# Estimation of Intrinsic Image Sequences from Image+Depth Video

Kyong Joon Lee[1,2], Qi Zhao[3], Xin Tong[1], Minmin Gong[1], Shahram Izadi[4], Sang Uk Lee[2], Ping Tan[5], and Stephen Lin[1]

[1] Microsoft Research Asia
[2] Seoul National University
[3] UC Santa Cruz
[4] Microsoft Research Cambridge
[5] National University of Singapore

**Abstract.** We present a technique for estimating intrinsic images from image+depth video, such as that acquired from a Kinect camera. Intrinsic image decomposition in this context has importance in applications like object modeling, in which surface colors need to be recovered without illumination effects. The proposed method is based on two new types of decomposition constraints derived from the multiple viewpoints and reconstructed 3D scene geometry of the video data. The first type provides *shading constraints* that enforce relationships among the shading components of different surface points according to their similarity in surface orientation. The second type imposes *temporal constraints* that favor consistency in the intrinsic color of a surface point seen in different video frames, which improves decomposition in cases of view-dependent non-Lambertian reflections. Local and non-local variants of the two constraints are employed in a manner complementary to local and non-local *reflectance constraints* used in previous works. Together they are formulated within a linear system that allows for efficient optimization. Experimental results demonstrate that each of the new constraints appreciably elevates the quality of intrinsic image estimation, and that they jointly yield decompositions that compare favorably to current techniques.

## 1 Introduction

Intrinsic image decomposition aims to separate an image into its reflectance and shading components. The reflectance component contains the intrinsic color, or albedo, of surface points independent of the illumination environment. On the other hand, the shading component consists of various lighting effects that include shadows and specular highlights in addition to shading. Decomposing an image into reflectance and shading can benefit computer vision algorithms such as segmentation and texture map recovery which are designed to analyze one of these components but are degraded by variations in the other. However, intrinsic image decomposition remains a difficult problem as it is highly underconstrained, with two quantities (reflectance and shading) to estimate at each pixel from a single input value (pixel color).

A. Fitzgibbon et al. (Eds.): ECCV 2012, Part VI, LNCS 7577, pp. 327–340, 2012.

### 1.1 Previous Work

Various approaches have been employed for intrinsic image estimation. Many works are based on local analysis of image derivatives. In these methods, each image derivative is attributed to either shading or reflectance change, and then the derivatives of each type are integrated to obtain the shading and reflectance images. A simple and common approach for derivative classification is to attribute large intensity or chromaticity derivatives to reflectance changes, and smaller derivatives to shading [1–3]. This approach presumes that scenes have piecewise constant reflectance and smooth shading variations, an assumption that often does not hold in natural scenes. Reflectance and shading derivatives have also been distinguished using classifiers trained on labeled edge data [4, 5]. However, an accurate classification may not always be possible from the local appearance of an edge.

Several works employ decomposition cues that extend beyond local analysis. In [6, 7], decompositions at a local level are constrained to be consistent globally to a plausible model of a simple scene. In [8], points from different parts of an image are constrained to have the same reflectance if they have the same local texture. The additional constraints from non-local analysis can lead to improved intrinsic image decompositions, but the availability of this additional information may be limited in many scenes. A couple of recent works employ a global constraint on the solution by assuming a sparse number of distinct reflectances in the scene [9, 10]. User interaction may alternatively be used to relate shading or reflectance among different image points [10–12], but may not be suitable as a preprocessing step for automatic computer vision algorithms.

Instead of estimating a decomposition from a single input image, some previous methods address the less ill-posed problem of estimating intrinsic images from a sequence of images captured at a fixed viewpoint and with multiple lighting conditions [13–16]. With the extra data from an image sequence, these methods generally produce decompositions of higher quality than from single-image techniques. However, these methods are applicable only in certain scenarios, e.g., outdoor cameras capturing image sequences or video over a long time period.

### 1.2 Our Approach

In this work, we address the problem of intrinsic image decomposition for a different type of image sequence – video from a moving image+depth camera – that provides multiple viewpoints and reliable 3D scene reconstruction. We show that such input contains information particularly useful for estimating the shading and reflectance components of the scene, and exploit this data to improve intrinsic image decomposition via new shading and temporal constraints.

**Shading Constraints.** Intrinsic image methods conventionally assume smoothness in shading and/or reflectance to constrain the solution. Smoothness constraints on reflectance often are relaxed when there exist chromaticity differences between adjacent pixels, as this often indicates a change in albedo [2, 3, 8]. However, there lacks such a cue for reducing smoothness in shading. This can lead to

significant errors in image areas where both shading and reflectance are discontinuous, such as at object and surface boundaries. We address this issue using surface normal data computed from the input sequence. Locally, we use this information to deemphasize shading smoothness between adjacent pixels with different surface normals, which are differently exposed to the scene's lighting. On the other hand, our method promotes shading smoothness when adjacent pixels have similar surface orientations.

