SAND2018-12632J

Polarimetric Interferometric SAR Change Detection Discrimination

R. Derek West and Robert M. Riley

***Abstract*—A coherent change detection (CCD) image, computed from a geometrically-matched, temporally separated pair of complex-valued synthetic aperture radar (SAR) image sets, conveys the pixel-level equivalence between the two observations. Low coherence values in a CCD image are typically due to either some physical change in the corresponding pixels, or a low signal-to-noise observation. A CCD image does not directly convey the nature of the change that occurred to cause low-coherence.**

In this paper, we introduce a mathematical framework for discriminating between different types of change within a CCD image. We utilize the extra degrees-of-freedom and information from polarimetric interferometric SAR (PolInSAR) data and PolInSAR processing techniques to define a 29-dimensional feature vector that contains information capable of discriminating between different types of change in a scene. We also propose two change-type discrimination functions, that can be trained with feature vector training data, and demonstrate change-type discrimination on an example image set for three different types of change. Furthermore, we also describe and characterize the performance of the two proposed change-type discrimination functions by way of receiver operating characteristic (ROC) curves, confusion matrices, and pass matrices.

***Keywords*—Polarimetric Interferometric SAR (PolInSAR), Coherent Change Detection (CCD), Optimum Coherence, H/A/α Decomposition, Feature Vector, H/A/α Filter, Probabilistic Feature Fusion (PFF) Model**

I. Introduction

Complex-valued synthetic aperture radar (SAR) images, when properly calibrated, represent measurements of the magnitude and phase of scattering mechanisms within a scene, as measured for a given nominal wavelength of the radio frequencies of the electromagnetic (EM) spectrum. The magnitude map directly communicates the radar cross-section (RCS) distribution across the imaged scene, and has a spatial structure that generally correlates with underlying scattering intensities; the corresponding phase of the complex-valued has the appearance of a uniform random field with no directly apparent correlation with underlying scattering physics. However, the utility of the phase of the measured data is manifested when interfering two complex-valued images of a scene, collected with nearly equivalent geometries. The magnitude of the interfered image product can provide an indication of the changes that have occurred in the scene between observation instances, and is traditionally known as a coherent change detection (CCD) image.

Traditional CCD images, computed from single-polarization complex-valued SAR data, indicate regions of a scene where a loss of coherence has occurred between observations. These low-coherence regions are generally interpreted as "changes" within the scene, and are typically the result of thermal noise, natural temporal decorrelation, wind, or actual changes to scattering surfaces caused by, for example, ground vehicles traversing over non-indelible surfaces. Traditional CCD images do not directly enable discrimination between the causes of decreased coherence. A value in the continuous interval $[0\,1]$ is assigned to each pixel in a CCD image to indicate the estimated coherence magnitude, but there is nothing to indicate the nature of, or underlying cause of, observed variations in coherence.

Several algorithms have been proposed for detecting changes in polarimetric SAR (PolSAR) and polarimetric interferometric SAR (PolInSAR) images sets. The methods have ranged from hypothesis testing based-on equal-scattering mechanisms and power-ratios [1], detecting and thresholding partial target observations [2], generalized likelihood-ratio test that can discriminate between actual changes and low-coherence due to noise by accounting for low signal-to-noise ratio, and change detection based on polarimetric contrast minimization [3].

In additional related work, researchers have used PolSAR and PolInSAR image sets to isolate specific types of change such as detecting regions that have changed from bare land to water and changes in land cover [4] and detecting buildings in urban areas [5]. Furthermore, other research has been conducted to use PolInSAR for improved unsupervised classification of image content [6], [7].

This paper introduces a framework and two methods for identifying, and discriminating between, different regions of low-coherence within a CCD image, such as identifying low-coherence due to radar shadow or vegetation. The proposed framework is built around the additional degrees of freedom that PolInSAR provides and enables a scattering-physics based discrimination between various regions of coherence loss.

II. Background

A. Polarimetric SAR

For the sake of this paper, it is assumed that the data are collected from a well-designed airborne fully-polarimetric radar sensor, that an appropriate image formation algorithm has been applied to the observations, and that fully-polarimetric calibration of the raw scattering observations has been performed.

A calibrated single-polarization SAR image conveys the RCS of the scattering in a scene for a selected polarization state. Examples of polarization states include: transmitting "vertically"-oriented EM fields and receiving vertically-oriented EM fields (VV polarization) and transmitting and receiving "horizontally"-oriented EM fields (HH polarization).

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.

The first observation is that of the "V"-pol co-polarization response, while the latter is the orthogonal "H"-pol co-polarization response. For each co-polarization observation, there is also a corresponding orthogonal cross-polarization response that can be observed. A calibrated fully-polarimetric SAR image set conveys the RCS of the scattering in a scene for a complete set of co- and cross-polarization bases. There is a continuum of possible orientation basis sets that can be selected from; the "standard" linear polarization bases will be used in this paper: VV, HH, HV, and VH, where XY indicates that a Y oriented EM field was transmitted and an X oriented EM field was received. (Note that any other polarization basis can be synthesized from any other fully-polarimetric basis.) Measured PolSAR data can be expressed 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; there are of course one of these observations at every pixel location within the formed image products. For this paper, it is posited that all observations were performed in a monostatic manner (i.e. stationary and time invariant, with no displacement between sensor position between illumination and reception), and hence the law of reciprocity at the scattering boundaries states that $\tilde{S}_{\mathrm{HV}} \approx \tilde{S}_{\mathrm{VH}}$; the approximation is asserted because in reality all scattering observations are in fact a superposition of the actual scattering signal and undesired independent additive noise contributions. Under the assumption of cross-polarization reciprocity, we define an alternate cross-polarization observation, $\tilde{S}_{\mathrm{CX}} = (\tilde{S}_{\mathrm{HV}} + \tilde{S}_{\mathrm{VH}})/2$.

The total power of scattering observations from all of the polarization channels can also be combined into a span image, with the span calculated as:

$$\text{Span} = |\tilde{S}_{\mathrm{HH}}|^2 + |\tilde{S}_{\mathrm{HV}}|^2 + |\tilde{S}_{\mathrm{VH}}|^2 + |\tilde{S}_{\mathrm{VV}}|^2. \tag{2}$$

B. Polarimetric SAR Decompositions

For clutter observations, the utility of a fully-polarimetric image set comes from the decomposition of coherent combinations of the polarization channels. Polarimetric decompositions have been an active area of research for the past three decades and a variety of decompositions exist in the literature such as the Yamaguchi four-component scattering model [8], the general four-component scattering power decomposition with unitary transformation of coherency matrix (G4U) [9], and the information theoretic H/A/α decomposition [10]. The general utility of a polarimetric decomposition is to map or distribute the raw observations to some specified categories or "lexicon" of specified scattering mechanisms. One of the simplest such mappings is the Pauli feature vector, which is constructed by projecting the 2×2 scattering matrix onto the Pauli-spin matrix bases [10]. For monostatic observations, the 3×1 Pauli feature vector has the following form:

$$\mathbf{k} = \frac{1}{\sqrt{2}} \begin{bmatrix} \tilde{S}_{\mathrm{HH}} + \tilde{S}_{\mathrm{VV}} \\ \tilde{S}_{\mathrm{HH}} - \tilde{S}_{\mathrm{VV}} \\ \tilde{S}_{\mathrm{CX}} \end{bmatrix}. \tag{3}$$

The general physical interpretation of the first entry of the Pauli feature vector is that it represents odd-bounce scattering mechanisms such as spheres and trihedral corner reflectors. The physical interpretation of the second entry is that it represents even-bounce scattering mechanisms such as vertically or horizontally rotated diplanes, while the third entry represents even bounce from dihedrals rotated $45°$; in this case, rotation is with respect to angle around the observation line-of-sight direction vector.

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

$$\mathbf{T} = \left\langle \mathbf{k}\mathbf{k}^H \right\rangle_N, \tag{4}$$

where $\langle\cdot\rangle$ denotes a spatial ensemble average over a neighborhood of N pixels, and the superscript ^{H} denotes conjugate transpose. This polarimetric coherency matrix is the preferred observation input form for most polarimetric decompositions.

