Detection theory project report

Calvin Galagain
ENS Paris Saclay - MVA
calvin.galagain@gmail.com

Abstract

The goal of our Detection Theory project is to study a paper and discuss the concepts discovered in the course in relation to this paper.

1. Introduction

The paper we chose is entitled "*Change detection for high resolution satellite images, based on SIFT descriptors and an a contrario approach*"[1], and was published in 2014 for IEEE Geoscience and Remote Sensing Symposium. It was written by F Dellinger, J Delon, Y Gousseau, J Michel, and F Tupin. This paper comes without code.

2. Problem definition

This paper discusses satellite image analysis and more specifically the capacity to quickly assess damages in an area after a disaster (e.g landslide, earthquake, industrial accident, plant explosion). Since there are usually only a few images available for a designated area, one must rely on several sensors and sources to run a comparison, which can lead to wasting time on 'false' differences between images (due to changing conditions or capture artifacts) and not focusing on the actual damage.

Some elements link the typical images that can be seen in the paper to Gestalt theory, even though the focus will be on shapes of buildings or landscapes seen from above, and almost exclusively focused on T junctions and various shades of color, as well as heavily relying on grouping to distinguish between elements. The particularity of satellite images removes a lot of the Gestalt theory elements since one is not considering a picture potentially often seen by the human eye. The most important point is that the notion of background is lost as it would now solely represent the actual earth texture under the buildings / trees, etc. Elements of symmetry can be used, but the forced focus on a restricted area removes the potential for closure characteristics. Finally lots of satellite images are not full color, restricting once more the potential to leverage additional elements of the Gestalt theory.

The first step is to detect **interest points** in provided images : this is done leveraging SIFT and SAR, as it will be discussed in further detail in the code and algorithm discussion part. This initial step allows to produce **detected keypoints** (each associated to a **descriptor** following SIFT methodology), and matched keypoints by computing the distance between the descriptors of detected keypoints of image A to image B (described in *Figure 2* of the paper [1]). This allows us to obtain **matched keypoints**.

The *a contrario* approach comes on top of this initial step. Following general principles of *a contrario* approach, one considers that it is unlikely that the density of detected keypoints is much higher than the density of matched keypoints, and that if such a difference is observed (appearing unusual to the regular phenomenon observed by the eye, following Helmholtz principle [2]), it probably means there is a difference between image A and image B worthy of interest. In our context, that can indicate potential damage after a natural disaster or another event.

Since the goal is to detect changes on satellite images thanks to detected and matched keypoints, the **background model** consists of locally considering the density of matched keypoints on an area, and considering that the density of detected keypoints in that area is the same (and those keypoints are distributed according to a Poisson process) ; if however one finds much more detected keypoints than matched keypoints locally, it would be considered “surprising” and indicative of a change area.

When surveying an area A (circle around a keypoint) of an image, one can record the number of locally detected keypoints (x), among the N total number of detected keypoints on the image. Under the background model, the event probability of finding x keypoints in A follows a binomial law of parameter p (intensity of the Poisson process locally, made of the ratio of matched keypoints in A over the entire quantity of matched keypoints in the image). This detection test can lead us to provide the formula of the NFA (based on the binomial tail of PH0, and number of tests being the number of keypoints, as we want to cover the whole image):

$$NFA = N \cdot P_{\mathcal{H}_0}(X \geq x) = N \cdot \sum_{i=x}^{N} \binom{N}{i} p^i (1-p)^{N-i}$$

This “detected keypoints” detection tests will be surprising when passing a certain threshold of detected keypoints in the area.

Proposition 4 (2.7) in the course gives us indeed that:

- with X random variable following a known distribution (Poisson) under H0
- sitting in the particular case when ni = N for all i

It immediately follows that the above function is indeed an NFA.

3. Code discussion

An algorithm to first identify detected and matched keypoints can be implemented, and has been done in the attached Colab notebook (SIFT-SAR + AC-RANSAC to suppress false matches like in the paper).

