Detection theory and industrial applications

Jean-Michel Morel jean-michel.morel@ens-paris-saclay.fr

With contributions of:

Agnès Desolneux, Lionel Moisan, Rafael Grompone, Thibaud Ehret, Gregory Randall, Jérémy Jakubowicz, Yiqing Wang, Mauricio Delbracio, Pablo Musé, Tina Nikoukhah, Marina Gardella, Miguel Colom, José Lezama

“*Detection* ” is the most frequent request made by researchers, industrials, police, press, defence for exploiting images, images series, video among other data.

“*Detection* ” means that an automatic decision must be made. A wrong decision may entail costs and false alerts if it is falsely positive, and worse costs, accidents and disasters if it is falsely negative.

Therefore *Detection* requests a general decision theory controlling the “number of false alarms” and giving tight detection thresholds

This theory exists, it uses simple (but sometimes subtle) probability arguments, mixed with a fine control of image and video features

References

Skolnik, M. I. **(1962).** Introduction to radar. *Radar handbook*, *2*, 21.

LSD: a Line Segment Detector (2012), www.ipol.im
Rafael Grompone von Gioi, Jérémie Jakubowicz, J.M.M., Gregory Randall

An Unsupervised Point Alignment Detection Algorithm (2015), www.ipol.im
José Lezama, Gregory Randall, J.M.M., Rafael Grompone von Gioi

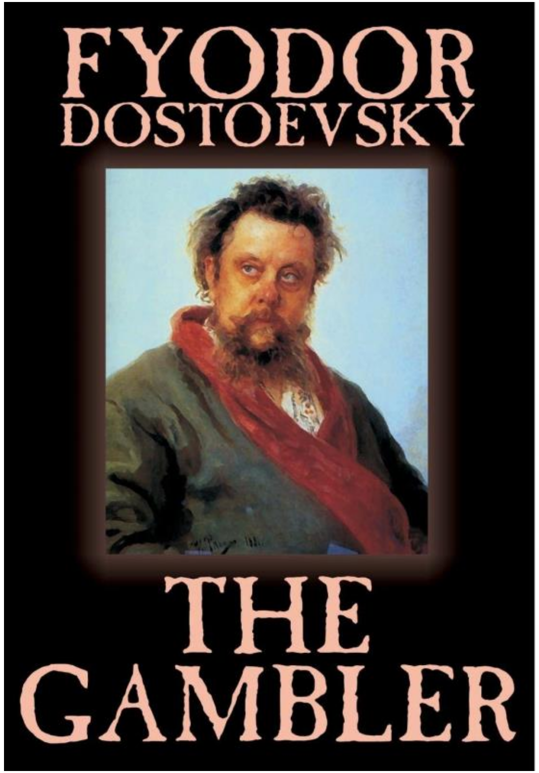
Cloud Detection by Luminance and Inter-band Parallax Analysis for Pushbroom Satellite Imagers (2020), www.ipol.im
Tristan Dagobert, Rafael Grompone von Gioi, Carlo de Franchis, J.M.M., Charles Hessel

Local JPEG Grid Detector via Blocking Artifacts, a Forgery Detection Tool (2020), www.ipol.im
2020-05-21 · Tina Nikoukhah, Miguel Colom, J.M.M., Rafael Grompone von Gioi

How to Reduce Anomaly Detection in Images to Anomaly Detection in Noise, www.ipol.im
2019-12-08 · Thibaud Ehret, Axel Davy, Mauricio Delbracio, J.M.M.

Agnès Desolneux, Lionel Moisan, & J.M.M. (2007). *From gestalt theory to image analysis: a probabilistic approach* (Vol. 34). Springer Science & Business Media.

Example 1: Playing Roulette with Dostoievski



Example 1: Playing Roulette with Dostoevski

Extract from the novel « The gambler »:

That time, as if on purpose, a circumstance arose which, incidentally, recurs rather frequently in gambling. Luck sticks, for example, with red and does not leave it for ten or even fifteen turns. Only two days before, I had heard that ***red had come out twenty two times in a row*** *in the previous week. One* ***could never recall a similar case at roulette*** *and it was spoken of with astonishment.*

Number of false alarms = expected number of occurrences of the event = (number of tests) x (probability of the rare event)

Why 22? The probability that red appears 22 times in a row of 22 is $\left(\frac{18}{37}\right)^{22}$, namely about 10^{-7}. The computation of the probability that this happens at least once in a series of n trials is doable but a bit intricate.

We can, instead, directly compute the *expected number of occurrence of the event*, or "Number of False Alarms" as

$$\mathrm{NFA}(n) = (n-21) \times \left(\frac{18}{37}\right)^{22}.$$

Number of false alarms = expected number of occurrences of the event = (number of tests) x (probability of the rare event)

$$\mathrm{NFA}(n) = (n - 21) \times \left(\frac{18}{37}\right)^{22}.$$

The event is likely to happen if its is larger than 1, which yields roughly $n \geq 10^7$. Thus, we are led to compute how many trials a passionate gambler may have done in his life.

Considering that a professional gambler would play roulette at 100 evenings of 5 hours a year for 20 years, estimating in addition that a roulette trial may take about 30 seconds, we deduce that an experienced gambler would observe at the most, in his gambling life span, about $n = 20 \times 100 \times 5 \times 120 \simeq 10^6$ trials. We deduce that 1 out of 10 professional gamblers can have observed such a series of 22. Actually, Dostoevsky's information about the possibility of 22 series is clearly based on conversations with specialists. The hero says:

I own a good part of these observations to Mr. Astley, who spends all of his mornings by the gambling tables but never gambles himself.

Example 2: birthdays in a class

Example 2: birthdays in a class

- C_n the number of n-tuples of students in the class having the same birthday (this number is computed exhaustively by considering all possible n-tuples. If, for example, students 1, 2, and 3 have the same birthday, then we count three pairs $(1, 2)$, $(2, 3)$, and $(3, 1)$).
- $\mathbb{P}_n$, the probability that "there is at least one group of n alumni having the same birthday": $\mathbb{P}_n$ is the probability of the event "$C_n \geq 1$";
- p_n, the probability that there is at least one n-tuple and no $(n + 1)$-tuple.

We are primarily interested in the evaluation of $\mathbb{P}_n$ and of the expectation $\mathbb{E}C_n$ as good indicators for the exceptionality of the event.

Example 2: birthdays in a class

Proposition 3. *The expectation of the number of pairs of alumni having the same birthday in a class of 30 is* $\mathbb{E}C_2 = \frac{30\times 29}{2\times 365} \approx 1.192$. *The expectation of the number of* n*-tuples is* $\mathbb{E}C_n = \frac{1}{365^{n-1}}\binom{30}{n}$. *By an easy calculation,* $\mathbb{E}C_3 \approx 0.03047$ *and* $\mathbb{E}C_4 \approx 5.6\times 10^{-4}$.

Proof — Enumerate the students from $i = 1$ to 30 and call E_{ij} the event "students i and j have the same birthday". Also, call $\chi_{ij} = \mathbb{1}_{E_{ij}}$. Clearly, $\mathbb{P}(E_{ij}) = \mathbb{E}\chi_{ij} = 1/365$. Thus, the expectation of the number of pairs of students having the same birthday is

$$\mathbb{E}C_2 = \mathbb{E}\left(\sum_{1\le i<j\le 30}\chi_{ij}\right) = \sum_{1\le i<j\le 30}\mathbb{E}\chi_{ij} = \frac{30\times 29}{2}\frac{1}{365} \approx 1.192.$$

The general formula follows by analogous reasoning. $\square$

Birthdays in a class: first the classic approach

$$p_2 = \frac{1}{365^{30}} \sum_{i=1}^{15} \left[\frac{\prod_{j=0}^{i-1} \binom{30-2j}{2}}{i!} \prod_{k=0}^{29-i} (365-k) \right] \approx 0.678$$

and, after a brave computation, $\mathbb{P}_3 \approx 0.0285$. In the same way,

$$p_3 = \frac{1}{365^{30}} \sum_{i=1}^{10} \frac{\prod_{j=0}^{i-1} \binom{30-3j}{3}}{i!} \left[\prod_{k=0}^{29-2i} (365-k) + \sum_{l=1}^{\left[\frac{30-3i}{2}\right]} \frac{\prod_{m=1}^{l} \binom{30-3i+2-2m}{2}}{l!} \prod_{n=0}^{29-2i-l} (365-n) \right]$$

so that $p_3 \approx 0.027998$ and $\mathbb{P}_4 \approx 5.4 \times 10^{-4}$.

$\mathbb{E}C_3 \approx 0.03047$ *and* $\mathbb{E}C_4 \approx 5.6 \times 10^{-4}$.

Take home message: for the detection of rare events, the computation of the expectation of the event, or Number of False Alarms is much easier than the computation of its probability of appearing, and it brings more information.

The steps to solve a detection problem

- Define the background in which interesting objects will be detected: background model H_0
- A definition of the object to be detected (exceptional under H_0)
- Make an *a priori* count N_{test} of the number of detection tests
- After each test of the presence of the object, compute $P_{H_0}(observed)$
- Deduce the Number of False Alarms of the test, $NFA(\text{observed}) := \#(\text{tests}) \times P_{H_0}(\text{observed})$
- If $NFA(\text{observed})$ is small, detection. Reliability of the detection measured by its NFA.

An aparte: Statistics in the wild, or how to fight illusory detections

Statistics in the wild, or how to fight illusory detections

Observations made by Dr Gastaldi and two other doctors (June 2020):

"For the past few weeks, all three of us have been prescribing this treatment to all our patients with coronavirus. For my part, this represents more than 200 patients. I have only had two serious cases that required hospitalization and have since been discharged. Obviously, this is not a multi-center, randomized study, but these are very interesting results. Based on the known data on the disease, out of at least 200 cases, we should have had at least two deaths and about 40 hospitalizations."

Exercise: finding on the internet the mortality rate among symptomatic patients and the number of medical doctors in France, compute the NFA of this event (> 200 saved patients and no death) and deduce how many such medical « discoveries » may have been done.

https://www.femmeactuelle.fr/sante/news-sante/coronavirus-trois-medecins-generalistes-pensent-avoir-trouve-un-traitement-contre-le-covid-19-2093814

Solution by Florian Laborde

Il y a en France plus de $N_c = 100000$ médecins généralistes. Même s'ils n'ont pas tous une exposition équivalente aux patients on considère qu'ils ont tous un nombre élevé de patients distincts. La crise du Covid 19 a un taux empirique de mortalité d'au moins $t = 2\%$ en France, en considérant que au moins toutes les personnes symptomatiques sont au moins des cas avérés de Covid. Le Dr Gastaldi se base sur une analyse de taille $T = 200$ patients sur lesquels il ne trouve aucune mortalité parmi les asymptomatique traités par ses soins. On se demande si ce test est significatif ? Combien de fois une fausse alarme de ce type avec cette taille d'échantillon est-elle possible en France ? i.e: **Combien y-a-t-il de traitements miracles ?**

$$NFA(n_{medecins}) = (1-t)^T \times (n - 199)$$

En effet, on regarde la probabilité que sur 200 patients consécutifs aucun ne décède. Indépendamment la probabilité de survie est de 98% (1- taux de mortalité). C'est donc une Bernoulli de paramètre $\mathcal{B}(0.98, 200)$ à celà on corrige par le nombre de médecins qui sont suceptibles de faire une expérience similaire $N_c = 100000$. D'où:

$$NFA(N_c) = (1-t)^T \times (N - T - 1) = 0.98^200 \times (10^5 - 199) = 1755$$

Le nombre de fausses alarme est donc d'environ 1755. **Il y a donc environ 1750 médecins en France qui en faisant un traitement quelconque n'ont relevé aucun décès sur un échantillon de 200 patients.** On peut se demander combien de patients il faudrait prendre pour que l'expérience soit pertinente à $\epsilon = 1$ près ? Il faudrait environ **650 patients** sans décès. (Cela pourrait arriver une fois en France statistiquement sans que ce soit significatif).

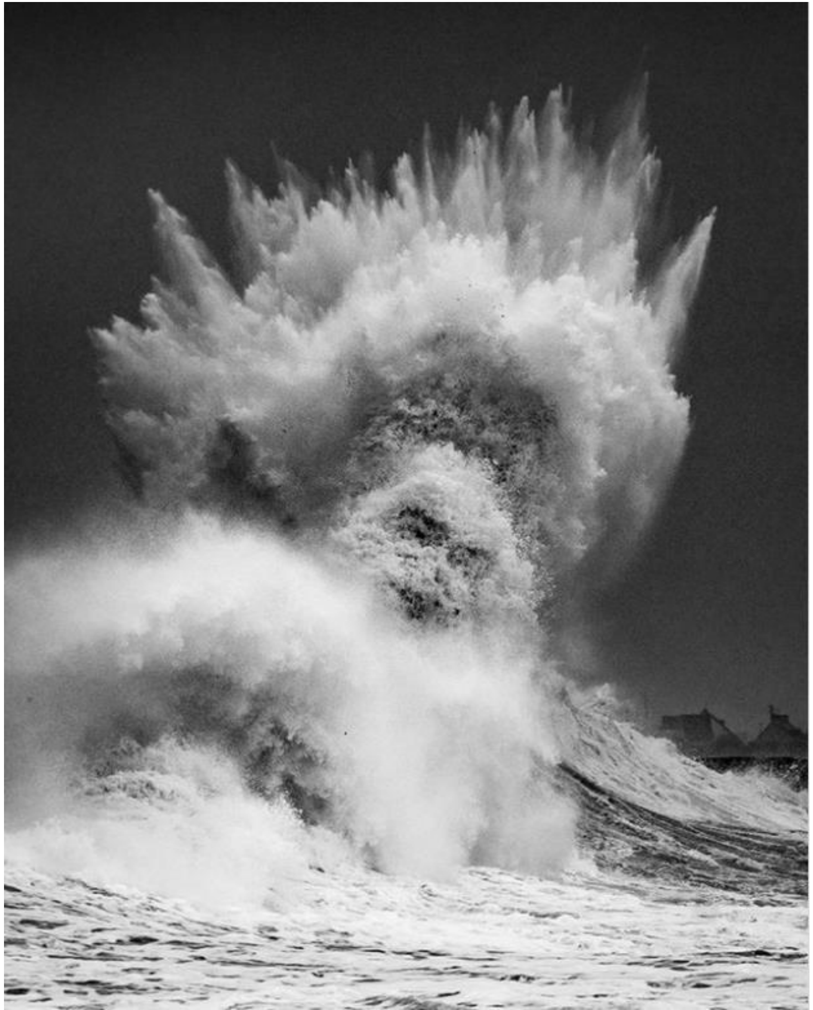
Perception analysis implies making statistics « in the wild » (a posteriori design of the testing set)

Danger of ignoring the number of tests to evaluate a *number of false alarms* (NFA), (also called *per family error rate* (PFER))

Neglecting this fact leads to discover crabs on Mars!

Mars Exploration Rover

Statistics in the wild, or how to fight illusory detections



Danger of ignoring the <u>number of tests</u> to evaluate a *<u>number of false alarms</u>* (NFA), (also called *per family error rate* (PFER))

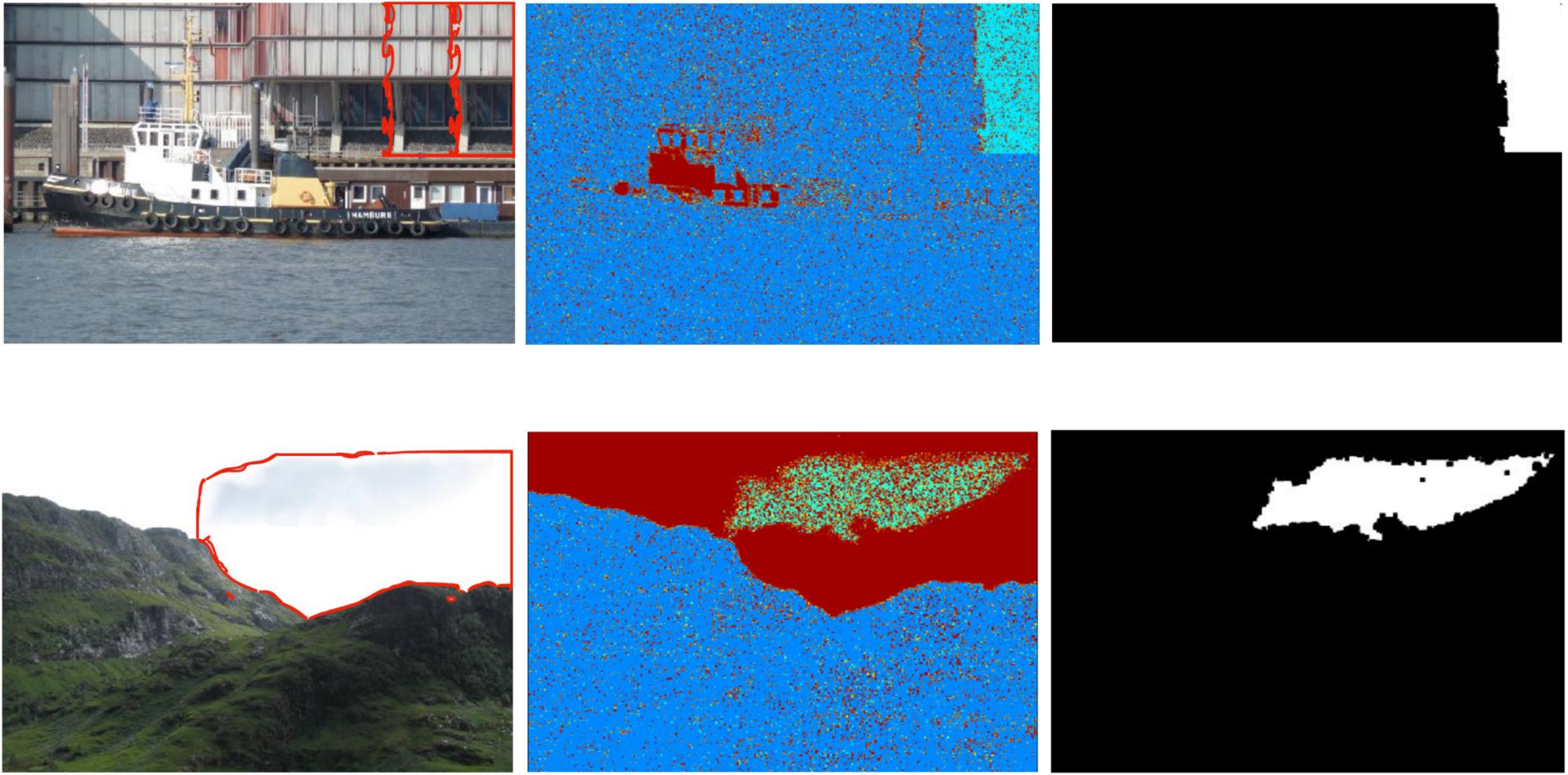
Neglecting this fact leads to discover <u>gods in the ocean!</u>

The image by photographer Mathieu Rivrin taken on January 30, 2021 shows the storm Justine with a face that could be that of the god of Greek mythology Poseidon (Neptune for the Romans).

