Image and Vision Computing 41 (2015) 42–53

Contents lists available at ScienceDirect

Image and Vision Computing

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

Chromatic shadow detection and tracking for moving foreground segmentation☆

Ivan Huerta [a,d,*], Michael B. Holte [b], Thomas B. Moeslund [b], Jordi Gonzàlez [c]

[a] Institut de Robòtica i Informàtica Industrial (CSIC-UPC), Parc Tecnològic de Barcelona, Llorens i Artigas 4-6, 08028 Barcelona, Spain
[b] Department of Architecture, Design and Media Technology, Aalborg Universtity, 9220 Aalborg, Denmark
[c] Computer Vision Center & Department of Computer Science (UAB), Edifici O, Campus UAB, 08193 Bellaterra, Spain
[d] DPDCE, University IUAV, Santa Croce 1957, 30135 Venice, Italy

ARTICLE INFO

Article history:
Received 15 November 2013
Received in revised form 6 October 2014
Accepted 14 June 2015
Available online 2 July 2015

Keywords:
Detecting moving objects
Chromatic shadow detection
Temporal local gradient
Spatial and temporal brightness and angle distortions
Shadow tracking

ABSTRACT

Advanced segmentation techniques in the surveillance domain deal with shadows to avoid distortions when detecting moving objects. Most approaches for shadow detection are still typically restricted to penumbra shadows and cannot cope well with umbra shadows. Consequently, umbra shadow regions are usually detected as part of moving objects, thus affecting the performance of the final detection. In this paper we address the detection of both penumbra and umbra shadow regions. First, a novel bottom-up approach is presented based on gradient and colour models, which successfully discriminates between chromatic moving cast shadow regions and those regions detected as moving objects. In essence, those regions corresponding to potential shadows are detected based on edge partitioning and colour statistics. Subsequently (i) temporal similarities between textures and (ii) spatial similarities between chrominance angle and brightness distortions are analysed for each potential shadow region for detecting the umbra shadow regions. Our second contribution refines even further the segmentation results: a tracking-based top-down approach increases the performance of our bottom-up chromatic shadow detection algorithm by properly correcting non-detected shadows. To do so, a combination of motion filters in a data association framework exploits the temporal consistency between objects and shadows to increase the shadow detection rate. Experimental results exceed current state-of-the-art in shadow accuracy for multiple well-known surveillance image databases which contain different shadowed materials and illumination conditions.

1. Introduction

A fundamental problem for all automatic video surveillance systems is detecting objects of interest in a given scene [1]. A commonly used technique for segmentation of moving objects is background subtraction [2,3]. Detection of moving regions (i.e., the foreground) is achieved by comparing the current image from a reference background image in a pixel-by-pixel manner. There have been many segmentation problems already tackled in the literature related to motion segmentation [4], such as bootstrapping [5,6], changing background [7–9] and sudden illumination changes [10], to cite but a few, but one of the critical challenges is still shadow detection. Although this issue has been widely studied in [11–13], shadow segmentation is still far from being solved. The focus of this paper is to cope with the shadow problem.

Shadows can be divided into two categories: *static* and *dynamic* (*moving*) shadows. Static shadows occur due to static background objects (e.g., trees, buildings, and parked cars) blocking the illumination from a light source. Static shadows can be successfully incorporated into the background model thus being properly detected. However, the impact of dynamic shadows is critical for foreground segmentation, since objects can be merged or hidden by other objects, and both their size and shape can be distorted. This results in a reduction of the performance of foreground detection approaches applied in scene monitoring, object recognition, target tracking and people counting.

Dynamic shadows are more problematic, since they are due to the moving objects (e.g., people and vehicles). Dynamic shadows can take any size and shape, and can be *penumbra* (soft) or *umbra* (hard) shadows. Penumbra shadows exhibit low values of intensity but similar chromaticity values w.r.t. the background. Umbra shadows can exhibit different chromaticity values than the background, and their intensity values can be similar to those of any new object appearing in a scene. When the chromaticity of umbra shadows differs enough from the

☆ This paper has been recommended for acceptance by Rita Cucchiara.

* Corresponding author at: Institut de Robòtica i Informàtica Industrial (CSIC-UPC), Parc Tecnològic de Barcelona, Llorens i Artigas 4-6, 08028 Barcelona, Spain.
E-mail addresses: ihuerta@iri.upc.edu, huertacasado@iuav.it (I. Huerta), mbh@create.aau.dk (M.B. Holte), tbm@create.aau.dk (T.B. Moeslund), jordi.gonzalez@cvc.uab.cat (J. Gonzàlez).

http://dx.doi.org/10.1016/j.imavis.2015.06.003
0262-8856/

chromaticity of the global scene illumination, we define this as *chromatic shadow*. Consequently, umbra shadows are significantly more difficult to detect, and therefore usually detected as a part of moving objects by current state-of-the-art approaches.

This paper presents an approach which successfully detects umbra and penumbra shadows. First a bottom-up approach for detection and removal of chromatic moving shadows in surveillance scenarios is presented based on our previous work [14]. We apply a multi-stage approach combining multiple cues, namely colour, gradient information, and shadow statistics. Secondly, a top-down architecture based on a tracking system is proposed to enhance the chromatic shadow detection presented in [14]. This step is required since shadows can be lost for a number of frames due to camouflage. In these cases the use of motion filters allows our proposed system to track shadows, thus improving the accuracy and robustness of the final foreground detection performance. Experimental results show that applying the top-down shadow tracking, the shadow detection rate is improved by approximately 13%.