The information-theoretic H/A/α polarimetric decomposition will be utilized in this paper [10]. The H/A/α decomposition utilizes the eigendecomposition of the polarimetric coherency matrix,

$$\mathbf{T} = \mathbf{U}\Lambda\mathbf{U}^{-1} \tag{5}$$

$$= \sum_{i=1}^{3} \lambda_i \mathbf{u}_i \mathbf{u}_i^H, \tag{6}$$

where $\mathbf{U}$ is the matrix of eigenvectors (note that $\mathbf{U}^{-1} = \mathbf{U}^H$ for Hermitian symmetric matrices) and Λ is a diagonal matrix of the corresponding eigenvalues; the eigenvectors $\mathbf{u}$ share the same polarimetric bases as the input Pauli feature vectors. The H/A/α decomposition computes the following quantities from the set of eigenvectors and eigenvalues (the eigenvalues are assumed ordered, $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$):

$$P_i = \frac{\lambda_i}{\sum_{i=1}^{3} \lambda_i}, \qquad 0 \leq P_i \leq 1, \tag{7}$$

$$H = -\sum_{i=1}^{3} P_i \log_3 P_i, \qquad 0 \leq H \leq 1, \tag{8}$$

$$A = \frac{\lambda_2 - \lambda_3}{\lambda_2 + \lambda_3}, \qquad 0 \leq A \leq 1, \tag{9}$$

$$\alpha = \sum_{i=1}^{3} P_i \cos^{-1}\left(|\mathbf{u}_i(1)|\right), \qquad 0° \leq \alpha \leq 90°, \tag{10}$$

The α parameter indicates the average scattering mechanism; H is the entropy parameter, indicating the dominance (or purity) of the indicated average scattering mechanism; and A is the anisotropy parameter, indicating the relative significance of the second and third eigenstates.

C. Optimum Coherence

Two complex-valued SAR images, from sufficiently similar observation geometries with highly overlapped spatial spectral support, can be spectrally trimmed and spatially co-registered at the sub-pixel level, and then "interfered" to produce a

corresponding complex-valued coherence image. For temporally separated observations, the magnitude of the resulting coherence estimate conveys the degree to which scattering observations in the scene have maintained coherence (i.e. information equivalence).

Given two such co-registered complex-valued images, $\tilde{\iota}_1$ and $\tilde{\iota}_2$, the pixel-wise complex-valued sample coherence estimate can be computed using a sample correlation function,

$$\tilde{\gamma} = \frac{\langle \tilde{\iota}_1 \tilde{\iota}_2^* \rangle_N}{\langle |\tilde{\iota}_1|^2 \rangle_N \langle |\tilde{\iota}_2|^2 \rangle_N}, \qquad 0 \leq |\tilde{\gamma}| \leq 1, \tag{11}$$

where the superscript * denotes a complex-conjugation, and N indicates the number of neighborhood image pixels included in the ensemble average operation [11]. A CCD image is formed from the magnitude of $\tilde{\gamma}$ and an interferogram is formed from the angle of $\tilde{\gamma}$.

Similar to the single-polarization case, temporally separated PolSAR observations can be co-registered and interfered. However, unlike the scalar case, PolSAR data sets are encapsulated in a complex-valued matrix or vector of observations. Within the past couple of decades, researchers in the PolInSAR community have developed various methodologies for producing coherence estimates from PolSAR image sets [12], [13], [14]. One class of such methods are based on a coherence optimizing (CO) operation, whereby the resulting coherence estimation is maximized by determination of appropriate complex-valued unit-length weighting vectors informed by underlying scattering processes, and provide improved quality interferograms over what can be produced with single-polarization observations [15], [16].

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

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

where $\mathbf{\Omega}_{12} = \left\langle \mathbf{k}_1 \mathbf{k}_2^H \right\rangle_N$, $\mathbf{T}_{11} = \left\langle \mathbf{k}_1 \mathbf{k}_1^H \right\rangle_N$, $\mathbf{T}_{22} = \left\langle \mathbf{k}_2 \mathbf{k}_2^H \right\rangle_N$, and the $\mathbf{w}_i$ vectors are the weighting vectors, to be determined, that steer the OC process to the maximum indicated coherence that the observations can support. These weighting vectors utilize the same bases as the input Pauli feature vectors, and hence correspond to equivalent scattering mechanism definitions.

All of the quantities in the right-hand side of equation (12) can be computed from the data, except the $\mathbf{w}_i$ vectors; to find the weighting vectors, equation (12) is re-cast as maximizing the following Lagrangian function:

$$\begin{aligned} L(\mathbf{w}_1, \mathbf{w}_2) \quad = \quad & \mathbf{w}_1^H \mathbf{\Omega}_{12} \mathbf{w}_2 + \mu_1 \left(\mathbf{w}_1^H \mathbf{T}_{11} \mathbf{w}_1 - C_1 \right) \\ & + \mu_2 \left(\mathbf{w}_2^H \mathbf{T}_{22} \mathbf{w}_2 - C_2 \right), \end{aligned} \tag{13}$$

where the objective is to maximize the first term, subject to the constraints given in the next two terms. This operation is "unconstrained" in that equality of $\mathbf{w}_1$ and $\mathbf{w}_2$ is not required; other maximization schemes enforce equality of the optimizing eigenvectors. Enforcement of optimizing eigenvector equality assumes no change in underlying scattering mechanisms, which is not in general an accurate assumption, especially when there is temporal separation between observations; furthermore, additional information may be contained in the unconstrained $\mathbf{w}$ vectors that can help to discriminate between different types of change.

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

$$\nu = \mu_1 \mu_2^* \tag{14}$$

$$\mathbf{T}_{11}^{-1} \mathbf{\Omega}_{12} \mathbf{T}_{22}^{-1} \mathbf{\Omega}_{12}^H \mathbf{w}_1 = \nu \mathbf{w}_1 \tag{15}$$

$$\mathbf{T}_{22}^{-1} \mathbf{\Omega}_{12}^H \mathbf{T}_{11}^{-1} \mathbf{\Omega}_{12} \mathbf{w}_2 = \nu \mathbf{w}_2, \tag{16}$$

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 ($0 \leq |\nu_3| \leq |\nu_2| \leq |\nu_1| \leq 1$), and two corresponding sets of 3×1 eigenvector triplets ($\{\mathbf{w}_{1,1}\ \mathbf{w}_{1,2}\ \mathbf{w}_{1,3}\}$ and $\{\mathbf{w}_{2,1}\ \mathbf{w}_{2,2}\ \mathbf{w}_{2,3}\}$). It can be shown that the optimum coherence estimates can be calculated via $|\tilde{\gamma}_{Opt,i}| = \sqrt{|\nu_i|}$, with $|\tilde{\gamma}_{Opt,1}|$ (further denoted as $|\tilde{\gamma}_{Opt}|$) as the maximum supported observation coherence estimation magnitude.

Typically, the weighting vectors are handled as utilitarian within the scope of determining the optimal estimate of complex interferometric coherence. At this point, the authors propose that these weighting vectors are not just a means to an end, but in fact contain information that enables categorization of underlying mechanisms for observed variations in temporal coherence.

III. Change Discrimination Framework

The change discrimination framework we propose utilizes the H/A/α decomposition parameters, as well as the optimum coherence values and the steering vectors from the OC algorithm.

Because the weighting vectors from the OC process have the same basis as the Pauli feature vectors, they can be formed into coherency matrices and processed through the H/A/α decomposition, along with the original Pauli feature vectors formed from the $\mathbf{S}_1$ and $\mathbf{S}_2$ measured scattering matrices. By doing so, H, A, α values can be produced for each of the six OC weighting vectors and the two original image sets. The H/A/α parameters can be stacked into a three-element vector as follows,

$$\mathbf{d}_X^T = [\ H \quad A \quad \alpha\ ], \tag{17}$$

where X represents the data processed through the H/A/α decomposition (*i.e.* $\mathbf{k}_1$, $\mathbf{w}_{1,1}$, etc.). A 29-dimensional feature vector can be formed by vertically concatenating the H/A/α 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,

$$\mathbf{d}^T = [\ \mathbf{d}_{\mathbf{k}_1}^T\ \mathbf{d}_{\mathbf{k}_2}^T\ \mathbf{d}_{\mathbf{w}_{1,1}}^T\ \mathbf{d}_{\mathbf{w}_{1,2}}^T\ \mathbf{d}_{\mathbf{w}_{1,3}}^T\ \mathbf{d}_{\mathbf{w}_{2,1}}^T\ \mathbf{d}_{\mathbf{w}_{2,2}}^T\ \mathbf{d}_{\mathbf{w}_{2,3}}^T\ \mathbf{e}_{\boldsymbol{\gamma},\mathbf{k}}^T\ ], \tag{18}$$

where

$$\mathbf{e}_{\boldsymbol{\gamma},\mathbf{k}}^T = [\ |\tilde{\gamma}_{Opt,1}| \quad |\tilde{\gamma}_{Opt,2}| \quad |\tilde{\gamma}_{Opt,3}| \quad ||\mathbf{k}_1||_2 \, ||\mathbf{k}_2||_2\ ]. \tag{19}$$

