# REAL-TIME CHANGE DETECTION METHODS FOR VIDEO-SURVEILLANCE SYSTEMS WITH MOBILE CAMERA

*Franco Oberti, Lucio Marcenaro and Carlo S. Regazzoni*
University of Genoa, Department of Biophysical and Electronic Engineering (DIBE)
Signal Processing and Telecommunications Group (SP&T)
Via Opera Pia 11/A, I-16145, Genoa (Italy)
e-mail: oberti@dibe.unige.it

## ABSTRACT

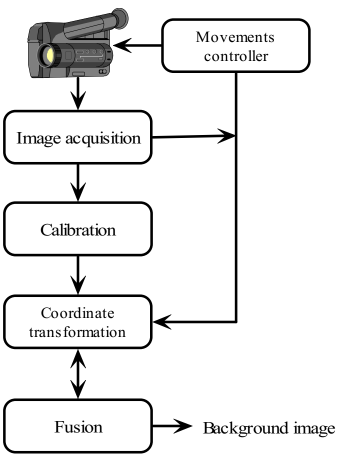
In this paper, a video-surveillance system based on a mobile-camera is presented. The proposed real-time method is able to detect an object in a video sequence from a non-static camera. In particular during an off-line phase the system creates a panoramic multi-layer background image. In the on-line phase the system compares the acquired images with a portion of the panoramic background. Different change detection methods are analyzed. Experimental results are presented in order to validate the proposed methods. The proposed method could be used for extending efficient algorithms for scene understanding already developed and tested for fixed cameras to a mobile camera environment.

## 1. INTRODUCTION

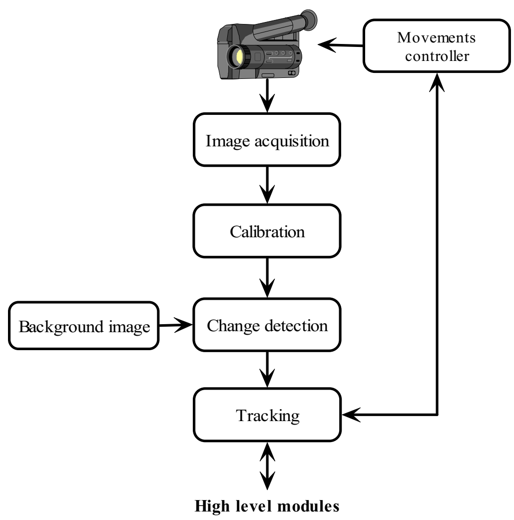
Due to the increasing demand of complex vision systems based on scene interpretation, detection and tracking of moving objects in wider areas is becoming a more and more important task. Different video-surveillance systems have been developed based on video processing and understanding techniques [1, 2]. These systems are able to recognize and classify potentially dangerous situation and to generate some kind of alarm to capture the attention of a human operator.
Most of the existing automatic video-surveillance systems are based on a given static camera configuration . In order to guard wider areas, multiple camera environments have been used [3]. This solution can be is too expensive. The problem could be solved by using a mobile camera with pan and tilt movements. Algorithms based in motion estimation and optical flow to detect and track the objects in the scene [4] could be used in this configuration. However this solution is computationally heavy and does not allow a real-time which is necessary for a video-surveillance system.
This paper shows a method for extending efficient algorithms already developed and tested for fixed cameras to a mobile camera environment. Since existing video-surveillance systems have to satisfy real-time constraints, change detection algorithms with low computational requirements need to be implemented even for systems using mobile-head cameras.
The proposed approach is based on two phases. The first one is an off-line step that does not have real-time requirements and is devoted to the generation of a panoramic multi-layer background image. The second one is aimed to generate a binary change detection image which identifies foreground regions in the guarded scene.
The paper is structured as follows: Section 2 describes the system architecture; in particular different change detection criterions are studied. Section 3 presents the results obtained and finally conclusions are drawn in Section 4.

## 2. SYSTEM DESCRIPTION

The presented approach is designed for the creation of a change detection algorithm for non-static cameras allowing the use of this method with traditional static video-surveillance modules.
The adopted mobile camera has integrated pan/tilt movements that can be controlled from a PC with a set of instructions [5]. The camera is calibrated by modeling the extrinsic parameters of the sensor by using the technique described in [6].




