INTRINSIC DECOMPOSITION FOR STEREOSCOPIC IMAGES

Dehua Xie[1], *Shuaicheng Liu*[1], *Kaimo Lin*[2], *Shuyuan Zhu*[1], *and Bing Zeng*[1]

[1]School of Electronic Engineering, University of Electronic Science and Technology of China
[2]Dept. of Electrical & Computer Engineering, National University of Singapore

ABSTRACT

Intrinsic image decomposition is an important technique that decomposes an image into reflectance and shading components. In this paper, we enable intrinsic decomposition for stereoscopic images. Traditional approaches cannot be directly applied to decompose stereoscopic images, yielding inconsistent reflectance and 3D artifacts after recoloring. To solve this problem, we propose a straight yet effective method for stereoscopic intrinsic decomposition, which consists of classical retinex constraint as well as disparity constraint. The former encodes the shading smoothness prior while the latter controls the reflectance similarity between two views. To further reduce ambiguity, we employ local and non-local texture cues by using superpixels within and across two views. The experiments show that our method can effectively decompose stereoscopic images with high quality and offer a comfortable 3D viewing experience.

***Index Terms*—** Stereoscopic, intrinsic image decomposition, retinex, disparity

1. INTRODUCTION

With the success of 3D movies, stereoscopic cameras and displays become more and more popular. In recent years, various methods have been proposed to target stereoscopic images/videos processing. To name a few, stereoscopic warping [1], cloning [2],inpainting [3], panorama [4],retargeting [5], stabilization [6] are all popular image/video editing research topics. However, among all these studies, fewer work has been proposed to address the intrinsic decomposition problem for stereoscopic images, which is one of the most classical and important topics in image editing.

Intrinsic image decomposition has decades of history in vision and graphics communities [7, 8, 9]. It aims at decomposing an image I into the reflectance (R) component and the shading (S) component as follows:

$$I = R \times S, \tag{1}$$

where "$\times$" denotes the pixel-to-pixel multiplication. Many existing problems can benefit from such a decomposition with the presence of these useful confounded information. For examples, resurfacing and material recognition prefer a reflectance image without appearance variations caused by shading [10]; and the shape-from-shading algorithm will be able to infer the surface geometry more accurately when less reflectance variation appears in the shading map [11].

Though there continuously are a number of reported works contributing to this field, e.g., [12, 13, 14], intrinsic decomposition still remains a challenging problem. It is highly ill-posed as the number of unknowns (R, S) is twice as many as the observations (I). Furthermore, these traditional monocular decompositions cannot be directly applied to stereoscopic images, due to the resulting inconsistent reflectance. By considering several extra constraints within and across views, our approach can deliver more pleasant results. Fig. 1 shows the effect of recoloring.

2. RELATED WORK

Intrinsic decomposition is first introduced in [7], which proposed the retinex algorithm to analyze local image deviations in shading and reflectance. Some refined retinex algorithms with different assumptions were reported in [14, 15, 16]. To further reduce the ambiguity, some methods [12, 17] tried to learn the priors to judge the image derivatives, while others [14, 18] proposed additional constraints to reduce the number of unknowns. Specifically, Weiss *et. al.* [15] used multiple images to regularize the system. However, it requires an accurate image registration, which is challenging in many scenarios. Kang *et. al.* [19] proposed a method for stereoscopic decomposition. However, their method mainly focuses on graphics images and is only conducted on synthetic data. In this paper, we propose a more general approach that can be applied to stereo pairs captured by consumer-level cameras.