Similarity in shading components may exist not only locally between adjacent pixels with similar surface normals, but also non-locally between other pixels in the image with the same surface orientation. This consistency can be enforced by constraining such pixels to have similar illumination components. These non-local shading constraints between possibly distant points in an image are complementary to the non-local constraints on reflectance proposed in [8]. Both forms of non-local constraints can be employed together in solving for the two intrinsic image components. Unlike the non-local reflectance cues which relate points within the same texture region, our non-local shading cues can also relate points between different texture regions. This is especially important when processing a scene containing separate objects, since their decompositions cannot be made consistent to each other (e.g., in terms of illumination intensity) using only texture or smoothness constraints.

To apply this non-local shading constraint between a pair of points, they not only must have consistent surface normals, but also must share the same lighting condition (i.e., angular distribution and intensities of incident illumination). Unlike neighboring pixels processed with the local shading constraint, distant pixels in the non-local case often have inconsistent lighting conditions due to different occlusions of light by objects in the scene. We address this issue by again taking advantage of reconstructed 3D scene geometry to determine whether the non-local points have similar *visibility* toward potential lighting directions, as well as similar surface normals, before enforcing the non-local shading constraint. In addition, the light sources are assumed to be distant from the examined scene areas, such that all points in the scene would have the same lighting condition if unoccluded.

**Temporal Constraints.** In single-image methods, view-dependent reflectance effects such as specular highlights cannot be processed correctly since Lambertian reflectance is typically assumed in their formulations. In our multi-view scenario, we make use of temporal consistency in each surface point's reflectance component throughout the video to identify outliers caused by specular highlights that shift in image position for different viewpoints. Decomposition errors that would normally result from specular highlights in one image are avoided by discarding these outliers with respect to the other images. Since a significant change in viewpoint generally exists only between temporally distant video frames, we refer to this use of reflectance consistency as a non-local temporal constraint. We moreover incorporate local temporal constraints on reflectance consistency among neighboring frames to reduce the effects of imaging noise on decomposition solutions.

We note that reflectance consistency over time is also utilized in techniques based on image sequences with fixed cameras and moving light sources [13–15]. However, since these methods act on reflectance derivatives between adjacent pixels, they are susceptible to the effects of biased illumination sampling as explained in [15]. Biased sampling is not an issue in our work, since we utilize moving cameras instead of moving light sources, and define the temporal constraints directly on individual surface points instead of on their derivatives.

To demonstrate the significance of the proposed shading and temporal constraints, we use the non-local texture method of [8] as a baseline algorithm and show that adding these new constraints to it leads to improvements in intrinsic image decomposition. Experiments are presented on several challenging scenes.

We note that the intrinsic colorization method of [17] also estimates intrinsic images from a set of images taken from different viewpoints. Specifically, it utilizes photographs of famous landmarks downloaded from the internet, which generally are captured by different cameras, from different viewpoints, and under different illumination conditions. However, [17] does not take advantage of the additional information available from multiple views for intrinsic image estimation. Instead, it merely aligns image regions taken from different viewpoints, and utilizes differences in lighting conditions among the images in a manner similar to [13].

## 2 Background

Intrinsic image estimation may be expressed as the decomposition of an image $I$ into the product of a shading image $S$ and a reflectance image $R$:

$$I_p = S_p R_p \tag{1}$$

where $p$ denotes a point in the image space. In the logarithmic domain, this equation becomes

$$i_p = s_p + r_p \tag{2}$$

where the lowercase labels denote the logarithms of image values. As evident in this formula, intrinsic image estimation is an ill-posed problem, with two unknowns $(s, r)$ at each pixel and only one measurement $(i)$.

To obtain a solution, most intrinsic image estimation methods employ a Retinex approach that models $s$ and $r$ as being smooth except between pixels where there exists a large difference in intensity and/or chromaticity [1–3]. For color images, an image derivative with a significant chromaticity change is attributed to a change in reflectance. With this *local reflectance constraint* and smoothness in $s$ and $r$, a decomposition is solved using an energy function such as

$$\sum_{(p,q)\in\aleph} \left[ (s_p - s_q)^2 + \omega^r_{p,q} \left( (i_p - s_p) - (i_q - s_q) \right)^2 \right] \tag{3}$$

where

$$\omega^r_{p,q} = \begin{cases} \omega_r & \text{if } (1 - \hat{c_p}^T \hat{c_q}) < \tau_r \\ 0 & \text{otherwise} \end{cases} . \tag{4}$$

Here, $\aleph$ is the set of all adjacent pixel pairs in $i$, $\hat{c_p}$ denotes the $3 \times 1$ normalized color vector of pixel $p$, $\omega_r$ is a constant weight, $\tau_r$ represents a given threshold, and $\omega_{p,q}$ represents a weight for reflectance smoothness.

In [18], this model is supplemented with *non-local reflectance constraints* [8] that are obtained through an examination of surface texture. If two pixels have matching local neighborhoods with respect to chromaticity values, their intrinsic reflectance values are considered to be the same. This property is derived from the theory of Markov Random Fields [19]. Pixels in an image are thus grouped according to their local neighborhoods, and the reflectance of all pixels within a group are constrained to be the same, i.e.,

$$r_q = r_p \;\text{ if }\; q \in G_r(p) \tag{5}$$

where $G_r(p)$ denotes the *reflectance group* of $p$ as determined by clustering of local chromaticity textures. In practice, we implemented this by adding the following term to Eq. (3):

$$\sum_{p \in \Omega} \sum_{q \in G_r(p)} \left[ \omega_{nlr} \left( (i_p - s_p) - (i_q - s_q) \right)^2 \right] \tag{6}$$

where $\Omega$ denotes the set of image pixels and $\omega_{nlr}$ represents a constant weight.

