Human Shadow Removal with Unknown Light Source

Chia-Chih Chen and J. K. Aggarwal
Computer & Vision Research Center / Department of ECE
The University of Texas at Austin
{*ccchen* | *aggarwaljk*}*@mail.utexas.edu*

Abstract

In this paper, we present a shadow removal technique which effectively eliminates a human shadow cast from an unknown direction of light source. A multi-cue shadow descriptor is proposed to characterize the distinctive properties of shadows. We employ a 3-stage process to detect then remove shadows. Our algorithm improves the shadow detection accuracy by imposing the spatial constraint between the foreground subregions of human and shadow. We collect a dataset containing 81 human-shadow images for evaluation. Both descriptor ROC curves and qualitative results demonstrate the superior performance of our method.

1. Introduction

The existence of human shadows is a general problem in tracking and recognizing human activities. Shadows not only distort the color properties of the area being shaded but also complicate the edge structure of the figure as a whole. There are several factors that together determine the appearance of a shadow, for example, the view point of camera, the angle of incidence, the light intensity, and the number of light sources, etc. Further, under the sun, the dominant orientation of a human shadow changes as a function of time. Therefore, a human tracker becomes more prone to miss the target, and the motion pattern of a single action varies considerably. For simplification, by human shadow we mean a human cast shadow in contrast with a human self shadow.

When comparing a shadow with the neighboring background, the most obvious observation is the difference in luminance. Shadows not only reduce the luminance of the shaded area but also distort its chromaticity. To lower the dependence of chromaticity on luminance, several methods [6, 9, 11] have been proposed to exploit the invariant color spaces. Besides color, other features have also been shown to be useful for shadow characterization. In [1], for example, an edge map is used to segment the image into edged, smooth, and textured regions. The smooth region is regarded as the candidate area of a shadow. Edge orientation and other shadow related geometric properties are adopted as features in [2]. The paper by Prati *et al.* [8] provides a comparative survey on the literature of shadow detection.

The purpose of this work is to replace the region of a human shadow with the estimation of underlying unshaded background. Without loss of generality, we assume that human figures are posed vertically and the foreground mask is available to us. We simplify this problem by taking advantage of the fact that both human and shadow regions within a foreground blob are connected components. The task of shadow detection can thus be posed as a search for the linear boundary which best separates the two connected subregions. We propose a bottom-up classification scheme to approximate the optimal boundary. The preliminary classification is to divide foreground pixels into the intermediate classes of *shadow* and *non-shadow*. Based on the pixel locations of the labeled pixels, the secondary classification segments a connected shadow region from the foreground blob. Finally, we inpaint the detected shadow region with a Gaussian spatial filter. For reliable characterization of a shadow pixel, we extract three types of features from each pixel and its neighborhood. These features include color, pixel location, and Histogram of Oriented Gradients (HOG)[5].

This paper is organized as follows: Section 2 introduces our shadow descriptor. The proposed process for shadow removal is presented in Section 3. We demonstrate our experimental results in Section 4 and conclude in Section 5.

2. Characterization of shadow pixel

Most existing work on shadow detection [6, 9, 11] uses color as the major or the only cue to character-

ize shadows. However, in real world imagery, shadows have a wide spectrum of luminance values. Therefore, shadow detectors that mainly rely on color information are more susceptible to the changes in lighting conditions. In this work, we use a multi-cue descriptor to represent human shadows. The proposed descriptor is the concatenation of three normalized shadow distinctive features, which are detailed as follows:

Color. For accurate detection of shadows, the choice of color space is important. Various color spaces have been explored to search for a transformation which provides better discrimination of projected pixels or least effects of shadows on chromaticity. Both HSI (hue, saturation, intensity) and HSV (V for value) are popular color models in literature. Particularly, in Tsai's [11] experiments on 6 color spaces, the highest detection rate is achieved by remapping color into HSI space. Following his results, we transform RGB color into HSI space by

$$\begin{bmatrix} I \\ V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{-\sqrt{6}}{6} & \frac{-\sqrt{6}}{6} & \frac{\sqrt{6}}{3} \\ \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & 0 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

$$S = \sqrt{V_1^2 + V_2^2} \quad (2)$$

$$H = \tan^{-1}(V_2/V_1), V_1 \neq 0 \quad (3)$$

Log-polar coordinates. Connecting to the bottom of human figures, shadows appear in various orientations and shapes. We find that pixel locations in Cartesian coordinates are less informative about the coverage of a shadow. Therefore, we devise a modified log-polar coordinate system to make better use of pixel location as a feature.

Motivated by the non-uniform mapping from a human retina to visual cortex [10], a log-polar coordinate system is preferable to a Cartesian system in certain applications. In log-polar coordinate system, the origin area has a higher resolution as compared to the periphery. We modify the system in a way that the distribution of coordinate resolution approximates the confidence map of shadow coverage. As shown in Figure 1, the modified log-polar coordinates are superimposed onto a human-shadow foreground blob. Via this representation, shadow pixels will occupy mostly the low-resolution peripheral area, where the certainty about a shadow's presence is low.

