Journal of Visual Communication and Image Representation 55 (2018) 504–517

Contents lists available at ScienceDirect

Journal of Visual Communication and Image Representation

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

Moving cast shadow detection using scale-relation multi-layer pooling features☆

Chih-Wei Lin[a,b,c,*]

[a] College of Computer and Information Sciences, Fujian Agriculture and Forestry University, China
[b] College of Forestry, Fujian Agriculture and Forestry University, China
[c] Forestry Post-doctoral Station of Fujian Agriculture and Forestry University, China

ARTICLE INFO

Keywords:
Shadow removal
Moving cast shadow
Scale-relation
Multi-layer pooling
Ensemble decision scheme

ABSTRACT

Moving cast shadows detection and removal are indispensable for object detection and are the problems in visual surveillance applications which have been studied over the years. However, finding an efficient model that can handle the issue of moving cast shadow in various situations is still challenging. Unlike prior methods, we use a data-driven method without strong parametric assumptions or complex models to address the problem of moving cast shadow. In this paper, we propose a novel feature-extracting framework called Scale-Relation Multi-Layer Pooling Feature Extracting (SMPF) which includes two main tasks: (1) Scale-Relation Scheme (SRS), (2) Multi-Layer Pooling Scheme (MLPS). By leveraging the scale space, SRS firstly decomposes feature images of each shadow properties into various scales and further considers the relationship between adjacent scaled feature images of each shadow properties to extract the scale-relation features. Then, we design the multi-layer pooling scheme (MLPS) to integrate the features in a local region and to reduce the dimension of extracting features. After that, the density map is generated for various properties of shadow with low dimension. Finally, to seek the criteria for discriminating moving cast shadow, we use random forest algorithm as the ensemble decision scheme. The main contributions of this study are (1) we design the features with multi-scale which can provide abundant information to describe the moving cast shadow, (2) the multi-layer pooling scheme generates the density map to integrate and reduce the dimensions of features. Experiments on the popular benchmarks and the proposed dataset with benchmarks demonstrate that the proposed method can achieve the performances of the popular methodologies.

1. Introduction

Visual surveillance related systems (e.g., smart home, security monitoring, and intelligent transportation systems) are important but challenging, and there has been much interest in automatically detecting objects in various environments [1–4]. For instance, when monitoring the abnormal behavior events, the useful information, such as locations, shapes, and sizes of the objects are extracted to exactly determine what the behavior is. The methodologies of object detection segment the image into foreground and background, in which the foreground also refers to moving objects and the shadow is an extension of the object. Shadow exists when the natural or artificial light is obscured by the objects and that can be divided into self and cast shadows. Self-shadow is cast on its surface which is the part of objects. Cast-shadow is cast on a surface in the direction of the direct light. Cast-shadow has similar characteristics of objects which cause the misclassification of objects detection and makes the object detection methods less reliable.

Numerous of studies have been presented for cast shadow detection [5,6]in visual surveillance system. According to the frequency of sampling, those studies can be classified into two categories, *the static non-successive image*, and *the image sequence*, according to the frequency of sampling. Although the methodologies of both categories might be similar, the applications are various. In *the static non-successive image*, various information is extracted to discriminate shadow and non-shadow regions, such as gradients [7,8], colour [9,10], and intensities [11]. In addition to hand-crafted features, learning features automatically by using convolutional deep neural networks (ConvNets) is developed in shadow detection and removal [12,13]. Although, the mentioned methodologies perform well in the static non-successive

☆ This paper has been recommended for acceptance by Anderson Rocha.

* Address: College of Computer and Information Sciences, Fujian Agriculture and Forestry University, China.
E-mail address: cwlin@fafu.edu.cn.

https://doi.org/10.1016/j.jvcir.2018.06.028
Received 16 December 2017; Received in revised form 26 April 2018; Accepted 30 June 2018
Available online 11 July 2018
1047-3203/

images, some of these techniques have high computational cost, and others are not applicable in surveillance system which lose the characteristic of the immediacy.

Image sequence is also called the video-based approaches and it has static and dynamic shadows either in indoors or outdoors. The static shadows are cast by the static objects, such as buildings, parked cars and tables, which are illuminated by a light source and are modeled as background [14,15]in visual surveillance system. In contrast, the dynamic shadows are also referred to the moving cast shadows which are generated by moving objects, such as vehicles and pedestrians, and also detected as the foreground. In the scope of object detection, the motion patterns of a moving object and its cast shadow are similar, and that significantly affects the performance and arises the paradox of detecting objects. Many studies have been presented to deal with moving cast shadows, some of those researches have the limitation of parameter assumption [16] and others can only be used in the specific scenes. The moving cast shadow detection and removal is a challenging issue in scene understanding; it has numerous factors, such as various light sources, intensities of light, and angles of incidence, that affect the appearance of shadow features. More specifically, the two critical challenges to develop an ideal cast shadow detection algorithm in visual surveillance system are: (1) the characteristics between moving objects and shadows are similar, such as geometries, chromaticities and motion properties, (2) foreground and shadow are similar to background, due to the problem of camouflage between parts of moving objects and its shadows and their corresponding backgrounds.

