Shadow Removal from Single RGB-D Images*

Yao Xiao Efstratios Tsougenis Chi-Keung Tang
The Hong Kong University of Science and Technology
{yxiaoab,tsougenis,cktang}@cse.ust.hk

Abstract

We present the first automatic method to remove shadows from single RGB-D images. Using normal cues directly derived from depth, we can remove hard and soft shadows while preserving surface texture and shading. Our key assumption is: pixels with similar normals, spatial locations and chromaticity should have similar colors. A modified nonlocal matching is used to compute a shadow confidence map that localizes well hard shadow boundary, thus handling hard and soft shadows within the same framework. We compare our results produced using state-of-the-art shadow removal on single RGB images, and intrinsic image decomposition on standard RGB-D datasets.

1. Introduction

Shadow removal from single images constitutes an ill-posed problem with more unknowns than equations to solve. State-of-the-art shadow removal methods operating on RGB images use custom capture (e.g., narrow-band camera [7]), user interaction [16], specialized algorithms using texture and gradient similarity [18], chromaticity and Euclidean distance [8] but *none* uses depth cues. With depth, surface normals can be computed and occluding relationship can be inferred, both of which are invaluable to robust shadow removal from single images. See Figure 1 for a comparison of shadow removal with and without depth cues using the present algorithm.

(a) (b) (c)

Figure 1. Shadow removal without and with depth cues. (a) is the input, (b) is the result without depth consideration, (c) result with depth cues, where the spatially varying shadow is seamlessly removed and surface shading is preserved.

While depth and 3D information definitely help, their robust estimation from single RGB images remain difficult. The recent emergence of low-cost depth sensors is likely to overcome this bottleneck. This paper proposes the first shadow removal algorithm from single RGB-D images, leveraging depth cues so that a simple and fully automatic method suffices for robust shadow removal.

Unfortunately, the problem still remains ill-posed since an RGB-D image is still formed by the complex interaction of unknown illumination, albedo and 3D geometry (depths are not true 3D). However, given the same surface, its image typically contains both shadowed and unshadowed pixels of the surface. Otherwise, we might not have perceived the surface as shadowed at all had it been completely under shadow. Using this observation, we translate the shadow removal problem into one of matching unshadowed samples to their shadowed counterparts for 'relighting' the latter.

Normals computed from depth makes a direct contribution to our matching problem: pixels with similar normals, spatial locations and chromaticity should have similar colors in the shadowless image as shown in Figure 1. This assumption has an inherent limitation: lack of unshadowed samples for removing attached shadows where all their normals point away from light. Notwithstanding, while normals provide useful information for each pixel's shading, the chromaticity level is strongly connected to the texture. With no other assumptions used in this paper, both hard and soft shadows with spatially varying intensity can be extracted, while the texture and shading under the shadow are preserved after shadow removal.

Inspired by the recent success of the nonlocal principle in image denoising [3] and matting [10], we introduce a modified nonlocal matching method that is normal-aware to sample relevant unshadowed and shadowed pixels. Using our feature similarity, a method is proposed to work in tandem with raw depth information to compute a shadow confidence map that localizes well hard shadow boundary, thus handling hard and soft shadows within the same framework. A standard energy minimization using the confidence map is then used to automatically produce the optimized shadowless image. Figure 2 previews some of our results.

*The research is supported by the Research Grant Council of the Hong Kong Special Administrative Region under grant no. 619112.

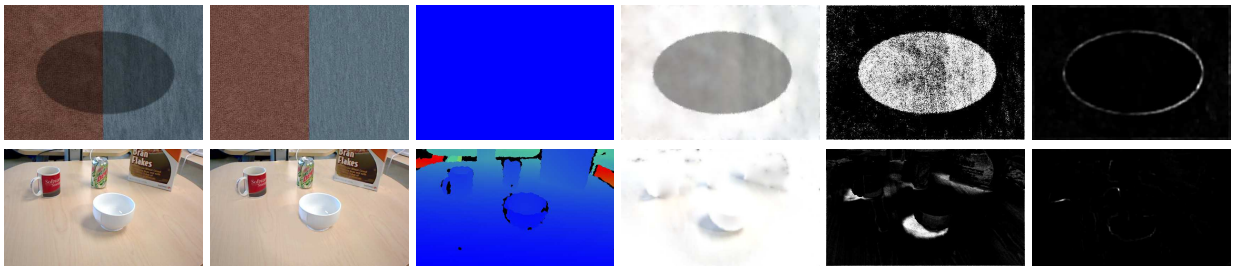


Figure 2. From left to right: input shadowed image, shadowless image, depth information, shadow image, shadow confidence and shadow boundary using our method.

