Interactive Shadow Removal and Ground Truth for Variable Scene Categories

Han Gong
http://www.cs.bath.ac.uk/~hg299
Darren Cosker
http://www.cs.bath.ac.uk/~dpc

Media Technology Research Centre
Department of Computer Science
University of Bath
Bath, UK

Abstract

We present an interactive, robust and high quality method for fast shadow removal. To perform detection we use an on-the-fly learning approach guided by two rough user inputs for the pixels of the shadow and the lit area. From this we derive a fusion image that magnifies shadow boundary intensity change due to illumination variation. After detection, we perform shadow removal by registering the penumbra to a normalised frame which allows us to efficiently estimate non-uniform shadow illumination changes, resulting in accurate and robust removal. We also present the first reliable, validated and multi-scene category ground truth for shadow removal algorithms which overcomes limitations in existing data sets – such as inconsistencies between shadow and shadow-free images and limited variations of shadows. Using our data, we perform the most thorough comparison of state of the art shadow removal methods to date. Our algorithm outperforms the state of the art, and we supply our P-code and evaluation data and scripts to encourage future open comparisons.

1 Introduction

Shadows are ubiquitous in image and video data, and their removal is of interest in both Computer Vision and Graphics. Although shadows can be useful cues, e.g. shape from shading, they can also affect the performance of algorithms (e.g. in segmentation and tracking). Their removal and editing is also often the pain-staking task of graphical artists. A successful shadow removal method should seamlessly relight the shadow area while keeping the lit area unchanged. A shadow is generally defined as having an umbra and penumbra area – denoted by the central shadow region and its border (penumbra) transitioning illumination between the fully dark and lit area. A shadow image I_c can be considered as a Hadamard product of a shadow scale layer $\mathcal{S}_c$ and a shadow-free image $\mathcal{I}_c$ as shown in Eq. 1 where c is a RGB channel. The scales of the lit area are 1 and other areas' scales are between 0 and 1.

$$I_c = \mathcal{I}_c \circ \mathcal{S}_c \tag{1}$$

In this paper, we propose an interactive, high-quality and robust method for fast shadow removal using two rough user-defined strokes indicating the shadow and lit image areas. Our approach sacrifices full autonomy for very broad and simple user input – contrasting with existing manual approaches that require fine-scale input (accurate shadow contours)

or highly simplistic inputs (single pixel) that can result in shadow detection artefacts. Our on-the-fly learning approach is robust to large variations in user input. Given detection, we deliver reliable shadow removal – verified with thorough quantitative tests for different types of shadow scene (for the first time in this area) comparing to previous state of the art approaches. We also present a large high-quality and multi-scene category ground-truth data set for the evaluation of shadow removal. This is quantitatively verified to ensure registration and illumination errors between test and ground truth images are eliminated. Our approach represents a state of the art method for shadow removal, with the most robust comparative evaluation of such approaches to date.

1.1 Related work

Approaches to shadow removal can be categorised as either automatic [5, 13] or user-aided [1, 7, 9, 10, 11, 12]. The differentiation between fully automatic or user-aided relates to initial detection of the shadow – with removal itself (after detection) being a largely automatic task. In any case, both removal and detection are ill-posed problems and difficult to reliably achieve. Intrinsic image based methods are one such popular approach to the problem, such as [5, 13]. The decomposition of intrinsic images provides shading and reflectance information but can be unreliable leading to over-processed results. Other approaches [8, 14] detect shadows by learning shadow features. However, detection is constrained by the range of training data and quality of image edge detection and segmentation there-in. User-aided methods generally achieve higher accuracy in shadow detection at the practical expense of varying degrees of manual assistance. Wu *et al.* [12] require the user to define multiple regions of shadow, lit area, uncertainty and exclusion. Others [9, 11] require input that carefully defines the shadow boundary. Shor and Lischinski [10] only require one shadow pixel as input, but have limitations in cases where the other shadowed surfaces are not surrounded by the initially detected surface. Arbel and Hel-Or [1] require users to specify multiple texture anchor points. [5, 10] apply native in-painting for penumbra recovery which result in penumbra artefacts. Most others [1, 7, 9] assume highly-constrained curve or surface functions for illumination change which limit their range of removable shadows.

To date, most shadow removal methods [1, 7, 11, 12] have been evaluated by visual inspection on some selected images – with only a few exceptions performing quantitative evaluation. This is in part due to a lack of high-quality, varied, and public ground truth data. [10] performs a quantitative test but comparison is difficult due to the their data not being publicly available. Guo *et al.* [8] provide the first public ground truth data set for shadow removal and perform quantitative testing. However, the difficultly of collecting such a data set is highlighted in their work, with the appearance some global illumination changes and mis-registration between the shadow and shadow-free images being a difficult factor to control. This can make quantitative testing on such data somewhat difficult, as these errors can influence shadow removal results.