(a)




(b)

**Figure 1.** *System description divided in two phases: a) off-line: generation of the multi-layer background, b) on-line: the video-surveillance system with mobile camera.*

The functioning of the proposed system can be divided in two subsequent stages. During the first off-line phase (fig. 1a) the camera acquires a certain number of images of the empty scene with different position of the camera's head. A set of images is acquired such that they cover the whole panoramic of the scene: the result of the processing of the acquired images is a global background image used by change detection module.
In the on-line phase (fig. 1b) the system processes the images acquired by the camera using the global image generated during the off-line stage as background. During this phase the system uses a link with the sensor for controlling the movement of the camera.
In order to provide a reference image of the background to the change detection module, a panoramic background model of the whole empty monitored scene is constructed. To obtain the panoramic image, a set of images are acquired with predefined positions of the camera. Due to superposition between the images, in some regions of the image it is possible to use the information from up to four different images for optimising the change detection performances. In order to memorize all the available information a multi-layer global background is generated. Figure 2 illustrates the multi-layer background in which the gray scale indicates the level of overlapping for each region of the panoramic view.

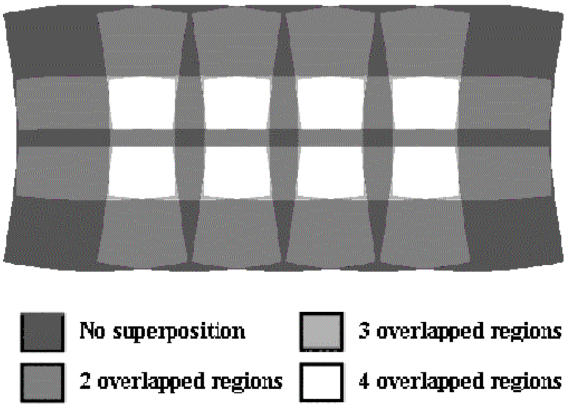



**Figure 2**: *overlapping regions in multi-layers background.*

The construction of the panoramic image derives from the union of a set of images, previously transformed in order to obtain a perfect alignment. Because of the movements allowed by the mobile camera a spherical projection has been used. The linearity of the correspondence between coordinates of the image and the coordinates of the camera movements is achieved by projecting images planes on the virtual sphere. This can be obtained by using a roto-translation matrix $T$ that depends by pan $(\vartheta)$ and tilt $(\varphi)$ angles and focal distance $F$.
If $(x, y)$ are the coordinates of a point in an acquired image, and $\vartheta$ and $\varphi$ are the pan and tilt angles used for acquiring the image and $F$ is the focus of the camera sensor, one can obtain the corresponding coordinates in the panoramic background $(x_s, y_s)$ by using the following transformation:

$$\begin{bmatrix} x_s \\ y_s \end{bmatrix} = T(\vartheta, \varphi, F) \cdot \begin{bmatrix} x \\ y \end{bmatrix} \qquad (1)$$

The matrix $T$ depends on three parameters $\vartheta$, $\varphi$ and $F$, corresponding to pan and tilt rotations and translations.

Using equation (1) it is possible to construct the multi-layer background. In the considered system, 15 images were used for generating the panoramic background. Figure 3 shows the result of the generation of the panoramic background. Since the background is a multi-layer structure, the mean value for each pixel was selected for visualization purposes.

**Figure 3**: *the panoramic background.*

**2.2 On-line phase**

During the on-line phase the system acquires a frame from the camera and using the information from the multi-layer panoramic background and the camera position performs a change detection step to find the objects present in the scene.

In order to obtain a real-time behavior a simple difference and thresholding algorithm has been selected. By using equation (1) each pixel of the acquired image $v_{current}(x, y)$ is projected into the reference system $(x_s, y_s)$ of the panoramic background. The difference is computed between the value of each point in the current image $v_{current}(x_s, y_s)$ and the correspondent value in the panoramic background image $v_{background}(x_s, y_s)$.

$$\forall(x_s, y_s) \to v_{diff}(x_s, y_s) = \left| v_{current}(x_s, y_s) - v_{background}(x_s, y_s) \right| \qquad (2)$$

The difference $v_{diff}(x_s, y_s)$ is then thresholded to generate the change detection image with a threshold $\eta$

