SAND2015-9431J

Single-Pol Synthetic Aperture Radar Terrain Classification using Multiclass Confidence for One-Class Classifiers

Mark W. Koch[1] , Mary M. Moya, and Ryan M. Steinbach

Sandia National Laboratories

Abstract

Except in the most extreme conditions, Synthetic aperture radar (SAR) is a remote sensing technology that can operate day or night. A SAR can provide surveillance over a long time period by making multiple passes over a wide area. For object-based intelligence it is convenient to segment and classify the SAR images into objects that identify various terrains and man-made structures that we call "static features." In this paper we introduce a novel SAR image product that captures how different regions decorrelate at different rates. Using superpixels and their first two moments we develop a series of one-class classification algorithms using a goodness-of-fit metric. P-value fusion is used to combine the results from different classes. We also show how to combine multiple one-class classifiers to get a confidence about a classification. This can be used by downstream algorithms such as a conditional random field to enforce spatial constraints.

1 Introduction

Synthetic aperture radar (SAR) [28] is an airborne sensing technology that can operate at all times of the day. It's an all-weather system that has a long standoff and can image either day or night, except in the most extreme conditions. A high resolution image of an area is created by combining multiple results from different viewing angles. SAR is also a coherent imager that measures both the phase and magnitude of the return in the area it

[1] Sandia National Laboratories, P.O. Box 5800, MS 1163, Sensor Exploitation Applications Department, Albuquerque, NM; mwkoch@sandia.gov.

is illuminating. For two registered images taken at different times, one can use coherent change detection (CCD) [28] to detect minute changes from one collection to the next.

By making multiple passes over a wide area, a SAR can provide surveillance over a long time period. One research area of interest is exploiting this large amount of imagery to learn about activities in the surveillance area. For this type of object based intelligence, one approach uses a geospatial-temporal semantic graph [5]. Here nodes of the graph represent fundamental objects in the image data such as buildings, road, water, etc., and the edges represent their spatial and temporal relationships. Image searches, such as “find all of the buildings at least 200m from a paved road,” can be accomplished using the graph framework.

Before a geospatial-temporal semantic graph can ingest this large amount of data, it is necessary to segment and classify the SAR images into objects that identify the various terrains and man-made structures. Here, we refer to these objects as *static-features*.

In this paper we concentrate on single-pol SAR imagery. While fully-polarimetric SAR can be decomposed into scattering mechanisms [9][10][13][38][40][57][67] and is very helpful in static feature extraction, many current SAR systems only have a single polarization. If multiple polarizations are available, our approach can be extended to incorporate the additional polarimetric information. Due to the complexity of registering multiple image sources, we also concentrate on using a single imaging modality. For example, optical imagery could be used as a complementary source of information in determining terrain classes.

2 Previous work on terrain classification using single-pol SAR

Terrain classification for SAR is a very important topic of research. With such a vast amount of data being generated that covers large areas of land, it can be difficult to comb through all of the raw data. In order to aid analysts and scientist in interpretation of this data it can be important to classify the images' terrain. The terrain classification problem is different depending on if the data is single-pol or multi-pol SAR data. Multi-pol SAR has many advantages in terrain classification because there is more available information. This extra information can be exploited to determine the scattering mechanisms in an image. With the knowledge of the scattering mechanisms, terrain types that appear the same in single-pol SAR will look unique in multi-pol SAR. However, obtaining multi-pol SAR data requires a more complex platform and is not always available. Since the technique presented here is focused on terrain classification utilizing only single-pol SAR, this literature review excludes multi-pol SAR techniques. Utilizing SAR backscatter imagery for classification can be complicated due to the coherent speckle inherent to the SAR imaging modality. Speckle filtering [11][21][22][34][37] is often employed as a preprocessor to segmentation. Without the filtering, the speckle can make it difficult to extract features such as edges, which can be exploited to perform segmentation.

For example, Chamundeeswari et al. [6] use an adaptive Lee filter [36] [43], and Yang et al. [69] used the Maximum Homogenous Region filter. Deng et al. [16] used multilook to reduce speckle. Another approach could perform superpixel segmentation on the raw data, [45] [66] and then perform speckle filtering on the superpixel.

Superpixel segmentation also has the benefits of reducing the complexity of the calculations and minimizing redundant information by grouping homogeneous pixels.

In order to classify the content within SAR images, the images first must be split into small windows or superpixels, as in the case of Xiang et al. [66]. Superpixels have an advantage over arbitrary rectangular windows because the superpixels group together like features and preserve the borders between regions. From these smaller segments a variety of different features can be calculated to perform classification. These features are either based on the texture or the distribution of intensities of the pixels in the small segments and include: co-occurrence probabilities, Gabor filter [8] [66] [69], Markov Random Fields [8], mean [6], [47], variance [6] [47], semi-variogram [47], lacunarity [47], wavelet components [47], histogram of oriented gradients [69], and aspect models [68]. These different calculations allow different intensity and texture features to be measured. Combinations of the above metrics can be utilized to achieve an increase in classification performance, assuming the metrics are not highly correlated [6] [8] [47]. For example, Chamundeeswari et al. fused the features using fusion based on principal component anaylsis [6].

Several different techniques can be used with the above features to perform the classification. Yang et al. [69] utilized the Extremely Randomized Clustering Forest (ERF) to cluster the features into classes. Following the ERF a Markov Random Field was used to smooth the results at the pixel level. Tison et al. [60] exploited a Bayesian classifier with a Markovian frame work to include spatial context information. In [68], a Hierarchical Markov Model was used with aspect models. Hierarchical Markov Models allow the classification to be guided by adjacent and multiscale features. Superpixels were exploited with a support vector machine to perform classifications in [66], while a K-means classifier was used in [6]. The type of classifier is less important in [6], [66] because the improvement of PCA fused features [6] and the improvements from superpixels [66] were being highlighted. Clasui performs classification using feature vectors, created from Fisher linear discriminant, with the Fisher criterion [8]. Maillard et al. also utilizes the Fisher criterion, but also uses the Mahalonobis distance and the Kolmogorov – Smirnov test [47].

In contrast to the above methods, our approach additionally uses temporal coherence statistics from multi-pass, high resolution SAR products taken over a long time period. A novel SAR product is developed called a long-term CCD (LCCD) image. Here, we assume man-made objects decorrelate slowly over time in comparison to natural features. Combining the LCCD with a time-averaged backscatter product provides high-confidence terrain classification. Following an offline training phase, our approach is fully-automated with high probability of correct detection.

Our approach is unique in that it not only makes use of higher-resolution and multiple SAR products, but takes advantage of the fact that man-made objects tend to decorrelate slowly over time. The classification algorithm is based on probabilistic fusion that produces a goodness-of-fit test statistic, which creates closed decision boundaries that surround classes-of-interest represented by the training set. It also excludes samples dissimilar to anything in the training set by declaring them as an “unknown” class. The input to the fusion algorithm is based on a Z-test for central tendency. This test captures the essence of the various SAR products while allowing for increased generalization to similar superpixels with similar statistics, but not necessarily matching distributions. Up to now the classification is made using just the pixels covered by the superpixel. There is also contextual information where adjacent superpixels tend to have the same label. We enforce these spatial constraints using a conditional random field.

3 Overall Approach

Figure 1 shows a block diagram for our approach. The input is based on what we call *SAR image products*. These products result from combining SAR images from multiple passes. Here we assume the images have been calibrated and are collected at approximately the same aspect and grazing angles. For example by registering multiple SAR images of the same area collected at different times, one can smooth out the image texture and reduce speckle [20], and reduce Doppler streaks by using the median value of each pixel across the time dimension. This produces cleaner images while maintaining spatial resolution.

Next a superpixel [1] segmentation algorithm is used to group pixels into statistically homogeneous regions. For each superpixel, feature extraction involves identifying the corresponding pixels in each SAR product. The classification stage has an array of statistical models. The rows represent the image products and the columns represent the feature vectors. The feature vectors are then combined using probabilistic fusion is used to combine the results across image products or for each column of the statistical models.

4 Multipass SAR image products

By exploiting temporal and spatial statistics, it is possible to derive a number of SAR image products designed to reduce speckle or enhance terrain features, some of which reduce the effective spatial resolution and some which do not.

4.1 Subaperture multilook

As suggested by the work of Lee [39] we create a subaperture multilook image [28] which is used by the superpixel oversegmentation algorithm. Note other products are used later in the classification process. Forming a subaperture multilook image requires Fourier transforming a complex-valued SAR image into its corresponding two-dimensional wavenumber / spatial-spectrum domain, partitioning the spectrum into non-overlapping 2D pieces, inverse Fourier transforming each sub-spectrum into its corresponding complex image, taking the magnitude, and non-coherently averaging the images formed from each piece of the spectrum [28] Even though the subaperture multilook image has the same pixel spacing, the result has an effective coarser spatial resolution (than a SAR backscatter magnitude image formed from the complete phase history), but also has reduced speckle. See Figure 2b for a four-look subaperture multilook and Figure 2a for the corresponding SAR image.

4.2 Mean and median images

We can use multiple passes of SAR images collected from the same scene to compute image statistics for speckle reduction. We calibrate each radar cross section (RCS) image and, from a stack of co-registered RCS images of the same scene, compute a median image to form the median-RCS image, which is a temporal multilook product. Figure 3a shows an example of an median RCS image.

Using a stack of CCD images co-registered to each other we can compute a temporal mean CCD. Figure 3b shows an example of a mean CCD image.

4.3 Long term CCD image

We define a long-term CCD (LCCD) image as a CCD image where the time elapsed between the first SAR image and the second SAR image is greater than a single day. In order to understand the overall decorrelation as time increases, LCCDs were generated for increasing time delta's from $\Delta t = 1 \text{ to } 14$ days. For each time delta, multiple LCCDs are possible. For example, a $\Delta t = 2$ LCCD can be generated for SAR images collected on March 1st and March 3rd. A $\Delta t = 2$ LCCD could also be generated for SAR images collected on March 2nd and March 4th.