We use this model of reflectance constraints as a baseline algorithm to show the contribution of the proposed shading and temporal constraints.

## 3 Shading Constraints

From the image+depth video, we formulate new decomposition constraints complementary to those described above for reflectance. The first of these impose relationships among shading components based on the orientations of surface points. To obtain these surface normals, we reconstruct the 3D geometry of the scene from the depth video, using the KinectFusion algorithm [20]. Though 3D reconstruction could instead be computed by structure-from-motion using only the image video, we found surface normal estimates by KinectFusion to be appreciably more accurate and reliable.

The *local shading constraint* accounts for the similarity in shading between adjacent pixels with matching normal orientations, as they both share the same incident lighting. We model this analogously to the *local reflectance constraint* in Eq. (4), based on an inner product of surface normals instead of chromaticity vectors:

$$\omega^s_{p,q} = \begin{cases} \omega_s & \text{if } (1 - \hat{n_p}^T \hat{n_q}) < \tau_s \\ 0.1\omega_s & \text{otherwise} \end{cases} \tag{7}$$

where $\tau_s$ is a threshold, and $\omega_s$ is a constant weight. This factor is used with the shading smoothness term in Eq. (3), giving a larger weight between adjacent pixels that have similar normal orientations. In contrast to Eq. (4), here we allow some amount of shading smoothness even if there is some difference in normal directions, as ambient illumination can contribute to smoothness in shading in

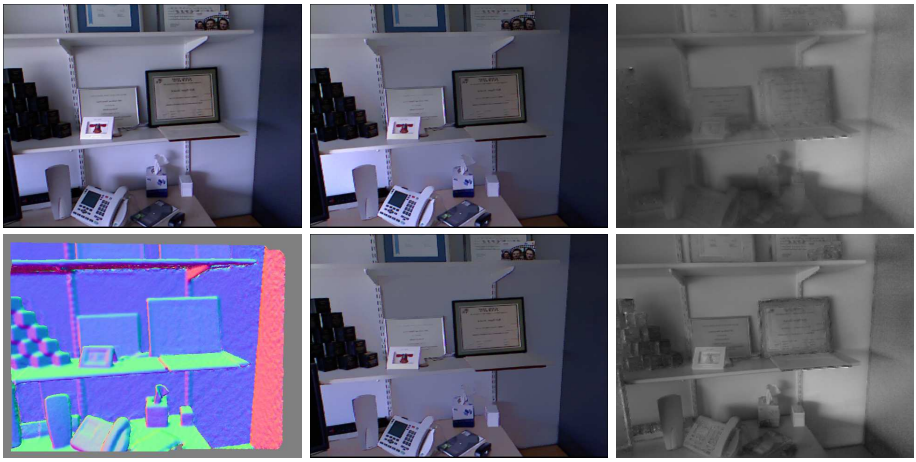

**Fig. 1.** Results of applying local shading constraint: **Top left** original scene image. **Top center & right** albedo & shading image from baseline algorithm. **Bottom left** normal maps used in the constraint. **Bottom center & right** albedo & shading image with the constraint.

such cases. Fig. 1 presents results of applying this constraint to the baseline algorithm. Improvements are found in the shading image, with sharper shading boundaries at object edges with significant surface normal variation.

A similar relationship may also exist between non-adjacent pixels that share the same surface normal direction. For this, we formulate a *non-local shading constraint* analogous to the *non-local reflectance constraint*, but operating on grouped neighborhoods of surface normals instead of chromaticities:

$$s_q = s_p \;\text{ if }\; q \in G_s(p) \;\text{ and }\; \|V(q) - V(p)\| < \tau_v \tag{8}$$

where $G_s(p)$ denotes the *shading group* of $p$ computed by clustering of local surface normal neighborhoods, $V$ represents the visibility of a pixel, and $\tau_v$ is a threshold. In computing shading groups, we opt to compare local neighborhoods rather than individual pixels to reduce the influence of measurement noise. Our implementation considers $11\times11$ patches and limits the search for matching patches to a $101\times101$ neighborhood to avoid substantial processing. Also, we consider the visibility of each point to account for differences in lighting condition due to occlusions. Here, the visibility at each surface point is computed from the reconstructed scene geometry by sampling different lighting directions and testing whether the light from that direction would be blocked by objects in the scene. This visibility test can be done efficiently by *shadow mapping*, a standard computer graphics technique that compares the actual depth of the point and the corresponding depth map value towards that point as seen from the lighting direction, illustrated for one direction in Fig. 2. In our experiments, we sample all combinations of polar and azimuth angles $(\theta, \phi)$ from $-\pi/4$ to $\pi/4$

