ROBUST REAL-TIME INTRUSION DETECTION WITH FUZZY CLASSIFICATION

Giovanni Milanesi, Augusto Sarti, Stefano Tubaro

Dip. di Elettronica e Informazione
Politecnico di Milano
Piazza Leonardo Da Vinci 32
I-20133 Milano, Italy

ABSTRACT

In this paper we propose a novel system for indoor video surveillance, which is able to detect and track moving objects even in the presence of significant variations of scene illumination. After a preliminary analysis and clustering of temporal changes in the video sequence, the algorithm performs a classification based on fuzzy logic, aimed at identifying moving regions that really correspond to unexpected objects in the scene. The proposed approach tends to discard shadows, reflections and luminance profile changes due to illumination variations. One key feature of our system is its modest computation complexity, which allows it to operate in real-time on a common PC platform. The system has been tested on a wide variety of situations, proving its effectiveness and robustness.

1. INTRODUCTION

Advancement in the area of image analysis and cost reductions in CCTV equipment are currently boosting a formidable growth in video surveillance systems. In particular, intrusion detection and classification in indoor and outdoor environments represents an application of great interest.

In order to guarantee the necessary level of safety and effectiveness, a video-surveillance system should be able to correctly respond to a wide range of complex situations. Typical scene chances that do not correspond to intrusions are changes in the environmental illumination, such as natural light dimming due to clouding or sun setting; flickering fluorescent tubes; lightbulbs switching on or off; car brights flashing through the windows; etc.). Also, in order for the system to be able to correctly analyse regions of detected motion, shadows and reflexions due to moving objects should be detected and treated separately from the actual objects in motion.

In this paper we propose a novel intrusion detection system that is particularly suitable for indoor use. The system is able to detect and track moving objects that appear in the field of view of a static camera and is able to robustly distinguish between luminance profile changes due to a moving object and those due to illumination changes that can normally occur in the environment. As we can see in Fig. 1, the system consists of three cascaded blocks: the first performs change detection and low-level analysis to extract the regions of change, and is able to roughly distinguish between illumination changes and geometrical scene changes. The goal of the attention focusing block is to perform temporal tracking of the regions of interest (bounding boxes) in order to regularise them and improve the preliminary classification performed by the previous block. The third block implements the final classification phase, based on the analysis of some carefully chosen image characteristics. The aim of the classification algorithm is to distinguish between areas where geometric scene changes (real intrusions) occurred and ares where the changes are to be attributed to variations in illumination, shadows, reflections, etc.

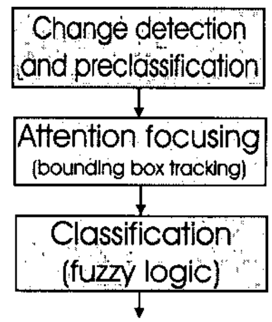

Fig. 1. Structure of the our video analysis algorithm.

This global approach is, in fact, scalable as it allows us to remove the last one or two blocks from the processing chain of Fig. 1, and end up with a reduced system that is still usable for intrusion detection purposes (of course with reduced performance). In the following three Sections, we will describe the three basic blocks of Fig. 1. Section 5 will provide more information on the global complexity of the system and present the results a series of tests conducted on real sequences.

2. CHANGE DETECTION

The first block of our video-surveillance system is aimed at an accurate frame-by-frame detection of the areas that exhibit significant changes between the current frame and previous frames (or some reference frame). In order to reduce the computational load, only changes in the luminance profile are considered. The algorithm, however, is designed in such a way to be relatively insensitive to changes in the global scene illumination.

The first step is to compute the difference between frames in order to localize those areas where significant luminance changes took place. There are several ways to do so, one very simple solution that takes into account both the differential change $(F_c(x,y) - F_p(x,y))$ between current and previous frames, and the absolute change $(F_c(x,y) - F_b(x,y))$ between the current frame and a reference one, which consists of a pixel-by-pixel computation of

$$\max\left[(F_c - F_p), (F_c - F_b)\right] \quad . \tag{1}$$

Here the reference frame (background) is a reasonably recent frame acquired knowing that there was no motion in the scene. If (1) exceeds a threshold T_h, then the corresponding pixel is labeled as "change point". At the end of this process we have a boolean mask $M_c(x,y)$ that specifies the presence of local changes. The threshold T_h is dynamically computed taking into account the local average and standard deviation of the samples extracted with the mask $M_c(x,y)$, in accordance with what proposed by Hamadami [3]. Pixels of the background frame F_b are not updated all in the same way. In fact, the update is faster for the pixels below threshold and slower for the others. This way it is still possible to keep track of small luminance changes that occur between frames, while structural scene changes (large objects that move) will not be treated like change areas for long.

The information contained in the change mask $M_c(x,y)$ is then improved through morphological closing, and the areas of interest (connected regions of change) are enclosed in bounding boxes. The result will be a set of partially overlapping rectangles which are finally fused together into a smaller number of larger non-overlapping boxes.