Once LCCDs are generated for a specific Δt, the resulting LCCDs are co-registered into an image stack. The median LCCD is then calculated for a Δt by finding the median value for each pixel from all corresponding pixels in the stack. The result is a single median-LCCD image for the given time delta. The median helps remove Doppler streaks that can appear in SAR images because of moving vehicles and minor weather disturbances such as motion of vegetation during and between wind events. Figure 3c shows an example of a median LCCD image.

To analyze the behavior of the LCCD for different static features, analysts selected representative superpixels for each statistic feature. These results were verified using site visits or overhead optical imagery. The plots in Figure 4 show the result of taking the mean of the median-LCCD values for these superpixels and plotting how the mean changes with Δt. Figure 4 shows that at $\Delta t = 1$ day, there are basically three ranges of LCCD values: high, medium, and low. High LCCD is around 0.9 and occurs for paved-roads, gravel, and desert static-features. Low LCCD is below 0.5 and occurs for shadows and trees. Here the pixels for $\Delta t = 1$ are essentially decorrelated. Medium LCCD occurs between 0.8 and 0.6. The static-features in this range are man-made, cement, and soil. As Δt increases soil and desert decorrelate at a high rate whereas paved, man-made, gravel and cement decorrelate at a much slower rate. This type of feature is very powerful in discriminating paved-road from hard packed soil. Here both have similar RCS, but the soil decorrelates faster.

5 SAR segmentation

Image segmentation is the process of dividing an image into homogeneous regions. Instead of classifying individual pixels or using sliding-window-processing of pixels, we oversegment the image into superpixels. A wide body of research has been developed for, and has applied superpixel algorithms to optical imagery [1][19][41][50]. In our work these superpixels form groups of approximately 500 pixels that have similar location and intensity. They facilitate statistical characterization of classes by providing self-similar regions, reducing computation complexity, and following natural boundaries instead of introducing artifacts as sliding window approaches often do.

Recent works [23][45][59][65][66][70] have allowed application of superpixel segmentation to SAR imagery, as well, in spite of SAR speckle. We follow the approach of [45] and use the Simple Linear Iterative Clustering (SLIC) algorithm. The SLIC algorithm implements a localized k-means algorithm [17][42] with a Euclidean distance metric that depends on both spatial and intensity differences. When applying SLIC to an RCS domain such as subaperture multilook, we use the log-magnitude domain. Here, Euclidean distance in the log-magnitude domain is equivalent to a ratio-intensity distance in the magnitude domain, as proposed by Xiang *et. al* [66]. Achanta, et.al. [1] use the LAB color space channels, with equal weights when applying SLIC to optical color imagery. We may select one, two or three of the coregistered SAR image products as input channels for the SLIC segmentation. If we had reason to believe that any channel(s) were more important than other(s), we could weight the images accordingly. Figure 5 shows an example of a superpixel segmentation using the subaperture multilook SAR image in Figure 2b.

6 Classification

Figure 6 shows the classification approach in more detail. The core of the approach is what we call a *match matrix*. Each row of the match matrix represents a superpixel from a SAR image product. Each column represents models for a static-feature of interest. Each entry in the match matrix represents the score of matching the appropriate model to the superpixel corresponding to the correct SAR image product. Sensor fusion then combines the scores together and the classification module selects the best score from the different fusion algorithms.

6.1 One-class classifiers

We use a one-class classifier to match terrain models to features from a superpixel. A popular assumption in most pattern recognition approaches is the “closed-world” assumption [14]. For example, in computer vision, all the objects are assumed to be known. This allows the use of “key features” to distinguish objects [14] and machine learning algorithms [7][12][26][49][51][64] to search for differences between objects. Unfortunately, this approach does not work in unconstrained environments where potentially any object can appear in a scene.

For unconstrained environments, a *one-class* classifier is an appropriate choice [44]. For this paper, we use the term *target* to describe the class the classifier is designed for and *nontarget* for the other classes. For one-class classifiers, we are interested in one specific target θ_1 represented by the alternative hypothesis H_1, and the null hypothesis H_0 represents the nontarget $\overline{\theta_1}$ class. While this might seem like an oversimplification, one could argue that for a multi-class problem with other targets of interest $\theta_2,...,\theta_N$ one would design a one-class classifier for each of them. For the θ_1 one-class classifier we can further divide the nontargets into two groups: the other targets of interest $\theta_2,...,\theta_N$ and the unknown class θ_0. This allows us to further distinguish between two types of false alarm errors: *between-class* and *out-of-class*. Between-class errors occur when alarming on another target $\theta_2,...,\theta_N$ by calling it the target θ_1. Out-of-class errors occur when alarming on an unknown signature θ_0 by calling it the target θ_1. In making any decision, we want to control two types of errors: *missed detection errors* and *false alarm errors*. Missed detection errors result from missing a target signature by calling it a nontarget, and false alarm errors result from alarming on a nontarget signature by calling it a target.

For example, suppose we are interested in classifying regions in SAR images into classes such as road or scrub-desert or unknown. Here, θ_1 would represent the road and θ_2 the scrub-desert class. The unknown class θ_0 would represent all regions not in θ_1 or θ_2. This could be other regions such as shadow, soil, gravel, etc.. These unknown regions, a possible source of the out-of-class errors, are a significant problem in real-world pattern recognition problems in unconstrained environments. A Bayesian classifier approach designed for the road vs. scrub-desert problem, while minimizing the between-class errors, would require models of all the possible objects that could be imaged by the sensor to control the out-of-class errors. Otherwise it would classify a shadow or soil regions as road or scrub-desert. Modeling “the whole world” of possible objects is untenable for most realistic systems deployed in unconstrained environments. Instead, we use a goodness of fit (GOF) classifier to control the out-of-class errors, and power analysis [48] to model the unknown class.

Whereas Bayesian classifiers minimize the between-class error, they do nothing to control the out-of-class errors. Figure 7 illustrates this potential problem. The Figure shows a two-dimensional feature space, with samples from two targets: Target A represented by stars and Target B represented by filled circles. Assuming normal distributions and equal covariance matrices for the targets, the Bayes decision boundary has a linear form, Figure 7a. Whereas the Bayes classifier minimizes the between-class errors of the A and the B targets, it does not control the out-of-class errors caused by unknown objects represented by “x” symbols. Depending on which side of the boundary the nontarget falls, the classifier will assign the unknown to one of the known classes and make an out-of-class error. Figure 7b shows a GOF classifier that tries to surround the target class. Here, the unknown objects, that have widely differing features from the target class (“x” symbols), will be classified correctly. In general, the GOF classifier has improved out-of-class errors, but the between-class errors will increase, since it is not necessarily an optimal Bayes classifier.

6.2 The one class problem and requirements

Assume we have a template T_i for the class θ_i and a set of sensor observations U. Let d_i be the GOF between U and T_i, i.e. $d_i = f(U,T_i)$. We call d_i *miniphilic* if d_i is small for when between U and T_i are close in some distance metric space and *maxiphilic* for d_i large when U and T_i are similar. For this paper we will assume miniphilic unless specifically stated. Let $\Theta = \{\theta_0, \theta_1, ..., \theta_N\}$ be the set of possible classes, where $\overline{\theta_i}$ is any class that isn't θ_i, i.e. $\overline{\theta_i} = \Theta - \{\theta_i\}$. Let $p(d_i \mid \theta_i)$ and $p(d_i \mid \overline{\theta_i})$ represent the probability density function (PDF) of the GOF d_i, given the class θ_i and the $\overline{\theta_i}$, respectively. The PDF's can be discrete or continuous and can be determined through theoretical means or empirical modeling.

The PDF $p(d_i \mid \theta_i)$ is usually straightforward to determine, since one has knowledge of the target of interest The PDF $p(d_i \mid \overline{\theta_i})$ is usually more problematic. One approach for modeling the nontarget class uses *statistical power analysis* [48] to model the *worst case nontarget*. The approach has some similarities to that taken by [25] for modeling composite hypotheses by determining the *least favorable choice*. Power analysis assumes the tested effect is linear and the measured effect size (small, medium or large) is known. Typically, power analysis allows the statistician to determine if enough samples were collected to give the test a high power.

While the exact form of $p(d_i \mid \overline{\theta_i})$ depends on $p(d_i \mid \theta_i)$, we show an example from [30] where $p(d_i \mid \theta_i)$ is $N(0,1)$. This is convenient in problems where the central limit theorem can be applied or for when $p(d_i \mid \theta_i)$ is a heavy tailed distribution such as

Rayleigh, Chi-Square, or Gamma and a square or cube root transform can approximate normal distribution. For example if $X_{r,\lambda}$ is a random variable with a Gamma distribution with a shape and scale of r and λ then $X_{r,\lambda}^{1/3}$ is approximately normal [46] and using the Wilson-Hilferty transform [63] has mean and standard deviation:

$$\mu_{wh} = \lambda^{1/3}\Gamma(r+1/3)/\Gamma(r) \quad \text{and} \quad \sigma_{wh}^2 = \lambda^{2/3}\Gamma(r+2/3)/\Gamma(r) - \mu_{wh}^2 \tag{1}$$

In [30], it was shown that the worst case $p(d_i \mid \overline{\theta}_i)$ PDF is $N(\tilde{\mu},1)$. Here the location parameter $\tilde{\mu}$ represents the smallest acceptable effective difference between class θ_i and class $\overline{\theta}_i$.