First real example : image forgery detection

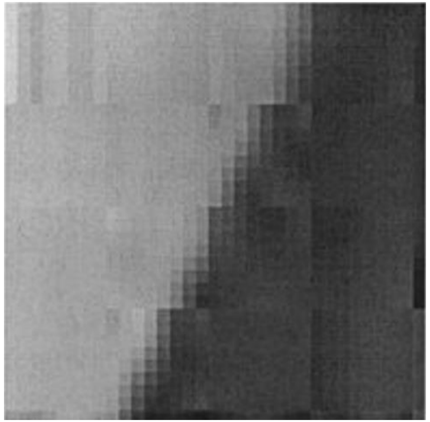
(Fake news debunking, work in collaboration with Agence France Presse)

Forgery detection



The cue to forgery detection is the number of zeros in a JPEG bloc. Each digital image is divided in 8x8 blocs. The high frequencies in each bloc are put to zero by JPEG: this allows one to retrieve the position of the blocs and therefore the original JPEG grid.

But if the image has been manipulated in parts, the JPEG grid will generally be shifted. Thus forgery detection amounts to find clusters of blocs where the grid is not aligned with the general grid.

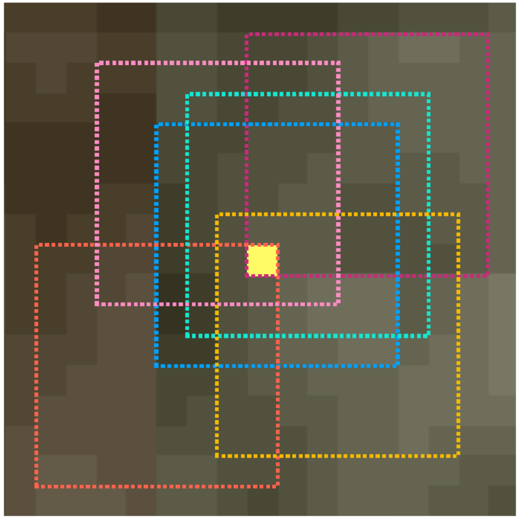


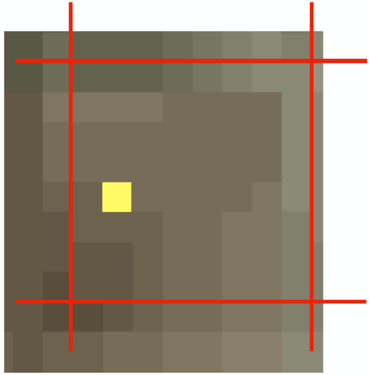
$$B = \begin{bmatrix} -26 & -3 & -6 & 2 & 2 & -1 & 0 & 0 \\ 0 & -2 & -4 & 1 & 1 & 0 & 0 & 0 \\ -3 & 1 & 5 & -1 & -1 & 0 & 0 & 0 \\ -3 & 1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

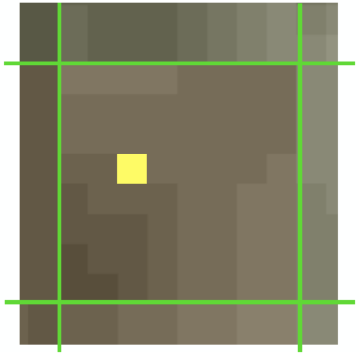
Forgery detection

- *Background model:* each pixel votes for one of the 64 possible JPEG grids. At taxi driver distance larger than 8 pixels, the votes are i.i.d. uniform with probability 1/64 for each grid
- *Event :* in a window containing n votes, more than k points distant by more than 8 from each other, vote for the same grid
- *Number of tests*= #(tested windows) $\times$ (#(block sizes)

Selecting the grid by the number of zeros







The number of zeros is larger when a 8x8 block is aligned with a previous JPEG compression grid

The tail of the binomial law

Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli variables, that is, independent variables such that $X_j \in \{0,\ 1\}$ and

$$\mathbb{P}\left[X_j = 1\right] = p, \quad \mathbb{P}\left[X_j = 0\right] = 1 - p.$$

We set $S_n = X_1 + \cdots + X_n$. Then for $0 \leq k \leq n$

$$\mathbb{P}\left[S_n = k\right] = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}.$$

This law is called the binomial law with size n and parameter p and we denote by $\mathcal{B}(n, p)$ the probability distribution defined by

$$p_k = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}.$$

The tail of the binomial law

The mean and variance of the binomial law $\mathcal{B}(n,p)$ are respectively np and $np(1-p)$. We now consider the "tail of the binomial law"

$$\mathcal{B}(n,k,p) = \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i}.$$

This tail measures the probability that, among n tests, the observed number of variables X_i with value 1 is above k. As soon as k exceeds significantly its expectation np, we can expect a detection.

The tail of the binomial law

Let us suppose that we are observing a square patch of an image where the number of votes for a given valid grid origin is counted at a distance of eight pixels. Let us say that k votes are counted for that valid grid among a total of n votes. Under the null hypothesis H_0, votes for the given grid origin become Bernoulli random variables with probability $\frac{1}{64}$. So under H_0, the number of votes becomes a random variable K and, given the independence of votes (at distance larger than eight), it follows a binomial distribution of parameter $p = \frac{1}{64}$. Thus,

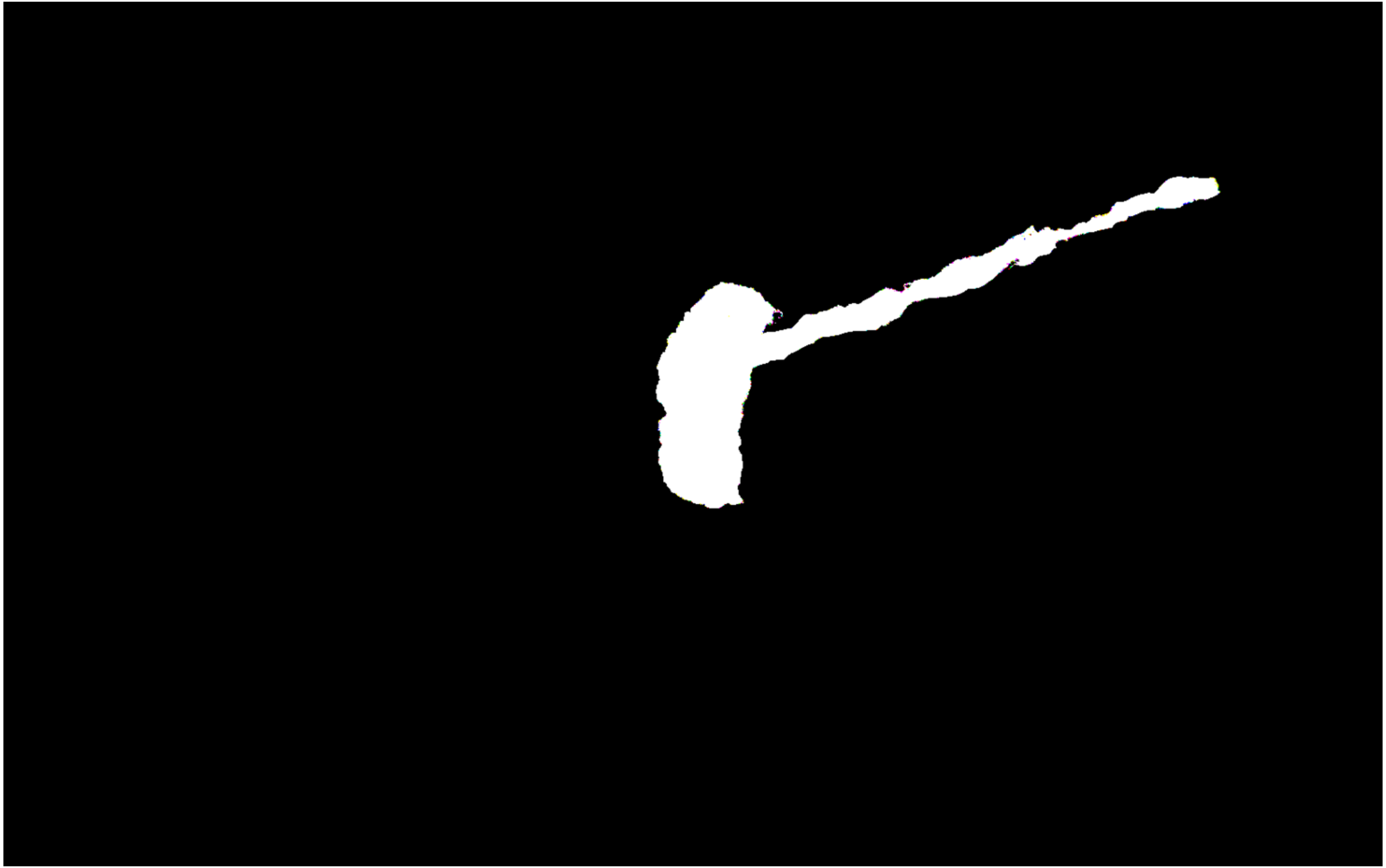
$$\mathbb{P}(K \geq k) = \mathcal{B}(n, k, p) = \sum_{j=k}^{n} \binom{n}{j} p^j (1-p)^{n-j},$$

Thus, every square window of a $X \times Y$ pixels image is included in the family of tests and the 64 grid origins are tested for each one. Then, the number of tests can be approximated by

$$N_T = 64 \cdot XY \cdot \sqrt{XY}, \tag{10.3}$$

where $\sqrt{XY}$ gives a rough estimation of the possible window sides, and XY gives the number of possible positions for the square windows of a given size. All in all, given a window to be analyzed, the grid origin with the maximum of votes is selected and its number of votes at distance eight pixels is counted. Then, the NFA is given by

$$\mathrm{NFA} = 64 \cdot XY \cdot \sqrt{XY} \cdot \mathcal{B}(n, k, p). \tag{10.4}$$



NFA= 2.E-86

Test the method here:

https://ipolcore.ipol.im/demo/clientApp/demo.html?id=77777000073

Real example 2: LSD, Line Segment Detector

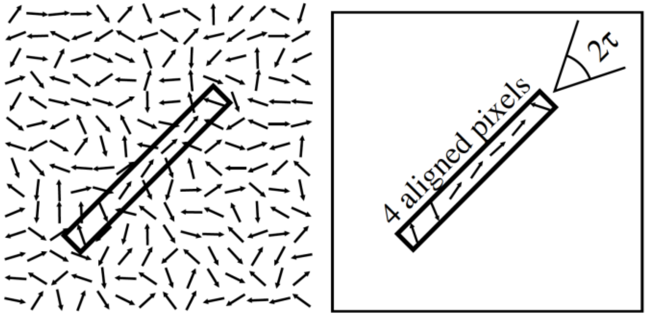
交通大学医学院附属新华医院）病人费用清单

:00 至 2014-07-17 08:30:00　　总费用：10379.33元
费别：社保病人　　预交金总额：2500.00元
姓名：刘彦　　科室：耳鼻喉颈外科　　床位号：24
打印时间：2014.07.17 08:12:44　　共2页，第1页

别	名称	代码	折扣	单价	数量	单	医保范	自费金	分类自	总金额
合计:					211.		5255.29	5056.90	67.14	10379.33
护理	小计:									68.00
护理	[甲]Ⅰ级护理	1201000		14.00	2.0	日	28.00			28.00
护理	[甲]Ⅱ级护理	1201000		12.00	2.5	日	30.00			30.00
护理	[甲]动静脉置管护理	1201000		5.00	2.0	日	10.00			10.00
化验	小计:									280.00
化验	[甲]活化部分凝血活酶时	2502030		15.00	1.0		15.00			15.00
化验	[甲]肌酐测定/干化学法	2503070		15.00	1.0		15.00			15.00
化验	[甲]尿素测定/干化学法	2503070		15.00	1.0		15.00			15.00
化验	[甲]尿液分析(9联及以	2501020		10.00	1.0		10.00			10.00
化验	[甲]凝血酶时间测定(TT)	2502030		15.00	1.0		15.00			15.00
化验	[甲]葡萄糖测定/干化学	2503020		15.00	1.0		15.00			15.00
化验	[甲]血浆凝血酶原时间测	2502030		15.00	1.0		15.00			15.00
化验	[甲]血浆纤维蛋白原测定	2502030		10.00	1.0		10.00			10.00
化验	[甲]血清r-谷氨酰转肽	2503050		15.00	1.0		15.00			15.00
化验	[甲]血清白蛋白测定/干	2503010		15.00	1.0		15.00			15.00
化验	[甲]血清丙氨酸氨基转移	2503050		15.00	1.0		15.00			15.00
化验	[甲]血清碱性磷酸酶测	2503050		15.00	1.0		15.00			15.00
化验	[甲]血清尿酸测定/干化	2503070		15.00	1.0		15.00			15.00
化验	[甲]血清天门冬氨酸氨基	2503050		15.00	1.0		15.00			15.00
化验	[甲]血消直接胆红素测	2503050		15.00	1.0		15.00			15.00
化验	[甲]血消总胆红素测定/	2503050		15.00	1.0		15.00			15.00
化验	[甲]血消总胆汁酸测定	2503050		15.00	1.0		15.00			15.00
化验	[甲]血消总蛋白测定/干	2503010		15.00	1.0		15.00			15.00
化验	[甲]血细胞分析/全血计	2501010		20.00	1.0		20.00			20.00
检查	小计:									102.00
检查	[甲]电脑多导联心电图	3107010		20.00	1.0		20.00			20.00
检查	[甲]手术标本检查与诊断	2703000		80.00	1.0		80.00			80.00
检查	[甲]血氧饱和度监测	3107010		2.00	1.0	小	2.00			2.00
麻醉	小计:									1125.00
麻醉	[甲]麻醉后复苏监护	3301000	100.0	280.0	1.0	次	280.00			280.00
麻醉	[甲]麻醉中监测　4.10导	3301000	100.0	45.00	1.0	小	45.00			45.00
麻醉	[甲]全身麻醉　4.10导入	3301000	100.0	800.0	1.0	次	800.00			800.00
拍片	小计:									70.00
拍片	[甲]数字化摄影	2101020		70.00	1.0	部	70.00			70.00
手术	小计:									1380.00
手术	[甲]腭咽成形术　4.10导	3306060	100.0	1380.	1.0	次	1380.00			1380.00
手术器	小计:									748.13
手术器	等离子体手术系统耳鼻喉	c0105-	100.0	2992.	0.25	支		748.13		748.13
手术特	小计:									100.00
手术特	[甲]双电极电凝器　4.10	s0026	100.0	100.0	1.0	次	100.00			100.00
输氧	小计:									4.00
输氧	[甲]加压给氧(指呼吸机	1203000	100.0	4.00	1.0	小	4.00			4.00
西药	小计:									5616.30
西药	◇丙帕他莫针　1g	0901244	100.0	85.00	2.0	支		170.00		170.00

[illegible]


粤B·B777G
206 CC



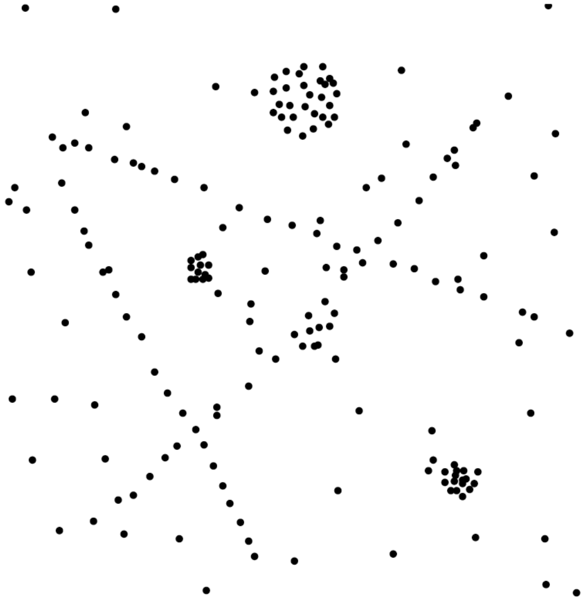
Straight edges in images

- Background model: the gradient orientation at each pixel is i.i.d. uniform on $[0, 2\pi]$
- Event: in a rectangle r, at least k among n pixels are aligned up to some precision p with the vector normal to the rectangle
- Number of tests: number of pairs of pixels in an image : $(MN)^2$
- $NFA(n,k,p) = (MN)^2\mathcal{B}(n,k,p) = (MN)^2 \sum_{l=k}^{n} C_n^l p^l (1-p)^{n-l}$

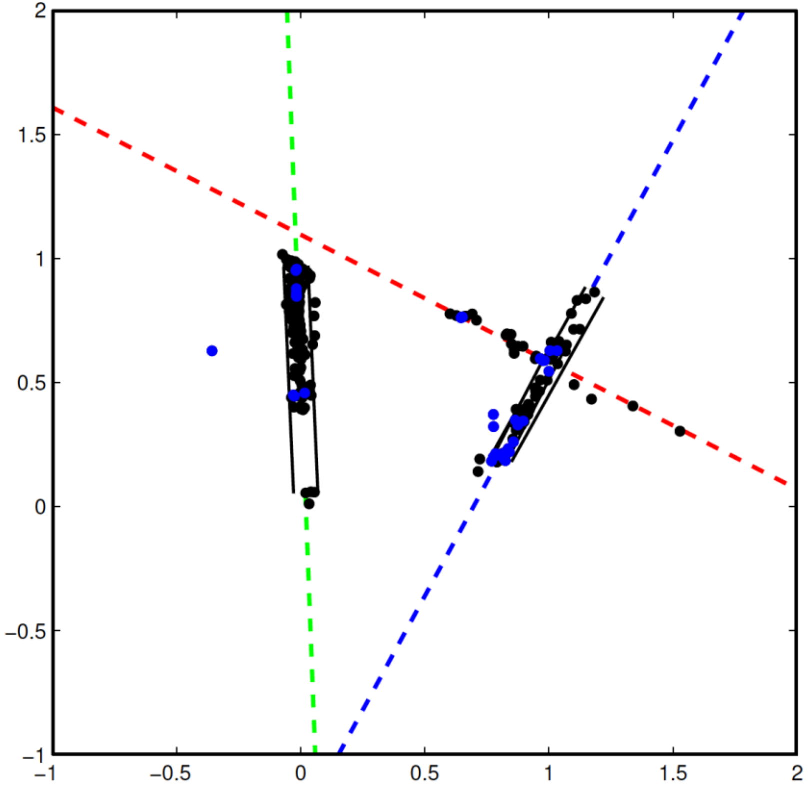
[Read the paper, have the code and test the online demo on any image here]