2. Related Work

In the absence of shadow removal with RGB-D images, this section reviews recent and representative works on shadow removal from single images. Intrinsic images separation and pertinent works using single RGB and RGB-D image input are also reviewed.

Shadow Removal. Shadow removal from a single image [5] uses entropy minimization to derive an illumination invariant grayscale image for shadow removal without resorting to any calibration; an advancement over the latter has been made in [7]. These techniques on the other hand make several assumptions, such as Planckian lighting (e.g., sunlight) and narrow-band cameras. Recent work [6] completed the earlier work [5] by introducing a new technique, quadratic entropy with fast Gauss Transform to minimize the entropy. In [16], the proposed interactive technique allows users to mark up shadowed and unshadowed samples with similar textures and then an energy minimization process solves for the underlying shadow. The user-interaction approach has been also followed by [11] and [1]. More recently in [18] the problem of recognizing shadows from monochromatic images was addressed. Variant and invariant cues are applied to a number of classifiers for shadow detection. In [9], hard shadow portions located on the ground are removed by training a decision tree classifier based on sensitive/variant features around edges. Recently, a region-based approach [8] was proposed where pair-wise classification of shadowed and unshadowed regions has led to successful shadow removal results.

Intrinsic Images. Shadow and shading removal are often addressed alongside with intrinsic image estimation. In [15], intrinsic images were separated from a single image by classifying image derivatives as changes due to either reflectance or shading, followed by belief propagation to correct ambiguous regions. The recent commercialization of low-cost RGB-D cameras (e.g., Microsoft Kinect) is likely to make the problem more tractable. In [2] a complex, non-convex optimization was used to obtain a smoothed depth map and a spatially varying illumination model from which the intrinsic images are decomposed. The same problem was addressed in [4] with a simpler approach using nonlocal regularizers for each decomposition component.

3. Shadow Removal with Depth Cues

We first present the image model, followed by presenting our normal-aware nonlocal neighbor and feature matching, which will be used to define the shadow confidence map for subsequent shadow removal.

3.1. Image Model

We use the same image formation equation in [16] which is derived from the image model used for intrinsic image decomposition, but with a different interpretation:

$$I = \beta\mathcal{F} \tag{1}$$

where $\mathcal{F}$ is the shadowless image which includes *shading*, and β is the shadow *only*, a three-channel fractional factor each in $[0, 1]$ for scaling the respective color channel. Since β can be different for different pixels, the equation can handle hard and soft shadows with spatially-varying intensity.

Without normals, in [16] their β cannot distinguish shadow and shading. Thanks to the use of normals from depth for feature matching, our shadow β excludes shading, or equivalently, the shadowless image $\mathcal{F}$ preserves shading as shown in Figure 1.

3.2. Nonlocal Feature Matching

We assign at each pixel a shadow confidence ranged in $[0, 1]$ to indicate the likelihood the pixel is shadowed. Similar to nonlocal denoising and matting [3, 10], a nonlocal neighborhood strategy is adopted to match shadowed and unshadowed pixels except that the region will adapt to the surface normal orientation which is described next.

3.2.1 Normal-Aware Nonlocal Neighborhood

In nonlocal denoising and matting, a spatial and isotropic window is used for matching nonlocal neighbors. In our case, we search within a sufficiently large window to include shadowed and unshadowed pixels which are similar in normals, chromaticity and spatial locations. In particular, we make the searching region normal-aware by orienting it according to the surface normal thus making it anisotropic in the screen space.

The bivariate normal distribution is defined as

$$f_{\mathbf{x}}(x_1, x_2) = \frac{1}{2\pi\sqrt{|\mathbf{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) \quad (2)$$

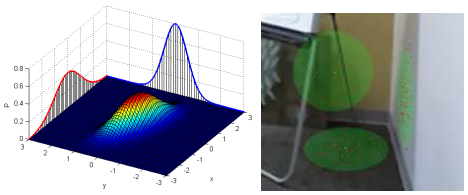
Suppose a pixel p with image coordinates $(x_p, y_p)^T$ has a neighbor pixel q of (x_q, y_q). Then

$$\mathbf{x} \sim \mathcal{N}_2(\boldsymbol{\mu}, \mathbf{\Sigma}) \quad (3)$$

where $\mathbf{x} = (x_q, y_q)^T$, $\boldsymbol{\mu} = E[(x_q, y_q)] = (x_p, y_p)^T$ and $\mathbf{\Sigma}$ is the covariance matrix. Consider the isotropic distribution in the image space,

$$\mathbf{\Sigma}_{iso} = \begin{pmatrix} r^2 & 0 \\ 0 & r^2 \end{pmatrix} \quad (4)$$

where r is the sampling radius for including both shadowed and unshadowed samples. While samples distribution should be spatially isotropic, since most of the surfaces in real scene are not facing the viewing direction, this will result in anisotropic distribution in image space.



