TIME-VARYING SEGMENTATION FOR MAPPING OF LAND COVER CHANGES

François Aspert[1], Meritxell Bach Cuadra [1], Alessio Cantone[2], Francesco Holecz[2], and Jean-Philippe Thiran[1]

[1]*Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, SWITZERLAND, Email: francois.aspert@epfl.ch*
[2]*Sarmap S.A., Cascine di Barico, 6989 Purasca, SWITZERLAND, Email: francesco.holecz@sarmap.ch*

ABSTRACT

We propose a new set of algorithms to analyze in an automated fashion multi-temporal (vector-valued) SAR sequences, taking advantage of information redundancies and complementarities. Our goal is twofold: to automatically extract coherent regions and to analyze backscattering coefficients across these consistent regions. The proposed approach allows to discriminate between natural (special weather conditions or to the growth of crops inside fields) and human (early harvest of cultures) changes which is not possible with existing techniques. To this end, several issues are tackled: filtering, edge detection and image segmentation methods capable of exploiting this multi-temporal information. Moreover the system is designed so as to be completely sensor independent and to merge information from different spaceborne systems without any a-priori knowledge about the type of sensor used.

1. INTRODUCTION

Synthetic Aperture Radar (SAR) images are corrupted with a significant amount of multiplicative speckle noise, reducing the achievable accuracy when aiming to perform monitoring by means of a segmentation stage. However, they can provide data at any time under all weather or illumination conditions which is highly suitable for monitoring purposes. In this paper, we propose a new framework that takes advantage of SAR multi-temporal sequences information redundancy for improving the segmentation step effectiveness. Moreover, the whole processing chain has been chosen to be fully sensor-independent as available data composing a data set can originate from any type of platform.

First we propose to apply anisotropic diffusion scheme based on a multi-temporal gradient formulation to remove the speckle noise present in the image while preserving and enhancing (in some cases) the relevant edges. We apply subsequently a multi-temporal extension of the Canny edge detector in order to generate an global edge map consistent across time. Using the information obtained in the previous step, we use the edge map as side information and the filtered image to extract a reliable region map for the scene pictured by the multi-temporal sequence. Finally, in order to retrieve the particular details and changes as well as the unnatural changes of the single images, we refine the global region map for each image through a generic statistical scene segmentation technique. In what follows, we will present in order each steps involved in the time-varying segmentation. Then, we will present and discuss the different results before giving the conclusions and the possible future orientations of the work.

2. FILTERING

The first task to perform for an efficient segmentation is a regularization step. A lot of research work have been conducted in the specific domain of SAR images speckle suppression. In [1], some local statistics are taken into account whereas [2]makes a Maximum A Posteriori (MAP) estimation of the true Radar Cross Section (RCS) value. Each of these filters possess its corresponding multichannel extension using either multivariate statistical description [3] or linear combination of multi-temporal images sequence as in [4], improving the global speckle reduction. However most of these types of filters can be strongly dependant of the type of sensors used (i.e: Gamma-Map) since based on statistical scene descriptors. Also, if features masks are used (i.e: structured Gamma-Map), a accuracy loss can be introduced when regarding particular shape preservation due to the lack of a priori information about size and type of the features present in the image. Therefore, in order to take advantage of the redundant information available when using multi-temporal sequences while being fully independent regarding the origin of images composing the input sequence, we choose here to use an hybrid multi-temporal anisotropic diffusion scheme.

2.1. Single Image Diffusion Filtering

First introduced in [5] for single optical images, this particular type of filtering allows a high level of regularization in homogenous areas while preserving the relevant features ultimately used for segmentation (edges or more

generally discontinuities). For a continuous image, diffusion on image $\mathbf{I}$ may be enacted by the partial differential equation 1:

$$\frac{\partial \mathbf{I}}{\partial t} = \text{div}[c(\|\nabla \mathbf{I}_\sigma\|) \cdot \nabla \mathbf{I}_\sigma], \quad (1)$$

where ∇ is the gradient , div is the divergence operator, and c, the conduction coefficient is a matrix of diffusion coefficients of the same size as $\mathbf{I}$. c is designed to be a non-linear function of the smoothed image gradient magnitude $\|\nabla \mathbf{I}_\sigma\|$. The design of c if extremely important. Black [6] made an in-depth study of the design of c and link the Perona diffusivity function to the weighting functions of robust statistical estimation. This led to another diffusivity function from the Tukeys biweight:

