ABSTRACT:
In this paper we propose to study the potential of jointly using polarimetric and tomographic SAR data to recognize and localize buildings in complex scenarios. We present extraction methods for both polarimetric and tomographic features. One the one hand, we propose to use the polarimetric bilateral filter that has proved to be a powerful tool to retrieve the polarimetric covariance matrix while reducing speckle and preserving edges. Thus, polarimetric decompositions can be used for physical interpretation. On the other hand, the TomoSNI and TomoSeed algorithms allow to respectively extract interest points and segment geometric primitives in 3D point clouds obtained with tomographic focusing methods. We show how the output of such algorithms could be combined in order to allow the extraction of buildings. We also analyze different issues related to complex scenarios that may impede a correct detection and discuss some possible solutions.

1 INTRODUCTION

Polarimetric SAR systems provide useful information related to the backscattering mechanisms resulting from the interaction between a polarized electromagnetic wave and physical media. Polarimetric images are collected by emitting a series of electromagnetic pulses and receiving backscattered signals through different polarized antennae. Polarimetric decomposition methods (Cloude and Pottier, 1996) allow a physical interpretation of the different scattering mechanisms inside a resolution cell. Thanks to such decompositions it is possible to extract information related to the intrinsic physical and geometrical properties of the imaged media.

SAR tomography is the extension of conventional two-dimensional SAR imaging principle to three dimensions and is achieved by the formation of an additional synthetic aperture in elevation, using a coherent combination of images acquired from parallel flight tracks. The introduction of tomographic SAR allows a direct localization of all scattering contributions in a volume (Reigber and Moreira, 2000). This is performed either by FFT-based beam-forming or spectral methods such as Capon, MUSIC (Guillaso and Reigber, 2005) and the more recent CS (compressive sensing) (Zhu and Bamler, 2012).

In this paper, we use the high resolution MUSIC method (Guillaso and Reigber, 2005) in order to detect scatterers associated to ground and buildings. Indeed, this method enhances the height estimation of objects.

We propose here to jointly use the information given by polarimetry and tomography in order to detect the presence of buildings in the scene. From a polarimetric point of view, it is potentially possible to detect points that belong to a building edge since they are characterized by a strong return. This may be performed by estimating the power (SPAN) of the polarimetric covariance matrix from a polarimetric SAR (PolSAR) image. In order to avoid false alarms due to speckle, the data has to be previously filtered by an adaptive method. Here, we propose to use the recently developed bilateral PolSAR filter (D’Hondt et al., 2013) to extract speckle-free polarimetric features.

However, locating the facade of a building allows only a partial detection of the object. Distinguishing the roof is not possible with polarimetry due to its scattering similarity with bare soil. However, thanks to tomography, it is possible to estimate the heights of the different scatterers present in the resolution cell. Furthermore, since parts of a roof can generally be approximated by a piecewise planar structure, it is possible to group all the pixels belonging to such a structure to obtain a segmentation of the scene into geometric primitives.

To perform this challenging task, we propose to combine recent developments in point extraction, characterization and segmenta-tion of polarimetric and tomographic SAR data (Guillaso et al., 2012, D’Hondt et al., 2012).

Therefore, considering the fact that the facade pixels should be spatially close to the roof ones, using both polarimetric and tomographic information should make it possible to retrieve the entire structure of the building, according to the simple representation facade/roof.

Our information extraction method is composed of three main steps:

- First, a statistical detection of interest points is performed. In fact, low SNR and phase calibration errors affect the performance of spectral methods, resulting in inaccurate localization of scatterers and isolated points with no physical meaning. Spurious points can be successfully discarded by the method of (Guillaso et al., 2012) that uses a robust statistical threshold detecting scatterers with a significant amplitude.
- Then, a segmentation of building heights in terms of geometric primitives is performed to group pixels into piecewise planar patches (D’Hondt et al., 2012). This method is based on region growing, is robust to outliers and allows an automatic adaptation to the local level of noise that is present in each of the structures to detect.
- Finally a polarimetric feature extraction allows to reduce speckle while preserving bright lines due to buildings. Those lines can then be detected by a statistical thresholding of the PolSAR covariance matrix power (SPAN). Besides this feature extraction should also allow to characterize surface and
Figure 1: Optical image representing the test area (top), corresponding SAR intensity image (bottom).

double bounce scattering by the use of PolSAR decomposition methods.