Since the multi-layer panoramic background may contain more than one value for the coordinates $(x_s, y_s)$, $v_{background}(x_s, y_s)$ can be selected according to different criterions. In the following, $v_{background}(x_s, y_s)_i$ is the value of the pixel with coordinates $(x_s, y_s)$ in the *i*-th overlapped background image; for each considered pixel, $i \in [1, \ldots, N]$ where $N \in [1, \ldots, 4]$ is the maximum number of overlapped layers for pixels with maximum superposition of background layers. Thus, $N$ different hypothesis $v_{diff}(x_s, y_s)_i$ can be obtained by substituting $v_{background}(x_s, y_s)_i$ in equation (2).
For a real system, the misalignment between the acquired image and the panoramic background has to be taken into account: this is caused by the limited precision in the mobile camera positioning that can produce unwanted noise in the change detection image. Thus, the acquired image projected in the spherical reference system can be expressed by the following:

$$v_{current}(x', y') = v_{current}(x_s + \xi_{x_i}, y_s + \xi_{y_i}) \qquad (3)$$

where the terms $\xi_{x_i}$ and $\xi_{y_i}$ take into account the misalignment along $x_s$ and $y_s$ axes.

An illumination variation between the panoramic background and the acquired image should be also considered; in this case, an illumination term $\Delta O_i = (O - O_i)$ expressing the global term of the intensity variation between each reference image in the multi-layer background and the acquired image can be included.
By expanding eq. (3) using the Taylor series truncated to the second order and neglecting the noise introduced by the acquisition process, the difference $v_{diff}(x_s, y_s)_i$ can be written as:

$$v_{diff}(x_s, y_s)_i = \left| \begin{array}{l} v_{current}(x_s, y_s) - v_{background}(x_s, y_s)_i + \\ \frac{\partial v_{current}}{\partial x_s}\xi_{x_i} + \frac{\partial v_{current}}{\partial y_s}\xi_{y_i} + \Delta O_i \end{array} \right| \quad (4)$$

Different fusion strategies can be used for integrating $v_{diff\,i}$ hypotheses; in the following, three different change detection criterions will be evaluated: Mean criterion, Minimum difference criterion and Minimum distance criterion.

*2.2.1. Mean criterion*

The mean method calculates a mean background pixel value from the multi-layer background: each pixel value is given by:

$$v_{diff}(x_s, y_s) = \frac{1}{N}\sum_{i=1}^{N} v_{diff}(x_s, y_s)_i =$$

$$\left| \begin{array}{l} v_{current}(x_s, y_s) - \frac{1}{N}\sum_{i=1}^{N} v_{background}(x_s, y_s)_i + \\ \frac{1}{N}\sum_{i=1}^{N}\left( \frac{\partial v_{current}}{\partial x_s}\xi_{x_i} + \frac{\partial v_{current}}{\partial y_s}\xi_{y_i} + \Delta O_i \right) \end{array} \right| \quad (5)$$

Considering that each $v_{background}(x_s, y_s)_i$ can be modeled as the ideal background value $v_{background}(x_s, y_s)$ plus a misalignment and illumination error, the mean of the backgrounds can be expressed as:

$$\frac{1}{N}\sum_{i=1}^{N} v_{background}(x_s, y_s)_i = v_{background}(x_s, y_s) + \alpha + \beta \quad (6)$$

By using this criterion, the obtained difference is equal to the ideal one plus an error proportional to the mean of the misalignment and illumination errors of the current value plus $\alpha$ and $\beta$. $\alpha$ and $\beta$ in eq. (6) are related to the misalignment and illumination errors between $v_{background}(x_s, y_s)_i$ and the ideal background value $v_{background}(x_s, y_s)$. Despite of introduced systematic errors, an advantage of this method is that a multi-layer background is not necessary. By applying a mean operation extended to the whole image, a new single layer background can be generated. As a consequence only one value is necessary for each pixel of the background image thus allocating less memory than a multi-layer background.

*2.2.2. Minimum distance criterion*

The minimum distance criterion, is intended to minimize the misalignment errors according to a simple and computationally efficient method. This is done by selecting the *i*-th background pixel which minimizes the term corresponding to misalignment errors:

$$v_{background}(x_s, y_s)_i \Rightarrow \min\left( \frac{\partial v_{current}}{\partial x_s}\xi_{x_i} + \frac{\partial v_{current}}{\partial y_s}\xi_{y_i} \right) \quad (7)$$

Using this background in eq. (4), the resulting change detection value is equal to:

$$v_{diff}(x_s, y_s) = \left| \begin{array}{l} v_{current}(x_s, y_s) - v_{background}(x_s, y_s)_i + \\ \min\left( \frac{\partial v_{current}}{\partial x_s}\xi_{x_i} + \frac{\partial v_{current}}{\partial y_s}\xi_{y_i} \right) + \Delta O_i \end{array} \right| \quad (8)$$

