# Project: An a-contrario approach for object detection in video sequence

Benoît REBER

April 3, 2023

**Article implemented:** M Ammar, S Le Hégarat-Mascle, "An a-contrario approach for object detection in video sequence", International Journal of Pure and Applied Mathematics 89 (2), 173-201, 2013

**Abstract**

This report details a method for object detection in video sequences, following an *a contrario* approach. The core idea is to identify abnormal concentrations of anomalies within an image, defined as the difference between a current observation and a background model. The process is divided into two main stages: first, detecting individual pixels that show significant change, and second, identifying rectangular zones with a statistically high density of these "object pixels". A key feature of this method is its ability to operate without requiring manual thresholding, relying instead on the principle of minimizing the Number of False Alarms (NFA). We present an experimental validation of the method in three scenarios: detecting an intruder object, motion detection, and a specific failure case, highlighting the importance of a robust background model.

## 1 Introduction

Object detection in video sequences without trying to characterize them is a common problem that often constitutes a pre-processing step in embedded systems or video surveillance applications.

The authors of the article propose a general framework that consists of reducing object detection to the detection at time $t$ of zones with abnormally high concentrations of anomalies in the image $\Delta_t$ defined as the difference between an observation $I_t$, also called innovation, and a model $B_t$ constructed from previous observations: $\Delta_t = I_t - B_t$.

An a contrario model is proposed in which our anomalies will correspond to pixels $s$ such that $\Delta(s)$ is significant, meaning that there is an "abnormal" variation of the innovation relative to the model at these points. An $NFA$

is then proposed as well as an algorithm that has the particularity of not using a detection threshold hyperparameter, as is often the case.

The article's approach is general in that the choice of the innovation $I_t$ and the model $B_t$ defines what will be detected. For example, in the case of a fixed camera, by taking an image of the place in normal time as the model $B_t$ and the image captured at time $t$ as the innovation, one will detect zones where objects of interest are potentially located. Also, if we take the image at time $t-1$ as the model $B_t$ and the image at time $t$ as the innovation $I_t$, one will detect zones where moving objects are potentially located. The authors propose combining different choices of $B_t/I_t$ pairs to detect different classes of objects in the practical case of cameras embedded in vehicles.

The act of searching for zones with an abnormal concentration of anomalies to find objects is in direct connection with Gestalt theory. Let's place ourselves in the case where the model $B_t$ is the empty scene and $I_t$ is the image actually observed. When the color difference is significant, by the law of color consistency, we tend to associate the points with different objects. Searching for zones where many of these differences are observed is in fact searching for zones where we are likely to associate these points by the law of proximity to form objects. We, of course, need to assume here that the image definition is sufficient for the objects to occupy several pixels. In the other useful cases, we can generally apply similar reasoning with these two laws of Gestalt theory.

We denote $\mathcal{D}_t$ the set of pixels whose value is known. (useful case of a moving camera if we have to project an image from a past time taking into account the movement, we lose pixels)

The approach adopted in the article consists of 3 steps:

1. Detect pixels where there are significant differences between the model $B_t$ and the innovation $I_t$. A first background model $\mathrm{H}_1$ and an $NFA_1$ criterion are then defined. This first step makes it possible to move with a low computational cost from an image with unknown pixels and pixels having a numerical value (gray level) to an image whose pixels take on three labels: *unknown*, *background*, *object*.

2. Detect zones of the image where there is an abnormally high concentration of points labeled *object*. A second background model $\mathrm{H}_1$ and an $NFA_2$ criterion are then defined. The authors choose to search for rectangular zones.

3. Elimination of redundant detections and selection of an optimal set of windows defined as the one maximizing the $NFA_2$ criterion.

In the following, we will often omit the index $t$ when it does not harm understanding.

## 2 Detection of points of interest

The authors propose a simple model for $\Delta$, the difference between the innovation $I_t$ and the model $B_t$.

**Background model 2.1** ($\mathrm{H}_1$)**.** *We assume that the $|\mathcal{D}|$ coordinates of the image $\Delta$ are independent and identically distributed with law $\mathcal{N}(0, \sigma^2)$.*

This choice seems reasonable since if we assume the model $B_t$ to be faithful enough to reality, there is no reason to expect with the innovation $I_t$ any differences other than noise due to the measurement. The noise is chosen to be centered Gaussian, which is a common choice, often encountered in nature and practical for calculations.