Once two images are obtained with a good list of detected and matched keypoints, the algorithm that follows could be:

Algorithm for NFA:
Input: a set of detected (D) and matched (M) keypoints
Output: NFA value for a keypoint k

In a radius of 60 pixels around k, count the number of detected keypoints (n) and matched keypoints (m).
Use N = D+M total number of points in the image.
Use p = m/M
Compute the NFA = $N * \mathrm{sum}_{(\text{from } i=n \text{ to } N)} C^i_N * p^i * (1-p)^{(N-i)}$

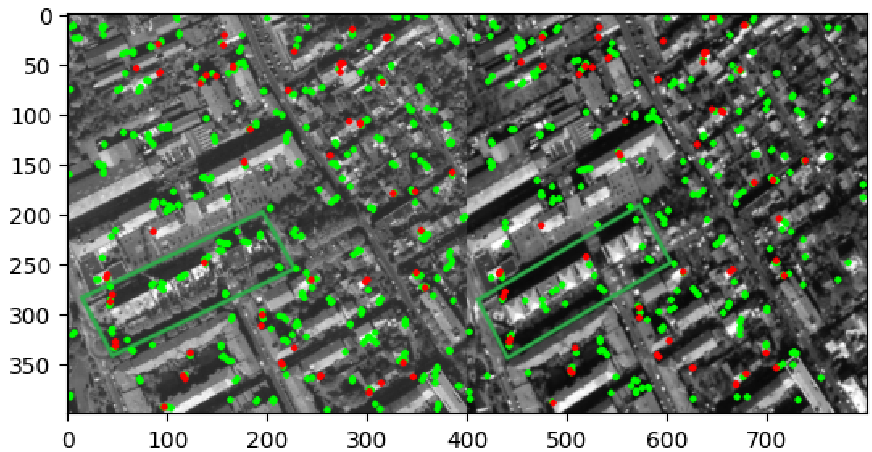
Applying this for each keypoint in the image will give a NFA for each, and the threshold ε should be kept below 10^{-10} to decide if a keypoint is classified as change. Further, enough points (around 30) within an area will constitute a detection.

<u>Exclusion principle:</u> few methods are used to reduce the redundancy in keypoints. First, an adapted version of the SIFT algorithm for SAR images is used (very logically called SAR-SIFT) to better handle noise from those non optical images. When applying this to the images, only detected keypoints found in the two images are considered to reduce the number of detections.
Afterwards, considering that images with / without change area are deformed by an affine transformation, the AC-RANSAC algorithm is used to remove false matches so it does not “pollute” the number of matches, since it is used to evaluate the density in the following steps when applying the *a contrario* approach. These two steps ensure that the number of detected features is reduced.

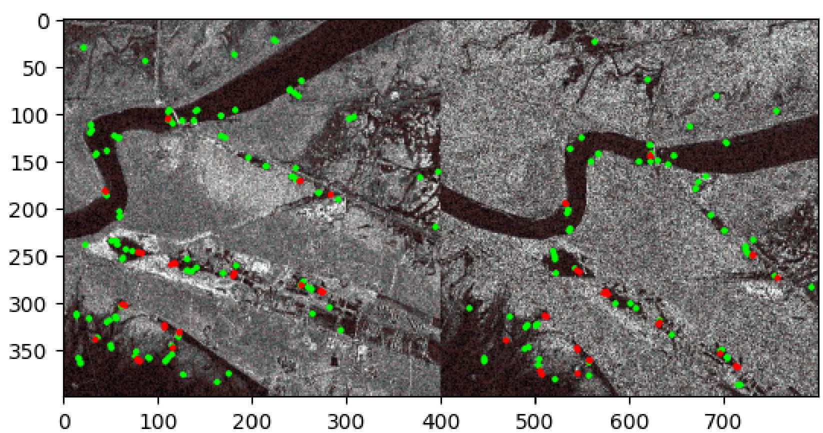
Further down the process, only a small area or radius *r* is considered each time, and an NFA is produced for each keypoint. The threshold on ε helps to classify as change only the areas where indeed a change happened between the two images. Finally, in practice, adding noise to the images will also help to avoid getting identical keypoints and descriptors.

