HOW TO COMBINE TERRASAR-X AND COSMO-SKYMED HIGH-RESOLUTION IMAGES FOR A BETTER SCENE UNDERSTANDING?

H. Sportouche, F. Tupin, J.-M. Nicolas, T. Perciano
Institut Telecom, Telecom ParisTech
LTCI, TSI Department
75634 Paris Cedex 13, France

C.-A. Deledalle
Ceremade, University Paris-Dauphine,
75775 Paris Cedex 16, France

ABSTRACT
This paper presents a processing chain to combine a CosmoSkyMed (CSK) image and a TerraSAR-X (TSX) image. After registration and calibration steps, processing at different levels is studied: pixel level for the detection of stable features and joint filtering, primitive level for stability analysis and object level (like roads) for joint interpretation.

Index Terms— SAR image combination, non-local filtering, change detection

1. INTRODUCTION
With the recent arrival of new high-resolution (HR) and very high resolution (VHR) optical and radar (SAR) sensors, researches led in the domain of data fusion are today booming, considering different kinds of multi-modality images, like multi-sensor images (eg: exploitation of a set of TerraSAR-X and CosmoSkyMed HR SAR images in [1]), multi-aspect images (eg: exploitation of four couples of VHR SAR airborne images acquired with orthogonal directions in [2]), multi-source images (eg: exploitation of a couple composed of an optical VHR Quickbird image and a TerraSAR-X HR SAR image in [3]), or multi-temporal images (eg : exploitation of four time series of TerraSAR-X HR SAR images in [4]).

Such a fusion framework is motivated by the ability to extract from these data complementary information, whose combination allows a better scene understanding, that turns out to be very useful for a lot of applications in remote sensing, such as object recognition [2, 5], 3D reconstruction [3, 6], and change detection [4, 7, 8].

Nevertheless, difficulties inherent to data fusion (absolute radiometric calibration, geometric differences due to conditions of acquisition, temporal decorrelation, requirement of fine registration) make the joint use of multi-modality images a very challenging task, especially in the case of multi-sensor HR SAR spaceborne images, that will be considered in this study and for which relatively few works have been proposed until now [1, 9, 10].

In this paper we investigate how two SAR images of two different sensors could be combined. We mainly focused on low-level step (pixel level) and the stability of features. We also propose a whole methodological processing chain although we only implemented some parts of it. Specially, the high level step, based on object recognition, has been developed only for the road network. The aim of a fusion approach of CosmoSkyMed (CSK) and TerraSAR-X (TSX) images could be an enhanced product of a denoised image with the identification of changed objects. Compared to usual change detection approach in multi-temporal SAR series, the difficulty is increased with the variability of sensors, thus raising some calibration aspects.

This paper is organized as follows: in section 2, we describe the preliminary steps before fusion of radar images of different sensors which are the registration and absolute calibration of the data ; in section 3, we present a global processing chain and describe some steps at the pixel, primitive or object level; in section 4, we analyze some preliminary results for a couple of TerraSAR-X and CosmoSkyMed stripmap images. The paper ends with some first conclusions and further works.

2. PRE-PROCESSING OF THE DATA: REGISTRATION AND CALIBRATION

2.1. Registration step
The sensor parameters are now available with a very good precision, allowing accurate projection on the ground of the acquired data. In our approach, one of the image, which will be refered as the “slave” image, will be projected in the geometry of the “master” image. This projection is done using the available sensor and platform parameters and a mean ground height (assuming a locally flat studied area). A re-sampling process (based on a zero-padding by a factor 4, followed by a bilinear pixel interpolation separately performed on the real and imaginary parts) is involved to compensate the slight resolution difference between both SAR images.
This first step is based on the sensor parameters and the result has to be refined with a translation. This translation is obtained by optimizing a global similarity criterion on calibrated images which is described in section 3.1. If the images are too dissimilar, it could be necessary to take manual ground control points, but the matching can be done automatically using the criterion proposed in section 3.1. Please note that the geometry of the “master” image is kept unchanged (particularly we stayed in slant range), while the second one is re-sampled in the first geometry.