Several methods are proposed to choose $\sigma^2$; for example, one can take $I_t$ and $B_t$ two arbitrary independent innovations and directly estimate the variance of the coordinates of $\Delta_t$.

The authors then propose to focus on $\Delta^2$ defined by $\Delta^2(s) = [\Delta(s)]^2$ for any $s$. The choice to square it allows, on the one hand, to move small values ($< 1$) closer to zero and large values ($\geq 1$) further away from it, and on the other hand, by setting, for $D \subset \mathcal{D}$, $\delta_D^2 = \sum_{s \in D} \Delta^2(s)$, we easily fall back to a well-known distribution whose quantiles are easy to calculate: $\delta_D^2/\sigma^2$ follows a $\chi^2$ distribution with $|D|$ degrees of freedom.

To find the subset of $\mathcal{D}$ containing the pixels with an abnormally high $\Delta^2$ value, the authors choose to try to detect its complementary in $D_t$, the set $\hat{D}$ of pixels with an abnormally low value - that is, those for which the model $B_t$ and the innovation $I_t$ agree and which will be labeled *background*.

This choice is justified for practical reasons: by searching for the set of *background* points, we know that we must make exactly one detection, whereas conversely, it is not reasonable to assume that we must necessarily make a detection by searching for the most significant set in terms of abnormally large values. We would have to choose a detection threshold $\epsilon$ below which we choose not to detect anything, which would have the disadvantage of adding a hyperparameter.

**Remark.** *There is therefore no problem of redundant detections since we assume that there is one and only one detection to be made and we then choose the most significant one.*

We denote $\mathbb{P}_{H_1}(\delta_D^2)$ the probability of observing by chance, under the model $\mathrm{H}_1$, a value of $\delta_D^2$ at most as large as the one observed. Since $\delta_D^2/\sigma^2$ follows a $\chi^2$ distribution with $|D|$ degrees of freedom, we then have

$$\mathbb{P}_{H_1}(\delta_D^2) = \mathbb{P}(\chi^2 \in [0, \delta_D^2/\sigma^2] | \delta_D^2) = \frac{1}{\Gamma(\frac{|D|)}{2}} \int_0^{\delta_D^2/\sigma^2} e^{-t} t^{\frac{|D|}{2}-1} dt$$

As there are $|\mathcal{D}|$ ways to choose the number of pixels in a part $D$ of $\mathcal{D}$, it appears that there are $2^{|\mathcal{D}|} = \sum_{k=0}^{|\mathcal{D}|} \binom{|\mathcal{D}|}{|D|}$ parts of $\mathcal{D}$ to test.

Following the article, we set $\eta_1(D) = |\mathcal{D}| \binom{|\mathcal{D}|}{|D|}$. This choice makes it possible to penalize parts that contain a number of pixels close to $|\mathcal{D}|/2$, which may seem reasonable because we expect to have a relatively small number of detections.

The authors then propose to define a first NFA as follows:

$$\begin{aligned} NFA_1(D) &= \eta_1(D)\mathbb{P}(\chi^2 \in [0, \delta_D^2/\sigma^2] | \delta_D^2) \\ &= \eta_1(D) \frac{1}{\Gamma(\frac{|D|}{2})} \int_0^{\delta_D^2/\sigma^2} e^{-t} t^{\frac{|D|}{2}-1} dt \end{aligned}$$

We can easily check that it is indeed an NFA thanks to proposition 4 on page 44 of the course. Indeed, it is enough to check $\sum_{D \subset \mathcal{D}} \frac{1}{\eta_1(D)} \leq 1$; a simple calculation shows that there is in fact equality.

$\hat{D}$ is then defined as the part of $\mathcal{D}$ whose $NFA_2$ is the smallest.

To avoid having to test each part of $\mathcal{D}$, the authors propose an algorithm based on a simple property.

