Cast Shadow Removing in Foreground Segmentation

A.Branca, G.Attolico, A.Distante
Istituto per lo Studio dei Sistemi Intelligenti per l’Automazione - CNR
Via Amendola 166/5, 70126 BARI - ITALY

Abstract

In this paper, we focus on the problem of foreground segmentation in outdoor environment with a static TV camera. Our application context is the visual surveillance of archaeological sites. In this context the main aim is to detect the presence of people and to recognize their gestures in order to individuate the illegal actions. In this paper we concentrate solely on the primary step of moving object detection. In particular, the system should be able of recovering the true shape of the moving objects in order to allow to a classifier to discriminate a people from any other moving object as car and animals. Moreover, the system should not be sensitive to changes in lighting, weather, number of people, etc., and it is required to work autonomously for long periods of time. A main problem in analyzing real outdoor daylight scenes is to deal with shadows cast by moving objects such as vehicles or pedestrians. Cast shadows are often detected as a part of the moving objects since they move in the same way. When the detected objects contain shadows, large errors may occur with respect their recognition.

In this paper a new approach for cast shadow removing is proposed. Our idea is to detect shadow points as points that are static for a short temporal sequence and that are characterized by a photometric gain, with respect the reference background image, that is lower than unit and that we estimate for each new image through an optimization approach.

1. Introduction

In this paper we address the foreground detection problem in the context of visual surveillance in outdoor environment. Surveillance applications are mainly carried out in uncontrolled environments. Therefore the foreground segmentation relies mostly on motion data, since these are less dependent on various assumptions such as known subject. Detecting the moving parts of the observed scene, without characterizing the kind is the main goal in this context. Foreground segmentation is mainly based on temporal information. The use of temporal data is mostly based on the assumption of a static background (and camera). In this case the differences between images from a sequence must originate from the movements of the subject. Two sub-classes may be introduced: subtraction and flow. Optical Flow [3, 4] is used to detect independently moving objects in the presence of camera motion; however, most optical flow computation methods are computationally complex, and cannot be applied to full-frame video streams in real-time without specialized hardware. Subtraction is widely used by simply subtracting the current image in a pixel by pixel fashion from the immediately precedent or successive frame (temporal difference) or from a reference image of the same scene without any movement (background subtraction) which is update during the processing [6, 8]. Generally, the background subtraction methods are implemented with statistical approaches. A sequence of background images of the scene is recorded and the mean and variance intensity or color of each pixel are calculated over time. In the current image each pixel is compared to the statistics of the background image and classified as belonging to the background or not. Temporal differencing is adaptive to dynamic environments, but generally it does a poor job of extracting all relevant feature pixels, while background subtraction [5, 6, 7, 9] provides the most complete feature data, but it is extremely sensitive to dynamic scene changes due to lighting and extraneous events. Finally, in order to solve both the problems, some authors have proposed to combine Temporal differencing and Background subtraction [1, 8, 10].

In our work we consider the motion information estimated between successive temporal images to isolate all moving objects, in addition a comparison with a background image without any foreground object is applied in order to extract the correct object contours (see Fig.(1)). A binary target representation is recovered for each moving object. Background is update iteratively for each new image by considering the incremental approach proposed in [16], that allows the adaptation of the whole background model to the current scene lighting apart from the presence and the motion velocity of foreground objects, or background motion.
In addition to being able to discriminate between background and foreground it is also necessary to detect shadows. Shadows occur when objects partially or completely occlude direct light from a light source. There are two parts in a shadow: the self-shadow and the cast shadow. The self-shadow is the part of object which is not illuminated by direct light. The cast shadow is the region projected by the object in the direction of direct light. Our goal is to detect the cast shadow from the object. We can interpret shadows in the image, and the effect they have on the pixels in the scene, as a semi-transparent region in which the scene reflectance undergoes a local attenuation. The problem of shadow detection is difficult to solve because similar photometric characteristics may also be exhibited by actual objects in the scene. Although numerous shadow detection methods have been proposed, they all suffer from certain limitations that make them ineffective in real outdoor environment. Under the constraint that the imaging sensor is not undergoing motion, numerous shadow detection methods have been proposed [11, 13, 12, 14] that identify shadow regions by analyzing their photometric properties, i.e. they have a constant photometric gain with respect the background image less than unit. The main drawback of most of previous approaches is to impose photometric constraints locally on individual points, then obtaining a lot of sparse points located on the shadow but also on the objects, and in addition they are based on local a priori thresholding. Finding a general threshold to be used on all different light conditions is a difficult task.

1.1 Method Overview

The main aim of this work is to remove cast shadows generated by moving objects for improving their classification. While an object is moving its shadow follows it, but only its edges appear to move frame by frame, while the inside shadow points, due to the transparency characteristic of the shadow, remain stationary until they become edge points of the shadow. In fact, a shadow is a background region in which the scene reflectance undergoes a local attenuation. In this paper, we propose an effective shadow detection method based on the basic concept of transparency. Our idea is to detect shadow points as points that are static for a short temporal sequence and that are characterized by a photometric gain with respect the reference background image lower than unit.

This imposes to compare each current image with the immediately precedent image for static point detection (section 2), and with the reference background image for both the photometric gain computation and shadow removing (section 3), and for the actual foreground moving point detection (section 4).