We present some experiments by applying our feature extraction to the POLTOMSAR dataset provided by the DLR over the Oberpfaffenhofen that is composed of 14 fully polarimetric tracks in L-Band. While our feature extraction results are potentially promising, we show that for the case of complex scenarios with interactions with multiple buildings and 45 degree orientation, polarimetry might not be sufficient to detect buildings and propose some possible further investigation.

2 PRESENTATION OF TEST AREA

Fig. 1 shows a manually selected area that will we use in our experiments. It has been extracted from the POLTOMSAR dataset that is composed of 14 fully polarimetric tracks in L-band and has been acquired by the ESAR sensor from DLR over the Oberpfaffenhofen area. The optical picture and the corresponding SAR intensity are shown. This area has been selected because it contains several buildings with complex structures and spatial proximity, that makes it challenging for detection algorithms.

3 EXTRACTION OF POLARIMETRIC FEATURES

3.1 Polarimetric SAR data

Polarimetric SAR systems measure the relation between the transmitted and received electromagnetic wave in two orthogonal polarizations in the form of a scattering matrix (Lee and Potter, 2009)

\[
S = \begin{bmatrix}
S_{hh} & S_{hv} \\
S_{vh} & S_{vv}
\end{bmatrix},
\]  

(1)

where \(h\) and \(v\) denote horizontal and vertical polarization, respectively. The reciprocity assumption in the mono-static case leads to \(S_{hv} = S_{vh}\). For analysing purpose, it is convenient to represent this scattering information in the form of a target vector

\[
k_p = \frac{1}{\sqrt{2}}[S_{hh} + S_{vv}, S_{hh} - S_{vv}, 2S_{hv}]^T,
\]  

(2)

where \(p\) denotes the Pauli basis.

On distributed targets, this information is generally considered in a statistical framework. Under the fully developed speckle hypothesis, the target vectors \(k\) follow a complex circular \(d\)-variates Normal distribution (Goodman, 1963)

\[
p(k) = \frac{1}{\pi^d|\Sigma|} \exp \left(-k^\dagger \Sigma^{-1}k\right),
\]  

(3)

where \(|.\|\) is the determinant of a matrix and \(\dagger\) represents the conjugate transpose of a complex vector. This distribution is characterized by its second order moment \(\Sigma = E[kkk^\dagger]\) that contains information about power and relative phase between the polarimetric channels. When the scattering vector is expressed in the Pauli basis, the matrix \(\Sigma = T\) is usually called coherency matrix.

Unfortunately, \(\Sigma\) has to be estimated from the data over several independent samples, resulting in a loss of spatial resolution. The sample covariance is expressed as

\[
\Sigma = \frac{1}{L} \sum_{i=1}^{L} k_i k_i^\dagger,
\]  

(4)

and is obtained by an operation called multi-looking that consists in taking the average of contiguous pixels.

Under this Gaussian hypothesis, the multi-look covariance matrix follows a complex Wishart density (Goodman, 1963). The estimation accuracy can be improved by increasing the number of samples. Unfortunately, increasing this number is only possible on homogeneous areas where pixels arise from identical scattering mechanisms.