2.2. Calibration step

Some steps of the processing chain are based on data comparison. For those steps, the values provided by both sensors have to be comparable. To do so, an absolute radiometric calibration of the data is performed. Using the parameter files of both data providing the calibration constants, the amplitude values are converted into backscattering coefficients or “sigma-zero” values (expressed in dB), allowing a direct numerical comparison as will be exploited further.

The calibration of the CosmoSkyMed image is done by the following formula:

$$\sigma_{0,db}^{CSK}(i,j) = 10 \log \left( \frac{|x_{(i,j)}|^2 \sin(\theta_{loc}^f) \times RSR^{2\alpha}}{CC \times RF^2} \right)$$

where log is the decimal logarithm, $x_{(i,j)}$ is the image value at pixel $(i,j)$, $\theta_{inc}^f$ is the reference incidence angle, $RSR$ is the reference slant range, $\alpha$ is the reference slant range exponent, $CC$ is the calibration constant, and $RF$ is the rescaling factor.

The calibration of the TerraSAR-X image is done by the following simplified formula (noise contribution is neglected):

$$\sigma_{0,db}^{TSX}(i,j) = 10 \log \left( |x_{(i,j)}|^2 \sin(\theta_{inc}^{loc}) \times k_s \right)$$

where $\theta_{inc}^{loc}$ is the local incidence angle and $k_s$ is the calibration and processor scaling factor. Due to some simplifying assumptions, some calibration errors may remain.

3. PROPOSED PROCESSING CHAIN

The comparison and combination of the two data can be done at different levels: pixel level, primitive level, and object level. The proposed chain aims at exploiting these different levels to build an enhanced image which could be a denoised image (ideally combining both information for stable areas) with changed objects. The synopsis of such a method is described in figure 1.

3.1. Pixel level

A first step of comparison of the two data is the detection of similar features (stable objects of the image). There have been many works for change detection between two SAR images of the same sensor, usually based on a pixel level comparison, for instance using thresholded ratio of images. In this paper, we propose to adapt the approach developed in [11] for two SAR images provided by two different sensors. In [11] the comparison between two noisy values is expressed as a generalized likelihood test ratio of two hypotheses: either the two values are provided by the same “true” underlying reflectivity, or they come from two distinct reflectivities. Since these “true” reflectivities are unknown, they are estimated with a maximum likelihood estimator. The expression for two SAR images of the same sensor is given in [11], but for different sensors, only the values after the absolute calibration step can be compared. The calibration step provides “sigma-zero” values after a logarithm transformation. Therefore, the expression has to be derived using Fisher-Tippett distributions and provides the following dissimilarity criterion for the comparison of two values:

$$d_{1,2} = 2 \ln \left( \cosh \left( \frac{\ln(10)}{20} \times (\sigma_{0,db}(i1,j1) - \sigma_{0,db}(i2,j2)) \right) \right)$$

where $\ln$ is the natural logarithm and $\cosh$ is the hyperbolic cosine.

A map of similarity is computed for each pixel by computing a similarity criterion for a small patch centered on this pixel, using the previous equation and an independence hypothesis of the pixels of the patch. To take into account the slight inaccuracy of the registration step, the best similarity value is kept for each pixel allowing plus or minus one pixel in both directions. Improved results should be obtained using locally adaptive patches. Indeed, in case of elevated objects and different incidence angles, the neighborhood of the pixel will be varying in the patch. The binary similarity map is then obtained by thresholding the patch similarity values. The non-stable areas identified by this approach may correspond to the two following situations: change between the two dates or different appearance of the objects due to the difference of incidence angles.