For a fixed $|D|$, the smallest value of $NFA_1$ is obtained for $D$ having the lowest value $\delta_D^2 = \sum_{s \in D} \Delta^2(s)$, meaning that D is exactly the subset of $\mathcal{D}$ associated with the $|D|$ smallest values of $\Delta^2$. It is then a matter of constructing $\hat{D}$ by successively adding the remaining pixel $s$ associated with the smallest value $\Delta^2(s)$ as long as this addition causes $NFA_1(\hat{D})$ to decrease.

We report here the algorithm proposed in the article:

## 3 Detection of zones of interest

At this stage, we have an image whose pixels can have three labels: *unknown* ($\bar{\mathcal{D}}$), *background* ($\hat{D}$), *object* ($\mathcal{D} \setminus \hat{D}$). We then search for zones with high concentrations of points labeled *object* in connection with Gestalt theory as explained in the introduction. We choose to restrict ourselves to searching for rectangular windows.

To do this, we note $p = \frac{|\mathcal{D} \setminus \hat{D}|}{|\mathcal{D}|}$ the empirical proportion of points of interest (*object*) among the known pixels of the image and we define the second background model proposed in the article.

**Background model 3.1** ($\mathrm{H}_2$)**.** *We assume that the points of $\mathcal{D}$ are independent (as for* $\mathrm{H}_1$*) and that each has the label object with probability $p$ and background with probability* $1 - p$.

**Algorithm 1** Estimation of *background* and *object* pixel sets.

Compute image $\Delta^2$
Sort increasingly $\Delta^2$ pixels in a vector $\left(\epsilon_i^2\right)_{1\leq i\leq|\mathcal{D}|}$
Initialize $\delta_D^2 \leftarrow 0$, $D \leftarrow \emptyset$ and $NFA_1^{min} \leftarrow +\infty$
**for** $i \in [1, |\mathcal{D}|]$ **do**
    Add $\epsilon_i^2$ to $\delta_D^2$
    $s \leftarrow$ the pixel corresponding to $i$ index in $\left(\epsilon_i^2\right)_{1\leq i\leq|\mathcal{D}|}$
    $D \leftarrow D \cup \{s\}$
    Compute $NFA_1\left(\delta_D^2, |D|, \sigma, |\mathcal{D}|\right)$
    **if** $NFA_1\left(\delta_D^2, |D|, \sigma, |\mathcal{D}|\right) < NFA_1^{min}$ **then**
        $NFA_1^{min} \leftarrow NFA_1\left(\delta_D^2, |D|, \sigma, |\mathcal{D}|\right)$
        $\hat{D} \leftarrow D$
    **end if**
**end for**
*point* domain $\leftarrow \mathcal{D}\backslash\hat{D}$

This background model allows us to return to a common framework in detection theory that was covered in the course. We will be led to work with tails of binomial distributions that we can estimate well when the number of tests is large.

For a window $W$, we note $\kappa_W = |W \cap \mathcal{D} \setminus \hat{D}|$ the number of *object* points in the window and $\nu_W = |W \cap \mathcal{D}|$ the total number of known pixels in the window. $p_W = \frac{\kappa_W}{\nu_W}$ is then the proportion that interests us.

We denote $\mathbb{P}_{\mathrm{H}_2}(p, \kappa_W, \nu_W)$ the probability of observing by chance, under the model $\mathrm{H}_2$, more than $\kappa_W$ *object* points in the window $W$. Since under $\mathrm{H}_2$, $\kappa_W$ follows a binomial distribution with parameters $\nu_W$ and p, we then have

$$\mathbb{P}_{\mathrm{H}_2}(p, \kappa_W, \nu_W) = \mathbb{P}(\mathcal{B}(\nu_W, p) \geq \kappa_W | \kappa_W)$$

We can easily count the number of rectangular windows we can choose and therefore the number of tests to perform. Indeed, a rectangular window of any size is defined entirely by two points in opposite corners. If we denote $l$ the width of the image and $L$ its length, then there are $\sum_{i=1}^{l}\sum_{j=1}^{L}(l+1-i)(L+1-j) = \sum_{i=1}^{l}\sum_{j=1}^{L} ij = \frac{l(l+1)}{2}\frac{L(L+1)}{2}$. This number is far too large in practice for real-time use; without specifying it, it seems that the authors then propose to restrict themselves to the following case. There is 1) a single dimension (unique length and unique width) of window with a fixed number of pixels and 2) windows with the same number of pixels are disjoint. For a window $W$ of $k$ pixels, there are then at most $\frac{|\mathcal{D}\cup\bar{\mathcal{D}}|}{k}$. As we are going to consider different window sizes, the article then proposes to set $\eta_2(W) = \frac{|\mathcal{D}\cup\bar{\mathcal{D}}|}{|W|}2^{|W|}$. We note that this choice penalizes windows that are too large.

