Image Based Robust Target Classification for Passive ISAR

Andrea Manno-Kovacs, *Member, IEEE,* Elisa Giusti, *Member, IEEE,* Fabrizio Berizzi, *Senior Member, IEEE,* and Levente Kovács, *Member, IEEE*

Abstract—The paper presents an automatic and robust, image feature based target extraction and classification method for multistatic passive Inverse Synthetic Aperture Radar (ISAR) range/cross-range images. The method can be used as a standalone solution or for augmenting classical signal processing approaches. By extracting textural, directional and edge information as low-level features, a fused saliency map is calculated for the images and used for target detection. The proposed method uses the contour and the size of the detected targets for classification, is lightweight, fast and easy to extend. The performance of the approach is compared with machine learning methods and extensively evaluated on real target images.

Index Terms—Target classification, passive radar, ISAR, ATR.

I. INTRODUCTION

FOR decades, passive radars (PR) have attracted the attention of the scientific community because of numerous advantages over conventional active radars, namely, low vulnerability to electronic countermeasure, counter-stealth advantage and no electro-magnetic (e.m.) emission, which have made PR an attractive technology especially for military applications.

Passive radars exploit available non-cooperative illuminators of opportunity (e.g., digital video broadcasts [1], mobile signals [2], digital or FM radio [3], etc.) as signal sources, and deploy one or more controlled receivers for target detection and imaging. There is continued scientific and industrial/defense interest towards such systems, especially since low cost passive radars [4] and real time processing are now possible. Moreover, PRs do not require frequency allocation neither produce e.m. pollution. These features have made the use of PRs attractive also for civilian applications such as maritime and aerial traffic surveillance.

As research progresses in this area, more radar techniques are added to PR systems to make them able to handle tasks like radar imaging [5] of non-cooperative targets using Inverse Synthetic Aperture Radar (ISAR) methods [6], [7], [8], [2]. The application of ISAR processing to passive radar systems enables the generation of 2D images of detected targets. The passive ISAR images represent an estimate of the target reflectivity function at those frequency bands where the PRs operate and may be used for Automatic Target Recognition (ATR) purposes. Using passive ISAR images for automatic target classification is a field still in development.

To be effective for ATR, Passive ISAR images should have fine enough spatial resolutions, namely range and cross-range resolutions. Range resolution depends on the signal instantaneous bandwidth, while cross-range resolution depends on both the operative frequency and the aspect angle changes between the radar and the target due to the target's own motions, which are however unknown. Broadcast signals, typically used by Passive radar systems as signal of opportunity, are narrow band signals and uses low operative frequencies that determine coarse range and cross-range resolutions, respectively. However finer range resolutions may be achieved by coherently adjoining more frequency channels, as demonstrated in [9]. This permits to gather a signal with a larger instantaneous bandwidth, approximately N times that of a single frequency channel signal, if N is the number of the adjoined frequency channels. Finer cross-range resolutions may be achieved by processing longer coherent processing time, which in turn means larger aspect angle changes.

Among related works, there are methods for target detection using passive radar signal processing [2], [10], [11], [12]. In this paper we focus on image-based target detection and classification. The are methods for image based ISAR target detection that require a priori target structure information [13]. Methods for passive ISAR target classification include an image based classification work using 3 classes, achieving 71% recognition rate [14]. Another target classification approach requires coordinated flight models, on-board recorded flight paths and profiles, altitude and velocity information, producing 21-100% rates. In [15] we presented a proof-of-concept generic passive ISAR image based classification with 61% average recognition rate that does not require a priori target models or other sensor information.

This paper's focus is on the investigation of automatic target classification capabilities for passive ISAR. The proposed method uses the range/cross-range images of targets produced by radar imaging, and presents a generic, purely image based target extraction and classification approach. The presented method is evaluated on real passive ISAR data.

The contributions of this paper are as follows. i). We present and evaluate a robust, generic, model-free, no-constraints approach for image based target classification that only uses image-based target features and no additional external data, for various target classes and image resolutions. ii). The method does not classify generic image features, but robustly extracts the target from the images and uses their shape information.

A. Manno-Kovacs and L. Kovács are with the Institute for Computer Science and Control - MTA SZTAKI, Hungarian Academy of Sciences, Budapest, Hungary, e-mail: {andrea.manno-kovacs, levente.kovacs}@sztaki.mta.hu

E. Giusti and F. Berizzi are with the Department of Information Engineering, University of Pisa, Italy, e-mail: elisa.giusti@cnit.it, f.berizzi@iet.unipi.it

F. Berizzi is with the CNIT Radar and Surveillance System (RaSS) National Laboratory, Pisa, Italy, e-mail: f.berizzi@iet.unipi.it
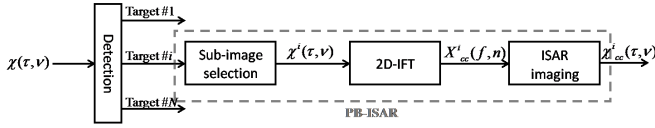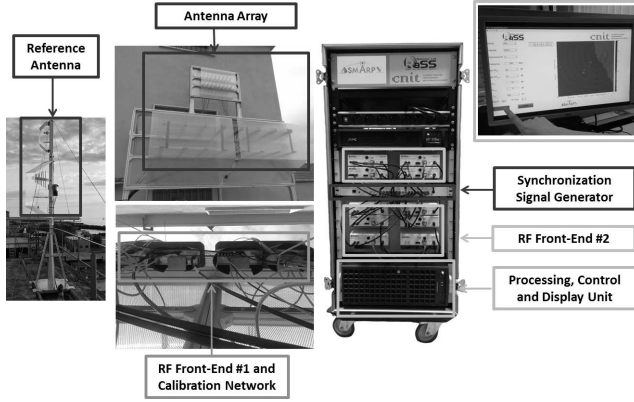
Fig. 1. Passive Radar geometry

Fig. 2. Graphical representation of both $s_{surv}(t)$ and $s_{ref}(t)$ according to the batch approach

iii). The method is usable with a low number of target samples but easy to extend with more samples and target classes. iv). It requires no re-training when extending the labeled target dataset. The current method improves on our previous proof-of-concept [15] with v). a target masking step during detection, by vi). using a better performing contour extraction and vii). an adapted and improved local edge orientation selection step.