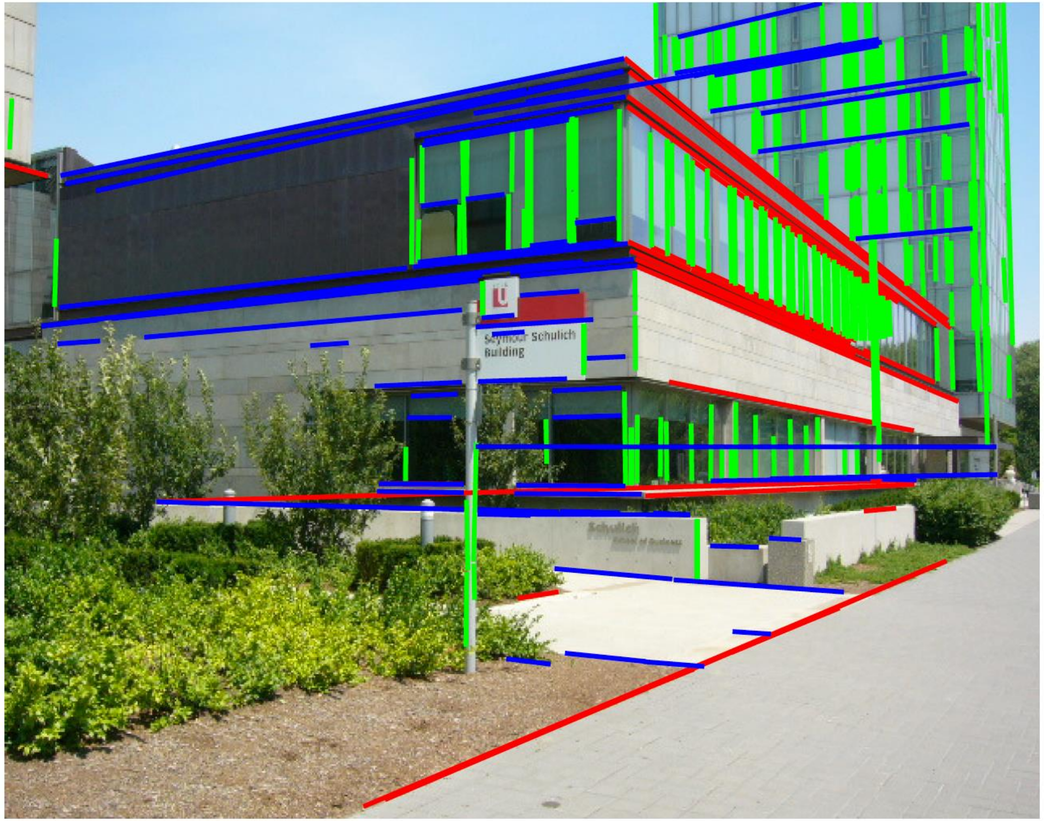
[LSD: a Line Segment Detector]
2012-03-24 · Rafael Grompone von Gioi, Jérémie Jakubowicz, Jean-Michel Morel, Gregory Randall

Real example 3: Detection of dot alignments



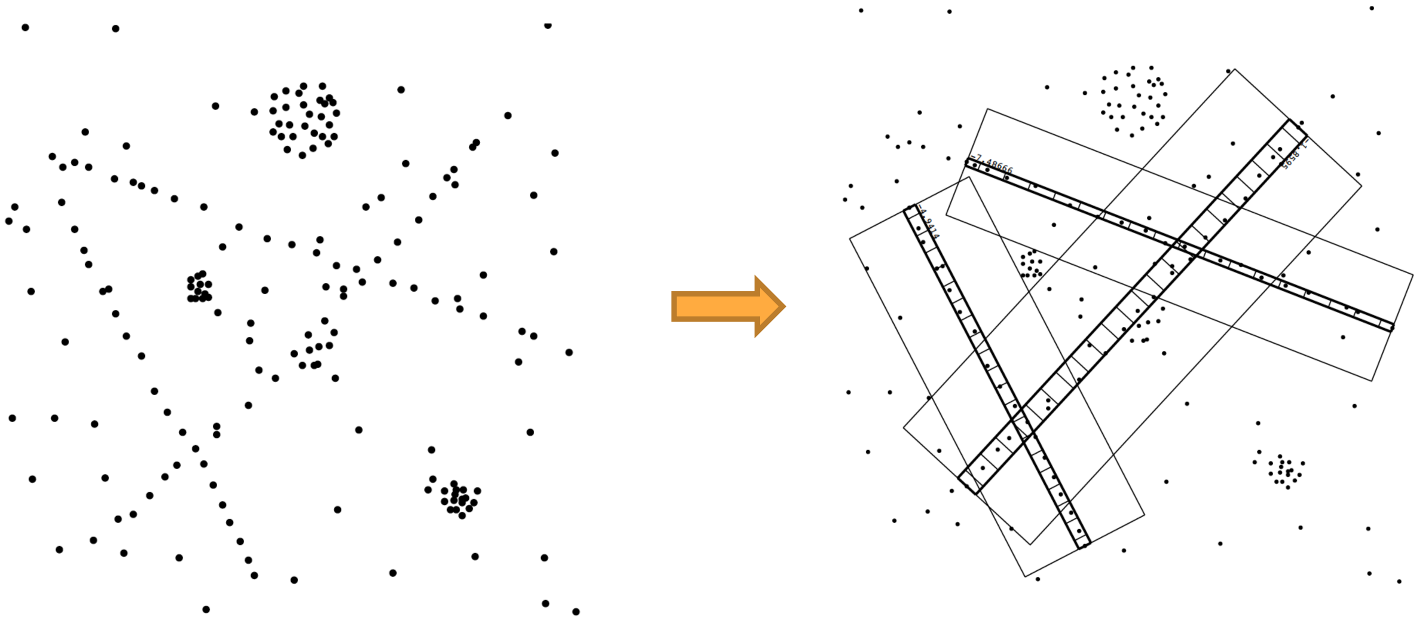
Real example 3: Detection of dot alignments: Using first LSD and then alignment of lines (which are dots in the dual space) leads to the detection of vanishing points and of the horizon





Dot alignments

- Background model: in each tested rectangle, the dots are uniformly spread (uniform Poisson point process)
- Event : in a rectangle r of some width divided in c uniform cells, at least $b(r, c, x)$ cells are occupied by some dot
- number of tests = #(pairs of dots)× #(widths)× #(cell aspect ratios)



Dot alignments

- Background model: in each tested rectangle, the dots are uniformly spread (uniform Poisson point process)
- Event : in a rectangle r of some width divided in c uniform cells, at least $b(r, c, x)$ cells are occupied by some dot
- number of tests = #(pairs of dots)× #(widths)× #(cell aspect ratios)

$$N_{tests} = \frac{N(N-1)}{2} WLC = \frac{N(N-1)}{2} WL\sqrt{N}. \tag{1.11}$$

The NFA of the new event definition is then

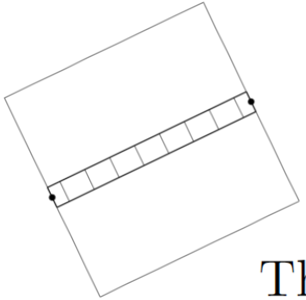
$$\begin{aligned} \mathrm{NFA}_4(r, R, c, \mathbf{x}) &= N_{tests} \cdot \mathbb{P}\Big[b(r, c, \mathbf{X}) \geq b(r, c, \mathbf{x}) \,\Big|\, n(R, \mathbf{X}) = n^*(R, \mathbf{x})\Big] \\ &= \frac{N(N-1)}{2} WLC \cdot \mathcal{B}\big(c, b(r, c, \mathbf{x}), p_1(R, c)\big). \end{aligned} \tag{1.12}$$

A set $\mathcal{C}$ of different values are tried for the number of boxes c into which the rectangle is divided, and the one producing the lowest NFA is taken. Thus, the number of tests must be multiplied by its cardinality $\#\mathcal{C} = C$. In practice we set $C = \sqrt{N}$ and that leads to

$$N_{tests} = \frac{N(N-1)}{2} WLC = \frac{N(N-1)}{2} WL\sqrt{N}. \tag{1.11}$$

The NFA of the new event definition is then

$$\begin{aligned} \mathrm{NFA}_4(r, R, c, \mathbf{x}) &= N_{tests} \cdot \mathbb{P}\Big[b(r, c, \mathbf{X}) \geq b(r, c, \mathbf{x}) \,\Big|\, n(R, \mathbf{X}) = n^*(R, \mathbf{x}) \Big] \\ &= \frac{N(N-1)}{2} WLC \cdot \mathcal{B}\big(c, b(r, c, \mathbf{x}), p_1(R, c)\big). \end{aligned} \tag{1.12}$$

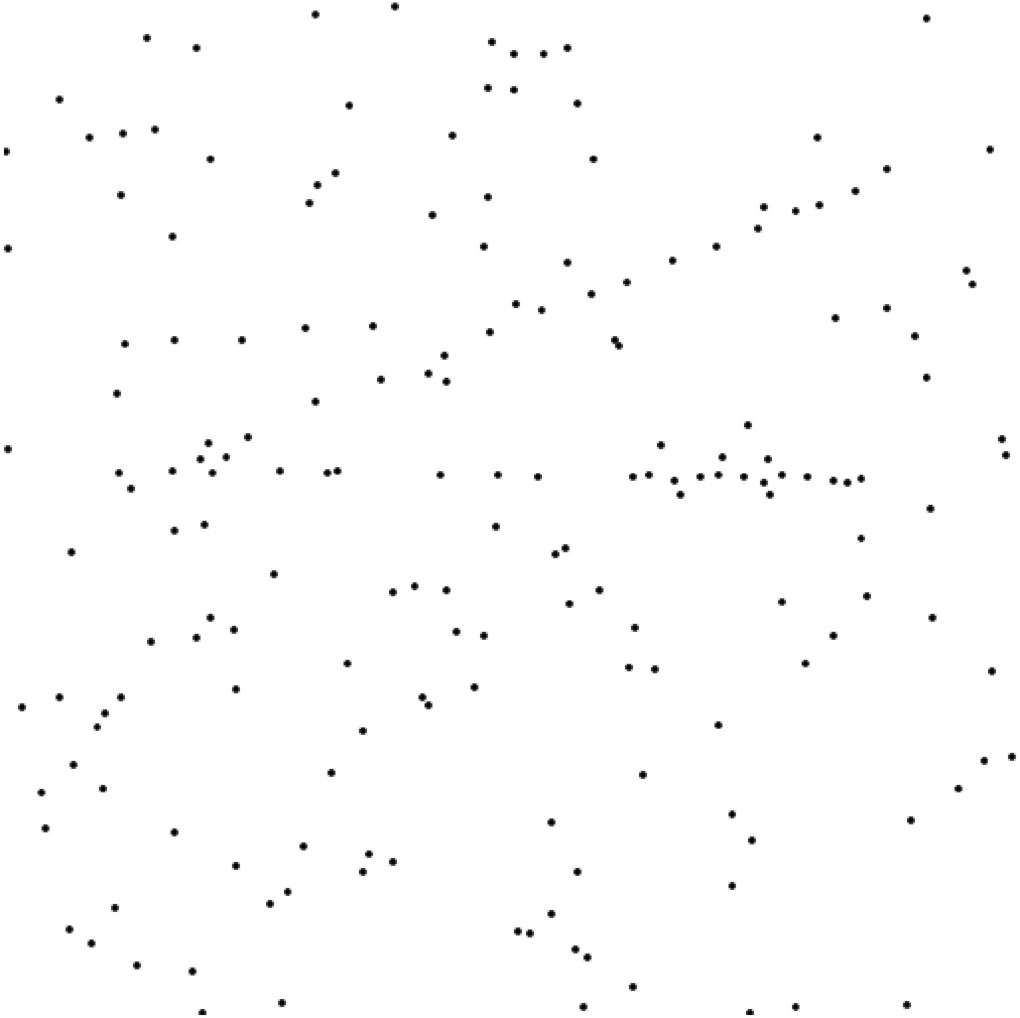


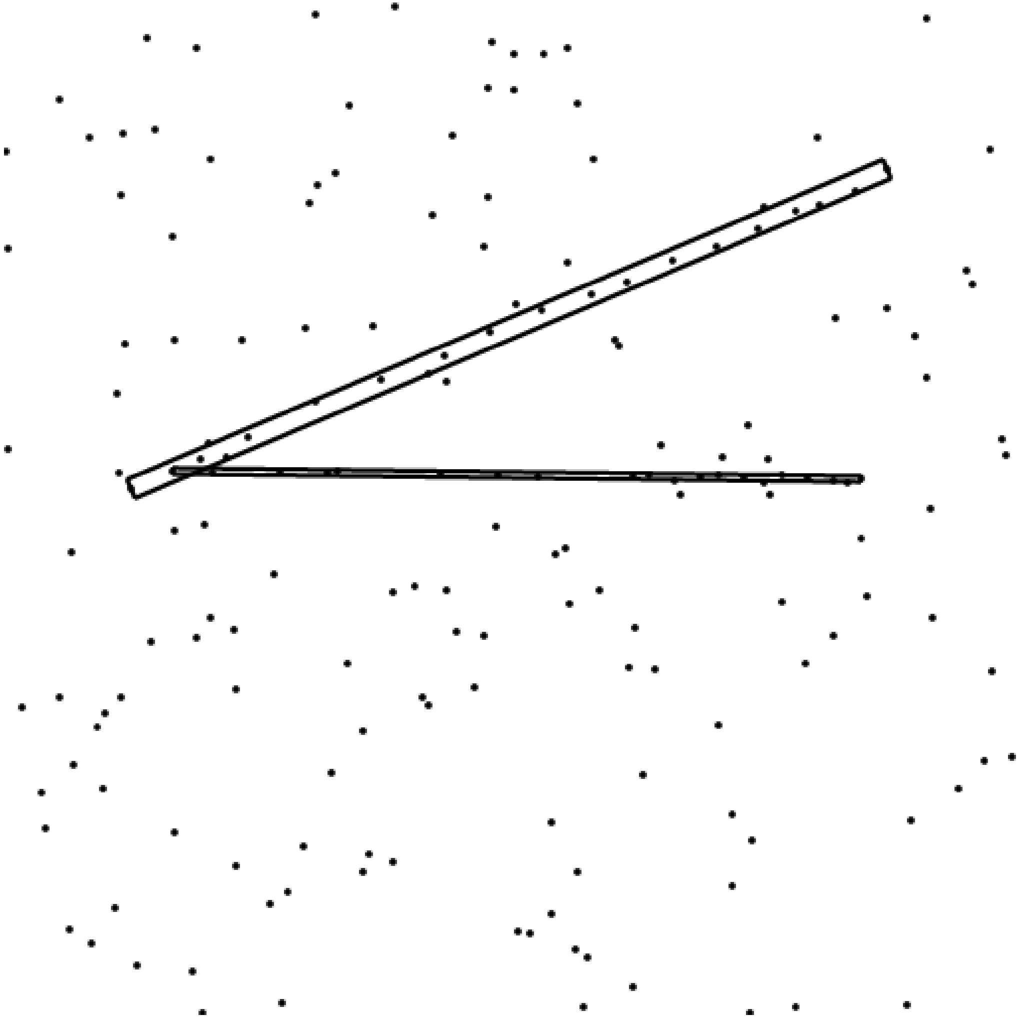
The probability of one point falling in one of the boxes is $p_0 = \frac{S_B}{S_L}$, where S_B and S_L are the areas of the boxes and the local window respectively. Then, the probability of having one box occupied by at least one of the $n^*(R, \mathbf{x})$ points (i.e., of an *occupied* box) is

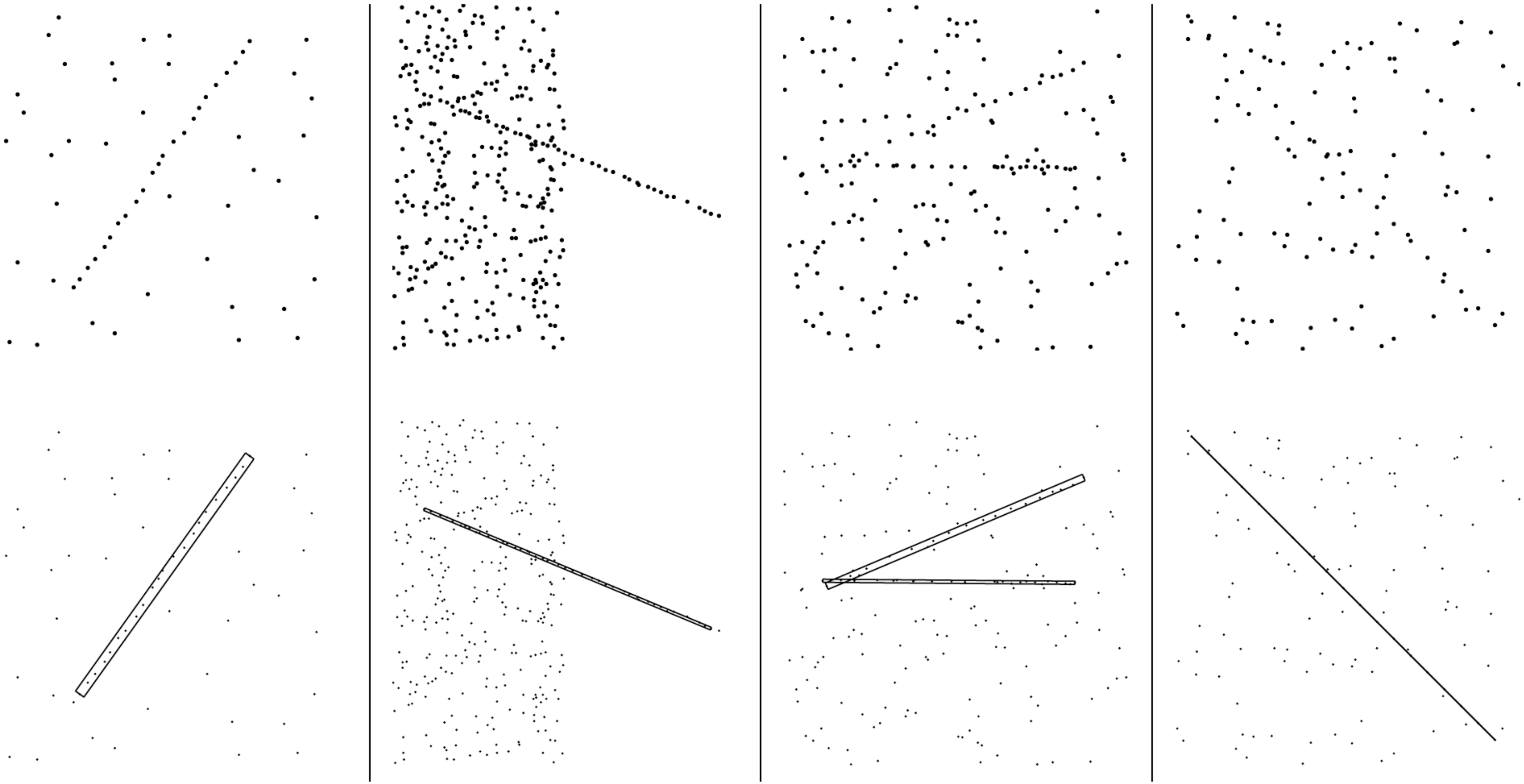
$$p_1(R, c) = 1 - (1 - p_0)^{n^*(R,\mathbf{x})}. \tag{1.9}$$