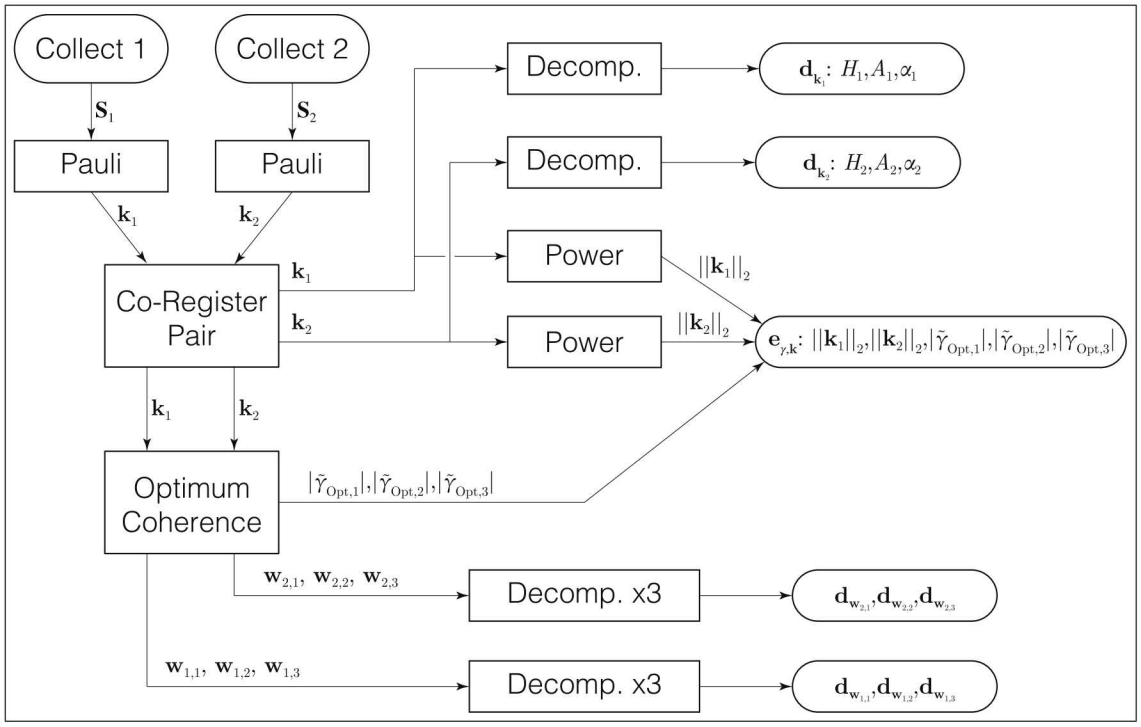


Fig. 1. Illustration of the processing steps required to form the 29-dimensional feature vectors used for change discrimination.

Figure 1 illustrates the processing steps to form the 29-dimensional feature vectors.

The H/A/α decomposition parameters computed from the two passes of the scene are natural features to use for discrimination between change-types; however, they do not include cross-coherent information in the parameters. The weighting vectors produced by the OC process are computed from the coherency matrices from both passes and from the cross-coherency matrix. The interpretation of the weighting vectors is that they extract the highest possible coherence from the cross-coherency matrix by adjusting where the magnitude and phase of their weights are placed. As an example, if a region of ground doesn't change between two passes, the steering vectors would place most of the weight on the first element of the Pauli feature vector. However, if the ground had changed between the two passes, the OC process would re-adjust the placement of the weights accordingly, in search of finding the maximum coherence. The H/A/α decomposition of the weighting vectors gives an indication of the average scattering mechanisms where coherence was still found ($\bar{\alpha}$) and the relative strength of the associated eigenvalues of the scattering mechanisms (H and A). Depending on the nature of the change in the scene, these could further discriminate change-types.

The inclusion of the set of OC coherence values, $|\tilde{\gamma}_{Opt,i}|$, for $i \in \{1,\ 2,\ 3\}$, is to separate feature vectors in an image set that have reduced coherence from feature vectors that do not contain change. Finally, the $||\mathbf{k}_i||_2$ terms, for $i \in \{1,\ 2\}$, include features that convey the magnitude values of the scene, which are otherwise normalized out in the computation of each of the other entries of the feature vector.

There are two main components to the change discrimination framework: 1) determining an exemplar feature vector for a given type of change, and 2) a function that can discriminate between feature vectors. The exemplar feature vectors can be defined from a scattering- and scene-change physics model, or alternatively can be determined from pertinent training data derived from observations of change types of interest. The behavior of the discriminating function can be mathematically stated as:

$$\gamma_{\text{ChangeD}} = f\left(\mathbf{d}, \mathbf{d}_0\right) = \begin{cases} |\gamma|, & \text{if } ||\mathbf{d} - \mathbf{d}_0|| < \varepsilon \\ 1, & \text{otherwise,} \end{cases} \tag{20}$$

where $|\gamma|$ is a selected coherence map, such as a CCD or OC image, and $\mathbf{d}_0$ is a feature vector that defines a particular type of change and $||\cdot|| < \varepsilon$ is some measure of "closeness". Functions with this behavior will isolate a particular type of change by preserving the low-coherence values for changes that are "close" to the change-type defined by the feature vector and assigning a high-coherence value to changes that are not "close" to the feature vector. The role of $|\gamma|$, which can be a CCD or OC image, is to be the canvas to display the isolated change-type. The selection of which coherence map to use will depend on user preference and does not impact the change discrimination algorithm.

IV. Data

The PolInSAR image sets used in this paper were collected in central New Mexico, USA, with the Sandia National Laboratories developed FARAD PhoeniX radar system (9.6 GHz center frequency), operated on a DHC-6 airplane. The data

acquisition geometry parameters for the image sets are given in table II. As can be seen, the repeat geometries for the image sets support interferometric processing. Furthermore, the short temporal baseline between passes limits natural decorrelation within the scene to the effects of wind and a slight amount of spatially-induced volumetric decorrelation due to the baseline between passes. Note, the noise figure σ_0 for each image varies across the scene, especially along the range direction; however, because the worst-case noise figure $\sigma_{0,\text{max}}$ is low, no first-order noticeable effects were observed for not compensating for noise.

In this paper, a data driven approach was 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 fifteen coherent, fully-polarimetric image sets. (Note, there are thirteen different scene centers in the image sets; the two image sets that share the same scene center are imaged at different aspect angles.) The training data were collected from nine of the image sets and the test data were collected from the remaining six image sets with separate scene centers from the training image sets. Within each image set, training and test feature vectors were collected for three different change types: tree (TRE), low-return (LRT), and ground (GRD). Table I summarizes the number of training and test pixels selected for each change-type, where each selected pixel corresponds to a 29-dimensional feature vector. Figure 2 illustrates example images where training data were selected. The solid colored blue, green, and red regions within the image denote selected pixels of LRT, TRE, and GRD change-types, respectively. Note, the illustrated images are pseudo-colored with the H/A/α and Span parameters in the following manner: the hue channel (in HSV space) contains a function of the α parameter, the saturation channel contains a function of the mix of the entropy and anisotropy parameters, and the value channel contains the Span image with a defined dynamic range. The resulting HSV structure is then transformed to RGB. The mapping of the α parameters was chosen to give a natural appearance to the image (*i.e.* trees are green, etc.).

TABLE I. SUMMARY OF THE NUMBER OF PIXELS SELECTED FOR TRAINING AND TESTING FOR TREE (TRE), LOW-RETURN (LRT), AND GROUND (GRD) CHANGE-TYPES. NOTE, EACH SELECTED PIXEL CORRESPONDS TO A 29-DIMENSIONAL FEATURE VECTOR.

Change Type	Training	Test
TRE	97,862	73,098
LRT	106,169	75,874
GRD	5,816	5,763

V. DISCRIMINATION FUNCTIONS

A practical and robust approach to change discrimination must be found that can take into account the nature of the entries of the feature vector. In this paper, we explore two discrimination functions and analyze their performance on blind test data. The selection of which feature vector elements to include in constructing the discrimination function can significantly change the corresponding performance of the discrimination function in isolating a change type of interest. The first discrimination function presented utilizes all of the elements of the feature vector and the second employs a simple decision rule to determine whether or not to include feature vector elements. The inclusion or exclusion of feature vector elements in the discrimination functions can be tailored to isolate a desired type of change.

A random selection of 5,000 feature vectors were drawn from the training data for each class described in table I. The training data for each class were further subdivided into disjoint sets of 2,500 samples. One set was used to train model parameters for each discrimination function approach and the other was used for determining empiric metric thresholds for performance evaluation.

A. H/A/α Filter Banks

The nature of the feature vector lends itself to designing H/A/α filters for a given type of change. That is, a filter can be designed for each $\mathbf{d}_X$ sub-vector of the feature vector $\mathbf{d}$, along with an optimum coherence filter and a Span filter. The filters can be implemented by way of a look-up table (LUT), where, for example, an H/A/α triplet can be used as the entry for a table and the corresponding filter value output.