The proposed method extracts the targets from the range/cross-range images, and uses their shape and length features for automatic classification. The segmentation of the target from the images is based on our previous results in saliency based feature extraction [16], [17]. Evaluations show that we can achieve classification rates up to 92% on real data. Application use cases include passive observation for force and area protection [4].

In the following we will describe the imaging process (Sec. II), the proposed target extraction and classification (Sec. III) and the evaluations (Sec. IV).

II. RADAR AND IMAGING

In this section the mathematical background of both passive radar (PR) processing and passive ISAR is provided. The PR prototype used for the measurements is also presented.

A. PR Geometry and Processing

Let the geometry be represented in Fig. 1, where T_ξ is a Cartesian reference system embedded on the IO. The receiver is composed of two antennas, one pointing toward the IO to gather the reference signal and the other pointing toward the area to be surveyed, R_{TxTg}, R_{TxRx} and R_{RxTg} are the transmitter to target, transmitter to receiver and receiver to target distances and β is the bistatic angle. A PR is intrinsically bistatic since the transmitter and the receiver are not co-located. According to the results in [18], a bistatic config-uration can be approximated with an equivalent monostatic configuration with a virtual sensor located along the bisector of the bistatic angle β. Therefore the Bistatic Equivalent Monostatic (BEM) radar can be defined as in Fig. 1, and the bistatic ISAR theory applied in this framework.

The target motions with respect to the BEM are described by the target total rotation vector $\mathbf{\Omega}_T$. The effective rotation vector $\mathbf{\Omega}_{eff}$ is the component of $\mathbf{\Omega}_T$ that contributes to the ISAR image formation, namely $\mathbf{\Omega}_{eff} = \mathbf{i}_{LoS}^{BEM} \times \left[\mathbf{\Omega}_T(t) \times \mathbf{i}_{LoS}^{BEM}\right]$.

At the receiver the cross-ambiguity function between the reference signal, $s_{ref}(t)$ and the surveillance signal, $s_{surv}(t)$ is computed that provides the range/Doppler (RD) map. Sub-optimal approaches are typically used instead of the standard CAF (Cross Ambiguity Function) algorithm to speed up the range/Doppler map formation. One of such approaches is the "CAF batches" algorithm [19]. Following this approach, both $s_{surv}(t)$ and $s_{ref}(t)$ are divided into batches of temporal length T_b as depicted in Fig. 2.

The CAF function can be then seen as a weighted sum of the cross-correlation calculated within each batch, as in Eq. (1). The "CAF batches" algorithm approaches the "standard CAF" algorithm performance when $T_b\nu_{max} \ll 1$:

$$\chi(\tau, \nu) = \sum_{n=1}^{N_b} e^{-j2\pi\nu n T_b} \cdot \int s_s(t;n)s_r^*(t-\tau;n)\,dt. \quad (1)$$

The system that we use (SMARP, Sec. II-C) is a software defined radar system, therefore equation (1) is performed dig-itally after the analog to digital converter (ADC), as follows:

$$\chi(\tau_p, \nu_q) = \sum_{n=1}^{N_b} e^{-j2\pi\nu_q n T_b} \cdot \sum_{m=1}^{M} s_s(t_m;n)s_r^*(t_m-\tau_p;n), \quad (2)$$

where $t_m = m\Delta t$, $\tau_p = p\Delta\tau$ and $\nu_q = q\Delta\nu$ are the fast time, delay-time and Doppler samples, $m = 1, \cdots, M$, $M\Delta t = T_b$, $p = 1, \cdots, P$, $q = 1, \cdots, Q$ and Δt, $\Delta\tau$ and $\Delta\nu$ are the sampling intervals. When no super-resolutions techniques are implemented, $P = M$ and $Q = N_b$.

B. Passive-Bistatic ISAR Processing

The passive ISAR processing chain is visualized in Fig. 3. The algorithm takes as input the RD map as produced by the "CAF batches" algorithm. The moving target echo is extracted from the RD map through a 2D filter applied to the RD map which aims at isolating the moving target sub image from the other targets and noise and clutter. This is a fundamental step since the ISAR processing can be applied to a target at a time and noise and clutter may affect the ISAR processing performance.

The target sub-image cropped from the RD map is then pro-jected back onto the data domain, namely the "frequency/slow-time" domain via a 2D Fourier transform. In [6] the authors demonstrate that by inverse Fourier transforming the RD sub-image of the moving target a bistatic ISAR-like signal in the "frequency/slow-time" domain can be obtained to which the ISAR processing can be applied.

Fig. 3. Passive ISAR processing block scheme
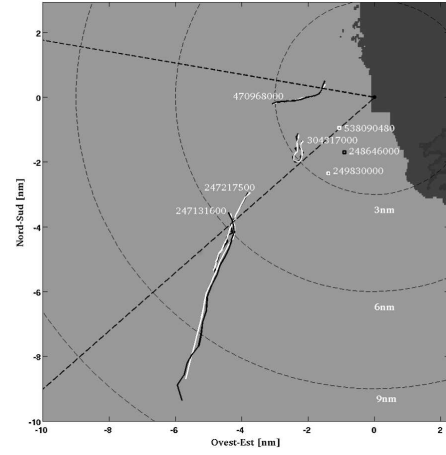

Fig. 4. The SMARP system.

Fig. 5. An example of the SMARP tracking results: AIS trajectories (white lines) and radar tracks (black line).
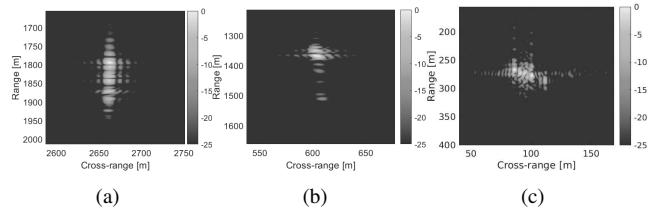

(a)   (b)   (c)

Fig. 6. Examples of P-ISAR images of three ships: (a) Cargo ship 200.63m long and 26.5m wide, (b) Cargo ship 184m long and 27m wide, (c) cooperative ship equipped with GPS and 32m long and 7m wide.

