On the possibility of automatic multisensor image registration

Jordi Inglada, Giros Alain

▶ **To cite this version:**

Jordi Inglada, Giros Alain. On the possibility of automatic multisensor image registration. IEEE Transactions on Geoscience and Remote Sensing, 2004, 42 (10), pp.2104-2120. hal-00569950

HAL Id: hal-00569950
https://hal.science/hal-00569950

Submitted on 1 Mar 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On the Possibility of Automatic Multisensor Image Registration

Jordi Inglada and Alain Giros

***Abstract*—Multisensor image registration is needed in a large number of applications of remote sensing imagery. The accuracy achieved with usual methods (manual control points extraction, estimation of an analytical deformation model) is not satisfactory for many applications where a subpixel accuracy for each pixel of the image is needed (change detection or image fusion, for instance). Unfortunately, there are few works in the literature about the fine registration of multisensor images and even less about the extension of approaches similar to those based on fine correlation for the case of monomodal imagery. In this paper, we analyze the problem of the automatic multisensor image registration and we introduce similarity measures which can replace the correlation coefficient in a deformation map estimation scheme. We show an example where the deformation map between a radar image and an optical one is fully automatically estimated.**

***Index Terms*—Image registration, multisensor, similarity measures.**

I. Introduction

THE PROBLEM we want to deal with is the one of the automatic fine registration of images acquired with different sensors. By different sensors, we mean sensors that produce images with different radiometric properties, i.e., sensors which measure different physical magnitudes: optical sensors operating in different spectral bands, radar and optical sensors, etc.

For this kind of image pairs, the classical approach of fine correlation [1], [2], cannot always be used to provide the required accuracy, since this similarity measure (the correlation coefficient) can only measure similarities up to an affine transformation of the radiometries.

There are two main questions which can be asked about what we want to do.

1) Can we define what the similarity is between, for instance, a radar and an optical image?
2) What does *fine registration* mean in the case where the geometric distortions are so big and the source of information can be located in different places (e.g., the same edge can be produced by the edge of the roof of a building in an optical image and by the wall–ground bounce in a radar image)?

We can answer by saying that the images of the same object obtained by different sensors are two different representations of the same reality. For the same spatial location, we have two different measures. Both items of information come from the same source, and thus, they have a lot of common information. This relationship may not be perfect, but it can be evaluated in a relative way: different geometrical distortions are compared, and the one leading to the strongest link between the two measures is kept.

The paper is organized as follows. Section II is a review of the existing remote sensing image registration literature. In Section III, we introduce a theoretical approach to image registration. The problem of modeling image deformations[1] is analyzed in Section IV. In Section V, we evaluate a set of similarity measures that can be used for the multisensor image registration problem, and we use one of them in Section VI in order to estimate the deformations between a radar image and an optical image of the same scene. Finally, in Section VII, we propose the use of deformation maps for the estimation of topography using radar and optical acquisitions.

II. Review of Existing Work

In this section, we will review the works published in the literature about the automatic multisensor image registration. The literature about the subject is rather limited in the field of remote sensing compared to what has been published in the fields of medical imaging and computer vision. An interesting survey of image registration techniques can be found in [3]. This survey poses the problem of image registration using the concepts of similarity measure, geometric transformations, and feature space. We will take a similar approach in Section III.

The approach taken in most of the works consists in automatically extracting homologous points (HPs) in both images and using them to estimate a parametric analytical deformation model.

Ton and Jain [4] were among the pioneers in the research of automatic algorithms to emulate photointerpreter-based registration. They proposed an algorithm for HP selection allowing for the estimation of rotational and translational transformations on Landsat images. Their approach is similar to the one by Li *et al.* [5], who used the salient points of active contours as HPs. The main problem of this approach is the heavy computation needed for the implementation of active contours.

Several authors, as for instance Cracknell and Paithoonwattanakij [6], combine an orbital model and a HP search using correlation and heuristic planning. The idea consists in searching for couples of HP with a good correlation. The set of HPs selected is used to solve a least squares estimation of a

Manuscript received October 6, 2003; revised July 6, 2004.

The authors are with the Centre National d'Études Spatiales, DCT/SI/AP, F-31401 Toulouse Cedex 9, France (e-mail: jordi.inglada@cnes.fr).

Digital Object Identifier 10.1109/TGRS.2004.835294

[1]We prefer the term *deformation* to the classical *disparity* because the latter is related to stereo-optical vision. In the case of optical-radar image pairs, the geometric relative deformations due to topography (depth) do not correspond to that kind of model. We will use the term *deformation* as a generalization of *disparity*.

0196-2892/04$20.00 © 2004 IEEE

parametric model. They claim to have subpixel accuracy on the National Oceanic and Atmospheric Administration's Advanced Very High Resolution Radiometer images.

Within the family of methods using automatic correlation, one can highlight the work of Foroosh *et al.* [7], where a closed-form expression for subpixel shift estimation is given. However, this approach can only be applied to translations. It is also interesting to point out the work of Stone *et al.* [8] where an algorithm for subpixel shift registration based on Fourier transforms is presented. This algorithm has the advantage of being very fast and robust to aliasing.

These approaches are not really multisensor in the sense that they use a similarity measure which is not. A way to transform these approaches into multisensor ones is by comparing extracted image primitives. For example, Inglada and Adragna [9] use a simple edge detection and a genetic algorithm in order to find the best set HP in the case of a Système Pour l'Observation de la Terre (SPOT)–European Remote Sensing Satellite (ERS) registration. The control points of the master image are randomly taken amongst the extracted edges.

Thépaut *et al.* [10] do a first geometrical correction using orbit information, and then a residual translation compensation using the correlation between the edges extracted from both, ERS synthetic aperture radar (SAR), and SPOT images. Other approaches exist for edge matching, as for instance the one of Wu and Maître [11], where a multiresolution analysis is used together with a hypothesis testing.

One of the first works on feature-based image registration was proposed by Ventura *et al.* [12]. They even applied it to the problem of image to map registration. The approach was also finding HP by matching extracted features.

Dai and Khorram [13] use a feature-based approach: they extract closed edges that are characterized using invariant moments. Then, the extracted areas are matched using their characterization. Finally, the centers of gravity of each area are used as HPs for the estimation of an affine transformation. They apply the approach to Landsat images, and they obtain an accuracy better than one pixel, which is similar to the accuracy obtained with manual registration.

Djamdji *et al.* [14] propose a multiresolution approach, where the discrete wavelet transform is used. The automatic extraction of HPs is done by comparing thresholded wavelet coefficients.

All these approaches try to extract HP in order to compute an analytical deformation model. On the other hand, when working with images acquired with the same (type of) sensor, one can use a very effective approach. Since a correlation coefficient measure is robust and fast for similar images, one can afford to apply it in every pixel of one image in order to search for the corresponding HP in the other image. One can, thus, build a deformation grid (a sampling of the deformation map). If the sampling step of this grid is short enough, the interpolation using an analytical model is not needed, and high-frequency deformations can be estimated. The obtained grid can be used as a resampling grid and, thus, obtain the registered images.

No doubt, this approach, combined with image interpolation techniques (in order to estimate subpixel deformations) and multiresolution strategies, allows for obtaining the best performances in terms of deformation estimation and, hence, for the automatic image registration.

Unfortunately, in the multisensor case, the correlation coefficient cannot be used. This will be justified in Section V-B. We will, thus, try to find similarity measures that can be applied in the multisensor case with the same approach as the correlation coefficient.

III. Model for the Image Registration Problem

In this section, we give several definitions that allow for the formalization of the image registration problem. First of all, we define the master image and the slave image.

Definition 1 (Master Image): Image to which other images will be registered. Its geometry is considered as the reference.

Definition 2 (Slave Image): Image to be geometrically transformed in order to be registered to the master image.

Two main concepts are the one of *similarity measure* and the one of *geometric transformation*.

Definition 3: Let I and J be two images, and let c be a similarity criterion. We call a similarity measure any scalar strictly positive function

$$S_c(I, J) = f(I, J, c). \tag{1}$$

S_c has an absolute maximum when the two images I and J are *identical* in the sense of criterion c.

Definition 4: A geometric transformation T is an operator that, applied to the coordinates (x, y) of a point in the slave image, gives the coordinates (u, v) of its HP in the master image

$$\begin{pmatrix} u \\ v \end{pmatrix} = T \begin{pmatrix} x \\ y \end{pmatrix}. \tag{2}$$

Finally, we introduce a definition for the image registration problem.

Definition 5 (Registraton Problem):

1) Determine a geometric transformation T that maximizes the similarity between a master image I and the result of the transformation $T \circ J$

$$\text{Arg} \max_T (S_c(I, T \circ J)). \tag{3}$$

2) Resampling of J by applying T.

We must note that Le Moigne *et al.* have proposed in a recent contribution [15] a modular approach for registration that allows the analysis of different similarity measures and different optimization strategies. The presented results, which are still preliminary, are very promising. The multisensor case has been dealt with, but only for optical images (Ikonos and Landsat Enhanced Thematic Mapper Plus). The case of very different images (e.g., optical and radar) has not been explored.

IV. Geometric Deformation Modeling

The geometric transformation of Definition 4 is used for the correction of the existing deformation between the two images to be registered. This deformation contains information, which

TABLE I
CHARACTERIZATION OF THE GEOMETRIC DEFORMATION SOURCES

	Intensity	Spatial Frequency
Mean Attitude	Strong	Low
Stereo	Medium	High and Medium
Attitude evolution	Low	Low to Medium

is linked to the observed scene and the acquisition conditions. The deformations can be classified into the following three classes, depending on their physical source:

1) deformations linked to the mean attitude of the sensor (incidence angle, presence or absence of yaw steering, etc.);
2) deformations linked to a stereo vision (mainly due to the topography);
3) deformations linked to attitude evolution during the acquisition (vibrations that are mainly present in pushbroom sensors).

These deformations are characterized by their spatial frequencies and intensities, which are summarized in Table I.

Depending on the type of deformation to be corrected, its model will be different. For example, if the only deformation to be corrected is the one introduced by the mean attitude, a physical model for the acquisition geometry (independent of the image contents) will be enough. If the sensor is not well known, this deformation can be approximated by a simple analytical model. When the deformations to be modeled are high frequency, analytical (parametric) models are not suitable for a fine registration. In this case, one has to use a fine sampling of the deformation, which means the use of deformation grids. These grids give, for a set of pixels of the master image, their location in the slave image.

The following points summarize the problem of the deformation modeling.

1) An analytical model is just an approximation of the deformation. It is often obtained as follows:
   a) directly from a physical model without using any image content information;
   b) by estimation of the parameters of an *a priori* model (polynomial, affine, etc.). These parameters can be estimated
      i) either by solving the equations obtained by taking HP. The HP can be manually or automatically extracted;
      ii) or by maximization of a global similarity measure.
2) A deformation grid is a sampling of the deformation map.

The last point implies that the sampling period of the grid must be short enough in order to account for high-frequency deformations (Shannon theorem). Of course, if the deformations are nonstationary (it is usually the case of topographic deformations), the sampling can be irregular.

