FUSION OF MULTI-VARIATE EDGE DETECTORS FOR HIGH-RESOLUTION POLARIMETRIC SAR IMAGES

D. Borghys, C. Perneel, M. Acheroy

Signal & Image Center, Royal Military Academy
Renaissancelaan 30, B-1000 Brussel, Belgium
Tel: +32 2 737 64 74, Fax: +32 2 737 64 72, E-mail: Dirk.Borghys@elec.rma.ac.be

ABSTRACT

Edge detection in SAR images is a difficult problem due to the presence of speckle. However, the statistical properties of speckle in uniform regions of a SAR image can be used for the development of edge detectors. For single-channel multi-look intensity images, the ratio-detector [1] is widely accepted to be the optimal edge detector. For multi-channel data, it is possible to apply the ratio-detector to each separate channel and fuse the results. Alternatively multi-variate methods can be used. They treat the different channels as a whole and there is no need for subsequent fusion. Furthermore they take the inter-channel correlation into account. We already proposed two edge detectors based on multi-variate statistical hypothesis tests. The first one is based on a test for the difference of variance and applied to SLC images, the second uses a test for the difference of means and is applied to log-intensity images. The two multi-variate edge detectors give complementary results. Hence the idea to fuse these results. Fusion of the results of both detectors for equivalent false alarm thresholds gave poor results. In the article we propose a new method to find the region of optimal fusion for the two edge detectors. The method is based on the combination of two statistical methods for investigating the complementarity of "experts" and a figure-of-merit for edge detection. Results of applying the proposed method to a high-resolution, polarimetric, L-band E-SAR image are shown.

Keywords: edge detection, polarimetric SAR images, multi-variate statistics, fusion

1 INTRODUCTION

Edge detectors that work well in optical images fail in SAR images due to the particular properties of the speckle in SAR images [1]. The speckle has the characteristics of a multiplicative noise (in intensity images) with a very low signal-to-noise ratio. This makes pixel-wise methods using a simple filtering mask, as used in optical images, inappropriate. A commonly adopted solution is to take into account larger neighbourhoods of each pixel to decide whether an edge passes through that pixel. This is done by scanning the image with a set of two adjacent rectangular windows and estimating some statistic in both windows. The difference between both estimations is an indication that the edge between the two windows corresponds to an edge in the image. This gives an answer for one possible edge orientation. The set of rectangles is rotated around the current pixel to verify the presence of an edge along other orientations. Mostly 2, 4 or 8 orientations are tested and results combined. The existing methods (e.g. [2, 1, 3]) were mostly applied on single-band multi-look intensity images and differ by the comparison criterion that is used (see [4] for an overview). When dealing with multi-channel SAR images, e.g. polarimetric images, existing methods can be applied to each separate channel and the results fused. In [5] we proposed two new edge detectors for multi-channel SAR images. Both are based on multi-variate statistical hypothesis tests. The first tests the difference in means in log-intensity images and the second uses a test for the difference in variance and is applied on the complex images. Both are applied on single-look SAR images. Results show the two detectors to be complementary [6]. Therefore fusion was investigated. The effectiveness of the fusion not only depends on the fusion algorithm, but also on the range of outcomes of the individual detector's responses that are fused. Fusion of the results of both detectors for equivalent false alarm thresholds gave poor results. The main topic of this paper is the development of a new method to determine for which ranges of thresholds the results of both detectors can be enhanced significantly by fusion. The proposed method is based on the combination of two statistical methods for investigating the complementarity of "experts". This combination allows to establish the combination of detector thresholds for which fusion should give the best results. Once the ranges of thresholds are determined, standard fusion methods can be applied. The proposed method is applied to an L-band full-polarimetric (HH,HV and VV polarisation) E-SAR image[1].