$$c(\|\nabla \mathbf{I}_\sigma\|, \lambda) = \begin{cases} \frac{1}{2}[1 - (\frac{\|\nabla \mathbf{I}_\sigma\|}{\lambda})^2]^2 & \|\nabla \mathbf{I}\| \leq \lambda \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Where λ refers to the sensitivity parameter. But, the main drawback of nonlinear diffusion is that such a technique leaves the edge features unfiltered. To overcome this situation, Weickert [7] introduced edge-direction sensitive diffusion. The amount of diffusion is controlled by a matrix $\mathbf{D}$ (also called diffusion tensor) of values specifying the diffusion importance in the features direction. The anisotropic diffusion is thus described by Eq. 3.

$$\frac{\partial \mathbf{I}}{\partial t} = div[\mathbf{D}(\nabla \|\mathbf{I}_\sigma\|) \cdot \nabla \mathbf{I}_\sigma], \qquad \mathbf{D} = \begin{pmatrix} a & b \\ b & c \end{pmatrix}, \quad (3)$$

where,

$$a = \phi_1 \cos^2 \alpha + \phi_2 \sin^2 \alpha, \quad (4)$$

$$b = (\phi_1 - \phi_2) \sin \alpha \cos \alpha, \quad (5)$$

$$c = \phi_1 \sin^2 \alpha + \phi_2 \cos^2 \alpha. \quad (6)$$

Where α is the direction of the gradient (maximum variation angle). ϕ_1 controls the diffusion along the gradient whereas $\phi 2$ will be in charge of the filtering process perpendicular to this gradient. Therefore, ϕ_1 h will be chosen to behave in the same way as c in nonlinear diffusion. ϕ_2 will be fixed to a constant value as we require edges to be smoothed uniformly.

2.2. Vector-Valued Diffusion Filtering

For vector-valued diffusion, the difference with the work done in [8] and explained in section 2.1 lies in the diffusion amount computation which is no longer varying with a single-image gradient. The choice is made to measure the gradient using the whole set of images. The most natural choice is then to use the reliable formulation for gradient computation with vector data stated in [9] used by Sapiro [10] which takes the gradient as a two dimensional manifold embedded in $\Re^m$. We obtain for the multi-temporal image sequences the following First Fundamental Form (FFF):

$$d\mathbf{f}^2 = \begin{pmatrix} dx \\ dy \end{pmatrix}^{\mathbf{T}} \begin{pmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{pmatrix} \begin{pmatrix} dy \\ dy \end{pmatrix} \quad (7)$$

where,

$$\begin{cases} g_{11} = \sum_{i=1}^{m} \nabla \mathbf{I}^2_{\sigma,(i,x)}, \\ g_{12} = \sum_{i=1}^{m} \nabla \mathbf{I}_{\sigma,(i,x)} \nabla \mathbf{I}_{\sigma,(i,y)}, \\ g_{22} = \sum_{i=1}^{m} \nabla \mathbf{I}^2_{\sigma,(i,y)}. \end{cases} \quad (8)$$

Where $\mathbf{I}^2_{\sigma,(i,x)}$ and $\mathbf{I}^2_{\sigma,(i,y)}$ stands respectively for gradient estimation along columns and lines. The direction and magnitude of the maximum and minimum rate of change corresponding to the computed gradient directions can be then extracted from the FFF eigenvalues and eigenvectors in Eq. 7. Finally, Eq. 9 gives the practical framework of the anisotropic diffusion process.

$$\begin{cases} \frac{\partial \mathbf{I}_1}{\partial t} = \text{div}[\mathbf{D} \cdot \nabla \overrightarrow{\mathbf{I}}], \\ \vdots \\ \frac{\partial \mathbf{I}_n}{\partial t} = \text{div}[\mathbf{D} \cdot \nabla \overrightarrow{\mathbf{I}}], \end{cases} \quad (9)$$

where $\overrightarrow{\mathbf{I}}$ corresponds to the whole multi-temporal image sequence and $\mathbf{I}_i$ is the i^{th} image in the sequence. Therefore, each image is filtered separately using the global sequence information, taking into account features from all images.

3. EDGE DETECTION

The method chosen here is derived from the well-known Canny edge detector [11]. In opposition to Ratio Edge Detector (RED) [12], Canny detection has the advantage of nor relying on statistical a-priori on regions distributions and to be a purely analytical scheme for detecting contours. Moreover, it is less sensitive to noise than other analytical methods such as the Marr-Hildreth operator. Finally, such a detector is valid for all possible edge orientations and subject to a limited number parameter estimation.

