Polarimetric Synthetic-Aperture-Radar Change-Type Classification with a Hyperparameter-Free Open-Set Classifier

Mark W. Koch, R. Derek West, Robert Riley, and Tu-Thach Quach
Sandia National Laboratories
Albuquerque, NM 87185-1163
mwkoch@sandia.gov
SAND2018-3569 C

Abstract

Synthetic aperture radar (SAR) is a remote sensing technology that can truly operate 24/7. It's an all-weather system that can operate at any time except in the most extreme conditions. Coherent change detection (CCD) in SAR can identify minute changes such as vehicle tracks that occur between images taken at different times. From polarimetric SAR capabilities, researchers have developed decompositions that allow one to automatically classify the scattering type in a single polarimetric SAR (PolSAR) image set. We extend that work to CCD in PolSAR images to identify the type change. Such as change caused by no return regions, trees, or ground. This work could then be used as a preprocessor for algorithms to automatically detect tracks.

1. Introduction

1.1. Synthetic aperture radar and change detection

Synthetic aperture radar (SAR) [4] is a remote sensing technology that provides its own illumination. Thus SAR is an all-weather system that can image at any time except in the most extreme conditions. It can operate either day or night and has a long standoff. SAR combines multiple results from different viewing angles to create a high-resolution image of an area. For example, Figure 1a shows a SAR image of a golf course containing roads, trees, sand traps, grass, and buildings.

For the area it is illuminating, SAR is a coherent imager that measures both phase and magnitude of the return. Using two registered images taken at different times, one can use coherent change detection (CCD) [4] to detect minute changes from one collection to the next, such as tracks left by a vehicle driving through the scene. Figure 1b shows the corresponding CCD image with examples of different types of change. Detection of tracks is useful for surveillance and search and rescue applications [18]. However, automatic detection of vehicle tracks in SAR CCD is difficult due to various sources of low coherence other than the vehicle track change we wish to detect, such as ground surface change due to weather effects and vegetation, registration errors, and radar shadows. Figure 1b shows not only low coherence for the vehicle tracks but also for trees and shadows.

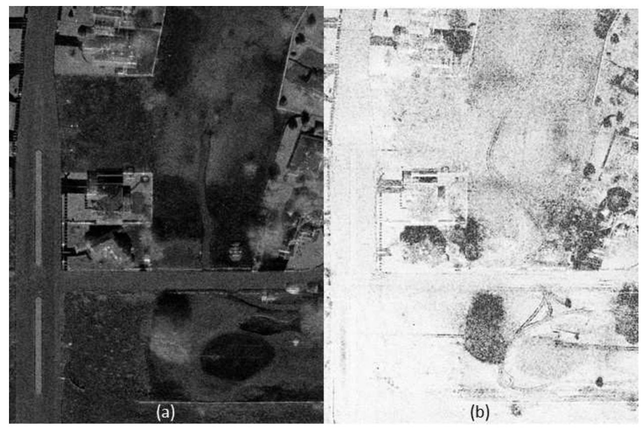


Figure 1: Example SAR and corresponding CCD image. (a) SAR image of golf course. (b) CCD image from two SAR images taken at different times.

1.2. Polarimetric SAR (PolSAR)

The radar cross section (RCS) of the scattering mechanisms in a scene for a selected polarization state, can be conveyed by a calibrated single-polarization SAR image Examples of polarization states include: VV, HH, HV, and VH. Here, XY indicates that a Y oriented EM field was transmitted and an X oriented EM field was received. Figure 2 shows four polarimetric images taken at different transmit and receive orientations.

Using second order statistical characterization of the polarmetric return, multiple decompositions [23] have been developed to extract the responsible scattering mechanism. Because of it's information-theoretic properites, we use the $H/A/\alpha$ polarimetric decomposition [1]. Here H represents the entropy or randomness of the scatter, A the anisotropy or the direction of scatter, and α the scattering mechanism. Figure 3 shows the H/α classification plane. For this classification plane, Cloude and Pottier have identified eight zones that correspond to different SAR scattering mechinisms and one zone (Z_3) that corresponds to an infeasible region indicated by gray area. For example, a low H and middle value of α would be in zone Z_8 and indicates a single dipole scatterer. As H increases to moderate values such as in zone Z_5, this would indicate scattering mechanisms from vegatation, and then even higher values such as in zone Z_2 would indicate scatters from tree canopies. See [1] for a complete description. These zones are not as crisp as indicated by the Figure. Different classes can straddle the boundaries and the boundaries can change with collection geometry and system noise [14].

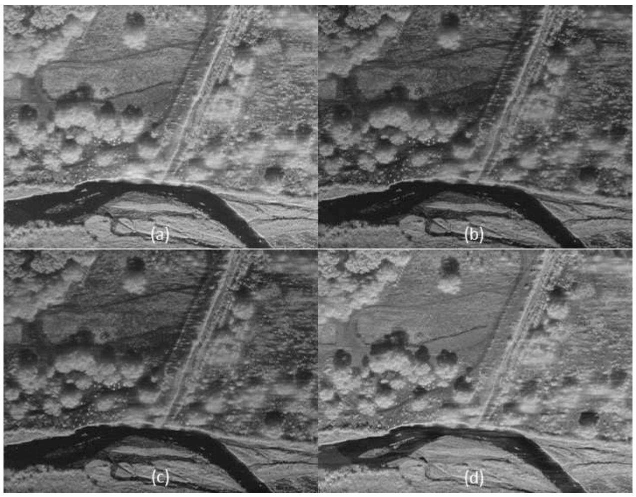


Figure 2. Illustration of (a) an HH polarization image, (b) an HV polarization image, (c) a VH polarization image, and (d) a VV polarization image.

