

Single-Image Shadow Detection and Removal using Paired Regions

Ruiqi Guo Qieyun Dai Derek Hoiem
University of Illinois at Urbana Champaign
`{guo29,dai9,dhoiem}@illinois.edu`

Abstract

In this paper, we address the problem of shadow detection and removal from single images of natural scenes. Different from traditional methods that explore pixel or edge information, we employ a region based approach. In addition to considering individual regions separately, we predict relative illumination conditions between segmented regions from their appearances and perform pairwise classification based on such information. Classification results are used to build a graph of segments, and graph-cut is used to solve the labeling of shadow and non-shadow regions. Detection results are later refined by image matting, and the shadow free image is recovered by relighting each pixel based on our lighting model. We evaluate our method on the shadow detection dataset in [19]. In addition, we created a new dataset with shadow-free ground truth images, which provides a quantitative basis for evaluating shadow removal.

1. Introduction

Shadows, created wherever an object obscures the light source, are an ever-present aspect of our visual experience. Shadows can either aid or confound scene interpretation, depending on whether we model the shadows or ignore them. If we can detect shadows, we can better localize objects, infer object shape, and determine where objects contact the ground. Detected shadows also provide cues for lighting direction [10] and scene geometry. On the other hand, if we ignore shadows, spurious edges on the boundaries of shadows and confusion between albedo and shading can lead to mistakes in visual processing. For these reasons, shadow detection has long been considered a crucial component of scene interpretation (e.g., [17, 2]). But despite its importance and long tradition, shadow detection remains an extremely challenging problem, particularly from a single image.

The main difficulty is due to the complex interactions of geometry, albedo, and illumination. Locally, we cannot tell if a surface is dark due to shading or albedo, as illustrated in Figure 1. To determine if a region is in shadow, we must compare the region to others that have the same material and orientation. For this reason, most research focuses on mod-

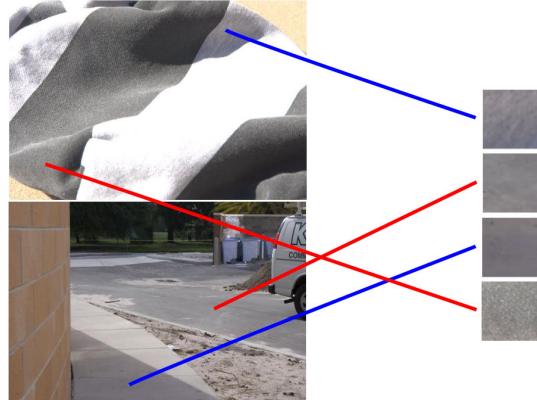


Figure 1. What is in shadow? Local region appearance can be ambiguous, to find shadows, we must compare surfaces of the same material.

eling the differences in color, intensity, and texture of neighboring pixels or regions. Many approaches are motivated by physical models of illumination and color [12, 15, 16, 7, 5]. For example, Finlayson et al. [7] compare edges in the original RGB image to edges found in an illuminant-invariant image. This method can work quite well with high-quality images and calibrated sensors, but often performs poorly for typical web-quality consumer photographs [11]. To improve robustness, others have recently taken a more empirical, data-driven approach, learning to detect shadows based on training images. In monochromatic images, Zhu et al. [19] classify regions based on statistics of intensity, gradient, and texture, computed over local neighborhoods, and refine shadow labels using a conditional random field (CRF). Lalonde et al. [11] find shadow boundaries by comparing the color and texture of neighboring regions and employing a CRF to encourage boundary continuity.

Our goal is to detect shadows and remove them from the image. To determine whether a particular region is shadowed, we compare it to other regions in the image that are likely to be of the same material. To start, we find pairs of regions that are likely to correspond to the same material and determine whether they have the same illumination conditions. We incorporate these pairwise relationships, together with region-based appearance features,

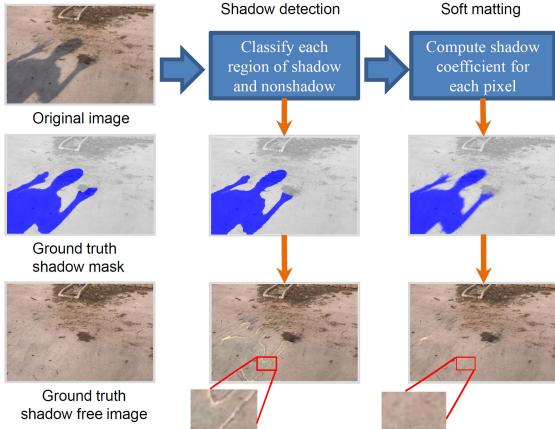


Figure 2. Illustration of our framework. First column: the original image with shadow, ground truth shadow mask, ground truth image; Second column, hard shadow map generated by our detection method and recovered image using this map alone. Note that there are strong boundary effects in the recovered image. Third column, soft shadow map computed using soft matting and recovery result using this map.

in a shadow/non-shadow graph. The node potentials in our graph encode region appearance; a sparse set of edge potentials indicate whether two regions from the same surface are likely to be of the same or different illumination. Finally, the regions are jointly classified as shadow/non-shadow using graph-cut inference. Like Zhu et al. [19] and Lalonde et al. [11], we take a data-driven approach, learning our classifiers from training data, which leads to good performance on consumer-quality photographs. Unlike others, we explicitly model the material and illumination relationships of pairs of regions, including non-adjacent pairs. By modeling long-range interactions, we hope to better detect soft shadows, which can be difficult to detect locally. By restricting comparisons to regions with the same material, we aim to improve robustness in complex scenes, where material and shadow boundaries may coincide.