3.1. Single image

For single image, the Canny edge detector consist in the sequential execution of five steps. First the image is smoothed to eliminate and noise. It then finds the image gradient to highlight regions with high spatial derivatives. The algorithm then tracks along these regions and suppresses any pixel that is not at the maximum (non-maximum suppression) in the gradient direction. The gradient array is further reduced by hysteresis thresholding.

3.2. Multi-temporal images

In order to process vector-valued data, we propose here to extend the algorithm presented for single image using the work done in section 2.2. Although some partial applications of the Canny technique to vector-valued data

has been found in [13] (regarding hysteresis thresholding) , as far as we know, the whole Canny algorithm has not been applied to multi-image yet. In the vector-valued data case, the main difference with original Canny lies in the computation of the direction and magnitude of gradient. For each pixel, a multi-temporal gradient magnitude λ can be extracted from image sequences FFF (Eq. 7) and computed as:

$$\lambda = \frac{\sqrt{(g_{11} - g_{22})^2 + 4g_{12}^2}}{2}. \quad (10)$$

The gradient orientation θ can be derived as

$$\theta = \frac{1}{2} \arctan \frac{2g_{12}}{g_{11} - g_{22}}. \quad (11)$$

Therefore, applying smoothing, non-maximum suppression and hysteresis thresholding with the gradient magnitude and directions defined by equations 10 and 11 leads to the definition of a multi-temporal Canny edge detector. It is important to notice that this approach still satisfy the criterions of detection, localization and clear response defining an optimal edge detector.

4. REGION SEGMENTATION

4.1. Multi-temporal Region Segmentation

Found in [14], this segmentation algorithm regards the task of finding regions within the image as a coding task. The approach is based on an hypothetical image coding scheme that decomposes a given image into homogeneous segments and encodes these segments independently of each other. Such an approach also supports situations in which the encoder and the decoder have some common information about the image, i.e. in our case an edge map extracted from the filtered images, to improve segmentation. Thus, the entire coding procedure can be then described as a two-part source channel coding with side information. The final region delineation is achieved by a region growing step satisfying the minimum code length constraint described.

The complete cost function that measures the expected code length is expressed in the following equation.

$$H_{MC} = \frac{1}{2} \left(\sum_{\nu \in \gamma} \frac{n_\nu}{d} \sum_{i \leq M} \log \sigma_{\nu,i}^2 + H^{BND} \right). \quad (12)$$

If M stands for the number of available images in the sequence and $\sigma_{\nu,i}$ denotes the variance of segment ν in image i. H^{BND} is the cost of encoding segments boundary

The final step for achieving the required segmentation is to find the segmentation corresponding to a minimal cost with equation 12. To this end, a region growing stage is taken: First, images are divided in one-pixels wide distinct regions. Then, all the merging costs between 4-connected regions are computed and stored in a list. Hence, the merging operation corresponding to the lower cost can be selected. Finally, the merge list is updated. The only parameter intervening here is the number of merge operations to perform. This is a crucial matter as it defines the final scale for segmentation. Here this step has been highly summarized, further information can be found in [14]

4.2. Time Varying Segmentation (TVS)

At this point, it is important to note that a first preprocessing chain allowing the generation of common edge and segmentation map for image sequences is now available. But such global maps are not suffcien when regarding the particular task of recovering the single-date specific changes occurring in each image. Indeed these changes can originate either from natural changes or from punctual human intervention. In the first case, the global segmentation will be accurate whereas in the second case, the potential variations appearing only in a minority of images within a sequence will not be detected and only highly redundant structures will be retrieved.

We here aim at tracking across time each potential abnormal radiometric changes for each segment. We thus propose to take the global region and contour map obtained previously and process each obtained segment separately. Starting with a high number of classes (k_{max} user defined), the algorithm will perform a Finite Gaussian Mixture Modeling (FGMM) fitting for the investigated segment for class numbers in the range$[1, k_{max}]$ at each date of the sequence. Then, using Minimum Distance Length (MDL) criterion, the optimal number representing the segment will be computed. Finally, if needed, a Maximum A Posteriori (MAP) segmentation will be performed to output the single-date segmentation. The whole processing is detailed in Algorithm 1.