Stereoscopic image/video processing is an important topic, which has attracted lots of attention in the community. Consistent manipulation on two different views is crucial for high quality stereo editing. To this end, Wang *et. al.* [20] explored an inpainting technique for stereoscopic images, where both color and depth can be recovered simultaneously. Image cloning methods [2] copies the content and pastes it into a new 3D environment under the stereoscopic 3D disparity constraints. Meanwhile, image warping is one of the most basic manipulates in image editing. For instance, Niu *et. al.* [1] extended the conventional 2D image warping to

Fig. 1. An illustration of recoloring on stereoscopic images.

handle 3D stereoscopic images; Lang *et. al.* [21] exploited a locally adaptive algorithm for nonlinear disparity mapping; Du *et. al.* [22] proposed to change the perspective for stereo pictures; and Didyk *et. al.* [23] introduced a perceptual model of disparity. More recently, stereo image cropping [24] and authoring [25] have also been explored.

3. OUR METHOD

Take the logarithm at both sides of Eq. (1) gives, $I = R + S$. For simplicity, we reuse I, R, and S to represent their log values. We decompose stereoscopic image pairs by optimizing over the following energy with respect to shading S:

$$\begin{aligned}\arg\min_S E(S) =& \lambda_r E_r(S) + \lambda_d E_d(S) \\ &+ \lambda_l E_l(S) + \lambda_{nl} E_{nl}(S) + \lambda_a E_a(S),\end{aligned} \quad (2)$$

where E_r, E_{nl}, E_d, E_l, and E_a are the retinex term, disparity term, local term, non-local term and absolute scale term, respectively, with λ_s, λ_r, λ_l, λ_{nl} and λ_a being the associated weights. Each of these terms will be elaborated in detail in the following. Notice that the reflectance R can be obtained by $I - S$. Through experiments, we set $\lambda_s = \lambda_r = \lambda_l = \lambda_{nl} = 1$ and $\lambda_a = 1000$ for all examples in our system.

3.1. Retinex Constraint

According to the retinex theory, large derivatives in image intensity attribute to reflectance changes, while small derivatives are given to shading variations. Thus, the retinex term $E_r(S)$ is formulated as a weighted sum of shading and reflectance of neighboring pixels over the whole image region [18]:

$$E_r(S) = \sum_m \sum_{n \in \mathcal{N}(m)} \left[(S_m - S_n)^2 + \omega_{m,n} (R_m - R_n)^2 \right], \quad (3)$$

where m stands for all pixels and $\mathcal{N}(m)$ denotes the four adjacent neighbors of m. Substituting R_m by $I_m - S_m$, yields:

$$\begin{aligned} E_r(S) = \sum_m \sum_{n \in \mathcal{N}(m)} & \left[(1 + \omega_{m,n}) \Delta S_{m,n}^2 + \Delta I_{m,n}^2 \right. \\ & \left. - 2\omega_{m,n} \Delta I_{m,n} \Delta S_{m,n} \right], \end{aligned} \quad (4)$$

where $\Delta S_{m,n} = S_m - S_n$, $\Delta I_{m,n} = I_m - I_n$, and $\omega_{m,n}$ is a balancing weight that can be calculated in a number of ways [7, 18]. Here, we evaluate the chromaticity distance of neighboring pixels. If the chromaticity distance is greater than a threshold τ[1], we set $\omega_{m,n} = 0$, otherwise it is set to 100. Large reflectance variations are allowed, only when its chromaticity varies dramatically. The chromaticity distance is defined as:

$$D(m, n) = 2 \times \left(1 - \frac{\langle \mathbf{m}, \mathbf{n} \rangle}{\|\mathbf{m}\|_2 \cdot \|\mathbf{n}\|_2} \right) \quad (5)$$

where $\mathbf{m}$ and $\mathbf{n}$ denote color values of pixels m and n.