1.2 Contributions

Given our overview of state of the art approaches, we propose three main contributions:
1) A rigorous, highly-varied and categorised shadow removal ground truth data set: Our quantitatively verified high quality data set contains a wide range of ground truth data organised into common shadow categories. Based on this data, we quantitatively evaluate our method against other state of the art algorithms on different shadow category types.
2) Simple user input based shadow detection: Our shadow detection component requires

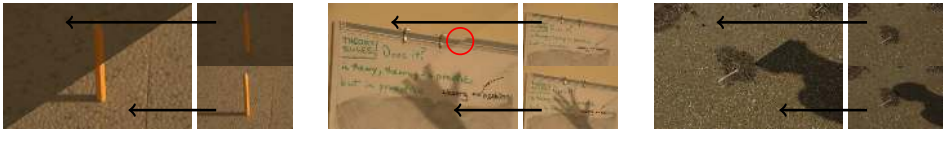
only two rough user scribbles marking samples of lit and shadow pixels. Our approach differs from previous work requiring more complex user-inputs or simpler inputs that compromise robustness and quality.

3) High quality and fast shadow removal: Unlike existing methods requiring slow pixel-wise optimisation or an inflexible fitting model, we introduce penumbra unwrapping to derive a few unified intensity samples across the penumbra for robust and efficient estimation of illumination changes without requiring prior training. Our method is simple and fast yet offers state of art shadow removal quality.

To summarise, we believe our contributions are important to this area of research due to our significant improvements over the state of the art in shadow removal in a wide range of thorough and repeatable tests – what we believe to be the most thorough to date.

2 Shadow removal ground truth

A thorough quantitative evaluation of shadow removal performance requires a high-quality, diverse shadow-free ground truth. The first public data set was presented by Guo *et al.* [8]. In our work, we propose a new data set that introduces multiple shadow categories, and overcomes potential environmental illumination and registration errors between the shadow and ground truth images. An example of comparison between our new data and a previous data set is shown in Fig. 1, which highlights the potential difficulties that mis-registration and illumination instability can present. In § 4, we quantitatively compare the quality of our ground truth data to the previous state of the art. We capture shadow images and their ground



(a) mismatched illuminaiton (b) unregistered pixels (c) our data (no artefacts)

Figure 1: For each image: top left segment – shadow-free image; bottom right segment – shadow image. (a) and (b) are taken from [8] which reflect the two annotated issues. An example from our data – which rejects image pairs with these properties is shown in (c).

truth using a camera with a tripod and a remote trigger. This setup minimises misalignment due to camera shake. To minimise illumination variance, we capture images in a very short interval of time using a manual capture mode with fixed ISO and exposure settings. When collecting data, environmental effects are often unavoidable, e.g., wind can cause camera shake or the sun might move behind the clouds. Such failed acquisitions are thus rejected from our data set using a quantitative assessment which we outline in §4. For evaluation, our shadow data are also categorised according to 4 different attributes: degree of texture, shadow softness, brokenness of shadow, and colour variation. In total, our final data set after rejection consists of 186 test cases. We compare our new data against a previous state of the art data set in our evaluation.

3 Interactive Shadow Removal

In this section, we overview our algorithm first in brief, and then expand on technical details for each of its components. Our algorithm consists of 3 steps (see Fig. 2):

1) Pre-processing (§3.1) We detect an initial shadow mask (Fig. 2(b)) using a KNN classifier trained from data from two rough user inputs (e.g. Fig. 2(a)). We generate a *fusion image*, which magnifies illumination discontinuities around shadow boundaries, by fusing channels of YCrCb colour space and suppressing texture (Fig. 2(c)).
2) Penumbra unwrapping (§3.2) Based on the detected shadow mask and fusion image, we sample the pixel intensities of sampling lines perpendicular to the shadow boundary (Fig. 2(d)), remove noisy ones and store the remaining as columns for the initial penumbra strip (Fig. 2(e)). We align the initial columns' illumination changes using its intensity conversion image (Fig. 2(f)). This results in an aligned penumbra strip (Fig. 2(g)) whose conversion image (Fig. 2(h)) exhibits a stabler profile.
3) Estimation of shadow scale and relighting (§3.3) From the penumbra strip, we synthesise a few unified samples, e.g. Fig. 2(i), of intensity change which allows for a fast and robust estimation of illumination change along sampling lines to derive the sparse scales for all sampled sites (Fig. 2(j)) which are propagated to form a dense scale field (Fig. 2(k)). We remove shadows by inverse scaling using this non-uniform field (Fig. 2(l)).