The authors then propose the following NFA:

$$
\begin{aligned}
NFA_2(W) &= \eta_2(W)\mathbb{P}_{\mathrm{H}_2}(p, \kappa_W, \nu_W) \\
&= \eta_2(W) \sum_{i=\kappa_W}^{\nu_W} \binom{\nu_W}{i} p^i (1-p)^{\nu_W - i}
\end{aligned}
$$

We then check $\sum_j \frac{1}{\eta_2(W_j)} \leq 1$ to conclude as before that it is indeed an NFA as defined in the course. Since there are at most $\frac{|\mathcal{D} \cup \bar{\mathcal{D}}|}{k}$ disjoint windows with $k$ pixels, we write:

$$
\begin{aligned}
\sum_j \frac{1}{\eta_2(W_j)} &\leq \sum_{k=1}^{|\mathcal{D} \cup \bar{\mathcal{D}}|} \frac{1}{2^k} \\
&\leq \sum_{k=1}^{\infty} \frac{1}{2^k} = 1
\end{aligned}
$$

**Remark.** *It is important to note that the $NFA_2$ proposed in the article is not an NFA if we do not make additional hypotheses compared to what is indicated in the article. Indeed, if we do not assume that there is a single dimension of window (unique length and width) with a fixed number of pixels, it is immediate that $\sum_j \frac{1}{\eta_2(W_j)} > 1$. In practice, since we will limit ourselves to a limited number of different window sizes and these windows will have a sufficiently large minimum number of pixels, we will in any case have an NFA as defined in the course (we directly adapt the proof of NFA verification from the next section).*

For the calculation of the tail of the binomial distribution, we will use the approximation seen in the course and constructed from Hoeffding's inequality.

# 4 Management of redundant detections and selection of an optimal set of zones

To avoid redundant detections, the authors propose an approach without a hyperparameter.

The first step of the method consists of eliminating redundant detections and is close to the exclusion principle of definition 18 on page 163 of the course. The authors choose to take windows that are pairwise disjoint. This seems a reasonable choice for the application it is intended for: detecting objects for pre-processing. When two windows overlap, we keep the one with the greatest significance (lowest $NFA_2$). In practice, we rank all the windows in ascending order of $NFA_2$ and by going through this list, we remove any window that has a non-empty intersection with a window already encountered.

We always compare the method of the authors of the article to that on page 163 and algorithm 7 proposed there.

The second step then differs from algorithm 7 proposed in the course in the section related to the exclusion principle. Instead of choosing an $NFA_2$ threshold below which we retain the detections, an approach without a hyperparameter is proposed. It consists of choosing the set of windows whose $NFA_2$ of the union is the lowest. This amounts to recursively constructing the set $\bigcup_{i\leq k-1}^{K} W_i$ by setting

$$W_k = argmin_W \left\{ NFA_2(p, \bigcup_{i\leq k-1} W_i \cup W \right\})$$

To extend the definition of $NFA_2$, we set $\kappa_{1\ldots k} = |\bigcup_{i\leq k} W_i \cap \mathcal{D} \setminus \hat{D}|$ and $\nu_{1\ldots k} = |\bigcup_{i\leq k} W_i \cap \mathcal{D}|$. These quantities can also be written as sums since after the first step we are working with disjoint windows. We finally set $\bar{p}_k = \frac{\kappa_{1\ldots k}}{\nu_{1\ldots k}}$ the proportion of points labeled *object* within the union of windows considered. We can then calculate $NFA_2$ as before using these new variables.