Besides the detection of stable features, this similarity criterion can also be used to obtain an extension of the spatial filter algorithm, proposed by Deledalle et al [12], in a simplified spatio-temporal version, but using the calibrated data. A discussion on multi-temporal filtering can be found in [13].
3.2. Primitive level

Some features, characteristic of objects of interest, are extracted from both registered images using un-calibrated data through the following sub-steps:

- Pre-processing: the data are independently spatially filtered using the patch-based non-local mean (NLM) denoising algorithm, proposed by Deledalle et al. [12] and dedicated to SAR imagery.
- Punctual target extraction: The extraction of bright points is performed on initial data, through the mean intensity ratio-based detector, proposed by Lopes et al. [14].
- Linear feature extraction: The extraction of bright lines and dark lines is performed on initial data, through the constant false alarm rate (CFAR) detector, proposed in [15].
- Homogeneous region extraction: Some sufficiently large homogeneous regions, likely to help to a better scene understanding, are extracted on filtered data, by combining the result of a markovian segmentation (using a k-means classification as initialization) with the result of the previous bright and dark network extraction [15, 16] (defining properly the region limits).

These primitives can be compared by direct logical AND, but this approach requires very fine registration and objects on the ground.

3.3. Object level

As previously said, due to the difference of incidence angles between the sensors and to their different sensibilities (although with the same band, probably due to the differences in the synthesis of the complex data - spectrum apodization, etc.), the low level comparison or direct comparison at primitive level usually gives limited results. Therefore the previous steps should be used as preliminary steps to build a joint interpretation of the image in terms of objects. For instance, building, bridge, river, etc., detection approaches have to be investigated and improved using both data.

In this paper, we will only show how the combined use of the two data can improve the road detection step. As proposed in [16], the information on linear structures is combined at low level. Then some structural information is used to retrieve the network. In this way, the false alarm rate can be reduced by using both data.

4. APPLICATION ON A COUPLE OF TERRASAR-X AND COSMOSKYMED STRIIPMAP IMAGES

4.1. Presentation of the available data set

The methodological chain is applied on a couple of two TerraSAR-X and CosmoSkyMed images (figure 2), acquired in stripmap mode (ascending pass, H-H polarization), imaged respectively at 31th May 2009 and at 12th March 2009 in the region of Saint-Gervais - Mont-Blanc in France (mean altitude of 638 meters) and presenting a respective spacing of 1.98 meter and 2.07 meters in azimuth and 1.37 meter and 1.18 meter in range.

They present visually a quite different content that can be explained by several causes: a large gap (about 10 degrees) between both incidence angles (implying important differences especially in mountaineous areas due to overlay phenomenon), a noticeable difference in terms of aesthetic quality (eg: shadow contrast, bright point focusing, etc.) proper to each sensor processing (specially SAR synthesis) and some true semantic changes due to a temporal gap (about 3 months) between both images.

4.2. Analysis of the results

At the low level, the similarity map is showed in figure 3 (top image). It depends on the chosen threshold but globally very few areas are considered as stable with a $3 \times 3$ analysis window. This is not suprising since there are many changes: in the moutainous part due to the relief; in urban areas due to the different incidence angles; in agricultural areas due to the
seasonal variations; and probably some calibration errors.

Besides, the joint filtering of the data did not bring a significant improvement compared to the filtering of a single image. This is mainly due to the changes between the two images (change in time and incidence angle).

At the primitive level, very few bright targets and bright lines have been identified as stable (see figure 3, middle image). A deeper analysis should be led, but the resolution is probably a limitation and more stable features could be obtained with spotlight data. The object level with interaction of primitives should permit to go a step further towards image interpretation. The combination of both images for road detection (see figure 3, bottom image) permits to reduce the false alarm rate and globally improves the results.

5. CONCLUSION AND FURTHER WORKS

In this paper, a first attempt for the joint exploitation of a multi-sensor SAR couple has been proposed. Different levels of comparison have been investigated and a similarity criterion for calibrated data has been derived. The necessity of high level tools able to detect objects of interest has been highlighted and will be subject of further works.

6. REFERENCES