Figure 4 shows the $H/A/\alpha$ decomposition of two passes of a scene. The colorspace of the image is defined in HSV space where the scattering mechanism parameter controls the hue in the image, a mixture of entropy and anisotropy control the saturation, and the quarter-power span image controls the value. The span is the total magnitude of the pixel across the four images with polarimeteric image set.

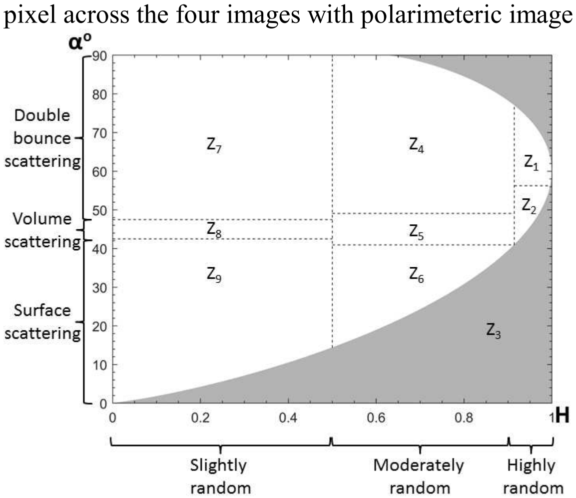


Figure 3. H/α classification plane

2. Track detection using Polarimetric Data

Automatic track segmentation can be view as a two step process. The first step is to identify CCD pixels that could be on a track called *tracklets*. Whereas the second step involves linking these tracklets together to create longer tracks and remove false tracks. In single pol imagery, tracks have been identified using thresholding [16], a radon transform [11], ridge features [18], and MR8 filters [19]. Currently, the most successful tracklet linking method is based on a 6-layer convolutional neural network (CNN) trained to find natural tracks where each succeeding layer has a larger receptive field to link tracks over a larger area [17].

In this paper we concentrate on finding tracklets. Past work, on using polarimetric data, has concentrated on discriminating different terrain types in SAR imagery, but here we extend this work and use polarimetric data to discriminate and classify different types of change and thereby identify tracklets. From Figure 1b, one can see that automatically detecting vehicle tracks is hampered by the fact that low coherence appears also in shadow and tree areas.

In the next section we will discuss the processing of the polarimetric imagery and the extraction of the tracklet feature vector. In Section 4, we discuss the classifier and in the last sections we discuss the data we use for training and testing and the performance of the system.

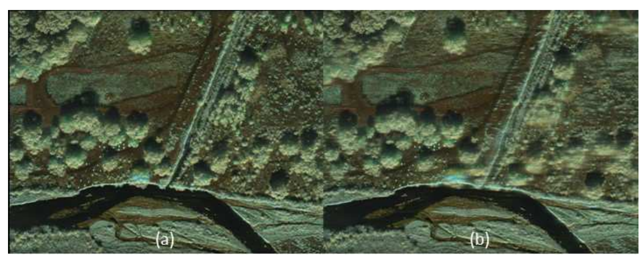


Figure 4. H/A/α images from (a) pass 1 and (b) pass 2. The images are defined in the HSV colorspace where α controls the hue, a mixture of *H* and *A* control the saturation, and the quarter-power span image controls the value.

3. Polarimetric SAR

Measured PolSAR data at each pixel can be written as a scattering matrix:

$$\mathbf{S} = \begin{bmatrix} \tilde{S}_{\mathrm{HH}} & \tilde{S}_{\mathrm{HV}} \\ \tilde{S}_{\mathrm{VH}} & \tilde{S}_{\mathrm{VV}} \end{bmatrix} \tag{1}$$

where $\tilde{S}_{\mathrm{XY}}$ is a complex-valued measurement. Since the observations were performed in a monostatic manner we assume that $\tilde{S}_{\mathrm{HV}} \approx \tilde{S}_{\mathrm{VH}}$.

The total power of scattering observations from all of the polarization channels can also be combined into a *span* magnitude image:

$$\text{Span} = |\widetilde{S}_{\text{HH}}|^2 + |\widetilde{S}_{\text{HV}}|^2 + |\widetilde{S}_{\text{VH}}|^2 + |\widetilde{S}_{\text{VV}}|^2 . \qquad (2)$$

3.1. Pauli feature vector

The objective of a polarimetric decomposition is to map raw observations into different types of scattering mechanisms. One mapping is called the Pauli feature vector, which is constructed by projecting the 2×2 scattering matrix onto the Pauli-spin matrix bases [1]. The 3×1 Pauli feature vector has the following form:

$$\mathbf{k} = \frac{1}{\sqrt{2}}[\widetilde{S}_{\text{HH}} + \widetilde{S}_{\text{VV}}, \widetilde{S}_{\text{HH}} - \widetilde{S}_{\text{VV}}, \widetilde{S}_{\text{CX}}]^T . \qquad (3)$$

where $\widetilde{S}_{\text{CX}} = (\widetilde{S}_{\text{HV}} + \widetilde{S}_{\text{VH}})/2$.