Secondly, a top-down architecture based on a tracking system is proposed to enhance the chromatic shadow detection, where motion filters are used for tracking. This step is required since shadows can be lost for a number of frames due to camouflage, so in these cases the use of basic motion filters allows our proposed system to track shadows thus improving the accuracy and robustness of the final foreground detection performance.

The remainder of the paper is organised as follows. Next section reviews the field in shadow detection and tracking, along with our contributions to this subject. In Section 3, the theoretical concept of our approach is outlined. The algorithm for foreground segmentation, along with the detection and removal of chromatic moving shadows, is described in Section 4. The top-down process used to enhance the shadow detection is described in Section 5. Finally, we present experimental results in Section 6 and concluding remarks in Section 7.

2. Related methodology

Shadow detection is a major field of research within computer vision. Even though many algorithms have been proposed [11–13], the problem of detection and removal of shadows in complex environments is still far from being completely solved. A common direction in the research is to assume that shadows decrease the luminance of an image, whilst the chrominance stays relatively unchanged [15,16]. However, this is not the case in many scenarios, e.g., in outdoor scenes. Other approaches apply geometrical information. Onoguchi [17] uses two cameras to eliminate the shadows of pedestrians based on object height, where objects and shadows must be visible to both cameras. Ivanov et al. [18] apply a disparity model, which is invariant to arbitrarily rapid changes in illumination, for modelling the background. However, to overcome rapid changes in illumination at least three cameras are required. In [19], Salvador et al. exploit the fact that a shadow darkens the surfaces on which it is cast, to identify an initial set of shadowed pixels. This set is then pruned by using colour invariance and geometric properties of shadows. Hsieh et al. [20] first separate the objects of interest and assume that the objects and their shadow have different orientations. Then, several features like orientation, mean intensity and centre position of a shadow region are used to parametrize a shadow model. It should be noted that most of the approaches which apply geometrical information require shadows to be cast on a flat plane, and give strong assumptions that need to be fulfilled.

Another popular approach is to exploit colour differences between shadow and background in different colour spaces. In [21], Cucchiara et al. consider the hypothesis that shadows reduce surface brightness and saturation whilst maintaining the hue properties in the HSV colour space. Liu et al. [22] propose another approach working in the HSV colour space, which combines local and global features for shadow removal. Schreer et al. [23,24] adopt the YUV colour space, whilst Horprasert et al. [25], Kim et al. [16] and [26] build a model in the RGB colour space to express normalised luminance variation and chromaticity distortions. However, these methods require illumination sources to be white, and assume that shadow and non-shadow have similar chrominance.

Several authors use textures to obtain a segmentation without shadows. The idea is that the structure of the texture/gradients/edges of regions lit by shadow is unchanged. Leone and Distante [27] propose a texture-based approach using a preliminary procedure in order to evaluate the photometric information for all pixels marked as foreground. This process shows how much darker the segmented regions are with respect to the background model. Next, texture analysis is performed by projecting the neighbourhood of pixels onto a set of Gabor functions. Another algorithm for detection of moving cast shadows, based on a local texture descriptor called Scale Invariant Local Ternary Pattern (SILTP), is presented by Qin et al. [28]. Zhang et al. [29] use ratio edges to detect and locate where the shadows are, and in Chen et al. [30] HOG descriptors of shadows are learned using SVMs to locate shadow regions. Sanin et al. [31] propose a method similar to Huerta et al. [14] where gradient magnitude and gradient orientation are used to detect shadows, but based on the gradient direction correlation. However, this paper [31] does not take into account cases where there are no texture/gradient in the bacground model. Whereas Heikkila and Pietikainen [32] apply Local Binary Patterns. Nevertheless, it still fails to detect umbra shadows.

To overcome these shortcomings, a number of approaches apply colour constancy methods, combine different techniques or use multi-stage approaches. In addition to scene brightness properties, Stauder et al. [33] extract edge width information to differentiate penumbra regions from the background. In [34], Finlayson et al. use shadow edges along with illuminant invariant (intrinsic) images to recover full colour shadow-free images, and in [35] the authors propose an entropy-based approach. Even so, a part of the colour information is lost in removing the effect of the scene illumination at each pixel in the image. Weiss [36] computes the reflectance edges of the scene to obtain an intrinsic image without shadows. However, this approach requires significant changes in the scene, and the reflectance image also contains the scene illumination. Huang and Chen [37] adopt a bi-illuminant model and apply a Gaussian Mixtures Model (GMM) and confidence-rated learning, whilst Martel et al. introduce a parametric approach based on Gaussian mixtures (GMSM) [38]. Additionally, they propose a non-parametric framework based on the physical properties of light sources and surfaces, and apply spatial gradient information to reinforce the learning of the model parameters [39]. Nadimi and Bhanu [40] propose a multi-stage approach for outdoor scenes, which is based on a spatio-temporal albedo test and dichromatic reflection model. Finally, a number of authors have introduced methods for shadow removal using only one single image [41–46], however, the focus of this work is shadow detection and removal in video sequences. Comparative studies of shadow detection techniques can be found in [11,47].

Exploiting temporal information for shadow detection has been rarely used. Liu et al. [22] use temporal information in order to avoid misclassifying the object under segmentation as shadow. To solve this, a nearest neighbour match method is used to track the object by checking if it is a foreground object in the previous frames. However, the authors only use tracking to recover misdetected objects and thus lose the shadow information. In our case, we take advantage of this information to track shadows, resulting in both improved object and shadow detection. In the previous papers, when a shadow has successfully been detected it is usually removed instantly, since only the object is of interest for further processing and not the shadow. As a result, the shadow information is lost. Our idea is to use this information to improve other aspects of object and shadow detection and tracking. Specifically, if a detected shadow is tracked over time instead of being discarded, it could be used to improve the shadow detection and possibly the object detection and tracking as well.