(a) (b)

Figure 3. The probability density function of bivariate distribution and the surface normal adaptive sampling are presented in (a) and (b), respectively.

Suppose the samples are isotropically distributed in surface tangent plane, then they are transformed to image space based on the following equation:

$$\begin{pmatrix} \mathbf{x}-\boldsymbol{\mu} \\ z \end{pmatrix} = \mathbf{M}\begin{pmatrix} \mathbf{x}' \\ 0 \end{pmatrix} \quad (5)$$

where $\mathbf{x}'$ denotes the coordinates in the surface tangent plane. $\mathbf{M}$ represents the rotation matrix transforming from the tangent plane to the image plane, which can be easily computed based on the surface normal. Then $\mathbf{x} - \boldsymbol{\mu} = \mathbf{M}'\mathbf{x}'$, where $\mathbf{M}'$ is formed by extracting the first two rows and columns from $\mathbf{M}$. The covariance matrix in the image plane $\mathbf{\Sigma}$ is given by

$$\begin{aligned}\mathbf{\Sigma} = E[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T] &= E[\mathbf{M}'\mathbf{x}'\mathbf{x}'^T\mathbf{M}'] \\ &= \mathbf{M}'\mathbf{\Sigma}_{iso}\mathbf{M}'^T. \end{aligned} \quad (6)$$

3.2.2 Feature Similarity

Feature similarity is used to estimate the visual distance between nonlocal neighbors p and q, denoted by $q \in N_p$ within the above-mentioned window support. For shadow removal, our basic assumption is pixels with similar chromaticity, normals and spatial locations should have similar colors:

$$\alpha^c_{pq} = \exp\left(\frac{\|\mathrm{ch}(I_p) - \mathrm{ch}(I_q)\|^2}{2\sigma_c^2}\right) \quad (7)$$

$$\alpha^n_{pq} = \exp\left(\frac{\|n(p) - n(q)\|^2}{2\sigma_n^2}\right) \quad (8)$$

$$\alpha^d_{pq} = 1 - \frac{\|\bar{p} - \bar{q}\|}{\max\limits_{q\in N_p}\|\bar{p} - \bar{q}\|} \quad (9)$$

where $\mathrm{ch}(I_p)$, $n(p)$ and $\bar{p}$ denote respectively the chromaticity, normal and spatial location of p.

Chromaticity. The chromaticity is adopted as a feature in order to handle the texture parts of the image. The absolute distance is applied *without* the normalization done in [4]. In contrast to [4] where the neighbors are sampled from the whole image, our sample distribution is less global and thus the chromaticity variation is likely to be smaller. The other reason lies on cases where extremely dark shadows slightly corrupt the background's chromaticity. This causes the difference in chromaticity to be magnified by such normalization. We set σ_c to be 0.15 for tolerating chromaticity corruption error.

Normals. Surface normals are estimated based on the depth information. Since the specific process is familiar to computer vision community, all details are provided in the supplementary file for gaining space purposes. Normals help to distinguish shading and shadow, both of which exist in low frequency domain, since normals can indicate whether the illumination change is caused by shading or irradiance blocking. Significant illumination variation for pixels with the same normal is likely to be caused by occlusion. The σ_n parameter is set to 0.5 in our experiments.

Spatial locations. Unlike α^c_{pq}, normalization is done for α^d_{pq} to account for scale variation. While such relative similarity linearly penalizes neighbors away from the center, the distribution radius r can be increased to compensate the effect.

3.2.3 Shadow Confidence

The feature similarity α_{pq} between nonlocal neighbors p and q is

$$\alpha_{pq} = \alpha^c_{pq}\alpha^n_{pq}\alpha^d_{pq} \quad (10)$$

The confidence of p being shadowed, denoted by $C(p)$, is estimated based on the feature similarity between nonlocal neighbors:

$$m_p = \frac{1}{\sum_{q\in N_p}\alpha_{pq}} \sum_{q\in N_p} \alpha_{pq} I_q \quad (11)$$

$$D_p = 1 - \exp\left(\frac{\max(m_p - I_p, 0)^2}{2\sigma^2}\right) \quad (12)$$

$$C(p) = \frac{D_p}{|N_p|} \sum_{q \in N_p} \alpha_{pq} \tag{13}$$

where m_p is the corresponding weighted average intensity based on their similarity. There are three cases:

Both shadowed and unshadowed pixels exist. If p is shadowed, then I_p will tend to be lower than the average thus yielding high confidence. Otherwise the confidence values will be clamped to 0 owing to the max function.

All neighbors are unshadowed pixels. Then I_p will be very close to the average leading to an extremely low confidence.

