Image Segmentations for Through-the-Wall Radar Target Detection

Cher Hau Seng, Student Member, IEEE, Moeness G. Amin, Fellow, IEEE, Fauzia Ahmad, Senior Member, IEEE, and Abdesselam Bouzerdoum, Senior Member, IEEE

Abstract

Detection of stationary targets using pixel-wise likelihood ratio test (LRT) detectors has been recently proposed for through-the-wall radar imaging (TWRI) applications. In this paper, we employ image segmentation techniques, in lieu of LRT, for target detection in TWRI. More specifically, the widely used between class variance thresholding, maximum entropy segmentation, and K-means clustering are considered to aid in removing the clutter, resulting in enhanced radar images with target regions only. For the case when multiple images of the same scene are available through diversity in polarization and/or vantage points around a building structure, we propose to use image fusion, following the image segmentation step, to generate an enhanced composite image. In particular, additive, multiplicative, and fuzzy logic fusion techniques are considered. Performance of the segmentation and fusion schemes is evaluated and compared to that of the assumed LRT detector using both EM modeling and real data collected in a laboratory environment. The results show that, although the principles of segmentation and detection are different, the segmentation techniques provide either comparable or improved performance over the LRT detector. Specifically, in the cases considered, the maximum entropy segmentation produces...
the best results for detection of targets inside building structures. For fusion of multiple segmented images of the same scene, the fuzzy logic fusion outperforms the other methods.

Index Terms

Image segmentation, image fusion, target detection, through-the-wall radar imaging

I. INTRODUCTION

Imaging of building interiors has been a subject of interest in many applications related to rescue missions, homeland security, and defense [1], [2], [3], [4], [5], [6], [7]. Indoor images are typically characterized by the presence of both spatially extended targets, like exterior and interior walls, and compact, point-like targets, such as humans. Also, near-field operations give rise to point spread functions that vary in range and cross-range. Accordingly, the same target can have different spatial distribution, depending on its position.

Detection of stationary targets in through-the-wall radar imaging (TWRI) using statistical detectors based on likelihood ratio tests (LRT) has been discussed in [8], [9]. Specifically, a Neyman-Pearson (NP) test was used in [8] to detect targets in indoor radar images by defining pixel-wise null and alternative hypotheses, coupled with a user-defined false alarm rate (FAR). However, in this test, the exact statistics of the radar images need to be known a priori. As this information is target and scene dependent, the NP test was therefore extended in [9] to iteratively adapt the test parameters to the radar image statistics. In order to improve and optimize the parameter estimates, morphological filtering in the image domain was introduced as a pre-processing stage in [10]. By adapting the target detector’s parameters to the changing characteristics of the pre-processed radar images, a more robust detection can be produced. To date, both NP and Bayesian tests have been employed for the detection and fusion of multi-view and multi-polarization through-the-wall radar (TWR) images [9], [11]. Generally, the statistical detectors that incorporate the LRT generate a binary mask that depicts the target locations in the image.

Although the LRT approach has been successfully applied to some scenes and several scenarios, it assumes particular probability density functions (PDFs) for all targets in the scene. The assumed PDFs may prove unsuitable for complex scenes that are acquired using real data or data from numerical EM modeling. The need to define appropriate PDFs for all targets and
clutter in the image, and to specify an appropriate FAR by the operator presents a shortcoming of the LRT detector, since in most cases, neither the PDFs nor the FAR is known a priori. In this paper, we propose the use of image segmentation methods as an alternative to target detection. In lieu of the LRT, we apply image segmentation methods that exploit information obtained from image intensities and histograms to separate targets from clutter, and thereby enhance the image quality.

There exist many different types of image segmentation methods [12], image thresholding methods and region based methods, to name a few. Region-based methods, such as the watershed, are not suitable for TWR images as the radar returns are very sparse and target regions tend to be very small. One of the most basic thresholding methods is to set a threshold at the valley of a bimodal intensity histogram [13]. However, TWR image histograms generally do not exhibit a traceable valley. There have been some automatic threshold selection techniques proposed in order to overcome this difficulty. Such techniques select an optimal threshold by a discriminant criterion, so as to maximize the separability of the resultant classes. We propose to examine the three most commonly used automatic image thresholding methods, namely, the between-class variance (BCV) thresholding [14], entropy-based segmentation [15], and K-means clustering [16] for TWR image segmentation. The segmentation process associated with these techniques produces threshold values, thus generating binary masks that are similar to that of the LRT detector. Using the generated binary image, the original image is masked, producing an enhanced image with the target regions only.

For the detection of targets in multiple images, the proposed image segmentation-based target detector involves a two-step process that predicates on the availability of TWR images acquired from different viewing angles and for different polarizations. Instead of performing the fusion and detection at the same time, like in the case of the LRT detector, image segmentation techniques are first applied to individual radar images to separate each image into several regions. The segmented results are then fused using pixel-wise addition, multiplication, or through the use of fuzzy logic [17]. The latter has proven to be a valuable tool since it deals with the fusion problem at a local level, i.e., different parts of the image can be fused differently, depending on the relative presence of the target and clutter. Reversing this sequence of operation, by applying fusion prior to segmentation, provides inferior results and does not fully utilize the potential enhancement of individual images before combinations. It is noted that image segmentation
methods have been applied for object detection in Synthetic Aperture Radar (SAR) aerial images [13], [18], [19], [20] and Ground Penetrating Radar (GPR) images [21]. They have also been considered as a pre-processing stage for change detection [22], [23] and object classification [24], [25], [26] in radar images. It is also noted that a similar segmentation-fusion approach has been proposed in [27] to improve the classification of SAR images. However, our work differs by applying the segmentation and fusion methods as an alternative to the LRT detector for both single and fused images, in the specific area of TWRI.

In this paper, we investigate the performances of the LRT detector and the proposed method for indoor radar imaging, both with and without image fusion. Both target detection schemes are evaluated and compared using real two-dimensional (2D) polarimetric and multi-view images collected at the Radar Imaging Lab, Center for Advanced Communications, Villanova University. These methods are also applied to numerical EM modeling data. For both real and modeled data, different scenarios that are representative of a variety of possible indoor scenes are constructed so as to strengthen our findings and conclusions. In order to compare methods within the same framework, the LRT detector is cast as another form of image segmentation method with a corresponding threshold.

The performance is assessed in terms of image enhancements after the separation of target from clutter in individual images. The results show that in most cases considered, the entropy-based segmentation technique outperforms other image segmentation techniques, including the LRT method, by providing higher clutter suppression, while successfully maintaining the target regions. The BCV thresholding and K-means clustering methods provide similar results to those of the entropy-based segmentation when considering target regions; however, their respective clutter levels are much higher. Nevertheless, the proposed segmentation-based detectors are more advantageous than the LRT detector since they do not require a predefined PDF and FAR.

