Pattern Recognition 63 (2017) 30–44

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Stochastic shadow detection using a hypergraph partitioning approach☆

Vitor Gomes, Pablo Barcellos, Jacob Scharcanski*

Institute of Informatics, Federal University of Rio Grande do Sul, 91501-970, Porto Alegre, RS, Brazil

A R T I C L E I N F O

Article history:
Received 9 July 2015
Received in revised form
14 September 2016
Accepted 15 September 2016
Available online 20 September 2016

Keywords:
Shadow detection
Chromaticity
Gradient correlation
Hypergraph segmentation

A B S T R A C T

Discriminating shadows from the objects casting them often is challenging in practice, since the moving targets and their shadows tend to present similar motion patterns, and foreground detection methods often confuse cast shadows with foreground objects. To overcome these shadow detection difficulties, we propose a new stochastic shadow detection approach. In the proposed method, chromatic and gradient information are integrated with image hypergraph segmentation using a cascade of shadow/non-shadow classifiers, and a stochastic majority voting scheme is used to detect the shadow regions. The proposed method receives as input the segmented foreground objects and their cast shadows (mask), and outputs the shadows detected in the foreground mask. The experimental results were obtained with seven well known datasets, and suggest that the proposed shadow detection scheme can be more robust to different video acquisition conditions than other shadow detection methods, that are representative of the state-of-the-art.

1. Introduction

Detection of moving objects (i.e. foreground detection) is an important step in many image- and video-based monitoring applications [1,2], like traffic monitoring [3–5] and people detection and/or tracking. Unfortunately, cast shadows have the same motion patterns as the objects casting them, and most foreground detection methods tend to confuse cast shadows with foreground objects [6], downgrading the performance of these methods. Besides, cast shadows usually are adjacent to moving objects, and the segmentation process tends to merge foreground objects separated by cast shadows, leading to erroneous object detection and recognition [7]. Therefore, the reliable detection of shadows is an important step in foreground and background detection.

Shadowing occurs when a light source is occluded by an object in the scene. The part of the object that is not illuminated is called self-shadow, and the area projected on the scene corresponding to the illumination occlusion is called cast shadow, or moving cast shadow if the object is in motion [8]. A substantial amount of work has been published on this topic and some representative recent contributions are discussed next.

Prati et al. [9] proposed two metrics to evaluate the performance of shadow detection methods, namely shadow detection rate (η) and shadow discrimination rate (ξ), that are calculated as follows:

$$\eta = \frac{TP_S}{TP_S + FN_S} \quad \text{and} \quad \xi = \frac{TP_F}{TP_F + FN_F}, \tag{1}$$

where TP and FN are the true positive and false negative pixels detected in shadow (S) and (non-shadow) foreground (F) regions. If the shadow detection rate η increases, the number of correctly detected cast shadow pixels increases. The shadow discrimination rate ξ tries to estimate the proportion of foreground pixels that are mistakenly labeled as background pixels, and increases when the number of pixels mistakenly assigned to the background decreases. There is a compromise between the two measures η and ξ.

A recent survey on shadow detection organizes the available methods in four categories [10]: (a) chromaticity-based methods; (b) physically inspired methods; (c) geometry-based methods; and (d) texture-based methods.

Cucchiara et al. proposed a chromaticity-based method [1] that tries to detect shadows by measuring the rate of change between the HSV components of a video frame and a background reference. This method assumes that shadow pixels in the video frame and in the background reference do not differ substantially in their hue component, and have low saturation and intensity values. However, just color information may not be able to discriminate

☆The authors would like to thank Digicon LTDA. and CAPES-Coordenação de Aperfeiçoamento de pessoal de Nível Superior-for financial support.

* Corresponding author.
E-mail addresses: vogomes@inf.ufrgs.br (V. Gomes), prmbarcellos@inf.ufrgs.br (P. Barcellos), jacobs@inf.ufrgs.br (J. Scharcanski).

http://dx.doi.org/10.1016/j.patcog.2016.09.008
0031-3203/© 2016 Elsevier Ltd. All rights reserved.

correctly shadows and foreground objects that have darker colors similar to shadows (e.g. dark cars and their cast shadows).

Physically based methods try to learn the appearance of shadow pixels, and may achieve higher accuracies than chromaticity-based methods [11]. Unfortunately, these methods tend to fail when dealing with objects having chromaticities similar to the background [10].

Geometry-based methods evaluate geometric features of shadows, such as the shadow orientation, size and shape [12]. The disadvantage of these methods is that previous knowledge of the scene is required, which often is unavailable [13]. These methods may fail when the shadow region and the foreground object have similar orientations [10].

Texture-based methods assume that shadows preserve most of the scene textures. For example, the method proposed in [7] uses chromaticity information to identify shadow regions, and then uses gradient information to refine the initial estimates of shadow or non-shadow pixels. This method provides good shadow detection results, and improves on the results of the chromaticity method in [1] by combining chromaticity and gradient information. However, significant color differences between foreground objects and cast shadows are required to obtain high quality shadow detection with this method [7].