All neighbors are shadowed. Then the confidence will be low either. However, the proposed sampling strategy with the right choice of r has inhibited this case. Moreover, if it does happen, the term $\frac{1}{|N_p|}\sum_{q \in N_p} \alpha_{pq}$ in Eq. (13), which regulates the shadow confidence by the average similarity, acts as the bootstrap when the estimated confidence is unreliable due to only few similar neighbors being present (i.e., low average similarity).

The complex nature of real scenes imparts unavoidable error during the estimation of the confidence map, which is caused by depth inaccuracy along object boundary and dark textures. Such error can be ameliorated by the smoothing constraint in the optimization which will be introduced in Section 3.3.

3.2.4 Shadow Boundary Confidence

The removal of *hard* shadow boundary over a textured area constitutes one of the most challenging tasks in shadow removal. Background texture and shadow boundary details co-exist locally in the high frequency domain making it difficult to identify the main cause for the irradiance change.

To deal with this complex situation, the prevailing approach is to isolate the boundary information from the texture details intersecting in high frequency in order to estimate the shadow limits. Segmentation-based method [6] is not general due to the *a priori* information provided clear boundaries, and probability-based method [16] relies on user hints. We propose a new approach, namely, shadow boundary confidence, which is based on the assumption that illumination change caused by shadows in a fairly large scale is greater than the one caused by texture.

To make the boundary confidence bias to neither shadow nor non-shadow area, we compute a confidence measure called "nonlocal bright" confidence:

$$B_p = 1 - \exp\left(\frac{\max(I_p - m_p, 0)^2}{2\sigma^2}\right) \tag{14}$$

$$C_p^B = \frac{B_p}{|N_p|} \sum_{q \in N_p} \alpha_{pq} \tag{15}$$

Note that the C previously defined can be regarded as the corresponding "nonlocal dark" version, now denoted as C^D. For both dark and bright confidence we compute the *windowed total variation* $\mathcal{D}$ and *windowed inherent variation* $\mathcal{L}$, which are introduced in [17]:

$$\mathcal{D}_{x,y}^{\{B,D\}}(p) = \sum_{q \in R(p)} g_{p,q} |\partial_{x,y} C_q^{\{B,D\}}| \tag{16}$$

$$\mathcal{L}_{x,y}^{\{B,D\}}(p) = |\sum_{q \in R(p)} g_{p,q} (\partial_{x,y} C_q^{\{B,D\}})| \tag{17}$$

where $R(p)$ is the rectangular region centered at p and $g_{p,q}$ represents a weighing function defined by a Gaussian filter. Intuitively shadow boundary contributes more directional gradients than textures, leading to a larger $\mathcal{L}$. Thus the overall shadow boundary confidence is defined as:

$$V^{\{B,D\}}(p) = \frac{\sqrt{\mathcal{L}_x^{\{B,D\}}(p)^2 + \mathcal{L}_y^{\{B,D\}}(p)^2}}{\sqrt{\mathcal{D}_x^{\{B,D\}}(p)^2 + \mathcal{D}_y^{\{B,D\}}(p)^2} + \epsilon} \tag{18}$$

$$C_p^{bound} = \sqrt{V^B(p) V^D(p)} \tag{19}$$

ϵ is set to prevent zero division. The estimated boundary confidence will be used in the regularization of the smoothing constraint during optimization.

3.3. Shadow Removal

The shadow confidence map is used to optimize the β (and $\mathcal{F} = I/\beta$). The input RGB image is first transformed to the logarithmic domain:

$$i_p = b_p + f_p \tag{20}$$

The energy minimization formulation is:

$$E(\mathbf{b}) = E_F(\mathbf{b}) + \lambda_S E_S(\mathbf{b}) + \lambda_A E_A(\mathbf{b}) \tag{21}$$

where $\mathbf{b}$ is the set of all β to be optimized, $E_F(\mathbf{b})$, $E_S(\mathbf{b})$ and $E_A(\mathbf{b})$ are respectively the shadowless constraint term, the smoothing constraint term and the absolute scale constraint term.

3.3.1 Shadowless Constraint Term

Recalling our basic assumption that unshadowed pixels with similar features are likely to have the same color or illumination. Two pixels p and q with a large similarity α_{pq} tend to have same shadowless image f, that is, $b_p - b_q = i_p - i_q$. The shadowless constraint is defined as

$$E_F(\mathbf{b}) = \sum_p C(p) \sum_{q \in N_p} \alpha_{pq} \|b_p - b_q - (i_p - i_q)\|^2 \tag{22}$$

Recall also that $C(p)$ contains the max function to truncate negative values to 0, which means that pixels brighter than the mean value will be assigned with a confidence value equal to 0. In practice, a threshold $thre_c$ is applied, typically set as 0.1, in order to exclude pixels with low confidence. This significantly reduces the computational load when the majority of the pixels are unshadowed.