As for multiple image target detection, the performance of the proposed detection scheme is compared with that of the LRT detector that jointly utilizes multi-view and multi-polarization signatures. The corresponding results show that the entropy-based segmentation method, coupled with the pixel-wise fuzzy logic fusion, outperforms the LRT detector by providing higher clutter suppression, while keeping the number of missed detections relatively low. In essence, the entropy-based segmentation method can be generally viewed as most suitable for target detection in TWRI.
The remainder of the paper is organized as follows. Section II details the TWR image formation process, where the wideband delay-and-sum beamforming is reviewed. Section III describes the LRT target detection method proposed in [9]. The alternative target detection method based on image segmentation techniques is then discussed in Section IV. Section V evaluates the performance of the LRT detector and image segmentation-based methods using real and numerical modeling data, and Section VI concludes the paper.

II. BEAMFORMING

In this section, we present the fundamental equations describing the delay-and-sum beamformer, considered in this paper, for imaging through walls. We consider a 2D SAR system in which a single antenna at one location transmits the signal and receives the radar return, and then moves to the next location and repeats the same operation in a plane parallel to the front wall [7], [28].

Consider the scenario where the wall is located in the $xy$-plane, and has a thickness $d$ and a dielectric constant $\varepsilon$. Assume that there are $M$ monostatic antenna locations, with a standoff distance $z_{off}$ from the wall. The region to be imaged is located beyond the wall along the positive $z$-axis, as shown in Fig. 1. Let the transceiver, placed at the $m$-th location $x_{tm} = (x_{tm}, y_{tm}, -z_{off})$, illuminate the scene with a wideband signal $s(t)$. For the case of a single point target located at $x_p = (x_p, y_p, z_p)$, the signal measured by the $m$-th transceiver is given by $a(x_p)s(t - \tau_{m,p})$, where $a(x_p)$ is the complex reflectivity of the point target and $\tau_{m,p}$ is the propagation delay encountered by the signal as it travels from the $m$-th transceiver to the target at $x_p$, and then back to the same transceiver. The delay $\tau_{m,p}$ is given by

$$\tau_{m,p} = \frac{2l_{mp,\text{air},1}}{c} + \frac{2l_{mp,\text{wall}}}{v} + \frac{2l_{mp,\text{air},2}}{c},$$

where $c$ is the speed of light and $v = c/\sqrt{\varepsilon}$ is the signal propagation speed through the wall. The variables $l_{mp,\text{air},1}, l_{mp,\text{wall}}$ and $l_{mp,\text{air},2}$ represent the traveling distances of the signal before, through and beyond the wall, respectively, from the $m$-th transceiver to the target at $x_p$. This process is evaluated for each transceiver location until all $M$ locations have been exhausted.

Although the corresponding $M$ outputs can be processed to generate a three-dimensional (3D) image of the scene, we restrict ourselves to B-scan (cross-range vs. downrange at a fixed height) images. This is because the images presented in this paper are either B-scan images or crossrange
vs. downrange images obtained with one-dimensional (1D) antenna arrays. The region of interest at height \( \bar{y} \) is divided into a finite number of pixels in downrange and cross-range, represented by the \( z \)- and \( x \)-coordinates, respectively. The complex composite signal corresponding to the image of the pixel located at \( x_q = (x_q, \bar{y}, z_q) \) is obtained by applying time delays and weights to the \( M \) measurements, and then summing the results. The output for a single target case is given by

\[
 r_q(t) = \sum_{m=1}^{M} w_m a(x_p) s(t - \tau_{m,p} + \tau_{m,q}).
\]  

(2)

Here, \( w_m \) and \( \tau_{m,q} \) are, respectively, the weight and the focusing delay applied to the output of the \( m \)-th transceiver. The focusing delay is given by (1) with the target voxel subscript \( p \) replaced by the focusing pixel subscript \( q \). The complex amplitude image value for the pixel located at \( x_q \) is obtained by passing the signal \( r_q(t) \) through a filter matched to the transmitted pulse and sampling the output of the filter at time \( t = 0 \) as follows:

\[
 I(x_q) = \sum_{m=1}^{M} w_m a(x_p) s(t - \tau_{m,p} + \tau_{m,q}) * h(t)|_{t=0},
\]  

(3)

where \( h(t) = s^*(-t) \) is the impulse response of the matched filter. The process described is performed for all \( N \) pixels in the region of interest at height \( \bar{y} \) to generate the composite image of the scene. The general case of multiple targets can be obtained by the superposition of target reflections.

It is noted that the output of the delay-and-sum beamformer is a complex valued image. The magnitude image, corresponding to the complex amplitude image, is given by

\[
 \bar{I}(x_q) = |I(x_q)|.
\]  

(4)

For target detection in the image domain, we deal with the magnitude image only.

### III. Likelihood Ratio Tests Detector

In this section, we review the general image domain-based framework for statistical target detection based on the LRT, which utilizes multiple images corresponding to different viewing angles and/or polarization. Specifically, we consider the iterative version of the pixel-wise NP detector introduced in [9] that adapts the test parameters to the radar image statistics.
Let \( \tilde{I}_j, j = 1, 2, \ldots, J \) be the set of acquired magnitude images corresponding to a total of \( J \) viewing angles and/or polarization. The pixel-wise NP test is given as
\[
\prod_{j=1}^{J} \frac{P_r(\tilde{I}_j|H_1)}{P_r(\tilde{I}_j|H_0)} \geq \gamma, \tag{5}
\]
where \( H_0 \) and \( H_1 \) denote, respectively, the null (target absent) and alternative (target present) hypothesis. The functions \( P_r(\cdot|H_0) \) and \( P_r(\cdot|H_1) \) are the conditional PDFs under the null and alternative hypothesis, respectively, and the parameter \( \gamma \) is the likelihood ratio threshold, which can be obtained by specifying a desired FAR, \( \alpha \), as
\[
\alpha = \int_{\gamma}^{\infty} P_\ell(\ell|H_0) d\ell, \tag{6}
\]
where \( P_\ell(\ell|H_0) \) denotes the distribution of the likelihood ratio under the null hypothesis.

Let \( \hat{\theta}^0_{H_0} \) and \( \hat{\theta}^0_{H_1} \) denote the initial estimates of the parameter vectors \( \theta_{H_0} \) and \( \theta_{H_1} \) describing the PDFs under \( H_0 \) and \( H_1 \), respectively. Given a FAR \( \alpha \), a binary image \( B^1_{NP} \), where superscript 1 represents the first iteration, can be obtained by evaluating (5). In order to enhance and optimize the estimation of the noise and test PDF parameters, morphological filtering is employed to obtain the binary image \( B^1_{MF} \) (see [10] for details). This image can be used as a mask on the original set of images in order to obtain the revised parameter estimates \( \hat{\theta}^1_{H_0} \) and \( \hat{\theta}^1_{H_1} \). These revised parameters are then fed back to the NP test to obtain an improved detection result. The iteration stops when convergence is achieved. Figure 2 shows the block diagram of the iterative target detection approach.

It is noted that the final output of the LRT detector described above is a single binary image that indicates the presence of the targets. For more detailed description of the LRT detector, the reader is referred to [8], [9], [10], [11].

IV. IMAGE SEGMENTATION-BASED TARGET DETECTION

Unlike the LRT detector, which performs simultaneous detection and fusion of a set of magnitude images, the proposed method is a two-step process. First, an image segmentation technique is applied to the individual input images. Then, the segmented images are fused to generate an enhanced single composite intensity image that has high target intensities and low clutter and noise levels. Figure 3 shows the block diagram of the proposed image segmentation-based target detector.
It should be noted that when there is only one input image, only the image segmentation method will be applied. For the case of multiple input images, both image segmentation and image fusion methods will be applied for detection.