2 MULTI-VARIATE EDGE DETECTORS

2.1 EDGE DETECTOR BASED ON A DIFFERENCE IN MEANS

The contour detector problem is transformed into a multi-variate hypothesis test, the null-hypothesis being that the pixels in the two scanning rectangles are samples from two populations with equal averages. This can then be tested using a Hotellings T^2 test [7] defined as:

$$T^2 = \frac{n_1 n_2 (\overline{\mathbf{X_1}} - \overline{\mathbf{X_2}})^t \mathbf{C}^{-1} (\overline{\mathbf{X_1}} - \overline{\mathbf{X_2}})}{n_1 + n_2}, \qquad (1)$$

where n_1 and n_2 are the number of observations in the two rectangles, $\overline{\mathbf{X_k}}$ is the average vector of the observations in window k and $\mathbf{C}$ is the pooled covariance matrix. In the null-hypothesis of equal population averages the trans-

[1] The presented work is the result of a collaboration with the Institute of Radio Frequency Technology and Radarsystems of the German Aerospace Center (DLR), who also provided the images.

formed statistic:

$$T_F = \frac{(n_1+n_2-p-1)T^2}{(n_1+n_2-2)p} \sim F(p, n_1+n_2-p-1), \quad (2)$$

where p is the number of variables and F is the Fisher-Snedecor distribution. The test is applied to the log-intensity images in which differences in radar reflectivity of uniform regions are reflected purely as variations of first order statistical moments. Therefore $p = 3$, i.e. we have one image per polarisation.

2.2 EDGE DETECTOR BASED ON A DIFFERENCE OF VARIANCE

The statistical hypothesis test for a difference of variance used here is the Levene test [7]. The samples from the two scanning windows are transformed in absolute deviations of sample means, e.g. for HH polarisation: $x_{ik}^{HH} \rightarrow | x_{ik}^{HH} - \overline{x}_k^{HH} |$ with k the index of the scanning rectangle and i the index of the observation within a scanning rectangle. The question whether two samples display significantly different amounts of variance is then transformed into a question of whether the transformed values show a significantly different mean [7] and the Hotellings T^2 test can again be used. This test is applied to the complex image, we thus have two component images per polarisation, i.e. the real and imaginary part of the SLC image and $p = 6$.

2.3 INFLUENCE OF SPATIAL CORRELATION

The derivation of Hotellings-T^2 test statistic (e.g. [8]) assumes that the covariance matrix of the mean is equal to the covariance matrix of a sample divided by the number of observations in the sample. This is only true when the observations are uncorrelated. However, in SAR images neighbouring pixels are correlated. This is partly due to the SAR system itself, in particular its Point Spread Function [4], and partly to texture in the terrain. The spatial correlation ρ causes the behaviour of the detectors based on the Hotellings test to deviate from the theoretical prediction, i.e. the test-statistic becomes too large and too many false alarms are found at a given theoretical constant false alarm (CFAR) threshold. For the part of ρ due to the SAR system, a correction factor can be determined for the test statistic [6]. However, even a slight terrain texture, increases the spatial correlation, resulting again in an increased number of false alarms. A solution is to sub-sample within the scanning rectangles. For fixed grid sub-sampling the theoretical CFAR thresholds can be derived from the average spatial correlation between points on that grid and when the sampling ratio is low enough slight terrain textures have only a minor influence on the detector.

3 FUSION OF EDGE DETECTORS

In order to determine for which ranges of thresholds the results of both detectors can be significantly improved by fusion, statistical methods for comparing experts are used. Each detector is considered as an expert that for each pixel gives its opinion on the presence of an edge. The comparison of the two "experts" is based on their results on two small regions of the SAR image. On the two regions the ground truth was delimited manually, i.e. the edges were indicated. This ground truth is used to determine the probability of detection P_d and false alarms P_{fa} for a given threshold for each detector. This information is used as input for the statistical methods that are used in the comparison. The statistical methods are described below. The two detectors are considered as two experts yielding a binary decision (edge, no-edge) at every pixel of the test images. The decision is based on the choice of a pair of thresholds (T_{Levene}, $T_{Hotellings}$). This decision can be validated using the ground truth and from this validation two contingency tables (see table 1) are determined that form the basis for the statistical comparison of the two experts. One table M_{Pd} describes the performance in terms of detection, the other one M_{FA} in terms of false alarms.