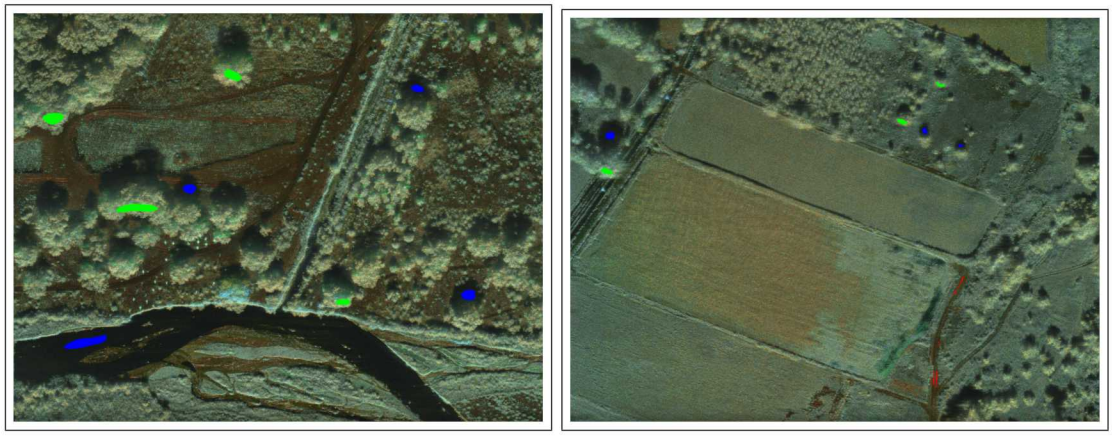
There are several different ways to design a LUT filter. For example, the steps for designing a data-driven filter are as follows:

- Gather training feature vectors for a change-type of interest
- Compute a 3-D histogram from the H/A/α and optimum coherence sub-vectors and a 2-D histogram from the Span values.
- Smooth the histogram an appropriate smoothing kernel to fill in the pass-band region and create transition regions to the stop-band
- Refine pass-band region by thresholding and saturating and normalizing the pass-band output to one
- Apply a scaling to shape the transition region output

Once the LUT filters are designed, they can be applied to data very efficiently. The input data need to be quantized to the bin resolution of the filter and then scaled to integer values that correspond to the indices of the LUT filter. The filter output is assigned by passing the integer-valued indices into the LUT filters. For the feature vector described in equation (18), this process will generate ten scalar values for each feature vector in the image; let F_{X_i} denote the output of the LUT filter X_i, for example:

$$F_{X_i} = \text{LUT}_{X_i}\left(\mathbf{d}_{X_i}\right). \tag{21}$$

There are many different ways to combine the output of the filter banks into a change-type image. One such way to produce a final output for a given change-type filter-bank consisting of the ten LUT filters is given by the following



(a) (b)

Fig. 2. Illustration of example images used for selecting training data. The solid colored blue, green, and red regions denote the where LRT, TRE, and GRD change-type pixels were selected, respectively.

change discrimination function:

$$\gamma_{\text{ChangeD}} = 1 - (1 - |\gamma|) \left(\prod_{i=1}^{10} F_{X_i} \right)^{1/10}, \qquad (22)$$

where $|\gamma|$ is the selected change map in equation (20). (For computational reasons the discrimination function is actually computed as

$$\gamma_{\text{ChangeD}} = 1 - (1 - |\gamma|)\, e^{\frac{1}{10} \sum_{i=1}^{10} \log\left(F_{X_i} + \mu\right)}, \qquad (23)$$

where μ =1E-30 is a small number to prevent the possibility of computing the natural logarithm of zero.)

Examples of the H/A/α LUT filters for GRD, TRE, and LRT change-types, for the $\mathbf{d}_{\mathbf{k}_1}$ sub-vector, are illustrated in figure 3. The passbands of the illustrated filters overlap. If this were the only filter used for change-type discrimination, then it is obvious that the performance would not be very good. However, the combination of the filter banks for the other elements of the feature vector work together to discriminate between change-types.

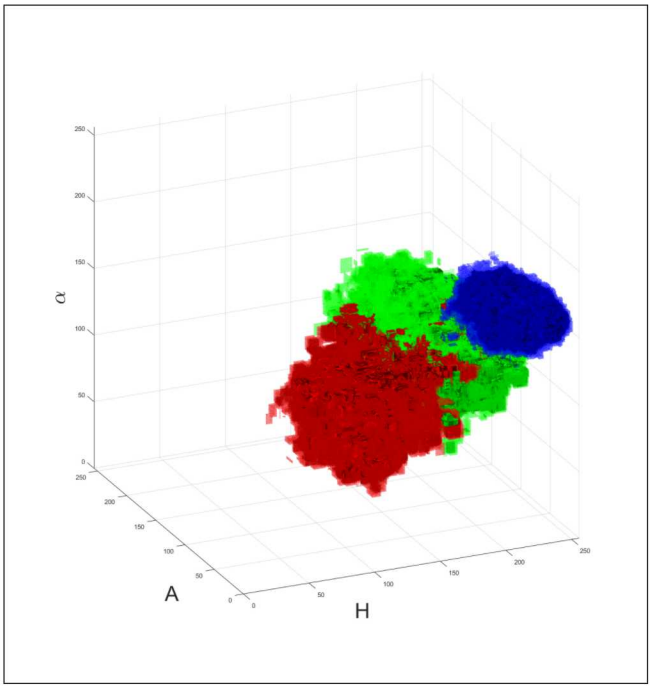


Fig. 3. Illustration of example $\mathbf{k}_1$ H/A/α LUT filters for ground (red), tree (green), and low-return (blue).

B. Probabilistic Feature Fusion

For a given change-type feature vector, it can be assumed that each component is distributed according to some probability distribution. Determining the true distribution for each component could be an intractable problem; however, if training feature vectors are available for a given change-type, a distribution can be fit to the histogram of the components of the vector; for example, if all of the feature vector components are scaled to the interval zero to one, a beta distribution may fit to them. (Note, some transformation of the components of the vector, such as the absolute distance of the training vector component values from their mean, can also be computed.)

If the empirical distributions fit the training data well, then a statistical/probabilistic method can be taken for change-type discrimination. One such technique, probabilistic feature fusion (PFF), provides a straightforward approach to this problem [17], [18]. The PFF framework allows for the modeling of features for in-class data and does not require any modeling of out-of-class data. Furthermore, the output of PFF is a p-value, which gives a measure of in-class model consistency. It is also

TABLE II. SUMMARY OF THE GEOMETRY PARAMETERS AND IMAGE RESOLUTION FOR THE TRAINING AND TESTING IMAGE SETS. THE TOP ROW GIVES THE PARAMETERS FOR THE FIRST PASS IN THE IMAGE SET AND THE BOTTOM ROW GIVES THE PARAMETERS FOR SECOND PASS. THE PARAMETERS ARE DEFINED AS FOLLOWS: ψ IS THE GRAZING ANGLE, θ IS THE SQUINT ANGLE ($+90^\circ$ DEFINES RIGHT-LOOKING BROADSIDE AND -90° DEFINES LEFT-LOOKING BROADSIDE), ϕ IS THE HEADING ANGLE, ρ_{Az} IS THE AZUMITH RESOLUTION, ρ_{Rg} IS THE RANGE RESOLUTION, $\sigma_{0,\text{MAX}}$ IS THE MAXIMUM NOISE LEVEL ACROSS THE IMAGE, RANGE IS THE DISTANCE OF THE SENSOR TO THE SCENE CENTER AT MID-APERTURE, BASELINE IS THE MID-APERTURE RANGE DIFFERENCE BETWEEN THE TWO PASSES, AND Δt IS THE TIME BETWEEN THE MID-APERTURE POINTS OF FIRST AND SECOND PASSES. NOTE THAT THE FIRST PASS IN IMAGE SET IMAGE SET 13 WAS COLLECTED LEFT-LOOKING AND SECOND PASS WAS COLLECTED RIGHT-LOOKING.