At this point we can limit our analysis to the detected bounding boxes in order to reduce the computational cost, and apply a robust algorithm that exhibits little sensitivity to luminance changes [1, 2]. We adopt a simple multiplicative model for the scene illumination, therefore the luminance profile $F(x,y)$ turns out to be the product between an "illumination profile" $I(x,y)$, which is assumed as slowly varying in space, and a "local texturing" $S(x,y)$, whose frequency content is more in the high range. As both reference frame $F_r(x,y)$ and current frame $F_c(x,y)$ are modeled as such a product

$$\begin{aligned} F_r(x,y) &= I_r(x,y)S_r(x,y) \\ F_c(x,y) &= I_c(x,y)S_c(x,y) \quad , \end{aligned} \tag{2}$$

the behavior of the ratio F_c/F_r will exhibit different frequency content, depending on what is changing in the scene. If the local change is purely in the illumination, then the ratio $V = F_c/F_r$ will correspond to I_c/I_r, therefore its frequency content will be in the low range. Conversely, if the change is purely geometric, then the ratio F_c/F_r will correspond to S_c/S_r, therefore it will be rapidly varying. This can be easily exploited using two filters that extract the two frequency components of interest, like the local average $m_v(x,y)$ (low-pass) and the local standard deviation $\sigma_v(x,y)$ (high-pass). The analysis of this information is, at this point, quite straightforward:

- $|m_v| \simeq 1$ and modest σ_v imply that no change occurred between F_r and F_c;
- $|m_v| >> 1$ and modest σ_v imply that there has been a diffuse change in the scene, which is likely to be due to the illumination;
- large σ_v implies that there has been a signficant variation in the local texturing, which is likely to be due to geometrical changes in the scene.

Indeed, there are many exceptions to this criterion, which are due to model failure. A multiplicative model of the illumination is, in fact, quite simplicistic, and is easy to fail, for example, in the presence of reflective surfaces and non-diffuse illumination. However, for matte surfaces and diffuse illumination, it performs quite nicely.

A quantized version of the local variance map M_σ is stored by the system for further analysis (see Fig. 2).

Fig. 2. Example of the variance map generated by the change detector. Darker regions correspond to a larger variance.

3. ATTENTION FOCUSING

The attention focusing block implements an object tracking phase by searching for correspondences between bounding boxes in consecutive frames. This is done by studying the similarity of both shape and motion. The tracking phase, as well as the classification phase, are based on the variance mask M_v generated by the previous block.

The tracking algorithm is based on the method proposed by Chetverikov [4]. The attention is focused on three consecutive frames (previous, current, and next), and the correspondence between bounding boxes is determined through the minimization of an appropriate cost function f. This cost function takes into account both motion compatibility (based on the motion of the centroids of the variance mask within the considered boxes) and shape compatibility (based on the zero-order moment of the variance masks).

This solution is characterized by a good computational efficiency, particularly in situations like ours, where the number of change areas typically no more than 6 or 7.

4. CLASSIFICATION

Once determined and tracked the change areas, we need to classify such changes according to their origin. In particular, we want to distinguish geometric changes (moving intruders, sometimes only partially visible) from any other type of change (typically shadows, reflexions, and noise sources of other nature). The parameters used to discriminate between such two categories are:

- **Weight ratio** – Ratio R between the zero-order moment of the variance mask $M_\sigma(x, y)$ and its perimeter. This parameter describes the "activity" of the luminance profile, after normalization on the part of the perimeter. This normalization action tends to make the parameter insensitive to the distance from the viewpoint.

- **Morphological index** – This parameter is based on the so-called morphological spectrum [5, 6], which is an operator that extracts the contribution of every structural element from an image through a series of operations of morphological opening

$$f_n = \frac{\mu(\Psi_n(M_v)) - \mu(\Psi_{n+1}(M_v))}{\mu(M_v)} \quad , \qquad (3)$$

  where Ψ_n is the morphological opening operator of order n; μ is the operator that computes the zero-order moment. Notice that n is the size of the kernel used for the morphological opening. The morphological spectrum of order n represents the contribution of the details of "size" n to the variance mask M_σ. In what follows we will use a morphological index that incorporates the information contained in several morphological spectrum coefficients.

In order to characterize such parameters, we run a series of tests with various types of intruders (completely visible, partially occluded, etc.) and scenes (strongly changing lighting conditions, presence of reflections, etc.). As we can see from Fig. 3, the parameter R associated to an intruder turned out to be always significantly higher than in the other cases, thus representing a strong discriminant for the purpose of intrusion detection. Furthermore, we noticed that the variance masks M_σ associated to intruders, tend to stimulate morphological indices of significant size ("relevant details"), such as f_2, while non-geometric changes usually excite smaller morphological indices ("irrelevant details"), such as f_1. After analyzing the response of such two morphological indices to intruders, partially-occluded intruders, and scene changes of other nature, we concluded that it would be more efficient and equally effective efficient to simply look at their difference. In fact, morphological indices corresponding to an intruder tend to stay above a 45-degree line on the (f_1, f_2) plane, while those of non-intruders tends to occupy the area below such line.