Our shadow detection provides binary pixel labels, but shadows are not truly binary. Illumination often changes gradually across shadow boundaries. We also want to estimate a soft mask of shadow coefficients, which indicate the darkness of the shadow, and to recover a shadow-free image that depicts the scene under uniform illumination. The most popular approach in shadow removal is proposed in a series of papers by Finlayson and colleagues, where they treat shadow removal as an reintegration problem based on detected shadow edges [6, 9, 8]. Our region-based shadow detection enables us to pose shadow removal as a matting problem, similarly to Wu et al. [18]. However, the method of Wu et al. [18] depends on user input of shadow and non-shadow regions, while we automatically detect and remove shadows in a unified framework (Figure 2). Specifically, after detecting shadows, we apply matting technique of Levin

et al. [13], treating shadow pixels as foreground and non-shadow pixels as background. Using the recovered shadow coefficients, we calculate the ratio between direct light and environment light and generate the recovered image by re-lighting each pixel with both direct light and environment light.

To evaluate our shadow detection and removal, we propose a new dataset with 108 natural scenes, in which ground truth is determined by taking two photographs of a scene after manipulating the shadows (either by blocking the direct light source or by casting a shadow into the image). To the best of our knowledge, our dataset is the first to enable quantitative evaluation of shadow removal on dozens of images. We also evaluate our shadow detection on Zhu et al.’s dataset of manually ground-truthed outdoor scenes, comparing favorably to Zhu et al. [19].

The main contributions of this paper are (1) a new method for detecting shadows using a relational graph of paired regions; (2) an automatic shadow removal procedure derived from lighting models making use of shadow matting to generate soft boundaries between shadow and non-shadow areas; (3) quantitative evaluation of shadow detection and removal, with comparison to existing work; (4) a shadow removal dataset with shadow free ground truth images. We believe that more robust algorithms for detecting and removing shadows will lead to better recognition and estimates of scene geometry.

2. Shadow Detection

To detect shadows, we must consider the appearance of the local and surrounding regions. Shadowed regions tend to be dark, with little texture, but some non-shadowed regions may have similar characteristics. Surrounding regions that correspond to the same material can provide much stronger evidence. For example, suppose region s_i is similar to s_j in texture and chromaticity. If s_i has similar intensity to s_j , then they are probably under the same illumination and should receive the same shadow label (either shadow or non-shadow). However, if s_i is much darker than s_j , then s_i probably is in shadow, and s_j probably is not.

We first segment the image using the mean shift algorithm [4]. Then, using a trained classifier, we estimate the confidence that each region is in shadow. We also find *same illumination pairs* and *different illumination pairs* of regions, which are confidently predicted to correspond to the same material and have either similar or different illumination, respectively. We construct a relational graph using a sparse set of confident illumination pairs. Finally, we solve for the shadow labels $\mathbf{y} = \{-1, 1\}^n$ (1 for shadow) that maximize the following objective:

$$\hat{\mathbf{y}} = \arg \max_{\mathbf{y}} \sum_{i=1} c_i^{\text{shadow}} y_i + \alpha_1 \sum_{\{i,j\} \in E_{\text{diff}}} c_{ij}^{\text{diff}} (y_i - y_j) \quad (1)$$

$$- \alpha_2 \sum_{\{i,j\} \in E_{\text{same}}} c_{ij}^{\text{same}} \mathbf{1}(y_i \neq y_j)$$

where c_i^{shadow} is the single-region classifier confidence weighted by region area; $\{i, j\} \in E_{\text{diff}}$ are different illumination pairs; $\{i, j\} \in E_{\text{same}}$ are same illumination pairs; c_{ij}^{same} and c_{ij}^{diff} are the area-weighted confidences of the pairwise classifiers; α_1 and α_2 are parameters; and $\mathbf{1}(.)$ is an indicator function.

In the following subsections, we describe the classifiers for single regions (Section 2.1) and pairs of regions (Section 2.2) and how we can reformulate our objective function to solve it efficiently with the graph-cut algorithm (Section 2.3).