		Levene Test		
		No-Edge	Edge	
Hotellings Test	No-Edge	n_{11}	n_{12}	n_{10}
	Edge	n_{21}	n_{22}	n_{20}
		n_{01}	n_{02}	n

Table 1: *Contingency table for the two detectors*

If the two detectors fully agree, the elements n_{21} and n_{12} of the table are zero. If this is true for both M_{FA} and M_{Pd}, fusion will not be useful and any of the two experts can be chosen. If n_{21} and n_{12} are non-zero and almost equal, the two experts are providing complementary information and fusion is useful. The symmetry of the table is thus a first characteristic that gives an idea of the usefulness of fusion. The second characteristic is the interdependence of the two experts. The two characteristics can be assessed by statistical tests. The Mc Nemar and the Kappa test investigate resp. the symmetry and the interdependence.

3.1 THE MC NEMAR TEST

The Mc Nemar test [9] was designed to test for the significance of changes, e.g. before and after a medical treatment. In our case it will test whether the opinions of the two experts differ significantly. The hypotheses are defined as:

$$H_0: \quad p_{21} = p_{12} \quad H_1: \quad p_{21} \neq p_{12} \qquad (3)$$

with e.g. p_{12} = *Prob(Hotelling Result = NoEdge, Levene Result=Edge)*. If the null-hypothesis is verified the expected value $e_{12} = e_{21} = \frac{n_{12}+n_{21}}{2}$. The test statistic:

$$d^2 \quad = \frac{(n_{12}-n_{21})^2}{n_{12}+n_{21}} \quad \sim \chi^2(\nu = 1) \qquad (4)$$

3.2 THE KAPPA TEST

The Kappa test [10] is applied to the elements of concordance between the two experts, i.e. the elements n_{11} and n_{22} of the contingency table. It compares the values of these elements with the value the elements would have if the two experts were independent, as predicted by the marginal probabilities:

$$H_0: \quad p_{ii} = p_{0i}p_{i0} \quad H_1: \quad p_{ii} \neq p_{0i}p_{i0} \qquad (5)$$

The observed concordance is $p_0 = \frac{n_{11}+n_{22}}{n}$ and

$$p_e = \frac{\frac{n_{10}n_{01}}{n} + \frac{n_{20}n_{02}}{n}}{n} \tag{6}$$

is its expected value. The Kappa coefficient is defined as: $\kappa = \frac{p_0 - p_e}{1-p_e}$. It can be shown that if n is large enough ($n > 8$), the variable

$$Z_\kappa = \frac{\kappa}{\frac{\sqrt{p_e}}{\sqrt{n(1-p_e)}}} \sim \mathcal{N}(0,1) \tag{7}$$

For a threshold $z_{\alpha/2}$ three cases are distinguished: the discordant ($Z_\kappa < -z_{\alpha/2}$), the concordant ($Z_\kappa > z_{\alpha/2}$) and the non-concordant case ($-z_{\alpha/2} < Z_\kappa < z_{\alpha/2}$).

3.3 COMBINING THE TWO STATISTICAL TESTS

The two tests can be applied for a set of combinations of thresholds for the Levene and Hotellings test. For each pair of thresholds the contingency tables are determined, in test images, for false alarms and detections and the tests are applied. Different classes are defined as a function of the results of the two tests. Combining the two tests gives 6 possible values, but the symmetric concordant case is subdivided in two sub-classes. For the symmetric concordant case results can be improved by fusion if $n_{12} > 0$ and $n_{21} > 0$ because this means that the two detector have a similar number of false classifications but the miss-classified elements are not the same. Therefore the symmetric concordant case is split in two sub-cases, one where the discordant elements are zero (class 2) and one where they are significantly different from zero (class 1). The latter is the ideal case for fusion because both experts globally agree but locally can give a different advice. The other classes are:

cl	McNemar Test	Kappa Test	Consequence for Fusion
	Symmetr.	Concordant	
1		$n_{12}, n_{21} > 0$	Fusion useful
2		$n_{12}, n_{21} = 0$	Fusion not useful
3		Non-Conc.	Fusion not useful
4		Discordant	NA
5	Non-Symm.	Concordant	Fusion maybe useful
6		Non-Conc.	Choose best expert
7		Discordant	NA

In class 3 fusion is not useful. The discordant case (classes 4 and 7) was not observed in the test images. In class 5 results could be improved by a fusion strategy in two successive steps. The first (fastest or cheapest) expert is used for the first classification and, depending on the confidence in the results, the second expert is used in the second stage. In class 6 the threshold for one detector is very low and the other is high. The experts thus disagree and depending on whether the false alarms or the probability of detection is important the expert with the highest or lowest threshold value should be chosen. Fig. 1 and 2 resp. show the results for P_{fa} and P_d. In each region the number of the corresponding class is shown.

