to some extent. Few of them can handle strong radiometric changes [5], such as those caused by highly different exposures of the images. Thus, to reduce the ambiguities in stereo matching,

color correction can be used as a preprocessing step to compensate the color discrepancy between stereo images.

The existing color correction methods can be classified into two categories [6]: parametric methods and non-parametric methods. The former is better than the latter in terms of extensibility [7]. Generally, parametric methods explicitly represent the relationship between the target image and the source image using the transformation  $I_s = M * I_t$ , where  $M$  represents the mapping of the three color channels, and  $I_s$  and  $I_t$  represent the source image and the target image, respectively. The parametric methods can be further divided into global parametric methods and local parametric methods. Global parametric methods generally provide a uniform mapping  $M$  between the colors of two images;  $M$  may be a diagonal matrix, an affine matrix or an arbitrary  $3 \times 3$  matrix, which corresponds to the diagonal model, the affine model and the linear model, respectively. However, local regions in the scene may have different color mappings; thus, global methods may fail to address these complicated cases. To address non-uniform color mappings, local methods are developed to produce a more elaborate mapping by transforming colors in the corresponding regions between a pair of images, for which establishing accurate regional correspondences is crucial.

Pixel values encode all the properties of the scene it captures. Brainard and Wandell [8] proposed that the distinct scene

\* Corresponding author.

E-mail address: [jqfeng@cad.zju.edu.cn](mailto:jqfeng@cad.zju.edu.cn) (J. Feng).

properties can be represented as a set of separate “intrinsic images”, which are decomposed from the original image. Currently, the common intrinsic decomposition methods focus on one particular case: the decomposition of an image into a reflectance layer and a shading layer, i.e., the surface reflectance and the spectral power distribution of the ambient light [9]. These two layers jointly reproduce the original RGB image.

Most intrinsic decomposition algorithms are based on the Lambertian model hypothesis. Under this model, the surface of an object is regarded as an ideal diffusion surface that appears the same from all directions. Thus, applying ideal intrinsic decomposition to a pair of stereo images with color discrepancy, the corresponding points in the resulting reflectance layers should have the same values. This indicates that the color discrepancy is mainly caused by the luminance changes in the shading layers. Thus, the benefit of intrinsic decomposition is that the initial color discrepancy could be eliminated by correcting luminance difference in the shading layers ideally. However, due to the intrinsic decomposition is an ill-posed problem, so far, it is hard to get the same reflectance between the stereo images by most of the state-of-art decomposition methods. On the other hand, a small amount of color discrepancy, caused by the object material, orientation, distance, etc., still exists in the decomposed reflectance layer. Hence, to make corrected image as similar as possible with the target one, the correction method should be performed on two intrinsic layers respectively. Note that both layers are approximately uncorrelated, so the corrected layers can faithfully reproduce an RGB image without introducing blending artifacts.

This paper proposes an intrinsic color correction method, which is applied to a pair of stereo images to improve the accuracy of stereo matching. Our method begins with applying intrinsic decomposition to both the target image and the source image. Then, to address non-uniform color mapping between two stereo images, a consistent segmentation algorithm is proposed to establish the region correspondences, which is based on the Mean-Shift segmentation and some highly reliable Scale-Invariant Feature Transform (SIFT) feature correspondences. Based on the consistent segmentation results, a local correction algorithm, which correct two intrinsic layers of the source image separately and region-wisely, is then proposed. The local differences of luminance contrast in the shading layer are corrected by applying an affine mapping matrix of luminance statistics for each region. And the color discrepancy in the reflectance layer is compensated by transferring the statistical properties of the colors from the source to the target. Moreover, to eliminate color discontinuities across regions, a weighted correction framework is utilized to generate a smooth corrected result. The two corrected intrinsic layers are then combined into the final corrected result.

The main contributions of this paper are listed below:

- A local color correction method is proposed to alleviate the color discrepancy between two stereo images and then improve the accuracy of their stereo matching. By utilizing the intrinsic decomposition, different color discrepancies caused by different factors (shading and reflectance) can be processed separately and appropriately.
- A practical consistent segmentation method that produce more accurate region correspondences between two stereo images is proposed, which can be also used in other region correspondence establishment between stereo images.

## 2. Related work

*Robust stereo matching.* Some robust stereo matching methods have considered color discrepancy. The AD-census [10] algorithm is one of the best-ranked stereo matching algorithms on the Mid-

dlebury Evaluation Website [11]. Its matching cost is a combination of the absolute difference and the census transform, thus, it is relatively insensitive to radiometric differences. Heo et al. [12] took radiometric variations into account and proposed a new measure, called the Adaptive Normalized Cross-Correlation (ANCC). In their following work [13], they proposed an iterative framework that alternates between stereo matching and color correction, which infers both accurate depth maps and color-consistent stereo images. Hirschmuller [1] described the Semi-Global Matching (SGM) stereo method, which uses Mutual Information based matching cost for compensating radiometric differences of input images. Other methods improved the robustness by introducing geometric information into the support windows [14,15]. Recently, with the process of deep learning, many methods [16,17] learned the disparity result directly without explicit matching cost. However, so far, only stereo image pairs which have similar color performance are used as training datasets.

*Color correction.* Color correction techniques can be classified into model-free non-parametric approaches and model-based parametric approaches according to the survey presented by Xu and Muligan [7]. Most model-free methods [18–22] build a look-up table to record the mapping of the entire range of the color value; the look-up table is usually computed via the corresponding pixel pairs in the overlapping area of two images. However, these methods may not be robust because the image data are prone to be influenced by noise or outliers.

Model-based parametric methods can be categorized into two classes: global approaches and local approaches. Reinhard et al. [23] proposed a global method, which performs color transfers using statistical information. In the un-correlated  $\alpha\beta$  color space, the original colors of the image are corrected by applying a global color transformation, which is derived by matching the means and the standard deviations of the global color distributions of the image to another. Xiao and Ma [24] proposed an approach that directly addresses color correction in correlated color spaces, e.g., the RGB color space. The method treats the color of each pixel as a three-dimensional stochastic variable. The color transformation is computed from a statistical perspective by fitting the random variable distributions of the target image to those of the source image. During the fitting process, the correlation between the three components is measured by the covariance matrix. These global methods are generally based on the assumption that the color mapping between different regions can be described by the same transformation, thus they are not suitable for complex scenes where the color mappings are distinct in different regions. Moreover, models based on weighted affine transformations [25], splines [26], Gaussian mixture models [27], and nonlinear manifold learning approaches [28] have also been proposed.

To address complex scenes, several local methods have been proposed. In the survey of Xu and Mulligan [7], the method proposed by Tai et al. [29] is highly recommended for local color correction for complex scenes. The method segments both source and target images into several regions using a probabilistic segmentation algorithm based on the Gaussian mixture model (GMM). Region correspondences are obtained by mapping the corresponding Gaussian components. However, the GMM model restricts the sensitiveness of segmentation because small regions may not be segmented properly if they do not contain a sufficient number of pixels to form a distinct Gaussian component. Oliveira et al. improved Tai’s method using the mean-shift algorithm; the improved method is less time-consuming and does not require a predefined number of regions [30]. Ly et al. [31] used the marker-controlled watershed transformation method to segment the two images; the method can improve the accuracy of region correspondences near object boundaries.