As a conclusion, we can say that Definition 5 poses the registration problem as an optimization problem. This optimization can be either global or local with a similarity measure, which can also be either local or global. All this is synthesized in Table II.

TABLE II
APPROACHES TO IMAGE REGISTRATION

Geometric model	Similarity measure	Optimization of the deformation
Physical model	None	Global
Analytical model with a priori HP	Local	Global
Analytical model without a priori HP	Global	Global
Grid	Local	Local

The ideal approach would consist in a registration that is locally optimized, both in similarity and deformation, in order to have the best registration quality. This is the case when deformation grids with dense sampling are used. Unfortunately, this case is the most computationally heavy, and one often uses either a low sampling rate of the grid or the evaluation of the similarity in a small set of pixels for the estimation of an analytical model.[2] Both of these choices lead to local registration errors, which, depending on the topography, can amount to several pixels.

Even if this registration accuracy can be enough in many applications, (orthoregistration, import into a geographic information system, etc.), it is not acceptable in the case of data fusion, multichannel segmentation, or change detection [16]. This is why we will focus on the problem of deformation estimation using dense grids.

None of the references presented in Section II uses the local optimization approach. We can also note that in the multisensor case only few authors [15] have used any similarity measure other than the correlation coefficient. However, in the medical imaging field, as we will see in Section V, a lot of similarity measures have been proposed as a generalization of the correlation coefficient. These measures enable the registration of very different imagery modalities. Nevertheless, these works are not directly usable in our problem, since the geometric deformations present in medical images can be easily represented by global analytical models. Indeed, often a rigid model (rotation, translation, scale) or slightly elastic (affine plus a $a \cdot x \cdot y$ term) is enough, since: 1) the sensors are stable; 2) the stereo effect is small; 3) and only the point of view changes. As we have noted above, deformations due to topography can locally have high frequencies for medium- and high-resolution sensors (30 m and better), thus our need for a fine modeling. We also point out that the problem of hidden faces is beyond the scope of this paper.

V. SIMILARITY MEASURES

The fine modeling of the geometric deformation we are looking for needs for the estimation of the coordinates of nearly every pixel in the master image inside the slave image. In the classical monosensor case where we use the correlation coefficient, we proceed as follows.

[2]However, computation time is not nowadays a real issue. As an illustration, we can give the example of the PAN+XS fusion procedure developed at the Centre National d'Études Spatiales (CNES) for the SPOT 5 ground segment where the subpixel registration between the Panchromatic band and the multispectral channels is done on 24 000 × 24 000 images in less than 1 h on a Sun Sparc Ultra-4 workstation.

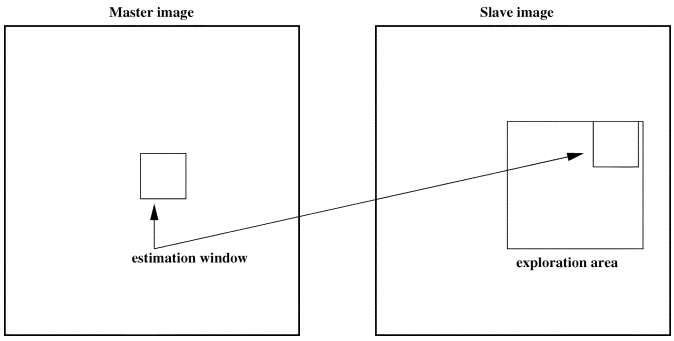


Fig. 1. Estimation of the correlation surface.

The geometric deformation is modeled by local rigid displacements. One wants to estimate the coordinates of each pixel of the master image inside the slave image. This can be represented by a displacement vector associated to every pixel of the master image. Each of the two components (lines and columns) of this vector field will be called deformation grid.

We use a small window taken in the master image, and we test the similarity for every possible shift within an exploration area inside the slave image (Fig. 1).

That means that, for each position, we compute the correlation coefficient. The result is a correlation surface whose maximum gives the most likely local shift between both images

$$\rho_{I,J}(\Delta x, \Delta y) = \frac{1}{N} \cdot \frac{\sum_{x,y}(I(x,y) - m_I)(J(x+\Delta x, y+\Delta y) - m_J)}{\sigma_I \sigma_J}. \quad (4)$$

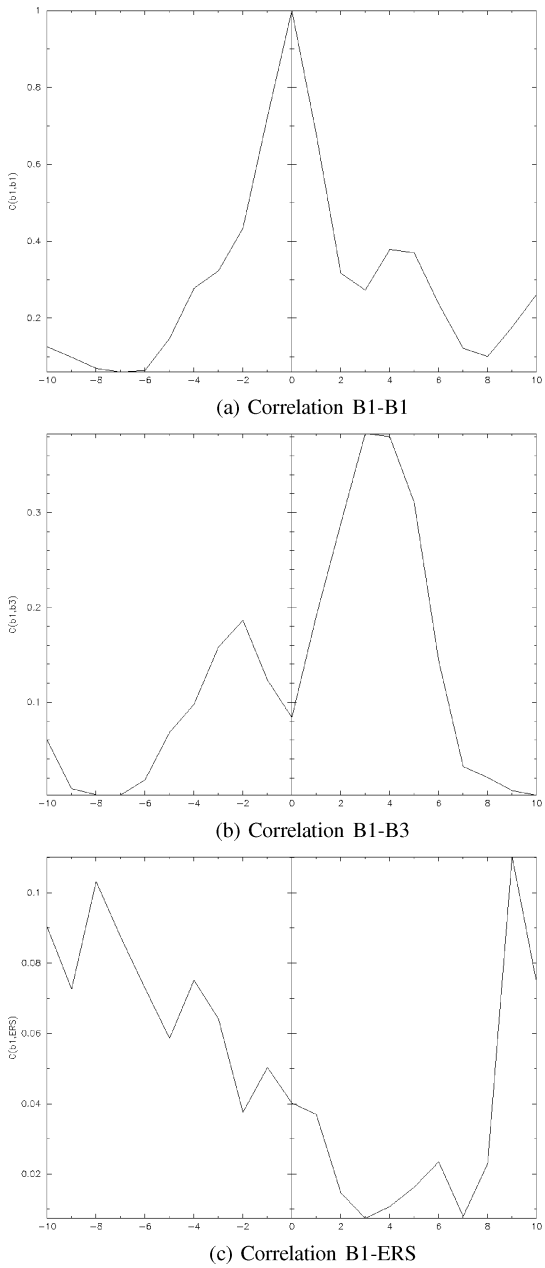
In this expression, N is the number of pixels of the analysis window; m_I and m_J are the estimated mean values inside the analysis window of, respectively, image I and image J; and σ_I and σ_J are their standard deviations.

Quality criteria can be applied to the estimated maximum in order to give a confidence factor to the estimated shift: width of the peak, maximum value, etc. Subpixel shifts can be measured by applying fractional shifts to the sliding window. This can be done by image interpolation.

The interesting parameters of the procedure are the following.

- *The size of the exploration area:* it determines the computational load of the algorithm (we want to reduce it), but it has to be large enough in order to cope with large deformations.
- *The size of the sliding window:* the robustness of the correlation coefficient estimation increases with the window size, but the hypothesis of local rigid shifts may not be valid for large windows.

The correlation coefficient cannot be used with original gray-level images in the multisensor case. It could be used on extracted features (edges, etc.), but the feature extraction can introduce localization errors. Also, when the images come from sensors using very different modalities, it can be difficult to find similar features in both images. In this case, one can try to find the similarity at the pixel level, but with other similarity measures and apply the same approach as we have just described.



(a) Correlation B1-B1

(b) Correlation B1-B3

(c) Correlation B1-ERS

Fig. 2. Measure of $\rho(\Delta x)$ for three different pairs of images.

The concept of similarity measure has been presented in Definition 3. The difficulty of the procedure lies in finding the function f, which properly represents the criterion c. We also need that f be easily and robustly estimated with small windows. We extend here what we proposed in [17].

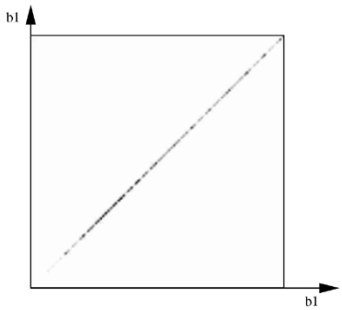
A. Correlation Coefficient

We remind here the computation of the correlation coefficient between two image windows I and J. The coordinates of the pixels inside the windows are represented by (x, y)

$$\rho(I,J) = \frac{1}{N} \frac{\sum_{x,y}(I(x,y) - m_I)(J(x,y) - m_J)}{\sigma_I \sigma_J}. \quad (5)$$

In order to qualitatively characterize the different similarity measures we propose the following experiment. We take two

(a) SPOT 5 B1



(b) Joint Histogram

Fig. 3. Joint histogram of an image with itself.

images that are perfectly registered, and we extract a small window of size $N \times M$ from each of the images (this size is set to 101×101 for this experiment). For the master image, the window will be centered on coordinates (x_0, y_0) (the center of the image), and for the slave image, it will be centered on coordinates $(x_0 + \Delta x, y_0)$. With different values of Δx (from -10 pixels to 10 pixels in our experiments), we obtain an estimate of $\rho(I, J)$ as a function of Δx, which we write as $\rho(\Delta x)$ for short. The obtained curve should have a maximum for $\Delta x = 0$, since the images are perfectly registered. We would also like to have an absolute maximum with a high value and with a sharp peak, in order to have a good precision for the shift estimate.

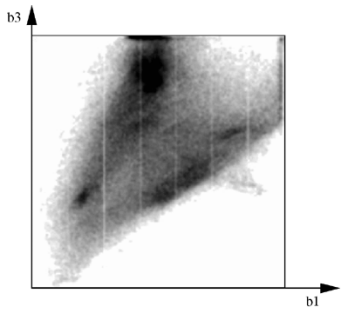
In the following, we will make this experiment with different image pairs and different similarity measures. Fig. 2 shows the results obtained when the correlation coefficient is applied to [Fig. 2(a)] one extract of the B1 channel of a SPOT 5 image with itself, [Fig. 2(b)] an extract of channel B1 with the extract of channel B3, and [Fig. 2(c)] the extract of channel B1 with an extract of an ERS-2 SAR image. The images are presented in Figs. 3–5.

We can see that the correlation coefficient has a good behavior for the first pair, but its performances are bad when the images radiometries are different. The correlation coefficient can be characterized as follows:

- well-known algorithm;
- fits the registration needs when using radiometrically similar images;
- simple and fast computation;
- high precision in the estimation of the deformation;
- robust to noise.

(a) SPOT 5 B1

(b) SPOT 5 B3



(c) Joint Histogram

Fig. 4. Joint histogram of two channels (B1-B3) of the same SPOT 5 image.