3.2. Disparity Constraint

Per-pixel dense disparity estimation methods are well documented in [26]. In our implementation, we calculate dense optical flow [27] between two views for disparity correspondances. We drop the inaccurate flow (e.g., occluded pixels) by checking the pixel color similarity following the flow. Other estimation methods can also be adopted [28]. We want to encourage the corresponding pixels between two views to have similar reflectance. Given the per-pixel correspondences, we define the disparity term $E_d(S)$ as:

$$E_d(S) = \sum_m (R_m^l - R_{m'}^r)^2, \quad (6)$$

where R^r and R^l refer to the left and right reflectance[2], m and m' are matched disparity point pair. By substituting R with $I - S$, Eq. (6) can be rewritten as:

$$E_d(S) = \sum_m (I_m^l - I_{m'}^r - S_m^l + S_{m'}^r)^2. \quad (7)$$

3.3. Local Constraint

Local texture cues have shown their effectiveness to reduce the ambiguity [18]. We employ superpixels [29] to represent image regions with similar color and textures. We encourage the local similarity within the superpixels on reflectance, which gives rise to the local term $E_l(S)$:

$$E_l(S) = \sum_{P_l^i \in \Gamma_l} \sum_{m,n \in P_l^i} (I_m - I_n - S_m + S_n)^2, \quad (8)$$

[1] τ is set to 0.0009 in our implementation.

[2] If no l and r are specified, we represent both.

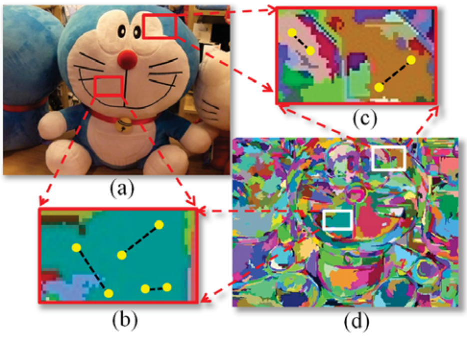

Fig. 2. An illustration of local constraints

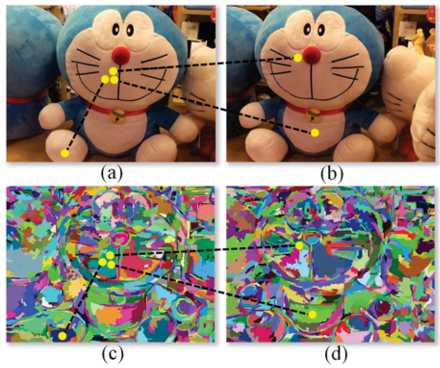

Fig. 3. An illustration of non-local constraints

where P_l^i represents a superpixel group, Γ_l denotes the set of superpixel groups, and m and n are two different pixels from the same superpixel.

An illustration of superpixels for local constraint is shown in Fig. 2. We sample several pairs of points (yellow dots highlighted in Fig. 2 (b), (c)) in each superpixel (Fig. 2 (d)). The number of sampled points is determined by the size of a superpixel. Notably, local constraints are applied to two views separately.

3.4. Non-Local Constraint

Figure 3 shows an example of non-local constraints. We encourage superpixels with similar image color to share a similar reflectance, which can effectively reduce the ambiguity of the system. Particularly, for each superpixel, we find the K nearest neighbors in terms of image color in both views and sample several pairs of locations inside them as illustrated in Fig. 3. We can easily spot that there are many similar non-adjacent patches within and across two views. The non-local term $E_{nl}(S)$ is then defined as:

$$E_{nl}(S) = \sum_{P_{nl}^i, P_{nl}^j \in \Gamma_{nl}} \sum_{m \in P_{nl}^i, n \in P_{nl}^j} (I_m - I_n - S_m + S_n)^2, \quad (9)$$

where Γ_{nl} is the union of all superpixels from both views. Compared to the local constraints, the difference between Eq. (9) and Eq. (8) is that pixels m and n are not in the same superpixel but different superpixels.

3.5. Absolute Scale Constraint

Finally, we add the absolute scale term as follows:

$$E_a(S) = \sum_{m \in P_a} (S_m - 1)^2, \quad (10)$$

where the set P_a denotes two brightest pixels, one for each view. We constrain the brightest pixels in two views to have a unit shading, which further regularizes the system.

It is important to point out that all terms in Eq. (2) are quadratic, yielding a sparse linear system which is minimized by the Matlab '\' solver.