Image Set	Collection Date	ψ (degrees)	θ (degrees)	ϕ (degrees)	ρ_{Az} (m)	ρ_{Rg} (m)	$\sigma_{0,\text{max}}$ (dBsm)	Range (km)	Baseline (m)	Δ t (seconds)
		34.97	-89.94	25.98	0.1117	0.1016	-35.48	4.082		
1	07/17/2013	35.06	-89.72	25.74	0.1117	0.1016	-35.49	4.080	6.79	667
		34.96	89.74	249.26	0.1117	0.1016	-32.32	4.133		
2	07/19/2013	35.01	89.42	249.61	0.1117	0.1016	-32.35	4.126	8.58	663
		35.10	-90.29	69.29	0.1117	0.1016	-35.37	4.121		
3	07/19/2013	35.03	-89.19	68.19	0.1117	0.1016	-35.36	4.125	6.12	661
		35.04	-90.97	15.99	0.1117	0.1016	-35.34	4.134		
4	07/22/2013	34.92	-91.52	16.55	0.1117	0.1016	-35.35	4.128	10.05	632
		34.99	-89.53	14.54	0.1117	0.1016	-35.33	4.135		
5	07/22/2013	35.08	-88.69	13.71	0.1117	0.1016	-35.34	4.132	6.57	633
		35.01	-90.08	15.10	0.1117	0.1016	-35.33	4.135		
6	07/22/2013	35.12	-88.66	13.66	0.1117	0.1016	-35.36	4.129	10.08	714
		35.03	90.30	194.72	0.1117	0.1016	-32.32	4.135		
7	07/22/2013	35.08	90.64	194.38	0.1117	0.1016	-32.34	4.130	5.62	677
		34.91	-90.10	69.10	0.1117	0.1016	-35.35	4.125		
8	07/24/2013	34.88	-93.11	72.11	0.1117	0.1016	-35.36	4.122	3.98	759
		34.95	-90.48	69.48	0.1117	0.1016	-35.37	4.121		
9	07/24/2013	35.12	-91.03	70.05	0.1117	0.1016	-35.37	4.123	12.53	759
		35.08	91.76	247.24	0.1117	0.1016	-32.36	4.123		
10	07/24/2013	35.13	90.41	248.60	0.1117	0.1016	-32.33	4.134	11.96	753
		34.94	-89.07	1.09	0.1117	0.1016	-35.33	4.135		
11	07/26/2013	34.99	-90.78	2.81	0.1117	0.1016	-35.34	4.130	5.69	608
		34.97	89.33	182.69	0.1117	0.1016	-32.32	4.135		
12	07/26/2013	34.99	90.18	181.85	0.1117	0.1016	-32.33	4.133	2.33	619
		33.14	-91.43	3.48	0.1015	0.1016	-32.79	3.976		
13	04/01/2015	33.48	91.55	180.47	0.1015	0.1016	-31.51	3.991	27.51	201
		29.97	-89.62	135.63	0.1121	0.1016	-38.28	4.431		
14	06/26/2013	29.90	-91.10	137.09	0.1121	0.1016	-39.48	4.445	14.92	527
		29.74	-87.78	43.81	0.1121	0.1016	-39.50	4.434		
15	06/26/2013	29.88	-88.38	44.40	0.1121	0.1016	-39.51	4.436	10.40	528

important to note that training a PFF model only requires tens, to hundreds of training data feature vectors, as opposed to other techniques that may require much more training data.

The steps to constructing a PFF model for change-type discrimination are straightforward; the high-level steps are as follows:

- Gather training feature vectors for a change-type of interest
- If necessary, transform the data in such a way that enables a good fit for a probability distribution
- Fit a selected probability distribution by estimating the distribution parameters from the data
- Compute p-values from the in-class data
- Compute p-values from the out-of-class training data from the in-class model
- Select the feature vector components that discriminate well between the in-class and out-of-class data
- Transform and fuse the p-values from the selected features
- Compute the p-values from the theoretical gamma distribution of the selected transformed and fused feature vector p-values

Once the PFF model is constructed, it is straightforward to implement on data. For a given feature vector from an image set, compute p-values from the selected components of the feature vector, fuse them, and compute a fused p-value. If the fused p-value is greater than a determined threshold then it is considered in-class, otherwise it is considered out-of-class.

As an example, consider the PFF model constructed for low-coherence regions due to trees. The training data for low-coherence caused due to trees consists of 97,862 feature vectors collected over several image sets, from which, 2,500 feature vectors were randomly selected to construct the PFF model. The following transformation was applied to the feature vector components of the training data:

$$\hat{d}_i = |d_i - \mu_i| \,, \tag{24}$$

where d_i is the i^{th} component of the feature vector and μ_i is the mean-value of the i^{th} component of the feature vector training data.

The histograms of the $\hat{d}_i$ values have a half-normal probability distribution shape; thus, the half-normal distribution parameter was estimated, for each $\hat{d}_i$, from the data. The subplots in figure 4 (a) illustrate the PFF modeling for one of the twenty-nine feature vector elements. As can be seen in the upper-left subplot, the half-normal distribution (red curve) is a good approximation to the histogram of the feature vector components. The upper-right plot illustrates that the computed in-class p-values are approximately uniformly distributed, as they should be for a good probability distribution fit. The lower-left plot in the figure on the left illustrates the separation

of the in-class (red curve) and out-of-class data. In addition, the computed out-of-class p-values plotted in the lower-right are very low.

In an effort to automate the selection of features, a twenty-bin empirical cumulative density function was computed from the histograms of the out-of-class p-value data; if the eleventh bin contained at least 75% of the probability mass, then the feature vector component was included in the model, otherwise it was discarded. From this process, and ensuring the three optimum coherence values are included, thirteen of the twenty-nine features were selected to be included in the PFF model. Figure 4 (b) summarizes the performance of the fused features. The upper-left subplot illustrates that the histogram of the fused data fits well to the theoretical distribution of the fused features; the upper-right subplot confirms the fit by illustrating the in-class fused p-values are uniformly distributed. The fusion of the selected features has excellent histogram separation from the out-of-class data and the computed out-of-class p-values are all very low.

Similar to the output of the H/A/α filter-banks, the output of a PFF model can also produce a change discrimination image. The change discrimination function for PFF is as follows:

$$\gamma_{\text{ChangeD}} = 1 - (1 - |\gamma|)\, P_{\text{Fused}}, \tag{25}$$

where $|\gamma|$ is the selected change map in equation (20) and P_{Fused} is the fused p-value map of the image.

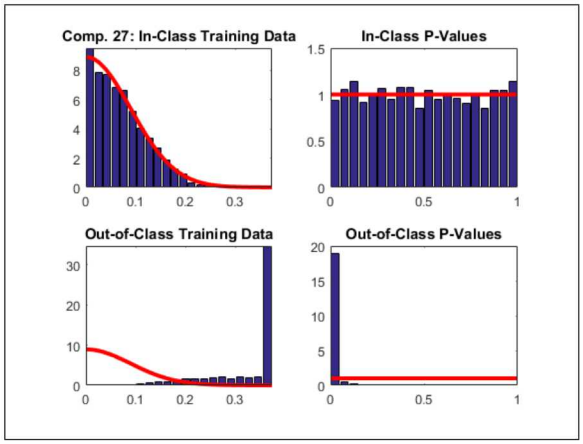
VI. Results

In this section, we characterize the relative performance of the change discrimination functions considered in this paper by computing a variety of quantitative metrics as well as by qualitatively analyzing images produced from the estimated discrimination functions.

A. Quantitative Results

To evaluate the relative performance between the approaches discussed, ROC curves were generated by drawing 2,500 feature vectors at random, for each class, from both the training and test data. The random draw was repeated five times to show the consistency of the models. The cyan and green plots illustrated in figure 5 illustrate the ROC curves generated for the training data for the PFF models and H/A/α filters, respectively; the blue and red plots illustrate the same for the test data for the PFF models and H/A/α filters, respectively.

For the TRE class, it is clear that the PFF model doesn't generalize very well from the training data to the test data. The TRE class potentially has high intra-class variability. Furthermore, the blind test data samples are not necessarily draws from the same distribution as the training data; thus, we attribute the lack of generalization to the blind test data containing samples not seen in the training data. The H/A/α model does generalize better from the training data to the test data for the TRE class, which indicates that for the TRE class it may be more robust to unseen samples in the blind test data; however, for reasonable operating points, it lacks the performance of the PFF model.



(a)

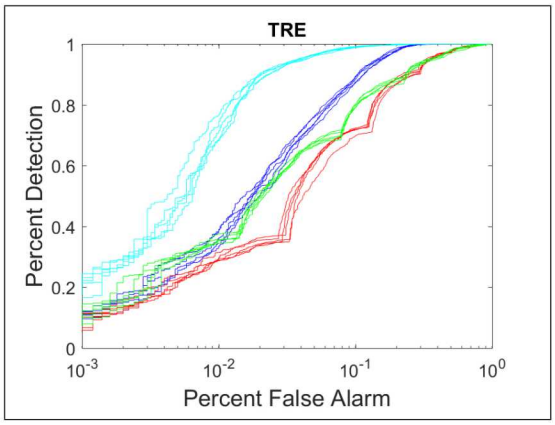
(b)

Fig. 4. Illustration of (a) the PFF modeling of the twenty-seventh feature vector element and (b) the fused selected features. The feature in (a) is an excellent feature to fuse in the PFF model, because it discriminates very well between in-class and out-of-class data. The theoretical distribution fits the data from the fused selected features very well and has excellent discrimination between in-class and out-of-class data.