2.1. Single Region Classification

When a region becomes shadowed, it becomes darker and less textured (see [19] for empirical analysis). Thus, the color and texture of a region can help predict whether it is in shadow. We represent color with a histogram in L*a*b space, with 21 bins per channel. We represent texture with the texton histogram provided by Martin et al. [14]. We train our classifier from manually labeled regions using an SVM with a χ^2 kernel (slack parameter $C = 1$). We define c_i^{shadow} as the output of this classifier times a_i , the pixel area of region i .

2.2. Pair-wise Region Relationship Classification

We cannot determine whether a region is in shadow by considering only its internal appearance; we must compare the region to others with the same material. In particular, we want to find *same illumination pairs*, regions that are of the same material and illumination, and *different illumination pairs*, regions that are of the same material but different illumination. Differences in illumination can be caused by direct light blocked by other objects or by a difference in surface orientation. In this way, we can account for both shadows and shading. Comparison between regions with different materials is uninformative because they have different reflectance.

We detect shadows using a relational graph, with an edge connecting each illumination pair. To better handle occlusion and to link similarly lit regions that are divided by shadows, we enable edges between regions that are not adjacent in the image. Because most pairs of regions are not of the same material, our graph is still very sparse. Examples of such relational graphs are shown in Figure 3.

We train classifiers (SVM with RBF kernel; $C = 1, \sigma = 1$) to detect illumination pairs based on comparisons of their color and texture histograms, the ratio of their intensities, their chromatic alignment, and their distance in the image. These features encode the intuition that regions of the same reflectance share similar texture and color distribution when viewed under the same illumination; when viewed under different illuminations, they tend to have similar texture but differ in color and intensity. We also take into account the distance between two regions, which greatly reduces false comparisons while enabling more flexibility than considering only adjacent pairs.

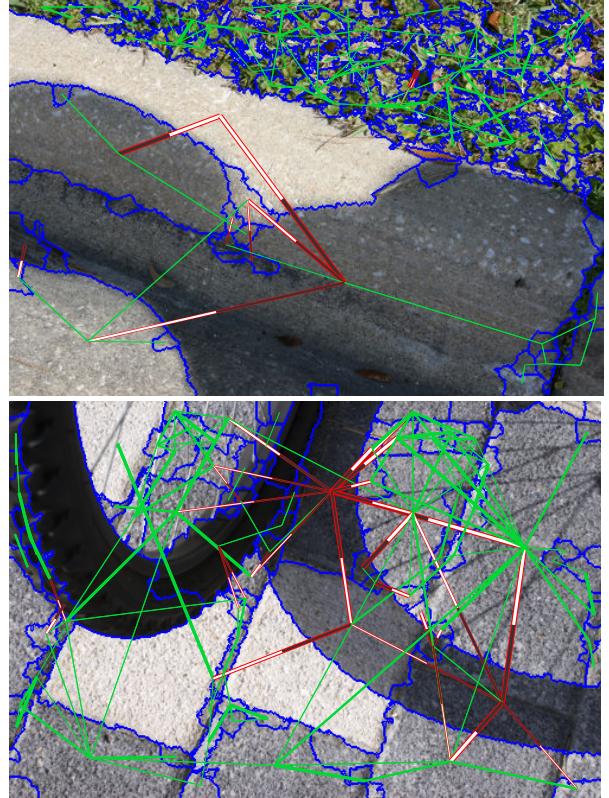


Figure 3. Illumination relation graph of two example images. Green lines indicate *same illumination pairs*, and red/white lines mean *different illumination pairs*, where white ends are the non-shadow regions and gray ends are shadows. The width shows the confidence of the pair.

χ^2 distances between color and texture histograms are computed as in Section 2.1. Regions of the same material will often have similar texture histograms, regardless of differences in shading. When regions have both similar color and texture, they are likely to be same illumination pairs.

Ratios of RGB average intensity are calculated as ($\rho_R = \frac{R_{avg1}}{R_{avg2}}, \rho_G = \frac{G_{avg1}}{G_{avg2}}, \rho_B = \frac{B_{avg1}}{B_{avg2}}$), where R_{avg1} , for example, is the average value of the red channel for the first region. For a shadow/non-shadow pair of the same material, the non-shadow region has a higher value in all three channels.

Chromatic alignment: Studies have shown that color of shadow/non-shadow pairs tend to align in RGB color space [1]. Simply put, the shadow region should not look more red or yellow than the non-shadow region. This ratio is computed as ρ_R/ρ_G and ρ_G/ρ_B .

Normalized distance in position: Because distant image regions are less likely to correspond to the same material, we also add the normalized distance as a feature, computing it as the Euclidean distance of the region centers divided by the square root of the geometric mean of the region

areas.

We define c_{ij}^{same} as the output of the classifier for same-illumination pairs times $\sqrt{a_i a_j}$, the geometric mean of the region areas. Similarly, c_{ij}^{diff} is the output of the classifier for different-illumination pairs times $\sqrt{a_i a_j}$. Edges are weighted by region area and classifier score so that larger regions and those with more confidently predicted relations have more weight. Note that the edges in E_{diff} are directional: they encourage y_i to be shadow and y_j to be non-shadow. In both cases, the 100 most confident edges are included if their classifier scores are greater than 0.6 (subsequent experiments indicate that including all edges with positive scores yields similar performance).