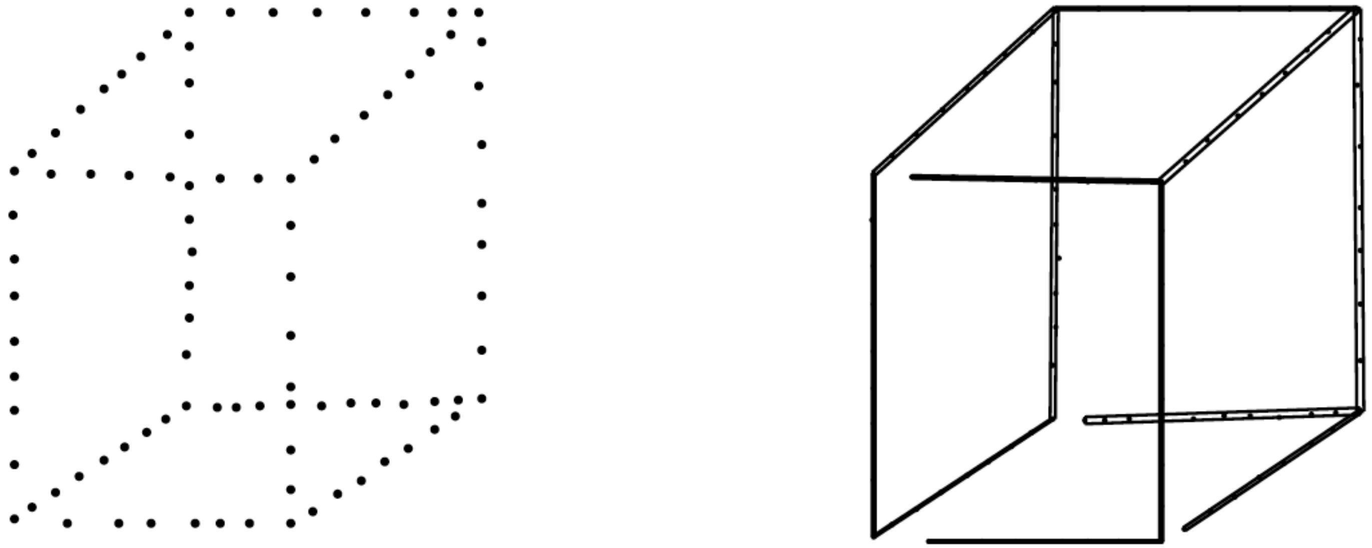
We will denote by $b(r, c, \mathbf{x})$ the observed number of occupied boxes in the rectangle r when divided into c boxes. Finally, the probability of having at least $b(r, c, \mathbf{x})$ of the c boxes occupied is

$$\mathcal{B}\big(c, b(r, c, \mathbf{x}), p_1(R, c)\big). \tag{1.10}$$









Read the paper, have the code and test the online demo on any image here

An Unsupervised Point Alignment Detection Algorithm (2015), www.ipol.im
José Lezama, Gregory Randall, J.M.M., Rafael Grompone von Gioi

Theory:

A general definition of NFA

How to estimate the binomial tail

The interpretation of multiple detections : nonmaxima suppression

A general definition of NFA

Definition 2. *[73] Given a set of random variables $(X_i)_{i\in[1,N]}$ with known distribution under a null-hypothesis $(\mathcal{H}_0)$, a multi-test function $f(i,x)$ is called an NFA if it guarantees a bound on the expectation of its number of false alarms under $(\mathcal{H}_0)$, namely:*

$$\forall \varepsilon > 0, \mathbb{E}[\#\{i, f(i,X_i) \leq \varepsilon\}] \leq \varepsilon.$$

To put it in words, raising a detection every time the test function is below ε should give under $(\mathcal{H}_0)$ an expectation of less than ε false alarms. An observation $\mathbf{x}_i$ is said to be "ε-meaningful" if it satisfies $f(i,\mathbf{x_i}) \leq \varepsilon$, where ε is the predefined target for the expected number of false alarms. The lower $f(i,\mathbf{x})$ the "stronger" the detection.

[73] A-contrario detectability of spots in textured backgrounds B Grosjean, L Moisan
Journal of Mathematical Imaging and Vision 33 (3), 313-337

A general definition of NFA

A common way to build an NFA is to take

$$f(i, \mathbf{x_i}) = N\mathbb{P}_{\mathcal{H}_0}(X_i \geq \mathbf{x}_i) \tag{2.3}$$

or

$$f(i, \mathbf{x_i}) = N\mathbb{P}_{\mathcal{H}_0}(|X_i| \geq |\mathbf{x}_i|), \tag{2.4}$$

where N is the number of tests, i goes over all tests, and $\mathbf{x}_i$ are the observations which excess should raise an alarm. These test functions are typically used when anomalies are expected to have higher values than the background in the first case, or when anomalies are expected to have higher modulus than the background. If for example the (X_i) represent the pixels of an image, there would be one test per pixel and per channel. Hence N would be the product of the image dimension by the number of image channels.

A general definition of NFA

Proposition 4. *Let* $(X_i)_{1\le i\le N}$ *be a set of random variables and* $(n_i)_{1\le i\le N}$ *a set of positive real numbers. Then the function* $NFA(i,x_i) = n_i \cdot \mathbb{P}(X_i \ge x_i)$ *is an NFA as soon as* $\sum_{i=1}^{N} \frac{1}{n_i} \le 1$, *and in particular if* $n_i = N$ *for all* i. *As a consequence, for any multi-dimensional random vector* X_i, *if* $g(X)$ *is any real valued function, then the function*

$$f(i,x_i) = N\mathbb{P}_{\mathcal{H}_0}(g(X_i) \ge x_i) \tag{1.7}$$

is an NFA.

$$\forall \varepsilon > 0, \mathbb{E}[\#\{i, f(i,X_i) \le \varepsilon\}] \le \varepsilon$$

Let us interpret Proposition 4. It basically says that the multiple test $\mathrm{NFA}(i, X_i) \leq \varepsilon$, $1 \leq i \leq N$, is controlled by ε, in the sense that under H_0 (naive model) there is no more than ε false detections on average. It can be related to the Bonferroni strategy for multiple tests in the following way. Since each test $\mathrm{NFA}(i, X_i) \leq \varepsilon$ has a confidence level $\alpha_i = \frac{\varepsilon}{n_i}$, (as a consequence of Lemma 1), the probability of having at least one false alarm is

$$\alpha = \mathbb{P}(\exists i, \mathrm{NFA}(i, X_i) \leq \varepsilon) \leq \sum_{i=1}^{N} \mathbb{P}(\mathrm{NFA}(i, X_i) \leq \varepsilon) \leq \sum_{i=1}^{N} \alpha_i \leq \varepsilon. \tag{2.8}$$

Estimating the binomial tail

$$NFA(l,k,p) = N_{test} \cdot \mathbb{P}\left[S_l \geq k\right],$$

where

$$S_l = \sum_{i=1}^{l} X_i$$

X_i are independent Bernoulli random variables with parameter p

$$\mathcal{B}(l,k,p) = \sum_{i=k}^{l} \binom{l}{i} p^i (1-p)^{l-i}.$$

In this estimation problem, p is fixed, l is rather large, and k in excess with respect to its expected value pl since we look for meaningful events. The number of tests usually will be very large and the NFA is interesting mainly when it is smaller than 1; so we are primarily interested in good estimates of $\mathcal{B}(l,k,p)$ when this quantity is very small. There are several tools available to do so.

Estimating the binomial tail

Proposition 9. *Let X_i, $i = 1, \dots, l$ be independent Bernoulli random variables with parameter $0 < p < \frac{1}{4}$ and let $S_l = \sum_{i=1}^{l} X_i$. Consider a constant $p < r < 1$ or a real function $p < r(l) < 1$. Then $\mathcal{B}(l, k, p) = \mathbb{P}\left[S_l \geq k\right]$ satisfies*

$$\text{(Slud)} \quad -\tfrac{1}{l} \log \mathbb{P}\left[S_l \geq rl\right] \leq \frac{(r-p)^2}{p(1-p)} + O\left(\frac{\log l}{l}\right), \tag{3.2}$$

$$\text{(Hoeffding-bis)} \quad -\tfrac{1}{l} \log \mathbb{P}\left[S_l \geq rl\right] \geq (r-p)^2 \frac{\log \frac{1-p}{p}}{1-2p} + O\left(\frac{1}{l}\right), \tag{3.3}$$

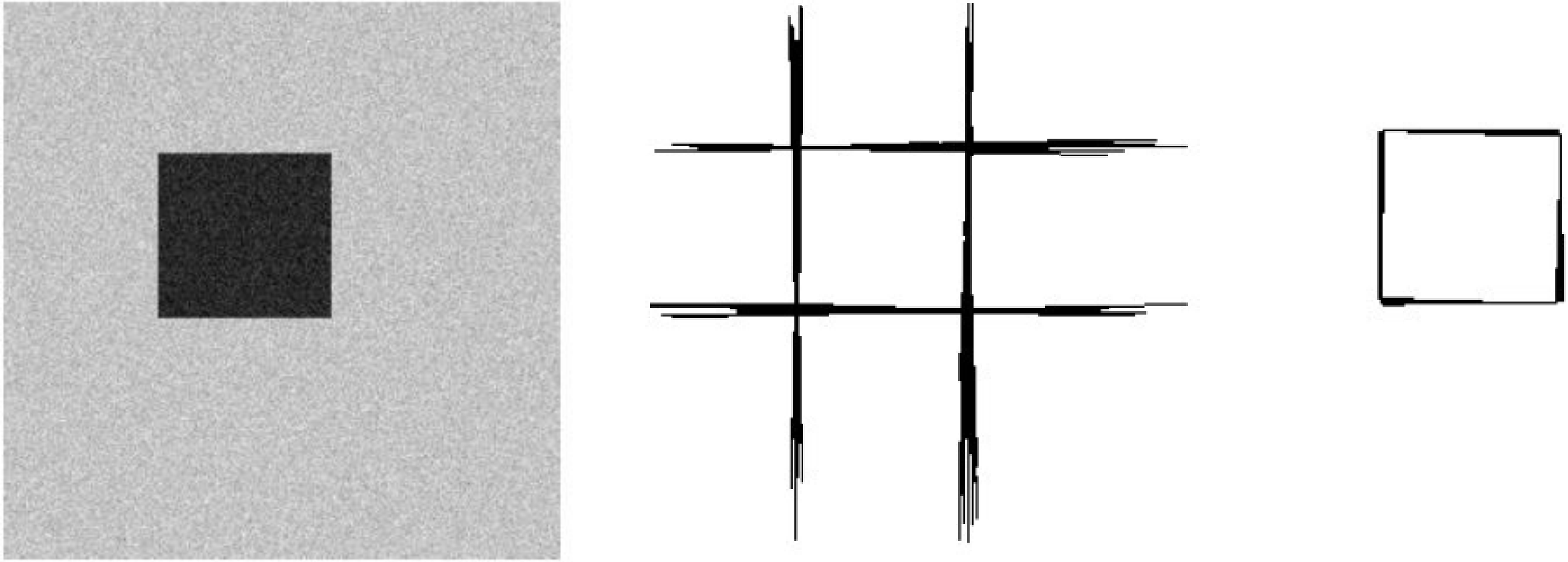
$$\text{(Central limit)} \quad -\tfrac{1}{l} \log \mathbb{P}\left[S_l \geq r(l)l\right] \sim \frac{(r(l)-p)^2}{p(1-p)} \quad \text{if} \quad (r(l)-p)l^{\frac{1}{3}} \xrightarrow[l\to\infty]{} 0, \tag{3.4}$$

$$\text{(Hoeffding)} \quad -\tfrac{1}{l} \log \mathbb{P}\left[S_l \geq rl\right] \geq r \log \frac{r}{p} + (1-p) \log \frac{1-r}{1-p}, \tag{3.5}$$

$$\text{(Large deviation)} \quad -\tfrac{1}{l} \log \mathbb{P}\left[S_l \geq rl\right] \sim r \log \frac{r}{p} + (1-p) \log \frac{1-r}{1-p}, \tag{3.6}$$

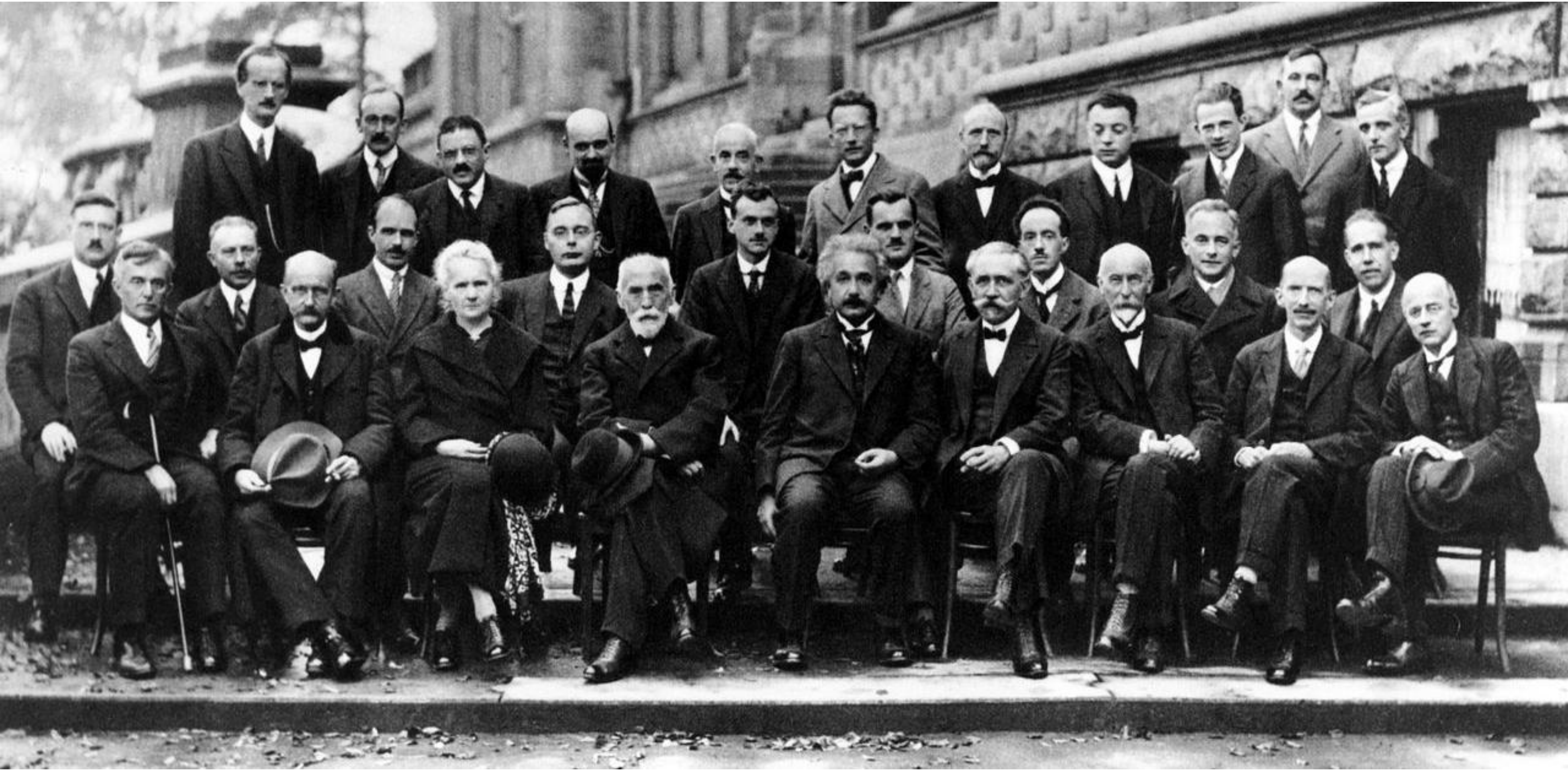
where the last equivalence holds when r is fixed and l tends to infinity.

Non maxima suppression of multiple detections



Noisy square, meaningful alignments, maximal meaningful alignments

Non maxima suppression of multiple detections



An Analysis of the Viola-Jones Face Detection Algorithm, Yiqing Wang IPOL

Non maxima suppression of multiple detections

An Analysis of the Viola-Jones Face Detection Algorithm, Yiqing Wang IPOL

Non maxima suppression of multiple detections

An Analysis of the Viola-Jones Face Detection Algorithm, Yiqing Wang IPOL

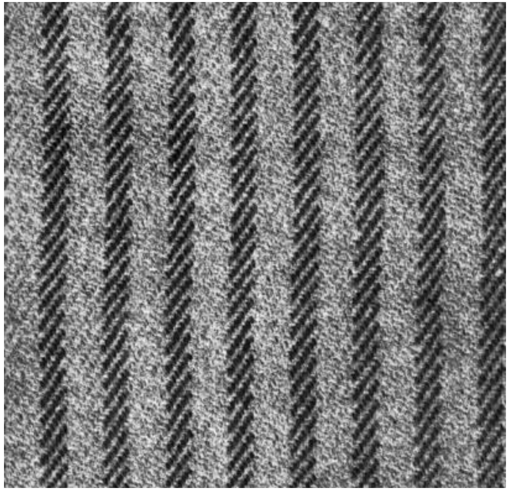
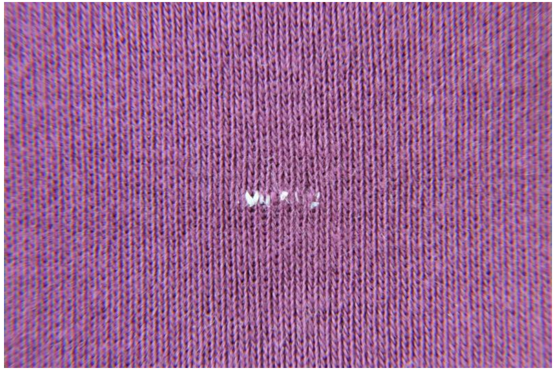
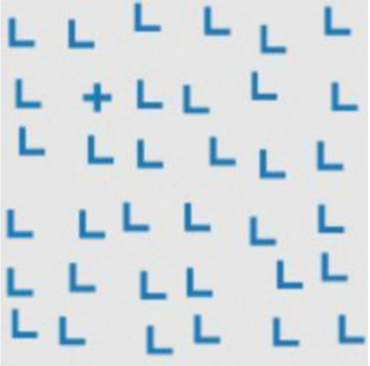
Anomaly detection in any image

Anomaly detection in any image

- *Background model:* Here, huge variety of (non uniform) backgrounds. Hence, detection made after *background subtraction* in the *residual noise image*. This noise image becomes the background model.
- *Event:* a block in which all pixels have a residual larger than $c \times \sigma$, where σ is the std of the residual noise (Reed-Xiaoli)
- *Number of tests* = #(pixels)×#(scales)

How to Reduce Anomaly Detection in Images to Anomaly Detection in Noise
2019-12-08 · Thibaud Ehret, Axel Davy, Mauricio Delbracio, Jean-Michel Morel