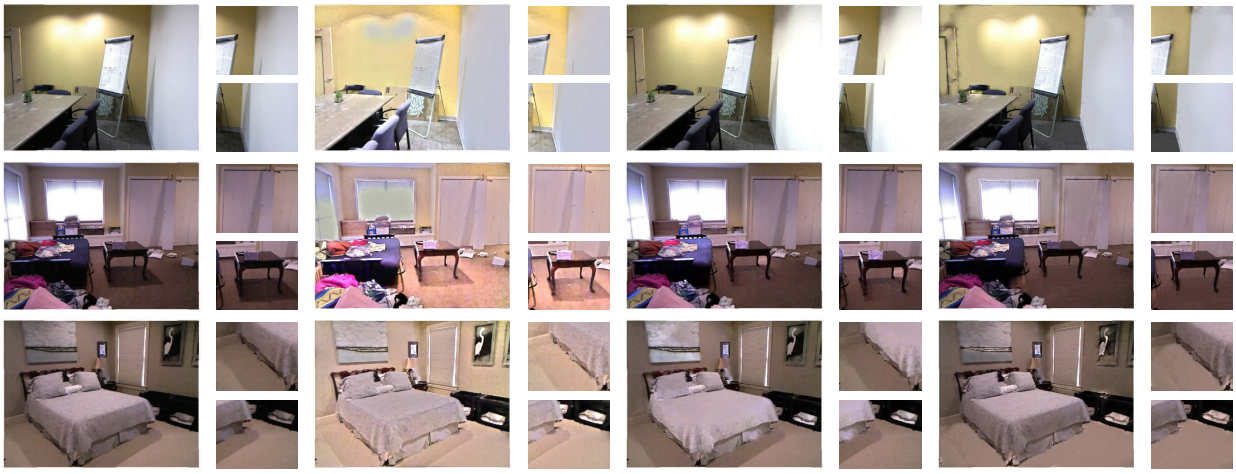


Figure 4. Indoor scenes. From left to right: input shadowed image, the albedo image by Chen and Koltun [4], the shadowless image of Guo et al. [8], and the shadowless image of our method. Specific areas are zoomed for better visualization. Refer to supplemental material for other results.

3.3.2 Smoothing Constraint Term

Consider soft shadows where β varies slowly, in contrast with hard shadows where a rapid change exists across the boundary. A smoothing constraint on β that is aware of hard shadow boundary is needed. The smoothing regularization should also ameliorate sparse errors in the confidence map, which is defined as

$$E_S(\mathbf{b}) = \sum_p (1 - C^B(p)) \sum_{q \in N_p^l} \|b_p - b_q\|^2 \qquad (23)$$

where N_p^l denotes the local spatial neighbors.

3.3.3 Absolute Scale Constraint Term

During the optimization process, there exist pixels that should not participate in the process: they are neither high confidence pixels ($N_C = \{p | C(p) > thre_c\}$) nor neighbors of high confidence pixels ($N_N = \{\bigcup_{p \in N_C} \{q | q \in N_p\}\}$). Pixels not existing in the union of the two sets are the ones that should maintain their colors (i.e., $\beta = 1$) and should not be involved in the shadowless constraint. Since pixels involved in this term are usually located relatively distant from the shadow area and only smoothing regularization will be imposed on them, they have little impact to the shadowless regularization.

$$E_A(\mathbf{b}) = \sum_{p \in N \setminus (N_C \cup N_N)} \|b_p - 1\|^2 \qquad (24)$$

where N represents the whole pixel set.

The absolute scale constraint is also essential because the shadowless constraint reconstructs β up to a scale, where in Eq. (22) $\|((b_p - \gamma) - (b_q - \gamma)) - (i_p - i_q)\|^2$ for any real and positive γ can be used to produce the same effect.

4. Experimental Results

This is a first major attempt to demonstrate depth cues can significantly improve shadow removal results with a simple strategy as described, and we expect other state of the art shadow removal algorithms to benefit when depth cues are taken into consideration. All softwares have been developed in the MATLAB 2012b environment, while all experiments were executed on a laptop with an Intel Core Duo 3.00GHz CPU with 8GB RAM. For an image of size 640 x 480 it typically takes 5 to 7 minutes, noting that the processing time varies with the total number of unshadowed pixels and those that are uninvolved in the energy minimization.