2.3. Graph-cut Inference

We can apply efficient and optimal graph-cut inference by reformulating our objective function (Eq. 1) as the following energy minimization:

$$\hat{\mathbf{y}} = \arg \min_{\mathbf{y}} \sum_k \text{cost}_k^{\text{unary}}(y_k) + \alpha_2 \sum_{\{i,j\} \in E_{\text{same}}} c_{ij}^{\text{same}} \mathbf{1}(y_i \neq y_j) \quad (2)$$

with

$$\begin{aligned} \text{cost}_k^{\text{unary}}(y_k) = & -c_k^{\text{shadow}} y_k - \alpha_1 \sum_{\{i=k,j\} \in E_{\text{diff}}} c_{ij}^{\text{diff}} y_k \quad (3) \\ & + \alpha_1 \sum_{\{i,j=k\} \in E_{\text{diff}}} c_{ij}^{\text{diff}} y_k. \end{aligned}$$

Because this is submodular (binary, with pairwise term encouraging affinity), we can solve for $\hat{\mathbf{y}}$ using graph cuts [3]. In our experiments, $\alpha_1 = \alpha_2 = 1$.

3. Shadow removal

Our shadow removal approach is based on a simple shadow model where lighting consists of directed light and environment light. We try to identify how much direct light is occluded for each pixel in the image and relights the whole image using that information. First, we use a matting technique to estimate a fractional shadow coefficient value. Then, we estimate the ratio of direct to environmental light, which together with the shadow coefficient, enables a shadow-free image to be recovered.

3.1. Shadow model

In our illumination model, there are two types of light sources: direct light and environment light. Direct light comes directly from the source (e.g., the sun), while environment light is from reflections of surrounding surfaces. Non-shadow areas are lit by both direct light and environment light, while for shadow areas, part or all of the direct light is occluded. The shadow model can be represented by the formula below.

$$\mathbf{I}_i = (t_i \cos \theta_i \mathbf{L}_d + \mathbf{L}_e) \mathbf{R}_i \quad (4)$$

where \mathbf{I}_i is a vector representing the value for the i -th pixel in RGB space. Similarly, both \mathbf{L}_d and \mathbf{L}_e are vectors of size 3, each representing the intensity of the direct light and environment light, also measured in RGB space. \mathbf{R}_i is the surface reflectance of that pixel, also a vector of three dimensions, each corresponding to one channel. θ_i is the angle between the direct lighting direction and the surface norm, and t_i is a value between $[0, 1]$ indicating how much direct light gets to the surface. The equations in Section 3 for matrix computation refers to a pointwise computation, except for the spectral matting energy function and its solution. When $t_i = 1$, the pixel is in a non-shadow area, and when $t_i = 0$, the pixel is in an umbra, otherwise, the area is in a penumbra ($0 < t_i < 1$). For a shadow-free image, every pixel is lit by both direct light and environment light and can be expressed as:

$$\mathbf{I}_i^{\text{shadow-free}} = (\mathbf{L}_d \cos \theta_i + \mathbf{L}_e) \mathbf{R}_i$$

We define $k_i = t_i \cos \theta_i$, which we will refer to k_i as the shadow coefficient for the i -th pixel in the rest of the paper; $k_i = 1$ for pixels in non-shadow regions.

3.2. Shadow Matting

The shadow detection procedure provides us with a binary shadow mask where each pixel i is assigned a \hat{k}_i value of either 1 or 0. However, illumination often changes gradually along shadow, and segmentation results are often inaccurate near the boundaries of regions. Using detection results as shadow coefficient values can create strong boundary effects. To get more accurate k_i values and get smooth changes between non-shadow regions and recovered shadow regions, we apply soft matting technique.

Given an image \mathcal{I} , matting tries to separate the foreground image \mathcal{F} and background image \mathcal{B} based on the following formulation.

$$\mathbf{I}_i = \gamma_i \mathbf{F}_i + (1 - \gamma_i) \mathbf{B}_i$$

\mathbf{I}_i is the RGB value of the i -th pixel of the original image \mathcal{I} , and \mathbf{F}_i and \mathbf{B}_i are respectively the RGB value of the i -th pixel of the foreground \mathcal{F} and background image \mathcal{B} . By rewriting the shadow formulation given in (4) as:

$$\mathbf{I}_i = k_i (\mathbf{L}_d \mathbf{R}_i + \mathbf{L}_e \mathbf{R}_i) + (1 - k_i) \mathbf{L}_e \mathbf{R}_i$$

an image with shadow can be seen as the linear combination of a shadow-free image $\mathbf{L}_d \mathcal{R} + \mathbf{L}_e \mathcal{R}$ and a shadow image $\mathbf{L}_e \mathcal{R}$ (\mathcal{R} is a three dimensional matrix whose i -th entry equals \mathbf{R}_i), a formulation identical to that of image matting.