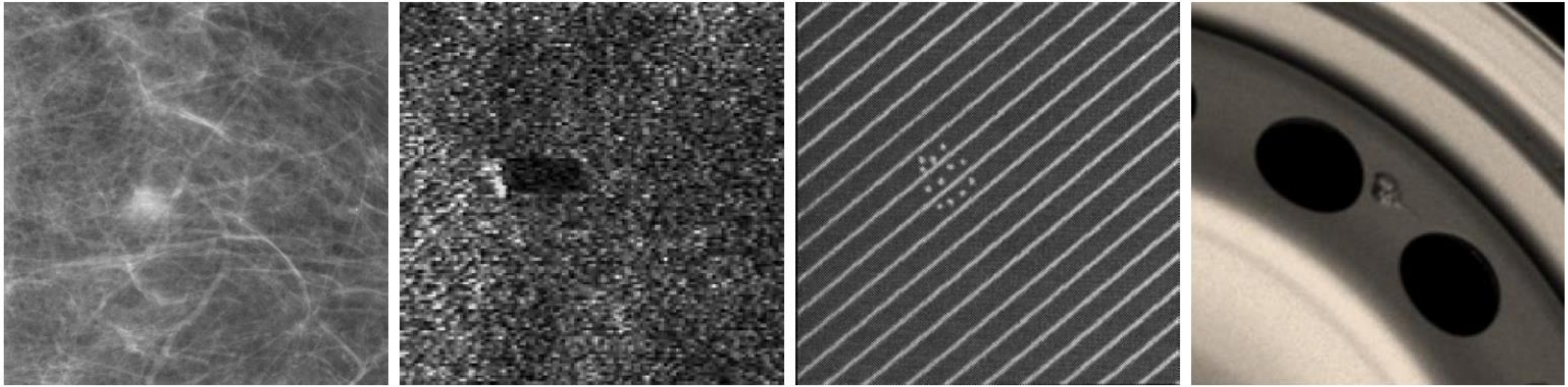
Anomaly detection in any image


“**Novelty (or anomaly) detection** is the task of classifying test data that differ in some respect from the data that are considered “normal”**.** This may be seen as “one-class classification”, in which a model is constructed to describe “normal” data. The novelty detection approach is necessary because the quantity of available “abnormal” data is insufficient to construct explicit models for non-normal classes. **Detection must work even in a single image with a single anomaly.”**

A review of novelty detection (2014) Marco A.F. Pimentel, David A. Clifton, Lei Clifton, Lionel Tarassenko

Anomaly detection in any image



Examples of industrial images with anomalies to detect. From left to right a suspicious mammogram, an undersea mine, a defective textile pattern and a defective wheel

“Novelty (or anomaly) detection is the task of classifying test data that differ in some respect from the data that are considered “normal”. This may be seen as “one-class classification”, in which a model is constructed to describe “normal” data. The novelty detection approach is necessary because the quantity of available “abnormal” data is insufficient to construct explicit models for non-normal classes. <u>**Detection must work even in a single image with a single anomaly.”**</u>

A review of novelty detection (2014) Marco A.F. Pimentel, David A. Clifton, Lei Clifton, Lionel Tarassenko

- Decompose the image u into all of its 8×8 patches P
- Find for each patch P the 16 **most similar** patches P_i, $i = 1, \ldots, 16$ (located elsewhere in the image)
- Find the **best estimate** of P from the P_i according to
$$\hat{P} = \frac{1}{Z} \sum_{i=1}^{n} \exp\left(-\frac{\|P - P_i\|_2^2}{h^2} \right) P_i \tag{1}$$

Anomaly detection in any image

- Reconstruct an **normal image model** $\tilde{u}$ by aggregating all patch estimates (a simple mean)

Self-similar part scale 0

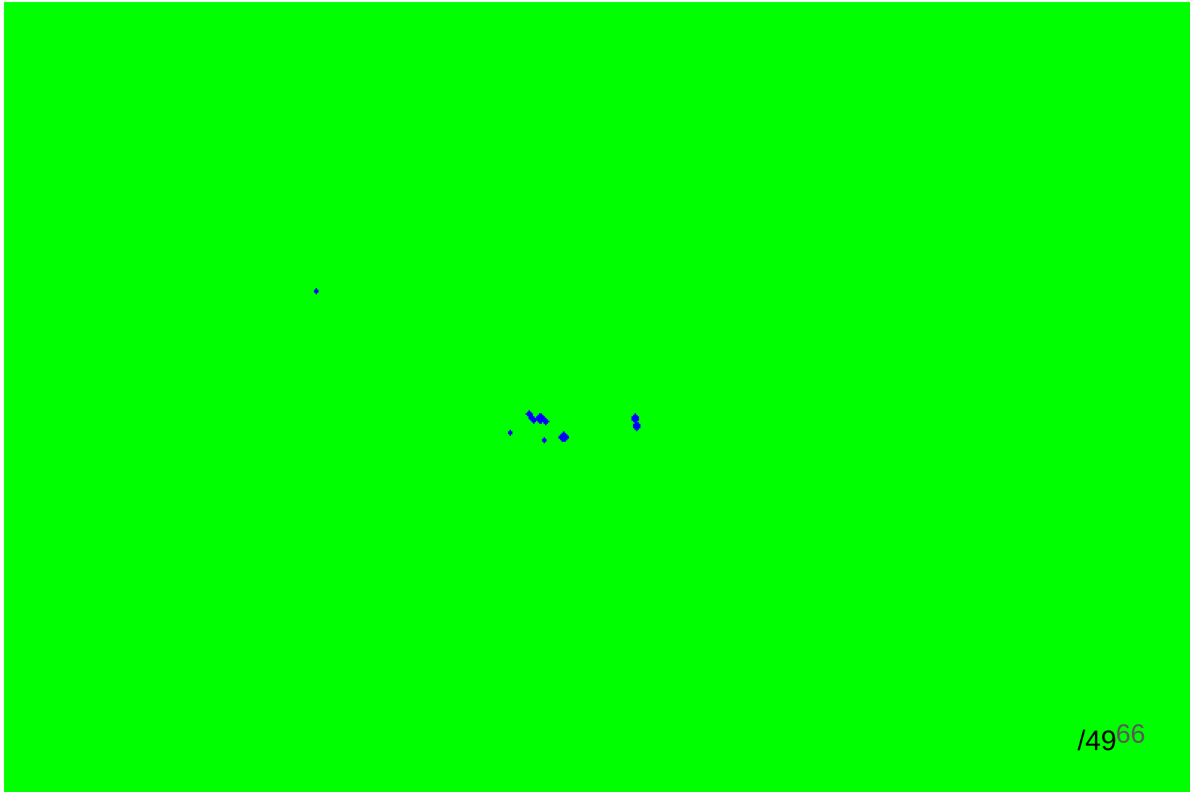
Anomaly detection in any image

- Compute the **noise difference** $N := \tilde{u} - u$

Non self-similar residue scale 0

Anomaly detection in any image

Detections in « noise », scale 0, min (log NFA) = -11,7



Anomaly detection in any image

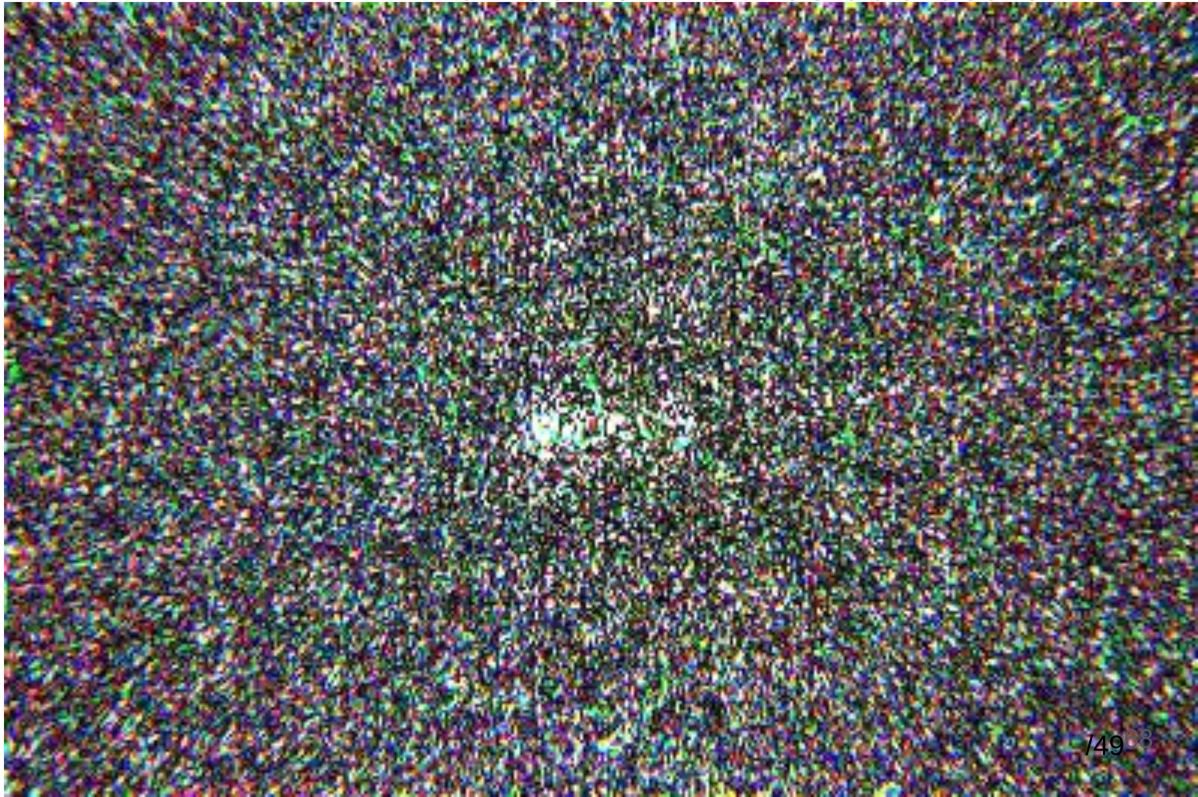
- Reconstruct an **normal image model** $\tilde{u}$ by aggregating all patch estimates (a simple mean)

Self-similar part scale 1

Anomaly detection in any image

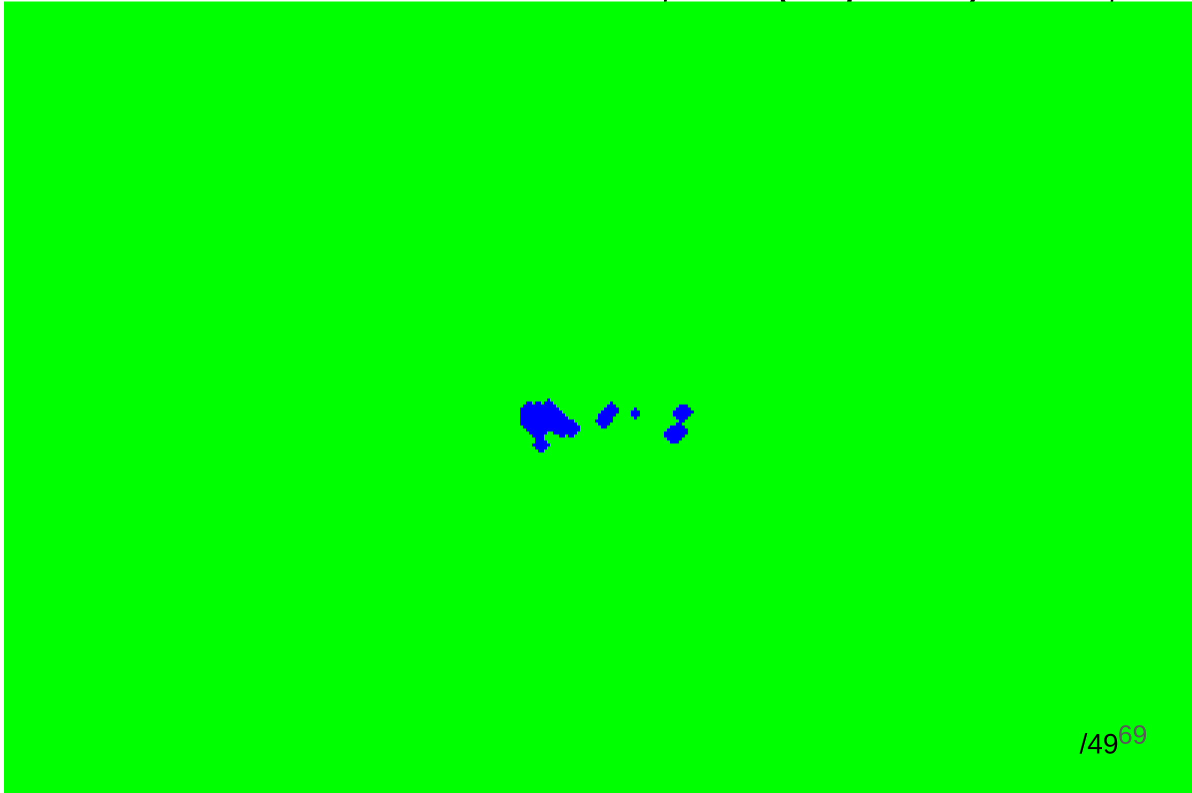
- Compute the **noise difference** $N := \tilde{u} - u$

Non self-similar residue scale 1



Anomaly detection in any image

Detections in « noise » scale 1, min (log NFA) = -26,2



Anomaly detection in any image

- Reconstruct an **normal image model** $\tilde{u}$ by aggregating all patch estimates (a simple mean)

