Contrast invariant image intersection

Coloma Ballester *
Eliane Cubero-Castan †
Manuel González ‡
Jean-Michel Morel §

Abstract

Given two images of the same scene taken at different times or with different cameras, we propose an "image intersection" method. Contrast invariant image intersection permits to generate a new image which does not depend upon the contrast of the cameras or the overall differences in illumination of the scene at different times. We call section of a digital image $u(x)$ any connected component of a bilevel set $\{x, \lambda \leq u(x) \leq \mu\}$. The sections are contrast invariant features of an image. The proposed method starts with two registered digital images 1 and 2 and constructs two intersection images, that is, a subimage of Image 1 whose sections are present in Image 2 and a subimage of Image 2 whose sections are present in Image 1.

*Dep. of Matematiques i Informatica, Universitat Illes Balears, Ctra. Valldemossa, Km. 7.5, 07071 Palma de Mallorca, Spain, dmicbn0@ps.uib.es

†CNES, CT/IA/QTIS, 18 avenue Edouard Belin 31055 Toulouse Cedex France, Eliane.Cubero-Castan@cst.cnes.fr

‡Dep. Matematiques i Informatica, Universitat Illes Balears, Ctra. Valldemossa, Km. 7.5, 07071 Palma de Mallorca, Spain, dmimgh0@ps.uib.es

§CMLA, Ecole Normale Supérieure de Cachan, 61 av du Président Wilson, 94 235 Cachan Cedex France, morel@ceremade.dauphine.fr

1 Introduction

Image comparison is one of the basic techniques of image analysis. In the case of satellite imaging, this task is both very relevant in cartography or change detection and very difficult because of the drastic changes of the ground aspect with time and weather. We shall see that image intersection can deal with images taken at time distances larger than months and even years. In addition, none of the techniques we shall discuss now yields an *image intersection*. Indeed, some of the existing techniques propose a registration, that is the search for common features in two images of the same scene. Now, these features (in general edges) do no permit an image reconstruction. Other techniques deal with image *fusion*, that is, the synthesis of several images into a more accurate or more informative one. Image *intersection* tries to define instead a common denominator to two images.

The most popular technique in order to compare two images is correlation [1], [7]. Using FFT, the correlation function is easily and fast computed and peak of the correlation function yields the right translation registering both images. [11] describes an efficient FFT based registration method between two images, one of them having been translated, rotated and rescaled. This kind of registration is difficult to apply to images taken at large time delays. Indeed, features of the image have changed and as a consequence peaks of the correlation function are smooth and do not even yield a clear registration. This is why many feature based image comparison techniques have been developped. The subjacent idea is the same as in pattern recognition : such features as edges, regions, lines are looked for in both images. Then a local or global correlation technique is applied to the feature maps [12]. As an example, in [3] and [6], registration of SPOT images is based on segmentation : a first edge detection followed by a connectivity algorithm is applied to each channel yielding regions. Then a distance energy between segmentations is minimized by simulated annealing to obtain a global matching of the three channels. In [8], instead, edge maps obtained by a standard Canny type edge detector of two images are compared by using a Haussdorff semidistance. Another variant is [5], where contours are obtained as zero-crossings of the Laplacian of a gaussian and edge comparison is based on maximal correlation along closed contours.

The mentionned feature based comparison techniques have several drawbacks and limitations. Indeed, most feature detections techniques assume a previous linear or nonlinear filtering of the image. In both cases, edge-sensitive filters highly depend upon the dynamics (grey level scale) of the image and therefore yield very different results for images differently contrasted. Now, multidate images have very different contrast and illumination. This makes feature detection based image comparison unfit for images with large time delays. This remark also applies to segmentation methods. All segmentation methods highly rely on the contrast of the image and detect a region only if its constrast with respect to neighboring regions is above an *a priori* threshold. In practice, this yields the so called *over- or subsegmentation* problem.

In addition to the contrast parameters, all mentionned feature or segmentation detection algorithms also introduce at least one additional scale parameter. The scale parameter fixes *grosso modo* the pass band of the filter, or equivalently, the spatial resolution of the method. In any case, this forces any feature based comparison method to compare not two single images, but two multiscale sequences of images. An interesting new class of features for comparison, the so-called (ϵ, δ)-connected components, has been proposed in [14]. It depends upon two parameters: ϵ is a contrast parameter and δ a smoothness parameter. This method avoids a previous filtering but is not contrast invariant if $\epsilon \neq \delta$. It is, however, the closest method to what is proposed here. We bring two structural improvements : contrast invariance and the possibility of constructing an intersection image.