6.3 Multiclass confidence for one class classifiers

It is often desirable to estimate the confidence of a decision given the GOF d_i, which would be inversely proportional to the uncertainty of a decision. This is useful when a human analyst needs to interpret the results or for further processing, such as generating the unary terms in a conditional random field (CRF). For the binary problem, with $\Theta = \{\theta_0, \theta_1\}$, one can use Bayes rule and the law of total probability to get

$$p(\theta_1 \mid d_1) = \frac{p(\theta_1)p(d_1 \mid \theta_1)}{p(\theta_1)p(d_1 \mid \theta_1) + (1 - p(\theta_1))p(d_1\overline{\theta}_1)} \tag{2}$$

Here $\overline{\theta}_1$ is equivalent to the unknown class θ_0 and we call $p(\theta_1 \mid x)$ the confidence of our decision that the model M_i matches the sensor observations U. Dividing the numerator and denominator by $(1 - p(\theta_1))p(d_1 \mid \overline{\theta}_1)$, equation (2) becomes:

$$p(\theta_1 \mid d_1) = \frac{\lambda_0 \Lambda(d_1)}{1 + \lambda_0 \Lambda(d_1)} \tag{3}$$

where $\lambda_0 = p(\theta_1)/p(\overline{\theta}_1) = p(\theta_1)/(1-p(\theta_1))$ is the prior odds ratio for class θ_1 and $\Lambda(d_1) = \dfrac{p(d_1 \mid \theta)}{p(d_1 \mid \overline{\theta})}$ is the likelihood ratio. A $\Lambda(x) > 1$ corresponds to an GOF d_i where the class θ_1 is more likely. As $\Lambda(x) \to \infty$, $p(\theta_1 \mid d_1)$ goes to one, indicating high confidence for class θ_1. As $\Lambda(x) \to 0$, $p(\theta_1 \mid d_1)$ goes to zero, indicating low confidence for class θ_1. For the special case of $\lambda_0 = \Lambda(x) = 1$, $p(\theta_1 \mid d_1)$ is ½ indicating we no evidence either way for class θ_1.

To extend the binary problem to a multiclass problem $\Theta = \{\theta_0, \theta_1, \ldots, \theta_n\}$ we use the Dempster Shafer theory of evidence. The Dempster Shafer (DS) theory [54] is a mathematical theory of evidence that allows combining evidence from different sources to arrive at a degree of belief. It models uncertainty by not requiring one to assign all of one's belief to a single proposition.

The main assumption we make is that evidence is *consonant*. This allows us to use the probabilistic framework that we established in section 6.2. Shafer defines consonant evidence as evidence that points in a single direction and only varies in the precision of focus [54]. This fits well with the GOF metric. The GOF describes the difference between stored knowledge, for example a template of the target, and the measured data. Thus it points only in the direction and focus of the hypothesis represented by the stored knowledge.

For the multiclass problem the frame of discernment or set of possible outcomes is the power set 2^{Θ}. For the two class problem the power set is $\{\phi, \theta_0, \theta_1, \Theta\}$, where ϕ represents the empty set. From [54] the support function for class θ_k is

$$\left.\begin{aligned} m(\bar{\theta}_k) &= 0 \\ m(\theta_k) &= 1 - \frac{p(x \mid \bar{\theta}_k)}{p(x \mid \theta_k)} \\ m(\Theta) &= \frac{p(x \mid \bar{\theta}_k)}{p(x \mid \theta_k)} \end{aligned}\right\} \text{if } \frac{p(x \mid \theta_k)}{p(x \mid \bar{\theta}_k)} \geq 1 \tag{4}$$

and the support for the nontarget $\bar{\theta}_k$ is

$$\left.\begin{aligned} m(\bar{\theta}_k) &= 1 - \frac{p(x \mid \theta_k)}{p(x \mid \bar{\theta}_k)} \\ m(\theta_k) &= 0 \\ m(\Theta) &= \frac{p(x \mid \theta_k)}{p(x \mid \bar{\theta}_k)} \end{aligned}\right\} \text{if } \frac{p(x \mid \bar{\theta}_k)}{p(x \mid \theta_k)} > 1 \tag{5}$$

where $m(\cdot)$ represents the DS basic probability assignment (BPA) function and $m(\Theta)$ represents the amount of uncertainty in the GOF d_k. Using the approach in [29], one can combine evidence additively by converting the BPA's to weights of evidence.

Two BPA's can be combined using Dempster's rule [15]:

$$(m_1 \oplus m_2)(A) = \frac{\sum_{A_1 \cap A_2 = A} m_1(A_1) m_2(A_2)}{\sum_{A_1 \cap A_2 \neq \phi} m_1(A_1) m_2(A_2)} \tag{6}$$

Unfortunately, for N classes the number of resulting BPA's is 2^N. This can get unwieldly as N gets much over say 10. Thus, as the number of classes increases the size of the frame of discernment Θ increases exponentially. This requires computing and storing a BPA for every possible subset. To overcome this problem we use the Bayesian approximation algorithm [3][62]. This approach reduces the number of focal elements (subsets of Θ) while retaining the essence of the information they represent. The Bayesian approximation reduces the BPA to a probability distribution where only singletons subsets of a frame are allowed to be focal points. Given a BPA $m(\cdot)$ the Bayesian approximation $\hat{m}(\cdot)$ is given as

$$\hat{m}(A) = \begin{cases} \dfrac{\sum_{B:A\subseteq B} m(B)}{\sum_{C:C\subseteq\Theta} m(C)\,|C|} & \text{If } |A| = 1 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

Thus for the consonant evidence (4) and (5) and for class θ_k the Bayesian approximation is:

$$\hat{m}(\theta_i) = \begin{cases} (m(\theta_k) + m(\Theta))/L & \text{If } i = k \\ (m(\bar{\theta}_k) + m(\Theta))/L & \text{otherwise} \end{cases} \tag{8}$$

where

$$L = m(\theta_k) + m(\bar{\theta}_k)N + m(\Theta)(N+1) \tag{9}$$

Here $m(\Theta)$ is spread equally over the N classes plus the unknown class and $m(\bar{\theta}_k)$ is spread equally over the all the classes except for class θ_k. We use $\hat{m}(\theta_i)$ for $p(\theta_i \mid d_i)$, and for $N = 2$, it's straightforward to show that $\hat{m}(\theta_1) = p(\theta_1 \mid d_1)$ in equation (3) for $\lambda_0 = 1$.

Table 1 shows some examples for a three class problem. To save space we don't show the uncertainty $m(\Theta)$ for each class's GOF BPA, since the sum of the BPA should sum to one, e.g. $m(\Theta) = 1 - m(\theta_k) - m(\bar{\theta}_k)$ and also all the shown BPA's are multiplied by 100.

The first row in the table shows the case where the GOF fails for all three classes and there is a high BPA $m(\bar{\theta}_k)$ for $k = 1,2,3$. Note due to the assumption of consonant evidence if $m(\bar{\theta}_k) > 0$ then $m(\theta_k) = 0$ or vice versa. This translates into combined confidence of the unknown class $\hat{m}(\theta_0)$ as 0.87. The other three classes have very low confidence. The second row shows an example where the goodness of fit is very good for class 1 and poor for the other classes. After combination the majority of confidence is placed on $\hat{m}(\theta_1)$. In row 3 the GOF increase for class 2, but remains the same for class 1. This is reflected in the confidence calculation with $\hat{m}(\theta_1)$ decreasing and $\hat{m}(\theta_2)$ increasing. In row 4 the GOF further decreases for class 1. This results in a confidence between $\hat{m}(\theta_1)$ and $\hat{m}(\theta_2)$ being very similar and also an increase in the unknown class $\hat{m}(\theta_0)$.

7 P-value fusion

We use p-value fusion [56] for not only fusing matching scores in the match matrix (Figure 6), but also to combine different statistical features (such as the 1st and 2nd moments) from the superpixels. For this section we will use the word "score" to not only mean GOF scores from the match matrix, but also statistical features that we extract from a superpixel. The important steps in the fusion process are to normalize the scores so they can be easily combined and to account for any dependence between the scores.

For ease of presentation we drop the index indicating class and assume without loss of generality that there is just one class. Let t_i, $i = 1,\ldots,q$, represent the scores to fuse. For the null distribution, let $f(t_i \mid \theta)$ represent the probability density function (PDF) of t_i conditioned on class θ and $F(t_i \mid \theta)$ represent the corresponding cumulative distribution function (CDF). Then $1 - F(t_i \mid \theta)$ is the p-value and is uniformly distributed [27]. It follows that

$$Y_i = -\log(1 - F(t_i \mid \theta)) \tag{10}$$

is exponentially distributed [18]. It is expected that Y_i will be large for scores that violate the null hypothesis. This process normalizes the scores so that they can be combined by addition:

$$\mathcal{F} = \sum_{i=1}^{q} Y_i \tag{11}$$

where $\mathcal{F}$ is the fused score. The distribution of $\mathcal{F}$ is approximated as a gamma for the null distribution with parameters

$$\hat{r} = \frac{q^2}{q + C}, \text{ and } \hat{\lambda} = \frac{q}{q + C} \tag{12}$$

where $\hat{r}$ and $\hat{\lambda}$ are the shape and scale parameters of the gamma distribution respectively, and

$$C = \sum_{i=1}^{q} \sum_{j \neq i} \hat{\rho}_{ij} \tag{13}$$

Here, $\hat{\rho}_{ij}$ represents the estimated correlation between Y_i and Y_j. The correlation coefficients $\hat{\rho}_{ij}$ are estimated from labeled training data. See [56] for more details. Intuitively, the parameter C accounts for the amount of redundant information.

Because each Y_i is expected to be large for scores that violate the null hypothesis, it is sensible to make the corresponding static feature identification decisions by setting a threshold on the value of $\mathcal{F}$. Fused scores at or below the chosen threshold are classified as the corresponding class, while those with scores above the threshold are classified as the unknown class.

8 GOF for SAR segmentation

For this section we assume we have access to a set of superpixels that have been assigned to a specific terrain class. We divide each class into three sets, one used for training, one for testing and one for validation. The training set is used to select the GOF classifier parameters, while the test set is used to select thresholds and features, and the validation set is used to create receiver operating characteristic (ROC) curves that measure feature performance.

SAR speckle is an artifact of coherent imaging. In the magnitude domain, speckle for a region with constant backscatter is typically modeled with a Rayleigh distribution [61]. The Rayleigh distribution is a good model for SAR backscatter from homogenous clutter regions such as bare ground, dense forest canopies, and snow covered ground [61]. For nonhomgenous regions other models have been proposed. For example, the lognormal or Weibull distributions have been used for sea ice [61]. Salazar proposed using the beta prime distribution as a unified model for describing homogeneous and heterogeneous clutter and extremely heterogeneous (man-made) areas [52][53]. The beta prime model is completely characterized by two parameters, which dictate the shape and scale of the SAR data. Unfortunately, estimating shape and scale parameters doesn't have a closed form solution. In [24] Gao gives a summary of the 40 year history of SAR modeling and gives more than a 100 references on the subject.