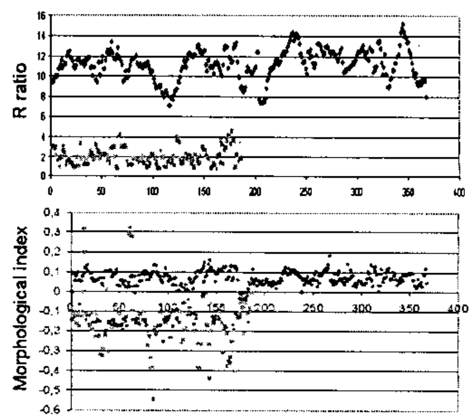

Fig. 3. Intrusion discrimination based on the selected parameter. The light dots represent non-geometrical changes, while the black dots correspond to intruders. The abscissa corresponds to the scene index.

Now that we have a pair of good discriminants, we can use them jointly throufsgh a properly defined classifier. Our approach to this problem is based on fuzzy logic [7] and the semantic rules used for classification are:

- *IF R is High AND Morphological Index is Relevant THEN geometric (human) intrusion;*

- *IF R is Low AND Morphological Index is Irrelevant THEN non-geometric intrusion.*

The membership function relative to the input linguistic variables were determined through statistical analysis (hystograms) of the available data.

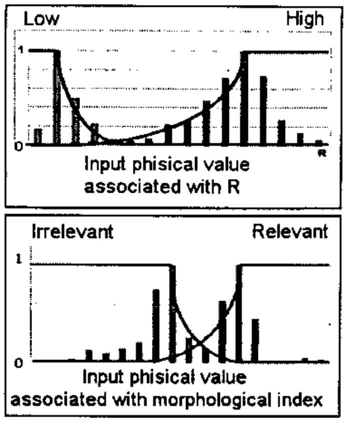

Fig. 4. Membership function relative to the input linguistic variables of the fuzzy logic classifier.

5. PERFORMANCE EVALUATION

The system performance was measured in terms of wrong classifications. We acquired our grayscale videos using both a low-quality webcam and a digital camera of good quality. The intruder, once in the scene, was allowed to change posture or partially hide behind furniture. The scene lighting was often changed during the video acquisition. The results of these experiments are collected in the following table, where the first column lists the content, while the first row lists the results of the classification.

	Intruder	Not an intruder	Uncertain
Visible intruder	175	10	8
Partially visible intruder	484	65	2
Not an intruder	20	206	0

In spite of the unfavorable selection of testing sequences, the percentage of correct classification is around 89%. If we had used only the ratio R, the percentage of success would have dropped of more than 10%. The column labeled as "uncertain" denotes the situations in which the classifier was unable to make a decision. In all considered videos, however, the intruder's trajectory was always uniquely identified. As far as computational efficiency is concerned, the system is able to work at 3 to 5 frames per second on a 800MHz PIII.

Fig. 5. A view of the working system. The bounding box has been correctly tracked and classified as associated to an intruder.

6. CONCLUSIONS

We proposed and implemented a video-surveillance system operating in real-time that turned out to be robust against illumination changes and shadows in the scene. The classifier proved able to correctly recognize intruders even in unstable and difficult illumination conditions.

7. REFERENCES

[1] T. Ebrahimi, E. Durucan, "Robust and Illumination Invariant Change Detection Based on Linear Dependence". *Proc. EUSIPCO 2000*, Tampere, Finland.

[2] K. Skifstad, R. Jain, "Illumination Independent Change Detection for Real World Image Sequences", *Proc. CVGIP*, Vol. 46, pp 387–399, 1989.

[3] G. X. Ritter, J. N. Wilson, *Handbook of Computer Vision Algorithms in Image Algebra*, CRC Press, 1996.

[4] D. Chetverikov, J. Verestoy, "Tracking Feature Points: a New Algorithm" *Proc. ICPR*, Vol. 2, pp. 1436–1439, 1998.

[5] P. Maragos "Pattern Spectrum and Multiscale Shape Representation", *IEEE Tr. PAMI*, Vol. 2, pp. 701–716, 1989.

[6] A. Asano "Texture Analysis Using Morphological Pattern Spectrum and Optimization of Structuring Elements" *Proc. ICIAP*, pp. 209–214, 1999.

[7] L. A. Zadeh "From Computing With Numbers to Computing with Words. From Manipulation of Measurements to Manipulation of Perceptions". *IEEE Tr. CAS I: Fund. Theory and Appl.*, Vol. 46, pp. 105–119, 1999.