For the calculation of $NFA_2$, we use the Hoeffding approximation seen in proposition 10 on page 68 of the course. The authors of the article then define *significance* as follows:

$$\begin{aligned} S(p, \bigcup_{i\leq k} W_i) &= -log(NFA_2(\bigcup_{i\leq k} W_i)) \\ &\simeq \nu_{1\ldots k} \left[ \bar{p}_k log(\frac{\bar{p}_k}{p}) + (1-\bar{p}_k) log() \frac{1-\bar{p}_k}{1-p} \right] \\ &\quad - log(\eta_2(\bigcup_{i\leq k} W_i)) \end{aligned}$$

In this new framework, we must check that we again have an NFA when testing sets of windows. This proof is provided in the article. We denote $N_W$ the total number of windows after step 1 and $a$ the minimum number of pixels in a window.

$$\begin{aligned} \frac{1}{|\mathcal{D}|} \sum_{k=1 window}^{N_W windows} \frac{\nu_{1\ldots k}}{2^{\nu_{1\ldots k}}} &\leq \frac{1}{|\mathcal{D}|} \sum_{k=1 window}^{N_W windows} \frac{ak}{2^{ak}} \\ &\leq \frac{1}{|\mathcal{D}|} \sum_{k=1}^{N_W} \frac{ak}{2^{ak}} \binom{N_W}{k} \\ &\leq \frac{a}{|\mathcal{D}|} \frac{N_W}{2^a} e^{\frac{N_W-1}{2^a}} \end{aligned}$$

We successively use the variations of $x \to \frac{x}{e^x}$, that $\forall x \in \mathbb{R}, \forall n \in \mathbb{N}, x\frac{d}{dx}(1+x)^n = \sum_{k=1}^{n} kx^k \binom{n}{k}$ and the inequality $\forall x, r \in \mathbb{R}_+, (1+x)^r \leq e^{rx}$. Since the term on the right is increasing in $N_W$, in practice it will be enough to check with the image format and the window sizes that we are testing that the right-hand side is indeed less than 1 to have an NFA. t In the particular case where all windows have the same number of pixels - that is, there exists $\alpha > 0$ such that $\nu_{1\ldots k} = k\alpha$ - we can write $\bar{p}_k = \frac{\sum_{i=1}^{k} \kappa_i}{k\alpha}$.

By replacing $\bar{p}_k$ by $\frac{\kappa_{1\ldots k}}{\nu_{1\ldots k}}$ in the expression for significance and studying it as a function of $\kappa_{1\ldots k}$ (i.e., we fix the number of windows to choose and their size, so only $\kappa_{1\ldots k}$ varies), we easily show that for a fixed number of windows $k$, the greatest significance is obtained by taking the $k$ windows with the largest proportion of pixels labeled *object*. The authors then propose to use the result when the windows are all of the same size as an approximation for the general case to avoid having to test all possible combinations of windows and choose the one minimizing the $NFA_2$. This choice is supported by their experiments which show that the correct set of windows is then generally found.

We report here the algorithm proposed in the article which we have corrected some typos from:

**Algorithm 2** Estimation of object windows.

Compute the parameter $p$ and the 3-label image
Initialize the set of windows $\{W\}_{1\leq i\leq N_W} = \{W_j, j \in [1, N_W]\}$ where $N_W$ is the initial number of considered windows
**for all** window $W$ in $\{W\}_{1\leq i\leq N_W}$ **do**
  Compute the parameters $(\kappa_W, \nu_W, p_W)$
  Compute the significance $S_W = S(p, p_W, \nu_W, |W|)$ (Eq. 5)
**end for**
Sort decreasingly $S_W$ values in a vector $(S_i)_{1\leq i\leq N_W}$
**for** $j \in [1, N_W]$ **do**
  $W_j$ such that $S_{W_j} = S_j$ in $(S_i)_{1\leq i\leq N_W}$
  **for** $k \in [j+1, N_W]$ **do**
    $W_k$ such that $S_{W_k} = S_k$ in $(S_i)_{1\leq i\leq N_W}$
    **if** $W_k \cap W_j \neq \emptyset$ **then**
      remove $W_j$ from $(S_i)_{1\leq i\leq N_W}$
      $N_W \leftarrow N_W - 1$
    **end if**
  **end for**
**end for**
Sort decreasingly $p_W$ values in a vector $(p_i)_{1\leq i\leq N_W}$
Initialize $S_2^{max} = -\infty$, $\nu_\cup = 0$, $\kappa_\cup = 0$ et $\hat{K} = 0$
**for** $k \in [1, N_W]$ **do**
  $W_k$ such that $p_{W_k} = p_k$ in $(p_i)_{1\leq i\leq N_W}$
  Add $\nu_{W_k}$ to $\nu_\cup$, and $\kappa_{W_k}$ to $\kappa_\cup$
  Compute $\overline{p_k} = \frac{\kappa_\cup}{\nu_\cup}$
  Compute $S_\cup = S\left(p, \overline{p_k}, \{W_j\}_{j=1}^{k}\right)$ (Eq. 6)
  **if** $S_\cup > S_2^{max}$ **then**
    $S_2^{max} = S_\cup$
    $\hat{K} \leftarrow k$
  **end if**
**end for**
$\widehat{D} = \bigcup_{j=1}^{\hat{K}} W_j$

