The Stability of UWB Low-Frequency SAR Images

Renato Machado, Viet Thuy Vu, Mats I. Pettersson, Patrik Dammert, and Hans Hellsten

Abstract—This letter presents an analysis of prefiltered clutter ultrawideband (UWB) very high frequency synthetic aperture radar (SAR) images. The image data are reorganized into subvectors based on the observation of the image-pair magnitude samples. Based on this approach, we present a statistical description of the SAR clutter obtained by the subtraction between two real SAR images. The statistical analysis based on bivariate distribution data organized into different intervals of magnitude can be an important tool to further understand the properties of the backscattered signal for low-frequency SAR images. In this letter, it is found that, for “good” image pairs, the subtracted image has Gaussian distributed clutter backscattering and that the noise mainly consists of the thermal noise and, therefore, speckle noise does not have to be considered. This is a consequence of the stable backscattering for a UWB low-frequency SAR system.

Index Terms—Bivariate distribution, CARABAS-II, clutter, statistical modeling, very high frequency synthetic aperture radar (VHF SAR).

I. INTRODUCTION

It is well known that low-frequency synthetic aperture radar (SAR), operating on the very high frequency (VHF) and ultrahigh frequency (UHF) band, is very efficient for change detection (CD). This capability of detecting changes is mainly related to high relative bandwidth and low frequency which give the long-time stability, precise georeferenced measurements, and small influences of weather [1]. Thus, by considering the same flight geometry, it is possible to acquire multiple images of a certain ground area at different times and still have high correlated images [1].

The detection of targets that appear or disappear between data acquisitions is a method that has been pioneered and well established during the last 16 years in Sweden. The algorithm must allow some small changes in the forest while highlighting the changes of interest (e.g., human activities, trucks, small huts, deforestation, etc.) [2], [3]. The research in this area has been largely stimulated by large-scale experiments using the Swedish airborne SAR system with the low-frequency wavelength-resolution CARABAS [1], [2], [4], [5].

An important observation that we must highlight here is that the ultrawideband (UWB) SAR has a different speckle process than the traditional SAR. In the traditional SAR, there are many scatters in the resolution cell; however, in the UWB SAR, there might only be one scatter in the resolution cell if the resolution is high. This may be illustrated very simply as follows: Backscattering in radar occurs basically on scatters that have the size of the wavelength or bigger. For UWB SAR systems like CARABAS, the wavelength is larger than the resolution cell. Specifically, the resolution of CARABAS is 2.5 m while the wavelength (at center frequency) is 5.5 m [3]. Hence, there is a single scatter per resolution cell, and small objects are weakly scattered, which means that it will not be sensitive to forest floor roughness, forest canopy, shrubs, leaves, and smaller tree branches, and speckle noise is therefore very low [6]. Additionally, low-frequency UWB SAR adds even more stability to the images, and for the CARABAS system, the scatters of size 5 m or bigger do not change between illuminations even if the separation in time is hours or days [3].

Along the last years, CD algorithms have been designed based on testing statistics, such as amplitude ratio and generalized likelihood ratio tests, which consider a pair of images by using the one-look data statistics [2], i.e., the probability density functions (pdfs) are based on the backscattering information provided by each image, where several different approaches for those amplitude clutter distributions have been proposed, such as Rayleigh, K-distribution, Weibull, Nakagami–Rice, Fisher, and lognormal [7]–[9].

In this letter, we present a statistical analysis of UWB VHF SAR experimental data. The data are reorganized into subvectors based on the observation of the image-pair magnitude samples. The results provide interesting insights about the change statistics in UWB low-frequency SAR images. The important finding is that, if the statistics are “stable” for the subvectors (for the different image-pair magnitude ranges), we understand that the noise is mainly caused by thermal noise and not scattering caused by speckle processes between illuminations. Otherwise, if the statistics are not “stable,” then it is most likely that the noise has contribution from the common speckle process, originally from other sources, such as clutter decorrelation, flight track, filtering process, and equalization.

The rest of this letter is organized as follows. Section II describes the data considered in this letter. Section III addresses the statistical analysis when a pair of images undergoes a preprocessing. Section IV gives the numerical results, which are used to feed the discussion presented at the end of this section. Finally, Section V presents some conclusions.