For the first entry of the Pauli feature vector, the physical phenomenon is an odd-bounce scattering mechanisms such as trihedral corner reflectors or spheres. The second entry represents even-bounce scattering mechanisms such as horizontally or vertically or rotated diplanes, while the third entry represents even-bounces from dihedrals rotated 45°.

3.2. The $\text{H/A/}\alpha$ polarimetric decomposition

The polarimetric coherency matrix can be formed by computing the spatial average of the outer-product of the Pauli feature vectors:

$$\mathbf{T} = < \mathbf{k}\mathbf{k}^H >_N , \qquad (4)$$

where $<\cdot>$ denotes a spatial ensemble average over a neighborhood of subscript N pixels.

The $\text{H/A/}\alpha$ decomposition utilizes the eigendecomposition of the polarimetric coherency matrix,

$$\mathbf{T} = \mathbf{U}\Lambda\mathbf{U}^{-1} = \sum_{i=1}^{3} \lambda_i \mathbf{u}_i \mathbf{u}_i^H , \qquad (5)$$

where $\mathbf{U}$ is the matrix of eigenvectors and Λ is a diagonal matrix of the corresponding eigenvalues. The $\text{H/A/}\alpha$ decomposition computes the following quantities from the set of eigenvectors and ordered eigenvalues ($\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$):

$$H = -\sum_{i=1}^{3} P_i \log_3 P_i , \qquad 0 \leq H \leq 1, \qquad (6)$$

$$A = (\lambda_2 - \lambda_3)/(\lambda_2 + \lambda_3) \qquad 0 \leq A \leq 1, \qquad (7)$$

$$\alpha = \sum_{i=1}^{3} P_i \cos^{-1}\left(|\mathbf{u}_i(1)|\right), \qquad 0^\circ \leq \alpha \leq 90^\circ , \qquad (8)$$

where $P_i = \lambda_i / \sum_{i=1}^{3} \lambda_i , \quad 0 \leq P_i \leq 1$. The description of parameters is as follows:

1. α is the scattering mechanism parameter and indicates the average scattering mechanism
2. H is the entropy parameter and indicates the purity of the scattering mechanism
3. A is the anisotropy parameter and gives the relative significance of the second and third eigenstates.

3.3. Polarimetric coherence

Two complex-valued SAR images with similar observation geometries can be co-registered at the sub-pixel level, and then *interfered* to produce a corresponding complex-valued coherence map. For observations seperated in time, the magnitude of the resulting coherence estimate conveys the degree to which scattering observations in the scene have maintained coherence.

Over the past couple of decades, researchers have developed various methodologies for producing coherence estimates from PolSAR image sets [1][24][27]. One method is called *optimum coherence* (OC). Here, the resulting coherence estimation is maximized by determining the appropriate weighting vectors informed by underlying scattering processes, and provides improved quality interferograms over what can be produced with just a single-polarization observation [2][9].

The fully-polarimetric coherence estimation can be computed from the equation:

$$\widetilde{\gamma}_{\text{Opt}} = \frac{\mathbf{w}_1^H \mathbf{\Omega}_{12} \mathbf{w}_2}{\sqrt{(\mathbf{w}_1^H \mathbf{T}_{11} \mathbf{w}_1)(\mathbf{w}_2^H \mathbf{T}_{22} \mathbf{w}_2)}} , \qquad (9)$$

where $\mathbf{\Omega}_{12} = < \mathbf{k}_1 \mathbf{k}_2^H >_N$, $\mathbf{T}_{11} = < \mathbf{k}_1 \mathbf{k}_1^H >_N$, $\mathbf{T}_{22} = < \mathbf{k}_2 \mathbf{k}_2^H >_N$, and the $\mathbf{w}_i$ vectors are complex-valued weighting vectors. The weighting vectors create the maximum indicated coherence that the observations can support. These weighting vectors utilize the same bases as the input Pauli feature vectors, and hence corresponding to equivalent scattering mechanism definitions.

To estimate the $\mathbf{w}_i$ vectors, equation (9) is recast as the following unconstrained Lagrangian maximization problem:

$$\begin{aligned} L(\mathbf{w}_1, \mathbf{w}_2) = \mathbf{w}_1^H \mathbf{\Omega}_{12} \mathbf{w}_2 + \mu_1\left(\mathbf{w}_1^H \mathbf{T}_{11} \mathbf{w}_1 - C_1\right) + \\ + \mu_2\left(\mathbf{w}_2^H \mathbf{T}_{22} \mathbf{w}_2 - C_2\right), \end{aligned} \qquad (10)$$

where the objective is to maximize the first term, subject to the constraints given in the next two terms.