A. Image Segmentation

Image segmentation based on intensity and histogram information is a simple technique which involves the basic assumption that the objects and the background in the sensed image have distinct gray level distributions [29]. Since objects in remotely sensed imagery are often homogeneous, threshold values separating two or more regions in the gray level histogram can be obtained. Threshold selection methods can be classified into two groups, namely, global methods and local methods. A global thresholding technique separates the entire image into target and background regions with a single threshold value, whereas local thresholding methods partition the given image into a number of sub-images and determine a threshold for each of the sub-images separately. As global thresholding methods are computationally less intensive, they have been more popular for radar image analysis [30]. In this paper, we consider two of the commonly applied global thresholding methods, namely, BCV thresholding [14] and entropy-based segmentation [15], as candidates for the image segmentation step of the proposed detection scheme. Since image segmentation can also be viewed as the partitioning of the observed intensities into groups, we also consider the application of K-means clustering [16].

Consider the histogram of an input image as a discrete PDF $\rho(i)$:

$$\rho(i) = \frac{f_i}{N}$$

with $\rho(i) \geq 0$ and $\sum_{i=0}^{L-1} \rho(i) = 1$, where $f_i$ is the frequency of intensity level $i$ and $N$ is the total number of pixels in the image. Each pixel in the image assumes an intensity level from the set $\{0, 1, ..., L - 1\}$, where $L$ denotes the number of intensity levels or histogram bins.

1) Between-Class Variance Thresholding: The BCV thresholding method segments an image into two regions by determining a threshold value $T_{BCV}$ that maximizes the sum of class variances:

$$T_{BCV} = \arg \max_d \left\{ p_{r1}(d) [m_{r1}(d) - m_j]^2 + p_{r2}(d) [m_{r2}(d) - m_j]^2 \right\},$$

(8)
where $m_i$ is the mean image intensity, $r_1$ and $r_2$ are the two regions of the image histogram relative to the intensity level $d$, $p_{r_1}(d)$ and $p_{r_2}(d)$ are the respective region probabilities, which are expressed as

$$
p_{r_1}(d) = \sum_{i=0}^{d} \rho(i) \tag{9}$$
$$
p_{r_2}(d) = \sum_{i=d+1}^{L-1} \rho(i). \tag{10}$$

and $m_{r_1}(d)$ and $m_{r_2}(d)$ are the means of the respective regions, which are given by

$$
m_{r_1}(d) = \sum_{i=0}^{d} \frac{i \cdot \rho(i)}{p_{r_1}(d)} \tag{11}$$
$$
m_{r_2}(d) = \sum_{i=d+1}^{L-1} \frac{i \cdot \rho(i)}{p_{r_2}(d)}. \tag{12}$$

All values of $d = 1, 2, ..., L - 2$ are considered and the corresponding functions (8 - 12) are evaluated. The intensity value, $d$, that produces the maximum sum of the class variances is chosen as the threshold value $T_{BCV}$.

2) Entropy-Based Segmentation: Similar to the BCV method, the entropy-based segmentation decides on the threshold value in an iterative fashion. Instead of maximizing the sum of class variances, the entropy-based segmentation maximizes the sum of class entropies. Based on the information derived from the image histogram, the entropy of two regions is maximized using the following equation:

$$
T_H = \arg \max_d \{ H_{r_1}(d) + H_{r_2}(d) \}, \tag{13}
$$

where $H_{r_1}(\cdot)$ and $H_{r_2}(\cdot)$ are the respective region entropies. Let $p_i$ be the probability of intensity level $i$ and $P_d = \sum_{i=0}^{d} p_i$ be the total probability. The entropy of each region can be expressed as

$$
H_{r_1}(d) = - \sum_{i=0}^{d} \frac{p_i}{P_d} \ln \frac{p_i}{P_d} \tag{14}
$$
$$
H_{r_2}(d) = - \sum_{i=d+1}^{L-1} \frac{p_i}{P_d} \ln \frac{p_i}{P_d}. \tag{15}
$$

Given that the entropy for a region can also be calculated as

$$
H_d = - \sum_{i=0}^{d} \rho(i) \ln \rho(i), \tag{16}
$$
the total entropy of the image can be expressed as

\[ H_{tot} = - \sum_{i=0}^{L-1} \rho(i) \ln \rho(i). \] (17)

Thus, (14 - 15) can be simplified as follows:

\[ H_{r1}(d) = - \sum_{i=0}^{d} p_i \ln \frac{p_i}{P_d} \]
\[ = - \frac{1}{P_d} \left[ \sum_{i=1}^{d} \rho(i) \ln \rho(i) - P_d \ln P_d \right] \]
\[ = \ln(P_d) + \frac{H_d}{P_d} \] (18)

\[ H_{r2}(d) = - \sum_{i=d+1}^{L-1} p_i \ln \frac{p_i}{P_d} \]
\[ = - \frac{1}{1-P_d} \left[ \sum_{i=d+1}^{L-1} \rho(i) \ln \rho(i) - (1-P_d) \ln(1-P_d) \right] \]
\[ = \ln(1-P_d) + \frac{H_{tot} - H_d}{1-P_d} \] (19)

Iterating \( d \) from 1 to \( L - 2 \), the intensity value, \( d \), that produces the maximum sum of the distribution entropies is chosen as the threshold value \( T_H \).

3) K-Means Clustering: Clustering methods partition the observed intensities into classes, and can also be used for segmenting images. Also known as unsupervised classification, the classes are generally unknown and are explored based on the data by using a similarity measure. Given \( N \) pixels, the K-means clustering method partitions the pixels into \( K \) clusters by minimizing the sum of the within-cluster variances (WCSS) [16]:

\[ WCSS = \sum_{k=1}^{K} \sum_{i=1}^{N} ||v_i^k - \mu_k||^2, \] (20)

where \( v_i^k \) is the \( i \)-th sample of the \( k \)-th class with centroid \( \mu_k \). The pseudo-code for the K-means clustering is given as follows:
1) Initialize the number of classes $K$ and centroids $\mu_k$.
2) Assign each pixel to the group whose centroid is the closest.
3) After all the pixels have been assigned, re-calculate the centroids.
4) Repeat Steps 2 and 3 until the centroids no longer change.

Although the K-means is computationally very efficient, the major disadvantage is the need to specify the number of classes a priori. In the absence of this knowledge, one may resort to measures that could estimate the number of classes automatically [31]. This approach is not considered here. Instead, we set the number of classes as two to ensure consistency with the other image segmentation methods.

For the case of only one input image, the binary image produced by the image segmentation methods is masked on the original input image to produce an enhanced image with the target regions only.