II. DATA DESCRIPTION

The data considered in this letter have been provided by the Swedish Defence Research Agency (FOI) [10] but can also be found in [11]. The raw radar data provided by the CARABAS-II
system were recorded to onboard hard drives, while all signal processing such as image formation, radiometric calibration, geocoding, and others was performed offline. Also, it is worth mentioning that the images used in this letter have been equalized. More information about the system parameters used by airborne CARABAS-II can be found in [3], [4], and [10].

For this letter, we have used SAR images where the phase information has been removed by taking the absolute value of the data. Each image has been georeferenced to the Swedish reference system RR92 [12] and covers an area of size 2 km × 3 km. The images cover the same ground area, but they were formed using the SAR data acquired during different flight passes. The experiment was performed in a restricted military zone, and no targets were added or removed from the scene during the flights.

The pixel size is 1 m × 1 m; thereby, each image has 6 million pixels, which is displayed in an image of 3000 rows by 2000 columns.

In Fig. 1, the bright objects are targets (vehicles), which can be identified by the 5 × 5 array of bright dots in the upper left of each image. Those vehicles were deployed under concealment of the foliage in the forest. The other bright parts of those images are the big structures, such as power lines, boulders, and fences. In this letter, we used six image pairs, i.e., totally 12 SAR images. Since each image has 6 million samples, the statistical analysis is based on 12 × 6 million samples.

### III. Statistical Analysis

The images were obtained from different flights over the same ground area. The images are geometrically and radiometrically equalized in such a way that their similarity is maximized [3].

We perform statistical analysis of preprocessed images which are the result of a subtraction between two SAR images, denoted by Images 1 and 2. The images are represented by two 3000 × 2000 real data matrices, namely, $H_1$ and $H_2$, respectively. The deployments (targets) are removed from the images before the statistical analysis is performed. When a pair of images with different deployments is considered, then the pixels referred to those deployments are removed from both images.

As mentioned earlier, in this letter, we are interested in the stability behavior of the UWB SAR data, not in the changes that can either appear or disappear in between the data acquisitions. However, CD methods that have been developed on the results of this letter can be found in [13].

The matrices $H_i$ are transformed into the vectors $X_i = \text{vec}(H_i)$, where $\text{vec}(\cdot)$ is an operator that transforms an $r \times c$ matrix into a $1 \times r \cdot c$ vector. The elements of vector $X_i$ are denoted by $x_i(1), x_i(2), \ldots, x_i(6 \cdot 10^6)$. A good way to visualize the similarity between two images is through the bivariate distribution, as presented in Fig. 2. Ideally, i.e., if the images were exactly the same, the bivariate distribution would correspond to a 45° inclination line. Clearly, it is impossible to obtain such an ideal bivariate distribution from two different images because of the presence of scattering and sources of noise which changes randomly from one image acquisition to another.

The magnitude vector of an image pair is defined as

$$M_p \triangleq [m_1, \ldots, m_{6 \cdot 10^6}]$$

in which

$$m_n = \sqrt{x_1(n)^2 + x_2(n)^2}$$

where $n = 1, \ldots, 6 \cdot 10^6$. In Fig. 2, we can see that $X_1$ and $X_2$ range from 0 to 4.8, approximately. Consequently, $m_n$ can assume values between 0 and 6.8, roughly.

The statistics presented in this letter are estimated by using the probabilistic approach, where $X_1$ and $X_2$ are assumed to be random variables defined on the same probability space. In order to proceed with this analysis, $X_i, i = 1, 2$, can be written as

$$X_i = S_i + \eta_i$$
where
\[ S_i = S + \eta_i. \]  

(4)

\( S \) is the desired SAR image, \( \eta_i \) denotes the clutter noise, and \( \eta_t \) is the thermal noise at the receiver. At VHF, the thermal noise is very high with noise temperatures 50–60 dB higher than at microwaves [4].

Now, let us define \( X_D \) as the resultant vector of the performed subtraction operation, which is given by
\[ X_D = X_1 - X_2. \]  

(5)

By using (3) and (4), (5) can be rewritten as
\[ X_D = \eta_c + \eta_t \]  

(6)

where \( \eta_c = \eta_{c1} - \eta_{c2} \) and \( \eta_t = \eta_{t1} - \eta_{t2} \).

