UPDATING OF THE REFERENCE IMAGE FOR VISUAL SURVEILLANCE SYSTEMS (IDSS 2003)

C J M TORNIERI F BREMOND M THONNAT

France, Institut National de Recherches en Informatique et Automatique

ABSTRACT

This paper presents a new algorithm for updating the reference image used in a Visual Surveillance platform. Updating the reference image is a crutial step of the motion detection. Most of the classical methods consist in slow integration of the current image into the reference image. We propose to model the integration process to discriminate parts of the image that correspond to illumination and environment changes from those that correspond to individuals evolving in the scene. We conclude this article by showing some results on subway video sequences.

1. INTRODUCTION

In this paper, we present a new algorithm for updating the reference image used in a Visual Surveillance platform (2). This work has been done in the framework of the european project ADVISOR dedicated to visual surveillance in subway stations using fixed monocular cameras. Motion detection is one of the main steps of the video analysis. This step is done by subtracting the current image from a reference image corresponding to the background image. In order to be robust, the platform must update at each image the reference image to take into account the illumination changes and also changes of the environment such as a moving door. Classical methods (1), (5) to update the reference image integrate gradually a portion of the current image into the reference image. These methods are well adapted for slow illumination changes. However they do not handle sudden illumination changes nor environment changes. We propose here a new algorithm that consists in the modelisation of the integration process to discriminate parts of the image that correspond to illumination and environment changes (opening a cupboard) from those that correspond to individuals. First we describe the Visual Surveillance platform and particulary the detection module. Then we explain how to build and update the reference image using moving and stationary regions. Finally we give some results of our algorithm applied on subway video sequences.

2. DETECTION OF MOVING REGIONS

2.1 Overview of the Visual Surveillance Platform

We use a Visual surveillance platform described in F. Cupillard et al (2) is composed of three modules in order to go up to a high level interpretation of video sequences. In a first stage, images are grabbed from a capture card. Then they are sent to a detection and frame to frame tracking module that is in charge of detecting moving regions. Finally in the last stage, all the tracked regions are used by an interpretation module to recognise high level information like person behaviours or pre−defined scenarios. In this paper, we are more interested in the detection and tracking module and especially how to build and update the reference image.

2.2 Overview of the Motion Detector

The detection module is composed of six steps. The segmentation step subtracts the current image with the reference image. The reference image corresponds to the empty scene and is also called the background image. Then the difference image is thresholded to obtain a binary image where white pixels represent moving pixels and black pixels represent non−moving pixels. The moving pixels are then grouped into connected regions. These regions are called blobs.

The next step consists in tracking the blobs from two successive images. A tracked blob is called a moving region and can be linked to zero, one or several blobs detected in the previous image. Usually a moving region corresponds to an individual evolving in the scene.

Another step consists in computing the stationary regions. A stationary region is a part of the current image that does not appear in the reference image and that is always detected at the same location in the current image. A stationary region usually corresponds to a change in the environment or a sudden illumination that lasts in the current image.

The last step consists in updating the reference image by integrating stationary regions into the reference image without integrating the individuals.

2.3 Shadow Removal

A first enhancement to compute the reference image is to better detect motion and in particular to be able to discriminate shadows from real individuals. To avoid detecting shadows during segmentation step, we use two shadow detection methods. The first one assumes that shadows only introduce intensity variation while preserving colour information as explained in A. Prati et al (3). These assumptions are particulary well defined in the HSV colour space. To prevent high processing time, only pixels detected in movement in the RGB colour space are considered as potential shadow pixels. The second method is based on texture variation. It assumes that shadows preserve texture information of overlapped surfaces. Combining these two methods, most of the shadows are successfully removed.

3. INITIALISATION OF THE REFERENCE IMAGE

When the platform is started, to initialise the reference image, it is necessary to have a background image that matches with the current scene. A first solution is to choose the first grabbed image as the background image. The drawback of this solution is to integrate into the reference image any people present at the beginning in the scene. Another solution is to choose a better background image among many background images from a library that correspond to different illumination conditions.