3.2 POLSAR Bilateral Filter

It has recently been shown (D’Hondt et al., 2013) that the PolSAR bilateral filter was a useful tool for the estimation of the covariance matrix from noisy data. In fact, this filter allows a high amount of speckle filtering in homogeneous areas and a good restoration of edges and deterministic scatterers while preserving the polarimetric information. The estimated covariance at location \(x_k\) is computed by performing a weighted sum of the sample covariance matrices at each position \(x_i\) inside a sliding window:

\[
\Sigma(x_k) = \sum_{x_i \in W} w_i \Sigma(x_i).
\]  

(5)

The weights \(w_i\) are defined by a similarity measure based on a matrix distance \(d(\ldots)\)

\[
w_i(x_k) = \frac{f_s(||x_i - x_o||^2)}{\sum_{x_i \in W} f_s(||x_i - x_o||^2)} f_r[d(\Sigma(x_k), \Sigma(x_o))],
\]  

(6)

where the function \(f_s\) and \(f_r\) have a Gaussian shape.

Since the covariance matrices lie on the manifold of Hermitian positive definite (HPD) matrices, recent studies showed it was natural to use appropriate distances (Pennec et al., 2006, Barbaresco, 2009). An affine invariant metric has been proposed as a replacement for the Euclidean metric to deal with HPD matrices. The corresponding distance between two matrices \(\Sigma_1\) and \(\Sigma_2\) is (Barbaresco, 2009)

\[
d_{\alpha}(\Sigma_1, \Sigma_2) = ||\log(\Sigma_1^{-\frac{1}{2}} \Sigma_2 \Sigma_1^{-\frac{1}{2}})||_F,
\]  

(7)

where \(\log(\cdot)\) is the matrix logarithm and \(||.||_F\) the Frobenius norm. This distance corresponds to the length of the geodesic (i.e. the shortest path) between two points on the manifold and is invariant by affine transformation.
3.3 Polarimetric analysis

Fig. 2 shows the Pauli RGB representation of the original polarimetric data and the results of applying the PolSAR bilateral filter. It may be observed that while allowing a strong reduction of speckle, the filter captures well the spatial properties of the scene allowing a significant visual enhancement. It may also be noticed that all the bright parts due to building facades are well preserved. In order to further analyze the polarimetric information, we have applied the widely used $H/\alpha$ or Entropy/Alpha decomposition to the filtered image. The entropy parameter is bounded between 0 and 1 and describes the random behavior of a target: if $H = 0$ the target is considered as purely deterministic and only one scattering mechanism is present inside the resolution cell. On the contrary, $H = 1$ means that the target is totally random and that 3 scattering mechanisms equally contribute to the pixel. This is generally the case with volume scattering due to vegetation. The $\alpha$ parameter characterizes the type of dominant scattering mechanism and has a range from 0 (tri/hedral or surface) to $\pi/2$ radians (dihedral or double bounce). Fig. 3 show images of those parameters over the selected area. Finally, Fig. 4 show the SPAN image (the sum of the diagonal elements of the covariance matrix) extracted from the filtered image and the extracted bright lines that are produced by the strong returns from building facades. The extraction of such lines is performed in two steps. First, following the observation of (Lee et al., 2009) that the strong intensities produced by deterministic scatterers correspond generally to a very small proportion of the pixels (typically the strong returns from building facades). The extraction of such lines is performed in two steps. First, following the observation of (Lee et al., 2009) that the strong intensities produced by deterministic scatterers correspond generally to a very small proportion of the pixels (typically 2 – 3%) we set a threshold corresponding to the 97-th percentile of the intensity of the image. The pixels above the threshold belong to the bright structures. In order to keep only long lines, we then perform a connected component labelling and remove the segments that have less than 30 pixels. It can be seen that this allows to extract the bright line corresponding to the biggest building but it is not clear how the other extracted lines are related to the other smaller buildings. This is due to the fact that the line/roof model is only valid for buildings with a very simple structure.