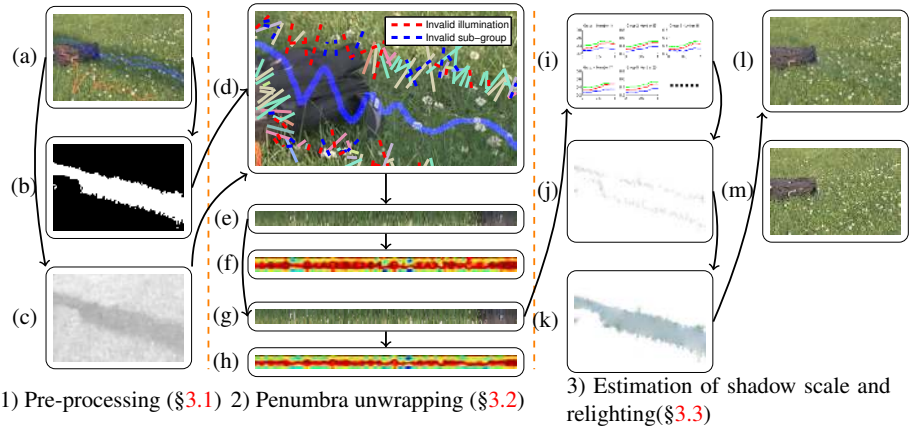


Figure 2: Our shadow removal pipeline. (a) input: a shadow image and user strokes (blue for lit pixels and red for shadowed pixels); (b) detected shadow mask; (c) fusion image; (d) initial penumbra sampling (solid lines in different colours indicate valid samples of different sub-groups. Dashed lines are invalid samples removed using two criteria); (e) initial penumbra regularisation; (f) initial penumbra conversion image; (g) final penumbra regularisation; (h) final penumbra conversion image; (i) penumbra illumination estimation; (j) sparse shadow scale; (k) dense shadow scale; (l) output; (m) ground truth.

3.1 Pre-processing

Pre-processing provides a detected shadow mask and a fusion image to assist penumbra unwrapping. Although there have been automatic methods for shadow detection, results are dependent on training data quality and variation. Instead, our method requires no prior training or learning – only two user-supplied rough inputs indicating sample lit and shadow pixels (Fig. 2(a)). We supply the marked pixels' RGB intensities in the Log domain as the training features and construct a KNN classifier. We choose the Euclidean distance as the distance measure and the majority rule with nearest point tie-break as the classification measure. We apply spatial filtering with a Gaussian kernel (size = 5, SD = 3) to the obtained

image of posterior probability and binarised filtered image using a threshold of 0.5. Although detection errors along the boundary, as well as post-filtering, can result in intensity samples with unsynchronised illumination changes at sharp boundary, our penumbra unwrapping and alignment step (§3.2) can compensate for this. Thus, our shadow *removal* method is somewhat robust to noise in the initially detected shadow mask, and would also be applicable to alternative (e.g. automatic) detection methods.

To assist unwrapping of the penumbra, we derive an image that magnifies illumination discontinuities around the shadow boundary – also assisting penumbra location – which we call the *fusion image* (e.g. Fig. 2(c)). There are two steps: **1) Magnification of illumination discontinuity** We derive an initial fusion image $\mathcal{F}$ that maximises the contrast between shadow and lit areas by linearly fusing the three channels ($\mathcal{C}_l$) of YCbCr space as follows:

$$\mathcal{F} = \sum_{l=1}^{3} a_l \mathcal{C}_l \text{ subject to } \sum_{l=1}^{3} a_l = 1 \tag{2}$$

where a_l is the fusing factor of $\mathcal{C}_l$ (positive). The best fusing factors are derived by minimising the following objective function E_b:

$$E_b(\mathbf{a}) = \mu(\mathcal{F}_\mathbf{S})/\mu(\mathcal{F}_\mathbf{L}) + (\sigma(\mathcal{F}_\mathbf{S}) + \sigma(\mathcal{F}_\mathbf{L}))/\sigma(\mathcal{F}_{\mathbf{S}\cup\mathbf{L}}) \tag{3}$$

where $\mathbf{a}$ is the vector of fusing factors and $\mathcal{F}_\mathbf{S}$ and $\mathcal{F}_\mathbf{L}$ are the two sets of shadow and lit pixels marked by user scribbles. In this paper, we define σ and μ as functions that respectively compute the standard derivation and mean of a set of values. The first term ensures larger distinction between pixels of lit and shadow regions and the second term ensures smaller variation for pixels of the same lit or shadow regions. **2) Suppression of texture** We reduce image texture by applying a median filter with a 10-by-10 neighbourhood to $\mathcal{F}$.

In our experiments, we found YCbCr colour space to offer more perceptually meaningful information, while illumination in RGB space can be affected by texture noise. An example comparison of fusing channels in YCrCb colour space compared to RGB colour space is shown in Fig. 3.

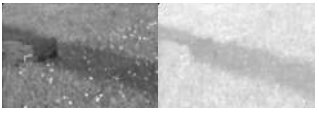


Figure 3: Comparison of colour space for fusion: The same optimisation scheme is applied to the image in Fig. 2(a) but using different colour spaces. Texture noise, such as in the flowers, appears more pronounced in the RGB fusion image (left) whilst the YCrCb fusion image (right) presents cleaner illumination information.

3.2 Penumbra unwrapping

The shadow boundary generally has a noisy profile with variable penumbra width. This can lead to inaccurate estimation of shadow scales and resulting artefacts. We therefore unwrap the penumbra into a strip and align its sampled columns of illumination change. This improves the detection of outliers and allows linearisation of processing in the penumbra – leading to significant gains in efficiency and speed (see Fig. 2(g)).