Due to space limitation we highlight a subset of our results in the paper. Refer to the supplemental material for all the results. A set of 30 images were collected from two different datasets commonly applied in RGB-D method evaluations. The NYU dataset [14] comprises of indoor scenes while the Cornell dataset contains both outdoor [13] and synthetic [12] RGB-D images. The goal here is to demonstrate that the proposed algorithm constitutes a generic solution independent of the scene nature (indoor, outdoor or synthetic) as compared to the latest state of the arts [8] and [4]. Chen and Koltun [4] is the most recent state of the art for intrinsic image separation from single RGB-D images. Their albedo image, which is supposed to be free of shading and shadow, is compared while noting that this is not exactly the shadowless image we optimize for. In the absence of shadow removal methods that use depth cues to the best of our knowledge, Guo et al. [8] is compared since it is the state-of-the-art shadow removal focusing on outdoor scenes with considerably high quality results. Moreover, since Finlayson et al. [5, 7] results for outdoor images are state of the art, comparison will be presented in supplementary material for gaining space. A comparison with interactive tech-

Figure 5. Outdoor scenes. From left to right: the house-yard, the road and the school-yard are depicted. For each scene, we respectively show the input shadowed image, the shadowless image by Guo et al. [8], the albedo image by Chen and Koltun [4], and the shadowless image by our method. Refer to supplemental material for other results.

niques [16, 1] will also be presented based on the same used images, since the high complexity of shadows and the scene in the indoor/outdoor images in [14, 13] does not allow easy user interaction. Although quantitative comparison is rare for previous shadow removal works, we give one based on the synthetic images, by comparing the Mean Square Error (MSE) and the Structural Similarity Index Method (SSIM) between the corresponding ground truth shadowless images and the shadowless results.

4.1. Indoor Scenes

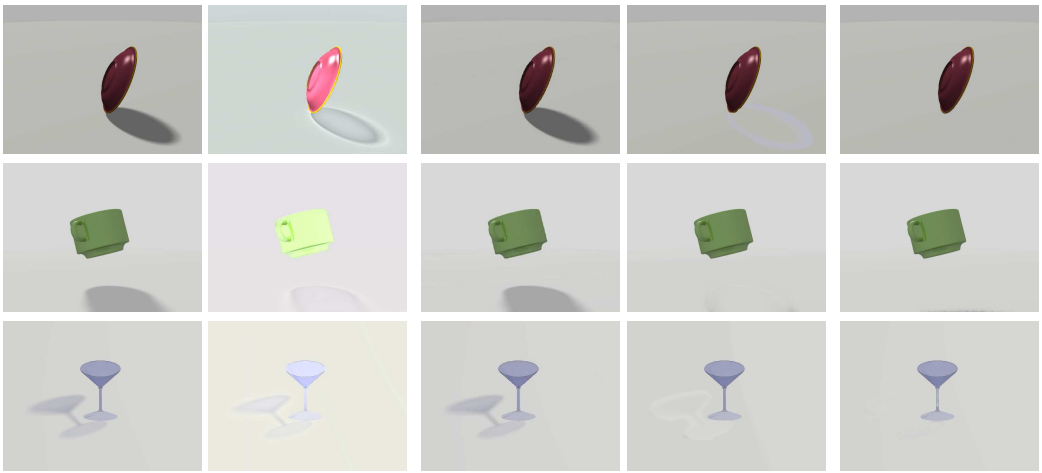
Ten indoor images are selected from the NYU dataset which was captured by both the RGB and depth camera using the Microsoft Kinect. The corresponding aligned depth maps are also provided and used in our method and [4]. Since [8] does not take into consideration the depth information, only the RGB image is used as the algorithm's input. Figure 4 shows the most representative results.

The complexity of the selected indoor images is obvious with all multi-sized objects casting shadows to the rest of the scene. The majority of soft or hard shadows are successfully detected and removed by our method. In [4], the shadow portions of the images are also detected, a fact that also justifies the significance of depth. However, there exist portions that have been missed basically due to its global nature. Specifically, a globally spatial smoothing constraint is imposed as regularization term of indirect illumination. Therefore, a locally rapid change of illumination cannot be handled successfully. In Guo et al. [8] the matching pair approach considering the RGB average intensity or chromaticity features partially removes the shadowed portions. However, this method takes into consideration the environmental light (reflections) and the direct light (sun), the latter missing in indoor scenes and thus the shadow detection performance is adversely affected. Conclusively, it should be noted that despite our high shadow removal performance, attached shadow areas (see bed sides in Figure 4) where normals point away from light are better handled by [8] and [4].

4.2. Outdoor Scenes

The complexity of outdoor scenes is widely known to the shadow detection community raising the challenge level. Since depth information is necessary in our method, the dataset published by Saxena et al. [13] has been selected for this experiment. The specific dataset contains numerous outdoor images along with the corresponding depth maps, 10 of which are selected for evaluating the tested algorithms.