4 EXTRACTION OF TOMOGRAPHIC FEATURES

4.1 Tomographic focusing by SP-MUSIC

SAR tomography consists in focusing several SAR images in the third dimension, in order to image volumetric areas, such as forests or cities. This means to form a synthetic aperture along the direction perpendicular to azimuth and to the radar line of sight. The geometry of a tomographic data acquisition uses typically $K$ parallel tracks non uniformly spaced, which observe the same scene. From the $K$ images, 3D profiles might be extracted. This makes it possible to detect targets under covered volume or to generate 3D representation of the scene under study.

A tomographic data acquisition system is constituted by $K$ sensors, or interferometric paths. The signals $x_k$ received by each sensor $k$ provided by $D$ scatterers localized at height $\{z_d\}_{d=1}^D$ are arranged in the $K \times 1$ vector $x$:

$$ x = A s + n $$

where $s$ represents the backscattered power of the $D$ scatterers and $n$ denotes a vector formed by scalar $n_k$ representing a circular Gaussian white noise. The $A$ matrix, with dimension $K \times D$, contains the phase response due to the sensor geometry only. This matrix is made up various vectors $a(z_d)$ representing the steering vector, which corresponds to the $d$-th scatterer:

$$ a(z_d) = [e^{i\phi_1(z_d)} \cdots e^{i\phi_K(z_d)}]^T $$

with $\phi_k(z_d)$ denoting the phase caused by the distance between a scatterer at height $z_d$ and antenna $k$.

To focus such tomographic SAR data, the range of valid heights is scanned using the steering vector $a(z)$ of (9). The power, estimated by the MUSIC method, is given by:

$$ \hat{P}_M(z) = \frac{1}{a^H(\hat{E}_N \hat{E}_N^H)a} $$

where $\hat{E}_N$ represents the noise subspace obtained after an eigendecomposition of the observed covariance matrix, $\hat{R} = \hat{E} \hat{E}^H$. 

Figure 2: Pauli RGB representation of the original image (top) and the bilateral filtered one (bottom).

Figure 3: Images of the $H/\alpha$ parameters computed on the filtered image.
The SPAN image is computed. Bright lines are detected by thresholding. Small segments are removed.

Figure 4: Extraction of bright lines corresponding to building corners.

The term pseudo-beamforming is employed here because \( \hat{P}_M(z) \) is only usable to localize the target height but not to measure its backscattering power.

Figure 5: Reflectivity reconstruction using SP-MUSIC-1 (1 scatterer assumed). Color corresponds to the height of the scatterers. Data are represented in a ground range geometry.

Fig. 5 shows the point cloud of our test site. We notice the presence of numerous artifacts, especially points located in the “air”. Despite these artifacts, it is possible to clearly distinguish our 4 main objects: main building with roof structure, meadows, secondary building as well as the fence.

4.2 Selection of points: the TomoSNI algorithm

Artifacts generated during tomographic processing using SP-MUSIC-1 are mainly due to the following factors:

- **Low reflectivity**, observed in the presence of roads, water, shadowed areas.
- **Wrong scatterer number assumption.** Indeed, the SP-MUSIC-1 will fail if we have more than one scatterer per resolution cell.
- **Orbit path estimation.** Phases have to be calibrated during the processing. Errors will arise when focusing far from the reference point.

\[
TomoSNI(x,y) = \text{median} \left( \frac{\hat{P}_M(x,y)}{\max(\hat{P}_M(x,y))} \right) \tag{11}
\]

where \( \hat{P}_M(x,y) \) is the pseudo-power over height given by (10). We calculate the median value of each normalized pseudo-spectrum and we select pixels where the TomoSNI < \( T \) is less than a threshold \( T \). The threshold is defined as follow:

\[
T = \text{median}(\text{TomoSNI}) + \text{MAD}(\text{TomoSNI}) \tag{12}
\]

where MAD stands for median absolute deviation. Up to now, this approach is only valid when one scatterer is present in the resolution cell.

The proposed TomoSNI algorithm is illustrated in Fig. 7. The amplitude of the peak in \( p_2 \) is comparable to the noise level, while the amplitude of the peak in \( p_1 \) is clearly higher. Fig. 8 shows the new point cloud distribution. Note that more than 95% of artifacts have been removed.