4.2.1. Finite Gaussian Mixture Model Fitting

Let us index n data points to be represented using k classes. Let us furthermore denote each data feature vector (observed data) by $x_j \in [0, 255]$ with $j \in [1, n]$. The family of random variables y_i, $i \in [1, k]$ represents the different required spectral classes.

Now, let us suppose that all the random variables x_j are identically and independently distributed. Then, the probability density function of the intensity pixel can be defined by:

$$P(x_j) = \sum_{i=1}^{k} P(y_i) P(x_j | y_i). \quad (13)$$

A density function of this form is called *finite mixture (FM)* density. $P(y_i)$ is the *prior* probability of each class y_i, also called *mixing* parameter. $P(x_j|y_i)$

Algorithm 1: Time-varying segmentation

INPUT: ANLD filtered data set $\mathbf{I} = \{I_1, \ldots, I_P\}$.
INPUT: Region map set $\mathbf{M} = \{R_1, \ldots, R_N\}$
INPUT: k_{max}
OUTPUT: $\mathbf{TVS} = \{TVS_1, \ldots, TVS_P\}$
PROCEDURE:
for i=1 to P **do**
  for j=1 to N **do**
    0: Initialize mean and variance for k-classes partition using k-means clustering with random clusters
    for k=k_{max} to 1 **do**
      1: Get region $\mathbf{R}$ as $I_i \cap R_j$
      2: Compute a k-classes FGMM on segment histogram.
      3: Compute k-classes MDL criterion
      4: Compute a $k-1$-classes FGMM from k-classes mixture using AEM.
    end for
    5: Using optimal FGMM (mixture with highest relevance), output the corresponding MAP segmentation $\mathbf{R}_{TVS}$
    6: $\mathbf{TVS}_i = \mathbf{TVS}_i \cup \mathbf{R}_{TVS}$
  end for
end for

is the probability density function of x_j given the class y_i, also named *transition* or *conditional* probability. A reasonably assumption for the probability density function of the observing data x_j given the underlying class y_i is the Gaussian function. If $\theta_{y_i} = \{\mu_{y_i}, \sigma_{y_i}\}$ represents the Gaussian PDF parameters for the y_i spectral class, the fitting process aims to find the the maximum likelihood (ML) estimation of $\mathbf{\Theta}_k = \{\theta_{y_1}, \ldots, \theta_{y_k}, P(y_1), \ldots, P(y_k)\}$ based on a set of N independent observations $\mathbf{x}_{obs} = \{x_1, \ldots, x_n\}$ leads to:

$$\widehat{\mathbf{\Theta}}_k = \arg\max_{\mathbf{\Theta}_k} L(\mathbf{\Theta}_k, \mathbf{x}_{obs}), \tag{14}$$

where $L(\mathbf{\Theta}_k, \mathbf{x}_{obs})$ is the log-likelihood function. In general, Equation 14 has no closed form solution but it can be approached with the EM algorithm. The algorithm stops when the relative difference between estimated means is lower than the stopping criterion ϵ.

4.2.2. MDL criterion

In EM based approaches, the estimate number $\hat{k}$ of classes can be defined as minimizing some cost function:

$$\hat{k} = \arg\min_k \{\mathcal{C}(\widehat{\mathbf{\Theta}}_k, k), k \in [1, k_{max}]\}, \tag{15}$$

where k_{max} is the spectral classes upper bound. Under this general formulation, we find the MDL criterion [15] in which the cost function is

$$\mathcal{C}_{MDL}(\widehat{\mathbf{\Theta}}_k, k) = -L(\widehat{\mathbf{\Theta}}_k, x_i) + \frac{N(k)}{2} \log n, \tag{16}$$

where $N(k)$ is the number of parameters needed to specify a k-component mixture. In this case, it reduces to $N(k) = (k-1) + 3k/2$. In our case, we recursively computes the MDL value for decreasing values of k to extract the optimal number of gaussian that will fit the histogram.