B. Image Fusion

When there are multiple input images, the segmented images are then fused to produce a single enhanced image. Here, we consider the commonly used image fusion techniques in TWRI, such as the pixel-wise additive fusion [32], multiplicative fusion [7], and fuzzy logic image fusion [17], as candidates for the image fusion step of the proposed detection scheme. The fuzzy logic scheme has proven valuable, since it adaptively fuses different parts of the image, depending on the relative presence of the target and clutter. However, as the fuzzy logic fusion approach is implemented through a fuzzy inference system that formulates a mapping from two inputs to one output, its use is restricted to a set of no more than two enhanced images. In general, the number of images acquired from different viewing angles and for different polarizations can, however, exceed this limit. To overcome this limitation, we implement the two-stage fuzzy fusion proposed in [33], which fuses the outputs of the additive and multiplicative fusion processes. It is noted that, prior to image fusion, the segmented images are first normalized to ensure that they all have the same dynamic range.

In the two-stage image fusion, arithmetic fusion methods for TWRI proposed in [7], [32] are first applied. The images are fused through the additive and multiplicative fusion methods, given
as

\[ \hat{I}_A(m,n) = \sum_{j=1}^{J} \hat{I}_j(m,n) \]  \hspace{1cm} (21)

\[ \hat{I}_M(m,n) = \prod_{j=1}^{J} \hat{I}_j(m,n), \]  \hspace{1cm} (22)

where \( \hat{I}_j(m,n) \) is the \( j \)-th normalized segmented image and \( \hat{I}_A(m,n) \) and \( \hat{I}_M(m,n) \) are the final images resulting from the additive and multiplicative fusion, respectively. Then, a second fusion stage that exploits the capabilities of both additive and multiplicative fusion is applied, where the fuzzy logic approach is used to fuse the outputs of the arithmetic fusion. The readers are referred to [17] for a detailed description of the fuzzy logic based fusion approach.

V. PERFORMANCE EVALUATION

We evaluate both the LRT detector and the proposed detection method using real 2D polarimetric and multi-view images collected at the Radar Imaging Lab, Center for Advanced Communications, Villanova University. The techniques are also applied to numerical EM modeling data provided by the US Army Research Lab. We first compare the performance of the image segmentation methods in terms of isolation of target regions in each individual image. The LRT detector with \( J = 1 \) is also applied as an image segmentation method to each individual image for comparison. The binary images generated by both image segmentation and LRT detector are used as a mask on the original image to produce a corresponding enhanced image.

Next, we compare the performance of the various image fusion techniques to determine their suitability for the proposed detection method. It is noted that image fusion is performed only on the images that were enhanced by the image segmentation techniques in the previous step. Fusion of the images produced by the LRT detector is not performed. Instead, the output of the LRT detector with \( J > 1 \) was used for comparison. In other words, the various combinations of the image segmentation and fusion techniques for the proposed detection method are compared with the LRT detector.

We note that the frequency bands of operation considered in these examples fall within the 0.5 to 3.5 GHz range. This frequency range is most amenable to signal propagation through various wall types [1]. Furthermore, all radar images presented in this Section, other than the binary ones, are plotted on a 35 dB log scale, and the vertical and the horizontal axes represent the
downrange and cross-range, respectively, with units in meters. The FAR for the LRT detector is fixed at 2.5%.

A. Experimental Setup

Two real data scenes and one numerical EM modeling scene are considered. The first real data scene consists of calibrated targets, with data acquired from a single viewpoint using multiple polarizations. The second real data scene is a populated scene with both calibrated targets and objects typically found in an office. The data corresponding to the populated scene are acquired from multiple views with a single polarization. The numerical EM modeling data consist of both multi-view and multi-polarization images.

1) Calibrated Scene: Both co-polarization (HH and VV) and cross-polarization (HV and VH) data sets were collected from a calibrated scene, containing a sphere, a top hat, a vertical dihedral, two dihedrals rotated at 22.5 and 45 degrees, respectively, and three trihedrals, all placed at different downrange, cross-range and elevations, as shown in Fig. 4. For each polarization setting, the scene was imaged with a 1 GHz bandwidth stepped-frequency signal centered at 2.5 GHz. Two horn antennas, model H-1479 by BAE Systems, were mounted side-by-side on a Field Probe Scanner, one oriented for horizontal polarization and the other for vertical polarization. The set up was used to synthesize a 57-element linear array with an inter-element spacing of 22 mm. The transmit power was set to 5 dBm. Data were collected through a 127 mm thick non-homogeneous plywood and gypsum board wall, positioned 1.27 cm in downrange from the front face of the antennas, as shown in Fig. 4. More detailed information about the experimental set up is provided in [34] and the electrical properties of the wall material are described in [35]. The four polarimetric images acquired from the scene are shown in Fig. 5. We note that because of the non-homogeneous nature of the wall, the correction for the wall effects in the beamforming process has not been applied. It can be observed from Fig. 5 that only two targets are detected in the HV and VH images. This is due to the fact that the rotated dihedrals produce a stronger cross-polarization return. Since the HV and VH images are almost identical, the performance of the various methods will be evaluated only for the HH, HV and VV images of the calibrated scene.

2) Populated Scene: Multi-view vertical polarization data sets were also collected from a populated scene, containing a vertical dihedral, a sphere, a table with metal legs, and a chair,
each placed at different downrange, cross-range and height, as shown in Fig. 6. A stepped-frequency signal, consisting of 801 frequency steps of size 3 MHz, covering the 0.7 – 3.1 GHz band, was used for data collection. The transmit power was set to 5 dBm. A quad-ridge horn antenna, model ETS-Lindgren 3164-04, was used as the transceiver and mounted on a Field Probe Scanner to synthesize a 57-by-57 element planar array with an inter-element spacing of 22 mm. Imaging was performed through a 140 mm thick solid concrete block wall from two vantage points, namely the front and the side views. The reader is referred to [7] for more details of the experimental set up and [35] for a description of the electrical properties of the wall material. The data were processed to produce B-scan images of size 117 x 117, corresponding to the front and side views, at the heights of the dihedral and the table. Since the wall is homogeneous, the beamforming process accounted for the wall effects on signal propagation. Figure 7 shows the registered input images from the front and side views, at the dihedral’s elevation. The dihedral elevation represents an example of a scene with high signal-to-clutter ratio. Figure 8 shows the images corresponding to the table elevation, which represent the case of low signal-to-clutter ratio. Although the metal legs of the table are present in the images (indicated by white circles), it is difficult to discern the target presence without any prior knowledge of the scene.