Fusion techniques have another applicative framework than the one proposed here. Fusion methods [2], [10], [13] can be in general subsumed under the name of *multichannel segmentation*. They create from several sources a segmented image where each region inherits some attribute computed from the different channels. This attribute may be of very different kinds, like color, vegetation, type of cultivation, water, urban zone, forest, ploughed area, etc.

2 Image intersection method

In contrast to the mentioned fusion and feature comparison methods, the method we propose here permits to create a digital image *common* to two initial images of the same object taken from the same angle, but not necessarily at the same distance or at the same time. If this condition is not realized, a previous registration step is required; now, a variant of the intersection method itself might yield this registration ([9]). One of the applications is satellite imaging, but we shall also experiment on other kinds of multiimages. The intersection method is easily described once we have defined some adapted terminology.

2.1 Method glossary

- *Pixels* of a digital image. We denote the pixels by their integer coordinates, (i, j) where i stands for the row index and j for the column index.

- *Grey level.* We denote the grey level at pixel (i, j) by $u(i, j)$.

- *Connected component of a given set of pixels E.* We call connected component of E any maximal subset of pixels of E which cannot be divided into two disconnected pieces M and M'. We say that M and M' are disconnected if no pixel of M has any side in common with a pixel of M' (in case of 4-connectedness). One can also choose the 8-connectedness : then M and M' are disconnected if and only if no pixel of M has any vertex in common with a pixel of M'.

- *Bilevel set.* Let v and w be two grey level values such that $v \leq w$. We call bilevel set with levels v and w the set of all pixels (i, j) such that $v \leq u(i, j) \leq w$. Notation :

$$X_{v,w}u = \{(i, j),\ v \leq u(i, j) \leq w\}$$

- *Section of an image.* Let $v \leq w$ be two grey level values. We call section of the image with levels v and w any connected component of the bilevel set $X_{v,w}$.

- *Contrast change.* Any nondecreasing function $g(s)$. If u is an image, one obtains a new image $g(u(i, j))$ by applying the contrast change g. Notice that if g is strictly increasing, bilevel sets of u and $g(u)$ are globally the same. Indeed, $X_{g(v),g(w)}g(u) = X_{v,w}u$.

- *Reverse contrast.* Any nonincreasing function g. Notice, again, that if g is strictly decreasing, then each bilevel set and each section of u is equal to a bilevel set or section of $g(u)$ and conversely. Indeed, we obviously have

$$X_{g(w),g(v)}g(u) = X_{v,w}u.$$

To summarize, if we change or reverse the contrast of an image, then we keep exactly the same sections : only, their levels change. *The same remark applies if we do local contrast or reverse* *contrast changes, that is, contrast changes which are restricted to one section.* Then all subsections of this section are kept, up to level changes.

Main assumption of the intersection method:
We model the time changes in the illumination response of a given object, or the changes arising from one channel to the other, as some local contrast or reverse contrast change. The main argument in favour of this assumption is that most objects change colour over time and with changes of illumination, but keep their shape. *According to this assumption, we deduce that stable shape information on objects (which are connected) should be contained in sections* (see [4]).

2.2 Description of the intersection method

The intersection method looks for all sections which have remained stable in shape in two images and reconstructs an image whose sections belong to both initial images.

Intersection method (Part 1 : matching of sections).
a- Make a list of all sections of Image 1 and all sections of Image 2.
b- Given a distance between sets, $d(O_1, O_2)$ and an adaptive threshold $\varepsilon(O_1, O_2)$, for every section O_1 of Image 1, look for a section O_2 of Image 2 such that $d(O_1, O_2) \leq \varepsilon(O_1, O_2)$. If a corresponding section O_2 has been found in Image 2, then put the section O_1 into the list of *sections of Image 1 belonging to the intersection.*
c- In the same way, establish a list of all sections of Image 2 belonging to Image 1.
The main point about the intersection is that we have an image reconstruction algorithm from all sections found in common to two or more images.

Intersection method (Part 2 : construction of the intersection images).
We define two "subimages" of Image 1 which are common to both Image 1 and Image 2. We denote in general by $O_{v,w}$ a section of an image whose upper level is w, and lower level is v. Calling $\{S_k\}_k = \{O^k_{v_k,w_k}\}_k$ the set of sections of Image 1 which match with a section of Image 2, we define the *inf-intersection* of both images by

$$\underline{u}(i,j) = \sup\{v_k, \text{ such that } (i,j) \in S_k\}$$

(default value: 0). In the same way, we define the *sup-intersection* of both images by

$$\overline{u}(i,j) = \inf\{w_k, \text{ such that } (i,j) \in S_k\}$$

(default value: 0). In the same way, we define two subimages of Image 2 which are intersections of both images. The two subimages display the same information (i.e. the shape of the sections common to both images). Only the contrast of sections changes. In most cases, the visual aspect of $\underline{u}$ and $\overline{u}$ is quite similar and, in the experiments, we will display only one of them.