A reference point, (x_{ref}, y_{ref}), is first located before the computation of the projected pixel location. Here y_{ref} is equal to the y-coordinate of the top of the foreground blob. x_{ref} corresponds to the x-intercept of the major axis of a person's silhouette. We compute x_{ref} by locating the peak of the horizontal projection histogram, which is an accumulation of foreground pixels

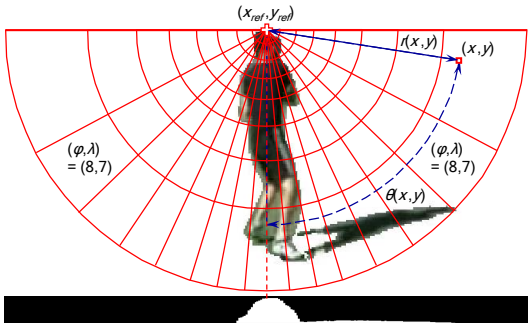

Figure 1. Top: diagram of the modified log-polar coordinate system. Bottom: horizontal projection histogram.

along the vertical axis (Figure 1). The radius, R, is the maximum distance from the reference point to the blob boundary. Θ is a fixed value which is set to $\pi/2$. The radial and angular resolution of the coordinate system are symbolized by Φ and Λ, respectively. The mapping from (x,y) to (ϕ,λ) is defined as

$$\phi(x,y) = \left\lceil \Phi + 1 - (\Phi+1)^{1-r(x,y)/R} \right\rceil \quad (4)$$

$$\lambda(x,y) = \left\lceil \Lambda + 1 - (\Lambda+1)^{1-\theta(x,y)/\Theta} \right\rceil \quad (5)$$

where

$$r(x,y) = \sqrt{(x-x_{ref})^2 + (y-y_{ref})^2} \quad (6)$$

$$\theta(x,y) = \tan^{-1}\left|(y-y_{ref})/(x-x_{ref})\right| \quad (7)$$

HOG. We use orientation transformed single-cell HOG as one component feature for two reasons. First, the dominant edge orientation of a human local silhouette is mostly close to vertical, while the dominant orientation of a shadow can be in all directions. Second, strong edge structure is not always available from the region of a shadow [4].

As shown in Figure 2, a human figure is connected with shadows oriented in various directions. The arrows and square areas represent gradient vectors and HOG cells, respectively. For the cells on the human figure (human HOG), the corresponding HOG vectors are expected to have greater values over the horizontal bins. However, the maximum bin of a shadow region HOG vector (shadow HOG) is closely related to the dominant orientation of the shadow.

In [5], unsigned gradient vectors are used for HOG computation. However, there is one problem with the original HOG representation, which adversely effects detection accuracy. In a human HOG, there are two bins (0 and π) that correspond to the horizontal direction.

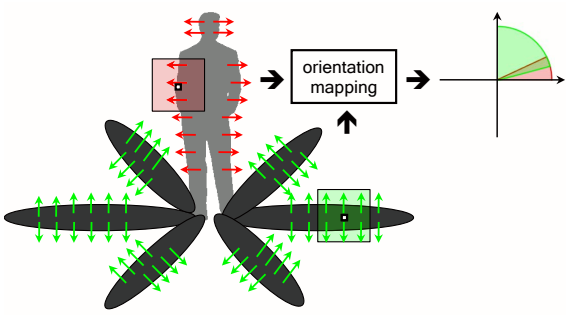

Figure 2. Left: Single-cell HOG on human and shadow subregions. Right: Schematic drawing of the Eq. (8) projected human (red) and shadow (green) gradient vectors on a polar coordinate.

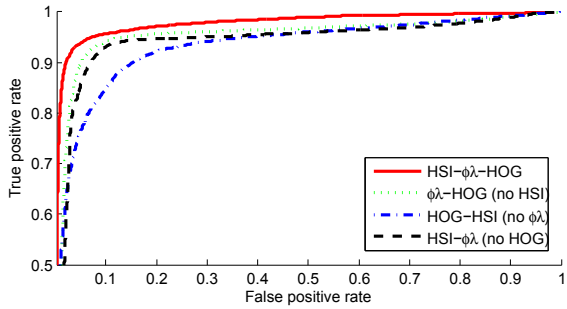

Figure 3. ROC curves of the proposed (solid line) and the reduced feature shadow descriptors (dashed lines).

Therefore, a horizontal gradient vector may vote for either of the two bins depending on the sign of human-background intensity difference and the cell location (left or right body parts specifically). To solve this problem, we apply the following mapping.

$$\phi = \frac{\pi}{2} - \mathrm{sgn}(\frac{\pi}{2} - \theta) \cdot (\frac{\pi}{2} - \theta) \qquad (8)$$

Here θ represents the unsigned angle of a gradient vector, which is transformed into ϕ for HOG computation.

3. Shadow detection and removal

In this work, we aim at finding a boundary which divides a human-shadow foreground blob into its ground truth subregions. We propose a 3-stage process to implement this idea. The first stage performs a binary classification on pixels of a foreground blob. A RBF kernel Support Vector Machines (SVM) classifier is trained with descriptors from the labeled shadow images.