We sample the intensity of sampling lines perpendicular to the shadow boundary (Fig. 2(d)) as columns for the initial penumbra strip. The length of a sampling line is determined by locating suitable start and end points guided by the fusion image $\mathcal{F}_p$. We start a bi-directional search from each boundary point that extends the sampling line towards the lit area (end point) and the shadow area (start point) as described in Algorithm 1. We initially set the start and end points as the boundary point (x_b, y_b) and the direction vector $\Delta\mathbf{v}$ as the normalised

Algorithm 1: Penumbra sample end point selection

input : boundary point (x_b, y_b), fusion image $\mathcal{F}_p$
output: two ends ($\mathbf{p}_s$, $\mathbf{p}_e$) of a sampling line
$\widetilde{F} \longleftarrow \nabla\mathcal{F}_p$; $\mathbf{p}_s \longleftarrow (x_b, y_b)$; $\mathbf{p}_e \longleftarrow (x_b, y_b)$; $\mathcal{L} \longleftarrow |\widetilde{F}(x_b, y_b)|$; $\Delta\mathbf{v} \longleftarrow \widetilde{F}(x_b, y_b)/\mathcal{L}$;
repeat
  $\mathbf{v}_s \longleftarrow \widetilde{F}([\mathbf{p}_s])$; $\mathbf{v}_e \longleftarrow \widetilde{F}([\mathbf{p}_e])$; $\mathcal{L}_s \longleftarrow \mathbf{v}_s \cdot \Delta\mathbf{v}$; $\mathcal{L}_e \longleftarrow \mathbf{v}_e \cdot \Delta\mathbf{v}$;
  $\mathbf{p}_s \longleftarrow \mathbf{p}_s - \Delta\mathbf{v}$; $\mathbf{p}_e \longleftarrow \mathbf{p}_e + \Delta\mathbf{v}$;
until $\mathbf{p}_s$ *or* $\mathbf{p}_e$ *is not within the range of* $\mathcal{F}$ ***or*** $10\mathcal{L}_s > \mathcal{L}$ *or* $10\mathcal{L}_e < \mathcal{L}$;

gradient vector of (x_b, y_b). To get the position for a start point, we iteratively subtract $\Delta\mathbf{v}$ from the start point until its projected gradient is small enough (vice versa for the end point).

To avoid outliers, e.g. sampling lines at occlusion boundaries, we filter invalid samples based on an assumption of similar shadow scales. We first compute a scale vector $Y_c = T_l - T_s$ where T_l and T_s are the average Log-domain RGB intensities of the lit and shadow halves of a sample. We convert Y_c to spherical coordinates as a feature vector Y_s. We apply DBSCAN clustering [4] (radius:0.2) to Y_s of all samples and store the samples that belong to the largest cluster as valid ones with valid illumination. For finer scale estimation, we divide the valid cluster into a few sub-groups using mean-shift [3] (band width:0.06) and discard the samples of the invalid sub-groups whose numbers are less than 10% of the largest sub-group's. Fig. 2(i) shows an example of the above sample categorisation.

As the lengths of samples are different, we normalise this by re-sizing all the samples to their average length n_a. The normalised samples are concatenated as columns to form the initial penumbra strip. The vertical illumination change of columns at this stage may still be misaligned. We therefore resolve this using fine-scale alignment focusing on vertical centre adjustments and scaling of the columns. Fig. 2(g) shows an example of the aligned strip and Fig. 4 visualises the process of alignment. To assist alignment, we amplify illumination changes by deriving a fusion image $\mathcal{W}_p$ of the strip. Similar to our previous formulation, we first fuse the strip using Eq. 2 with factors $\mathbf{a}$ and apply an ideal low-pass frequency filter using a cut-off $k_w = \max([n_a/8], 2)/n_a$. To locate the centre, we generate a conversion image $\mathcal{O}(i) = -|\mathcal{W}_p(i) - \mu(\mathcal{W}_p(i))|$, e.g. Fig. 2(f), where i is the column number. In $\mathcal{O}$, the centres of the columns of strips appear as peaks. We can find the peaks of columns from $\mathcal{O}$ and vertically align based on these.

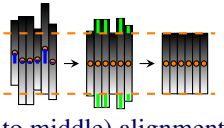


Figure 4: Alignment of penumbra strip: The orange circles are the centres of columns in the penumbra strip. The orange dashed lines indicate the desired column length for a strip. The blue and green lines indicate the shifts required. The alignment is in two steps: (left to middle) alignment of centre; (middle to right) alignment of illumination change.

To further ensure the rates of illumination change in columns are the same, we vertically stretch each column about its centre by shifting the two ends of a column. Before alignment, we derive another fusion image $\mathcal{W}_{cf}$ from the centre-aligned strip $\mathcal{W}_c$ repeating the same procedure of centre alignment. We compute the normalised shifts (the green lines in Fig. 4) of the upper-ends $\mathbf{B}_s = (1 - \sigma_a/\mu(\sigma_a))/2$ where σ_a is a vector of the column-wise standard derivations of $\mathcal{W}_{cf}$. The shifts at the bottom-ends have the same magnitude but in the opposite direction. Although our previous gradient-assisted sampling already offers a clean strip, minor alignment errors can still noticeable from the intensity distance images . These extra alignment steps alleviate these (e.g. Fig. 2(f) and Fig. 2(h)).