4.2.3. AEM algorithm

The $k-1$-component mixture is obtained from the k-component one by selecting and merging the two closest distributions.

$$(y_i, y_j) = \arg\min_{(i,j)} (\hat{P}(y_i) + \hat{P}(y_j))$$
$$\mathcal{D}_s[P(\mathbf{x}_{obs}|\hat{\theta}_{y_i}), P(\mathbf{x}_{obs}|\hat{\theta}_{y_j})]. \tag{17}$$

This is done using the symmetric Kullback-Leibler divergence measure [16]. It can be expressed in our univariate case for two Gaussian probability density functions y_1 and y_2 as:

$$\mathcal{D}_s[P(\mathbf{x}_{obs}|\hat{\theta}_{y_1}), P(\mathbf{x}_{obs}|\hat{\theta}_{y_2})] \tag{18}$$
$$= \frac{1}{2}\left[(\sigma_{y_1} - \sigma_{y_2})(\sigma_{y_1}^{-1} - \sigma_{y_2}^{-1}) + \frac{(\mu_{y_1} - \mu_{y_2})^2}{\sigma_{y_1}^{-1} - \sigma_{y_2}^{-1}}\right]$$

Finally, $\widehat{\Theta}_{k-1}$ is computed as in [17]:

$$P(y_m) = \begin{cases} P(y_m), & m \neq 1 \\ P(y_1) + P(y_2), & m = 1 \end{cases} \tag{19}$$

$$\mu_{y_m} = \begin{cases} \mu_y, & m \neq 1 \\ \frac{P(y_1)\mu_{y_1} + P(y_2)\mu_{y_2}}{P(y_1)+P(y_2)}, & m = 1 \end{cases} \tag{20}$$

$$\sigma_{y_m} = \begin{cases} \sigma_{y_m}, & m \neq 1 \\ \frac{\Gamma(y_1,y_2)}{P(y_1)+P(y_2)} - \mu_{y_1}^2, & m = m_1 \end{cases} \tag{21}$$

where,

$$\Gamma(y_1, y_2) = P(y_1)(\sigma_{y_1} + \mu_{y_1}^2) + P(y_2)(\sigma_{y_2}\mu_{y_2}^2). \tag{22}$$

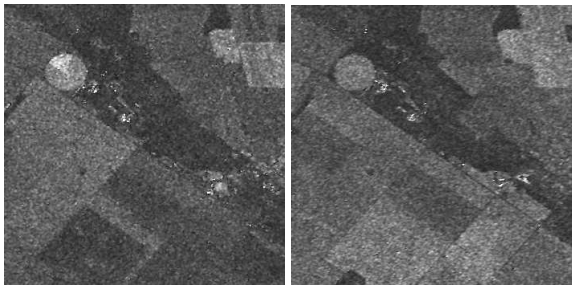
5. EXPERIMENTS AND RESULTS

5.1. Datasets

This section presents the dataset that will be used in this study for testing and validation purposes. Figure 1 shows a 350*350 pixels wide cut from a set of 16 images of South-Africa agricultural land acquired between 07/12/2004 and 10/08/2005. Images from this dataset originates from different sensors (Envisat ASAR and Radarsat1). We will further refer to this set as South-Africa.

5.2. Results

Here, results from each addressed topic are presented. Figure 2 highlights the performance of multi-temporal



(a) 06 January 05 (b) 11 June 05

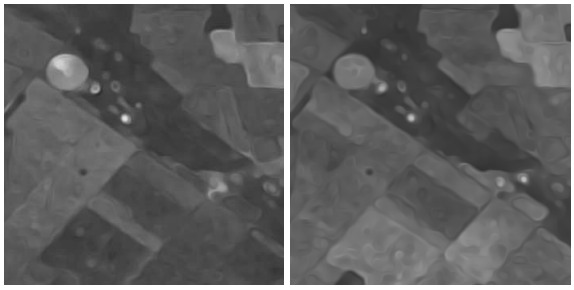
Figure 1. South-Africa original multi-temporal sequence

anisotropic diffusion at two different dates. To this end, we used: 150 diffusion iterations of the algorithm, an anisotropy parameter $\phi_2 = 0.5$ and a gaussian blurring kernel of variance $\sigma = 3$. The threshold λ was fixed as the median value of the gradient magnitude computed over 75×75 non-overlapping windows. Here,we can observe the high regularization power of anisotropic diffusion. The speckle is greatly decreased while important features are preserved. Most of small artifacts are removed, but the high level of anisotropy introduces a rounding effect when filtering sharp corners.