3.1 Selection of the best Reference Image

Our platform is designed to work in predefined environments (subway stations, bank agencies). For each camera it is possible to store offline several background images grabbed with different scene illumination conditions before using the platform. During the initialisation stage, the platform chooses among these images the one which best matches the current scene. The choice is based on several statistical variables: mean intensity on the whole image, standard deviation of intensity, mean colour value on the UV channels and the standard deviation of the colour values. First the platform grabs N current images (for example N=20) to build a mean image corresponding to the scene. Then the statistical variables are computed on this mean image and compared with the statistical variables of the stored images. The image with the nearest statistical variables is selected as the reference image as shown in figure 1.

This initialisation of the reference image has two drawbacks. First the quality of the reference image depends on the number of stored background images and on their proximity with the current scene illumination conditions. Second, the selection of the reference image is a global approximation process. For example a part of the image containing a different scene object (an opened cupboard) or a different illumination (a new source of light) constitutes a bad approximation of the reference image.

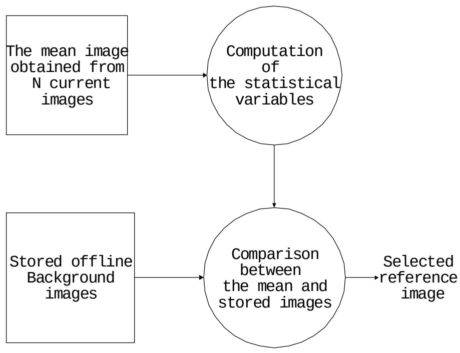

Figure 1. Overview of the algorithm to initialise the reference image

3.2 Enhancement of the Reference Image

To improve the algorithm, we combine the selected background image with a mean image obtained after accumulating N current images (for example N=100) as shown in figure 2. This mean image corresponds to the reference image wherever nobody was detected. When someone is detected among the N current images, the mean image is a mixed of this person and of the background. So the mean image is an approximation of the reference image. There are two criteria to distinguish a person from the background : 1) even if a person is motionless, he or she moves enough to produce some intensity variations in the image, 2) the colour of a person is usually different from the colour of the background. On the basis of these criteria, we try to identify which pixels belong to the reference image. A pixel is considered as a moving pixel (foreground) if its intensity or colour standard deviations are large enough. Thus, the intensity or colour standard deviations are computed for each pixel of the mean image. If these standard deviations are under a threshold, the corresponding pixel of the initial reference image is replaced by the pixel of the mean image else it remains unchanged.

So thanks to this method, if there is a part of the image with no moving objects, the corresponding parts of the reference image corresponds exactly to the empty scene. When there is a moving object, we select a part of a background image previously stored.

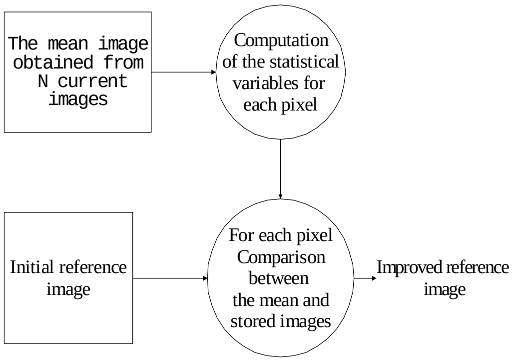

Figure 2. Overview of the algorithm to improve the reference image

4. UPDATE OF THE REFERENCE IMAGE

The update of the reference image as proposed by A. Elgammal and al (1) and T. Horprasert (5) must take into account environment changes by integrating noise while preventing people integration. The main idea of our algorithm is to compute moving regions corresponding to individuals evolving in the scene and to compute stationary regions corresponding to noise and to integrate the stationary regions into the reference image without integrating the individuals.