The variance of \( X_j \) can be defined by [14]
\[ \sigma_j^2 \triangleq E \left[ (X_j - E[X_j])^2 \right] = E \left[ X_j^2 \right] - \mu_j^2 \]  

(7)

where \( E \) denotes the expectation operation and \( \mu_j \) is the mean of \( X_j, j = 1, 2, \) and \( D \).

The mean of \( X_D \) is given by
\[ \mu_D = \mu_1 - \mu_2 \]  

(8)

and assuming that the two noise sources are independent, it follows that
\[ \sigma_D^2 = \sigma_c^2 + \sigma_t^2 \]  

(9)

where \( \sigma_c^2 \) and \( \sigma_t^2 \) are the variance of \( \eta_c \) and \( \eta_t \), respectively.

Additionally, we evaluate the cross-correlation coefficients for the three magnitude ranges. The cross-correlation coefficient presented in this letter is estimated by using the traditional probabilistic approach. Therefore, \( X_1 \) and \( X_2 \) are assumed to be real random variables, and the centered normalized correlation \( \rho \) can be defined as [14]
\[ \rho \triangleq \frac{\text{cov}(X_1, X_2)}{\sigma_1 \sigma_2} = \frac{E \left[ X_1 X_2 - X_2E(X_1) - X_1E(X_2) + E(X_1)E(X_2) \right]}{\sigma_1 \sigma_2} \]  

(10)

where \( \mu_i \) and \( \sigma_i \) are respectively the mean and standard deviation of \( X_i, i = 1, 2, \).

Since \( X_1 \) and \( X_2 \) are discrete random variables, \( \rho \) can be computed as follows:
\[ \hat{\rho} = \frac{1}{N} \sum_{n=1}^{N} x_1(n)x_2(n) - \hat{\mu}_1 \hat{\mu}_2 \]  

(11)

where
\[ \hat{\mu}_i = \frac{1}{N} \sum_{n=1}^{N} x_i(n) \quad \hat{\sigma}_i = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (x_i(n) - \hat{\mu}_i)^2}. \]  

(12)

We use (12) for estimating the mean and standard deviation of \( X_D \).

A. Preprocessed Analysis

Vector \( X_D \), defined in (5), is presented in Fig. 3. As one can see, \( X_D \) ranges from \(-0.55 \) to \( 0.55 \), roughly.

The histogram of \( X_D \) is given in Fig. 4. The histogram presents a symmetrical distribution around zero. Thus, in order to verify which theoretical pdf demonstrates a good fit to the histogram of \( X_D \), we considered some symmetrical pdfs, i.e., Gaussian, Laplace, and hyperbolic secant, all normalized to the area presented by the histogram.

The theoretical pdfs considered in this letter are given by
\[ \text{pdf}_{\text{Gaus}} = \frac{1}{\sigma_D \sqrt{2\pi}} \exp \left\{ -\frac{(X_D - \mu_D)^2}{2\sigma_D^2} \right\} \]  

(13)

\[ \text{pdf}_{\text{Hip}} = \frac{1}{2\sigma_D} \text{sech} \left( \frac{X_D - \mu_D}{\sigma_D} \right) \]  

(14)

\[ \text{pdf}_{\text{Lap}} = \frac{\sqrt{\pi}}{2\sigma_D} \exp \left\{ -\frac{|X_D - \mu_D|}{\sigma_D} \right\} \]  

(15)

where \( \sigma_D \) and \( \mu_D \) are the standard deviation and the median value, respectively.

Based on the mean and variance obtained from the assessed data, we could, by inspection, verify that the Gaussian pdf is the one that fits best for the \( X_D \) data distribution. It is worth to mention that the main region of the distribution which must be taken into account is the tail of the distribution, which represents the most important part of the pdf as a CD algorithm perspective.
As mentioned before, we separated $M_p$ in subvectors as a function of intervals of magnitude. In Fig. 2, the plots represent the sample observation of the bivariate distribution for the three magnitude ranges considered in this letter.

Figs. 5–7 present the corresponding distribution for the $m_n$ range presented in Fig. 2. Again, the Gaussian pdf is the one that fits best for the $X_D$ data histograms associated with the $m_n$ ranges.