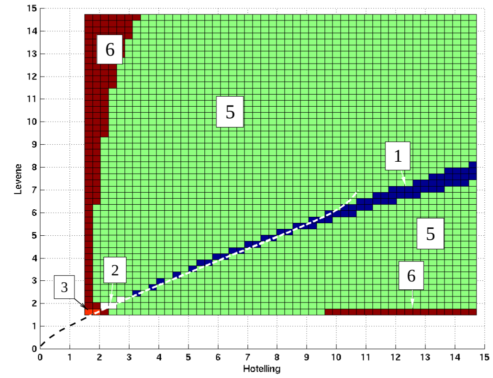

Figure 1: *Classification of 2D threshold space for P_{fa} (top) and P_d (bottom)*

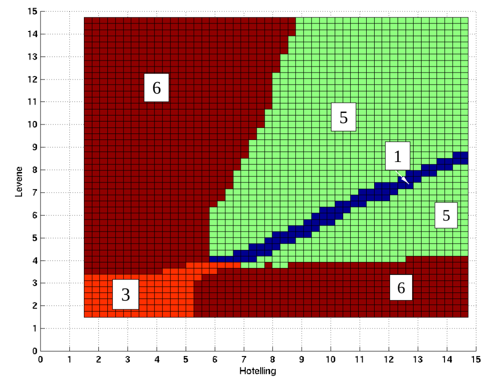

Figure 2: *Classification of 2D threshold space for P_d*

Class 1 is the most interesting for fusion: the two detectors have globally a similar performance and yet their opinions differ in individual pixels of the test image. The area is linear and the area for P_d overlaps with the one for P_{fa}. The dashed line in fig. 1 is the line of equal P_{fa} for the two detectors for sampling on a fixed grid and taking the maximum of the result of the 8 orientations of the scanning rectangles. The optimal area for fusion only partly coincides with this line, it deviates at high thresholds. Fusing the results of the two detectors by combining responses of each detector corresponding to the same P_{fa} is thus not optimal as confirmed by the results we obtained.

3.4 A FOM FOR EDGE DETECTION

Any detection algorithm makes a compromise between P_d and P_{fa}. A ROC curve can be used to see the evolution of the compromise when the threshold of the algorithm is varied. An alternative is to define a figure of merit (FOM) for detection. Based on the FOM for target detection introduced in [11] we define a FOM for edge detection as:

$$FOM = \frac{P_d}{\alpha P_d + \beta P_{fa}} \tag{8}$$

with $\alpha, \beta \in [0,1]$, $\alpha + \beta = 1$. α and β allow to vary the importance attached to false alarms or detected points. For a given α and β, the maximum of the FOM as a function of the detector threshold indicates the optimal threshold for the detector. The location of the maximum shifts when α

and β are varied. For two detectors, the FOM can be calculated for each and their product and sum resp. give the result for fusion with an "AND" and an "OR" operator. The maximum of the 2-D FOM gives the optimal combination of the two thresholds for the corresponding operator. Calculating the FOM for different values of α and β and different ground-truthed test images showed that:

- the maximum of the FOMs coincides with a point of the "optimal fusion line" for any choice of α and β
- when varying α and β, the maximum of the 2D FOM moves along the optimal fusion line
- the position of the maximum is the same for the AND and OR operators
- for fixed α and β the position of the maximum of the 2D FOM is different for different test images

These observations confirm that the linear regions in fig. 1 and fig. 2 indeed correspond to an optimal range of threshold combinations for fusion.

4 RESULTS AND DISCUSSION

The two edge detectors were applied using 11×51 windows and sub-sampling on a fixed grid such that $\rho < 0.1$ between neighbouring grid points. The maximum of the results for 8 orientations of the scanning windows was taken. Figs. 3 and 4 show the results of the two multi-variate detectors.