4.1 Slow Illumination changes

To take into account slow illumination changes, we continuously update the reference image with the current image as explained in A. François and G. Medioni (4). Each pixel of the background image is slightly blended (a small integration coefficient such as 0.01) with its corresponding pixel of the current image. This method has two major issues. On one hand, motionless persons should not be integrated in the reference image (null integration coefficient). On the other hand, sudden illumination changes or noise require a direct integration in the reference image (a high integration coefficient).

4.2 Moving Regions

A moving region is a tracked blob from two consecutive images. It corresponds to an individual evolving in the scene and should not be integrated in the reference image. This section explain how the moving regions are computed.

- Creation and Update

Moving regions are characterised by their position in the image, their rectangular bounding box, the total trajectory length since their creation, their life time and their type.

The tracking process is just to establish if whether the blobs are moving. It only needs two lists of moving regions. The first one, *Lp* represents previous moving regions and the second one, *Lc* current moving regions.

The update of moving regions consists of selecting for each blob of the current image the previous moving region that gives the best overlapping rate. Let's A*i* represents the intersection area between a moving region *m(t−1)* and a blob *b(t)* and A*b* represents the blob area. The overlapping rate is defined as : $overlapping = \frac{Ai}{Ab}$. Once the new detected blob is linked to the previous moving region, it is added into the list *Lc*. If a blob has no intersection with any previous moving region then a new moving region without link is inserted into the list *Lc*.

- Types of Moving Regions

The type of a moving region is determined by the total trajectory length and its life time. There are four different types:

- Newly detected: The life time of the moving region is short
- Noise: The life time of the moving region is large and its trajectory length is short
- Person: The life time of the moving region is large and its trajectory length is long
- Unknown: Ambiguous tracking situation. We do not know for sure which moving region is the father of this one. Its life time and trajectory length cannot be conputed.

- Unknown moving regions

Several blobs newly detected may be linked with the same previous moving region. This situation occurs when the previous moving region splits into the newly detected blobs. In this case, these newly detected blobs become new moving regions (called splitted regions) with the same father. While these splitted regions overlap the bounding box of the previous moving region, they are ambiguous in the sense that we do not know if they correspond to a noise or to the previous moving region. When one of the splitted region stop overlapping the bounding box, we consider that this splitted region corresponds to the previous moving region that moves away and we update its life time. All the other splitted regions remain ambiguous and are considered as new moving regions.

4.3 Stationary Regions

A stationary region is a part of the current image that does not appear in the reference image and that is always detected at the same location in the current image. A stationary region usually corresponds to a noise such as a sudden illumination change that lasts in the current image. It is represented by a rectangular zone in the image to which we associate a template. The template gives the precise shape of the pixels constituting the stationary zone.

A stationary region S(t) is created as shown in figure 3 with any blob M1(t) detected at time t that is supposed to be stationary. If the blob M1(t) corresponds to a moving region with the type person then M1(t) is not considered as stationary so no stationary region is created. The stationary region S(t) is updated at time t+1 with the set of overlapping blobs M2(t+1): the bounding box of the stationary region is replaced by the bounding box of the intersection of all blobs M2(t+1) with the stationary region S(t) (see figure 3). At each step the template corresponding to the stationary region is updated.

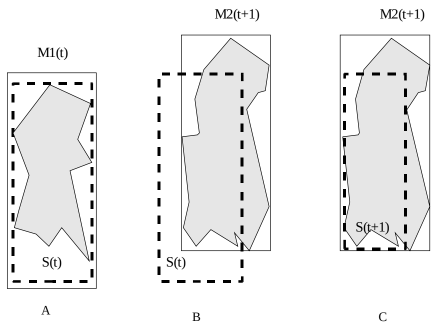

Figure 3. Construction at t and update at t+1 of the stationary region S.

A) Creation of the stationary region S(t) from a blob M1(t) at time t.

B) Detection of a blob M2(t+1) overlapping S(t).

C) Update of the stationary region S(t+1).