In a recent survey presented by Al-Najdawi et al. [14], a different taxonomy was proposed to organize shadow detection methods. The shadow methods are classified based on object/environment dependency, or based on the implementation domain, which can be pixel domain or transform domain. According to [14], object/environment dependent methods are designed to detect a particular type of shadow (e.g. vehicle or human cast shadows) in a specific environment (e.g. indoor or outdoor scenes). Pixel domain methods are further divided into monochrome domain and color domain methods, and use pixel information (e.g. color) for shadow detection. Transform domain methods use different types of information for shadow detection, such as frequency domain, texture, or yet geometric information. Some transform domain methods are illustrated next.

The method proposed in Al-Najdawi et al. [15] removes insignificant cast shadows in video sequences based on edge and region information in multiple frames. A shadow is called insignificant when edges of the shadow region are not as sharp as the edges of the corresponding object. First, a mask containing moving objects and cast shadows is obtained. Then, the Canny edge detector is used to generate an edge map. The shadow regions are then removed using edge matching and region growing in multiple frames. Finally, the boundaries of the objects are improved and noise is removed by using a post-processing procedure. The motivation for this approach is that an insignificant shadow region often appears in an area where the gray levels change gradually from the background to the shadowed area. A disadvantage of this method is that if the object moves slowly, there is little change at the boundary between the object and the background, affecting negatively the results [15].

Guo et al. [16] proposed a single-image shadow detection and removal method based on a paired-regions approach. First, the image is segmented using the mean-shift algorithm [17], then a trained classifier based on color and texture information is used to estimate the confidence that each region is a shadow region. This classifier is trained using manually labeled regions, and a Support Vector Machine (SVM) is used to find similar regions under different illumination conditions. Next, regions with the same and different illuminations are represented by a relational graph, which is partitioned using graph-cuts in shadow and non-shadow regions. Finally, the results are improved by using image matting to smooth the transitions between shadow and non-shadow regions, and shadow-free images are obtained by relighting. A disadvantage of this method is that it requires training a shadow classifier, which may involve manual labeling of shadow regions for generic scenes.

Another typical transform domain method [13] handles shadow detection using the robust wavelet watershed segmentation algorithm [18–20]. The segmented image regions are classified as shadow and non-shadow regions using a HSV-based approach, similar to the method proposed by Cucchiara et al. [1]. A problem with this method is that using only HSV color information to classify shadow and non-shadow regions may lead to erroneous results when the foreground objects colors are similar to shadows.

The method proposed in [21] uses the MTM (Matching by Tone Mapping) transformation as the distance between image patches of a video frame and a background reference. Shadowing effects are assumed to be non-linear tone mappings of the background gray levels. Since the MTM distance is invariant to non-linear mappings between corresponding image patches in shadow and non-shadow-regions, this MTM distance results in small values for shadowed areas and in large values for foreground patches that differ from the background [21]. To detect shadows, the MTM transformation is applied to the spatial neighborhood of each pixel in the foreground, generating a MTM distance map. The Otsu thresholding method [22] is then applied to the MTM distance map to discriminate between shadow and non-shadow regions. Unfortunately, thresholding the MTM distance map to guarantee an accurate discrimination between shadow and non-shadow regions is not trivial.

Khare et al. [23] proposed a shadow detection method based on the dual-tree complex wavelet transform to measure the difference between a video frame and the background reference in the HSV color space. The dual-tree complex wavelet transform of the difference images in the saturation and value channels is calculated. The standard deviation of the wavelet coefficients is computed, each coefficient is adaptively thresholded, and the image is reconstructed by discarding the wavelet coefficients that are associated to shadows. Finally, morphological operations are used as post-processing. This method provides interesting results, but since the wavelet coefficients are calculated based on color information only, the method performance tends to decrease if the foreground objects have chromaticities similar to shadows.

The method in Lalonde et al. [24] detects automatically shadows of objects on the ground, from a single image. They assume that the types of materials constituting the ground in outdoor scenes are limited (e.g., asphalt, brick, stone, mud, grass, concrete). Thus, the appearances of the shadows on the ground are not as widely varying and can be learned from a set of training images. The first stage of the detector consists in training a decision tree classifier on a set of sensitive shadow features based on each edge of the image. Then, a Conditional Random Field (CRF) based optimization is used to enforce a grouping of the shadow edges, creating longer contours. Finally, a global scene layout descriptor, specifically trained to detect grounds in images, is incorporated within the CRF. A disadvantage of this method is its focus only on outdoor scenes. The ground material in indoor scenes may vary greatly, resulting in classification errors. Also, using only a single image (i.e. no background model) sometimes do not provide enough information to accurately classify cast shadows.

Our proposed method was designed to overcome the common drawbacks found in most shadow detection methods available. In the proposed method, chromatic and gradient information are integrated with the image (or video frame) hypergraph segmentation, and finally a stochastic majority voting scheme is used to detect shadow regions. It is assumed that the foreground objects (mask) have already been detected by other methods, and that the foreground mask contains both the object and its shadow. Initially, a weighted image hypergraph is partitioned into K sub-