# 5 Experiments

We propose to test the method of the article in 3 cases:

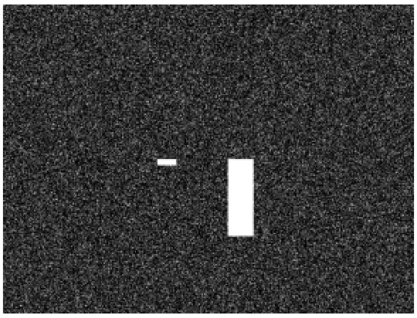
1. detection of an intruder object: we have a reliable model $B_t$ and we then search in the observation $I_t$ for an object that should not be there
2. motion detection: for a fixed observer, we compare the image captured at time $t$ with that captured at time $t-1$. We set $B_t = I_{t-1}$.
3. failure of the method: by taking a context similar to the previous framework, we simulate the effect of turning on a light which will disrupt the detection of the moving object

We work with simple simulated data and we use the same window and image dimensions as in the article (so the $NFA_2$ is indeed an NFA). Image of dimension 640x480, window sizes 20x20, 20x40, 40x20, 40x40?

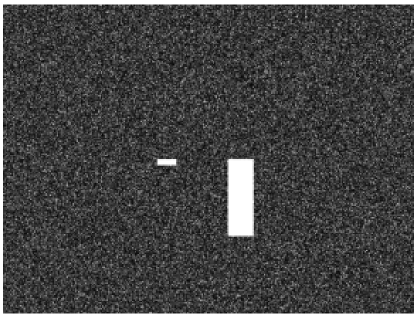
## 5.1 Detection of an intruder object

We simply choose $B_t$ to be the matrix with only zeros. For $I_t$, we take $B_t$ to which we have added an independent $\mathcal{N}(0, 1)$ noise on each coordinate to represent potential measurement errors, and two intruder objects: these are two rectangles, one of which does not fit into any window, and another that fits into the 40x40 window.

(a) $B_t$



(b) $I_t$



(c) $\Delta^2$

(d) *object* points

We observe that step 1 allows us to correctly assign the *object* label to the two rectangles added in addition to the noise.