	#1	#2	#3	#4	#5
without	0.045	0.153	0.045	0.044	0.083
with	0.023	0.054	0.044	0.036	0.042

Table 1. The APD values of examples in Fig. 4 under with/out disparity constraints

4. EXPERIMENTS

Among all the five terms in Eq. (2), the retinex, local and absolute are quite classical. Here, we would like to validate the effectiveness of disparity and non-local terms, both visually and numerically, following which we present our final reuslts in Fig. 4.

4.1. With and Without Disparity Constraint

To evaluate the effect of the disparity term on reflectance consistency numerically, we define the *average pixel distance* (APD) as follows:

$$APD(R_m^l, R_{m'}^r) = \frac{\sum_{m,m' \in \Phi} \|R_m^l - R_{m'}^r\|_2}{N(\Phi)}, \quad (11)$$

where Φ is the set of all matched pixels in two views and $N(\Phi)$ is the total number of matched pixel pairs. We compare the color difference of corresponding reflectance numerically (note that the color is normalized to $[0, 1]$).

Table 1 lists APD values with and without the disparity constraints on the examples in Fig. 4. Without the disparity constraints, the reflectance is inconsistent between two views, which is indicated by the larger APD values. Moreover, we use the method of [30] to decompose our 5th example in Fig. 4. As it is a monocular method, it is applied to two views separately. The reflectances are shown in Fig. 5, which is inconsistent between two views. Notably, we adopt the same parameters for all examples.

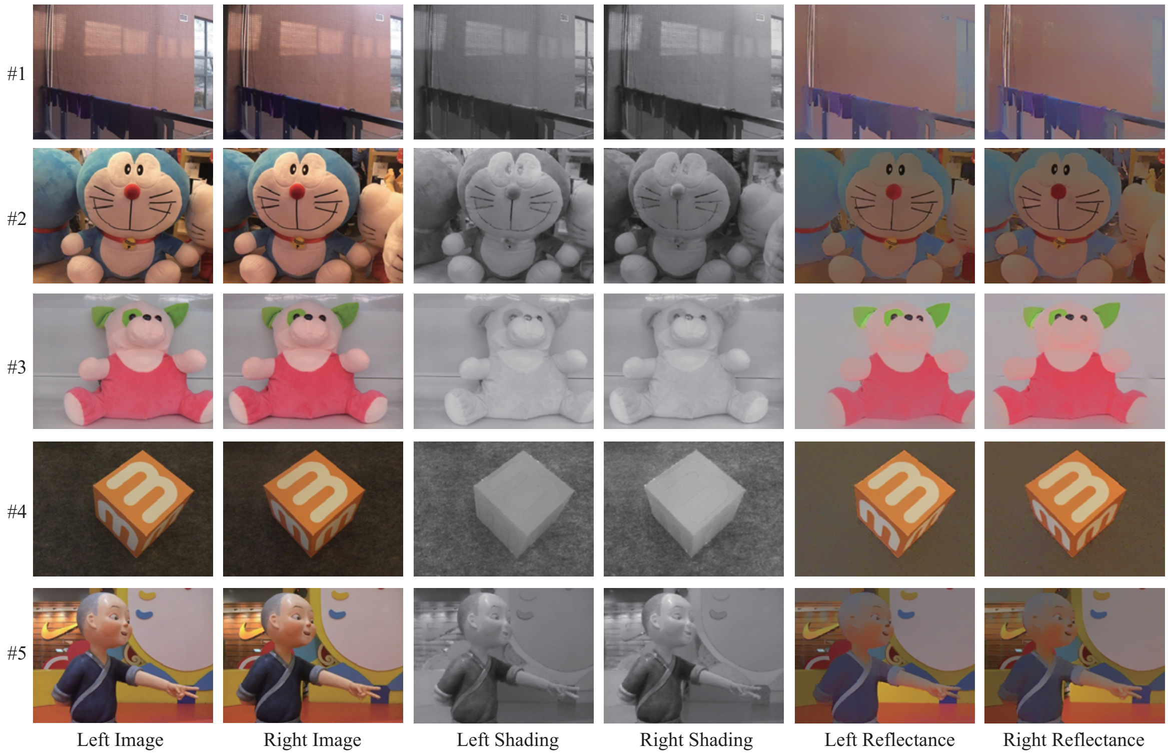

Fig. 4. Our results. More results are provided in the supplementary. Images are captured by a steroscopic camera of ourselves.