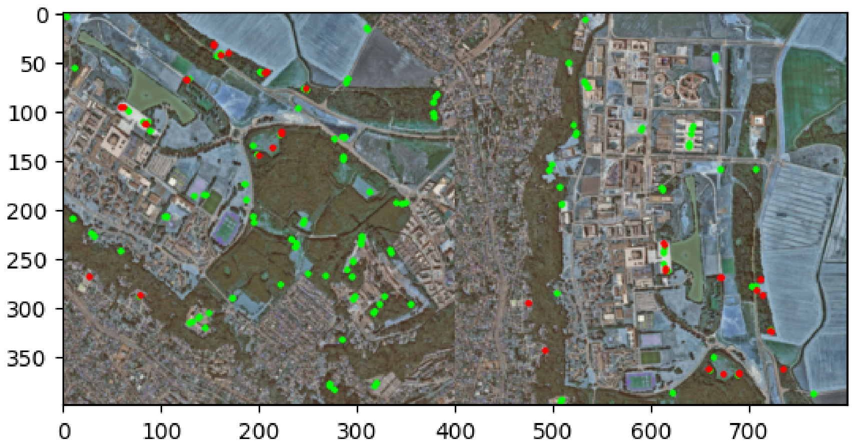
<u>Dataset for successful application:</u> As this is a two-step process (first identify the keypoints and then apply the *a contrario* approach), anything that would disturb the first part would render the calculation of an NFA useless. It is therefore interesting when comparing datasets to check the implementation we have made and display the keypoints first.

When reproducing for comparison the images from the paper with a lower resolution :

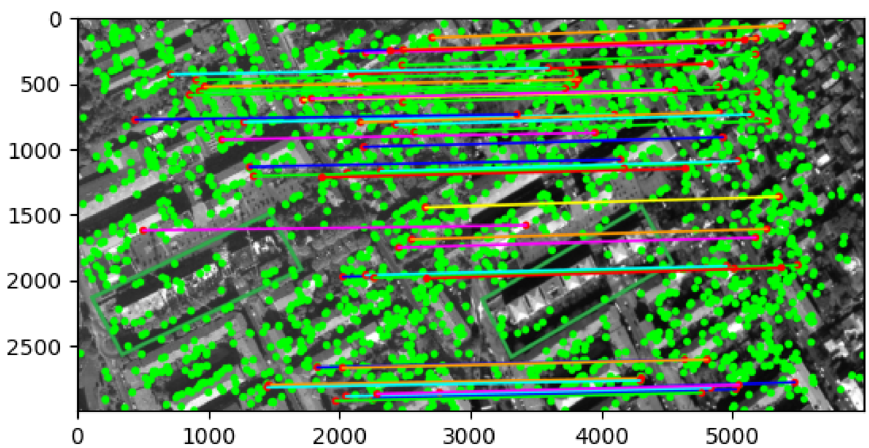


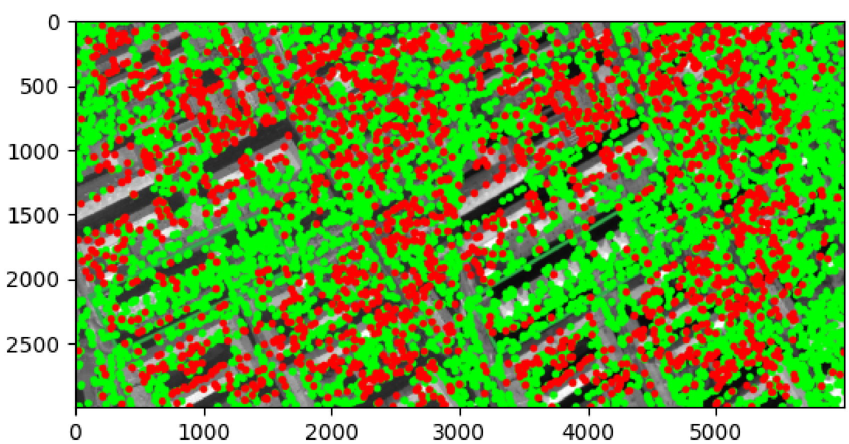
Compared to another couple of images of our choosing:

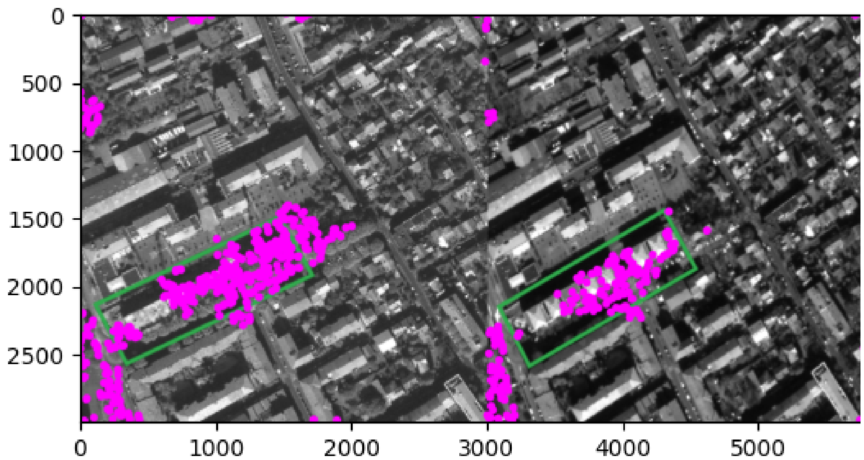




<u>Algorithm application, results and analysis:</u> to get comparable results to the paper, we need to increase the resolution and hence the number of descriptors (we get raw descriptors first, then match them from one image to another, and finally link them):