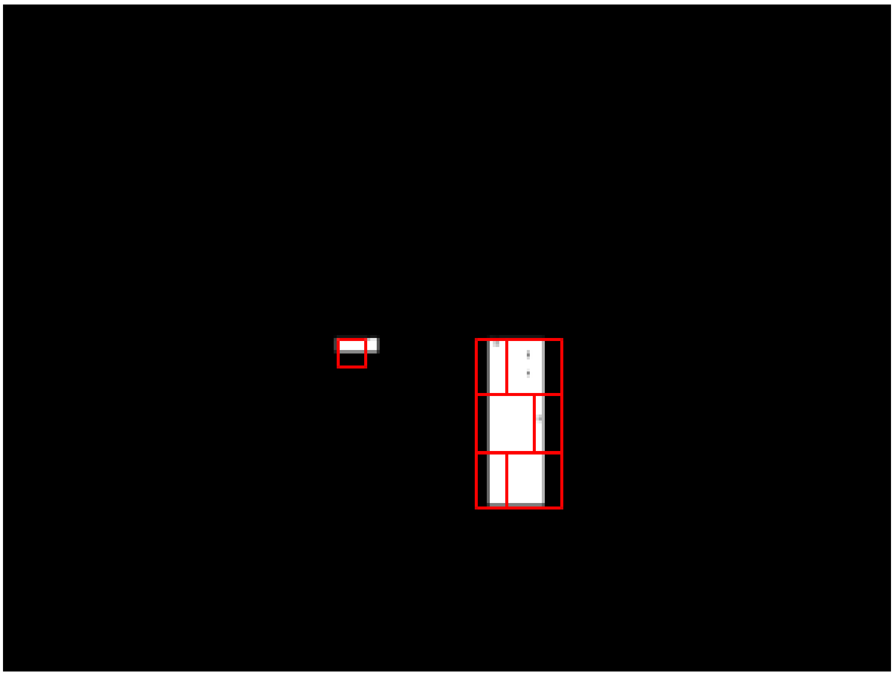


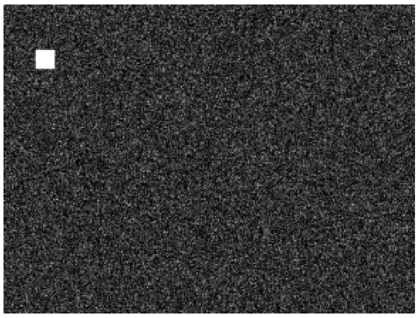
Figure 1: Detections on the *object* points

The algorithm correctly selects windows containing the two objects as sought.

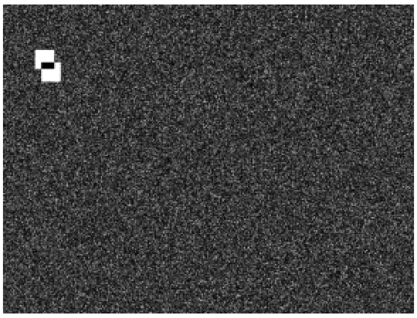
### 5.2 Detection of a moving object

We now choose $B_t$ to be the matrix with only zeros to which we have added a square with a side of 30. For $I_t$, we take the matrix with only zeros and add the same square but slightly translated.

(a) $B_t$



(b) $I_t$



(c) $\Delta^2$

(d) *object* points

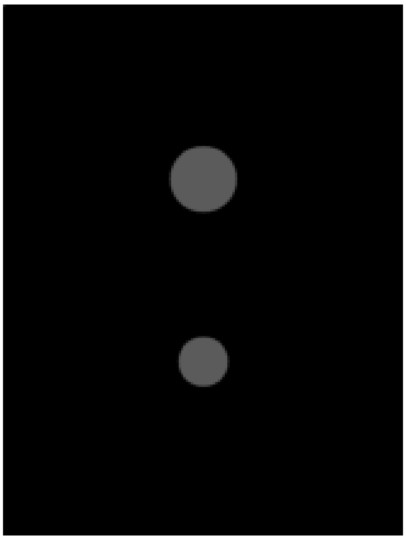
Again, we observe that step 1 correctly assigns the *object* label to the desired zones; here to zones containing the moving object at time $t$ or at time $t-1$ but not both.

Figure 2: Detections on the *object* points

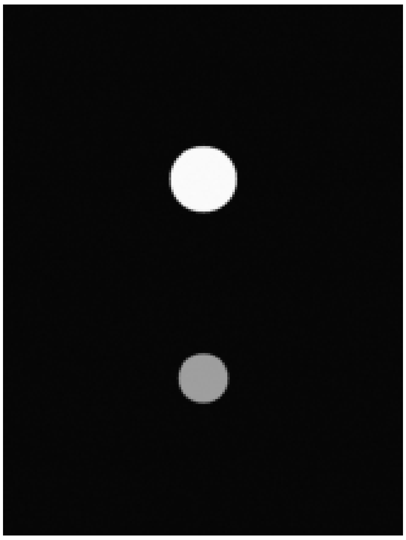
The algorithm correctly selects windows containing the object at time $t$ or at time $t-1$.