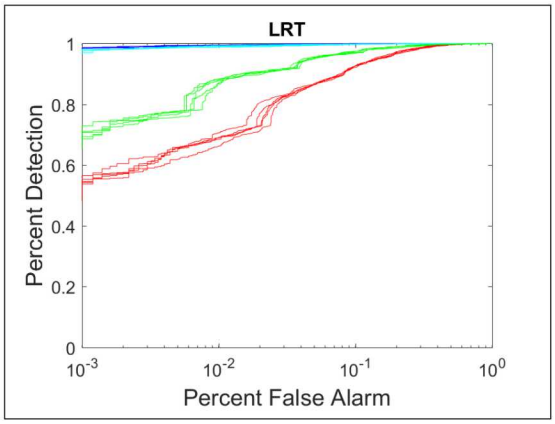
The PFF model performs exceptionally well for the LRT class on the training data and it generalizes well to the blind test data. The H/A/α filters model lacks the performance of the PFF model and doesn't generalize nearly as well, either.

Finally, the PFF model generalizes well to the blind test data for the GRD class and has good points of operation on the ROC curve. The H/A/α filters, on the other hand, do not generalize well to the test data and they lack good operating points on the ROC curve. Overall, the PFF models seem to generalize better from training to blind test data and tend to have better percent detection (PD) and percent false-alarm (PFA) performance over reasonable operating points on the ROC curves.

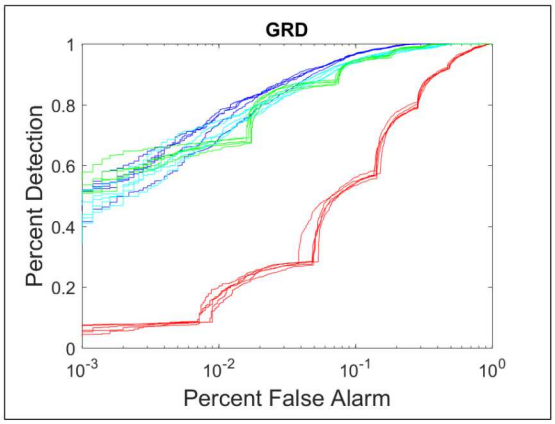
Confusion matrices were computed from the full test set



(a)



(b)



(c)

Fig. 5. Illustration of the ROC curves for the (a) TRE, (b) LRT, and (c) GRD change types. The cyan and blue curves illustrate the performance of the PFF models on the training and test data, respectively; similarly, the green and red curves illustrate the performance of the H/A/α filters for the training and test data, respectively.

as an additional measure of performance. The PD $= 90\%$ thresholds were determined empirically from the partitioned training data as described above, for each class, and were then applied to the test data. The confusion matrices for the test data for each discrimination function are given in table III. The labels on the left in the tables are the true change-type and the labels above the array are the declared change-types. Also, it should be noted that the unknown (UNK) category in the confusion matrix is populated with the maximum values that were not above the corresponding class thresholds. The computed percentages were rounded to the nearest hundredth of a percent and a '-' was placed if a computed percentage was less than 0.1%. In the confusion matrices, the TRE and LRT data perform well for both approaches; however, the GRD data were often mis-labeled as TRE, especially for the H/A/α filters.

TABLE III. THESE TABLES SHOW THE CONFUSION MATRICES GENERATED FROM THE TEST DATA FOR THE H/A/α FILTERS AND PFF CHANGE DISCRIMINATION FUNCTIONS. THE PD $= 90\%$ THRESHOLD FOR THE UNKNOWN CATEGORY WAS DETERMINED EMPIRICALLY FROM THE TRAINING DATA.

H/A/α	TRE	LRT	GRD	UNK
TRE	89.3	2.7	1.2	6.8
LRT	5.4	82.6	0.2	11.7
GRD	39.1	0.1	33.1	27.8

PFF	TRE	LRT	GRD	UNK
TRE	91.5	-	0.4	8.1
LRT	0.1	85.1	0.9	13.9
GRD	21.0	-	70.2	8.8

So-called pass matrices were also computed from the full test set. A pass matrix gives the percentage of both in-class and out-of-class test data that pass a given PD threshold. Ideally, only the PD threshold percentage will pass for in-class test data and no out-of-class data will pass the threshold. The percentage of out-of-class test data that pass the threshold gives an indication of the class similarity of an out-of-class category. (Note that there are no constraints that the rows or columns of a pass matrix must sum to 100%.) Table IV gives the pass matrices for the PFF and H/A/α filter models, where the PD $= 90\%$ threshold was determined empirically from the training data. In the H/A/α filter pass matrix, it can be seen that 23.7% of the LRT and 40.1% of the GRD test data pass the TRE threshold and that only 44.8% of the GRD test data passed its own threshold. The TRE and LRT test data perform very well for the PFF models; however, there is still a fair amount of the GRD test data that passed the TRE threshold, as well.

B. Qualitative Results

To visualize the ability to discriminate different types of change, we applied the PFF models and H/A/α filters to one of the training image sets. Figure 6(a) and (b) illustrate the span images from the first and second passes of the scene, respectively, and figures 6(c) and (d) illustrate the pseudo-colored H/A/α decomposition images for the first and second passes, respectively. Notice in the second image there is a fair

TABLE IV. THESE TABLES SHOW THE PASS MATRICES GENERATED FROM THE TEST DATA FOR THE H/A/α FILTERS AND PFF CHANGE DISCRIMINATION FUNCTIONS. THE PD $= 90\%$ THRESHOLD WAS DETERMINED EMPIRICALLY FROM THE TRAINING DATA.

H/A/α	TRE	LRT	GRD
TRE	92.0	15.4	10.9
LRT	23.7	85.8	0.2
GRD	40.1	0.1	44.8

PFF	TRE	LRT	GRD
TRE	91.7	-	1.0
LRT	-	85.5	-
GRD	26.3	-	78.8

amount of vegetation that is not focused well, especially on the right hand side of the image, which given the short temporal baseline between images, is likely due to wind and volumetric de-correlation. Figure 6(e) illustrates the optimum coherence image computed between the complex-valued PolInSAR image sets of the two passes. There are several regions where the coherence is low due to radar shadow, regions of de-correlation due to vegetation (trees and shorter vegetation), and de-correlation on a dirt road where a vehicle seems to have driven.

The data from the training image set was processed through both the PFF and H/A/α filter-bank models. Figure 7 illustrates the change discrimination images formed by evaluating equations (23) and (25) for the three classes under consideration. (Note, the H/A/α change discrimination images were formed by converting the output of the filters to empiric quantile values to increase the dynamic range.) Comparing the output of the TRE class between the PFF and H/A/α filter models in figures 7 (a) and (b), it is clear that the PFF model still contains some ground change, but overall does better at isolating the changes due to trees than the H/A/α filter model, which agrees with results in the pass matrices. Both methods perform well on the LRT class, as can be seen in figures 7 (c) and (d); notice that in the section of the river, both methods do well discriminating the low-coherence due to the water from the low-coherence due to the sand bars. Finally, figures 7 (e) and (f) illustrate the results on the GRD class. There are not many ground changes in the scene and in this case, the H/A/α approach produces a cleaner change isolation map than the PFF model.

The output of the different H/A/α filters 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 need to be evaluated. (For the sake of this paper, we are considering a low-coherence value to be $|\tilde{\gamma}_{Opt}| \leq 0.7$.) For a given pixel with low-coherence, the change type is declared by the largest output of the functions, similarly for the PFF models. Figure 8 illustrates the confusion images for (a) the H/A/α filters and (b) the PFF models. The varying light gray to white regions in the figure are the original coherence values for $|\tilde{\gamma}_{Opt}| \geq 0.7$, the red, green, and blue regions represent the GRD, TRE, and LRT classes respectively. If the largest output value is below its corresponding metric threshold, determined from the training data empirically from the PD $= 90\%$ thresholds, then it is declared unknown (UNK) and is colored cyan.