3 Fast Implementation

3.1 Comparison of the sections

Experimentations show that the simplest criterion to match sections also is the most efficient. We define the L^1-distance of two sections O_1 and O_2 as the area of their difference, that is, we take $d(O_1, O_2) = d^1(O_1, O_2)$ and $d^1(O_1, O_2) \approx |(O_1 \cup O_2) \setminus (O_1 \cap O_2)|$, where $|O|$ denotes the area of O.

We say that two sections O_1 and O_2 are comparable if

$$d^1(O_1, O_2) \le m_{D_1}$$

where m_{D_1} is a tolerance factor proportional (for regions large enough) to the sum of lengths of boundaries of O_1 and O_2. As an approximation we take

$$m_{D1} = \min\left\{15\%\left(|O_1| + |O_2|\right), 3.5\left(|O_1|^{1/2} + |O_2|^{1/2}\right)\right\}.$$

3.2 Fast preselection of sections

Each section has characteristics which we can use in a fast implementation to make a preselection of candidates matching pairs of sections. At an algorithmic level, we have chosen several of them in order to characterize the section: its area, its barycenter and its minimum and maximum coordinates (that is, if O is a section of a given image, we consider $x_M^O = \max\{x \ : \ (x,y) \in O\}$, $x_m^O = \min\{x \ : \ (x,y) \in O\}$, $y_M^O = \max\{y \ : \ (x,y) \in O\}$ and $y_m^O = \min\{y \ : \ (x,y) \in O\}$).

To sum up, we say that the sections O_1, O_2 are good candidates to matching (noted by $O_1 \sim O_2$) if

$$\begin{array}{c} |O_1| - m_A \le |O_2| \le |O_1| + m_A, \quad \|\text{barycenter}_{O_1} - \text{barycenter}_{O_2}\| \le m_B, \\ x_M^{O_1} - m_{mM} \le x_M^{O_2} \le x_M^{O_1} + m_{mM}, \quad x_m^{O_1} - m_{mM} \le x_m^{O_2} \le x_m^{O_1} + m_{mM}, \\ y_M^{O_1} - m_{mM} \le y_M^{O_2} \le y_M^{O_1} + m_{mM}, \quad y_m^{O_1} - m_{mM} \le y_m^{O_2} \le y_m^{O_1} + m_{mM}, \\ d^1(O_1, O_2) \le m_{D1}, \end{array}$$

where m_A, m_B, and m_{mM} are real tolerance factors in function of the square root of the O_1 area, excepting m_A which depends upon the O_1 area. Obviously, the relation $O_1 \sim O_2$ defined by these tolerance factors is only reflexive, but not symmetric or transitive. Thus, given O_1, in order to find the sections O_2 candidates to match with O_1, we take the tolerance factors done by: $m_A = \%|O_1|$, $m_B = \%\sqrt{|O_1|}$, $m_{mM} = \%\sqrt{|O_1|}$, and
$m_{D1} = \min\left\{15\%\left(|O_1| + |O_2|\right), 3.5\left(|O_1|^{1/2} + |O_2|^{1/2}\right)\right\}$, where the percentages in m_A, m_B and m_{mM} are fixed at the beginning of the algorithm.

Given O_1, in order to find the sections O_2 that are candidates to match with O_1, it is advantageous to have previously ordered all the sections according to one characteristic so that we can find in $\log N$ operations the set $\mathcal{S}$ of sections that have a similar characteristic, and to compare their characteristics with O_1. This last search has to be done for each section of $\mathcal{S}$, so that its cardinality should be the smallest possible. Thus, it is interesting to order the sections according to the characteristic that has the smallest associated tolerance factor: in practice, the area.

3.3 Fast image intersection

Let us explain the details for a computation architecture. This architecture permits to overcome two pitfalls: to store in memory all the sections of an image, which leads to storage problems which cannot be solved with the actual technology, and to avoid to repeat computation of the same connected component.

We start by describing an algorithm computing the connected components of a bilevel set $X_{v-t,w}$ from the connected components of $X_{v,w}$.