Computing gradients, with respect to the weighting vectors, gives the following coupled eigenvalue problem:

$$\nu = \mu_1 \mu_2^* \qquad (11)$$

$$\mathbf{T}_{11}^{-1} \mathbf{\Omega}_{12} \mathbf{T}_{22}^{-1} \mathbf{\Omega}_{12}^H \mathbf{w}_1 = \nu\, \mathbf{w}_1 \qquad (12)$$

$$\mathbf{T}_{22}^{-1} \mathbf{\Omega}_{12}^H \mathbf{T}_{11}^{-1} \mathbf{\Omega}_{12} \mathbf{w}_2 = \nu\, \mathbf{w}_2 , \qquad (13)$$

where the vectors $\mathbf{w}_1$ and $\mathbf{w}_2$ are the eigenvectors of $\mathbf{T}_{11}^{-1} \mathbf{\Omega}_{12} \mathbf{T}_{22}^{-1} \mathbf{\Omega}_{12}^H$ and $\mathbf{T}_{22}^{-1} \mathbf{\Omega}_{12}^H \mathbf{T}_{11}^{-1} \mathbf{\Omega}_{12}$, respectively. Since the matrices have dimension 3×3, there are a total of three ordered eigenvalues ($0\leq|\nu_3|\leq|\nu_2|\leq|\nu_1|\leq1$), and two corresponding sets of 3×1 eigenvector triplets: ($\{\mathbf{w}_{1,1}\,\mathbf{w}_{1,2}\,\mathbf{w}_{1,3}\}$ and $\{\mathbf{w}_{2,1}\,\mathbf{w}_{2,2}\,\mathbf{w}_{2,3}\}$). Through algebraic operations, it can be shown that the maximum supported observation coherence estimation magnitude can be calculated via $|\tilde{\gamma}_{\mathrm{Opt}}|=\sqrt{|\nu_1|}$.

These weighting vectors not only get us the OC maps, but in fact they contain information that enables categorization of underlying mechanisms for observed variations in temporal coherence.

3.4. Change detection feature vector

The change discrimination feature vector uses the $\mathrm{H/A/}\alpha$ decomposition parameters, as well as the optimum coherence values and the steering vectors from the OC algorithm [25]. Here, H, A, α values can be produced for each of the six weighting vectors and the two original image sets. The $\mathrm{H/A/}\alpha$ parameters can be stacked into a three-element vector as follows,

$$\mathbf{d}_X^T=[H \quad A \quad \alpha], \tag{14}$$

where X represents the data processed through the $\mathrm{H/A/}\alpha$ decomposition (*i.e.* $\mathbf{k}_1$, $\mathbf{w}_{1,1}$, etc.). A 29-dimensional feature vector can be formed by vertically concatenating the $\mathrm{H/A/}\alpha$ vectors computed from the original image sets and the weighting vectors from the OC algorithm, along with the optimum coherence values and the two square-root span values computed from the two image sets. To be explicit, the feature vector has the form,

$$\mathbf{d}^T=\left[\mathbf{d}_{\mathbf{k}_1}^T\,\mathbf{d}_{\mathbf{k}_2}^T\,\mathbf{d}_{\mathbf{w}_{1,1}}^T\,\mathbf{d}_{\mathbf{w}_{1,2}}^T\,\mathbf{d}_{\mathbf{w}_{1,3}}^T\,\mathbf{d}_{\mathbf{w}_{2,1}}^T\,\mathbf{d}_{\mathbf{w}_{2,2}}^T\,\mathbf{d}_{\mathbf{w}_{2,3}}^T\,\mathbf{e}_{\gamma,\mathbf{k}}^T\right] \tag{15}$$

where

$$\mathbf{e}_{\gamma,\mathbf{k}}^T=\left[|\tilde{\gamma}_{Opt,1}| \quad |\tilde{\gamma}_{Opt,2}| \quad |\tilde{\gamma}_{Opt,3}|\,\|\mathbf{k}_1\|_2\,\|\mathbf{k}_2\|_2\right] \tag{16}$$

Figure 5 illustrates the processing steps to form the 29-dimensional feature vectors.

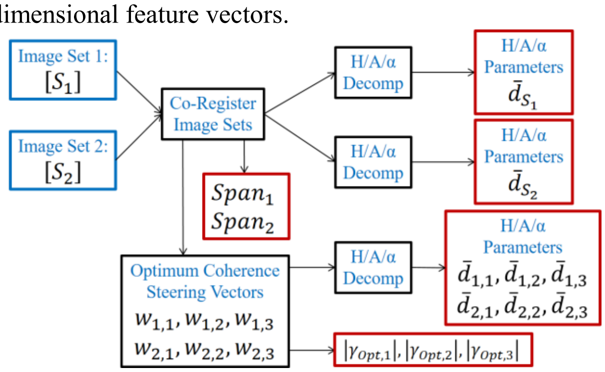


Figure 5. Illustration of the processing steps required to form the 29-dimensional feature vectors used for change discrimination

4. Hyperparameter-free open-set classifier

There are two main components to the change discrimination framework: 1) determining a prototype feature vector for a given type of change, and 2) a distance metric that can discriminate between the prototype and feature vectors. We use a data driven approach and derive the prototype feature vectors from hand annotated training data derived from observations of change types of interest.