Once the data is transformed into a domain compatible with a typical ISAR processor input, any ISAR processor may be used to form a focused ISAR image of the target. Without loosing any generality, we will make use of the Image Contrast Based Autofocus (ICBA) technique followed by a Range-Doppler (RD) image formation [6]. Finally, the ISAR image is rescaled along the cross-range direction to obtain an image displayed in a spatial coordinate system [20].

C. The SMARP System

The proposed approach has been tested on data acquired with the SMARP (Software-defined Multiband Array Passive Radar) passive radar demonstrator. SMARP has been developed by the Radar and Surveillance Systems Laboratory (RaSS Lab.) of the Italian National Inter-University Consortium for Telecommunications (CNIT). SMARP is a dual band and dual polarization passive radar operating at UHF (470-790 MHz) and S-band (2100-2200 MHz). In its current version SMARP is able to acquire up to 25 MHz bandwidth signal at UHF [4]. A picture of SMARP is shown in Fig. 4.

The SMARP system radio frequency (RF) front-end is composed of the reference and surveillance antennas, the "RF front-end 1" that includes the calibration network, and the "RF front-end 2" that includes the synchronization network. The picture on the right hand-corner shows a picture of the workstation monitor on which the "processing, control and display unit" block is implemented.

An example of the detection and tracking results is shown in Fig. 5. Examples of passive ISAR images of ships obtained with SMARP system are shown in Fig. 6.

The dotted lines in Fig. 5 represent the main beam of the surveillance antenna and the black dot represents the SMARP geographical location. The IO is approximately $30kms$ away from the receiver inland. The bistatic angle is determined by the receiver, transmitter and target location, therefore it may

change a little depending of the target coordinates with respect to both the transmitter and the receiver. However the bistatic angle was in the range $[10° - 40°]$. The monostatic range resolution was $6.25m$, whilst the bistatic range resolution is coarser as it is worsened by the cosine of half the bistatic angle. The coherent processing interval (CPI) also differs among different targets. The CPI used to generate images in Fig. 6 was $5s$. Similar CPIs were used to generate the other ISAR images that compose the used dataset. The typical ISAR processing output is an image in the range/Doppler map. The algorithm proposed in [20] has been used to convert the Doppler axis to cross-range axis, therefore from Hertz to meters. Fully scaled ISAR images as those in Fig. 6 permit to estimate the target length.

During trials both targets of opportunity and cooperative targets were present. Targets of opportunity are cargo ships approaching or leaving the Livorno harbor, which is close to the SMARP system location, with similar trajectories and mainly along the SMARP Line of Sight. This explains why images in Fig. 6 (a) and (b) that represent cargo, have a similar pose and therefore a similar ISAR image. Conversely, the cooperative target was travelling on a predefined trajectory that included circumferences and during the acquisition was travelling approximately perpendicular to the line of sight of the radar. This explains why the image in Fig. 6(c) has a different pose with respect to the others.

The target size can be estimated from a 2D ISAR image. It is worth to mention that, since the 2D ISAR image is the result

TABLE I
TARGET CLASSES IN THE DATASET.

Class nr	1	2	3	4	5	6	7	8	9	10	11	12
Class label	A	B	C	D	E	F	G	H	I	J	K	L
Elements	14	16	15	11	9	18	31	18	26	35	38	63

of a projection of the three-dimensional target reflectivity function onto an unknown 2D plane, the target size could be underestimated and may change over time. This may lead to an incorrect estimation of the target's geometrical features, such as the target size. The use of a sequence of ISAR images instead of only a single one may partially overcome this issue and could provide a better estimate of the target's length.

III. TARGET EXTRACTION AND CLASSIFICATION

In the target extraction step we use a fused textural and directional feature map to detect target candidates and extract the contour and the length of the target. In the classification step we use these features to estimate the class of an unknown target based on the a priori labeled dataset. The dataset used for evaluations was obtained with the described radar (Sec. II-C), and contains 294 range/cross-range images of 12 target classes labeled A-L (3 planes, 9 ships, containing 9-63 class elements as shown in Table I, and some samples in Fig. 6).

The proposed method processes the radar-generated ISAR images, and it does not depend on the ISAR imaging process. Currently, the method processes single images independently, thus it does not depend on or use target motions, trajectories, dynamics. It handles the classification process from a purely content-based retrieval point of view.

A. Detection and Feature Extraction

The proposed method does not use a priori target information, it only relies on discriminative image features. A benefit of such an approach is flexibility and independence from target model constraints. The target detection and extraction uses fused morphological, textural and edge feature maps.

The input images can have different pixel resolutions, and they can also have different the spatial resolutions. Also, the two dimensions of one image can have different spatial resolutions (as shown in the examples of Fig. 7 (a, c, e)). Thus, as a first step before processing, we rescale the input images so each image will have the same meters/pixel (m/px) resolution along both of its axes (e.g., Fig. 7 (b, d, f)). The range of different spatial resolution values in the dataset is 0.81-11.72 m/px. These rescaled images will be the inputs of the detection (Fig. 8(a)), which begins with an adaptive Otsu thresholding [21] step for noise reduction. Generally, Otsu's method performs clustering-based thresholding to binarize gray level images: by assuming that the image contains only two classes (foreground and background), a threshold is calculated adaptively to minimize intra-class variance and maximize inter-class variance between them based on the intensity. In [15] such a filtered image was used as input, but in this work we improved the detection process by using the thresholded image as region of interest mask on the input (Fig. 8(b)).
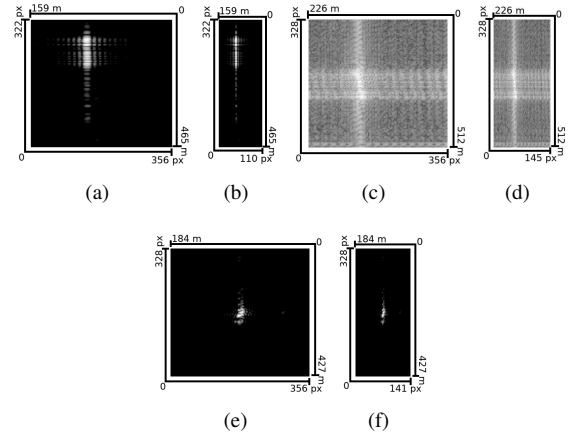
Fig. 7. Input images (a, c, e) and rescaled versions (b, d, f) for processing.