Algorithm 1

Let $X_{v,w}$ be the bilevel set with levels v and w of Image 1 (the same for Image 2). We assume that all connected components of $X_{v,w}$ are already computed and stored on RAM memory. Also we assume already stored in the dictionnary the parameters of the sections of $X_{v,w}$. Those parameters are stored also on RAM memory in order to be modified by giving the parameters of the sections

of $X_{v-t,w}$. The description of the $X_{v,w}$ sections can be accomplished in the following way: for each one of the connected components, say $O_1, O_2, \ldots, O_N$, of $X_{v,w}$, we choose a label. For instance, one can choose a particular pixel belonging to the section. Let $x_1, x_2, \ldots, x_N$ be its pairs of coordinates. We construct an image $z(x)$ of the bilevel set in such a way that $z(x) = 0$ if x does not belong to the bilevel set and $z(x) = x_k$ if x belongs to O_k, $k \in \{1, \ldots, N\}$. In order to compute the bilevel set $X_{v-t,w}$ we proceed as follows:
1- Generate an image $z'(x)$ containing the pixels with grey level value between $v - t$ and v, which are going to be added to the bilevel $X_{v,w}$ in order to give $X_{v-t,w}$. For instance, let $z'(x) = 1$ if $v - t \leq \text{Image } 1(x) < v$ and let $z'(x) = 0$ in other case.
2- For each x such that $z'(x) = 1$:

- If all the neighbors of x have value in z equal to 0, create a new connected component and assign to it the coordinates of x as label. Create and compute its parameter.
- If at least one neighbor of x has non zero value in z and all the neighbors of x have the same value (that is, they belong to the same connected component, say O_1), then add x to this connected component and actualize the parameter of O_1.
- If two neighbors of x belong to different connected components, say O_1 and O_2, then proceed to the fusion of those connected components, that is, set $z(x) = x_1$ for all the pixels in O_2 and compute again the parameter of the new region (it is basically accomplished by a simple addition of the parameters of O_1, O_2 and x: for instance, the new area is the sum of the areas of O_1, O_2 plus 1).

When we have considered all the pixels x such that $z'(x) = 1$, we obtain the connected components of the bilevel set $X_{v-t,w}$ The connected components of $X_{v,w}$ have been deleted but the parameters (characteristics) of each section have been stored in memory.
Algorithm iteration.
We select a quantification step QS, 1 or a larger amount. As an example, let us take 12.

Initialization
Apply algorithm 1 with $z(x) = 0$, $v = 255 - 12 = 243$ and $w = 255$. We obtain the sections of bilevel set $(243, 255)$.

Iteration
Compute step by step, using algorithm 1, the sections and the characteristics of the bilevel sets from $(255-2\times12, 255)$, $(255-3\times12, 255)$, to $(3, 255)$. Next, change 255 into $255-12$ and repeat the process to extract the sections of bilevel sets from $(255-24, 255-12)$, $(255-36, 255.-12)$, to $(3, 255-12)$. And so on.

4 Image intersection experiments

4.1 Image description

In Experiment 1, we took two gray-level images (of size 320×240) of a room where some objects have been removed or added, taken with 49 hours difference. Experiment 2 and Experiment 3 are led with gray-level images (of size 150×150) of multispectral SPOT images (XS1,XS2,XS3). The images used in the experiments are a image subset of the original Spot images (of size 2000×2000 pixels) of places around Colmar and Farsala and taken at differents days and months from 1994 to 1996. We compared two different spectral components of the same satellite image or the same

	Experiment 1		Experiment 2		Experiment 3	
Images/channel date	Image 1	Image 2	XS1 94/3/20	XS2 94/3/20	XS1 95/7/7	XS1 96/7/19
nb. of sections	12819	10753	14064	18843	6202	3944
correspondances (c)-(d) correspondances (e)-(f)	3942	3892	4851	5858	348 216	294 213
percentage (c)-(d) percentage (e)-(f)	30.75	36.19	34.49	31.09	5.6 3.48	7.45 5.4

Table 1: The number of sections, of correspondances and the percentage of localized sections for each pair of images and each experiment

spectral component of two satellite images taken at one year delay. In the experiments below the quantification step was $QS = 2$.

4.2 Representative sections

Because of the sampling effects, too small sections are not relevant, and the same is true for too large sections because of boundary effects. So we fix (optionally in the program) two thresholds Area_{min} and Area_{max} on the area of the sections to be considered as representative. We took $\text{Area}_{\text{min}} = 6, \quad \text{Area}_{\text{max}} = \frac{\text{Area(image)}}{2}$. In the following, section should be understood as representative section.

On the other hand, notice that the same section can appear in different bilevel sets if it is sufficiently contrasted relative to its neighborhood. We have the choice to consider it as one section, or to consider it as n different sections, where n is the number of bilevel sets it appears in. In Table 1, we consider it as an unique section. The number of sections for some of the images used in the experiments is given in Table 1.