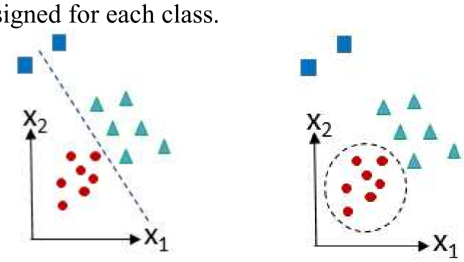
The behavior of the discriminating function can be stated as:

$$\gamma_{\mathrm{ChangeD}}=f(\mathbf{d},\mathbf{d}_0)=\begin{cases}\gamma, & \text{if } \|\mathbf{d}-\mathbf{d}_0\|<\varepsilon\\ 1, & \text{otherwise},\end{cases} \tag{17}$$

where γ is a selected coherence map and $\mathbf{d}_0$ is a feature vector prototype that defines a particular type of change and $\|\cdot\|<\varepsilon$ is some measure of "closeness" to the feature vector $\mathbf{d}$. This is known as a goodness-of-fit classifier or an open-set classifer (for more details see next section). We also use a hyperparameter free classifier. This prevents overtraining, because the designer has no parameters to "tweak" in order to improve the classifier performance.

4.1. Open-set classifier

One can view a feature vector as a point in $\Re^D$ where D is the dimension of the signature in the feature space. Figure 6 shows an example of a 2D features space (x_1,x_2). The red circles represent signatures from class 1 and the green triangles represent signatures from class 2. The blue squares are the unknown class. Figure 6a shows a linear decision boundary for the discriminator, but similar problems occur with nonlinear ones. This linear decision boundary separates the two classes and performs great if we are in a constrained environment with a closed set of classes [21]. But if there is an unknown class like the blue squares then the classifier will make errors and assign it to class 1 or 2. The solution is to use what was called *one-class* [7] [12][13] or *goodness-of-fit* [5] classifiers before around 2010, but now are referred to as open set classifiers [21]. These classifiers have closed decision boundary and allow the classifier to reject the unknown class as shown in Figure 6b. To handle multiple classes an open set classifier is designed for each class.



(a) (b)

Figure 6. Classifier examples. (a) Closed set. (b) Open set.

4.2. Feature quantization

In this paper, we use feature quantization for creating a compressed feature vector and giving a hyperparameter-free classifier. Here, we create a feature vector that captures the essential information and clusters similar classes with high probability, but still discriminates between the various classes. Let d_i and d_j represent components of the feature vector $\boldsymbol{d}$ in equation (15). Let $\boldsymbol{y}$ represent the compressed feature vector with components y_k. Then

$$y_k = \begin{cases} 2 & \text{if } d_i > d_j \\ 1 & \text{otherwise} \end{cases}, \text{ for } i > j . \tag{18}$$

Even though $\boldsymbol{y}$ has 406 components the components are binary so that 32 components can be packed into a small number of 32 bit words. Figure 7 shows an example of a quantized feature vector where yellow represents a 2 and blue represents a 1. Only the upper half of the comparison matrix is shown, since both halves are redundant.

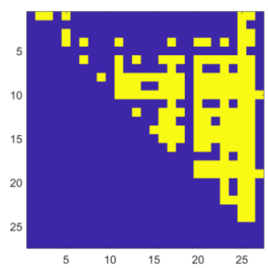
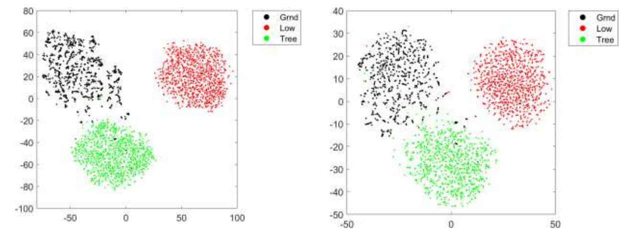


Figure 7. Quantized feature vector where yellow represents a 2 and blue represents a 1.

We use t-distributed stochastic neighbor embedding (t-SNE) [10] to visualize the raw and quantized feature space. The t-SNE approach is an unsupervised algorithm for dimensionality reduction and for visualizing high-dimensional data on a 2 or 3-dimensional manifold. It works by embedding the high-dimensional points in a low dimension such that the similarities between points are respected. Here, nearby points in the high-dimensional space correspond to nearby embedded low-dimensional points and distant points in high-dimensional space correspond to distant embedded low-dimensional points. Figure 8a shows the t-SNE results for the raw feature vector with the points color coded by class. Figure 8b shows the t-SNE plot for the quantized feature space. Both plots show similar separation and groupings of the classes, so we conclude that we do not lose very much information in going to the quantized representation.

4.3. Multinomial pattern matching

An ideal open-set classifier for a quantized feature vector is multinomial pattern matching (MPM) [22][6]. The MPM test statistic uses a multinomial indexing transform $\Gamma : \Re \to \mathrm{I}$, $\mathrm{I} = \{1,2,\ldots,Q\}$, that maps the feature vector into a discrete index representing group membership. In our case $Q = 2$ and uses (18) for the mapping Γ, but the following mathematics applies for any $Q \geq 2$.



(a) (b)

Figure 8. The t-SNE visualization of the feature space. (a) Raw feature vector. (b) Quantized feature vector.