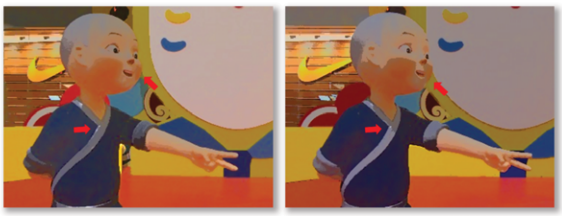

Fig. 5. The reflectances produced by method [30] on Fig. 4 #5. The inconsistence is highlighted by the red arrows.

4.2. With and Without Non-Local Constraint

Figure 6 shows a visual comparison of with and without the non-local constraints. Here, we conduct the experiment on a single image. The input image (Fig. 6(a)) is segmented by superpixles (Fig. 6(b)). Notice the color difference in the reflectance image (Fig. 6(c)) as highlighted by the red arrow. Our result (Fig. 6(d)) is more consistent.

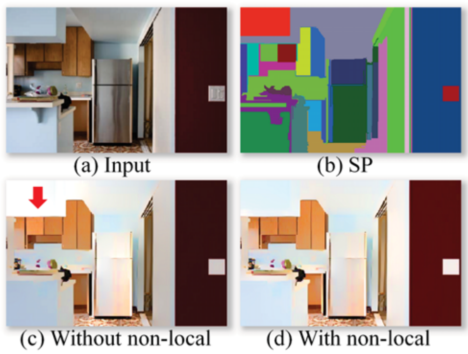

Fig. 6. With and without non-local constraints. Image credits: is licensed under *Creative Commons*.

5. CONCLUSION

We have presented a method for intrinsically decomposing stereoscopic images. Our method combines several constraints, namely retinex, disparity, local, non-local, and absolute scale constraints, into a unified framework to produce a consistent reflectance decomposition. Our method has been validated on various stereoscopic images to produce a better performance as compared to the conventional approaches that are developed for monocular images.

6. ACKNOWLEDGE

We would like to thank all reviewers for their comments to improve our paper. This work has been supported by National Natural Science Foundation of China (61502079, 61370148 and 61300091)

References

[1] Y. Niu, W. Feng, and F. Liu, "Enabling warping on stereoscopic images," *ACM TOG*, vol. 31, no. 6, pp. 183, 2012.

[2] S. Luo, I. Shen, B. Chen, W. Cheng, and Y. Chuang, "Perspective-aware warping for seamless stereoscopic image cloning," *ACM TOG*, vol. 31, no. 6, pp. 182:1–182:8, 2012.

[3] T. Mu, J. Wang, S. Du, and S. Hu, "Stereoscopic image completion and depth recovery," *The Visual Computer*, vol. 30, no. 6-8, pp. 833–843, 2014.

[4] F. Zhang and F. Liu, "Casual stereoscopic panorama stitching," in *CVPR*, 2015.

[5] K. Lee, C. Chung, and Y. Chuang, "Scene warping: Layer-based stereoscopic image resizing," in *CVPR*, 2012, pp. 49–56.

[6] F. Liu, Y. Niu, and H. Jin, "Joint subspace stabilization for stereoscopic video," in *ICCV*, 2013, pp. 73–80.

[7] E. Land and J. McCann, "Lightness and retinex theory," *Journal of the Optical Society of America A,*, vol. 61, no. 1, pp. 1–11, 1971.

[8] H. Barrow and J. Tenenbaum, "Recovering intrinsic scene characteristics from images," *Computer Vision Systems*, p. 2, 1978.