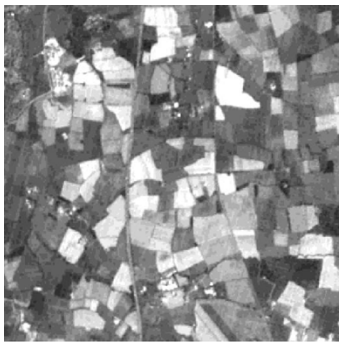
However, its main disadvantage is that it can only take into account affine transformations between radiometries ($j = \alpha i + \beta$), so it cannot be used in the general multisensor case.

B. Generalization: Probabilistic Interpretation

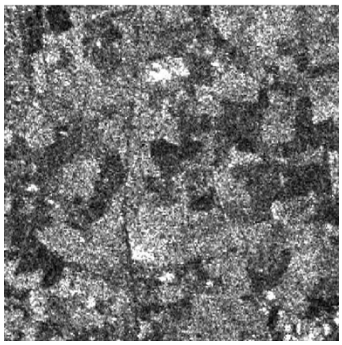
The correlation coefficient formulation [see (5)] can be revisited with a probabilistic interpretation

$$\rho(I,J) = \frac{1}{N} \frac{\sum_{x,y} (I(x,y) - m_I)(J(x,y) - m_J)}{\sigma_I \sigma_J} = \sum_{(i,j)} \frac{(i - m_I)(j - m_J)}{\sigma_I \sigma_J} p_{ij} \tag{6}$$

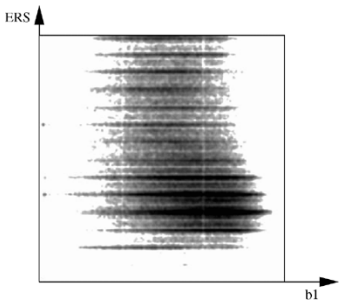
where the sum is taken over the list of radiometry pairs (i, j), and p_{ij} is the value of the joint normalized histogram (estimation of the joint probability density function (pdf) $f_{ij}(i, j)$) of the pair of images.



(a) SPOT 5 B3



(b) ERS-2



(c) Joint Histogram

Fig. 5. Joint histogram of a SPOT 5 B3 image and a ERS-2 image.

That means that we are assuming a linear model

$$j = (i - m_I)\frac{\sigma_J}{\sigma_I} + m_J \tag{7}$$

and we evaluate its likelihood by weighting with the probability of each radiometry couple p_{ij}.

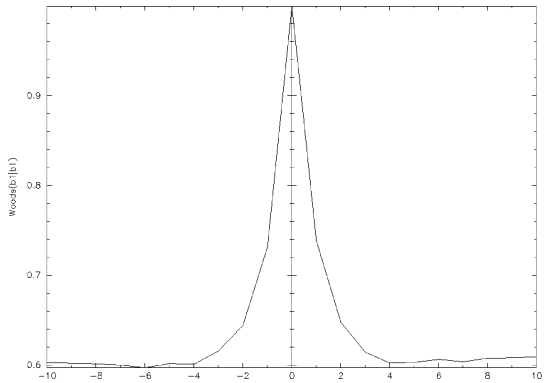
One could assume different models for the radiometry pairs leading to different measures as, for instance, the identity model $i = j$, which leads to the L_n norm

$$L_n(I,J) = \sum_{i,j} |i-j|^n \, p_{ij} \tag{8}$$

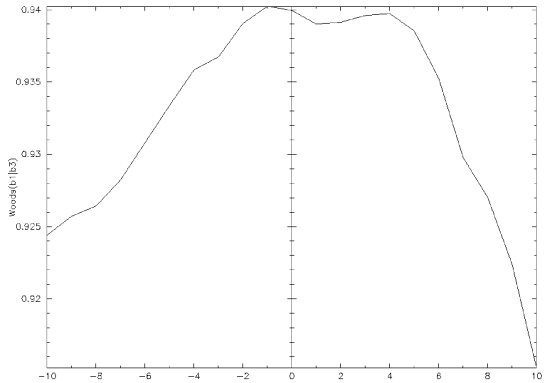
or more complex models based on textural approaches, as follows:

Diagonal moment:

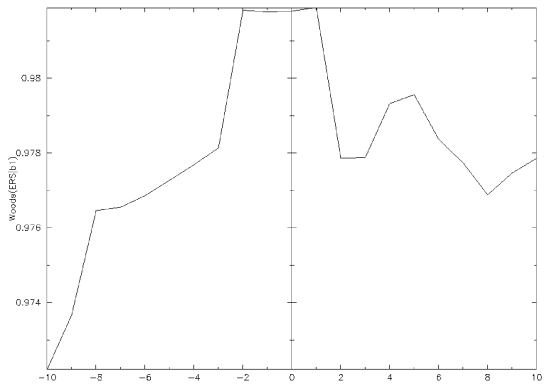
$$\mathrm{MD}(I,J) = \sum_{i,j} |i-j| \, (i+j-\sigma_I-\sigma_J) p_{ij}. \tag{9}$$



(a) B1 with B1



(b) B1 with B3



(c) B1 with ERS

Fig. 6. Image shift experiment: Woods criterion.

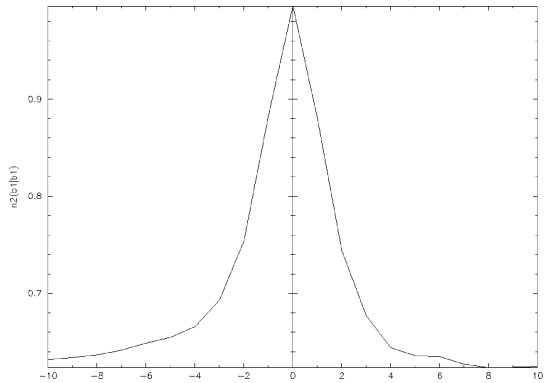
Cluster Shade:

$$C_{\mathrm{shade}}(I,J) = \sum_{i,j} (i+j-\sigma_I-\sigma_J)^3 p_{ij}. \tag{10}$$

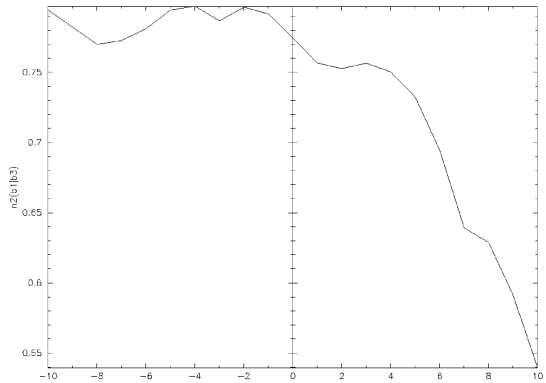
Cluster Prominence:

$$C_{\mathrm{pro}}(I,J) = \sum_{i,j} (i+j-\sigma_I-\sigma_J)^4 p_{ij}. \tag{11}$$

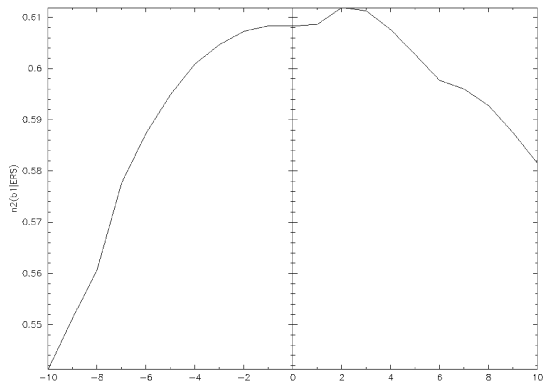
An assessment of these measures for image registration can be found in [18]. They are very sensitive to noise and are not useful for the comparison of gray levels of multisensor image pairs.



(a) B1 with B1

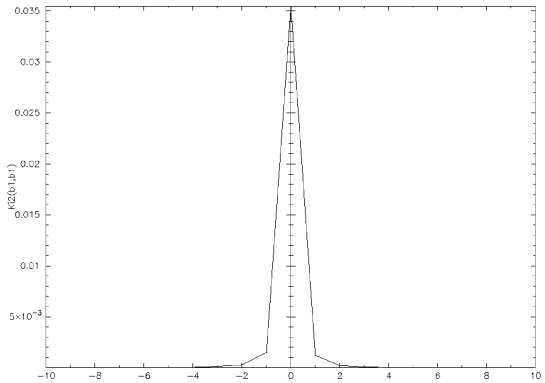


(b) B1 with B3

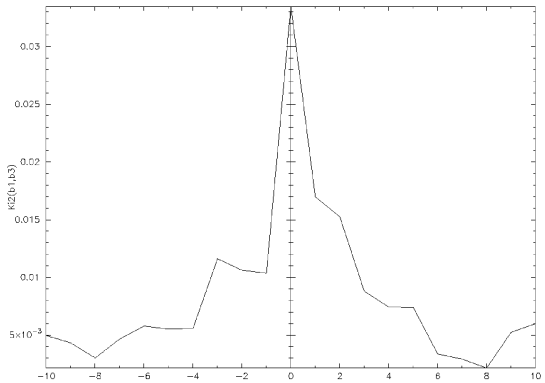


(c) B1 with ERS

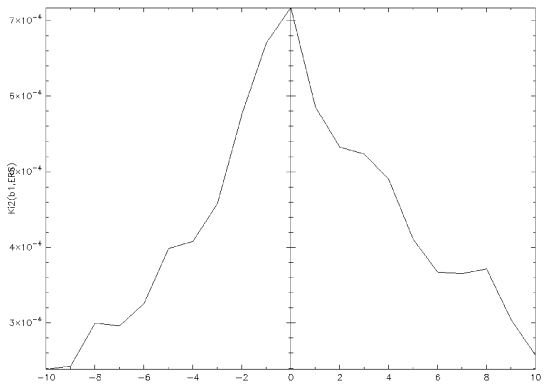
Fig. 7. Image shift experiment: Correlation ratio.



(a) B1 with B1



(b) B1 with B3



(c) B1 with ERS

Fig. 8. Image shift experiment: Distance to independence.

C. Estimation of Similarity Measures and p_{ij}

In the expression of the correlation coefficient the term p_{ij} is an estimation of the joint pdf of the radiometries of the images we are comparing. It can be seen as a link (transfer function) between both radiometries.

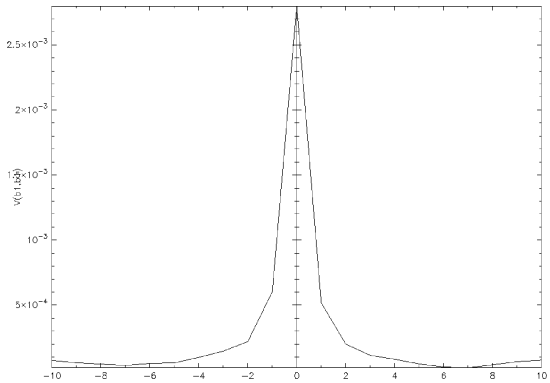
We show here several examples of the estimation of the joint histogram. In Figs. 3–5 are shown, respectively, the joint histograms of one image with itself (B1-B1), two different channels of the same SPOT 5 image (B1-B3), and a SPOT 5 B3–ERS-2 pair.