4.4 Integrating Regions in the Reference Image

Depending on the properties of a stationary region, we have to decide whether the region should be integrated in the reference image. There are three cases:

1. Noise: The life time of the stationary region is large (time > 40). No moving region overlaps the stationary region except moving region of type noise. Changes in the template are not significant. A high blending coefficient (close to 1.0) is used to integrate the region into the reference image.
2. Person: The life time of the stationary region is short (time < 40). A moving region of type person is on the stationary region. A null blending coefficient is used.
3. Ambiguous: The life time of the stationary region is large (time > 100). One or more unknown moving regions overlap the stationary region. A low blending coefficient (around 0.05) is used to allow more time to take a decision.

4.5 Handling Movable Scene Objects

Some scene objects such as doors or chairs may move and be wrongly considered as individuals. Although these objects are not individuals, they are moving and have the size of a person, so classical algorithm cannot manage such situations and integrate the scene objects in the reference image. We propose here to use a colour distribution associated to each scene object in order to be able to differentiate them from individuals and from the background.

Thus we store for each interesting scene object (doors for instance) several colour histograms that represent the different states of the object (closed door for the foreground, opened door for the background) and we compute at each image the histograms corresponding to all individuals close to the scene object. Then to determine if a moving pixel is belonging to a scene object, to the background or to an individual we compare the probability associated to the colour of the moving pixel with the colour histogram of the scene object, of the background and of the individuals. The best probability determines the owner of the pixel. If the pixel belongs to a scene object then it is integrated in the reference image and the scene object is not anymore detected as a moving region.

5. RESULTS

We have tested our new algorithm on different video sequences: six videos of subway station, one video of an outdoor railway (1 hour), four videos of a bank office (up to two hours) and live camera in a office. Results are very promising. Even if there are a lot of people, our algorithm manages to include only illumination and environment changes. As you can see in figure 5, a new object is present at the beginning of the video sequence while it is not on the initial reference image shown in figure 4. In figure 7, the new scene object has been correctly integrated in the reference image whereas nobody has been integrated.

6. CONCLUSION

In this paper we have presented a new algorithm that build and update the reference image used to segment moving regions in a video sequence. This algorithm is able to manage slow and sudden illumination changes and also environement changes. It also prevents integration of tracked individuals. The building stage uses offline stored background images. The update stage combines moving regions and stationary regions to detect illumination and environment changes. We have tested it on many video sequences and obtained very promising results even in difficult situations. However there are still incorrect integrations of individuals who do not moved during the whole sequence. The next step in our work will be to use feedback from the interpretation module to help to decide what to do when a stationary region is detected.

Figure 4. The initial reference image initialised from previously stored background images.

Figure 5. First image of a subway video sequence. A new scene object is present in this image.

Figure 6. The new scene object is detected as a stationary region represented by a blue bounding box.

Figure 7. Fifty images later: the new scene object has been integrated into the reference image.

(1) Ahmed Elgammal; David Harwood; Larry Davis, ’’Non−parametric Model for Background Subtraction’’, ECCV 2000

(2) Cupillard Fréderic; Bremond François; Thonnat Monique, ’’Tracking Groups Of People For Video Surveillance’’, 2nd European workshop on AVBS Systems 2001

(3) Andrea Prati; Ivana Mikic; Costantino Grana; Mohan M. Trivedi, ’’Shadow Detection Algorithms for Traffic Flow Analysis: a Comparative Study’’, ITSC 2001

(4) Alexandre R.J. François and Gérard G. Medioni, ’’Adaptive Color Background Modeling for Real−Time Segmentation of Video Streams’’, CISST 1999

(5) Thanarat Horprasert, David Harwood, Larry S. Davis, ’’A Robust Background Subtraction and Shadow Detection’’, Proc. ACCV’2000, Taipie, Taiwan, January 2000

Article published in Intelligent Distributed Surveillance Systems (IDSS) February 2003 IEE, Savoy Place, London, UK