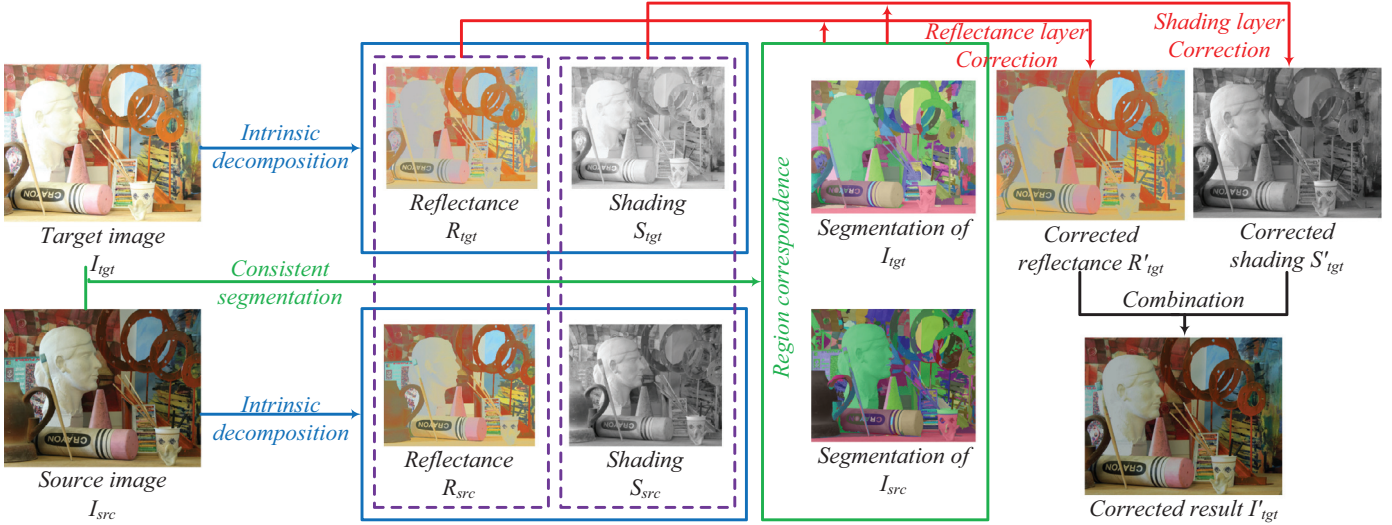


Fig. 1. A flowchart of the proposed intrinsic color correction method.

*Color correction for stereo matching.* Several color correction methods have been proposed to correct colors of stereo images in stereo matching. A region-based correction method is proposed in [32]. Ran et al. [33] propose a local color correction method for alleviating color discrepancy between two stereo images. The region correspondences are generated by projecting the segmentation of the target image onto the source image via homography matrices. Due to occlusion and perspective distortion, the correspondences may be imprecise.

*Intrinsic image decomposition.* Intrinsic image decomposition is an intensively investigated problem in computer graphics and vision. The retinex theory for decomposing an image into two intrinsic layers, of which the large and small intensity gradients are assumed to correspond to reflectance and shading respectively, was proposed by Land and McCann [34]. Grosse et al. [35] implemented the theory and provided a ground-truth dataset for intrinsic image decomposition. The retinex theory works well in many cases but may fail in the presence of strong shadows or smoothly varying planar textures because of the unexpected frequencies of the changes in such areas. Many other algorithms have been proposed under different assumptions. Shen and Yeo assumed that neighboring pixels with similar chromaticities have the same reflectance values [36]. Barron and Malik [37] assumed that the reflectance is piecewise-constant and that certain shading values are more likely to appear than others. Garces et al. [38] assumed that neighboring pixels have similar shading. Bell et al. [39] incorporated all of the above ideas in a global manner; their method can be regarded as one of the state-of-the-art methods. Recently, Beigpour et al. [40] proposed an intrinsic decomposition method, which is driven by multi-view light-field appearance editing. With conducting pre-processing operations (including white balancing and specular removal) and regarding light field as a generalization of stereoscopic images, reflectance layers between views are extracted as similar as possible, which is very impressive.

### 3. The proposed method

Fig. 1 shows the overview of the proposed intrinsic color correction method. We take a pair of stereo images with color discrepancies as input, one as the target image  $I_{tgt}$  (to be corrected), and the other as the source image  $I_{src}$  (unchanged).

The proposed approach consists of four stages. First, the two stereo images are decomposed into their intrinsic layers under the Lambertian Model hypothesis, i.e.,  $I_{tgt}$  is decomposed into

the reflectance layer  $R_{tgt}$  and the shading layer  $S_{tgt}$ , and  $I_{src}$  is decomposed into  $R_{src}$  and  $S_{src}$ . Second, a consistent segmentation algorithm is adopted to establish the consistent region correspondences between the two stereo images. Third, a region-based correction method, which is based on the consistent region correspondences, is applied to the intrinsic layers of the two images with different correction strategies. The corrected reflectance layer and the corrected shading layer are then combined into the final corrected result  $I'_{tgt}$ .

#### 3.1. Intrinsic decomposition

In this stage, a color image  $I$  is decomposed into two intrinsic image layers using the method proposed by Bell et al. [39]. Bell's method incorporates many important priors that have appeared in the literature and can perform intrinsic image decomposition well and effectively for real-world photographs. It attempts to find the optimal three-channel reflectance layer  $R^*$  and the single-channel shading layer  $S^*$  that is the most likely under probability distribution  $p$ :

$$R^*, S^* = \arg \max_{R, S} p(R, S|I) \quad (1)$$

The decomposition result satisfies  $I_i^c = R_i^c \cdot S_i$ , where  $I_i^c$  is the channel  $c \in \{r, g, b\}$  of pixel  $i$ , as well as for  $R_i^c$ . Note that  $S_i$  has only one channel. The probability distribution  $p$  is defined according to a set of priors related to  $R$  and  $S$ , which are based on some assumptions about the nature of the scene and the physics of the imaging process. Considering several important priors in a global sense, Bell's decomposition method is modeled as an energy minimization in a fully connected conditional random fields (CRF) [41]. The energy function consists of two types of terms. The unary term  $\psi_i$  for each pixel is used to make the shading smooth and avoid extreme values. The binary term  $\psi_{ij}$  for each pair of pixels is used to make the reflectance piecewise constant for all pairs of similar pixels. The energy to minimize is defined as follows:

$$E(x) = \underbrace{\omega_r E_r(x)}_{\text{binary } \psi_{ij}} + \underbrace{\omega_s E_s(x) + \omega_t E_t(x)}_{\text{unary } \psi_i} \quad (2)$$

where  $x$  indicates a pixel.  $E_r(x)$  is the binary reflectance term, which is based on the prior that if neighboring pixels have similar chromaticities or intensities, they should also have similar reflectance;  $E_s(x)$  is a unary shading term, which is based on the prior that the shading channel tends to vary smoothly across a



**Fig. 2.** An intrinsic decomposition example of two stereo images (the Art dataset). The layers of reflectance and shading have different degrees of color discrepancies.

smooth surface;  $E_t(x)$  is unary term for each pixel, which makes the optimizer not select extreme values of shading for too many pixels;  $\omega_r$ ,  $\omega_s$  and  $\omega_t$  are the weighting coefficients. The overall energy is minimized by iterative optimization that alternates between reflectance and shading. The reflectance term is optimized using the mean-field approximation to the CRF distribution [41], and the shading term is optimized using an iteratively re-weighted least squares method. After a certain number (empirically, 25) of iterations, both the target image and the source image are decomposed into two intrinsic layers. Other details can be found in [39].

To guarantee the consistency of the color value range in the two stereo images, the reflectance and shading layers are normalized. An example of the intrinsic decomposition is shown in Fig. 2. As shown in this figure, the reflectance image and the shading image have different degrees of color discrepancy; therefore, they should be corrected separately.