As expected, the joint histogram of an image with itself is a straight line with slope 1. It shows the full correlation between the two images: the identity transfer function. This kind of situation is well dealt with by the correlation coefficient.

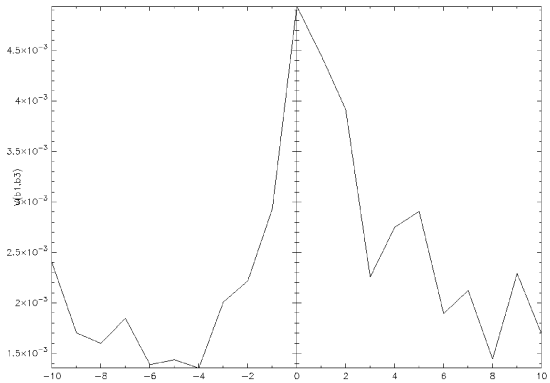
TABLE III
EXPRESSIONS OF FUNCTION f IN THE f-*DIVERGENCE* FAMILY

Measure	$f(x)$
Kolmogorov distance	$\frac{1}{2}\|x-1\|$
Mutual information	$x\log x$
Kullback divergence	$(x-1)\log x$
χ^2-divergence	$\frac{1}{2}(x-1)^2$
Hellinger distance	$\frac{1}{2}(\sqrt{x}-1)^2$
Toussaints distance	$x\frac{x-1}{x+1}$
Lin K-divergence	$x\log\frac{2x}{1+x}$

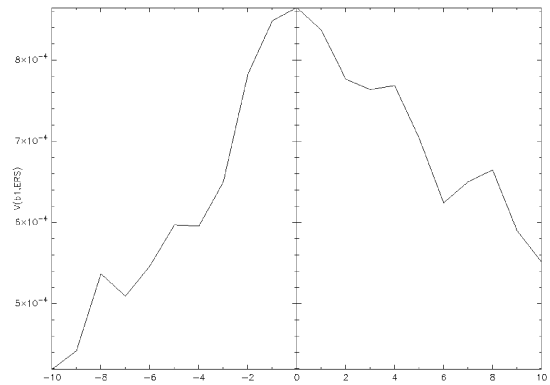
The B1-B3 case (Fig. 4) shows two nearly linear tendencies that are mixed up. This case cannot be dealt with by the correlation coefficient.



(a) B1 with B1



(b) B1 with B3



(c) B1 with ERS

Fig. 9. Image shift experiment: Kolmogorov distance.

Finally, Fig. 5 shows the impossibility of finding any correlation link between two sensors, which are as different as an optical and a radar one.

1) Computation Time: The main difference between the two expressions of the correlation coefficient given by (5) and (6) is the estimation of the joint pdf needed in the second expression. This estimation is usually done by computing the joint histogram. The joint histogram can be computed with different methods, but their discussion is beyond the scope of this paper. However, it is important to note that the method chosen for histogram computation may induce significant changes in the computation cost of the similarity measure. As an example, with our implementation (counting with optimization of the number of classes), there is an increase factor of 4 in computation time between (5) and (6).

The multisensor measures that will be introduced in the next section need the estimation of the joint histogram. Hence, their computation time is comparable to the one of (6). The differences of computation complexity between these measures are negligible, since the longest part of the algorithm is taken by the joint histogram estimation.

D. Multisensor Measures

We introduce here several similarity measures that have been proved useful in the problem of multimodality medical image registration [19].

In the following, the sums will be computed over radiometry values. We will use the conditional mean

$$m_{I|j} = \frac{1}{p_j} \sum_i i p_{ij} \tag{12}$$

and the conditional variance

$$\sigma^2_{I|j} = \frac{1}{p_j} \sum_i (i - m_{I|j})^2 p_{ij}. \tag{13}$$

For each of the following measures, we will perform the experiment described in Section V-A.

1) Measures Using the Radiometry Values and the Probabilities: Within this class, we will not take into account the measures that are based on the differences of radiometries (L_n norm of the difference) [20]–[22] or textural measures, since they give low-quality results.

a) Normalized sandard deviation or Woods criterion: The work by Woods *et al.* first on monomodal registration [23] and then on multimodal registration [24] lead to the elaboration of this similarity measure. Given the intensity on one image, i.e., the set of pixels having this value, this measure takes into account the variability of the intensities of the homologous pixels in the other image. The underlying hypothesis is that this variability (which is actually a measure of variance) will be minimum when the images are registered

$$\text{Woods}(I|J) = \sum_j \frac{\sigma_{I|j}}{m_{I|j}} p_j. \tag{14}$$

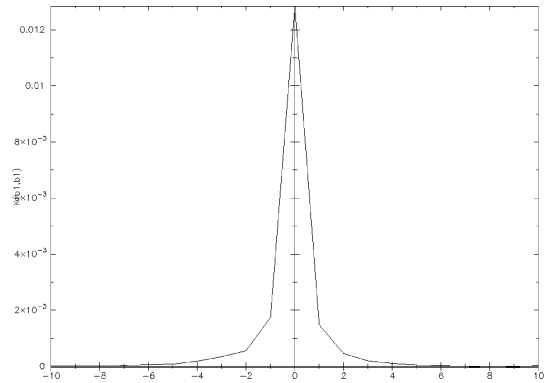
In order to have a criterion which has to be maximized, we will use

$$S_{\text{Woods}}(I|J) = 1 - \sum_j \frac{\sigma_{I|j}}{m_{I|j}} p_j. \tag{15}$$

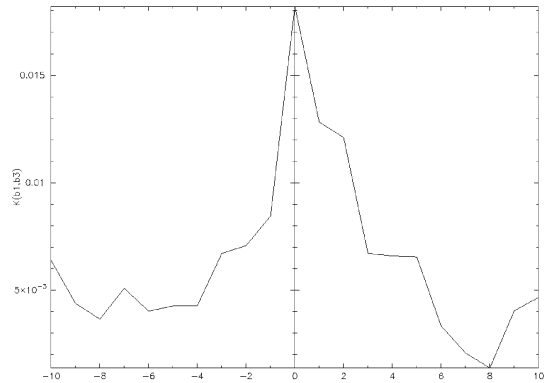
The results on our three test image pairs are shown in Fig. 6. We see that for the monosensor case, the results are similar to those of the correlation coefficient. For the two multisensor examples, we obtain high values near the zero-shift, but the location of these maxima is not accurate.

b) Correlation ratio: This is a very well known measure in statistics. It has been first proposed in the framework of image registration by Roche *et al.* [25]. It is defined as follows:

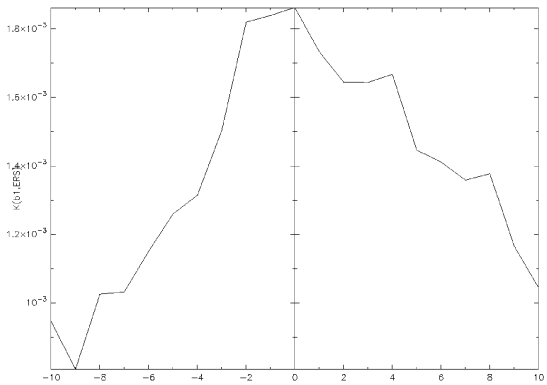
$$\eta^2(I|J) = 1 - \frac{1}{\sigma^2_I} \sum_j \sigma^2_{I|j} p_j. \tag{16}$$



(a) B1 with B1

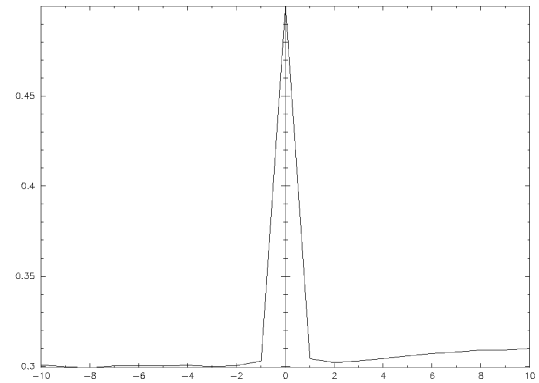


(b) B1 with B3

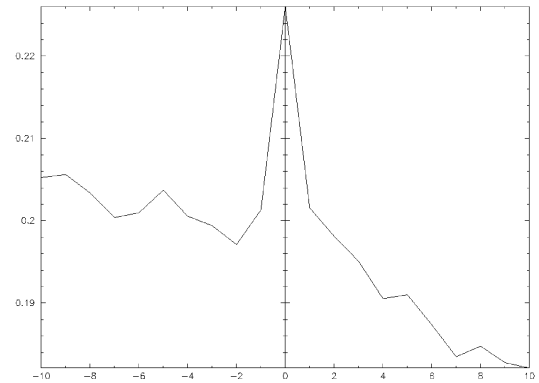


(c) B1 with ERS

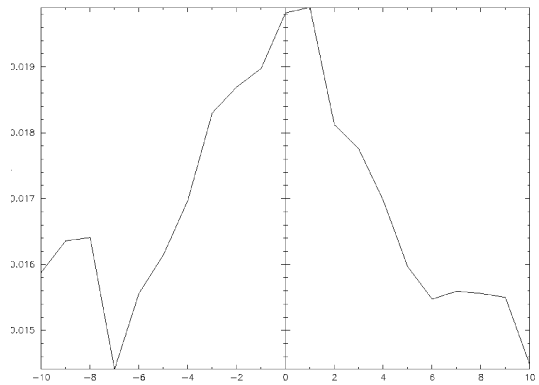
Fig. 10. Image shift experiment: Mutual information.



(a) B1 with B1



(b) B1 with B3



(c) B1 with ERS

Fig. 11. Image shift experiment: CRA.

Its interpretation is similar to the one of the Woods criterion. The results are shown in Fig. 7, and they are worse than those of the Woods criterion.

2) Measures Using Only the Probabilities: This class of measures does not directly use the radiometries of the pixels, but only the estimation of the joint pdf. Of course, the pixel pairs are used for the estimation of this probability.

a) Distance to independence: It is a normalized version of the χ^2 test

$$\chi^2(I,J) = \sum_{i,j} \frac{(p_{ij} - p_i p_j)^2}{p_i p_j}. \tag{17}$$

It measures the degree of statistical dependence between both images, since for two independent random variables, the joint pdf is equal to the product of the marginals. The correlation coefficient is a test of independence of order 2, and this one is the generalization to any order. The results are shown in Fig. 8. In this case, for the three pairs, we obtain an absolute maximum for the zero-shift case, which is sharp enough for a robust automatic detection.