4.3 Experiments

In all the experiments that we displayed, the tolerance factor m_{D1} is the one described in Subsection 3.1 and the tolerance factors are $m_A = 20\%|O_1|$, $m_B = 20\%\sqrt{|O_1|}$, $m_{mM} = 20\%\sqrt{|O_1|}$. The number of correspondances for each pair of images and each experiment is shown in Table 1.

In Experiment 1 (Figure 1), we took two images (of size 320×240) of a room where some objects have been removed or added, taken with 49 hours difference. In Experiment 2 (Figure 4), we make the intersection of XS1 and XS2 SPOT channels image of Farsale, date 94/3/20. In Experiment 3 (Figure 5), we intersect the XS1 channel of two images (of Colmar) taken with one year of difference (one date is 95/7/7 and the other is 96/7/19). In Table 1 we display the number of sections, the number of correspondances and the percentage of localized sections.

Execution time is less than 1 minute with a standard workstation.

References

[1] P.E. Anuta, Spatial Registration of Multispectral and Multitemporal Digital Imagery Using Fast Fourier Transform Techniques. GeoEl(8), October 1970, pp. 353-368.

[2] A. Baraldi, F. Parmiggiani. Single Linkage Region Growing Algorithms Based on the Vector Degree of Match. *IEEE Trans. Geosci. Remote Sensing*, 34(1):137–148, 1996.

[3] P. Bourret, B. Cabon, E. Cubero-Castan, E. Schaeffer, J.-Y. Rousselot. A neural approach for satellite image registration and pairing segmented areas. In *EUROPTO*, pages 25–28, Paris, Sep. 1995.

[4] V. Caselles, B. Coll, J.-M. Morel. Topographic maps. Preprint, CEREMADE, 1997.

[5] X. Dai, S. Khorram, H. Cheshire. Automated Image Registration for Change Detection from Landsat Thematic Mapper Imagery. In *Proc. Int. Geoscience and Remote Sensing Symp.* 1996, vol. 3, pp. 1609-1611.

[6] M.-L. Duplaquet, E. Cubero-Castan. Updating cartographic models by Spot images interpretation. In *European Symposium on Satellite Remote Sensing*, Roma, Sep 26-30, 1994.

[7] J. Hsieh, H. Liao, K. Fan, M. Ko. A Fast Algorithm for Image Registration without Predetermining Correspondences. ICPR96(A90.1). 9608.

[8] D.P. Huttenlocher, G.A. Klanderman, W.J. Rucklidge. Comparing Images Using the Hausdorff Distance. *IEEE Trans. on Pattern Analysis and Machine Intell.*, IEEE-PAMI(15), 9:850–863, 1993

[9] P. Monasse, F. Guichard. Fast computation of a contrast-invariant image representation. To appear.

[10] A. Murni, A.K. Jain, J. Rais. A Framework for Multi-date Multi- sensor Image Interpretation. In *Proc. Int. Geoscience and Remote Sensing Symp.* 1996, vol. 3, pp. 1851-1854.

[11] B.S. Reddy, B.N. Chatterji. An FFT based technique for translation, rotation and scale-invariant image registration. IEEE-IP (5), n8, August 1996, pp. 1266-1271.

[12] J. Sato, R. Cipolla. Image Registration Using Multiscale Texture Moments. IVC(13), No. 5, June 1995, pp. 341-353. (A2)

[13] A.H.S. Solberg, T. Taxt, A.K. Jain. A Markov Random Field Model for Classification of Multisource Satellite Imagery. *IEEE Trans. Geosci. Remote Sensing*, 34(1):100–113, 1996.

[14] Yang Wang, Prabir Bhattacharya. Hierarchical stereo correspondence using features of gray connected components. *ICIP, Santa Barbara*, October 26-29, 1997, pp.264-267.

(a) Image 1 : original room image, $(320, 240)$

(b) Image 2 : original room image, $(320, 240)$

Figure 1: Experiment 1: in this experiment we make the intersection of two images of a room with 49 hours difference. In (a), (b) we display the original images.

(c) *sub_intersection* $\bar{u}$, from Image 1

(d) *sub_intersection* $\bar{u}$, from Image 2

Figure 1: Experiment 1: In (c), (d) we display the *sub_intersection* of both images.