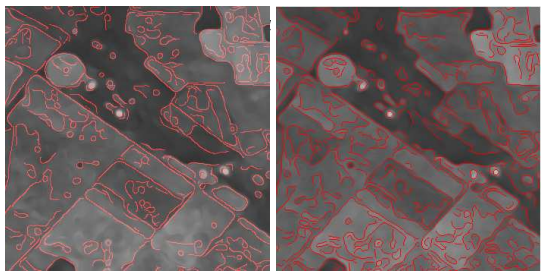
For edge detection we used a spatially varying hysteresis threshold T_1 computed as the 80% percentile over 75×75 non-overlapping windows. T_2 has been obtained using $T_2 = 0.1 * T_1$. In Figure 3, the edge maps obtained with the multi-temporal Canny detector 3(a) and single image Canny detector 3(b) are shopn. We observe that good results are obtained with the multi-temporal Canny edge detection. Indeed, most of the important structures are detected. Compared to the result obtained when performing the single date detection, we can see that much less false-edge detection are present. This also corroborate the efficiency of the anisotropic filtering.

Figure 4 shows a portion of the segmentation achieved at two different scales: 150 final segments in Fig. 4(a) and 250 final segments in Fig. 4(b). We clearly notice the importance of selecting the appropriate level for the segmentation as too few final segments can induce severe deviation of algorithm and to the merge of distinct segments. However in the case of 250 final segments we visually obtain an accurate segmentation where homogenous regions have been retrieved.

For TVS segmentation, we fixed $k_{max} = 15$ for AEM and $\epsilon = 0.001$ for FGMM. Figures 5 and 6 presents the result of the time varying segmentation algorithm used on two particular regions obtained in Figure 4. Figures 5(a) - 5(d) and 6(a) - 6(d) are the original images whereas Figures 5(e) - 5(h) and 6(e) - 6(h) is the corresponding time varying segmentation. In this experiment, we can verify for two different regions the behavior of TVS algorithm. It successfully estimated the number of classes in when a particular segment is constant and split it into subsegments. In Fig. 5(e), we also see the limitation of MAP segmentation which can lead to the creation of small isolated segments.

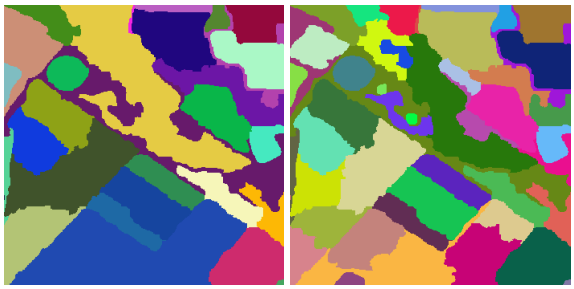


(a) 06 January 05 (b) 11 June 05



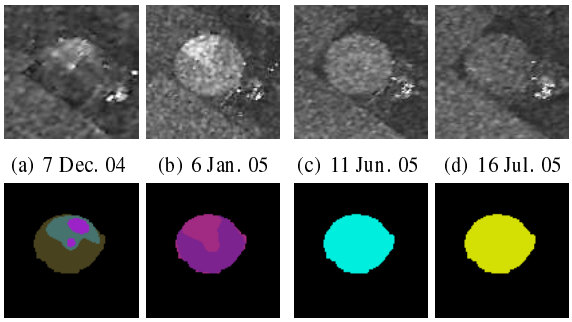
(a) 11 June 05 multi-temporal Canny edge detection (b) 11 June 05 Canny edge detection

Figure 3. Filtered images with superposed edge map



(a) 150 segments (b) 250 segments

Figure 4. Multi-temporal Region Segmentation



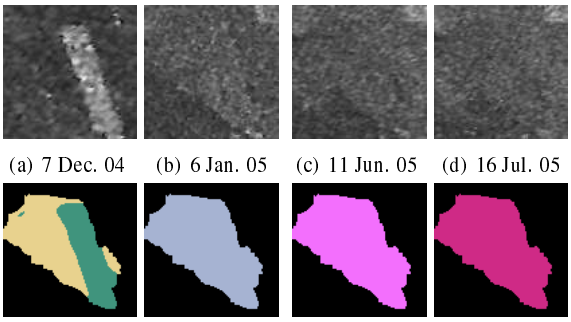
(a) 7 Dec. 04 (b) 6 Jan. 05 (c) 11 Jun. 05 (d) 16 Jul. 05

(e) 7 Dec. 04 (f) 6 Jan. 05 (g) 11 Jun. 05 (h) 16 Jul. 05

Figure 5. Time varying Segmentation for region 1

6. CONCLUSION

A complete tool-chain has been designed and tested for retrieving multi-temporal SAR sequences segmentations by applying successively a) Multi-temporal ANLD