To solve the matting problem, we employ the spectral matting algorithm from [13], minimizing the following energy function:

$$E(\mathbf{k}) = \mathbf{k}^T \mathcal{L} \mathbf{k} + \lambda (\mathbf{k} - \hat{\mathbf{k}})^T D (\mathbf{k} - \hat{\mathbf{k}})$$

where $\hat{\mathbf{k}}$ indicates the estimated shadow label (Section 2), with $\hat{k}_i = 0$ being shadow areas and $\hat{k}_i = 1$ being non-shadow. D is a diagonal matrix where $D(i, i) = 1$ when

the k_i for the i -th pixel should agree with \hat{k}_i and 0 when the k_i value is to be predicted by the matting algorithm. In our experiments, we set $D(i, i) = 0$ for pixels within a 5-pixel distance of the detected label boundary, and $D(i, i) = 1$ for all other pixels. \mathcal{L} is the matting Laplacian matrix proposed in [13], aiming to enforce smoothness over local patches. In our experiments, a patch size of 3×3 is used.

The optimal k value is the solution to the following sparse linear system:

$$(\mathcal{L} + \lambda D)\mathbf{k} = \lambda \mathbf{d}\hat{\mathbf{k}}$$

where \mathbf{d} is the vector comprising of elements on the diagonal of the matrix D . In our experiments, we empirically set λ to 0.01.

3.3. Ratio Calculation and Pixel Relighting

Based on our shadow model, we can relight each pixel using the calculated ratio and k value. The new pixel value is given by:

$$\mathbf{I}_i^{\text{shadow-free}} = (\mathbf{L}_d + \mathbf{L}_e)\mathbf{R}_i \quad (5)$$

$$= (k_i \mathbf{L}_d + \mathbf{L}_e)\mathbf{R}_i \frac{\mathbf{L}_d + \mathbf{L}_e}{k_i \mathbf{L}_d + \mathbf{L}_e} \quad (6)$$

$$= \frac{\mathbf{r} + 1}{k_i \mathbf{r} + 1} \mathbf{I}_i \quad (7)$$

where $\mathbf{r} = \frac{\mathbf{L}_d}{\mathbf{L}_e}$ is the ratio between direct light and environment light and \mathbf{I}_i is the intensity of the i -th pixel in the original image. For each channel, we recover the pixel value separately. We now show how to recover \mathbf{r} from detected shadows and matting results.

To calculate the ratio between direct light and environment light, our model checks for adjacent shadow/non-shadow pairs along the shadow boundary. We believe these patches are of the same material and reflectance. We also assume direct light and environment light is consistent throughout the image. Based on the lighting model, for two pixels with the same reflectance, we have:

$$\begin{aligned} \mathbf{I}_i &= (k_i \mathbf{L}_d + \mathbf{L}_e)\mathbf{R}_i \\ \mathbf{I}_j &= (k_j \mathbf{L}_d + \mathbf{L}_e)\mathbf{R}_j \end{aligned}$$

with $\mathbf{R}_i = \mathbf{R}_j$.

From the above equations, we can arrive at:

$$\mathbf{r} = \frac{\mathbf{L}_d}{\mathbf{L}_e} = \frac{\mathbf{I}_j - \mathbf{I}_i}{\mathbf{I}_i k_j - \mathbf{I}_j k_i}$$

To estimate \mathbf{r} , we sample patches from shadow/non-shadow pairs and vote for values of \mathbf{r} based on average RGB intensity and k values within the pairs of patches. Votes in the joint RGB ratio space are accumulated with a histogram, and the center value of the bin with the most votes is used for \mathbf{r} . The bin size is set to 0.1, and the patch size is 4×4 pixels.

4. Experiments and Results

In our experiments, we evaluate both shadow detection and shadow removal results. For shadow detection, we measure how explicitly modeling the pairwise region relationship affects detection results and how well our detector can generalize cross datasets. For shadow removal, we evaluate the results quantitatively on our dataset by comparing the recovered image with the shadow-free ground truth and show qualitative results on both our dataset and the UCF shadow dataset [19].

4.1. Dataset

Our shadow detection and removal methods are evaluated on the UCF shadow dataset [19] and our proposed new dataset. Zhu et al. made available a set of 245 images they collected themselves and from Internet, with manually labeled ground truth shadow masks.

Our training set consists of 32 images with manually annotation of shadow regions and pairs illumination conditions between regions. Our test set contains 76 image pairs, collected from common scenes or objects under a variety of illumination conditions, both indoor and outdoor. Each pair consists of a shadowed image (the input to the algorithm) and a ground truth image where every pixel in the image has the same illumination. For 46 image pairs, we taken one image with shadow and then another shadow-free ground truth image by removing the source of the shadow. The light sources for both images remain the same. One disadvantage of this approach is that it does not include self shadows of objects. To account for that, we collect another set of 30 images where shadows are caused by objects in the scene. To create an image pair for this set of images, we block the light source so that the whole scene is in shadow. We automatically generate the ground truth shadow mask by thresholding the ratio between the two images in a pair. This approach is more accurate and robust than manually annotating shadow regions.