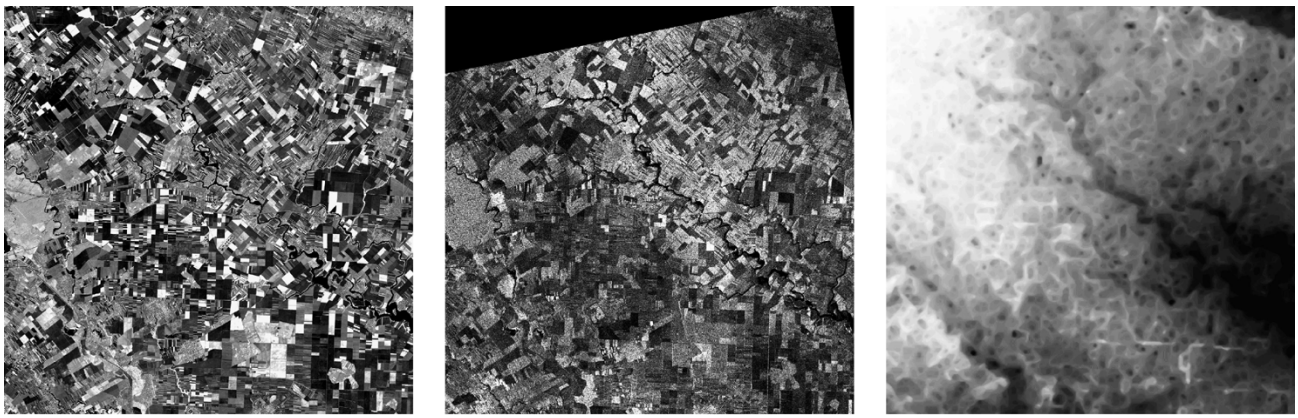
b) f-divergence family: An *f-divergence* [26] measures the expectation of the diversity of the likelihood ratio between two distributions P and Q

$$D_f(P,Q) = E_Q\left[f\left(\frac{dp(x)}{dq(x)}\right)\right] = \int f\left(\frac{p(x)}{q(x)}\right) q(x)dx. \tag{18}$$

E_Q is the expectation with respect to Q; $dp(x)/dq(x)$ is the derivative with respect to a density; and f is continuous and

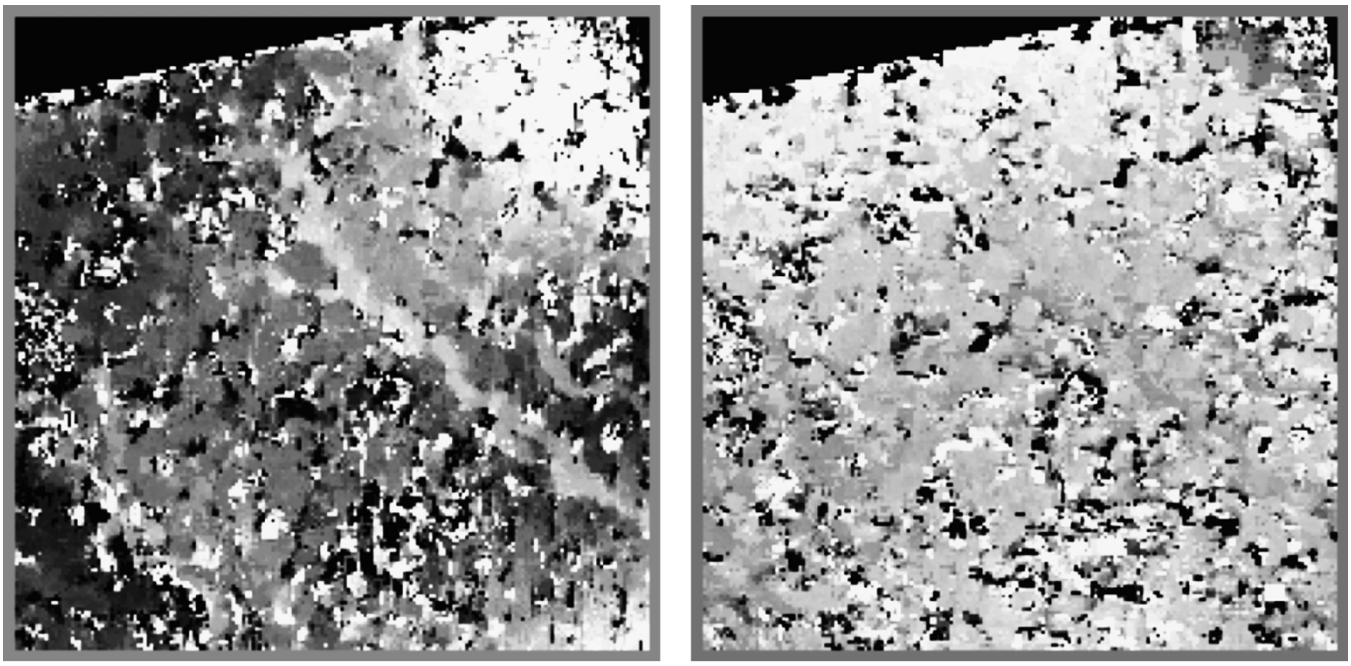
TABLE IV
DEFORMATION GRID ESTIMATION MEAN SQUARE ERROR

Mutual information, B1-B3. MSE in $pixel^2$						
Window size	$T = \infty$	$T = 4000$	$T = 2000$	$T = 1000$	$T = 500$	$T = 200$
20×20	5.4	7.9	8.8	8.8	8.3	10.5
40×40	0.5	0.9	1.1	2.4	2.3	3.6
60×60	0.1	0.2	0.2	0.1	0.3	3.2
80×80	0..0	0.0	0.0	0.1	0.5	3.9
100×100	0.0	0.0	0.0	0.1	0.7	5.5



(a) SPOT 4 B3 (b) ERS-2 SAR (c) DEM

Fig. 12. Images and DEM for the test area.



(a) Columns (b) Lines

Fig. 13. Deformation grid. CRA, estimation window is 51 × 51 pixels, and the sampling rate is ten pixels.

convex on $[0, +\infty)$. A divergence can be seen as a relative entropy. In order to simplify the notation, we will use $p = p_{ij}$, $q = p_i p_j$, and $\int = \sum_{i,j}$.

Depending on the choice of f (see Table III), we can obtain several interesting cases, as follows.

1) *Kolmogorov distance:*

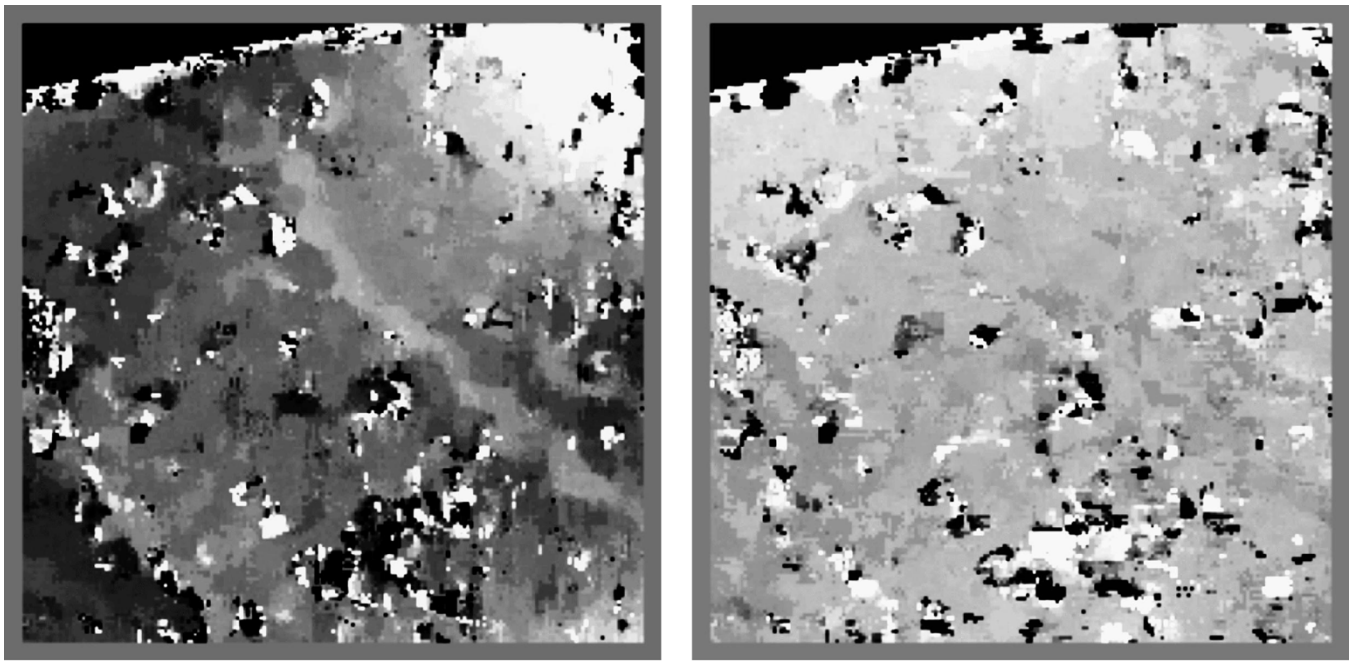
$$V(P,Q) = \frac{1}{2}\int |p - q|\,. \tag{19}$$

2) *Kullback information or mutual information:*

$$K(P,Q) = \int p \log \frac{p}{q}. \tag{20}$$

3) *Kullback divergence:*

$$K'(P,Q) = \int (q - p)\,(\log q - \log p)\,. \tag{21}$$



(a) Columns (b) Lines

Fig. 14. Deformation grid. CRA, estimation window is 81 × 81 pixels, and the sampling rate is ten pixels.

4) χ^2*-divergence:*

$$R(P,Q) = \frac{1}{2}\int \frac{(p-q)^2}{q}. \tag{22}$$

5) *Hellinger distance:*

$$\mathcal{H}^2(P,Q) = \frac{1}{2}\int (\sqrt{p}-\sqrt{q})^2. \tag{23}$$

6) *Toussaints distance:*

$$T(P,Q) = \int p - \frac{2pq}{p+q}. \tag{24}$$

7) *Lin K-divergence:*

$$K_{\text{div}}(P,Q) = \int p \log \frac{2p}{p+q}. \tag{25}$$

All these measures give very similar results [27]. We study two of them.

1) *Kolmogorov distance:*

$$V(P,Q) = \frac{1}{2}\int |p_{ij} - p_i p_j|. \tag{26}$$

It can be seen as a L_1 norm version of the χ^2 criterion. The results are shown in Fig. 9.

2) *Mutual information:*

$$K(P,Q) = \int p_{ij} \log \frac{p_{ij}}{p_i p_j}. \tag{27}$$

The results are shown in Fig. 10.

Both measures give satisfactory results, which are very similar to the ones obtained with the distance to the independence.

c) Cluster reward algorithm: Let $H_{IJ}(k,l)$ be the joint histogram of the pair of images, and let $H_I(k)$ and $H_J(k)$, respectively, be the marginal histograms and P the number of pixels. We define

$$I_{\text{CRA}} = \frac{\frac{\Phi}{F} - \frac{F}{P^2}}{1 - \frac{F}{P^2}} \tag{28}$$

where

$$\Phi = \sum_{k=0}^{N-1}\sum_{l=0}^{N-1} H_{IJ}^2(k,l) \tag{29a}$$

$$F = \sqrt{h_I h_J} \tag{29b}$$

$$h_I = \sum_{k=0}^{N-1} H_I^2(k) \tag{29c}$$

$$h_J = \sum_{k=0}^{N-1} H_J^2(k). \tag{29d}$$

The I_{CRA} index will have a high value when the joint histogram has little dispersion. This lack of dispersion can be due to a correlation (histogram distributed along a line) or to the clustering of radiometry values within the histogram. In both cases, one can predict the values of one image from the values of the other.