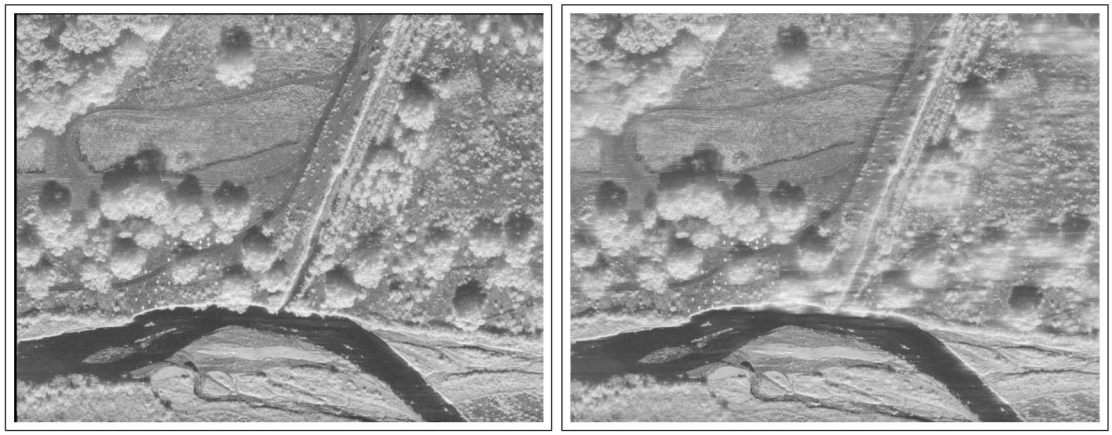
There are a few interesting regions to point out in the confusion images:

- There are multiple contributors of low-coherence in the river; there is low-coherence due to the river and also due to the sand bars that are potentially wet and densely packed, thus would have a low-radar return. However, the optimum coherence image indicates the river as a single low-coherence region. The H/A/α filter model labels the sand bars as predominantly the TRE class, whereas the PFF model declares them as mostly unknown.
- There are transition regions between the trees and the radar shadows they cast. Both the PFF and H/A/α filter models declare much of these regions as unknown. The overlap between the TRE and LRT filters illustrated in figure 3 may be a clue as to why the H/A/α filter model has less of these transition regions labeled as unknown as the PFF model.
- Many of the trees are out of focus in figure 6 (b), especially on the right hand side of the image. These regions are labeled as unknown in the PFF model confusion image, however, they seem to be primarily labeled as TRE in the H/A/α filter confusion image (more than likely for the same reason as the tree and radar shadow transition regions).
- The are regions of low-vegetation throughout the scene. Depending on the density of the vegetation, there may be a mixture of both vegetation and ground in the region. The H/A/α method seems to be finding these regions more consistent with the TRE class, which may indicate that the model for the TRE class may not be as selective (rejecting out-of-class) given that the TRE model was trained on tree data similar to what is illustrated in figures 2(a) and (b). The PFF model labels these regions as a mixture of GRD, TRE, and UNK. The inclusion of the UNK class in some of these regions may indicate that the PFF model is giving these regions low scores, which is likely due to lack of training data in such regions.
- Finally, observe that the PFF models label more of the low-coherence on the dirt road as GRD change than the H/A/α approach.

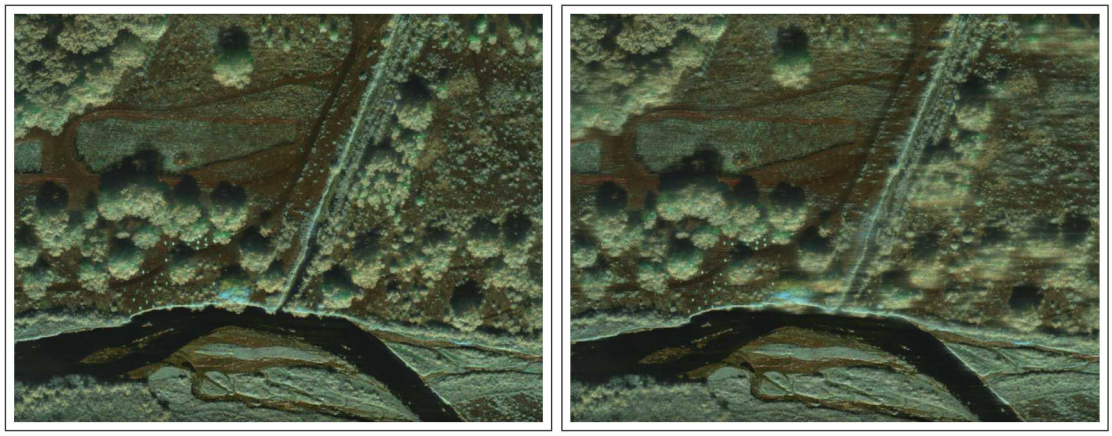
VII. CONCLUSION

In this paper we introduced and successfully demonstrated the concept of change discrimination in coherence images. We described and characterized two functions that can discriminate between different types of change. We demonstrated the ability to isolate three different types of changes contained in the low-coherence regions of an example optimum coherence image.

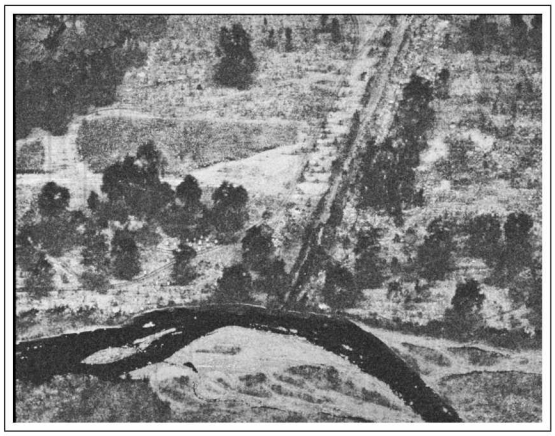
We characterized the performance of these methods using quantitative measures: ROC curves, confusion matrices, and pass matrices. We also illustrated the results qualitatively by applying the proposed discrimination functions to an image set to produce change discrimination images. Furthermore, we also created colorized confusion images where the color is determined by the highest function scores across the classes of the respective change-type discrimination functions.



(a) (b)

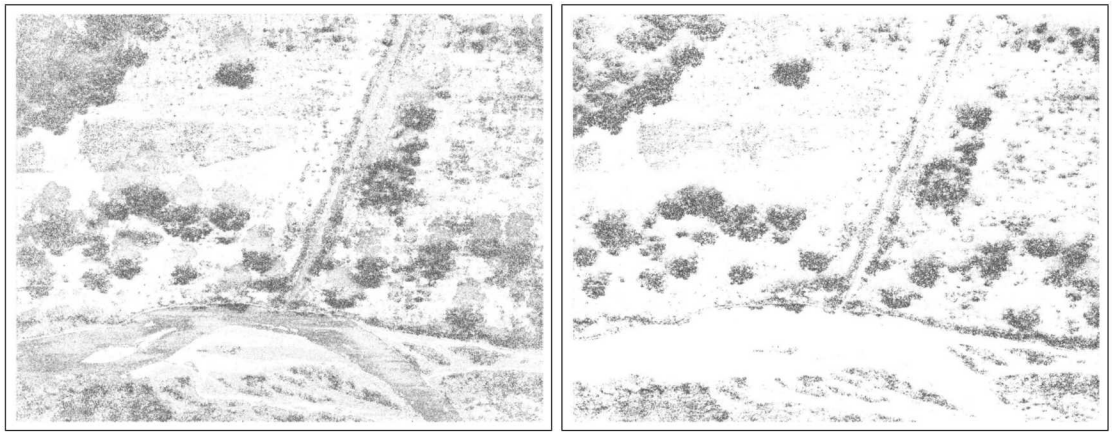


(c) (d)



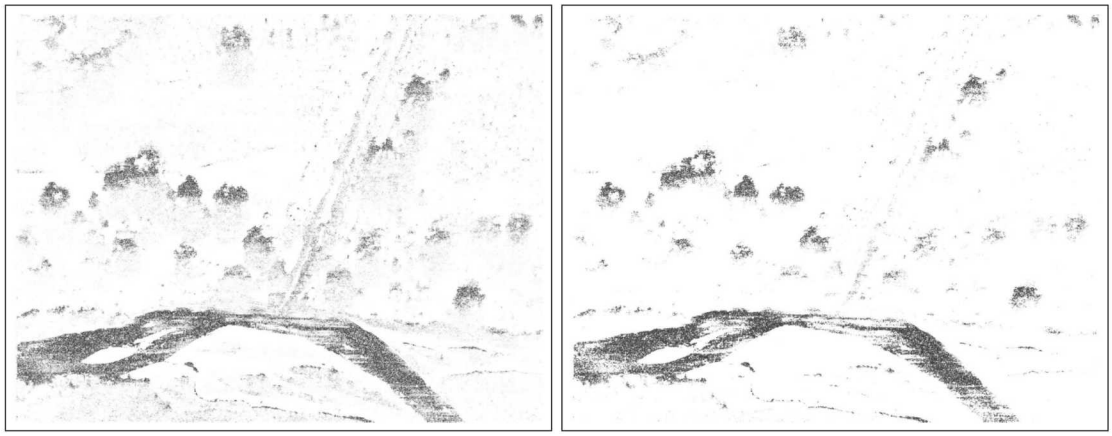
(e)

Fig. 6. Illustration of image products from two passes of an example scene. Images (a) and (b) show the Span images of the scene for the first and second passes, respectively. Images (c) and (d) show the H/A/α images of the scene for the first and second passes, respectively. Image (e) shows the optimum coherence image formed from the two passes. (Note, the values in the optimum coherence image have been squared to improve the image contrast.)