4.2. Shadow Detection Evaluation

Two sets of experiments are carried out for shadow detection. First, we try to compare the performance when using only the unary classifier, only the pairwise classifier, and both combined. Second, we conduct cross dataset evaluation, training on one dataset and testing on the other. The per pixel accuracy on the testing set is reported in Figure 4(c) and the qualitative results are shown in Figure 5.

4.2.1 Comparison between unary and pairwise information

Using only unary information, the performance on the UCF dataset is 87.5%, versus a 83.4% achieved by classifying everything to non-shadow, and 88.7% reported in [19]. Different from our approach which makes use of color information, [19] conducts shadow detection on gray scale

images. By combining unary information with pairwise information, we achieved an accuracy of 90.0%. Note that we are using a simpler set of features and simpler learning method than [19].

The pairwise illumination relations are especially important on our dataset. Using them, the overall accuracy increases by more than 8%, and 50% more shadow areas are detected than with the single region classifier.

4.2.2 Cross dataset evaluation

The result in Figure 4(d) indicate that our proposed detector can generalize across datasets. This is especially notable since the two datasets are very different in nature, with [19] containing more large scale scenes and hard shadows. As shown in Figure 4(d), the unary and pairwise classifiers trained on [19] performs well on both datasets. This is understandable since their dataset is more diverse and contains more training images.

4.3. Shadow Removal Evaluation

To evaluate shadow free image recovery, we used as measurement the root mean square error (RMSE) in L*a*b color space between the ground truth shadow free image and the recovered image, which is designed to be locally perceptually uniform. We evaluate our results on the whole image as well as shadow and non-shadow regions separately.

The quantitative evaluation is performed on the subset of images with ground truth shadow free image (a total of 46 images). Shadow/non-shadow regions are given by the ground truth shadow mask introduced in the previous section. As shown in Table 4(e), our shadow removal procedure based on image matting yields results that are perceptually close to ground truth.

We show results overall and individually for shadow and non-shadow regions (according to the binary ground truth labels). The “non-shadow” regions may contain light shadows, so that error between original and ground truth shadow-free images is not exactly zero for these regions. To show that matting helps achieve smooth boundaries, we also compare the recovery results using only the detected hard mask. We also show results using soft matte generated from the ground truth hard mask, which provides a more accurate evaluation of the recovery algorithm.

The qualitative results for shadow removal are shown in Figure 5: 5(a) shows the detection and removal results on the UCF shadow dataset [19]; 5(b) demonstrates results on our dataset; and 5(c) is an example where our shadow detector successfully detects the self-shadow on the box. An interesting failure example is shown in Fig. 5(d) where the darker parts of the checkerboard are paired with the lighter parts by the pairwise detector and thus removed in the recovery stage.

5. Conclusion and Discussion

In conclusion, we proposed a novel approach to detect and remove shadows from a single still image. For shadow detection, we have shown that pairwise relationship between regions provides valuable additional information about illumination condition of regions, compared with simple appearance-based models. We also show that by applying soft matting to the detection results, the lighting conditions for each pixel in the image are better reflected, especially for those pixels on the boundary of shadow areas. Our conclusions are supported by quantitative experiments on shadow detection and removal in Figure 4(c) and 4(e).

Currently our detection method relies on the initial segmentation which may group soft shadows with non-shadow regions. In our model, we do not differentiate whether illumination changes are caused by shadows or shading due to orientation discontinuities, such as building walls.

To further improve detection, we could incorporate more sophisticated features, such as those in Zhu et al. [19]. We could also incorporate geometry estimates into our detection framework, as in Lalonde et al. [11], which could help to remove false pairings between regions and to indicate the source of the shadow. We will make available our dataset and code, which we hope will facilitate the use of shadow detection and removal in scene understanding.

Acknowledgements

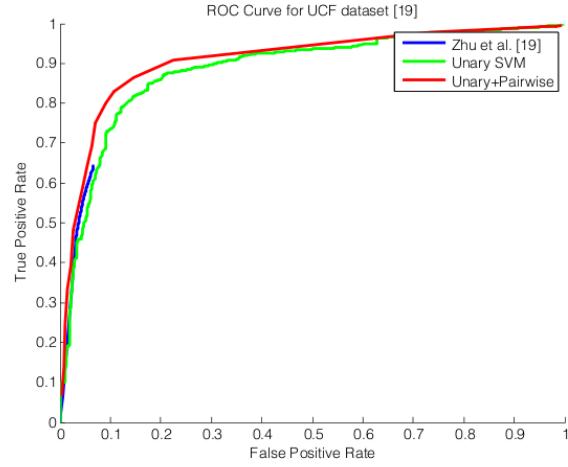
This work was supported in part by the National Science Foundation under IIS-0904209 and by a Google Research Award.