In order to compare I_{CRA} with the *f-divergences*, we can rewrite (28) as

$$I_{\text{CRA}} = \frac{\int p_{ij}^2 - \int p_i^2 \int p_j^2}{\sqrt{\int p_i^2 \int p_j^2} - \int p_i^2 \int p_j^2}. \tag{30}$$

If we consider the denominator as a normalization term, we can focus only in the numerator. This numerator contains the same terms as the *f-divergences*, i.e., a term that depends on the joint pdf and a term that depends on the product of the marginals.

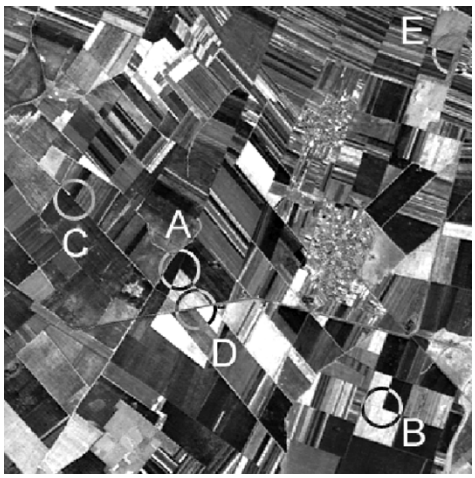


Fig. 15. Region of interest and test points for the validation.

TABLE V
RESULTS OF THE FIRST VALIDATION METHOD. SEE TEXT AND FIGS. 15–20 FOR THE DETAILS

Point		ERS	SPOT	Deformation		
				Manual	Computed	Error
A	Column	153.00	150.20	2.80	2.56	0.24
(fig. 16)	Line	226.50	225.12	1.38	1.50	-0.12
B	Column	326.56	323.22	3.34	3.32	0.02
(fig. 17)	Line	345.00	343.44	1.56	1.44	0.12
C	Column	62.00	60.11	1.89	1.48	0.41
(fig. 18)	Line	168.33	165.89	2.44	2.31	0.13
D	Column	166.56	163.44	3.12	2.56	0.56
(fig. 19)	Line	259.78	258.00	1.78	1.57	0.21
E	Column	390.20	385.10	5.10	4.57	0.53
(fig. 20)	Line	42.50	40.75	1.75	1.37	0.38

We can, thus, make an interpretation that is similar to the independence tests. As expected, the obtained results (Fig. 11) are similar to those obtained with the *f-divergences*.

The main interest of the CRA with respect to the *f-divergence* family is that the joint histogram noise due to estimation has less influence in the similarity measure. This allows us to use smaller estimation windows. The only drawback in doing this is that the peak of the measure will be less sharp.

E. Characterization of Similarity Measures

Similarity measures can be characterized using the following criteria:

- *the geometric resolution:* maximum frequency of the deformations;
- ability to deal with images acquired from different sensors.

The first point is directly linked to the number of pixels needed to obtain a good estimate of the measure. Indeed, since when estimating deformation grids we assume that the deformation can be decomposed on local shifts, that means that there should be no deformation inside the estimation window. However, if we need large windows for the estimation, the likelihood for our assumption to be true can be very low.

In order to illustrate this problem, we have made the following experiment. We have taken the pair B1-B3 of a SPOT 5 image, and we have applied a sinusoidal deformation, with period T pixels to the slave (B3) image in the horizontal direction. We have made the estimation of the deformation using the mutual information, and we have computed the mean square error between the estimated and the real deformation grids

$$E = \sum_i \sum_j \left(G(i,j) - \widehat{G}(i,j) \right)^2$$

where $G(i,j)$ is the real sinusoidal deformation, and $\widehat{G}(i,j)$ is the estimated deformation grid. The experiment has been made for different values of the period T and the estimation window size. The results, which are measured in square pixels, are shown in Table IV. We observe the following.

- For medium to long periods, the quality of the estimation increases with the window size.
- For short periods ($T = 500$ and $T = 200$), the increase of the window size produces a decrease of the performances. This is due to the fact that geometrical deformations are strong inside the estimation window.

F. Behavior in Presence of Noise

One could also analyze how the different similarity measures behave when noise is present in the data. The presence of additive noise in the data produces a dispersion of the joint histogram of the images. For the *f-divergence* family, since the estimation windows used in the experiments contain a high number of samples, noisy data does not cause noisy estimations of the similarity, but rather wider similarity peaks. However, the location of the similarity optimum is not affected by noise. This behavior is confirmed by the SPOT-ERS couple, since the radar image contains a strong multiplicative noise.

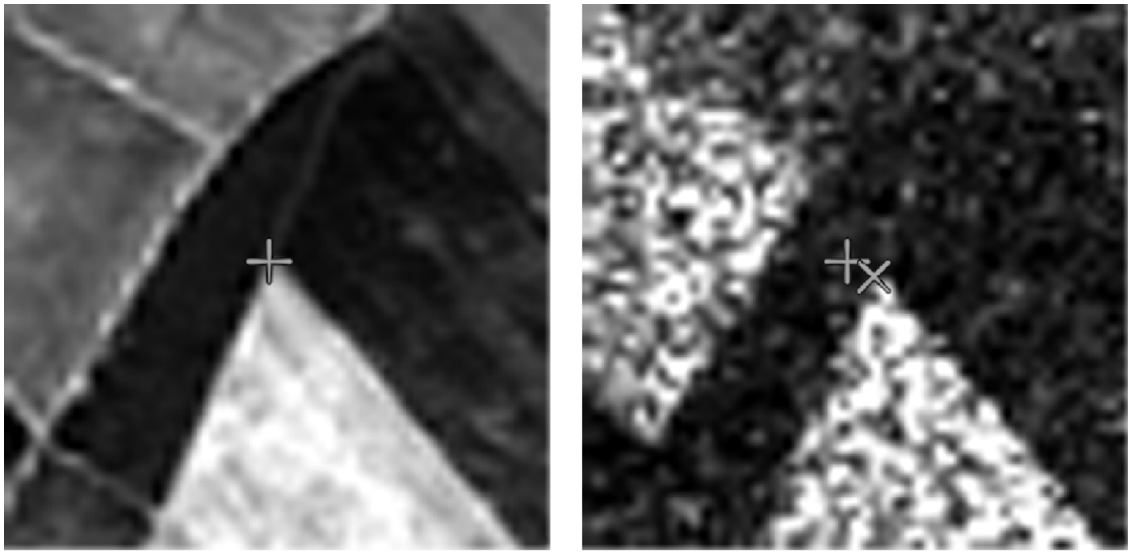
VI. GRID ESTIMATION: A REAL CASE

The similarity measures introduced in the previous sections have been tested in a simple framework of integer shifts in one dimension and with low frequency (Table IV). In order to further test the performances of these methods, we will apply them in the same way as we would do using the correlation coefficient, i.e., estimating a deformation grid between a radar and an optical image.

Our dataset consists in the following pair (a region of a size of 2000×2000 pixels is used for our tests):

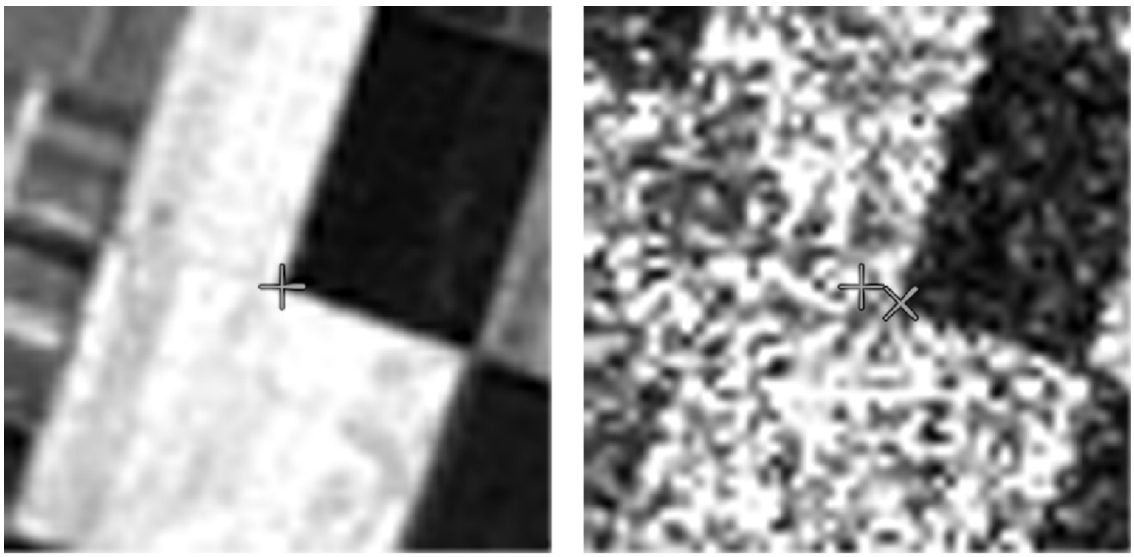
- the B3 channel of a SPOT 4 image (20-m pixel resolution) acquired on June 24, 2001 over the East of the Bucharest area [Fig. 12(a)];
- a ERS-2 SAR three-looks intensity image (12.5-m pixel size and approximately 20-m pixel resolution) acquired on May 10, 2001 over the same area [Fig. 12(b)].

Both images were orthorectified: for the SPOT 4 image, a digital elevation model (DEM) [Fig. 12(c)] with an altimetric precision better than 10 m and a planimetric precision around 10 m has been used, together with the acquisition model (orbits, attitude) for the satellite; for the ERS-2 image, no DEM was used, but a constant altitude and homologous points manually taken on the SPOT 4 image were used in the orthorectification process. Globally, the images show a good superposition, but local errors exist, which can amount to several pixels due to the simple geometric modeling of the deformation of the radar image. We have discussed these problems in Section IV.



(a) Point A in the master image (b) Point A in the slave image

Fig. 16. Location of point A for a null displacement (+) and for the measured displacement (×).



(a) Point B in the master image (b) Point B in the slave image

Fig. 17. Location of point B for a null displacement (+) and for the measured displacement (×).