For road finding in SAR images, Koch et al. [31] use the nonparameteric two sample Kolmogorov-Smirnov (KS) test [33][49][58]. This approach could easily be extended to multiple classes by designing a KS model for each combination of class and image product. The KS test compares the CDF of pixels within a test superpixel against each trained model CDF and produces a match statistic for each. This allows the algorithm to adapt to any distribution for a given static-feature and SAR image-product. One concern is that the KS test is too sensitive and does not generalize well to similar terrains.

In our approach we use the 1st and 2nd moments, γ_1 and γ_2, of the superpixel pixel intensities, respectively. These features are descriptive enough to describe the distribution of different terrain pixels in a superpixel, but not as sensitive as the CDF used by the KS test. By the central limit theorem, the probability density function for these moments is normal with mean and variance of μ_i and σ_i^2 ($i = 1,2$), respectively. This is conditioned on the knowledge of class θ from the training data. Using the training data and the leave-one-out technique we estimate $\hat{\mu}_i$ and $\hat{\sigma}_i^2$ given θ.

The goodness of fit test is built on the two-tailed Z test for both γ_1 and γ_2 and combining their Z test p-values using p-value fusion. For the Z test, the null hypothesis H_0 is that γ_i is $N(\mu_i, \sigma_i^2)$ or belongs to class θ and the alternative hypothesis is that γ_i is not $N(\mu_i, \sigma_i^2)$ or belongs to $\overline{\theta}$. The test statistic is

$$z_i = (\gamma_i - \hat{\mu}_i)/\sigma_i \tag{14}$$

and the two-tail p-value is

$$2(1 - \Phi(|z_i|)) = -2\Phi(|z_i|) \tag{15}$$

where $\Phi(.)$ represents the $N(0,1)$ CDF. Using p-value fusion we can combine the p-values to get

$$z = -\log(-2\Phi(|z_1|)) - \log(-2\Phi(|z_2|)) \tag{16}$$

The fused score z has a gamma distribution with parameters $(\hat{r},\hat{\lambda})$ given by (12). We assume $\mathrm{E}\{x_i x_j^2\} << 1$ for $i \neq j$, so $C << 1$ for equation (12). Since $q = 2$, the gamma shape and scale parameters become $\hat{r} = 2$ and $\hat{\lambda} = 1$. The p-value for the combined GOF test is

$$1 - \Gamma(z \mid \hat{r},\hat{\lambda}) \tag{17}$$

where $\Gamma(.)$ is gamma CDF. This can be used in the p-value fusion when fusing GOF tests from multiple products.

Figure 8 shows the CDF of the p-values (17) for two different classes and products. The theoretical distribution for the p-values is uniform so the CDF is a ramp with a slope of one. The dotted lines in Figure 8 show the theoretical distribution, while the solid lines show the actual distribution of p-values. Figure 8a shows the p-values for the tree class using the median RCS product. The p-values are estimated using a leave-one-out technique using the tree-class training superpixels. For K superpixels, $K-1$ are used to estimate μ_i and σ_i^2 ($i = 1,2$) and the one left out is used to compute a p-value (17). This is done K times leaving a different superpixel out each time. The resulting CDFs are shown in Figure 8. There is a reasonable match between the actual and the theoretical distributions. Figure 8b shows the p-value CDF's for the tree class, but for the mean CCD product. Figure 8c and Figure 8d show the bright man-made class for products median RCS and 21 day LCCD, respectively.

To create performance curves we hand pick superpixels belonging to the classes of interest. The truth of the data is verified using optical imagery and site visits. Figure 9 show performance curves for two of the classes and differing products. The metric for measuring performance is the receiver operating characteristic (ROC) curves. The test superpixels were used to create these curves that plot the probability of false alarm vs. the probability of detection at different operating thresholds. There are two types of curves in Figure 9. One shows the performance for a single product score and the other shows the performance for the results for multiple products. For the single product result the ROC curve is obtained by applying a threshold to the combine moment score (16) and estimating correct detections and false alarms. For the multiple products ROC curve the threshold is applied to the fused score (11). Figure 9a and Figure 9b show the ROC curves for the tree class using the single products median RCS and the mean CCD, respectively. Figure 9c shows the fused result which is much better than either product by itself. Figure 9d and Figure 9e show the ROC curves for the bright man-made class using the single products median RCS and 21 day LCCD. Figure 9c shows the fused result which improves the PD at high false alarm rate, but not low false alarm rate. The man-made class is the most difficult probably due to the superpixel algorithm including other classes, besides the man-made class, in a group with it and layover from man-made structures mixing with the background.

9 Conditional random field

To enforce spatial consistency we represent the labeling of the superpixels as a conditional random field (CRF) [35] and infer the most likely labeling by maximize the posterior probability of the random field. Graph cuts with alpha expansion are used for this multilabel optimization problem [2]. CRF's are a special case of a Markov random filed (MRF). An MRF describes the $\Pr(A,S)$ using a graphical model. Here, $S=\{s_1,\ldots,s_M\}$ represents the pixel values underneath the M superpixels. Each s_k

represents the data in a contiguous group of pixel locations in the product images. The variable $A = \{a_1,...,a_M\}$ represents the set of class assignments. Where a_k is the label for superpixel s_k. Thus each a_k is a number between 0 and M, where 0 represents the unknown class. The MRF can be used as generative or inferential model. The problem is that s_i and features derived from s_i can be highly correlated and/or redundant and an MRF is poor at describing these correlations and requires a densely connected graph.

The CRF is an MRF conditioned on the data: $\Pr(A \mid S)$ [35]. This removes the correlations and redundancies in S, since S is now assumed given. The model is no longer generative, but that is not a problem, since we want to do inference.

Typically the CRF objective function is represented as $-\log(\Pr(A \mid S))$ and is called the energy function. Thus, instead of maximizing the posterior probability we want to find the labeling that minimizes the energy function. The second order energy function for a CRF is given as:

$$E(A) = \sum_{i \in V} \psi_i(a_i) + \sum_{i \in V, j \in N_i} \psi_{ij}(a_i, a_j) \tag{18}$$

Here V corresponds to the set of superpixel indicies and the neighbor set N_i corresponds to the set of superpixel indices that are adjacent to superpixel i. The term $\psi_i(a_i)$ describes the energy of assigning superpixel i the label a_k and is call the *unary* term. The pairwise term $\psi_{ij}(a_i, a_j)$ represents the energy of the similarity between superpixels i and j. A low energy for $a_i = a_j$ would favor the two superpixels to have the same label.

9.1 Unary term

For each superpixel $j = 1,...,M$ and terrain class $k = 1,...,N$ the unary term we want to compute is $-\log(p(\theta_k \mid \mathcal{F}_{jk}))$. To accomplish that, we first find $p(\mathcal{F}_{jk} \mid \theta_k)$ and $p(\mathcal{F}_{jk} \mid \overline{\theta}_k)$. From the construction of the p-value fusion algorithm $p(\mathcal{F}_{jk} \mid \theta_k)$ is gamma with parameters $(\hat{r}, \hat{\lambda})$ (12). Using (1) and applying the transform:

$$\hat{\mathcal{F}}_{jk} = (\mathcal{F}_{jk}^{1/3} - \mu_{wh}) / \sigma_{wh} \tag{19}$$

$p(\hat{\mathcal{F}}_{jk} \mid \theta_j)$ is approximately $N(0,1)$. From section 6.2 $p(\hat{\mathcal{F}}_{jk} \mid \overline{\theta}_k)$ is $N(\tilde{\mu}_k, 1)$. Where the location parameter $\tilde{\mu}_k$ represents the smallest acceptable effective difference between class θ_k and class $\overline{\theta}_k$. This can be determined by knowledge of the power of the test or empirically using training and testing data. Using (4) and (5) followed by (8) and then (6) we can get $\hat{m}(\theta_k \mid \hat{\mathcal{F}}_{jk})$ for each superpixel which can be used as $p(\theta_k \mid \hat{\mathcal{F}}_{jk})$ in the unary function, $-\log(p(\theta_k \mid \mathcal{F}_{jk}))$.

9.2 Pairwise term

The pairwise terms takes the form of a contrast sensitive Pott's model [32]

$$\psi_{ij}(a_i, a_j) = \begin{cases} 0 & \text{if } a_i = a_j \\ g(i,j) & \text{Otherwise} \end{cases} \tag{20}$$

where $g(i,j)$ is a contrast feature based on the ratio of the average median RCS between two adjacent superpixels. Typically it is defined as

$$g(i,j) = \theta_p + \theta_v \exp(-\theta_b c(i,j)) \tag{21}$$

Here, we define $c(i,j) = \min(\mu_i / \mu_j, \mu_j / \mu_i)$ where μ_i is the average RCS value in the magnitude domain, and $\theta_b^2 = \mathrm{VAR}\{c(i,j)\}$ [55]. The variables θ_p and θ_v are selected to give the correct tradeoff between the unary and pairwise terms.

10 Results

Using just the unary term, Figure 11 shows a labeled image of the SAR image in Figure 10, which contains data from the validation set. The labeling is done by selecting the class with highest confidence after combining all the GOF class results. The black labeled superpixels indicate regions where none of the GOF models fit very well and after Dempster/Shafer combination (6) the confidence $\hat{m}(\theta_0)$ is greater than all the other confidences, $\hat{m}(\theta_k)$ $k = 1,...,N$. There are some obvious false alarms where the hardpacked dirt is misclassified as paved and cement. Also, some of the returns from power lines (below the road) get misclassified as tree or gravel.

Using graph cuts and alpha expansion, Figure 12 shows the labeling from minimizing the CRF energy function (18). The result is an image with more spatial consistency. A large number of isolated unknown superpixels where switched over to the neighboring class majority for those areas with low contrast, but the false alarms in the unary terms still exist. To improve those one would need higher order constraints that incorporate shape and context.