3) EM Modeling of Complex Scene: A complex room, constructed using a 200 mm thick exterior brick wall with glass windows and a wooden door, and having an interior room, with a 50 mm thick sheetrock wall and a door, was simulated using numerical EM modeling software. In addition to wooden furniture, namely a bed with generic fabric mattress, a couch with generic fabric cushions, a bookshelf, a dresser, and a table with four chairs, four humans are also present at different positions in the scene. Both co-polarization (VV) and cross-polarization (HV) data sets were collected from the left and bottom views. For more details on EM modeling, the reader is referred to [36]. The schematic of the complex scene is shown in Fig. 9 and the four input images corresponding to the scene are provided in Fig. 10. No correction for the wall effects was applied during the beamforming process. The complex scene is an example which consists of targets with both high and low signal-to-clutter ratio. For instance, the exterior walls have high returns which overwhelm the returns from the interior walls and humans. Hence, the exterior walls have high signal-to-clutter ratio, while the interior walls and humans have low signal-to-clutter ratio.
B. Performance Measure

The resulting images corresponding to the aforementioned scenes are assessed both qualitatively and quantitatively. Visual inspection is used to assess how well the targets are maintained in the image with respect to the ground truth. The image enhancements in terms of clutter suppression are assessed using the Improvement Factor in the Target-to-Clutter Ratio, denoted as IF. Let $P_{\mathcal{R},I}$ denotes the average power of region $\mathcal{R}$ in image $\tilde{I}$. The IF is given as

$$IF = 10\log_{10} \left[ \frac{P_{\mathcal{R}_t,\tilde{I}_e} \times P_{\mathcal{R}_c,\tilde{I}_i}}{P_{\mathcal{R}_t,\tilde{I}_i} \times P_{\mathcal{R}_c,\tilde{I}_e}} \right],$$

where $\tilde{I}_i$ is the input image and $\tilde{I}_e$ is the enhanced image. $P_{\mathcal{R}_r,\tilde{I}_m}$ can be expressed as

$$P_{\mathcal{R}_r,\tilde{I}_m} = \frac{1}{N_{\mathcal{R}_r}} \sum_{x_q \in \mathcal{R}_r} (\tilde{I}_m(x_q))^2,$$

where $\tilde{I}_m(x_q)$ is the $q$-th pixel of region $\mathcal{R}_r$ in image $\tilde{I}_m$, and $N_{\mathcal{R}_r}$ and $N_{\mathcal{R}_c}$ are the number of pixels in the target region, $\mathcal{R}_t$, and clutter region, $\mathcal{R}_c$, respectively. The pre-defined target and clutter regions for each of the scenes are provided in Figs. 11 to 14. It should be noted that the target mask for the EM modeling, complex scene, as shown in Fig. 14, only includes the targets of interest, which are the room layout and human targets. We consider furniture reflections as unwanted returns and accordingly, they are treated as clutter.

C. Image Segmentation

In this section, the performance of the image segmentation techniques and LRT method with $J = 1$ are compared. For each individual image, the respective methods are used to obtain a binary mask that depicts the target locations. After obtaining an enhanced image by applying the binary mask to the original input image, the IF is calculated.

1) Calibrated Scene: The results of applying the image segmentation methods to the images of the calibrated scene in Fig. 5 are presented in Fig. 15. It can be observed that the LRT detection method applied to the individual images does not yield image enhancements. While the targets are retained, the noise and clutter present in the original images persist even after application of the LRT detector. As for the BCV thresholding, the clutter is reduced in comparison to that of the LRT detector. It is also observed that the K-means clustering method produces a similar result to that of the BCV thresholding. Although there are some missed detections, visually, the
entropy-based segmentation outperforms all the other methods by producing images with low clutter levels.

The IFs for the results obtained from the calibrated scene are provided in Table I. With the exception of the cross-polarization case, where the BCV thresholding and the K-means clustering removed all the clutter to obtain the highest IF, it is evident that the entropy-based segmentation generally outperforms the other methods by producing enhanced images with high IF; albeit with some missed detections.

### Table I
**Improvement Factor in Target-to-Clutter Ratio of Calibrated Scene Images after Enhancements through Target Detection**

<table>
<thead>
<tr>
<th>Method</th>
<th>HHH</th>
<th>HV</th>
<th>VV</th>
</tr>
</thead>
<tbody>
<tr>
<td>LRT Detector 2.5% FAR</td>
<td>1.962</td>
<td>2.5124</td>
<td>3.0411</td>
</tr>
<tr>
<td>BCV Thresholding</td>
<td>2.5501</td>
<td>17.7063</td>
<td>2.1898</td>
</tr>
<tr>
<td>Entropy-based Segmentation</td>
<td>4.3516</td>
<td>16.2246</td>
<td>6.3979</td>
</tr>
<tr>
<td>K-Means Clustering</td>
<td>2.571</td>
<td>17.7063</td>
<td>2.2344</td>
</tr>
</tbody>
</table>

2) Populated Scene: Figures 16 and 17 show the results of applying the image segmentation methods to the populated scene images in Figs. 7 and 8, respectively. At the dihedral level, it can be observed from Fig. 16 that the BCV thresholding and K-means clustering have the best performance in terms of enhancing the images from the front view. Although the entropy-based segmentation and LRT detector are able to detect the dihedral, the clutter levels are higher in comparison. However, both BCV thresholding and K-means clustering do not perform well in enhancing the images acquired from the side view. Although the LRT detector also does not perform well, comparatively less clutter is retained. Thus, the LRT produces an output image with a higher IF than the BCV and K-means for the side view. It is observed that the entropy-based segmentation outperforms all the other methods in the side view image by successfully detecting the dihedral and producing an output image with the least amount of clutter. The corresponding IFs are tabulated in Table II.

Figure 17 shows the image enhancement results at the table’s elevation. It can be observed that all the methods under investigation are able to maintain the targets, except for the LRT detector which fails to retain all four legs of the table (depicted in white circles). While two targets are
TABLE II
IMPROVEMENT FACTOR IN TARGET-TO-CLUTTER RATIO OF POPULATED SCENE IMAGES AFTER ENHANCEMENTS THROUGH TARGET DETECTION (DIHEDRAL LEVEL)

<table>
<thead>
<tr>
<th>Dihedral Level</th>
<th>VV Front View</th>
<th>VV Side View</th>
</tr>
</thead>
<tbody>
<tr>
<td>LRT Detector 2.5% FAR</td>
<td>4.4331</td>
<td>8.6943</td>
</tr>
<tr>
<td>BCV Thresholding</td>
<td>8.2646</td>
<td>1.4556</td>
</tr>
<tr>
<td>Entropy-based Segmentation</td>
<td>6.7664</td>
<td>11.2706</td>
</tr>
<tr>
<td>K-Means Clustering</td>
<td>8.2162</td>
<td>1.5151</td>
</tr>
</tbody>
</table>

detected by the LRT detector in the front view image, it detects only one target in the side view image. Amongst the image segmentation techniques that are able to maintain the four targets, it can be observed from Table III that the entropy-based segmentation produces enhanced images with the highest IF. This is due to the fact that the entropy produces an image with less false detections than those of the BCV and K-means.

TABLE III
IMPROVEMENT FACTOR IN TARGET-TO-CLUTTER RATIO OF POPULATED SCENE IMAGES AFTER ENHANCEMENTS THROUGH TARGET DETECTION (TABLE LEVEL)

<table>
<thead>
<tr>
<th>Table Level</th>
<th>VV Front View</th>
<th>VV Side View</th>
</tr>
</thead>
<tbody>
<tr>
<td>LRT Detector 2.5% FAR</td>
<td>5.6081</td>
<td>8.5922</td>
</tr>
<tr>
<td>BCV Thresholding</td>
<td>1.3736</td>
<td>1.3044</td>
</tr>
<tr>
<td>Entropy-based Segmentation</td>
<td>6.1886</td>
<td>3.0734</td>
</tr>
<tr>
<td>K-Means Clustering</td>
<td>1.3682</td>
<td>1.3237</td>
</tr>
</tbody>
</table>

3) EM Modeling of Complex Scene: The results of applying the image segmentation methods to the complex scene images (Fig. 10) are presented in Fig. 18. From the image enhancement results shown in Fig. 18, the LRT detection method produces the best result in detecting both exterior and interior walls in the co-polarization (VV) images. However, the clutter from the original images are also maintained. The entropy-based segmentation manages to detect the interior walls with less clutter in the bottom view of the co-polarization image. However, it fails to detect the interior walls from the left view image. Although the BCV thresholding and K-means clustering have higher IFs, as shown in Table IV, both methods also remove the interior walls and only the exterior walls are detected in the co-polarization images.
For the cross-polarization (HV) images, the LRT detector does not perform well as there are missed detections of the human targets. It is observed that the LRT detector fails to detect the targets in the left view image, while only one target is detected in the bottom view image. On the other hand, the BCV thresholding, entropy-based segmentation, and K-means clustering produce good detection of the human targets from the bottom view, with the entropy having the advantage of lower clutter levels. This is validated from Table IV, which shows that the entropy-based segmentation has the highest IF. However, the entropy segmentation has two missed detections in the left view image.