This criterion produces good results in case of considerable misalignment errors with small illumination variations. Since this method does not take into account the $\Delta O_i$ term, it can generate detection errors when the minimum of the alignment error corresponds to a high illumination variation.
The implementation of this method selects the background reference pixel value on the basis of its position in the image and independently from its intensity. Given a pixel value in the acquired image $v_{current}(x, y)$, the correspondent reference value $v_{background}(x_s, y_s)$ is selected between the candidates as the overlapping value $v_{background}(x_s, y_s)_i$ with the minimum spatial distance from the considered pixel.
In this approach, the camera position of each overlapping image used when creating the background should be stored. In the following, $(\vartheta, \varphi)$ and $(\vartheta_i, \varphi_i)$ correspond to the camera position of the currently acquired image and the *i*-th background image respectively. The reference value is then selected as the overlapping value $v_{background}(x_s, y_s)_i$ which minimizes the spatial distance $\sqrt{(\vartheta - \vartheta_i)^2 + (\varphi - \varphi_i)^2}$. This method is the one which requires higher allocation of memory, since it is necessary to associated a camera coordinate to each pixel used for creating the panoramic background.

*2.2.3. Minimum difference criterion*

Previous methods produce good results for certain working conditions only. It is then necessary to introduce a method taking into account both the misalignment and illumination errors. However real-time and low-memory-usage requirements need to be met. The purpose of the following method is to select the difference value minimizing misalignment and illumination errors as indicated in the following equation:

$$v_{background}(x_s, y_s)_i \Rightarrow \min\left( \frac{\partial v_{current}}{\partial x_s}\xi_{x_i} + \frac{\partial v_{current}}{\partial y_s}\xi_{y_i} + \Delta O_i \right) \quad (9)$$

Substituting eq. (9) into eq. (4) the obtained change detection value is given by:

$$v_{diff}(x_s, y_s) = \left| \begin{array}{l} v_{current}(x_s, y_s) - v_{background}(x_s, y_s)_i + \\ \min\left( \frac{\partial v_{current}}{\partial x_s}\xi_{x_i} + \frac{\partial v_{current}}{\partial y_s}\xi_{y_i} + \Delta O_i \right) \end{array} \right|$$

(10)

In order to find possible solutions, let's analyze the following two hypothesis:

- H0: $v_{current}(x_s, y_s)$ corresponds to the background.
- H1: $v_{current}(x_s, y_s)$ corresponds to a foreground object.

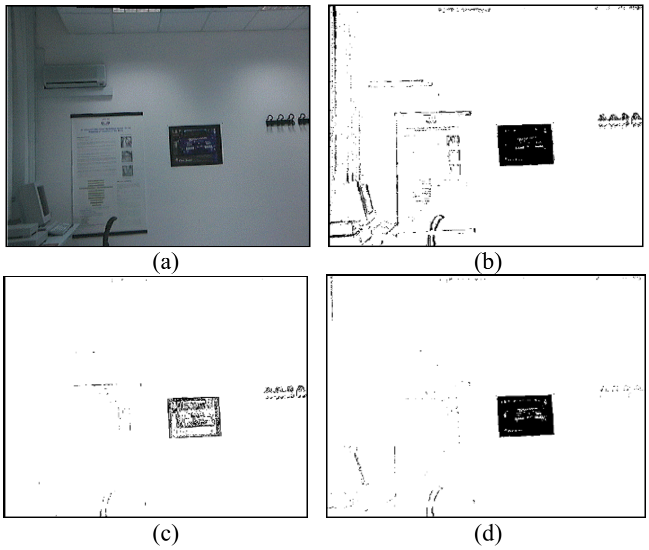



**Figure 4**: *(a) original image; change detection images: (b) mean value, (c) minimum distance, (d) minimum difference.*

| Criterion | False detection probability ($P_f$) | Correct detection probability ($P_d$) |
|---|---|---|
| **Mean** | 0.04 | 0.92 |
| **Minimum distance** | 0.04 | 0.26 |
| **Minimum difference** | 0.04 | 0.96 |

**Table 1:** *Probabilities of correct detection and false alarm for the evaluated criterions.*

1. In the ideal H0 case, $v_{current}(x_s, y_s) = v_{background}(x_s, y_s)_i$ and then the eq. (13) becomes:

$$v_{diff}(x_s, y_s) = \left| \min\left( \frac{\partial v_{current}}{\partial x_s} \xi_{x_i} + \frac{\partial v_{current}}{\partial y_s} \xi_{y_i} + \Delta O_i \right) \right| \quad (11)$$

Then the *i*-th background that minimizes the illumination changes plus the misalignment error is the one minimizing also the difference value. Using this criterion, the minimum false detection probability is guaranteed.