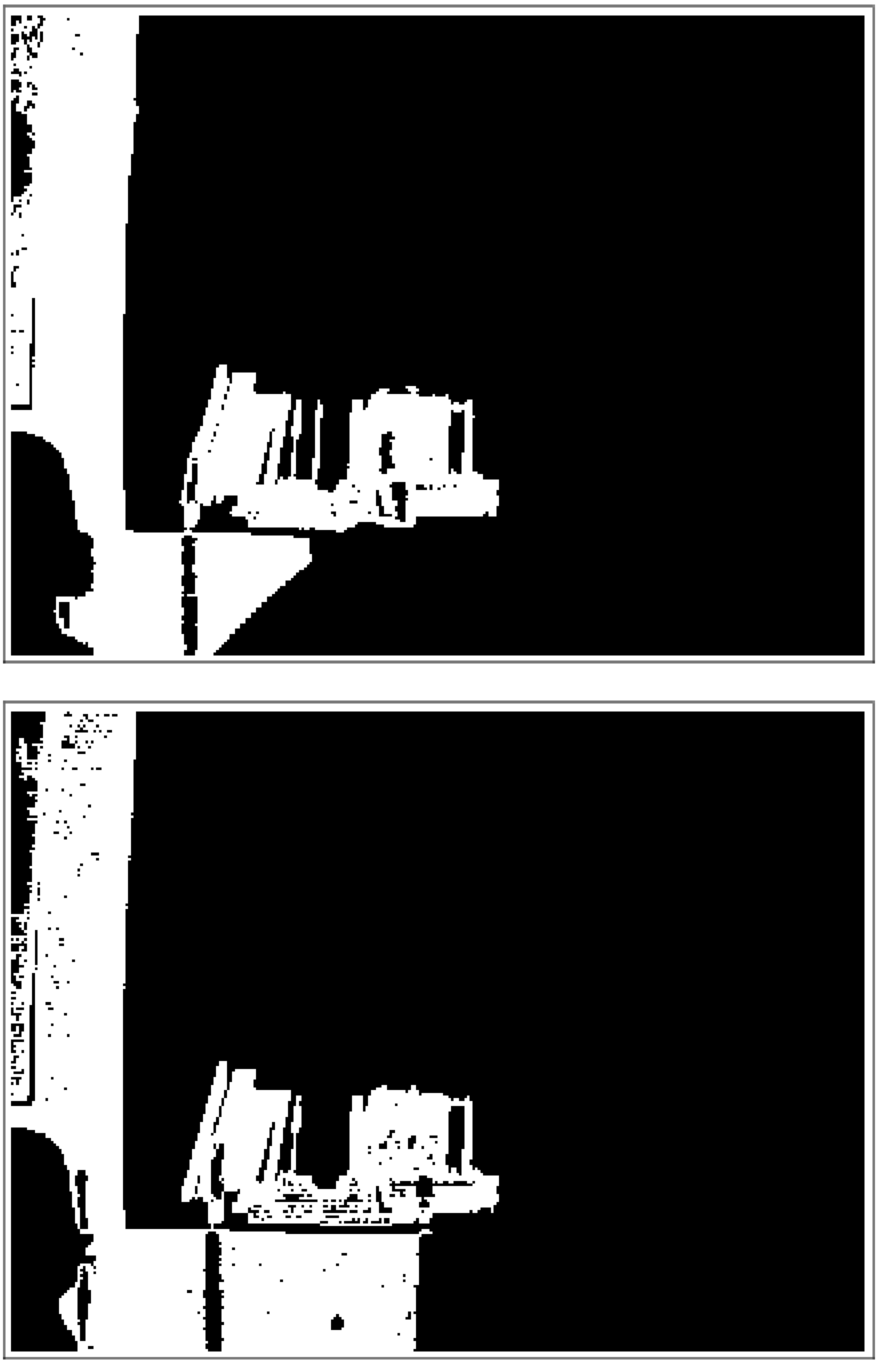


Figure 2: Two sections not matched in experiment 1. Upper: section of the bilevel set $\{x\ :\ 128 \leq u(x) \leq 224\}$ from Figure 1 (a). Down: section of the bilevel set $\{x\ :\ 136 \leq u(x) \leq 216\}$ from Figure 1 (b).