The results depicted in Figure 5 justify the good performance of [8], since the majority of the shadow portions are detected. However the chromaticity of the shadowless images seems to be locally affected. In [4], even the most complex shadowed areas are detected but the removal process is not satisfying. On the contrary, the proposed method manages to detect and remove the majority of the shadowed portions producing high quality results favorably compared to [8]. However, for outdoor scenes under strong sunlight, the chromaticity of shadow regions is usually corrupted making it difficult to strike a good balance between removing shadows and preserving texture. It should be also noted though that raw outdoor depth information can be often inaccurate affecting the algorithm's performance i.e. missing depth information in distant areas like background trees, building's facade and school building's passageway in Figure 5 leads to failure shadow removal cases. Although [8] does not have to cope with this issue using only the RGB



$\mathcal{I}$ [4] [8] [16] Ours

Figure 6. Synthetic scenes. The plate, the mug and the martini glass are depicted. For each scene, we respectively show the input shadowed image, the albedo image by Chen and Koltun [4], the shadowless image by Guo et al. [8], the shadowless image by Wu et al. [16] and the shadowless image by our method. Refer to supplemental material for other results.

image, still our results are convincing in most cases.

4.3. Synthetic Scenes

A dataset created by Saxena et al. [12] provides us with numerous RGB-D synthetic images illustrating single objects along with their corresponding shadows. Considering their shadow complexity, 10 images have been selected for the experimental evaluation of our algorithm. Some sample results are depicted in Figure 6.

Although the synthetic scenes are arguably simple, they still present challenges to the state-of-the-art methods. The absence of direct lighting condition seriously affects the results produced by Guo et al. [8] which is designed to operate mainly on outdoor images. On the contrary, Chen and Koltun [4] successfully detects and partially removes the shadow parts of the image. However, the boundary effect cannot be properly smoothed the same as [16]. Since [4] is focused on distinguishing the reflectance variation from shading, the aforementioned phenomenon is not considered within the optimization process. In [16], the boundary effect is considerably better but still noticeable. On the other hand, the shadow boundary confidence (Eq. 13) in our method can alleviate the boundary artifact.

Since the ground-truth images can be easily created for the synthetic images, a quantitative comparison is also presented hereafter. The perceptual quality of the shadowless image $\mathcal{F}$ is assessed based on SSIM which takes the HVS into consideration. In addition, the MSE measuring the intensity distance between $\mathcal{F}$ and the ground truth images is calculated for: the local foreground object, the local background under the extracted shadow, and the complete image. The results indicate that our proposed algorithm satisfies the perceptual requirement preserving also the error in low levels. Note that the foreground area in [16] is masked off from any processing achieving slightly better perceptual results.

	MSE Foreground	MSE Background	MSE $\mathcal{F}$	SSIM $\mathcal{F}$
[16]	-	2.791	4.667	0.9974
[8]	39.540	342.473	312.560	0.9846
Ours	32.630	1.139	3.675	0.9956

Table 1. Quantitative results on synthetic images. The average MSE and SSIM values of [16], [8] and our method.

4.4. User-assisted Methods

Shadow removal using user-assisted methods are hard to be applied in the outdoor images [13] and indoor scenes [14] due to their high complexity. Therefore, it seems fair to have a comparison with [1] and [16] based on their image results and in particular, on textured images. Since they do not use depth, we turn off the depth cue in our algorithm for comparison. Figure 7 shows that our algorithm produces the shadowless images which are comparable if not of higher quality, justifying our texture treatment using chromaticity, windowed total variation and windowed inherent variation in [17]. Moreover, our method is fully automatic and can be applied to complex indoor or outdoor scenes.

4.5. Application

Using our shadow removal as preprocessing, we found that intrinsic image separation can be significantly enhanced. For instance, since the goal in [4] is to decompose a single RGBD image into the corresponding albedo and shading image, the use of the shadowless image produced by our algorithm contributes considerably in their final results as shown in Figure 8. We expect other state-of-the-art methods can directly benefit using our shadowless images.

(a) (b) (c) (d)

Figure 7. Comparison with user-assisted shadow removal on textured surfaces. (a) is the input. (b), (c) and (d) are shadowless images respectively produced by [1], [16] and our method.