2. Temporal Image Analysis

The comparison of two successive images $I_{t-1}$ and $I_t$, acquired at two consecutive time instants, for static point detection is not performed by classical frame difference, as performed by most of previous approaches proposed in literature. Frame difference approach imposes to solve the difficult problem of setting a threshold and has the disadvantage of measuring only the point to point intensity similarity, without considering the neighbouring points. On the contrary, we propose to estimate the similarity between two points $p_i$ and $q_i$ by means of their small neighborhood similarity determined by the radiometric similarity (eq.1):

$$R(p_i, q_i) = \frac{m[W_1(p_i)W_2(q_i)] - m[W_1(p_i)]m[W_2(q_i)]}{\sqrt{v[W_1(p_i)]v[W_2(q_i)]}}$$  \hspace{1cm} (1)

where $m$ and $v$ represent respectively the mean and variance functions estimated into small windows $(W_1, W_2)$, of size $5 \times 5$ pixels, centered respectively on the points $(p_i, q_i)$ on the two successive images to be compared.

The estimated similarity can take values in the range between 0 and 1, we consider similar (i.e. static) two points if their radiometric similarity is greater than 0.9.

3. Shadow Removing

Shadows are represented by stationary points $\{s_i\}$ in the temporal sequence $(I_{t-1}, I_t)$ with a local constant intensity attenuation with respect the reference background image. The shadow points $\{s_i\}$ are stationary points in the temporal sequence $(I_{t-1}, I_t)$ that, with respect the corresponding background points $\{b_i\}$, differ by a constant factor $\Lambda$ due to the change of luminosity (photometric gain), i.e.

$$\Lambda = \lambda_i = s_i/b_i$$ \hspace{1cm} (2)
on a shadow region all points will have the same gain $\Lambda$ with respect the corresponding background points.

Among all temporal static points recovered, as above described, by analyzing a pair of successive frames, the candidate shadow points will be those with a photometric gain less of 0.9. Actually also some other no shadow static points could satisfy this constraint, for example the points of a foreground object that are not moving between the two considered temporal successive frames. Our aim is to search for all neighbouring points mutually compatible according to the photometric gain similarity. We obtain this by applying an optimization approach that will select all points with photometric gain with low variance.

We impose a global constraint including the properties of photometric gain constancy with respect the background among all candidate neighbouring shadow points. In particular, we consider a binary constraint imposing to any pair of neighbouring candidate shadow points $s_i$ and $s_j$ to have the same photometric gain with respect the corresponding points on the background, that is:

$$\lambda_i = \lambda_j$$

shortly, we can define the following binary constraint function:

$$C(i,j) = e^{-\tau(\lambda_i - \lambda_j)}$$

A shadow region will be represented by the largest set of mutually compatible points according to the imposed constraint $C$. An iterative relaxation labeling approach [15] has been used for determining all image points mutually compatible according to the imposed constraint $C$.

The optimal photometric gain $\Lambda$ is estimated as the mean value among all the photometric gain $\{\lambda_i\}$ selected by the relaxation process. $\Lambda$ should correspond to the optimal gain exhibited by all cast shadow points.

The goodness of $\Lambda$ is measured by its variance $\sigma$, i.e. its mean distance from each selected $\lambda_i$. The variance should be very low because the photometric gain should be a constant factor.

When a large number of object points are static and are considered as candidate shadow points, the final variance measure is influenced to increase to highest values. In this case, the last photometric gain successively estimated in the temporal sequence is considered.

The optimal estimated photometric gain $\Lambda$ is used for detecting, in a small neighborhood of each point selected by the relaxation process, all points whose photometric gain $\lambda_i$ differs, in absolute value, from $\Lambda$ no more than two times the estimated variance. This additional step is necessary for recovering all moving contour shadow points that have not been selected by the static point detection step as candidate shadow points.

4. Background Subtraction

The shadow points detected in the current image $I_i$ are replaced by the corresponding points in the reference background image $B_k$. The resulting image $I'_i$ will now differ from the background only by the actual foreground points. We compare the two images $I'_i$ and $B_k$ for searching all foreground objects. The same radiometric similarity (eq.1), above used for detecting the static points between two temporal frames, is used for detecting all moving points of the current image $I'_i$ with respect the reference image $B_k$. We select the actual points moving with respect the background among those moving also in the temporal sequence (previous selected by the temporal image analysis process). For each moving point so selected we search in a small neighborhood for additional points with a photometric gain less of 0.9 or greater of 1.1. In this way the moving points characterized by low radiometric similarity are incremented by all neighbouring foreground points that for some reason do not satisfy this constraint, but that can be considered as foreground points for their photometric gain different from the unit.

5. Experimental Results

The method has been tested on real image sequences acquired with a static camera on an archeological site by simulating the true movements performed by intruders. In fig.2-3 we report two results obtained by applying the above technique. In particular it is highlighted as the method is able of cleaning an image of the cast shadow to allow the recovering of the true shape of the foreground objects.

References


1051-4651/02 $17.00 (c) 2002 IEEE
Figure 2. Foreground segmentation of a human figure

Figure 3. Foreground segmentation of two human figures