11 Conclusion

Diversity in SAR imagery is very important for classification. Since we don't have polarimetric or frequency diversity we use temporal diversity. With the availability of a large amount of high resolution surveillance data collected at different times we have a unique opportunity to build on current approaches and find new ways to segment SAR images, specifically automatically identifying classes such as paved, shadow, desert, soil, etc. in SAR imagery. The phenomenon of SAR speckle requires careful consideration on using standard optical approaches for image segmentation. With the availability of a long time history of SAR imagery for a specific area, we have introduced a new SAR-derivative product based on a long term CCD. Usually the time separation between successive SAR collections for a CCD creation is on the order of hours, but we have investigated how different regions decorrelate over many days. We have found that the type of terrain often determines the rate of decorrelation. For example, the paved-road static feature maintains a high degree of coherence even after many days. However, an LCCD image by itself fails to provide sufficient discrimination between certain high desert areas, man-made objects and paved roads. Instead, if we apply probabilistic fusion to combine multiple data sources, LCCD and SAR RCS for this case, significantly reduces between-class confusion. Because p-value fusion models the class of interest and tests for fidelity to the model, it allows an unknown class, which is difficult or impossible with many machine learning approaches [7][12][26][49]. We show how to combine the output of many one-class classifiers to get a confidence for each class and the unknown class. This is important for follow-on algorithms such as a CRF.

For features, we use the first two moments of the superpixel. These features are descriptive enough to describe the distribution of different terrain pixels in a superpixel, but not as sensitive as a CDF used by the KS test. To get the combined p-value for these moment based features, we use the Z-test and p-value fusion. From these p-values, for a single class and single image product, we can fuse multiple products to get a confidence for the class of that superpixel.

To enforce spatial consistency constraints we use a CRF. A labeling is achieved by minimizing the CRF energy function using graph-cuts and alpha-expansion. The energy function is created using the multilabel confidences from the combination of one-class classifiers and contrast function based on the ratio of the average intensity for adjacent superpixels.

12 Future Work

While we investigated temporal diversity in SAR imagery, it would of interest to look at other types of diversity. The coherence values are dependent on the geometry and material. For example, at shallow grazing angles, paved roads or metal roofs may scatter all their energy away from the sensor, resulting in low coherence. On the other hand, terrain like shrubs, might give more consistent returns, as the scattering mechanism volumetric. Thus the multi-passes at different grazing angle might add another layer of diversity to the SAR terrain classification problem.

One problem with an approach using single-pol SAR is the variation of the backscatter with local grazing angle. Elevation changes can create deviations from the idealized planar surface. This causes the backscatter of the same terrain class to vary for other surfaces such ridges, ravines, peaks, and valleys. One approach is to use the local grazing angle to compensate for these elevation changes. To estimate the local grazing angle one needs an elevation map. Interferometric SAR (InSAR) is one approach, but the multiple SAR collections need to be planned with InSAR in mind. Another idea is to use lidar to get the elevation information.

Another area for future work is the selection of superpixels for training. It's difficult for an analyst to know what types of terrain are discriminated by a sensor and which are confused. Therefore unsupervised segmentation of SAR images is important even though the ultimate goal maybe a result with classes identified. One problem with supervised segmentation is the scale of the approach. Specific parameter selection and techniques tend to be very good at finding large or small regions, but not both. An attractive approach is to combine segmentations based on different parameters and algorithms to get a single overall best segmentation [4].

Acknowledgements

This work was supported by PANTHER, a Laboratory Directed Research and Development (LDRD) Project at Sandia National Laboratories. For additional information about PANTHER, please contact Kristina Czuchlewski, Ph.D., krczuch@sandia.gov.

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000 and SAND#2014-2731C.

References

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. *34,* no. *11*, pp. *2274-2282,* Nov. 2012.

[2] Y. Boykov, O. Veksler, and R. Zabin, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. *23,* no. *11,* pp. *1222-1239,* Nov. 2001.

[3] M. Bauer, "Approximation algorithms and decision making in the Dempster-Shafer theory of evidence—an empirical study," *Int. J. Approximate Reasoning,* vol. *17,* no. *2 – 3,* pp. *217-237,* Aug. – Oct. 1997.

[4] W. Brendel and S. Todorovic, "Segmentation as maximum-weight independent set", *Advances in Neural Information Processing Systems 23,* Eds: J. Lafferty and C. Williams and J. Shawe-Taylor and R. Zemel and A. Culotta, pp 307-315, 2010.

[5] R. C. Brost, W. C. McLendon III, O. Parekh, M. D. Rintoul, D. R. Strip, and D. M. Woodbridge, "A computational framework for ontologically storing and analyzing very large overhead image sets," presented at

the 3rd ACM SIGSPATIAL Int. Workshop on Analytics for Big Geospatial Data (BigSpatial-2014), Dallas, TX, USA, Nov. 4, 2014.
[6] V. V. Chamundeeswari, D. Singh, and K. Singh, "An analysis of texture measures in PCA-based unsupervised classification of SAR images," *IEEE Geosci. Remote Sens. Lett.,* vol. *6,* no. *2,* pp. *214-218,* Apr. 2009.
[7] D. Ciresan, U. Meier, J. Masci, L. M. Gambardella and J. Schmidhuber, "Flexible, High Performance Convolutional Neural Networks for Image Classification," *IJCAI*, 2011.
[8] D. A. Clausi, "Comparison and fusion of co-occurrence, Gabor and MRF texture features for classification of SAR sea-ice imagery," *Atmosphere-Ocean,* vol. *39,* no. *3,* pp. *183-194,* 2001.
[9] S.R. Cloude, "Depolarisation and scattering entropy," in *Polarisation: Applications in Remote Sensing*, Oxford, UK: Oxford University Press, 2010.
[10] S.R. Cloude and E. Pottier, "An entropy based classification scheme for land applications of polarimetric SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. *35,* no. *1,* pp. *68-78,* Jan. 1997.
[11] T. R. Crimmins, "Geometric filter for speckle reduction," *Applied Optics*, vol. *24*, no. *10*, pp. *1438-1443*, May 1985
[12] C. Corte and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. *20*, no. *3*, pp. *273-297*, Sept. 1995. ??
[13] Y. Cui, Y. Yamaguchi, J. Yang, H. Kobayashi, S. Park, and G. Singh, "On complete model-based decomposition of polarimetric SAR coherency matrix data," *IEEE Trans. Geosci. Remote Sens.*, vol. *52*, no. *4*, pp. *1991-2001*, Apr. 2014.
[14] J. Davis and A. Tyagi, "Minimal-latency human action recognition using reliable-inference," Image and Vision Computing, vol. *24*, no. *5,* pp. *455-472*, May 2006.
[15] A. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Ann. Math. Statist.*, vol. *38*, no. *2,* pp. *325-339,* 1967.
[16] H. Deng and D. A. Clausi, "Unsupervised segmentation of synthetic aperture Radar sea ice imagery using a novel Markov random field model," *IEEE Trans. Geosci. Remote Sens.,* vol. *43,* no. *3,* pp. *528-538,* Mar. 2005.
[17] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification,* New York, NY: John Wiley & Sons, 2012.
[18] M. Evans, N. Hastings, and B. Peacock, *Statistical Distributions,* New York, NY: John Wiley & Sons, 1993.
[19] B. Fulkerson, A. Vedaldi, and S. Soatto, "Class segmentation and object localization with superpixel neighborhoods," presented at the 2009 IEEE 12th Int. Conf. Comput. Vis., Kyoto, Jpn., 2009, pp. *670-677*.
[20] A. Ferretti, et al., "InSAR Principles: Guidelines for SAR interferometry Processing Interpretation," TM-19, Feb 2007.
[21] V. S. Frost, J. A. Stiles, K. S. Shanmungan, and J. C. Holtzman, "A Model for Radar Images and its Application for Adaptive Digital Filtering of Multiplicative Noise," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. *PAMI-4,* no. *2,* pp. *157-165,* Mar. 1982.
[22] L. Gagnon and A. Jouan, "Speckle filtering of SAR images: a comparative study beteen complex-wavelet-based and standard filters," *Proc SPIE*, vol. *3169*, pp. *80-91*, Aug. 1997.
[23] L. Gan, Y. Wu, M. Liu, P. Zhang, H. Ji, and F. Wang, "Triplet Markov fields with edge location for fast unsupervised multi-class segmentation of synthetic aperture radar images," *IET Image Proc.,* vol. *6,* no. *7,* pp. *831-838,* Oct. 2012.
[24] G. Gao, "Statistical modeling of SAR images: a survey." *Sensors,* vol. *10,* no. *1,* pp. *775-795,* 2010.
[25] J. Gibson and J. L. Melsa, *Introduction to Nonparametric Detection with Applications*, New York, NY, USA: IEEE Press, 1996, pp. *25*.
[26] G. E. Hinton, S. Osindero, and Y. The, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. *18,* no. *7,* pp. *1527-1554*, *2006.*
[27] R. Hogg and A. Craig, *Introduction to Mathematical Statistics*, 4th ed. New York, NY, USA: Macmillan, 1978.
[28] 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*, New York, NY, USA: Springer, 1996.
[29] M. W. Koch, "One-class multiple-look fusion: a theoretical comparison of different approaches with examples from infrared video," in *2013 IEEE Conf. on Comput. Vis. Pat. Recogn. Workshops,* Portland, OR, Jun. 2013, pp. *342-347*.
[30] M. Koch, G. Haschke, and K. Malone, "Classifying acoustic signatures using the sequential probability ratio test," *Sequential Analysis J.*, vol. *23*, no. *4*, pp. *557-583*, 2004.
[31] M. Koch, M. Moya, J. Chow, J. Goold, and R. Malinas, "Road Segmentation Using Multipass Single-Pol Synthetic Aperture Radar Imagery," in *2015 IEEE Conf. on Comput. Vis. Pat. Recogn. Workshops,* Boston, MA, Jun. 2015, pp. *151-160*.
[32] P. Kohli, L. Ladický, P.H.S Torr, "Robust higher order potentials for enforcing label consistency," in *2008 IEEE Conf. on Comput. Vis. Pat. Recogn., Anchorage, AK, Jun. 2008,* pp.*1-8*.
[33] A. Kolmogorov, "Sulla determinazione empirica di una legge di distribuzione," *G. Ist. Ital. Attuari,* vol. *4,* pp. *83-91,* 1933.
[34] D. T. Kuan, A. A. Sawchuk, and P. Chavel, "Adaptive Restoration of Images with Speckle," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. *35,* no. *3,* pp. *373-383,* Mar. 1987.
[35] J. Lafferty, A. McCallum, and F. Pereira "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. 18th Int. Conf. Machine Learning,* 2001, pp. *282-289*.