Self-similar part scale 2

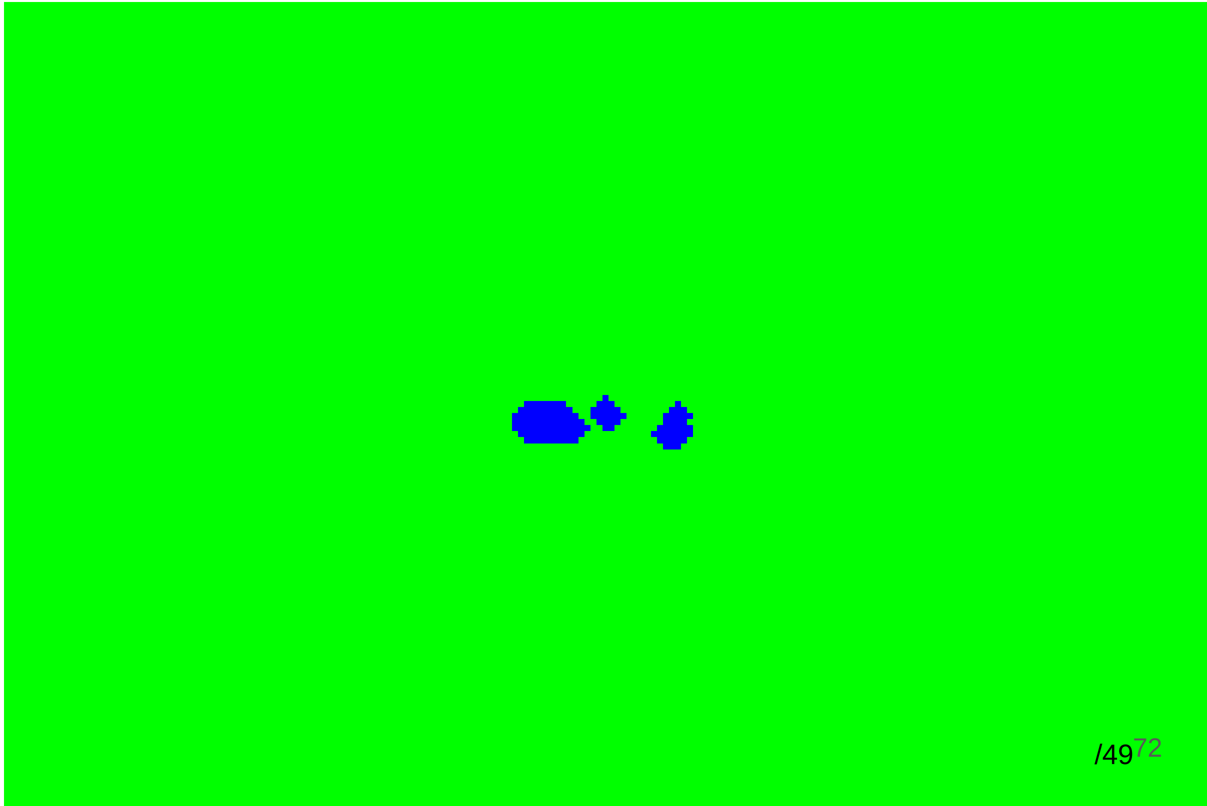
Anomaly detection in any image

- Compute the **noise difference** $N := \tilde{u} - u$

Non self-similar residue scale 2

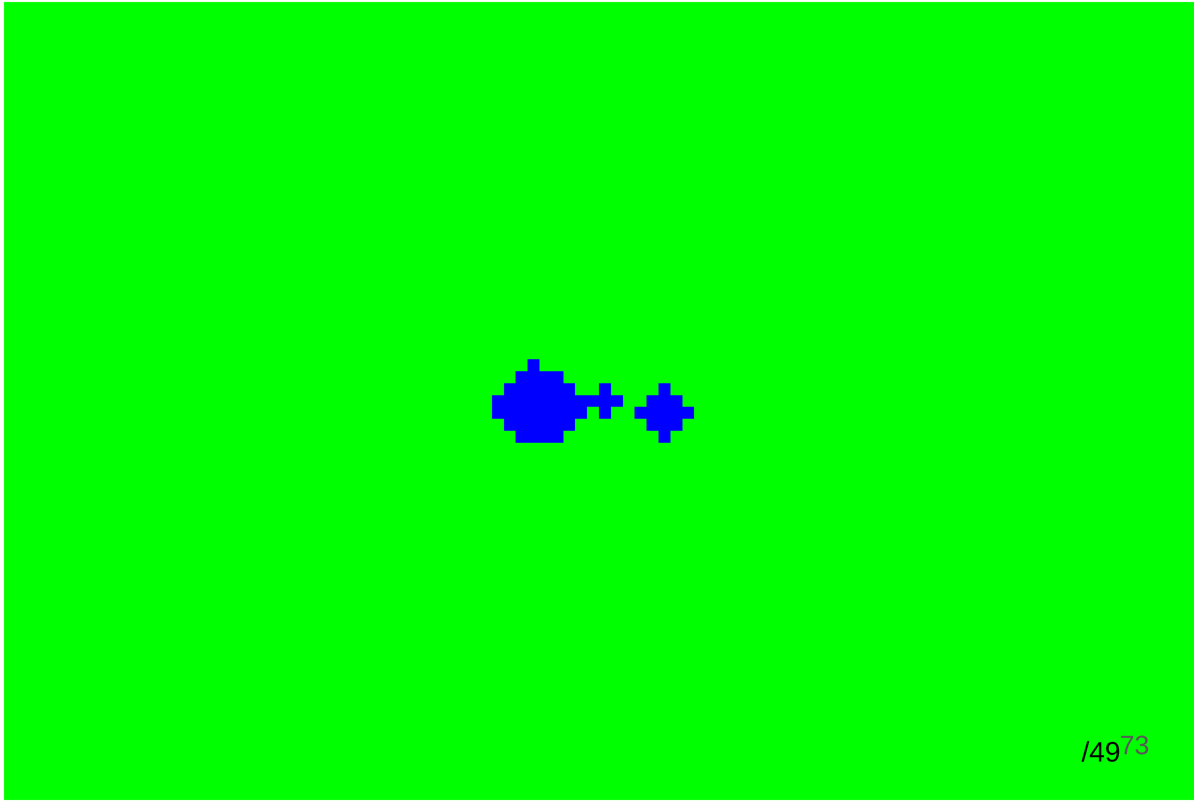
Anomaly detection in any image

Detections in « noise » scale 2, min (log NFA) = -11,7



Anomaly detection in any image

Detections in « noise » scale 3 , min (log NFA) = -19,9



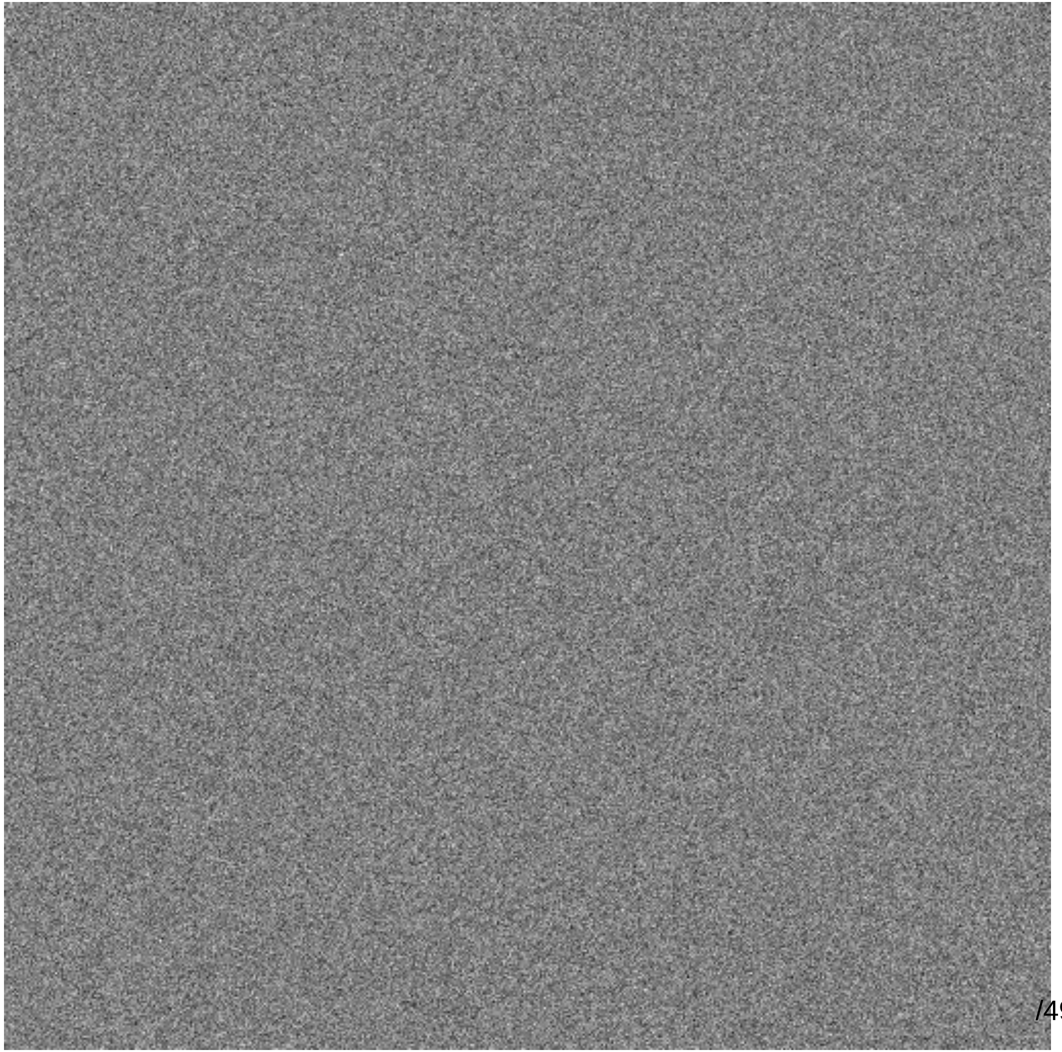
Sanity check 1 : No detection white noise!

Scale 0
Minimum log10NFA region = 1.03

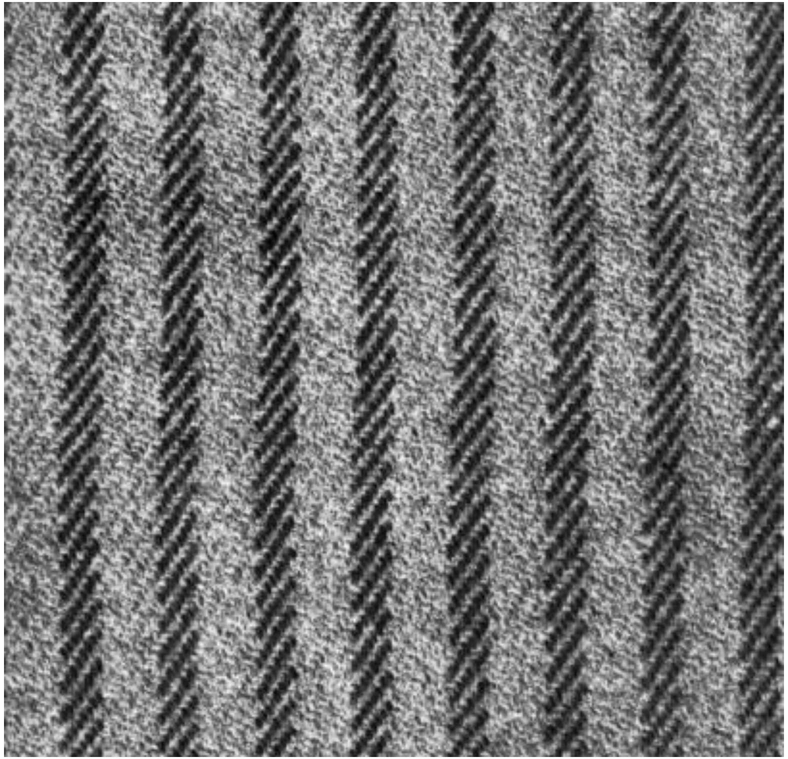
Scale 1
Minimum log10NFA region = 0.04

Scale 2
Minimum log10NFA region = 2.66

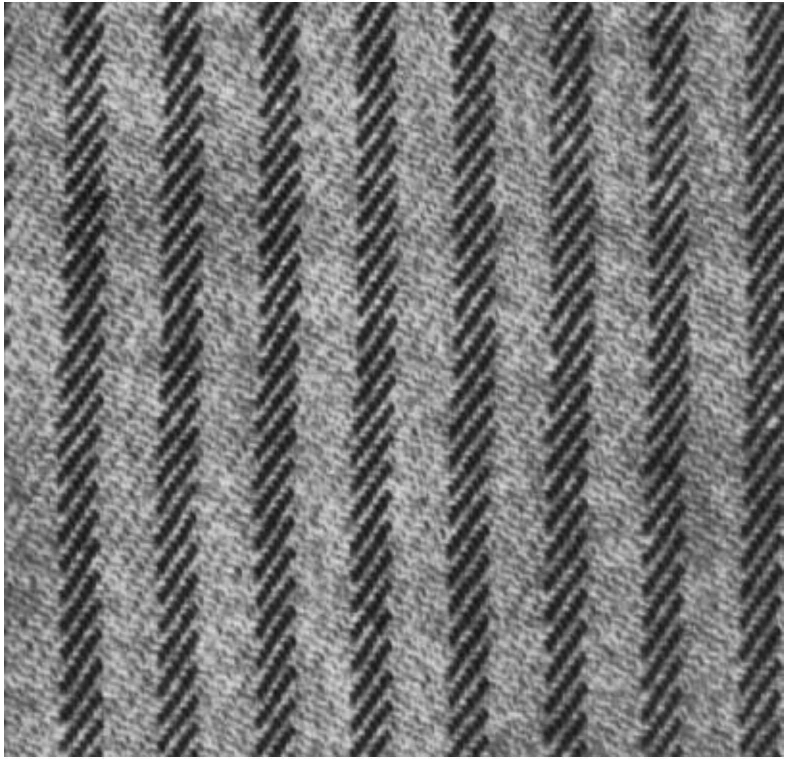
Scale 3
Minimum log10NFA region = -0.11



Sanity check 2: No detection in homogeneous texture



ORIGINAL



SELF-SIMILAR PART

Sanity check 3: working on the residual increases the « NFA gap » between false alarms and detections

mammography with tumor | new detection (log NFA=-12) | Detections in [14]: log NFA = 0

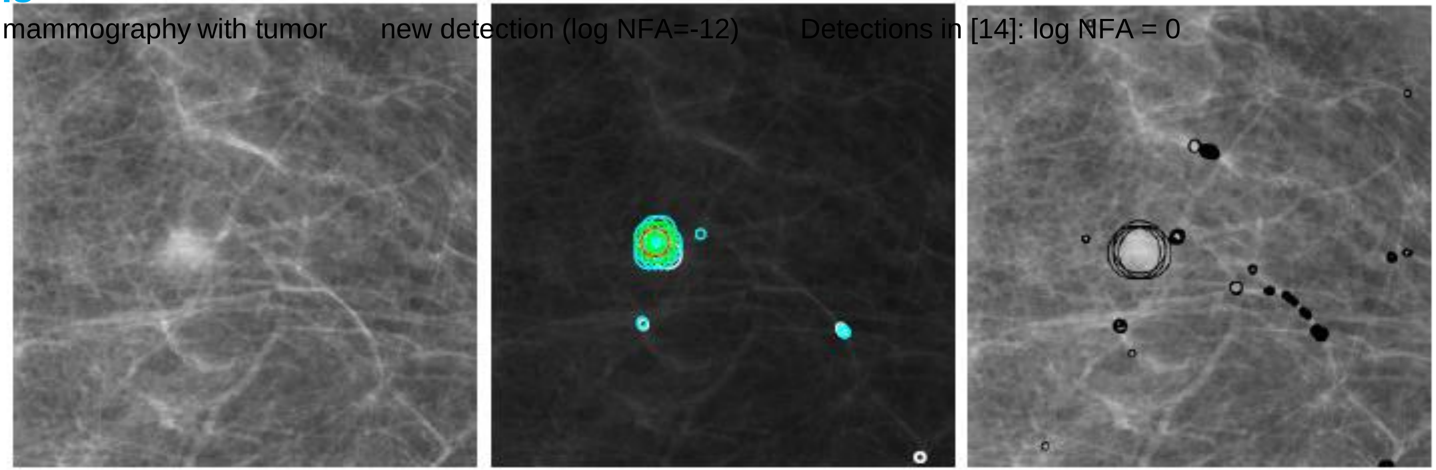
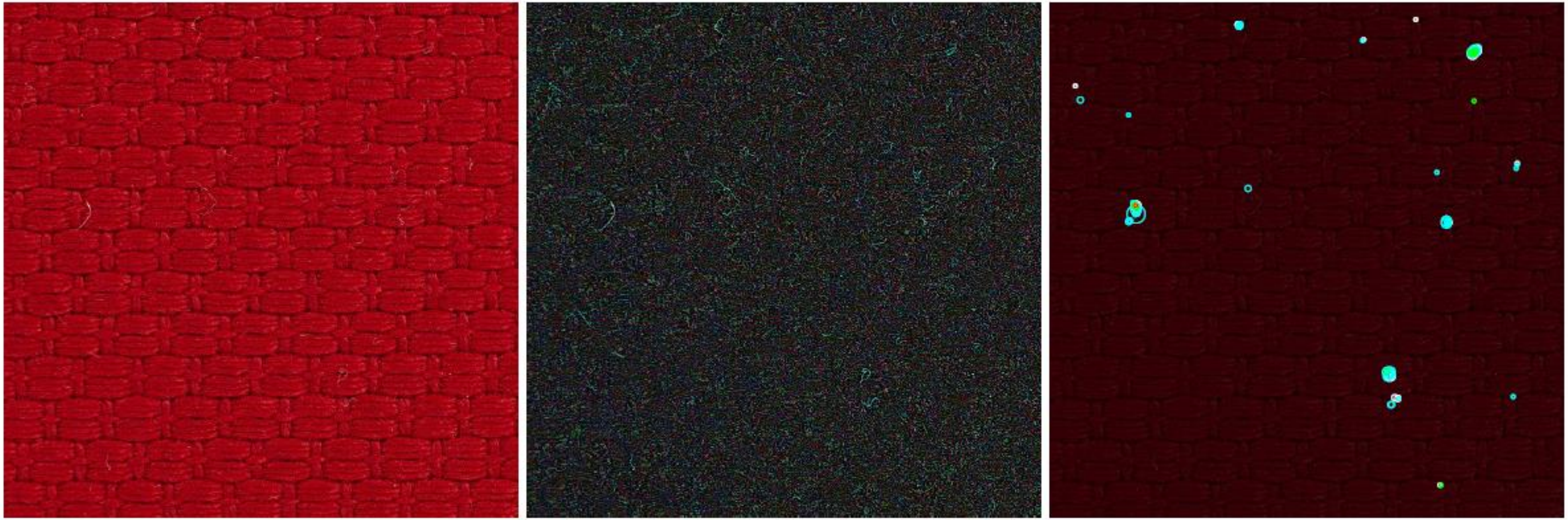


Figure 5. The region represented by the large white spot in the left image is a tumor. The proposed self-similarity anomaly detector successfully detects the tumor with a much significant NFA than the one from Grosjean and Moisan [24] (an NFA of 10^{-12} versus their reported NFA of 0.15), while making fewer false detections.

B. Grosjean and L. Moisan. A-contrario detectability of spots in textured backgrounds. Journal of Mathematical Imaging and Vision, 33(3):313–337, 2009.

Example on a real scene with no ground truth

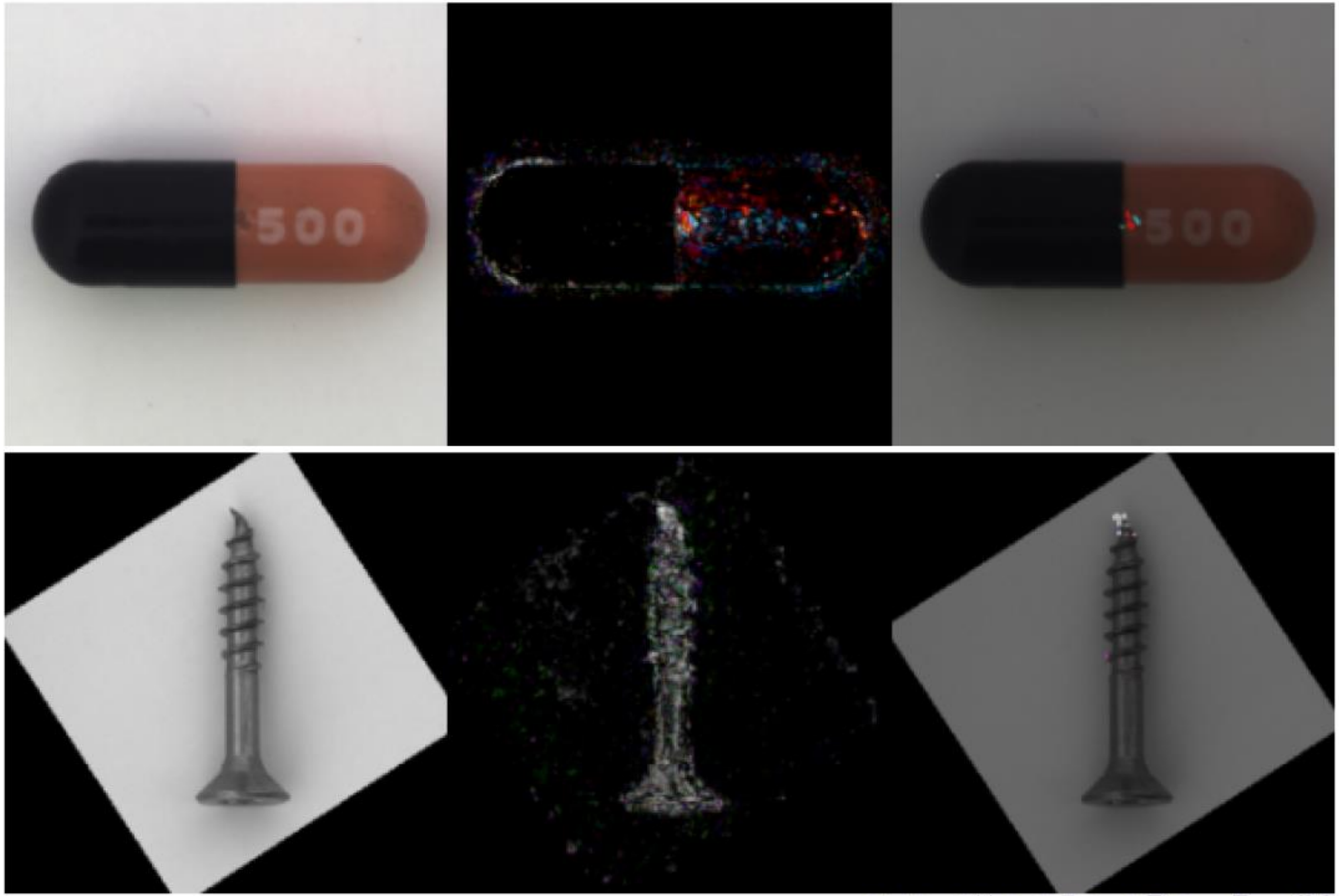


Left: Picture of textile, right: The residual for pixels and the detections. All the textile impurities are highlighted on the residual.

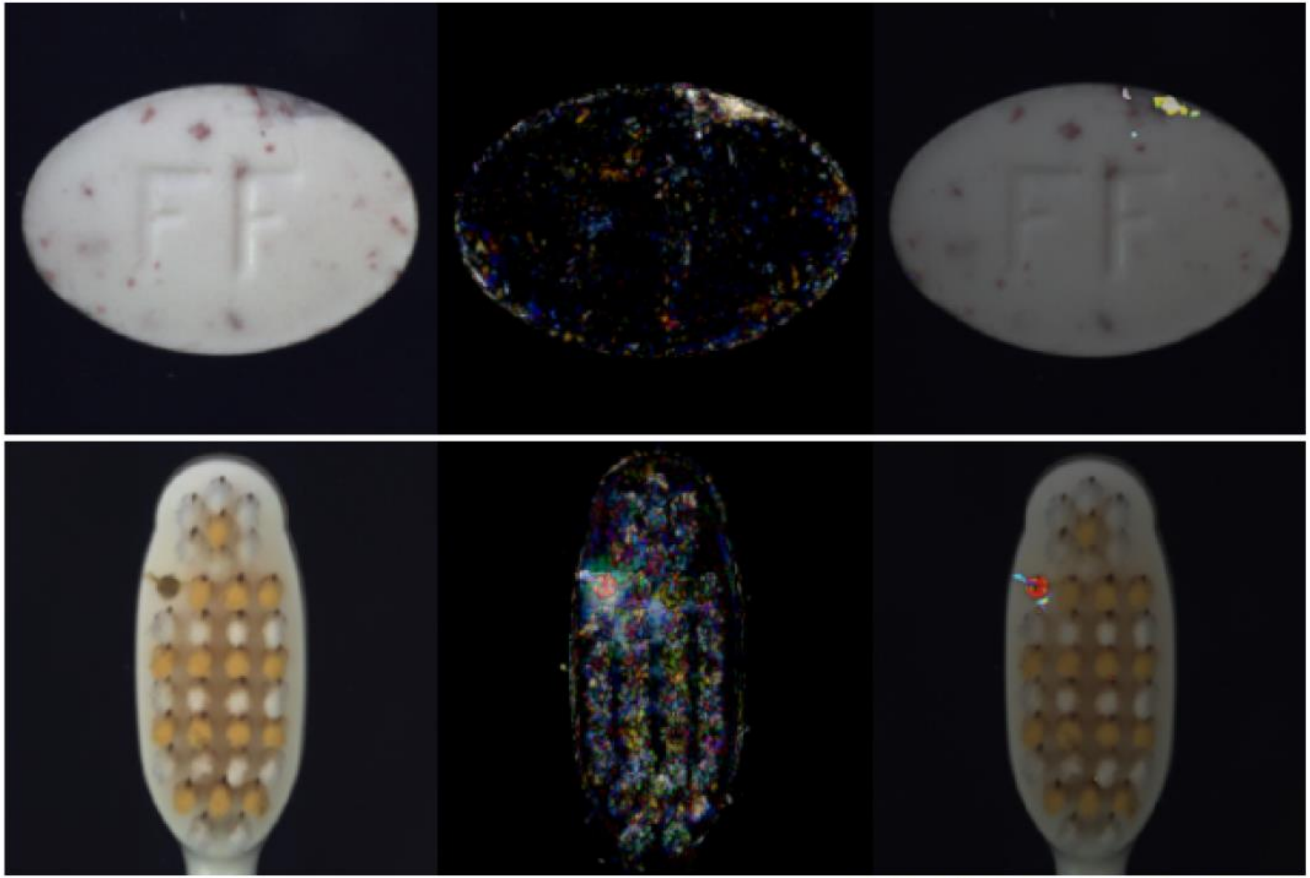
Example on a real scene with no ground truth

Left: Input image, Right: detections with pixels. The method successfully detects a tank hidden in the landscape. This example is one of the examples provided by Itti et al.
L. Itti and C. Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. Vision research, 2000.