### 3.2. Consistent segmentation of stereo images

It is necessary to determine the regions correspondences at first since colors are transferred between corresponding local regions in our method. In this section, a consistent segmentation method is proposed to establish region correspondences between two stereo images.

A consistent segmentation between two images is defined as: if two pixels belong to a region in the segmented source image, their corresponding pixels also belong to the corresponding region in the segmented target image [42]. Therefore, we aim to assign the same labels to the same objects segmented from a pair of stereo images. However, due to color discrepancies, a straightforward color-based labeling is often inaccurate. Instead of color, we use reliable feature matches as constraints to establish accurate region correspondences. Moreover, we define the Accurate Correspondences (AC) requirements to be fulfilled by the consistent segmentation as well as possible. The definition is as follows:

1. Every region belongs to a single object.
2. Every region contains a certain number of features.

The flowchart shown in Fig. 3 gives a brief description of our method. Details of each step are described below. First, the SIFT features [43] are extracted and matched between two stereo images. The invariance of the SIFT features to illumination variations ensures that a sufficient number of feature matches can be obtained even if there is a significant color discrepancy between the two input images. In addition, the epipolar geometry constraints [44] are adopted to remove outliers in the initial SIFT matches, which is defined as follows:

$$(X^{src})^T F X^{tgt} = 0 \quad (3)$$

where  $F$  is the fundamental matrix, and  $X^{src}$  and  $X^{tgt}$  are the positions of the matched SIFT features in  $I_{src}$  and  $I_{tgt}$ , respectively. SIFT matches that do not satisfy Eq. (3) within a given threshold of precision are considered as outliers and are rejected. The random sample consensus (RANSAC) [45] algorithm is applied to estimate  $F$ . For robustness and accuracy, we use the Sampson distance instead of the algebraic distance in the estimation of the fundamental matrix. Finally, a set of reliable SIFT features matches are retained.

Second, the source image and target image are initially segmented into small regions via the Mean-Shift segmentation algorithm [46]. Due to the limitation of the Mean-Shift segmentation algorithm, no parameter can produce segmentation results that simultaneously satisfy both of the AC requirements above.

For example, some regions may have no feature points if the segmentation makes every region belong to a single object, as shown in the red rectangle marked in Fig. 4(a); on the other hand, if every segmented region contains a certain number of features, some regions may not be segmented properly, as shown in yellow rectangle marked in Fig. 4(b). Considering this situation, we set the parameters of the Mean-Shift segmentation algorithm as  $\sigma_s = 25$ ,  $\sigma_r = 9$ ,  $M = 500$  for the source image to satisfy requirement 1, as shown in Fig. 4(a) and as  $\sigma_s = 7$ ,  $\sigma_r = 4$ ,  $M = 300$  for the target image to satisfy requirement 2, as shown in Fig. 4(b); the segmentations of both images will eventually satisfy both of the AC requirements through the following correspondence-based merging process.

Third, a correspondence-based merging process is proposed to generate consistent region correspondences after the initial segmentations. Fig. 5 is an illustration of this process. The input is the reliable SIFT matches and the initial segmentation results. The details of this process are described as follows:

**Step 1.** The initial region correspondences are established according to this principle: If the number of SIFT matches between two regions of the two images is greater than a threshold  $t$ , these two regions are marked as a correspondence.

**Step 2.** In the target image, we merge each region that has no correspondence into one of its nearby regions that have correspondences. The merging algorithm is described as follows: for each target region  $P_i^{tgt}$  that has no correspondence, we search for the region  $P_j^{tgt}$  in its 2-ring neighborhood, which has the corresponding region and minimizes the following distance term  $D$ :

$$D = \alpha |C_i - C_j| + \beta |B_i - B_j| \quad (4)$$

where  $C_i$  and  $C_j$  are the average colors of the region  $P_i^{tgt}$  and the region  $P_j^{tgt}$ , respectively;  $B_i$  and  $B_j$  are the barycenter of the region  $P_i^{tgt}$  and the region  $P_j^{tgt}$ , respectively;  $\alpha$  and  $\beta$  are weights. Then, we merge the region  $P_i^{tgt}$  into the region  $P_j^{tgt}$ . In our experiments, we set  $\alpha$  to 0.7 and  $\beta$  to 0.3. As shown in Fig. 5, region  $a$  is merged into region  $b$ .

**Step 3.** In the source image, we merge the regions that have the same corresponding region in the target image. As shown in Fig. 5, region  $c$  and region  $d$  are merged as a single region.

**Step 4.** In the source image, we merge each region that has no correspondence into one of its nearby regions that have correspondences, similar to the method in Step 2.

**Step 5.** In the target image, we merge the regions that have the same corresponding region in the source image, similar to the method in Step 3.

Through the above merging process, new segmentation results  $\{\tilde{P}_i^{src}\}$ ,  $\{\tilde{P}_j^{tgt}\}$  are obtained, and a bijective mapping between them is constructed. The result of the consistent segmentation is shown in the last step in Fig. 3. We can observe that the corresponding regions basically represent the same objects in two stereo images.



Fig. 3. The three steps of the proposed consistent segmentation method.

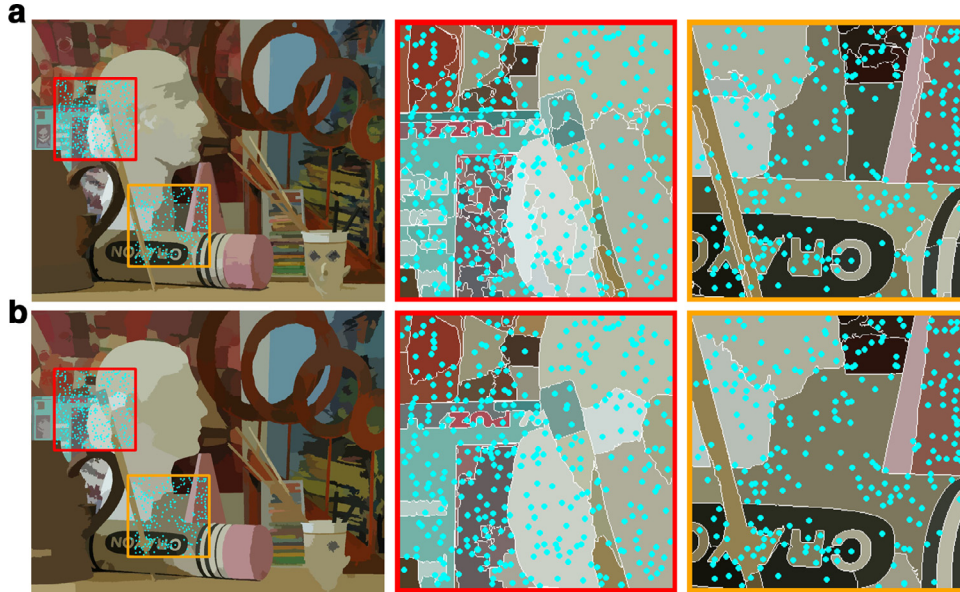


Fig. 4. The Mean-Shift segmentation with different parameters of the same image. (a) is the segmentation result satisfy requirement 1 ( $\sigma_s = 25$ ,  $\sigma_r = 9$ ,  $M = 500$ ); (b) is the segmentation result satisfy requirement 2 ( $\sigma_s = 7$ ,  $\sigma_r = 4$ ,  $M = 300$ ). Features within the rectangles are shown in color cyan. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

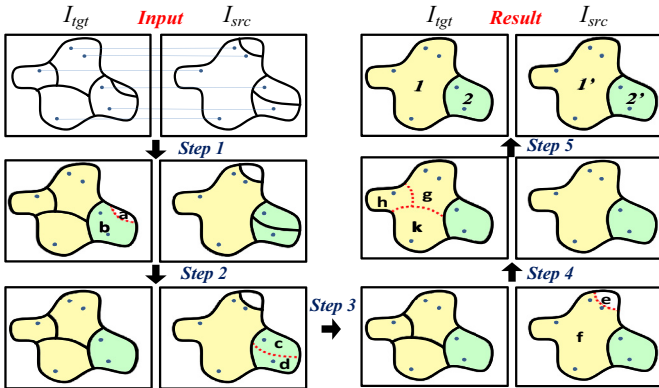


Fig. 5. An illustration of the correspondence-based merging process.