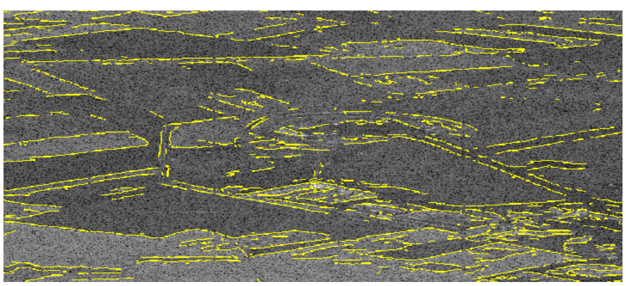

Figure 3: *Results of Levene test-based detector*

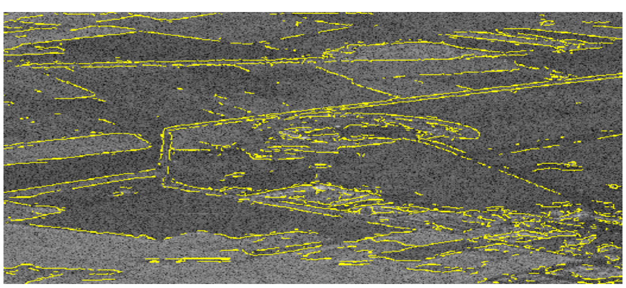

Figure 4: *Results of the Hotellings test-based detector*

To apply the fusion methods, the results of the two detectors are first rescaled in a "working region". The working region is a part of the linear region of optimal fusion in the 2D space of detector thresholds. On the rescaled image different information fusion operators [12] were applied. The non-associative sum operator [12] gave the best results (fig. 5). Results after fusion are globally better than before fusion but some linear features detected by one detector are lost after fusion. A possible improvement can be to detect linear features in each detector's response and restore them after the proposed fusion in a subsequent object-level fusion step.

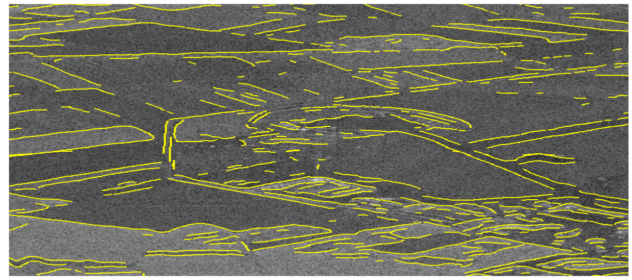

Figure 5: *Fused results using a non-associative sum*

5 REFERENCES

[1] R. Touzi, A. Lopes, and P. Bousquet. A statistical and geometrical edge detector for sar images. *IEEE-GRS*, 26(6):764–773, November 1988.

[2] V.S. Frost, K.S. Shanmugan, and J.C. Holtzman. Edge detection for sar and other noisy images. In *Proc. IGARSS (Munich)*, pages 4.1–4.9, 1982.

[3] F. Tupin. *Reconnaissance des Formes et Analyse de Scènes en Imagerie Radar à Ouverture Synthétique*. PhD thesis, ENST, Paris, September 1997.

[4] R. Fjörtoft. *Segmentation d'images radar par detection de contour*. PhD thesis, Institut National Polytechnique de Toulouse, Toulouse, March 1999.

[5] D. Borghys, C. Perneel, and M. Acheroy. Contour detection in high-resolution polarimetric sar images. In *SPIE Conference on SAR Image Analysis,Modelling and Techniques III; Barcelona*, sept 2000.

[6] D. Borghys. *Interpretation and Registration of High-Resolution Polarimetric SAR Images*. PhD thesis, ENST, Dépt. TSI, Paris, Nov 2001.

[7] B.F.J Manly, editor. *Multivariate Statistical Methods*. Chapman and Hall, 1995.

[8] T.W. Anderson. *Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, 1958.

[9] S. Siegel. *Nonparametric Statistics for the behavioral Sciences*. Mc Graw-Hill Inc., New-York, 1956.

[10] J. Cohen. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.*, 20:27–46, 1960.

[11] L. Sévigny et al. *Autonomous Long Range IR Target Acquistion (NATO Unclassified). Final report of AC243/P3/RSG9, Project 5*. NATO, Brussels, 1995.

[12] I. Bloch. Information combination operators for data fusion: A comparative review with classification. *IEEE-SMC(Part A)*, 26(1):52–67, January 1996.