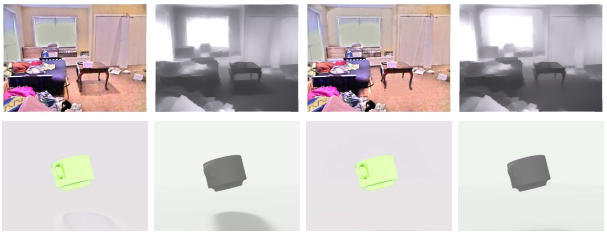
5. Conclusions and Limitations

We have presented a novel method that capitalizes depth cues to successfully detect and remove shadows from still images. Surface normals are estimated from depth obtained from low-cost sensor for the automatic shadow detection and removal on real-world photos. Our method uses a modified nonlocal matching where feature similarity is defined by normals, chromaticity, and spatial locations. The proposed shadow confidence boundary model detects both hard and soft shadows and adaptively applies smoothing along the boundary. The performance of our algorithm has been evaluated using indoor, outdoor and synthetic datasets. A comparison with four related algorithms have demonstrated the high quality results.

Generally our method's performance can be affected by the non-existence of corresponding unshadowed samples. A satisfying ratio between shadowed and unshadowed pixels has been experimentally justified to be at least 1:1. However, failure cases are generated by dealing with attached shadows where the corresponding normals point away from light and areas and completely under shadow areas, both of them lacking of unshadowed samples. Moreover, the algorithm's performance is connected to the accuracy of the input depth map, while noting that we do not need very accurate depth to produce good results as shown in our experimental section. Aforementioned limitations/failure cases for single-image RGB-D shadow removal can in fact be overcome, e.g., with better depth sensors and relaxing full automation: one mouse click to cluster normals of the same material for removing attached shadows. Conclusively, datasets with outdoor and synthetic RGB-D images with accurately aligned depth maps are still missing. With the growing popularity of RGB-D sensors, we expect a more thorough RGB-D datasets for shadow removal is underway.

References

[1] E. Arbel and H. Hel-Or. Shadow removal using intensity surfaces and texture anchor points. *TPAMI*, 33(6):1202–1216, 2011. 2, 6, 7, 8



[4] [4] + Ours

Figure 8. Albedo and shading images from intrinsic image decomposition, without and with shadow removal preprocessing.

[2] J. T. Barron and J. Malik. Intrinsic scene properties from a single rgb-d image. *CVPR*, 2013. 2

[3] A. Buades, B. Coll, and J.-M. Morel. Nonlocal image and movie denoising. *IJCV*, 76(2):123–139, 2008. 1, 2

[4] Q. Chen and V. Koltun. A simple model for intrinsic image decomposition with depth cues. In *ICCV*, 2013. 2, 3, 5, 6, 7, 8

[5] G. D. Finlayson, M. S. Drew, and C. Lu. Intrinsic images by entropy minimization. In *ECCV*, pages 582–595. 2004. 2, 5

[6] G. D. Finlayson, M. S. Drew, and C. Lu. Entropy minimization for shadow removal. *IJCV*, 85(1):35–57, 2009. 2, 4

[7] G. D. Finlayson, S. D. Hordley, C. Lu, and M. S. Drew. On the removal of shadows from images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(1):59–68, 2006. 1, 2, 5

[8] R. Guo, Q. Dai, and D. Hoiem. Paired regions for shadow detection and removal. *TPAMI*, Preprint, 2012. 1, 2, 5, 6, 7

[9] J.-F. Lalonde, A. A. Efros, and S. G. Narasimhan. Detecting ground shadows in outdoor consumer photographs. In *ECCV*, pages 322–335. 2010. 2

[10] P. Lee and Y. Wu. Nonlocal matting. In *CVPR*, pages 2193–2200, 2011. 1, 2

[11] F. Liu and M. Gleicher. Texture-consistent shadow removal. In *Computer Vision–ECCV 2008*, pages 437–450. Springer, 2008. 2

[12] A. Saxena, J. Driemeyer, J. Kearns, C. Osondu, and A. Y. Ng. Learning to grasp novel objects using vision. In *ISER*, 2006. 5, 7

[13] A. Saxena, M. Sun, and A. Y. Ng. Make3d: Learning 3d scene structure from a single still image. *TPAMI*, 31(5):824–840, 2009. 5, 6, 7

[14] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus. Indoor segmentation and support inference from rgbd images. In *ECCV*, 2012. 5, 6, 7

[15] M. F. Tappen, W. T. Freeman, and E. H. Adelson. Recovering intrinsic images from a single image. *TPAMI*, 27(9):1459–1472, 2005. 2

[16] T.-P. Wu, C.-K. Tang, M. S. Brown, and H.-Y. Shum. Natural shadow matting. *TOG*, 26(2):8, 2007. 1, 2, 4, 6, 7, 8

[17] L. Xu, Q. Yan, Y. Xia, and J. Jia. Structure extraction from texture via relative total variation. *ACM Transactions on Graphics (TOG)*, 31(6):139, 2012. 4, 7

[18] J. Zhu, K. G. Samuel, S. Z. Masood, and M. F. Tappen. Learning to recognize shadows in monochromatic natural images. In *CVPR*, pages 223–230, 2010. 1, 2