### 3.3. Local correction of intrinsic layers

We now have the normalized reflectance and shading layers of the two stereo images and the region correspondences between them. In this section, the two layers of the target image are corrected separately according to the corresponding layers of the source image. Based on the region correspondences, we perform different local correction methods for the intrinsic layers region by region.

#### 3.3.1. Local correction functions

Considering that the reflectance layer has a different number of channels and different physical properties from the shading layer, we adopt two different correction methods to process the two layers, which are introduced as follows:

**Shading layer correction.** Our method to correct the shading layer is inspired by the global method proposed by Reinhard et al. [23]. First, for each pair of corresponding regions,  $\{P_k^{src}, P_k^{tgt}\}$ , we compute their distribution statistics  $\mu_k$  and  $\sigma_k$ , i.e., the mean and standard deviation. Then, for each region  $P_k^{tgt}$  in the target shading layer, we apply the following correction function:

$$S_k^{tgt'}(i, j) = \mu_k^{src} + \frac{\sigma_k^{src}}{\sigma_k^{tgt}} (S_k^{tgt}(i, j) - \mu_k^{tgt}) \quad (5)$$

where  $S_k^{tgt}(i, j)$  and  $S_k^{tgt'}(i, j)$  denote the initial shading value and corrected shading value of pixel  $(i, j)$  in region  $P_k^{tgt}$ , respectively.

**Reflectance layer correction.** Reinhard's method can produce reliable results, but it must be performed in the  $\alpha\beta$  color space, whose three channels are uncorrelated. However, the three channels of the reflectance layers are not uncorrelated. Furthermore, the dimension "l" in  $\alpha\beta$  which used in [23] represents lightness while the reflectance layer represents the surface reflectance of the objects without the effect of lightness. In consideration of these factors, we adopt a correlated color transfer method, which can achieve color transfer directly in a correlated color space. The basic idea of this method is that color correction is performed by transferring the statistical properties of the colors. And we introduce the covariance matrices, which are related to the three channels, to the transferring.

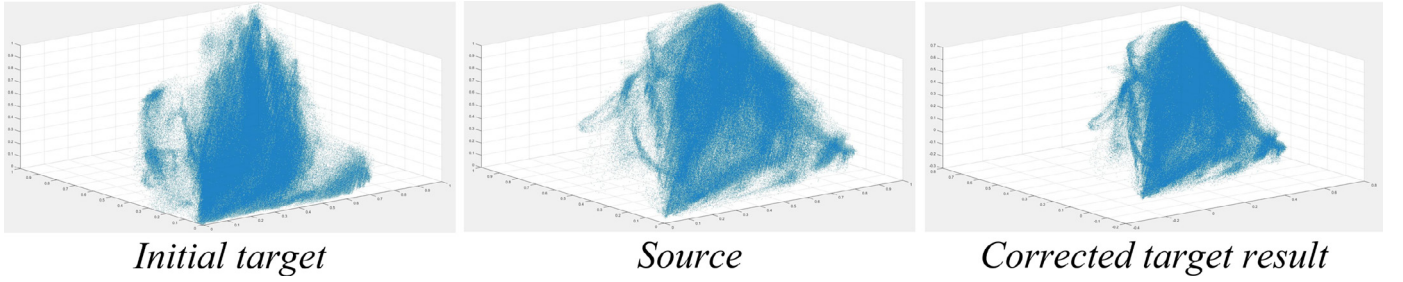


Fig. 6. 3D visualization of the reflectance distribution of the Art dataset.

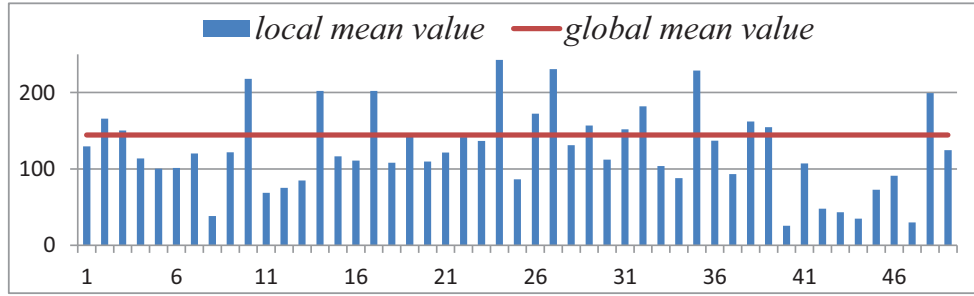


Fig. 7. Global mean value and local mean values of the 50 regions of the target shading layer of the Art dataset.

Fig. 6 shows the distribution of the original target reflectance, the source reflectance, and the corrected target reflectance. It can be seen the reflectance distribution of the corrected target image is similar to that of the source image. The detailed correction process is presented as follows:

For the reflectance layers of two stereo images, we first calculate the mean reflectance and the covariance matrices of the three components. For a corresponding region pair  $\{P_k^{tgt}, P_k^{src}\}$ , a local correction function is defined as:

$$R_k^{tgt'}(i, j) = Q_k^{src} \cdot Q_k^{tgt} \cdot R_k^{tgt}(i, j) \quad (6)$$

where  $R_k^{tgt}$  and  $R_k^{tgt'}$  denote the initial reflectance and the corrected reflectance in the homogeneous form of the pixel point  $(i, j)$  in region  $P_k^{tgt}$ .  $Q_k^{tgt}$  and  $Q_k^{src}$  represent two transformations, respectively, where  $Q_k^{tgt} = TS_k^{tgt} \cdot RT_k^{tgt} \cdot SL_k^{tgt}$  and  $Q_k^{src} = SL_k^{src} \cdot RT_k^{src} \cdot TS_k^{src}$ . And  $TS$ ,  $RT$  and  $SL$  denote the translation, rotation and scaling matrices, respectively. These transformations can be derived using the mean reflectance and the covariance matrices. Detailed derivation can be found in [47].

In global approaches, the parameters of the above two functions are constant for the whole image. However, different regions may have different statistical properties, so the correction function parameters should be distinct for different regions. For example, Fig. 7 shows the local mean value in each segmented region and the global mean value of the shading layer. The local mean values are not equal to the global mean value, and they vary among the segmented regions, which is why local methods perform better.

### 3.3.2. Weighted correction frame

Although the corrected intrinsic layers of target image have a similar appearance to those of the source image in the corresponding regions, the region boundaries may be discontinuous. To eliminate these discontinuities caused by local transformations, the local correction functions are blended to produce a smooth correction. The blending is weighted by the influence masks (IM) [42]. The IM weights for pixel  $(i, j)$ , denoted as  $IM_k(i, j)$  is defined as:

$$\exp\left(-\frac{\|C_k^{tgt}(i, j) - \mu_k^{tgt}\|^2}{2\alpha^2}\right) \times \exp\left(-\frac{dist((i, j), P_k^{tgt})^2}{2\beta^2}\right) \quad (7)$$

Table 1

Running time of each step for several tested examples.