As it can be observed from the samples in Fig. 7, the images can have quite limited features, the marginal color information and the often unclear contours both make target extraction challenging. Therefore, our aim is to perform a salient object localization and outline approximation by exploiting the reduced feature space, applying the fusion of available features, like texture, edge and orientation.

As a first step, textural features are extracted. The image is analyzed on the pixel level, then partitioned into texture atoms on a regional level, using a sparse texture model [22]. A texture distinctiveness value is assigned to the texture atoms, representing the relevance of the texture in the image. The assigned distinctiveness value is calculated using both the occurrence frequency and visual attention-based rules concerning the position of the texture atom (i.e., the center region of the image draws higher visual attention), resulting in a T texture map. In the T map more distinct textures are assigned a higher distinctiveness value (in Fig. 8(c) darker color means higher distinctiveness). By Otsu-thresholding the T map we obtain an estimate for the location of the salient (Fig. 8(d)). The number of texture atoms has to be set beforehand: after testing different values, the recommendations of the original method were followed and the initial atom number was set to 20.

Beside textural features, limited contour (edge) information is also available, which can be applied for salient object detection. In case of the passive ISAR images, there are usually no clear contours, therefore it is more efficient to represent the outline by salient keypoints and defining salient object features based on this point set. To extract more representative outline information, an object-specific main orientation is calculated based on the main orientation of the gradient in the close proximity of the salient keypoints. This technique was previously introduced for general imagery [16], [23] and satellite imagery [24], where the P salient keypoint set is calculated as the local maxima of the Modified Harris for Edges and Corners (MHEC) characteristic function [25], which is a modification of the Harris detector's characteristic function [26] for noisy and high curvature boundaries. The Harris detector was introduced for detecting corner points, by calculating changes of intensity for a small shift in the image. In case of a corner point, this change is expected to

be large. The change can be approximated by the so-called Harris matrix, which is based on the second-order image derivatives. Using the eigenvalues of the Harris matrix, a corner response is generated to distinguish between corner, edge and flat (homogeneous) regions. The local maxima of the corner response indicates the location of corner points.

When representing objects, beside corners, edges may also help, therefore a modification of the corner response is applied here. Based on the λ_1 and λ_2 eigenvalues of the Harris matrix, the modified characteristic function is calculated as their maximum: $H_{\mathrm{mod}} = \max(\lambda_1, \lambda_2)$. Local maxima of H_{mod} define the P salient keypoint set, which represent the object outline in the subsequent feature extraction step. P set is applied in the mentioned earlier works for defining the dominant orientations of the currently processed region of interest (ROI), by calculating the local gradient orientation density (LGOD) [27] in the small neighborhood around the points of P. A $W_n(i)$, $n \times n$ ($n = 3$) window is defined around the ith point of the point set (P_i), and for each $r \in W_n(i)$ pixel, the local gradient orientation is calculated:

$$\varphi_i = \underset{\varphi \in [-90, +90]}{\mathrm{argmax}} \left\{ \frac{1}{N_i} \sum_{r \in W_n(i)} \frac{1}{h} \cdot \|\nabla g_r\| \cdot \kappa \left(\frac{\varphi - \varphi_r^{\nabla}}{h} \right) \right\}, \tag{3}$$

where $\nabla g_r = [g_{r,x} \ g_{r,y}]$ is the intensity gradient vector at (x_r, y_r) image pixel, with $\|\nabla g_r\| = \sqrt{g_{r,x}^2 + g_{r,y}^2}$ magnitude and $\varphi_r^{\nabla} = tan^{-1}[\frac{g_{r,y}}{g_{r,x}}]$ orientation. To get a more balanced data for φ_i selection, the weighted orientation histogram in the $W_n(i)$ is smoothed. To achieve this, the κ Gaussian smoothing kernel is applied, which is a non-negative, symmetric function and represents the shape of the Gaussian (bell-shaped) hump. The h bandwidth parameter is responsible for the scale of the smoothing, larger values may result in smoother function and there is always a trade-off between the preservation of the original data characteristics and the smoothing effect. By following the earlier works on the LGOD [27], [24], $h = 0.7$ bandwidth parameter is applied. Finally, $N_i = \sum_{r \in W_n(i)} \|\nabla g_r\|$ is defined.

After calculating all the φ_i values for all keypoints, a $\vartheta(\varphi)$ orientation histogram is defined from them. Previously [15] the dominant peaks of the histogram were analyzed by correlating a series of Gaussian functions to the orientation histogram iteratively. However, in the present case, the salient objects usually have one dominant direction (like in Fig. 7), therefore the iterative search for the main orientations can be simplified to a maximum-search on the $\vartheta(\varphi)$ orientation histogram. The main direction will be:

$$\alpha = \underset{\vartheta \in [-90, +90]}{\mathrm{argmax}} \{\vartheta(\varphi)\}. \tag{4}$$

Extensive evaluations also confirmed that the detection with this simplification results in higher classification performance, moreover the detection process becomes faster. The P salient keypoint set is shown in white in Fig. 9(a) for the cropped image, the calculated $\vartheta(\varphi)$ orientation histogram is in Fig. 9(b) with the α dominant orientation marked with the black line.