(a) 7 Dec. 04 (b) 6 Jan. 05 (c) 11 Jun. 05 (d) 16 Jul. 05

(e) 7 Dec. 04 (f) 6 Jan. 05 (g) 11 Jun. 05 (h) 16 Jul. 05

Figure 6. Time varying Segmentation for region 2

b)Multi-temporal edge detection c) Multi-temporal Region segmentation and d) Single image time-varying segmentation. It has been shown that the use of multi-temporal data to perform speckle filtering clearly improved the global robustness and effectiveness of the whole processing even if some non-redundant image features are lost. The multi-temporal canny edge detection clearly outperformed the traditional edge detectors such as single image Canny algorithm with much less of false edges detection and better behavior for retrieving global image sequences structures. A direct application for the multi-temporal region segmentation has also been investigated with TVS segmentation. It enabled to determine a precise profile of radiometric changes within time series. Compared to existing change detection algorithms, it has the advantage of being insensitive to region consistent radiometric changes. Also, it does not require the use of multivariate statistics which is extremely costly in term of computation. Since it is a region-based approach it not subject to pixel-based variations techniques.

It is important to note here that even if the building blocks of our tool chain are not theoretically new, they have been assembled and designed in a fully operational and user-ready way. They also have been integrated into the Sarscape [1] framework with which hundreds of images have been processed.

As a conclusion, we can say that the proposed methods are promising and even if some improvements can be taken to enhance performance(using Median Absolute Deviation thresholding for filtering, replacing MAP in d) by another region growing algorithm...) it can already be used for multi-temporal SAR data processing.

REFERENCES

1. V. S. Frost, J. A. Stiles, K. S. Shanmugan, and J. C. Holtzman. A Model for Radar Images and its Application to Adaptive Digital Filtering of Multiplicative Noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4:157–166, 1982.

2. A. Lopès, E. Nezry, R. Touzi, and H. Laur. Maximum a Posteriori Speckle Filtering and First Order Texture Models in SAR Images. In *Proceedings of IGARSS*, 1990.

3. E. Nezry, F. Zagolski, A. Lopès, and F. Yakam-Simen. Bayesian Filtering of Multi-Channel SAR Images for Detection of Thin Structures and Data Fusion. In *Proceedings of SPIE*, volume 2958, pages 130–139, 1996.

4. S. Quegan and J.J. Yu. Filtering of Multichannel SAR Images. *IEEE Transactions on Geoscience and Remote Sensing*, 39(11), November 2001.

5. P. Perona and J. Malik. Scale-Space and Edge Detection Using Anisotropic Diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990.

6. M.J. Black, G. Sapiro, D.H. Marimont, and D. Heeger. Robust Anisotropic Diffusion. *IEEE Tansactions on Image Processing*, 7(3), March 1998.

7. J. Weickert. *Anisotropic Diffusion in Image Processing*. B.G.Teubner, 1998.

8. S. T. Acton and J. Landis. Multi-Spectral Anisotropic Diffusion. *International Journal of Remote Sensing*, 18(13):2877–2886, January 1997.

9. S. Di Zenzo. A Note on the Gradient of a Multi-Image. *Computer Vision, Graphics, and Image Processing*, 33:116–125, January 1986.

10. G. Sapiro and D.L. Ringach. Anisotropic Diffusion of Multivalued Images with Applications to Color Filtering. *IEEE Transactions on Image Processing*, 5(11), November 1996.

11. J. Canny. A Computational Approach to Edge Eetection. *IEEE Tansactions on Pattern Analysis and Machine Learning*, PAMI-8:679–698, 1986.

12. R. Touzi, A. Lopes, J. Bruniquel, and P. Vachon. Coherence Estimation for SAR Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 37(1):135–149, 1999.

13. R. Nevatia and K. Ramesch Babu. Linear Feature Extraction and Description. *Computer Graphics and Image Processing*, 13:257–269, 1980.

14. L. Hermes and J.Buhmann. Boundary-Constrained Agglomerative Segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 42(9), September 2004.

15. J. Rissanen. *Stochastic Complexity in Statistical Inquiry Theory*. World Scientific Publishing Co., Inc., River Edge, NJ, USA, 1989.

16. S. Kullback. *Information Theory and Statistics*. John Wiley and Sons., New York, 1959.

17. Mário A. T. Figueiredo, José M. N. Leit ao, and Anil K. Jain. On Fitting Mixture Models. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 54–69, 1999.

[1] http://sarmap.ch