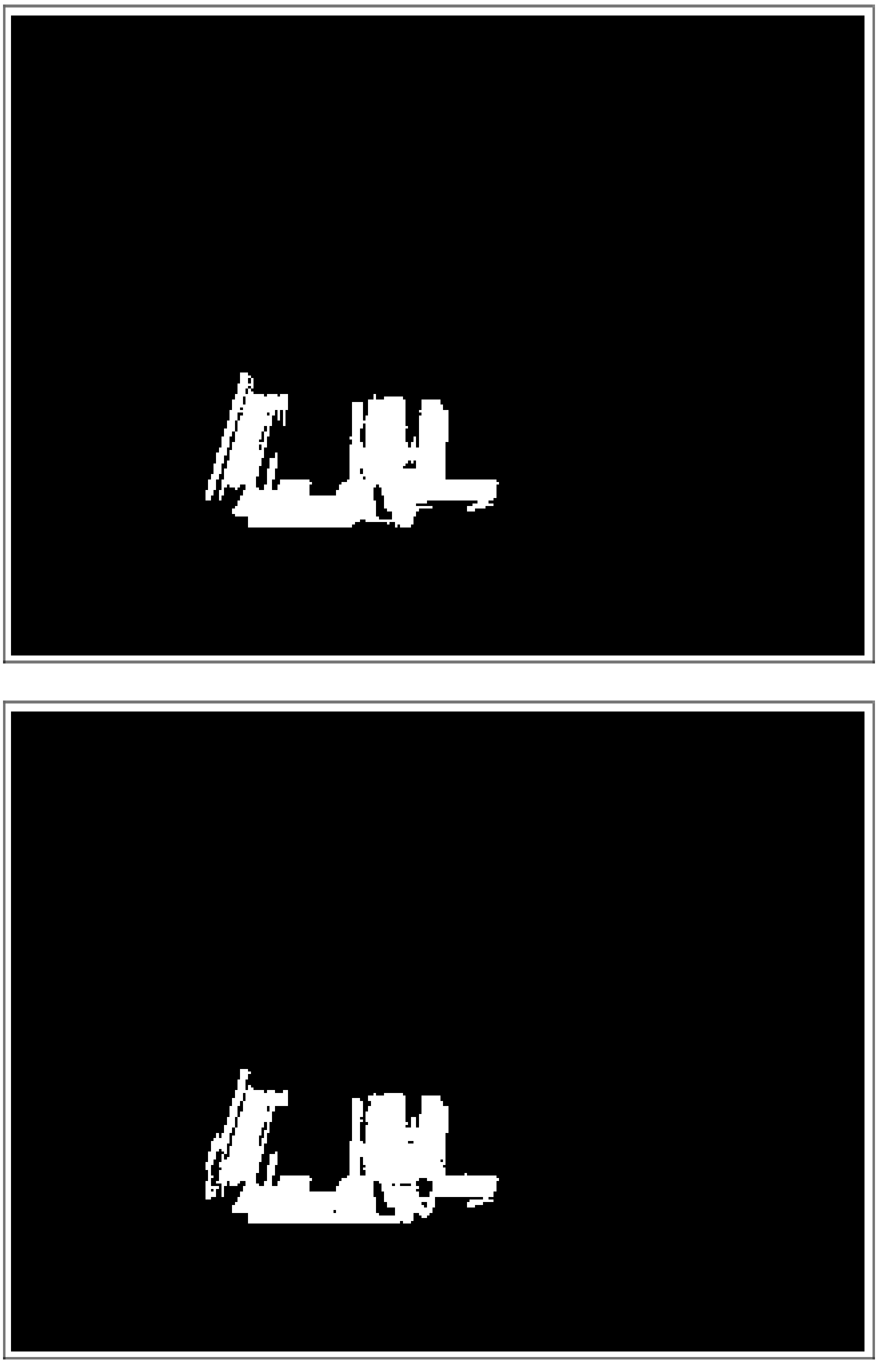
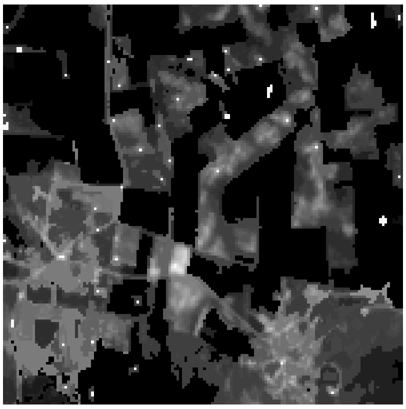


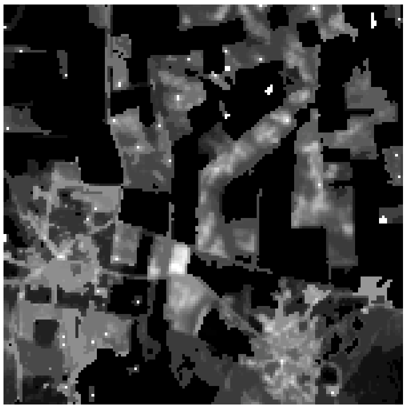
Figure 3: Two matched sections in experiment 1 (Figure 1). Upper: section of bilevel set $\{x : 176 \leq u(x) \leq 232\}$ from Figure 1 (a). Down: section of bilevel set $\{x : 176 \leq u(x) \leq 224\}$ from Figure 1 (b).