4.3 Primitive extraction: the TomoSeed algorithm

In order to retrieve high level information of tomograms, it is useful to group the points that belong to the same geometrical primitive. In the case of buildings, it is reasonable to assume that...
Figure 7: Normalized pseudo-spectrum for $p_1$ and $p_2$. The red line indicates the median value for each spectrum and the blue line represents the threshold.

Figure 8: Reflectivity reconstruction after using the TomoSNI approach. Color corresponds to the height of the scatterers. Data are represented in ground range geometry.

3D planes will be able to capture most of the present information. Moreover, even in the case of a non-planar structure, it is always possible to provide a piecewise planar approximation. The local curvature is not too high. Grouping the points according to a parametric model is not only useful for detection, but it also naturally reduces noise present in the elevations due to phase errors and low coherence. Therefore, our method can also be used for reconstruction.

The TomoSeed algorithm (D’Hondt et al., 2012) takes advantage of the spatial connectivity of points in range/azimuth. In fact, tomograms retrieved with the MUSIC approach provide a high degree of the spatial connectivity of points in range/azimuth. In fact, the tomograms retrieved with the MUSIC approach provide a high degree of the spatial connectivity of points in range/azimuth.

Our algorithm is composed of two main steps. The first one is called seed selection and collects all the samples contained in a 3D sphere of radius $R$ around each point of the data set. For each potential initial ”seed” patch, a 3D plane is fitted by mean square minimization. The variance $\sigma^2$ obtained from the residual gives a measure of non-planarity, which is expected to be high if the patch is located over several planes. Once all the neighborhoods have been examined, the retained starting seed is the one that correspond to the lowest $\sigma^2$.

The second step called patch extraction progressively adds points to the initial patch by examination of its spatial neighbors, in the spirit of region growing. A point $p_i$ is accepted if the condition $d(z_i, P(x_i, y_i)) < 3.5\sigma$ is satisfied, where $d(,)$ is the distance between the point $z_i$ and the plane. Here, the model is a simple plane described by the function $P(x, y) = ax + by + c$ and the elevation noise is assumed to be white Gaussian for the sake of simplicity. However, the approach could be extended to other models. As new points are added, the plane parameters and the variance are updated. In practice, to keep a low computational cost, these quantities are updated only if the area of the region doubles.

**5 DISCUSSION**

The previously introduced methods allow to extract features from two different types of data that carry complementary information. On the one hand, the tomographic point clouds provide...
In this paper, we have studied the complementarity of polarimetry and tomography in order to retrieve buildings from multi-channel SAR data. We have presented some recently developed methods that can be used to extract polarimetric and tomographic features in order to feed a detection algorithm. The PolSAR bilateral filter allows a strong de-speckling of the image, allowing to apply polarimetric decompositions and detect bright structures. The TomoSNI and TomoSeed algorithms allow the detection of points of interest from tomographic point clouds as well as their geometric grouping and de-noising. We have shown that although these features contain a significant amount of information, they may not be sufficient in the case of complex scenarios. Complexity of building structures and ambiguity in the polarimetric parameters show that more information should be extracted from the scene in order to allow an automatic detection.

6 CONCLUSION

In this paper, we have studied the complementarity of polarimetry and tomography in order to retrieve buildings from multi-channel SAR data. We have presented some recently developed methods that can be used to extract polarimetric and tomographic features in order to feed a detection algorithm. The PolSAR bilateral filter allows a strong de-speckling of the image, allowing to apply polarimetric decompositions and detect bright structures. The TomoSNI and TomoSeed algorithms allow the detection of points of interest from tomographic point clouds as well as their geometric grouping and de-noising. We have shown that although these features contain a significant amount of information, they may not be sufficient in the case of complex scenarios. Complexity of building structures and ambiguity in the polarimetric parameters show that more information should be extracted from the scene in order to allow an automatic detection.