### D. Fusion of Enhanced Images

After segmenting the individual images, the enhanced images are normalized and fused to compare the detection results with the LRT detector that incorporates the multi-view and multi-polarization information. Here, pixel-wise additive fusion, multiplicative fusion, and two-stage fuzzy logic fusion are used to fuse the enhanced images.

Through empirical observations of the three experimental data sets, the radar images are separated into four regions for the fuzzy logic fusion. The pixel intensities, ranging from 0 to 105 for Region1, 105 to 165 for Region2, 165 to 225 for Region3 and the remaining 225 to 255 for Region 4, are used to form the membership functions. As for the fuzzy rules, it is generally observed that target objects tend to have high pixel intensities (Region4) and background noise has low pixel intensities (Region1). Thus, input pixels with higher intensities are preferred. For instance, if an input pixel is labeled as Region4, the output pixel will be set to Region4 automatically. When input pixels are from different regions, such as from Region1 and Region3,
the average value from both regions is set as the output value. When the input pixels originate from Region1 and Region2, the output values could be suppressed to Region1, as in the case of [17]. However, in our experiments, we maintain the output value as Region2 to avoid over-suppression of information. The defined set of non-overlapping logical rules is shown in Table V.

### TABLE V

<table>
<thead>
<tr>
<th>Input 1</th>
<th>Input 2</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region 1</td>
<td>Region 1</td>
<td>Region 1</td>
</tr>
<tr>
<td>Region 2</td>
<td>Region 2</td>
<td>Region 2</td>
</tr>
<tr>
<td>Region 3</td>
<td>Region 3</td>
<td>Region 3</td>
</tr>
<tr>
<td>Region 4</td>
<td>Any</td>
<td>Region 4</td>
</tr>
<tr>
<td>Region 2</td>
<td>Region 1</td>
<td>Region 2</td>
</tr>
<tr>
<td>Region 3</td>
<td>Region 1</td>
<td>Region 2</td>
</tr>
</tbody>
</table>

1) **Calibrated Scene:** Figure 19 shows the fusion results of the enhanced images from the calibrated scene (Fig. 15). We observe that the additive fusion of the enhanced images maintains all the targets, clutter, and false detections. Among the three combinations of the image segmentation methods with additive fusion, the entropy-based additive fusion produces the best result with the detection of all eight targets and an IF of 8.36 dB, as shown in Table VI. The multiplicative fusion fails to detect all eight targets because only the rotated dihedrals, which are mutual to both co- and cross-polarization images, survive the fusion process while the remaining targets are suppressed. The two-stage fuzzy fusion, which fuses the outputs of the additive and multiplicative fusion, produces a balanced output with high targets and low clutter levels. Between the additive and fuzzy logic fusion that maintains most of the targets, it can be observed that the entropy segmented image, coupled with the two-stage fuzzy fusion, produces the best result with an IF of 11.37 dB.

In comparison to the LRT detector that incorporates the multi-polarization signatures to produce the output shown in Fig. 20, the fuzzy fused composite image based on entropy segmentation outperforms the multi-polarization LRT detector by detecting seven targets versus the LRT detector’s six. The false detections in the image produced by the LRT detector are also
TABLE VI
IMPROVEMENT FACTOR IN TARGET-TO-CLUTTER RATIO OF FUSED CALIBRATED SCENE IMAGES

<table>
<thead>
<tr>
<th></th>
<th>Additive Fusion</th>
<th>Multiplicative Fusion</th>
<th>Fuzzy Logic Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCV Thresholding</td>
<td>4.2907</td>
<td>2.6817</td>
<td>8.6974</td>
</tr>
<tr>
<td>Entropy-based Segmentation</td>
<td>8.3592</td>
<td>-0.0420</td>
<td>11.3735</td>
</tr>
<tr>
<td>K-Means Clustering</td>
<td>4.3279</td>
<td>2.6813</td>
<td>8.6961</td>
</tr>
</tbody>
</table>

more significant than the fuzzy logic fusion. Thus, the entropy-based segmentation coupled with the fuzzy logic fusion outperforms the other methods by producing an image with high target and low clutter levels.

2) Populated Scene: Similar fusion results to that of the calibrated scene are also observed for the populated scene. As can be observed from Fig. 21 for the high signal-to-clutter ratio scenario, although the additive fusion generally retains most of the noise and clutter present in the individual images, the corresponding composite image based on the entropy segmented images produces an output image with minimal clutter. The image segmentation processes also help improve the multiplicative and fuzzy logic fusion results by suppressing all the clutter. As shown in Table VII, both multiplicative and fuzzy logic fusion produce composite images with similar IFs. Generally, it can be observed that all three fusion methods, employing the entropy-segmented images, produce an output with low clutter levels. Comparison to the multi-view LRT detector, whose output is shown in Fig. 22, reveals that the composite images produced by the proposed methods outperform the LRT detector. This is because the LRT detector fails to distinguish between targets and clutter in its detection mask.

TABLE VII
IMPROVEMENT FACTOR IN TARGET-TO-CLUTTER RATIO OF FUSED POPULATED SCENE IMAGES (DIHEDRAL LEVEL)

<table>
<thead>
<tr>
<th></th>
<th>Additive Fusion</th>
<th>Multiplicative Fusion</th>
<th>Fuzzy Logic Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCV Thresholding</td>
<td>2.6245</td>
<td>16.8401</td>
<td>16.7586</td>
</tr>
<tr>
<td>Entropy-based Segmentation</td>
<td>10.7841</td>
<td>14.8618</td>
<td>14.2677</td>
</tr>
<tr>
<td>K-Means Clustering</td>
<td>2.6968</td>
<td>16.8511</td>
<td>16.7585</td>
</tr>
</tbody>
</table>

In the low signal-to-clutter ratio case, it can be observed from Fig. 23 that the composite images produced by the multiplicative and fuzzy logic fusion based on the entropy segmented
images clearly depict the four legs of the metal table, even though there are some false detections close to the table. The LRT detector that incorporates the multi-view information misses detection of one of the table legs, but has fewer false alarms compared to that of the composite images. Table VIII shows that the fusion of the entropy segmented images produces composite images with the highest IF.