3.3 Estimation of shadow scale and relighting

Using the penumbra strip, we can now derive a single unified intensity sample of each valid sub-group for estimation of penumbra illumination as opposed to performing computational-costly pixel-wise optimisation, e.g.[7, 9, 11]. Each sub-group's unified sample of intensity change G can be computed as the mean of its columns in the aligned penumbra strip such that white noise is cancelled. Unlike previous work, e.g. [7, 8, 9], assuming a constrained model of illumination change, our illumination model is adaptive to G. We fit G using a piecewise cubic Hermite polynomial [6] which preserves the shape of the data and its monotonicity. Our illumination model is variable for different penumbra profiles as we cluster and re-scale the intensity samples in the previous steps. For each RGB channel, the normalised estimated scales are computed by dividing each estimated curve by its maximum. As the sampling sites have previously been amended during alignment of the strip, we update the two ends of each sampling line as follows:

$$\begin{bmatrix}\mathbf{p}_s^n(i)\\ \mathbf{p}_e^n(i)\end{bmatrix} = \begin{bmatrix}\mathbf{p}_s^o(i)\\ \mathbf{p}_e^o(i)\end{bmatrix} + (\mathbf{C}(i) + \begin{bmatrix}-1\\ 1\end{bmatrix}\mathbf{B}_s(i))\mathbf{V}_r(i) \tag{4}$$

where i is the sampling line index, $\mathbf{p}_s^o(i)$ and $\mathbf{p}_e^o(i)$ are the previous start and end points of a sampling line, $\mathbf{p}_s^n(i)$ and $\mathbf{p}_e^n(i)$ are the updated points, $\mathbf{V}_r(i)$ is the vector of boundary normal, $\mathbf{C}(i)$ and $\mathbf{B}_s(i)$ are the normalised centre and column width shifts. According to the updated sampling lines, we distribute the unified scales back to the their original positions in the image (e.g. Fig. 2(j)). To obtain a dense scale field (e.g. Fig. 2(k)), we interpolate the sparse scales in the penumbra region by smoothly interpolating and extrapolating the scales in other regions using spring-metaphor based in-painting [2]. The shadow-free image can be obtained by inverse scaling according to Eq. 1.

4 Evaluation

In this section, we evaluate our algorithm versus other state of the art shadow removal methods. We also evaluate the quality of our new ground truth versus existing ground truth. For fair comparison in the community: our P-code, user input data, output data and scripts for ground truth evaluation and quantitative evaluations of shadow removal are also made public along with the release of our data set.

Evaluation of ground truth quality Ideal pairs of ground truth images should have a minimum intensity difference in the common lit area – which will also indicate whether registration is poor (due to camera shake or scene movement – which we wish to reject). We utilise this to assess the quality of ground truth candidates. We first compute the error image $\Delta I = I_s - I_g$ and the ratio image $I_r = \Phi(I_s) \oslash \Phi(I_g)$, where I_s and I_g are shadow and shadow-free images, $\oslash$ is element-wise division and Φ is a function that converts RGB image to grey-scale image. The set of pixels P_r of I_r that satisfies $I_r(P_r) > 1$ are regarded as lit pixels. We compute the ground truth error $Q_d = \mu(|\Delta I(P_r)|) + \sigma(\Delta I(P_r))$ and remove ground truth pairs when $Q_d > 0.05$. Using this measure, our initial data capture of 195 test cases results in 186 test cases with stable illumination changes between the shadow and ground truth images. Comparing to the quality of a previous ground truth data set [8] results in a mean error of 0.18 (leaving 28 out of 79 test cases) while our error is 0.02.

Quantitative evaluation of shadow removal In previous work [8, 10], the quality of shadow removal is measured by directly using the per-pixel error between the shadow removal result and shadow-free ground truth. However, in our work we also consider the size

of the shadow, and the fact that some shadows may be darker than others. We therefore compute the error ratio $\mathbf{E}_r = \mathbf{E}_n/\mathbf{E}_o$ as our quality measurement where $\mathbf{E}_n$ is the error between the ground truth (no shadow) and shadow removal result, and $\mathbf{E}_o$ is the error between the ground truth (no shadow) and the original shadow image. This normalised measure better reflects removal improvements towards the ground truth independent of original shadow intensity and size. We assess $\mathbf{E}_n$ and $\mathbf{E}_o$ using Root-Mean-Square-Error (RMSE) of RGB intensity. To test robustness, we also compute the standard derivation for each measurement.