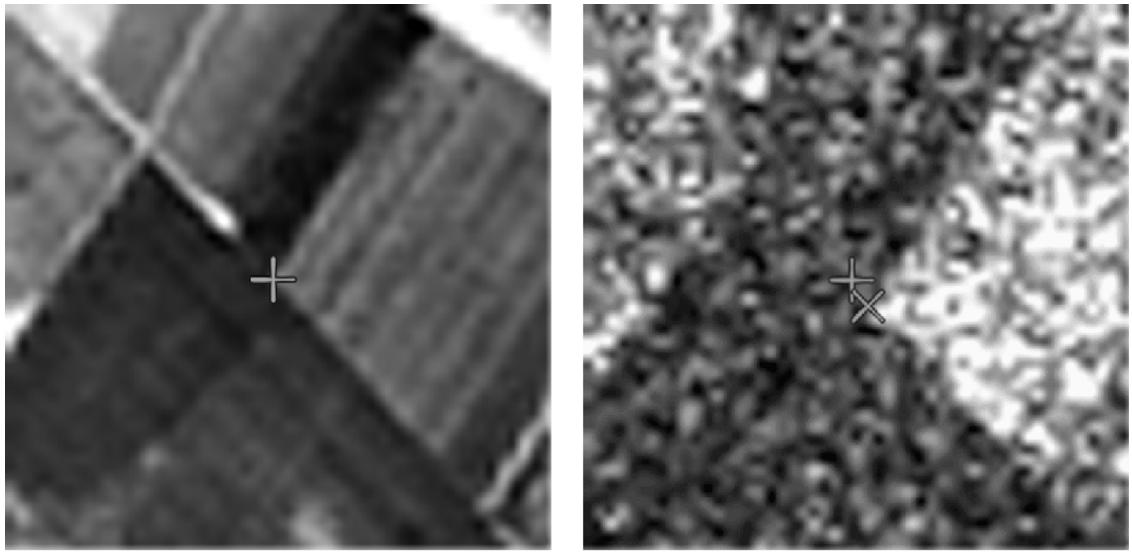
If we analyze the DEM, we see that a gentle slope descending from northwest to southeast exists and that abrupt topography features appear in the northeast and the southwest. The shape of the river can also be identified in the DEM.

We will estimate the residual deformation between the two images with the cluster reward algorithm (CRA). We will estimate the local shifts (lines and columns) every ten pixels, thus building the deformation grids. Using noninteger shifts for the estimation window, we can estimate the deformation with subpixel accuracy. The noninteger shifts are applied by interpolating the slave image using a *sinc* function weighted by a Gaussian function whose variance is chosen so as to keep 90% of the energy in a filter with a length of 13 samples.

First of all, we will use an estimation window of size 51×51 pixels and an exploration area of ± 4 pixels around each pixel. The map of the measured deformations is shown in Fig. 13. Applying quality criteria to the peak of the measure (value and shape), we find that 70% of the measured points are considered as valid and that 24% of the remaining pixels are not valid because the peak is in the limit of the exploration area.

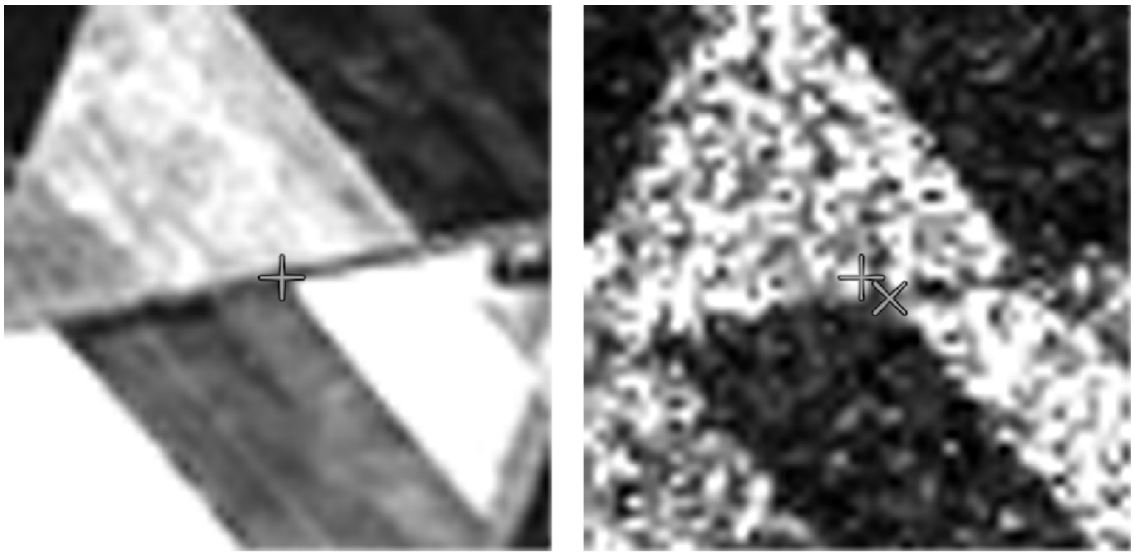
Analyzing these grids we can draw the following conclusions.

- The grids are noisy. This is due to the size of the estimation window.
- The main deformations are measured in the column directions, i.e., the direction for which there is an stereoscopic effect between the two acquisitions (ERS-2 and SPOT 4 have polar orbits).
- The measured deformations are strongly correlated to the topography, mainly near the river area and at the two abrupt changes in the northeast and the southwest areas. This shows the limitations of the analytical models for the topographic deformations.



(a) Point C in the master image (b) Point C in the slave image

Fig. 18. Location of point C for a null displacement (+) and for the measured displacement (×).



(a) Point D in the master image (b) Point D in the slave image

Fig. 19. Location of point D for a null displacement (+) and for the measured displacement (×).

- Low-frequency gradients of small amplitude appear on both grids. This may be due to a bad choice of the degree of the polynomial modeling.

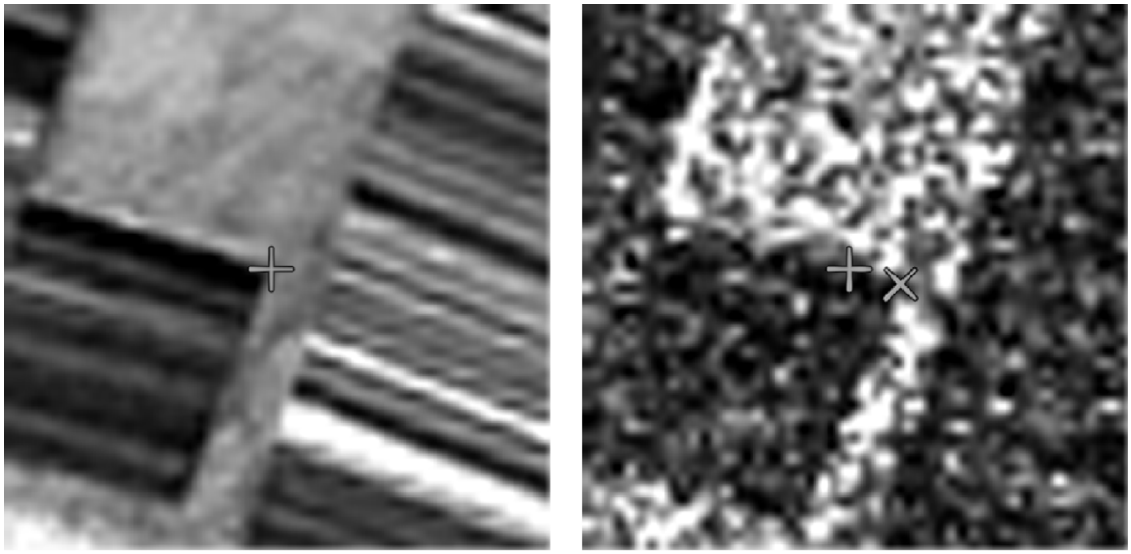
In order to validate the assumption that links the estimation window size to the noisy aspect of the grid, we perform a new test with a window size of 81×81 pixels. We also increase the exploration area to ± 8 pixels in order to increase the number of valid points. We obtain the following results: 78% of valid points; 21% of the points have the peak in the limit of the exploration area. The deformation grids are shown in Fig. 14. We observe a decrease of the noise.

A. Validation of the Results

In order to validate the displacements obtained in the deformation grids, we have used three different approaches.

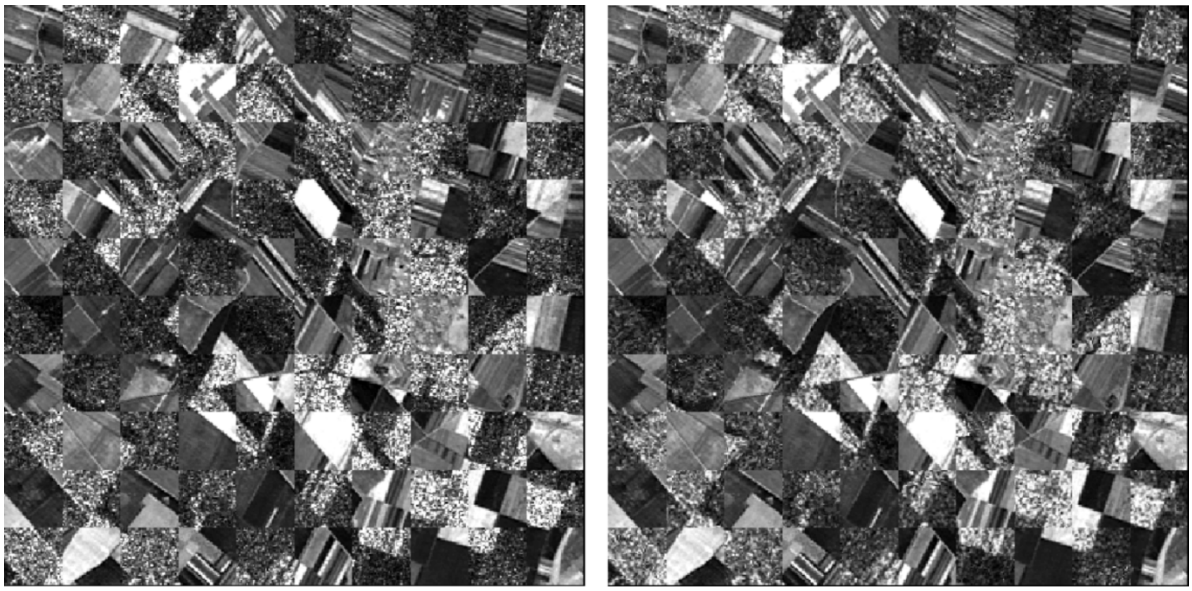
1) We select a set of points in the master (SPOT 4) image, and we manually find the HP in the slave (ERS-2) image and compare the obtained displacements with the ones measured in the grids. This method compares the result of the automatic processing with a manual HP selection.
2) For the same set of points of the master image, we point out the two points in the slave image: the point with the same coordinates (zero-shift) and the point that results in applying the shift measured in the grid. This method shows the relative improvement in registration with respect to the original image pair.
3) We resample the slave image using the measured grids and visually check the quality of the registration before and after resampling.

The set of test points for the first and second validation methods are shown in Fig. 15. The results of method 1 are



(a) Point E in the master image (b) Point E in the slave image

Fig. 20. Location of point E for a null displacement (+) and for the measured displacement (×).