	Intrinsic decomposition	Consistent segmentation	Color correction
Art	402.743 s	119.114 s	26.72 s
Baby1	235.947 s	136.611 s	7.085 s
Bowling1	296.817 s	146.84 s	3.868 s
Reindeer	347.011 s	165.068 s	12.777 s

where  $\|C_k^{tgt}(i, j) - \mu_k^{tgt}\|$  is the value distance between the shading/reflectance value of pixel  $(i, j)$  of the target image and the mean shading/reflectance value  $\mu_k^{tgt}$  of region  $P_k^{tgt}$ , and  $dist((i, j), P_k^{tgt})$  is the Euclidean distance between pixel  $(i, j)$  and the center of region  $P_k^{tgt}$ . Examples of some IM are shown in Fig. 8. The color correction functions weighted by the IM are defined as follows:  $(\sum_{k=1}^N R_k^{tgt'} \times IM^k) / (\sum_{k=1}^N IM^k)$ ,  $(\sum_{k=1}^N S_k^{tgt'} \times IM^k) / (\sum_{k=1}^N IM^k)$ .

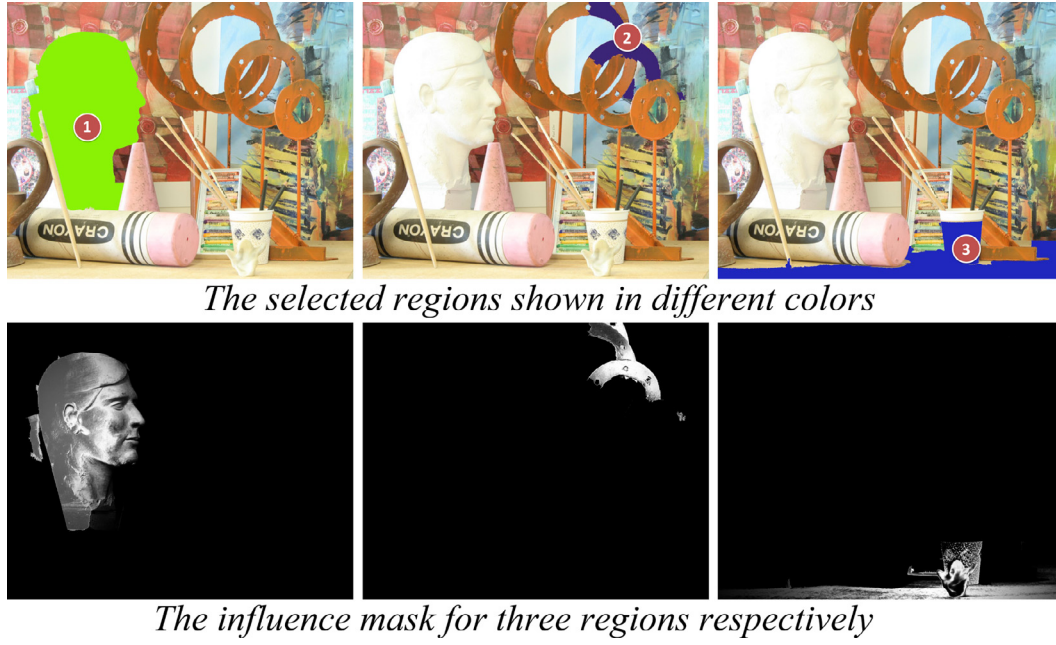
### 3.4. Intrinsic combination

After the local corrections for the intrinsic layers, we obtain the corrected reflectance layer  $R^{tgt'}$  and shading layer  $S^{tgt'}$  for the target image, respectively. The final corrected target image is then obtained by intrinsic combination, which is defined as:  $I^{tgt'} = R^{tgt'} \cdot S^{tgt'}$ .

## 4. Results, evaluation and discussion

The Middlebury stereo datasets [11] are adopted to test the proposed algorithm. The datasets in 2005 and 2006 provide stereo pairs captured under different lighting conditions or with different exposure settings. The proposed color correction method is tested on several stereo pairs with color discrepancy. All the experiments are conducted on a desktop with an Intel Core i7 4.00GHz CPU and with 16GB of memory.

The running times for four stereo pairs with resolution  $1390 \times 1110$  are listed in Table 1. Most of the time is consumed by



**Fig. 8.** Visualization of the influence masks.

intrinsic decomposition. However, considering the distinctive factors to cause the color discrepancy of the reflectance layer and the shading layer, better improvement would be produced by handling two layers separately, which is also proven by the following experimental results. We compare our method with several previous methods, including the global approach (GL) [23], Expectation–Maximization based local approach (EM) [29], and Region–Wise local color correction (RW) [33]. The abbreviations “IN” and “n-IN” refer to two versions of our proposed method. “IN” is the full implementation of the method, and “n-IN” does not include intrinsic decomposition and combination, and is directly applied to the original image. In this way, we try to discuss the role of intrinsic decomposition in the proposed method.

#### 4.1. Subjective evaluation

First, we subjectively assess the color correction results. Human perception can serve as a straightforward evaluation. Fig. 9 shows the corrected results of the Reindeer dataset using the five correction methods. All the methods can improve the color consistency of the target image with the source image to some extent. Thereinto, n-IN and IN show relatively greater improvements. Observing the last row in Fig. 9, our results are very consistent with the ground truth, where color changes and edges are noticeable in the results of other methods.

#### 4.2. Color similarity

The color similarity (CS) criterion, which is defined as the  $\alpha\beta$  color distance between the colors of the corrected (or original) image and the ground truth image, is adopted to quantitatively evaluate the experimental results. The distance function is defined as follows:

$$CS_{org} = \sum \|c_{gt}(i, j) - c_{org}(i, j)\|$$

$$CS_{crt} = \sum \|c_{gt}(i, j) - c_{crt}(i, j)\|$$

where  $c_{gt}(i, j)$  denotes the color of pixel  $(i, j)$  in the ground truth image,  $c_{org}(i, j)$  denotes its color in the original target image and

$c_{crt}(i, j)$  denotes its color in the corrected target image.  $CS_{org}$  measures the CS between the original target image and the ground truth image, and  $CS_{crt}$  measures the CS between the corrected target image and the ground truth image. The CS metric proposed by Oliveira et al. [30] is adopted to evaluate the improvement ratio of the correction method, which is defined as:

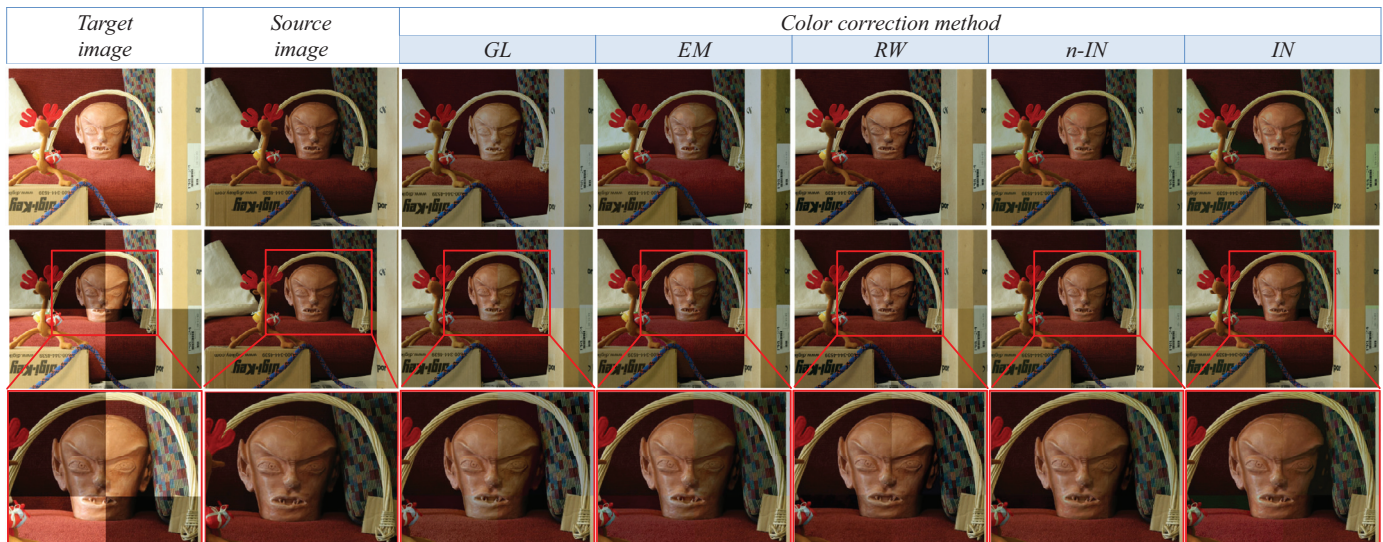
$$RCS = \frac{CS_{org} - CS_{crt}}{CS_{org}} \quad (8)$$

where  $RCS$  represents the color correction improvement ratio of the improvement in  $CS_{crt}$  relative to  $CS_{org}$ . A larger  $RCS$  ratio indicates a better color correction ability.

Table 2 shows the CS ratios of the five methods, which are applied to eight datasets. The local correction methods have a better performance than the global algorithm [23] because the global method changes the overall mean and variance of the color distribution, which may lead to biased results, especially when the color discrepancy between the stereo images is particularly large. Among the five color correction methods, the proposed correction methods with (IN) and without (n-IN) the intrinsic decomposition have the highest RCS ratios because of the accurate region correspondences. “IN” performed better than “n-IN”, which means that intrinsic decomposition improves the corrected results. The EM algorithm also performs well for most of the images. RW performs better than EM for images containing abundant color variations. The RW method does not account for the region correspondences in the occluded areas, so the corrections in those areas are inaccurate.

#### 4.3. Improvement on stereo matching

The proposed intrinsic color correction scheme can serve as a pre-processing step for stereo matching algorithms to improve the disparity accuracy, which is the main purpose. Two strategies are used to evaluate the improvement of the proposed algorithm on stereo match. First, with a baseline stereo matching algorithm, the proposed method is compared with different color correction methods (i.e., GL, EM, RW). Second, the proposed color correction method is tested with different stereo matching algorithms, of which the matching costs take account for the color discrepancy.

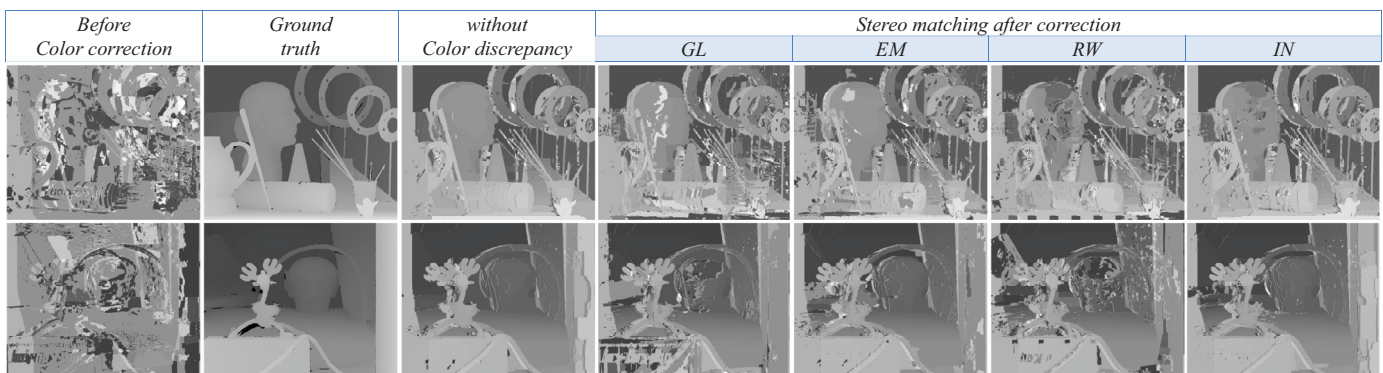


**Fig. 9.** The corrected results of the Reindeer dataset. The seven images from left to right in each row are the original target image, the ground truth image, and the corrected results obtained by GL [23], EM [29], RW [33], n-IN (our method without intrinsic decomposition) and IN (our method). The last two rows are generated by covering the corrected images with the ground truth in the top-left corner and the bottom-right corner. (Other results can be found in the supplemental files).

**Table 2**

The RCS ratios for eight datasets by the five methods (Other results can be found in the supplemental files). The unit is percentage.

	Aloe	Art	Baby1	Books	Bowling1	Cloth1	Dolls	Reindeer
<b>GL</b>	79.7561	85.2062	84.3083	77.4789	79.4131	85.1659	81.4288	83.6869
<b>EM</b>	90.6878	90.4002	93.4579	78.9137	84.7834	92.4639	87.4502	85.8269
<b>RW</b>	91.4118	85.2062	92.3952	78.9137	80.6024	92.3775	81.4288	82.6257
<b>n-IN</b>	94.3648	91.9891	93.1057	78.2551	93.0264	94.5844	89.2728	91.6097
<b>IN</b>	<b>95.3008</b>	<b>93.4809</b>	<b>93.6775</b>	<b>79.2133</b>	<b>93.0965</b>	<b>94.6832</b>	<b>89.9839</b>	<b>91.7756</b>



**Fig. 10.** The comparisons of the stereo matching results with different color correction methods applied to the Art and Reindeer datasets.

#### 4.3.1. Different color correction methods

The baseline stereo matching algorithm we used is the graph-cut algorithm proposed by Kolmogorov and Zabih [48], which employs the commonly used absolute difference as the matching cost [48]. The assumption is that the corresponding pixels have the same color, which tend to be influenced by color discrepancies. We compute the disparity maps from the original stereo images and the corrected stereo images to compare the proposed method with other color correction methods. The disparity maps for the Art and Reindeer datasets are shown in Fig. 10. Due to the color discrepancy, the disparity map computed without color correction contains a large amount of noise and errors compared with the ground truth disparity. After color correction, the disparity maps are more accurate and close to the ground truth of disparity maps. To evaluate the disparity maps quantitatively, pixels that differ from the

ground truth disparity map by more than two pixels are treated as bad pixels. We then calculate the bad pixel rates, denoted as  $BP$ , in stereo matching when different correction methods are used to correct the target images. Table 3 shows the bad pixels rates for the tested examples. The “Tgt” row shows the results obtained from uncorrected stereo images; the “Tgt-n” row shows the results obtained from stereo images that originally have no color discrepancy. The “IN” method often has the lowest bad pixel rate among the tested methods.