(a) Image 1 : original SPOT XS1 image, $(150, 150)$-7-95

(b) Image 2 : original SPOT XS2 image, $(150, 150)$-7-96

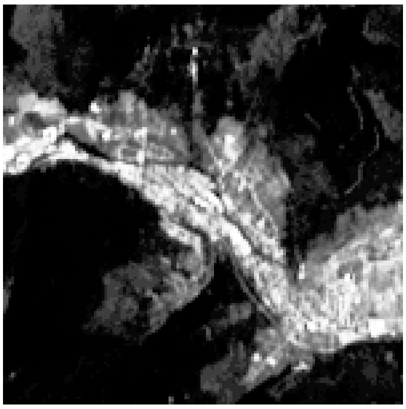


(c) *sup-intersection* $\overline{u}$, from Image 1

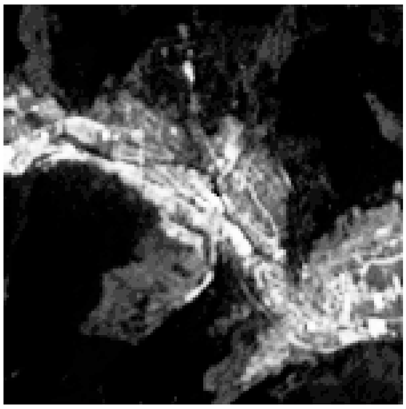


(d) *sup-intersection* $\overline{u}$, from Image 2

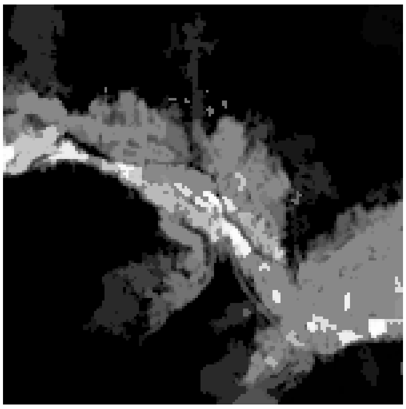
Figure 4: Experiment 2: intersection of XS1 and XS2 SPOT channel image of Farsala, date 94/3/20. Upper: The original images. Down: the *sup_intersection* of both images.



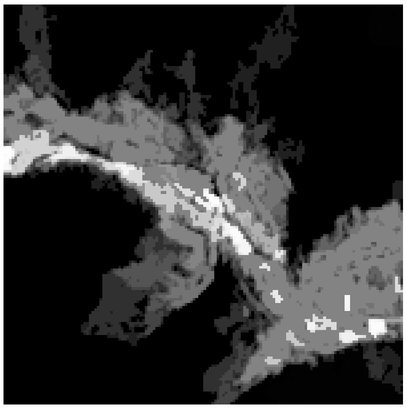
(a) Image 1 : original SPOT XS1 image, (150, 150)-7-95



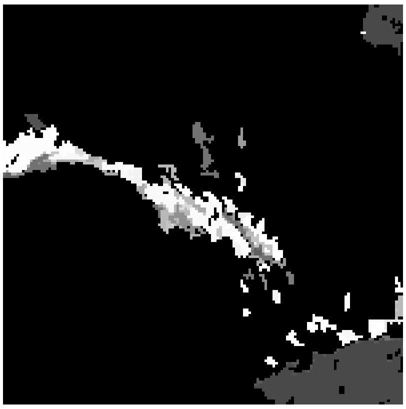
(b) Image 2 : original SPOT XS2 image, (150, 150)-7-96



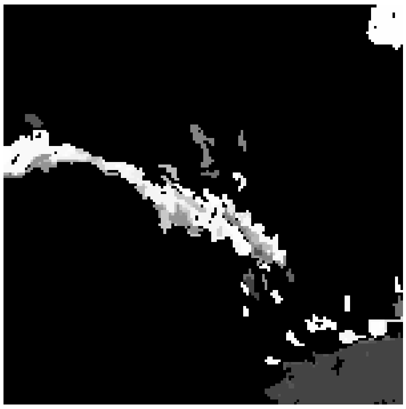
(c) *sup-intersection* $\overline{u}$, from Image 1



(d) *sup-intersection* $\overline{u}$, from Image 2



(e) *sup-intersection* $\overline{u}$, from Image 1



(f) *sup-intersection* $\overline{u}$, from Image 2

Figure 5: Experiment 3: intersection of XS1 channel of two images of Colmar taken with one year of difference. Upper: the original images. Middle: the *sup_intersection* of both images using a fixed tolerace factor $m_{D_1} = \frac{15}{100}(|O_1| + |O_2|)$. Down: the *sup_intersection* of both images using the tolerance factor described in Section 3.