| TABLE VIII |
| Improvement Factor in Target-to-Clutter Ratio of Fused Populated Scene Images (Table Level) |

<table>
<thead>
<tr>
<th>Method</th>
<th>Additive Fusion</th>
<th>Multiplicative Fusion</th>
<th>Fuzzy Logic Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCV Thresholding</td>
<td>2.3180</td>
<td>7.1884</td>
<td>9.9716</td>
</tr>
<tr>
<td>Entropy-based Segmentation</td>
<td>5.2657</td>
<td>15.0991</td>
<td>16.1727</td>
</tr>
<tr>
<td>K-Means Clustering</td>
<td>2.3241</td>
<td>7.1969</td>
<td>9.9856</td>
</tr>
</tbody>
</table>

3) EM Modeling of Complex Scene: Figures 25 and 26 show the image fusion results of the segmented multi-view co-polarization (VV) and cross-polarization (HV) images acquired from the complex scene, respectively. In the co-polarization composite images, it is evident from Fig. 25 that the additive fusion based on the entropy segmented images produces the best visual result by detecting both exterior and interior walls. It also has the highest IF, as tabulated in Table IX.

| TABLE IX |
| Improvement Factor in Target-to-Clutter Ratio of Fused Multiview Complex Scene Images |

<table>
<thead>
<tr>
<th>Co-Polarization (VV)</th>
<th>Additive Fusion</th>
<th>Multiplicative Fusion</th>
<th>Fuzzy Logic Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCV Thresholding</td>
<td>6.8322</td>
<td>-20.1781</td>
<td>4.4134</td>
</tr>
<tr>
<td>Entropy-based Segmentation</td>
<td>7.0535</td>
<td>-18.7161</td>
<td>4.4921</td>
</tr>
<tr>
<td>K-Means Clustering</td>
<td>6.8318</td>
<td>-20.1781</td>
<td>4.4124</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cross-Polarization (HV)</th>
<th>Additive Fusion</th>
<th>Multiplicative Fusion</th>
<th>Fuzzy Logic Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCV Thresholding</td>
<td>3.3666</td>
<td>9.4389</td>
<td>7.2484</td>
</tr>
<tr>
<td>Entropy-based Segmentation</td>
<td>7.8120</td>
<td>8.1383</td>
<td>7.5353</td>
</tr>
<tr>
<td>K-Means Clustering</td>
<td>3.4099</td>
<td>9.4347</td>
<td>7.2403</td>
</tr>
</tbody>
</table>

For the cross-polarization composite images (Fig. 26), the fuzzy logic fusion produces images that balance the output of the additive and multiplicative fusion. Although the IF is lower than that of the multiplicative, visually, the entropy-segmented image coupled with the fuzzy fusion
produces a result with the least clutter. The additive fusion of the entropy segmented images also produces a similar result, where the clutter level is lower compared to that of BCV thresholding and K-means clustering. When only considering the human targets, it is evident that, for the cross-polarization images, the multiplicative fusion of the BCV thresholding and K-means clustering images produces the best result. However, it can be observed that the entropy-based segmentation, coupled with the fuzzy logic fusion, produces the best result that detects both human targets and room layout.

We observe from Fig. 27 that the multi-view LRT detector, when applied to the co-polarization images, produces a similar output to that of additive fusion based on the entropy segmented images. While the LRT detector manages to detect the four humans in the cross-polarization images, the clutter levels are higher than that of the multiplicative fusion with the image segmentation techniques.

Figures 28 and 29 show the image fusion results of the segmented polarimetric images from the left and bottom views of the complex scene, respectively. From the left view (Fig. 28), it can be observed that, while the additive fusion and fuzzy logic fusion coupled with BCV thresholding and K-means clustering are able to maintain the targets, clutter levels are still quite high. Similar results are also observed for the bottom view (Fig. 29). Although the composite images based on the entropy segmented images do not detect all the targets, it generally produces an image with high target and low clutter levels. The IFs, as tabulated in Table X are also the highest. In both cases, the multiplicative fusion fails to detect the human targets and the room layout, due to the suppression of non-mutual pixels in the enhanced images. Figure 30 shows that the multi-polarization LRT detector, which is evaluated on the polarimetric images both from the left and bottom views, does not perform well enough to detect the walls and human targets. In both cases, high clutter levels are retained.

E. Discussions

From the experimental results, it can be generally observed that the BCV thresholding and K-means clustering yield similar segmentation performances and their differences in the IF are almost negligible. This is due to the fact that the K-means clustering’s minimization of the WCSS is equivalent to the BCV thresholding’s maximization of the sum of class variances when there are only two regions [14]. It can also be observed that the BCV thresholding generally has
a poorer performance than the entropy-based segmentation when the image has more clutter. This is because the sparsity of the high pixel values (target pixels) in a TWR image forces the BCV thresholding method to also include low pixel values (clutter pixels) in the target class, in order to obtain a high class variance. Unlike the BCV method, the entropy-based segmentation is less affected by the sparse target pixels in the image and is only dependent on the pixel value probability. That is, low pixel values will affect thresholding if they assume similar probabilities as those of the target pixels. In the case when the scene has less clutter, it is observed that a similar performance will be produced by both entropy-based segmentation and BCV thresholding methods.

Using the calibrated scene as an example, the corresponding co-polarization images contain more clutter than the cross-polarization image. Hence, in the case of co-polarization images, the entropy-based segmentation outperforms the BCV, which choses a lower thresholding value. This is evident from Fig. 31, wherein the maximum class variances for both regions in the co-polarization image are skewed towards the lower pixel regions. As a result, the summation of both class variances produces a threshold value that is also located in the low pixel values region, causing the segmented image to include most of the clutter. As can be observed from Fig. 32, the respective class entropies are less affected by the skewness towards the low pixel values region. As the threshold value is increased, the entropy of one region will generally decrease while the entropy of the other region will increase. Thus, the maximization of the sum of class entropies will produce a balanced threshold where the information content in both regions are almost

---

**TABLE X**  
**IMPROVEMENT FACTOR IN TARGET-TO-CLUTTER RATIO OF FUSED POLARIMETRIC COMPLEX SCENE IMAGES**

<table>
<thead>
<tr>
<th></th>
<th>Left View</th>
<th>Bottom View</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Additive Fusion</td>
<td>Multiplicative Fusion</td>
</tr>
<tr>
<td>BCV Thresholding</td>
<td>3.6250</td>
<td>-6.1226</td>
</tr>
<tr>
<td>Entropy-based Segmentation</td>
<td>11.5689</td>
<td>-13.2830</td>
</tr>
<tr>
<td>K-Means Clustering</td>
<td>3.6476</td>
<td>-6.1682</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Additive Fusion</td>
<td>Multiplicative Fusion</td>
</tr>
<tr>
<td>BCV Thresholding</td>
<td>2.3547</td>
<td>-11.7157</td>
</tr>
<tr>
<td>Entropy-based Segmentation</td>
<td>7.3869</td>
<td>-6.3540</td>
</tr>
<tr>
<td>K-Means Clustering</td>
<td>2.4264</td>
<td>-11.7157</td>
</tr>
</tbody>
</table>
equal. For the case of a less cluttered image, such as the one acquired using cross-polarization, a higher threshold value is produced by the BCV method since the maximum class variance for the second region is located at a much higher pixel value region. This dominance allows the summation of both class variances to produce a high threshold value.