[36] J. S. Lee, “Digital image enhancement and noise filtering by use of local statistics,” *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. *PAMI-2,* no. *2,* pp. *214-218,* Mar. 1980.
[37] J. S. Lee, "Speckle analysis and smoothing of synthetic aperture radar images," *Computer Graphics and Image Processing*, vol. *17,* no. *1,* pp. *24-32,* Sept. 1981.
[38] J. S. Lee and T. L. Ainsworth, “The effect of orientation angle compensation on coherency matrix and polarimetric target decompositions,” *IEEE Trans. Geosci. Remote Sens.*, vol. *49,* no. *1,* pp. *53-64,* Jan. 2011.
[39] J. S. Lee and I. Jurkevich, “Segmentation of SAR images,” *IEEE Trans. Geosci. Remote Sens.*,” vol. *27,* no. *6,* pp. *674-680,* Nov. 1989.
[40] J. S. Lee and E. Pottier, *Polarimetric Radar Imaging: From Basics to Applications*, Boca Raton, FL, USA: CRC Press, 2002, ch. 7.
[41] A. Levinshtein, A. Stere, K. N. Kutulakos, D. Fleet, S. J. Dickinson, and K. Siddiqui, “Turbopixels: Fast superpixels using geometric flows,” *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. *31,* no. *12,* pp. *2290-2297,* Dec. 2009.
[42] S. P. Lloyd, “Least squares quantization in PCM,” *IEEE Trans. Inf. Theory,* vol. *28,* no. *2,* pp. *129-137,* Mar. 1982.
[43] A. Lopes, R. Touzi, and E. Nezry, “Adaptive speckle filters and scene heterogeneity,” *IEEE Trans. Geosci. Remote Sens.,* vol. *28,* no. *6,* pp. *992-1000,* Nov. 1990.
[44] M. Moya and D. Hush, “Network constraints and multi-objective optimization for one-class classification,” *Neural Networks*, vol. *9,* no. *3,* pp. *463-474,* Apr. 1996.
[45] M. M. Moya, M. W. Koch, D. N. Perkins, R. D. West, “Superpixel segmentation using multiple SAR image products,” in *Radar Sensor Technology XVIII (DS 108), SPIE DSS Symp,* Baltimore, MD, May 2014, pp. .
[46] K. Krishnamoorthy, T. Mathew, and S. Mukherjee, “Normal-based methods for a gamma distribution: prediction and tolerance intervals and stress-strength reliability,” *Technometrics*, vol. *50,* no. *1,* pp. *69–78,* 2008.
[47] P. Maillard, D. Clausi, and H. Deng, “Operational map-guided classification of SAR sea ice imagery,” *IEEE Trans. Geosci. Remote Sens.,* vol. *43,* no. *12,* pp. *2940-2951,* Dec. 2005.
[48] K. Murphy, and B. Myors, *Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests*, New Jersey, USA: Lawrence Erlbaum Associates, 1998.
[49] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes 3rd Editions: The Art of Scientific Computing*, New York, NY, USA: Cambridge University Press, 2007, pp. *736-738*.
[50] X. Ren and J. Malik, “Learning a classification model for segmentation,” in *Proc. 9th IEEE Int. Comput. Vis.,* Nice, France, 2013, vol. *1,* pp. *13-16*.
[51] D. E. Rumelhart and J. McClelland, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge: MIT Press, 1986.
[52] J. S. Salazar, “Detection Schemes for Synthetic Aperture Radar Imagery Based on a Beta Prime Statistical Model,” Ph.D. dissertation, ECE, UNM, Albuquerque, NM 1999.
[53] J. S. Salazar, D. R. Hush, M. W. Koch, R. J. Fogler, and L. D. Hostetler, “Statistical modeling of target and clutter in single-look non-polorimetric SAR imagery,” *Proc. Int. Conf. Signal Image Processing*, Las Vegas, NV, USA, Oct. 1998.
[54] G. Shafer, *A Mathematical Theory of Evidence*, Princeton, NJ, USA: Princeton University Press, 1976.
[55] J. Shotton, J. Winn, C. Rother, and A. Criminisi, “TextonBoost: Joint Appearance, Shape and Context Modeling for Multi-Class Object,” in *Computer Vision – ECCV*, Springer Berlin Heidelberg, 2006, pp. *1-15*.
[56] K. Simonson, “Probabilistic fusion of ATR results,” Sandia, Albuquerque, NM, USA, Sandia Report, SAND98-1699, August 1998.
[57] G. Singh, Y. Yamaguchi and S.E. Park, “General four-component scattering power decomposition with unitary transformation of coherency matrix,” *IEEE Trans. Geosci. Remote Sens.*, vol. *51,* no. *5,* pp. *3014-3022,* May 2013.
[58] N. Smirnov, "Table for estimating the goodness of fit of empirical distributions,” Ann. Math. Statist., vol. *19*, no. *2,* pp. *279-281,* 1948.
[59] X. Su, C. He, Q. Feng, X. Deng and H. Sun, “A supervised classification method based on conditional random fields with multiscale region connection calculus model for SAR image,” *IEEE Geosci. Remote Sens. Lett.,* vol. *8,* no. *3,* pp. *497-501,* May 2011.
[60] C. Tison, J.-M. Nicolas, F. Tupin, and H. Maitre, “A new statistical model for Markovian classification of urban areas in high-resolution SAR images,” *IEEE Trans. Geosci. Remote Sens.,* vol. *42,* no. *10,* pp. *2046-2057,* Oct. 2004.
[61] F. T. Ulaby and M. C. Dobson, *Handbook of Radar Scattering Statistics for Terrain*, Norwood, MA, USA: Atrech House, 1989.
[62] F. Voorbraak, "A computationally efficient approximation of Dempster-Shafer theory," *Int. J. Man-Machine Studies,* vol. *30,* no. *5,* pp. *525-536,* May 1989.
[63] E. Wilson and M. Hilferty, “The Distribution of Chi-Squares,” *Proc. Nat. Academy Sci.*, vol. *17,* no. *12,* pp. 684-688, Dec. 1931.
[64] P. J. Werbos, *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*, 1975.

[65] Y. Wu, X. Wang, P. Xiao, L. Gan, C. Y. Liu, and M. Li, "Fast algorithm based on triplet Markov fields for unsupervised multi-class segmentation of SAR images," *Sci. China Info. Sci.,* vol. *54,* no. *7,* pp. *1524-1533*, Jul. 2011.
[66] D. Xiang, T. Tang, L. Zhao, and Y. Su, "Superpixel generating algorithm based on pixel intensity and location similarity for SAR image classification," *IEEE Geosci. Remote Sens. Lett.,* vol. *10,* no. *6,* pp. *1414-1418,* Nov. 2013.
[67] Y. Yamaguchi, T. Moriyama, M. Ishido, and H Yamada, "Four-component scattering model for polarimetric SAR image decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. *43,* no. *8,* pp. *1699-1706,* Aug. 2005.
[68] W. Yang, D. Dai, B. Triggs, and G.-S. Xia, "SAR-based terrain classification using weakly supervised hierarchical Markov aspect models," *IEEE Trans. Image Process.,* vol. *21,* no. *9,* pp. *4232-4243,* Sept. 2012.
[69] W. Yang, T. Zou, D. Dai, and Y. Shuai, "Supervised land-cover classification of TerraSAR-X imagery over urban areas using extremely randomized clustering forests," in *2009 Joint Urban Remote Sensing Event,* Shanghai, China, pp. *1-6*.
[70] H. Yu, X. Zhang, S. Wang, and B. Hou, "Context-Based Hierarchical Unequal Merging for SAR Image Segmentation," IEEE Trans. Geosci. *Remote* Sens., vol. *51,* no. *2,* pp. *995-1009*, Feb 2013.

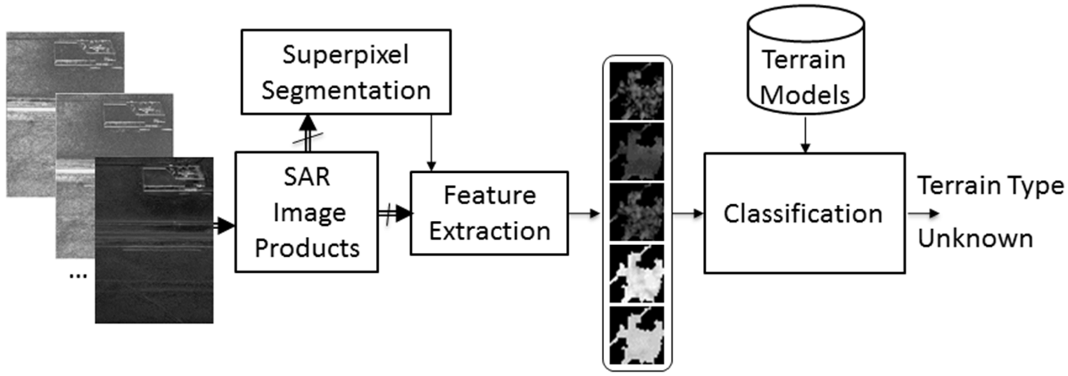
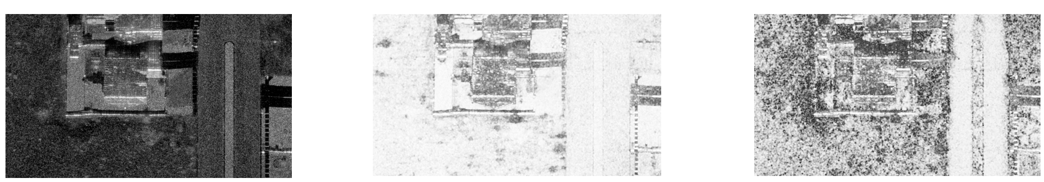


Figure 1: Block diagram of SAR terrain classification approach.