Anomaly detection in industrial parts



Anomaly detection in industrial parts



Thank you, questions?

Cloud detection in time series of satellite images

- *Background model:* the images are different (because of cloud cover): in two successive registered images of the ground, the difference of gradient orientation is uniform i.i.d.
- *Event :* a "large enough" 4-connected component of pixels where the sum of differences of orientations is small enough
- *Number of tests*= #(pixels)×#(number of connected components)

Cloud detection in time series of satellite images

Local Image Comparison

Given two images u and v defined on the same domain Ω (of size $X \times Y$), and a set of pixels R, we would like to know whether both images are similar in the region R. To this aim, we will use the distance

$$s_{u,v}(R) = \sum_{\omega \in R} \frac{|\mathrm{Angle}(\nabla u(\omega), \nabla v(\omega))|}{\pi}, \tag{11.1}$$

namely the sum of all normalized gradient angle errors in R

Cloud detection in time series of satellite images

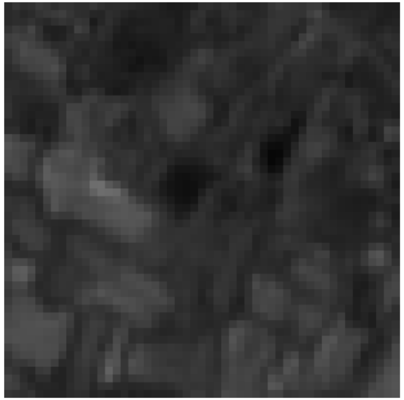
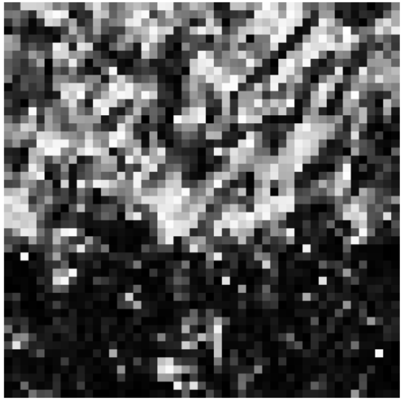
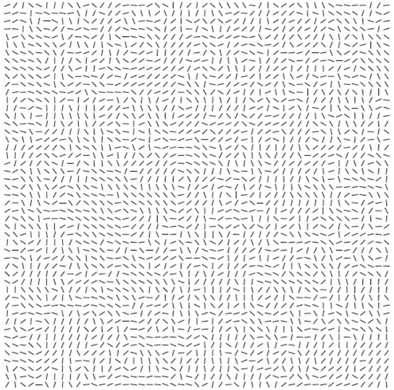


Image a

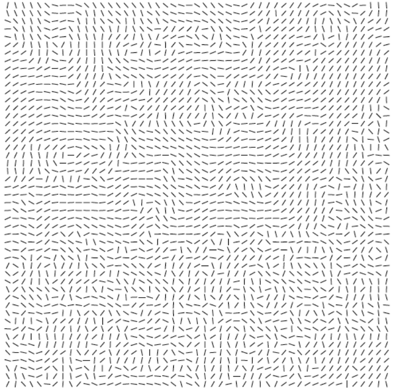
Image b



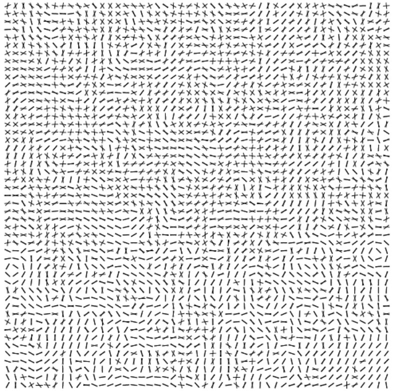
Normalized angle error



Gradient orientation a



Gradient orientation b



Superposed gradient orientations

Cloud detection in time series of satellite images

A contrario formulation

For a given region R, we need to decide whether the distance $s_{u,v}(R)$ is small enough, indicating whether the corresponding parts of the images are similar or not. We propose to use for this an *a contrario* formulation. A natural background model $\mathcal{H}_0$ is that the gradient orientations at each pixel are independent random variables, uniformly distributed in $[-\pi, \pi)$. (This will happen, for example, if one of the images contains a cloud covering the region.) Following the *a contrario* framework, we will define the NFA associated to a candidate region match as

$$\mathrm{NFA}(u, v, R) = N_T \cdot \mathbb{P}\Big[S_{\mathcal{H}_0}(R) \leq s_{u,v}(R)\Big], \tag{11.2}$$

where $S_{\mathcal{H}_0}(R)$ is a random variable corresponding to the distance $s_{U,V}(R)$ for random images U and V whose gradient orientation follow $\mathcal{H}_0$.

But under $\mathcal{H}_0$ the gradient orientations are uniformly distributed in all directions, which implies that the normalized angle error at each pixel are independent random variables following a uniform distribution in $[0,1]$. As a result, $S_{\mathcal{H}_0}(R)$ corresponds to the sum of $|R|$ independent and uniformly distributed random variables taking values in $[0,1]$. Thus, $S_{\mathcal{H}_0}(R)$ follows the Irwin-Hall distribution [149] and for a given s, with $0 \leq s \leq |R|$, we obtain:

$$\mathbb{P}\left[S_{\mathcal{H}_0}(R) \leq s\right] = \frac{1}{|R|!} \sum_{i=0}^{\lfloor s \rfloor} (-1)^i \binom{|R|}{i} (s-i)^{|R|}, \tag{11.3}$$

where $\lfloor s \rfloor$ is the integer part of s and $\binom{n}{i}$ is the binomial coefficient.

Cloud detection in time series of satellite images

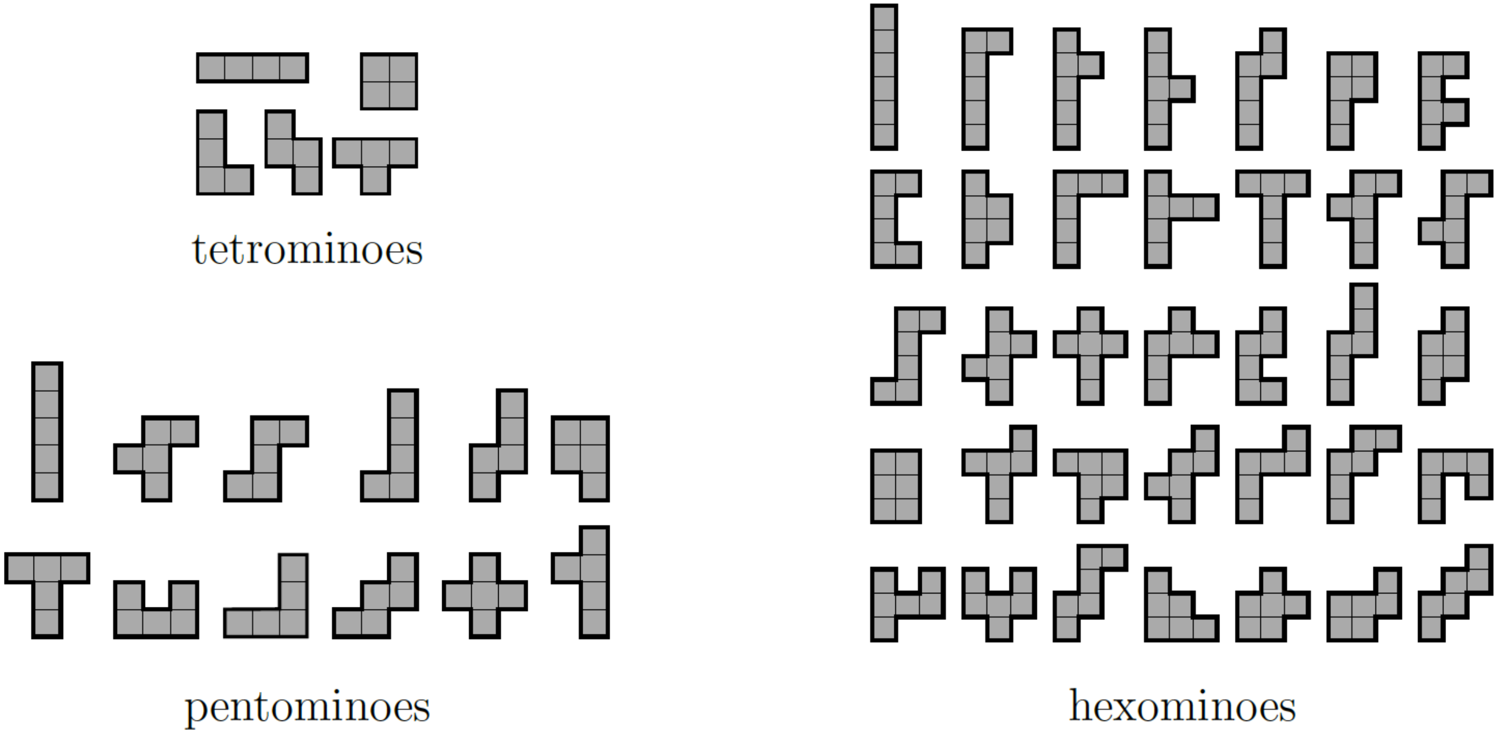


Figure 11.2: Polyominoes of four (tetrominoes), five (pentominoes) and six (hexominoes) elements. When rotations and reflections are not considered, there are 5 tetrominoes, 12 pentominoes and 35 hexominoes, as shown here. When rotations and reflections are also considered distinct, there are 19 tetrominoes, 63 pentominoes and 216 different hexominoes.

Polyominoes

The exact number b_n of different polyomino configurations of given size n is not known in general, but there are good approximations of this number [147]. In our case, it is enough to use an estimate of the order of magnitude, so the approximate formula given in [147] is sufficient for our needs. It reads

$$b_n \approx \alpha \frac{\beta^n}{n}, \qquad (11.4)$$

where $\alpha \approx 0.316915$ and $\beta \approx 4.062570$.

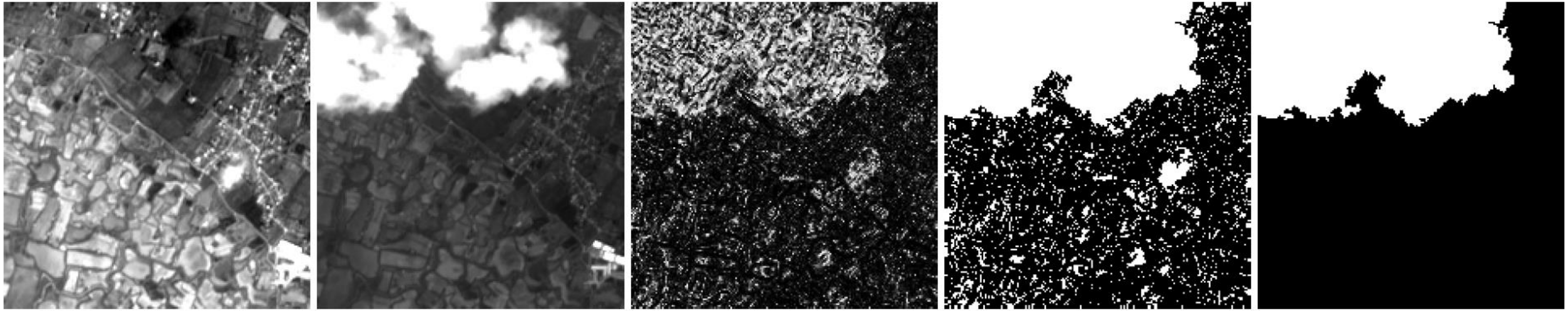


image A | image B | normalized angle error | initial visible ground map | final visible ground map

Anomaly detection theory: Gaussian models and background subtraction

Detection in hyperspectral images

Matteoli, S. ; Diani, M. ; Corsini, G., *A tutorial overview of anomaly detection in hyperspectral images*

I. S. Reed, X.(iaoli) Yu, *Adaptive Multiple-Band CFAR Detection of An Optical Pattern with Unknown Spectral Distribution*, IEEE Trans. Acoust. Speech Signal Process., 38(10) (1990) 1760-1770.

Gaussian vectors and the Mahalanobis distance

Definition 4. *We say that* $\mathbf{x}$ *is a (non degenerate) Gaussian vector in* $\mathbb{R}^k$ *if* $\Sigma_{\mathbf{x}}$ *is nondegenerate and if* $\mathbf{x}$ *has density*

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{k}{2}} (\det \Sigma_{\mathbf{x}})^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\mu_{\mathbf{x}})^t \Sigma_{\mathbf{x}}^{-1} (\mathbf{x}-\mu_{\mathbf{x}})}. \tag{1.11}$$

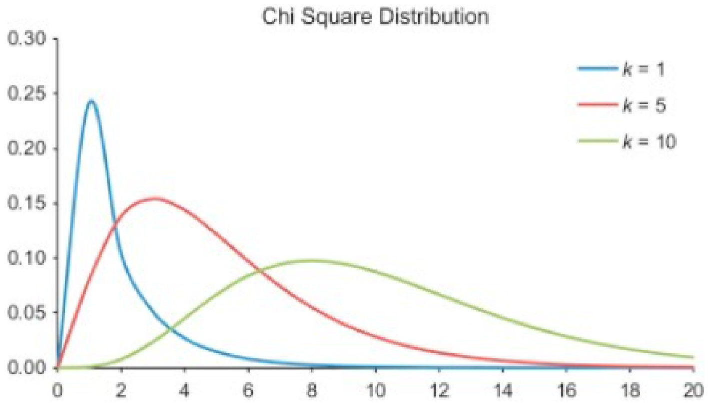
Conversely a random vector with this density has mean $\mu_{\mathbf{x}}$ *and covariance matrix* $\Sigma_{\mathbf{x}}$.

Gaussian vectors and the Mahalanobis distance

The χ^2 law If $x_1, \cdots x_k$ are independent, standard normal random variables, then the sum of their squares, $\sum_{i=1}^{k} x_i^2$ is distributed according to the χ^2 distribution with k degrees of freedom denoted as χ_k^2. The decay of the χ_k^2 law is rapid, as shown by its probability density function

$$f(x;\, k) = \begin{cases} \dfrac{x^{\frac{k}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{k}{2}} \Gamma\left(\frac{k}{2}\right)}, & x > 0; \\ 0, & \text{otherwise.} \end{cases}$$

where $\Gamma(k/2)$ denotes the gamma function, which has a closed-form expression when its argument is an integer.



Gaussian vectors and the Mahalanobis distance

The Mahalanobis distance from a vector $x = (x_1, x_2, x_3, \dots, x_k)$ with respect to a random vector $\mathbf{x}$ with mean $\mu_{\mathbf{x}} = (\mu_1, \mu_2, \mu_3, \dots, \mu_k)^T$ and covariance matrix $\Sigma_{\mathbf{x}}$ is defined by

$$D_M(x) = \sqrt{(x - \mu_{\mathbf{x}})^T \Sigma_{\mathbf{x}}^{-1} (x - \mu)}. \tag{1.14}$$

The link with the χ^2 law is as follows. If $\mathbf{x} \sim \mathcal{N}_k(\mu, \Sigma_{\mathbf{x}})$ is a random Gaussian vector with expectation $\mu_{\mathbf{x}}$ and positive definite covariance matrix $\Sigma_{\mathbf{x}}$ then

$$\tilde{\mathbf{x}} =: \Sigma_{\mathbf{x}}^{\frac{1}{2}} \mathbf{x} \sim \mathcal{N}(0, I_k)$$

is a normal variable of dimension k and therefore

$$D_M^2(\mathbf{x}, \mu) = ||\tilde{\mathbf{x}}||^2 \sim \chi_k^2,$$

namely the law of the sum of k independent normal laws of dimension 1.

Gaussian vectors and the Mahalanobis distance

Consequence : testing anomalies with respect to a Gaussian background

In short, the Mahalanobis distance between a Gaussian vector $\mathbf{x}$ and its expectation follows a χ^2 law with k degrees of freedom. Hence, if $0 < p << 1$ is a fixed p-value and $\chi^2_{k;1-p}$ denotes the $1-p$ quantile of χ^2_k then

$$\mathbb{P}\left[D_M^2(X,\mu) \geq \chi^2_{k;1-p}\right] = p = \mathbb{P}\left[X \in A_p\right]$$

where

$$A_p := \left\{x \in \mathbb{R}^k | D_M^2(x,\mu) \geq \chi^2_{k;1-p}\right\}$$

is by definition the anomalous region with p-value p.

Anomaly detection in hyperspectral images

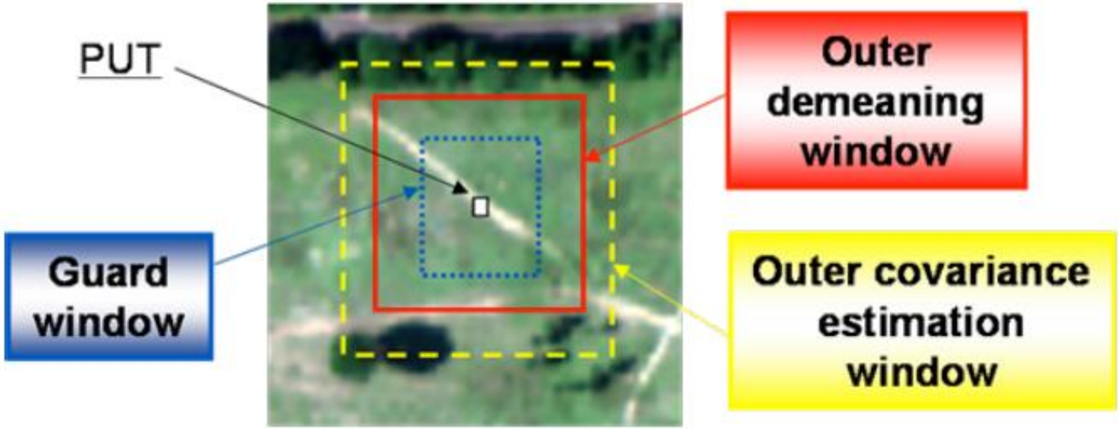


Figure 1. Spatial windows used in the RX implementation: outer demeaning window (red, solid line), outer covariance estimation window (yellow, dashed line), guard window (blue, dotted line).