By defining the main direction representing the salient object, an improved edge map can be constructed, where edges
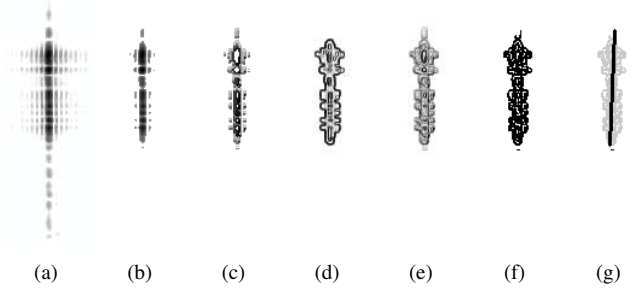

Fig. 8. The feature map generation and target extraction steps: (a) input; (b) filtered input; (c) texture feature image; (d) directional image; (e) fused feature image; (f) final target candidates; (g) length of final target (all inverted for visualization). Images are inverted for better visibility.

in the main direction are further emphasized. To achieve this, the Morphological Feature Contrast (MFC) operator [28] is applied with a linear extension to extract features in the defined orientation. Beside other direction-based edge detection methods [29], [30], MFC is selected because of its ability to handle proper orientation information (not just histogram binning) on a contour level (not just pixel level) at corners as well. MFC extracts bright and dark individual features separately by applying morphological opening and closing. The size of the structural elements are selected to suppress background textures and to enhance important features. To detect specific types of features (in the dominant direction), a subsequent ρ_{lin} filter is applied in the main α direction, resulting in a feature-enhanced M_α map. To combine the different salient edge information, M_α and H_{mod} are fused to achieve the salient structural information (Fig. 8(d)):

$$S = \max(\max(0, \log(M_\alpha)), \max(0, \log(H_{\mathrm{mod}}))). \tag{5}$$

At this point, the T texture map represents the salient texture information (Fig. 8(c)), S represents the salient structural information (Fig. 8(d)), which are fused to detect the salient object:

$$C = \gamma \left| \nabla(S(x,y)) \right| + (1 - \gamma) \left| \nabla(T(x,y)) \right|, \tag{6}$$

where a fixed $\gamma = 0.3$ parameter was set based on experiments. The fused feature image is shown in Fig. 8(e).

To obtain the final target region, we apply adaptive Otsu thresholding on the C feature map (Fig. 8(f)) and extract the remaining largest blob as the target.

For the classification process we previously [15] used the raw contour points $H = \{h_i, i > 1\}$ of this blob as the first feature. Here we use the convex hull of the obtained contour points and the blob's main length $L = \max \{d(h_i, h_j), i \neq j\}$ as features during classification (Fig. 8(g)).

B. Classification

During classification we estimate the class of an unknown target based on a priori samples. We handle this process as a top-N retrieval task, where we first index the labeled dataset, then perform a retrieval step, looking for the best match to the query target and proposing a class based on the most similar results. The used index structure is a BK*-tree [31], which is a
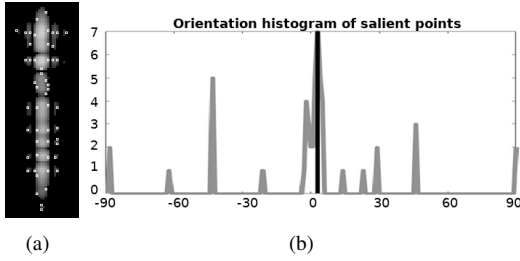
Fig. 9. Calculation of the main direction of the object: (a) the salient points and (b) the orientation histogram. The main orientation (in black) is defined as the orientation with the maximum value in the histogram.

Fig. 10. Normalized confusion matrix. Class labels (A-L) follow Table I.

	A	B	C	D	E	F	G	H	I	J	K	L
A	0.8	0.1	0	0.1	0	0	0	0	0	0	0.1	0
B	0	0.7	0	0	0	0	0	0.1	0	0	0.2	0.1
C	0.1	0	0.4	0.1	0	0	0.1	0.1	0	0	0.1	0.1
D	0.3	0	0	0.5	0	0	0	0.1	0	0.1	0	0.1
E	0	0	0	0	0.4	0.2	0	0	0.3	0	0	0
F	0	0	0	0.1	0.8	0	0	0.1	0	0	0	0
G	0	0	0	0	0	0	0.8	0	0	0	0	0.2
H	0	0.1	0	0	0	0	0.1	0.6	0	0.2	0.1	0.1
I	0	0	0	0	0.1	0	0	0	0.9	0	0	0
J	0.1	0	0	0	0	0	0	0	0	0.8	0.1	0
K	0	0	0	0	0	0	0	0.1	0	0	0.8	0.1
L	0	0	0	0	0	0	0	0	0	0	0.1	0.8

metric tree where a node can have multiple children, each child falling into a specific distance interval from its parent. The tree can be built starting with any random dataset sample, adding the rest sequentially. The comparison of two targets is based on the so called tangent or turning function representation [32] of the extracted contours (difference or target contours q, r denoted by $DT(q,r)$), weighted by the length difference of the two targets ($DL(q,r)$): $D(q,r) = DT(q,r) \cdot DL(q,r)$.

When looking for similar targets, we retrieve the top 5 matches from the index and propose the class of the most frequent result as the class of the unknown query target. In a real system implementation this would be done continuously, classifying based on the updated statistics of previous class proposals. The retrieval process follows the steps from [31], given an unknown target q, and n_R the root node of the index:

1) If $d_0 = D(q, n_R) < t$ (t sensitivity threshold constant), then n_R is a result. Let $t_1 = d_0 - t$, $t_2 = d_0 + t$.
2) For each node P_i ($P_0 = n_R$) having $c_j, j = 1...M$ children,
   a) Select c_j which overlap with $[t_1, t_2]$,
   b) If $d_j = D(q, c_j) < t$, c_j is a result,
   c) Update $t_1 = d_j - t$ and $t_2 = d_j + t$, iterate step 2.