(a) (b)

Figure 2: Example SAR and multilook images. (a) Calibrated SAR image. (b) Four look subaperture multilook.



(a) (b) (c)

Figure 3: Example SAR product images. (a) Median RCS (b) Mean CCD for Δt=1 day. (c) Median-LCCD for Δt=14 days.

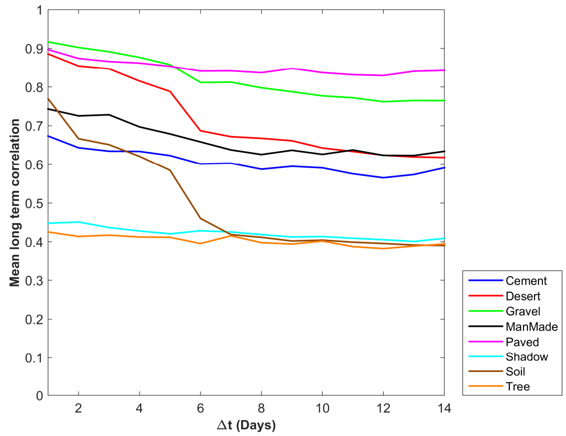


Figure 4: Mean LCCD for $\Delta t = 1\,\text{to}\,14$ days and different types of static-features (best viewed in color).

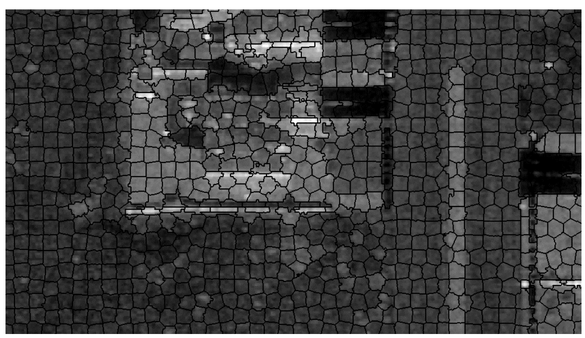


Figure 5: Example superpixel image.

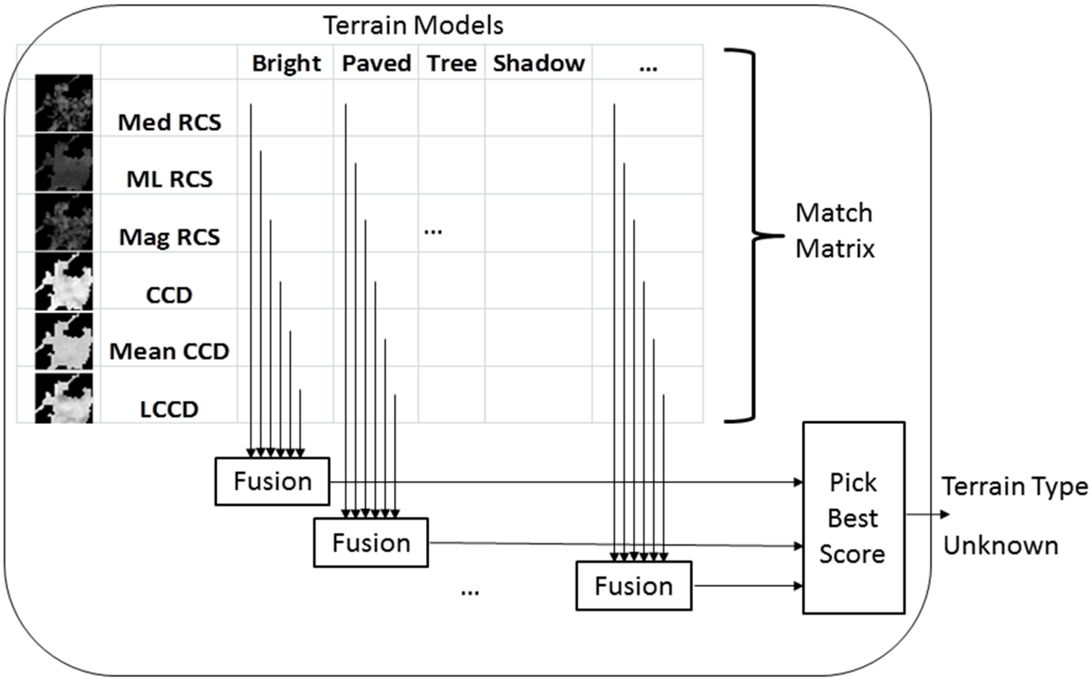


Figure 6: Classification approach.

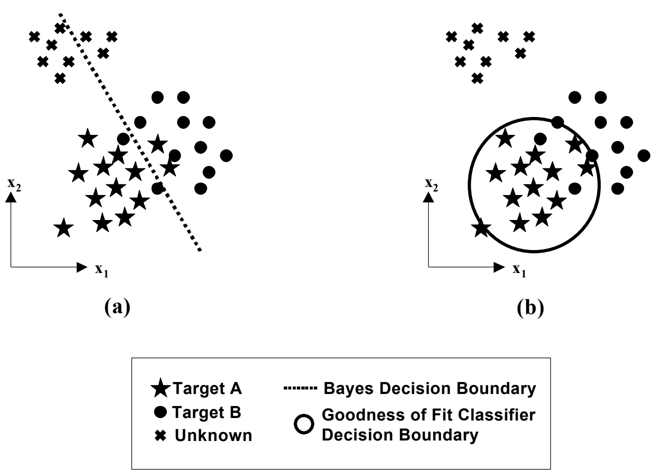


Figure 7: Comparison of Bayes and goodness of fit (GOF) classifiers. (a) Bayes classifier. (b) GOF classifier.

Table 1. Example one-class combination to get multiclass confidence.

	Class 1		**Class 2**		**Class 3**		**Confidence**			
	$m(\theta_1)$	$m(\bar{\theta}_1)$	$m(\theta_2)$	$m(\bar{\theta}_2)$	$m(\theta_3)$	$m(\bar{\theta}_3)$	$\hat{m}(\theta_0)$	$\hat{m}(\theta_1)$	$\hat{m}(\theta_2)$	$\hat{m}(\theta_3)$
1	0	91	0	99	0	95	87	8	1	4
2	85	0	0	90	0	85	13	84	1	2
3	85	0	50	0	0	99	10	69	21	0
4	60	0	50	0	0	99	18	45	36	0

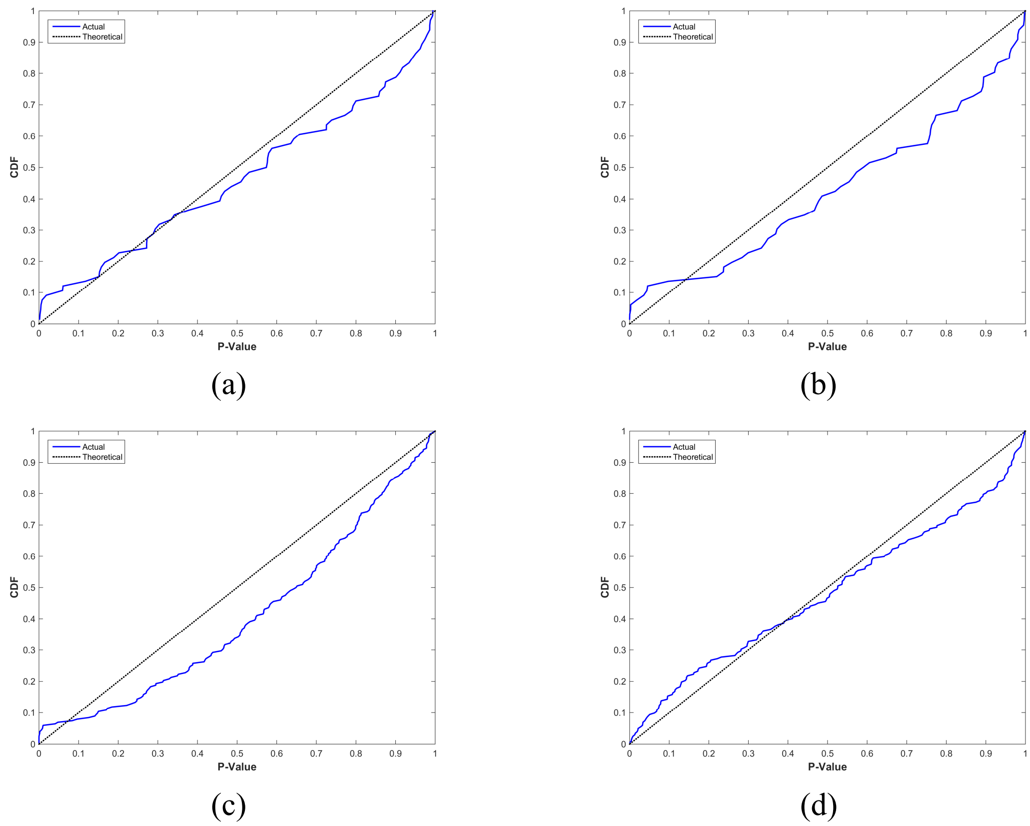


Figure 8. CDF of GOF p-values for different classes and products. Blue line shows the actual CDF and the dotted black line shows the theoretical uniform distribution. (a) and (b) tree class. (c) and (d) man-made bright class. (a) and (c) median RCS. (b) mean CCD. (d) 21 day LCCD.