		Yang *et al.*[13]		Guo *et al.*[8]		Gong *et al.*[7]		Ours	
		E_r	E_r*	E_r	E_r*	E_r	E_r*	E_r	E_r*
Tex.	1	1.85 (1.48)	1.81 (2.65)	0.53 (0.50)	0.42 (0.57)	0.32 (0.19)	0.16 (0.18)	**0.26 (0.15)**	**0.10 (0.10)**
	2	1.23 (0.74)	0.82 (0.89)	0.59 (1.09)	0.47 (1.15)	0.38 (0.33)	0.27 (0.35)	**0.26 (0.11)**	**0.13 (0.10)**
	3	1.75 (0.89)	1.37 (1.64)	0.71 (0.60)	0.64 (1.03)	0.70 (0.42)	0.65 (0.55)	**0.50 (0.38)**	**0.39 (0.45)**
	M	1.61 (1.04)	1.49 (1.36)	0.61 (0.73)	0.51 (0.92)	0.47 (0.31)	0.36 (0.36)	**0.21 (0.22)**	**0.34 (0.21)**
Sof.	1	1.27 (0.75)	0.80 (0.87)	0.52 (1.08)	0.39 (1.13)	0.33 (0.31)	0.21 (0.32)	**0.23 (0.09)**	**0.10 (0.09)**
	2	1.89 (1.60)	2.09 (2.85)	0.70 (0.36)	0.64 (0.43)	0.44 (0.21)	0.29 (0.25)	**0.35 (0.15)**	**0.16 (0.11)**
	3	1.86 (1.12)	1.86 (1.27)	1.09 (0.75)	1.01 (0.97)	0.76 (0.35)	0.71 (0.53)	**0.63 (0.25)**	**0.44 (0.25)**
	M	1.67 (1.16)	1.58 (1.66)	0.77 (0.73)	0.68 (0.84)	0.51 (0.29)	0.40 (0.37)	**0.40 (0.17)**	**0.23 (0.15)**
Bro.	1	1.47 (1.13)	1.19 (1.83)	0.59 (0.98)	0.48 (1.04)	0.36 (0.29)	0.23 (0.31)	**0.26 (0.13)**	**0.12 (0.10)**
	2	1.07 (0.17)	0.66 (0.19)	0.42 (0.29)	0.27 (0.35)	0.44 (0.25)	0.29 (0.28)	**0.28 (0.11)**	**0.11 (0.08)**
	3	2.32 (0.96)	2.56 (1.48)	1.42 (1.06)	1.55 (1.84)	0.98 (0.31)	1.05 (0.50)	**0.72 (0.29)**	**0.53 (0.28)**
	M	1.62 (0.75)	1.47 (1.17)	0.81 (0.78)	0.76 (1.08)	0.59 (0.29)	0.52 (0.36)	**0.42 (0.17)**	**0.25 (0.15)**
Col.	1	1.20 (0.48)	0.73 (0.45)	0.48 (0.64)	0.36 (0.78)	0.32 (0.18)	0.19 (0.20)	**0.25 (0.10)**	**0.11 (0.08)**
	2	4.19 (2.02)	6.07 (3.28)	1.67 (2.29)	1.56 (2.07)	0.83 (0.67)	0.67 (0.73)	**0.48 (0.16)**	**0.27 (0.17)**
	3	2.95 (1.95)	3.53 (3.59)	1.20 (0.99)	1.34 (2.33)	1.10 (0.68)	1.20 (1.18)	**0.59 (0.27)**	**0.49 (0.39)**
	M	2.78 (1.48)	3.44 (2.44)	1.12 (1.31)	1.09 (1.73)	0.75 (0.51)	0.69 (0.70)	**0.44 (0.18)**	**0.29 (0.21)**
Other		1.72 (1.26)	1.56 (2.05)	0.72 (0.89)	0.65 (1.26)	0.57 (0.46)	0.48 (0.64)	**0.40 (0.29)**	**0.26 (0.32)**