Matteoli, S. ; Diani, M. ; Corsini, G., *A tutorial overview of anomaly detection in hyperspectral images*

I. S. Reed, X.(iaoli) Yu, *Adaptive Multiple-Band CFAR Detection of An Optical Pattern with Unknown Spectral Distribution*, IEEE Trans. Acoust. Speech Signal Process., 38(10) (1990) 1760-1770.

Anomaly detection in hyperspectral images

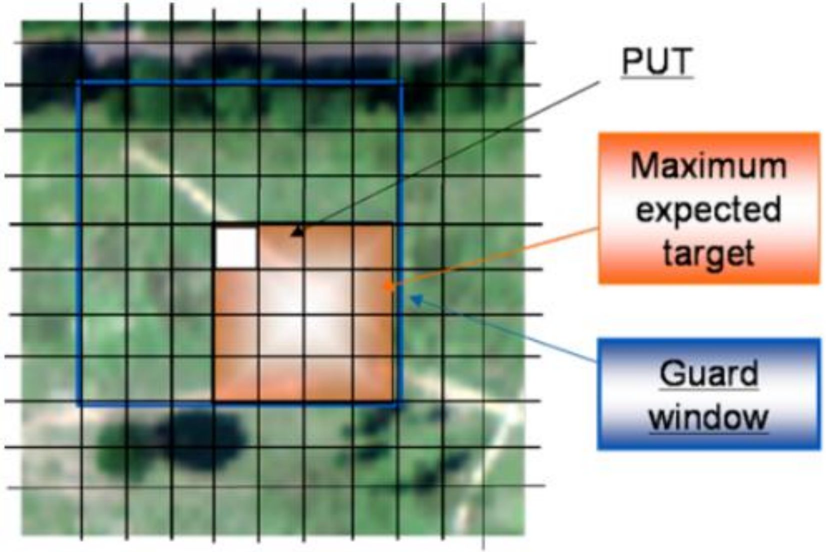


Figure 1.7: For a 4 x 4 pixels expected target size, the guard window should be at least of 7 x 7 pixels in order to not include target pixels in the background parameters estimation windows. Figure borrowed from Matteoli et al. [70]

Matteoli, S. ; Diani, M. ; Corsini, G., *A tutorial overview of anomaly detection in hyperspectral images*

I. S. Reed, X.(iaoli) Yu, *Adaptive Multiple-Band CFAR Detection of An Optical Pattern with Unknown Spectral Distribution*, IEEE Trans. Acoust. Speech Signal Process., 38(10) (1990) 1760-1770.

Anomaly detection in hyperspectral images

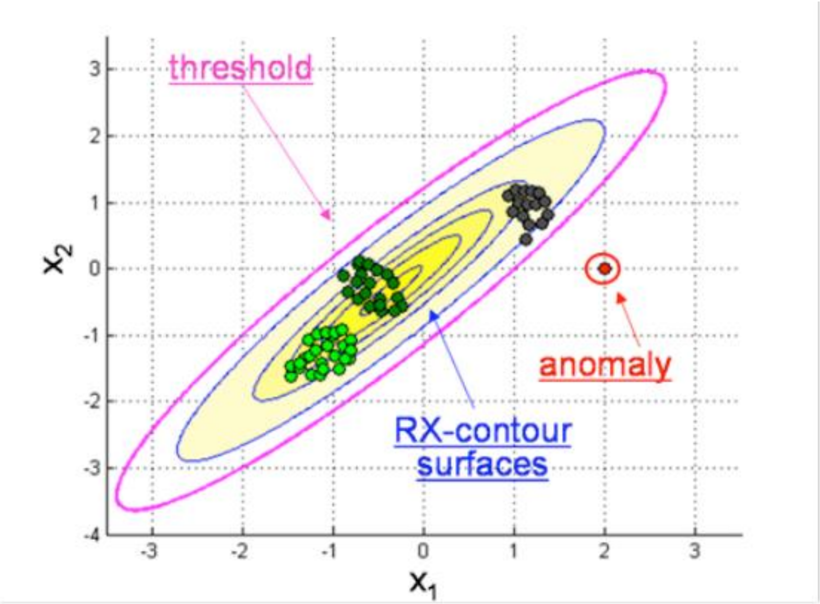


Figure 2. Graphical interpretation in a two-dimensional space for the RX AD decision rule. The RX AD decision surfaces are ellipsoids in the multidimensional space.

Matteoli, S. ; Diani, M. ; Corsini, G., *A tutorial overview of anomaly detection in hyperspectral images*

I. S. Reed, X.(iaoli) Yu, *Adaptive Multiple-Band CFAR Detection of An Optical Pattern with Unknown Spectral Distribution*, IEEE Trans. Acoust. Speech Signal Process., 38(10) (1990) 1760-1770.

Anomaly detection in hyperspectral images

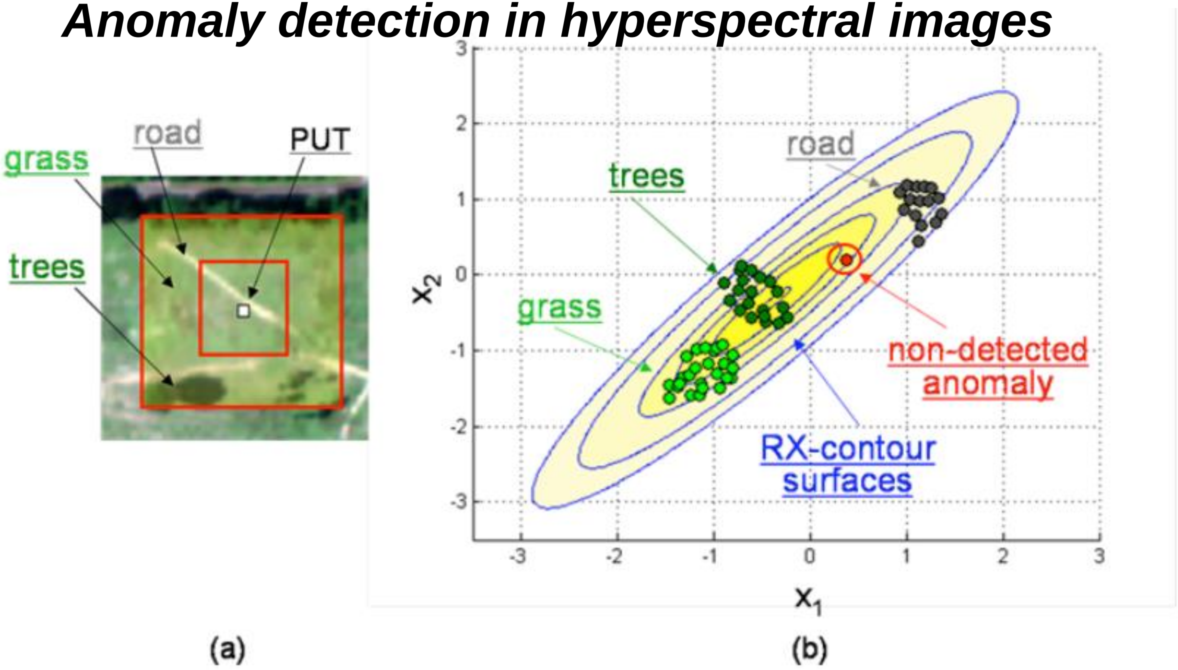
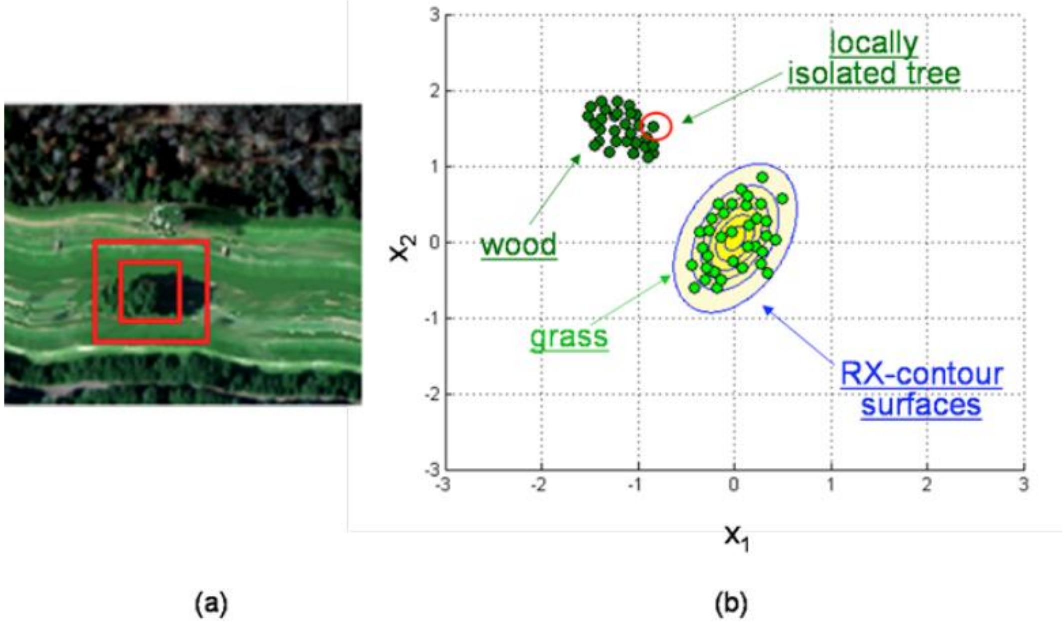


Figure 4. Graphical representation of a non-homogeneous background condition. (a) spatial domain. (b) a simplified two-dimensional spectral domain. The presence of multiple classes in the immediate vicinity of the PUT could prevent an anomaly from being detected. In these cases the LNM is inadequate.

Matteoli, S. ; Diani, M. ; Corsini, G., *A tutorial overview of anomaly detection in hyperspectral images*

I. S. Reed, X.(iaoli) Yu, *Adaptive Multiple-Band CFAR Detection of An Optical Pattern with Unknown Spectral Distribution*, IEEE Trans. Acoust. Speech Signal Process., 38(10) (1990) 1760-1770.

Anomaly detection in hyperspectral images



Graphical representation of the detection of a local anomaly that is not anomalous in the whole scene. (a) spatial domain. (b) simplified two-dimensional spectral domain. The scene reported contains a forest and a locally isolated tree. The RX sliding window is represented in red. The samples captured by this window are pixels of a homogeneous background of grass, and hence the locally isolated tree is detected even if it is not anomalous in the scene.

Matteoli, S. ; Diani, M. ; Corsini, G., *A tutorial overview of anomaly detection in hyperspectral images*

I. S. Reed, X.(iaoli) Yu, *Adaptive Multiple-Band CFAR Detection of An Optical Pattern with Unknown Spectral Distribution*, IEEE Trans. Acoust. Speech Signal Process., 38(10) (1990) 1760-1770.

Anomaly detection in hyperspectral images

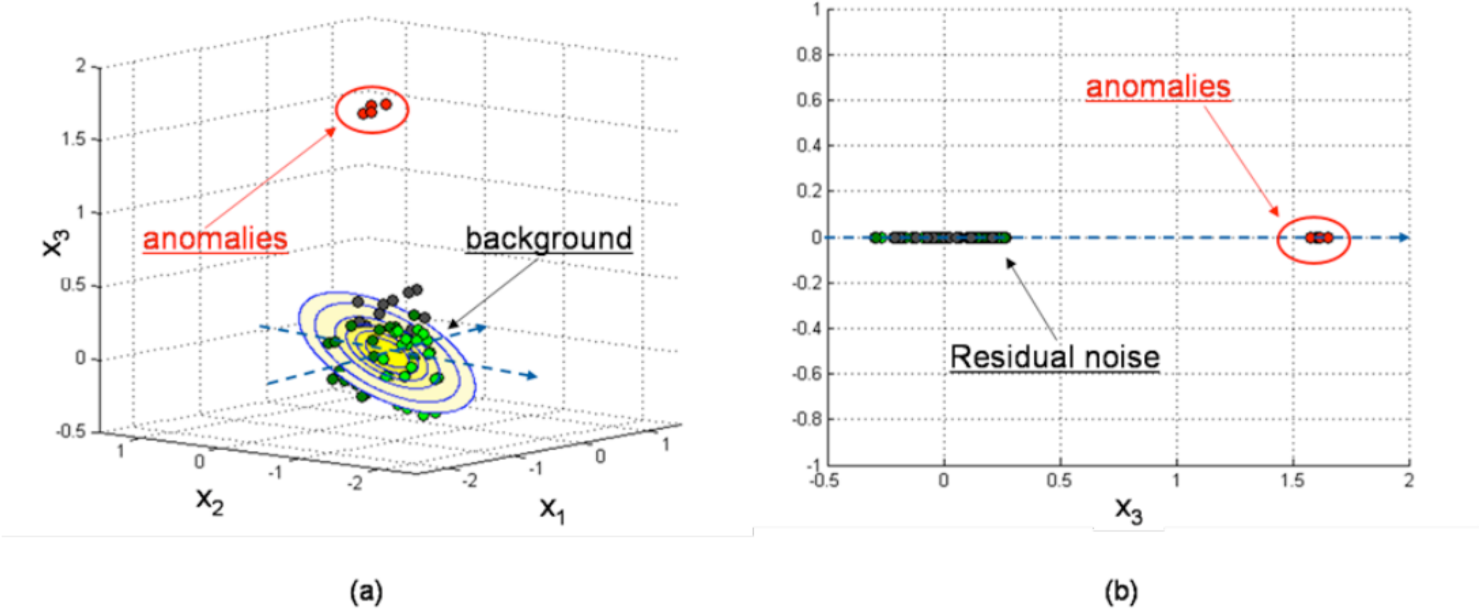


Figure 12. Graphical interpretation of the OSP concept. (a) a simplified three-dimensional spectral domain. The two principal directions (dashed blue lines) that address the background have been identified by a linear transformation. (b) The data have been projected onto the subspace orthogonal to the one spanned by these two identified background directions. Along this third orthogonal direction (dashed blue line) the *detectability* of anomalies is clearly improved, as the background has been nearly suppressed.

Matteoli, S. ; Diani, M. ; Corsini, G., *A tutorial overview of anomaly detection in hyperspectral images*

J. C. Harsanyi, C-I. Chang, *Hyperspectral Image Classification and Dimensionality Reduction: An Orthogonal Subspace Projection Approach*, IEEE Trans. Geosci. Remote Sens., 32(4) (1994) 779-785.

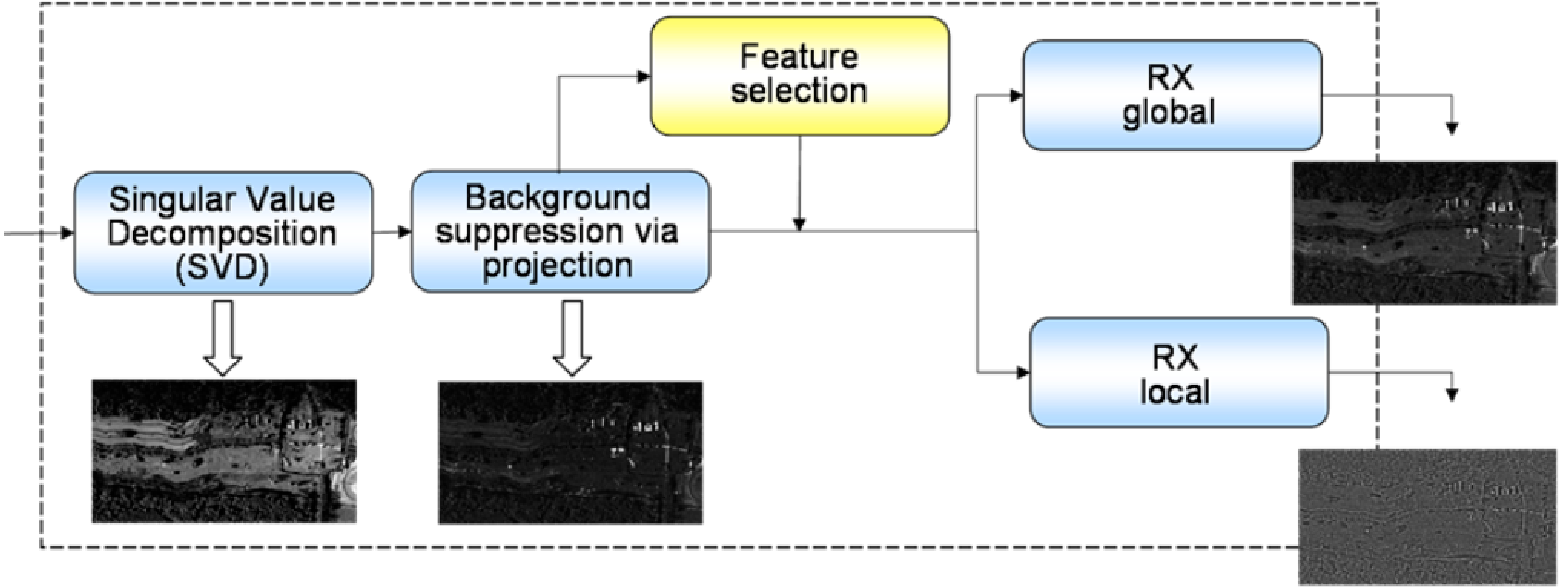


Figure 13. Block diagram of the possible configurations allowed by the OSP approach presented here. After the SVD, the background is suppressed by orthogonal projection. The energies of the corresponding images are represented that show the effect of the background suppression. The yellow block is an optional step that can further improve the performance by selecting the components that address the most anomalous pixels. These components can then be processed by RX, either in a global or local application, providing in output the test statistic.

Matteoli, S. ; Diani, M. ; Corsini, G., *A tutorial overview of anomaly detection in hyperspectral images*

J. C. Harsanyi, C-I. Chang, *Hyperspectral Image Classification and Dimensionality Reduction: An Orthogonal Subspace Projection Approach*, IEEE Trans. Geosci. Remote Sens., 32(4) (1994) 779-785.

Anomaly detection in RGB images by background subtraction and final detection in noise

[How to Reduce Anomaly Detection in Images to Anomaly Detection in Noise]

2019-12-08 · Thibaud Ehret, Axel Davy, Mauricio Delbracio, Jean-Michel Morel

How to Reduce Anomaly Detection in Images to Anomaly Detection in Noise

Anomaly detectors address the difficult problem of detecting automatically exceptions in a background image, that can be as diverse as a fabric or a mammography. Detection methods have been proposed by the thousands because each problem requires a different background model.

Anomaly detection cannot be formulated in a Bayesian framework: this would require to simultaneously learn a model of the anomaly, and a model of the background.

(In the case where there are plenty of examples of the background and for the object to be detected, neural networks may provide a practical answer, but without explanatory power). In the case of anomalies, we often dispose of only one image as unique informer on the background, and of no example at all for the anomaly.

The problem can be reduced to detecting anomalies in residual images (extracted from the target image) in which noise and anomalies prevail. Hence, the general and impossible background modeling problem is replaced by a simple noise model, and allows the calculation of rigorous detection thresholds.

Our approach is therefore unsupervised and works on arbitrary images. The residual images can favorably be computed on dense features of neural networks. Our detector is powered by the a contrario detection theory, which avoids over-detection by fixing detection thresholds taking into account the multiple tests.