In this paper, we design a novel feature-extracting method called Scale-Relation Multi-Layer Pooling Feature Extracting (SMPF) based on the concept of data driven approach to instead of detecting shadow with strong parametric assumptions or complex model. SMPF firstly decomposes the feature images of each shadow property into various scales and leverages the relationships between adjacent scales to extract the scale-relation features for various characteristics of shadow based on our previous work (SRF) [17]. Then, we design a multi-layer pooling scheme to increase the invariance of the feature and to generate the density map for various characteristics of properties of shadow which concisely describe the feature of shadow and efficiently discriminate the moving cast shadow. Finally, we associate an ensemble decision scheme with density map of various properties to construct the model for moving cast shadow detection. This study is the extension of our previous conference paper [17]. There are two main tasks: (1) Scale-Relation Scheme (SRS), (2) Multi-Layer Pooling Scheme (MLPS), in which the MLPS is the extension compared with our previous conference paper. SRS focuses on generating feature with abundant information to discriminate foreground and shadow. However, the dimensions of the generated features is high and that has high computation cost when predicting. This study maintains the advantage of our previous conference paper, besides, we adjust the framework to reduce dimensions of feature. The proposed multi-layer pooling scheme (MLPS) can efficiently reduce the dimensions of the generated features and improve the computation performance in prediction. In the experiment, we strengthen the analyzation by comparing the proposed method with several popular methods. Moreover, we provide a new dataset with various scenes to fill the gap of the widely used dataset and further use CDnet dataset which has a large amount of benchmarks in experiments.

The rest of this paper is organized as follows: First, we present the works which are related to our study in Section 2. Then, the framework of the proposed method are described in Section 3, after which we introduce the extraction of scale-relation features of various characteristics of each shadow properties and present the multi-layer pooling scheme to generate the density maps which concisely describe the shadow characteristics with low dimensions in Section 4. Experimental results are presented in Section 5, and Section 6 gives the conclusions.

2. Related work

Cast shadow is considered as part of the moving object and that affects the results of object detection. Numerous frameworks dedicate to solve the problem of moving cast shadow.

The background subtraction algorithms are the popular techniques to extract moving objects but have the issue of moving cast shadow [18,15]. According to the assumptions that the chromaticity of background is similar to that of shadow, and the illumination of shadow is darker than the background, various color spaces are used to analyze the properties of luminance and chromaticity, such as HSV, HSI, YUV, C1C2C3, and RGB. To remove the cast shadow, the statistical nonparametric is exploited to represent the RGB-color ratios (SNP2) [19]. Moreover, the RGB color space is transferred into HSV and that is used to evaluate the properties of the illumination and chromaticity (SNP1) [20–22](DNM) [23] (ASE) [24] (NTM) [25]. However, the colour space transformation is a time-consuming process. Therefore, some researchers use the YUV-values to substitute the RGB color space. The YUV-values can be directly provided from the image sequence without colour space transformation and those are linearly reduced in the shadows [26]. In addition, to further analyze the spectral characteristics of shadows, the combinations of various color space and local color constancy based on region are considered, such as the integration of HSV and C1C2C3 (CCM) [27] and local color constancy of luminance with region (LCC) [28]. The common drawback of aforementioned methodologies is the decision scheme of shadow which is sensitive to the threshold.

In the outdoor environments, two major illumination sources, ambient light (blue light) and the sun (white light), are useful to discriminate moving objects and their shadows [29,26]. Those properties are called physical properties in the issue of moving cast shadow detection. Although the moving objects block the illumination source of the sun, the ambient light is always irradiating on the region. When the ratio of ambient light is increasing, the color distribution of shadow tends to be blue color. The physical property is the ratio between blue and brightness values, its accuracy is higher than the other color features in outdoors and has similar performance in indoors, because the light source differs in indoors and outdoors. However, when the foreground has the similar color with the shadow, the physical property is still misclassified, because the color information is adopted.

The orientation, size and shape of the specific objects, such as vehicles and standing people, are considered as the geometry properties in the moving cast shadow detection (GBM) [30]. The geometry can overcome the problem of the similarity of the colors between foreground and background because it is not color based. However, it has some limitations, for example, the object and shadow must have different orientations and only accept the unitary light source. The geometrical property of shadow is exploited to segment the boundary between self-shadow and cast shadow [31].

Although the region of shadow is darker than the background, the texture information still exists (LRT) [32]. There are two common texture properties, the difference of gradient and the difference of angle between a pixel and its neighboring. The texture properties of shadow are similar with the background and those are different from the foreground. Texture properties perform well, because it does not rely on color information and does not have the limitations as described in geometry properties. However, it is misclassified when texture properties are similar between foreground and background, for example, the smooth object on the floor. The assumption of the similarity of the textural information between shadow and background regions is used to detect shadows (SRT) [33].

However, using a single property can only partly solve the problem of shadow. Therefore, some researchers integrated various properties to improve the performance of shadow detection and removal. The potential shadow regions can be detected by using the RGB color model and refine the potential shadow regions by considering physical and