References

- [1] M. Baba and N. Asada. Shadow removal from a real picture. In *SIGGRAPH*, 2003. 3
- [2] H. Barrow and J. Tenenbaum. Recovering intrinsic scene characteristics from images. In *Comp. Vision Systems*, 1978. 1
- [3] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 23(11):1222–1239, 2001. 4
- [4] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *PAMI*, 24(5):603–619, 2002. 2
- [5] G. D. Finlayson, M. S. Drew, and C. Lu. Entropy minimization for shadow removal. *IJCV*, 85(1):35–57, 2009. 1
- [6] G. D. Finlayson, S. D. Hordley, and M. S. Drew. Removing shadows from images using retinex. In *Color Imaging Conference*. IS&T - The Society for Imaging Science and Technology, 2002. 2
- [7] G. D. Finlayson, S. D. Hordley, C. Lu, and M. S. Drew. On the removal of shadows from images. *PAMI*, 28:59–68, Jan 2006. 1
- [8] C. Fredembach and G. Finlayson. Hamiltonian path-based shadow removal. In *BMVC*, volume 2, pages 502–511, Oxford, U.K., 2005. 2
- [9] C. Fredembach and G. D. Finlayson. Fast re-integration of shadow free images. In *Color Imaging Conference*, pages

Our dataset (unary)	Shadow	Non-shadow
Shadow(GT)	0.469	0.531
Non-shadow(GT)	0.091	0.909
Our dataset (unary+pairwise)	Shadow	Non-shadow
Shadow(GT)	0.709	0.291
Non-shadow(GT)	0.057	0.943
UCF (unary)	Shadow	Non-shadow
Shadow(GT)	0.515	0.485
Non-shadow(GT)	0.053	0.947
UCF (unary+pairwise)	Shadow	Non-shadow
Shadow(GT)	0.750	0.250
Non-shadow(GT)	0.070	0.930
UCF (Zhu et al. [19])	Shadow	Non-shadow
Shadow(GT)	0.639	0.361
Non-shadow(GT)	0.067	0.934

(a) Detection confusion matrices



(b) ROC Curve on UCF dataset

	UCF shadow dataset	Our dataset
BDT+BCRF [19]	0.887	-
<i>Our method</i>		
Unary SVM	0.875	0.796
Pairwise SVM	0.673	0.794
Unary SVM + adjacent Pairwise	0.897	0.872
 Unary SVM + Pairwise	0.900	0.883

(c) Shadow detection evaluation (per pixel accuracy)

Training source	pixel accuracy on UCF dataset	pixel accuracy on our dataset
Unary UCF	0.875	0.818
Unary UCF,Pairwise UCF	0.900	0.864
Unary UCF,Pairwise Ours	0.879	0.884
Unary Ours	0.680	0.796
Unary Ours,Pairwise UCF	0.752	0.861
Unary Our,Pairwise Our	0.794	0.883

(d) Cross dataset shadow detection

Region Type	Original	No matting	Automatic matting	Matting with Ground Truth Mask
Overall	13.7	8.7	8.3	6.4
Shadow regions	42.0	18.3	16.7	11.4
Non-shadow regions	4.6	5.6	5.6	4.8

(e) Shadow removal evaluation on our dataset(pixel intensity RMSE)

Figure 4. (a) Confusion matrices for shadow detection. (b) ROC Curve on UCF dataset. (c) The average per pixel accuracy on both dataset. (d) cross dataset tasks, testing the detector trained on one dataset on another one. (e) The per pixel RMSE for shadow removal task. First column shows the error when no recovery is performed; the second column is when detected shadow masks are directly used for recovery and no matting is applied; the third column is the result of using soft shadow masks generated by matting; the last column shows the result of using soft shadow masks generated from ground truth mask.

- 117–122. IS&T - The Society for Imaging Science and Technology, 2004. 2
- [10] J.-F. Lalonde, A. A. Efros, and S. G. Narasimhan. Estimating natural illumination from a single outdoor image. In *ICCV*, 2009. 1
- [11] J.-F. Lalonde, A. A. Efros, and S. G. Narasimhan. Detecting ground shadows in outdoor consumer photographs. In *ECCV*, 2010. 1, 2, 6
- [12] E. H. Land, John, and J. Mccann. Lightness and retinex theory. *Journal of the Optical Society of America*, pages 1–11, 1971. 1
- [13] A. Levin, D. Lischinski, and Y. Weiss. A closed-form solution to natural image matting. *PAMI*, 30(2):228–242, 2008. 2, 4, 5
- [14] D. R. Martin, C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color, and