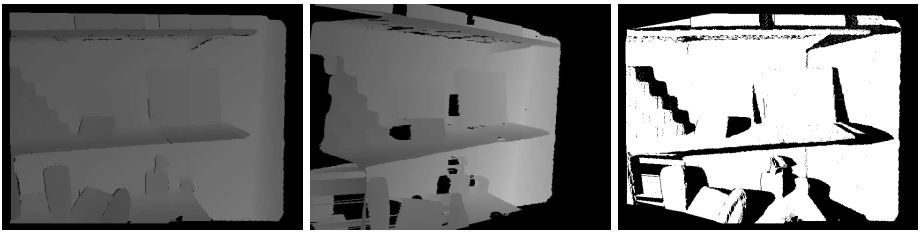

**Fig. 2.** Calculating visibility: **Left** captured depth map. **Center** synthesized depth map from a sampled lighting direction. **Right** visibility of points in the original depth map to the sampled light direction, where black indicates pixels that are blocked.

at intervals of $\pi/12$, except for $(0, 0)$, and organize the visibility results into a 48-D binary feature vector for each pixel.

To compute distance between the binary vectors, we use the Hamming distance, i.e.,

$$\|V(q) - V(p)\| = \sum_{i=1}^{48} |V_i(q) - V_i(p)|$$

where $V_i$ denotes the $i^{th}$ element of a vector.

Fig. 3 shows the effect of incorporating visibility into the non-local shading constraint. Utilizing normal groups without visibility can improve shading consistency among disjoint surfaces that share the same normal orientation, such as the wall sections separated by brackets. However, not accounting for visibility may lead to over-smoothing in the shading component (which causes shading effects to appear in the reflectance image as highlighted by the red boxes), since the difference in lighting condition between two points with the same normal is not taken into account. Adding visibility into the constraint helps to include cast shadows in the shading image while maintaining shading consistency among disjoint surfaces.

The grouping of shading values is incorporated into the energy function in a manner analogous to the *non-local reflectance constraint*, by adding the following term to Eq. (3):

$$\sum_{p \in \Omega} \sum_{q \in G_s(p)} \left[ \omega_{nls} \left( s_p - s_q \right)^2 \right] \tag{9}$$

where $\omega_{nls}$ represents a constant weight. This establishes relationships between the shading components of distant points that otherwise might not be preserved.

## 4 Temporal Constraints

The second set of constraints make use of the multiple observations and viewpoints captured in the video over time. The first of these, the *local temporal constraint*, aims to reduce degradation in the decomposition solution caused by imaging noise. This is done by first identifying the correspondences for each

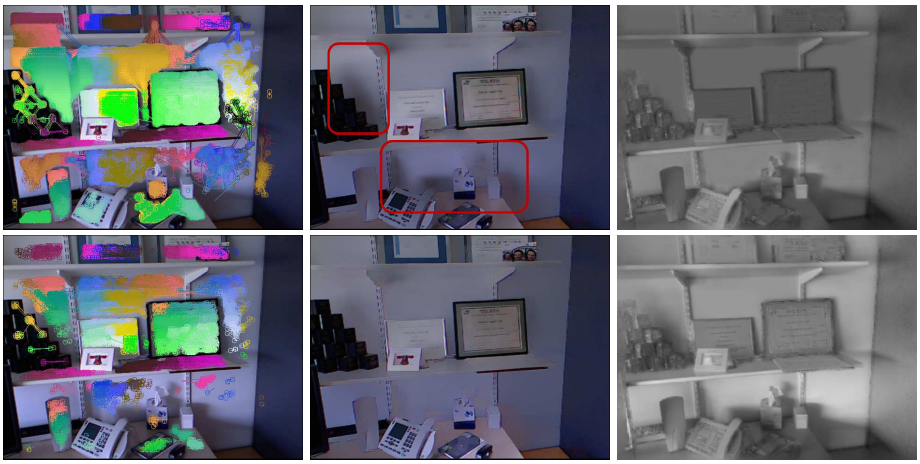

**Fig. 3.** Effect of visibility in non-local shading constraint: **Top left** shading groups for several sampled pixels, without considering visibility. **Top center & right** resulting albedo & shading image without visibility. **Bottom left** shading groups constrained by visibility. **Bottom center & right** resulting albedo & shading image.

point in the preceding ten frames, and then using the average color among the corresponding points for processing in the decomposition algorithm:

$$c_{p(t)} = \frac{1}{11} \sum_{k=0..10} c_{\varsigma_p(t-k)} \tag{10}$$

where $c_{p(t)}$ is the $3\times1$ color vector of a point $p$ in frame $t$, and $\varsigma_p(t-k)$ is the point corresponding to $p$ in frame $t-k$. In computing the correspondences, we make use of the reconstructed scene geometry by obtaining the 3D coordinate of each pixel in frame $t$ and projecting it to the image plane in frame $t-k$, using reliable camera parameters provided by the KinectFusion algorithm [20]. As shown in Fig. 4, this constraint reduces the impact of noise and enhances intrinsic image quality, especially in dark regions with a low signal-to-noise ratio.