There are several benefits of such a retrieval-based approach. The indexing and retrieval can be fast (shown later); when looking for similar nodes and traversing the tree, large parts of it can be disregarded at every level (only retaining subtrees of some of the children); the index is easy to extend with new elements, which only need to be added to the tree without the need for full reconstruction (i.e., re-training), and it is also easy to parallelize, since we can run multiple retrieval steps on the index simultaneously.

IV. EVALUATION

For evaluation, we used the above mentioned dataset of 294 images. As a first step, Fig. 10 shows the normalized confusion matrix of the full dataset with the proposed method. This matrix's goal is to show the intra- and inter-class similarity/variability of the used target classes, i.e., how similar or different the classes are from a feature-based discrimination point of view. As it shows, some targets can more easily be mistaken for another, similar target. Our goal is to try to get close to these values when only a partial dataset is used for training and evaluated with distorted versions of excluded target sample images. The average recognition rate (diagonal mean of the matrix) is 69%.

During evaluations we followed the same rules: randomly withhold 40-70% from each class - used for testing, denoted by RW (ratio withheld) - and use the remaining 60-30% of elements for training (or indexing in our case). We repeated each test 10 times, and averaged the results over all classes and tests. The methods we compared with are HOG (histogram of oriented gradients) and LBP (local binary patterns) based support vector machines (SVM) using linear (SVML), Gaussian (SVMG), RBF (SVMR) and polynomial (SVMP) kernels, k-nearest neighbors (KNN) and decision trees (Dec.tree).

We measured performance by calculating the average recognition rate (ARR): the percentage of queries when the respective method correctly estimates the class.

The first test (Fig. 11(a)) shows ARR results (averaged over all classes for each method over all repetitions) when testing the classification using randomly withheld dataset elements ("Prop." denotes the proposed method). These results show that in general, when using test images similar to previously known dataset elements, some machine learning approaches perform better: these approaches are good in extracting general image statistics and matching these statistics to a priori samples, while the proposed method (Prop.) deals with the extracted target regions only. However, in the following tests we concentrate on more realistic scenarios, where the targets can be distorted versions of the known samples. In Fig. 11(a) we also included results using [15] ("Prev") to show the changes in this proposed method have improved the approach.

Fig. 11(b) presents ARR results for tests augmented with 90°, 180° and 270° rotated and 10% cropped (along each border) versions: for a test image, its rotated and cropped versions are also generated and tested. As the results show, in a more realistic setting the proposed method performs better.

For Fig. 11(c) we further augmented the previous test by adding 45° and 135° rotated versions of the test images and the proposed method performed better than all compared approaches. For further insight, Fig: 12 shows interim ARR results for the best performers from Fig. 11(c), showing ARR averaged for each class over all runs (column colors representing different withholding ratios). These figures show that the proposed method has fairly low variation vs. different withholding ratios and can reach over 90% averages.

We also performed evaluations to showcase one of the main benefits of the proposed method, that it is robust against rotation changes and it can classify images of known targets containing previously unseen distortions with higher reliability than the other approaches. Thus, we performed tests using the whole dataset for training/indexing and only using the rotated and cropped (as described above) versions of target images
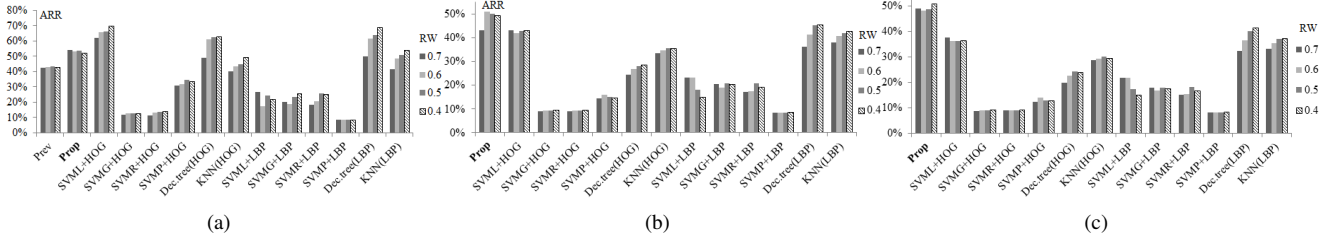
Fig. 11. (a) Average recognition rates (ARR) when withholding RW ratio of each class for testing. (b) Augmenting (a) with 90°, 180°, 270° rotated and 10% cropped versions. (c) Augmenting (b) with 45°, 135° rotated versions.
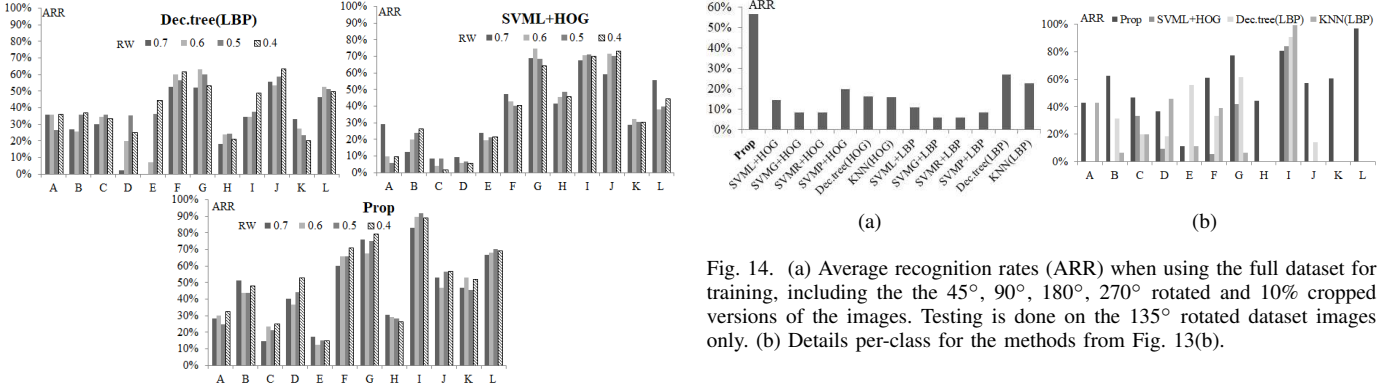