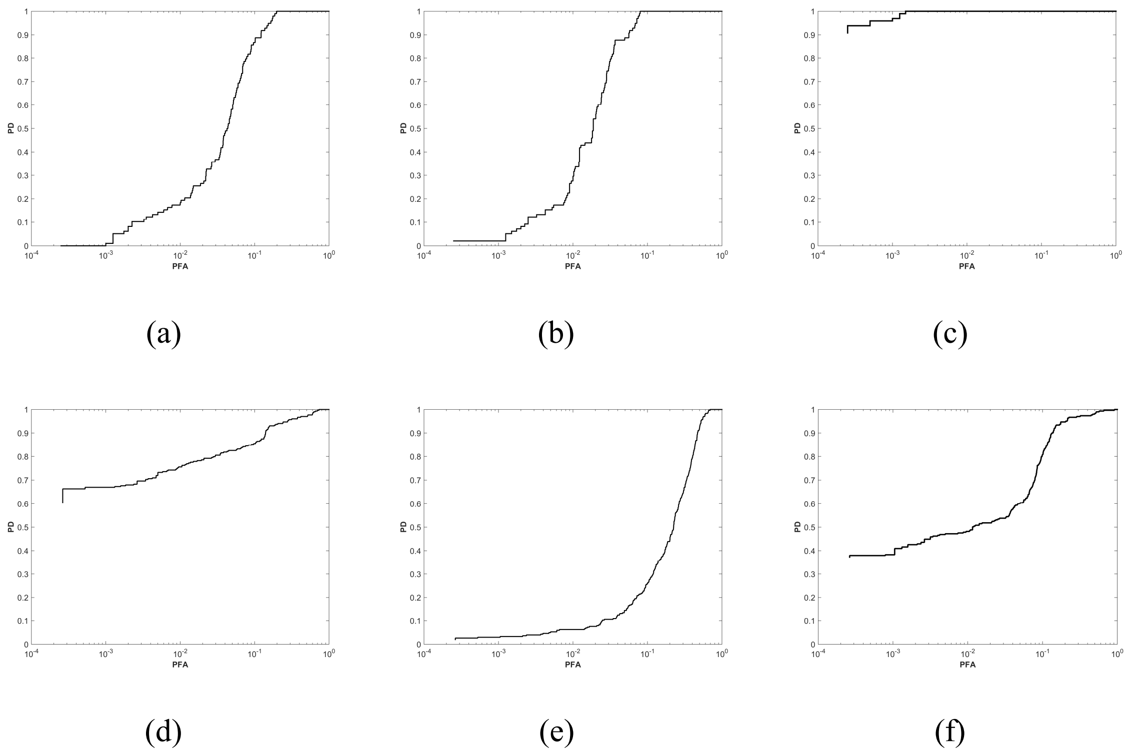


Figure 9. ROC curves for different classes and products. (a)-(c) is for the tree class with (a) the median RCS, (b) the mean CCD, an (c) the fusion of the two products from (a) and (b). (d)-(f) is the for the man-made class. (d) the median RCS, (e) the 21 day LCCD and (f) the fusion of the two products from (d) and (e).

Figure 10. Example SAR image

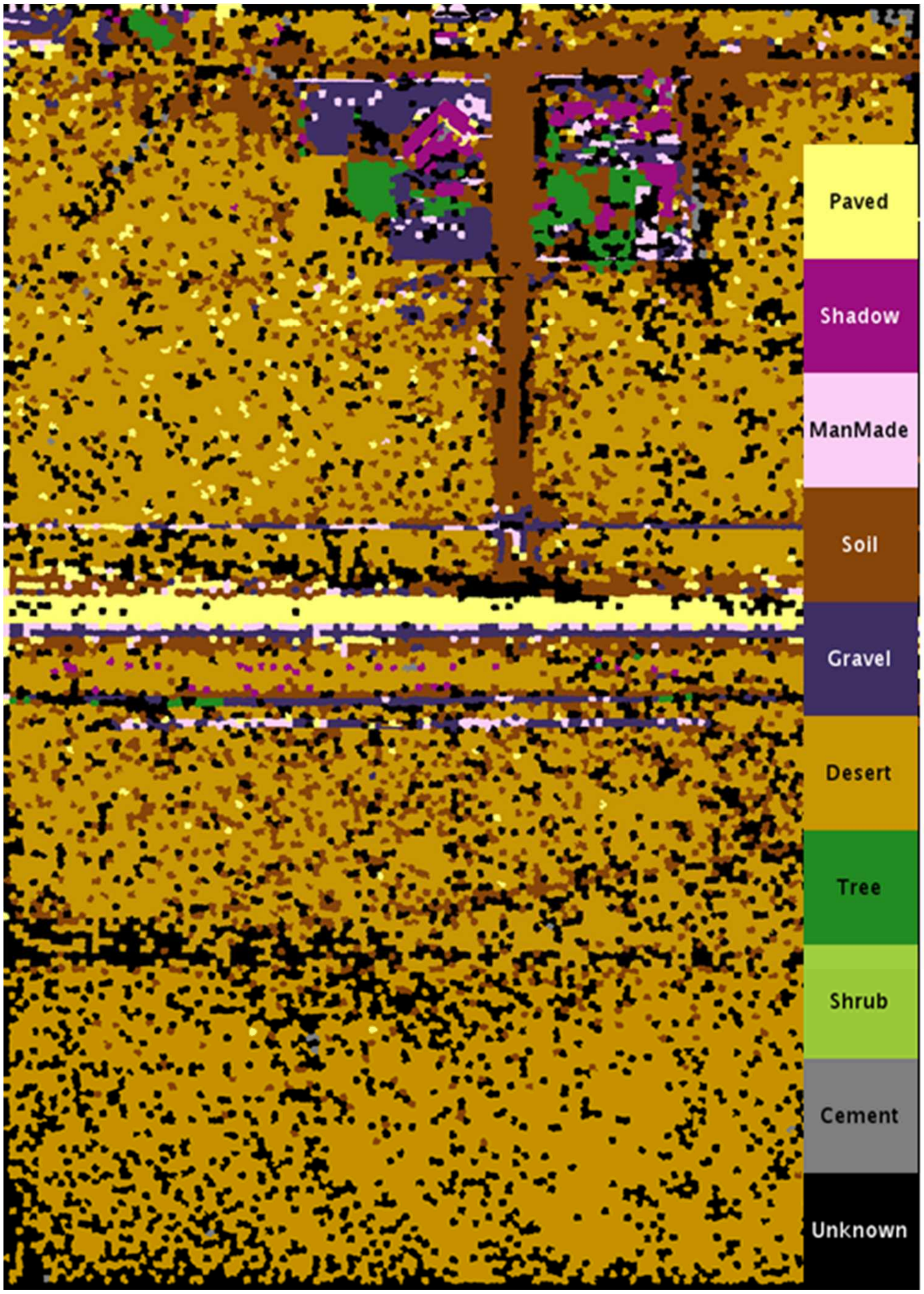


Figure 11. Labeling before applying the CRF (best viewed in color)

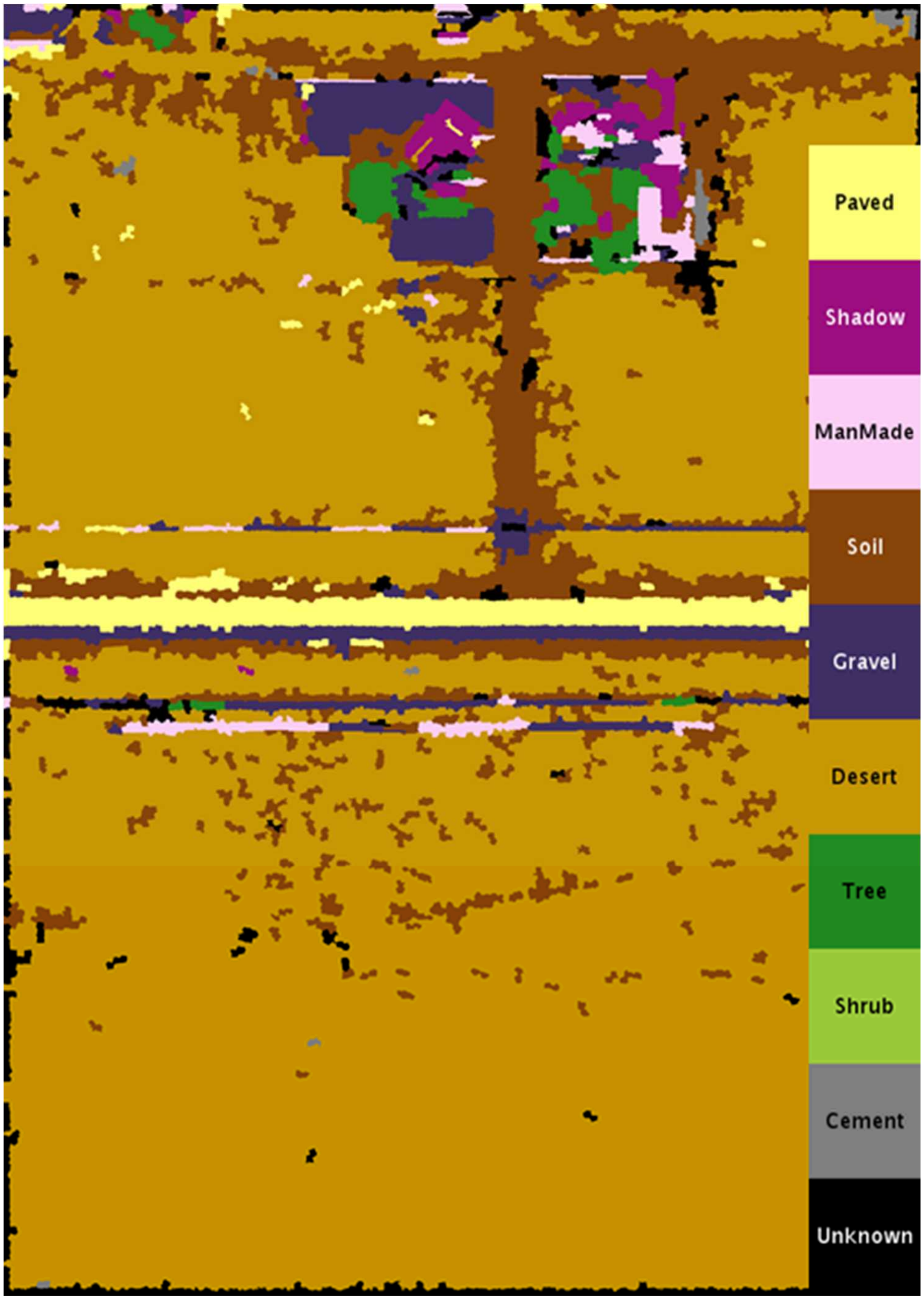


Figure 12. After applying CRF (best viewed in color).