We additionally take advantage of multiple viewpoints in the video to reduce the effects of specular highlights on the decomposition. In this *non-local temporal constraint*, the correspondences of a point at temporally distant earlier frames are examined to determine whether the image intensity of the point is an outlier, i.e., contains specular reflection. If so, then its color is replaced with that from a non-specular corresponding pixel:

$$c_{p(t)} = \arg \min_{c_{\varsigma_p(t-20*k)}} \left\| c_{\varsigma_p(t-20*k)} \right\| \ : \ k = 1..5. \tag{11}$$

If $t - 20 * k < 0$, then the non-existent frame is disregarded. Likewise, a term is ignored if its frame contains no correspondence. The effect of this constraint is exemplified in Fig. 5, where the specular highlight is properly handled. In our

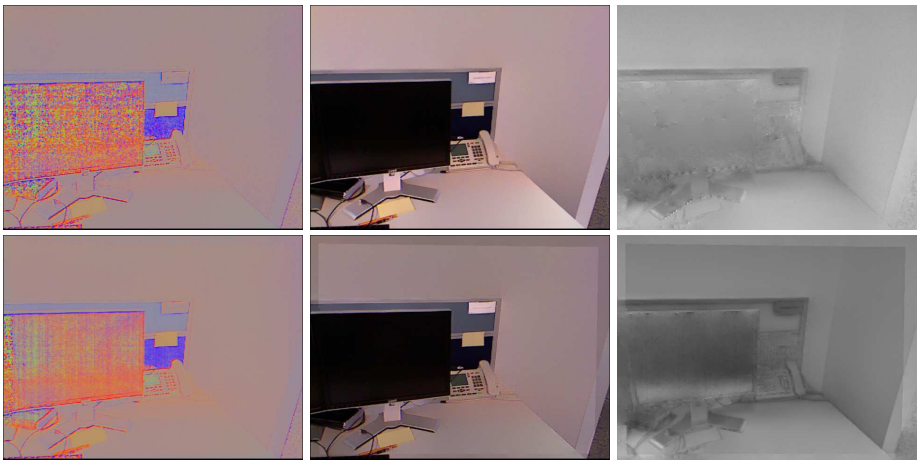

**Fig. 4.** Result of applying local temporal constraint: **Top left** original color, shown as chromaticity to better visualize noise effects in dark regions. **Top center & right** resulting albedo & shading image. **Bottom left** chromaticity after applying the constraint. **Bottom center & right** resulting albedo & shading image.

implementation, the local temporal constraint is applied before the non-local constraint.

## 5 Optimization

To estimate intrinsic images with local and non-local reflectance, shading and temporal constraints, we minimize the following energy function:

$$\arg\min_{\mathbf{s}} \sum_{(p,q)\in\aleph} \left[\omega^s_{p,q}(s_p - s_q)^2 + \omega^r_{p,q}\left((i_p - s_p) - (i_q - s_q)\right)^2\right] \tag{12}$$

$$+ \sum_{p\in\Omega}\sum_{q\in G_r(p)} \left[\omega_{nlr}\left((i_p - s_p) - (i_q - s_q)\right)^2\right] + \sum_{p\in\Omega}\sum_{q\in G_s(p)} \left[\omega_{nls}\left(s_p - s_q\right)^2\right]$$

after applying the temporal constraints in Eq. (10)-(11).

Eq. (12) is a quadratic function with respect to a vector $\mathbf{s}$ containing all the unknown variables $s_p$ for shading. We represent this function in a standard quadratic form as follows:

$$\arg\min_{\mathbf{s}} \frac{1}{2}\mathbf{s^T A s} - \mathbf{b^T s} + c \tag{13}$$

where $\mathbf{A}$ is an $M \times M$ symmetric positive-definite matrix and $\mathbf{b}$ is an $M \times 1$ vector. $M$ indicates the number of variables to be calculated, i.e., the number of pixels in the image. Although $\mathbf{A}$ is a very large matrix, most of its elements are zero. A non-diagonal element $a_{pq}$ is non-zero only when pixels $p, q$ are adjacent

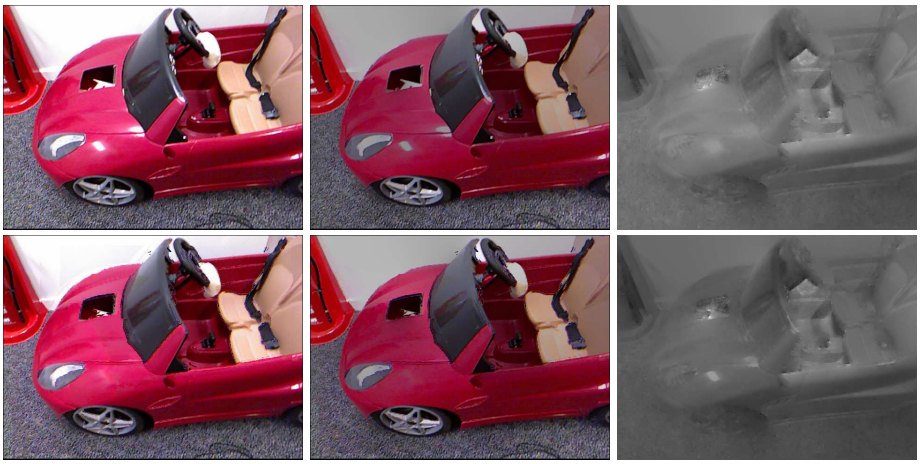

**Fig. 5.** Result of applying non-local temporal constraint: **Top left** original color. **Top center & right** resulting albedo & shading image. **Bottom left** color after applying the constraint. **Bottom center & right** resulting albedo & shading image.

or grouped together by non-local constraints. Thus Eq. (13) is straightforward to optimize using a sparse linear solver. In experiments we use the preconditioned conjugate gradient technique implemented on parallel graphics hardware, which requires only 0.1 second per $640 \times 480$ image frame.