2. If H1 is considered, eq. (11) can be expressed as:

$$v_{diff}(x_s, y_s) = \left| \Omega_i + \frac{\partial v_{current}}{\partial x_s} \xi_{x_i} + \frac{\partial v_{current}}{\partial y_s} \xi_{y_i} + \Delta O_i \right| \quad \text{with}$$

$$|\Omega_i| = \left| v_{current}(x_s, y_s) - v_{background}(x_s, y_s)_i \right| >> 0 \quad (12)$$

In this case, choosing the minimum value of $v_{diff}(x_s, y_s)$ does not guarantee the errors minimization, since $v_{diff}(x_s, y_s)$ depends also on $\Omega_i$. Thus this decision can lead to the minimization of the $\Omega_i$ instead of $\left( \frac{\partial v_{current}}{\partial x_s} \xi_{x_i} + \frac{\partial v_{current}}{\partial y_s} \xi_{y_i} + \Delta O_i \right)$.

However, one can suppose that in the case of H1, $|\Omega_i| >> \left| \frac{\partial v_{current}}{\partial x_s} \xi_{x_i} + \frac{\partial v_{current}}{\partial y_s} \xi_{y_i} + \Delta O_i \right|$.

Then, even if the minimization of the alignment and illumination errors is not guaranteed, the hypothesis H1 results in a small increase of the probability of misdetection.

The minimum difference criterion evaluates the smaller difference between a pixel in the current image and the correspondent pixel in each overlapped background image and selects this one as the reference one.

## 3. EXPERIMENTAL RESULTS

In order to test the proposed methods a rectangular object is inserted in the scene. The shape of the object facilitates the building of ideal change detection images for comparison of performance and does not influence the behavior of the methods. The system has been implemented on a 400MHz PC using Linux operating system. The processing times for this architecture are: 15 sec. off-line phase, 0.5 sec. on-line phase. In this case the system is able to work at 2 frames per second. The pan and tilt angles used for the test were in the range (-15$^o$, 20$^o$) for PAN and (-10$^o$, 10$^o$) for TILT. In figure 4 an example of the original image and the change detection images with the three methods are showed. Table 1 presents the performances of the methods, considering a fixed value of Probability of false alarm. In particular the highest probability of detection is obtained by the minimum difference criterion. Presented results demonstrate that the minimum difference criterion provides the best performances among the presented approaches. Obtained results with this method satisfy requirements of video-surveillance systems such as high probability of correct detection, low false detection probability and real-time behavior. The resulting image of change detection can be used in existing high level image processing modules of existing video-surveillance systems.

## 4. CONCLUSIONS

A real-time change detection method for mobile cameras has been presented. The algorithm is based in two phases: during an off-line phase a panoramic multilayer background image is generated. For the on-line phase three different change detection criterions have been studied. The generated change detection image can be used as an input to the higher level modules of a standard video-surveillance system based on a static camera. Results show that among the proposed change detection methods, the minimum difference criterion gives high performances in terms of probabilities of correct and false detection, allowing the use of this method with traditional static video-surveillance modules.

## REFERENCES

[1] C.S. Regazzoni, "Recognition and Tracking of Multiple Vehicles from Complex Image Sequences", *Road Vehicle Automation II*, pp.297-306, O.Nwagboso ed., Wiley, London, 1997.

[2] P.Remagnino, A.Baumberg, T.Grove, T.Tan, D.Hogg, K.Baker and A.Worrall. "An integrated traffic and pedestrian model-based vision system", in *Proceedings of British Machine Vision Conference*, pages 380-389, 1997.

[3] V. Kettnaber, and R. Zabih, Counting people from multiple cameras, Proc.1999 IEEE Int. Conf. on Multimedia Computing and Systems, Florence (I), June 1999, Vol. 2, pp. 267-271.

[4] *M.Middendorf and H.-H.Nagel,* "Vehicle Tracking Using Adaptive Optical Flow Estimation", in First IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS2000), March 31, 2000, pp. 38-45, Grenoble, France.

[5] *Marcus Breig* "Sony EVI D30/31 Pan/Tilt Unit" University of Dortmund.

[6] R.Y.Tsai "An efficient and accurate camera calibration technique for 3D machine vision". *Proc. IEEE Comp. Soc. Conf. On CVPR*, Miami Beach, 1986.