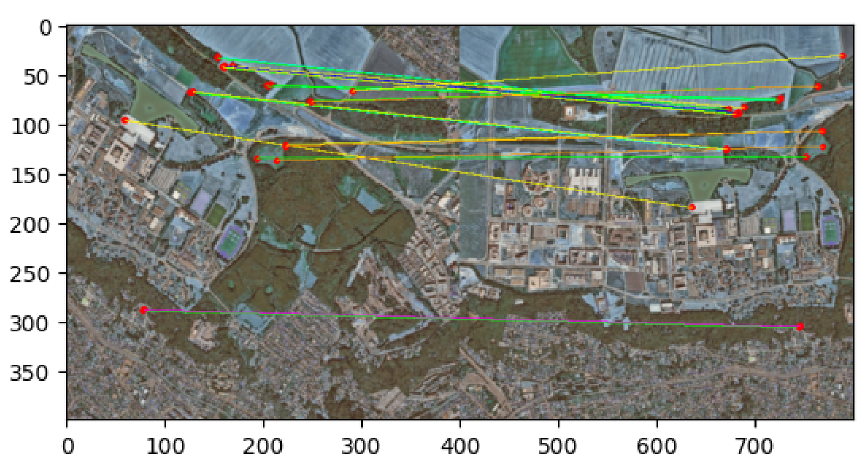


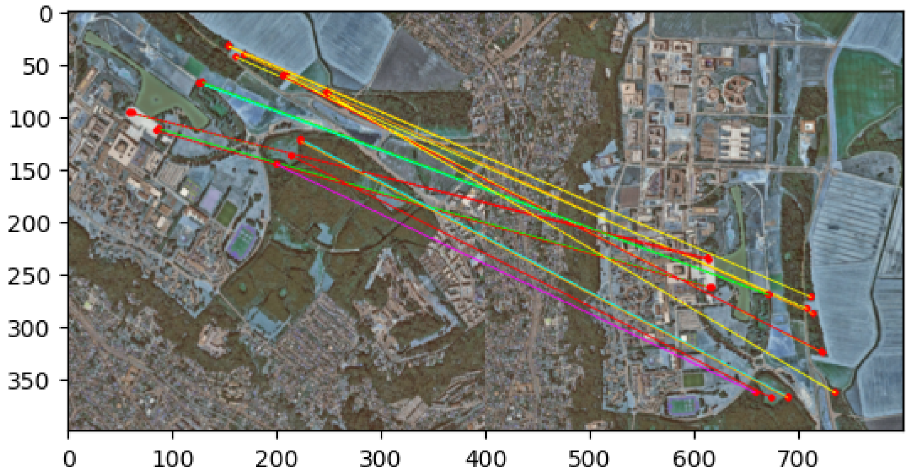


This yields very similar results to the paper we studied, using only KD trees and a localized approach of the ratio calculation provided by the *a contrario* methodology..

It should be noted that applying the NFA with the raw formula (on the global image) generates complex and (very high values) and would require heavy simplification to be calculated.

<u>Complex application / edge cases:</u> Since the main issue is first to get enough descriptors and then produce the right number of matches, we can imagine that a few cases will be tricky. For example if one uses images of different formats, the approach will not work properly. Low resolutions will also be an issue as we showed that higher resolution generated enough descriptors to apply the *a contrario* approach, and any resizing and operation that change proportions or detection parameters that would be too high would render this approach inefficient. The below shows examples where the keypoints identification does not work because of too much distortion from an image to another :





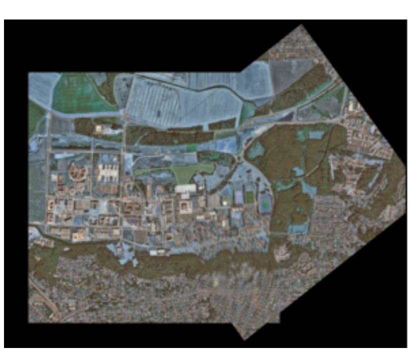
Trying to apply the *a contrario* approach here would not work.

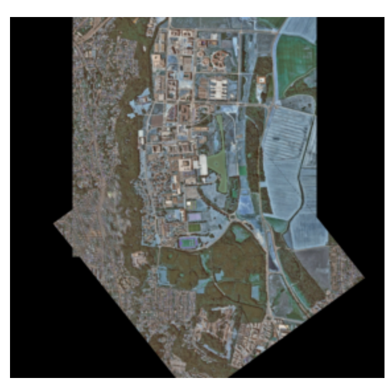
Furthermore, in the context of satellite images, any cloudy background could be considered as faulty detection and create a drop in matched keypoints that would follow in a wrong NFA calculation and a false change area to be detected.

<u>Limitations and potential improvements:</u> the paper suggests that automatic adaptation of the parameters could be an interesting new development. It is also worth noting that relying on SIFT is not necessarily the most effective feature representation method (a comparison with SURF or BRISK could be of interest). Here some other limitations could be linked to illumination artifacts or “changes” that are simply occlusion, as the image need to be taken in similar conditions, despite the potential use of various sensors/ angle of captures for those. where objects can be occluded by other objects.

Lastly, execution times that we found were logically directly linked to the size of the image, it should be noted that while a small image took around 90 seconds to have the descriptors processed and filtered, a higher resolution image (necessary to get enough points and proceed with the approach) took around 2 hours (with CPU) to process the various elements. Once the table of descriptors is obtained however the application of the local KD tree approach only took a few seconds.

Finally, since the results are good enough, another example of application can be to fuse two images based on the matches when the area to be covered can be extended based on the common area between the two images, such as displayed here:





References

[1] *Change detection for high resolution satellite images, based on SIFT descriptors and an a contrario approach*", 2014 for IEEE Geoscience and Remote Sensing Symposium, by F Dellinger, J Delon, Y Gousseau, J Michel, and F Tupin

[2] Agnes Desolneux, Lionel Moisan, and Jean-Michel Morel, *From Gestalt Theory to Image Analysis: A Probabilistic Approach*, Springer Publishing Company, Incorporated, 2007.