## 6 Results

We validate our decomposition method on image sets captured with a Microsoft Kinect camera. These sets contain several image+depth videos of indoor scenes containing various static objects. As a preprocessing step, we aligned the depth and image frames since they are captured with slightly offset cameras. The algorithm parameters, which affect the relative influence and strictness of the different constraints, were empirically fixed to $\omega_r = 10, \omega_s = 1, \tau_r = 0.001 = tau_s = 0.001, \omega_{nlr} = \omega_{nls} = 0.1, \tau_v = 1$ throughout our experiments.

A benchmark dataset for evaluating intrinsic image algorithms was presented in [21]. However, it does not provide multiple viewpoints of a scene or geometry information, and so it is not applicable to our method. Since techniques do not exist for obtaining ground truth intrinsic images for the general large-scale scenes considered in this work, we evaluate our method through qualitative comparisons.

### 6.1 Effects of Individual Constraints

To show the effects of each proposed constraint, Fig. 6 presents results obtained with our full algorithm and compares them to decompositions computed with

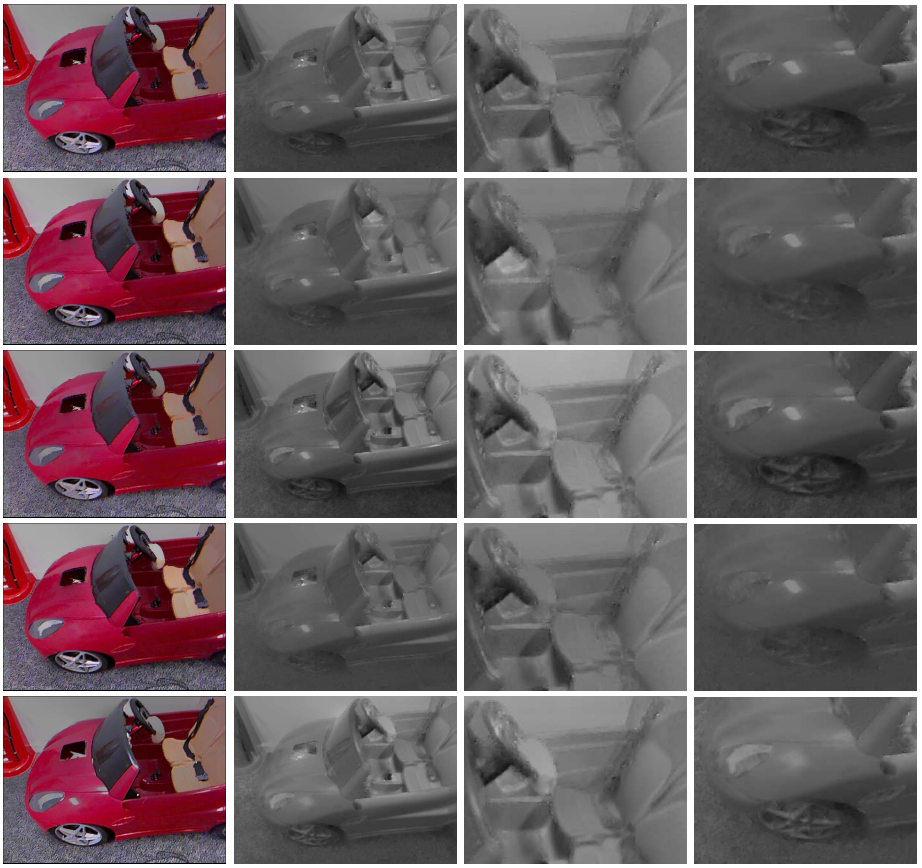

**Fig. 6.** Decompositions computed using **Row 1** the full algorithm; **Row 2** without local shading; **Row 3** without non-local shading; **Row 4** without local temporal; **Row 5** without non-local temporal.

various constraints removed. Please see the supplementary material for additional examples. Removing the local shading constraint (row 2) causes shading boundaries to become less sharp, e.g., in the magnified view of the wheel. Not using the non-local shading constraints (row 3) can lead to parallel surfaces having less consistent shading, e.g., the inner side of the far door and the wall. Without the local temporal constraints (row 4), the increase in image noise degenerates shading quality, especially in dark areas with low signal-to-noise ratio such as around the wheel. Leaving out the non-local temporal constraint causes sharp specular highlights to appear in the reflectance image, e.g., on the bonnet and the near door.