Using the training data from each class, we estimate a class prototype ($\mathbf{d}_0$ in equation (17)) of the quantile probabilities $\hat{P}_{bq}$. Here, the probabilities $\hat{P}_{bq}$ represent the observed proportion of changed detection training signatures for which quantized feature component b in the class signature maps to quantile q. Thus, the template M_T is a $B \times Q$ matrix of quantile probabilities $\hat{P}_{bq}$. Equation (19) gives the MPM test statistic for quantized feature vector component y_b:

$$Z_{MPM} = \sum_{k=0}^{B-1} \frac{(1-\hat{P}_{k,\Gamma(y_k)})^2 - \hat{E}_k}{\sqrt{C \times \hat{V}_k}} . \tag{19}$$

Here, the quantities $\hat{E}_k$ and $\hat{V}_k$ represent the estimated expected value and variance of the quadratic penalty $(1-\hat{P}_{k,\Gamma(y_k)})^2$, and C accounts for the correlations between feature vector components. Low Z_{MPM} scores are consistent with a good match to the class. Using the central limit theorem [15], as the number of components in the quantized feature vector $\boldsymbol{y}$ increases, we can show that Z_{MPM} approximates a normal distribution with zero mean and unit variance ($N(0,1)$) conditioned on the target data.

The estimates of the mean and variance of the quadratic penalty $\left(1-\hat{P}_{kq}\right)^2$ are:

$$\hat{E}_k = \sum_{q=0}^{Q-1} \tilde{P}_{kq}(1-\hat{P}_{kq})^2 \tag{20}$$

and

$$\hat{V}_k = \sum_{q=0}^{Q-1} \tilde{P}_{kq}(1-\hat{P}_{kq})^4 - \hat{E}_k^2 \tag{21}$$

Here, $\tilde{P}_{bq}$ estimates the quadratic penalty probabilities for component b and quantile q. To compute $\tilde{P}_{bq}$, we use a Bayes estimator of the form [20]:

$$\tilde{P}_{bq} = \alpha\, p_{bq}^{0} + (1-\alpha)\hat{P}_{bq} \tag{22}$$

where $0 \le \alpha \le 1$ represents a weight, $\hat{P}_{bq}$ represents the maximum likelihood estimation of the probability p_{bq}, and p_{bq}^{0} represents a-priori information about p_{bq} and satisfies the properties of a probability. The Bayes estimator prevents the case of estimating a zero probability for a specific feature component b and quantile q. For the multinomial distribution, the Dirichlet distribution is the conjugate prior [26]. Using a symmetric Dirichlet prior gives $p_{bq}^{0} = 1/Q$, and $\alpha = Q\upsilon/(n+Q\upsilon)$ [20], where n represents the number of training signatures and υ is a single user specified parameter for the Dirichlet distribution. Thus, the Bayes estimation equation (22) for the quantile probabilities becomes [20][22][6]:

$$\tilde{P}_{bq} = (\upsilon + n\hat{P}_{bq})/(n+Q\upsilon)\,. \tag{23}$$

As $n \to \infty$, the Bayes estimate $\tilde{P}_{bq}$ approaches the maximum likelihood estimate $\hat{P}_{bq}$. We estimate $\hat{P}_{bq}$ and υ using the training data and a leave-one-out (LOO) estimation technique. The υ parameter is selected to give the LOO Z_{MPM} scores a zero mean. With an automated approach for selecting υ we have a hyperparameter free classifier.

5. Data selection

In this paper, a data driven approach is taken to estimate the change-discrimination functions for three different types of change by selecting training feature vectors, for each change type, from homogeneous regions within a variety of image sets.

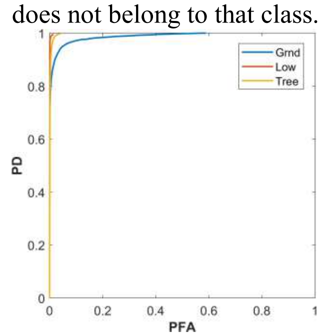
Both training and test feature vectors were collected for different change-types from a variety of coherent, fully-polarimetric image sets; the training data were collected from nine different image sets and the test data were collected from six different image sets, separate from the training image sets. Within each image set, training and test feature vectors were collected for three different change types: tree (*Tree*), low-return (*Low*), and ground (*Grnd*).

6. Results

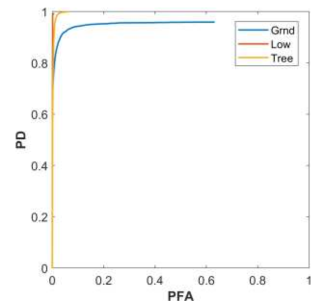
In this section, we show performance results of the training and testing data using receiver operating characteristic (ROC) curves, confusion matrices and then a CCD image color coded by the detected classes.

6.1. ROC curves and confusion matrices

Figure 9 shows the receiver operating characteristic (ROC) curves for the training and testing data. Here we plot the probability of false alarm (PFA) vs. probability of detection (PD). Each point on the ROC is determined by a threshold τ on the score Z_{MPM}. If $Z_{MPM} \le \tau$ then we decide that the pixel belongs to the class of the associated MPM classifier. If $Z_{MPM} > \tau$ then we decide the pixel does not belong to that class.



(a)



(b)

Figure 9. ROCs. (a) Training data. (b) Test data.