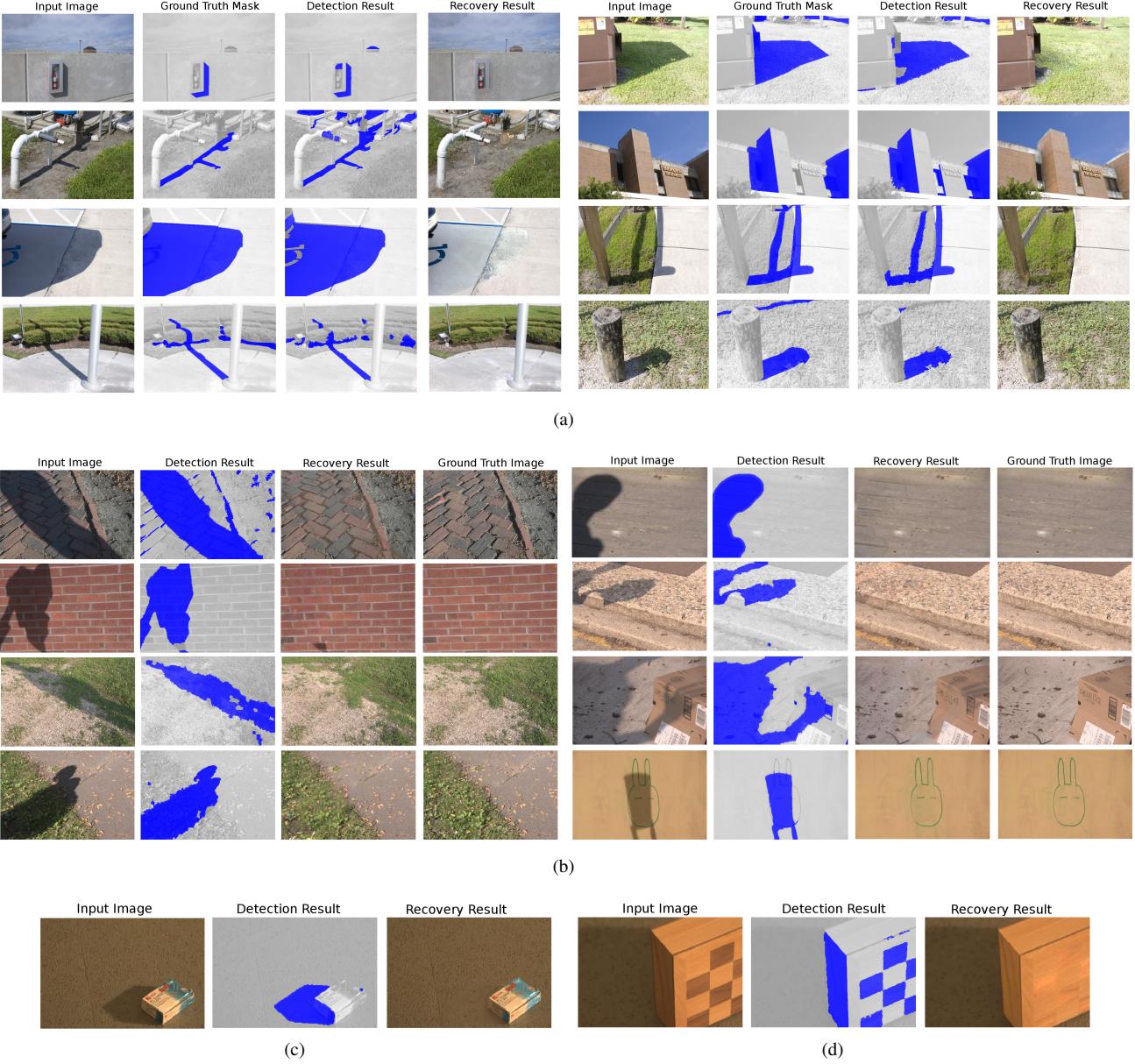


Figure 5. (a)Detection and recovery results on UCF dataset [19]. These results show that our detection and recovery framework also works well in complicated scenes. (b) Detection and recovery results on our dataset. (c) Example of detection and recovery on scenes with self-shadow. The self-shadow is correctly detected by our detector. (d) Failure example. The darker parts of the chessboard are mistakenly detected as shadow and as a result, removed in the recovery process.

texture cues. *PAMI*, 26(5):530–549, 2004. 3

- [15] B. A. Maxwell, R. M. Friedhoff, and C. A. Smith. A bi-illuminant dichromatic reflection model for understanding images. In *CVPR*, 2008. 1
- [16] S. G. Narasimhan, V. Ramesh, and S. K. Nayar. A class of photometric invariants: Separating material from shape and illumination. In *ICCV*, 2003. 1
- [17] D. L. Waltz. Generating semantic descriptions from drawings of scenes with shadows. Technical report, Cambridge, MA, USA, 1972. 1

- [18] T.-P. Wu, C.-K. Tang, M. S. Brown, and H.-Y. Shum. Natural shadow matting. *ACM Trans. Graph.*, 26(2), 2007. 2
- [19] J. Zhu, K. G. G. Samuel, S. Masood, and M. F. Tappen. Learning to recognize shadows in monochromatic natural images. In *CVPR*, 2010. 1, 2, 3, 5, 6, 7, 8