[9] Christophe Schlick, "A survey of shading and reflectance models," in *Computer Graphics Forum*, 1994, vol. 13, pp. 121–131.

[10] G. Ye, E. Garces, Y. Liu, Q. Dai, and D. Gutierrez, "Intrinsic video and applications," *ACM TOG*, vol. 33, no. 4, pp. 80, 2014.

[11] R. Zhang, P.-S. Tsai, J. Cryer, and M. Shah, "Shape from shading: a survey," *IEEE TPAMI*, vol. 21, no. 8, pp. 690–706, 1999.

[12] T. Narihira, M. Maire, and S. Yu, "Learning lightness from human judgement on relative reflectance," in *CVPR*, 2015, pp. 2965–2973.

[13] S. Bi, X. Han, and Y. Yu, "An l_1 image transform for edge-preserving smoothing and scene-level intrinsic decomposition," *ACM TOG*, vol. 34, no. 4, pp. 78, 2015.

[14] L. Shen, P. Tan, and S. Lin, "Intrinsic image decomposition with non-local texture cues," in *CVPR*, 2008, pp. 1–7.

[15] Y. Weiss, "Deriving intrinsic images from image sequences," in *ICCV*, 2001, vol. 2, pp. 68–75.

[16] A. Bousseau, S. Paris, and F. Durand, "User-assisted intrinsic images," in *ACM TOG*, 2009, vol. 28, p. 130.

[17] J. Barron and J. Malik, "Shape, illumination, and reflectance from shading," *IEEE TPAMI*, vol. 37, no. 8, pp. 1670–1687, 2015.

[18] Q. Zhao, P. Tan, Q. Dai, L. Shen, E. Wu, and S. Lin, "A closed-form solution to retinex with nonlocal texture constraints," *IEEE TPAMI*, vol. 34, no. 7, pp. 1437–1444, 2012.

[19] J. Kang, B. Jiang, J. Chen, and X. Liu, "Stereoscopic image recoloring via consistent intrinsic decomposition," in *ICVRV*. IEEE, 2014, pp. 272–277.

[20] L. Wang, H. Jin, R. Yang, and M. Gong, "Stereoscopic inpainting: Joint color and depth completion from stereo images," in *CVPR*, 2008, pp. 1–8.

[21] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. Gross, "Nonlinear disparity mapping for stereoscopic 3d," *ACM TOG*, vol. 29, no. 4, pp. 75, 2010.

[22] S. Du, S. Hu, and R. Martin, "Changing perspective in stereoscopic images," *IEEE TVCG*, vol. 19, no. 8, pp. 1288–1297, 2013.

[23] P. Didyk, T. Ritschel, E. Eisemann, K. Myszkowski, and H. Seidel, "A perceptual model for disparity," *ACM TOG*, vol. 30, no. 4, pp. 96, 2011.

[24] F. Zhang, Y. Niu, and F. Liu, "Making stereo photo cropping easy," in *ICME*, 2013, pp. 1–6.

[25] F. Liu, Y. Niu, and H. Jin, "Casual stereoscopic photo authoring," *IEEE TMM*, vol. 15, no. 1, pp. 129–140, 2013.

[26] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *IJCV*, vol. 47, no. 1-3, pp. 7–42, 2002.

[27] C. Liu, *Beyond pixels: exploring new representations and applications for motion analysis*, Ph.D. thesis, MIT, 2009.

[28] H. Hirschmüller and D. Scharstein, "Evaluation of stereo matching costs on images with radiometric differences," *IEEE TPAMI*, vol. 31, no. 9, pp. 1582–1599, 2009.

[29] P. Felzenszwalb and D. Huttenlocher, "Efficient graph-based image segmentation," *IJCV*, vol. 59, no. 2, pp. 167–181, 2004.

[30] S. Bell, K. Bala, and N. Snavely, "Intrinsic images in the wild," *ACM TOG*, vol. 33, no. 4, pp. 159, 2014.