Table 1 and

Table **2** show the confusion matrices for all MPM classifiers working together for the training and test data, respectively. The thresholds were selected to give 90% PD for the classifiers working independently. If multiple MPM classifiers have a score less than τ then we assign the pixel to the class with the lowest MPM score. If all MPM classifiers have a score greater than τ then we assign the pixel to an *unknown* change class. The rows of the confusion give actual change types and the columns give the change types declared by MPM. Again, if none of the classifier outputs surpass the threshold then the *unknown* class is decided (last column). From the confusion results one can see that *Grnd* classifier has difficulty distinguishing ground from trees. This is illustrated more in the test confusion matrix. Since the classifier is hyperparameter free we know that this is an inherent anomaly in the data and not in the parameter selection for the classifier.

Table 1. Confusion matrix for training data

	Grnd	**Low**	**Tree**	**UNK**
Grnd	93%	0%	6%	1%
Low	0%	97%	0%	3%
Tree	2%	0%	96%	3%

Table 2. Confusion matrix for test data

	Grnd	Low	Tree	UNK
Grnd	74%	0%	17%	9%
Low	0%	96%	0%	4%
Tree	0%	0%	99%	1%

6.2. Confusion image

Figure 10 shows the CCD image created from two polarmetric images taken at different times. The values in the CCD image range from 0 to 1 where 0 indicates low coherence and 1 indicates high coherence. One can see a lot of false change due to the trees and shadow. From the OC and H/A/α decomposition of it's parent images we can create at feature vector (15). This vector is then quantized (18) and then processed by the three MPM classifiers, one classifer for each class of interest. The output of the three MPM classifiers can be compared across classes to make decisions about what type of change is contained within a pixel. Furthermore, only the pixels that have a low-coherence value are evaluated. We consider a low-coherence value to be $|\gamma|$ <0.7. For a given pixel with low-coherence, the change type is declared by the minimum of the classifiers except when all the outputs are greater than the decision threshold, then the unknown class is declared. Figure 11 illustrates the *confusion image*. The gray/white regions in the Figure are the no change CCD values for $|\gamma|$ ≥0.7, the red, green, and blue regions represent the classes: *Grnd*, *Tree*, and *Low* classes respectively. The *unknown* class is given by the cyan color.

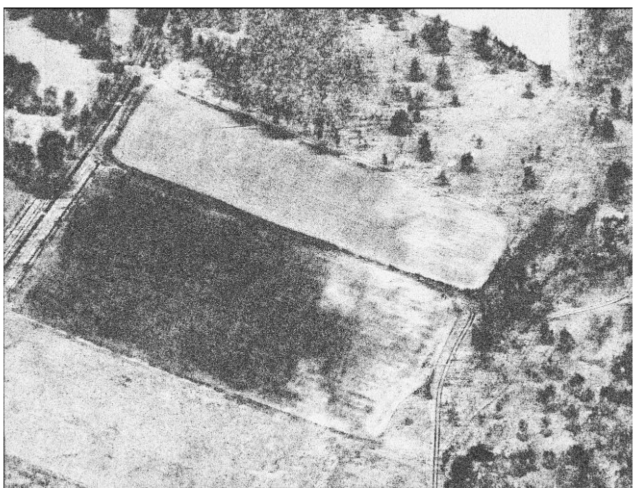


Figure 10. CCD image. Darker values indicate change or no return.

Note the transition regions between the trees and the radar shadows they cast. The MPM model declares much of these transition regions as unknown.

7. Conclusion

As more SAR radars become available with polarimetric capabilities it's important to be able to exploit these data to their full advantage. Many researchers have shown how to use polarimetric SAR to determine the scattering type in a single polarimetric image set. We have extended the H/A/α polarimetric decomposition to detect the type of change in an optimal coherence image. We can sucessfully identify three different types of low coherence: low return areas, ground and trees. We stress the importance of developing an open-set classifier, so that we do not have to train with all possible changes that produce low coherence. This gives rise to an unknown class which represents low coherence types that were not part of the training processes.

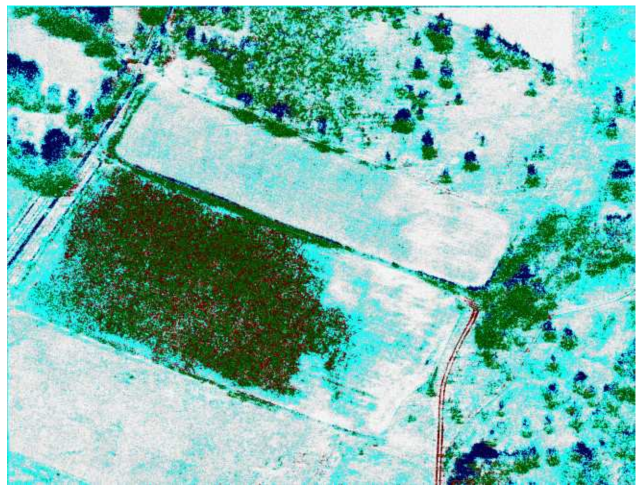


Figure 11. CCD image in Figure 10 color coded by change type detected by MPM. The classes are *Tree* (green), *Low* (blue) and *Grnd* (red). If the largest selected value is greater than the associated change-type threshold, then it is labeled as unknown (UNK) and is colored cyan. No change is shown as grayscale.