		Yang *et al.*[13]		Guo *et al.*[8]		Gong *et al.*[7]		Ours	
		E_r	E_r*	E_r	E_r*	E_r	E_r*	E_r	E_r*
Tex.	1	1.26 (0.43)	0.70 (0.33)	0.32 (0.25)	0.18 (0.29)	0.26 (0.10)	0.11 (0.08)	**0.23 (0.11)**	**0.08 (0.05)**
	2	1.16 (0.46)	0.73 (0.49)	0.43 (0.71)	0.31 (0.88)	0.35 (0.27)	0.23 (0.31)	**0.25 (0.09)**	**0.12 (0.08)**
	3	1.71 (0.84)	1.43 (1.18)	0.64 (0.50)	0.48 (0.56)	0.68 (0.42)	0.65 (0.57)	**0.50 (0.40)**	**0.40 (0.47)**
	M	1.38 (0.58)	0.95 (0.67)	0.46 (0.49)	0.32 (0.58)	0.43 (0.26)	0.33 (0.32)	**0.33 (0.20)**	**0.20 (0.20)**
Sof.	1	1.19 (0.46)	0.70 (0.45)	0.37 (0.65)	0.24 (0.81)	0.31 (0.24)	0.19 (0.28)	**0.23 (0.08)**	**0.10 (0.07)**
	2	1.20 (0.42)	0.83 (0.41)	0.51 (0.35)	0.42 (0.42)	0.39 (0.17)	0.23 (0.19)	**0.30 (0.13)**	**0.13 (0.07)**
	3	1.65 (0.83)	1.60 (1.02)	0.96 (0.49)	0.80 (0.65)	0.79 (0.40)	0.76 (0.61)	**0.66 (0.30)**	**0.47 (0.29)**
	M	1.35 (0.57)	1.04 (0.63)	0.62 (0.50)	0.49 (0.62)	0.49 (0.27)	0.39 (0.36)	**0.40 (0.17)**)	**0.23 (0.15)**
Bro.	1	1.21 (0.47)	0.73 (0.46)	0.41 (0.63)	0.29 (0.79)	0.32 (0.24)	0.19 (0.27)	**0.24 (0.10)**	**0.10 (0.08)**
	2	1.03 (0.19)	0.64 (0.20)	0.29 (0.11)	0.11 (0.06)	0.34 (0.17)	0.22 (0.19)	**0.28 (0.10)**	**0.10 (0.07)**
	3	2.07 (0.67)	2.18 (0.92)	1.10 (0.53)	0.93 (0.72)	0.98 (0.35)	1.06 (0.55)	**0.73 (0.32)**	**0.53 (0.31)**
	M	1.43 (0.44)	1.18 (0.53)	0.60 (0.43)	0.44 (0.52)	0.55 (0.25)	0.49 (0.34)	**0.41 (0.17)**	**0.25 (0.15)**
Col.	1	1.19 (0.45)	0.71 (0.43)	0.40 (0.62)	0.27 (0.76)	0.30 (0.14)	0.17 (0.15)	**0.24 (0.09)**	**0.10 (0.07)**
	2	1.47 (0.32)	1.33 (0.64)	0.45 (0.00)	0.34 (0.11)	1.32 (0.82)	1.37 (0.95)	**0.38 (0.10)**	**0.17 (0.05)**
	3	3.03 (1.95)	3.63 (3.64)	1.21 (1.01)	1.35 (2.38)	1.12 (0.69)	1.22 (1.20)	**0.59 (0.27)**	**0.49 (0.40)**
	M	1.90 (0.91)	1.89 (1.57)	0.69 (0.54)	0.66 (1.09)	0.91 (0.55)	0.92 (0.77)	**0.40 (0.16)**	**0.26 (0.17)**
Other		1.58 (1.07)	1.31 (1.74)	0.59 (0.68)	0.48 (1.10)	0.54 (0.47)	0.47 (0.67)	**0.39 (0.30)**	**0.26 (0.34)**

Table 1: Shadow removal errors according to 4 attributes (top: results for all (214) test cases; bottom: results for 177 test cases without detection failures). The non-stared and stared columns indicate the error score where all pixels in the image are used, and just shadow area pixels respectively. For each score of each attribute, the images with other predominant attributes (degree = 3) are not used. Hence, test cases have a strong single bias towards one of the attributes. "Other" refers to a set of shadow cases showing no markedly predominant attributes (degree = 1). "M" refers to the average score for each category. Standard derivations are shown in brackets. Method [8] is trained using a large shadow detection data set from [14]. The best scores are made bold and the second best scores are underlined. As E_o for the whole image is lower than E_o for the shadow area only – and E_n for both cases are similar after shadow removal – E_r for the shadow area only is thus generally lower.

Improving on previous work [8, 10], our removal test is based on our data set of 186 cases, which contains challenging categories for soft, broken and colour shadows and shadows cast

on strong textured surfaces as well as simpler shadows, plus 28 examples from [8] – resulting in 214 test cases in total. Each case is rated according to 4 attributes, which are *texture*, *brokenness*, *colourfulness* and *softness*, in 3 perceptual degrees from weak to strong which were aggregated by five users. In Table 1, we show combined shadow removal error results from both automatic and semi-automatic shadow removal algorithms (all 214 cases). However, as automatic algorithms can often fail in detection – leading to artefacts unfairly biasing the removal error score – we also show results where we have removed detection failures from other algorithms – leaving 177 cases in total. This second test therefore concentrates solely on the quality of removal given accurately detected shadow inputs. In both experiments, our method shows leading performance across all comparisons.

Tab. 2 shows some typical visual results of shadow removal on various scenarios from our data set. Our supplementary material shows a wide range of other removal results with higher resolution images. We therefore encourage readers to examine these figures as compelling evidence of the strength of our approach. As is the case with all current shadow removal methods, our method has most difficulty in extreme cases, e.g. Tab. 3, where shadows are highly broken, colourful, or soft.