#### 4.3.2. Different stereo matching algorithms

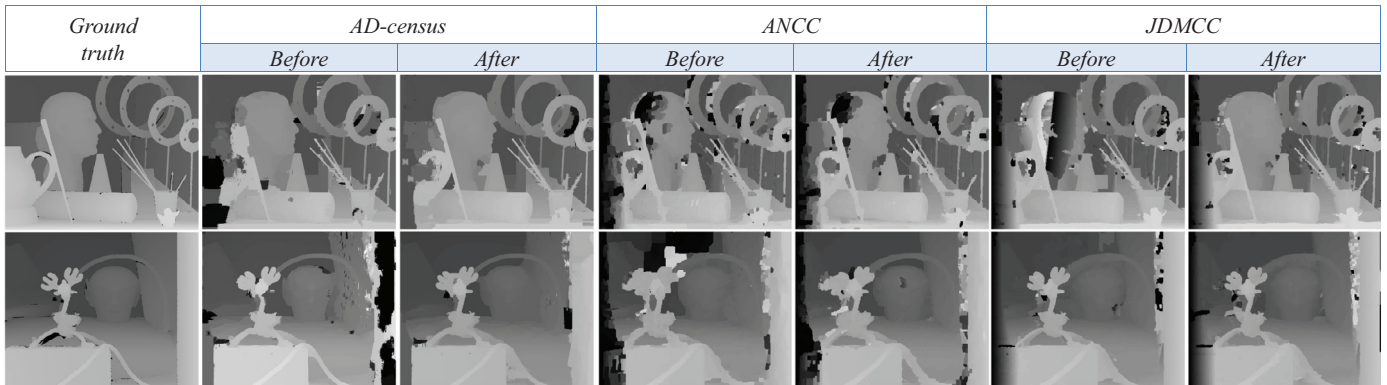
Using the proposed color correction method, we test several robust stereo matching algorithms that can handle a certain degree of color discrepancy, such as AD-census [10], ANCC [12] and JDMCC [13]. The results are shown in Fig. 11. For the method AD-census,



**Table 3**

The stereo matching bad pixel statistics. (Other results can be found in the supplemental files.)

	Aloe	Art	Baby1	Books	Bowling1	Cloth1	Dolls	Reindeer
<b>Tgt</b>	0.8500	0.7462	0.7208	0.8545	0.8152	0.8561	0.8053	0.8245
<b>Tgt-n</b>	0.0688	0.3496	0.0840	0.2559	0.3061	0.0158	0.0853	0.1823
<b>GL</b>	0.2412	0.4298	0.3214	0.5141	0.6114	0.1638	0.4228	0.4048
<b>EM</b>	0.1535	0.4413	0.1676	0.4403	0.5618	0.0372	0.1943	0.2479
<b>RW</b>	0.0792	0.3641	0.2280	0.4341	0.3531	0.0398	0.4249	0.3845
<b>IN</b>	<b>0.0676</b>	<b>0.3553</b>	<b>0.0857</b>	<b>0.3688</b>	<b>0.3297</b>	<b>0.0139</b>	<b>0.1159</b>	<b>0.1933</b>

**Fig. 11.** The “AD-census”, “ANCC” and “JDMCC” stereo matching results for the Art and Reindeer datasets.

the bad pixel rate generated from the corrected images is lower than those without color correction. The other two methods always produce good results for stereo image pairs with color discrepancy; whereas, when the input images are in full resolution, these two methods will cost a large amount of memory, which may beyond the capacity of general computers. Thus, our proposed method is a good choice to compensate for the color discrepancy in stereo matching.

#### 4.4. Robustness and boundaries

To show the robustness and boundaries of our method, we will discuss the impact of consistent segmentation errors and dis-occlusion over the final results.

First, the consistent segmentation error (CSE), is defined as the percentage of pixels that are not satisfied with the definition of consistent segmentation (in Section 3.2). The consistent segmentation ground-truth is obtained from disparity ground-truths provided by Middlebury Dataset [11]. Qualitatively, the color discrepancy could be compensated obviously. However, to some extents, there is color discrepancy in areas containing segmentation errors. And, we calculate the CSE values, the RCS values and the BP values of several test data, respectively, for quantitative analysis. It shows that the smaller the segmentation error (CSE), the greater the color similarity improvement, which indicates the relationship between CSE and RCS is negative. On the other hand, the relationship between CSE and the percentage of bad pixels (BP) is positive.

Second, to discuss the impact of dis-occlusion factor over the final results, pixels are divided into occlusion pixels and non-occlusion pixels, based on the ground-truth of visibility (given by the Middlebury Dataset). The color corrected target image and the disparity results show that the proposed approach will benefit the color similarity and disparity computing of both occlusion pixels and non-occlusion pixels. For quantitative analysis, the RCS values and the BP values for occlusion pixels and non-occlusion pixels are calculated for test data, respectively. We find that after color correction, the color similarity of all pixels, including occlusion pixels and non-occlusion pixels, could be improved, i.e., greater than

90%. And, the percentage of bad pixels of occlusion areas and non-occlusion areas could be reduced about 47% and 38%, respectively. Readers can refer to the supplementary file for detailed figures and statistics.

## 5. Conclusions and future work

In this paper, we presented a local intrinsic color correction method to address the color discrepancy between two stereo images. Consistent segmentation, intrinsic decomposition, and local region-based color correction of the intrinsic layers are the three major steps of our method. A consistent segmentation scheme was proposed to establish accurate region correspondences. With the intrinsic scheme and more accurate region correspondences, our approach improves the color similarity as well as the accuracy of the disparity computed by stereo matching. These improvements are demonstrated by a detailed discussion and comparison with other methods.

Rather than the efficiency, we discuss the probability of using intrinsic images to reduce color dissimilarity between a pair of stereo images. Our method is also useful in other image processing manipulations or computer vision applications, such as image stitching and multi-view geometry. All the original images and data can be found in the supplemental files.

## Acknowledgment

This work was supported jointly by the National Natural Science Foundation of China under Grants No. 61732015 and No. 61472349, Key Research and Development Program of Zhejiang Province, China under Grant No. 2018C01090.

## Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi: [10.1016/j.cag.2019.05.008](https://doi.org/10.1016/j.cag.2019.05.008)