### 6.2 Comparison to Other Methods

We additionally validate our technique through comparisons with recent intrinsic image estimation methods. Fig. 7 exhibits results from conventional Retinex

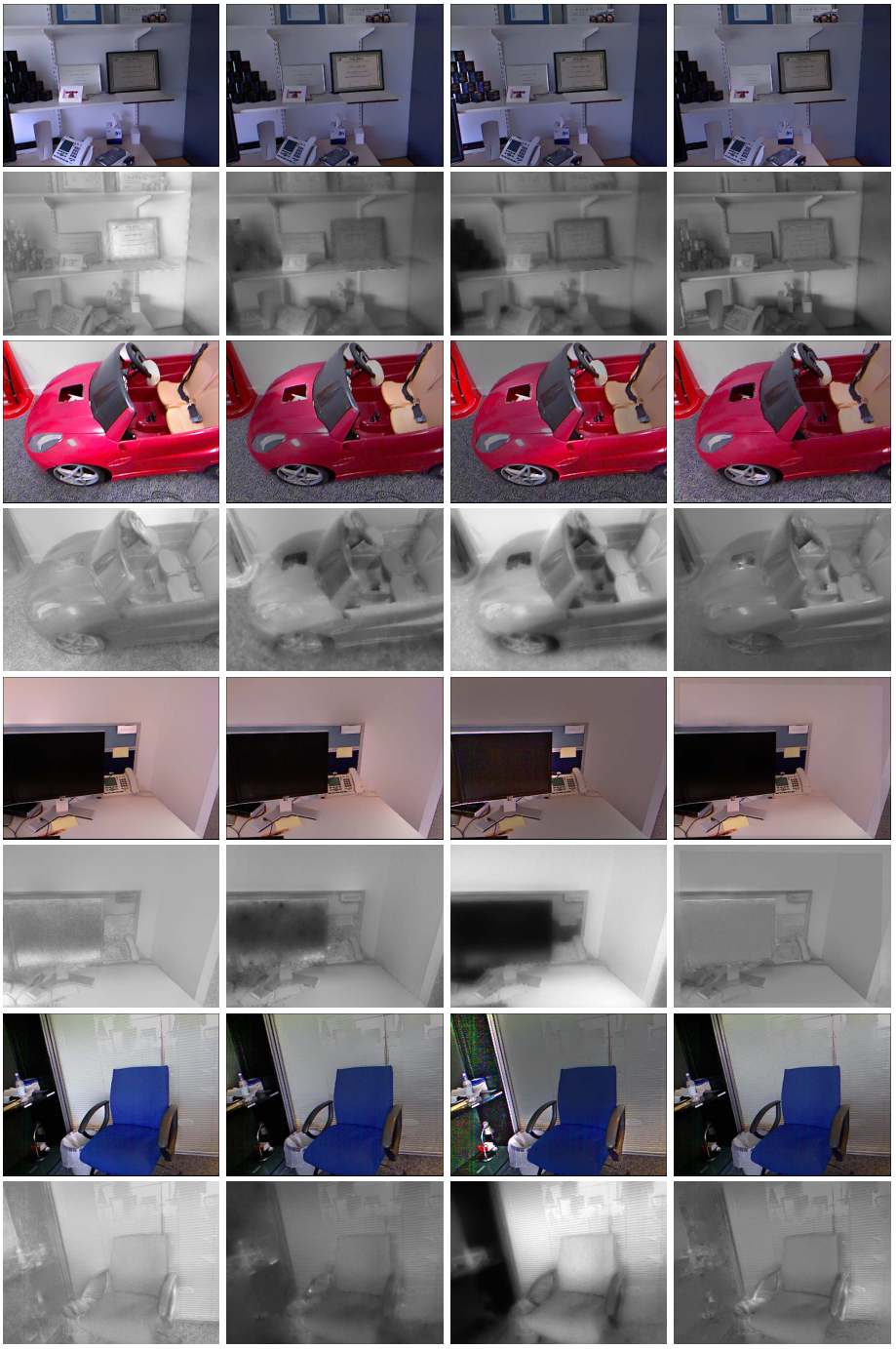

**Fig. 7.** Comparison with other methods. **Column 1** conventional Retinex [21], **Column 2** Shen et al. [8], **Column 3** Gehler et al. [9]. **Column 4** Ours.

(with code downloaded from [21]), Shen et al. [8], Gehler et al. [9], and the proposed method. To view larger images with greater detail, please see the supplementary material. In conventional Retinex, the shading results appear to capture many of the shading variations. However, the lack of shading constraints leads to cast shadows and inconsistent shading in reflectance images (e.g., on the wall of the desk/shelf scene). In addition, a lack of temporal constraints leaves highlights (e.g., car scene) and sharp reflections (e.g., glass wall in chair scene) in reflectance images, and maintains strong noise in shading images (e.g., monitor scene). The results of Shen et al. appear similar to those of conventional Retinex, due to little repetitive texture in the scenes. The decompositions of Gehler et al. also show the effects of not having shading and temporal constraints. For example, the shading on disjoint parallel surfaces (e.g., the far door of the car and the wall; and also the monitor and the wall) are significantly different from each other. We note that our algorithm also suffers from this inconsistency when the non-local shading constraint is not included, such as between the monitor and wall in Fig. 4.

## 7 Conclusion

We presented a technique for intrinsic image estimation based on shading and temporal constraints derived from image+depth video. These constraints come in both local and non-local forms that complement existing local and non-local constraints on reflectance. In our experiments, this framework is shown to yield state-of-the-art decomposition results.