### IV. Numerical Results

In Fig. 2, one can identify the three magnitude ranges considered in this letter and note that the amplitude interval with the lowest magnitude has the highest number of samples. We could have divided the bivariate distribution into more than three magnitude ranges; however, we understand that the approach adopted here is good enough to perform the desired investigation. For the statistical analysis, we separate the image pairs into good and bad image pair groups. We considered the following criterion. By using the data with high magnitude for $0.65 < m_n \leq 6.8$, we computed the correlation coefficient between $X_1$ and $X_2$ ($\rho$). For this magnitude, if $\rho$ is greater than 0.60, then the image pair was considered as a good pair, and if $\rho$ is smaller than 0.20, then it was labeled as a bad pair.

Tables I and II present some results obtained from this investigation. The statistical approach adopted for this analysis seems to be adequate since $\mu_D$ calculated from $X_D$ and $\mu_D$ calculated by (8) are approximately the same. We presented the statistics for six image pairs only due to space constraints. The labels used for the image pairs ($\#1, \ldots, \#6$) are not relevant, and they simply were used to organize the data and the analysis results.

Observing the results presented in Table I, one can say that Images 1 and 2 were obtained from two stable flights. Moreover, the standard deviation presents only a little difference for the different ranges of magnitude, which represents a very similar variance for the entire image. Based on these observations, it is possible to say that $S_1$ and $S_2$ are very similar and, therefore,
\( \sigma_D^2 \approx \sigma_t^2 \) is a good assumption. If this statement in regard to \( \sigma_D^2 \) is true, then it means that the main contribution of the change observed between those image pairs comes mainly from the thermal noise. This result is unique because, for all other radar systems, clutter speckle noise is the main contribution to differences between images. It is a consequence of the high resolution in comparison to the operational wavelength, and that backscattering is very stable in time at VHF. Thus, there is no much image equalization filter approach to do in order to improve the quality of \( X_D \), except to increase the transmit power of the system. On the other hand, in Table II, we observe that the standard deviation increases as the \( m_m \) range contains higher values, which means that the signal variance is not the same for the whole range of amplitudes anymore. Thus, we can infer that the background signal has contribution from other sources, i.e., filtering process, flight tracking, etc., which are independent from the thermal noise. Therefore, based on the analysis presented in the previous section, we can say that the variance of the background signal for the bad image pairs can be represented by \( \sigma_D^2 = \sigma_c^2 + \sigma_t^2 \). In that case, it is worth to apply some additional signal processing technique in order to improve the image quality, which is justified by the presence of additional undesirable sources of noise.

V. CONCLUSION

In this letter, we have presented the statistical analysis of a linear combination between two equalized images. By dividing the images into different ranges of magnitude, we observed that, for image pairs with high correlation, i.e., \( \rho > 0.60 \), the variance \( \sigma_D \) over all magnitude ranges is stable, and \( |\mu_D| \approx 0 \), which can be associated to a signal with those statistical characteristics, i.e., the thermal noise. This is a unique result because, normally, speckle noise is the main contributor in high clutter backscatter areas. Thus, it is possible to assume the following approximation: \( \sigma_D^2 \approx \sigma_t^2 \). On the other hand, for the image pairs with low correlation, i.e., \( \rho < 0.20 \), variance \( \sigma_D \) is not stable (or constant) for the whole range of amplitudes anymore, which means that there are some other sources of signal distortion and, therefore, the approximation \( \sigma_D^2 \approx \sigma_t^2 \) is not possible to be applied anymore. In that case, we suggest a more appropriate model for the difference image variance, which is now given by \( \sigma_D^2 = \sigma_c^2 + \sigma_t^2 \). With the right information in hand, we can take a better decision on how to improve the system performance, for example. The idea and approach to analyze statistics based on bivariate distribution data organized into different intervals of magnitude can be an important tool to further understand the properties of the backscattered signal from a scene and how it can vary in an image pair, which can be a valuable premise for CD processing.

ACKNOWLEDGMENT

The authors would like to thank the Swedish Defence Research Agency (FOI) for providing the CARABAS-II SAR data, the Brazilian National Council for Scientific and Technological Development (CNPq), the Swedish-Brazilian Research and Innovation Centre (CISB), and SAAB AB for the financial support.

REFERENCES