Acknowledgements

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

References

[1] J. L. Alvarez-Perez, “A multidimensional extension of the concept of coherence in polarimetric sar interferometry,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 3, pp. 1257–1270, March 2015.

[2] S. R. Cloude and K. P. Papathanassiou, “Polarimetric sar interferometry,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 5, pp. 1551–1565, Sep 1998.

[3] S. R. Cloude and E. Pottier, “A review of target decomposition theorems in radar polarimetry,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, no. 2, pp. 498–518, Mar 1996.

[4] C. V. Jakowatz, Jr., D. E.Wahl, P. E. Eichel, D. C. Ghiglia, and P. A. Thompson. *Spotlight-mode Synthetic Aperture Radar: A Signal Processing Approach*. Springer, 1996.

[5] Koch, Mark W., Greg B. Haschke, and Kevin T. Malone. "Classifying acoustic signatures using the sequential probability ratio test." *Sequential Analysis* 23, no. 4 (2004): 557-583

[6] Koch, Mark W., and Kevin T. Malone. "Vehicle classification in infrared video using the sequential probability ratio test." In *Augmented Vision Perception in Infrared*, pp. 63-84. Springer London, 2009.

[7] Koch, Mark W., Mary M. Moya, Larry D. Hostetler, and R. Joseph Fogler. "Cueing, feature discovery, and one-class learning for synthetic aperture radar automatic target recognition." *Neural Networks* 8, no. 7 (1995): 1081-1102.

[8] Kolmogorov, Vladimir, and Ramin Zabin. "What energy functions can be minimized via graph cuts?" *IEEE transactions on pattern analysis and machine intelligence* 26, no. 2 (2004): 147-159.

[9] J. Lee and E. Pottier, *Polarimetric Radar Imaging: From Basics to Applications*. CRC Press, 2002.

[10] Maaten, Laurens van der, and Geoffrey Hinton. "Visualizing data using t-SNE." *Journal of Machine Learning Research* 9, no. Nov (2008): 2579-2605.

[11] Malinas, Rebecca, Tu-Thach Quach, and Mark W. Koch. "Vehicle track detection in CCD imagery via conditional random field." In *Signals, Systems and Computers, 2015 49th Asilomar Conference on*, pp. 1571-1575. IEEE, 2015.

[12] Moya, Mary M. and Larry D. Hostetler. “One-class generalization in second-order backpropagation networks for image classification.” *Proc. Int’l Joint Conf. Neural Networks*, IEEE, 1990.

[13] Moya, Mary M., and Don R. Hush. "Network constraints and multi-objective optimization for one-class classification." *Neural Networks* 9, no. 3 (1996): 463-474.

[14] Park, Sang-Eun, and Wooil M. Moon. "Unsupervised classification of scattering mechanisms in polarimetric SAR data using fuzzy logic in entropy and alpha plane." *IEEE Transactions on Geoscience and Remote Sensing* 45, no. 8 (2007): 2652-2664.

[15] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill Book Company, New York, 1965.

[16] Quach, Tu-Thach, Rebecca Malinas, and Mark W. Koch. "A model-based approach to finding tracks in SAR CCD images." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 41-47. 2015.

[17] Quach, Tu-Thach. "Convolutional networks for vehicle track segmentation." *Journal of Applied Remote Sensing* 11, no. 4 (2017): 042603.

[18] Quach, Tu-Thach, Rebecca Malinas, and Mark W. Koch. "Low-level track finding and completion using random fields." *Electronic Imaging* 2016, no. 14 (2016): 1-5.

[19] Quach, Tu-Thach. "Vehicle track segmentation using higher order random fields." *IEEE Geoscience and Remote Sensing Letters* 14, no. 3 (2017): 369-373.

[20] T. J. Santner and D. E. Duffy, *The Statistical Analysis of Discrete Data*, Springer-Verlag, New York, 1989.

[21] Scheirer, Walter J., Anderson de Rezende Rocha, Archana Sapkota, and Terrance E. Boult. "Toward open set recognition." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, no. 7 (2013): 1757-1772.

[22] K. M. Simonson, “Multinomial Pattern Matching: A Robust Algorithm for Target Identification,” *Proceedings of Automatic Target Recognizer Working Group*, Huntsville, AL, 1997.

[23] G. Singh, Y. Yamaguchi, and S. E. Park, “General four-component scattering power decomposition with unitary transformation of coherency matrix,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 5, pp. 3014–3022, May 2013.

[24] M. Qong, “Coherence optimization using the polarization state conformation in polinsar,” *IEEE Geoscience and Remote Sensing Letters*, vol. 2, no. 3, pp. 301–305, July 2005.

[25] R. D. West and R. M. Riley, “Polarimetric Interferometric SAR Change Detection Discrimination,” submitted to *IEEE Transactions on Geoscience and Remote Sensing*.

[26] S. Wilks, *Mathematical Statistics*, John Wiley & Sons Inc, New York, 1962.

[27] B. Wu, L. Tong, Y. Chen, and L. He, “New methods in multibaseline polarimetric sar interferometry coherence optimization,” *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 10, pp. 2016–2020, Oct 2015.