## References

- [1] Hirschmuller H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans Pattern Anal Mach Intell* 2008;30(2):328–41.
- [2] Viola P, Wells WM III. Alignment by maximization of mutual information. *Int J Comput Vis* 1997;24(2):137–54.
- [3] Zabih R, Woodfill J. Non-parametric local transforms for computing visual correspondence. In: Proceedings of the third European conference on computer vision ECCV'94, II; 1994. p. 151–8. Stockholm, Sweden, May 2–6, 1994.
- [4] Zbontar J, LeCun Y. Computing the stereo matching cost with a convolutional neural network. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015. p. 1592–9.
- [5] Hirschmüller H, Scharstein D. Evaluation of cost functions for stereo matching. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition (CVPR); 2007. 18–23 June 2007, Minneapolis, Minnesota, USA.
- [6] Faridul HS, Pouli T, Chamaret C, Stauder J, Reinhard E, Kuzovkin D, et al. Colour mapping: a review of recent methods, extensions and applications. *Comput Graph Forum* 2016;35:59–88. Wiley Online Library.
- [7] Xu W, Mulligan J. Performance evaluation of color correction approaches for automatic multi-view image and video stitching. In: Proceedings of the twenty-third IEEE Conference on computer vision and pattern recognition, CVPR; 2010. p. 263–70. San Francisco, CA, USA, 13–18 June 2010.
- [8] Fischler MA, Witkin AP. Recovering intrinsic scene characteristics from images. *Tech. Rep.*; 1981. DTIC Document.
- [9] Brainard DH, Wandell BA. Analysis of the retinex theory of color vision. *JOSA A* 1986;3(10):1651–61.
- [10] Mei X, Sun X, Zhou M, Jiao S, Wang H, Zhang X. On building an accurate stereo matching system on graphics hardware. In: Proceedings of the IEEE international conference on computer vision workshops (ICCV workshops). IEEE; 2011. p. 467–74.
- [11] Middlebury stereo website; 2018. [vision.middlebury.edu/stereo/](http://vision.middlebury.edu/stereo/).
- [12] Heo YS, Lee KM, Lee SU. Robust stereo matching using adaptive normalized cross-correlation. *IEEE Trans Pattern Anal Mach Intell* 2011;33(4):807–22.
- [13] Heo YS, Lee KM, Lee SU. Joint depth map and color consistency estimation for stereo images with different illuminations and cameras. *IEEE Trans Pattern Anal Mach Intell* 2013;35(5):1094–106.
- [14] Bleyer M, Rhemann C, Rother C. Patchmatch stereo-stereo matching with slanted support windows. In: Proceedings of the BMVC, 11; 2011. p. 1–11.
- [15] Zhang C, Li Z, Cheng Y, Cai R, Chao H, Rui Y. MeshStereo: a global stereo model with mesh alignment regularization for view interpolation. In: Proceedings of the IEEE international conference on computer vision; 2015. p. 2057–65.
- [16] Luo W, Schwing AG, Urtasun R. Efficient deep learning for stereo matching. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 5695–703.
- [17] Ummenhofer B, Zhou H, Uhrig J, Mayer N, Ilg E, Dosovitskiy A, et al. Demon: depth and motion network for learning monocular stereo. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 5; 2017. p. 6.
- [18] Fecker U, Barkowsky M, Kaup A. Histogram-based prefiltering for luminance and chrominance compensation of multiview video. *IEEE Trans Circuits Syst Video Technol* 2008;18(9):1258–67.
- [19] Jia J, Tang C. Tensor voting for image correction by global and local intensity alignment. *IEEE Trans Pattern Anal Mach Intell* 2005;27(1):36–50.
- [20] Kim SJ, Pollefeys M. Robust radiometric calibration and vignetting correction. *IEEE Trans Pattern Anal Mach Intell* 2008;30(4):562–76.
- [21] Pitié F, Kokaram AC, Dahyot R. N-dimensional probability density function transfer and its application to colour transfer. In: Proceedings of the 10th IEEE international conference on computer vision (ICCV); 2005. p. 1434–9. 17–20 October 2005, Beijing, China.
- [22] Bellavia F, Colombo C. Dissecting and reassembling color correction algorithms for image stitching. *IEEE Trans Image Process* 2018;27(2):735–48.
- [23] Reinhard E, Ashikhmin M, Gooch B, Shirley P. Color transfer between images. *IEEE Comput Graph Appl* 2001;21(5):34–41.
- [24] Xiao X, Ma L. Color transfer in correlated color space. In: Proceedings of the ACM international conference on virtual reality continuum and its applications. ACM; 2006a. p. 305–9.
- [25] Hwang Y, Lee J-Y, So Kweon I, Joo Kim S. Color transfer using probabilistic moving least squares. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2014. p. 3342–9.
- [26] Faridul H, Stauder J, Kervec J, Trémeau A. Approximate cross channel color mapping from sparse color correspondences. In: Proceedings of the IEEE international conference on computer vision workshops; 2013. p. 860–7.
- [27] Grogan M, Dahyot R. Robust registration of gaussian mixtures for colour transfer; 2017. arXiv: 1705.06091.
- [28] Liao D, Qian Y, Li Z-N. Semisupervised manifold learning for color transfer between multiview images. In: Proceedings of the 23rd international conference on pattern recognition (ICPR). IEEE; 2016. p. 259–64.
- [29] Tai Y, Jia J, Tang C. Local color transfer via probabilistic segmentation by expectation-maximization. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition (CVPR); 2005. p. 747–54. 20–26 June 2005, San Diego, CA, USA.
- [30] Oliveira M, Sappa AD, Santos V. Unsupervised local color correction for coarsely registered images. In: Proceedings of the 24th IEEE conference on computer vision and pattern recognition; 2011. p. 201–8. Colorado Springs, CO, USA, 20–25 June 2011.
- [31] Ly D-S, Beucher S, Bilodeau M. Color correction through region matching leveraged by point correspondences. In: Proceedings of the IEEE international conference on image processing (ICIP). IEEE; 2014. p. 640–4.
- [32] Wang Q, Yan P, Yuan Y, Li X. Robust color correction in stereo vision 2011;6626(1):965–8.
- [33] Ran Q, Zhao W, Feng J. Robust region-wise colour correction method for stereo matching. *IET Comput Vis* 2016.
- [34] Land EH, McCann JJ. Lightness and retinex theory. *JOSA* 1971;61(1):1–11.
- [35] Grosse RB, Johnson MK, Adelson EH, Freeman WT. Ground truth dataset and baseline evaluations for intrinsic image algorithms. In: Proceedings of the IEEE 12th international conference on computer vision, ICCV; 2009. p. 2335–42. Kyoto, Japan, September 27 - October 4.
- [36] Shen L, Yeo C. Intrinsic images decomposition using a local and global sparse representation of reflectance. In: Proceedings of the 24th IEEE conference on computer vision and pattern recognition, CVPR; 2011. p. 697–704. Colorado Springs, CO, USA, 20–25 June 2011.
- [37] Barron JT, Malik J. Shape, illumination, and reflectance from shading. *IEEE Trans Pattern Anal Mach Intell* 2015;37(8):1670–87.
- [38] Garces E, Muñoz A, Lopez-Moreno J, Gutierrez D. Intrinsic images by clustering. *Comput Graph Forum* 2012;31(4):1415–24.
- [39] Bell S, Bala K, Snavely N. Intrinsic images in the wild. *ACM Trans Graph* 2014;33(4):159:1–159:12.
- [40] Beigpour S, Shekhar S, Mansouryar M, Myszkowski K, Seidel H-P. Light-field appearance editing based on intrinsic decomposition. *J Percept Imaging* 2018;1(1):1–15 10502.
- [41] Krähenbühl P, Koltun V. Efficient inference in fully connected CRFs with gaussian edge potentials. In: Shawe-Taylor J, Zemel RS, Bartlett PL, Pereira F, Weinberger KQ, editors. Advances in Neural Information Processing Systems 24. Curran Associates, Inc.; 2011. p. 109–17.
- [42] Wang H, Dai L, Zhang X. Consistent segmentation based color correction for coarsely registered images. In: Proceedings of the 2nd IAPR Asian conference on pattern recognition (ACPR). IEEE; 2013. p. 319–24.
- [43] Lowe DG. Object recognition from local scale-invariant features. In: Proceedings of the ICCV; 1999. p. 1150–7.
- [44] Hartley A, Zisserman A. Multiple view geometry in computer vision. 2nd ed. Cambridge University Press; 2006. ISBN 978-0-521-54051-3.
- [45] Fischler MA, Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun ACM* 1981;24(6):381–95.
- [46] Comaniciu D, Meer P. Mean shift: a robust approach toward feature space analysis. *IEEE Trans Pattern Anal Mach Intell* 2002;24(5):603–19.
- [47] Xiao X, Ma L. Color transfer in correlated color space. In: Proceedings of the ACM international conference on virtual reality continuum and its applications. ACM; 2006b. p. 305–9.
- [48] Kolmogorov V, Zabih R. Computing visual correspondence with occlusions using graph cuts. In: Proceedings of the eighth IEEE international conference on computer vision, ICCV, 2. IEEE; 2001. p. 508–15.