Fig. 12. Per-class (A-L) average recognition rates (ARR) from Fig. 11(c) for the proposed (Prop.), SVML+HOG and Dec.tree(LBP) methods.
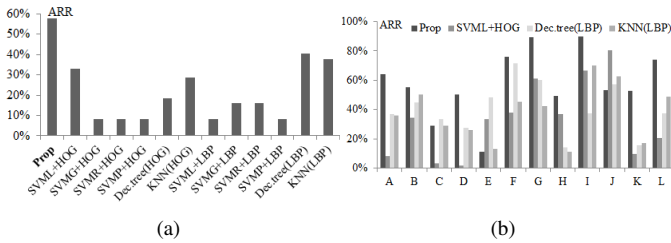

Fig. 13. (a) Average recognition rates (ARR) when using the full dataset for training, and only the 45°, 90°, 135°, 180°, 270° rotated and 10% cropped versions of images for testing. (b) Details per-class for the best performers.
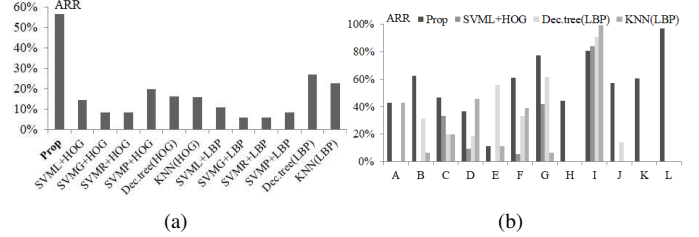

Fig. 14. (a) Average recognition rates (ARR) when using the full dataset for training, including the the 45°, 90°, 180°, 270° rotated and 10% cropped versions of the images. Testing is done on the 135° rotated dataset images only. (b) Details per-class for the methods from Fig. 13(b).

for testing, for each method. Fig. 13 shows the results (overall averages in (a) and per-class in (b)). To further support this property, we performed another test, where we used the whole dataset for training, including the rotated and cropped versions, but excluding the 135° rotation. Then, we only used these 135° rotated images as a test set, and the results are shown in Fig. 14(a, b). As we expected, in case of the proposed approach the inclusion of the rotated versions of the dataset images did not provide considerable changes, some of the other approaches improved, but overall the proposed method remained the better performer.

Overall, the proposed method shows a stable performance and high robustness against input variations and distortions.

We also measured the time performance of the methods. Our intent was to show that the proposed method is viable from a practical usage point of view, and its processing times make it suitable for implementation in a practical system. Table II

shows runtime results for training (a) and testing (b).

We would like to note here, that although some of the methods show lower testing times in some cases - e.g., the Dec.tree(LBP) method can produce generic retrieval rates close to our approach in Fig. 11 -, the benefits of the proposed method can still make it more preferable:

- It is rotation independent, and it does not need different rotations to be included in the training dataset. Other approaches would need all possible (or preferred) rotated versions to be included in training.
- It can robustly extract the contours of the target and use it as a basis for classification, thus it is more robust than other approaches that classify based on overall image statistics, and it can better handle the classification of unknown variations.
- Adding a new element to the dataset means one addition to the index, while the other methods would need to be retrained (including all preferred distortions) when extending the dataset.

Training was measured by training with the full dataset (294 elements) while in testing we measured the classification time for 1176 queries (which corresponds to the tests in Fig. 13). The proposed method runs single threaded, the others multithreaded (hardware: dual Intel Xeon E5645 2.4GHz, 12HT cores; Prop.: C++, others: Matlab R2017a).

V. CONCLUSION

The paper presents and evaluates robust automatic target extraction and classification capabilities in passive ISAR range/cross-range images using textural and structural feature maps, without a priori information about target types or characteristics. The method approaches the problem from a

TABLE II
TIME MEASUREMENTS (S) FOR TRAINING AND TESTING.

	Prop	SVML+HOG	SVMG+HOG	SVMR+HOG	SVMP+HOG	Dec.tree(HOG)	KNN(HOG)	SVML+LBP	SVMG+LBP	SVMR+LBP	SVMP+LBP	Dec.tree(LBP)	KNN(LBP)
Train	0.02	145.8	64.3	63.5	141.9	54.9	52.9	89.6	37.1	37.1	7.2	0.6	0.6
Test	1.14	124.5	364.1	365.2	217.9	0.1	350.3	0.13	0.2	0.2	0.1	0.1	0.5

content-based image retrieval point of view, is lightweight, can easily incorporate extended datasets without retraining. It also provides better recognition rates than compared approaches in realistic use cases when unknown targets can be distorted versions of previously known labeled samples. The method was evaluated on a 294 real image dataset containing targets with resolutions of 0.81-11.72 meters/pixel, obtained with the Software-defined Multiband Array Passive Radar (SMARP) demonstrator of the CNIT-RaSS Lab., operating at UHF (470-790 MHz) and S-band (2100-2200 MHz) frequencies. Further improvements would include image sequence processing, and further increase in robustness and classification performance.

REFERENCES

[1] D. Olivadese, E. Giusti, D. Petri, M. Martorella, A. Capria, and F. Berizzi, "Passive ISAR with DVB-T signals," *IEEE Tr. on Geoscience and Remote Sensing*, vol. 51, no. 8, pp. 4508–4517, 2013.

[2] P. Krysik, K. Kulpa, P. Samczynski, K. Szumski, and J. Misiurewicz, "Moving target detection and imaging using GSM-based passive radar," in *Proc. of IET Intl. Conf. on Radar Systems*, 2012.

[3] C. Coleman and H. Yardley, "Passive bistatic radar based on target illuminations by digital audio broadcasting," *IET Radar, Sonar & Navigation*, vol. 2, no. 5, pp. 366–375, 2008.

[4] A. Capria, E. Giusti, C. Moscardini, M. Conti, D. Petri, M. Martorella, and F. Berizzi, "Multifunction imaging passive radar for harbour protection and navigation safety," *IEEE Aerospace and Electronic Systems Magazine*, vol. 32, no. 2, pp. 30–38, 2017.