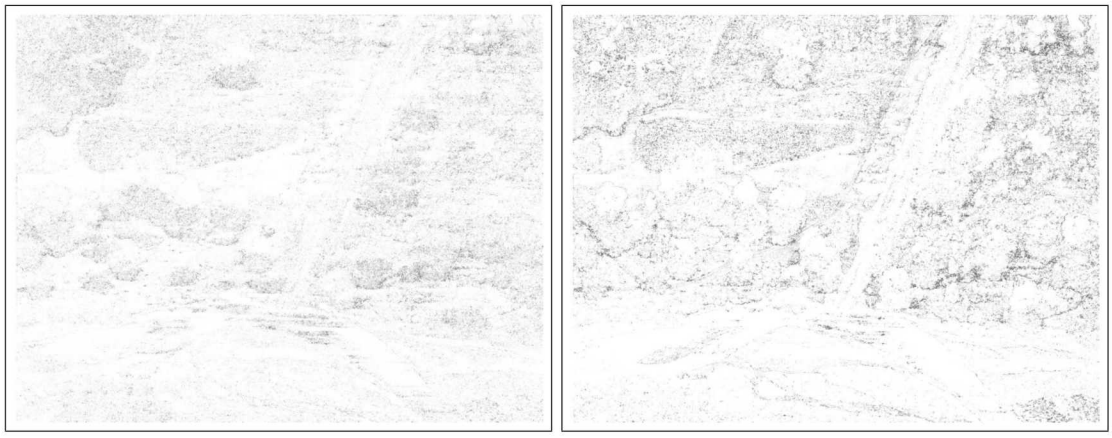
(a) H/A/α Filter: TRE

(b) PFF: TRE



(c) H/A/α Filter: LRT

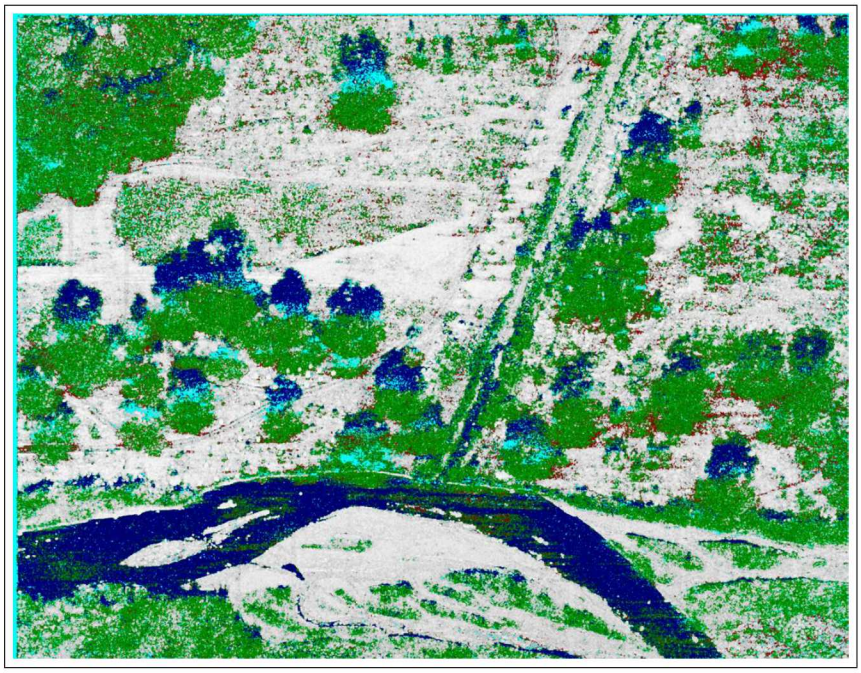
(d) PFF: LRT



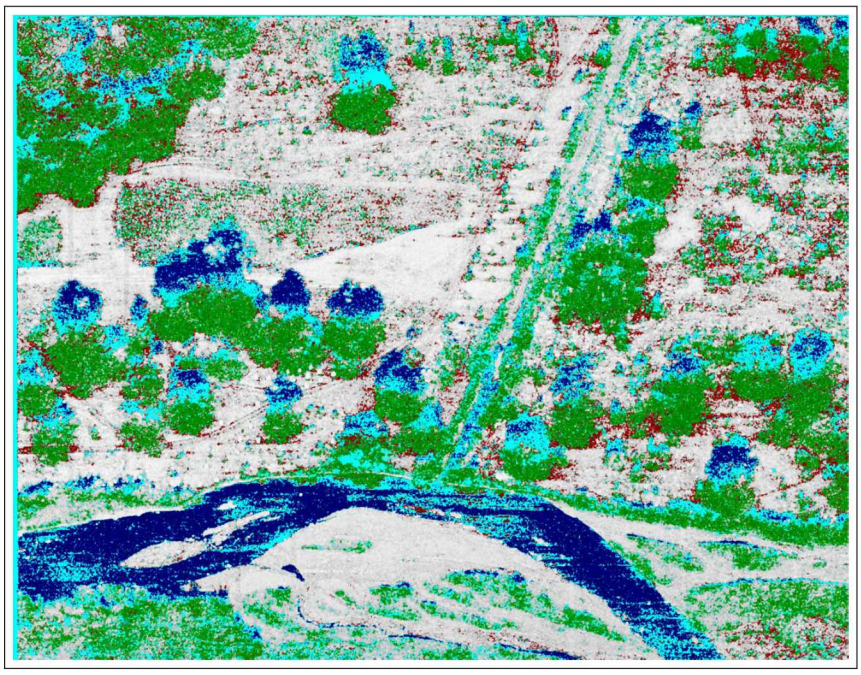
(e) H/A/α Filter: GRD

(f) PFF: GRD

Fig. 7. Illustration of the change discrimination images of the three classes under consideration from the two discrimination functions. The H/A/α filter change discrimination images are on the left and the PFF are on the right. Note that the values in the change discrimination images have been squared to improve the dynamic range for display.



(a)



(b)

Fig. 8. Illustration of the confusion images formed from (a) the H/A/α filters and (b) the PFF models for the TRE (green), LRT (blue) and GRD (red) change types. The cyan color indicates regions where the largest metric value was below the corresponding metric threshold and are labeled as unknown (UNK).

Our current research is to explore other change discrimination functions, such as machine learning methods, and to determine which methods generalize well to unseen test data and have good performance characteristics. Additionally, we are trying to understand the scattering-physics, from first-principles, to determine how to construct a feature vector for different change-types.

VIII. Acknowledgments

This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

This work was supported by the Laboratory Directed Research and Development program at Sandia National Laboratories. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and 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] A. Marino and I. Hajnsek, "A change detector based on an optimization with polarimetric SAR imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 4781–4798, Aug 2014.

[2] A. Marino, S. R. Cloude, and J. M. Lopez-Sanchez, "A new polarimetric change detector in radar imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 5, pp. 2986–3000, May 2013.

[3] J. Yin and J. Yang, "A new change detector in PolSAR imagery," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2016, pp. 5662–5665.

[4] M. Liu, H. Zhang, C. Wang, and Y. Tang, "PolSAR change detection for specific land cover type by testing equality of two PolInSAR coherency matrixes," in *2012 International Conference on Computer Vision in Remote Sensing*, Dec 2012, pp. 371–376.

[5] L. Zhang, B. Zou, and W. Tang, "Building detection based on polarimetric interferometric eigenvalue similarity parameter," in *2011 IEEE RadarCon (RADAR)*, May 2011, pp. 658–661.

[6] L. Xu, S. Li, Y. Deng, and R. Wang, "Unsupervised classification of polarimetric synthetic aperture radar interferometry using polarimetric interferometric similarity parameters and SPAN," *IET Radar, Sonar Navigation*, vol. 8, no. 9, pp. 1135–1144, 2014.

[7] M. Jager, M. Neumann, S. Guillaso, and A. Reigber, "A self-initializing PolInSAR classifier using interferometric phase differences," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 11, pp. 3503–3518, Nov 2007.

[8] Y. Yamaguchi, T. Moriyama, M. Ishido, and H. Yamada, "Four-component scattering model for polarimetric SAR image decomposition," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 43, no. 8, pp. 1699–1706, 2005.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

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

[17] K. M. Simonson, "Probabilistic fusion of ATR results," Sandia National Laboratories, Albuquerque, New Mexico 87185 and Livermore, California 94550, SAND Report SAND98-1699, August 1998.

[18] K. M. Simonson, R. D. West, R. L. Hansen, T. E. LaBruyere, and M. H. Van Benthem, "A statistical approach to combining multisource information in one-class classifiers," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 10, no. 4, pp. 199–210, 2017. [Online]. Available: http://dx.doi.org/10.1002/sam.11342