In future work, we plan to roughly estimate the illumination conditions using the reconstructed scene geometry and estimated shading map, and then use this information to bootstrap our intrinsic image estimation process. Rough knowledge of lighting conditions would allow for improvement in the visibility component of the non-local shading constraint, by densely sampling about the major lighting directions and giving greater weight to those visibilities. The non-local temporal constraint may also benefit from rough illumination estimation, e.g., by identifying previous frames with more appropriate viewpoints for reducing the effects of specular highlights at each point.

Our current implementation solves the decomposition of a $640 \times 480$ image frame in 91.7 seconds on average, with only 0.1s for solving the system of equations on the GPU and the rest of time for determining reflectance and shading groups on the CPU. We believe that significantly faster processing could be obtained with a full implementation on parallel graphics hardware that precomputes an initial set of non-local reflectance/shading groups on the GPU, incrementally updates these groups on the fly with newly visible scene points, and initializes the decomposition of each frame by mapping the solution from the previous frame. With this scheme, we hope to compute intrinsic images near or at video frame rates.

**Acknowledgement.** We thank Carsten Rother of MSRC for valuable discussions. Ping Tan is supported by the Singapore MOE grant R-263-000-555-112.

## References


1. Land, E., McCann, J.: Lightness and retinex theory. Journal of the Optical Society of America A 3, 1684–1692 (1971)
2. Funt, B.V., Drew, M.S., Brockington, M.: Recovering Shading From Color Images. In: Sandini, G. (ed.) ECCV 1992. LNCS, vol. 588, pp. 124–132. Springer, Heidelberg (1992)
3. Kimmel, R., Elad, M., Shaked, D., Keshet, R., Sobel, I.: A variational framework for retinex. International Journal of Computer Vision 52, 7–23 (2003)
4. Bell, M., Freeman, W.T.: Learning local evidence for shading and reflectance. In: ICCV, vol. 1, pp. 670–677 (2001)
5. Tappen, M.F., Freeman, W.T., Adelson, E.H.: Recovering intrinsic images from a single image. IEEE Trans. Pattern Analysis and Machine Intelligence 27, 1459–1472 (2001)
6. Sinha, P., Adelson, E.: Recovering reflectance and illumination in a world of painted polyhedra. In: ICCV, pp. 156–163 (1993)
7. Freeman, W., Pasztor, E., Carmichael, O.: Learning low-level vision. International Journal of Computer Vision 40, 24–57 (2000)
8. Shen, L., Tan, P., Lin, S.: Intrinsic image decomposition with non-local texture cues. In: CVPR (2008)
9. Gehler, P.V., Rother, C., Kiefel, M., Zhang, L., Schölkopf, B.: Recovering intrinsic images with a global sparsity prior on reflectance. In: Neural Info. Proc. Systems, NIPS (2011)
10. Shen, L., Yeo, C.: Intrinsic images decomposition using a local and global sparse representation of reflectance. In: CVPR (2011)
11. Bousseau, A., Paris, S., Durand, F.: User-assisted intrinsic images. ACM Trans. Graphics (SIGGRAPH Asia 2009 Issue) 28, 1–10 (2009)
12. Shen, J., Yang, X., Jia, Y., Li, X.: Intrinsic images using optimization. In: CVPR (2011)
13. Weiss, Y.: Deriving intrinsic images from image sequences. In: ICCV, vol. 2, pp. 68–75 (2001)
14. Matsushita, Y., Nishino, K., Ikeuchi, K., Sakauchi, M.: Illumination normalization with time-dependent intrinsic images for video surveillance. In: CVPR, vol. 1, pp. 3–10 (2003)
15. Matsushita, Y., Lin, S., Kang, S.B., Shum, H.-Y.: Estimating Intrinsic Images from Image Sequences with Biased Illumination. In: Pajdla, T., Matas, J. (eds.) ECCV 2004, Part II. LNCS, vol. 3022, pp. 274–286. Springer, Heidelberg (2004)
16. Agrawal, A., Raskar, R., Chellappa, R.: Edge suppression by gradient field transformation using cross-projection tensors. In: CVPR, vol. 2, pp. 2301–2308 (2006)
17. Liu, X., Wan, L., Qu, Y., Wong, T.T., Lin, S., Leung, C.S., Heng, P.A.: Intrinsic colorization. ACM Trans. Graphics (SIGGRAPH Asia 2008 Issue) 27, 152:1–152:9 (2008)
18. Zhao, Q., Tan, P., Dai, Q., Shen, L., Wu, E., Lin, S.: A closed-form solution to retinex with non-local texture constraints. PAMI 34, 1437–1444 (2012)
19. Zhu, S., Wu, Y., Mumford, D.: Filters, random fields and maximum entropy (frame): Towards a unified theory for texture modeling. International Journal of Computer Vision 27, 107–126 (1998)
20. Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S., Freeman, D., Davison, A., Fitzgibbon, A.: Kinectfusion: Real-time 3D reconstruction and interaction using a moving depth camera. In: ACM Symp. User Interface Software and Technology, UIST (2011)
21. Grosse, R., Johnson, M.K., Adelson, E.H., Freeman, W.T.: Ground truth dataset and baseline evaluations for intrinsic image algorithms. In: ICCV (2009)