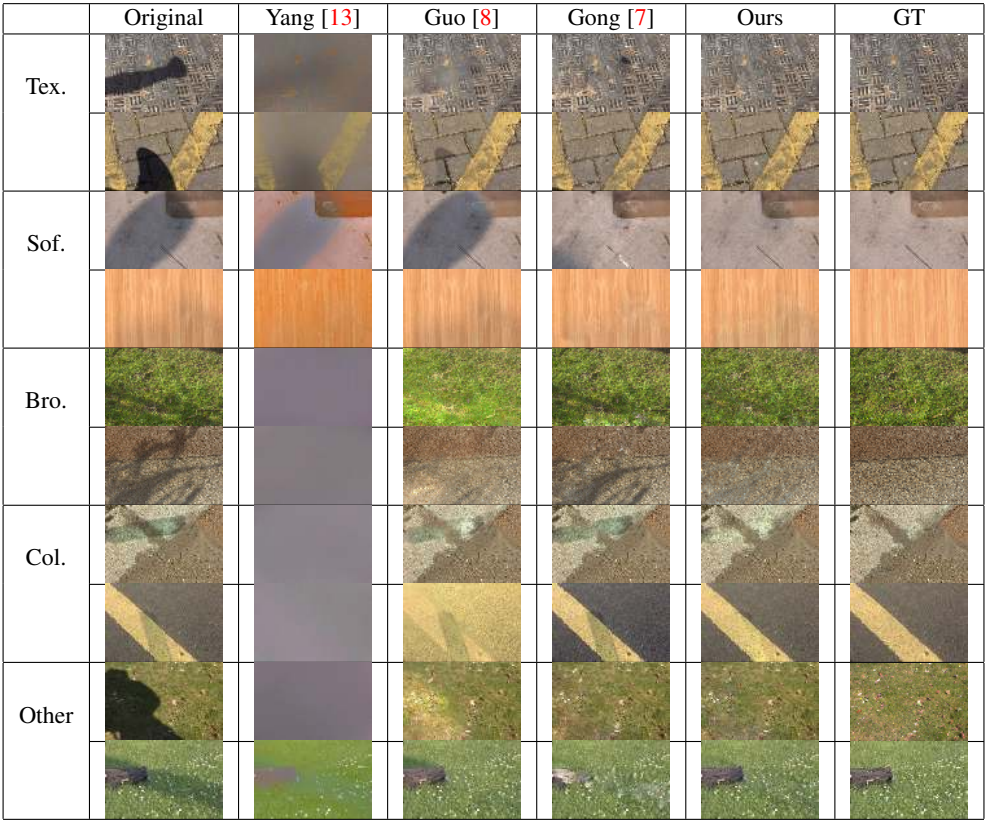


Table 2: Comparisons using images from our data set. The table shows our results given test cases with strong degrees of the corresponding attribute except for "Other", which refers to cases where there is no predominantly strong attribute.

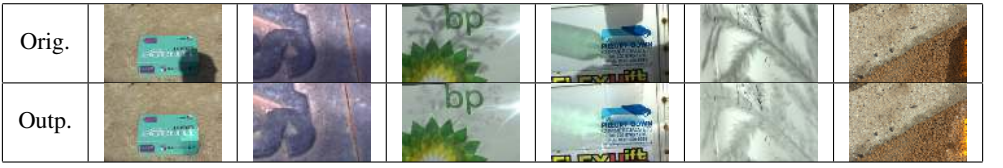


Table 3: Our failure cases – where all the leading methods we tested create highly visible image artefacts and fail to remove the shadow.

5 Conclusions

We have presented an interactive method for fast shadow removal together with a state of the art ground truth. Our method balances the complexity of user input with robust shadow removal performance. Our quantitatively-verified ground truth data set overcomes potential issues relating to mismatched illumination and registration. We have evaluated our method against several state of the art methods using a thorough quantitative test and shown leading state of the art performance. Our P-code, evaluation data and scripts will be made public to the community to encourage future comparison of techniques.

Acknowledgements

This work was funded by the China Scholarship Council and the University of Bath. We also thank the authors of previous work in the area for sharing their implementations for our comparisons.

References

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

[2] Marcelo Bertalmio, Guillermo Sapiro, Vincent Caselles, and Coloma Ballester. Image inpainting. In *SIGGRAPH*, pages 417–424, 2000.

[3] Dorin Comaniciu and Peter Meer. Mean shift: A robust approach toward feature space analysis. *PAMI*, 24(5):603–619, 2002.

[4] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD*, volume 96, pages 226–231, 1996.

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

[6] Frederick N Fritsch and Ralph E Carlson. Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, 17(2):238–246, 1980.

[7] H. Gong, D. Cosker, C. Li, and M. Brown. User-aided single image shadow removal. In *ICME*, pages 1–6, 2013.

[8] R. Guo, Q. Dai, and D. Hoiem. Paired regions for shadow detection and removal. *PAMI*, PP(99):1–1, 2012.

[9] F. Liu and M. Gleicher. Texture-consistent shadow removal. In *ECCV*, pages 437–450, 2008.

[10] Y. Shor and D. Lischinski. The shadow meets the mask: Pyramid-based shadow removal. *CGF*, 27(2):577–586, 2008.

[11] Y.-F. Su and H. H. Chen. A three-stage approach to shadow field estimation from partial boundary information. *IEEE Trans. on Image Proc.*, 19(10):2749–2760, 2010.

[12] Tai-Pang Wu, Chi-Keung Tang, Michael S Brown, and Heung-Yeung Shum. Natural shadow matting. *ACM TOG*, 26(2):8, 2007.

[13] Q. Yang, K.-H. Tan, and N. Ahuja. Shadow removal using bilateral filtering. *IEEE Trans. on Image Proc.*, 21(10):4361–4368, 2012.

[14] J. Zhu, K. G. G. Samuel, S. Masood, and M. F. Tappen. Learning to recognize shadows in monochromatic natural images. In *CVPR*, 2010.