### 5.3 Failure of the method

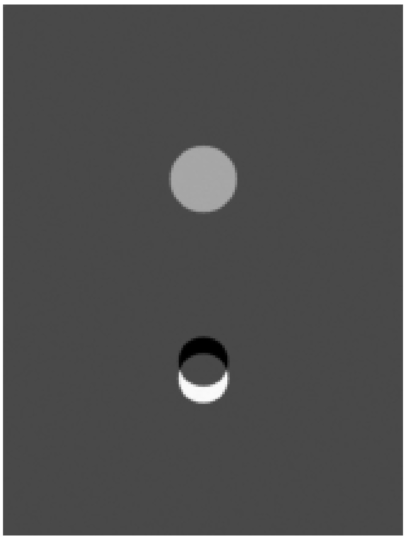
We place ourselves in a context similar to the previous case, $B_t = I_{t-1}$. We have two objects initially at time $t-1$. At time $t$, one of them is translated while the other object becomes a light source that radiates over the entire scene.



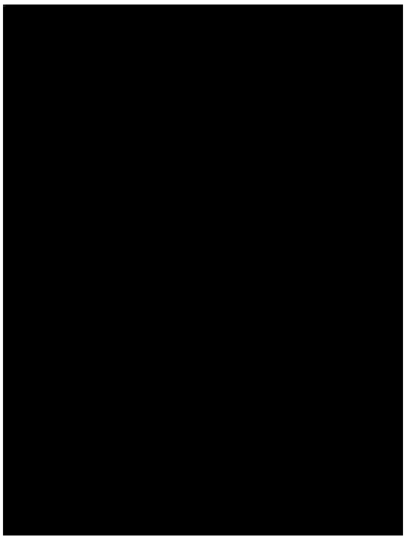
(a) $B_t$



(b) $I_t$



(c) $\Delta^2$



(d) *object* points

In this case, we fail to detect any points labeled *object*.
The algorithm is in complete failure. The reason is quite obvious: since $\Delta^2$ contains only large values, the probability of observing, under the model $\mathrm{H}_1$, sets of pixels with smaller values is always almost equal to 1. This case of failure illustrates well the importance of the model $B_t$. Here in the case of a sudden change in luminosity, it is erroneous and we cannot detect anything. This example also illustrates the lesser importance of the $\sigma^2$ parameter mentioned by the authors; it must be chosen reasonably but

it is better to have a good model $B_t$ than a good estimate of $\sigma^2$.

**Remark.** *In practice, we find that the first step of the method (detection of pixels labeled background) is very computationally expensive. This is mainly due to the fact that there are generally few points that change significantly. Several avenues for improvement are possible. For example (and this is what we have implemented), since there are generally very few points to detect compared to the total number of pixels, we can start from* $\delta_D^2 = \sum_{s \in \mathcal{D}} \Delta^2(s)$ *and progressively remove the pixels with the largest* $\Delta^2$ *value to search for the smallest* $NFA_1$*. Since the evolution of* $NFA_1$ *as a function of the number of points (selected as in algorithm 1) has a minimizer and seems to be convex, many approaches are in fact conceivable. For example, a very simple solution not tested could be to search for the minimum of the function by dichotomy.*

# 6 Conclusion

The a contrario approach to object detection, as presented in the reference article and detailed in this report, provides a robust framework that sidesteps the common problem of parameter tuning. By defining objects as statistically significant anomalies, the method eliminates the need for arbitrary thresholds. The two-stage process—first detecting anomalous pixels, then grouping them into significant regions—is both theoretically sound and practically effective. Experiments confirm its success in classic scenarios like detecting intruders or motion, which underscores the method's versatility and reliability. However, the failure case under a sudden change in ambient lighting highlights a crucial limitation: the method's performance is highly dependent on the quality of the background model. While a good estimation of the variance $\sigma^2$ is important, a well-chosen and adaptable model $B_t$ is paramount for robust detection. Moreover, the computational cost of the first detection stage presents a practical challenge, suggesting avenues for future optimization. Potential improvements could involve more efficient algorithms for finding the set of "background" pixels or adapting the method to handle dynamic changes in the scene more gracefully.

# 7 Bibliography


(a) M Ammar, S Le Hégarat-Mascle, "An a-contrario approach for object detection in video sequence", International Journal of Pure and Applied Mathematics 89 (2), 173-201, 2013

(b) Jean-Michel Morel, Rafael Grompone von Gioi, "Detection Theory and Industrial Applications", 2024