(a) Before registration (b) After registration

Fig. 21. Checkerboard visualization of the registration.

shown in Table V where, for each test point, we give the following:

- its coordinates (line and column) in both (ERS and SPOT) images as obtained by manual point selection;
- the deformation as a result of the difference of coordinates above;
- the deformation computed by the automatic algorithm;
- and the deformation error, i.e., the difference between manual and automatic approaches.

We see that the automatic processing yields results that are very close to manual operation. Indeed, the Euclidean distance between manual and computed deformations is always less than 0.65 pixels. One has to bear in mind that the manual measure is not perfect, so this can be considered a very good result.

The results of the second method are shown on Figs. 16–20. One can see that for every point, the measured displacement gives a better relative position of the studied point than the null displacement.

Fig. 21 shows the results of the third validation method. One can see that the resampling of the radar image using the estimated deformation grids [Fig. 21(b)] leads to a better registration of the images [Fig. 21(a)].

VII. DEM Estimation

Besides the image registration approach, the results of the previous section allow for the estimation of topography using an optical-radar image pair. In this section, we show how this could be done using a single satellite. This procedure is given here as

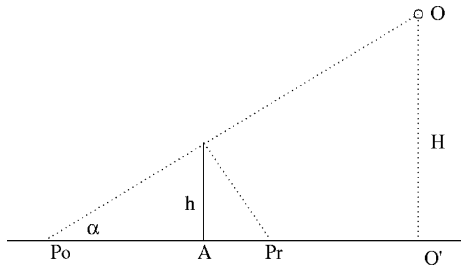


Fig. 22. DEM estimation from a single platform.

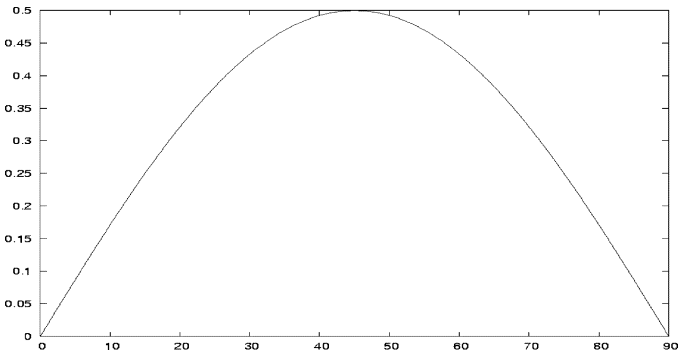


Fig. 23. Factor $\sin 2\alpha/2$ for α in degrees.

an illustration and should not be considered as a substitution of parametric sensor models when they are available.

Fig. 22 shows a geometric model for the problem. Both sensors (optical and radar) are onboard of the same satellite at point O whose vertical projection to the ground is point O', and H is the height of the orbit. Both image acquisitions are simultaneous and the incidence angle is α. We want to measure the height h of point A. We can locally make the assumption of flat earth.

In the optical image, the point A is actually seen at point Po, while it is projected into point Pr in the radar image, if we assume a flat wavefront. If we call $\Delta x = \overline{APo} + \overline{APr}$ the measured shift of the point between both images, we obtain

$$h = \Delta x \frac{\tan\alpha}{\tan^2\alpha + 1} = \Delta x \frac{\sin 2\alpha}{2}. \tag{31}$$

That means that the height of the point is proportional to the measured shift. It is interesting to note that it does not depend neither on any baseline between the two instruments nor on their altitude. Fig. 23 shows the dependency of the proportionality factor with the incidence angle. Typically, for optical sensors $\alpha > 60^\circ$ and for radar acquisitions, $\alpha < 70^\circ$. That means that the proportionality factor will be between 0.3–0.45.

In the case where the acquisitons are made with different incidence angles α_1 and α_2 (different platforms, for example), (31) becomes

$$h = \Delta x \frac{\tan\alpha_1}{\tan\alpha_1 \tan\alpha_2 + 1}. \tag{32}$$

VIII. Conclusion

We have formalized the problem of remote sensing image registration. We also have described which are the limitations of the correlation coefficient in the multisensor case, and we have shown how other similarity measures can be used to extend the correlation coefficient approach to the multisensor problem.

These measures have been widely used in the medical imaging field, but they have always been used together with analytical models and global similarity research. We have shown that these approaches are not enough in the case of remote sensing.

A simple case study with an optical-radar pair has allowed us to show that it is possible to automatically measure high-frequency deformations between multisensor images by using a local similarity estimation leading to the estimate of deformation grids of subpixel accuracy. This result opens the possibility of automatic optical-radar DEM estimation from a single satellite.

Acknowledgment

The authors would like to thank their colleagues at CNES/QTIS, for all the fruitful discussions about the fine-image registration problem.

References

[1] C. Kuglin and D. Hines, "The phase correlation image alignment method," in *Proc. IEEE Conf. Cybernetics Soc.*, 1975, pp. 163–165.

[2] V. N. Dvorchenko, "Bounds on (deterministic) correlation functions with applications to registration," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-5, pp. 206–213, Feb. 1983.

[3] L. Brown, "A survey of image registration techniques," *ACM Comput. Surv.*, vol. 24, no. 4, pp. 325–376, 1992.

[4] J. Ton and A. K. Jain, "Registering Landsat images by point matching," *IEEE Trans. Geosci. Remote Sensing*, vol. 27, pp. 642–651, Sept. 1989.

[5] H. Li, B. Manjunath, and S. K. Mitra, "A contour-based approach to multisensor image registration," *IEEE Trans. Image Processing*, vol. 4, pp. 320–334, Mar. 1995.

[6] A. Cracknell and K. Paithoonwattanakij, "Fast automatic registration for satellite imagery by heuristic planning," *Proc. Inst. Elect. Eng.*, vol. 136, no. 5, pp. 221–225, 1989.

[7] H. Foroosh, J. B. Zerubia, and M. Berthod, "Extension of phase-correlation to sub-pixel registration," *IEEE Trans. Image Processing*, vol. 11, pp. 188–200, Mar. 2002.

[8] H. Stone, M. T. Orchard, E.-C. Cheng, and S. A. Martucci, "A fast direct fourier-based algorithm for subpixel registration of images," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, pp. 2235–2243, Oct. 2001.

[9] J. Inglada and F. Adragna, "Automatic multi-sensor image registration by edge matching using genetic algorithms," in *Proc. IGARSS*, 2001.

[10] O. Thépaut, K. Kpalma, and J. Ronsin, "ERS SAR and SPOT images automatic registration in a multichannel consensual segmentation scheme," in *Proc. IGARSS*, 1998.

[11] Y. Wu and H. Maître, "Registration of a SPOT image and a SAR image using multiresolution representation of a coastline," in *Proc. ICPR*, 1990, pp. 913–917.

[12] A. Ventura, A. Rampini, and R. Schettini, "Image registration by recognition of corresponding structures," *IEEE Trans. Geosci. Remote Sensing*, vol. 28, pp. 305–314, May 1990.

[13] X. Dai and S. Khorram, "An feature-based image registration algorithm using improved chain-code representation combined with invariant moments," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, pp. 2351–2362, Sept. 1999.

[14] J. Djamdji, A. Bijaoui, and R. Maniere, "Geometrical registration of images: The multiresolution approach," *Photogramm. Eng. Remote Sens. J.*, vol. 59, no. 5, pp. 645–653, May 1993.

[15] J. Le Moigne, A. Cole-Rhodes, R. Eastman, K. Johnson, J. Morisette, N. Netanyahu, H. Stone, and I. Zavorin, "Earth science imagery registration," in *Proc. IGARSS*, 2003.

[16] J. Townshend, C. Justice, C. Gurney, and J. McManus, "The impact of misregistration on change detection," *IEEE Trans. Geosci. Remote Sensing*, vol. 30, pp. 1054–1060, Sept. 1992.
[17] J. Inglada, "Similarity measures for multisensor remote sensing images," in *Proc. IGARSS*, 2002.
[18] M. Bro-Nielsen, "Medical image registration and surgery simulation," Ph.D. dissertation, Tech.Univ. Denmark, Lyngby, Denmark, 1996.
[19] D. Sarrut, "Recalage multimodal et plate-forme d'imagerie médicale à accès distant," Ph.D. dissertation, Univ. Lumière Lyon 2, Lyon, France, 2000.
[20] P. Thévenaz, U. Ruttimann, and M. Unser, "Iterative multi-scale registration without landmarks," *IEEE Trans. Image Processing*, vol. 7, pp. 27–41, Jan. 1998.
[21] E. de Castro and C. Morandi, "Registration of translated and rotated images using finite Fourier transforms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, pp. 700–703, May 1987.
[22] A. Apicella, J. Kippenhan, and J. Nagel, "Fast multi-modality image matching," *Proc. SPIE*, vol. 1092, pp. 252–263, 1989.
[23] R. Woods, S. Cherry, and J. Mazziota, "Rapid automated algorithm for aligning and reslicing PET images," *J. Comput. Assist. Tomogr.*, vol. 16, no. 4, pp. 620–633, 1992.
[24] R. Woods, J. Mazziota, and S. Cherry, "MRI-PET registration with automated algorithm," *J. Comput. Assist. Tomogr.*, vol. 17, no. 4, pp. 536–546, 1993.
[25] A. Roche, G. Malandain, X. Pennec, and N. Ayache, "The correlation ratio as a new similarity measure for multimodal image registration," in *Proc. MICCAI*, Cambridge, MA, Oct. 1998, pp. 1115–1124.
[26] I. Csiszar, "Information type measures of difference of probability distributions and indirect observations," *Studia Sci. Math. Hungar.*, vol. 2, pp. 299–318, 1967.
[27] D. Sarrut and S. Miguet, "Similarity measures form image registration," in *Proc. Eur. Workshop on Content-Based Multimedia Indexing*, Toulouse, France, Oct. 1999, pp. 263–270.

Jordi Inglada received the telecommunications engineer degree from both Universitat Politècnica de Catalunya, Catalunya, Spain, and the École Nationale Supérieure des Télécommunications de Bretagne, Brest, France, in 1997, and the Ph.D. degree in signal processing and telecommunications from Université de Rennes 1, Rennes, France, in 2000.

He has been since been with the Centre National d'Études Spatiales, Toulouse, France, working in the field of remote sensing image processing. He is in charge for the development of image processing algorithms for the operational exploitation of earth observation images, mainly in the fields of image registration, change detection, and object recognition.

Alain Giros received the physics engineer degree from École Nationale Supérieure de Physique de Marseille, Marseille, France, in 1982, with a specialization in optics and image processing.

He has since been with the Centre National d'Études Spatiales (CNES), Toulouse, France. He began his carrier in the SPOT 1 project team where he was responsible of the whole image processing software. Then, he was responsible for the development of several image processing systems, including DALI (the worldwide catalog of SPOT images), several image quality assessment systems for SPOT, and the Vegetation User Services Facility. Interested in automated image registration since the mid-1990s, he has been promoting this research subject at CNES. He is now in charge of the development of image processing techniques mainly in the field of image time series processing, automated registration, and image information mining.