[5] J. L. Garry, C. J. Baker, G. E. Smith, and R. L. Ewing, "Investigations toward multistatic passive radar imaging," in *Proc. of IEEE Radar Conference*, 2014, pp. 607–612.

[6] M. Martorella and E. Giusti, "Theoretical foundation of passive bistatic ISAR imaging," *IEEE Tr. on Aerospace and Electronic Systems*, vol. 50, no. 3, pp. 1647–1659, 2014.

[7] D. Gromek, M. Baczyk, P. Samczynski, K. Kulpa, and P. Baranowski, "A concept of the parametric autofocus method for passive ISAR imaging," in *Proc. of Intl. Radar Symposium*, 2017, pp. 1–10.

[8] W. Qiu, E. Giusti, A. Bacci, M. Martorella, F. Berizzi, H. Zhao, and Q. Fu, "Compressive sensing-based algorithm for passive bistatic ISAR with DVB-T signals," *IEEE Tr. on Aerospace and Electronic Systems*, vol. 51, no. 3, pp. 2166–2180, 2015.

[9] M. Conti, F. Berizzi, M. Martorella, E. D. Mese, D. Petri, and A. Capria, "High range resolution multichannel DVB-T passive radar," *IEEE Aerospace and Electronic Systems Magazine*, vol. 27, no. 10, pp. 37–42, 2012.

[10] T. Martelli, F. Colone, E. Tilli, and A. D. Lallo, "Multi-frequency target detection techniques for DVB-T based passive radar sensors," *Sensors*, vol. 16, no. 10, p. 1594, 2016.

[11] G. Cui, J. Liu, H. Li, and B. Himed, "Target detection for passive radar with noisy reference channel," in *Proc. of IEEE Radar Conference*, 2014.

[12] N. del Rey-Maestre, M. Jarabo-Amores, D. Mata-Moya, J. Bárcena-Humanes, and P. G. del Hoyo, "Machine learning techniques for coherent CFAR detection based on statistical modeling of UHF passive ground clutter," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 104–118, 2018.

[13] C. Benedek and M. Martorella, "Ship structure extraction in isar image sequences by a Markovian approach," in *Proc. of IET Intl. Conf. on Radar Systems*, 2012, pp. 62–66.

[14] J. Pisane, S. Azarian, M. Lesturgie, and J. Verly, "Automatic target recognition for passive radar," *IEEE Tr. on Aerospace and Electronic Systems*, vol. 50, no. 1, pp. 371–392, 2014.

[15] A. Manno-Kovacs, E. Giusti, F. Berizzi, and L. Kovács, "Automatic target classification in passive ISAR range-crossrange images," in *Proc. of IEEE Radar Conference*, 2018, pp. 206–211.

[16] A. Manno-Kovacs, "Direction selective vector field convolution for contour detection," in *Proc. of IEEE Intl. Conf. on Image Processing (ICIP)*, 2014, pp. 4722–4726.

[17] A. Kovács and T. Szirányi, "Improved Harris feature point set for orientation sensitive urban area detection in aerial images," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 4, pp. 796–800, 2013.

[18] C. Moscardini, D. Petri, A. Capria, M. Conti, M. Martorella, and F. Berizzi, "Batches algorithm for passive radar: a theoretical analysis," *IEEE Tr. on Aerospace and Electronic Systems*, vol. 51, no. 2, pp. 1475–1487, April 2015.

[19] D. Petri, C. Moscardini, M. Martorella, M. Conti, A. Capria, and F. Berizzi, "Performance analysis of the batches algorithm for range-Doppler map formation in passive bistatic radar," in *Proc. of IET Intl. Conf. on Radar Systems*, 2012, pp. 1–4.

[20] M. Martorella, "Novel approach for ISAR image cross-range scaling," *IEEE Tr. on Aerospace and Electronic Systems*, vol. 41, no. 1, pp. 281–294, 2008.

[21] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Tr. on Systems, Man and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[22] C. Scharfenberger, A. Wong, K. Fergani, J. S. Zelek, and D. Clausi, "Statistical textural distinctiveness for salient region detection in natural images," in *Proc. of IEEE Intl. Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 979–986.

[23] A. Manno-Kovacs, "Direction selective contour detection for salient objects," *IEEE Tr. on Circuits and Systems for Video Technology*, vol. PP, no. 99, pp. 1–15, 2018.

[24] A. Manno-Kovacs and T. Szirányi, "Orientation-selective building detection in aerial images," *ISPRS J. of Photogrammetry and Remote Sensing*, vol. 108, pp. 94–112, 2015.

[25] A. Kovacs and T. Szirányi, "Harris function based active contour external force for image segmentation," *Pattern Recognition Letters*, vol. 33, no. 9, pp. 1180–1187, 2012.

[26] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. of the 4th Alvey Vision Conf.*, 1988, pp. 147–151.

[27] C. Benedek, X. Descombes, and J. Zerubia, "Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics," *IEEE Tr. on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 33–50, 2012.

[28] I. Zingman, D. Saupe, and K. Lambers, "A morphological approach for distinguishing texture and individual features in images," *Pattern Recognition Letters*, vol. 47, pp. 129–138, 2014.

[29] P. Perona, "Orientation diffusions," *IEEE Tr. on Image Processing*, vol. 7, no. 3, pp. 457–467, 1998.

[30] S. Yi, D. Labate, G. R. Easley, and H. Krim, "A shearlet approach to edge analysis and detection," *IEEE Tr. on Image Processing*, vol. 18, no. 5, pp. 929–941, 2009.

[31] L. Kovács, "Parallel multi-tree indexing for evaluating large descriptor sets," in *Proc. of IEEE Intl. Workshop on Content-Based Multimedia Indexing (CBMI)*, 2013, pp. 173–178.

[32] L. J. Latecki and R. Lakamper, "Application of planar shape comparison to object retrieval in image databases," *Pattern Recognition*, vol. 35, no. 1, pp. 15–29, 2002.