Using the first stage classification results as intermediate ground truth, the second stage computes the linear boundary within the foreground blob that minimizes the classification error. For this purpose, pixel coordinates alone are used as a feature. We adopt a linear classifier to avoid the overfitting problem from a complex decision boundary. In other words, the linear classifier divides a foreground blob into human and shadow subregions by referring to stage one labeled pixel locations. In the third stage, we inpaint the detected shadow region with the estimation of an unshaded background. A 2D spatial filter is applied to replace the color of each detected pixel with the Gaussian-weighted average of neighboring background pixel values.

As can be imagined, both feature extraction and nonlinear classification are time consuming processes. Therefore, without loss of accuracy, we use a downsampled human bounding box to speed up the process. We first compute the linear boundary from the resized foreground image, and then apply the interpolated boundary to the original image.

4. Experimental results

To evaluate our method, we compose a dataset which includes 81 human-shadow images from UCF YouTube Action Dataset [7]. As shown in Figure 4, the selected images are the single frames from 27 YouTube human action videos. We extract 3 nonconsecutive frames from each video. The foreground mask of each image is manually segmented into *shadow* and *non-shadow* subregions. We perform two types of experiments to show the detailed performance of the proposed descriptor and the accuracy of our method as a whole. We randomly divide the videos into two parts, which contain 14 and 13 videos respectively. We use all the 42 images from the 14-video part for training and the rest of the 39 images for testing.

In the first experiment, we compare the ROC curve of the proposed shadow descriptor with those of the reduced feature descriptors. The 4 descriptors in comparison are HSI-$\phi\lambda$-HOG, $\phi\lambda$-HOG, HOG-HSI, and HSI-$\phi\lambda$, where $\phi\lambda$ represents the modified log-polar coordinates. We evaluate the performance by computing their area under the ROC curves (AUC) in Figure 3. As expected, the proposed descriptor (solid line) contains all the features and outperforms others. To measure the contribution from each feature, we use the AUC difference between the full feature descriptor and the reduced feature descriptor as a measurement. For example, the

Figure 4. Processing sequences of our method. The images in each sequence correspond to (a) the original, (b) detected pixels marked, (c) detected region marked, and (d) shadow removed image.

importance of HOG feature is measured by the AUC difference between the red and black curve. We are surprised to find that the HSI color feature contributes the least to the detection. The pixel location in modified log-polar coordinates is the most discriminative feature for shadow detection.

For the second experiment, we average the per image detection accuracy over 5 rounds of random video partitions. The average accuracy is *96.37%*. Moreover, we measure the accuracy improvement by imposing the spatial constraint. That is, we compare the detection accuracy before and after the line fitting. The average accuracy improvement per image is *1.3%*, while the accuracy improvement is brought to *84.6%* of the testing images. In MATLAB® implementation on a Pentium D 2.8GHz PC, the average time required to process a 10,000-pixel foreground blob is about 3 seconds. Figure 4 demonstrates the qualitative results. We show 7 sets of the processing sequence.

5. Conclusions

We present an effective technique to remove human shadows. The major contribution of this work is twofold. First, we propose a multi-cue shadow descriptor which provides more reliable characterization of shadows. Our shadow detector is able to achieve high accuracy, although the classifier is trained and tested on images from different sets of videos. Second, the proposed 3-stage process largely reduces the risk of classifying human pixels as *shadow* and leaves the shadow removed figure as an intact region. Our method has led to accurate recognition of activities in [3].

6. Acknowledgement

This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No. HR0011-08-C-0135.

References

[1] S. Bi, D. Liang, X. Shen, and Q. Wang. Human cast shadow elimination method based on orientation information measures. *In ICAL*, 2007.

[2] C. Chang, W. Hu, J. Hsieh, and Y. Chen. Shadow elimination for effective moving object detection with gaussian models. *In ICPR*, 2002.

[3] C.-C. Chen and J. K. Aggarwal. Recognizing human action from a far field of view. In *IEEE Workshop on Motion and Video Computing (WMVC)*, 2009.

[4] H.-T. Chen, H.-H. Lin, and T.-L. Liu. Multi-object tracking using dynamical graph matching. *In CVPR*, 2001.

[5] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. *In CVPR*, 2005.

[6] G. D. Finlayson, M. Drew, and C. Lu. Intrinsic images by entropy minimization. *In ECCV*, 2004.

[7] J. Liu, J. Luo, and M. Shah. Recognizing realistic actions from videos "in the wild". *In CVPR*, 2009.

[8] A. Prati, I. Mikic, R. Cucchiara, and M. M. Trivedi. Comparative evaluation of moving shadow detection algorithms. In *In CVPR-EEMCV*, 2001.

[9] E. Salvador, A. Cavallaro, and T. Ebrahimi. Shadow identification and classification using invariant color models. *In ICASSP*, 2001.

[10] E. Schwartz. Spatial mapping in primate sensory projection: analytic structure and relevance to perception. *Biological Cybernetics*, 25:181–194, 1977.

[11] V. Tsai. A comparative study on shadow compensation of color aerial images in invariant color models. *IEEE Trans. on Geoscience and Remote Sensing*, 2006.