The observations that i) the entropy-based segmentation outperforms the BCV thresholding method when there is more clutter, and ii) both methods have similar performance when there is less clutter in the image, are further supported by Fig. 33. The histograms in the left column are produced from images with more clutter, and those in the right column are histograms of images with less clutter. The thresholds associated with the various segmentation methods are plotted in red, green, and black, respectively, for the entropy-based segmentation, BCV thresholding, and K-means clustering. As can be seen, the threshold values corresponding to entropy are much higher in case of high clutter and comparable to those of K-means and BCV thresholding in case of low clutter.

In this paper, we have considered three different scenes of varying complexity in an attempt to have a broad representation of a variety of possible indoor scenes. In all cases, the corresponding image histograms are uni-modal, as depicted in Figure 33. Thus, it can be generally concluded that most TWR images have uni-modal histograms. In order to automate the target detection process for uni-modal images, an image segmentation method that is not affected by the uni-modality and sparsity of the targets needs to be taken into consideration. Through the experimental results, it has been shown that the entropy-based segmentation, coupled with the fuzzy logic-based image fusion, is the most effective and viable alternative to the LRT for target detection in TWR images.

VI. CONCLUSION

In this paper, we examined image processing approaches for target detection in TWR. Image segmentation techniques were first applied to enhance the images by distinguishing between the target and clutter regions. Image fusion techniques were then used to fuse the enhanced images. Performances of various candidate segmentation and fusion techniques were evaluated using real 2D polarimetric and multi-view images, as well as numerical EM modeling data. The results showed that the entropy-based segmentation technique produced, in most of the cases considered, an output better than those produced by the other segmentation schemes, as well as
the LRT detector. Although there were some missed detections, the entropy-based segmentation consistently provided high target and low clutter levels. While the BCV thresholding and K-means clustering methods also maintained most of the targets, the clutter levels were much higher by comparison. For the fusion of multi-view and multi-polarization images, it was generally observed that the fuzzy logic fusion of the segmented images revealed important information about the indoor targets and walls.

ACKNOWLEDGMENT

The authors would like to thank Dr. Traian Dogaru from the US Army Research Lab for providing the numerical EM modeling data.
REFERENCES


Fig. 1. Geometry on transmit for two-dimensional imaging.

Fig. 2. Block diagram of the iterative statistical target detector.

Fig. 3. Block diagram of the proposed image segmentation-based target detector.
Fig. 4. Calibrated scene, showing the scene that is imaged through a non-homogenous plywood and gypsum wall (left), and schematic diagram of the scene (right).

Fig. 5. Images acquired from the calibrated scene.
Fig. 6. Populated scene, showing the scene that is imaged through a homogenous concrete wall from two vantage points (left), and schematic diagram of the scene (right).

Fig. 7. Registered input images obtained from the dihedral's elevation.

Fig. 8. Registered input images obtained from the table's elevation.
Fig. 9. Schematic of the complex scene.

Fig. 10. Images of a complex scene produced by numerical EM modeling.

Fig. 11. Target and clutter mask of the calibrated scene.
Fig. 12. Target and clutter mask of the populated scene at the dihedral’s elevation.

Fig. 13. Target and clutter mask of the populated scene at the table’s elevation.

Fig. 14. Target and clutter mask of the complex scene.
Fig. 15. Target detection in the calibrated scene: from top to bottom, statistical LRT detector, BCV thresholding, entropy-based segmentation and K-means clustering.
Fig. 16. Image enhancement results from the populated scene (dihedral level).
Fig. 17. Image enhancement results from the populated scene (table level).
Fig. 18. Image enhancement results from the complex scene.
Additive Fusion after BCV Thresholding

Additive Fusion after Entropy Segmentation

Additive Fusion after K-Means Clustering

Multiplicative Fusion after BCV Thresholding

Multiplicative Fusion after Entropy Segmentation

Multiplicative Fusion after K-Means Clustering

Fuzzy Logic Fusion after BCV Thresholding

Fuzzy Logic Fusion after Entropy Segmentation

Fuzzy Logic Fusion after K-Means Clustering

Fig. 19. Image fusion of the segmented images from the calibrated scene.
Fig. 20. Multi-polarization statistical LRT detection on calibrated scene images.
Fig. 21. Image fusion of the segmented images from the populated scene (dihedral level).
Fig. 22. Multiview statistical LRT detection on populated scene images (dihedral level).
Fig. 23. Image fusion of the segmented images from the populated scene (table level).
Fig. 24. Multiview statistical LRT detection on populated scene images (table level).

Additive Fusion of VV scenes after BCV Thresholding
IF: 6.8322

Additive Fusion of VV scenes after Entropy Segmentation
IF: 7.0535

Additive Fusion of VV scenes after K−means Clustering
IF: 6.8318

Multiplicative Fusion of VV scenes after BCV Thresholding
IF: −20.1781

Multiplicative Fusion of VV scenes after Entropy Segmentation
IF: −18.7161

Multiplicative Fusion of VV scenes after K−means Clustering
IF: −20.1781

Fuzzy Logic Fusion of VV scenes after BCV Thresholding
IF: 4.4134

Fuzzy Logic Fusion of VV scenes after Entropy Segmentation
IF: 4.4921

Fuzzy Logic Fusion of VV scenes after K−means Clustering
IF: 4.4124

Fig. 25. Image fusion of the segmented multiview co-polarization images from the complex scene.
Additive Fusion of HV scenes after BCV Thresholding
IF: 3.3666
Additive Fusion of HV scenes after Entropy Segmentation
IF: 7.812
Additive Fusion of HV scenes after K−means Clustering
IF: 3.4099

Multiplicative Fusion of HV scenes after BCV Thresholding
IF: 9.4389
Multiplicative Fusion of HV scenes after Entropy Segmentation
IF: 8.1383
Multiplicative Fusion of HV scenes after K−means Clustering
IF: 9.4347

Fuzzy Logic Fusion of HV scenes after BCV Thresholding
IF: 7.2484
Fuzzy Logic Fusion of HV scenes after Entropy Segmentation
IF: 7.5353
Fuzzy Logic Fusion of HV scenes after K−means Clustering
IF: 7.2403

Fig. 26. Image fusion of the segmented multiview cross-polarization images from the complex scene.

Fig. 27. Multiview statistical LRT detection on co-polarization and cross-polarization complex scene images.
Fig. 28. Image fusion of the segmented polarimetric images from the left view of the complex scene.
Fig. 29. Image fusion of the segmented polarimetric images from the bottom view of the complex scene.

Fig. 30. Multi-polarization statistical LRT detection on co-polarization and cross-polarization complex scene images.
Fig. 31. Class variances with respect to the threshold values for co-polarization images (left) and cross-polarization image (right) of the calibrated scene.

Fig. 32. Class entropies with respect to the threshold values for co-polarization images (left) and cross-polarization image (right) of the calibrated scene.
Fig. 33. Histogram of different images along with the thresholds associated with the segmentation methods, showing calibrated scene (top row), populated scene (middle row) and complex scene (bottom row).